

Tesis doctoral

---

Departament de Bioquímica i de Biologia Molecular

DEVELOPMENT AND APPLICATION OF  
COMPUTATIONAL TECHNIQUES TO DRUG  
DISCOVERY AND STRUCTURE-FUNCTION  
RELATIONSHIPS

---

Jose Carlos Gómez Tamayo

Mireia Dunach Masjuan  
Tutor  
Gianluigi Caltabiano  
Arnau Cordoní Montoya  
Director

# ***Table of contents***

---



1	Introduction	
1.1	G-protein coupled receptors (GPCRs)	3
1.2	The structure of GPCRs	4
1.3	The transmembrane functional water molecules	7
1.4	Pharmacology and signaling	8
1.5	Mechanisms of receptor activation	9
2	Methods	
2.1	Introduction	23
2.2	Virtual screening	24
2.3	Ab initio methods	25
2.4	Molecular dynamics simulations	26
2.5	Molecular docking	28
2.6	Structure prediction.	29
2.7	Databases	30
3	Objectives	33

# ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



4	Results	
4.1	Overview	35
4.2	Summary and contributions	37
4.3	LigandFinder webserver	41
4.4	Sulfur-containing amino acids in membrane proteins	
	<b>Part I</b>	59
	<b>Part II</b>	89
4.5	GPCR-SAS webserver	117
4.6	The extracellular entrance provides selectivity to serotonin 5-HT <sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.	135
5	Conclusions	155



## 1 Introduction.

### 1.1 G-protein coupled receptors

G-protein coupled receptors (GPCRs) form a superfamily of membrane proteins with current estimates of about 1,000 genes (2%-3% of the human proteome) (1). These receptors, located at the cell surface, transduce extracellular signals into the interior of the cell through second messenger cascades controlled by different heterotrimeric guanine nucleotide-binding proteins (G-proteins) coupled at their intracellular regions (1, 2). The GPCR family responds to sensory signals of external origin such as light, odors, pheromones, or tastes; and endogenous signals such as neurotransmitters, (neuro) peptides, lipids, proteases, glycoprotein hormones, and purine ligands, among others (3). This diversity makes GPCRs to be involved in many physiological processes such as the regulation of the cardiovascular function, neurotransmission, and endocrine signaling. Due to their accessibility from the extracellular environment, most of the commercial drugs target GPCRs (4).

Based on sequence similarity methods, GPCRs have been grouped into five main families named: Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2, and Secretin, forming the GRAFS classification system (1). The Glutamate family is formed by 15 members, the Adhesion family is formed by 24 members, the Frizzled/taste family is formed by 24 members, the Secretin family is formed by 15 members and the Rhodopsin family, which is the largest group, is formed by 701 proteins (Figure 1). Alternatively, GPCRs have also been classified into three major classes or clans (5). Class A, B, C and F correspond to rhodopsin, adhesion and secretin, glutamate, and Frizzled/taste families respectively.



## ***Introduction***

---



To date, 142 GPCR crystal structures (81 of unique ligand-receptor complexes and 37 of unique receptors) are deposited in the Protein Data Bank (7). Available crystal structures include receptors from classes A, B (8), C (9), and F (10), in complex with agonists, antagonists, inverse agonists (11), allosteric modulators (12), or biased ligands (13), in complex with a G protein (14) or with beta-arrestin (15), and in the form of monomers or homo-oligomers (10, 16–19).

Multiple sequence alignment has shown that, despite the percentage of sequence identity is very low, there are highly conserved positions and motifs. This feature was used by Ballesteros and Weinstein to define a general numbering scheme based on the most conserved positions in the TM helices among the rhodopsin (Class A) family (20). Each residue is defined by two numbers: the first (1 through 7) corresponds to the helix in which the amino acid of interest is located; the second indicates its position relative to the most conserved residue in the helix, arbitrarily assigned to 50. The Ballesteros-Weinstein notation allows an easy comparison among residues in the TM segments of different receptors.

Crystallization of GPCRs has been possible thanks to the advances in protein engineering techniques and crystallographic methods (21). These techniques include creating receptor-T4 lysozyme and receptor-apocytochrome chimaeras, co-crystallization with monoclonal antibody fragments, and thermo-stabilization. However, despite this continuous increase of solved GPCRs structures and the enormous advances in understanding GPCR structure-function relationships, there are key limitations. First, crystal structures capture a static view of the receptor, while GPCRs are known to exist in different conformational states (22). Also, truncations of the N- and C-terminus, and key loops limit the understanding of these parts in the structure-function relationships. Moreover, the use of antibodies, mutations, and insertion of T4 lysozyme could bias the conformation ensemble of the receptor (23).

## Development and application of computational techniques to drug discovery and structure-function relationships

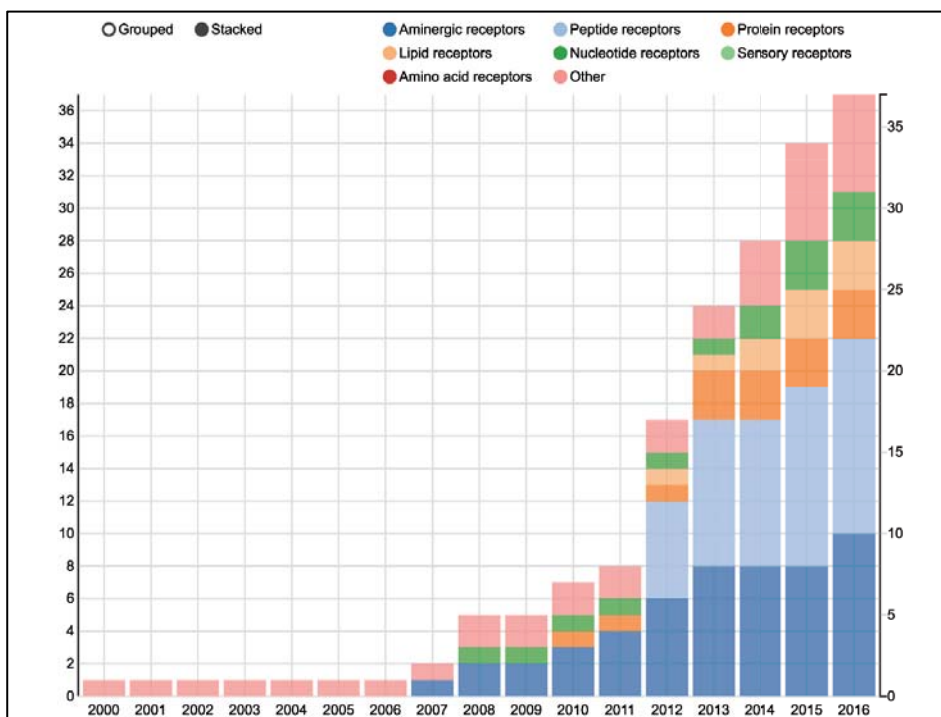
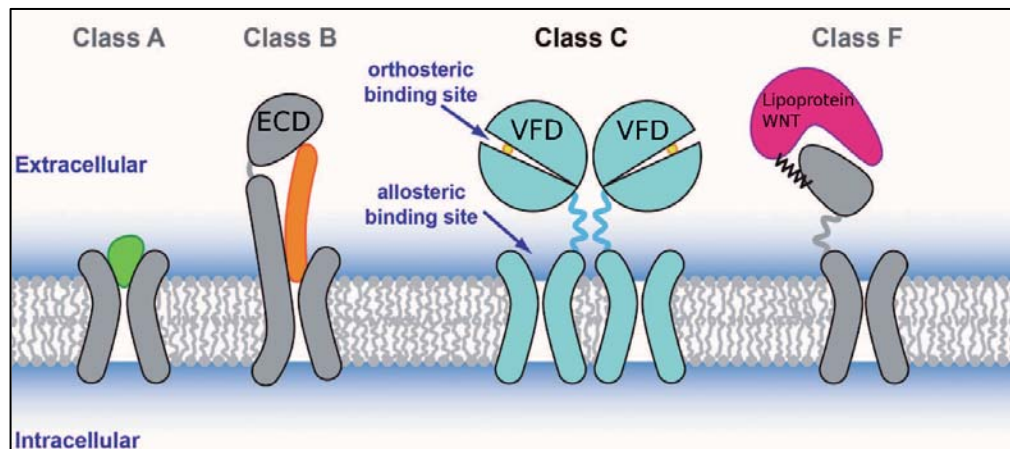


Figure 2. Distribution of available GPCR crystal structures along the past 16 years. Adapted from gpcrd (5)

Though different GPCRs classes share a conserved 7TM domain, there are important key differences at the extracellular domains (ECD). Class B receptors are characterized by a long TM 1 and a large N-terminal ECD involved in the binding of the peptide ligand (Figure 3). Class C receptors are characterized by a huge ECD, called Venus flytrap domain (VFD), which contains the orthosteric binding site (24). Class C receptors exist as homo-dimers, which are covalently linked via a disulfide bond in the VFD, forming the functional unit (Figure 3); Class F shows an ECD formed by a cysteine rich domain (CRD) as well as a linker domain (Figure 3). In contrast, class A GPCRs lacks a large ECD (Figure 3), with some exceptions (glycoprotein hormone receptors also contain a large ECD for hormone binding). The orthosteric binding site for the endogenous ligand varies among the different families. In class A GPCRs, the endogenous ligand bind in a cavity within the 7TM bundle, whereas in class B GPCRs, the endogenous ligand (peptide) bind the ECD and the 7TM domain (8). In class C, the orthosteric site is located in the VFDs. Finally, class F receptors are activated by the lipoprotein WNT which binds the CRD (10).

## Introduction



**Figure 3. Representation of the GPCR classes and the mode of binding of the endogenous ligand.** For class A GPCRs, the endogenous ligand binds the 7TM domain. For class B, the endogenous peptide ligand binds the 7TM and ECD domains. For class C, the endogenous ligands are recognized by the VFDs. Finally, for class F, the lipoprotein WNT binds the CRD domain. Adapted from (24).

### 1.3 The transmembrane functional water molecules

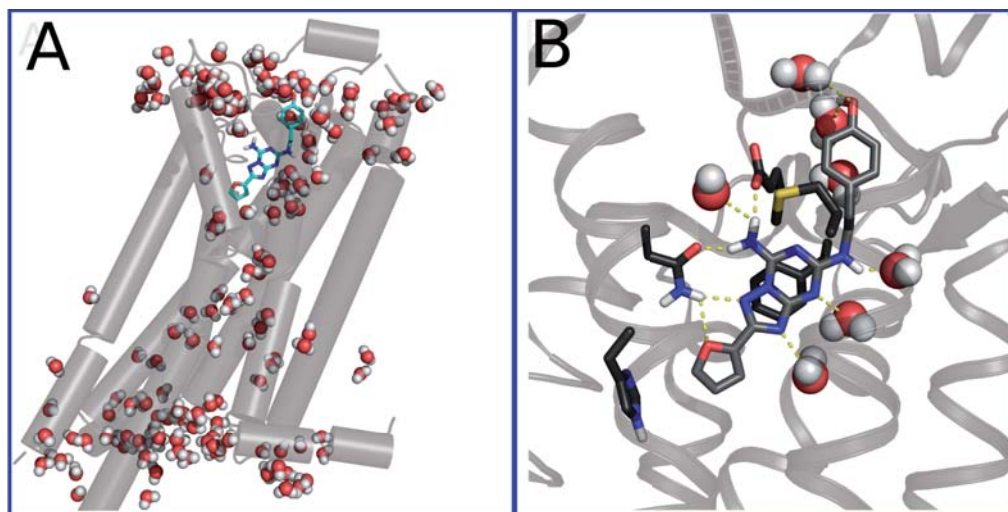
The majority of the TM regions deviate from ideal  $\alpha$ -helices, showing structural anomalies like kinks and bulges. These distortions are energetically stabilized through complementary intra- and inter-helical interactions involving polar side chains, backbone carbonyls, and, in some cases, specific structural and functional water molecules embedded in the TM bundle (25). Interestingly, as the resolution of the crystal structures improves, the number of observed internal waters increases (26). Recently, the publication of the adenosine A2a receptor in complex with the antagonist ZM241385 (26) has shown a receptor practically filled with water molecules (185 in total) where the ligand and G-protein binding sites are connected through a quasi-continuous channel of waters (Figure 4). Yuan et al. (27) have also shown that activation correlates with the formation of this continuous internal water pathway. Similarly, the recent crystal structure of the M2 muscarinic acetylcholine receptor has shown the presence of a continuous aqueous channel extending from the extracellular surface to a depth



## Development and application of computational techniques to drug discovery and structure-function relationships



of approximately the middle of the membrane (28). Moreover, water molecules are also involved in ligand binding as shown in the crystal structure of A2a in complex with ZM241385 (26) (Figure 4).



**Figure 4.** A) Water molecules present in the Adenosine A2a receptor (PDB id 4ejy). B) The binding of the ZM241385 antagonist involves, in addition of receptor side chains, waters molecules.

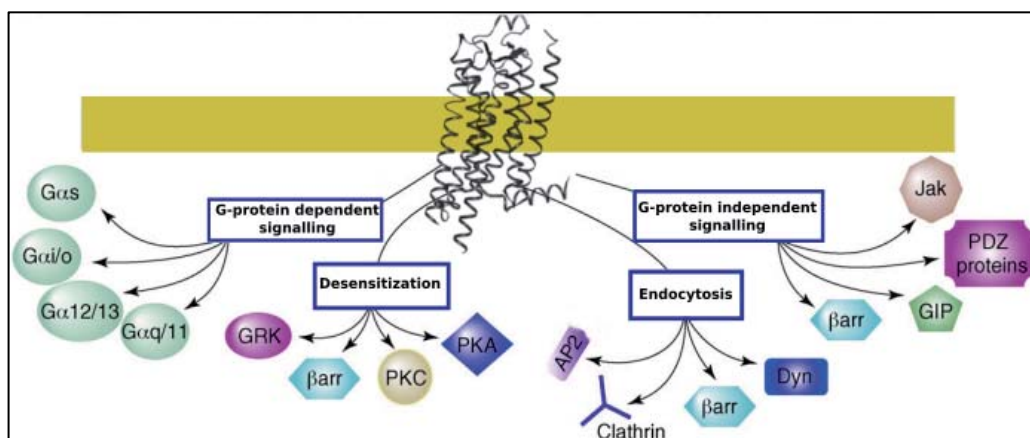
### 1.4 Pharmacology and Signaling

Recent advances in G protein-coupled receptors have shown that the mechanism they transduce extracellular signals is far more complex than previously imagined. Indeed their pharmacological regulation now extends far beyond the simple concept of agonism and antagonism to incorporate concepts of constitutive activity, inverse agonism, allosteric modulation and signal bias. Historically, GPCR receptors were believed to be in equilibrium between two states (inactive and active) in which agonists stabilized the active conformation while antagonist's stabilized the inactive conformation. New experimental observations found that the efficacy of a panel of drugs may differ for different intracellular signaling pathways, meaning a single ligand could act as agonist, antagonist, inverse agonist or partial agonist on one pathway while having a different efficacy in another (29–31), a phenomenon known as biased signaling, functional selectivity, stimulus trafficking or collateral efficacy (29, 32, 33). This remarkably versatility is incompatible with a simple on-off signaling model and is more compatible with a complex system in which many different conformations

## Introduction



of the receptor co-exist, and ligand binding stabilize one of them (22). Upon agonist stimulation, GPCRs undergo conformational changes, the most relevant being the outward movement of TM6, which opens a cleft in the 7TM domain allowing G-protein's coupling. Binding of heterotrimeric G-protein leads to exchange of GDP for GTP on the  $G\alpha$  subunit initiating the dissociation of G-protein  $\alpha$  and  $\beta\gamma$  dimers that act as signaling units and activate various effectors such as cyclic monophosphate (cAMP), calcium ion, or phosphoinositides. Agonist-occupied GPCRs, become substrate for G protein-coupled receptor kinases (GRKs) that phosphorylate GPCRs, which become target of  $\beta$ -arrestins.  $\beta$ -arrestins binding blocks further G-protein activation. This easy and linear signaling pathway has been recently challenged by various experiments that showed GPCRs can independently interact with other proteins (see Figure 5) as  $\beta$ -arrestins and G protein-coupled receptor kinases (GRKs) (34–36), which also led to the discovery of above discussed signal biased ligands.



*Figure 5. Representation of effectors pathways mediated by GPCRs.*

### 1.5 Activation

The increasing number of crystal structures of GPCRs during the last decade, have offered invaluable information to understand the complex world of GPCR activation at the molecular level. Still crystal structures show a static picture of the receptor and cannot unveil the complex activation mechanism of GPCRs. Binding of ligands to specific sites in the receptor induces a series of small-scale

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---

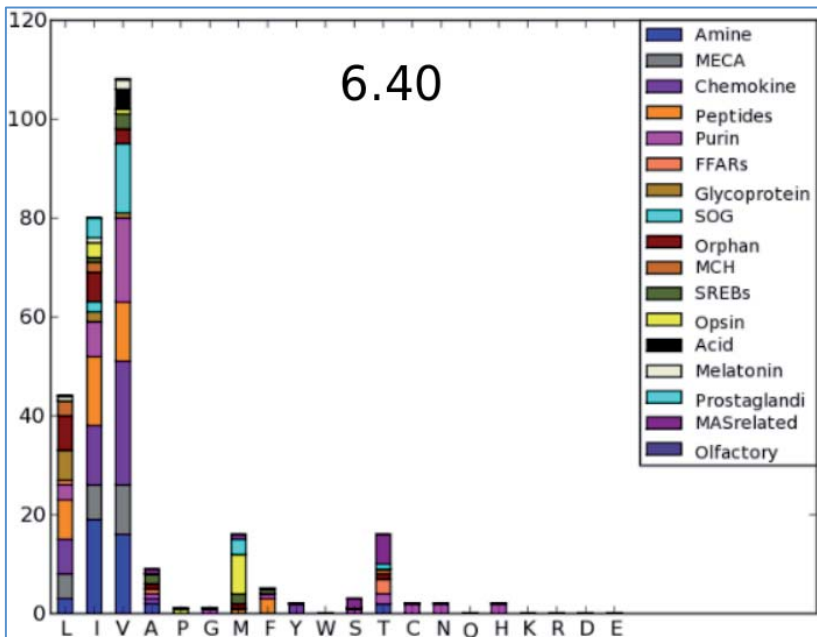
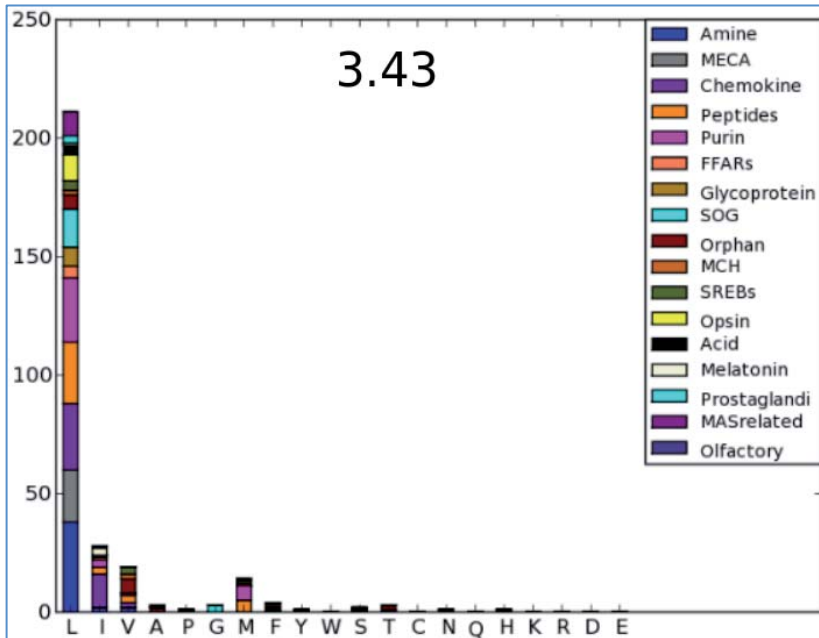


conformational changes that lead to the TM6 outward movement and subsequently G-protein binding and activation. The details of all these events are still mostly unclear. Nonetheless intense use of biochemical (site-directed mutagenesis, radio labeled residues, etc) and theoretical techniques, such as sequence analysis and molecular dynamics simulations, have allowed the identification of some highly conserved sequence regions involved in small-scale conformational changes, named “molecular switches”, which have a key role in receptor activation (37). GPCR’s ligand binding alters the conformation of “molecular switches” inducing the activation of one or more signaling pathways. Known switches include: 1) the hydrophobic core, 2) the “ionic lock” switch, 3) the TM3-7 lock, and 4) the tyrosine toggle switch linked to the NPxxY motif in TM7. Not all molecular switches are conserved among all classes indeed some of them are specific to one or more sub-class. The hydrophobic core is the only molecular switch conserved among all class A family members (38).

### **The hydrophobic core**

The residues of the hydrophobic core associated to the inactive state anchoring are: L<sup>3.43</sup>, F<sup>6.44</sup> and X<sup>6.40</sup> (38). The high conservation of these positions (see Figure 6) reveals their importance. L<sup>3.43</sup> is conserved in around 75% of class A GPCRs and is a hydrophobic bulky residue (Leu, Val, Ile, or Met) in approximately 95% of them. F<sup>6.44</sup> is conserved in around 80% of the receptors while residue at position 6.40 is a bulky hydrophobic residue in more than 80% of class A receptors. Agonist binding triggers a conformational rearrangement of the hydrophobic core, which ultimately leads to the outward movement of TM5 and TM6, and the inward movement of TM7 and TM3. Site-directed mutagenesis studies shows that mutated residues enhancing the strength of hydrophobic interactions decrease constitutive activity (inactive state more stable) while unfavorable mutations has the opposite effect (increase basal activity, making inactive state less stable), see figure 7A (37, 38).

# Introduction



## Development and application of computational techniques to drug discovery and structure-function relationships

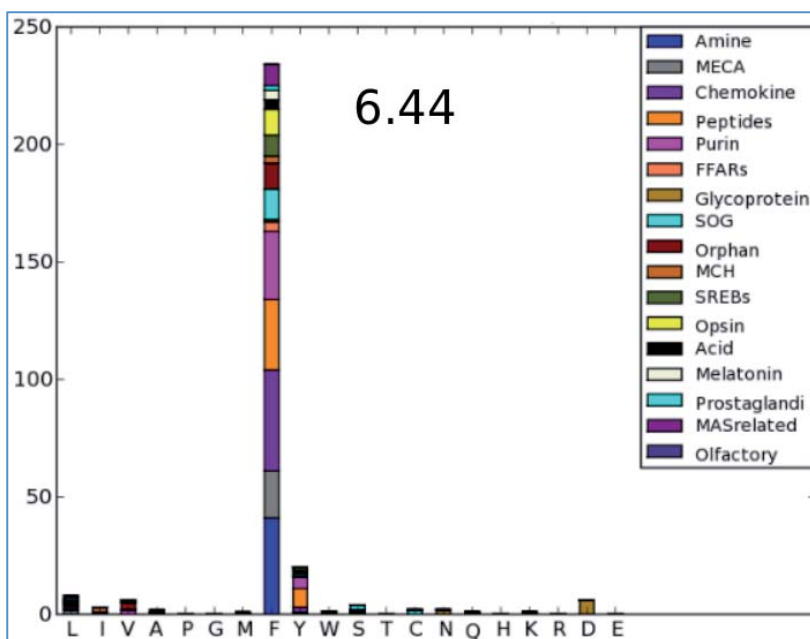
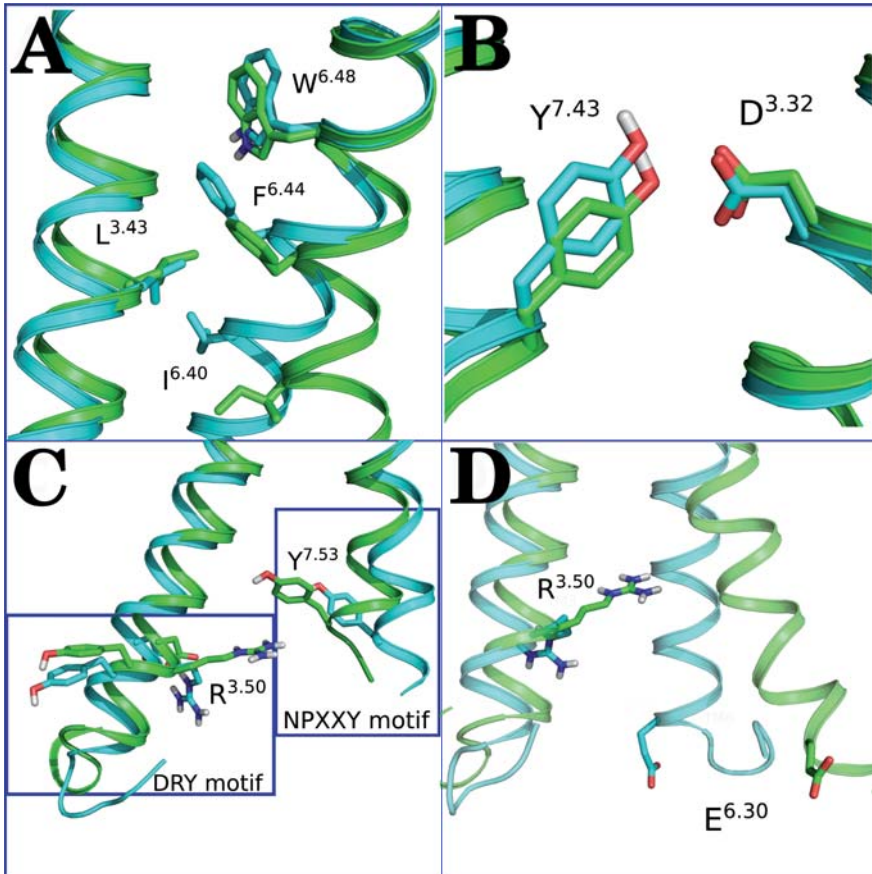


Figure 6. Distribution of residues in positions 3.43, 6.40 and 6.44 among class A GPCR families.

### The TM3-TM7 switch

The TM3-7 switch is an interaction between TM3 and TM7. In rhodopsin, disruption of the interaction of K<sup>7.43</sup> (covalent bonded to retinal) with E<sup>3.28</sup> is the first event occurring upon retinal isomerization. Though TM3-TM7 switch is characteristic of rhodopsin, some aminergic receptors exhibit a similar interaction (through D<sup>3.32</sup> and Y<sup>7.43</sup>, see figure 7B) as revealed in the crystal structures of histamine H1 receptor and dopamine D3 receptor in complex with antagonists (39). Opioids receptors also present this interaction (16, 17, 40).

## Introduction



**Figure 7. Molecular switches in class A GPCR activation (Adrenergic beta 2 receptor).** Inactive and active conformations are colored in cyan and green respectively. A) Residues involved in the hydrophobic core. B) TM3-TM7 lock switch. C) Superposition of active and inactive tyrosine toggle switch. D) Superposition of active and inactive residues forming the ionic lock in adrenergic beta 2 receptor.

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



### **The tyrosine switch**

Inactive state-crystals show a hydrophobic region formed by Y<sup>7.53</sup>, Y<sup>5.58</sup>, V<sup>1.53</sup>, V<sup>2.43</sup>, and I<sup>6.40</sup>, which isolate the residues of the “ionic lock” from the quasi-continuous channel of waters that connect the orthosteric binding site to the cytoplasmic part of the receptor. Outward movement and rotation of TM6 following agonist’s binding leads to a conformational change of Y<sup>7.53</sup> (from highly conserved NPxxY motif) which open the gate to water entrance (27, 41), thus connecting the cytosol with the orthosteric binding site by a continuous channel of water molecules.

### **The ionic lock**

The ionic lock is an ionic interaction between residues R<sup>3.50</sup> of the conserved (D/E)R(W/Y) sequence motif in TM3 and residues D/E<sup>6.30</sup> in TM6 (11, 42) . It is present in trace amine, purinergic, somatostatin, opsin and aminergic subfamilies of class A GPCRs. Despite the ionic lock interaction is important in a small set of GPCRs, this interaction has been emphasized because the first resolved crystal structures showed it. Disruption of the ionic lock is mandatory for maximal receptor activation and is triggered by almost all agonists (43) of the above-mentioned sub-families.

## Introduction

---



### References

1. Fredriksson, R., Lagerström, M. C., Lundin, L.-G., and Schiöth, H. B. (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272
2. Pierce, K. L., Premont, R. T., and Lefkowitz, R. J. (2002) Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.* **3**, 639–650
3. Lagerström, M. C., and Schiöth, H. B. (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat. Rev. Drug Discov.* **7**, 339–357
4. Ma, P., and Zimmel, R. (2002) Value of novelty? *Nat. Rev. Drug Discov.* **1**, 571–572
5. Isberg, V., Mordalski, S., Munk, C., Rataj, K., Harpsoe, K., Hauser, A. S., Vroiling, B., Bojarski, A. J., Vriend, G., and Gloriam, D. E. (2015) GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res.* **44**, 356–364
6. Stevens, R. C., Cherezov, V., Katritch, V., Abagyan, R., Kuhn, P., Rosen, H., and Wüthrich, K. (2013) The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. *Nat. Rev. Drug Discov.* **12**, 25–34
7. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242
8. Hollenstein, K., Kean, J., Bortolato, A., Cheng, R. K. Y., Doré, A. S., Jazayeri, A., Cooke, R. M., Weir, M., and Marshall, F. H. (2013) Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature.* **499**, 438–43
9. Wu, H., Wang, C., Gregory, K. J., Han, G. W., Cho, H. P., Xia, Y., Niswender, C. M., Katritch, V., Meiler, J., Cherezov, V., Conn, P. J., and Stevens, R. C. (2014) Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. *Science.* **344**, 58–64
10. Wang, C., Wu, H., Katritch, V., Han, G. W., Huang, X.-P., Liu, W., Siu, F. Y., Roth, B. L., Cherezov, V., and Stevens, R. C. (2013) Structure of the human smoothed receptor bound to an antitumour agent. *Nature.* **497**, 338–43



## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



11. Kobilka, B. K., and Deupi, X. (2007) Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **28**, 397–406
12. Christopher, J. A., Aves, S. J., Bennett, K. A., Doré, A. S., Errey, J. C., Jazayeri, A., Marshall, F. H., Okrasa, K., Serrano-Vega, M. J., Tehan, B. G., Wiggin, G. R., and Congreve, M. (2015) Fragment and Structure-Based Drug Discovery for a Class C GPCR: Discovery of the mGlu5 Negative Allosteric Modulator HTL14242 (3-Chloro-5-[6-(5-fluoropyridin-2-yl)pyrimidin-4-yl]benzotrile). *J. Med. Chem.* **58**, 6653–6664
13. Saulière, A., Bellot, M., Paris, H., Denis, C., Finana, F., Hansen, J. T., Altié, M.-F., Seguelas, M.-H., Pathak, A., Hansen, J. L., Sénard, J.-M., and Galés, C. (2012) Deciphering biased-agonism complexity reveals a new active AT1 receptor entity. *Nat. Chem. Biol.* **8**, 622–630
14. Rasmussen, S. G. F., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T. a., Lyons, J. a., Caffrey, M., Gellman, S. H., Steyaert, J., Skinotitis, G., Weis, W. I., Sunahara, R. K., and Kobilka, B. K. (2011) Crystal structure of the  $\beta$ 2 adrenergic receptor–Gs protein complex. *Nature.* **477**, 549–555
15. Kang, Y., Zhou, X. E., Gao, X., He, Y., Liu, W., Ishchenko, A., Barty, A., White, T. A., Yefanov, O., Han, G. W., Xu, Q., de Waal, P. W., Ke, J., Tan, M. H. E., Zhang, C., Moeller, A., West, G. M., Pascal, B. D., Van Eps, N., Caro, L. N., Vishnivetskiy, S. A., Lee, R. J., Suino-Powell, K. M., Gu, X., Pal, K., Ma, J., Zhi, X., Boutet, S., Williams, G. J., Messerschmidt, M., Gati, C., Zatsepin, N. A., Wang, D., James, D., Basu, S., Roy-Chowdhury, S., Conrad, C. E., Coe, J., Liu, H., Lisova, S., Kupitz, C., Grotjohann, I., Fromme, R., Jiang, Y., Tan, M., Yang, H., Li, J., Wang, M., Zheng, Z., Li, D., Howe, N., Zhao, Y., Standfuss, J., Diederichs, K., Dong, Y., Potter, C. S., Carragher, B., Caffrey, M., Jiang, H., Chapman, H. N., Spence, J. C. H., Fromme, P., Weierstall, U., Ernst, O. P., Katritch, V., Gurevich, V. V., Griffin, P. R., Hubbell, W. L., Stevens, R. C., Cherezov, V., Melcher, K., and Xu, H. E. (2015) Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature.* **523**, 561–7
16. Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., Pardo, L., Weis, W. I., Kobilka, B. K., and Granier, S. (2012) Crystal structure of the  $\mu$ -opioid receptor bound to a morphinan antagonist. *Nature.* **485**, 321–326
17. Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G. W., Vardy, E., Liu, W., Thompson, A. a., Huang, X.-P., Carroll, F. I., Mascarella, S. W., Westkaemper, R. B., Mosier, P. D., Roth, B. L., Cherezov, V., and Stevens, R. C. (2012) Structure of the human  $\kappa$ -opioid receptor in complex with JDTic. *Nature.* **485**, 327–332

## Introduction

---



18. Wu, B., Chien, E. Y. T., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., Hamel, D. J., Kuhn, P., Handel, T. M., Cherezov, V., and Stevens, R. C. (2010) Structures of the CXCR4 Chemokine. *Science*. **330**, 1066–1071
19. Huang, J., Chen, S., Zhang, J. J., and Huang, X.-Y. (2013) Crystal structure of oligomeric  $\beta$ 1-adrenergic G protein-coupled receptors in ligand-free basal state. *Nat. Struct. Mol. Biol.* **20**, 419–25
20. Ballesteros, J. A., and Weinstein, H. (1995) [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428
21. Tate, C. G., and Schertler, G. F. (2009) Engineering G protein-coupled receptors to facilitate their structure determination. *Curr. Opin. Struct. Biol.* **19**, 386–395
22. Kobilka, B. K., and Deupi, X. (2007) Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **28**, 397–406
23. Venkatakrishnan, A. J., Deupi, X., Lebon, G., Tate, C. G., Schertler, G. F., and Babu, M. M. (2013) Molecular signatures of G-protein-coupled receptors. *Nature*. **494**, 185–194
24. Wu, H., Wang, C., Gregory, K. J., Han, G. W., Cho, H. P., Xia, Y., Niswender, C. M., Katritch, V., Meiler, J., Cherezov, V., Conn, P. J., and Stevens, R. C. (2014) Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. *Science*. **344**, 58–64
25. Pardo, L., Deupi, X., Dölker, N., López-Rodríguez, M. L., and Campillo, M. (2007) The role of internal water molecules in the structure and function of the rhodopsin family of G protein-coupled receptors. *Chembiochem*. **8**, 19–24
26. Liu, W., Chun, E., Thompson, A. a, Chubukov, P., Xu, F., Katritch, V., Han, G. W., Roth, C. B., Heitman, L. H., IJzerman, A. P., Cherezov, V., and Stevens, R. C. (2012) Structural basis for allosteric regulation of GPCRs by sodium ions. *Science*. **337**, 232–6
27. Yuan, S., Filipek, S., Palczewski, K., and Vogel, H. (2014) Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway. *Nat. Commun.* **5**, 4733
28. Haga, K., Kruse, A. C. A., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C.,

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



- Weis, W. I., Okada, T., Kobilka, B. K., Haga, T., and Kobayashi, T. (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature*. **482**, 547–551
29. Wei, H., Ahn, S., Shenoy, S. K., Karnik, S. S., Hunyady, L., Luttrell, L. M., and Lefkowitz, R. J. (2003) Independent beta-arrestin 2 and G protein-mediated pathways for angiotensin II activation of extracellular signal-regulated kinases 1 and 2. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10782–10787
30. Gesty-Palmer, D., Chen, M., Reiter, E., Ahn, S., Nelson, C. D., Wang, S., Eckhardt, A. E., Cowan, C. L., Spurney, R. F., Luttrell, L. M., and Lefkowitz, R. J. (2006) Distinct  $\beta$ -arrestin- and G protein-dependent pathways for parathyroid hormone receptor-stimulated ERK1/2 activation. *J. Biol. Chem.* **281**, 10856–10864
31. Azzi, M., Charest, P. G., Angers, S., Rousseau, G., Kohout, T., Bouvier, M., and Piñeyro, G. (2003) Beta-arrestin-mediated activation of MAPK by inverse agonists reveals distinct active conformations for G protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 11406–11411
32. Galandrin, S., Oligny-Longpré, G., and Bouvier, M. (2007) The evasive nature of drug efficacy: implications for drug discovery. *Trends Pharmacol. Sci.* **28**, 423–430
33. Violin, J. D., and Lefkowitz, R. J. (2007)  $\beta$ -Arrestin-biased ligands at seven-transmembrane receptors. *Trends Pharmacol. Sci.* **28**, 416–422
34. Lefkowitz, R. J. (2005) REVIEW Transduction of Receptor Signals. *Science*. **512**, 512–518
35. Reiter, E., and Lefkowitz, R. J. (2006) GRKs and  $\beta$ -arrestins: roles in receptor silencing, trafficking and signaling. *Trends Endocrinol. Metab.* **17**, 159–165
36. Penela, P., Murga, C., Ribas, C., Lafarga, V., and Mayor, F. (2010) The complex G protein-coupled receptor kinase 2 (GRK2) interactome unveils new physiopathological targets. *Br. J. Pharmacol.* **160**, 821–832
37. Trzaskowski, B., Latek, D., Yuan, S., Ghoshdastider, U., Debinski, a, and Filipek, S. (2012) Action of molecular switches in GPCRs--theoretical and experimental studies. *Curr. Med. Chem.* **19**, 1090–109
38. Tehan, B. G., Bortolato, A., Blaney, F. E., Weir, M. P., and Mason, J. S. (2014) Unifying Family A GPCR Theories of Activation. *Pharmacol. Ther.* **143**, 51–60

## Introduction

---



39. Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V., Abagyan, R., Cherezov, V., Liu, W., Han, G. W., Kobayashi, T., Stevens, R. C., and Iwata, S. (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature*. **475**, 65–70
40. Granier, S., Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Weis, W. I., and Kobilka, B. K. (2012) Structure of the  $\delta$ -opioid receptor bound to naltrindole. *Nature*. **485**, 400–404
41. Standfuss, J., Edwards, P. C., D'Antona, A., Fransen, M., Xie, G., Oprian, D. D., and Schertler, G. F. X. (2011) The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature*. **471**, 656–660
42. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Trong, I. Le, Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000) Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science*. **289**, 739–745
43. Kobilka, B. K. (2007) G protein coupled receptor structure and activation. *Biochim. Biophys. Acta*. **1768**, 794–807
44. Fenton, A. W. (2008) Allostery: an illustrated definition for the “second secret of life.” *Trends Biochem. Sci.* **33**, 420–425
45. May, L. T., Leach, K., Sexton, P. M., and Christopoulos, A. (2007) Allosteric modulation of G protein-coupled receptors. *Annu. Rev. Pharmacol. Toxicol.* **47**, 1–51
46. Christopoulos, A., and Kenakin, T. (2002) G protein-coupled receptor allostery and complexing. *Pharmacol. Rev.* **54**, 323–74
47. Kenakin, T. P. (2009)  $\gamma$ TM Receptor Allostery: Putting Numbers to Shapeshifting Proteins. *Trends Pharmacol. Sci.* **30**, 460–469
48. Keov, P., Sexton, P. M., and Christopoulos, A. (2011) Allosteric modulation of G protein-coupled receptors: A pharmacological perspective. *Neuropharmacology*. **60**, 24–35
49. Chan, W. Y., McKinzie, D. L., Bose, S., Mitchell, S. N., Witkin, J. M., Thompson, R. C., Christopoulos, a, Lazareno, S., Birdsall, N. J. M., Bymaster, F. P., and Felder, C. C. (2008) Allosteric modulation of the muscarinic M4 receptor as an approach to treating schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10978–10983

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



50. Leach, K., Sexton, P. M., and Christopoulos, A. (2000) Quantification of allosteric interactions at G protein-coupled receptors using radioligand binding assays. *Curr. Protoc. Pharmacol.* **Chapter 1**, Unit 1.22
51. Kenakin, T. (2005) New concepts in drug discovery: collateral efficacy and permissive antagonism. *Nat. Rev. Drug Discov.* **4**, 919–927
52. Stevenson, L. a, Easson, M., Price, M. R., Baillie, G. L., Goodwin, R., Mclean, A., Mcintosh, L., Goodwin, G., Walker, G., Westwood, P., Marrs, J., Thomson, F., Cowley, P., Christopoulos, A., Pertwee, R. G., and Ross, R. a (2005) Allosteric Modulation of the Cannabinoid CB 1 Receptor. *Mol. Pharmacol.* **68**, 1484–1495
53. Angers, S., Salahpour, A., and Bouvier, M. (2002) Dimerization: an emerging concept for G protein-coupled receptor ontogeny and function. *Annu. Rev. Pharmacol. Toxicol.* **42**, 409–35
54. Birdsall, N. J. M. (1982) Can different receptors interact directly with each other? *Trends Neurosci.* **5**, 137–138
55. Milligan, G. (2007) G protein-coupled receptor dimerisation: Molecular basis and relevance to function. *Biochim. Biophys. Acta - Biomembr.* **1768**, 825–835
56. Pin, J. P., Neubig, R., Bouvier, M., Devi, L., Filizola, M., Javitch, J. A., Lohse, M. J., Milligan, G., Palczewski, K., Parmentier, M., and Spedding, M. (2007) International Union of Basic and Clinical Pharmacology. LXVII. Recommendations for the recognition and nomenclature of G protein-coupled receptor heteromultimers. *Pharmacol Rev.* **59**, 5–13
57. Ferré, S., Baler, R., Bouvier, M., Caron, M. G., Devi, L. a, Durroux, T., Fuxe, K., George, S. R., Javitch, J. a, Lohse, M. J., Mackie, K., Milligan, G., Pflieger, K. D. G., Pin, J.-P., Volkow, N. D., Waldhoer, M., Woods, A. S., and Franco, R. (2009) Building a new conceptual framework for receptor heteromers. *Nat. Chem. Biol.* **5**, 131–134
58. Albizu, L., Moreno, J. L., González-Maeso, J., and Sealfon, S. C. (2010) Heteromerization of G protein-coupled receptors: relevance to neurological disorders and neurotherapeutics. *CNS Neurol. Disord. Drug Targets.* **9**, 636–50
59. Birdsall, N. J. M. (2010) Class A GPCR heterodimers: Evidence from binding studies. *Trends Pharmacol. Sci.* **31**, 499–508
60. Rozenfeld, R., Bushlin, I., Gomes, I., Tzavaras, N., Gupta, A., Neves, S., Battini, L., Gusella, G. L., Lachmann, A., Ma'ayan, A., Blitzer, R. D., and Devi, L. a. (2012)

## ***Introduction***

---



Receptor heteromerization expands the repertoire of cannabinoid signaling in rodent neurons. *PLoS One*. 10.1371/journal.pone.0029239

61. González-Maeso, J. (2011) GPCR oligomers in pharmacology and signaling. *Mol. Brain*. **4**, 20
62. Cordoní, A., Navarro, G., Aymerich, M. S., and Franco, R. (2015) Structures for G-Protein-Coupled Receptor Tetramers in Complex with G Proteins. *Trends Biochem. Sci.* **40**, 548–551

***Development and application of computational techniques to drug discovery and structure-function relationships***

---





## 2 Methods.

### 2.1 Computational Methods

Computational techniques can be applied to different disciplines including chemistry, physics and biology, among others. All these studies have been facilitated by the availability of experimental information, but also by a steady increase in compute power which is provided either by continuous advancements of hardware performance including GPUs, tailored computer architectures (1), or cloud computing approaches (2). They take advantage of computer performance in order to predict chemical, physical, and biological processes. Moreover, as biological entities can be studied from a complex system point of view, and everyday lots of new information come to light, the use of network analysis and big data analysis have also become powerful tools. All together form a framework, which complements and try to shed light at molecular level to the results obtained from experiments.

The theoretical background behind each technique narrows down the size of the system and accuracy of the calculation. Generally, the higher is the level of theory of the technique, the smaller is the system. From a bottom-up view (the smallest to the biggest), quantum-mechanical calculations involve explicit consideration of electrons making them computationally expensive and hence only suitable to small systems (hundred of atoms). Molecular mechanics uses classical mechanics to model molecular systems. Atomistic molecular mechanics considers each atom as a single particle while coarse-grain molecular mechanics simplifies the system by treating groups of atoms as particles (3) allowing optimization of systems over the million of atoms. Molecular dynamics simulations solve Newton equations of motion providing information on the dynamical behavior of the system (3). The range of molecular dynamics simulations applicability is wide and does not depend only on the system size, but also on the desired time of simulation and taken approximations. In our systems, integrated by a bilayer membrane, a GPCR, a G-protein and waters, the number of atoms can reach hundred of thousands.



# ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



On the other hand, other computational techniques are based in the comparison of available information rather than the application of chemical/physical formulas. For example, virtual screening usually used to find possible drug candidates in databases. Virtual screening can deal with million of compounds which can be represented as atoms in molecular docking (applying a scoring function, see section 2.5), pharmacophores (pharmacophore docking), descriptors (comparing physicochemical properties), or even text (calculating similarity between text chains) (4).

The following subsections describe the main techniques employed in this thesis in a summarized form. These techniques are further detailed in results section.

## **2.2 Virtual screening**

Virtual screening (VS) is a computational technique used to theoretically identify possible molecules, called “hits”, prone to be starting points of a drug discovery process. Hit compounds usually are submitted to modification in order to identify “lead” compounds, which are the starting point to a both time and economic-consuming process called “lead optimization”. It is been estimated that the process from lead identification to clinical trials are costs about 800 million US dollars and 14 years (5). The advances during the 90’s decade in combinatorial chemistry and high-throughput screening supposed important drug discovery acceleration. Nevertheless, many of the identified leads fall in the lead optimization process due to absorption, distribution, metabolism, excretion and toxicity (ADMET) inadequacy. The need for cheaper methodology to refine lead selection, neglecting unsuitable structures, opened the way to the development of virtual screening.

Virtual screening is based on the computational search of novel compounds in pre-compiled libraries of compounds. There exist two main virtual screening techniques: ligand-based (LBVS) and structure-based (SBVS) (6). Ligand based virtual screening relies on the use of compounds of known activity and includes approaches such as similarity and substructure searching, quantitative structure-activity relationships (QSAR), pharmacophore and three-dimensional shape matching (7, 8). On the other hand, SBVS uses the three-dimensional structure of a biological target (determined experimentally or modeled) to dock (see section 2.5)

## Methods

---



a panel of new molecules and rank them based on their suitability according to their predicted binding affinity and/or complementarity to the binding site.

Due to the historical difficulty to crystallize GPCRs (see introduction, section 1.2), LBVS approaches were the most used to infer binding properties of new compounds and despite the to-date availability of 142 crystal structures they still are a top methodology thanks mainly to their low computational cost, which make them the only suitable methodology to screen databases of millions of compounds. Besides, they can be complemented with three-dimensional information from structures or models (see section 2.5 and 2.6) in order to refine the results. As part of this doctorate thesis we designed a new VS approach and created a web application, LigandFinder, developed as a tool to perform LBVS for computationally-untrained scientist.

### 2.3 Ab initio methods.

Computational chemistry is the branch of the chemistry that uses computer science in order to theoretically solve chemistry problems. Among others, one of the most exact calculations involves the use of quantum mechanics laws used to solve chemical problems. The need of these methods arises from the impossibility to analytically solve Schrödinger equation on systems with more than one electron. Some of the properties these methods are able to calculate include structure, absolute and relative energy, distribution of electronic charge, dipoles and multipole momentums, vibrational frequencies, reactivity, UV-Vis spectrum prediction, magnetic properties, transition states etc.

Many methods exist to implement the mentioned properties but those based on molecular orbital theory achieve more realistic results. These are ab initio, semi-empirical, and density functional theory methods. However, despite of the great accuracy achieved by these methods, computational cost grows exponentially with the size of the system. In this thesis, quantum chemistry methods have commonly used in three-dimensional structure optimization of ligands, the parameterization of ligands (needed to perform molecular dynamics simulations), and specially in sulfur-containing amino acids where high level quantum calculations were used in order to optimize and evaluate interaction energies of clustered geometries.

# ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



## **2.4 Molecular dynamics simulations**

Molecular dynamics (MD) permits to simulate the time-evolution of a molecular system, which is represented classically considering a set of particles (atoms) defined by their positions and momenta. This technique has provided many insights related with the dynamical behavior of macromolecules such as conformational changes, free energy of binding, and stability. This information, often inaccessible to wet lab experiments, and certainly cheaper, complements experimental data helping to understand macromolecules nature.

The first MD simulation was performed in the fifties. It described a system of 32 molecules interacting like billiard balls represented by a “hard sphere” model (9). Martin Karplus, Michael Levitt, and Arieh Warshel, which were pioneer in the field, got the Nobel Prize in Chemistry in 2013 “for the development of multiscale models for complex chemical systems”. (10). Molecular dynamics importance is fast growing thanks to the advances in computational power and algorithms. Nowadays, it is possible to run simulations on systems of over a billion particles for microseconds.

Starting from a coordinate set -taken for example from a crystal structure- and assigning velocities to each atom -typically from a Boltzmann distribution at a given temperature-, successive coordinates and velocities are obtained by integrating the Newton’s equation for the motion in each coordinate direction. In one dimension. The evolution of the system in both coordinates and velocities are obtained by integrating the Newton’s equation for the motion in each coordinate direction.

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i}$$

**Formula 1.** One dimension Newton’s equation for movement.  $m_i$  and  $x_i$  are the mass and the coordinate for each atom.  $F_{x_i}$  is the derivative of the potential energy according to a force field equation.

## Methods



The forces on the atoms are estimated by the derivative of the energy. Once the forces are assigned, the Newton law of motion can be used to solve the molecular motion.

The calculation of the interaction energy within a classical description of a molecular system requires a forcefield. A forcefield consists of a set of equations used to generate the potential energies (and their derivatives, the forces) and the parameters used in these equations. The basic functional form of potential energy in molecular mechanics includes bonded interactions and non-bonded interactions. Bonded interactions define bonds stretching, angle bending and dihedrals rotation. Non-bonded interactions include an electrostatic term (Coulomb's law) and a van der the van der Waals term (usually computed with a Lennard-Jones potential) that accounts for both Pauli exclusion and dispersion. Forefield parameters can be obtained from different techniques: 1) theoretical such as ab initio or semi-empirical quantum mechanical calculations or 2) experimental data such as x-ray, electron diffraction, and NMR, Raman and neutron spectroscopy.

The majority of force fields rely on a simple potential energy function:

For bonded interactions:

$$V(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} k_\phi[\cos(n\phi + \delta) + 1] + \sum_{non-bonded\ pairs} \left[ \frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} + \frac{C_{ij}}{r_{ij}^6} \right]$$

For non-bonded interactions:

$$V(r) = \sum_{non-bonded\ pairs} \left( \epsilon_{ij} \left[ \left( \frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 * \left( \frac{R_{min,ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}} \right)$$

**Formula 2.** General form of the potential energy function  $V(r)$ .  $b$  is the interatomic distance,  $k_b$  and  $b_0$  estimate the strength and the equilibrium length of the bond,  $\vartheta$  is the angle formed by the two bond vectors, and  $\vartheta_0$  and  $k_\theta$  are characterizes the strength and equilibrium geometry of the angle. The torsional potential in the equilibrium is given by cosine function, where  $\phi$  is the torsion angle,  $\delta$  is the phase, and  $n$  represents the dihedral potential. The remaining part of the equation refers to the electrostatic interaction, where  $\epsilon_{ij}$  is a parameter related with the two interacting atoms ( $i, j$ ), and  $q_j$  and  $q_i$  are the effective charge on each atom.  $R_{min,ij}$  is the distance making Lennard-Jones equation energy the minimum.  $A$  and  $C$  are the bending of angles.

# ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



The total energy of the system is the sum of the kinetic and potential components displayed in Formula 2. Hence, molecular dynamics simulations can be used to optimize models generated by homology modeling, and dockings where the stability of the ligand binding to the protein can be analyzed along the time (11). Because of the typically large number of atoms involved in calculations, Newton's equation can not be resolved analytically, and several algorithms using different numerical methods have been developed to efficiently solve the equations (12).

The total energy is given by:

$$E_{total} = E_{bonded} + E_{non-bonded} + E_{other}$$

**Formula 3.** Bonded and non-bonded total energy. Includes repulsive, van der Waals, and coulombic interaction.

## **2.5 Molecular docking.**

Molecular docking is a receptor-based technique that theoretically infers the mode(s) of binding of a ligand in to a larger protein of known structure (either crystal data or homology model). The simplest case, in which proteins residues are maintained rigid,, obtaining the best pose for a compound imply the resolution of a problem with six degrees of freedom, three translational and three rotational, making clear that a complete conformational space screening must be performed before the docking itself. Different methodologies have been developed in order to get a reasonable sampling of conformational space in terms of energy and geometric diversity (13). Once all conformations of all ligands have been fitted within the binding site a scoring function evaluates the potential ligands poses in order to create a rank of the most likely mode of binding. Most scoring functions are physics-based and estimate the energy of each pose: the Gibbs free energy ( $\Delta G$ ), which estimation is challenging because of the many contributions to the interaction between small ligand and proteins. Depending on the contributors



taken in consideration and the way their  $\Delta G$ s are estimated, scoring functions are generally divided in three types: force field based, empirical based and knowledge based. The more a scoring function is accurate the more computationally expensive is.

### 2.6 Structure prediction: homology modeling and de novo modeling

Homology modeling consists of building a structural model of a target receptor based on the 3D structure of a phylogenetically related template. Despite of the GPCRs low sequence identity, they share a common 7TM-helices core structure, which facilitates homology modeling. It is been demonstrated the quality of a model depends on the used template (14). Special attention has to be paid to variable regions like receptor loops and ligand binding site. Intracellular and extracellular loops connecting the helices present the higher variability among GPCRs. They change in both space and sequence, making necessary to use more advanced modeling techniques such as *de novo* modeling, knowledge methods or their combination (15). The difficulty to achieve reliable loops was reflected in the last three modeling competition DOCK (2008,2010, and 2013) where none of the presented models reached an RMSD lower than 2 Angstroms for the extracellular loop 2. On the other hand, a good representation of binding site is necessary to obtain confident docking of compounds. As the number of available crystal structures has increased during the last decade, multiple template homology modeling has become more popular. Is been demonstrated multiple template can improve results when no close receptor is available, but no further improvement was found when close templates are available (14).

This was demonstrated in a docking experiment where the retrieve of active compounds depended on the proximity of the model binding site to the crystal structure (14).

# ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



## **2.7 Databases**

Available data from databases is essential for computational biochemistry and bioinformatics research. Databases provide organized and ready to use data from many sources. For example, all crystalized structures are uploaded to the Protein data bank. Some of the most important databases are listed below:

-ZINC database (16): ZINC is a database that contains more than 35 millions of purchasable compounds. This database has been used during the thesis many times, especially to build up ligandfinder (see section 3.2).

-Protein data bank (PDB) (17): This database contains all crystalized structures. It has been used repeatedly during this thesis.

-IUPHAR/BPS guide to pharmacology database (18): Contains information about GPCRs among other receptors. It covers data such as gene information, names, database links, ligands with biological data and structure, transduction mechanisms, tissue distribution, assays and other. This database has also been used in ligandfinder build up.

-UniProt (19): Provides a collection of protein sequences and their annotations. This database has been used to build up GPCR-SAS.



### References

1. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., and Shaw, D. E. (2012) Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* **41**, 429–52
2. Kohlhoff, K. J., Shukla, D., Lawrenz, M., Bowman, G. R., Konerding, D. E., Belov, D., Altman, R. B., and Pande, V. S. (2014) Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat. Chem.* **6**, 15–21
3. Jensen, F. (2007) *Introduction to Computational Chemistry*, 10.1007/s00214-013-1372-6
4. A. Lavecchia, and C.D. Giovanni (2013) Virtual screening strategies in drug discovery: A critical review. *Curr. Med. Chem.* **20**, 2839–2860
5. Song, C. M., Lim, S. J., and Tong, J. C. (2009) Recent advances in computer-aided drug design. *Brief. Bioinform.* **10**, 579–591
6. Dror, O., Shulman-peleg, A., Nussinov, R., and Wolfson, H. J. (2004) Predicting Molecular Interactions
7. Jahn, A., Hinselmann, G., Fechner, N., and Zell, A. (2009) Optimal assignment methods for ligand-based virtual screening. **23**, 1–23
8. Villoutreix, B. O., Renault, N., Lagorce, D., Sperandio, O., Montes, M., and Miteva, M. A. (2007) Free Resources to Assist Structure-Based Virtual Ligand Screening Experiments
9. Alder, B. J., and Wainwright, T. E. (1957) Phase Transition for a Hard Sphere System. *J. Chem. Phys.* **27**, 1208–1209
10. Royal, T. H. E., Academy, S., and Sciences, O. F. (2013) Development of multiscale models for complex chemical systems. *Nobel Prize Chem. 2013*
11. Potter, M. J., Kirchhoff, P. D., Carlson, H. A., and McCammon, J. A. (1999) Molecular dynamics of cryptophane and its complexes with tetramethylammonium and neopentane using a continuum solvent model. *J. Comput. Chem.* **20**, 956–970
12. Vlachakis, D., Bencurova, E., Papangelopoulos, N., and Kossida, S. (2014) *Current state-of-the-art molecular dynamics methods and applications*, 1st Ed., Elsevier Inc., 10.1016/B978-0-12-800168-4.00007-X

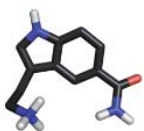


## ***Development and application of computational techniques to drug discovery and structure-function relationships***

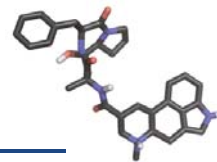
---



13. Lu, J., Kong, X., Liang, Z., Luo, C., and Jiang, H. (2012) Computational drug discovery. *Acta Pharmacol. Sin.* **33**, 1131–1140
14. Beuming, T., and Sherman, W. (2012) Current assessment of docking into GPCR crystal structures and homology models: Successes, challenges, and guidelines. *J. Chem. Inf. Model.* **52**, 3263–3277
15. Goldfeld, D. A., Zhu, K., Beuming, T., and Friesner, R. A. (2013) Loop prediction for a GPCR homology model: Algorithms and results. *Proteins Struct. Funct. Bioinforma.* **81**, 214–228
16. Irwin, J. J., and Shoichet, B. K. (2005) ZINC – A Free Database of Commercially Available Compounds for Virtual Screening ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **45**, 177–182
17. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242
18. Pawson, A. J., Sharman, J. L., Benson, H. E., Faccenda, E., Alexander, S. P. H., Buneman, O. P., Davenport, A. P., Mcgrath, J. C., Peters, J. A., Southan, C., Spedding, M., Yu, W., and Harmar, A. J. (2014) The IUPHAR / BPS Guide to PHARMACOLOGY : an expert-driven knowledgebase of drug targets and their ligands. **42**, 1098–1106
19. The UniProt Consortium (2014) UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–12



## Objectives



### 3 Objectives

Computer-aided drug discovery techniques have played a major role in the development of therapeutically important small molecules. These methods are classified as either ligand-based or structure-based methods.

Ligand-based methods use only ligand information for predicting activity depending on its similarity/dissimilarity to known structures. Thus, an objective of the thesis is:

- The development of LigandFinder, a web-based virtual screening application. The aim of this tool is to screen a library of compounds to find new, commercially available, ligands similar to a given set of known structure. This application aims to offer to the scientific community a free user-friendly tool to seek potentially active drugs.

Structure-based methods can be used when both target and ligand structures are known. These approaches mainly include homology modeling, ligand docking, and structure-based pharmacophore modeling, among others. Recently, it has become state of the art to use additional computational tools such as molecular dynamics and molecular mechanics to simulate and evaluate the conformational space of a protein. Thus, another objective of the thesis is:

- Design selective, versus the 5-HT<sub>1A</sub> receptor, ligands for the serotonin 5-HT<sub>7</sub> receptor using homology modeling, ligand docking, and molecular dynamics simulations.

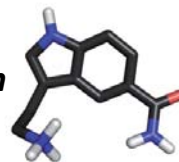
For closely related protein families, key structurally conserved and functional regions can be identified from multiple sequence analysis. Thus, in this thesis we also aim at the:

- Development of GPCR-SAS, a web-based application for statistical analysis of G protein-coupled receptor sequences (classes A, B, C, and F).

Over 60% of currently marketed drugs target membrane proteins. Due to the lipidic environment of membrane proteins, dispersion forces (mainly aromatic-aromatic, aromatic-aliphatic or aliphatic-aliphatic) are involved in stabilizing the tertiary structure of the protein or in structural changes. Surprisingly, non-

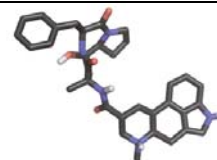
***Development and application of computational techniques to drug discovery and structure-function relationships***

---



bonded interactions involving sulfur-containing amino acids (Met and Cys) have received little attention, in contrast to interactions involving aromatic or hydrophobic amino acids. Thus, in this thesis we aim to:

- Evaluate the occurrence of interactions involving Met and Cys side-chains in crystal structures of membrane proteins and to characterize their strength in small-molecule model systems at the ab-initio level.



## 4 Results

### 4.1 Overview

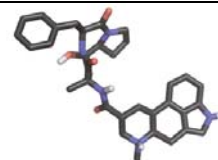
Launching a new drug into the market is an extremely expensive (about 1 billion \$) and long (more than 10 years) process. In average, for every 10.000 compounds tested, five enter clinical trials, and only one is eventually approved for patient use. However, and in spite of these difficulties, the investigation in new drugs is intense, since the economical and social benefits are very high. In fact, the top 50 pharma companies accounted for \$593.4 billion of prescription drug sales in 2010, and their investment in research reached a record \$65.3 billion in 2009.

Therefore, it is important to use new technologies in the R&D process of obtaining new compounds. By means of theoretical studies of ligand-target interactions, computational techniques have a fundamental role in this slow process as they help speeding up the early stage of identify “hits compounds” (new compounds with limited efficacy) and accelerate the “hit-to-lead” (lead compound is a chemically optimized form of the hit compounds) process, reducing both costs and time of the drug discovery and optimization. High-throughput virtual screening, where chemical libraries of million of compounds are virtually tested against the target of interest to find potential active molecules, is the first approximation of a computational drug design study. LigandFinder is a tool designed to let computationally unskilled scientists to easy perform 2D virtual screening. The algorithm filters compounds from a database of more than 20 millions of molecules to find the most similar according to a new concept “centroid”.

Once potential hits are detected, the next step in the drug discovery pathway is a more exhaustive screening procedure, using more computationally demanding techniques. Molecular docking fits compounds into the binding site of the receptor. The obtained docking poses are sorted by a scoring function that measures the quality of the interactions between the ligand and the receptor. However, due to the dynamic nature of proteins, it is insufficient to use a single static structure to predict putative binding modes. Molecular dynamics simulations can be used for

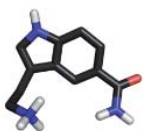
## *Results: Overview*

---

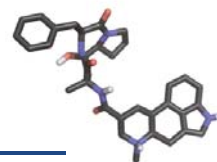


obtaining an ensemble of target conformations. We used these techniques to explore the binding modes of designed ligands into the serotonin 5-HT<sub>7</sub> and 5-HT<sub>1A</sub> receptors.

On the other hand, integral polytopic membrane proteins mediate the interaction of the cell with its surroundings. Because of their relevance to cellular physiology and their accessibility from the extracellular environment, membrane proteins represent a significant portion of therapeutic drug targets. Particularly G protein-coupled receptors (GPCRs), transport proteins and ion channels are among the most prominent target families for the pharmaceutical industry. Biological function of these membrane proteins involves conformational rearrangement of the transmembrane (TM) regions. For example, activation of the GPCR family requires the binding of the C-terminal  $\alpha$ -helix of the G protein to the intracellular cavity that is opened by the conformational rearrangement of TM 6. Similarly, multidrug transporters are flexible proteins that switch from outward-open to inward-open conformations, facilitating the release of the substrate. Such conformational changes require disruption and formation of key inter-helical interactions. Here, we have studied the role of sulfur-containing amino acids (Met and Cys) in forming inter-helical interactions. These residues are unique as they are highly polarizable due to the sulfur atom. Surprisingly, non-bonded interactions involving Met and Cys have received little attention in contrast to interactions involving aromatic amino acids.



## Results: Summary and contribution



### 4.2 Summary and contribution

**LigandFinder: A user-friendly virtual-screening web server.** The objective of this work is the development of a web-accessible tool to search for similar compounds. Users can either introduce a list of compounds (smiles or by drawing) or select them from a database of GPCRs ligands classified by receptor. Selected compounds are fused in a “super-ligand” which integrates all properties. This ligand is then confronted with a subset (size of the subset depends on the super-ligand properties) and a list of compounds sorted by the similarity to the super-ligand is returned. The application can be also used to explore GPCRs ligands properties.

Ligandfinder is available at <http://lmc.uab.cat/ligandfinder> and the manuscript is submitted to *Bioinformatics* journal.

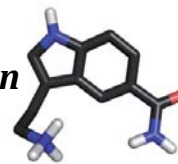
#### Study of sulfur-containing amino acids

This work arose after finding the importance in binding affinities of sulfur-containing amino acids in mutation experiments on cholecystokinin receptors (CCK1R and CCK2R) (1, 2). First, we published a review about the prevalence of these residues in GPCRs together with several examples found in literature where the presence of these residues were critical for protein functioning/ligand binding. During the analysis, we found that 47% of the aromatic residues in GPCRs are interacting with sulfur-containing amino acids. We then wondered if sulfur, the biggest essential element with fully filled 3p and empty 3d orbitals, could be conferring additional properties to these residues with respect to other residues. In order to further explore the nature of this interactions in membrane proteins, we analyzed the complete Protein Data Bank looking for interactions where these residues were involved, clustered them geometrically, and applied quantum mechanical calculations to obtain energies.

This work resulted in two publications: *Sulfur-containing amino acids in 7TMRs: Molecular gears for pharmacology and function* (3) and *Analysis Of The Interactions Of Sulfur-Containing Amino Acids in Membrane Proteins* (4).

# ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



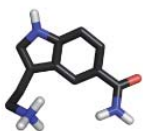
## **GPCRSAS**

GPCRSAS (G-protein coupled receptor sequence analysis and statistic) is a web application, which takes advantage of the structural similarity among GPCRs transmembrane regions to perform statistical analysis of sequence positions or motifs within the transmembrane helices of GPCR A, B, C, and F classes. GPCRSAS provides different types of analysis such as position/positions/range of positions conservation, entropy, co-evolutionary (co-variance) and correlation analysis. All this, applied to set of receptors selected through a hierarchical dropdown menu with different scheme of classifications and their respective levels and sublevels.

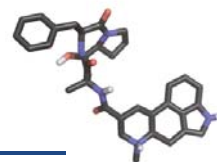
The application is available at <http://lmc.uab.cat/gpcrsas/> and the manuscript “GPCR-SAS: G protein-coupled receptors Sequence Analysis and Statistics” is under preparation for submission to *Plos Computational Biology* journal.

## **Selectivity in serotonin 5-HT7 receptor**

The objective of this project was to determine the structural basis of selectivity of a series of compounds towards 5-HT7. The difficulty to address this problem comes from the high similarity between 5-HT7 and 5-HT1A receptors binding site. Activation of 5-HT7 receptor has been related with nociceptive processing, learning and memory. On the other hand, blockage of the receptor leads to an antidepressant activity. The group of Maria L. López-Rodríguez, from Universidad Complutense de Madrid, synthesized the compounds and tested their affinity towards 5-HT7 and 5-HT1A receptors. Compounds share two moieties and a spacer of variable length. Results showed higher selectivity of the compounds towards 5-HT7 receptor. Interestingly, compounds with spacer length equal or longer than 6 carbons, become fully selective antagonists to 5-HT7. We proposed these compounds are no longer able to occupy the whole orthosteric binding site, triggering a reverse binding mode towards the extracellular entrance. This binding mode, similar to the ergotamine binding in 5-HT<sub>1b</sub> crystal structure, involves the non-conserved extracellular part of the receptor thus providing full selectivity. In order to further characterize the residues involved in recognition as well as stability of the compounds, receptors were modeled, compounds docked, and molecular dynamics simulations performed.



## Results: Summary and contribution



This work resulted in a publication “*The extracellular entrance provides selectivity to serotonin 5-HT 7 receptor antagonists with antidepressant-like behavior in vivo*” published in *Journal of Medicinal Chemistry* (5).

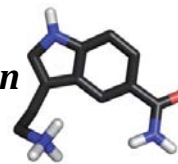
### References

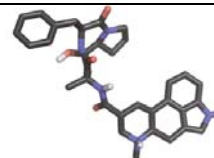
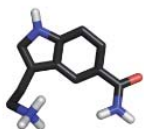
1. Gigoux, V., Escriet, C., Silvente-Poirot, S., Maigret, B., Gouilleux, L., Fehrentz, J. A., Gully, D., Moroder, L., Vaysse, N., and Fourmy, D. (1998) Met-195 of the cholecystokinin-a receptor interacts with the sulfated tyrosine of cholecystokinin and is crucial for receptor transition to high affinity state. *J. Biol. Chem.* **273**, 14380–14386
2. Escriet, C., Gigoux, V., Archer, E., Verrier, S., Maigret, B., Behrendt, R., Moroder, L., Bignon, E., Silvente-Poirot, S., Pradayrol, L., and Fourmy, D. (2002) The biologically crucial C terminus of cholecystokinin and the non-peptide agonist SR-146,131 share a common binding site in the human CCK1 receptor. Evidence for a crucial role of Met-121 in the activation process. *J. Biol. Chem.* **277**, 7546–7555
3. Cordoní, A., Gómez-Tamayo, J. C., Gigoux, V., and Fourmy, D. (2013) Sulfur-containing amino acids in 7TMRs: Molecular gears for pharmacology and function. *Trends Pharmacol. Sci.* **34**, 320–331
4. Gómez-Tamayo, J. C., Cordoní, A., Olivella, M., Mayol, E., Fourmy, D., and Pardo, L. (2016) Analysis Of The Interactions Of Sulfur-Containing Amino Acids In Membrane Proteins. *Protein Sci.* **00**, 1–8
5. Medina, R. A., Vázquez-Villa, H., Gómez-Tamayo, J. C., Benhamú, B., Martín-Fontecha, M., De La Fuente, T., Caltabiano, G., Hedlund, P. B., Pardo, L., and López-Rodríguez, M. L. (2014) The extracellular entrance provides selectivity to serotonin 5-HT 7 receptor antagonists with antidepressant-like behavior in vivo. *J. Med. Chem.* **57**, 6879–6884



***Development and application of computational techniques to drug discovery and structure-function relationships***

---





---

### **4.3 LigandFinder: A user-friendly virtual-screening web server.**

#### **Abstract**

LigandFinder is a flexible user-friendly web application that allows fast virtual screening to find new (commercially available) compounds similar to a set of compounds of known structure. It explores the chemical space of a database with more than 20M compounds and has been designed for users with no computational background, which can easily find new different compounds with chemical features similar to the input compound(s). To our knowledge it is the first free web service that allows the use of multiple ligands (instead of just one) as input, of which it detects their shared chemical features and performs a 2D virtual screening accordingly. Additionally, using the same tool a pre-computed database of GPCRs ligands have been designed, thus allowing the users to quickly explore new possible GPCRs ligands.

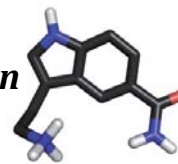
#### **Introduction**

The growth of available compounds databases and the advances in computational performance has enhanced the role of virtual screening in the process of the discovery of new pharmacologically active compounds. Indeed it helps the drug discovery process to lower its cost and to boost the speed in obtaining a hit compound (1, 2).

Virtual screening is generally classified in two main methods, depending on whether the structure of the receptor is available (structure-based or SBVS) or not (ligand-based or LBVS). SBVS involves the generation and docking of hundreds of conformations of known ligands making the experiment efficient but computationally expensive. On the other hand, among the ligand-based virtual screening methods, which include pharmacophores, shaped-based similarity and fingerprint similarity, those based on similarity comparison of compounds properties (fingerprints) are computationally inexpensive and can be used to fast explore huge databases (1). All these methods rely on the assumption that similar molecules share a similar biological activity (3). Unfortunately all virtual

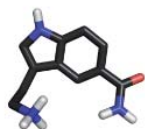
## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---

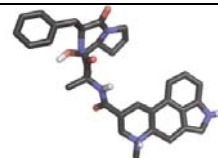


screening methods need the use of often-expensive programs, designed to be performed by users highly skilled in computational chemistry, making them inaccessible to the vast scientific community.

Herein, we present a new web application, LigandFinder, which allows computationally untrained chemists/biologists to perform a virtual screening of the full ZINC database (4), the biggest free database of commercially available compounds, using either one or a set of ligands as input. LigandFinder calculate more than 150 chemical properties of each input ligand, detect those shared by all input compounds and browse the full database in search of chemically similar compounds. In few minutes a list of potential ligands is given as output, each one identified by its ZINC id. The list of calculated parameters (see supplemental table 2 for a detailed list) is highly focused on compounds' chemical properties and on their functional groups and includes constitutional descriptors, ring descriptors and molecular properties. Advanced users will also be free to modify the descriptors boundaries and adapt their search according to their knowledge of the system. Additionally, a database of pre-calculated parameters for known GPCRs ligands is available (classified by family, receptor and activity). This database can be used to see statistics of ligand properties for desired receptors as well as to search for similar, still unknown, compounds to a subset of receptor ligands. This verified and known set of ligands was also used to validate our VS algorithm, which performance is shown in supplementary table 1.



## Results: LigandFinder



**LIGANDFINDER**

COMPOUND SEARCH    GPCRs LIGAND SEARCH    ABOUT

LigandFinder is a flexible user-friendly virtual screening web application to find new ligands similar to one or more set of input compounds. It explores the chemical space of a database of more than 20M commercially available compounds. Users can also explore pre-computed GPCRs ligands and search for similar compounds.

**DRAW SKETCH**

**SMILES**

INSERT AN SMILE AND CLICK TO ADD

**TYPE OF SEARCH**

Strict  
More info

Similarity  
More info

CLICK TO ADD AN SMILE FROM DRAW

SUBMIT SMILES

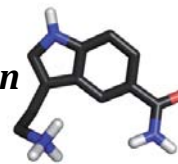
*Figure 1. LigandFinder main page.*

## Methods

### Dataset Building

3D structures were downloaded from ZINC database “all clean” (2013-12-18) and “all boutique” (2012-11-27) and parameterized using DRAGON6 (5) (detailed list of parameters in Supplemental Material). The database was curated removing erroneous molecules (duplicates and/or wrongly parameterized) and by adding the formal charge to the parameter list. The GPCR database, was built downloading ligands from IUPHAR/BPS database (October 2015) (6) and organized by the receptor(s) they bind to, its sub-family, ligand activity (agonist

## Development and application of computational techniques to drug discovery and structure-function relationships



or antagonist), type (small ligand or peptide). These ligands were converted to 3D structure using Babel (7), parameterized with DRAGON6 (5) and curated as described above.

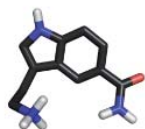
The screenshot displays a web interface for ligand parameterization. At the top, three 3D ball-and-stick molecular models are shown in separate boxes. Below them is a red 'SUBMIT' button. Underneath is a blue header for 'PARAMETERS'. The 'BASIC' section contains a table with the following parameters and their min/max values:

Parameter	Min	Max
MW (molecular weight)	122.18	436.92
charge (formal charge)	0	0
nBnz (number of benzene-like rings)	0	2
nHDon (number of donor atoms for H-bonds (N and O))	0	10
nHAcc (number of acceptor atoms for H-bonds (N,O,F))	0	6

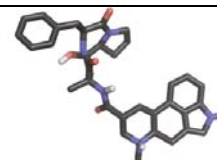
**Figure 2.** 3D structure and parameter limits for three example molecules.

### Ligand Similarity Search

The set of input ligands allows building an “ideal ligand”, called centroID, which set of parameters will be used to browse our parameterized ZINC database (see supplementary fig 1 and table 2). These parameters consist of ranges of values,



## Results: LigandFinder



based on observed input values. For example, input compound 1 has 3 hydrogen acceptors (HA), compound 2 has 1 and compound 3 has 4: our centroid's HA parameter will be 1-4 (2 being penalized since it is not observed). LigandFinder allows the user to perform searches using three level of flexibility: strict, semi-strict and flexible (with caveats in case the search's output will results in too many hits). For each parameter, strict search considers as range limits the lowest and the maximum value observed in input compounds (as described in above example).

The screenshot displays the LigandFinder interface with four main sections:

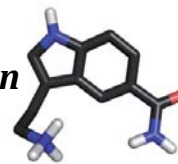
- LIGANDS**: A list of ligands with checkboxes for selection. The selected ligand is LY293284.
- 3D STRUCTURE**: A 3D ball-and-stick model of the selected ligand, LY293284.
- LIGANDS STATISTICS**: A table showing average, standard deviation (Sd), maximum (Max), and minimum (Min) values for various parameters across the selected ligands.
- SELECTED LIGAND PARAMETERS**: A table showing the specific values for the selected ligand (S-14671) for the same parameters.

Parameter	Average	Sd	Max	Min
MW	351.36	87.93	655.68	177.25
RBN	4.63	2.21	9.0	0.0
charge	1.06	0.24	2.0	1.0
nDB	1.14	1.08	5.0	0.0
nAB	11.72	4.47	23.0	0.0
nTB	0.06	0.24	1.0	0.0
nBnz	1.42	0.76	3.0	0.0
nHDon	2.38	1.17	6.0	1.0

Parameter	Value
Name	S-14671
MW	396.580
RBN	6
charge	1
nDB	1
nAB	16
nTB	0
nBnz	2

**Figure 3.** Serotonin receptor A ligands summary. Top left square contains the list of ligands. Clicking over the name changes right squares to the 3D structure (top) and parameter list (bottom) of the selected ligand. Bottom left square shows average parameter values.

# Development and application of computational techniques to drug discovery and structure-function relationships



Semi-strict search, suggested if strict search gives poor results, increases the range limits of the five most likely parameters (probability of a parameter to be different to zero) by  $\pm$  one unit. Finally, flexible search, suggested if the others options give poor results or the user prefers a bigger list of compounds, modifies all parameters limits by  $\pm$  one unit. Molecular weight is used as pre-filter to focus the search due to computational resources performance. If the difference between upper and lower molecular weight limits is less than 30M for strict search, less than 50 for semi-strict search, and less than 60 for flexible search, molecular weight limits are enhanced by  $\pm 15M$ ,  $\pm 25M$ , and  $\pm 30M$  from the compounds' MW means respectively.

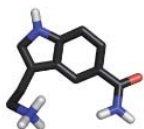
## Similarity

Global similarity between ligands and centroID is calculated (equation a) as the average of all parameters similarity. Single parameter similarity among hit ligands and the centroID is calculated as shown in eq. 1b, in which each Tanimoto-like parameter's similarity value is weighted by the distribution of values observed in the input set of ligands.

$$S = \sum_1^{150} S_i \quad (a)$$

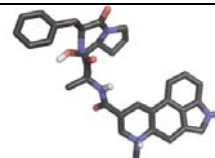
$$S = 1 - \sum_i [ (1 - L/nQi) * P_t ] \quad (b)$$

**Equation 1.** a) Global similarity between a hit ligand and the centroID (Query). b) Tanimoto similarity weighted coefficient for each parameter comparison. Since centroid may have different values for each property instead of a single one, parameter's similarity must be weighed according to the values observed in the query.  $P_i$  is the theoretical probability for each event to occur in parameter  $i$  as observed in the centroID (Query),  $nQi$  is the sum of observed events for parameter  $i$  in query, and  $L$  is the value of the event found in query ligand in parameter  $i$ . An Excel file with an example is provided in Supplemental Material.



## Results: LigandFinder

---

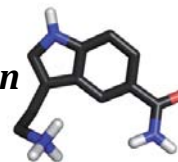


### Web interface

LigandFinder's main page shows the general compound search option as default. Ligands can be loaded both *via* a JSME chemical drawer (8) or pasted in smile format. A window allows the user to define the type of search (strict, semi-strict, and flexible), and sorting scheme of output ligands (by similarity or dissimilarity). Once the ligands are submitted a new page with the 3D structure representation of the molecules (9), their statistics and the Centroid descriptors limits is loaded. The user can either run the search or manually modify descriptors' limits. Once the calculation is finished, a list of ligands (up to 2000) is given as result (2D sketch, ZINC id and link to zinc database) sorted either by their similarity or dissimilarity with respect to the centroid. Results are shown in lists of 100 compounds, more molecules can be loaded clicking on "next 100" button. Additionally, a pdf (with the molecules displayed on the page) and bash (linux/MAC OS) and microsoft vbscript scripts to download smiles or sdf files of the whole search is provided to the user.

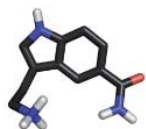


# Development and application of computational techniques to drug discovery and structure-function relationships



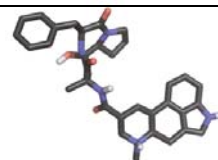
**Figure 4.** Search results. Found ligands are sorted by their similarity with respect to the centroid.

GPCRs' "new hits" service offers the user the possibility to search for similar compounds for GPCRs known ligands (small molecules). The user is required to select the family, the receptor and ligand's activity (agonist or antagonists). Upon submission, a page with the available ligands, statistics, 3D conformer and descriptors ranges for selected ligand is loaded. Users can select the ligands required for the similarity search (if none is selected the search will run using all) and select the type of search as previously described. The user can modify centroid's limits or just submit the calculation. The result is given in the same way as previously described. Once the configuration is set, the user can submit advanced search to modify parameters limits or run directly the search.



## *Results: LigandFinder*

---

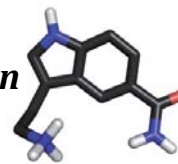


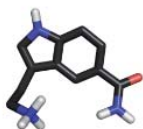
### **Conclusion**

With LigandSuit we aim to offer the user the experience to run his own 2D virtual screening without being specifically trained for it, still giving them the possibility to modify all descriptors, which are simple, understandable and easy to set for any user.

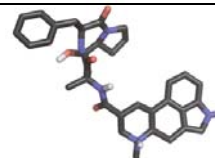
***Development and application of computational techniques to drug discovery and structure-function relationships***

---





## Results: LigandFinder

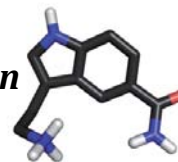


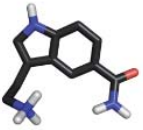
### References

1. A. Lavecchia, and C.D. Giovanni (2013) Virtual screening strategies in drug discovery: A critical review. *Curr. Med. Chem.* **20**, 2839–2860
2. Lu, J., Kong, X., Liang, Z., Luo, C., and Jiang, H. (2012) Computational drug discovery. *Acta Pharmacol. Sin.* **33**, 1131–1140
3. Martin, Y. C., Kofron, J. L., and Traphagen, L. M. (2002) Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **45**, 4350–4358
4. Irwin, J. J., and Shoichet, B. K. (2005) ZINC – A Free Database of Commercially Available Compounds for Virtual Screening ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **45**, 177–182
5. Mauri, a, Consonni, V., Pavan, M., and Todeschini, R. (2006) Dragon software: An easy approach to molecular descriptor calculations. *Match Commun. Math. Comput. Chem.* **56**, 237–248
6. Pawson, A. J., Sharman, J. L., Benson, H. E., Faccenda, E., Alexander, S. P. H., Buneman, O. P., Davenport, A. P., Mcgrath, J. C., Peters, J. A., Southan, C., Spedding, M., Yu, W., and Harmar, A. J. (2014) The IUPHAR / BPS Guide to PHARMACOLOGY : an expert-driven knowledgebase of drug targets and their ligands. **42**, 1098–1106
7. O'Boyle, N. M., Banck, M., James, C. a., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011) Open Babel: An Open chemical toolbox. *J. Cheminform.* **3**, 1–14
8. Bienfait, B., and Ertl, P. (2013) JSME: A free molecule editor in JavaScript. *J. Cheminform.* **5**, 1
9. Rego, N., and Koes, D. (2014) 3Dmol . js : Molecular Visualization with WebGL

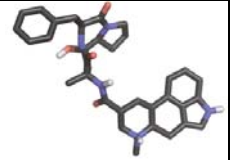
***Development and application of computational techniques to drug discovery and structure-function relationships***

---

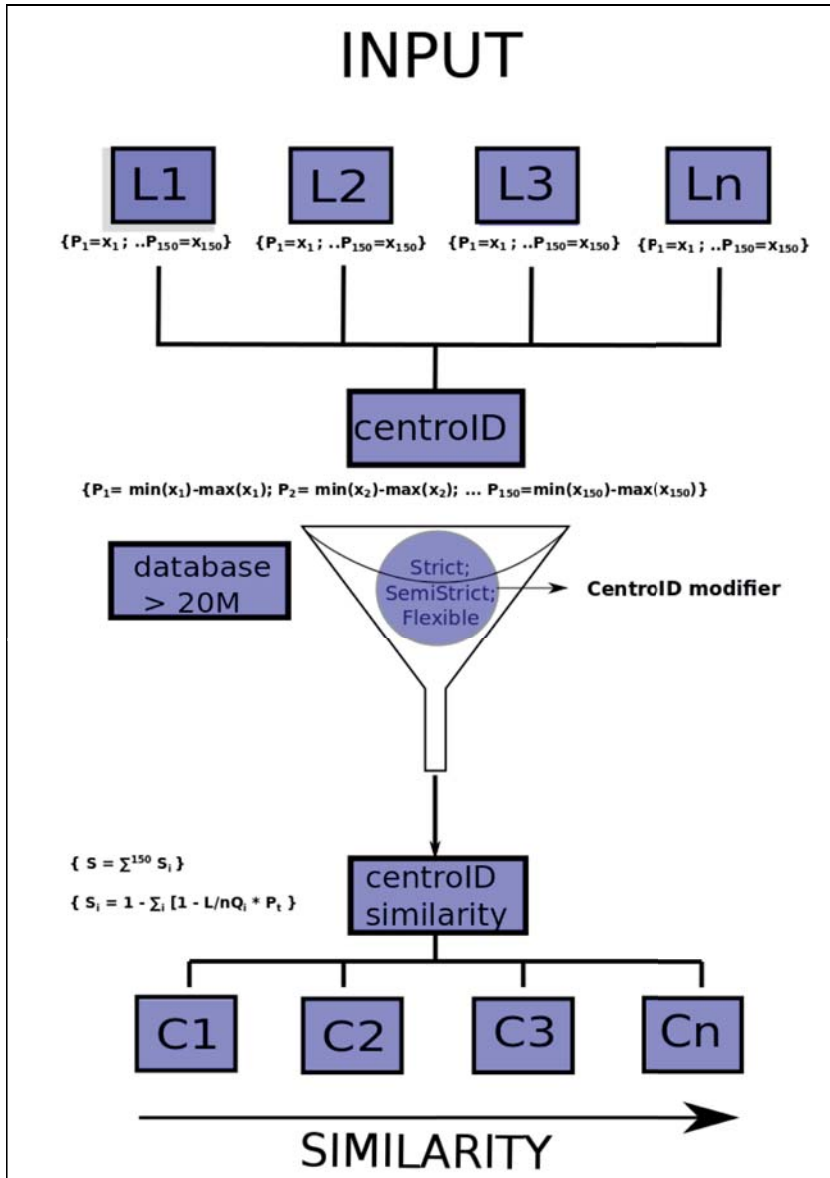




## Results: LigandFinder

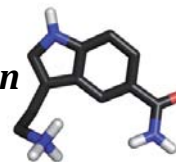


### Supplementary information



Supplementary Figure 1. Flowchart of LigandFinder virtual screening algorithm.

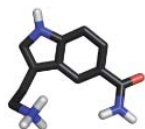
# Development and application of computational techniques to drug discovery and structure-function relationships



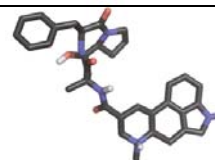
Receptor	Activity	3 ligands	6 ligands	9 ligands	Total
ADORA2	agonists	6	13	25	29
ADRB2	agonists	8	11	14	18
TACR1	antagonists	4	12	15	21

**Supplementary table1.** Compound retrieval using 3, 6 and 9 randomly selected ligands as training set from a total number of ligands (last column) reported by IUPHAR to bind 3 GPCR receptors (Adenosine receptor A2 agonists, adrenergic beta 2 agonists and Tachykinin Receptor 1 antagonists).

Parameter	Description
MW	molecular weight
RBN	number of rotatable bonds
charge	formal charge
nDB	number of double bonds
nAB	number of aromatic bonds
nTB	number of triple bonds
nBnz	number of benzene-like rings
nHDon	number of donor atoms for H-bonds (N and O)
nHAcc	number of acceptor atoms for H-bonds (N,O,F)
nX	number of halogen atoms
nR03	number of 3-membered rings
nR04	number of 4-membered rings
nR05	number of 5-membered rings
nR06	number of 6-membered rings
nR07	number of 7-membered rings
nR09	number of 9-membered rings
nR10	number of 10-membered rings
nR11	number of 11-membered rings
nR12	number of 12-membered rings
nR=Cp	number of terminal primary C(sp <sup>2</sup> )
nR=Cs	number of aliphatic secondary C(sp <sup>2</sup> )
nR=Ct	number of aliphatic tertiary C(sp <sup>2</sup> )
n=C=	number of allenes groups
nR#CH/X	number of terminal C(sp)
nR#C-	number of non-terminal C(sp)
nROCN	number of cyanates (aliphatic)
nArOCN	number of cyanates (aromatic)
nRNCO	number of isocyanates (aliphatic)
nArNCO	number of isocyanates (aromatic)
nRSCN	number of thiocyanates (aliphatic)
nArSCN	number of thiocyanates (aromatic)



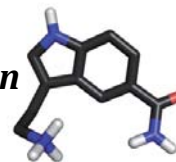
## Results: LigandFinder



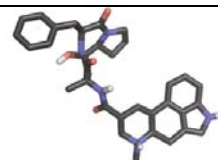
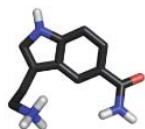
nRNCS	number of isothiocyanates (aliphatic)
nArNCS	number of isothiocyanates (aromatic)
nRCOOH	number of carboxylic acids (aliphatic)
nArCOOH	number of carboxylic acids (aromatic)
nRCOOR	number of esters (aliphatic)
nArCOOR	number of esters (aromatic)
nRCONH2	number of primary amides (aliphatic)
nArCONH2	number of primary amides (aromatic)
nRCONHR	number of secondary amides (aliphatic)
nArCONHR	number of secondary amides (aromatic)
nRCONR2	number of tertiary amides (aliphatic)
nArCONR2	number of tertiary amides (aromatic)
nROCON	number of (thio-) carbamates (aliphatic)
nArOCON	number of (thio-) carbamates (aromatic)
nRCOX	number of acyl halogenides (aliphatic)
nArCOX	number of acyl halogenides (aromatic)
nRCSOH	number of thioacids (aliphatic)
nArCSOH	number of thioacids (aromatic)
nRCSSH	number of dithioacids (aliphatic)
nArCSSH	number of dithioacids (aromatic)
nRCOSR	number of thioesters (aliphatic)
nArCOSR	number of thioesters (aromatic)
nRCSSR	number of dithioesters (aliphatic)
nArCSSR	number of dithioesters (aromatic)
nRCHO	number of aldehydes (aliphatic)
nArCHO	number of aldehydes (aromatic)
nRCO	number of ketones (aliphatic)
nArCO	number of ketones (aromatic)
nCONN	number of urea (-thio) derivatives
nC=O(O)2	number of carbonate (-thio) derivatives
nN=C-N<	number of amidine derivatives
nC(=N)N2	number of guanidine derivatives
nRC=N	number of imines (aliphatic)
nArC=N	number of imines (aromatic)
nRCNO	number of oximes (aliphatic)
nArCNO	number of oximes (aromatic)
nRNH2	number of primary amines (aliphatic)
nArNH2	number of primary amines (aromatic)
nRNHR	number of secondary amines (aliphatic)
nArNHR	number of secondary amines (aromatic)
nRNHR2	number of tertiary amines (aliphatic)
nArNR2	number of tertiary amines (aromatic)
nN-N	number of N hydrazines
nN=N	number of N azo-derivatives
nRCN	number of nitriles (aliphatic)
nArCN	number of nitriles (aromatic)
nN+	number of positively charged N
nNq	number of quaternary N
nRNHO	number of hydroxylamines (aliphatic)
nArNHO	number of hydroxylamines (aromatic)



# Development and application of computational techniques to drug discovery and structure-function relationships



nRNN0x	number of N-nitroso groups (aliphatic)
nArNNOx	number of N-nitroso groups (aromatic)
nRNO	number of nitroso groups (aliphatic)
nArNO	number of nitroso groups (aromatic)
nRNO2	number of nitro groups (aliphatic)
nArNO2	number of nitro groups (aromatic)
nN(CO)2	number of imides (-thio)
nC=N-N<	number of hydrazones
nROH	number of hydroxyl groups
nArOH	number of aromatic hydroxyls
nOHp	number of primary alcohols
nOHs	number of secondary alcohols
nOHt	number of tertiary alcohols
nROR	number of ethers (aliphatic)
nArOR	number of ethers (aromatic)
nROX	number of hypohalogenides (aliphatic)
nArOX	number of hypohalogenides (aromatic)
nO(C=O)2	number of anhydrides (-thio)
nH2O	number of water molecules
nSH	number of thiols
nC=S	number of thioketones
nRSR	number of sulfides
nRSSR	number of disulfides
nSO	number of sulfoxides
nS(=O)2	number of sulfones
nSOH	number of sulfenic (thio-) acids
nSOOH	number of sulfinic (thio-/dithio-) acids
nSO2OH	number of sulfonic (thio-/dithio-) acids
nSO3OH	number of sulfuric (thio-/dithio-) acids
nSO2	number of sulfites (thio-/dithio-)
nSO3	number of sulfonates (thio-/dithio-)
nSO4	number of sulfates (thio-/dithio-)
nSO2N	number of sulfonamides (thio-/dithio-)
nPO3	number of phosphites/thiophosphites
nPO4	number of phosphates/thiophosphates
nPR3	number of phosphanes
nP(=O)O2R	number of phosphonates (thio-)
nP(=O)R3/nPR5	number of phosphoranes (thio-)
nCH2RX	number of CH2RX
nCHR2X	number of CHR2X
nCR3X	number of CR3X
nR=CHX	number of R=CHX
nR=CRX	number of R=CRX
nR#CX	number of R#CX
nCHRX2	number of CHRX2
nCR2X2	number of CR2X2
nR=CX2	number of R=CX2
nCRX3	number of CRX3
nArX	number of X on aromatic ring



## Results: LigandFinder

---

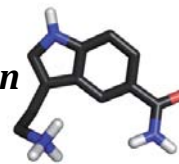
nCXr	number of X on ring C(sp3)
nCXr=	number of X on ring C(sp2)
nCconjX	number of X on exo-conjugated C
nAziridines	number of Aziridines
nOxiranes	number of Oxiranes
nThiranes	number of Thiranes
nAzetidines	number of Azetidines
nOxetanes	number of Oxetanes
nThioethanes	number of Thioethanes
nBeta-Lactams	number of Beta-Lactams
nPyrrolidines	number of Pyrrolidines
nOxolanes	number of Oxolanes
nH-Thiophenes	number of tetrahydro-thiophenes
nPyrroles	number of Pyrroles
nPyrazoles	number of Pyrazoles
nImidazoles	number of Imidazoles
nFuranes	number of Furanes
nThiophenes	number of Thiophenes
nOxazoles	number of Oxazoles
nIsoxazoles	number of Isoxazoles
nThiazoles	number of Thiazoles
nIsothiazoles	number of Isothiazoles
nTriazoles	number of Triazoles
nPyridines	number of Pyridines
nPyridazines	number of Pyridazines
nPyrimidines	number of Pyrimidines
nPyrazines	number of Pyrazines
n135-Triazines	number of 1-3-5-Triazines
n124-Triazines	number of 1-2-4-Triazines

---

**Supplementary table 2.** List of Dragon6's parameters used as fingerprints in our database.

***Development and application of computational techniques to drug discovery and structure-function relationships***

---





---

## **4.4 Sulfur-containing amino acids**

### **PART I**

#### **Sulfur-containing amino acids in 7TMRs: molecular gears for pharmacology and function**

##### **Abstract**

Seven-transmembrane receptors (7TMRs) mediate the majority of physiological responses to hormones and neurotransmitters in higher organisms. Tertiary structure stability and activation of these versatile membrane proteins require formation or disruption of complex networks of well-recognized interactions (such as H-bonds, ionic or aromatic-aromatic) but also of other type of interactions which have been less studied. In this review article we compile evidences from crystal structure, biophysical and site-directed mutagenesis data that indicate or support the importance of interactions involving Met and Cys in 7TMRs in terms of pharmacology and function. We show examples of Met/Cys-aromatic and Met-Met interactions participating in ligand binding, in tuning the orientation of functionally important aromatic residues during activation or even in modulating the type of signaling response. Collectively, data presented enlarge the repertoire of interactions governing 7TMR functioning.

##### **General considerations about structural features of 7TMRs**

7TMRs, also known as G protein-coupled receptors (GPCRs), mediate nearly all-human cellular responses to hormones and neurotransmitters. Therefore, they are one of the most attractive target families for drug discovery and already comprise about 30% of current therapeutic agents on the market Overington et al. (1). Significant advances in crystallization of 7TMRs have permitted elucidation of the structures of many receptors (2-22). These represent invaluable tools for understanding how 7TMRs function at residue and molecular levels. There is considerable evidence that, despite a remarkable diversity in ligands, the ligand-encoded extracellular signal is propagated from the binding site into intracellular domains through a common activation

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



mechanism. These would subsequently trigger further signaling pathways. For example, when 7TMRs activate G proteins, an outward tilt of the intracellular part of transmembrane (TM) 6 occurs together with a movement of TM5 towards TM6 and a side-chain extension of Arg<sup>3.50</sup> within the (E/D)RY motif in TM3 towards the protein core, to interact with the highly conserved Tyr<sup>5.58</sup> in TM5 and with Tyr<sup>7.53</sup> of the NPxxY motif in TM7 (superscripts refers to the Ballesteros and Weinstein numbering scheme (23)) (24-28). All these steps require formation and breakage of non-covalent interactions such as ionic, hydrogen bond and dispersion-stabilized interactions (including aliphatic-aliphatic, aromatic-aromatic, aromatic-aliphatic), most of which are well described (glossary and Table 1). There are, however, other types of dispersive interactions present in 7TMRs that are yet not well characterized and less recognized. This is the case of interactions involving sulfur-containing amino acids (Cys and Met) and in particular between Met/Cys and aromatic and between two Met or Cys residues. The nature of such interactions is primarily dispersive, although it is generally considered to involve a significant electrostatic component as well (29). In fact, the range of geometries observed for these interactions in crystal structures reveals the existence of contributions with different physicochemical origin: S $\cdots$  $\pi$  (in Met-aromatic and Cys-aromatic), C-H $\cdots$  $\pi$  interactions (in Met-aromatic), C-H $\cdots$ S hydrogen bonds (in Met-Met, Cys-aromatic and Met-aromatic), S-H $\cdots$ S hydrogen bonds (in Cys-Cys) and S-H $\cdots$  $\pi$  (in Cys-aromatic) (30-36). It is important to outline that the large polarizability of sulfur enhances also the interactions involving -CH<sub>2</sub>- and -CH<sub>3</sub> groups attached to sulfur compared to aliphatic chains (as for instance in Ile or Leu).

## Results: Sulfur-containing amino acids Part I



	Type of Interaction	Factors Responsible for Interaction	Energy Range (kcal/mol)	Distance dependency on energy	Example
Coulombic	ion-ion	ion charge	20-40	$1/r$	$-\text{NH}_3^+ \cdots \text{OOC}-$
	ion-dipole (H-bond)	ion charge, dipole magnitude	10-25	$1/r^2$	$-\text{NH}_3^+ \cdots \text{O}=\text{C}<$
	dipole-dipole (H-bond)	dipole magnitude, electronegativity	2-7	$1/r^3$	$>\text{C}=\text{O} \cdots \text{HN}<$
van Der Waals	dipole-induced dipole	dipole magnitude, polarizability	0.5-2.5	$1/r^4$	$-\text{OH} \cdots -\text{CH}_3$
	<u>dispersion</u>	polarizability	0.1-3	$1/r^6$	$-\text{CH}_3 \cdots -\text{CH}_3$

**Table 1.** Summary of Intermolecular Forces in Proteins

In the late '70s, Morgan and coworkers observed a high frequency of contacts between sulfur containing residues and aromatic residues in proteins, and even found large stacked arrangements composed of aromatic and Met or Cys residues (37). Further studies also demonstrated that Cys/Met-aromatic interactions were quite common in protein crystal structures and that Met was as likely as Phe or Trp to be near another Trp, with the majority of the interactions facing the ring (30-32). Recently, several attempts to determine binding energies for model systems representing sulfur-aromatic interactions have been performed (33,34,38). From experimental studies of modeled peptides in water, sulfur-aromatic interactions were estimated to contribute up to  $\sim 2$  kcal/mol (39). However, these interactions are probably even stronger when they occur in non-polar environments, as for instance the protein interior. Recent high-level *ab*

## Development and application of computational techniques to drug discovery and structure-function relationships



*initio* calculations gave interaction energies up to ~3-3.5 kcal/mol for side-chain analogs of Met-Phe, Cys-Phe, Cys-Cys and Met-Met interactions (Gómez-Tamayo, J.C., et al., in preparation). This puts these interactions at similar level (and even higher) in terms of strength to other commonly accepted interactions types such as aromatic-aromatic (the energy of interaction between two Phe residues is 2.4 kcal/mol (40-42)).

receptor	organism	PDB id	Met-Aro	Met-Met	Cys-Aro	Cys-Cys	Cys-Met	total <sup>a</sup>
rhodopsin	bovine	1GZM <sup>(2)</sup>	25	2	17	0	1	45
rhodopsin	squid	2Z73 <sup>(4)</sup>	22	7	12	1	2	44
β <sub>2</sub> adrenergic	human	2RH1 <sup>(5)</sup>	9	3	11	0	2	25
β <sub>1</sub> adrenergic	turkey	2VT4 <sup>(7)</sup>	9	2	7	0	2	20
H1 histamine	human	3RZE <sup>(8)</sup>	13	1	11	0	3	28
D3 dopamine	human	3PBL <sup>(9)</sup>	9	1	11	0	0	21
M2 muscarinic	human	3UON <sup>(10)</sup>	9	3	12	1	1	26
M3 muscarinic	rat	4DAJ <sup>(11)</sup>	8	0	10	1	0	19
κOR opioid	human	4DJH <sup>(12)</sup>	18	0	14	0	2	34
μOR opioid	mouse	4DKL <sup>(13)</sup>	18	0	15	0	2	35
δOR opioid	mouse	4EA3 <sup>(14)</sup>	12	0	12	0	3	27
NO/FQ opioid	human	4EJ4 <sup>(15)</sup>	14	0	10	0	1	25
A <sub>2A</sub> adenosine	human	4E1Y <sup>(16)</sup>	12	1	18	0	1	32
CXCR4 chemokine	human	3ODU <sup>(19)</sup>	3	0	10	0	0	13
S1PR1 sphingolipid	human	3V2Y <sup>(20)</sup>	9	1	12	0	0	22
NTSR1 neurotensin	human	4GRV <sup>(21)</sup>	17	1	12	0	2	32
PAR1 protease-activated	human	3VW7 <sup>(22)</sup>	9	0	16	0	0	25
		<b>Total</b>	<b>216</b>	<b>22</b>	<b>210</b>	<b>3</b>	<b>22</b>	<b>473</b>

**Table 2.** List of all interactions involving sulfur containing amino acids in the crystal structures of 7TMRs.

### Interactions involving sulfur-containing amino acids in 7TMR crystal structures

The analysis of the seventeen 7TMR crystal structures corresponding to distinct members of the family available revealed the existence of 216 Met-aromatic, 210 Cys-aromatic, 22 Met-Met, 22 Cys-Met and 3 Cys-Cys (excluding disulfide) interactions (Table 2 and Table 3). This means that, on average, each receptor contains 25 sulfur-aromatic and 3 sulfur-sulfur interactions. The low prevalence of Cys-Cys interactions is due to involvement of Cys in disulfide bridges or

## Results: Sulfur-containing amino acids Part I



hydrogen bond interactions and also to its shorter side-chain (and thus, smaller surface accessibility) compared to Met. Remarkably, it turns out that 47% of the aromatic residues present in 7TMRs are involved in interactions with Met and Cys residues, mostly Met-Phe, Met-Tyr and Cys-Phe pairs. Furthermore, Met/Cys-aromatic and Met/Cys-Met/Cys interactions in 7TMRs are often alternated with aromatic-aromatic interactions, forming large stacked arrangements such as those described by Morgan (37). Although their functional significance is not fully understood in 7TMRs, it is likely that these may constitute molecular machineries stabilizing specific receptor conformations or transmitting structural changes from the extracellular to the cytoplasmic interface, promoting or inhibiting binding to G proteins and/or other signaling proteins.

TM1								
res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>	res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>	
1.34	1.3	M/W	1GZM	1.42	1.38	C/F	2Z73	
	7.4	C/F	3VW7	1.43	2.57	C/Y	3V2Y	
1.35	1.31	M/W	2RH1 2VT4			M/F	4DJH	
	1.39	M/M	2RH1 2VT4		2.58	M/F	4GRV	
1.39	1.35	M/M	2RH1 2VT4		7.4	C/W	3RZE	
	1.38	M/Y	1GZM	1.44	1.4	M/F	1GZM	
	2.57	M/Y	4DJH		1.47	M/F	1GZM	
	2.58		M/F	1GZM	1.47	1.44	M/F	1GZM
			M/Y	3PBL 4GRV		2.51	C/F	2Z73
	7.4		M/F	1GZM	1.54	2.44	M/F	3UON 4DJH 4EA3 4EJ4
			M/W	2RH1 2VT4	1.57	2.43	M/F	3VW7
			C/Y	3VW7		7.53	M/Y	3ODU

Table 3 ....



# Development and application of computational techniques to drug discovery and structure-function relationships



TM2								
res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>	res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>	
2.38	2.42	M/Y	3ODU	2.57	1.39	M/Y	4DJH	
2.39	2.4	M/Y	3V2Y		1.43	C/Y	3V2Y	
	2.42	M/F	3V2Y			M/F	4DJH	
2.41	4.39	M/H	2Z73		3.32	M/Y	3V2Y	
		C/Y	3PBL		7.43	M/Y	4DJH	
2.43	4.43	M/F	2Z73		2.58	1.39	M/F	1GZM
	1.57	M/F	3VW7			2.53	M/F	1GZM
	7.53	M/Y	3VW7			1.39	M/Y	3PBL 4GRV
2.48	2.44	C/F	3UON		1.43	M/F	4GRV	
	3.38	M/F	2Z73		7.4	M/W	3RZE 3UON 4DAJ	
2.53	4.5	C/W	3UON	2.60	7.43	M/Y	3PBL 3RZE 3UON 4DAJ 4GRV	
	2.58	M/F	1GZM					
	3.31	M/F	1GZM					
	6.48	M/W	1GZM 2RH1					
2.62	7.43	M/Y	2RH1	3.24	C/W	3PBL		
	7.46	C/F	3ODU 3VW7	3.28	M/W	3RZE		
				3.28	M/Y	2Z73		
				7.4	M/M	2Z73		
				2.66	1.32	C/Y	2Z73	

Table 3 ....

## Results: Sulfur-containing amino acids Part I



TM3								
res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>	res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>	
	3.23	C/H	4EIY		3.32	M/F	3VW7	
3.22	3.25	C/C	4EIY		3.33	M/Y	3VW7 4DJH 4EJ4	
		C/F	2Z73			M/Y	3VW7	
3.24	2.6	C/W	3PBL		3.37	M/F	4DJH 4EA3 4EJ4	
3.25	2.64	C/Y	3RZE		5.47	M/F	4EA3	
	3.22	C/C	4EIY	3.36	6.44	C/F	3PBL	
		C/F	2Z73			M/F	4EA3	
	3.28	C/W	2RH1 2VT4			C/W	3PBL	
		C/F	3PBL		6.48	M/F	3VW7	
	3.29	C/H	3ODU				M/W	4DJH 4EA3 4EJ4
		C/Y	4GRV			6.52	C/F	3PBL
	5.3	C/C	2RH1 2VT4				M/H	4DJH 4EJ4
3.30	5.37	C/Y	4EIY			2.48	M/F	2Z73
	2.53	M/F	1GZM	3.31	3.38	M/F		
3.31	3.27	M/F	3RZE		4.5	M/W	2Z73	
	4.54	M/F	3RZE		3.44	C/W	3PBL	
3.32	2.57	M/Y	3V2Y		3.45	M/M	2Z73	
	3.36	M/F	3VW7	3.34	3.41	4.48	M/F	2Z73
3.35	M/M	3PBL	4.49			C/F	3V2Y 4EIY	
4.5	M/W	3PBL				4.52	M/W	2Z73
	C/W	3VW7					M/W	1GZM
3.34	4.54	M/F	3PBL		6.44	M/F	2Z73 3VW7	
	4.61	M/Y	4EJ4	3.35	3.43	M/F	3VW7	
3.35	3.34	M/M	3PBL		3.44	3.41	C/W	3PBL

**Development and application of computational techniques to drug discovery and structure-function relationships**



3.38	M/F	2Z73	5.54	M/F	3RZE	
4.5	C/W	2RH1 2VT4 4GRV	2.42	M/F	2Z73	
	M/W	3PBL		M/Y	4DJH 4EJ4	
			3.45	3.41	M/M	2Z73
				4.45	M/M	2Z73
			4.48	M/F		
			3.46	2.42	M/Y	4DJH 4EA3 4EJ4
				7.53	M/Y	4EA3
			3.47	5.57	C/F	3RZE
				5.58	C/Y	3RZE
			3.54	3.51	M/Y	3V2Y
			3.55	3.51	C/Y	1GZM 4EA3 4EJ4 4GRV

Table 3 ....

## Results: Sulfur-containing amino acids Part I



TM4							
res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>	res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>
4.37	2.42	M/F	2Z73 2VT4	4.48	3.41	M/F	2Z73
	4.43	M/M	2Z73		3.45	M/F	2Z73
4.39	2.41	M/H	2Z73	4.49	3.41	C/F	3V2Y 4EIY
		C/Y	3PBL		4.5	C/W	3V2Y
4.43	2.41	M/F	2Z73	4.52	3.41	M/W	1GZM
	4.44	M/M	3UON		4.48	M/F	1GZM
		M/H	1GZM		5.46	M/H	1GZM
	4.45	M/M	3UON		5.38	C/F	1GZM
4.45	2.42	M/F	2Z73	4.56	5.41	C/Y	1GZM
	3.45	M/M	2Z73		5.46	C/H	1GZM
	4.37	M/M	2Z73		4.54	C/F	3PBL
4.44	4.41	M/M	3UON	4.58	4.62	C/F	3PBL
		M/H	1GZM			M/F	2VT4
	4.45	M/M	3UON	4.6	3.29	M/Y	4GRV
4.45	2.42	M/F	3UON 4DAJ	4.61	3.23	M/H	4EIY
	4.41	M/M	3UON		3.27	M/F	4EIY
	4.44	M/M	3UON		3.34	M/Y	4EJ4
4.46	2.42	C/Y	3VW7	4.62	3.23	M/F	2VT4
	2.45	C/H	3VW7		4.58	C/F	3PBL
	4.5	C/W	3VW7			M/F	2VT4
				4.63	5.38	M/F	4EA3
				4.64	5.41	C/W	2Z73

Table 3 ....

**Development and application of computational techniques to drug discovery and structure-function relationships**



TM5								
res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>	res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>	
5.3	2.64	C/H	2RH1	5.43	3.33	C/F	3V2Y	
	3.28	C/W	2RH1 2VT4		5.39	C/Y	3V2Y	
5.35	5.38	M/M	4E1Y		5.42	C/F	3V2Y	
	6.59	M/F	4E1Y		5.42	M/F	1GZM	
	7.29	M/H	4E1Y		5.47	C/F	3V2Y	
5.38	3.33	M/Y	4DJH		6.56	C/F	4E1Y	
	4.56	C/F	1GZM		5.44	5.4	M/F	3ODU
	4.63	M/F	4EA3			5.45	M/F	4GRV
	5.35	M/M	4E1Y			6.56	C/F	4E1Y
		M/F	4DJH		5.45	3.37	M/Y	4GRV
	5.37	M/Y	4E1Y	5.41		C/F	4DJH 4EA3 4EJ4	
		M/F	3RZE	5.44		M/F	4GRV	
	5.41	C/F	4EA3	5.49		M/F	4GRV	
	5.42	M/F	1GZM	5.46	4.52	M/H	1GZM	
	6.52	M/H	4E1Y		4.56	C/H	1GZM	
	4.56	C/W	2Z73		5.41	C/F	4E1Y	
	4.56	C/Y	1GZM	5.51	5.42	M/H	1GZM	
	4.64	C/W	2Z73		5.47	M/F	4GRV	
5.41		M/W	3RZE	5.51	6.44	M/F	4GRV	
		C/F	4DJH		3.44	M/F	3RZE	
		M/F	3RZE	5.54	5.58	M/Y	2RH1 3UON 4E1Y	
		5.38	C/F		4EA3	6.41	M/M	2RH1 2VT4 3RZE
			M/Y		1GZM	6.44	M/F	2RH1 2VT4 3RZE 3UON 4DAJ

## Results: Sulfur-containing amino acids Part I



	5.45	C/F	4DJH 4EA3 4EJ4		3.47	C/F	3RZE
	5.46	C/F	4EIY		5.57	3.51	C/Y
5.42	3.37	M/F	2Z73			C/F	3VW7
	5.38	M/F	1GZM		5.56	C/F	1GZM 2Z73
	5.41	M/Y	1GZM		5.58	C/Y	4DJH
		M/F	1GZM	5.61	3.51	M/Y	4DJH 4EA3 4EJ4
	5.43	C/F	3V2Y		5.58	M/Y	4DJH 4EA3 4EJ4
	5.46	M/H	1GZM		5.63	5.59	M/F

Table 3 ...

TM6							
res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>	res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>
6.27	6.26	M/H	3RZE	6.54	6.51	M/Y	3UON 4DAJ
6.36	6.4	M/M	1GZM		7.3	C/F	3RZE
	7.53	M/Y	1GZM 3PBL 4EJ4		7.34	M/Y	4DAJ
	7.56	M/M	1GZM		7.35	M/W	3UON 4DAJ
		M/F	3PBL		7.38	C/F	4DJH 4EA3 4EJ4
6.4	6.36	M/M	1GZM		7.39	M/Y	3UON
6.41	5.54	M/M	2RH1 2VT4 3RZE	6.56	5.43	C/F	4EIY
	5.58	M/Y	2RH1 3RZE		5.44	C/F	4EIY
	6.42	C/F	3VW7		5.48	M/Y	3RZE
6.42	6.41	C/F	3VW7		6.6	C/F	4EIY
6.47	6.48	C/W	1GZM 3PBL 3V2Y 4DJH 4EA3 4EJ4 4GRV	6.57	6.61	C/F	4EIY
		C/F	3VW7		7.3	M/F	4GRV

## Development and application of computational techniques to drug discovery and structure-function relationships



	7.41	C/F	1GZM		7.31	M/Y	4GRV	
	7.44	C/F	30DU		7.34	M/F	4GRV	
	7.48	C/Y	1GZM		6.59	5.35	M/F	4EIY
		C/F	2RH1			6.58	C/F	4GRV
					6.61	6.6	C/Y	4GRV
						6.57	C/F	4EIY
			6.6	C/F		3UON		
				7.29	C/C	3V2Y		

Table 3 ....

TM7							
res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>	res. <sup>a</sup>	res. <sup>a</sup>	pair	PDB id <sup>b</sup>
7.29	5.35	M/H	4EIY		1.43	C/W	3RZE
	6.61	C/C	3V2Y		2.58	M/W	3RZE 3UON 4DAJ
	7.35	M/H	4EIY		2.62	M/M	2Z73
7.30	6.54	C/F	3RZE		7.36	M/W	3RZE
	6.57	M/F	4GRV	7.42	6.48	C/W	3UON 4DAJ
7.35	6.51	M/Y	1GZM		6.51	C/Y	3UON 4DAJ
	6.54	M/W	3UON 4DAJ	7.39	C/Y	4DAJ	
	7.29	M/H	4EIY	7.46	2.53	C/F	30DU 3VW7
7.36	M/Y	4EIY 4GRV	7.43		C/F	30DU	
7.36	7.33	M/Y	4GRV		7.45	C/H	30DU
	7.35	M/Y	4EIY	7.47	7.44	C/F	30DU
		M/Y	4GRV	7.51	7.48	M/H	2Z73
7.38	6.54	C/F	4DJH 4EA3 4EJ4	7.52	7.53	C/Y	3UON
					7.56	C/C	3UON 4DAJ
7.4	1.34	C/F	3VW7	7.56	6.36	M/M	1GZM
						M/F	3PBL

## Results: Sulfur-containing amino acids Part I



1.38	M/F	2Z73	6.39	C/F	3RZE
	M/F	1GZM		7.52	C/C
1.39	M/W	2RH1 2VT4			
	C/Y	3VW7			

**Table 3.** List of all interactions involving sulfur containing amino acids in the crystal structures of 7TMRs. <sup>a</sup> Interactions were considered for all pairs of residues having at least two side-chain atoms closer than 6.0 Å. Such large distance cutoff is justified by the long-range nature of interactions involving sulfur and aromatic group

One of the most remarkable examples of such a stacked arrangement can be found in the  $\beta_2$  adrenergic receptor ( $\beta_2$ AR). Specifically, a network of interactions extends from Cys116<sup>3,35</sup> to Trp32<sup>1,31</sup> and involves consecutive Met-aromatic, Met/Cys-Met and aromatic-aromatic interactions besides the ligand binding site (Figure 1a). In the core of this network Tyr316<sup>7,43</sup> participates in a Met-aromatic interaction with Met82<sup>2,53</sup> (see below) and also forms a hydrogen bond with Asp113<sup>3,32</sup>, the most critical residue for ligand binding in amine receptors. This suggests that this series of interaction could serve to modulate ligand binding, being sensitive for instance to allosteric modulators (43). In the  $\beta_1$ AR, M82<sup>2,53</sup>V mutation in the analogous residue had a thermo-stabilizing effect and was employed to successfully crystallize the receptor (44). This could be explained by a different conformation of Tyr316<sup>7,43</sup> that would restrain the receptor in an inactive state. Interestingly, crystal structures of histamine H1, dopamine D3, muscarinic M2, muscarinic M3 and  $\kappa$ OR feature Met<sup>2,58</sup> (instead of Met<sup>2,53</sup>) interacting with Tyr<sup>7,43</sup>. Figure 1b shows that residue Met83<sup>2,58</sup> in D3R structure can clearly modulate the conformation of Tyr373<sup>7,43</sup> in the same manner as Met<sup>2,53</sup> in  $\beta_1$ - and  $\beta_2$ -ARs.

The crystal structure of squid rhodopsin contains also a large patch of consecutive interactions involving various Met-Met extending across the membrane along TM4 and reaching intracellular loop (ICL) 2 (Figure 1c). In the same structure, a group of three Met residues connects C-terminal region of the unusually large TM5 and ICL2 featuring a triangle of Met-Met interactions that involves also H230 (Figure 1d).

In the following subsections we will show that interactions involving this type of residues are often present in the available repertoire of crystal structures of



## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---

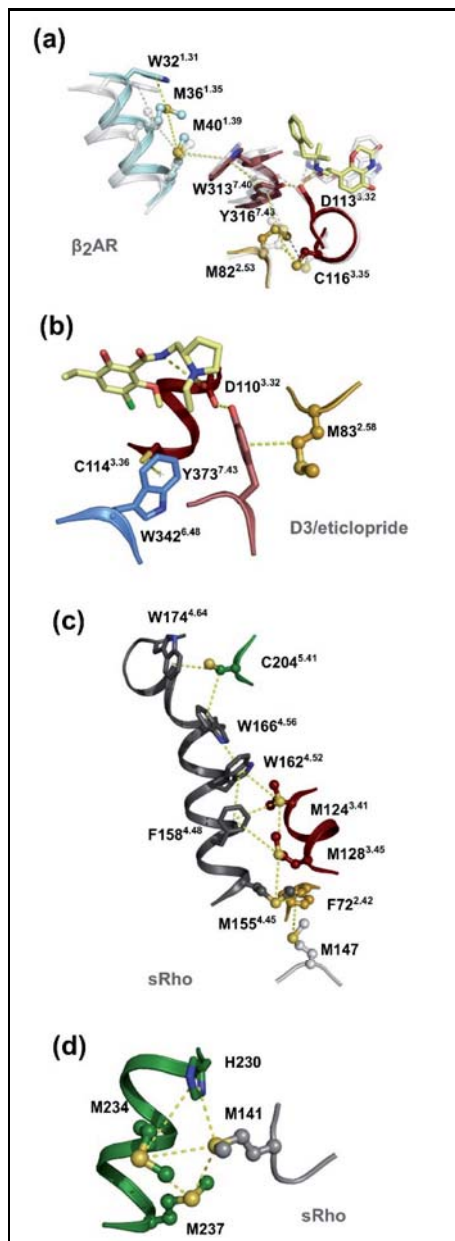


7TMRs and are involved in critical aspects of receptor pharmacology and functioning such as receptor-ligand interactions or activation micro-switches.

### **Met-aromatic interactions relevant for ligand binding**

Various examples of Met-Aromatic interactions are involved in ligand recognition. Figure 2a displays the binding of JD<sup>Tic</sup> to the  $\kappa$ -opioid receptor (OR) as a representative model for OR/ligand complexes. It can be seen that besides Asp138<sup>3.32</sup> (the main anchoring point for opioid ligands), Met142<sup>3.36</sup> forms a Met-aromatic interaction with the ligand that might contribute significantly to binding energy and might equally be important in setting ligand orientation. Interestingly, Met<sup>3.36</sup> is fully conserved in OR of all species and the four structures for this subfamily of receptors available exhibit similar Met-aromatic interactions with their respective ligands (12-15). In fact, it is known from structure/activity relationship studies that the aromatic moiety of peptidic or non-peptidic opioid ligands involved in this interaction is crucial for recognition by cognate receptors (45,46). Sequence analysis of human Class A 7TMRs (performed with the program G<sub>MoS</sub>, available at <http://lmc.uab.cat/gmos>) shows that Met<sup>3.36</sup> is present in 31 additional 7TMRs for peptides and 9 7TMRs for nucleotides. In the human melanocyte-stimulating hormone receptor, Met128<sup>3.36L</sup> substitution significantly decreased agonist potency for an endogenous peptide analog (47). This suggests that the recognition of ligand aromatic residues by Met<sup>3.36</sup> could also be relevant in additional receptors.

## Results: Sulfur-containing amino acids Part I



**Figure 1. Examples of large Met/Cys-aromatic and Met/Cys-Met interactions patches taken from 7TMRs crystal structures.** **a)** A patch of interactions connect Asp113<sup>3.32</sup> (the most important ligand binding element) and the extracellular parts of TMs 1 and 7 in  $\beta_2$ AR bound to carazolol (light-grey; PDB code: 2RH1 (5)) and BI-167107 (colored; PDB code: 3SN6 (6)). **b)** The crystal structure of D3R in complex with eticlopride reveals the presence of two sulfur-aromatic interactions in the vicinity of the ligand binding pocket: Cys114<sup>3.36</sup>-Trp342<sup>6.48</sup> and Met82<sup>2.58</sup>-Y316<sup>7.43</sup> (PDB code: 3PBL (9)). **c)** A patch of interactions in squid rhodopsin propagates along TM4 involving residues from TMs 2, 3 and 5 and from ICL2 (PDB code: 2Z73 (4)). **d)** A cluster of three Met-Met (involving also H230) in the intracellular part of TM5 and ICL2 in squid rhodopsin (PDB code: 2Z73 (4)). Pieces of TM helices and loops are shown as cartoon, relevant residues are shown as sticks and sulfur-containing residues as balls and sticks. The color code for helices and loops is TM1: cyan, TM2: gold, TM3: red, TM5: green, TM6: blue, TM7: pale-red, loops: grey). Ligands are shown in pale-yellow with non-carbon atoms colored by atom type.

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



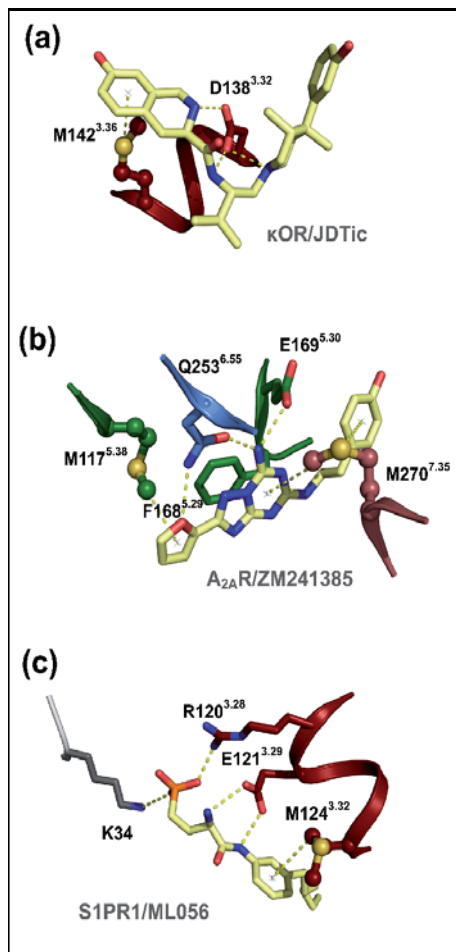
The large repertoire of A<sub>2A</sub> adenosine receptor structures offers also beautiful examples of Met-aromatic interactions within the orthosteric binding sites (17,18). Specifically, Met117<sup>5,38</sup> and Met270<sup>7,35</sup> interact with all crystallized A<sub>2A</sub>R ligands. Figure 2b displays the binding mode for agonist ZM241385 to A<sub>2A</sub>R as a representative case. Interestingly, M270<sup>7,35I</sup> mutation in dog A<sub>1</sub>R (the only A<sub>1</sub>R receptor containing Met at this position) changed binding affinities of specific agonist and antagonists ligands (48). The sequence analysis shows that Met<sup>5,38</sup> is conserved in 85% adenosine receptors and 12% purinoceptors, whereas Met<sup>7,35</sup> is mostly found in A<sub>2</sub>-type receptors (92% conserved), in 13% peptide receptors and 44% rhodopsins. In angiotensin II type 1 receptor, mutations of residue Met284<sup>7,35</sup> to Ala or Cys resulted in significantly impaired binding of antagonists (49,50).

The structure of the sphingosine 1-phosphate receptor 1 (S1PR1) in complex with a sphingolipid mimic (20) provides an additional example of Met-aromatic interaction between receptor and ligand (Figure 2c). Here the ligand central phenyl ring interacts with Met124<sup>3,32</sup>. The sequence analysis shows that this amino acid is also present in most of prostanoid receptors, S1PR3, melatonin MT<sub>1</sub>, and cholecystokinin receptors 1 and 2 (CCK1R and CCK2R). In CCK1R and CCK2R Met3.32 is a critical residue for binding and activation (see below).

### **Met-Met and Met-aromatic interactions in 7TMR activation**

In  $\beta_2$ ARs, comparison between carazolol and BI-167107 bound structures (representing inactive and active-like states, respectively) reveals a rearrangement of TM5-TM6 interface that involves two Met residues (Met215<sup>5,54</sup> and Met279<sup>6,41</sup>) that accompany TM6 opening during activation (5,6). These residues are part of a network of aromatic-aromatic and Met-aromatic interactions connecting residues Tyr219<sup>5,58</sup> and Phe282<sup>6,44</sup>. This serves ultimately to link the functionally relevant Trp286<sup>6,48</sup> and Tyr219<sup>5,58</sup>, producing a conformational change in Tyr219<sup>5,58</sup> (Figure 3a). The importance of interactions involving Met215<sup>5,54</sup> and Met279<sup>6,41</sup> is supported by their simultaneous presence in more than one third of the amine receptor sequences and in more than half of the chemokine receptors.

## Results: Sulfur-containing amino acids Part I



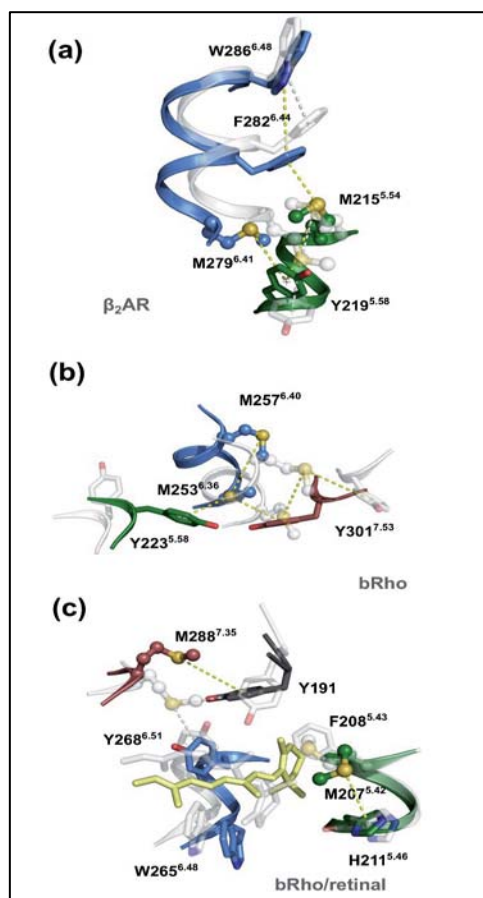
**Figure 2. Met-aromatic interactions between ligand and receptor in 7TMRs crystal structures.** **a)** The four opioid receptors ( $\kappa$ OR/JDTic structure is displayed as representative) exhibit an interaction between Met142<sup>3.36</sup> and an aromatic group in the ligand (PDB code: 4DKL (12)). **b)** Two Met-aromatic interactions involving Met117<sup>5.38</sup> and Met270<sup>7.35</sup> are present in the A<sub>2A</sub>R/ZMA241385 complex (PDB code: 4E1Y (16)). **c)** The crystal structure of S1PR1 in complex with a sphingolipid mimic reveals the presence of an interaction between Met124<sup>3.32</sup> and an aromatic group in the ligand (PDB code: 3V2Y (20)). Protein representations and colors are the same as in Figure 1. The inactive structure is shown in light-grey.

Despite Met<sup>5.54</sup>-Phe<sup>6.44</sup> pair is not present in rhodopsins, comparison between dark-state bovine rhodopsin and metarhodopsin II structures reveals also changes in Met-aromatic and Met-Met interactions between TMs 5 and 6 (Figure 3b) (2,3). In the dark-state structure, Met253<sup>6.36</sup> interacts with Tyr301<sup>7.53</sup> (of the NPxxY motif), whereas Met257<sup>6.40</sup> (located one turn after Met253<sup>6.36</sup>) does not participate in any Met-aromatic interaction. In metarhodopsin II, Tyr301<sup>7.53</sup> has lost the interaction with Met253<sup>6.36</sup> and becomes the partner of Met257<sup>6.40</sup>. In turn, Met257<sup>6.40</sup> interacts also with Tyr223<sup>5.58</sup> and forms a Met-Met interaction with Met253<sup>6.36</sup>. The importance of Met257<sup>6.40</sup> is manifested by the observation that M257<sup>6.40</sup>Y mutation is a constitutively active receptor (51). The recent

## Development and application of computational techniques to drug discovery and structure-function relationships



crystal structure of this mutant receptor shows that the two residues (Met and Tyr) are able to interact with Tyr223<sup>5.58</sup> and Tyr301<sup>7.53</sup> in a similar manner, with only minor changes to the overall structure (52). This triad of residues has been proposed to be responsible for breaking the Arg135<sup>3.50</sup>-Glu247<sup>6.30</sup> ionic lock and for opening up the G-protein binding site (27). The most notable difference is the formation of a hydrogen bond between Tyr257<sup>6.40</sup> and Arg135<sup>3.50</sup> that would be the responsible for displacing the equilibrium from inactive to active state. Interestingly, the sequence analysis shows that Met<sup>6.36</sup> and Met<sup>6.40</sup> are simultaneously present in 85% of vertebrate rhodopsins, a fact that reinforces the functional role of the two amino acids. In addition, Met<sup>6.36</sup> alone is present in 20% human 7TMRs, mostly peptide receptors.



**Figure 3. Met-aromatic and Met-Met interactions changing from inactive to active-like crystal structures.** **a)** Changes in the conformations of Met215<sup>5.54</sup> and Met279<sup>6.41</sup> in  $\beta_2$ AR crystal structures that could modulate the TM5/TM6 interface (PDB codes: 2RH1 (5) and 3SN6 (6)). **b, c)** Changes on Met-Met and Met-aromatic interactions (Met257<sup>6.40</sup> and Met253<sup>6.36</sup>) in rhodopsin structures in the cytoplasmic side of TMs 5-7 (b) and around retinal's binding site (PDB codes: 1GZM (2) and 3PQR (3)). Protein representations and colors are the same as in Figure 1. The inactive structures are shown with light-grey

## **Results: Sulfur-containing amino acids**

### **Part I**



Also in rhodopsin, Met207<sup>5.42</sup> and Met288<sup>7.35</sup> interact with various aromatic residues surrounding retinal's  $\beta$ -ionone ring (Figure 3c). The fact that Met<sup>7.35</sup> is common to A<sub>2A</sub>R and that the side-chain of Met207<sup>5.42</sup> is able to reach the same location as Met<sup>5.38</sup> in A<sub>2A</sub>R suggests an analogous role for these residues in the two receptors (see Figure 2b). In the inactive bovine rhodopsin structure, Met207<sup>5.42</sup> interacts with Phe208<sup>5.43</sup> and Tyr191 in ECL2, whereas Met288<sup>7.35</sup> is surrounded by Tyr268<sup>6.51</sup>, Tyr190 and Tyr191. In metarhodopsin-II, Met207<sup>5.42</sup> establishes a new interaction with His211<sup>5.46</sup> to the detriment of previous interactions with aromatic residues. In parallel, Met288<sup>7.35</sup> loses the interactions with Tyr268<sup>6.51</sup>. These changes are consistent with NMR signals associated to Met288<sup>7.35</sup> and with the suggestion that translation of Phe208<sup>5.43</sup> after retinal's  $\beta$ -ionone ring relocation favors the conformational change at the cytoplasmic side of TM5 that ends in Tyr223<sup>5.58</sup> (27,28). The sequence analysis shows that 53% of the vertebrate rhodopsins have both Met<sup>5.42</sup> and Met<sup>7.35</sup>. As an indication of specific functional role of Met<sup>5.42</sup>, rhodopsins containing M<sup>5.42</sup>L mutation give pigments with blue-shifted  $\lambda_{\max}$  (53).

### **Met/Cys-aromatic interactions in 7TMRs suggested by biophysical studies**

At least two groups have recently employed biophysical techniques to elegantly address the importance of interactions involving Met and Cys. First, fluorination experiments were performed on the indole ring of Trp386<sup>6.48</sup> in the dopamine D2 receptor (D2R), with the aim of characterizing the role of this residue in receptor activation (54). Progressive fluorination of Trp386<sup>6.48</sup> diminished the negative electrostatic potential at the aromatic ring surface. The observed trend regarding activation was indicative of a functionally important interaction at the face of the aromatic ring. In particular, tetrafluorination of Trp386<sup>6.48</sup> resulted in a 300-fold shift in activation potency. Using the crystal structure of D3R in their search for potential partners of Trp<sup>6.48</sup> the authors identified Cys<sup>3.36</sup> as the only residue that could be responsible for the interaction impaired by Trp386<sup>6.48</sup> fluorination. Thus, it was proposed that this Cys-aromatic interaction between Cys118<sup>3.36</sup> and Trp338<sup>6.48</sup> constituted a micro-domain regulating D2R activation (see Figure 1b, corresponding to Cys114<sup>3.36</sup> and Trp342<sup>6.48</sup> in D3R).

As a second example, <sup>1</sup>H-<sup>13</sup>C NMR resonances were utilized in two independent reports to monitor signals from Met residues in the  $\beta_2$ AR receptor (55,56). In

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



these two works, the authors found that receptors occupied by pharmacologically distinct ligands featured changes in chemical shifts and intensities in an efficacy-dependent manner for the resonances corresponding to Met82<sup>2.53</sup>, Met215<sup>5.54</sup> and Met279<sup>6.41</sup>. Furthermore, their signals correlated with the changes on TM helices 5-7 from inactive to active crystal structures (Figure 1), supporting the view that these residues participate in receptor activation. For Met82<sup>2.53</sup> (whose involvement in a large network of interactions has been discussed above, see Figure 1a), comparison between inactive and active crystal structures shows a conformational change where the interaction between Met82<sup>2.53</sup> and Tyr316<sup>7.43</sup> is preferentially enabled in the carazolol bound structures, whereas Met82<sup>2.53</sup>-Cys116<sup>3.35</sup> interaction is enhanced in BI-167107 bound structure (Figure 2a). Interestingly, changes on Met215<sup>5.54</sup> and Met279<sup>6.41</sup> resonances are also compatible with the structural changes in the carazolol and BI-167107 crystal structures discussed in the previous section.

### **Met-aromatic interactions suggested by site-directed mutagenesis and molecular modeling**

Several examples of Met-aromatic interactions in cholecystokinin receptors (CCK1R and CCK2R) that have remarkable pharmacological and functional importance were suggested on the basis of works from our group (23,57-59). Cholecystokinin peptide (CCK) exhibits a posttranslational sulfation of a Tyr residue. This sulfate moiety is essential for biological activity of CCK at the CCK1R, contributing very strongly (500-1000 fold) to CCK binding to CCK1R high affinity state. For receptors that have not yet been crystallized, it is still possible to derive insights from models based on x-ray crystal structures. It is reasonable to expect that these models will be accurate in the TM domains considering the current status of 7TMRs molecular modeling (60). Site-directed mutagenesis studies conducted in synergy with molecular modeling suggested that Tyr-SO<sub>3</sub><sup>-</sup> of CCK likely interacts with Arg197 (ECL2) of CCK1R. In addition, the Tyr side-chain of this Tyr-SO<sub>3</sub><sup>-</sup> of CCK was found to interact with Met195 in ECL2 (Figure 4a) (57). The analysis of Met195-mutated CCK1R indicated that the exchange of Met for Leu caused a minor decrease (3-fold) on the affinity of the high affinity sites for CCK, but a strong drop (75%) on the number of sites, despite the number of low-affinity binding sites remaining unchanged. Thus, Met195 is not important for binding affinity of CCK but rather dramatically influenced the amount of

## **Results: Sulfur-containing amino acids**

### **Part I**



CCK1R which could be converted to a high affinity state for CCK. A plausible interpretation for these data was that this Met-aromatic interaction between Met195 and Tyr-SO<sub>3</sub><sup>-</sup> defined a specific positioning of the sulfate group relative to Arg197, stabilizing ligand-bound CCK1R in a high affinity state. Again in the CCK1R, Met131<sup>3.23</sup> was suggested to interact with the C-terminal aromatic group of agonist peptides since it appeared to be essential for Gq dependent production of the second messengers, inositol phosphates (Figure 4a) (58).

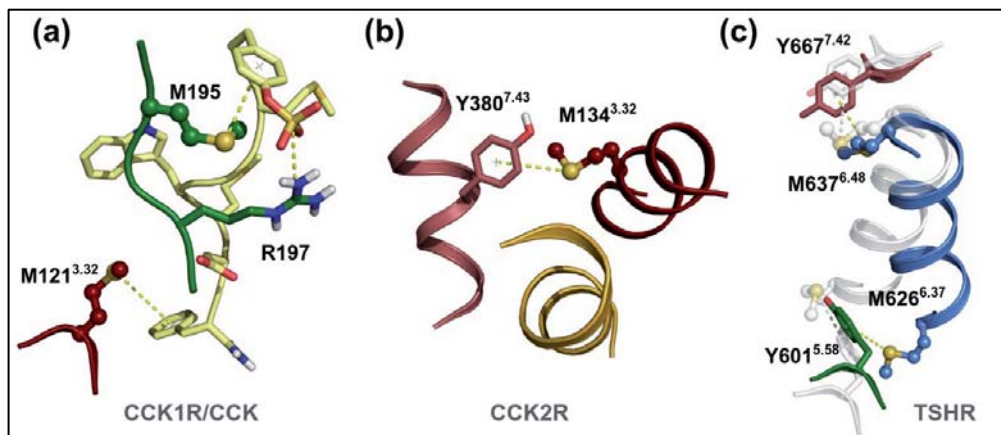
Recently, we have shown in CCK2R that the interaction between Met134<sup>3.32</sup> and Tyr380<sup>7.43</sup> governs the equilibrium between two CCK2R states: either coupling to G<sub>q</sub> or recruiting β-arrestin-2 (Figure 4b) (59). This conclusion was reached by mutating Met134<sup>3.32</sup> and Tyr380<sup>7.43</sup>, which dramatically affected CCK2R efficacy to recruit β-arrestin-2 whereas the mutations did not affect CCK2R efficacy to activate phospholipase-C. The existence of distinct CCK2R conformations associated to G-protein-dependent or β-arrestin-2-dependent signaling pathways was further pharmacologically demonstrated using a biased competitive antagonist which inhibited G-protein-dependent signaling but not β-arrestin-2-dependent pathways.

As a third example, the thyroid-stimulating hormone receptor (TSHR) exhibits elevated cAMP signaling in the basal state and becomes fully activated by thyrotropin. Two independent studies reported Met mutations generating constitutive active receptors (61,62). First, M626<sup>6.37</sup>I produced basal cAMP levels 13-fold higher than wild-type TSHR (61). TSHR structural models suggest that Met626<sup>6.37</sup> in TSHR likely participates in Met-aromatic interactions that switch from the inactive to the active state (Figure 4c). More precisely, Met626<sup>6.37</sup> would interact with one face of Tyr601<sup>5.58</sup> in the inactive state, whereas upon activation (and the outward movement of TM6), it would interact with the opposite face of Tyr601<sup>5.58</sup>. Thus, M626<sup>6.37</sup>I mutant would hinder these interactions. The second constitutively active mutant was M637<sup>6.48</sup>W (62). All glycoprotein hormone receptors contain Met<sup>6.48</sup> instead of the conserved Trp present in 80% 7TMRs. This mutant also displayed atypical pharmacology in that thyrotropin activated it with the same potency as the wild-type, whereas the small non-peptide NCGC00161870-01 agonist activated this mutant with a 14-fold decreased potency. The authors proposed that M637<sup>6.48</sup>W was important for stabilizing the active state by means of Met-aromatic interactions between residue Tyr667<sup>7.42</sup> and an aromatic moiety in the ligand. Sequence analysis indicates that



## Development and application of computational techniques to drug discovery and structure-function relationships

Met637<sup>6.48</sup>/Tyr667<sup>7.42</sup> pair is conserved in glycoprotein hormone receptors, suggesting that this interaction could have functional importance.



**Figure 4. Molecular models for Met-aromatic interactions evidenced from site-directed mutagenesis.** a) CCK binding to CCK1R involves two Met-aromatic interactions between Phe/Tyr residues of CCK and two Met residues Met195 (in ECL2) and Met121<sup>3.32</sup>. A ionic interaction is formed between Arg197 and the sulfate group of CCK's Tyr residue, whose aromatic side-chain participates in the aromatic-sulphur interaction with Met195 (model CCK1R\_1, adapted from (57,58); see Suppl. Data). b) The interaction Met134<sup>3.32</sup> and Tyr380<sup>7.43</sup> in CCK2R is required for receptor signaling through  $\beta$ -arrestin2 recruitment but not through G-protein coupling (model CCK2R\_1, adapted from (59); see Suppl. Data). c) Comparison between inactive (light-grey transparency) and active-like (colored) TSHR models (TSHR\_1 and TSHR\_2; see Suppl. Data) M648W and M637I are two constitutively active mutations whose behavior could be due to sulfur aromatic interactions. Protein representations and colors are the same as in Figure 1. The inactive structure is shown in light-grey.

### Polymorphisms associated with introduction/disappearance of Met residues

Various polymorphisms in 7TMRs that affect or introduce Met residues have been described. Because of the specific interactions between Met and aromatic residues, these polymorphisms are likely to modify receptor dynamics and thus raise to significant changes in phenotype. T280<sup>7.40</sup>M mutation in the CX3CR1 chemokine receptor constitutes a first example. Patients having this mutation who were infected with HIV-1 virus showed a rapid progression to AIDS (63). A homology model of CX3CR1 (Figure 5A) suggests that Met280<sup>7.40</sup> would exhibit

## **Results: Sulfur-containing amino acids**

### **Part I**



two Met-aromatic interactions with Phe371<sup>38</sup> and Tyr381<sup>39</sup> that may stabilize a specific receptor conformation promoting activation.

In TSHR, M453<sup>2.43</sup>T and M463<sup>2.53</sup>V are two activating mutations causing non-autoimmune hyperthyroidism (64,65). Homology models of the inactive TSHR (Figure 5B) show that Met453<sup>2.43</sup>, located at the cytoplasmic side of TM2, interacts with the conserved Tyr678<sup>7.53</sup> and Phe685<sup>8.50</sup>. Thus, it is likely that the M453<sup>2.43</sup>V mutation affects the equilibrium between inactive and active states, leading to constitutive activity. By contrast, Met463<sup>2.53</sup>, located at the region of closest contact between TMs 2, 3 and 7 could participate in a Met-aromatic interaction with Tyr667<sup>7.42</sup>. This interaction could play a similar role as the Met134<sup>3.32</sup>-Tyr380<sup>7.43</sup> interaction in CCK2R, which was found to govern G-protein-dependent signaling but not  $\beta$ -arrestin2-dependent pathways.

### **Concluding remarks**

7TMRs feature several examples of Met/Cys-aromatic and Met-Met interactions that are, in some cases, responsible for important pharmacological, signaling or functional events. In the present review we have described many examples of such interactions occurring between natural or synthetic ligands and their receptors, and between two or more residues within the receptor (see Figures 1-4 and Table 4). In ligand-protein complexes, interactions involving Met and Cys could, for instance, provide additional binding energy and also modulate ligand orientation within the binding site. Within the receptor, we have shown that these interactions might help to stabilize specific conformations. In fact, for receptors for which an active crystal structure is available, comparison between inactive and active structures shows changes associated to formation/breakage of interactions involving Met residues. Met residues can easily change side-chain conformation within a packed environment and, thus, offer extra versatility to respond to structural changes compared to aromatic rings (66). In addition, the intrinsic flexibility of Met might enable optimization towards energetically favorable Met-Met or Met-aromatic arrangements. These interactions are more directional than the interactions between aliphatic chains, a fact that makes them very useful to keep aromatic residues in well-defined orientations. In terms of strength, Met/Cys-aromatic and Met-Met interactions are stronger than other already widely-accepted interactions such as those between two aromatic residues.

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



Analysis of crystal structures for the 17 members of the 7TMR family has identified a large number of such interactions. Because of the relatively low prevalence of positions containing Cys or Met in 7TMRs, the interactions described in this report are highly subfamily-specific, suggesting that Met-Met and Met/Cys-aromatic interaction are often employed for fine-tuning receptor function within subfamilies. The Met<sup>5.54</sup>-Phe<sup>6.44</sup> interaction is a remarkable exception, being present in up to 25% of human Class A 7TMRs. Thus, this constitutes the most representative example of a Met/Cys-aromatic interaction associated to residues with known functional implication in 7TMRs. Also, involvement of such interactions in ligand binding to 7TMRs, suggests that exploiting this relatively unused Met-aromatic interaction may be an original way to design new ligands that may have therapeutic interest. Finally, the recent discovery of a Met-aromatic interaction in the CCK2R that was crucial for selective stabilization of protein state associated with  $\beta$ -arrestin-2 recruitment could be anticipating that interactions of moderate strength such as Met/Cys-aromatic and Met-Met may play a role in selectively targeting specific signaling pathways, such as those preferentially triggering G-protein-dependent or  $\beta$ -arrestin-dependent signaling pathways.

## Results: Sulfur-containing amino acids Part I



### References

1. Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006) How many drug targets are there? *Nat Rev Drug Discov* **5**, 993-996
2. Li, J., Edwards, P. C., Burghammer, M., Villa, C., and Schertler, G. F. X. (2004) Structure of bovine rhodopsin in a trigonal crystal form. *J Mol Biol* **343**, 1409-1438
3. Choe, H. W., Kim, Y. J., Park, J. H., Morizumi, T., Pai, E. F., Krauss, N., Hofmann, K. P., Scheerer, P., and Ernst, O. P. (2011) Crystal structure of metarhodopsin II. *Nature* **471**, 651-655
4. Murakami, M., and Kouyama, T. (2008) Crystal structure of squid rhodopsin. *Nature* **453**, 363-367
5. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K., and Stevens, R. C. (2007) High-Resolution Crystal Structure of an Engineered Human beta2-Adrenergic G Protein Coupled Receptor. *Science* **318**, 1258-1265
6. Rasmussen, S. G., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K., and Kobilka, B. K. (2011) Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549-555
7. Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G., Tate, C. G., and Schertler, G. F. (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486-491
8. Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V., Abagyan, R., Cherezov, V., Liu, W., Han, G. W., Kobayashi, T., Stevens, R. C., and Iwata, S. (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* **475**, 65-70
9. Chien, E. Y., Liu, W., Zhao, Q., Katritch, V., Han, G. W., Hanson, M. A., Shi, L., Newman, A. H., Javitch, J. A., Cherezov, V., and Stevens, R. C. (2011) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330**, 1091-1095
10. Haga, K., Kruse, A. C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C., Weis, W. I., Okada, T., Kobilka, B. K., Haga, T., and Kobayashi, T. (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **482**, 547-551
11. Kruse, A. C., Hu, J., Pan, A. C., Arlow, D. H., Rosenbaum, D. M., Rosemond, E., Green, H. F., Liu, T., Chae, P. S., Dror, R. O., Shaw, D. E., Weis, W. I., Wess, J., and Kobilka, B.

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



- K. (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **482**, 552-556
12. Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G. W., Vardy, E., Liu, W., Thompson, A. A., Huang, X. P., Carroll, F. I., Mascarella, S. W., Westkaemper, R. B., Mosier, P. D., Roth, B. L., Cherezov, V., and Stevens, R. C. (2012) Structure of the human kappa-opioid receptor in complex with JDTic. *Nature* **485**, 327-332
13. Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., Pardo, L., Weis, W. I., Kobilka, B. K., and Granier, S. (2012) Crystal structure of the  $\mu$ -opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321-326
14. Granier, S., Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Weis, W. I., and Kobilka, B. K. (2012) Structure of the delta-opioid receptor bound to naltrindole. *Nature* **485**, 400-404
15. Thompson, A. A., Liu, W., Chun, E., Katritch, V., Wu, H., Vardy, E., Huang, X. P., Trapella, C., Guerrini, R., Calo, G., Roth, B. L., Cherezov, V., and Stevens, R. C. (2012) Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* **485**, 395-399
16. Liu, W., Chun, E., Thompson, A. A., Chubukov, P., Xu, F., Katritch, V., Han, G. W., Roth, C. B., Heitman, L. H., AP, I. J., Cherezov, V., and Stevens, R. C. (2012) Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **337**, 232-236
17. Lebon, G., Warne, T., Edwards, P. C., Bennett, K., Langmead, C. J., Leslie, A. G., and Tate, C. G. (2011) Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **474**, 521-525
18. Xu, F., Wu, H., Katritch, V., Han, G. W., Jacobson, K. A., Gao, Z. G., Cherezov, V., and Stevens, R. C. (2011) Structure of an agonist-bound human A2A adenosine receptor. *Science* **332**, 322-327
19. Wu, B., Chien, E. Y. T., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., Hamel, D. J., Kuhn, P., Handel, T. M., Cherezov, V., and Stevens, R. C. (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330**, 1066-1071
20. Hanson, M. A., Roth, C. B., Jo, E., Griffith, M. T., Scott, F. L., Reinhart, G., Desale, H., Clemons, B., Cahalan, S. M., Schuerer, S. C., Sanna, M. G., Han, G. W., Kuhn, P., Rosen, H., and Stevens, R. C. (2012) Crystal structure of a lipid G protein-coupled receptor. *Science* **335**, 851-855
21. White, J. F., Noinaj, N., Shibata, Y., Love, J., Kloss, B., Xu, F., Gvozdenovic-Jeremic, J., Shah, P., Shiloach, J., Tate, C. G., and Grishammer, R. (2012) Structure of the agonist-bound neurotensin receptor. *Nature* **490**, 508-513
22. Zhang, C., Srinivasan, Y., Arlow, D. H., Fung, J. J., Palmer, D., Zheng, Y., Green, H. F., Pandey, A., Dror, R. O., Shaw, D. E., Weis, W. I., Coughlin, S. R., and Kobilka, B. K.

## Results: Sulfur-containing amino acids Part I



- (2012) High-resolution crystal structure of human protease-activated receptor 1. *Nature* **492**, 387-392
23. Ballesteros, J. A., and Weinstein, H. (1995) Integrated Methods for Modeling G-Protein Coupled Receptors. *Methods Neurosci.* **25**, 366-428
24. Audet, M., and Bouvier, M. (2012) Restructuring g-protein- coupled receptor activation. *Cell* **151**, 14-23
25. Schwartz, T. W., Frimurer, T. M., Holst, B., Rosenkilde, M. M., and Elling, C. E. (2006) Molecular mechanism of 7TM receptor activation - A global toggle switch model. *Annu Rev Pharmacol Toxicol* **46**, 481-519
26. Liapakis, G., Cordini, A., and Pardo, L. (2012) The G-protein coupled receptor family; actors with many faces. *Curr Pharm Des.* **18**, 175-185
27. Goncalves, J. A., Ahuja, S., Erfani, S., Eilers, M., and Smith, S. O. (2010) Structure and function of G protein-coupled receptors using NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **57**, 159-180
28. Deupi, X., Standfuss, J., and Schertler, G. (2012) Conserved activation pathways in G-protein-coupled receptors. *Biochem Soc Trans* **40**, 383-388
29. Tauer, T. P., Derrick, M. E., and Sherrill, C. D. (2005) Estimates of the ab initio limit for sulfur-pi interactions: the H<sub>2</sub>S-benzene dimer. *J. Phys. Chem. A* **109**, 191-196
30. Zauhar, R. J., Colbert, C. L., Morgan, R. S., and Welsh, W. J. (2000) Evidence for a strong sulfur-aromatic interaction derived from crystallographic data. *Biopolymers* **53**, 233-248
31. Pal, D., and Chakrabarti, P. (2001) Non-hydrogen bond interactions involving the methionine sulfur atom. *J Biomol Struct Dyn* **19**, 115-128
32. Samanta, U., Pal, D., and Chakrabarti, P. (2000) Environment of tryptophan side chains in proteins. *Proteins* **38**, 288-300
33. Duan, G. L., Smith, V. H., and Weaver, D. F. (2001) Characterization of aromatic-thiol pi-type hydrogen bonding and phenylalanine-cysteine side chain interactions through ab initio calculations and protein database analyses. *Mol. Phys.* **99**, 1689-1699
34. Ringer, A. L., Senenko, A., and Sherrill, C. D. (2007) Models of S/pi interactions in protein structures: comparison of the H<sub>2</sub>S benzene complex with PDB data. *Protein Sci* **16**, 2216-2223
35. Cabaleiro-Lago, E. M., Hermida-Ramon, J. M., and Rodriguez-Otero, J. (2004) Computational study of the interaction in (CH<sub>3</sub>)<sub>2</sub>X dimer and trimer (X = O, S). *J. Phys. Chem. A* **108**, 4923-4929
36. Ringer, A. L., Figgs, M. S., Sinnokrot, M. O., and Sherrill, C. D. (2006) Aliphatic C-H/pi interactions: Methane-benzene, methane-phenol, and methane-indole complexes. *J. Phys. Chem. A* **110**, 10822-10828

## Development and application of computational techniques to drug discovery and structure-function relationships



37. Morgan, R. S., Tatsch, C. E., Gushard, R. H., McAdon, J., and Warne, P. K. (1978) Chains of alternating sulfur and pi-bonded atoms in eight small proteins. *Int. J. Pept. Protein Res.* **11**, 209-217
38. Pranata, J. (1997) Sulfur aromatic interactions: A computational study of the dimethyl sulfide benzene complex. *Bioorg. Chem.* **25**, 213-219
39. Viguera, A. R., and Serrano, L. (1995) Side-chain interactions between sulfur-containing amino acids and phenylalanine in alpha-helices. *Biochemistry* **34**, 8771-8779
40. Sinnokrot, M. O., and Sherrill, C. D. (2006) High-accuracy quantum mechanical studies of pi-pi interactions in benzene dimers. *J. Phys. Chem. A* **110**, 10656-10668
41. Janowski, T., and Pulay, P. (2007) High accuracy benchmark calculations on the benzene dimer potential energy surface. *Chem. Phys. Lett.* **447**, 27-32
42. Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M., and Tanabe, K. (2002) Origin of attraction and directionality of the pi/pi interaction: model chemistry calculations of benzene dimer interaction. *J Am Chem Soc* **124**, 104-112
43. Bokoch, M. P., Zou, Y., Rasmussen, S. G., Liu, C. W., Nygaard, R., Rosenbaum, D. M., Fung, J. J., Choi, H. J., Thian, F. S., Kobilka, T. S., Puglisi, J. D., Weis, W. I., Pardo, L., Prosser, R. S., Mueller, L., and Kobilka, B. K. (2010) Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* **463**, 108-112
44. Warne, T., Serrano-Vega, M. J., Tate, C. G., and Schertler, G. F. (2009) Development and crystallization of a minimal thermostabilized G protein-coupled receptor. *Protein Expr. Purif.*
45. Guerrini, R., Calo, G., Rizzi, A., Bigoni, R., Rizzi, D., Regoli, D., and Salvadori, S. (2000) Structure-activity relationships of nociceptin and related peptides: comparison with dynorphin A. *Peptides* **21**, 923-933
46. Eguchi, M. (2004) Recent advances in selective opioid receptor agonists and antagonists. *Med. Res. Rev.* **24**, 182-212
47. Yang, Y., Hruby, V. J., Chen, M., Crasto, C., Cai, M., and Harmon, C. M. (2009) Novel binding motif of ACTH analogues at the melanocortin receptors. *Biochemistry* **48**, 9775-9784
48. Tucker, A. L., Robeva, A. S., Taylor, H. E., Holeton, D., Bockner, M., Lynch, K. R., and Linden, J. (1994) A1 adenosine receptors. Two amino acids are responsible for species differences in ligand recognition. *J Biol Chem* **269**, 27900-27906
49. Perlman, S., Costa-Neto, C. M., Miyakawa, A. A., Schambye, H. T., Hjorth, S. A., Paiva, A. C., Rivero, R. A., Greenlee, W. J., and Schwartz, T. W. (1997) Dual agonistic and antagonistic property of nonpeptide angiotensin AT1 ligands: susceptibility to receptor mutations. *Mol Pharmacol* **51**, 301-311
50. Boucard, A. A., Roy, M., Beaulieu, M. E., Lavigne, P., Escher, E., Guillemette, G., and Leduc, R. (2003) Constitutive activation of the angiotensin II type 1 receptor alters

## Results: Sulfur-containing amino acids Part I



- the spatial proximity of transmembrane 7 to the ligand-binding pocket. *J Biol Chem* **278**, 36628-36636
51. Han, M., Smith, S. O., and Sakmar, T. P. (1998) Constitutive activation of opsin by mutation of methionine 257 on transmembrane helix 6. *Biochemistry* **37**, 8253-8261
  52. Deupi, X., Edwards, P., Singhal, A., Nickle, B., Oprian, D., Schertler, G., and Standfuss, J. (2012) Stabilized G protein binding site in the structure of constitutively active metarhodopsin-II. *Proc Natl Acad Sci U S A* **109**, 119-124
  53. Yokoyama, S., Zhang, H., Radlwimmer, F. B., and Blow, N. S. (1999) Adaptive evolution of color vision of the Comoran coelacanth (*Latimeria chalumnae*). *Proc Natl Acad Sci U S A* **96**, 6279-6284
  54. Daeffler, K. N., Lester, H. A., and Dougherty, D. A. (2012) Functionally Important Aromatic-Aromatic and Sulfur-pi Interactions in the D2 Dopamine Receptor. *J Am Chem Soc* **134**, 14890-14896
  55. Kofuku, Y., Ueda, T., Okude, J., Shiraishi, Y., Kondo, K., Maeda, M., Tsujishita, H., and Shimada, I. (2012) Efficacy of the beta(2)-adrenergic receptor is determined by conformational equilibrium in the transmembrane region. *Nature communications* **3**, 1045
  56. Nygaard, R., Zou, Y., Dror, R. O., Mildorf, T. J., Arlow, D. H., Manglik, A., Pan, A. C., Liu, C. W., Fung, J. J., Bokoch, M. P., Thian, F. S., Kobilka, T. S., Shaw, D. E., Mueller, L., Prosser, R. S., and Kobilka, B. K. (2013) The Dynamic Process of beta(2)-Adrenergic Receptor Activation. *Cell* **152**, 532-542
  57. Gigoux, V., Escrieut, C., Silvente-Poirot, S., Maigret, B., Gouilleux, L., Fehrentz, J. A., Gully, D., Moroder, L., Vaysse, N., and Fourmy, D. (1998) Met-195 of the cholecystokinin-A receptor interacts with the sulfated tyrosine of cholecystokinin and is crucial for receptor transition to high affinity state. *J Biol Chem* **273**, 14380-14386
  58. Escrieut, C., Gigoux, V., Archer, E., Verrier, S., Maigret, B., Behrendt, R., Moroder, L., Bignon, E., Silvente-Poirot, S., Pradayrol, L., and Fourmy, D. (2002) The biologically crucial C terminus of cholecystokinin and the non-peptide agonist SR-146,131 share a common binding site in the human CCK1 receptor. Evidence for a crucial role of Met-121 in the activation process. *J Biol Chem* **277**, 7546-7555
  59. Magnan, R., Escrieut, C., Gigoux, V., De, K., Clerc, P., Niu, F., Azema, J., Masri, B., Cordomi, A., Baltas, M., Tikhonova, I. G., and Fourmy, D. (2013) Distinct CCK-2 receptor conformations associated with beta-arrestin-2 recruitment or phospholipase-C activation revealed by a biased antagonist. *J Am Chem Soc* **135**, 2560-2573
  60. Kufareva, I., Rueda, M., Katritch, V., Stevens, R. C., Abagyan, R., and participants, G. D. (2011) Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* **19**, 1108-1126



## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



61. Ringkunanont, U., Van Durme, J., Montanelli, L., Ugrasbul, F., Yu, Y. M., Weiss, R. E., Refetoff, S., and Grasberger, H. (2006) Repulsive separation of the cytoplasmic ends of transmembrane helices 3 and 6 is linked to receptor activation in a novel thyrotropin receptor mutant (M626I). *Mol Endocrinol* **20**, 893-903
62. Kleinau, G., Hoyer, I., Kreuchwig, A., Haas, A. K., Rutz, C., Furkert, J., Worth, C. L., Krause, G., and Schulein, R. (2011) From molecular details of the interplay between transmembrane helices of the thyrotropin receptor to general aspects of signal transduction in family a G-protein-coupled receptors (GPCRs). *J Biol Chem* **286**, 25859-25871
63. Becker, Y. (2007) The spreading of HIV-1 infection in the human organism is caused by fractalkine trafficking of the infected lymphocytes--a review, hypothesis and implications for treatment. *Virus genes* **34**, 93-109
64. Ferrara, A. M., Capalbo, D., Rossi, G., Capuano, S., Del Prete, G., Esposito, V., Montesano, G., Zampella, E., Fenzi, G., Salerno, M., and Macchia, P. E. (2007) A new case of familial nonautoimmune hyperthyroidism caused by the M463V mutation in the TSH receptor with anticipation of the disease across generations: a possible role of iodine supplementation. *Thyroid : official journal of the American Thyroid Association* **17**, 677-680
65. Kraemer, S., Rothe, K., Pfaeffle, R., Fuehrer-Sakel, D., Till, H., and Muensterer, O. J. (2009) Activating TSH-receptor mutation (Met453Thr) as a cause of adenomatous non-autoimmune hyperthyroidism in a 3-year-old boy. *J. Pediatr. Endocrinol. Metab.* **22**, 269-274
66. Gellman, S. H. (1991) On the role of methionine residues in the sequence-independent recognition of nonpolar protein surfaces. *Biochemistry* **30**, 6633-6636



## **4.4 Sulfur-containing amino acids**

### **PART II**

#### **Analysis of the interactions of sulfur-containing amino acids in membrane proteins**

##### **Abstract**

The interactions of Met and Cys with other amino acid side chains have received little attention, in contrast to aromatic-aromatic, aromatic-aliphatic or/and aliphatic-aliphatic interactions. Precisely, these are the only amino acids that contain a sulfur atom, which is highly polarizable and, thus, likely to participate in strong Van der Waals interactions. Analysis of the interactions present in membrane protein crystal structures, together with the characterization of their strength in small-molecule model systems at the *ab-initio* level, predicts that Met-Met interactions are stronger than Met-Cys  $\approx$  Met-Phe  $\approx$  Cys-Phe interactions, stronger than Phe-Phe  $\approx$  Phe-Leu interactions, stronger than the Met-Leu interaction, and stronger than Leu-Leu  $\approx$  Cys-Leu interactions. These results show that sulfur-containing amino acids form stronger interactions than aromatic or aliphatic amino acids. Thus, these amino acids may provide additional driving forces for maintaining the 3D structure of membrane proteins and may provide functional specificity.

##### **Introduction**

Non-bonded interactions are crucial for protein stability, function and ligand binding. These comprise electrostatic (including hydrogen bonds) and van der Waals (dipole-dipole, dipole-induced dipole and induced dipole-induced dipole) interactions (1). All these type of interactions have been extensively characterized in terms of strength, directionality, and physicochemical properties(2). However, their prevalence and importance vary depending on whether or not they occur in membrane or globular proteins due to their different environment. Both globular and membrane proteins position hydrophobic amino acid side chains toward the protein core, and maximize hydrogen bond interactions among backbone atoms. However, in contrast to

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



soluble proteins, the hydrophobic nature of the lipid bilayer imposes that residues pointing towards the membrane are also hydrophobic. Thus, dispersion forces (mainly aromatic-aromatic, aromatic-aliphatic or aliphatic-aliphatic) are involved in stabilizing the tertiary structure of the protein or in structural changes (2-7).

As polarizabilities of the two interacting partners become larger, van der Waals forces become stronger. For example, the aromatic ring of aromatic amino acids has a quadrupole  $\pi$  system that is highly polarizable, providing strong aromatic-aromatic dispersion interactions. Thus, aromatic side chains importantly contribute to the folding and thermodynamic stability of proteins (8). Similarly, sulfur-containing amino acids are also highly polarizable, as sulfur has filled 3p and empty 3d orbitals and contain a permanent dipole (9). Surprisingly, non-bonded interactions (dipole-induced dipole or dispersion) involving sulfur-containing amino acids (Met and Cys) have received little attention (10-12) in contrast to interactions involving aromatic amino acids (2). More than 30 years ago, Morgan and coworkers observed a high frequency of contacts between sulfur-containing residues and aromatic residues in proteins, and identified large stacked arrangements composed of aromatic and Met or Cys residues (13). Further studies also demonstrated that Cys- and Met-aromatic interactions were fairly common in protein crystal structures (12,14,15).

In the present work we aim to evaluate the occurrence of interactions involving Met and Cys side-chains in crystal structures of membrane proteins and to characterize their strength in small-molecule model systems at the *ab-initio* level. The employed level of theory improves previous calculations in analogous systems.(16-19) Our results show that Met-Met, Met-Phe, Met-Leu and Cys-Phe interactions are stronger in magnitude than Phe-Phe interactions.

### **Material and methods**

#### ***Analysis of membrane protein structures***

A non-redundant dataset of 327  $\alpha$ -helical transmembrane bundles were taken from TMalphaDB (20). This data set consists of crystallographic structures deposited in the Protein Data Bank (21) with resolution  $<3.5$  Å. Residues were

## Results: Sulfur-containing amino acids Part II



classified, based on their circular variance (22) (CV) of vectors drawn from the C $\alpha$  atom of a given residue to the C $\alpha$  atoms of neighbor residues, as exposed (CV > 0.7) or buried (CV  $\leq$  0.7) to the membrane. Met-Met, Met-Phe, Met-Leu, Cys-Met, Cys-Phe, and Cys-Leu interactions were considered if the distance  $d$  between the two side-chains (measured as the distance between the atoms S $\delta$  of Met, S $\gamma$  of Cys or C $\gamma$  of Leu or the centroid of the aromatic ring of Phe) was < 6Å. At longer distances the interaction energy becomes negligible (see Supplementary Figure S1). The relative orientation of the interacting side-chains was defined by the distance  $d$ , the angle  $P$  between side-chain planes (each plane defined by atoms C $\gamma$ , S $\delta$  and C $\epsilon$  of Met; C $\alpha$ , C $\beta$  and S $\gamma$  of Cys; C $\delta_1$ , C $\gamma$  and C $\delta_2$  of Leu; and the aromatic ring of Phe), and the angle  $\theta$  between the plane defined by side-chain A and the vector connecting the central atoms of each side-chain (A and B). Definitions of  $P$  and  $\theta$  angles were those used by Chakrabarti et al. to describe benzene dimer geometries (23). These interactions were clustered according to the conformational space defined by the distance  $d$  and the angles  $P$  and  $\theta$  (see Supplementary Tables S1-S3 for a detailed description).

### **Quantum mechanical calculations**

For the representative structure of each cluster, the energy of interaction between side chains was calculated using *ab initio* methods on small-molecule models systems: Met was mimicked by dimethyl sulfide (DMS), Cys by methanethiol (MT), Leu by propane (PRP), and Phe by benzene (BNZ). All chosen model structures were optimized at the MP2/6-31+G(d,p) level of theory, which has been shown to provide reasonably good geometries (24,25). Next, single point energy calculations were performed at the CCSD(T)/6-311+G(3df,2p) level. In order to minimize the basis set superposition error, counterpoise method by Boys and Bernardi (26) was utilized. Moreover, in order to evaluate the ability of the molecular mechanic AMBER force field to properly represent these interactions, AMBER energies were calculated for optimized geometries (Supplementary Figures S2-S6). All calculations were performed using GAUSSIAN09 program (27).



## Results and discussion

### *Structural bioinformatics analysis of the presence of Cys and Met in membrane proteins*

Table 1 summarizes the occurrence of the most frequent amino acids, together with Cys and Met, in the transmembrane region (i.e. excluding water-soluble domains or loops) of  $\alpha$ -helix bundles of membrane proteins with known crystal structure (see Methods). Amino acids such as Leu, Ile, Val and Phe are the most frequent at the membrane-embedded region, without preference for being localized in the protein core or in the membrane-exposed region. In contrast, sulfur-containing amino acids (i.e. Met and Cys) show lower frequencies and are mostly found buried in the core of the protein. Analysis of inter-residue interaction of Met and Cys reveals that a significant percentage of Met residues form interactions with aliphatic residues (Leu, Ala, Ile, Val) and Phe, while a smaller proportion interact with Met and Cys (Table 1). Thus, we aim to determine the orientation and strength of these interactions of sulfur-containing amino acids.

	Amino acid distribution			Side chain-side chain interactions						
	Total	Buried	Exposed	Cys	Met	Phe	Val	Ile	Ala	Leu
<b>Leu</b>	8.896 (17%)	3.787 (43%)	5.111 (57%)	474	1.129	2.415	3.427	2.491	3.764	2.319
<b>Ala</b>	6.198 (12%)	3.988 (64%)	2.209 (36%)	396	795	1.616	3.437	2.42	2.039	
<b>Ile</b>	5.801 (11%)	2.493 (43%)	3.309 (57%)	401	806	1.671	2.126	723		
<b>Val</b>	5.761 (11%)	2.826 (49%)	2.935 (51%)	357	700	1.276	1.363			
<b>Phe</b>	4.651 (9%)	2.464 (53%)	2.188 (47%)	276	707	859				
<b>Met</b>	1.933 (4%)	1.336 (69%)	597 (31%)	111	199					
<b>Cys</b>	768 (1.4%)	557 (73%)	211 (27%)	37						

**Table 1.** Structural bioinformatics analysis of membrane proteins. Amino acid type distribution (absolute frequencies and relative frequencies in percentage in parenthesis) observed in the survey of transmembrane domains of  $\alpha$ -helix bundles (the five most frequent residues and Met and Cys) classified as buried or exposed to the membrane. The most significant side-side chain interactions of the five most frequent amino acids and Met and Cys.

## Results: Sulfur-containing amino acids Part II



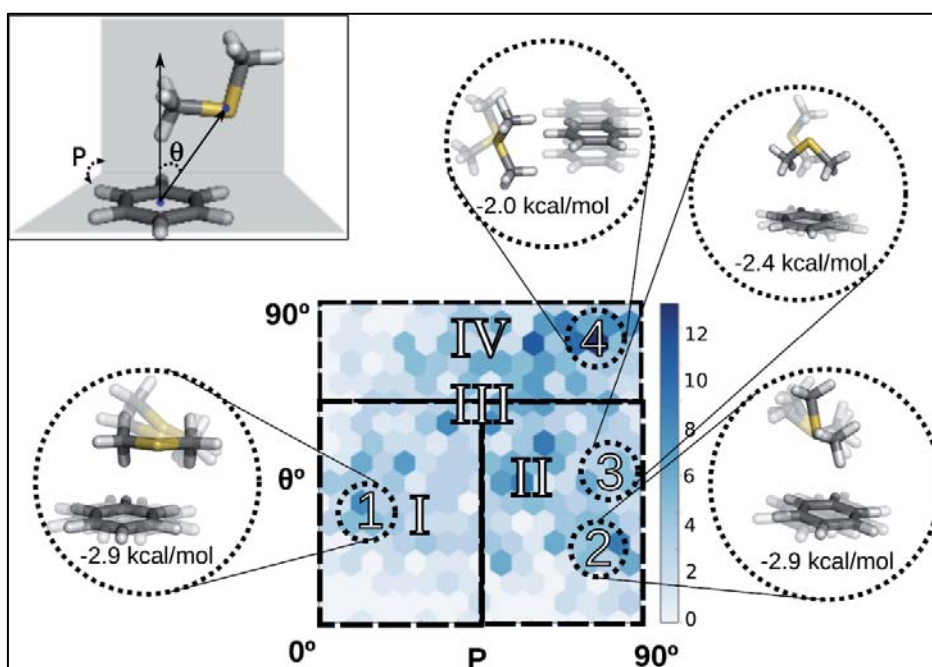
### *The orientation and strength of Met-Phe interactions*

The Met-Phe interactions identified in the crystal structures of membrane proteins (a total of 707 pairs, Table 1) were clustered based on relative distances and angles between the two amino acids (see Methods). Figure 1 shows the 2D histograms with the distribution of the Met-Phe interactions projected on the conformational space defined by  $\phi$  and  $\theta$  angles (see Methods and Supplementary Table S1). In order to evaluate the magnitude of the energy of interaction between both side chains, we performed high level *ab initio* calculations (see Methods) in small-molecule models systems (Met and Phe were mimicked by dimethyl sulfide (DMS) and the benzene ring (BNZ), respectively, Supplementary Figure S2). Clusters **I** (containing 24% of the observed interactions) and **II** (37%) reproduce the most favorable arrangements of the side chains (-2.9 kcal/mol) as calculated in comparable model systems **1** and **2**. Arabic numbers depict optimized *ab initio* models whereas roman numbers represent clusters observed in crystal structures. The two planes defined by DMS and BNZ molecules are almost parallel in model **1** and perpendicular in **2**, but in both cases a methyl group of DMS is located on top of the negative charge density at the center of BNZ ring ( $\pi$  electrons) and the sulfur atom on top of the positive charged density at the exterior of BNZ ring (-CH groups). In Cluster **III** (12%) the CH atoms of Met are pointing to the aromatic ring of Phe and the sulfur atom is pointing toward opposite direction, which results in an interaction of -2.4 kcal/mol in model **3**. Finally, cluster **IV** (28%) accounts for interactions in which the sulfur atom of Met acts as hydrogen bond acceptor for a -CH group from the phenyl ring of Phe, being the interaction energy -2.0 kcal/mol in the comparable model **4**. Interestingly, these computed energies are of the same magnitude as the values experimentally obtained for peptides in water (28).

In order to study the influence in the energy of interaction of the highly polarizable sulfur, compared to oxygen or the methylene group, we performed analogous *ab initio* calculations with model compounds that replace the sulfur atom (dimethyl sulfide, DMS, mimicking Met) by oxygen (dimethyl ether, DME) or -CH<sub>2</sub>- (propane, PRP, mimicking aliphatic amino acids) (Supplementary Figure S2). Comparison of these energies of interactions of model compounds **1-4** in DMS-BNZ complexes with analogous conformations of DME-BNZ shows that in all cases the sulfur-containing molecule (DMS) interacts stronger with the aromatic ring (BNZ) than in the oxygen-containing one (DME) with the exception of model **4**. This suggests that the induced positive charge density on the methyl group,

## Development and application of computational techniques to drug discovery and structure-function relationships

involved in the interaction with the  $\pi$  electrons of the ring in models **1-3**, is larger in the presence of the sulfur atom than in the presence of oxygen. Reasonably, because sulfur is a poorer hydrogen bond acceptor, in model **4** the sulfur atom forms weaker  $S\cdots HC$  hydrogen bond interaction with the CH group of the ring than oxygen ( $O\cdots HC$ ). Importantly, the energies of interactions of model compounds **1-4** in DMS-BNZ complexes are always more stable than in PRP-BNZ complexes, indicating that the interaction of aromatic rings with sulfur-containing groups is always stronger than with aliphatic groups.

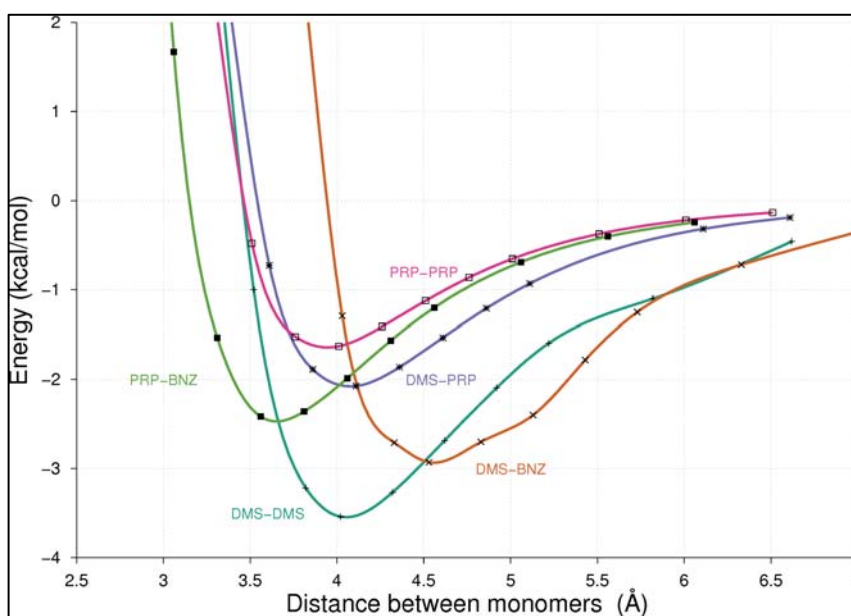


**Figure 1. The orientation and strength of Met-Phe interactions.** 2D histograms of the frequencies of occurrence of these interactions, clustered according to the conformational space defined by the distance  $d$  and the angles  $P$  and  $\theta$  (see Supplementary Table S1 for a detailed description). Roman and arabic numbers indicate the position in the 2D histogram of the most representative structure in the cluster and the energy-minimized structure, respectively. *Ab initio* geometry optimization at the MP2/6-31+G(d,p) level and calculated energy of interaction at the CCSD(T)/6-311+G(3df,2p) level (see Methods) are shown inside dotted circles as solid sticks. Representative structures obtained in the cluster analysis are shown as transparent sticks.

## Results: Sulfur-containing amino acids Part II



Because aromatic-aromatic interactions are considered key in the stability of membrane proteins,(8) we next compared the energies of interaction in DMS-BNZ complexes with those in BNZ-BNZ complexes. Comparison with the well-characterized (29-32) lowest energy configurations of BNZ-BNZ (Supplementary Figure S3), the T-shaped (-2.4 kcal/mol) and parallel displaced (-2.1 kcal/mol), indicates that Met forms more stable interactions with aromatics rings (DMS-BNZ) than aromatic-aromatic interactions (BNZ-BNZ).



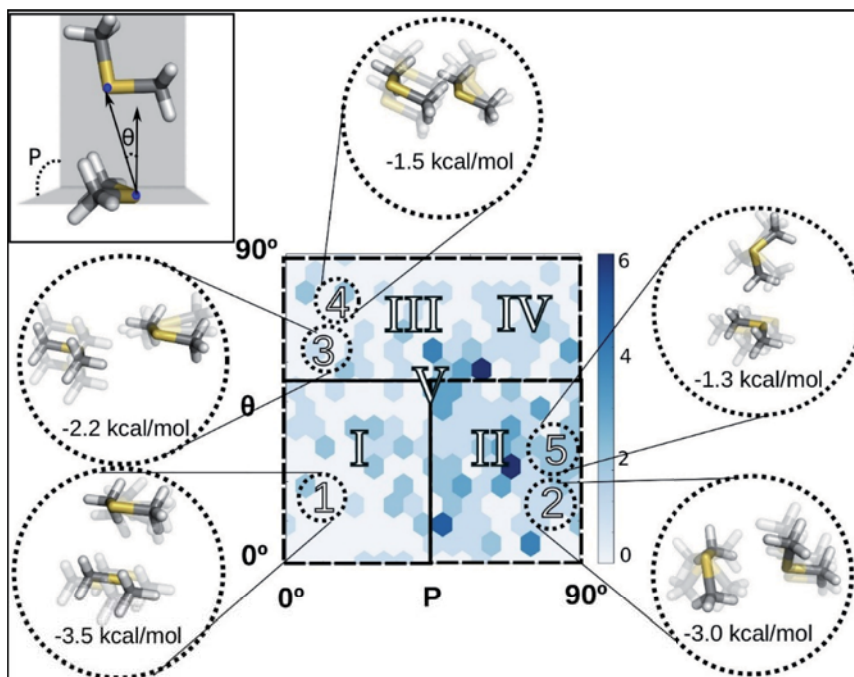
**Figure 2. Influence of the distance on the interaction energy.** Calculations were done on the DMS-DMS (Met-Met), DMS-BNZ (Met-Phe), DMS-PRP (Met-Leu), PRP-PRP (Leu-Leu), and PRP-BNZ (Leu-Phe) model systems with the lowest energy.  $d$  refers either to the sulfur-sulfur or the sulfur-benzene (centroid) distance.





***The orientation and strength of Met-Met interactions***

Clusters **I-V** in Figure 2, obtained from the 199 Met-Met interactions present in membrane proteins (Table 1), are calculated in a similar way to the clusters of Met-Phe (see above). Cluster **I**, containing 11% of the interactions, corresponds to an anti-parallel orientation, exhibiting the largest interaction energy (-3.5 kcal/mol) in the comparable model system **1**. The orientation of the Met side chains in cluster **II** (46%) is in the T-shaped configuration, being the interaction energy of -3.0 kcal/mol in the comparable model system **2**. In these configurations **1** and **2** each sulfur atom interacts respectively with four and



**Figure 3. The orientation and strength of Met-Met interactions.** 2D histograms of the frequencies of occurrence of these interactions, clustered according to the conformational space defined by the distance  $d$  and the angles  $P$  and  $\theta$  (Supplementary Table S2). See Legend of Figure 1 for further details.

## Results: Sulfur-containing amino acids Part II



three hydrogen atoms of the methyl groups ( $S\cdots HC$  interactions) that have positive charge density. The Met side chains in cluster **III** (5%) are in a parallel-displaced orientation, in a head-to-head configuration with both sulfur atoms engaged in the interaction (-2.2 kcal/mol in model **3**). Clusters **IV** (15%) and **V** (25%) account for the least favored Met-Met interactions (-1.5 and -1.3 kcal/mol in models **4** and **5**, respectively). Models mimicking these clusters reproduce a structure with a single sulfur atom interacting with four and two CH hydrogen atoms ( $S\cdots HC$  interactions), respectively. Cluster **IV** shows a parallel orientation in a head-to-tail configuration of the Met side-chains, while cluster **V** shows a T-shaped orientation in which the interactions occur through the methyl groups.

The influence of the highly polarizable sulfur atom in these interactions was evaluated by performing analogous *ab initio* calculations with model compounds that replace the sulfur atom (Supplementary Figure S4) (dimethyl sulfide, DMS) by oxygen (dimethyl ether, DME). The anti-parallel orientation of model **1** (-3.5 vs. -2.7 kcal/mol) and the T-shaped configuration of model **2** (-3.0 vs. -2.8 kcal/mol) are more stable in the sulfur-containing DMS-DMS complex than in the oxygen-containing DME-DME complex. The opposite is observed for models **3** (-2.2 vs. -2.4 kcal/mol), **4** (-1.5 vs. -1.6 kcal/mol) and **5** (-1.3 vs. -1.5 kcal/mol).

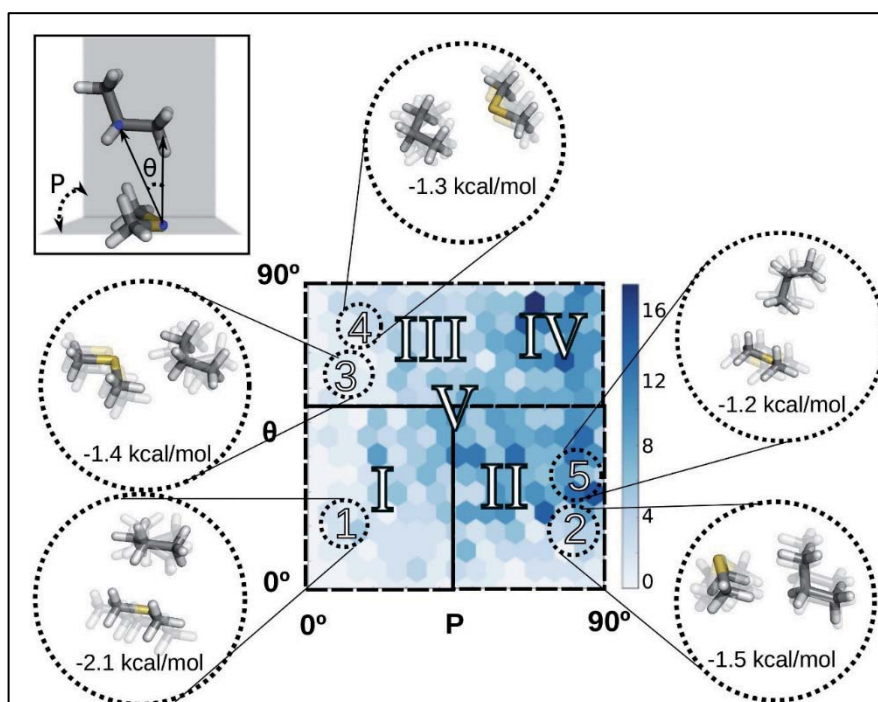
### ***The orientation and strength of Met-Leu interactions***

We have selected Leu as a representative residue to study the interactions of Met with aliphatic amino acids. The 1,129 Met-Leu interactions present in membrane proteins (Table 1) were clustered in a similar manner as in Met-Met interactions (see above). Because the  $C_\gamma$ ,  $S_\delta$  and  $C_\epsilon$  atoms of Met are analogous to the  $C_{\delta 1}$ ,  $C_\gamma$  and  $C_{\delta 2}$  atoms of Leu, the relative orientation of Met-Leu residues in clusters **I-V** (Figure 3) were taken in analogy with clusters **I-V** of Met-Met (Figure 2).

The computed interaction energies in comparable model systems **1-5** (Supplementary Figure S5) show that the anti-parallel orientation in model **1** exhibits the largest interaction energy (-2.1 kcal/mol). Comparison of the interaction energies in DMS-DMS (Met-Met clusters), DMS-PRP (Met-Leu clusters) and PRP-PRP (Leu-Leu clusters, not shown) models allow us to study the influence of the sulfur atom in the interaction energy. Clearly, the rank order of energies on interaction is DMS-DMS (2 sulfur atoms) < DMS-PRP (1 sulfur atom) < PRP-PRP (0 sulfur atom) in models **1** (-3.5 < -2.1 < -1.7 kcal/mol,



respectively), **2** ( $-3.0 < -1.5 < -1.4$  kcal/mol), **3** ( $-2.2 < -1.4 < -1.0$  kcal/mol), **4** ( $-1.5 < -1.3 < -1.1$  kcal/mol), and **5** ( $-1.3 < -1.2 < -1.0$  kcal/mol).



**Figure 4.** The orientation and strength of Met-Leu interactions. 2D histograms of the frequencies of occurrence of these interactions, clustered according to the conformational space defined by the distance  $d$  and the angles  $P$  and  $\theta$  (Supplementary Table S3). See Legend of Figure 1 for further details.

### **The interactions of Cys**

Cys interacts with Phe (a total of 276 pairs), Met (111 pairs) and hydrophobic amino acids such as Leu (474 pairs), Ala (396 pairs), Ile (401 pairs), and Val (357 pairs) (Table 1). In addition, Cys can interact with other Cys through a covalent disulfide bridge. Excluding disulfide bridges ( $S-S$  distances  $< 3\text{\AA}$ ) only 37 Cys-Cys pairs were observed in crystal structures (Table 1) in which Cys was acting as

## Results: Sulfur-containing amino acids Part II



hydrogen bond donor and/or acceptor. These Cys-Cys interactions were not further analyzed, as they belong to the common hydrogen bond interaction. Analysis of the crystal structures of Cys-Phe interactions in membrane proteins revealed three main interaction modes (Figure 4). In cluster **I** (12%) the sulfur  $S_{\gamma}$  atom is located on top of the aromatic ring, in cluster **II** (49%) the  $C_{\beta}$  atom is located on top of the ring, and in cluster **III** (39%) the sulfur  $S_{\gamma}$  atom is coplanar to the phenyl ring. *Ab initio* energy optimizations of model compounds (Cys and Phe were mimicked by methanethiol (MT) and the benzene ring (BNZ), respectively, Supplementary Figure S6) positioned the sulfhydryl hydrogen absent in the crystal structures. In model **1** (-3.0 kcal/mol) MT forms a S-H $\cdots$  $\pi$  hydrogen bond with the phenyl ring, whereas in model **2** (-2.9 kcal/mol) the S atom of MT is located on top of the positive charged density at the exterior of the BNZ ring and the methyl group on top of the negative charge density at the center of the ring. Importantly, *ab initio* energy minimization of model compound **3**, mimicking cluster **III**, led either to models **1** or **2**.

Clustering of the 111 Cys-Met and 474 Cys-Leu interactions (Table 1) was challenging due to the absence of the sulphhydryl hydrogen in the crystal structures. Thus, we performed *ab initio* energy optimizations of model compounds (MT-DMS or MT-PRP, Supplementary Figure S6) in which one of the methyl groups of DMS, in the reported Met-Met (DMS-DMS, Supplementary Figure S4) and Met-Leu (DMS-PRP, Supplementary Figure S5) interactions, was replaced by hydrogen. Comparison of these energies of interaction in MT-DMS complexes (-3.0, -2.9, -1.4, -1.7, -1.3 kcal/mol for models **1-5**, respectively) with analogous conformations of DMS-DMS (-3.5, -3.0, -2.2, -1.5, -1.3 kcal/mol for **1-5**, respectively) shows that DMS-DMS interactions are stronger with the exception of model **4**.

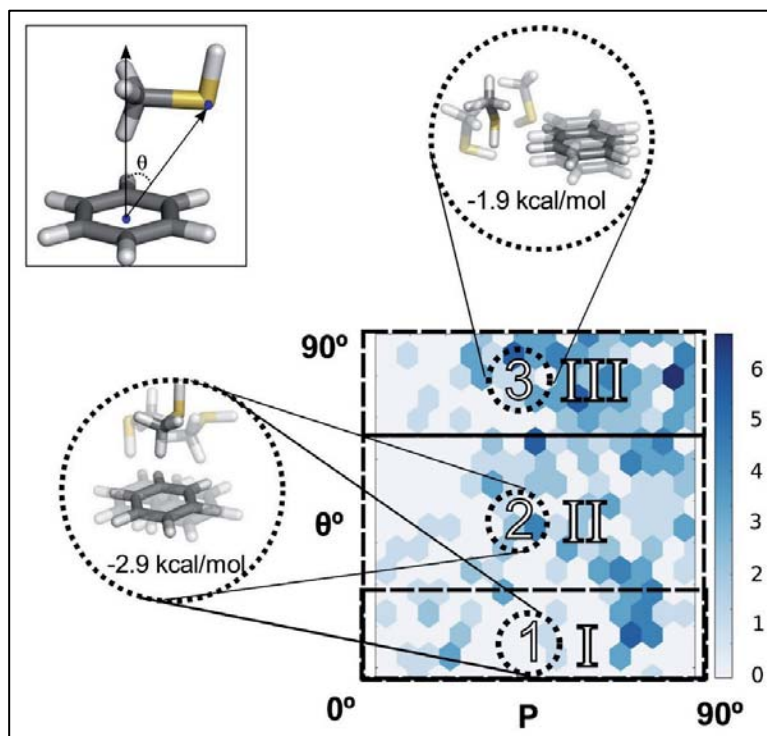
### ***Interaction energy comparison with the AMBER force field***

We computed the interaction energies of the small-molecule model compounds using the AMBER force field (33), with the aim of assessing the accuracy of this force field in reproducing the interactions of Met or Cys (see Methods). The results shown in Supplementary Figure S5 indicate a reasonable quantitative agreement in the interactions of Cys (mimicked by MT) with Met (DMS), Leu (PRP), and Phe (BNZ). The deviations are larger for Met (Supplementary Figures

## Development and application of computational techniques to drug discovery and structure-function relationships



S1, S3-S4) with an average difference relative to CCSD(T) of  $\sim 0.5$  kcal/mol. The larger deviations correspond to conformations in which sulphur atoms are in close proximity or when the sulphur atom is located on top of the benzene ring. Moreover, the rank order of Met-Phe interactions is not fully reproduced: CCSD(T) predicts  $1 = 2 < 3 < 4$  in DMS-BNZ models, while AMBER predicts  $3 < 1 < 2 < 4$  (Supplementary Figure S1). Similarly, CCSD(T) predicts  $1 < 2 < 3 < 4 < 5$  in DMS-DMS models, while AMBER predicts  $1 < 2 < 4 < 3 = 5$  (Supplementary Figure S3). In contrast, the interactions of Met with Leu are highly consistent, both in magnitude and rank order (Supplementary Figure S4). Overall, these results are in line with a recent report on  $\pi$ - $\pi$ , CH/ $\pi$ , and SH/ $\pi$  interactions (34).



**Figure 5.** The orientation and strength of Cys-Phe interactions. 2D histograms of the frequencies of occurrence of these interactions, clustered according to the conformational space defined by the distance  $d$  and the  $\theta$  angle. See Legend of Figure 1 for further details



---

## Conclusions

We and others have previously outlined the structural and functional role of Met-aromatic and Met-Met interactions in the family of G protein-coupled receptors (35-37). The present analysis of the inter-residue interactions in crystal structures of membrane protein reveals that Met and Cys often interact with Leu, Ile, Val, Phe, and other Met or Cys. The characterization of their strength using *ab-initio* calculations in small-molecule model systems, predicts that Met-Met, Met-Phe, Cys-Phe, Met-aliphatic and Cys-aliphatic interactions are stronger in magnitude than aliphatic-aliphatic interactions. Remarkably, Met-Met, Met-Phe, and Cys-Phe interactions are stronger than aromatic-aromatic. Thus, these types of interactions, which have often been misled, need to be taken into account when considering the forces that stabilize the overall fold in membrane proteins. In addition to the stronger interactions of Met and Cys, their more flexible side-chains may provide extra versatility and adaptation to conformational changes. We believe that these interactions are also likely to be important in the interior of globular proteins or in the formation of protein-ligand or protein-protein complexes. However, further studies would be necessary in these regard.

***Development and application of computational techniques to  
drug discovery and structure-function relationships***

---



## Results: Sulfur-containing amino acids Part II



### References

1. Muller-Dethlefs, K., and Hobza, P. (2000) Noncovalent interactions: A challenge for experiment and theory. *Chem. Rev.* **100**, 143-167
2. Meyer, E. A., Castellano, R. K., and Diederich, F. (2003) Interactions with aromatic rings in chemical and biological recognition. *Angewandte Chemie* **42**, 1210-1250
3. Nishio, M. (2004) CH/pi hydrogen bonds in crystals. *CrystEngComm* **6**, 130-158
4. Brandl, M., Weiss, M. S., Jabs, A., Suhnel, J., and Hilgenfeld, R. (2001) C-H...pi-interactions in proteins. *J. Mol. Biol.* **307**, 357-377
5. Kim, K., Karthikeyan, S., and Singh, J. (2011) How Different Are Aromatic pi Interactions from Aliphatic pi Interactions and Non-pi Stacking Interactions? *J. Chem. Theory Comput.*
6. Tsuzuki, S., Honda, K., Uchamaru, T., and Mikami, M. (2006) Estimated MP2 and CCSD(T) interaction energies of n-alkane dimers at the basis set limit: comparison of the methods of Helgaker et al. and Feller. *J. Chem. Phys.* **124**, 114304
7. Ringer, A. L., Figgs, M. S., Sinnokrot, M. O., and Sherrill, C. D. (2006) Aliphatic C-H/pi interactions: Methane-benzene, methane-phenol, and methane-indole complexes. *J. Phys. Chem. A* **110**, 10822-10828
8. Hong, H., Park, S., Jimenez, R. H., Rinehart, D., and Tamm, L. K. (2007) Role of aromatic side chains in the folding and thermodynamic stability of integral membrane proteins. *J. Am. Chem. Soc.* **129**, 8320-8327
9. Grossfield, A., and Woolf, T. B. (2002) Interaction of tryptophan analogs with POPC lipid bilayers investigated by molecular dynamics calculations. *Langmuir* **18**, 198-210
10. Valley, C. C., Cembran, A., Perlmutter, J. D., Lewis, A. K., Labello, N. P., Gao, J., and Sachs, J. N. (2012) The methionine-aromatic motif plays a unique role in stabilizing protein structure. *J. Biol. Chem.* **287**, 34979-34991
11. Pal, D., and Chakrabarti, P. (1998) Different types of interactions involving cysteine sulfhydryl group in proteins. *J. Biomol. Struct. Dyn.* **15**, 1059-1072
12. Pal, D., and Chakrabarti, P. (2001) Non-hydrogen bond interactions involving the methionine sulfur atom. *J. Biomol. Struct. Dyn.* **19**, 115-128
13. Morgan, R. S., Tatsch, C. E., Gushard, R. H., McAdon, J., and Warme, P. K. (1978) Chains of alternating sulfur and pi-bonded atoms in eight small proteins. *Int. J. Pept. Protein Res.* **11**, 209-217
14. Zauhar, R. J., Colbert, C. L., Morgan, R. S., and Welsh, W. J. (2000) Evidence for a strong sulfur-aromatic interaction derived from crystallographic data. *Biopolymers* **53**, 233-248
15. Samanta, U., Pal, D., and Chakrabarti, P. (2000) Environment of tryptophan side chains in proteins. *Proteins* **38**, 288-300



## Development and application of computational techniques to drug discovery and structure-function relationships



16. Pranata, J. (1997) Sulfur aromatic interactions: A computational study of the dimethyl sulfide benzene complex. *Bioorg. Chem.* **25**, 213-219
17. Cabaleiro-Lago, E. M., Hermida-Ramon, J. M., and Rodriguez-Otero, J. (2004) Computational study of the interaction in (CH<sub>3</sub>)<sub>2</sub>X dimer and trimer (X = O, S). *J. Phys. Chem. A* **108**, 4923-4929
18. Duan, G. L., Smith, V. H., and Weaver, D. F. (2001) Characterization of aromatic-thiol pi-type hydrogen bonding and phenylalanine-cysteine side chain interactions through ab initio calculations and protein database analyses. *Mol. Phys.* **99**, 1689-1699
19. Ringer, A. L., Senenko, A., and Sherrill, C. D. (2007) Models of S/pi interactions in protein structures: comparison of the H<sub>2</sub>S benzene complex with PDB data. *Protein Sci.* **16**, 2216-2223
20. Perea, M., Lugtenburg, I., Mayol, E., Cordomi, A., Deupi, X., Pardo, L., and Olivella, M. (2015) TAlphaDB and TMBetaDB: web servers to study the structural role of sequence motifs in alpha-helix and beta-barrel domains of membrane proteins. *BMC Bioinformatics* **16**, 266
21. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242
22. Mezei, M. (2003) A new method for mapping macromolecular topography. *J. Mol. Graph. Model.* **21**, 463-472
23. Chakrabarti, P., and Bhattacharyya, R. (2007) Geometry of nonbonded interactions involving planar groups in proteins. *Prog. Biophys. Mol. Biol.* **95**, 83-137
24. Riley, K. E., Platts, J. A., Rezac, J., Hobza, P., and Hill, J. G. (2012) Assessment of the performance of MP2 and MP2 variants for the treatment of noncovalent interactions. *J. Phys. Chem. A* **116**, 4159-4169
25. Hobza, P., Selzle, H. L., and Schlag, E. W. (1996) Potential energy surface for the benzene dimer. Results of ab initio CCSD(T) calculations show two nearly isoenergetic structures: T-shaped and parallel-displaced. *J. Phys. Chem.* **100**, 18790-18794
26. Boys, S. F., and Bernardi, F. (1970) The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **19**, 553-566
27. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N.,

## Results: Sulfur-containing amino acids Part II



- Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) Gaussian 09, Revision B.01. Wallingford CT
28. Viguera, A. R., and Serrano, L. (1995) Side-chain interactions between sulfur-containing amino acids and phenylalanine in alpha-helices. *Biochemistry* **34**, 8771-8779
  29. Janowski, T., and Pulay, P. (2007) High accuracy benchmark calculations on the benzene dimer potential energy surface. *Chem. Phys. Lett.* **447**, 27-32
  30. DiStasio, R. A., Jr., von Helden, G., Steele, R. P., and Head-Gordon, M. (2007) On the T-shaped structures of the benzene dimer. *Chem. Phys. Lett.* **437**, 277-283
  31. Sinnokrot, M. O., and Sherrill, C. D. (2006) High-accuracy quantum mechanical studies of pi-pi interactions in benzene dimers. *J. Phys. Chem. A* **110**, 10656-10668
  32. Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M., and Tanabe, K. (2002) Origin of attraction and directionality of the pi/pi interaction: model chemistry calculations of benzene dimer interaction. *J. Am. Chem. Soc.* **124**, 104-112
  33. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **117**, 5179-5197
  34. Sherrill, C. D., Sumpter, B. G., Sinnokrot, M. O., Marshall, M. S., Hohenstein, E. G., Walker, R. C., and Gould, I. R. (2009) Assessment of standard force field models against high-quality ab initio potential curves for prototypes of pi-pi, CH/pi, and SH/pi interactions. *J Comput Chem* **30**, 2187-2193
  35. Cordomi, A., Gomez-Tamayo, J. C., Gigoux, V., and Fourmy, D. (2013) Sulfur-containing amino acids in 7TMRs: molecular gears for pharmacology and function. *Trends Pharmacol. Sci.* **34**, 320-331
  36. Nygaard, R., Zou, Y., Dror, R. O., Mildorf, T. J., Arlow, D. H., Manglik, A., Pan, A. C., Liu, C. W., Fung, J. J., Bokoch, M. P., Thian, F. S., Kobilka, T. S., Shaw, D. E., Mueller, L., Prosser, R. S., and Kobilka, B. K. (2013) The Dynamic Process of beta(2)-Adrenergic Receptor Activation. *Cell* **152**, 532-542
  37. Magnan, R., Escrieut, C., Gigoux, V., De, K., Clerc, P., Niu, F., Azema, J., Masri, B., Cordomi, A., Baltas, M., Tikhonova, I. G., and Fourmy, D. (2013) Distinct CCK-2 receptor conformations associated with beta-arrestin-2 recruitment or phospholipase-C activation revealed by a biased antagonist. *J. Am. Chem. Soc.* **135**, 2560-2573

***Development and application of computational techniques to  
drug discovery and structure-function relationships***

---



**Results: Sulfur-containing amino acids**  
**Part II**



**Supplementary information**

Cluster	Number	P	$\theta$	Distance criteria
I	191 (24%)	0-45°	0-70°	$d(R-S\delta) < d(R-C\gamma) \vee d(R-S\delta) < d(R-C\epsilon)$
II	131 (37%)	45-90°	0-70°	$d(R-S\delta) < d(R-C\gamma) \vee d(R-S\delta) < d(R-C\epsilon)$
III	181 (12%)	0-90°	0-90°	$d(R-S\delta) > d(R-C\gamma) \wedge d(R-S\delta) > d(R-C\epsilon)$
IV	204 (28%)	0-90°	70-90°	-

**Table S1. Cluster analysis of Met-Phe interactions in crystal structures.** Number (and percentage) of Met-Phe interactions and values of P (angle between the planes defined by C, S, and C atoms of Met and the aromatic ring of Phe) and  $\theta$  (angle between the normal vector of the plane defined by the aromatic ring of Phe and the vector connecting the centroid R of the aromatic ring of Phe and S, of Met), and distance criteria (R accounts for the centroid of the aromatic ring of Phe, and S, C $\gamma$  and C $\epsilon$  represent the atoms of the Met side-chain), in clusters I-IV (see Fig 1).

## Development and application of computational techniques to drug discovery and structure-function relationships



Cluster	Number	P	$\theta$	Distance criteria
I	26 (11%)	0-45°	0-60°	$[d(S\delta_A-S\delta_B) < d(S\delta_A-C\gamma_A) \vee d(S\delta_A-S\delta_B) < d(S\delta_A-C\epsilon_A)] \wedge [d(S\delta_A-S\delta_B) < d(S\delta_B-C\gamma_A) \vee d(S\delta_A-S\delta_B) < d(S\delta_B-C\epsilon_A)]$
II	71 (46%)	45°-90°	0-60°	$[d(S\delta_A-S\delta_B) < d(S\delta_A-C\gamma_A) \vee d(S\delta_A-S\delta_B) < d(S\delta_A-C\epsilon_A)] \wedge [d(S\delta_A-S\delta_B) < d(S\delta_B-C\gamma_A) \vee d(S\delta_A-S\delta_B) < d(S\delta_B-C\epsilon_A)]$
III	12 (5%)	0-90°	60-90°	$[d(S\delta_A-S\delta_B) < d(S\delta_A-C\gamma_A) \wedge d(S\delta_A-S\delta_B) < d(S\delta_A-C\epsilon_A)] \wedge [d(S\delta_A-S\delta_B) < d(S\delta_B-C\gamma_A) \wedge d(S\delta_A-S\delta_B) < d(S\delta_B-C\epsilon_A)]$
IV	43 (15%)	0-90°	60-90°	$[d(S\delta_A-S\delta_B) > d(S\delta_A-C\gamma_A) \vee d(S\delta_A-S\delta_B) > d(S\delta_A-C\epsilon_A)] \wedge [d(S\delta_A-S\delta_B) < d(S\delta_B-C\gamma_A) \vee d(S\delta_A-S\delta_B) < d(S\delta_B-C\epsilon_A)]$
V	47 (25%)	0-90°	0-90°	$[d(S\delta_A-S\delta_B) > d(S\delta_A-C\gamma_A) \wedge d(S\delta_A-S\delta_B) > d(S\delta_A-C\epsilon_A)] \wedge [d(S\delta_A-S\delta_B) > d(S\delta_B-C\gamma_A) \vee d(S\delta_A-S\delta_B) > d(S\delta_B-C\epsilon_A)]$

**Table S2. Cluster analysis of Met-Met interactions in crystal structures.** Number (and percentage) of Met-Met interactions and values of P (angle between the planes defined by C, S, and C, atoms of Met) and  $\theta$  (angle between the normal vector of the plane defined by the C, S, and C, atoms of Met and the vector connecting the S, atoms of Met), and distance criteria (sub indexes A and B refer to atoms in distinct side-chains) in clusters I-V (see Fig 2).

## Results: Sulfur-containing amino acids Part II



Cluster	Number	P	$\theta$	Distance criteria
I	26 (11%)	0-45°	0-60°	$[d(S\delta_A-S\delta_B) < d(S\delta_A-C\gamma_s) \vee d(S\delta_A-S\delta_B) < d(S\delta_A-C\epsilon_s)] \wedge$ $[d(S\delta_A-S\delta_B) < d(S\delta_B-C\gamma_A) \vee d(S\delta_A-S\delta_B) < d(S\delta_B-C\epsilon_A)]$
II	71 (46%)	45°-90°	0-60°	$[d(S\delta_A-S\delta_B) < d(S\delta_A-C\gamma_s) \vee d(S\delta_A-S\delta_B) < d(S\delta_A-C\epsilon_s)] \wedge$ $[d(S\delta_A-S\delta_B) < d(S\delta_B-C\gamma_A) \vee d(S\delta_A-S\delta_B) < d(S\delta_B-C\epsilon_A)]$
III	12 (5%)	0-90°	60-90°	$[d(S\delta_A-S\delta_B) < d(S\delta_A-C\gamma_s) \wedge d(S\delta_A-S\delta_B) < d(S\delta_A-C\epsilon_s)] \wedge$ $[d(S\delta_A-S\delta_B) < d(S\delta_B-C\gamma_A) \wedge d(S\delta_A-S\delta_B) < d(S\delta_B-C\epsilon_A)]$
IV	43 (15%)	0-90°	60-90°	$[d(S\delta_A-S\delta_B) > d(S\delta_A-C\gamma_s) \vee d(S\delta_A-S\delta_B) > d(S\delta_A-C\epsilon_s)] \wedge$ $[d(S\delta_A-S\delta_B) < d(S\delta_B-C\gamma_A) \vee d(S\delta_A-S\delta_B) < d(S\delta_B-C\epsilon_A)]$
V	47 (25%)	0-90°	0-90°	$[d(S\delta_A-S\delta_B) > d(S\delta_A-C\gamma_s) \wedge d(S\delta_A-S\delta_B) > d(S\delta_A-C\epsilon_s)] \wedge$ $[d(S\delta_A-S\delta_B) > d(S\delta_B-C\gamma_A) \vee d(S\delta_A-S\delta_B) > d(S\delta_B-C\epsilon_A)]$

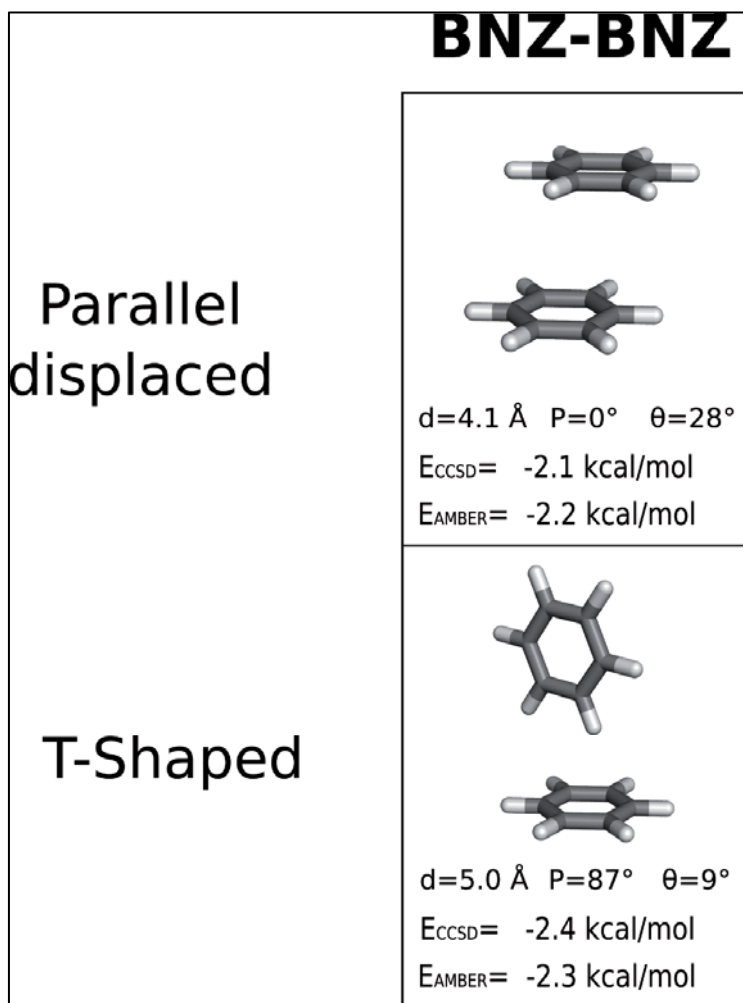
**Table S3. Cluster analysis of Met-Leu interactions in crystal structures.** Number (and percentage) of Met-Leu interactions and values of P (angle between the planes defined by the C<sub>s</sub>, S, and C<sub>i</sub> atoms of Met and the C<sub>ss</sub>, C<sub>i</sub>, and C<sub>sz</sub> atoms of Leu),  $\theta$  (angle between the normal vector of the plane defined by the C<sub>s</sub>, S, and C<sub>i</sub> atoms of Met and the vector connecting the S<sub>i</sub> atom of Met and the C<sub>i</sub> atom of Leu) and distance criteria (subindexes A and B refer to atoms in distinct side-chains) in clusters I-V (see Fig 3).



	DMS-BNZ	DME-BNZ	PRP-BNZ
1	  $d=4.3 \text{ \AA}$ $P=6^\circ$ $\theta=32^\circ$ $E_{\text{CCSD}} = -2.9 \text{ kcal/mol}$ $E_{\text{AMBER}} = -2.7 \text{ kcal/mol}$	  $d=3.9 \text{ \AA}$ $P=16^\circ$ $\theta=36^\circ$ $E_{\text{CCSD}} = -2.4 \text{ kcal/mol}$	  $d=3.6 \text{ \AA}$ $P=10^\circ$ $\theta=5^\circ$ $E_{\text{CCSD}} = -2.4 \text{ kcal/mol}$ $E_{\text{AMBER}} = -2.2 \text{ kcal/mol}$
2	  $d=4.0 \text{ \AA}$ $P=90^\circ$ $\theta=33^\circ$ $E_{\text{CCSD}} = -2.9 \text{ kcal/mol}$ $E_{\text{AMBER}} = -1.8 \text{ kcal/mol}$	  $d=4.3 \text{ \AA}$ $P=90^\circ$ $\theta=70^\circ$ $E_{\text{CCSD}} = -2.3 \text{ kcal/mol}$	  $d=3.6 \text{ \AA}$ $P=89^\circ$ $\theta=1^\circ$ $E_{\text{CCSD}} = -1.8 \text{ kcal/mol}$ $E_{\text{AMBER}} = -2.1 \text{ kcal/mol}$
3	  $d=4.9 \text{ \AA}$ $P=90^\circ$ $\theta=1^\circ$ $E_{\text{CCSD}} = -2.4 \text{ kcal/mol}$ $E_{\text{AMBER}} = -2.8 \text{ kcal/mol}$	  $d=4.7 \text{ \AA}$ $P=90^\circ$ $\theta=17^\circ$ $E_{\text{CCSD}} = -2.1 \text{ kcal/mol}$	  $d=4.7 \text{ \AA}$ $P=90^\circ$ $\theta=18^\circ$ $E_{\text{CCSD}} = -1.8 \text{ kcal/mol}$ $E_{\text{AMBER}} = -2.4 \text{ kcal/mol}$
4	  $d=5.2 \text{ \AA}$ $P=90^\circ$ $\theta=90^\circ$ $E_{\text{CCSD}} = -2.0 \text{ kcal/mol}$ $E_{\text{AMBER}} = -1.3 \text{ kcal/mol}$	  $d=4.5 \text{ \AA}$ $P=90^\circ$ $\theta=90^\circ$ $E_{\text{CCSD}} = -2.1 \text{ kcal/mol}$	  $d=5.3 \text{ \AA}$ $P=90^\circ$ $\theta=90^\circ$ $E_{\text{CCSD}} = -1.4 \text{ kcal/mol}$ $E_{\text{AMBER}} = -1.4 \text{ kcal/mol}$

**Figure S1. Small-molecule models systems mimicking Met-Phe interactions.** Geometry optimized, at the *ab-initio* MP2/6-31+G(d,p) level of theory, of the interactions between benzene (BNZ, mimicking Phe) and dimethyl sulfide (DMS, mimicking Met), dimethyl ether (DME), and propane (PRP, mimicking Leu). Each optimized structure is designated by an arabic number that corresponds to a roman number of the obtained clusters in crystal structures (see Fig 1). The values of  $d$ ,  $P$ , and  $\theta$  (see Suppl. Table 1 and Fig 1 for definition), and single point energy calculations at the *ab-initio* CCSD(T)/6-311+G(3df,2p) level of theory ( $E_{\text{CCSD}}$ ) and by molecular mechanics using the AMBER99 forcefield ( $E_{\text{AMBER}}$ ) are shown.

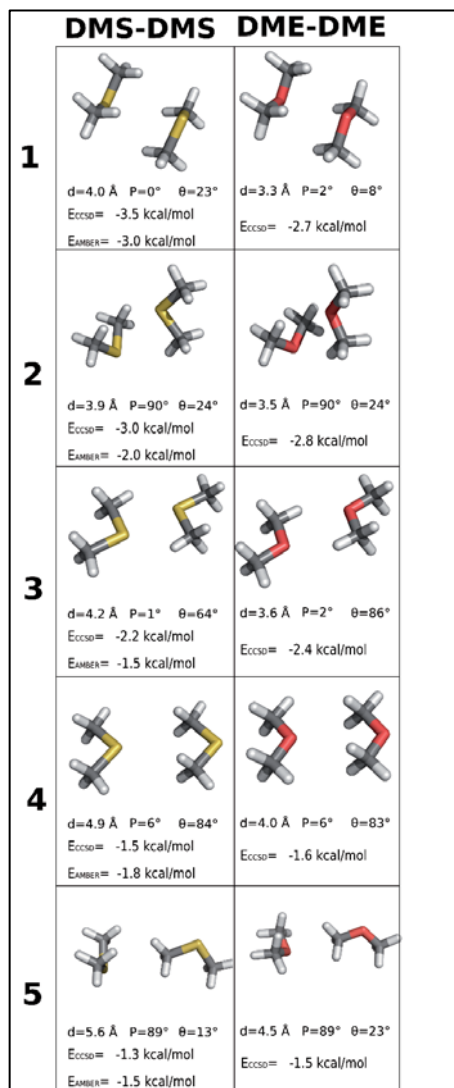
**Results: Sulfur-containing amino acids**  
**Part II**



**Figure S2. Small-molecule models systems mimicking Phe-Phe interactions.** Geometry optimized models, at the *ab-initio* MP2/6-31+G(d,p) level of theory, of benzene-benzene (BNZ, mimicking Phe) interactions in the lowest parallel displaced and T-shaped energy configurations. The values of  $d$  (calculated as the distance between the centroid  $R$  of the aromatic ring of BNZ),  $P$  (calculated as the angle between the planes defined by the aromatic rings of BNZ), and  $\theta$  (angle between the normal vector of the plane defined by the aromatic ring of Phe and the vector connecting the centroids  $R$  of the aromatic rings of BNZ), and single point energy calculations at the *ab-initio* CCSD(T)/6-311+G(3df,2p) level of theory ( $E_{\text{CCSD}}$ ) and by molecular mechanics using the AMBER99 forcefield ( $E_{\text{AMBER}}$ ).



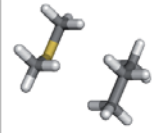
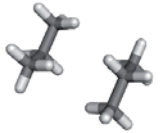
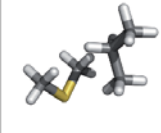
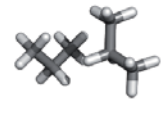




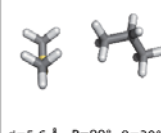
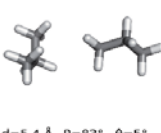
# Development and application of computational techniques to drug discovery and structure-function relationships



**Figure S3. Small-molecule models systems mimicking Met-Met interactions.** Geometry optimized models, at the *ab-initio* MP2/6-31+G(d,p) level of theory, of dimethyl sulfide (DMS, mimicking Met)-DMS and dimethyl ether (DME)-DME interactions. Each energy-minimized structure is designated by an arabic number that corresponds to a roman number of the obtained clusters in crystal structures (see Fig 2). The values of  $d$ ,  $P$ , and  $\theta$  (see Suppl. Table 2 and Fig 2 for definition), and single point energy calculations at the *ab-initio* CCSD(T)/6-311+G(3df,2p) level of theory ( $E_{CCSD}$ ) and by molecular mechanics using the AMBER99 forcefield ( $E_{AMBER}$ ) are shown.

## Results: Sulfur-containing amino acids Part II



	DMS-PRP	PRP-PRP
<b>1</b>	 $d=4.1 \text{ \AA}$ $P=3^\circ$ $\theta=21^\circ$ $E_{\text{CCSD}} = -2.1 \text{ kcal/mol}$ $E_{\text{AMBER}} = -2.1 \text{ kcal/mol}$	 $d=4.1 \text{ \AA}$ $P=0^\circ$ $\theta=15^\circ$ $E_{\text{CCSD}} = -1.7 \text{ kcal/mol}$ $E_{\text{AMBER}} = -1.5 \text{ kcal/mol}$
<b>2</b>	 $d=4.6 \text{ \AA}$ $P=89^\circ$ $\theta=38^\circ$ $E_{\text{CCSD}} = -1.5 \text{ kcal/mol}$ $E_{\text{AMBER}} = -1.4 \text{ kcal/mol}$	 $d=4.4 \text{ \AA}$ $P=90^\circ$ $\theta=11^\circ$ $E_{\text{CCSD}} = -1.4 \text{ kcal/mol}$ $E_{\text{AMBER}} = -1.2 \text{ kcal/mol}$
<b>3</b>	 $d=3.7 \text{ \AA}$ $P=0^\circ$ $\theta=90^\circ$ $E_{\text{CCSD}} = -1.4 \text{ kcal/mol}$ $E_{\text{AMBER}} = -1.2 \text{ kcal/mol}$	 $d=4.4 \text{ \AA}$ $P=90^\circ$ $\theta=90^\circ$ $E_{\text{CCSD}} = -1.0 \text{ kcal/mol}$ $E_{\text{AMBER}} = -0.8 \text{ kcal/mol}$
<b>4</b>	 $d=4.7 \text{ \AA}$ $P=2^\circ$ $\theta=88^\circ$ $E_{\text{CCSD}} = -1.3 \text{ kcal/mol}$ $E_{\text{AMBER}} = -1.3 \text{ kcal/mol}$	 $d=4.7 \text{ \AA}$ $P=2^\circ$ $\theta=89^\circ$ $E_{\text{CCSD}} = -1.1 \text{ kcal/mol}$ $E_{\text{AMBER}} = -0.9 \text{ kcal/mol}$
<b>5</b>	 $d=5.6 \text{ \AA}$ $P=90^\circ$ $\theta=20^\circ$ $E_{\text{CCSD}} = -1.2 \text{ kcal/mol}$ $E_{\text{AMBER}} = -1.2 \text{ kcal/mol}$	 $d=5.4 \text{ \AA}$ $P=83^\circ$ $\theta=5^\circ$ $E_{\text{CCSD}} = -1.0 \text{ kcal/mol}$ $E_{\text{AMBER}} = -0.8 \text{ kcal/mol}$

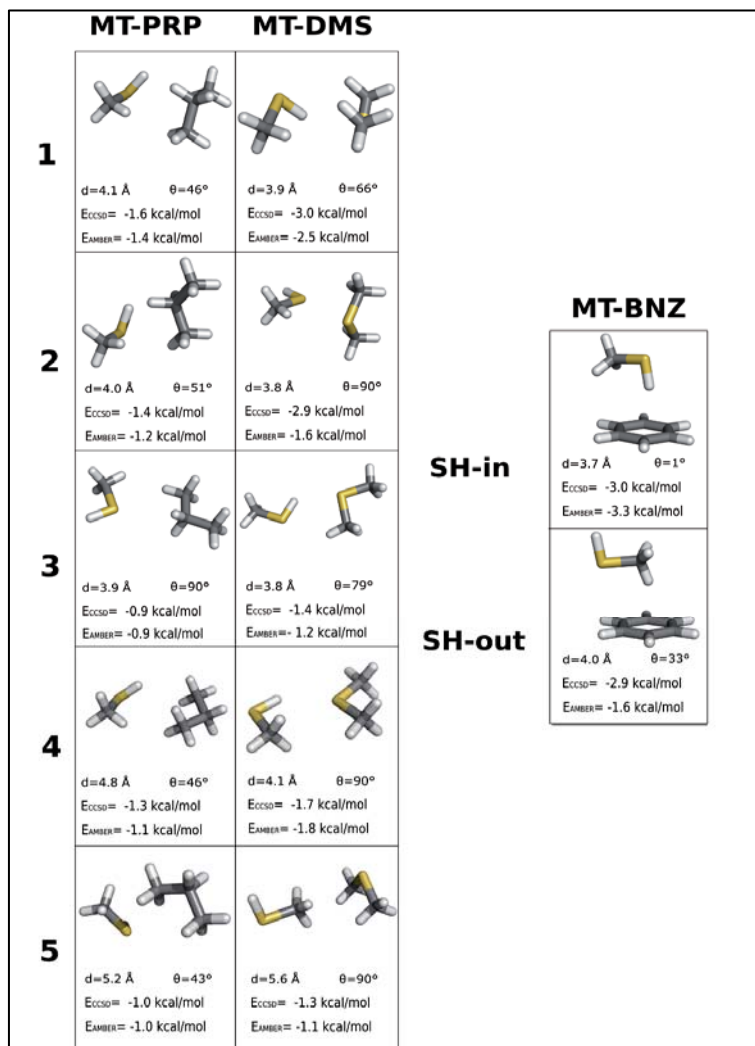
**Figure S4.** Small-molecule models systems mimicking Met-Leu interactions. Geometry optimized models, at the *ab-initio* MP2/6-31+G(d,p) level of theory, of dimethyl sulfide (DMS, mimicking Met) and propane (PRP, mimicking Leu) and PRP-PRP interactions. Each optimized structure is designated by an arabic number that corresponds to a roman number of the obtained clusters in crystal structures (see Fig 3). The values of  $d$ ,  $P$ , and  $\theta$  (see Suppl. Table 3 and Fig 3 for definition), and single point energy calculations at the *ab-initio* CCSD(T)/6-311+G(3df,2p) level of theory ( $E_{\text{CCSD}}$ ) and by molecular mechanics using the AMBER99 forcefield ( $E_{\text{AMBER}}$ ) are shown.

***Development and application of computational techniques to  
drug discovery and structure-function relationships***

---



## Results: Sulfur-containing amino acids Part II

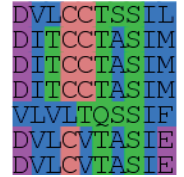


**Figure S5.** Small-molecule models systems mimicking Cys-Phe, Cys-Met and Cys-Leu interactions. Geometry optimized models, at the *ab-initio* MP2/6-31+G(d,p) level of theory, of the interactions between methanethiol (MT, mimicking Cys) and benzene (BNZ, mimicking Phe), dimethyl sulfide (DMS, mimicking Met) and propane (PRP, mimicking Leu). The values of  $d$  (calculated as the distance between S and Cy in PRP or the centroid R of the aromatic ring in BNZ, respectively) and  $\theta$  (angle between the normal vector of the plane defined by the aromatic ring of Phe or the plane defined by  $C_\beta$ ,  $S_\gamma$  and HS atoms and the vector connecting the centroid R of the aromatic ring of BNZ or the central atom of the side-chain to the center of the other side-chain), and single point energy calculations at the *ab-initio* CCSD(T)/6-311+G(3df,2p) level of theory ( $E_{\text{CCSD}}$ ) and by molecular mechanics using the AMBER99 force field ( $E_{\text{AMBER}}$ ).

***Development and application of computational techniques to  
drug discovery and structure-function relationships***

---





---

## **4.5 GPCR-SAS: G protein-coupled receptors Sequence Analysis and Statistics**

### **Abstract**

G protein coupled receptors (GPCRs) are one of the largest protein families in mammals. They mediate signal transduction across cell membranes and are important targets for the pharmaceutical industry. The GPCR-SAS web server provides a set of tools to analyze conservation of residues or sequence motifs across TM segments of GPCRs, identify correlations in substitutions and give statistical information of such correlations in sequence alignments. Thus, GPCR-SAS is a useful instrument that permits to study for instance family-specific mechanisms of activation and ligand or G protein selectivity.

### **Introduction**

G protein coupled receptors (GPCRs) are one of the most prevailing protein families in mammalian genomes (1). They are involved in most signal transduction processes across membranes, including the response to hormones and neurotransmitters and the senses of sight, smell and taste. GPCRs transduce extracellular signals across the cell membrane through G protein dependent but also through G protein independent processes. GPCRs are classified according to sequence similarities into six families or classes named A to F (2-4). Only classes A, B, C and F are present in humans and class A or rhodopsin-like comprises, by far, the largest number of members (5). The crystal structures showed a conserved TM structure with a common fold formed by an extracellular N-terminus, seven transmembrane helices (TM1-7), connected by alternating intracellular (i1 to i3) and extracellular (e1 to e3) hydrophilic loops, and a cytoplasmic C-terminus. In classes A, B and F the C-terminus contains an  $\alpha$ -helix (Hx8) parallel to the cell membrane (6-8).

The GPCR-SAS (GPCR Sequence Analysis and Statistics) server takes advantage of this structural similarity in the TM domain and allows performing statistical analysis of sequence positions or motifs within the TM helices and Hx8 of GPCRs of classes A, B, C and F. This statistical analysis can be used to rationalize ligand selectivity, G protein recognition or receptor activation, among others (9). The

# Development and application of computational techniques to drug discovery and structure-function relationships



algorithm is able to classify the results according to the occurrence in GPCR subfamilies.

## Design and Implementation

GPCR-SAS is a web application freely accessible at <http://lmc.uab.cat/gpcrsas/>. It is implemented in Python language and employs Django framework. The application relies on a MySQL database that contains the sequence alignments and various classification schemes. The main page shows two main sections (see Figure 1): *Positions and Sequence* and *Classification*. *Positions and Sequence* consists of four input forms (four boxes), two for positions and another two for amino acids. Depending on the input provided in the panel *Positions and Sequence*, GPCR-SAS can perform conservation, co-variance and correlation analyses. *Classification* provides navigable multilevel hierarchical classification systems of GPCRs in families and subfamilies according to different schemes (see below).

**Figure 1.** GPCR-SAS main page. Main page consists in two sections: 1) receive position(s) and motif(s) inputs and determines the type of calculation; 2) the hierarchical dropdown lists which selects the receptor families in which to perform the analysis. The type of analysis to perform depends on the fields which are filled (see below).

## Results: GPCR-SAS

DVLCCTSSIL  
DITCCTASIM  
DITCCTASIM  
DITCCTASIM  
VLVLTQSSIF  
DVLCVTASIE  
DVLCVTASIE

### Numbering scheme

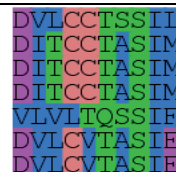
GPCR-SAS uses the Ballesteros & Weinstein numbering scheme designed for class A GPCRs (10), where the position of each residue is described by two numbers: the helix in which the residue is located and the position relative to a conserved residue in that helix, arbitrarily assigned to 50. In class A, these amino acids are: N1.50 in TM1 (97.6% conserved in non-olfactory human class A), D2.50 in TM2 (92.1%), R3.50 in TM3 (94.8%), W4.50 in TM4 (95.8%), P5.50 in TM5 (76.0%), P6.50 in TM6 (98.3%), and P7.50 in TM7 (93.7%). Although non-class A GPCRs most often do not have such conserved amino acids at these positions (11), extrapolation of the class A numbering-scheme is now possible thanks to structure-based sequence alignment between classes (12,13). Due to the lack of homology out of the TM domains GPCR-SAS cannot provide statistics for residues located on loops. Valid search ranges for the helices in each family are displayed in Table 1.

Helix	Start	End	# residues
TM1	1.23	1.61	39
TM2	2.37	2.72	36
TM3	3.23	3.63	41
TM4	4.40	4.71	32
TM5	5.33	5.69	37
TM6	6.28	6.61	34
TM7	7.28	7.53	26
sH8	8.40	8.57	18

*Table 1. Helix limits and length as implemented in GPCR-SAS*



# *Development and application of computational techniques to drug discovery and structure-function relationships*



## Sequence alignment

Amino acid sequences for GPCRs belonging to classes A, B, C and F were retrieved from the UniProtKB/Swiss-Prot database (<http://www.uniprot.org>). Alignments were performed independently for the four classes using (see table 2) ClustalO (14) as follows: *i*) Sequences corresponding to the available crystal structures were used to determine the length of the TM helices and to generate a profile, *ii*) Human sequences were aligned to the sequences of the crystal structures using a gap-opening penalty of 40 and a gap extension penalty of 0.1 (these parameters limit insertion of gaps in the TM segments). *iii*) The remaining vertebrate and invertebrate sequences were subsequently aligned. At each step the alignment was manually verified and no gaps were tolerated on TM regions other than those previously described at TMs 2 and 5 (15,16), which can be extrapolated to non-crystalized receptors without ambiguity, or those associated to trivial deletions.

	Class A		Class B	Class C	Class F
	Non-olfactory	Olfactory			
Human	291	425	49	22	36
Mammalian	1465	549	150	85	240
Vertebrata	1698	557	165	87	261
Eukaryota	1801	561	203	109	292
All	1816	561	203	109	293

*Table 2. Number of sequences for each class and subset of organism.*

## Results: GPCR-SAS

DVLCCTSSIL  
DITCCTASIM  
DITCCTASIM  
DITCCTASIM  
VLVLTQSSIF  
DVLCVTASIE  
DVLCVTASIE

### Classification schemes

GPCR-SAS database contains a set of classification systems based on different properties. Fredrickson classification system (17) is based on a phylogenetic analysis of human GPCR sequences. GPCRdb (18) (gpcr database) uses a pharmacologic classification of the receptors. BIAS-PROF GDS classification system (19) is based on physicochemical properties of the sequence. Rognan classification system (20) relies on the phylogenetic analysis of 30 residues around the ligand binding site. Chabbert classification system (21) is obtained from the calculation of Neighbor joining trees to sequence similarity of the class A transmembrane helices.<

	A	B	C	F
Fredrickson (17)				
GPCRdb (18)	Default	Default	Default	Default
BIAS-PROF (GDS) (19)				
Iuphar (22)				
Rognan (20)				
Chabbert (21)				
GPCR-safari (23)				

Table 3. Classification schemes available at GPCR-SAS.

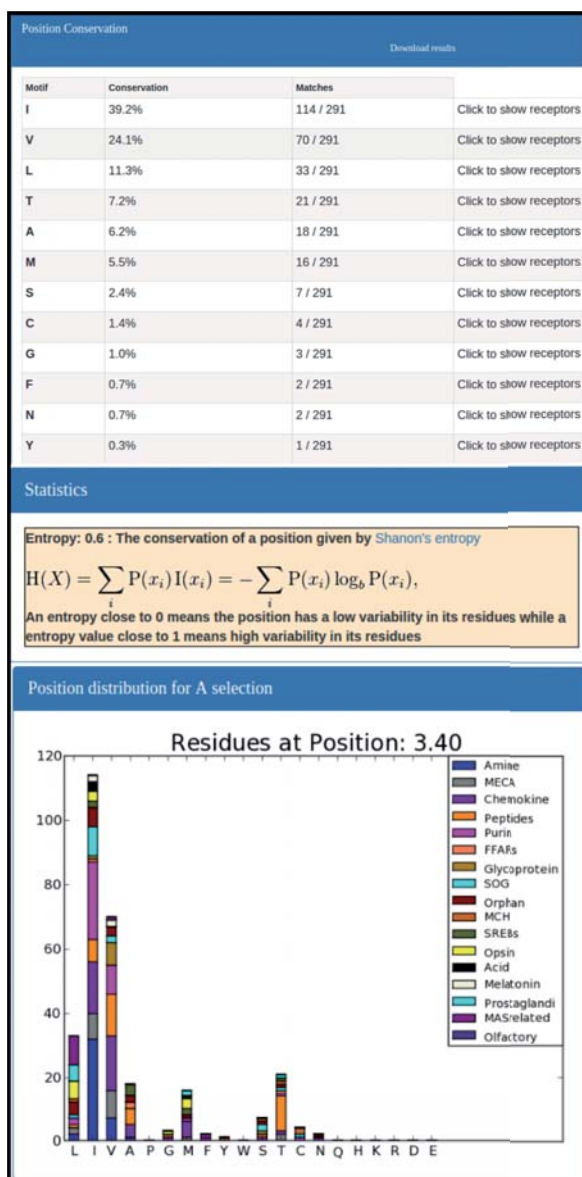
### Conservation analysis for a position or a set/range of positions.

If the given as input consists of only one single position (i.e. 3.50), consecutive positions (i.e. 3.50-3.54) or non-consecutive positions (3.50, 4.50, 5.50) the output is a list of amino acids or sequence motifs sorted by their occurrence on the active receptor selection and (only for searches with one single position) a bar plot that displays the amino acid distribution grouped by subfamilies. The results page also shows the Shannon entropy for the requested position or motif as a quantitative measure of variability (24). As a second step, the user may

# Development and application of computational techniques to drug discovery and structure-function relationships

DVLCCCTSSIL  
 DITCCTASIM  
 DITCCTASIM  
 DITCCTASIM  
 VLVIITQSSIF  
 DVLCVITASIE  
 DVLCVITASIE

request to display the list of receptors that contain each of the amino acids or motifs by clicking on “Click to show receptors” (see figure 2).



**Figure 2. Example of GPCR-SAS conservation analysis.**

## ***Results: GPCR-SAS***

DVLCCTSSIL  
DITCCTASIM  
DITCCTASIM  
DITCCTASIM  
VLVLTQSSIF  
DVLCVTASIE  
DVLCVTASIE

The user may provide both a positional input and an amino acid or sequence motif. In this case the output consists of the observed frequency and two drop down menus that help to compare the observed frequencies for the other (sub) families at the chosen level of classification and the subfamilies one level below.

### **Covariance analysis of two positions**

If two positions are given as input without specifying any amino acid, the calculation returns a list of amino acid or motif pairs sorted by their percentage of occurrence together with the entropy, the Observed Minus Expected Squared (OMES) value for the positions, which is based on a  $\chi^2$  test (25) Employing this method Pelé et al. were able to identify evolutionary hubs between pairs of residues in GPCRs (26). To evaluate the significance for the computed OMES value, GPCR-SAS computes the Z-score and the associated p-value based on the OMES values computed for all possible combinations of amino acids.

### **Correlation analysis**

When the user provides data for all input fields (two positions and two amino acids or motifs), the output is a table displaying how presence of the amino acid or motif at the first position is associated with presence of the amino acid or motif at the second position. The obtained values are used to compute an odds ratio that estimates how strongly the presence/absence of one of the amino acids is correlated with the presence/absence of the other amino acid. Two drop down lists preview the same information for the other groups at the current level of classification and one level below as the selection are displayed by clicking on the respective titles.

# Development and application of computational techniques to drug discovery and structure-function relationships

D V L C C T S S I L  
 D I T C C T A S I M  
 D I T C C T A S I M  
 D I T C C T A S I M  
 V L V I T Q S S I F  
 D V L C V T A S I E  
 D V L C V T A S I E

Position/s conservation		
Motif	Conservation	Matches
WN	60.1%	175 / 291
FD	14.1%	41 / 291
WD	3.8%	11 / 291
FN	2.1%	6 / 291
QN	2.1%	6 / 291
YN	1.7%	5 / 291
WT	1.4%	4 / 291
AN	1.4%	4 / 291
GN	1.4%	4 / 291
SD	1.4%	4 / 291
WH	1.0%	3 / 291

Statistics
<p>Entropy: 0.3 : The conservation of a position/s given by Shanon's entropy</p> $H(X) = \sum_i P(x_i) I(x_i) = - \sum_i P(x_i) \log_2 P(x_i),$ <p>An entropy close to 0 means the position has a low variability in its residues while a entropy value close to 1 means high variability in its residues</p>
<p>OMES: 12.6 : Mesures the correlation between two positions</p> $OMES(i, j) = \frac{1}{N(i, j)} \sum_{x,y} (N_{x,y}^{obs}(i, j) - N_{x,y}^{ex}(i, j))^2$ <p>OMES (observed minus Expected Squared) calculates the difference between the observed and expected occurrences of each possible pair of amino acids (x,y) at positions i and j of the alignment</p> <p>Z-Score : 35.8            P-value &lt; 0.001 P-value &lt; 0.0</p> <p>Z-score and P-value give the significance of the correlation by comparing the obtained OMES for the two positions with the OMES mean for all combinations of two positions.</p>

**Figure 2.** Covariance analysis for positions 6.48 and 7.49 in class A human receptors.

## Results: GPCR-SAS

DVLCCTSSIL  
 DITCCTASIM  
 DITCCTASIM  
 DITCCTASIM  
 VLVLQSSIF  
 DVLCVTASIE  
 DVLCVTASIE



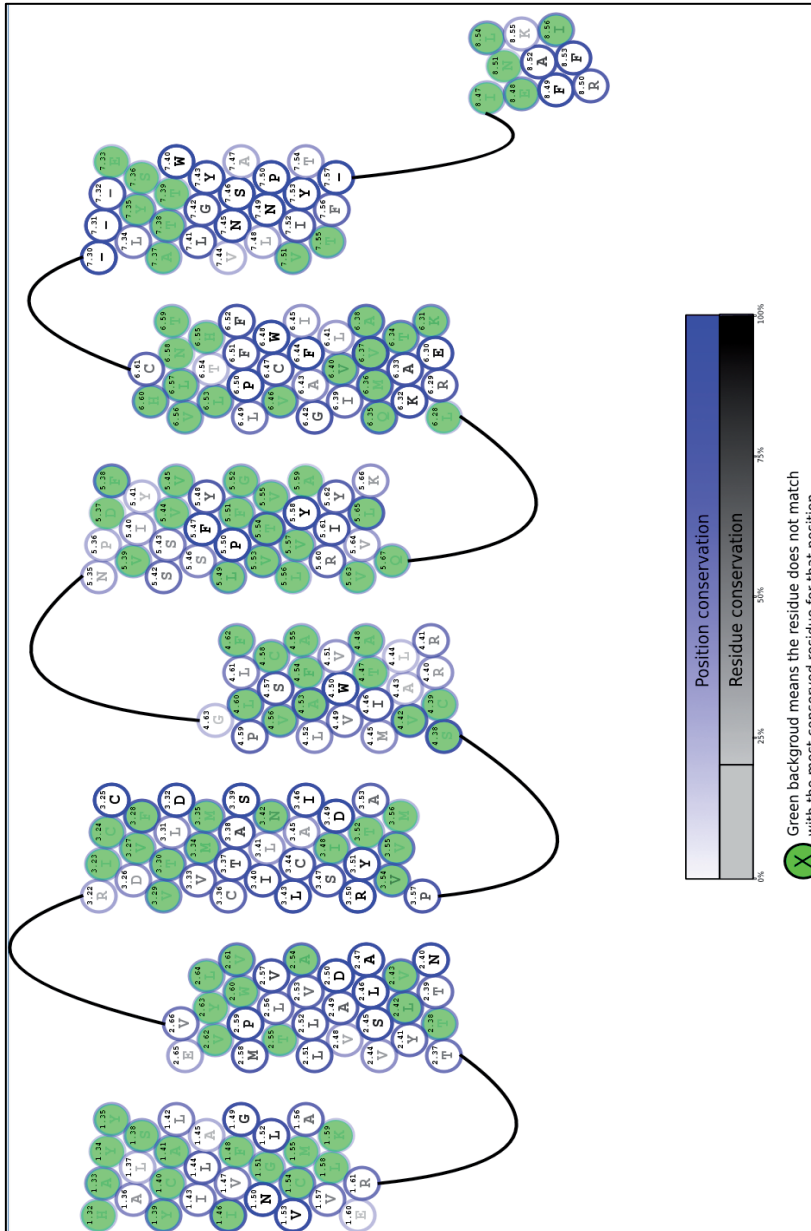
Figure 3. Correlation analysis example.

## Snake-plot representations

GPCR-SAS can also provide snake-plot representations for a particular receptor or for the consensus sequence of a group of selected receptors. In the consensus snake-plots produced by GPCR-SAS, each residue is represented by a circle, where both the amino acid letter and the contour are colored according to conservation (blue and black gradients, respectively). In single receptor snake-plots, if a residue does not match the most conserved residue for the active selection (i.e. branch, subfamily ...) its circle is colored in green and the contour indicated conservation of the consensus residue.

# Development and application of computational techniques to drug discovery and structure-function relationships

DVLCCTSSIL  
 DITCCTASIM  
 DITCCTASIM  
 DITCCTASIM  
 VLVIITQSSIF  
 DVLCVITASIE  
 DVLCVITASIE



**Figure 5.** Snakeplot representation of dopamine 3 receptor. Blue scale and grey scale maps the conservation of the position and the conservation of the specific residue in the position. Green background corresponds to residues which does not match with the consensus residue for the selected family.

## Results: GPCR-SAS

DVLCCTSSIL  
DITCCTASIM  
DITCCTASIM  
DITCCTASIM  
VLVLTQSSIF  
DVLCVTASIE  
DVLCVTASIE

### Results

The examples below illustrate how GPCR-SAS can be used, for instance, to speculate if some interactions observed in GPCR crystal structures are likely to exist in other GPCRs, if residues involved in the binding site of a receptors are common to others or not or to identify common activation elements.

#### The transmission switch

In class A GPCRs, rearrangement of the packing between residues 3.40, 5.50 and 6.44 leads to a weakening of the TM5-TM6 interface and local conformational changes that are transmitted towards the cytoplasmic ends of TM5 and TM6. These residues form the named transmission switch (Deupi Biochem Soc Trans 40 2012 383). We know that in the beta2-adrenergic receptors the residues at positions 3.40, 5.50 and 6.44 are Ile, Pro and Phe. Let us examine the content of position 3.40 in (human-non olfactory) class A GPCRs. A GPCR-SAS search for position 3.40 in the default sequence set (human-non olfactory class A GPCRs) returns the frequencies for all the observed amino acids in the left panel. The right panel displays the entropy of this position 0.6, which means a moderate variability (entropy ranges between 0 and 1) and a histogram displaying the distribution of the observed amino acids in the different subfamilies for this position. For example we can observe that most Amine receptors feature Ile at this position. Indeed, if we repeat the search selecting the subfamily Amine (GPCRdb classification scheme) we see that the frequencies of Ile raises to 76.2%. By clicking on “Show receptors” we can get the list of the receptors that contain each of the amino acids. For instance, HRH3\_HUMAN is the only Amine receptor that contains Ala in this position.

Similar GPCR-SAS searches for positions 5.50 and 6.40 show that most receptors class A receptors contain P5.50 (78.0%) and 6.44: F (80.4%). It looks like the triad 3.40, 5.50 and 6.44 is highly conserved in class A GPCRs. We can search for the three positions together to identify the most common triads. These are I3.40-P5.50-F6.44 (32.3%), V3.40-P5.50-F6.44 (16.5%) and other triads with frequencies < 5%.

Ile at position 3.40 with tells that 39.2% of the receptors contain Ile. Adding Ile in the first motif input (one letter amino acid code) and updating the calculation shows the conservation for Ile in the position 3.40. By clicking on the “Same-



## *Development and application of computational techniques to drug discovery and structure-function relationships*

DVLCCTSSIL  
DITCCTASIM  
DITCCTASIM  
DITCCTASIM  
VLVLTQSSIF  
DVLCVTASIE  
DVLCVTASIE

level” drop-down box we can compare to the other categories at the same level of classification (in this case we have not selected any level and thus the comparison is with the other classes). Ile is rarely/never found at position 3.40 in classes B, C and F. By clicking on the “Sub-level” drop-down box we compare within the subcategories of the parent category (in this case, Branches within class A). Almost half of the receptors at the alpha branch according the Fredrikson contain Ile3.40. By contrast the beta branch is the category with less frequency of Ile3.40 (16.7%). We may repeat the search using the wildcard for apolar amino acids instead of Ile. GPCR-SAS tells that the sum of frequencies for aliphatic residues adds to 86.9%.

If we search for the frequencies of amino acids at position 5.50 in other families, GPCR-SAS tells us that Pro is also the most conserved residue in the class F, although there are no class B or class C receptors containing Pro at this position. Still, we have suggested, based on mutagenesis studies, that 3.40 and 6.44 also form a transmission switch in the class B B1 Peptide GPCRs (13). GPCR-SAS tells us that 3.40 is a conserved aromatic position (Y:66.7% and F:33.3%), whereas 6.44 contain Ile (50.0%) and Phe (43.8%). Thus, B1 Peptide receptors have switched residue types at positions 3.40 and 6.44 compared to class A receptors. May this be a common feature of all class B GPCRs? GPCR-SAS suggests that there is a significant correlation between presence of an aromatic residue at position 3.40 and an aliphatic residue at position 6.44, with an odds ratio of 22.0.

What happens with the 3.40-6.44 pair in the Class C? GPCR-SAS tells that 3.40 contains mostly aromatic residues {Y:50%, F:27%}. However, 6.44 has small frequency of aromatic or aliphatic. The preferred residues are Thr (36.4%) and Ser (18.2%). Interestingly the analysis of the correlation between both positions in Class C shows a strong significant correlation (P-value < 0.001). This correlation does not exist for class F (P-value < 0.001).

## Results: GPCR-SAS

DVLCCTSSIL  
DITCCTASIM  
DITCCTASIM  
DITCCTASIM  
VLVLTQSSIF  
DVLCVTASIE  
DVLCVTASIE

### ***From ligand binding to conformational changes***

19.9% of human non-olfactory class A GPCRs exhibit D at position 3.32 -known to constitute the main anchoring point of aminergic ligands- and 31.3% have Y at position 7.43. For aminergic GPCRdb subfamily, a correlation analysis using GPCR-SAS shows that having D3.32 is always associated with the presence of either Y (93%) or W (7%) at position 7.43. This suggests that an aromatic residue that can form a hydrogen bond with D3.32 is essential for some receptors. D-Y is present not only in 90% of aminergic receptors, but also in 19% of peptide receptors (using GPCRdb classification scheme) including, for example, all somatostatin, opioid and urotensin II receptors. On the other hand D-W is specific of about half of the histamine and dopamine receptors. With an odds ratio of 71.5 (taking the whole class A human GPCRs set), the probability of having one of these residues when the other is present is 71.5 times higher than without its presence. The fact the anchoring point performed by these positions could form an evolutionary hub in class A GPCRs was previously reported (27). In order to estimate the probability of these positions forming an evolutionary hub, a covariance analysis was performed for class A human GPCRs. Resulting z-score value for OMES calculation is 16. This value means the OMES value for this pair of positions is shifted 16 standard deviations from the mean making it a high score evolutionary hub as reported by Pelè et al (26).

The crystal structures of  $\beta_2$ -adrenergic receptors revealed that two aromatic residues in TM6 participate in the binding of both agonists and antagonists: F6.51 and F6.52. These two residues could be active modulators of the toggle switch associated to CWxP (6.47-6.52). A GPCR-SAS search for CWxPFF shows that this motif is present in 10.7% human non-olfactory class A GPCRs, in particular in 61.1% of the aminergic -including all adrenergic and dopamine- and 11.7% of the peptide receptors.

### ***Y5.58 and the NPXXY motif***

Crystal structures of opsin and  $\beta_2$ -adrenergic receptor bound to a G-protein (or a peptide derived from the  $G\alpha$  C-terminus) revealed that Y5.58 and R3.50 exhibit important conformational changes from the inactive to the active structures, suggesting an important functional role for them (28,29). A GPCR-SAS search for

*Development and application of computational techniques to drug discovery and structure-function relationships*

---

DVLCCTSSIL  
DITCCTASIM  
DITCCTASIM  
DITCCTASIM  
VLVLTQSSIF  
DVLCVTASIE  
DVLCVTASIE

position 5.58 showed that Y5.58 is one of the most highly conserved residues in TM5, being present in 74.9% of human sequences. 16,5% of human sequences feature residues with hydrogen bonding capability (N, S, T, Q and H), which most likely, would preserve the role of Y in the GPCR activation mechanism. F is the non-polar most prevalent residue at this position. A subsequent detailed GPCR-SAS analysis shows that F5.58 is conserved in the SREB subfamily. Consequently, these receptors are likely to exhibit significant differences in the activation mechanism through TM5.

## Results: GPCR-SAS

DVLCCTSSIL  
DITCCTASIM  
DITCCTASIM  
DITCCTASIM  
VLVLTQSSIF  
DVLCVTASIE  
DVLCVTASIE

### References

1. Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006) How many drug targets are there? *Nat Rev Drug Discov* **5**, 993-996
2. Kolakowski, L. F. (1994) GCRDB - A G-protein-coupled receptor database. *Receptors & channels* **2**, 1-7
3. Attwood, T. K., and Findlay, J. B. (1994) Fingerprinting G-protein-coupled receptors. *Protein Eng.* **7**, 195-203
4. Horn, F., Bettler, E., Oliveira, L., Campagne, F., Cohen, F. E., and Vriend, G. (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.* **31**, 294-297
5. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G., and Schioth, H. B. (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256-1272
6. Liapakis, G., Cordomi, A., and Pardo, L. (2012) The G-protein coupled receptor family; actors with many faces. *Curr Pharm Des.* **18**, 175-185
7. Katritch, V., Cherezov, V., and Stevens, R. C. (2013) Structure-function of the G protein-coupled receptor superfamily. *Annual review of pharmacology and toxicology* **53**, 531-556
8. Venkatakrisnan, A. J., Deupi, X., Lebon, G., Tate, C. G., Schertler, G. F., and Babu, M. M. (2013) Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185-194
9. Deupi, X., Dolker, N., Lopez-Rodriguez, M. L., Campillo, M., Ballesteros, J. A., and Pardo, L. (2007) Structural models of class a G protein-coupled receptors as a tool for drug design: insights on transmembrane bundle plasticity. *Curr. Top. Med. Chem.* **7**, 991-998
10. Ballesteros, J. A., and Weinstein, H. (1995) Integrated Methods for Modeling G-Protein Coupled Receptors. *Methods Neurosci.* **25**, 366-428
11. Isberg, V., de Graaf, C., Bortolato, A., Cherezov, V., Katritch, V., Marshall, F. H., Mordalski, S., Pin, J. P., Stevens, R. C., Vriend, G., and Gloriam, D. E. (2015) Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends in pharmacological sciences* **36**, 22-31
12. Siu, F. Y., He, M., de Graaf, C., Han, G. W., Yang, D., Zhang, Z., Zhou, C., Xu, Q., Wacker, D., Joseph, J. S., Liu, W., Lau, J., Cherezov, V., Katritch, V., Wang, M. W., and Stevens, R. C. (2013) Structure of the human glucagon class B G-protein-coupled receptor. *Nature* **499**, 444-449
13. Spyridaki, K., Matsoukas, M. T., Cordomi, A., Gkoutelias, K., Papadokostaki, M., Mavromoustakos, T., Logothetis, D. E., Margioris, A. N., Pardo, L., and

## Development and application of computational techniques to drug discovery and structure-function relationships



- Liapakis, G. (2014) Structural-Functional Analysis of the Third Transmembrane Domain of the Corticotropin-releasing Factor Type 1 Receptor: Role In Activation And Allosteric Antagonism. *The Journal of biological chemistry* **289**, 18966-18977
14. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**, 539
  15. Gonzalez, A., Cordomi, A., Caltabiano, G., and Pardo, L. (2012) Impact of helix irregularities on sequence alignment and homology modeling of G protein-coupled receptors. *Chembiochem : a European journal of chemical biology* **13**, 1393-1399
  16. Becu, J. M., Pele, J., Rodien, P., Abdi, H., and Chabbert, M. (2013) Structural evolution of G-protein-coupled receptors: a sequence space approach. *Methods in enzymology* **520**, 49-66
  17. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G., and Schioth, H. B. (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Molecular Pharmacology* **63**, 1256-1272
  18. Vroling, B., Sanders, M., Baakman, C., Borrmann, A., Verhoeven, S., Klomp, J., Oliveira, L., de Vlieg, J., and Vriend, G. (2011) GPCRDB: information system for G protein-coupled receptors. *Nucleic Acids Research* **39**, D309-319
  19. Davies, M. N., Secker, A., Freitas, A. A., Mendao, M., Timmis, J., and Flower, D. R. (2007) On the hierarchical classification of G protein-coupled receptors. *Bioinformatics* **23**, 3113-3118
  20. Surgand, J. S., Rodrigo, J., Kellenberger, E., and Rognan, D. (2006) A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins* **62**, 509-538
  21. Deville, J., Rey, J., and Chabbert, M. (2009) An indel in transmembrane helix 2 helps to trace the molecular evolution of class A G-protein-coupled receptors. *Journal of molecular evolution* **68**, 475-489
  22. Pawson, A. J., Sharman, J. L., Benson, H. E., Faccenda, E., Alexander, S. P., Buneman, O. P., Davenport, A. P., McGrath, J. C., Peters, J. A., Southan, C., Spedding, M., Yu, W., Harmar, A. J., and Nc, I. (2014) The IUPHAR/BPS Guide to PHARMACOLOGY: an expert-driven knowledgebase of drug targets and their ligands. *Nucleic Acids Res* **42**, D1098-1106
  23. Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., Bellis, L., and Overington, J. P. (2015) ChEMBL web services:

## Results: GPCR-SAS

DVLCCTSSIL  
DITCCTASIM  
DITCCTASIM  
DITCCTASIM  
VLVLTQSSIF  
DVLCVTASIE  
DVLCVTASIE

- streamlining access to drug discovery data and utilities. *Nucleic Acids Res* **43**, W612-620
24. Shannon, C. E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal* **27**, 379-423
  25. Fodor, A. A., and Aldrich, R. W. (2004) On evolutionary conservation of thermodynamic coupling in proteins. *J Biol Chem* **279**, 19046-19050
  26. Pele, J., Moreau, M., Abdi, H., Rodien, P., Castel, H., and Chabbert, M. (2014) Comparative analysis of sequence covariation methods to mine evolutionary hubs: examples from selected GPCR families. *Proteins* **82**, 2141-2156
  27. Pele, J., Abdi, H., Moreau, M., Thybert, D., and Chabbert, M. (2011) Multidimensional scaling reveals the main evolutionary pathways of class A G-protein-coupled receptors. *PloS one* **6**, e19094
  28. Scheerer, P., Park, J. H., Hildebrand, P. W., Kim, Y. J., Krauss, N., Choe, H. W., Hofmann, K. P., and Ernst, O. P. (2008) Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497-502
  29. Rasmussen, S. G., Devree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K., and Kobilka, B. K. (2011) Crystal structure of the beta(2) adrenergic receptor-Gs protein complex. *Nature*

*Development and application of computational techniques to drug discovery and structure-function relationships*

---

DVLCCTSSIL  
DITCCTASIM  
DITCCTASIM  
DITCCTASIM  
VLVLTQSSIF  
DVLCVTASIE  
DVLCVTASIE

***Results: The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.***



---

## **4.6 The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.**

### **Abstract**

The finding that ergotamine binds serotonin receptors in a less conserved extended binding pocket close to the extracellular entrance, in addition to the orthosteric site, allowed us to obtain 5-HT<sub>7</sub>R antagonist **6** endowed with high affinity ( $K_i = 0.7$  nM) and significant 5-HT<sub>1A</sub>R selectivity (ratio > 1428). Compound **6** exhibits in vivo antidepressant-like effect (1 mg/kg, ip) mediated by the 5-HT<sub>7</sub>R, which reveals its interest as a putative research tool or pharmaceutical in depression disorders.

### **Introduction**

The serotonin (5-hydroxytryptamine, 5-HT) 5-HT<sub>7</sub> receptor (5-HT<sub>7</sub>R), first cloned in 1993,(1,2) is distributed in discrete areas of the brain and in the periphery.(3) Within the central nervous system (CNS), particularly high levels have been detected in thalamus, hippocampus and hypothalamus (especially within the suprachiasmatic nucleus). To date, the functional role of the 5-HT<sub>7</sub>R in various pathophysiological processes has been described.(3-6) For instance, activation of the 5-HT<sub>7</sub>R is involved in nociceptive processing, and it has been proposed that agonists of the receptor could be used as adjuvant drugs in pain treatment.(7,8) Results from animal models of learning and memory have suggested that the 5-HT<sub>7</sub>R represents a potential therapeutic target for the treatment of memory dysfunction in cognitive disorders (schizophrenia, Alzheimer's disease, and age-related decline).(9,10) On the other hand, pharmacological blockade or genetic inactivation of the 5-HT<sub>7</sub>R leads to an antidepressant-like behavioral profile in vivo.(11,12) It has also been suggested that the clinically established antidepressant effect of atypical antipsychotic drugs amisulpride and lurasidone is due to their blockade of the 5-HT<sub>7</sub>R.(13,14)

Major depression is a common psychiatric disorder associated with high symptomatic and functional burdens. Pharmacological treatment is often effective, but there remain substantial unmet needs in the form of non-responders, delayed



## ***Development and application of computational techniques to drug discovery and structure-function relationships***



onset of clinical effect, and side effects. In recent decades, an increasing demand has emerged for the discovery of new types of antidepressants. The therapy of this psychiatric disease is dominated by selective serotonin reuptake inhibitor (SSRI) antidepressants. However, the main drawback of this class of drugs is to produce a therapeutic effect only after several weeks of treatment. Importantly, a recent study showed that the pharmacological blockade of the 5-HT<sub>7</sub>R produces a faster antidepressant-like response than the 5-HT uptake inhibitor fluoxetine, commonly used as an antidepressant agent.(15) Therefore, 5-HT<sub>7</sub>R antagonists could represent a possible alternative to SSRIs as antidepressant drugs with a faster onset of action. All together, these studies have positioned the 5-HT<sub>7</sub>R as a promising drug target in the search for new antidepressant agents. Indeed, several antagonists of the receptor induce an antidepressant-like behavior in preclinical models of depression.(16-18) Interestingly, 5-HT<sub>7</sub>R antagonist SB-269970 (3-((2R)-2-[2-(4-methylpiperidin-1-yl)ethyl]pyrrolidin-1-yl)sulfonyl)phenol) also potentiates the effect of clinically used antidepressants such as citalopram, imipramine, desipramine, and moclobemide;(19,20) and JNJ-18038683 (3-(4-chlorophenyl)-1,4,5,6,7,8-hexahydro-1-(phenylmethyl)pyrazolo[3,4-*d*]azepine 2-hydroxy-1,2,3-propanetricarboxylate) has been evaluated in a clinical trial (phase 2) for the treatment of major depressive disorder.(21) This boosts the efforts toward the development of new 5-HT<sub>7</sub>R antagonists that are still needed to validate their suitability as therapeutic agents.

In this context, our research group has previously identified a structurally new family of 5-HT<sub>7</sub>R ligands represented by **1** and **2** (Chart 1).(22,23) However, the demonstrated close similarities between the binding sites of 5-HT<sub>7</sub> and 5-HT<sub>1A</sub> receptors have resulted in difficulties to develop selective 5-HT<sub>7</sub>R agents.(18,24-26) Thus, in this work we have focused our efforts in obtaining ligands endowed with 5-HT<sub>7</sub>/5-HT<sub>1A</sub> receptor selectivity. We took advantage of the increasing number of available crystal structures of the G protein-coupled receptor (GPCR) family,(27) to which 5-HT<sub>1A</sub> and 5-HT<sub>7</sub> receptors belong. All these structures share the common architecture of seven transmembrane domains (TMs),(28) forming a water-filled binding-site crevice. Moreover, the recent crystal structures of ergotamine bound to 5-HT<sub>1B</sub> and 5-HT<sub>2B</sub> receptors have revealed a large ligand binding cavity defined by the regular orthosteric pocket within TMs 3, 5, 6, 7 and extracellular loop (ECL) 2, embedded deep in the 7TM core, and an extended binding pocket close to the extracellular entrance.(29,30) Because this extracellular entrance is less conserved than the orthosteric binding site,(31) this finding opens the opportunity to obtain ligands with significant subtype selectivity.

**Results: The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.**



Compound **2**, previously reported in our laboratory,(23) showed high binding affinity for the 5-HT<sub>7</sub>R ( $K_i = 7$  nM) and moderate selectivity (5-HT<sub>7</sub>/5-HT<sub>1A</sub> ratio = 31). This compound was predicted to bind in the conserved orthosteric pocket within TMs 3, 5, and 6.(23) Because the 7TM bundle limits the length of this orthosteric pocket between TMs 3 (D3.32) and 5 (S5.42), we designed compounds **3-10** (Chart 1) that contain increasing number of methylene units in the spacer with the aim of improving receptor selectivity. In particular, analogue **6** has been characterized as a potent and selective 5-HT<sub>7</sub>R antagonist that exhibits in vivo antidepressant activity in the tail suspension and the forced swim tests.

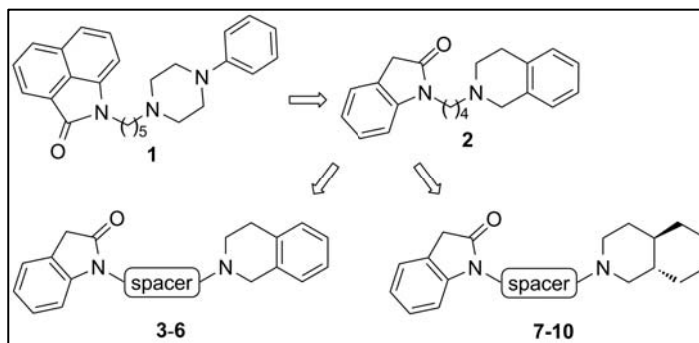


Chart 1

## Results and discussion

In the design of compounds **3-10** (Chart 1, Table 1), the main idea was that derivatives with a small number of methylene units (short spacers) would fit the protonated amine and the indolone moieties of the ligand within the orthosteric site, showing moderate selectivity due to a conserved pocket (Fig. 1A, top panel). In contrast, compounds with a large number of methylene units (long spacers) would reverse the binding mode so that the protonated amine binds D3.32 in the orthosteric site and the indolone moiety would expand toward the extracellular entrance where sequence divergences between 5-HT<sub>7</sub> and 5-HT<sub>1A</sub> receptors are observed (Fig. 1A, bottom panel).

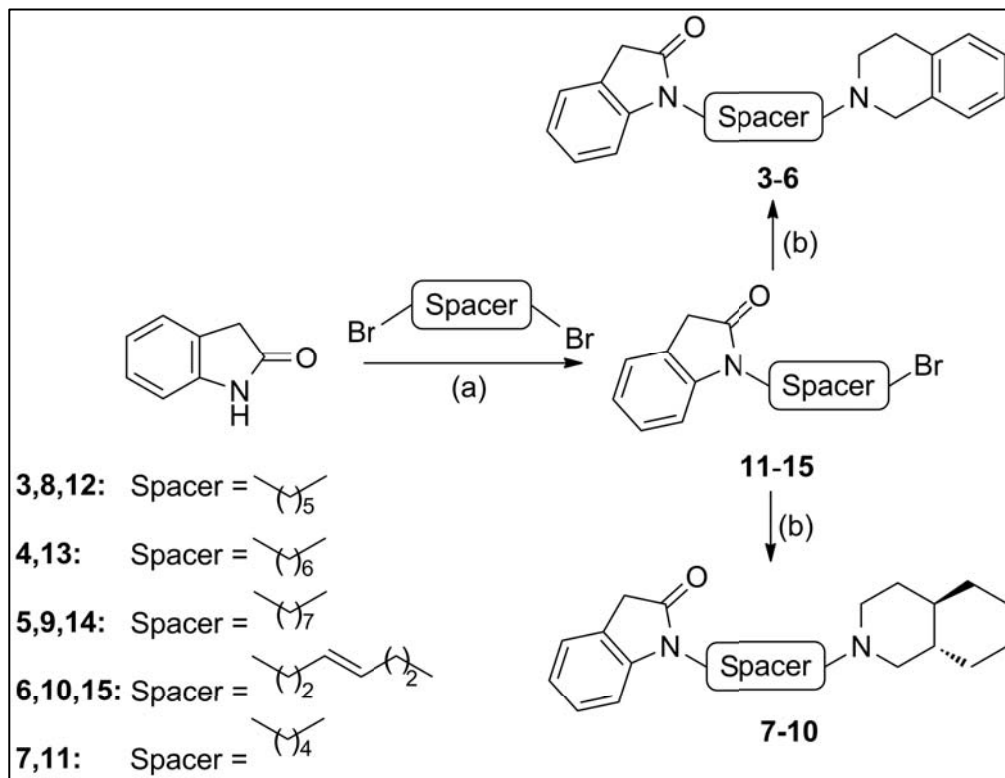
## Development and application of computational techniques to drug discovery and structure-function relationships



Target compounds **3-10** were synthesized starting from commercially available 1,3-dihydro-2*H*-indol-2-one. Alkylation with the appropriate dibromide derivative in the presence of  $K_2CO_3$  afforded bromoalkanes **11-15**. These intermediates were subsequently treated with 1,2,3,4-tetrahydroisoquinoline or ( $\pm$ )-*trans*-decahydroisoquinoline using triethylamine and acetonitrile as solvent, to obtain final compounds **3-6** or **7-10**, respectively (Scheme 1).

Affinity of compounds **3-10** for studied serotonin receptors was evaluated by radioligand competitive binding assays using membranes of CHO-K1 or HEK-293 cells transfected with human 5-HT<sub>7</sub> or 5-HT<sub>1A</sub> receptors, and [<sup>3</sup>H]LSD or [<sup>3</sup>H]-8-OH-DPAT, respectively. The affinity constants  $K_i$  calculated from the inhibitory concentration 50 (IC<sub>50</sub>) are shown in Table 1. All synthesized compounds bind the 5-HT<sub>7</sub>R and tetrahydroisoquinoline derivatives (**2-6**) display higher affinity than the corresponding decahydroisoquinoline analogues (**7-10**). Notably, an increase in the number of methylene units in the spacer in compounds **2-5** and **7-9** leads to an improvement of the selectivity against the 5-HT<sub>1A</sub>R. Moreover, compounds **6** and **10** containing a 3-hexenylene spacer exhibit higher 5-HT<sub>7</sub>R affinity ( $K_i = 0.7$  and 2.7 nM, respectively) than a hexylene analogue **4** ( $K_i = 24$  nM). Clearly, **6** and **10** displayed the best profiles of both 5-HT<sub>7</sub>R affinity and 5-HT<sub>7</sub>/5-HT<sub>1A</sub> selectivity (ratio > 1428 and 370, respectively). In these ligands the double bond of the spacer mimics the planar conformation of the carboxamide group of ergotamine (Fig. 1B). Thus, these experimental data suggest that 3-hexenylene spacer, as well as seven-methylene units, trigger the reverse binding mode towards the extracellular entrance. Moreover, the enhanced binding affinity of **6** ( $K_i = 0.7$  nM), relative to **2** ( $K_i = 7$  nM), shows that the indolone moiety optimally binds this cavity. In an attempt to characterize the amino acid residues of the extracellular entrance involved in the recognition of the indolone moiety, we performed molecular dynamics (MD) simulations of the complex between compound **6** and a 5-HT<sub>1BR</sub>-based homology model of the 5-HT<sub>7</sub>R (see Experimental Section and Figure 1C). Importantly, the interaction between the protonated amine and D3.32 remains stable through the simulation time, while the indolone moiety accomplishes key interactions with R6.58 and R7.36 at the extracellular entrance. The fact that the 5-HT<sub>1AR</sub> replaces these charged amino acids by Leu and Ala, respectively, explains the observed selectivity. In detail, E7.35 and Q235<sup>i+4</sup> in ECL 2 maintain R6.58 toward the extracellular entrance to interact with the carbonyl group of the indolone moiety, while the aromatic ring is located between L232<sup>i+1</sup> and L7.32, and forms a cation- $\pi$  interaction with R7.36 (Figure 1D).

**Results: The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.**



**Scheme 1.a Synthesis of Target Compounds 3-10.** a Reagents and conditions: (a)  $\text{K}_2\text{CO}_3$ ,  $\text{CH}_3\text{CN}$ , reflux, 20 h, 30-77%. (b) 1,2,3,4-tetrahydroisoquinoline or ( $\pm$ )-trans-decahydroisoquinoline,  $\text{Et}_3\text{N}$ ,  $\text{CH}_3\text{CN}$ , 60 °C, 24 h, 67-81%.

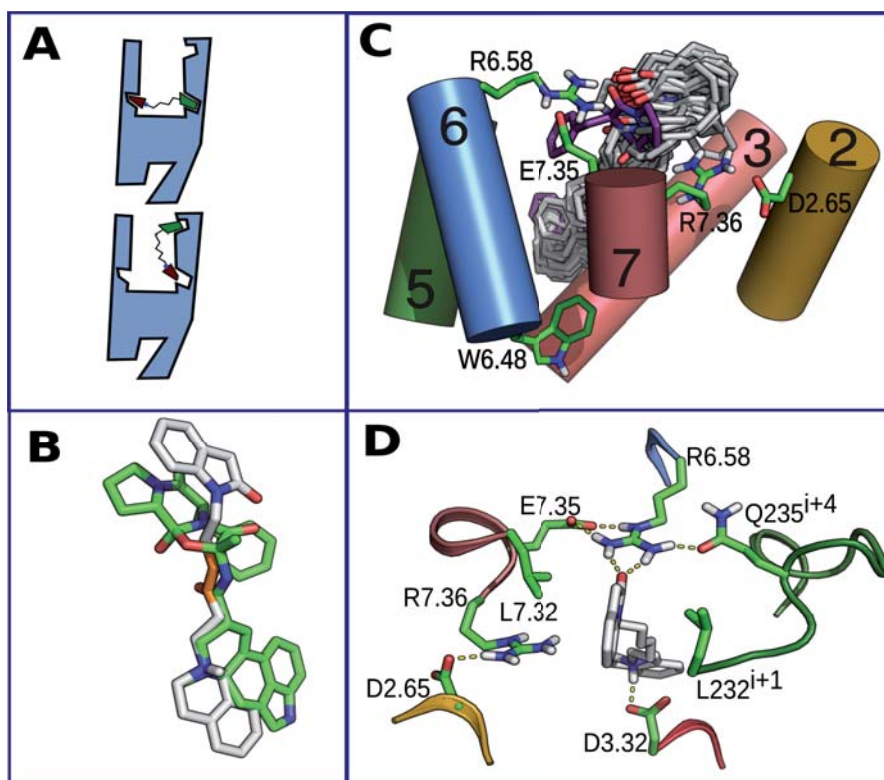
Subsequently, functional characterization of selective ligands **6** and **10** was assessed by evaluating their effect on adenylate cyclase (AC) activity in CHO cells expressing the human 5-HT<sub>7</sub>R. Treatment with the ligand caused a dose-dependent increase in cAMP levels, though a low effect in maximum activation was observed in both cases ( $E_{\text{max}}$  = 21% and 36%, respectively, Table 2). After pre-treatment with serotonin the new compounds induced a dose-dependent decrease of cAMP concentration, and maximum inhibitory effects ( $I_{\text{max}}$ ) of 81 and 66% were attained, respectively (Table 2). The concentration values that produce half of that effect,  $\text{IC}_{50}$ , and the calculated dissociation constants,  $K_B$ , are shown in Table 2. These data

# Development and application of computational techniques to drug discovery and structure-function relationships



indicate that ligands **6** and **10** exhibit in vitro antagonist character at the human 5-HT<sub>7</sub>R, in the AC assay.

Altogether, the newly identified 5-HT<sub>7</sub>R antagonist **6** displays an optimized profile in terms of affinity ( $K_i = 0.7$  nM), selectivity (5-HT<sub>7</sub>/5-HT<sub>1A</sub> ratio > 1428), and functional activity ( $K_B = 84$  nM,  $I_{max} = 81\%$ ).



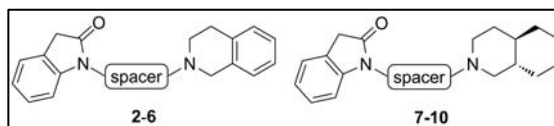
**Figure 1.** MD simulations of the 5-HT<sub>7</sub>R in complex with compound **6**. (A) Potential modes of binding of ligands 2-10. Derivatives with short spacers would bind within the orthosteric site, between TMs 2 and 5(22,23) (top panel), whereas derivatives with large spacers would bind both at the orthosteric site and at the extracellular entrance (bottom panel). (B) Superimposition of ergotamine (in green) and compound **6** (in white). The position of the double bond of **6**, to mimic the planar conformation of the carboxamide group of ergotamine, is shown in orange. (C) The structures of **6** (in white) computed during the simulation (50 structures collected every 2 ns). Superimposition of ergotamine (in purple), as found in the crystal structure of the 5-HT<sub>1B</sub>R, shows that both compounds bind the receptor in a similar manner. (D) Detailed view of the interaction of **6** with D3.32, L232<sup>i+1</sup> (at position *i*+1 relative to the conserved C231<sup>i</sup> engaged in a disulfide bond with C3.25 in TM 3), R6.58, L7.32, and R7.36.

***Results: The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.***

---



# Development and application of computational techniques to drug discovery and structure-function relationships



Compd	Spacer	$K_i \pm \text{SEM (nM)}^a$	
		5-HT <sub>7</sub> <sup>b</sup>	5-HT <sub>1A</sub> <sup>c</sup>
2	(CH <sub>2</sub> ) <sub>4</sub>	$7 \pm 2^d$	$219 \pm 11^d$
3	(CH <sub>2</sub> ) <sub>5</sub>	$28 \pm 6$	$523 \pm 7$
4	(CH <sub>2</sub> ) <sub>6</sub>	$24 \pm 3$	$576 \pm 8$
5	(CH <sub>2</sub> ) <sub>7</sub>	$36 \pm 3$	> 1000
6	(CH <sub>2</sub> ) <sub>2</sub> -CH=CH-(CH <sub>2</sub> ) <sub>2</sub>	$0.7 \pm 0.3$	> 1000
7	(CH <sub>2</sub> ) <sub>4</sub>	$29 \pm 5$	$470 \pm 5$
8	(CH <sub>2</sub> ) <sub>5</sub>	$41 \pm 8$	$491 \pm 3$
9	(CH <sub>2</sub> ) <sub>7</sub>	$62 \pm 7$	> 1000
10	(CH <sub>2</sub> ) <sub>2</sub> -CH=CH-(CH <sub>2</sub> ) <sub>2</sub>	$2.7 \pm 0.7$	> 1000

**Table 1.** 5-HT<sub>7</sub> and 5-HT<sub>1A</sub> Receptors Affinities of New Synthesized Compounds 3-10. *a* Values are the mean of two to four experiments performed in triplicate. *b* 5-CT was used as a reference compound ( $K_i = 1.8 \pm 0.6$  nM). *c* 8-OH-DPAT was used as a reference compound ( $K_i = 1.02 \pm 0.08$  nM). *d* Value from ref 23.

**Results: The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.**



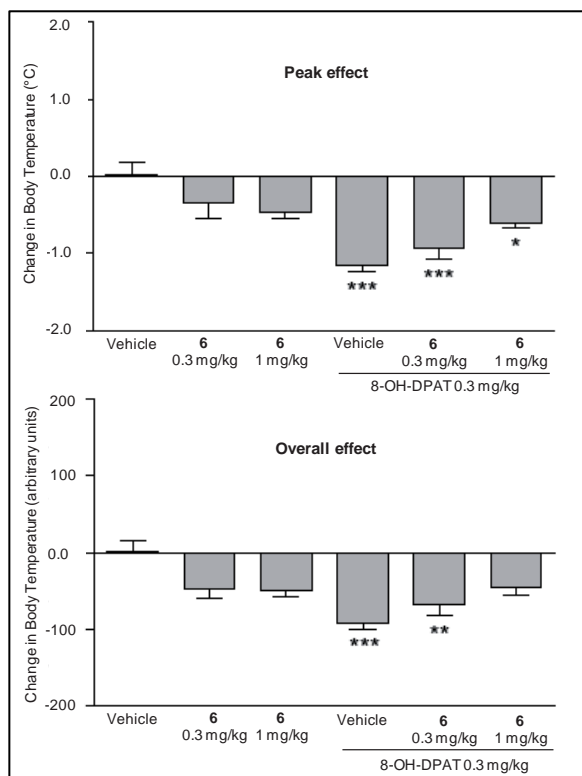
Ligand	EC50 ( $\mu$ M)	E <sub>max</sub> (%)	IC50 ( $\mu$ M)	KB (nM)	I <sub>max</sub> (%)
6	0.23	21	2.2	84	81
10	0.06	36	2.3	88	66
5-HT	0.017	—	—	—	—
Mesulergine	—	—	0.29	11	—

**Table 2. Functional Activity of High-Affinity 5-HT<sub>7</sub>R Ligands**

To further characterize compound 6, we assessed its antagonist character in vivo by determining its effect on hypothermia induced by reference agonist 8-OH-DPAT in mice. Core body temperature was measured using a rectal probe thermometer. A basal value was measured, immediately before any intraperitoneal (ip) injection, and changes in body temperature were determined as peak effect (maximum change during the first 30 min) and overall effect (area under the curve for the total registration period of 2 h). As shown in Figure 2, 8-OH-DPAT (0.3 mg/kg, ip) induced hypothermia, as expected, in contrast to tested compound 6 (0.3 and 1 mg/kg) that had no effect on body temperature by itself. Importantly, the hypothermic effect of 8-OH-DPAT could be counteracted by compound 6 when administered at the dose of 1 mg/kg 30 min before 8-OH-DPAT (Figure 2). It has been previously found that at the dose used (0.3 mg/kg) 8-OH-DPAT acts mainly as a 5-HT<sub>7</sub>R agonist.<sup>32</sup> Therefore, the ability of 6 to inhibit the hypothermic effect of 8-OH-DPAT indicates that the high-affinity ligand identified herein acts as a 5-HT<sub>7</sub>R antagonist in vivo.



# Development and application of computational techniques to drug discovery and structure-function relationships



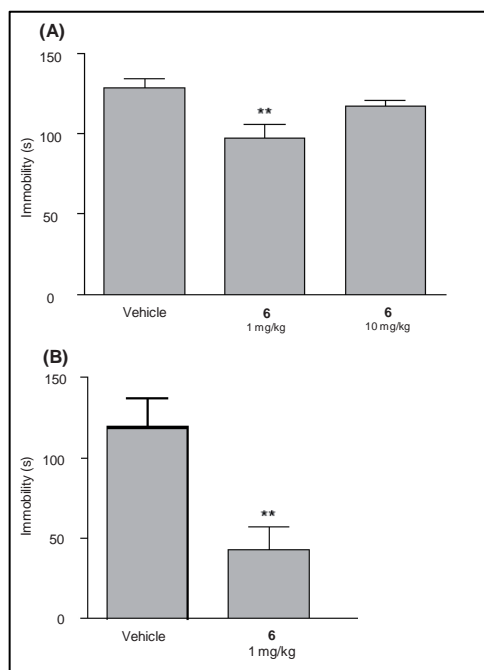
**Figure 2. Peak and overall effects of compound 6 in the 8-OH-DPAT-induced hypothermia assay in mice.** \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ; one-way analysis of variance followed by Bonferroni's post-hoc test.  $N = 5-7$  animals/group.

The pharmacological action of the new 5-HT<sub>7</sub>R antagonist **6** was evaluated in the tail suspension and forced swim tests, two mouse models in which an antidepressant behavior is attributed to a decrease in the immobility of the animal. The tail suspension test was performed for 6 min as previously described,(32,33) and the duration of immobility was determined for the last 4 min of the test. Immobility was defined as the absence of all except respiratory movement of the mouse. Statistical analysis of the results revealed a significant effect for treatment with tested compound **6**, as shown in Figure 3A. Specifically, ip administration of the antagonist at the dose of 1 mg/kg, but not at 10 mg/kg, induced a decrease in the immobility compared to vehicle.

## **Results: The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.**



Likewise, the forced swim test was conducted for 6 min as previously described,(19,32) and the duration of immobility was scored for the last 4 min of the test. In this assay, immobility was defined as the absence of all motion except minor movement required for the mouse to keep its head above the water surface. Again, compound **6** significantly reduced immobility at the dose of 1 mg/kg compared to vehicle (Figure 3B).



**Figure 3. Effects of compound 6 on mouse immobility in the tail suspension test (A) and the forced swim test (B).** \*\* $P < 0.01$ ; one-way analysis of variance followed by Bonferroni's post-hoc test or Student's two-tailed unpaired  $t$ -test, respectively.  $N = 7-9$  animals/group.

Therefore, ip administration of compound **6** at the dose of 1 mg/kg induces a decrease in the immobility of mice in the tail suspension and forced swim tests. The observed effects indicate 5-HT<sub>7</sub>R antagonist-like properties for ligand **6**, as previous studies have shown that inhibition or inactivation of the 5-HT<sub>7</sub>R, using selective antagonists or mice lacking the receptor, leads to reduced immobility in these tests.(12,19,32,34) In addition, it appears that ligand **6** exhibits similar behavior than other 5-HT<sub>7</sub>R antagonists in these in vivo models, where an attenuation of the antidepressant activity was observed after administration of a

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



higher dose.(13,19,34) It should also be pointed that **6** displays a high affinity for the 5-HT<sub>7</sub>R ( $K_i = 0.7$  nM), and is able to inhibit the in vivo hypothermic effect of 8-OH-DPAT. The specificity of antagonist **6** was further confirmed by the assessment of binding affinity toward other depression-related GPCRs such as the serotonin transporter and 5-HT<sub>1B</sub>, 5-HT<sub>2A</sub>, and 5-HT<sub>2C</sub> receptor subtypes, as well as at melatonin MT1 and corticotropin releasing factor CRF1 receptors.<sup>35</sup> Indeed, in all cases compound **6** (at a concentration of 1  $\mu$ M) displaced less than the 15% of the corresponding radioligand binding (see data in Table S1 of the Supporting Information), which supports that the observed antidepressant-like effect of compound **6** is mediated by the 5-HT<sub>7</sub>R.

In conclusion, the promising in vivo properties exhibited by the new high-affinity and selective 5-HT<sub>7</sub>R antagonist described herein reveal its interest as a putative research tool or pharmaceutical in depression disorders.

### **Experimental section**

#### **Computational model of the complex between compound **6** and the 5-HT<sub>7</sub>R.**

MODELLER v9.7(35) was used to build a homology model of human 5-HT<sub>7</sub>R (Uniprot code P34969) using the crystal structure of human 5-HT<sub>1B</sub> (PDB code 4IAR)(30) as template. Compound **6** was docked into the receptor model in such a manner that the protonated amine of the ring forms an ionic interaction with D3.32, while the indolone moiety expands toward the extracellular entrance. This structure was placed in a rectangular box containing a lipid bilayer (183 molecules of POPC) with explicit solvent (14334 water molecules) and a 0.15 M concentration of Na<sup>+</sup> and Cl<sup>-</sup> ions. This initial complex was energy-minimized and subsequently subjected to a 10 ns MD equilibration, with positional restraints on protein coordinates, to remove possible voids present in protein/lipids or proteins/water interfaces. These restraints were released, and two different replicas of 100 ns MD trajectories were produced at constant pressure and temperature, using the particle mesh Ewald method to evaluate electrostatic interactions. Computer simulations were performed with the GROMACS 4.6.3 simulation package,(36) using the AMBER99SB force field as implemented in GROMACS, Berger parameters for POPC lipids, and the general Amber force field

**Results: The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.**



(GAFF) and HF/6-31G\*-derived RESP atomic charges for the ligand. This procedure has been previously validated.(37)

**General Procedure for the Synthesis of Intermediates 11-15.**

To a stirred solution of 1,3-dihydro-2*H*-indol-2-one (0.80 g, 6 mmol) and the appropriate dibromide derivative (8 mmol) in anhydrous acetonitrile (90 mL), K<sub>2</sub>CO<sub>3</sub> (1.66 g, 12 mmol) was added. The reaction mixture was heated at reflux under an argon atmosphere for 20 h, then cooled to room temperature and filtered. The resulting solution was evaporated under vacuum, and the crude material was resuspended in water and extracted with dichloromethane (3 × 30 mL). The organic layers were dried (Na<sub>2</sub>SO<sub>4</sub>), filtered and evaporated under reduced pressure. The obtained residue was purified by column chromatography using the appropriate eluent, to afford pure **11-15**.

**1-[(3*E*)-6-Bromohex-3-en-1-yl]-1,3-dihydro-2*H*-indol-2-one (15).**

Obtained from 1,3-dihydro-2*H*-indol-2-one and (3*E*)-1,6-dibromohex-3-ene in 60% yield. Chromatography: hexane/EtOAc, from 9:1 to 8.5:1.5; IR (ATR)  $\nu$  1711, 1614, 1489, 1466; <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>)  $\delta$  2.40 (app q, *J* = 6.8, 2H, CH<sub>2</sub>CH=), 2.51 (app q, *J* = 6.8, 2H, CH<sub>2</sub>CH=), 3.29 (t, *J* = 7.1, 2H, CH<sub>2</sub>Br), 3.53 (s, 2H, CH<sub>2</sub>CO), 3.75 (t, *J* = 7.2, 2H, CH<sub>2</sub>N), 5.29-5.60 (m, 2H, CH=CH), 6.82 (d, *J* = 7.7, 1H, CH<sub>Ar</sub>), 7.02 (t, *J* = 7.6, 1H, CH<sub>Ar</sub>), 7.22-7.28 (m, 2H, 2CH<sub>Ar</sub>); <sup>13</sup>C NMR (50 MHz, CDCl<sub>3</sub>)  $\delta$  29.7, 30.7, 32.4, 35.8, 39.6 (5CH<sub>2</sub>), 108.7, 122.2, 124.5 (3CH), 124.6 (C), 127.8, 129.3, 129.8 (3CH), 144.5 (C), 175.0 (CO).

**General Procedure for the Synthesis of Final Compounds 3-10.**

To a suspension of the corresponding intermediate **11-15** (0.9 mmol) and 1,2,3,4-tetrahydroisoquinoline or (±)-*trans*-decahydroisoquinoline (1.5 mmol) in anhydrous acetonitrile (4 mL), triethylamine was added (0.2 mL, 1.5 mmol). The reaction mixture was heated at 60 °C under an argon atmosphere for 24 h. Upon cooling to room temperature, the solvent was evaporated under reduced pressure and the crude material was resuspended in water and extracted with dichloromethane (3 × 10 mL). The organic layers were dried (Na<sub>2</sub>SO<sub>4</sub>), filtered and evaporated, and the resulting oil was purified by column chromatography using the appropriate eluent, to provide pure final compounds **3-10**.

The free amine was characterized (yield, IR, NMR, MS), dissolved in anhydrous Et<sub>2</sub>O (6 mL/mmol), and a commercial 1 M HCl(g)/Et<sub>2</sub>O solution (1 mL/mmol) was added. The hydrochloride salt was isolated by filtration or evaporation,

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



washed with anhydrous Et<sub>2</sub>O, dried under high vacuum, and characterized (mp, elemental analysis).

### **1-[(3E)-6-(3,4-Dihydroisoquinolin-2(1H)-yl)hex-3-en-1-yl]-1,3-dihydro-2H-indol-2-one(6).**

Obtained from **15** and 1,2,3,4-tetrahydroisoquinoline in 80% yield. Chromatography: from EtOAc to EtOAc/EtOH, 9:1; mp 109-111 °C; IR (ATR)  $\nu$  1712, 1614, 1489, 1466, 1464; <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>)  $\delta$  2.24-2.32 (m, 2H, CH<sub>2</sub>CH=), 2.34-2.42 (m, 2H, CH<sub>2</sub>CH=), 2.50 (t, *J* = 6.7, 2H, NCH<sub>2</sub>), 2.73 (t, *J* = 5.9, 2H, CH<sub>2</sub>isoquin.), 2.90 (t, *J* = 5.8, 2H, NCH<sub>2</sub>isoquin.), 3.51 (s, 2H, CH<sub>2</sub>CO), 3.63 (s, 2H, NCH<sub>2</sub>isoquin.), 3.75 (t, *J* = 7.3, 2H, CH<sub>2</sub>NCO), 5.50-5.55 (m, 2H, CH=CH), 6.84 (d, *J* = 7.8, 1H, CH<sub>ind.</sub>), 7.00-7.05 (m, 2H, CH<sub>ind.</sub>, CH<sub>isoquin.</sub>), 7.08-7.16 (m, 3H, 3CH<sub>isoquin.</sub>), 7.23-7.31 (m, 2H, 2CH<sub>ind.</sub>); <sup>13</sup>C NMR (50 MHz, CDCl<sub>3</sub>)  $\delta$  29.0, 30.5, 30.8, 35.8, 39.8, 50.8, 56.0, 58.0 (8CH<sub>2</sub>), 108.5, 122.1, 124.5 (3CH), 124.6 (C), 125.6, 126.1, 126.6, 127.3, 127.8, 128.7, 131.0 (7CH), 134.3, 135.0, 144.6 (3C), 175.0 (CO); ESI-MS 347.3 (M+H)<sup>+</sup>. Anal. (C<sub>23</sub>H<sub>26</sub>N<sub>2</sub>O·HCl·H<sub>2</sub>O) C, H, N.

### **In vivo Evaluation of Compound 6.**

Male and female 8-12 weeks-old C57BL/6J mice obtained from the breeding facilities of The Scripps Research Institute were used in all experiments. The animals were group housed and had free access to standard food pellets and water. The experiments were done in accordance with the Guide for the Care and Use of Laboratory Animals as adopted and promulgated by the US National Institutes of Health, and were approved by the Animal Care and Use Committee at TSRI. Every effort was made to reduce the number of animals used and to minimize potential suffering. Compound **6** was dissolved in a small amount of DMSO and then in 0.9% saline. 8-OH-DPAT was dissolved in 0.9% saline. The vehicle 0.9% saline was used as a control. The drugs were administered by single ip injections in the doses indicated. All experiments were started at 9:00 h, and are briefly detailed below.

### **Agonist-Induced Hypothermia Assay.**

Core body temperature was measured using a rectal probe thermometer (Physitemp BAT-7001H, Physitemp Instruments, Clifton, NJ) as previously described.<sup>(33)</sup> A basal value was measured immediately before any injection and measurements were then made 15 and 30 min after injection, and subsequently every 30 min for a total registration period of 2 h. When the antagonistic

***Results: The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.***

---



properties of **6** were evaluated, it was administered 30 min before the agonist (8-OH-DPAT). Changes in body temperature were determined as peak effect (maximum change during the first 30 min) and overall effect (area under the curve for the entire registration period).

**Tail Suspension Test.**(19,32)

An individual mouse was suspended from the tip of its tail attached with a piece of tape to a metal bar placed horizontally 50 cm above the tabletop. The duration of the test was 6 min and behavior was scored by a trained investigator. Duration of immobility was determined for the last 4 min of the test. Immobility was defined as the absence of all except respiratory movement. After the test the animal was returned to its home cage.

**Forced Swim Test.**(19,32) An individual mouse was placed in a clear plastic cylinder with a diameter of 16 cm. The height of the cylinder was 25 cm and it was filled with 10 cm of clear water at 25 °C. The duration of the test was 6 min and behavior was scored by a trained investigator. Duration of immobility was determined for the last 4 min of the test. Immobility was defined as the absence of all movement except minor movement required for the mouse to keep its head above the surface. Afterward the mouse was towel dried and returned to its home cage. The water was replaced between each animal.

***Development and application of computational techniques to drug discovery and structure-function relationships***

---



**Results: The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.**

---



## References

1. Ruat, M., Traiffort, E., Leurs, R., Tardivel-Lacombe, J., Diaz, J., Arrang, J. M., and Schwartz, J. C. (1993) Molecular cloning, characterization, and localization of a high-affinity serotonin receptor (5-HT<sub>7</sub>) activating cAMP formation. *Proc. Natl. Acad. Sci. U S A* **90**, 8547-8551
2. Plassat, J. L., Amlaiky, N., and Hen, R. (1993) Molecular cloning of a mammalian serotonin receptor that activates adenylate cyclase. *Mol. Pharmacol.* **44**, 229-236
3. Hedlund, P. B., and Sutcliffe, J. G. (2004) Functional, molecular and pharmacological advances in 5-HT<sub>7</sub> receptor research. *Trends Pharmacol. Sci.* **25**, 481-486
4. Hedlund, P. B. (2009) The 5-HT<sub>7</sub> receptor and disorders of the nervous system: an overview. *Psychopharmacology* **206**, 345-354
5. Matthys, A., Haegeman, G., Van Craenenbroeck, K., and Vanhoenacker, P. (2011) Role of the 5-HT<sub>7</sub> receptor in the central nervous system: from current status to future perspectives. *Mol. Neurobiol.* **43**, 228-253
6. Gellynck, E., Heyninx, K., Andressen, K. W., Haegeman, G., Levy, F. O., Vanhoenacker, P., and Van Craenenbroeck, K. (2013) The serotonin 5-HT<sub>7</sub> receptors: two decades of research. *Exp. Brain Res.* **230**, 555-568
7. Brenchat, A., Ejarque, M., Zamanillo, D., Vela, J. M., and Romero, L. (2011) Potentiation of morphine analgesia by adjuvant activation of 5-HT<sub>7</sub> receptors. *J. Pharmacol. Sci.* **116**, 388-391
8. Brenchat, A., Zamanillo, D., Hamon, M., Romero, L., and Vela, J. M. (2012) Role of peripheral versus spinal 5-HT<sub>7</sub> receptors in the modulation of pain undersensitizing conditions. *Eur. J. Pain* **16**, 72-81
9. Cifariello, A., Pompili, A., and Gasbarri, A. (2008) 5-HT<sub>7</sub> receptors in the modulation of cognitive processes. *Behav. Brain Res.* **195**, 171-179
10. Roberts, A. J., and Hedlund, P. B. (2012) The 5-HT<sub>7</sub> receptor in learning and memory. *Hippocampus* **22**, 762-771
11. Wesolowska, A., Nikiforuk, A., Stachowicz, K., and Tatarczynska, E. (2006) Effect of the selective 5-HT<sub>7</sub> receptor antagonist SB 269970 in animal models of anxiety and depression. *Neuropharmacology* **51**, 578-586
12. Guscott, M., Bristow, L. J., Hadingham, K., Rosahl, T. W., Beer, M. S., Stanton, J. A., Bromidge, F., Owens, A. P., Huscroft, I., Myers, J., Rupniak, N. M., Patel, S., Whiting, P. J., Hutson, P. H., Fone, K. C., Biello, S. M., Kulagowski, J., and McAllister, G. (2005) Genetic knockout and pharmacological blockade studies of the 5-HT<sub>7</sub> receptor suggest therapeutic potential in depression. *Neuropharmacology* **48**, 492-502
13. Abbas, A. I., Hedlund, P. B., Huang, X. P., Tran, T. B., Meltzer, H. Y., and Roth, B. L. (2009) Amisulpride is a potent 5-HT<sub>7</sub> antagonist: relevance for antidepressant actions in vivo. *Psychopharmacology* **205**, 119-128



## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



14. Cates, L. N., Roberts, A. J., Huitron-Resendiz, S., and Hedlund, P. B. (2013) Effects of lurasidone in behavioral models of depression. Role of the 5-HT<sub>7</sub> receptor subtype. *Neuropharmacology* **70**, 211-217
15. Mnie-Filali, O., Faure, C., Lambas-Senas, L., El Mansari, M., Belblidia, H., Gondard, E., Etievant, A., Scarna, H., Didier, A., Berod, A., Blier, P., and Haddjeri, N. (2011) Pharmacological blockade of 5-HT<sub>7</sub> receptors as a putative fast acting antidepressant strategy. *Neuropsychopharmacology* **36**, 1275-1288
16. Volk, B., Gacsalyi, I., Pallagi, K., Poszavacz, L., Gyonos, I., Szabo, E., Bako, T., Spedding, M., Simig, G., and Szenasi, G. (2011) Optimization of (arylpiperazinylbutyl)oxindoles exhibiting selective 5-HT<sub>7</sub> receptor antagonist activity. *J. Med. Chem.* **54**, 6657-6669
17. Berrade, L., Aisa, B., Ramirez, M. J., Galiano, S., Guccione, S., Moltzau, L. R., Levy, F. O., Nicoletti, F., Battaglia, G., Molinaro, G., Aldana, I., Monge, A., and Perez-Silanes, S. (2011) Novel benzo[*b*]thiophene derivatives as new potential antidepressants with rapid onset of action. *J. Med. Chem.* **54**, 3086-3090
18. Leopoldo, M., Lacivita, E., Berardi, F., Perrone, R., and Hedlund, P. B. (2011) Serotonin 5-HT<sub>7</sub> receptor agents: Structure-activity relationships and potential therapeutic applications in central nervous system disorders. *Pharmacol. Ther.* **129**, 120-148
19. Sarkisyan, G., Roberts, A. J., and Hedlund, P. B. (2010) The 5-HT<sub>7</sub> receptor as a mediator and modulator of antidepressant-like behavior. *Behav. Brain Res.* **209**, 99-108
20. Wesolowska, A., Tatarczynska, E., Nikiforuk, A., and Chojnacka-Wojcik, E. (2007) Enhancement of the anti-immobility action of antidepressants by a selective 5-HT<sub>7</sub> receptor antagonist in the forced swimming test in mice. *Eur. J. Pharmacol.* **555**, 43-47
21. Bonaventure, P., Dugovic, C., Kramer, M., De Boer, P., Singh, J., Wilson, S., Bertelsen, K., Di, J. N., Shelton, J., Aluisio, L., Dvorak, L., Fraser, I., Lord, B., Nepomuceno, D., Ahnaou, A., Drinkenburg, W., Chai, W. Y., Dvorak, C., Sands, S., Carruthers, N., and Lovenberg, T. W. (2012) Translational evaluation of JNJ-18038683, a 5-hydroxytryptamine type 7 receptor antagonist, on rapid eye movement sleep and in major depressive disorder. *J. Pharmacol. Exp. Ther.* **342**, 429-440
22. López-Rodríguez, M. L., Porras, E., Morcillo, M. J., Benhamú, B., Soto, L. J., Lavandera, J. L., Ramos, J. A., Olivella, M., Campillo, M., and Pardo, L. (2003) Optimization of the pharmacophore model for 5-HT<sub>7</sub>R antagonism. Design and synthesis of new naphtholactam and naphthosultam derivatives. *J. Med. Chem.* **46**, 5638-5650

**Results: The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo.**



23. Medina, R. A., Sallander, J., Benhamú, B., Porrás, E., Campillo, M., Pardo, L., and López-Rodríguez, M. L. (2009) Synthesis of new serotonin 5-HT<sub>7</sub> receptor ligands. Determinants of 5-HT<sub>7</sub>/5-HT<sub>1A</sub> receptor selectivity. *J. Med. Chem.* **52**, 2384-2392
24. Pittala, V., and Pittala, D. (2011) Latest advances towards the discovery of 5-HT<sub>7</sub> receptor ligands. *Mini-Rev. Med. Chem.* **11**, 1108-1121
25. Pittala, V., Salerno, L., Modica, M., Siracusa, M. A., and Romeo, G. (2007) 5-HT<sub>7</sub> receptor ligands: Recent developments and potential therapeutic applications. *Mini-Rev. Med. Chem.* **7**, 945-960
26. López-Rodríguez, M. L., Benhamú, B., Morcillo, M. J., Porrás, E., Lavandera, J. L., and Pardo, L. (2004) Serotonin 5-HT<sub>7</sub> receptor antagonists. *Curr. Med. Chem. - C.N.S.A.* **4**, 203-214
27. Venkatakrisnan, A. J., Deupi, X., Lebon, G., Tate, C. G., Schertler, G. F., and Babu, M. M. (2013) Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185-194
28. Gonzalez, A., Cordomi, A., Caltabiano, G., and Pardo, L. (2012) Impact of helix irregularities on sequence alignment and homology modeling of G protein-coupled receptors. *ChemBioChem.* **13**, 1393-1399
29. Wacker, D., Wang, C., Katritch, V., Han, G. W., Huang, X. P., Vardy, E., McCorvy, J. D., Jiang, Y., Chu, M., Siu, F. Y., Liu, W., Xu, H. E., Cherezov, V., Roth, B. L., and Stevens, R. C. (2013) Structural features for functional selectivity at serotonin receptors. *Science* **340**, 615-619
30. Wang, C., Jiang, Y., Ma, J., Wu, H., Wacker, D., Katritch, V., Han, G. W., Liu, W., Huang, X. P., Vardy, E., McCorvy, J. D., Gao, X., Zhou, E. X., Melcher, K., Zhang, C., Bai, F., Yang, H., Yang, L., Jiang, H., Roth, B. L., Cherezov, V., Stevens, R. C., and Xu, H. E. (2013) Structural basis for molecular recognition at serotonin receptors. *Science* **340**, 610-614
31. Gonzalez, A., Perez-Acle, T., Pardo, L., and Deupi, X. (2011) Molecular basis of ligand dissociation in b-adrenergic receptors. *PLoS one* **6**, e23815
32. Hedlund, P. B., Huitron-Resendiz, S., Henriksen, S. J., and Sutcliffe, J. G. (2005) 5-HT<sub>7</sub> receptor inhibition and inactivation induce antidepressant-like behavior and sleep pattern. *Biol. Psychiatry* **58**, 831-837
33. Hedlund, P. B., Kelly, L., Mazur, C., Lovenberg, T., Sutcliffe, J. G., and Bonaventure, P. (2004) 8-OH-DPAT acts on both 5-HT<sub>1A</sub> and 5-HT<sub>7</sub> receptors to induce hypothermia in rodents. *Eur. J. Pharmacol.* **487**, 125-132
34. Wesolowska, A., Nikiforuk, A., and Stachowicz, K. (2006) Potential anxiolytic and antidepressant effects of the selective 5-HT<sub>7</sub> receptor antagonist SB 269970 after intrahippocampal administration to rats. *Eur. J. Pharmacol.* **553**, 185-190
35. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291-325

## ***Development and application of computational techniques to drug discovery and structure-function relationships***

---



36. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435-447
37. Cordomi, A., Caltabiano, G., and Pardo, L. (2012) Membrane protein simulations using AMBER force field and Berger lipid parameters. *J. Chem. Theory Comput.* **8**, 948-958



## 5 Conclusions

This thesis comprises both the development and application of computational techniques. Web applications have become very useful tools for the scientific community as they provide a platform-free ready-to-use interface. This thesis presents two web applications, LigandFinder (virtual screening) and GPCR-SAS (sequence analysis):

### LigandFinder

LigandFinder is a flexible user-friendly web application that allows fast virtual screening to find new (commercially available) compounds similar to a set of compounds of known structure. It explores the chemical space of a database with more than 20M compounds. It has been designed for users with no computational background, who can easily find new compounds with chemical features similar to the input compound(s). To our knowledge, LigandFinder is the first free web service that allows the use of multiple ligands (instead of just one) as input in the 2D virtual screening exercise. Additionally, using the same tool, a pre-computed database of GPCRs ligands have been designed, allowing the users to quickly explore new possible GPCRs ligands.

### GPCR-SAS

For closely related protein families, key structurally conserved and functional regions can be identified from multiple sequence analysis. This web application takes advantage of the structural similarity among GPCR transmembrane regions to perform statistical analysis of key positions or motifs of families A, B, C, and F of GPCRs. GPCR-SAS provides different types of analysis such as the conservation of position(s) or range of positions, as well as the entropy, co-evolutionary (co-variance) and correlation analysis. Additionally, a snakeplot representation with this information can be drawn.

## Conclusions

---



The knowledge of the structure of both the ligand and the target-protein facilitates the drug-discovery process. We have applied structure-based methods to design selective, versus the 5-HT<sub>1A</sub> receptor, ligands for the serotonin 5-HT<sub>7</sub>:

### **The extracellular entrance provides selectivity to 5-HT<sub>7</sub> receptor antagonists with antidepressant-like behavior in vivo**

The less conserved extracellular part of serotonin receptors is critical for ligand selectivity. This work describes the binding mode of a designed selective 5HT<sub>7</sub> receptor antagonist with antidepressant activity opening a gate for a new generation of antidepressant drugs.

The biological function of membrane proteins, triggered or inhibited by drugs, requires conformational rearrangement of the transmembrane regions. These conformational changes require disruption and formation of key inter-helical interactions. Here, we have studied the role of sulfur-containing amino acids (Met and Cys) in forming inter-helical interactions.

### **The role of sulfur-containing amino acids in the structure and function of membrane proteins**

7TMRs feature several examples of Met/Cys-aromatic and Met-Met interactions that are, in some cases, responsible for important pharmacological, signaling or functional events. In the present review we have described many examples of such interactions occurring between natural or synthetic ligands and their receptors, and between two or more residues within the receptor

The analysis of the inter-residue interactions in crystal structures of membrane protein reveals that Met and Cys often interact with Leu, Ile, Val, Phe, and other Met or Cys residues. The characterization of their strength using *ab-initio* calculations in small-molecule model systems, predicts that Met-Met, Met-Phe, Cys-Phe, Met-aliphatic and Cys-aliphatic interactions are stronger in magnitude than aliphatic-aliphatic interactions. Remarkably, Met-Met, Met-Phe, and Cys-Phe interactions are stronger than aromatic-aromatic interactions.

## *Conclusions*

---



## *Conclusions*

---

