UAB

Universitat Autònoma de Barcelona

STRUCTURE AND FUNCTION OF GPCRs

MIREIA JIMÉNEZ ROSÉS

MAY 2019

UAB

Universitat Autònoma de Barcelona

Doctoral Thesis

Structure and Function of GPCRs

Mireia Jiménez Rosés

May 2019

This work has been carried out at Laboratory of
Computational Medicine, Biostatistics Unit of the Medicine
Faculty, in order to obtain the Degree of Doctor in
Biochemistry, Molecular Biology and Biomedicine

Directors:
Dr. Leonardo Pardo   Dr. Arnau Cordomí

Tutor:
Dra. M. Victòria Noguès

ABSTRACT

G protein-coupled receptors (GPCRs) are the largest and most diverse superfamily of transmembrane proteins in Eukaryotes. GPCRs transduce a huge variety of exogenous and endogenous signals such as photons, hormones or neurotransmitters to initiate biological responses in the cell interior. Therefore, they are very interesting therapeutic targets.

This Doctoral Thesis focusses on the understanding of the structure and function of GPCRs, by applying computational chemistry techniques such as homology modelling, docking and molecular dynamics simulations. Particularly, the thesis addresses the structural determinants associated to the activation mechanism, the regulation by allosteric modulators, the oligomerization with other GPCR or additional proteins and the coupling to transducers (G proteins or arrestins).

ACKNOWLEDGMENT

In almost four years, there have been good and bad moments, including periods of motivation and disappointment, but finally, the elaboration of this thesis has been possible. For this reason, I would like to express my gratitude to all the people who have accompanied me throughout this personal and professional learning process.

First of all, I want to thanks my thesis directors Arnau and Leonardo for giving me the opportunity to do the PhD. They have guided me in the discovery of the research with computational methods and, of course, the GPCRs' world.

Thanks to those who were or are part of the LMC group for all time we spent together, not only sharing scientific knowledge, but also talking about anything, listening to my complaints, having good advice, laughing…

I would like to thank Laura, Gary and Herman for giving me the opportunity to make a stay in Janssen Pharmaceutica (J&J in Belgium), where I have learned a lot both personally and professionally.

Thanks also to all my friends for helping me to make this process more bearable. And particularly to Albert, Alba, Alicia, Jara, Lorena, Marc and Vicky for all *Budas*' dinners; Jose, Sergi and company for organizing our cinema, board games, escape rooms… evenings; and Oscar&co and my teammates for all the hours playing padel.

Finally, I would like to thanks my family for supporting me, day-to-day, in all my decisions. Without you, I would have never been possible to start my career as a researcher.

TABLE OF CONTENTS

PART 1.

INTRODUCTION

1. INTRODUCTION

1.1. G PROTEIN-COUPLED RECEPTORS

G protein-coupled receptors (GPCRs), also known as seven transmembrane helical receptors (7TMRs), are the largest and most diverse superfamily of membrane proteins in Eukaryotes. It is estimated that the human genome encodes more than 700 functional GPCRs, which means approximately 3% of the total genome (Fredriksson *et al.*, 2003).

The first three-dimensional (3D) crystallographic structure of a GPCR, the bovine rhodopsin at 2.8Å, came in 2000 (Palczewski *et al.*, 2000). It is not until 2007 that a structure of another receptor, the β_2-adrenergic receptor, came out (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2007). Since then, the number of crystal structures of GPCRs (and other membrane proteins) has been growing due to advances in crystallization and high-resolution X-ray and more recently also in cryogenic electron microscopy (cryo-EM) techniques (Figure 1.1) (Munk *et al.*, 2019). These advances in GPCRs crystallization include stabilization via *chimeric proteins* such as T4 lysozyme (T4L) or BRIL in place of intracellular loop 3 (ICL3) or at N-terminus, thermostabilization by *point mutations*, use of *high affinity ligands* and/or a *nanobody*, which is a single domain of the antigen-binding fragment (Fab fragment) of antibodies, which mimic G proteins (Chun *et al.*, 2012; Trzaskowski *et al.*, 2012; Manglik, Kobilka and Steyaert, 2017). Cryo-EM is suitable for the structure determination of GPCRs–G protein complexes (more than 64 kDa), and these constructs require minimal stabilization modifications compared with crystallography (Renaud *et al.*, 2018).

To date, according to the GPCRdb (Pándy-Szekeres *et al.*, 2018; Munk *et al.*, 2019) (http://gpcrdb.org/structure/statistics), 321 GPCR crystal structures (Figure 1.1) (62 of unique receptors) are deposited in the Protein Data Bank (Berman *et al.*, 2000). Available crystal structures include receptors from classes A (285 structures for 52 unique receptors), B (16 structures for 6 receptors), C (8 structures for 2 receptors) and F (12 structures for 2 receptor).
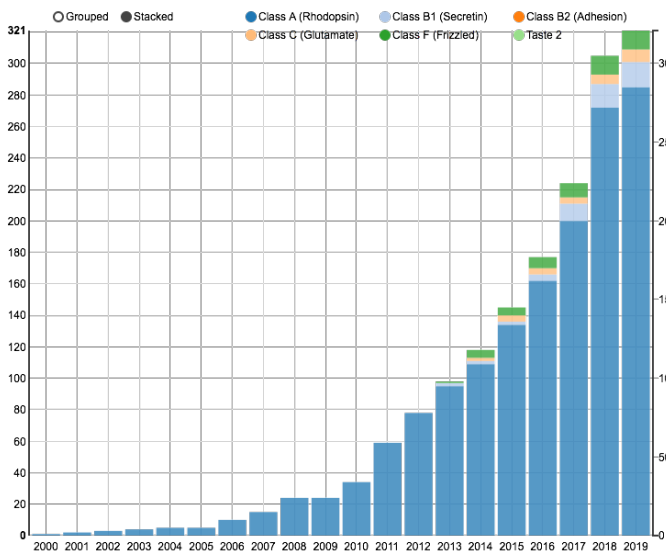
Figure 1.1| **Evolution of the available crystal structures along the past 19 years.** Figure taken from GPCRdb (http://gpcrdb.org/structure/statistics) (Pándy-Szekeres *et al.*, 2018; Munk *et al.*, 2019).

GPCRs transduce a variety of exogenous and endogenous signals as photons, odours, pheromones, hormones, peptides, neuro-transmitters and ions, etc. towards the interior of the cells (Palczewski and Orban, 2013). For this reason, they are fundamental in biological/physiological processes –from vision, smell and taste to

4

neurological, cardiovascular, endocrine and reproductive functions–required to maintain cellular homeostasis and to coordinate cellular activity. This makes them highly interesting therapeutic drug targets for their importance in the organism and also, because they contain not only endogenous ligand-binding (orthosteric), but also allosteric modulatory (allosteric) sites (see section *1.4.2. Allosteric modulators)*, both accessible to pharmacological agents (Wootten, Christopoulos and Sexton, 2013).

About 30-40% of all modern drugs act on approximately 40 GPCRs and it is estimated that about 400 receptors are potentially druggable (portion of the human genome that is susceptible to pharmacological interaction and simultaneously involved in pathological mechanisms leading to disease) (Wise, Gearing and Rees, 2002; Hauser *et al.*, 2017). Moreover, GPCRs account for approximately 19% of the 2000 and 3000 genes in the genome that are established drug targets (Russ and Lampel, 2005; Rask-Andersen, Masuram and Schiöth, 2014).

1.2. STRUCTURE OF GPCRs

Despite the very low sequence identity (SI) in the TM domains specially in the extracellular part (Gonzalez *et al.*, 2012), all GPCRs' share the common architecture, that is characterized by the presence of seven transmembrane (7 TMs) α-helices bundle linked by an extracellular N-terminal domain, three intracellular loops (ICL1 to ICL3), three extracellular loops (ECL1 to ECL3) and an intracellular C-terminus usually containing an α-helix (H8). In addition, they present a characteristic disulphide bridge between Cys residues at ECL2 and TM3 (Figure 1.2).

This α-helical TM bundle structure is one of the typical folds (together with β-barrels) observed in membrane proteins which satisfies the physical constraints imposed by the lipid bilayer, and allow strong conservation of the TM structure even at low sequence identity (<20%) (Olivella *et al.*, 2013).
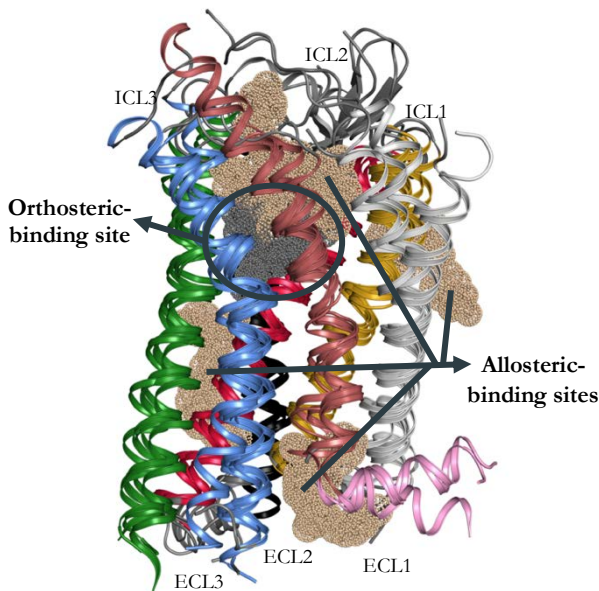
Figure 1.2 | Location of orthosteric- and allosteric-binding sites in some available crystal structures of class A. Cartoon representation shows TM1 (white), TM2 (yellow), TM3 (red), TM4 (grey), TM5 (green), TM6 (dark blue), TM7 (brown) and H8 (pink); grey dots represents the orthosteric-binding site and wheat dots are the allosteric-binding sites.

Superposition of the known crystal structures of class A GPCRs (Figure 1.2) shows that ***orthosteric ligand-binding site*** is located within the cavity formed in the extracellular side of the TM bundle (TMs 3, 5, 6 and 7), although each ligand penetrates to different

depths. It is now recognized that some GPCRs possess topographically distinct binding sites, known as **_allosteric-binding sites_**, which represent a promising opportunity to modulate the receptor function, potentiate a differential selectivity and improve the safety against traditional orthosteric ligands.

In recent years, the repertoire of GPCRs crystal structures with allosteric modulators has largely increased (more than 20 structures, see Table 1.1). The reported allosteric-binding sites are located at the extracellular side (entrance of the orthosteric site), the intracellular side (at the G protein- and β-arrestin-binding site) and the lipid-facing exterior of the 7TM helix bundle.

TABLE 1.1| **Solved crystal structures of GPCRs in complex with small-molecule allosteric modulators.** Adapted from Congreve, Oswald and Marshall, 2017; and Lu and Zhang, 2018.

GPCR	released year	PDB code	resolution (Å)	allosteric modulator	allosteric-binding site
class A					
CCR5	2013	4MBS	2.71	maraviroc	7TM helical bundle, extracellular side
FFAR1	2014	4PHU	2.33	TAK-875	outside 7TM helical bundle, extracellular side
P2Y1	2015	4XNV	2.30	BPTU	outside 7TM helical bundle, intracellular side
CCR2	2016	5T1A	2.81	CCR2-RA-[R]	7TM helical bundle, intracellular side
CCR9	2016	5LWE	2.80	vercirnon	7TM helical bundle, intracellular side
M2	2016	4MQT	3.70	LY2119620	7TM helical bundle, extracellular side
FFAR1	2017	5TZR	2.20	MK-8666	outside 7TM helical bundle, extracellular side
FFAR1	2017	5TZY	3.22	MK-8666, AP8	outside 7TM helical bundle, extracellular side
β2AR	2017	5X7D	2.71	Cmp-15PA	7TM helical bundle, intracellular side
PAR2	2017	5NDZ	3.60	AZ3451	7TM helical bundle, extracellular side
class B					
CRF1R	2013	4K5Y	2.98	CP-376395	7TM helical bundle, intracellular side
GCGP	2015	5EE7	2.50	MK-0893	outside 7TM helical bundle, intracellular side
GLP-1R	2017	5VEW	2.70	PG-06372222	outside 7TM helical bundle, intracellular side
GLP-1R	2017	5VEX	3.00	NNC0640	outside 7TM helical bundle, intracellular side
class C					
mGLU5	2014	4OO9	2.60	mavoglurant	7TM helical bundle, extracellular side
mGLU5	2014	4OR2	2.60	FITM	7TM helical bundle, extracellular side
mGLU5	2015	5CGD	2.60	HTL14242	7TM helical bundle, extracellular side
mGLU5	2018	6FFH	2.70	Fenobam	7TM helical bundle, extracellular side
mGLU5	2018	6FFI	2.20	M-MPEP	7TM helical bundle, extracellular side
class F					
SMO	2016	5L7I	3.30	vismodegib	7TM helical bundle, extracellular side

It has been shown that this evolutionarily conserved structural scaffold of non-covalent contacts for the GPCRs fold is due to a consensus network of inter-TM contacts, which are localized at the central and cytoplasmic side of the TM bundle, mainly at the interfaces of TM1-TM2, TM3-TM4, TM3-TM5 and TM3-TM6-TM7 (Venkatakrishnan *et al.*, 2013) (Figure 1.3).
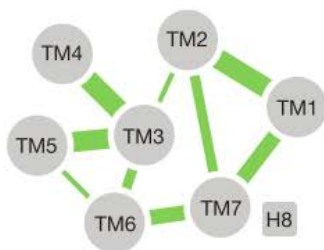
Figure 1.3 | Schematic representation of the consensus scaffold of non-covalent contacts in GPCRs. TM helices are represented as circles and H8 as a square. The lines indicate the presence of contacts and the thickness of the line is proportional to the number of contacts. Adapted from Venkatakrishnan *et al.*, 2013.

Furthermore, GPCR sequences contain at least one highly conserved amino acid in each helix (Mirzadegan *et al.*, 2003): Asn in TM1 (98%), Asp in TM2 (93%), Arg in TM3 (95%), Trp in TM4 (96%), Pro in TM5 (76%), Pro in TM6 (98%) and, Pro in TM7 (93%). These conserved amino acids, arbitrarily assigned to 50, were used by Ballesteros and Weinstein (Ballesteros and Weinstein, 1995) to define a general numbering scheme for GPCRs that consists in two numbers (*amino acid* $^{X.Y}$): the first number (X) corresponds to the helix (1 to 7) in which the amino acid is located and, the second number (Y) indicates its position relative to the amino acid 50, decreasingly the number toward the N-terminus and increasingly toward the C-terminus.

1.3. CLASSIFICATIONS OF GPCRs

Numerous classifications schemes have been proposed to sort out the superfamily of GPCRs.

The first classification, the **A-F CLASSIFICATION SYSTEM**, which is based in sequence homology, was proposed by Kolakowski (Kolakowski, 1994) and classifies receptors in six *classes* (Figure 1.4):

- **Class A** (*rhodopsin-like family*) includes > 700 members with 197 receptors with known ligands, >400 olfactory receptors and 87 orphans. This class is the largest and for that reason, it is sub-classified according the nature of the ligand in: alicarboxylic acid, aminergic, lipid, melatonin, nucleotide, peptide, protein, sensory, steroid, orphan and other receptors.

- **Class B** consist of a large N-terminal extracellular domain (ECD) and the GPCR common structure (7TMs domain and intracellular domain), and are divided in:
  - **Class B1** (*secretin receptor family*) includes 15 receptors with known ligands and 26 orphans. This class include the calcitonin receptors, corticotropin-releasing factor receptors, glucagon receptor family, parathyroid hormone receptors and vasoactive intestinal peptide, pituitary adenylate cyclase-activating polypeptide and growth-hormone-releasing hormone receptors.
  - **Class B2** (*adhesion receptor family*) includes 34 receptors, which possess a large extracellular N-termini that autoproteolytically cleaved from their 7TMs domains.

- **Class C** (*metabotropic glutamate/ pheromone receptors family*) include 12 receptors (calcium-sensing receptor, $GABA_B$ receptors and

metabotropic glutamate receptors) with known ligands and 8 orphans. They are composed of an exceptionally large extracellular domain, which include the Venus flytrap module and a cysteine rich domain, and the 7TMs domain.

- **Class F** (*frizzled and smoothened*) include 11 receptors with known ligands: 10 frizzled receptors (FZD$_{1-10}$) and smoothened receptor (SMO).

These are the classes present in vertebrates; **class D** (fungal mating pheromone receptors) and **class E** (cyclic AMP receptors) are only in invertebrates.

Class A
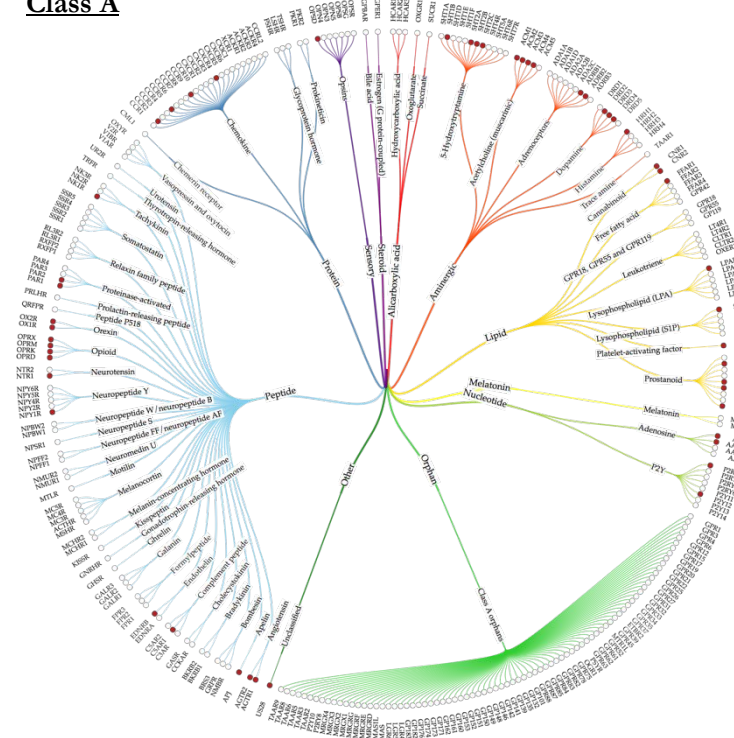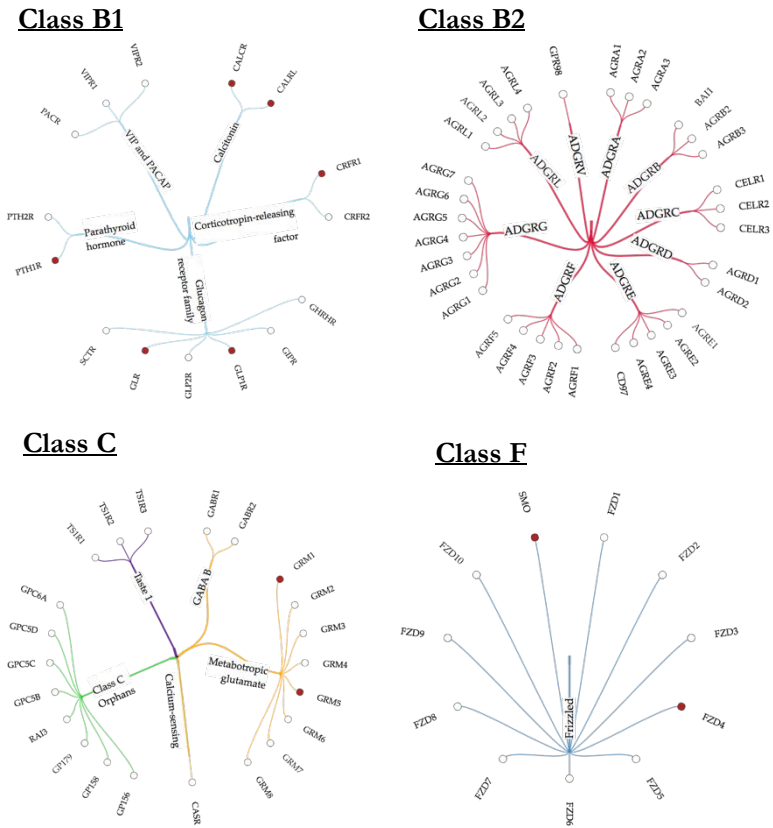
Figure 1.4| Structural coverage of the GPCRs class A, B1, B2, C and F. The red dot indicated the receptors that have been crystallized. Figures taken from GPCRdb (Pándy-Szekeres *et al.*, 2018; Munk *et al.*, 2019).

In 1999, Bockaert and Pin introduced a similar but extended nomenclature system, based on sequence homology and structural and ligand-binding criteria. This system is divided in five families, known as ***FAMILY 1-5 SYSTEM*** (Bockaert and Pin, 1999): **family**

1 which include *small ligands* like catecholamines and short peptides; **family 2** which are activated by *large peptides* like glucagon or secretin; **family 3** comprises the *metabotropic glutamate receptors* (mGluRs) and the *GABA_B receptor*; **family 4** includes *pheromone receptors* (also known as vomeronasal -VNs-) and **family 5** includes *frizzled* and *smoothened receptors.*

In 2003, Fredriksson and colleagues, on the basis of a phylogenetic study, classified 802 human GPCRs into five main families: **Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2** and **Secretin**, shortened with their acronym in **GRAFS SYSTEM** (Fredriksson *et al.*, 2003). The <u>*Rhodopsin family*</u> is the largest family and, according to Fredrikson *et al.,* it can be further divided into four main branches:

- − **α branch**: prostaglandin, amine, opsin, melatonin and MECA receptors (which include *Melanocortin, Endothelial, Cannabinoid and Adenosine receptors*),
- − *β* **branch:** most peptide receptors,
- − *γ* **branch:** SOG (for *Somatostatin, Opioid and Galanin*) receptors, melanin-concentrating hormone receptors and chemokine receptors, and
- − *δ* **branch:** MAS-related receptors, glycoprotein receptors, purin receptors and olfactory receptors.

The *International Union of Pharmacology, Committee on Receptor Nomenclature and Classification* (NC-IUPHAR) (Foord *et al.*, 2005; Alexander *et al.*, 2017) (see http://www.guidetopharmacology.org/) recommends to use this A-F system classification.

1.4. ALLOSTERIC NATURE OF GPCRs

GPCRs are considered allosteric proteins in different senses (Monod, Wyman and Changeux, 1965; Thal *et al.*, 2018; Weis and Kobilka, 2018). This includes the *allosteric mechanism linking agonist binding to G protein activation* (activation mechanism*)*, the *allosteric modulation* by molecules and also allosteric modulation within protomers via *dimerization/oligomerization*.

1.4.1. ALLOSTERIC MECHANISM LINKING AGONIST BINDING TO G PROTEIN ACTIVATION

GPCRs have evolved to transmit external environmental signal from the extracellular part of protein to their intracellular part, to produce a response through a transducer (which include *G proteins*, *G protein-coupled receptor kinases* -GRKs-, *arrestins…*) (see *1.4.1.2. Signal transduction of GPCRs*). Therefore, **GPCR signal transduction** involves allosteric signalling with long-range communication between two spatially distinct binding sites, the orthosteric-binding site and transducers-binding site (Bhattacharya and Vaidehi, 2014).

Most GPCRs are *constitutive active*, which means that they exhibit a certain level of basal activity without agonist binding, the apo form of the receptor. This is thought to be caused by thermal excitation of the environment providing enough energy to surmount the energy barrier to change from one state to other. This implies the existence of **different dynamic conformational states in GPCRs** (Figure 1.5), where each conformational state correspond to a minima in the free energy landscape: **inactive state** (R), **intermediate states** (R' -without ligand- and R" -with agonist ligand bound) and **active state** (R* -with agonist and transducer bound) (Miao and McCammon,

2016; Latorraca, Venkatakrishnan and Dror, 2017; Weis and Kobilka, 2018).
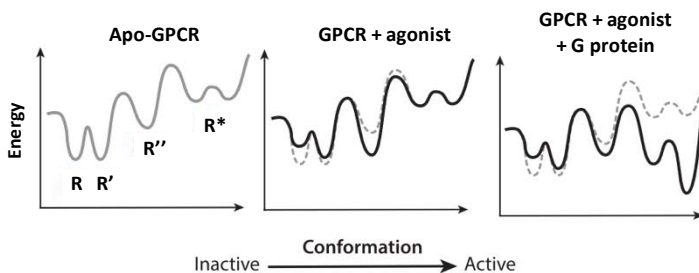
Figure 1.5| **Schematic illustration of the conformational dynamics occurring in GPCRS.** The grey lines (solid and dashed) indicated the energy landscape of the apo-state, and the solid black line indicated the effect of the agonist ligand and G protein. Adapted from Weis and Kobilka, 2018**.**

The notion of an equilibrium between conformational states and the basal activity is essential for understanding the basic pharmacology of GPCRs ligands. **Orthosteric ligands** can be grouped into classes according to its efficacy (Figure 1.6): **_full agonist_** that maximally stimulate the receptor, **_partial agonist_** that are unable to elicit fully stimulation even at saturating concentration, **_antagonist_** that occupy the orthosteric binding site but do not affect basal activity, and **_inverse agonist_** that reduce the basal or constitutive activity.
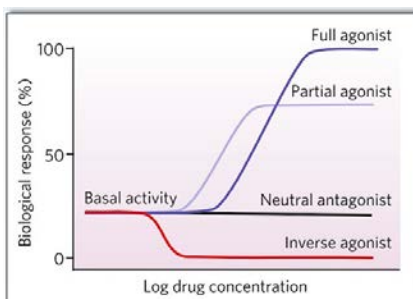
Figure 1.6| **Classification of ligand efficacy for GPCRs.** Adapted from Rosenbaum et al. (2009)**.**

Differences in both the type and the strength of chemical interactions between distinct ligands and an individual GPCR can affect ligand residence times or drive changes in receptor conformation (for example, differential phosphorylation patterns), which determine greater efficacy toward engaging a specific transducer (different G protein subtypes or arrestins). This is known as <u>biased signalling or functional selectivity</u>. The ability of some ligands to stimulate various pathways may be responsible for many of the undesired effects of drugs targeted to GPCRs.

1.4.1.1. *Structural insights of the activation mechanism*

Early insight regarding the mechanism of GPCRs activation and signal transduction came from the active structures of rhodopsin (Standfuss *et al.*, 2011) and of the β_2 adrenergic receptor in complex with nanobody (Rasmussen *et al.*, 2011) or with Gs protein (Rasmussen *et al.*, 2011), both members of the *Rhodopsin*-like receptors family.

These structures indicate a common evolutionary origin of the **activation mechanism** in GPCRs. When the agonist ligand bound to the GPCR, it affect mainly some residues at TM3 and TM6, starting the common movements in the extracellular part of GPCRs (Figure 1.7): the inward movement of TM5, the slight rotation and upward movement of TM3, the rotation of TM6 and the inward movement of TM7 and TM1. And it also lead the opening of intracellular cavity that involves an outward movement of helices TM6 (about 10Å) and TM5 (about 2Å), for the subsequently transducer binding (Tehan *et al.*, 2014; Manglik and Kruse, 2017).
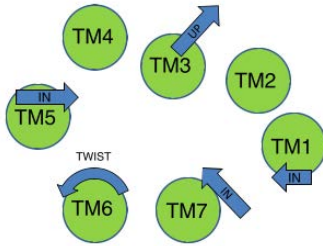
Figure 1.7| **Overview of the conformational changes seen in the extracellular part of GPCR during the activation mechanism.** Adapted from Tehan et al. (2014).

These rearrangement of the helices occur through series of changes in residue interactions called **_molecular switches_** (Trzaskowski *et al.*, 2012; Lee, Choi and Hyeon, 2015):

- **Transmission switch** (Sansuk *et al.*, 2011) involves the conserved residues Ile/Leu/Val/Met$^{3.40}$, Pro$^{5.50}$, Phe$^{6.44}$ and Trp$^{6.48}$ (Figure 1.8). The movement of TMs 3 and 6 could be caused by the rearrangement of conserved residues Phe$^{6.44}$ and Trp$^{6.48}$ (of the *FxxCWxP motif* in TM6) allowing the upward movement of TM3 along its axis (Ile/Leu/Val/Met$^{3.40}$ moves away from Pro$^{5.50}$). The movement of Trp$^{6.48}$, which forms an hydrophobic barrier between TMs 2, 3 and 6, open a gate that allows diffusion of water molecules from the bulk phase towards the internal space of the receptor (Yuan *et al.*, 2015).
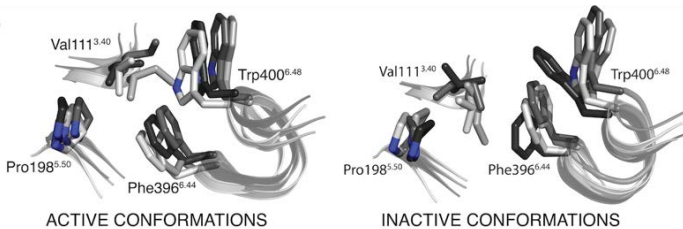
Figure 1.8| **Comparison of the transmission switch in several crystal structures.** It includes active and inactive crystal structures: M$_2$ acetylcholine receptor, A$_{2A}$ adrenergic receptor, β2 adrenergic receptor and rhodopsin.

- **Ionic lock.** This was one of the first molecular switches studied. The breaking of the highly conserved residue Arg[3.50] of the *E/DRY motif* in TM3 and an acid residue in position 6.30 (in TM6) may facilitate the movement of TMs 3 and 6 (Figure 1.9). This switch is critical for Rhodopsin, but it is uncertain if it is an important switch in GPCRs in general.

Figure 1.9| Diagram of the ionic lock in inactive and active states. Adapted from Lee et al. (2015).

- **Tyrosine cluster**. The upward movement of TM3 is partially stabilized by the optimal hydrophobic packing with the highly conserved Leu[2.46], which occupy the space of Asn[7.49] (NPxxY motif in TM7), and Asn[7.49] moves to interacts with Asp[2.50] and the water channel within the GPCR. This movement enables Tyr[7.53] to move into the cytoplasmic cleft between helices TM3 and form a hydrogen bond with Tyr[5.58] via a water molecule.

1.4.1.2. *Signal transduction of GPCRs*

Activation of GPCRs, mainly with the outward movement of TM6, exposes an intracellular pocket that can effectively couple signal transducers (G protein, GRKs and arrestins) to form functional signal complexes. The signal transduction of GPCRs can be divided in (Hilger, Masureel and Kobilka, 2018) (Figure 1.10):

- **G protein-dependent signalling pathways** when the heterotrimeric *guanine nucleotide-binding protein* (commonly known as *G protein*) is the transducer. G proteins are composed

of three subunits: alpha (Gα), beta (Gβ) and gamma (Gγ) subunits; and the Gβ and Gγ subunits form a stable dimeric complex, known as Gβγ. When Gα bound GDP (guanosine diphosphate), it associates with Gβγ dimer to form the inactive heterotrimer. GPCR activation promotes the engagement of the inactive heterotrimer that accelerates GDP dissociation and facilitates rapid GTP (guanosine triphosphate) binding which undergoes Gα conformational changes that result in the dissociation of the Gα and Gβγ subunits. Both subunits modulates the activity of different effector proteins (including adenylyl cyclases, cGMP phosphodiesterase, phospholipase C…) to modulate the second messengers (for example: cAMP, Ca^{2+}, K^+, diacylglycerol…). Finally, the cellular response is terminated when the Gα subunit hydrolyses GTP to GPD (GTPase activity) and reassociates with Gβγ subunit.

There are many classes or subtypes of Gα subunits (Lodish H, 2000; Offermanns, 2003):

- $Gα_s$ mainly activates the cAMP-dependent pathway and the influx of Ca^{2+} by stimulating the adenylyl cyclase and Ca^{2+} channels, respectively.

- $Gα_t$ stimulates the production of cGMP by the stimulation of the cGMP phosphodiesterase.

- $Gα_{i/o}$ inhibits the adenylyl cyclase (inhibition of the production of cAMP) and Na^+ channels (efflux of Na^+ changing the membrane potential). But it also stimulates phospholipase C (PLCs) which hydrolyse the phosphatidylinositol 4,5-bisphosphate (PIP_2) into

inositol 1,4,5-trisphosphate (IP$_3$) and diacylglycerol (DAG).

- – G$\alpha_{q/11}$ stimulates PLCs to product IP$_3$ and DAG.
- – G$\alpha_{12/13}$ are involved in Rho family GTPase signalling, which are involved in control of cell cytoskeleton remodelling.

G$\beta\gamma$ signalling is diverse and may results in activation of G protein-regulated inward rectifier K$^+$ (GIRK) channels, G protein-coupled receptor kinases (GRKs), PLCs, phosphoinoisite-3-kinase (PI-3-K) and voltage-dependent Ca^{2+} channels (VDCC) or inhibition of adenylyl cyclase.

- • **G protein-independent signalling pathways** when GPCRs signal through other transducers such as:
    - – *G protein-coupled receptor kinases* (GRKs) family (Komolov and Benovic, 2017) is formed by 7 members of serine/threonine protein kinases (GRK1-GRK7) that phosphorylate serine and threonine residues at intracellular loops (ICLs) and C-terminal of the activated GPCR, that will act as binding sites for arrestins.

    - – *Arrestins* (Gurevich and Gurevich, 2019) is a family formed by four subtypes: *arrestin-1* (also known as *visual rod arrestin* which binds light-activated phosphorylated rhodopsin), *arrestin-2* (*β-arrestin-1*), *arrestin-3* (*β-arrestin-2*) and arrestin-4 (*visual cone arrestin* which binds light-activated phosphorylated cone opsins)*, where 2 and 3 are ubiquitously. After the GPCR phosphorylation mediated by GRKs, arrestin undergoes several

conformational changes, its activation, which expose positively charged residues to interact with the phosphorylated residues. This process is the most stablished function of arrestins, the GPCR desensitization, which is the suppression of GPCR coupling to G proteins by direct competition. But the active arrestin must promote or suppress the binding of other effectors proteins: tyrosine kinase SRCs, cytokine receptors -JAK/STATs, mitogen-activated protein kinase -MAPK or MAP kinase which are pathways involved in regulation of cell functions, or clathrin, clathrin adaptor AP2… which are part of the internalization of the GPCR.
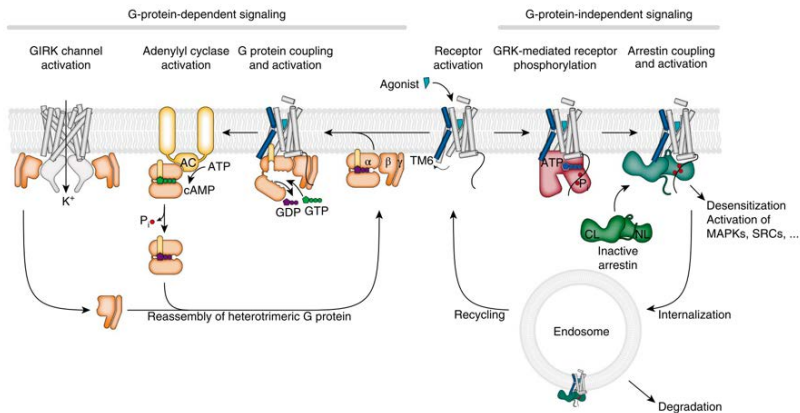
Figure 1.10| Scheme of GPCRs signal transduction. Intracellular transducers include G proteins (orange), GRKs (red) and arrestins (green: inactive arrestin and teal: active arrestin). Effector proteins included are adenylyl cyclase (AC, yellow) and G protein-coupled inwardly rectifying K^+ channels (GIRK, light grey). Taken from Hilger, Masureel and Kobilka, 2018.

1.4.2. ALLOSTERIC MODULATORS

Apart from the orthosteric ligands, there are certain ligands known as **allosteric ligands** or **allosteric modulators**, which bind to a different binding sites. Allosteric modulators can increase or decrease the affinity and/or efficacy of the orthosteric ligand. So, pharmacologically, they can be divided in two groups (Figure 1.11): *positive allosteric modulators* (**PAMs**) that enhance the response of the orthosteric ligand and *negative allosteric modulators* (**NAMs**) that inhibit it.
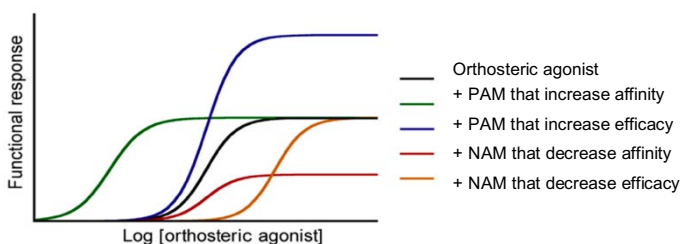
Figure 1.11| Functional responses of model allosteric modulators. Adapted from Brogi et al. (2014).

Some allosteric modulators have *intrinsic activity* (have a functional effect in the absence of orthosteric ligand), the **ago-PAMs**. Ligands that bind at the orthosteric-binding site and at the allosteric-binding site are known as **bitopic molecules**.

Several endogenous species (ligands, proteins and ions) can also act as allosteric modulators. For example, *sodium* is known to disfavour agonist binding and activation at many GPCRs (Katritch *et al.*, 2014); *lipids*, the lipidic composition of the GPCRs environment -the membrane composition- would modulate the kinetics of ligand response (Dawaliby *et al.*, 2016) and *G proteins* or *G protein mimetic*

(nanobody) stabilizes the active conformation and enhance agonist-binding affinity (DeVree *et al.*, 2016). Allosteric modulators are attractive for pharmaceutical companies as a strategy to achieve selectivity between receptor subtypes with nearly identical orthosteric sites, and reduce the side effects.

1.4.3. OLIGOMERIZATION OF GPCRs

GPCRs have been traditionally described as monomeric transmembrane (TM) receptors that form a ternary complex: a ligand, the GPCR and its associated G protein. Nevertheless, it is now well accepted that many GPCRs homo- and hetero-dimerize and even form higher-order oligomers. These complexes have been proposed to modulate trafficking, ligand binding cooperativity, and signalling efficacy (Farran, 2017).

Class C receptors form obligate dimers (Pin and Bettler, 2016), but new combinations of class A and class B receptors has been discovered (including adrenergic receptors, opioid receptors, dopaminergic receptors, etc.) forming homo-, hetero-dimers and oligomers in natural tissues (Navarro *et al.*, 2013, 2016; Medrano *et al.*, 2018). See http://gpcr-hetnet.com/ for more information on experimentally reported heteromers.

There are evidences than in class A GPCR dimers, the receptor activation is modulated by the allosteric communication between protomers (Farran, 2017; Navarro *et al.*, 2018).

The growing pool of high-resolution crystal structures available for GPCRs often reveal protomer-protomer contacts that may be feasible in the cell membrane (including homodimers, homotrimers and

homotetramers). These interface contacts can occur through all the TMs of the receptor, except for TM3 which is surrounded by others TMs. The interfaces more present are TM1-TM1, TM4-TM4, TM4/5-TM4/5, TM5-TM5 and TM5/6-TM5/6 (Figure 1.12). Although these structures do not necessary represent biologically relevant states, they could be used as a starting point to understand the interactions that determine the formation of dimer.
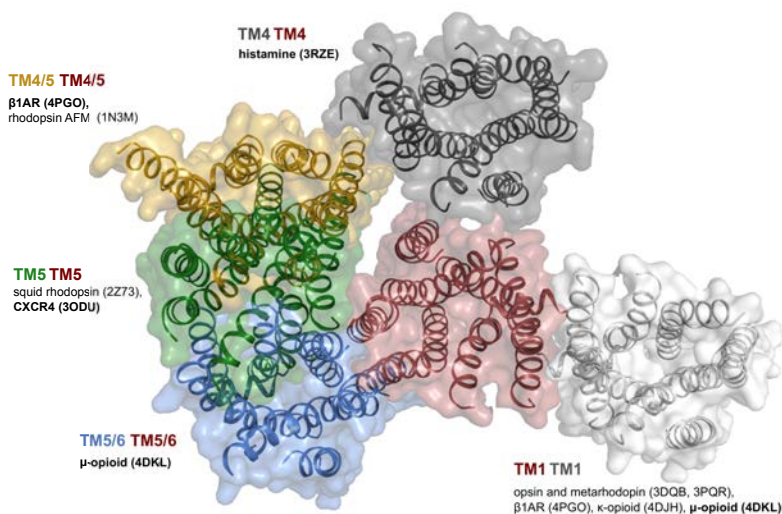
Figure 1.12| Examples of homomeric crystalized interfaces of several GPCRs. Adapted from Cordomí *et al.*, 2015.

1.5. REFERENCES

Alexander, S. P. H. *et al.* (2017) 'The concise guide to pharmacology 2017/18: G protein-coupled receptors', *Br. J. Pharmacol.*, 174, pp. S17–S129.

Ballesteros, J. and Weinstein, H. (1995) 'Integrated methods for the construction of three-dimensional models of structure−function relations in G protein-coupled receptors.', *Methods Neurosci.*, 25, pp. 366–428.

Berman, H. M. *et al.* (2000) 'The Protein Data Bank.', *Nucleic Acids Res.*, 28(1), pp. 235–42.

Bhattacharya, S. and Vaidehi, N. (2014) 'Differences in allosteric communication pipelines in the inactive and active states of a GPCR', *Biophys. J.*, 107(2), pp. 422–434.

Bockaert, J. and Pin, J. P. (1999) 'Molecular tinkering of G protein-coupled receptors: an evolutionary success.', *EMBO J.*, 18(7), pp. 1723–1729.

Cherezov, V. *et al.* (2007) 'High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor.', *Science*, 318(5854), pp. 1258–65.

Chun, E. *et al.* (2012) 'Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors', *Structure*, 20(6), pp. 967–976.

Cordomí, A. *et al.* (2015) 'Structures for G-Protein-Coupled Receptor Tetramers in Complex with G Proteins', *Trends Biochem. Sci.*, 40(10), pp. 548–551.

Dawaliby, R. *et al.* (2016) 'Allosteric regulation of G protein-coupled receptor activity by phospholipids.', *Nat. Chem. Biol.*, 12(1), pp. 35–9.

DeVree, B. T. *et al.* (2016) 'Allosteric coupling from G protein to the agonist-binding pocket in GPCRs', *Nature*, 535(7610), pp. 182–186.

Farran, B. (2017) 'An update on the physiological and therapeutic relevance of GPCR oligomers', *Pharmacol. Res.*, 117, pp. 303–327.

Foord, S. M. *et al.* (2005) 'International Union of Pharmacology. XLVI. G Protein-Coupled Receptor List', *Pharmacol. Rev.*, 57(2), pp. 279–288.

Fredriksson, R. *et al.* (2003) 'The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints.', *Mol. Pharmacol.*, 63(6), pp. 1256–1272.

Gonzalez, A. *et al.* (2012) 'Impact of Helix Irregularities on Sequence Alignment and Homology Modeling of G Protein-Coupled Receptors', *ChemBioChem*, 13(10), pp. 1393–1399.

Gurevich, V. V. and Gurevich, E. V. (2019) 'The structural basis of the arrestin binding to GPCRs', *Mol. Cell. Endocrinol.*, 484(2018), pp. 34–41.

Hauser, A. S. *et al.* (2017) 'Trends in GPCR drug discovery: new agents, targets and indications', *Nat. Rev. Drug Discov.*, 16(12), pp. 829–842.

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

Katritch, V. *et al.* (2014) 'Allosteric sodium in class A GPCR signaling', *Trends Biochem. Sci.*, pp. 233–244.

Kolakowski, L. F. (1994) 'GCRDb: a G-protein-coupled receptor database.', *Receptors Channels*, 2(1), pp. 1–7.

Komolov, K. E. and Benovic, J. L. (2017) 'G protein-coupled receptor kinases: Past, present and future', *Cell. Signal.*, 41(7), pp. 17–24.

Latorraca, N. R., Venkatakrishnan, A. J. and Dror, R. O. (2017) 'GPCR dynamics: Structures in motion', *Chem. Rev.*, 117(1), pp. 139–155.

Lee, Y., Choi, S. and Hyeon, C. (2015) 'Communication over the Network of Binary Switches Regulates the Activation of A2A Adenosine Receptor', *PLOS Comput. Biol.*, 11(2), p. e1004044.

Lodish H, et al. (2000) 'Section 20.7, Interaction and Regulation of Signaling Pathways', in *Mol. Cell Biol.* Available at: http://www.ncbi.nlm.nih.gov/books/NBK21659/.

Lu, S. and Zhang, J. (2018) 'Small Molecule Allosteric Modulators of G-Protein-Coupled Receptors: Drug–Target Interactions', *J. Med. Chem.*

Manglik, A., Kobilka, B. K. and Steyaert, J. (2017) 'Nanobodies to Study G Protein-Coupled Receptor Structure and Function.', *Annu. Rev. Pharmacol. Toxicol.*, 57(1), pp. 19–37.

Manglik, A. and Kruse, A. C. (2017) 'Structural Basis for G Protein-Coupled Receptor Activation', *Biochemistry*, pp. 5628–5634.

Medrano, M. *et al.* (2018) 'Orexin A/Hypocretin Modulates Leptin Receptor-Mediated Signaling by Allosteric Modulations Mediated by the Ghrelin GHS-R1A Receptor in Hypothalamic Neurons', *Mol. Neurobiol.*, 55(6), pp. 4718–4730.

Miao, Y. and McCammon, J. A. (2016) 'Graded activation and free energy landscapes of a muscarinic G-protein–coupled receptor', *Proc. Natl. Acad. Sci.*, 113(43), pp. 12162–12167.

Mirzadegan, T. *et al.* (2003) 'Sequence analyses of G-protein-coupled receptors: Similarities to rhodopsin', *Biochemistry*, pp. 2759–2767.

Monod, J., Wyman, J. and Changeux, J. P. (1965) 'On the nature of allosteric transitions: a plausible model.', *J. Mol. Biol.*, 12, pp. 88–118.

Munk, C. *et al.* (2019) 'An online resource for GPCR structure determination and analysis', *Nat. Methods*, 16(2), pp. 151–162.

Navarro, G. *et al.* (2013) 'Cocaine inhibits dopamine D2 receptor signaling via sigma-1-D2 receptor heteromers.', *PLoS One*, 8(4), p. e61245.

Navarro, G. *et al.* (2016) 'Quaternary structure of a G-protein-coupled receptor heterotetramer in complex with Gi and Gs.', *BMC Biol.*, 14(1), p. 26.

Navarro, G. *et al.* (2018) 'Cross-communication between Gi and Gs in a G-protein-coupled receptor heterotetramer guided by a receptor C-terminal domain', *BMC Biol.*, 16(1), p. 24.

Offermanns, S. (2003) 'G-proteins as transducers in transmembrane signalling', *Prog. Biophys. Mol. Biol.*, 83(2), pp. 101–130.

Olivella, M. *et al.* (2013) 'Relation between sequence and structure in membrane proteins', *Bioinformatics*, 29(13), pp. 1589–1592.

Palczewski, K. *et al.* (2000) 'Crystal structure of rhodopsin: A G protein-coupled receptor.', *Science*, 289(5480), pp. 739–45.

Palczewski, K. and Orban, T. (2013) 'From atomic structures to neuronal functions of g protein-coupled receptors.', *Annu. Rev. Neurosci.*, 36, pp. 139–64.

Pándy-Szekeres, G. *et al.* (2018) 'GPCRdb in 2018: Adding GPCR structure models and ligands', *Nucleic Acids Res.*, 46(D1), pp. D440–D446.

Pin, J. P. and Bettler, B. (2016) 'Organization and functions of mGlu and GABA B receptor complexes', *Nature*, 540(7631), pp. 60–68.

Rask-Andersen, M., Masuram, S. and Schiöth, H. B. (2014) 'The Druggable Genome: Evaluation of Drug Targets in Clinical Trials Suggests Major Shifts in Molecular Class and Indication', *Annu. Rev. Pharmacol. Toxicol.*, 54(1), pp. 9–26.

Rasmussen, S. G. F. *et al.* (2007) 'Crystal structure of the human β2 adrenergic G-protein-coupled receptor', *Nature*, 450(7168), pp. 383–387.

Rasmussen, Søren G. F. *et al.* (2011) 'Crystal structure of the β2 adrenergic receptor–Gs protein complex', *Nature*, 477(7366), pp. 549–555.

Rasmussen, Søren G F *et al.* (2011) 'Structure of a nanobody-stabilized active state of the β2 adrenoceptor', *Nature*, 469(7329), pp. 175–180.

Renaud, J.-P. *et al.* (2018) 'Cryo-EM in drug discovery: achievements, limitations and prospects', *Nat. Rev. Drug Discov.*, 17(7), pp. 471–492.

Rosenbaum, D. M., Rasmussen, S. G. F. and Kobilka, B. K. (2009) 'The structure and function of G-protein-coupled receptors', *Nature*, 459(7245), pp. 356–363.

Russ, A. P. and Lampel, S. (2005) 'The druggable genome: An update', *Drug Discov. Today*, pp. 1607–1610.

Sansuk, K. *et al.* (2011) 'A structural insight into the reorientation of transmembrane domains 3 and 5 during family A G protein-coupled receptor activation.', *Mol. Pharmacol.*, 79(2), pp. 262–9.

Standfuss, J. *et al.* (2011) 'The structural basis of agonist-induced activation in constitutively active rhodopsin.', *Nature*, 471(7340), pp. 656–660.

Tehan, B. G. *et al.* (2014) 'Unifying Family A GPCR Theories of Activation', *Pharmacol. Ther.*, 143(1), pp. 51–60.

Thal, D. M. *et al.* (2018) 'Structural insights into G-protein-coupled receptor allostery', *Nature*, 559(7712), pp. 45–53.

Trzaskowski, B. *et al.* (2012) 'Action of Molecular Switches in GPCRs - Theoretical and Experimental Studies', *Curr. Med. Chem.*, 19(8), pp. 1090–1109.

Venkatakrishnan, A. J. *et al.* (2013) 'Molecular signatures of G-protein-coupled receptors', *Nature*, 494(7436), pp. 185–194.

Weis, W. I. and Kobilka, B. K. (2018) 'The Molecular Basis of G Protein–Coupled Receptor Activation', *Annu. Rev. Biochem.*, 87(1), pp. 897–919.

Wise, A., Gearing, K. and Rees, S. (2002) 'Target validation of G-protein coupled receptors', *Drug Discov. Today*, 7(4), pp. 235–246.

Wootten, D., Christopoulos, A. and Sexton, P. M. (2013) 'Emerging paradigms in GPCR allostery: implications for drug discovery', *Nat. Rev. Drug Discov.*, 12(8), pp. 630–644.

Yuan, S. *et al.* (2015) 'W246 6.48 Opens a Gate for a Continuous Intrinsic Water Pathway during Activation of the Adenosine A 2A Receptor', *Angew. Chemie Int. Ed.*, 54(2), pp. 556–559.

PART 2.

OBJECTIVES

2. OBJECTIVES

The general aim of this Doctoral Thesis is to provide insight into the structure and function of G protein-coupled receptors (GPCRs). This includes understanding of the structural determinants associated to the different conformational states (activation mechanism after binding of agonist ligand), allosteric modulators regulation, oligomerization with other GPCR or additional proteins and coupling to transducers (G proteins or arrestin).

The specific aims addressed are:

GPCRs signal transmission

- Characterization of a novel missense mutation ($I131^{ICL2}T$) at the intracellular part of tyrothropin-releasing hormone receptor (TRHR) related to moderate central hypothyroidism.
- Understanding the structural determinants of the M_2 muscarinic acetylcholine receptor -M_2 mAChR-, for activation and allosterism.

GPRCs oligomerization

- Understanding the modulation of the anorexigenic effects of cocaine via the hetero-oligomer between ghrelin receptor 1a (GHS-R1a) and 1b (GHS-R1b) isoforms and sigma-1 receptor (σ_1R).

All this information will enable to design more efficient and selective therapeutically small molecules.

PART 3.

METHODOLOGY

3. METHODOLOGY

The aim of this chapter is to make a general overview of the **computational chemistry techniques** employed through the present thesis. **Computational chemistry** studies molecular systems through the application of computational models and simulations (numerical algorithms) with the aim of understanding their structure and/or properties. It uses theories, concepts and models of physical chemistry (Leszcynski, 2012; Piela, 2013) derived from *quantum mechanics* (QM), *classical molecular mechanics* (MM), *statistical mechanics* or *statistical ensembles*.

Molecular modelling includes all computational methods to mimic and study the structure and the behaviour of molecules. When focusing on protein/ligand complexes the methods can be classified into:

- **Ligand-based** (LG) **methods. These** rely on information from molecules that bind (and also from similar molecules that do not bind) to the target of interest, and include pharmacophore models and Quantitative Structure-Activity relationships.

- **Structure-based** (SB) **methods. These** rely on the knowledge of the 3D structure of the target (protein, DNA, …), and include homology modelling, molecular docking and molecular dynamics (MD) simulations.

This thesis uses SB methods. Figure 3.1 shows a typical workflow to choose the most suitable technique.
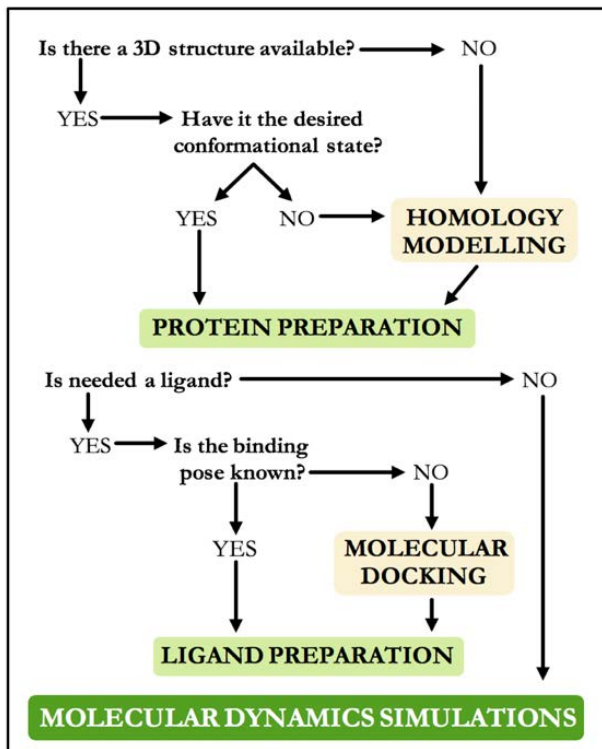
Figure 3.1| Structural-based computational techniques workflow.

3.1. HOMOLOGY MODELLING

Homology modelling is a structure prediction method that is used to construct a three-dimensional (3D) model of a *target protein* from a *template structure* of a related protein with known structure, and a sequence alignment between both proteins (*target* and *template*). The method relies on the fact that during evolution structure is more conserved than amino acid sequence. The quality of a model is linked with the sequence identity (SI, number of amino acids that match at

a given position of the alignment) between template and target sequences (Chothia and Lesk, 1986; Rost, 1999). Usually the worst parts of a model are the loop regions because they tend to be less conserved (insertions or deletions in the sequence).

Development of homology model is a multistep process (Figure 3.2).
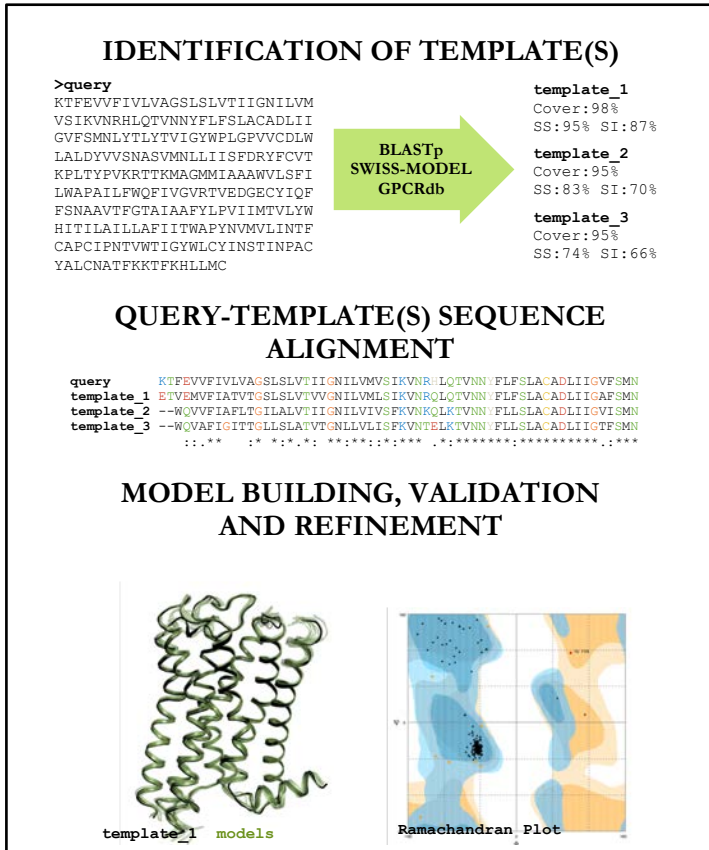
Figure 3.2| Summary of the homology modelling process. **KEY:** conserved sequence (*), conservative mutations (:), semi-conservative mutations (.), and non-conservative mutations ( ).

The first step is the ***identification of the template(s)*** and ***query-template(s) sequence alignments***. It is essential to compare the sequence of unknown structure (query) with known structures stored in the Protein Data Bank (PDB database) (Berman *et al.*, 2000), to align their sequences and choose the best candidate(s). A great tool to find suitable template(s) is **BLASTp** (*Basic Local Alignment Search Tool for proteins*, https://blast.ncbi.nlm.nih.gov/) (Altschul *et al.*, 1990). In this thesis, *PSI-BLAST* (Altschul *et al.*, 1997), which is a more sensitive alignment method based on iterative search for position-specific score matrix (PSSM) permitting the identification of distant relatives for a protein family. Because this thesis deals with GPCRs, we have also used the ***Template Selection tool*** included at **GPCRdb** (http://gpcrdb.org/structure/template_selection). The **GPCRdb** is an specific resource for GPCRs (Pándy-Szekeres *et al.*, 2018). The Template Selection tool uses a structure-based alignment, which is more accurate than a pure sequence alignment. Both BLAST and GPCRdb provide a list of PDBs ordered by sequence coverage, sequences similarity (SS) or sequence identity (SI) and query-template(s) sequence alignment to download.

The second step is ***Model building, validation and refinement***. The *model building* typically consists on comparative protein structure modelling by satisfaction of spatial restraints (atomic distances, dihedral angles and statistical distributions for them). MODELLER software (Sali, 1995) has been mostly used in this thesis to prepare protein models. This step often requires various iterations in which the alignment needs to be changed. This can be done using sequence editing tools such as Jalview (Waterhouse et al., 2009). The quality of models can be *validated* using the Ramachandran plot (Ramachandran, Ramakrishnan and Sasisekharan, 1963); the evaluation of the RMSD

of C_α atoms, and/or the use of a scoring function capable of discriminating good and bad models. Model building, validation and refinement are repeated until the best possible model (model that reproduce the experimental results) is obtained.

3.2. MOLECULAR DOCKING

Molecular docking aims at predicting the orientation of the ligand (small-molecule, peptide or protein) when it interacts with a protein using shape complementarity and interaction energy terms. Docking is a SB technique, so is necessary to have a 3D structure available (by X-ray or cryo-electron microscopy -cryo-EM- or Nuclear Magnetic Resonance -NMR) or a structure obtain by homology modelling (*see section 3.1.1*).

The docking process involves two steps:

1. Explore the conformational landscape of the small-molecule to find the best candidate binding modes (poses) within the receptor. This is done using **sampling methods** (Salmaso and Moro, 2018). Sampling methods are classified according to the degree of flexibility of the molecules involved in the calculation in *rigid* (both molecules rigid), *semi-flexible* (protein is rigid and molecules, flexible) and *flexible docking* (both molecules flexible).

2. Rank the putative binding modes of all poses generated and assess their binding affinity most programs employ a **scoring function** (Leelananda and Lindert, 2016). There are three types of scoring functions:

- *Force-field based*, which are the sum of intermolecular and intramolecular interactions,
- *Empirical*, which are the sum of various energy terms weighted by coefficients to fit them in a multiple linear regressions and,
- *Knowledge based scoring functions*, which use statistical analysis of intermolecular contacts in large 3D databases to derive it into potentials using Boltzmann principle.

In this thesis, ligand-protein dockings were performed using **Autodock Vina** (Trott and Olson, 2010) *with* semi-flexible docking and a hybrid scoring function (empirical and knowledge based), or **MOE software** (Corbeil, Williams and Labute, 2012) *with* sidechain flexible docking and the GBVI/WSA dG scoring function. Protein-protein dockings (in *sub-chapter 4.2.2*) were performed using **HADDOCK 2.2** (De Vries, Van Dijk and Bonvin, 2010). This software permits to incorporate experimental or predicted information (the interface region between the molecular components and/or their relative orientations) to drive the flexible docking process and using an empirical scoring function.

3.3. MOLECULAR DYNAMICS (MD) SIMULATIONS

Molecular dynamics (**MD**) is a deterministic method that provides the time-evolution of a molecular system represented by a set of particles with defined positions (an initial X-ray crystallographic, NMR or homology model structure).

Starting from this initial structure and assigning initial velocities to each atom (from a *statistical mechanics* approach, the Boltzmann distribution at a given temperature), successive coordinates and velocities are obtained by integrating **Newton's equation of motion** (equation 1) for each particle i:

$$\frac{d^2 \vec{x_i}(t)}{dt^2} = \frac{\vec{F_{x_i}}}{m_i}$$
equation 1

where x_i are the coordinates for the ith particle at time t, Fx_i is the force exerted to the ith particle (this force is referred as *force field*, and will be described in the *sub-section 3.3.1. Force Fields*) and m_i is the mass of the ith particle.

The result is a trajectory that shows how atomic positions and velocities evolve over time according to the influence of all atoms in the system. However, due to the large number of particles interacting with each other this equation is difficult to solve analytically, so it is necessary to perform **numerical integration** using methods, such as *Leapfrog integrator (*update the position and velocity at interleaved time points*).* The selection of the *time interval* (*dt*) is crucial for the simulation stability. The use of constraints in bonds involving hydrogen atoms permit, using the LINCS algorithms permits to use time-steps of 2 fs (Hess *et al.*, 1997).

The approach to simulate the behavior of a transmembrane protein complex is typically to embed the structure in a box that contains a lipid bilayer and water solvent. To avoid problems with boundary effects caused by the finite size of the box (unit cell), **periodic boundary conditions** (PBCs) are used, where the unit cell is surrounded by infinite replicas of itself (González, 2011). MD simulations are typically run in the NVT (also known as canonical

ensemble) or the NPT (or isothermal-isobaric ensemble) statistical mechanical ensembles. For simulations that contain a lipid bilayer the *NPT ensemble* is generally employed. To keep pressure and temperature constant, simulations are performed with a barostat (one for each system component: protein, lipids and solvent), and thermostats, respectively. The *Berendsen* (Berendsen *et al.*, 1984) or *Parrinello-Rahman* (Parrinello and Rahman, 1981) *barostats* (for the pressure, 0.1MPa in xyz coordinates) and *v-rescale thermostat* (for the temperature, 300K) (Bussi, Zykova-Timan and Parrinello, 2009) were used for all simulations performed in this thesis.

3.3.1. FORCE FIELDS

The molecular features that characterize each atom of the system determine their interactions and dynamics throughout the simulation. The mathematical function and parameters to allow the calculation of the potential energy (E) of the system according the position of the atoms is the so-called **force field** (FFs) (González, 2011)**.** The most used force fields in molecular dynamics of biological macromolecules are AMBER (Cornell *et al.*, 1995), CHARMM (MacKerell *et al.*, 1998), GROMOS (Oostenbrink *et al.*, 2004) and OPLS (Jorgensen and Tirado-Rives, 1988).

A typical force field function (equation 2) consists in **bonded terms** that define the *intramolecular interactions* (equation 3) in the system and **non-bonded terms** that mostly capture *inter-molecular interactions* (equation 4):

$$E = E^{BONDED} + E^{NON\text{-}BONDED} \qquad \text{equation 2}$$

$$E^{BONDED} = E^{bonds} + E^{angles} + E^{torsions} \qquad \text{equation 3}$$

$$E^{\text{NON-BONDED}} = E^{\text{van der Waals}} + E^{\text{electrostatic}} \qquad \text{equation 4}$$

Both bonds (E^{bonds}, equation 5) and angles (E^{angles}, equation 6) terms are described by harmonic potentials (Hooke's law):

$$E^{\text{bonds}} = \sum_{\text{bonds}} \frac{k_{b_{i,j}}}{2} \left(b_{i,j} - b_{0_{i,j}} \right)^2 \qquad \text{equation 5}$$

$$E^{\text{angles}} = \sum_{\text{angles}} \frac{k_{\theta_{i,j}}}{2} \left(\theta_{i,j} - \theta_{0_{i,j}} \right)^2 \qquad \text{equation 6}$$

where b and b_0 are the current bond length and its equilibrium value, $\theta_{i,j}$ and $\theta_{0_{i,j}}$ are the current angle value and its equilibrium value and $k_{b_{i,j}}$ and $k_{\theta_{i,j}}$ are the bond and angle force constants, between atoms *i* and *j,* respectively. The torsion term, also referred as dihedral term (E^{torsions}, equation 7) is often represented as a sum of cosine functions:

$$E^{\text{torsions}} = \sum_{\text{torsions}} k_{\varphi_{i,j}} \left[1 + \cos(m_{i,j}\, \varphi_{i,j} - \delta_{i,j}) \right] \qquad \text{equation 7}$$

where $\varphi_{i,j}$ is the torsion angle, $m_{i,j}$ is the multiplicity, $\delta_{i,j}$ is the phase angle and $k_{\varphi_{i,j}}$ is the dihedral force constant between atoms *i* and *j*.

As it was advanced in the equation 4, the non-bonded terms consist of two energy terms: van der Waals and electrostatic interactions, which are implemented through the Lennard-Jones 12-6 potential ($V^{\text{van der Waals}}$ or $V^{\text{Leonard-Jones}}$, equation 8) and the Coulombic potential ($V^{\text{electrostatic}}$ or V^{Coulomb}, equation 9), respectively. The Lennard-Jones potential is a mathematical model that approximates the interaction between a pair of atoms, accounting for two distinct forces, one attractive and another repulsive:

$$E^{\text{Leonard-Jones}} = \sum_{i,j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \qquad \text{equation 8}$$

where ε_{ij} is a parameter defining the depth of the energy minimum, r_{ij} is the distance between the atoms i and j, and σ_{ij} is the energy expressed as an inverse power function of the distance between the considered atoms i and j. Finally, the last term in the force field equation is the Coulombic potential:

$$E^{\text{Coulomb}} = \sum_{i,j} \frac{q_i\, q_j}{4\pi\, \varepsilon_0 r_{ij}} \qquad \text{equation 9}$$

where q_i and q_j are the partial charges assigned to atoms i and j, ε_0 is the dielectric constant and r_{ij} is the relative distance between these atoms.

The computation of non-bonded interactions is the most time-consuming part of a MD simulation as the evaluation of the forces scales quadratically with the number of atoms in the system if no approximation is used. Therefore, a distance *cut-off* of about 1nm is typically used for *LJ* (the maximum distance at which LJ interactions between atom pairs are computed). The same truncation strategy of Coulomb interactions originates problems in simulations. Therefore long-ranged electrostatic interactions beyond a certain cutoff (typically 1 nm) are not truncated but taken into account using *Particle Mesh Ewald* (PME) (Darden, York and Pedersen, 1993) methods. All simulations here performed used a spacing of about 0.15 nm for the PME grid.

The simulations performed as part of this thesis were all run using **Amber ff99sb-ildn** **protein force field** for the protein (an all-atom force field) (Lindorff-Larsen *et al.*, 2010) with **Berger lipid**

parameters (Berger, Edholm and Jähnig, 1997) for the membrane (united-atom force field: does not treat explicitly non-polar hydrogen atoms of CH_2 and CH_3 groups). Because we are interested in the protein, Berger lipids provide a reasonable description of the membrane at a lower computational cost compared to all-atom alternatives as CHARMM (Pastor and Mackerell, 2011), Stockholm lipids -Slipids- (Jämbeck and Lyubartsev, 2012) or AMBER lipids -Lipid14- (Dickson *et al.*, 2014); see Pluhackova *et al.*, 2016 for a comparison of biomembrane force fields. The compatibility of *Amber ff99sb-ildn protein force field* and *Berger lipid parameters* was studied in our group (Cordomí, Caltabiano and Pardo, 2012). The **water model** chosen was *TIP3P* (Price and Brooks, 2004), TIP model with three interaction sites, corresponding to the three atoms of the water molecule with rigid geometry.

All MD simulations performed as part of this thesis were conducted with GROMACS 4.6.7 and 5.1.4 simulation packages (Abraham *et al.*, 2015). Generally, in all projects of this thesis each system is simulated multiple times under the same conditions (same atomic coordinates but with random initial velocities). This process helps us to compare similarities and/or variations between these trajectories.

3.3.2 LIGAND AND PROTEIN SET-UP FOR RUN MD SIMULTIONS

One of the most powerful techniques used within this thesis is molecular dynamics (MD). Before running MD simulations, it is necessary to *pre-process both ligand and protein*, the **ligand and protein set-up process**. This step is crucial to reach accurate computational results (Madhavi Sastry *et al.*, 2013).

Description of each of these processes:

- **Ligand set-up**. The chemical and physical characteristics of the proteins are reasonably constant (see *3.3.1. Force Field subsection*). Thus, amino acids and other common molecules or building blocks are parametrized in all molecular modelling programs. However, most ligands (also non-natural amino acids) require explicit parametrization: assign bonded and van der Waals parameters and derive atomic partial charges. In AMBER this can be done with Antechamber program, part of AMBER Tools (Salomon-Ferrer, Case and Walker, 2013) using the General AMBER Force Field (common known as GAFF) (Wang et al., 2004). For all the ligands used in this thesis, **RESP** (Restrained Electrostatic Surface Potential) **charges** (Bayly *et al.*, 1993) were adjusted based on an electrostatic potential computed using **ab initio Hartree-Fork** methods (HF/6-31G*) using the program Gaussian (Wallingford CT, 2009).

- **Protein set-up**. The protein preparation process typically includes adding unresolved residue, hydrogen atoms and other missing atoms, assignment of tautomers and ionization states, and relaxation (minimization) of the protein structure. Assignment of the protonation state to ionizable groups at pH 7, was performed using **PROPKA** (Søndergaard, Olsson and Rostkowski Michał and Jensen, 2011). PROPKA also assigned the most likely His tautomers, and the right orientation of the amide group in Asp and Gln -the NH_2 and CO groups, which cannot be unambiguously assigned in the structures. After that, proteins were prepared using **AMBER Tools** (Salomon-Ferrer, Case and Walker, 2013), the LeaP program, to generate AMBER topology file and coordinate file. This program adds

hydrogens, generates the missing atoms (when possible) and caps the N- and C-termini of the structure (when needed).

3.3.3. VISUALIZATION AND ANALYSIS OF THE MD TRAJECTORIES

The molecular dynamic (MD) trajectory is the result of using Newton's laws of motion to predict the position of each atom of the system at every point during the simulated time interval, so, we obtain a *three-dimensional (3D) movie* that describes it. In order to carefully visualise and analyse all the MD trajectories generated in this thesis, GROMACS built-in analysis tools, PyMOL 1.7 and 2.0 (Schrödinger, 2014) and Visual Molecular Dynamics (VMD) 1.9 (Humphrey, Dalke and Schulten, 1996) programs were employed.

3.3.4. STATE-OF-THE-ART OF MD SIMULATIONS

In the middle of the 20th century, MD simulations theories and techniques were introduce in the scientific community, and the first MD simulation of simple gasses were performed (Alder and Wainwright, 1957). But it was not until in the late 1970s when the first MD simulation of a protein was performed (McCammon, Gelin and Karplus, 1977). In the last 25 years, they have begun to appear frequently in interdisciplinary papers to interpret experimental results and to guide experimental work (Hollingsworth and Dror, 2018).

The increasing utilization of MD simulations has been due to an explosion of experimental structures of certain protein families, especially in membrane proteins (for example: ion channels, transporters and GPCRs) whose structure was difficult to determine.

And also, MD simulation themselves have become much powerful (improvements in computing hardware, for example using Graphics Processing Units -GPUs- (Kutzner *et al.*, 2015)), more accurate (substantially improvement of force fields to reproduce the experimental data (Martín-García *et al.*, 2015) and give more reliable results) and more accessible to more researchers, for example: using detailed tutorials for GROMACS (Lemkul, 2019).

MD simulations can provide information to many interesting biological questions, and complement experimental results, as it has been done in this thesis. Some examples of the application of MD simulations (Figure 3.3) can be the study of:

- ***Transport across a membrane.*** In Latorraca *et al.*, 2017, MD simulations capture the spontaneous transition of the sugar transporter SemiSWEET (a symmetric dimer of three-helix bundles) from its outward-open state (where the substrate-binding pocket is accessible to the outside of the cell) to its inward-open state, along with the accompanying substrate translocation process.

- ***Protein folding,*** which is the process where a disordered polypeptide chain (primary -unfolded- structure) is molecular self-assembly to form a well-defined native 3D structure (tertiary structure). In the paper of Lindorff-Larsen *et al.*, 2011, the study of several small protein (less than 80 residues) reveals that the first step is the formation of the secondary structure (α-helices or β-sheets), which it is determined by the amino acid sequence of each protein; and finally, the nature of secondary structure aids the full folding of the protein.
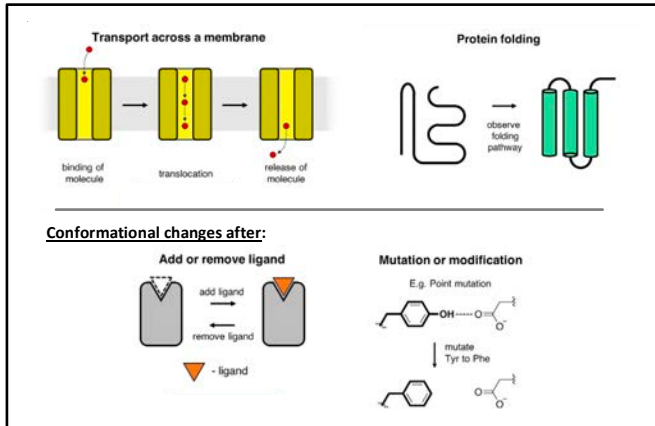
Figure 3.3| **Applications of MD simulations.** Adapted from: Hollingsworth and Dror, 2018.

- ***Conformational changes.*** MD simulations are employed in many equilibrium processes, so *addition, replacement or removal a ligand* in a determined protein structure can favour one of the equilibrium states (Dror *et al.*, 2013; Korczynska *et al.*, 2018). Usually, one direction of the process occur more quickly than the other, for example: in GPCRs, the deactivation mechanism is easier to achieve than the activation mechanism in GPCRs (Dror *et al.*, 2011; Miao *et al.*, 2013). In order to understand the key residues of binding pocket or in the function of a protein, *mutations* are made in specific amino acids in the receptor that allow the comparison with wild type protein (Piechnick *et al.*, 2012; Keränen, Åqvist and Gutiérrez-de-Terán, 2015).

Although MD simulations are getting longer -microsecond to millisecond timescales- and cheaper (system of 50000 atoms can be simulated for a microsecond using a GPU in few days), there are important biomolecular processes that occur on longer-time scales

that are not yet accessible with classical MD simulations. To overcome this limitation, several alternative methods have been developed during the last years:

- **Enhanced sampling techniques** (De Vivo *et al.*, 2016; Salmaso and Moro, 2018) consist on applying a bias force/potential to the system to increase the accessible timescale of slow processes (to escape from a local minima), usually using collective variables (CVs, which is a reaction coordinate and is often plotted against free energy) to introduce the bias. Some of the most common are:
    - *Replica Exchange Molecular Dynamics* (REMD) where independent simulations of the same system are run in parallel but at different temperatures, and exchanged at different points in time (Sugita and Okamoto, 1999).
    - *Steered Molecular Dynamics* (SMD) where an external force is applied to a molecular system in order to accelerate the kinetics of the conformational change of interest (Izrailev *et al.*, 1999).
    - *Umbrella Sampling* (US) (Torrie and Valleau, 1977) enforce sampling along the chosen CV by performing different simulations -windows- to an energetic bias, the harmonic potential -umbrella.
    - *Metadynamics* introduce a bias to the Hamiltonian of the system as a Gaussian-shaped function of one or more CVs (Laio and Parrinello, 2002).

- **Coarse-Grained (CG) MD simulations**, in which groups of atoms are condensed into a single particle, reducing the degrees of freedom of the system (Kmiecik *et al.*, 2016). The most

popular and widely used is MARTINI (Marrink *et al.*, 2007), in which four heavy atoms (and their hydrogens) are grouped into a single particle.

3.2.   REFERENCES

Abraham, M. J. *et al.* (2015) 'Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers', *SoftwareX*, 1–2, pp. 19–25.

Alder, B. J. and Wainwright, T. E. (1957) 'Phase Transition for a Hard Sphere System', *J. Chem. Phys.*, 27(5), pp. 1208–1209.

Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool.', *J. Mol. Biol.*, 215(3), pp. 403–10.

Altschul, S. F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: A new generation of protein database search programs', *Nucleic Acids Res.*, 25(17), pp. 3389–3402.

Bayly, C. I. *et al.* (1993) 'A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model', *J. Phys. Chem.*, 97(40), pp. 10269–10280.

Berendsen, H. J. C. *et al.* (1984) 'Molecular dynamics with coupling to an external bath', *J. Chem. Phys.*, 81(8), pp. 3684–3690.

Berger, O., Edholm, O. and Jähnig, F. (1997) 'Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature.', *Biophys. J.*, 72(5), pp. 2002–2013.

Berman, H. M. *et al.* (2000) 'The Protein Data Bank.', *Nucleic Acids Res.*, 28(1), pp. 235–42.

Bussi, G., Zykova-Timan, T. and Parrinello, M. (2009) 'Isothermal-isobaric molecular dynamics using stochastic velocity rescaling', *J. Chem. Phys.*, 130(7), p. 074101.

Chothia, C. and Lesk, A. M. (1986) 'The relation between the divergence of sequence and structure in proteins.', *EMBO J.*, 5(4), pp. 823–6.

Corbeil, C. R., Williams, C. I. and Labute, P. (2012) 'Variability in docking success rates due to dataset preparation', *J. Comput. Aided. Mol. Des.*, 26(6), pp. 775–786.

Cordomí, A., Caltabiano, G. and Pardo, L. (2012) 'Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters', *J. Chem. Theory Comput.*, 8(3), pp. 948–958.

Cornell, W. D. *et al.* (1995) 'A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules', *J. Am. Chem. Soc.*, 117(19), pp. 5179–5197.

Darden, T., York, D. and Pedersen, L. (1993) 'Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems', *J. Chem. Phys.*, 98(12), p. 10089.

Dickson, C. J. *et al.* (2014) 'Lipid14: The amber lipid force field', *J. Chem. Theory Comput.*, 10(2), pp. 865–879.

Dror, R. O. *et al.* (2011) 'Activation mechanism of the β2-adrenergic receptor', *Proc. Natl. Acad. Sci.*, 108(46), pp. 18684–18689.

Dror, R. O. *et al.* (2013) 'Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs.', *Nature*, 503(7475), pp. 295–9.

Gaussian, I. and Wallingford, C. (2009) 'Gaussian 09'.

González, M. (2011) 'Force fields and molecular dynamics simulations', *École thématique la Société Française la Neutron.*, 12, pp. 169–200.

Hess, B. *et al.* (1997) 'LINCS: A linear constraint solver for molecular simulations', *J. Comput. Chem.*, 18(12), pp. 1463–1472.

Hollingsworth, S. A. and Dror, R. O. (2018) 'Molecular Dynamics Simulation for All', *Neuron*, 99(6), pp. 1129–1143.

Humphrey, W., Dalke, A. and Schulten, K. (1996) 'Visual Molecular Dynamics', *J. Mol. Graph.*, 14(1), pp. 33–38.

Izrailev, S. *et al.* (1999) 'Steered molecular dynamics simulation of the Rieske subunit motion in the cytochrome bc1 complex', *Biophys. J.*, 77(4), pp. 1753–1768.

Jämbeck, J. P. M. and Lyubartsev, A. P. (2012) 'Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids', *J. Phys. Chem. B*, 116(10), pp. 3164–3179.

Jorgensen, W. L. and Tirado-Rives, J. (1988) 'The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin', *J. Am. Chem. Soc.*, 110(6), pp. 1657–1666.

Keränen, H., Åqvist, J. and Gutiérrez-de-Terán, H. (2015) 'Free energy calculations of A(2A) adenosine receptor mutation effects on agonist binding.', *Chem. Commun.*, 51(17), pp. 3522–5.

Kmiecik, S. *et al.* (2016) 'Coarse-Grained Protein Models and Their Applications', *Chem. Rev.*, 116(14), pp. 7898–7936.

Korczynska, M. *et al.* (2018) 'Structure-based discovery of selective positive allosteric modulators of antagonists for the M 2 muscarinic acetylcholine receptor', *Proc. Natl. Acad. Sci.*, p. 201718037.

Kutzner, C. *et al.* (2015) 'Best bang for your buck: GPU nodes for GROMACS biomolecular simulations', *J. Comput. Chem.*, 36(26), pp. 1990–2008.

Laio, A. and Parrinello, M. (2002) 'Escaping free-energy minima', *Proc. Natl. Acad. Sci.*, 99(20), pp. 12562–12566.

Latorraca, N. R. *et al.* (2017) 'Mechanism of Substrate Translocation in an Alternating Access Transporter', *Cell*, 169(1), pp. 96–107.

Leelananda, S. P. and Lindert, S. (2016) 'Computational methods in drug discovery', *Beilstein J. Org. Chem.*, 12, pp. 2694–2718.

Lemkul, J. (2019) 'From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]', *Living J. Comput. Mol. Sci.*, 1(1), pp. 1–52.

Leszcynski, J. (2012) *Handbook of Computational Chemistry*. Edited by J. Leszczynski. Dordrecht: Springer Netherlands.

Lindorff-Larsen, K. *et al.* (2010) 'Improved side-chain torsion potentials for the Amber ff99SB protein force field', *Proteins Struct. Funct. Bioinforma.*, 78(8), pp. 1950–1958.

MacKerell, A. D. *et al.* (1998) 'All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †', *J. Phys. Chem. B*, 102(18), pp. 3586–3616.

Madhavi Sastry, G. *et al.* (2013) 'Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments', *J. Comput. Aided. Mol. Des.*, 27(3), pp. 221–234.

Marrink, S. J. *et al.* (2007) 'The MARTINI force field: Coarse grained model for biomolecular simulations', *J. Phys. Chem. B*, 111(27), pp. 7812–7824.

Martín-García, F. *et al.* (2015) 'Comparing molecular dynamics force fields in the essential subspace', *PLoS One*, 10(3), pp. 1–16.

McCammon, J. A., Gelin, B. R. and Karplus, M. (1977) 'Dynamics of folded proteins', *Nature*, 267(5612), pp. 585–590.

Miao, Y. *et al.* (2013) 'Activation and dynamic network of the M2 muscarinic receptor', *Proc. Natl. Acad. Sci. U. S. A.*, 110(27), pp. 10982–10987.

Oostenbrink, C. *et al.* (2004) 'A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6', *J. Comput. Chem.*, 25(13), pp. 1656–1676.

Pándy-Szekeres, G. *et al.* (2018) 'GPCRdb in 2018: Adding GPCR structure models and ligands', *Nucleic Acids Res.*, 46(D1), pp. D440–D446.

Parrinello, M. and Rahman, A. (1981) 'Polymorphic transitions in single crystals: A new molecular dynamics method', *J. Appl. Phys.*, 52(12), pp. 7182–7190.

Pastor, R. W. and Mackerell, A. D. (2011) 'Development of the CHARMM Force Field for Lipids.', *J. Phys. Chem. Lett.*, 2(13), pp. 1526–1532.

Piechnick, R. *et al.* (2012) 'Effect of channel mutations on the uptake and release of the retinal ligand in opsin', *Proc. Natl. Acad. Sci. U. S. A.*, 109(14), pp. 5247–5252.

Piela, L. (2013) *Ideas of Quantum Chemistry*. 2nd Editio. Elsevier.

Pluhackova, K. *et al.* (2016) 'A Critical Comparison of Biomembrane Force Fields: Structure and Dynamics of Model DMPC, POPC, and POPE Bilayers.', *J. Phys. Chem. B*, 120(16), pp. 3888–903.

Price, D. J. and Brooks, C. L. (2004) 'A modified TIP3P water potential for simulation with Ewald summation', *J. Chem. Phys.*, 121(20), pp. 10096–10103.

Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. (1963) 'Stereochemistry of polypeptide chain configurations.', *J. Mol. Biol.*, 7, pp. 95–9.

Rost, B. (1999) 'Twilight zone of protein sequence alignments', *Protein Eng. Des. Sel.*, 12(2), pp. 85–94.

Sali, A. (1995) 'Comparative protein modeling by satisfaction of spatial restraints', *Mol. Med. Today*, 1(6), pp. 270–277.

Salmaso, V. and Moro, S. (2018) 'Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview', *Front. Pharmacol.*, 9(August), pp. 1–16.

Salomon-Ferrer, R., Case, D. A. and Walker, R. C. (2013) 'An overview of the Amber biomolecular simulation package', *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 3(2), pp. 198–210.

Schrödinger, L. (2014) *The PyMOL Molecular Graphics System, Version 1.7.*

Søndergaard, C. R., Olsson, M. H. M. and Rostkowski Michał and Jensen, J. H. (2011) 'Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values', *J Chem Theory Comput*, 7(7), pp. 2284–2295.

Sugita, Y. and Okamoto, Y. (1999) 'Replica-exchange molecular dynamics method for protein folding', *Chem. Phys. Lett.*, 314(1–2), pp. 141–151.

Torrie, G. M. and Valleau, J. P. (1977) 'Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling', *J. Comput. Phys.*, 23(2), pp. 187–199.

Trott, O. and Olson, A. (2010) 'NIH Public Access', *J. Comput. Chem.*, 31(2), pp. 455–461.

De Vivo, M. *et al.* (2016) 'Role of Molecular Dynamics and Related Methods in Drug Discovery', *J. Med. Chem.*, 59(9), pp. 4035–4061.

De Vries, S. J., Van Dijk, M. and Bonvin, A. M. J. J. (2010) 'The HADDOCK web server for data-driven biomolecular docking', *Nat. Protoc.*, 5(5), pp. 883–897.

Wang, J. *et al.* (2004) 'Development and testing of a general amber force field.', *J. Comput. Chem.*, 25(9), pp. 1157–74.

Waterhouse, A. M. *et al.* (2009) 'Jalview Version 2-A multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25(9), pp. 1189–1191.

PART 4.

RESULTS

4. RESULTS

The work presented as part of this Thesis includes three projects that provide insight into the following two main topics: **GPCRs signal transmission** (*chapter 4.1*) and **GPCRs oligomerization** (*chapter 4.2*). Each chapter is divided in subchapters (one subchapter per project) where each project contains its own introduction, methodology, results, discussion and references. My contribution in all these projects has been employing computational techniques (explained at *part 3. Methodology*). In almost all the projects, it is also included the experimental data obtained by our collaborators that is necessary to understand the overall context/outcome of the work.

What follows is a brief description of each project, people involved, and derived publications:

- **Tyrothropin-Releasing Hormone Receptor** (TRHR) (**subchapter 4.1.1**). Central hypothyroidism due to a THRH mutation causing impaired ligand affinity and transactivation of Gq. This work was carried out in collaboration with *Marta García* and *José C. Moreno* at Hospital Universitario La Paz (Madrid); *Jesús González de Buitrago* at Hospital San Pedro de Alcántara (Cáceres) and *Patricia M. Hinkle* at University of Rochester Medical Center Rochester (New York). It was published at *The Journal of Clinical Endocrinology & Metabolism 2017*, 102(7), pp. 2433–2442 (doi: 10.1210/jc.2016-3977).

- **M₂ Muscarinic Acetylcholine Receptor** (M₂ mAChR) at *sub-chapter 4.1.2.* Ligand-triggered structural changes in the M₂ muscarinic acetylcholine receptor. This section is a collaboration with *Minos-Timotheos Matsoukas* at University of Patras (Patras, Greece) and *Gianluigui Caltabiano* at Universitat Autònoma de Barcelona. This work is published at *Journal of Chemical Information and Modeling 2018*, 58(5), pp. 1074–1982 (doi: 10.1021/acs.jcim.8b00108).

- Hetero-oligomer of **Growth Hormone Secretagogue** (GHS) *or* **Ghrelin Receptor** (GHSR) subtypes 1a and 1b and the **Sigma-1 receptor**, at s*ub-chapter 4.2.2.* Cocaine blocks ghrelin effects via interaction with sigma-1 receptors. It is a collaboration with *David Aguinaga, Mireia Medrano, Edgar Angelats, Mireia Casanovas, Enric I. Canela, Rafael Franco* and *Gemma Navarro* at Universitat de Barcelona (Barcelona); *Ignacio Vega-Quiroga* and *Katia Gysling* at Pontificia Universidad Católica de Chile (Santiago, Chile); and *Milos Petrovic* at University of Central Lancashire (Preston, United Kingdom). This work is published at *Molecular Neurobiology 2018,* 56(2), pp. 1196–1210   (doi: 10.1007/s12035-018-1140-7).

4.1. GPCRs SIGNAL TRANSMISSION

4.1.1. CENTRAL HYPOTHYROIDISM DUE TO A TRHR MUTATION CAUSING IMPAIRED LIGAND AFFINITY AND TRANSACTIVATION OF Gq

4.1.1.1. *Background*

Central Congenital Hypothyroidism (CCH) is caused by deficient production of thyroid hormones (T4 and T3) due to low synthesis, secretion or bioactivity of thyrotropin (TSH) at the pituitary. CCH is an underdiagnosed disorder, since CCH patients are not detected by TSH-based neonatal screening programs for CH implemented in most countries (Persani, 2012). However, uncommon T4-based CH screening programs in few countries recently estimated the prevalence of CCH in 1 in 16,000-30,000 new-borns (van Tijn *et al.*, 2005; Adachi *et al.*, 2012). The paucity of clinical cases identified and the complexity of hypothalamic-pituitary regulation of the thyroid axis leaves the molecular mechanisms underlying CCH largely unknown (García, Fernández and Moreno, 2014). At present, genetic defects in only 3 genes were identified in patients with isolated CCH: *TSHB* (encoding the B-subunit of the TSH glycoprotein hormone), *TRHR (the specific 7-transmembrane domain receptor of hypothalamic TRH)* and *IGSF1 (a protein regulating the expression of TRHR at pituitary thyrotropes)*. Most CCH patients described with CCH harboured defects in *TSHB* and *IGSF1*, but only three families with *TRHR* defects have been so far identified (Collu *et al.*, 1997; Bonomi *et al.*, 2009; Koulouri *et al.*, 2016).

The thyrotropin-releasing hormone receptor (TRHR) is a G protein-coupled receptor (GPCR) located at pituitary thyrotropes and

activated by hypothalamic TRH. TRH interacts with amino acids of the ECLs and then moves into the TM binding pocket (Bílek and Stárka, 2005; Engel and Gershengorn, 2007). These interactions at the extracellular side trigger small local structural changes near the binding site that are translated into larger-scale helix movements at the intracellular site, mainly TMs 5 and 6, opening a cavity for the binding of the C-terminal α5 helix of the G-protein (Rasmussen *et al.*, 2011). The formation of the TRH-TRHR-Gq complex triggers the activation of phosphatidylinositol-calcium-protein kinase C (Gq-IP-PKC) pathway (Hinkle, Gehret and Jones, 2012).

TRH-TRHR signalling promotes the synthesis, secretion and bioactivity of TSH, all necessary for the proper synthesis of T4 and T3 in the thyroid gland (Weintraub *et al.*, 1989; Fares, 2006). None of three unrelated patients so far known with a TRHR defect was detected in TSH-based neonatal screening programs (Collu *et al.*, 1997; Bonomi *et al.*, 2009; Koulouri *et al.*, 2016). The first two patients described were referred to the clinician at the age of 9 and 11 years, respectively, with similar manifestations of short stature and variable symptoms consistent with hypothyroidism (lethargy, fatigue, poor school performance), while the third case was diagnosed and treated for CH at 2 months of life with normal growth (Koulouri *et al.*, 2016). In all cases, thyroid hormone profiles revealed normal TSH (with suspected low bioactivity) and the presence of moderate hypothyroidism. Heterozygous carriers were reported euthyroid.

The first patient described was compound heterozygote for an early stop codon in the *TRHR* (p.R17X) and an inframe deletion added to a missense change (p.S115-T117del + p.A118T) in the other allele (Collu *et al.*, 1997). The same p.R17X mutation was found in the

second patient in homozygous state (Bonomi *et al.*, 2009), while the third patient presents a homozygous missense mutation (p.P81R) (Koulouri *et al.*, 2016). All TRHR mutations identified so far severely impair TRHR signalling (Collu *et al.*, 1997; Bonomi *et al.*, 2009; Koulouri *et al.*, 2016).

Here we present a novel missense mutation in TRHR located at a highly conserved hydrophobic position (Φ) at the $(E/D)R^{3.50}YX_5P\Phi XY$ motif of GPCRs, which reduces the affinity for TRH and impairs, but not fully abrogates, signal transduction of the receptor. In consistence with its residual function, the mutation causes moderate CH in the homozygous state and central hyperthyrotropinemia in heterozygotes, suggesting compensatory elevation of TSH with reduced biopotency.

4.1.1.2. *Materials and Methods*

Hormonal determinations and TRH test. TSH, Free thyroxine (FT4), LH, FSH and prolactin were determined in serum by electrochemiluminescence with Elecsys-170 platform (Roche, Switzerland). IGF-1, IGFBP-3, ACTH and cortisol were measured by chemiluminescence with the Immulite 2000 system (Siemens, Germany). TRH stimulation test was performed as previously reported (van Tijn, de Vijlder and Vulsma, 2008).

Mutation screening. All coding regions of *TRH*, the gene encoding the thyrotropin-releasing hormone, *TRHR,* the gene encoding the TRH receptor, *TSHB,* the gene encoding the specific TSH beta subunit and IGSF1, the gene encoding immunoglobulin superfamily factor 1 were amplified by PCR using appropriate primers flanking

each exon. PCR products were purified and directly sequenced on an automated DNA sequencer (3100 Genetic Analyzer, Applied Biosystems).

Expression plasmids and reporter constructs. A plasmid encoding the long form of the rat TRH receptor with two HA tags on the amino-terminus has been described (Zhu, Cook and Hinkle, 2002). This pcDNA3-based plasmid was mutated to convert Isoleucine 131 to Threonine in pcDNA3 (QuikChange Lightning Site-Directed Mutagenesis Kit, Agilent Technologies). A reporter plasmid encoding AP1 (c-jun)-luciferase was kindly provided by Dr. Marvin Gershengorn at the NIH (Bethesda, MD) (Colson *et al.*, 1998). An HA-tagged M_3 muscarinic receptor plasmid was obtained from the cDNA Resource Center. $[^3H]N_3$-methyl-His_2-TRH was from Perkin Elmer. Because the human TRH receptor signals weakly in cell culture, studies were performed using the rat TRHR. Rat and human TRHR are highly homologous and identical from TM3 through most of TM4, fully encompassing ICL2 containing I131.

Fixed-cell ELISA. Human embryonic kidney (HEK293) cells were grown in DMEM medium with 5% fetal bovine serum and transfected with Lipofectamine 2000 as recommended by the manufacturer. The relative expression of HA-tagged TRH receptors on the plasma membrane was quantified by a previously described fixed-cell ELISA protocol using fixed but non-permeabilized cells (Jones *et al.*, 2007). Cells grown on a 48-well dish were transfected, grown for 24h, then washed and fixed for 10 min in 3% paraformaldehyde in PBS at room temperature. Cells were washed with PBS and 1:5000 anti-HA antibody (HA-11 from Covance) was added for 1 h in PBS/milk. Following additional washes, 1:5000 HRP-anti-mouse IgG (BioRad) in PBS/milk was added for 45 min

and the cells were washed and incubated for 2.5 min in TMB substrate (Sigma). The reaction was stopped with sulfuric acid and A_{450} measured. Background signal was determined in wells transfected with GFP in place of receptor and has been subtracted.

[³H]MeTRH binding assay. To measure the affinity of wild-type and mutant receptors, cells were plated in 6 or 12-well plates, transfected, and grown for 24 or 48 h. The media was then replaced with media containing 2nM [³H]MeTRH and various concentrations of unlabelled TRH and plates returned to the incubator for 1.5 h. Dishes were washed 3 times with PBS and cells dissolved in 0.1% SDS and radioactivity was counted. Non-specific binding was determined in parallel cultures transfected with GFP and has been subtracted; nonspecific binding was less than 5% of total. The relative affinity of receptors was calculated from the IC_{50} values using the Cheng-Prushoff equation and 3 nM as K_d for [³H]MeTRH. Binding and surface receptor determinations were performed in triplicate in each experiment and activity measurements were carried out in duplicate or triplicate. Figures show mean ± SEM or range of representative experiments; where not visible, error bars were within symbol size. All experiments were replicated at least 3 times. EC_{50} and IC_{50} values were determined using Prism Graphpad software.

Transient transfections and luciferase assays. Human embryonic kidney (HEK293) cells were grown in DMEM medium with 5% fetal serum and transfected with Lipofectamine 2000 as recommended by the manufacturer. For activity measurements, cells were grown on white 96-well plates and transfected with 50 ng/well containing approximately equal amounts of plasmids encoding AP1-luciferase reporter and receptor. In some experiments an M_3 muscarinic receptor was co-transfected as a control using 1/10 as much DNA as

the TRH receptor plasmid. After 24 h, fresh media containing TRH was added and incubation continued for 4 h when media was replaced with 50 µl luciferase substrate (Nanolight Firefly Reagent). Relative light units were recorded on a BioTek plate reader after 5-10 min.

Computational model of wild type and I131T mutant TRH receptors in complex with TRH and Gq. The "active-like" state of human TRHR (UniProt entry P34981) was built using a combination of structural templates. The crystal structure of active µ opioid receptor (PDB id 5C1M) (Huang *et al.*, 2015) was used for the construction of active TRHR (these receptors share 24% of sequence identity and 50% of sequence similarity). Such values are considered above the twilight zone because membrane proteins maintain a strong conservation of TMs even at low sequence identity (<20%) (Olivella *et al.*, 2013). Because the extracellular part of TM2 of TRHR contains P81$^{2.60}$ (Ballesteros and Weinstein, 1995) and ECL 1 contains W91VYG94, forming the highly conserved (W/F)x(F/L)G motif (Gonzalez *et al.*, 2012), amino acids V76$^{2.55}$-G94^{ECL1} were modelled using squid rhodopsin (PDB id 2Z73) (Murakami and Kouyama, 2008) as template. To study the effect of the I131^{ICL2}T mutation in the structure and function of TRHR an "active-like" model in complex with Gq was constructed. Thus, this "active-like" conformation of TRHR was modelled from the crystal structure of the β$_2$-adrenergic receptor (β$_2$-AR) in complex with Gs (PDB ID 3SN6) (Rasmussen *et al.*, 2011) by changing the conformation of the intracellular part of TMs 5 and 6 and ICL 2 of TRHR (A207$^{5.54}$ – N221$^{5.68}$, M267$^{6.36}$ – I273$^{6.42}$, P130^{ICL2} – F139^{ICL2}) for the active conformation of β$_2$-AR (M215$^{5.54}$ – E237$^{5.76}$, C265$^{6.27}$ – G280$^{6.42}$, P138^{ICL2} – K147^{ICL2}). These replacements open the intracellular cavity required for the binding of the C-terminal α5 helix of the G-protein.

Therefore, this "active-like" model of TRHR includes Gq, which was modelled in the "closed" conformation using the crystal structure of YM-254890-inactive Gq (PDB id 3AH8) (Nishimura *et al.*, 2010).

Modeller 9.12 was used to build this homology model (Martí-Renom *et al.*, 2000). TRH was docked into the "active-like" conformation of TRHR using MOE (Chemical computing group Inc., Montreal, QC, Canada) in such a manner that the pyroglutamic acid of TRH interacts with Y106$^{3.33}$ and N110$^{3.37}$ of TRHR, His with Y282$^{6.51}$, and the terminal Pro-NH$_2$ with R306$^{7.39}$, as has been proposed by site-directed mutagenesis (Perlman *et al.*, 1994, 1995).

Wild-type and I131^{ICL2}T mutation of the "active-like" models of TRHR were embedded in a pre-equilibrated box containing 263 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine (POPC) lipids with explicit solvent (~73000 water molecules) and a 0.15 M concentration of Na$^+$ and Cl$^-$ ions.

Each system was energy minimized and subjected to a 5 step molecular dynamics (MD) equilibration (10+5+2+2+2 ns). In the first step the whole system was fixed except hydrogen atoms; in the second step, the protein loops were released from restraints; and in the last three steps the restraints on the ligand and protein atoms were relaxed from 100, 50 to 10 kJ.mol^{-1}nm^{-2}, respectively. Unrestrained MD trajectories were produced for 500 ns using a 2 fs time step, constant pressure under semi-isotropic condition and constant temperature at 300 K. MD simulations were conducted with GROMACS 5.1.2 simulation package (Pronk *et al.*, 2013), using the AMBER ff99SB-ILDN force field for the protein, the parameters described by Berger for POPC lipids, and the Generalized Amber force field (GAFF) and HF/6-31g*-derived RESP atomic charges for

TRH. This combination of protein and lipid parameters has recently been validated (Cordomí, Caltabiano and Pardo, 2012).

4.1.1.3. *Results*

Clinical Case

The index case is a male of Roma descent, the third sibling of a consanguineous kindred (Figure 4.1). He was not detected at TSH-based neonatal screening program (using TSH threshold >7 mIU/L). At the age of 8 years he was referred to the pediatrician for evaluation of abnormal thyroid function. Hormonal tests revealed mild hypothyroidism (FT4: 9.52 pmol/L -N: 10.9-25.7 pmol/L- and TSH: 2.61 mIU/L -N: 0.27-4.2 mIU/L-). He presented overweight (BMI: 20.4 kg/m², +1.64 SD), but normal stature (122 cm, -0.58 SD) (Figure 4.1). The patient did not present symptoms of hypothyroidism, but thyroid hormone profiles persistently showed abnormally elevated TSH/FT4 ratio (0.274, N: 0.027-0.131), suggesting partial TSH insufficiency (Dietrich, Landgrafe and Fotiadou, 2012). The patient displayed normal TSH response to TRH. The ACTH dynamic test and basal PRL, FSH, LH, IGF-1 and IGFBP-3 determinations were normal. His testicular volume was 3 ml, consistent with a prepubertal stage. Brain MRI showed normal size and shape of the pituitary. Thyroid ultrasounds showed normal thyroid size and structure (data not shown). The patient was started on L-thyroxine (L-T4) replacement (50 μg/day) and one month later his FT4 levels reached the normal range (13.77 pmol/L) at the expense of suppressed TSH (0.07 mIU/L), a characteristic feature of treated central hypothyroidism (Persani, 2012).

Case	I.1	III.1	IV.1	IV.2	IV.3	IV.4	III.2
TSH mU/L (0.27-4.2)	5.17	1.3	2.95	8.05	2.61	4.65	2.71
FT4 pmol/L (10.9-25.7)	N	13.88	17.73	13.62	9.52	12.46	14.03
TSH/FT4 (0.027-0.131)	-	0.094	0.166	0.591	0.274	0.373	0.193
Height cm (-2…+2 SD)	-	171 -0.76	-	-	122 -0.58	-	155.4 -1.02
Genotype p.I131T (c.392T>C)	+/-	+/-	+/-	+/-	+/+	+/+	+/-
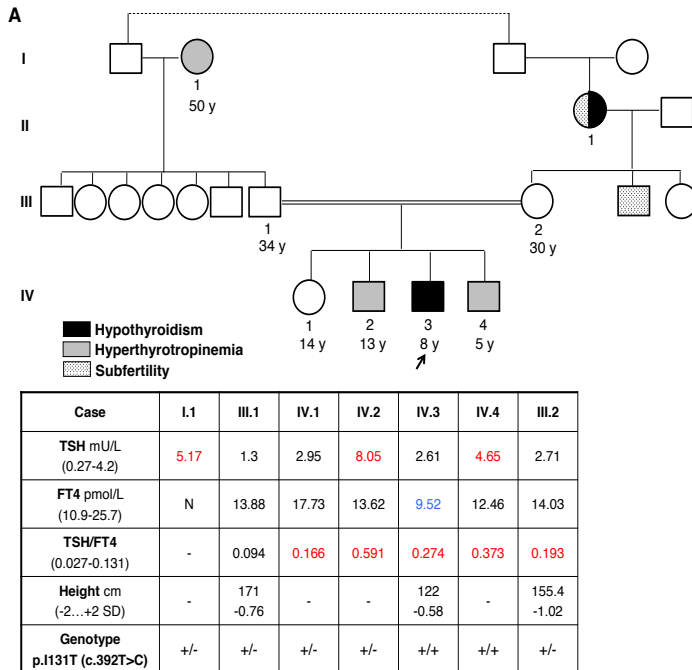
Figure 4.1| Clinical and biochemical features of a family with TRHR defect. Phenotype of members of the pedigree showing consanguinity, expressed as a double line linking symbols for mother and father of the index patient (arrow). Thyroid hormone profile of several members of the family. In blue and red represent hormone values or index below and above normal ranges, respectively.

Remarkably, several members of the family (two brothers, the mother and the paternal grandmother of index patient) showed hyperthyrotropinemia with normal FT4 (Figure 4.1). All (heterozygotes) relatives studied, with the exception of the patient's father, had abnormally elevated TSH/FT4 ratio, consistent with a compensation of the failure at the expense of (low-bioactivity) TSH increase (Figure 4.1) (Dietrich, Landgrafe and Fotiadou, 2012). In

order to complete (new sibiling) and re-evaluate the thyroid function in the patient's family, a TRH test was performed in four siblings and parents of the patient showing normal TSH and prolactin responses to TRH in all relatives (Figure 4.2).

Case	III.1	IV.1	IV.2	IV.3	IV.4	IV.5	III.2
TSH mU/L (0.27-4.2)	0.98	2.09	8.34	5.68	2.77	1.29	1.93
FT4 pmol/L (10.9-25.7)	14.67	14.93	13.13	10.55	13.51	14.03	13.13
TSH/FT4 (0.027-0.131)	0.067	0.140	0.635	0.538	0.205	0.090	0.147
Genotype p.I131T (c.392T>C)	+/-	+/-	+/-	+/+	+/+	+/-	+/-
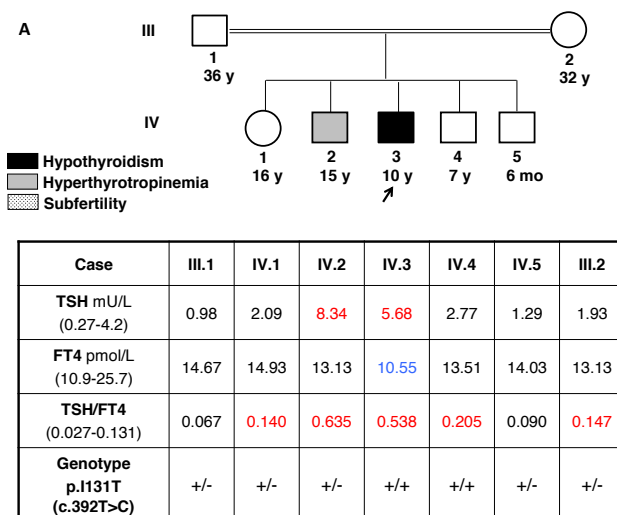
Figure 4.2 | Re-evaluation of the family members' thyroid function two years after diagnosis. Thyroid hormone profile of the individuals belonging family nucleus. L-T4 was withdrawn during 1 month in the proband. In blue and red represent hormone values or index below and above normal ranges, respectively.

Identification of the TRHR mutation

Direct sequencing of the coding exons of four candidate genes for central hypothyroidism (*TRH, TRHR, TSHB and IGSF1*) revealed a homozygous missense mutation in the *TRHR* gene of the patient

(c.392T>C), changing Isoleucine 131 into Threonine (p.I131T). The patient's mutation was inherited from his parents, who are both heterozygous carriers. Moreover, the paternal grandmother and three siblings of the patient are heterozygotes for the mutation, while one of his brothers is homozygous carrier (Figures 4.1, 4.2 and 4.3).
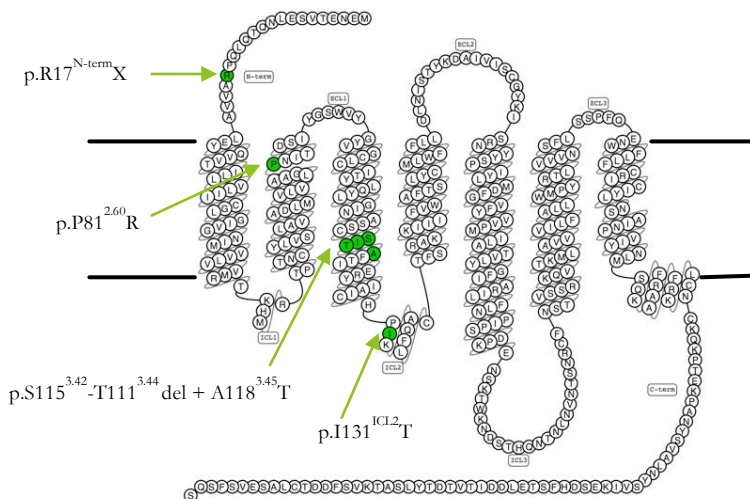
Figure 4.3| TRHR mutations. Scheme showing the location of the I131T mutation at the second intracellular loop of the TRHR (highlighted) and other TRHR mutations previously described. Figure generated using GPCRdb (Pándy-Szekeres *et al.*, 2018; Munk *et al.*, 2019).

Functional characterization of the I131T-TRHR mutant

To determine whether the Isoleucine to Threonine mutation affected the expression and trafficking of the receptor, the relative amount of HA-tagged receptors at the plasma membrane was measured on fixed but non-permeabilized cells. Wild-type TRHR were strongly expressed on the plasma membrane and there was no significant difference in receptor density when cells were transfected with equal

amounts of cDNA encoding the I131T mutant or a 1:1 mixture of wild-type and mutant TRHR (Figure 4.4A). The molecular basis for the signalling defect in the I131T-TRHR was investigated by measuring the affinity of wild-type and mutant TRHR for a radiolabelled agonist, [³H]Me-TRH, a high-affinity TRH analog. Cells were incubated with tracer [³H]Me-TRH and different concentrations of unlabelled TRH under equilibrium conditions. Significantly higher concentrations of unlabelled TRH were required to decrease [³H]Me-TRH binding to the I131T-TRHR, consistent with lower affinity for the natural ligand (Figure 4.4B). The relative affinity of the two receptors for TRH was 3.1 ± 0.3 and 9.1 ± 0.4nM for wild-type and I131T mutant TRH receptors, respectively (p<0.05).
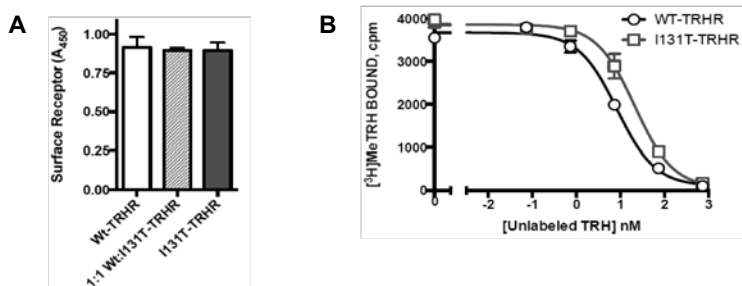
Figure 4.4| **I131T-TRHR functional studies.** (**A**) I131T-TRH receptors are located at the cell membrane at the same density as the wild-type. HA-tagged TRH receptors on the cell surface were measured by ELISA in the experiment depicted in panel B. (**B**) I131T-TRHR has reduced affinity for [³H]Me-TRH. Cells transfected with control plasmid or TRH receptors were incubated with 2 nM [³H]methyl-TRH and concentrations of unlabeled TRH shown and specific binding was measured after 1.5 h.

The effect of the I131T mutation in ICL2 of TRHR on its signalling capacity was tested by expressing wild-type and I131T-TRHR in

HEK293 cells, a widely used system for evaluating GPCR signalling. Because the human TRH receptor signals weakly in standard cell culture models, the studies were performed using the rat TRHR. Rat and human TRHR are highly homologous and identical from TM3 through most of TM4, fully encompassing ICL2 containing I131. The TRH receptor signals through a classical Gq-coupled pathway, stimulating an increase in intracellular calcium and activation of PKC (Hinkle, Nelson and Ashworth, 1996). Receptor activity was determined using an AP1-luciferase reporter containing a c-fos promoter sequence activated by TRH-Gq-IP-PKC pathway. Stimulation with TRH induced more than a 10-fold increase in AP1-luciferase activity in cells expressing the wild-type TRHR. The average EC_{50} for TRH was 2.8 ± 0.9nM (n=7). I131T-TRHR was capable of generating the same maximal response, however significantly higher concentrations of TRH were required: average $EC_{50} = 20.4 \pm 0.8$nM (n=6, p<0.05 vs. wild-type) (Figure 4.5A).

Signalling via the TRH receptor was also tested in cells co-expressing wild-type and I131T-TRHR mutant receptors in an attempt to mimic the situation in individuals heterozygous for the mutation. Because activity of the two receptors differed more at low concentrations of TRH, responses were determined either without any stimulus (constitutive activity), low TRH (1 and 5 nM), and maximally effective TRH (1 µM) (Figure 4.5B). As expected, the I131T-TRHR signalled weakly compared to wild-type at 1 and 5 nM TRH (Figure 4.4B). At maximal TRH concentrations, no significant differences in activity between mutant and wild-type (p>0.05) were present, consistent with the capacity of the mutant to generate maximal responses (Figure 4.5B). Cells co-expressing wild-type and mutant receptors responded as strongly as wild-type receptors. Activity in the absence of TRH, i.e.

constitutive activity, was not affected by mutation of I131, indicating that functional impairment of I131T-TRHR involves the ligand-activation mechanism of the receptor (Figure 4.5B).
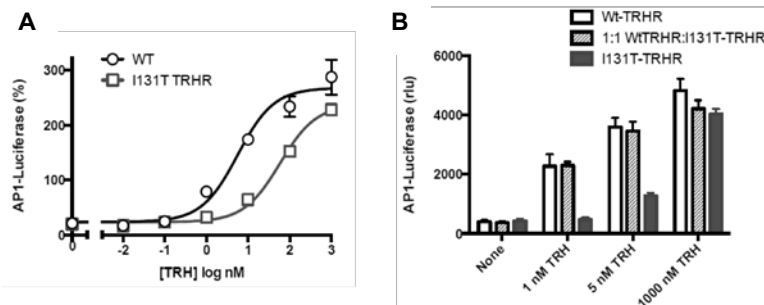
Figure 4.5| I131T-TRHR functional studies. Cells were transfected to express wild-type and/or mutant TRH receptors and an AP1-luciferase reporter. After 24 h cells were stimulated for 4 h with the concentrations of TRH shown and luciferase activity was quantified. (**A**) Response to different concentrations of TRH showing that the I131T TRHR has a higher EC_{50} but similar maximum response compared to the wild-type receptor. The M3 muscarinic receptor was co-transfected as a control for downstream effects; results are expressed as percent of the response to 10 μM carbachol. (**B**) I131T- and wild type-TRHR were transfected alone or together using a 1:1 DNA ratio and stimulated with the concentrations of TRH shown; luciferase activity is not normalized.

I131T TRHR mutant disrupts the interaction with Gq in the active-like TRH-TRHR-Gq model

An "active-like" model of the TRHR in complex with Gq was built up (Figure 4.6A, see *sub-section 4.1.1.2. Materials and Methods*). I131[ICL2] is located in ICL2 and pointing towards Gq. The molecular interface between TRHR and Gq is mainly formed by the interaction of TMs 3, 5 and 6 of TRHR with the C-terminal α5 helix of Gq, and ICL 2

of TRHR with the N-terminal $\alpha1$ helix and $\beta1$ and $\beta3$ strands of Gq. In detail, the hydrophobic side chain of I131^{ICL2} forms hydrophobic-hydrophobic interactions with L40$^{\beta1}$ in the $\beta1$ strand, V199$^{\beta3}$ in the $\beta3$ strand, and V344$^{\alpha5}$ and I348$^{\alpha5}$ in the $\alpha5$ helix of Gq (Figure 4.6B). These interactions remain stable during the MD simulation (see below and Figure 4.6C). Statistical analysis of sequence conservation at this I131^{ICL2} position, in the non-olfactory class A GPCR superfamily, was performed in a refined multiple sequence alignment. Among the 287 sequences reported in GPCRdb (Isberg *et al.*, 2014), we selected only 247 sequences that contain at least one of the two characteristic signatures of ICL2: a Pro residue at position (E/D)R$^{3.50}$YX$_5$P (Ballesteros and Weinstein, 1995) that starts the two-turn α-helix conformation of ICL2 and a Tyr residue at position (E/D)R$^{3.50}$YX$_5$PX$_2$$\underline{Y}$ that interacts with the negative charge of the (E/D)RY motif in TM3 (Venkatakrishnan *et al.*, 2013). Clearly, this refined sequence alignment shows that most GPCRs contain a hydrophobic residue at the homologous I131^{ICL2} position (L, 39%; F, 14%; V, 11%; I, 9%). Similar analysis of sequence conservation in the G protein family, at the homologous L40$^{\beta1}$, V199$^{\beta3}$, V344$^{\alpha5}$ and I348$^{\alpha5}$ positions (inset of Figure 4.6A), shows that this hydrophobic pocket is conserved but with some degree of variation.

In order to evaluate the effect of the I131^{ICL2}T mutation in the TRHR-Gq interface we performed MD simulations of wild-type (Figure 4.6C) and mutant (Figures 4.6D-E) receptors of the "active-like" model of TRHR in complex with Gq (see sub-section *4.1.1.2. Materials and Methods*). Replacement of I131^{ICL2} by Thr adds a polar hydroxyl group at this TRHR-Gq interface. However, this small change in polarity causes a significant distortion. While the

hydrophobic I131^{ICL2} side chain maintains the interactions with L40^{β1}, V199^{β3}, V344^{α5} and I348^{α5} of Gq during the MD simulation (Figure 4.6C), the polar side chain of T131^{ICL2} partly modifies the interaction of the α5 helix with TRHR (Figure 4.6E) to interact with polar side chain of K3445^{α5} (Figure 4.6D). Therefore, insertion of a polar Thr side chain into this hydrophobic pocket disrupts TRHR-Gq coupling.
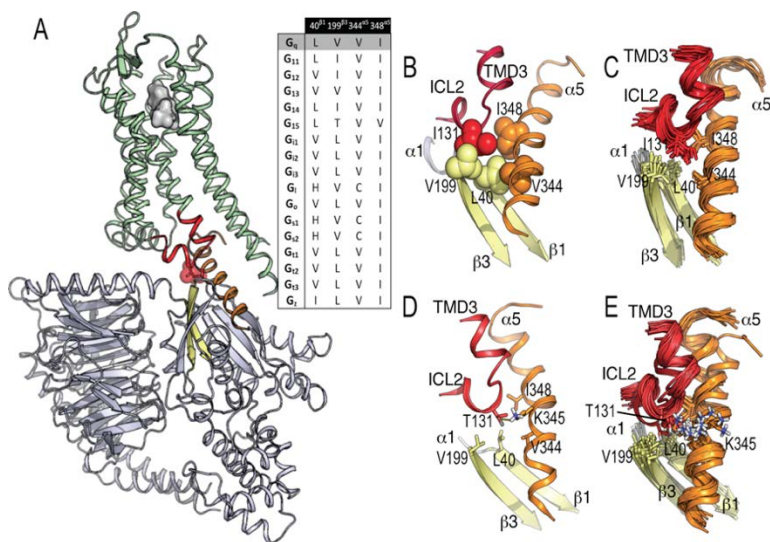
Figure 4.6| I131T-TRHR Computational model of the TRH-TRHR-Gq complex. (A) General view of the "active-like" model of TRHR in complex with Gq at the intracellular site and TRH at the extracellular site. TRH is shown as a gray surface, TRHR is shown as ribbons in green (intracellular TM3 and ICL2 are shown in red), Gq in blue white (β1 and β3 strands in yellow and the α5 helix in orange), and I113 in red spheres. Statistical analysis of sequence conservation in the G protein family at the homologous L40^{β1}, V199^{β3}, V344^{α5} and I348^{α5} positions of Gq. **(B)** Detailed view of the interaction of

I131ICL2 (wild-type) in ICL 2 of TRHR (in red) with $L40^{\beta1}$ in the β1 strand (in yellow), $V199^{\beta3}$ in the β3 strand (in yellow), and $V344^{\alpha5}$ and $I348^{\alpha5}$ in the α5 helix (in orange) of Gq. (**C**) Cartoon representation of twenty snapshots extracted from 500 ns MD trajectory of wild-type TRHR. The side chains of $I113^{ICL2}$, $L40^{\beta1}$, $V199^{\beta3}$, $V344^{\alpha5}$ and $I348^{\alpha5}$ are shown as sticks. (**D**) Detailed view of the interaction of $T131^{ICL2}$ (TRHR mutant) with $L40^{\beta1}$, $V199^{\beta3}$, $V344^{\alpha5}$, $K345^{\alpha5}$, and $I348^{\alpha5}$ of Gq. (**E**) Cartoon representation of twenty snapshots extracted from 500 ns MD trajectory of $I131^{ICL2}T$ mutant receptor.

4.1.1.4. *Discussion*

A novel missense TRHR mutation was identified in a consanguineous family causing central hypothyroidism in homozygotes and isolated TSH elevation in heterozygous carriers of the defect. The study suggests that two affected TRHR alleles with such mutation are necessary to develop the full hormone phenotype of hypothyroidism, and expands the scope of thyroid hormone derangements associated with TRHR mutations to include hyperthyrotropinemia, when one allele is affected.

TRHR defects are extremely rare and only three unrelated patients with central hypothyroidism have been described with TRHR mutations (Collu *et al.*, 1997; Bonomi *et al.*, 2009; Koulouri *et al.*, 2016). The three patients were missed by TSH-based CH screening programs and came to clinical attention at varying ages of 2 months, 7 and 9 years, respectively. They were biochemically hypothyroid, but clinical hypothyroidism was not overt or expressed through unspecific symptoms like short stature, neonatal jaundice, fatigue, poor school performance.

The index patient came to medical attention at the age of 8 years because of abnormal thyroid function tests in a routine check-up. Besides overweight, features of hypothyroidism were not overt, but low FT4 and normal TSH suggested central hypothyroidism. Interestingly, overweight or obesity was present in 38% of children with central hypothyroidism due to *IGSF1* defects (Sun *et al.*, 2012). Unlike previous cases with biallelic TRHR mutations, our patient showed normal TSH and PRL responses to TRH (Collu *et al.*, 1997; Bonomi *et al.*, 2009).

Retrospectively investigated, at 6 years of age the patient yet showed subtle hypothyroidism and hyperthyrotropinemia. Strikingly, the youngest sibling of the proband, also homozygous for the defect, shows isolated hyperthyrotropinemia at the age of 5 years. This suggests that TSH elevation may precede the development of overt hypothyroidism in homozygotes and represent a compensatory state which eventually fails along with increased demands for thyroid hormones. Three further family members, heterozygotes for the defect also present isolated hyperthyrotropinemia with elevated TSH/FT4 ratios (Dietrich, Landgrafe and Fotiadou, 2012).

I131T is the first missense mutation identified in the ICL2 of the TRHR and locates at a highly conserved hydrophobic position (Φ) of the class A GPCRs at the $(E/D)R^{3.50}YX_5P\Phi XY$ motif, which is important for the essential conformational changes required for G proteins activation (Venkatakrishnan *et al.*, 2013).

The three known mutations in TRHR (R17X, S115-T117del+A118T and P81R) severely impair TRH binding and completely abrogate signal transduction (Collu *et al.*, 1997; Bonomi *et al.*, 2009; Koulouri *et al.*, 2016). In contrast, the novel I131T TRHR mutant causes a 3-fold

reduction in TRH affinity and a 7.3-fold shift in the concentration-response curve for activation of the Gq-IP$_3$-PKC pathway yet binds and responds normally to high concentrations of TRH, which might relate with the milder phenotype of hypothyroidism in the family.

The functional differences between the novel I131T mutation and the previously described TRHR mutations correlate well with the signalling capacity of the affected receptors (Collu *et al.*, 1997; Bonomi *et al.*, 2009; Koulouri *et al.*, 2016). The previous mutations all showed deleterious effects on receptor function. R17X results in a prematurely truncated protein, missing all seven TMs. S115-T117del+A118T deeply alters the tertiary structure of the TM3 (essential to receptor function) and P81R allegedly alters the conformation of TM2 and therefore the TRH binding pocket. The model of TRHR in complex with Gq (Figure 4.6) shows that the hydrophobic I131^{ICL2} side chain of TRHR interacts with a hydrophobic pocket formed by L40$^{\beta 1}$, V199$^{\beta 3}$, V344$^{\alpha 5}$ and I348$^{\alpha 5}$ of Gq. Thus, it seems reasonable to propose that insertion of a polar Thr side chain into this hydrophobic pocket of Gq disrupts TRHR-Gq coupling. Accordingly, higher concentrations of TRH are necessary to achieve the same maximal transcriptional response by the I131T mutant with respect to wild type TRHR (Figure 4.4A). Mutation of this hydrophobic Φ residue in other GPCRs has similar effects but with different extent of G protein uncoupling (Moro *et al.*, 1993; Arora, Sakai and Catt, 1995; Burstein, Spalding and Brann, 1998; Gáborik *et al.*, 2003).

The mechanism by which the I131^{ICL2}T mutation at the intracellular site influences the binding of TRH to TRHR at the extracellular site is clearly indirect. Two related mechanisms have been yet proposed

to explain this effect. First, the constitutive activity of the receptor (equilibrium between inactive and active states in the absence of ligand) modulates ligand affinity and selectivity (Montanelli *et al.*, 2004). Second, there is an allosteric coupling from G protein to the agonist-binding pocket in GPCRs (DeVree *et al.*, 2016). Since the I131T mutation does not influence the constitutive activity of TRHR (Figure 4.5B), we suggest that the uncoupling of Gq to I131T-TRHR at the intracellular site, causes the decrease of TRH binding to TRHR at the extracellular site (Figure 4.4B). This work provides a structural explanation for the G-protein-mediated enhancement of agonist affinity.

Besides *TSHB* transcription, TRHR regulates glycosylation of TSH, which is required for the full biopotency of the TSH dimer (Weintraub *et al.*, 1989; van Tijn, de Vijlder and Vulsma, 2008). Patients with central hypothyroidism, especially of hypothalamic origin, have been described with low bioactive TSH and sometimes slightly elevated immunoreactive TSH (Persani, 2012). The I131T TRHR mutant seems to hinder the expected increase of FT4 after TRH stimulation in vivo, suggesting impaired TSH bioactivity. As reported in *Trh*[-/-] knockout mice (Rabeler *et al.*, 2004) (and also *Trh*[+/-] mice to a lesser extent), decreased TRH-TRHR signalling in thyrotropes might be responsible for the development of hyperthyrotropinemia with low TSH biopotency in carriers of I131T (Yamada *et al.*, 1997). In the absence of TRH action, hyperthyrotropinemia exhibited by *Trh*-deficient mice is explained by decreased negative feedback of thyroid hormones at the pituitary, leading to increased synthesis of a TSH with low biological activity. We propose a similar situation in human pituitaries with a partially defective I131T TRHR mutant.

In summary, we identified a novel missense mutation (p.I131T) in *TRHR* associated with overt central hypothyroidism in the biallelic state. Although inheritance of TRHR defects is typically recessive, we described for the first time the presence of central hyper-thyrotropinemia in heterozygous carriers of this mutation. Hyper-thyrotropinemia is proposed as a compensatory state preceding hypothyroidism in homozygotes. In individuals heterozygous for I131T TRHR, central hyperthyrotropinemia is present with normal T4. However, during pregnancy development of hypothyroidism should be specifically ruled out. The I131T mutation in the second intracellular loop of the receptor partially impairs TRH binding and coupling to Gq. Undiagnosed central hypothyroidism in children calls for a higher degree of suspicion from pediatric endocrinologists dealing with patients with borderline-low T4 and normal TSH or with isolated TSH elevations associated with abnormally elevated TSH/T4 ratios due to thyrotropic failure (Dietrich, Landgrafe and Fotiadou, 2012).

4.1.1.5. *References*

Adachi, M. *et al.* (2012) 'Mass screening of newborns for congenital hypothyroidism of central origin by free thyroxine measurement of blood samples on filter paper.', *Eur. J. Endocrinol.*, 166(5), pp. 829–38.

Arora, K. K., Sakai, A. and Catt, K. J. (1995) 'Effects of Second Intracellular Loop Mutations on Signal Transduction and Internalization of the Gonadotropin-releasing Hormone Receptor', *J. Biol. Chem.*, 270(39), pp. 22820–22826.

Ballesteros, J. and Weinstein, H. (1995) 'Integrated methods for the construction of three-dimensional models of structure−function relations in G protein-coupled receptors.', *Methods Neurosci.*, 25, pp. 366–428.

Bílek, R. and Stárka, L. (2005) 'The computer modelling of human TRH receptor, TRH and TRH-like peptides.', *Physiol. Res.*, 54(2), pp. 141–50.

Bonomi, M. *et al.* (2009) 'A family with complete resistance to thyrotropin-releasing hormone.', *N. Engl. J. Med.*, 360(7), pp. 731–4.

Burstein, E. S., Spalding, T. A. and Brann, M. R. (1998) 'The second intracellular loop of the m5 muscarinic receptor is the switch which enables G-protein coupling', *J. Biol. Chem.*, 273(38), pp. 24322–24327.

Collu, R. *et al.* (1997) 'A novel mechanism for isolated central hypothyroidism: inactivating mutations in the thyrotropin-releasing hormone receptor gene.', *J. Clin. Endocrinol. Metab.*, 82(5), pp. 1561–5.

Colson, a O. *et al.* (1998) 'A hydrophobic cluster between transmembrane helices 5 and 6 constrains the thyrotropin-releasing hormone receptor in an inactive conformation.', *Mol. Pharmacol.*, 54(6), pp. 968–978.

Cordomí, A., Caltabiano, G. and Pardo, L. (2012) 'Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters', *J. Chem. Theory Comput.*, 8(3), pp. 948–958.

DeVree, B. T. *et al.* (2016) 'Allosteric coupling from G protein to the agonist-binding pocket in GPCRs', *Nature*, 535(7610), pp. 182–186.

Dietrich, J. W., Landgrafe, G. and Fotiadou, E. H. (2012) 'TSH and Thyrotropic Agonists: Key Actors in Thyroid Homeostasis.', *J. Thyroid Res.*, 2012, p. 351864.

Engel, S. and Gershengorn, M. C. (2007) 'Thyrotropin-releasing hormone and its receptors--a hypothesis for binding and receptor activation.', *Pharmacol. Ther.*, 113(2), pp. 410–419.

Fares, F. (2006) 'The role of O-linked and N-linked oligosaccharides on the structure-function of glycoprotein hormones: development of agonists and antagonists.', *Biochim. Biophys. Acta*, 1760(4), pp. 560–7.

Gáborik, Z. *et al.* (2003) 'The role of a conserved region of the second intracellular loop in AT1 angiotensin receptor activation and signaling', *Endocrinology*, 144(6), pp. 2220–2228.

García, M., Fernández, A. and Moreno, J. C. (2014) 'Central hypothyroidism in children.', *Endocr. Dev.*, 26, pp. 79–107.

Gonzalez, A. *et al.* (2012) 'Impact of Helix Irregularities on Sequence Alignment and Homology Modeling of G Protein-Coupled Receptors', *ChemBioChem*, 13(10), pp. 1393–1399.

Hinkle, P. M., Gehret, A. U. and Jones, B. W. (2012) 'Desensitization, Trafficking, and Resensitization of the Pituitary Thyrotropin-Releasing Hormone Receptor', *Front. Neurosci.*, 6, p. 180.

Hinkle, P. M., Nelson, E. J. and Ashworth, R. (1996) 'Characterization of the calcium response to thyrotropin-releasing hormone in lactotrophs and GH cells.', *Trends Endocrinol. Metab.*, 7(10), pp. 370–4.

Huang, W. *et al.* (2015) 'Structural insights into μ-opioid receptor activation', *Nature*, 524(7565), pp. 315–321.

Isberg, V. *et al.* (2014) 'GPCRDB: an information system for G protein-coupled receptors.', *Nucleic Acids Res.*, 42, pp. D422-5.

Jones, B. W. *et al.* (2007) 'Phosphorylation of the endogenous thyrotropin-releasing hormone receptor in pituitary GH3 cells and pituitary tissue revealed by phosphosite-specific antibodies.', *J. Biol. Chem.*, 282(17), pp. 12893–906.

Koulouri, O. *et al.* (2016) 'A Novel Thyrotropin-Releasing Hormone Receptor Missense Mutation (P81R) in Central Congenital Hypothyroidism.', *J. Clin. Endocrinol. Metab.*, 101(3), pp. 847–51.

Martí-Renom, M. A. *et al.* (2000) 'Comparative protein structure modeling of genes and genomes.', *Annu. Rev. Biophys. Biomol. Struct.*, 29, pp. 291–325.

Montanelli, L. *et al.* (2004) 'Modulation of ligand selectivity associated with activation of the transmembrane region of the human follitropin receptor.', *Mol. Endocrinol.*, 18(8), pp. 2061–73.

Moro, O. *et al.* (1993) 'Hydrophobic amino acid in the i2 loop plays a key role in receptor-G protein coupling', *J. Biol. Chem.*, 268(30), pp. 22273–22276.

Murakami, M. and Kouyama, T. (2008) 'Crystal structure of squid rhodopsin.', *Nature*, 453(7193), pp. 363–7.

Nishimura, A. *et al.* (2010) 'Structural basis for the specific inhibition of heterotrimeric G q protein by a small molecule', *Proc. Natl. Acad. Sci. U. S. A.*, 107(31), pp. 13666–71.

Olivella, M. *et al.* (2013) 'Relation between sequence and structure in membrane proteins', *Bioinformatics*, 29(13), pp. 1589–1592.

Pándy-Szekeres, G. *et al.* (2018) 'GPCRdb in 2018: Adding GPCR structure models and ligands', *Nucleic Acids Res.*, 46(D1), pp. D440–D446

Perlman, J. H. *et al.* (1994) 'A Model of the Thyrotropin-releasing Hormone (TRH) Receptor Binding Pocket', *J. Biol. Chem.*, 269, pp. 23383–23386.

Perlman, J. H. *et al.* (1995) 'Distinct roles for arginines in transmembrane helices 6 and 7 of the thyrotropin-releasing hormone receptor.', *Mol. Pharmacol.*, 47(3), pp. 480–4.

Persani, L. (2012) 'Clinical review: Central hypothyroidism: pathogenic, diagnostic, and therapeutic challenges.', *J. Clin. Endocrinol. Metab.*, 97(9), pp. 3068–78.

Pronk, S. *et al.* (2013) 'GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.', *Bioinformatics*, 29(7), pp. 845–54.

Rabeler, R. *et al.* (2004) 'Generation of thyrotropin-releasing hormone receptor 1-deficient mice as an animal model of central hypothyroidism.', *Mol. Endocrinol.*, 18(6), pp. 1450–60.

Rasmussen, S. G. F. *et al.* (2011) 'Crystal structure of the β2 adrenergic receptor–Gs protein complex', *Nature*. Nature Publishing Group, 477(7366), pp. 549–555.

Sun, Y. *et al.* (2012) 'Loss-of-function mutations in IGSF1 cause an X-linked syndrome of central hypothyroidism and testicular enlargement.', *Nat. Genet.*, 44(12), pp. 1375–81.

van Tijn, D. A. *et al.* (2005) 'Neonatal detection of congenital hypothyroidism of central origin.', *J. Clin. Endocrinol. Metab.*, 90(6), pp. 3350–9.

van Tijn, D. A., de Vijlder, J. J. M. and Vulsma, T. (2008) 'Role of the thyrotropin-releasing hormone stimulation test in diagnosis of congenital central hypothyroidism in infants.', *J. Clin. Endocrinol. Metab.*, 93(2), pp. 410–9.

Venkatakrishnan, A. J. *et al.* (2013) 'Molecular signatures of G-protein-coupled receptors', *Nature*. Nature Publishing Group, 494(7436), pp. 185–194.

Weintraub, B. D. *et al.* (1989) 'Effect of TRH on TSH glycosylation and biological action.', *Ann. N. Y. Acad. Sci.*, 553(7), pp. 205–13.

Yamada, M. *et al.* (1997) 'Tertiary hypothyroidism and hyperglycemia in mice with targeted disruption of the thyrotropin-releasing hormone gene.', *Proc. Natl. Acad. Sci. U. S. A.*, 94(20), pp. 10862–7.

Zhu, C. C., Cook, L. B. and Hinkle, P. M. (2002) 'Dimerization and phosphorylation of thyrotropin-releasing hormone receptors are modulated by agonist stimulation', *J. Biol. Chem.*, 277, pp. 28228–28237.

4.1.2. LIGAND-TRIGGERED STRUCTURAL CHANGES IN THE M₂ MUSCARINIC ACETYLCHOLINE RECEPTOR

4.1.2.1. *Background*

Muscarinic acetylcholine receptors (mAChRs) compose a subfamily of receptors for acetylcholine (five distinct subtypes M_1-M_5) that are widely distributed throughout the peripheral and the central nervous system in humans and other organisms including invertebrates (Gregory, Sexton and Christopoulos, 2007; Langmead, Watson and Reavill, 2008). Their function involves the increase of exocrine secretions, contraction of the cardiac and smooth muscles in the gastrointestinal tract and lung and reduction of the heart rate. They are important for learning, memory and attention mechanisms, motor control, nociception and regulation of the sleep-wake cycle (Jakubík and El-Fakahany, 2010). Muscarinic receptors belong to the family of *Rhodopsin*-like G protein-coupled receptors (GPCRs) (Lagerström and Schiöth, 2008). They can signal both through G protein-dependent or independent pathways (Hulme, Birdsall and Buckley, 1990; Felder, 1995).

Advances in crystallization and high-resolution X-ray techniques (Trzaskowski *et al.*, 2012) have allowed a great increase in the number of crystallographic structures of GPCRs (Katritch, Cherezov and Stevens, 2013). These structures confirmed the common architecture of seven TM α-helices bundle linked by an extracellular N-terminal domain, three intracellular loops (ICL1 to ICL3), three extracellular loops (ECL1 to ECL3) and an intracellular C-terminus containing an α-helix (H8) parallel to the cell membrane (Liapakis *et al.*, 2012). They have also provided direct demonstration of the conformational

changes that occur in the process of GPCR signal transduction (Rasmussen *et al.*, 2011; Ring *et al.*, 2013; Mahoney and Sunahara, 2016). Such global conformational changes upon activation include the TM6 intracellular outward movement (Scheerer *et al.*, 2008), the transmission switch (Deupi and Standfuss, 2011) and distinct TM5-TM7 interactions (Park *et al.*, 2008; Goncalves *et al.*, 2010).

For the mAChRs, the structures of human M_1, M_2 and M_4 (Haga *et al.*, 2012; Kruse *et al.*, 2013; Thal *et al.*, 2016) and rat M_3 (Kruse *et al.*, 2012; Thorsen *et al.*, 2014) receptors have been resolved. Moreover, for the M_2R the available structures represent different conformational states. In 2012, the M_2R was first crystallized in complex with the high-affinity inverse agonist R-(−)-3-quinuclidinyl-benzilate (QNB) (Haga *et al.*, 2012) (PDB ID 3UON). In 2013, crystal structures of the M_2R were released in complex with a "G-protein mimetic" nanobody (Nb) bound to the high-affinity agonist iperoxo alone (PDB ID 4MQS) and in combination with LY2119620 (PDB ID 4MQT) (Kruse *et al.*, 2013), a positive allosteric modulator that binds in an extracellular site just above the orthosteric cavity. The allosteric modulator LY2119620 (3-amino-5-chloro-*N*-cyclopropyl-6-methoxy-4-methyl-thieno[2,3-b]pyridine-2-carboxamide) (Croy *et al.*, 2014; Schober *et al.*, 2014) displays strong positive cooperativity with iperoxo and mild negative cooperativity with the inverse agonist [3H]-N-methyl scopolamine ([3H]-NMS). Moreover, it enhances the affinity of M_2R for the orthosteric agonist iperoxo without changing its efficacy and furthermore may also directly activate the M_2R but with low potency and efficacy (Kruse *et al.*, 2013).

GPCRs are allosteric proteins that respond to extracellular stimuli via the orthosteric site and transmit the stimulus to an intracellular domain, the G protein–binding site, to transduce the signal

(Christopoulos and Kenakin, 2002). The reciprocal effects that orthosteric ligands and G proteins exert on one another support this allosterism (Nygaard *et al.*, 2013; van der Westhuizen *et al.*, 2015) a fact that it is also compatible with the existence of strong patterns of co-evolution between aforementioned sites (Süel *et al.*, 2003). The availability of M_2R crystallographic structures with agonist-bound active (with and without an allosteric modulator) and inverse-agonist bound inactive conformations, renders the M_2R an interesting receptor to study such molecular mechanisms of activation and allosterism at the molecular level. Accordingly, in the present study we have employed unbiased molecular dynamics (MD) simulations with pharmacologically distinct M_2R ligands. Since activation of GPCRs is difficult to achieve in MD simulations (Dror *et al.*, 2011; Miao *et al.*, 2013; Miao, Nichols and McCammon, 2014; Miao and McCammon, 2016), we chose to assess the reverse process: receptor deactivation by analysing conformational changes triggered by different types of ligands. We show that binding of the inverse-agonist to the M_2R's active state initiates structural changes that lead to the already structurally characterized inactive state. The results illustrate a putative sequence of events that follow the binding of a ligand towards receptor activation or vice versa and the effect of an allosteric modulator on the increase in affinity for the orthosteric ligand.

4.1.2.2. *Methods*

System preparation. The coordinates for the active M_2R (including the nanobody) were taken from the crystal structure of the M_2R in complex with agonist iperoxo (PDB ID 4MQS). This structure lacks the 178 amino acids-long ICL3 loop, which in our model was replaced

by 6 alanines, modeled using the loop routines of Modeller 9.12 software (Sali, 1995). Acidic and basic residues were modeled in the protonation state they would exhibit at pH 7 using PDB2PQR (Dolinsky *et al.*, 2004). The only exception was Asp103[3.32] (the counter-ion for the charged amine group present in all orthosteric ligands), which was considered protonated in the systems without orthosteric ligand.

Each ligand-receptor complex was embedded in a pre-equilibrated cubic bilayer box containing 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine (POPC) lipids. The initial orientation of the receptor relative to the membrane was guided based on the Orientations of Proteins in Membranes (OPM) and the PDB_TM databases (Tusnády, Dosztányi and Simon, 2004; Lomize *et al.*, 2006). Each systems' size was set to 8.5 x 8.1 x 10.3 Å (in absence of the nanobody) or 8.5 x 8.5 x 11.5 Å (in presence of the nanobody) using ~180 POPC molecules, ~14000 or ~16000 (in absence or presence of the nanobody, respectively) water molecules and a 0.15 M concentration of Na^+ and Cl^- ions.

Simulation details. Receptor-membrane systems were energy-minimized for 500 steps using Steepest Descent algorithm. Following energy minimization, each system was equilibrated using molecular dynamics (MD) simulations for a total of 27 ns in three steps, using position restraints 1) on non-hydrogen atoms (10 ns), 2) on backbone atoms and ligand (10 ns) and 3) only in TMs and ligand (7 ns). During this last step, forces were progressively decreased. Subsequently, production runs were launched without restraints.

Table 4.1 lists a summary of the ligand-receptor complexes simulated. For every system 2 or 3 replicates were run extending 1 µs each. MD

simulations were produced employing the Leapfrog integrator and a 2 fs time step. Pressure was kept at 0.1 MPa under semi-isotropic conditions and temperature was maintained constant at 300K using v-rescale thermostats (Bussi, Zykova-Timan and Parrinello, 2009). All bonds involving hydrogen atoms were constrained using the LINCS algorithm (Hess *et al.*, 1997). Non-bonded interactions were computed using a real-space cut-off of 1.0 nm, and the electrostatic interactions beyond the cut-off were treated using particle-mesh Ewald (PME) method (Darden, York and Pedersen, 1993) with a grid spacing of 0.15 nm. The protein force field used was Amber ff99sb-ildn (Lindorff-Larsen *et al.*, 2010) that permits the combination with the Berger (Berger, Edholm and Jähnig, 1997) parameters employed for the lipids, as previously described (Cordomí, Caltabiano and Pardo, 2012). The water model chosen was TIP3P (Price and Brooks, 2004). Ligands were parametrized using the Generalized Amber force field (GAFF) and HF/6-31g*-derived RESP atomic charges (Case *et al.*, 2005). All simulations were conducted with the GROMACS v4.6.7 simulation package (Abraham *et al.*, 2015; Kutzner *et al.*, 2015). Analyses of trajectories were done using GROMACS tools and VMD (Humphrey, Dalke and Schulten, 1996).

TABLE 4.1| Summary of systems simulated.

Ligand−Receptor Complex	Replicates	Time of Simulation	RMSD C_α (Å)	RMSD lig (Å)
Inverse agonist QNB	3	1 μs	2.15 ± 0.29	1.16 ± 0.23
Apo receptor	3	1 μs	2.21 ± 0.17	
Agonist iperoxo	2	1 μs	2.28 ± 0.38	1.21 ± 0.28
Agonist iperoxo + Nanobody	2	1 μs	1.83 ± 0.12	1.32 ± 0.24
Allosteric modulator LY2119620 + agonist iperoxo	2	1 μs	2.35 ± 0.46	1.66 ± 0.63
				1.44 ± 0.29
Allosteric modulator LY2119620 + agonist iperoxo + Nanobody	2	1 μs	1.61 ± 0.21	1.47 ± 0.54
				1.42 ± 0.24

The root mean square deviation (RMSD) values shown are the average of the replicates in all systems during the last 800 ns. The RMSD C_α and RMSD *lig* were calculated after fitting of the protein alpha carbons (C_α).

4.1.2.3. *Results and Discussion*

Based on the M₂R active state in complex with the high-affinity agonist iperoxo, we have set up several simulation systems (see *4.1.2.2. Methods* and Table 4.1). Starting from the aforementioned active state, in a second system, the agonist was replaced by the high-affinity inverse agonist R-(−)-3-quinuclidinyl-benzilate (QNB). As reference, the M₂R active structure bound to the agonist iperoxo with and without the nanobody that mimics the G protein were also simulated. Additionally, we simulated the M₂ active state in complex with iperoxo and the allosteric modulator LY2119620, as well as the apo state. Each system was simulated for 1 µs using MD with at least two replicates in a total simulation time of 14 µs.

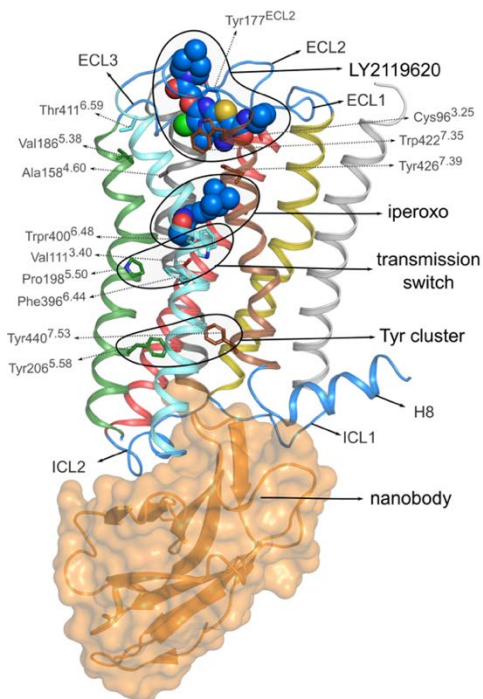**Figure 4.7 | Active structure of the M₂ receptor.** It illustrates the localization of the ligands binding pockets, the G-protein binding occupied by a nanobody (orange surface) site, the most important activation switches connecting them (the transmission switch and the Tyr cluster) plus key residues. The color code for TM helices of the receptor is 1 = white, 2 = gold, 3 = red, 4 = gray, 5 = green, 6 = cyan, 7 = pale-red.

Figure 4.7 displays schematically the localization of the orthosteric (iperoxo) and allosteric (LY2119620) ligand binding pockets, the G-protein binding site occupied by a nanobody and the most important activation switches connecting them.

In the simulations with the inverse agonist QNB (in red), however, specific movements of the extracellular halves of TMs 5 and 6 expanding the ligand-binding pocket were observed (Figure 4.8, central panel). To quantify these movements, the relative distances between TM4 and TM5, as well as TM3 and TM6 were monitored in the MD trajectories, and the values compared to the ones observed in the crystal structures (Table 4.2 and Figure 4.8A and 4.8B, respectively).

The three simulations of the inverse agonist QNB showed a progressive increase of these distances from the values of the initial structure (~9 Å for Ala158$^{4.60}$–Val186$^{5.48}$ and ~20 Å for Cys96$^{3.25}$–Thr411$^{5.98}$ in M$_2$ active) to larger distances characteristic of the inactive crystal structure (~13 Å for Ala158$^{4.60}$–Val186$^{5.48}$ and ~24 Å for Cys96$^{3.25}$–Thr411$^{5.98}$ in M$_2$ inactive). The change occurred in two steps. First a 3 Å increase in the distance Ala158$^{4.60}$–Val186$^{5.38}$ and 2 Å in Cys96$^{3.25}$–Thr411$^{6.59}$ in the first 100 ns or less of trajectory. This metastable intermediate conformation remained up to 400 ns in all simulations until the complete opening occurred, which implied additional 1 Å increase in the distance Ala158$^{4.60}$–Val186$^{5.38}$ and 3 Å in the distance Cys96$^{3.25}$–Thr411$^{6.59}$. In contrast, the simulations of the M$_2$-LY2119620-iperoxo-Nb complex (in dark grey) and the apo receptor (in orange) displayed constant distances for the extracellular cavity along the trajectories, and close to those measured in the crystallographic structures resembling active states (Table 4.2).

TABLE 4.2 | Residue-residue distances and χ$_2$ side chain rotamer associated to the deactivation mechanism in the M$_2$.

|	Cys96$^{3.25}$–Thr411$^{6.59}$ Distance (Å)	Ala158$^{4.60}$–Val186$^{5.38}$ Distance (Å)	Trp400$^{6.48}$ χ$_2$ (deg)	Tyr206$^{5.58}$–Tyr440$^{7.53}$ Distance (Å)
Crystallographic Structures				
M$_2$ INACTIVE (PDB ID 3UON)	24.8	13.3	52	12.6
M$_2$ ACTIVE (PDB ID 4MQS)	20.5	9.1	109	4.2
M$_2$ ACTIVE + PAM (PDB ID 4MQT)	20.5	9.1	101	4.4
Molecular Dynamics Simulations (last 800 ns)				
Inverse agonist QNB	23.8 ± 1.5	12.2 ± 0.8	56 ± 15.3	12.8 ± 3.8
Apo receptor	21.2 ± 1.4	8.4 ± 0.7	75 ± 19.2	10.3 ± 3.9
Agonist iperoxo	22.4 ± 1.6	8.8 ± 0.6	107 ± 9.6	7.2 ± 3.2
Agonist iperoxo + Nanobody	21.3 ± 0.7	9.8 ± 0.7	104 ± 10.3	4.3 ± 0.7
Allosteric modulator LY2119620 + agonist iperoxo	21.4 ± 0.9	10.2 ± 0.7	103 ± 14.2	7.7 ± 3.3
Allosteric modulator LY2119620 + agonist iperoxo + Nanobody	21.2 ± 0.8	9.7 ± 0.5	106 ± 9.9	4.3 ± 0.8

All distances were measured between the C$_\alpha$ (except for Tyr206$^{5.58}$-Tyr440$^{7.53}$ in which it was measured between hydroxyl oxygens) and averaged over the last 800n. The atoms that define the χ$_2$ side chain dihedral angle for the Trp400$^{6.48}$ are C$_\alpha$-C$_\beta$-C$_\gamma$-C$_{\delta1}$.

We next analysed the functionally important amino acids that participate in transmitting the signal from the ligand towards the cytoplasmic region of the receptor where the G protein binds. We focused on the "transmission switch" (Deupi and Standfuss, 2011), which consists of conformational rearrangements of residues Val111$^{3.40}$, Pro198$^{5.50}$, Phe396$^{6.44}$ and Trp400$^{6.48}$ as well as the Tyr cluster that comprises the interaction between Tyr206$^{5.58}$ and Tyr440$^{7.53}$ (see Figure 4.7), which facilitates the opening of the G protein binding cavity (Park *et al.*, 2008; Goncalves *et al.*, 2010).

Figure 4.8 | Opening of the extracellular part of the M$_2$-QNB complex (in red) during deactivation. (**Central panel**): Superposition of the active M2R crystal structure colored in blue, inactive M2R crystal structure colored in salmon and the end of the simulation of the M2-QNB complex in red. Despite the simulation of the M2-QNB complex was started in the active conformation (PBD id 4MQS), at the end of the trajectory, the conformations of TMs 5 and 6 for M2-QNB are close to the inactive crystal structure (in salmon, PDB id 3UON). The black

arrows show the changes from active to inactive conformation. (**A-B**): Examples of time-evolution of the distance between the Cα's of Ala158⁴·⁶⁰ and Val186⁵·³⁸ (**A**) and between the Cα's of Cys96³·²⁵ and Thr411⁶·⁵⁹ (**B**) in the different systems; the values of distances in the inactive (PDB ID 3UON) and active crystal structures (PDB ID 4MQT and 4MQS) are shown with dotted and dashed lines, respectively.
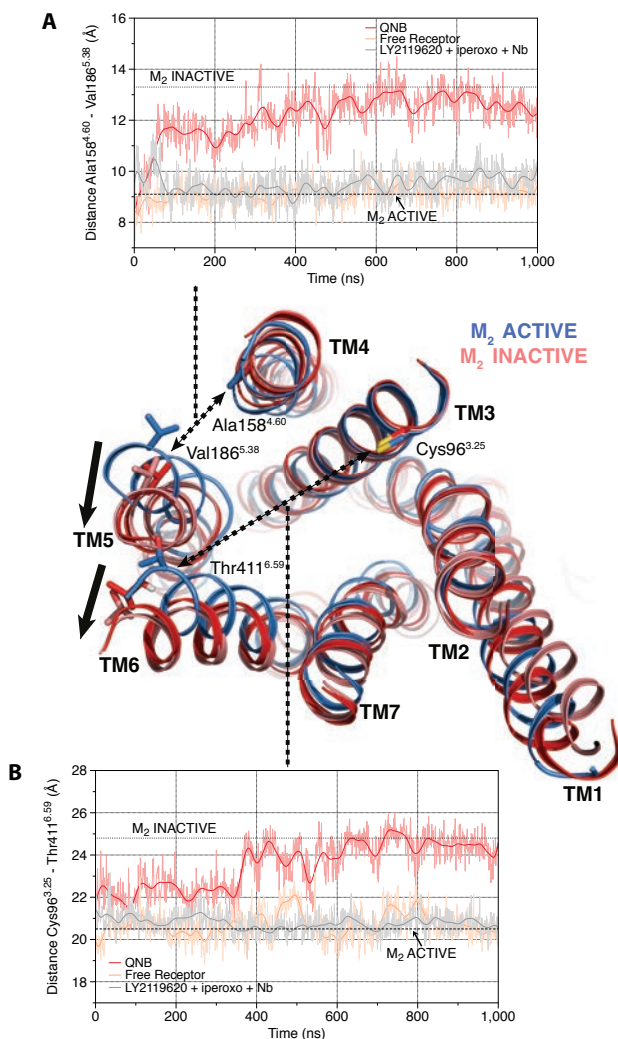
As observed in most active GPCR structures, the M_2 crystal features compact interactions between Val111[3.40], Trp400[6.48], Pro198[5.50] and Phe396[6.44], whereas inactive conformations have a more linear sequence of interactions (Trp400[6.48] – Phe396[6.44], Phe396[6.44] – Val111[3.40] and Val111[3.40] – Pro198[5.50]) (Figure 4.9A). We monitored the χ_2 side-chain dihedral angles of Trp400[6.48] of the transmission switch (Figure 4.9B-C) in all simulated systems and compared the values with those observed in the crystal structures (Table 4.2). The χ_2 dihedral angle changed very rapidly during the simulations and showed remarkable differences between systems, also in agreement with the values observed in the crystal structures.

In the active structures of the M_2R, Tyr206[5.58] and Tyr440[7.53] of the NPxxY motif form a water-mediated hydrogen bond which stabilizes the active conformation (Kruse *et al.*, 2013), as also seen in the active state structures of opsin (Park *et al.*, 2008) and β_2-AR (Ring *et al.*, 2013). Figure 30D shows the monitoring of the distance between Tyr206[5.58] and Tyr440[7.53] hydroxyl oxygens along the trajectories (only one replicate is shown) for each system. The M_2-active complexes display a constant distance along the simulation with values close to those in the crystallographic structures in the active conformations (4.2-4.4 Å, Table 4.2). In contrast, the distance in the apo receptor (in orange) and in the inverse agonist QNB complex (in red) typically increased by 4 to 6 Å. The distance Tyr206[5.58]-Tyr440[7.53] reflects two events: the first is the conformational change of Tyr440[7.53] from the active to the inactive form (4 Å increase) and the second is the posterior rigid-body movement of TM5, with Tyr206 retaining the initial conformation (additional 2 Å increase). In the simulations of the QNB systems we observed both events in two simulations and none of them in a third, whereas in the simulation of the apo we

observed the first event in two simulations and the second in only one simulation. In all observed cases, this second event occurs relatively late 600-850 ns.
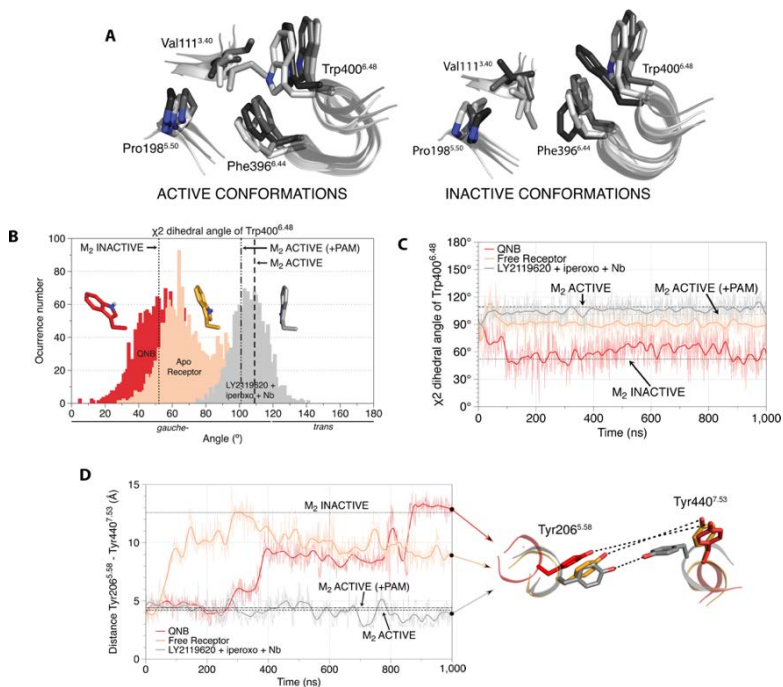
Figure 4.9| **Transmission switch and intracellular part of the M$_2$-QNB complex (in red) during deactivation. (A)** Comparison of the transmission switch in different crystal structures. M$_2$ receptor (in black), A$_{2A}$ adrenergic receptor (in dark-grey), β$_2$ adrenergic receptor (in grey) and rhodopsin (in light grey). **(B)** Histogram of the χ$_2$ side chain dihedral angle of Trp400[6.48] in the MD trajectories. **(C, D)** Examples of the time-evolution of **(C)** of the χ$_2$ side chain dihedral angle of Trp400[6.48] and **(D)** the distance between Tyr206[5.58] and Tyr440[7.53] in different systems (same replicates as in Figure 4.14). The values of distances in the inactive (PDB ID 3UON) and active crystal structures (PDB ID 4MQT and 4MQS) are shown with dotted and dashed lines, respectively.

Overall, the complete picture of the sequence of events for deactivation of active M_2R after binding of the inverse agonist QNB involves (Figure 4.10):

1) the rearrangement of the transmission switch adjacent to the orthosteric ligand (40-70 ns) monitored through the χ_2 side chain dihedral angle of $Trp400^{6.48}$,

2) the subsequent opening of the extracellular portion of TMs 5 and 6 (100-200 ns) monitoring the distance between $Ala158^{4.60}$ − $Val186^{5.38}$ and $Cys96^{3.25}$ − $Thr411^{6.59}$ and finally,

3) closure of the intracellular portion of TMs 5 and 6 (150->1000 ns) using $Tyr206^{5.58}$ – $Tyr440^{7.53}$ distance.
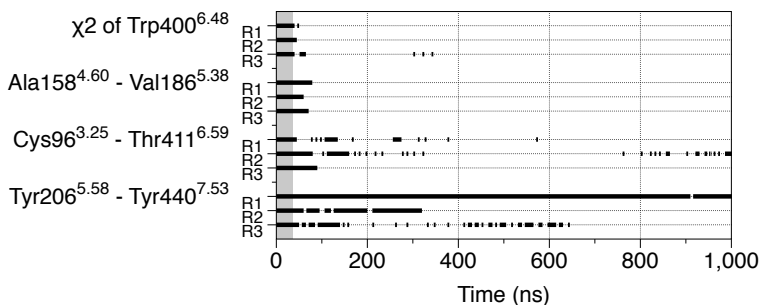
Figure 4.10 | Schematic representation of the deactivation process triggered by the inverse agonist QNB in three different replicates (R1-R3). Black vertical lines represent snapshots with active features in terms of χ_2 side chain dihedral angle of $Trp400^{6.48}$ (active $> 89°$) and distances between $Ala158^{4.60}$ − $Val186^{5.38}$ (active < 10.5 Å), $Cys96^{3.25}$ − $Thr411^{6.59}$ (active < 22 Å), and $Tyr206^{5.58}$ – $Tyr440^{7.53}$ (active < 6 Å).

Allosterism

In the simulations of the M_2 in complex with both the orthosteric agonist iperoxo and the allosteric modulator LY2196220, RMSD values for LY2196220 after least-square fit to protein Cα carbons ranged between 1.5 ± 0.5 Å (with iperoxo + nanobody) and 1.7 ± 0.6 Å (with iperoxo) (Table 4.1). These values (slightly larger than the ones for the orthosteric ligand) reflect the systematic shift of the piperazine group of LY2119620 from the conformation determined in the crystallographic structure of M_2. In fact, the electron density map for the crystal structure did not show clear density for the piperazine or Tyr177[ECL2], which interacts with LY2196220, suggesting that the movement observed in the simulations would not be an artefact.

We monitored the interactions formed by LY2196220 and the receptor (summarized in Figure 4.11) along the trajectories as well as the possible changes triggered by the allosteric modulator. Based on the present simulations we suggest that the piperazine ring forms water-mediated interactions with Glu172[ECL2], Tyr177[ECL2] and the backbone of ECL2 and ECL3.

Superposition of the M_2-iperoxo complexes crystal structures with and without LY2119620 (Kruse *et al.*, 2013) (RMSD = 0.16 Å on Cα excluding the Nb), indicates that the allosteric-binding site is pre-formed in the presence of the agonist, with the most remarkable difference being the rotameric change of Trp422[7.35], which adopts a vertical orientation (*gauche+* conformation in χ1 and *trans* in χ2) in presence of LY2119620 and a horizontal orientation (*trans* conformation in χ1 and *gauche+* in χ2) with iperoxo alone. Together with Tyr177[ECL2], both residues are crucial for allosteric ligand

recognition given that together with the aromatic ring of LY2119620, they form a three-layered aromatic stack (Gregory, Sexton and Christopoulos, 2007; Abdul-Ridha *et al.*, 2014) (Table 4.3 and Figure 4.12, central panel).
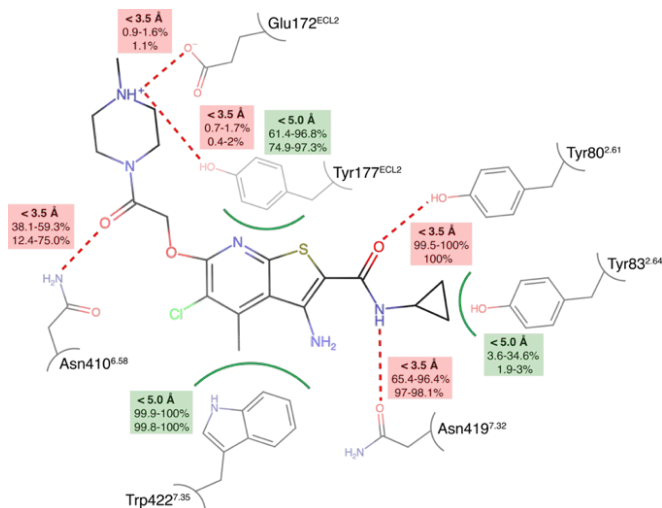
Figure 4.11| **Schematic representation of the interactions between the allosteric modulator LY2119620-iperoxo M$_2$R complexes during the MD simulations.** Polar contacts and hydrophobic contacts are shown as red dotted lines and in green solid lines, respectively. The percentages in which each interaction occurs in the simulations of LY2119620-iperoxo (first line) and LY2119620-iperoxo-Nb (second line) complexes, respectively, are also shown.

Figure 4.12A shows a histogram for the χ_1 of Tyr177[ECL2] in the M$_2$-LY2119620-iperoxo and M$_2$-iperoxo simulations with and without the nanobody (in dark and light green –with Nb- and dark and light grey –without Nb-, respectively). In the presence of the allosteric

modulator, χ_1 exhibits a unimodal rotameric distribution with the maximum centred around 50° (*gauche-* conformation). In the absence of the allosteric modulator, χ_1 adopts a multimodal variable distribution, with the largest population centred at ±180° (*trans* conformation) in the M$_2$-iperoxo complexes. A similar effect occurs on the distribution of χ_1 angles of Trp422[7.35] (top panel of Figure 4.12B), which exhibits a unimodal distribution centred around -75° (*gauche+* conformation) only when the allosteric modulator is present. By contrast, in absence of allosteric modulator χ_1 of Trp422[7.35] varies between -150° and 150° (*trans* conformation) whereas the χ_2 of Trp422[7.35] (bottom panel of Figure 4.12B) shows two possible rotamers centred at 120° (*gauche-* and *trans* conformation) or at -90° (*gauche+* conformation) in presence or absence of ligand, respectively.

TABLE 4.3| Side chain dihedral angles (χ_1 and χ_2) of amino acid residues involved in allosterism in the M$_2$ crystal structures.

	Tyr177[ECL2]	Trp422[7.35]		Tyr426[7.39]	
	χ_1	χ_1	χ_2	χ_1	χ_2
Crystallographic Structures					
M$_2$ INACTIVE (PDB ID 3UON)	62° (g−)	−74° (g+)	106° (g−)	−101° (g+)	−92° (g+)
M$_2$ ACTIVE (PDB ID 4MQS)	52° (g−)	178° (t)	−104° (g+)	−120° (g+/t)	−107° (g+)
M$_2$ ACTIVE + PAM (PDB ID 4MQT)	49° (g−)	−107° (g+)	123° (t)	−112°(g+)	−118° (g+)
Molecular Dynamics Simulations (last 800 ns):					
Inverse agonist QNB	g−: 64.5%	0.2%	18.5%	0%	95.3%
	g+: 0.9%	19.8%	72.7%	0%	
	t: 34.6%	80%	8.8%	100%	4.7%
Apo receptor	g−: 60.5%	22.3%	21.9%	0%	76.2%
	g+: 24.6%	41.6%	45.8%	35.6%	
	t: 15.0%	36.1%	32.3%	64.4%	23.8%
Agonist iperoxo	g−: 18.9%	0%	1.4%	0%	98.8%
	g+: 12.2%	0.6%	80.0%	5.4%	
	t: 68.9%	99.4%	18.6%	94.6%	1.2%
Agonist iperoxo + Nanobody	g−: 7.9%	0%	2.2%	0%	97.4%
	g+: 3.9%	0.1%	93.9%	0.6%	
	t: 88.2%	99.9%	3.9%	99.4%	2.6%
Allosteric modulator LY2119620 + agonist iperoxo	g−: 99.1%	0%	38.8%	0%	100%
	g+: 0.9%	100%	0%	99.6%	
	t: 0%	0%	61.2%	0.4%	0%
Allosteric modulator LY2119620 + agonist iperoxo + Nanobody	g−: 99.9%	0%	46.1%	0%	100%
	g+: 0.1%	100%	0%	99.4%	
	t: 0.0%	0%	53.9%	0.6%	0%

The atoms that define the χ_1 and χ_2 side chain dihedral angles are N-C$_\alpha$-C$_\beta$-C$_\gamma$ and C$_\alpha$-C$_\beta$-C$_\gamma$-C$_{\delta1}$ respectively. For the percentages, it is assumed that *gauche+* is -120° ≤ x < 0°, *gauche-* is 0° ≤ x < 120° and *trans* is x < -120° and 120° ≤ x.

Residues Trp422[7.35] and Tyr426[7.39], the latter being part of an aromatic lid/cage together with Tyr104[3.33] and Tyr403[6.51], are located in the interface between the allosteric and the orthosteric binding sites (Hibert *et al.*, 1991; Lu, Saldanha and Hulme, 2001). The importance of these residues, which are completely conserved in the family of muscarinic acetylcholine receptors, has been shown by mutagenesis studies accompanying the recent structure of the M_4 receptor in complex with the allosteric modulator LY2033298 (Thal *et al.*, 2016). Alanine mutations of Trp422[7.35] (Trp422[7.35]Ala) led to a complete loss in LY2033298 binding and Tyr426[7.39]Ala led to decrease in cooperativity between the allosteric modulator and the orthosteric ligand.

Accordingly, Tyr426[7.39] was monitored to examine how the allosteric modulator could provoke a conformational change in the orthosteric ligand-binding pocket. The simulations show that the conformation of Trp422[7.35] affects the rotameric conformation of Tyr426[7.39]. The time-evolution of the trajectories (not shown) reveals a rapid change of the conformation of Trp422[7.35] which is followed by the rotameric change of Tyr426[7.39] in < 50 ns. Due to poor electron density in this region, this change is not clearly observed in the M_2 active crystallographic structures (Table 4.3 and Figure 4.12, central panel). The side chain dihedral angles of Tyr426[7.39] also reveal differences associated to the presence or not of the allosteric modulator LY2119620 (Figure 4.12C). The dihedral angle $\chi 1$ of Tyr426[7.39] in M_2-LY2119620-iperoxo and M_2-LY2119620-iperoxo-Nb complexes is centred at -90º (*gauche+* conformation) while in M_2-iperoxo and M_2-iperoxo-Nb complexes it is centred at -155º (*trans* conformation). The dihedral angle $\chi 2$ in presence of LY2119620 is centred at 55º while in absence of it is centred at 100º (both *gauche* conformation).

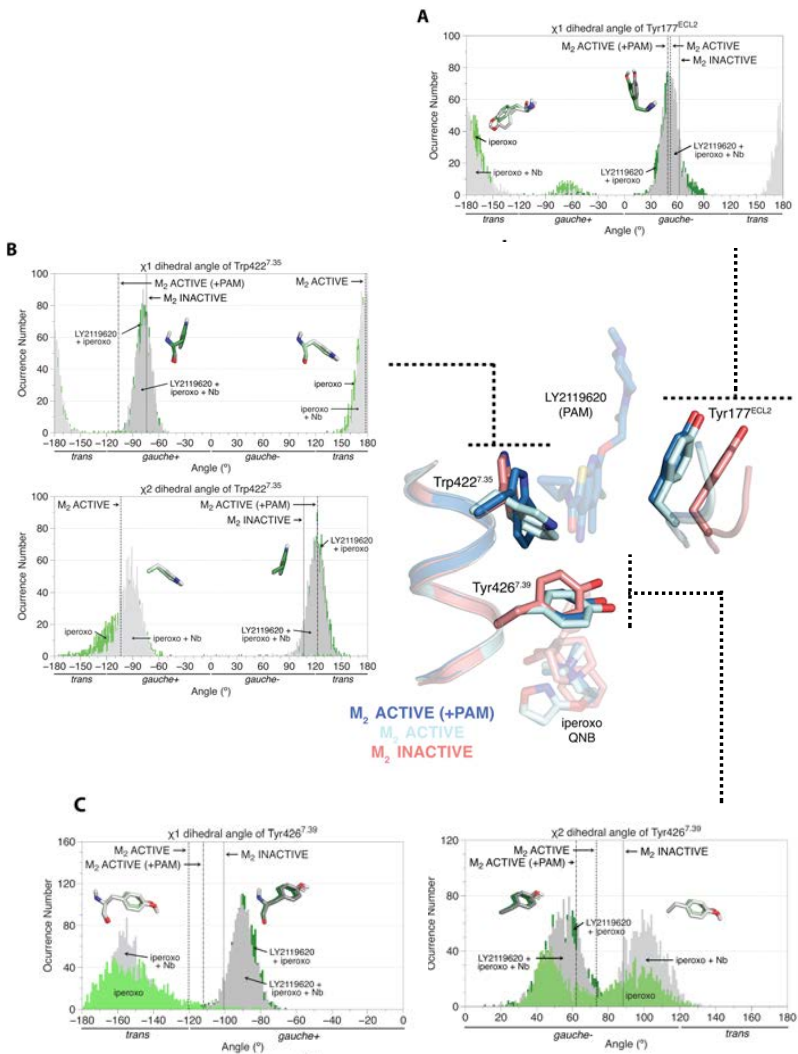**Figure 4.12| The orientation of Tyr177[ECL2], Trp422[7.35] and Tyr426[7.39] near the allosteric ligand. (Central panel):** Superposition of the M_2R crystal structures (active with PAM: dark blue, active: light blue, and inactive: salmon). **(A-C):** Histograms of the χ_1 and/or χ_2 side chain dihedrals of Trp422[7.35] (**A**), Tyr177[ECL2] (**B**) and Tyr426[7.39] (**C**).

We suggest that this difference in the orientation of Tyr426[7.39] with the presence of the allosteric modulator, which affects the rotameric conformation of Trp422[7.35], is the reason for the increase in affinity and potency for orthosteric agonist iperoxo.

4.1.2.5. *References*

Abdul-Ridha, A. *et al.* (2014) 'Molecular determinants of allosteric modulation at the M1 muscarinic acetylcholine receptor.', *J. Biol. Chem.*, 289(9), pp. 6067–79.

Abraham, M. J. *et al.* (2015) 'Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers', *SoftwareX*, 1–2, pp. 19–25.

Berger, O., Edholm, O. and Jähnig, F. (1997) 'Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature.', *Biophys. J.*, 72(5), pp. 2002–2013.

Bussi, G., Zykova-Timan, T. and Parrinello, M. (2009) 'Isothermal-isobaric molecular dynamics using stochastic velocity rescaling', *J. Chem. Phys.*, 130(7), p. 074101.

Case, D. A. *et al.* (2005) 'The Amber biomolecular simulation programs', *J. Comput. Chem.*, 26(16), pp. 1668–1688.

Christopoulos, A. and Kenakin, T. (2002) 'G protein-coupled receptor allosterism and complexing.', *Pharmacol. Rev.*, 54(2), pp. 323–74.

Cordomí, A., Caltabiano, G. and Pardo, L. (2012) 'Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters', *J. Chem. Theory Comput.*, 8(3), pp. 948–958.

Croy, C. H. *et al.* (2014) 'Characterization of the novel positive allosteric modulator, LY2119620, at the muscarinic M(2) and M(4) receptors.', *Mol. Pharmacol.*, 86(1), pp. 106–15.

Darden, T., York, D. and Pedersen, L. (1993) 'Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems', *J. Chem. Phys.*, 98(12), p. 10089.

Deupi, X. and Standfuss, J. (2011) 'Structural insights into agonist-induced activation of G-protein-coupled receptors', *Curr. Opin. Struct. Biol.*, 21(4), pp. 541–551.

Dolinsky, T. J. *et al.* (2004) 'PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations', *Nucleic Acids Res.*, 32(Web Server), pp. W665–W667.

Dror, R. O. *et al.* (2011) 'Activation mechanism of the β2-adrenergic receptor', *Proc. Natl. Acad. Sci.*, 108(46), pp. 18684–18689.

Felder, C. C. (1995) 'Muscarinic acetylcholine receptors: signal transduction through multiple effectors.', *FASEB J.*, 9(8), pp. 619–625.

Goncalves, J. a *et al.* (2010) 'Highly conserved tyrosine stabilizes the active state of rhodopsin', *Proc. Natl. Acad. Sci.*, 107(46), pp. 19861–19866.

Gregory, K. J., Sexton, P. M. and Christopoulos, A. (2007) 'Allosteric modulation of muscarinic acetylcholine receptors.', *Curr. Neuropharmacol.*, 5(3), pp. 157–167.

Haga, K. *et al.* (2012) 'Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist', *Nature*, 482(7386), pp. 547–551.

Hess, B. *et al.* (1997) 'LINCS: A linear constraint solver for molecular simulations', *J. Comput. Chem.*, 18(12), pp. 1463–1472.

Hibert, M. F. *et al.* (1991) 'Three-dimensional models of neurotransmitter G-binding protein-coupled receptors.', *Mol. Pharmacol.*, 40(1), pp. 8–15. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1649965.

Hulme, E. C., Birdsall, N. J. M. and Buckley, N. J. (1990) 'Muscarinic Receptor Subtypes', *Annu. Rev. Pharmacol. Toxicol.*, 30(1), pp. 633–673.

Humphrey, W., Dalke, A. and Schulten, K. (1996) 'Visual Molecular Dynamics', *J. Mol. Graph.*, 14(1), pp. 33–38.

Jakubík, J. and El-Fakahany, E. E. (2010) 'Allosteric modulation of muscarinic acetylcholine receptors', *Pharmaceuticals*, 3(9), pp. 2838–2860.

Katritch, V., Cherezov, V. and Stevens, R. C. (2013) 'Structure-Function of the G Protein–Coupled Receptor Superfamily', *Annu. Rev. Pharmacol. Toxicol.*, 53(1), pp. 531–556.

Kruse, A. C. *et al.* (2012) 'Structure and dynamics of the M3 muscarinic acetylcholine receptor', *Nature*, pp. 552–556.

Kruse, A. C. *et al.* (2013) 'Activation and allosteric modulation of a muscarinic acetylcholine receptor.', *Nature*, 504(7478), pp. 101–6.

Kutzner, C. *et al.* (2015) 'Best bang for your buck: GPU nodes for GROMACS biomolecular simulations', *J. Comput. Chem.*, 36(26), pp. 1990–2008.

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(4), pp. 339–357.

Langmead, C. J., Watson, J. and Reavill, C. (2008) 'Muscarinic acetylcholine receptors as CNS drug targets.', *Pharmacol. Ther.*, 117, pp. 232–243.

Liapakis, G. *et al.* (2012) 'The G-protein coupled receptor family: actors with many faces.', *Curr. Pharm. Des.*, 18(2), pp. 175–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22229577 (Accessed: 17 February 2016).

Lindorff-Larsen, K. *et al.* (2010) 'Improved side-chain torsion potentials for the Amber ff99SB protein force field', *Proteins Struct. Funct. Bioinforma.*, 78(8), pp. 1950–1958.

Lomize, M. A. *et al.* (2006) 'OPM: orientations of proteins in membranes database.', *Bioinformatics*, 22(5), pp. 623–5.

Lu, Z.-L. L., Saldanha, J. W. and Hulme, E. C. (2001) 'Transmembrane domains 4 and 7 of the M(1) muscarinic acetylcholine receptor are critical for ligand binding and the receptor activation switch.', *J. Biol. Chem.*, 276(36), pp. 34098–104.

Mahoney, J. P. and Sunahara, R. K. (2016) 'Mechanistic insights into GPCR–G protein interactions', *Curr. Opin. Struct. Biol.* Elsevier Ltd, 41, pp. 247–254.

Miao, Y. *et al.* (2013) 'Activation and dynamic network of the M2 muscarinic receptor', *Proc. Natl. Acad. Sci. U. S. A.*, 110(27), pp. 10982–10987.

Miao, Y. and McCammon, J. A. (2016) 'Graded activation and free energy landscapes of a muscarinic G-protein–coupled receptor', *Proc. Natl. Acad. Sci.*, 113(43), pp. 12162–12167.

Miao, Y., Nichols, S. E. and McCammon, J. A. (2014) 'Free energy landscape of G-protein coupled receptors, explored by accelerated molecular dynamics.', *Phys. Chem. Chem. Phys.*, 16(14), pp. 6398–406.

Nygaard, R. *et al.* (2013) 'The dynamic process of β2-adrenergic receptor activation', *Cell*, 152(3), pp. 532–542.

Park, J. H. *et al.* (2008) 'Crystal structure of the ligand-free G-protein-coupled receptor opsin', *Nature*, 454(7201), pp. 183–187.

Price, D. J. and Brooks, C. L. (2004) 'A modified TIP3P water potential for simulation with Ewald summation', *J. Chem. Phys.*, 121(20), pp. 10096–10103.

Rasmussen, S. G. F. *et al.* (2011) 'Crystal structure of the β2 adrenergic receptor–Gs protein complex', *Nature*. Nature Publishing Group, 477(7366), pp. 549–555.

Ring, A. M. *et al.* (2013) 'Adrenaline-activated structure of β2-adrenoceptor stabilized by an engineered nanobody', *Nature*. Nature Publishing Group, 502(7472), pp. 575–579.

Sali, A. (1995) 'Comparative protein modeling by satisfaction of spatial restraints', *Mol. Med. Today*, 1(6), pp. 270–277.

Scheerer, P. *et al.* (2008) 'Crystal structure of opsin in its G-protein-interacting conformation', *Nature*, 455(7212), pp. 497–502.

Schober, D. A. *et al.* (2014) 'Development of a Radioligand, [3H]LY2119620, to Probe the Human M2 and M4 Muscarinic Receptor Allosteric Binding Sites', *Mol. Pharmacol.*, 86(1), pp. 116–123.

Süel, G. M. *et al.* (2003) 'Evolutionarily conserved networks of residues mediate allosteric communication in proteins.', *Nat. Struct. Biol.*, 10(1), pp. 59–69.

Thal, D. M. *et al.* (2016) 'Crystal structures of the M1 and M4 muscarinic acetylcholine receptors.', *Nature*. Nature Publishing Group, 531(7594), pp. 335–40.

Thorsen, T. S. *et al.* (2014) 'Modified T4 Lysozyme Fusion Proteins Facilitate G Protein-Coupled Receptor Crystallogenesis', *Structure*. Elsevier Ltd, 22(11), pp. 1657–1664.

Trzaskowski, B. *et al.* (2012) 'Action of Molecular Switches in GPCRs - Theoretical and Experimental Studies', *Curr. Med. Chem.*, 19(8), pp. 1090–1109.

Tusnády, G. E., Dosztányi, Z. and Simon, I. (2004) 'Transmembrane proteins in the Protein Data Bank: identification and classification.', *Bioinformatics*, 20(17), pp. 2964–72.

van der Westhuizen, E. T. *et al.* (2015) 'Endogenous Allosteric Modulators of G Protein-Coupled Receptors', *J. Pharmacol. Exp. Ther.*, 353(2), pp. 246–260.

4.2. GPCR OLIGOMERIZATION

4.2.1. COCAINE BLOCKS GHRELIN EFFECTS VIA INTERACTION WITH SIGMA-1 RECEPTORS

4.2.1.1. *Background*

Used today as recreational drug, cocaine was first consumed by humans in the form of coca leaves. For example, indigenous peoples of South America, especially those living at higher altitude, used it to travel long distances across the Andes with reduced weight and little food. Despite such ancient knowledge, i.e. the appetite suppressant action of cocaine, the molecular basis of hunger dissipation remains unknown. This report was undertaken to test the hypothesis of whether the well-known anorexic effect of cocaine is mediated by growth hormone secretagogue (GHS) receptors, also known as ghrelin receptors, which are key players in the central control of food/energy intake (Howick *et al.*, 2017).

Ghrelin, the "hunger" endocrine hormone, is involved in the control of food intake and energy homeostasis (Cassidy and Tong, 2017). Its action is mediated by, up to date, only one specific ghrelin receptor that belongs to the superfamily of G-protein-coupled receptors (GPCRs) (Conn and Bowers, 1996; Geelissen *et al.*, 2003; Chan and Cheng, 2004). In humans, alternative splicing leads to: isoform 1a that contains seven transmembrane (TM) domains (GHS-R1a, 366 amino acids) and isoform 1b that lacks TMs 6 and 7 (GHS-R1b, 289 amino acids). These TM domains are required for ligand binding and coupling to heterotrimeric G proteins and, therefore, ghrelin cannot signal via GHS-R1b receptors (Mary *et al.*, 2013). GHS-R1b seems to serve as modulator of GHS-R1a surface expression and signalling

(Navarro *et al.*, 2016). In fact, GHS-R1b is expressed in the same cells as GHS-R1a and both isoforms interact to form heteromer receptor signalling units (Mary *et al.*, 2013). We have previously shown that GHS-R1b guides surface expression of functional GHS-R1a, acting as a dual modulator: low relative GHS-R1b expression potentiates GHS-R1a function, while high relative GHS-R1b expression inhibits GHS-R1a function (Navarro *et al.*, 2016). GHS-R1b negatively influences ghrelin action by allosteric interactions with the GHS-R1a-GHS-R1b heteromer that reduce the efficacy of the hormone (Chow *et al.*, 2012; Mary *et al.*, 2013). Although the purified GHS-R1a assembled into lipid discs is reportedly coupled to $G_{q/11}$ (Damian *et al.*, 2015) the receptor may couple to non-$G_{q/11}$ heterotrimeric G proteins (see www.guidetopharmacology.org). Indeed, we previously found preferential $G_{i/o}$ coupling of the GHS-R1a-GHS-R1b complex in HEK-293T cells and, preferential $G_{s/olf}$ coupling in both striatal and hippocampal neurons in culture (Navarro *et al.*, 2016). Heteromerization of GHS-R1a and GHS-R1b in heterologous expression systems is often needed for proper ghrelin-induced signalling. It should be noted that ghrelin receptors may form direct protein-protein interactions with a variety of GPCRs, *inter alia* with dopamine, melanocortin, prostanoid, serotonin, somatostatin, neurotensin, and GPR83 receptors (Borroto-Escuela *et al.*, 2014); see www.gpcr-hetnet.com and references therein).

The sigma-1 receptor (σ_1R) is an atypical type of endoplasmic reticulum-resident membrane protein whose exact function remains unknown. It has been proposed as a pluripotent modulator in living cells (Su *et al.*, 2016) that is gaining interest due to its potential as target against neuropathic pain (Mei and Pasternak, 2002; Corbera *et al.*, 2006; Sun *et al.*, 2016). While its physiological function remains elusive

and the endogenous ligand is yet to be discovered, σ_1R is the protein target for cocaine (McCracken, Bowen and Matsumoto, 1999; Skuza, 1999; Lever *et al.*, 2016). Thus, drugs blocking the interaction of cocaine with σ_1R is proposed to reduce drug-seeking behaviour (Matsumoto *et al.*, 2001). σ_1R-mediated cocaine actions in the central nervous system are dependent on its interactions with GPCRs. For instance, we have identified that cocaine binding to σ_1R receptors potentiates dopamine D_1R mediated adenylate cyclase signalling (Navarro *et al.*, 2013; Moreno *et al.*, 2014) and inhibits D_2R mediated signalling in striatal neurons (Navarro *et al.*, 2013). These actions disrupts the delicate balance between inputs of reward seeking controlled by D_1R-containing neurons and inputs of aversion coming from D_2R-containing neurons. In addition, cocaine binding to σ_1R also induces disruption of the orexin and corticotropin releasing factor receptor negative cross-talk, playing an important role in the stress-induced cocaine-seeking behaviour (Navarro *et al.*, 2015).

We show in this manuscript that σ_1R interacts and modulates the GHS-R1a, mediating the anorexigenic effects of cocaine in the central nervous system. Moreover, because structural information on GPCR macromolecular complexes already exist (Cordomí *et al.*, 2015) and the crystal structure of σ_1R has been recently elucidated (Schmidt *et al.*, 2016), we have devised the molecular architecture of the GHS-R1a-GHS-R1b heteromer in complex with σ_1R. We have taken advantage of the recent knowledge showing that σ_1R arranges as a homotrimer (each protomer with a single TM). We used synthetic peptides with the amino acid sequences from the TM helices of GHS-R1a, together with bimolecular fluorescence complementation assays and computer modelling, to find the oligomerization interfaces

between GHS-R1a, GHS-R1b, and σ1R, and to propose a structural model of the macromolecular complex.

4.2.1.2. *Materials and Methods*

Reagents. Cocaine-chlorhydrate was provided by the Spanish *Agencia del Medicamento* (Ref n°: 2003C00220). $\sigma_1 R$ ligand, PRE-084, and GHs-R1a ligands, ghrelin and YIL-781, were purchased from Tocris, Bristol, UK.

Fusion proteins and expression vectors. Sequences encoding amino acid residues 1–155 and 155–238 of the Venus variant of yellow fluorescence protein (Venus) were subcloned in the pcDNA3.1 vector to obtain complementary Venus n- and c-hemiproteins. Human cDNAs for GHS-R1a, GHS-R1b, or $\sigma_1 R$, were amplified without their stop codons using sense and antisense primers harboring: EcoRI and KpnI sites to be subcloned in pcDNA3.1RLuc vector (p*RLuc*-N1 PerkinElmer, Wellesley, MA) or in a GFP^2 containing vector (p-GFP^2, Packard BioScience, Meridien, CT) to provide $\sigma_1 R$-Rluc, GHS-R1a-Rluc, GHS-R1a-YFP, $\sigma_1 R$-YFP or GHS-R1a-GFP^2 plasmids. Human cDNA for $A_{2A}R$ was subcloned into p-GFP^2 vector harboring HindIII and BamHI sites to provide the encoding A_{2A}-GFP^2 plasmid. For bimolecular fluorescence complementation (BiFC) experiments, cDNA for GHS-R1b, GHS-R1a and $\sigma_1 R$ were subcloned into pcDNA3.1-nVenus and pcDNA3.1-cVenus harboring EcoRI and KpnI sites to provide plasmids encoding GHS-R1b-nYFP, GHS-R1b-cYFP, GHS-R1a-nYFP and $\sigma_1 R$-cYFP.

Cell lines, neuronal primary cultures and transient transfection. HEK-293T human embryonic kidney cells were grown in Dulbecco's

modified Eagle's medium (DMEM) (Gibco) supplemented with 2 mM L-glutamine, 100 µg/ml sodium pyruvate, 100 U/ml penicillin/streptomycin, MEM Non-Essential Amino Acid Solution (1/100) and 5% (v/v) heat inactivated Fetal Bovine Serum (FBS) (all supplements were from Invitrogen, Paisley, Scotland, UK). Primary cultures of striatal neurons were obtained from fetal Sprague Dawley rats of 19 days. Cells were isolated as described in Hradsky et al. (Hradsky *et al.*, 2013) and plated at a confluence of 40,000 cells/0.32 cm². Cells were maintained for 12 days in Neurobasal medium supplemented with 2 mM L-glutamine, 100 U/ml penicillin/ streptomycin, and 2% (v/v) B27 supplement (GIBCO) in 6-well microplates. Cells were transiently transfected with the corresponding cDNAs using the PEI (polyethylenimine, Sigma-Aldrich, St. Louis, MO, USA) method or, in the case of the anti-σ_1R siRNA, with lipofectamine 2000 (Thermo Fisher Scientific). After transfection cells were incubated in serum-free medium that after 4 h was replaced by complete medium. Experiments were carried out 48 h later (unless otherwise indicated).

Immunocytochemistry. HEK-293T cells were fixed in 4% paraformaldehyde for 15 min and washed with PBS containing 20 mM glycine to quench free aldehyde groups. After permeabilization with PBS-glycine buffer containing 0.2% Triton X-100 for 5 min, cells were blocked with PBS containing 1% bovine serum albumin (1 h at room temperature). σ_1R-YFP was detected by its own fluorescence (wavelength 530 nm), and Rluc-containing proteins were stained using a mouse monoclonal anti-Rluc antibody (1/200, 1 h, room temperature, Millipore, CA, USA) and a Cyn3-conjugated donkey anti-mouse antibody (1/200, Jackson Immunoresearch Laboratories, West Grove, PA, USA). Nuclei were stained with Hoechst (1/100,

SigmaAldrich, St. Louis, USA) and samples were mounted with Mowiol 30% (Calbiochem) and observed in a Leica SP2 confocal microscope (Leica Microsystems, Mannheim, Germany).

Resonance energy transfer. For Bioluminescence energy transfer (BRET) assays, HEK-293T cells were transiently transfected with a constant amount of cDNA for $\sigma_1 R$-Rluc and increasing amounts of cDNAs for GHS-R1a-GFP[2] or A_{2A}-GFP[2]. To normalize the number of cells, protein concentration was determined using a Bradford assay kit (Bio-Rad, Munich, Germany) using bovine serum albumin dilutions as standards. To quantify fluorescence, cell suspensions were distributed in 96-well microplates (black with transparent bottom) and fluorescence was read in a Fluostar Optima Fluorimeter (BMG Labtech, Offenburg, Germany) equipped with a high-energy xenon flash lamp, using a 10 nm bandwidth excitation filter (400 nm). For BRET measurements, cell suspensions (20 μg protein) were distributed in 96-well white microplates (Corning 3600, Corning, NY), and 5 μM DeepBlueC (Molecular Probes, Eugene, OR) was added right before BRET signal acquisition using a Mithras LB 940 reader (Berthold Technologies, DLReady, Germany). To quantify receptor-Rluc expression, luminescence readings were performed after 10 minutes of adding 5 μM coelenterazine H. Net BRET is defined as [(long-wavelength emission)/(short-wavelength emission)]-C_f, where C_f corresponds to [(long-wavelength emission)/(short-wavelength emission)] for the Rluc protein when expressed individually. For bimolecular complementation (BiFC) assays, HEK-293T cells were transiently transfected with a constant amount of cDNA encoding for proteins fused to nVenus or cVenus and incubated for 4 h in complete DMEM containing the interfering TAT peptides (with similar sequences to those in TM1 to TM7 of GHS-

R1a). YFP resulting from complementation was detected by placing cells (20 μg protein) in 96-well microplates (black plates with a transparent bottom) and reading the fluorescence in a Fluostar Optima Fluorimeter (BMG Labtech, Offenburg, Germany) using a 30 nm bandwidth excitation filter (485 nm).

Cytosolic cAMP determination. Forskolin dose-response curves in different density of cells were performed to select the most appropriate conditions of the assay, which resulted in 5,000 HEK-293T cells, 7,500 neurons and 0.5 μM forskolin. Subsequently, assays were performed in medium containing 50 μM zardeverine, placing cells in 384-well microplates. Then, was done the preincubation with reagents (the σ_1R agonist, PRE-084, the GHS-R1a antagonist, YIL-781, or cocaine) for 15 min followed by Ghrelin addition (100 nM final concentration) and, after 15 min incubation period 50 μM forskolin was added. Readings were performed 15 min later using a homogeneous time-resolved fluorescence energy transfer (HTRF) method requiring the Lance Ultra cAMP kit (PerkinElmer) and fluorescence readings (at 665 nm) in a PHERAstar Flagship micro-plate reader equipped with an HTRF optical module (BMG Labtech).

MAPK activation. To determine ERK1/2 phosphorylation, 40,000 HEK-293T cells/well or 50,000 neurons/well were plated in transparent Deltalab 96-well microplates and kept in the incubator for 48 h. The medium was substituted by serum-free DMEM medium 2 to 4 h before starting the experiment. Before addition of 100 nM ghrelin cells were pre-treated (10 min at 25°C) in serum-free medium with different reagents (the σ_1R agonist, PRE-084, the GHS-R1a antagonist, YIL-781, or cocaine). After 7 min of ghrelin-induced activation, cells were washed twice with cold PBS before the addition of 30 μL of Lysis Buffer (20 min). Supernatants (10 μL) were placed

in white ProxiPlate 384-well microplates, and ERK1/2 phosphorylation was determined using the AlphaScreen®SureFire® kit (Perkin Elmer) and the EnSpire® Multimode Plate Reader (PerkinElmer, Waltham, MA, USA).

Computational model of the GHS-R1a-σ_1R heteromer.

The structure of σ_1R was modelled based on the recently released crystal structure (PDB id 5HK1) (Schmidt *et al.*, 2016). The inactive conformation of human GHS-R1a [UniProt: Q92847] was built using crystal structures of the neurotensin 1 receptor (PDB id 4XES for all parts of the receptor, and PDB id 3ZEV for the C-terminal part of TM7 and helix 8) (Egloff *et al.*, 2014; Krumm *et al.*, 2015). The human neurotensin 1 receptor and GHS-R1a share 33% of sequence identity and 51% of sequence similarity. The "active-like" form of GHS-R1a was modelled by incorporating the active features present in the crystal structure of the β_2-adrenergic receptor in complex with G_s (PDB id 3SN6) (Rasmussen *et al.*, 2011). The "active-like" model of GHS-R1a contains G_i (PDB id 1AGR) (Tesmer *et al.*, 1997). The GHS-R1a homodimer was constructed based on the symmetric TM5/6 protein-protein interface observed in the crystal structure of the µOR (PDB id 4DKL) (Manglik, et al., 2012). The GHS-R1a-σ_1R complex was constructed using protein-protein docking with HADDOCK (van Zundert *et al.*, 2016), under the imposed experimental restrains that TMs 1 and 2 of GHS-R1a contact the single TM of σ_1R.

The molecular model of GHS-R1a homodimer in complex with G_i and σ_1R homotrimer was embedded in a pre-equilibrated box (20.0 x 15.0 x 15.0 Å) containing ~750 1-palmitoyl-2-oleoyl-sn-glycero-3-

phosphatidylcholine (POPC) lipids with explicit solvent (~ 106000 water molecules) and a 0.15 M concentration of Na^+ and Cl^- ions.

The complex was energy minimized and subjected to a 27 ns molecular dynamics (MD) equilibration, with decreasing positional restraints on protein coordinates. The production run without restraints was launched for 500 ns using a 2 fs time step, constant pressure under semi-isotropic condition at 0.1 MPa and constant temperature at 300 K. MD simulations were conducted with GROMACS 5.1.4 simulation package (Abraham *et al.*, 2015; Kutzner *et al.*, 2015), using the AMBER ff99SB-ILDN force field for the protein (Lindorff-Larsen *et al.*, 2010), the parameters described by Berger for POPC lipids (Berger, Edholm and Jähnig, 1997), and the water model chosen was TIP3P (Price and Brooks, 2004). This procedure has been previously validated (Cordomí, Caltabiano and Pardo, 2012).

4.2.1.3. *Results*

GHS-R1a forms heteromeric complexes with $\sigma_1 R$

Immunocytochemical assays were performed to detect whether colocalization between GHS-R1a and $\sigma_1 R$ occurred in transfected HEK-293T cells. Cells were transfected with cDNAs for $\sigma_1 R$ fused to YFP (0.75 μg cDNA) and for GHS-R1a fused to Rluc (1.66 μg cDNA). In cells expressing only $\sigma_1 R$-YFP, the receptor was detected by YFP fluorescence, identifying $\sigma_1 R$-YFP in intracellular structures. In HEK-293T cells expressing GHS-R1a-Rluc, the GHS-R1a was detected by a specific primary anti-Rluc and secondary Cy3 antibodies, being detected in intracellular structures and at the plasma membrane level. Interestingly, in HEK-293T cells coexpressing $\sigma_1 R$-

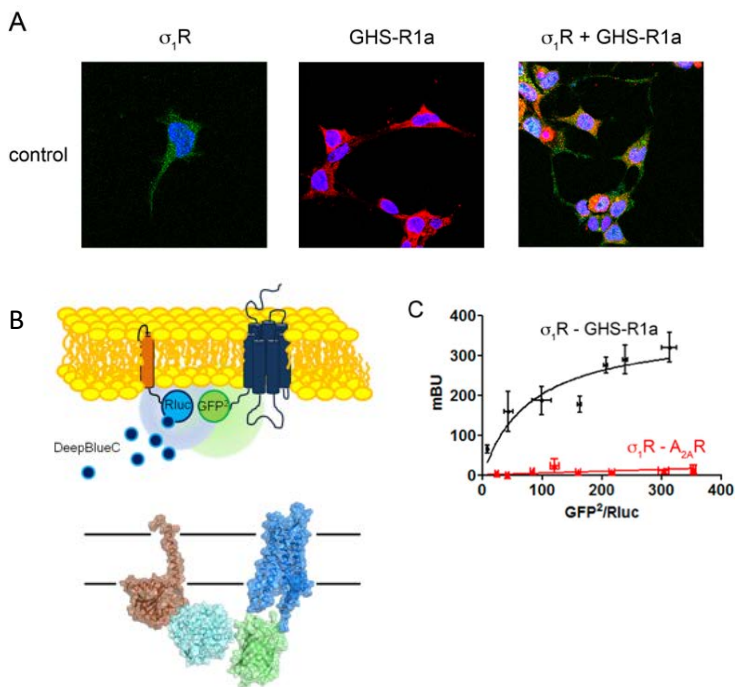YFP (0.75 µg cDNA) and GHS-R1a-Rluc (1.66 µg cDNA), colocalization of both receptors was observed (Figure 4.13A).

Figure 4.13 | GHS-R1a interacts with σ_1R forming σ_1R-GHS-R1a heteroreceptor complexes. (**A**) HEK-293T cells expressing σ_1R-YFP, GHS-R1a-Rluc or both, were monitored by the YFP fluorescence (*green*) or using a monoclonal anti-Rluc primary antibody and a cyanine-3-conjugated secondary antibody (*red*). Colocalization is shown in *yellow*. Nuclei were stained with Hoechst (1/100) (*blue*). Scale bar 10µm. (**B**) Scheme of BRET^2 assay using σ_1Rluc and GHS-R1a-GFP^2. (**C**) HEK-293T cells were transfected with a constant amount of cDNA (0.075µg) for σ_1Rluc and increasing amounts of GHS-R1a-GFP^2 (from 0.5-3µg cDNA) or A$_{2A}$R-GFP^2 (from 0.5- 2.5µg cDNA) as negative control. Values are the mean (in milliBRET units: mBU) ± S.E.M. from 6 to 8 different experiments.

To identify a potential direct interaction between $\sigma_1 R$ and GHS-R1a, we developed BRET experiments, transfecting a constant amount of cDNA for $\sigma_1 R$-Rluc (0.075 µg cDNA) and increasing amounts of cDNA for GHS-R1a-GFP^2 (0.5 to 3 µg cDNA). A saturation BRET curve was obtained thus indicating a specific interaction between $\sigma_1 R$ and GHS-1a (BRET$_{max}$ 371 $\pm$ 38 mBU, BRET$_{50}$ 68 $\pm$ 23) (Figures 4.13B-C). In contrast, when adenosine $A_{2A}R$-GFP^2 (0.5 to 2.5 µg cDNA) was used as negative control instead of GHS-R1a receptor, a linear plot with low BRET values was obtained.

Quaternary structure of the heteromeric complex between $\sigma_1 R$ and GHS-R1a

We next addressed the quaternary structure of $\sigma_1 R$-GHS-R1a complexes taking advantage of the recent publication of the crystal structure of $\sigma_1 R$ in a trimeric arrangement. This complex cannot be understood without considering that the GHS-R1a receptor form homomeric interactions with GHS-R1a and/or heteromeric interactions with GHS-R1b (Chow *et al.*, 2012; Mary *et al.*, 2013; Navarro *et al.*, 2016). Thus, to identify the TM interfaces involved in GHS-R1a-GHS-R1a homodimerization, GHS-R1a-GHS-R1b heterodimerization, and their TM interacting interfaces with $\sigma_1 R$, we used synthetic peptides with the amino acid sequence of TMs 1-7 of GHS-R1a fused to the transactivator of transcription (TAT, with GRKKRRQRRR sequence) peptide. These cell-penetrating peptides interact with the TM domain of membrane proteins and can selectively disrupt interactions between proteins, i.e. GPCR protomers (Hebert *et al.*, 1996; Ng *et al.*, 1996).

These peptides were first tested in HEK-293T cells expressing GHS-R1a-nYFP (0.75 µg cDNA) and GHS-R1a-cYFP (0.5 µg cDNA) to

find Bimolecular Fluorescense Complementation (BiFC) (approx. 4,000 units of fluorescence). Notably, in the presence of interference peptides, we observed that fluorescence decreased by two-fold only with TM5 and TM6 (Figure 4.14A). These results pointed to the TM 5/6 interface for GHS-R1a-GHS-R1a homodimerization. Similar results were obtained for GHS-R1a-GHS-R1b heterodimerization (TM 5/6 interface, Figure 4.14B), despite the GHS-R1b isoform lacks TMs 6 and 7 relative to GHS-R1a. Interestingly, when cells were transfected with cDNAs for GHS-R1a-nYFP, GHS-R1b-cYFP, and non-fused GHS-R1a, fluorescence was reduced in the presence of TM4, TM5, and TM6 peptides (Figure 4.14C). These results suggest an arrangement of protomers in which homodimerization of GHS-R1a occurs via the TM5/6 interface and heterodimerization of GHS-R1a and GHS-R1b occurs via the TM4/5 interface (Figure 4.14D). The fluorescence decrease induced by the TM6 peptide of GHS-R1a (in addition to TM4 and TM5, Figure 4.14C) could indicate that this peptide also restricts the interactions with TM 4 of GHS-R1b (Figure 4.14D).

Next, we investigated the TM domains of GHS-R1a involved in the interaction with σ_1R. Remarkably, in HEK-293T cells co-expressing GHS-R1a-nYFP (0.75 µg cDNA) and σ_1R-cYFP (0.5 µg cDNA), fluorescence complementation (4,000 units, which confirms the formation of GHS-R1a-σ_1R complexes) was significantly reduced in the presence of TM1, TM2, TM5 or TM6 interference peptides (Figure 4.14E). This clearly indicates the existence of two different interacting interfaces between GHS-R1a and σ_1R, involving either TM1/2 or TM5/6 interfaces of GHS-R1a and the single TM helix of σ_1R. When similar experiments were performed with σ_1R-cYFP (0.5 µg cDNA) and GHS-R1b-nYFP (0.5 µg cDNA), fluorescent signal

(3,500 units) decreased in the presence of TM1 and TM2 but not TM5 or TM6 (Figure 4.14F) because GHS-R1b lacks TMs 6 and 7 relative to GHS-R1a. Thus we hypothesized that TM1/2 or TM5/6 interfaces could be involved in the interaction between the GPCRs and σ_1R. To obtain data in a more physiological set-up, HEK-293T cells were transfected with cDNAs for GHS-R1a-nYFP and σ_1R-cYFP in the presence of non-fused GHS-R1b. As fluorescence complementation (5,000 fluorescence units) was reduced only by TM1 and TM2 (Figure 4.14G). So, the formation of the GHS-R1a-GHS-R1b heterotetramer via TM4/5 and TM5/6 interfaces (Figure 4.14D) only permits σ_1R to interact with GHS-R1a via the free TM1/2 interface.

Using structural details on TM interfaces of GPCR oligomers (Cordomí *et al.*, 2015) and the crystal structure of σ_1R (Schmidt *et al.*, 2016), together with the results from BiFC experiments performed in the absence and presence of disrupting TM peptides, we constructed a computational molecular model (see *sub-section 4.2.1.2. Material and Methods*) of the GHS-R1a in complex with G_i and σ_1R (Figure 4.14H). To simplify, GHS-R1b, which lack of TMs 6 and 7, was not included in the model. We assumed that at normal expression levels, TM 1/2 is the only possible interface for the GHS-R1–σ_1R complex as TM 5/6 is occupied by the GHS-R1 homo/heterodimer. This model tells that a single GPCR protomer of GHS-R1a cannot simultaneously bind σ1R via the TM 1/2 interface and a G protein due to the steric clash between the intracellular voluminous C-terminal tail of σ_1R, containing a rigid cupin-like β-barrel fold that forms the buried ligand-binding site and the βγ subunits of G_i (Figure 4.14H).
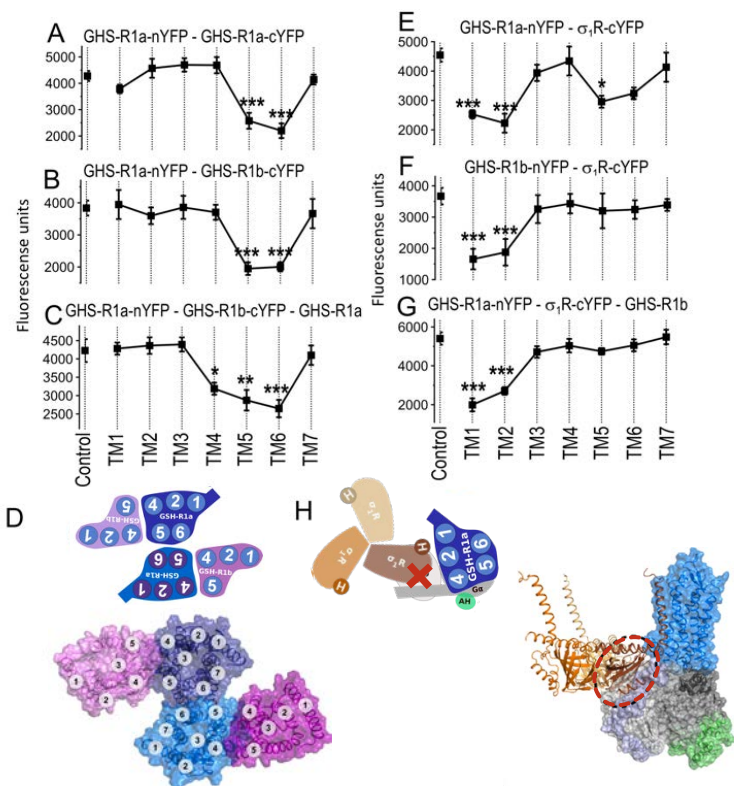
Figure 4.14| Effect of interfering peptides on the σ_1R-GHS-R1a interaction. Panels A-C, E-G: BiFC complementation experiments were performed in HEK-293T cells transfected with cDNA for GHS-R1a-nYFP and GHS-R1a-cYFP (**A**), with cDNA for GHS-R1a-nYFP and GHS-R1b-cYFP (**B**) with cDNA for GHS-R1a-nYFP and GHS-R1b-cYFP in the presence of 1.5 µg cDNA for GHS-R1a-nYFP (**C**) with cDNA for GHS-R1a-nYFP and σ_1R-cYFP (**E**), with cDNA for GHS-R1b-nYFP and σ_1R-cYFP (**F**) or with cDNA for GHS-R1a-nYFP and GHS-R1b-cYFP in the presence of 1.5 µg cDNA of GHS-R1a not fused (**G**). Prior to fluorescence determination, cells were treated with each of the interfering peptides (TM1 to TM7, 4 µM) during 4 h. Values are the mean ± S.E.M. from 8 to 10 different experiments. One-way

ANOVA followed by Dunnett's *post-hoc* test showed a significant effect of treatments versus control conditions. * p<0.05, ** p<0.01 and *** p<0.001. **Panel D**: Structural model for the GHS-R1a-GHS-R1b heterotetramer (GHS-R1a: *blue*, GHS-R1b: *purple*) viewed from the extracellular side (schematical and 3D models at the top and bottom, respectively). TM helices are indicated by circles (GHS-R1a: 1-7, σ_1R: H). **Panel H**: Structural model consisting of a GHS-R1a (*light blue*) coupled to G_i ($G\alpha$ Ras-like domain: *light grey*, $G\alpha$ alpha helical domain: *green*, $G\beta$: *dark grey* and $G\gamma$: *purple*) and in complex with a σ_1R homotrimer (in *red*, *orange* and *yellow*) viewed from the extracellular side (**left;** schematical model) or from the membrane (**right;** 3D model). It illustrates that a GHS-R1 protomer bound to the $G\alpha$ subunit of G_i cannot simultaneously bind σ_1R via the TM1/2 interface. Proteins are displayed with a transparent surface and cartoon, except σ_1R which is displayed with a cartoon only. The red cross and dashed ellipse outline the regions that clash.

Thus, we constructed a computational model consisting of the GHS-R1a homodimer, a σ_1R homotrimer and a G protein that fits with the requirements of the biochemical data and takes into account all available structural constraints (Figure 4.15). At the GHS-R1a homodimer, one protomer binds the G protein and the second protomer is responsible for the binding of the σ_1R TM helix (Figure 4.15). Interestingly, this model predicts that the cytoplasmic domain of one protomer of the σ_1R trimeric structure contact the α-subunit of G_i, suggesting that the functional protein of σ_1R is seemingly constituted by an homotrimer (Figure 4.15).

We used molecular dynamics (MD) simulations (see *sub-section 4.2.1.2. Material and Methods*) to evaluate the stability of GHS-R1a homodimer in complex with G_i and σ_1R homotrimer and the proximity between the α-subunit of Gi and σ_1R. We monitored the root mean square deviation (RMSD) of the alpha carbons (C_α)

of the whole complex (in *black*), the C_α of TM 5/6 interface of GHS-R1a homodimer (in *blue*) and the C_α of TM 1/2 interface of GHS-R1a and σ_1R (in *orange*) (Figure 4.16A), suggesting that all achieve stability at the last 400 ns (at ~6 Å, ~2.5 Å and ~2.5 Å, respectively). We also measured the distance between the helical domain of α-subunit of G_i (αAH, in *green*) and the closer protomer of σ_1R homotrimer (in *yellow*), which decreased from 32 Å to 25 Å (Figure 4.16B).
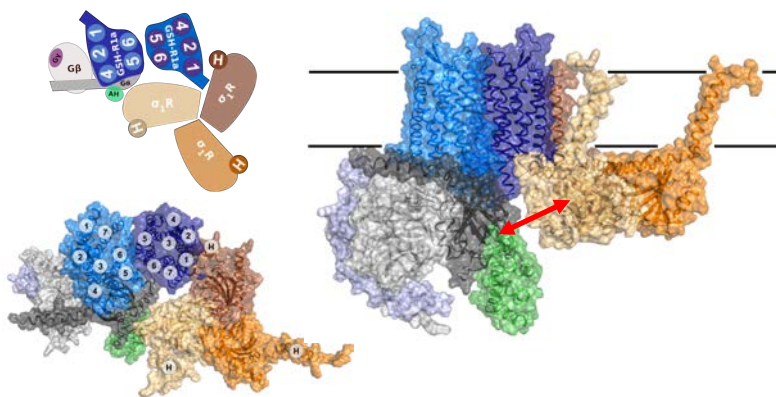

Figure 4.15 | GHS-R1a homodimer bind to σ_1R homotrimer and G_i protein. Computational structural model (see *sub-section 4.2.2.2. Material and Methods*) consisting of a GHS-R1a homodimer (GHS-R1a G_α-bound: *light blue*, GHS-R1a G_α-unbound: *dark blue*) in complex with a σ_1R homotrimer (in *red*, *orange* and *yellow*) coupled to G_i ($G\alpha$ Ras-like domain: *light grey*, $G\alpha$ helical domain: *green*, $G\beta$: *dark grey* and $G\gamma$: *purple*) viewed from the extracellular side (**left**; schematical and 3D models at the top and bottom, respectively) or from the membrane (**right**). Proteins are displayed with a transparent surface and cartoon, except σ_1R which is displayed with a cartoon only. TM helices are indicated by circles (GHS-R1a: 1-7, σ_1R: H).
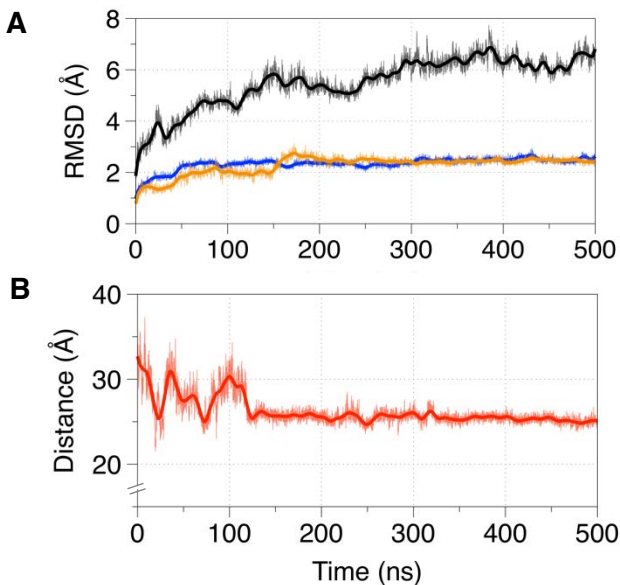
Figure 4.16 | Analysis of the MD simulation of GHS-R1a homodimer bind to σ_1R homotrimer and G_i protein. (**A**) Time-evolution of the RMSD of the C_α of the GHS-R1a homodimer bind σ_1R and G_i protein complex (in *black*), TM 5/6 interface of GHS-R1a homodimer (in *blue*) and TM 1/2 interface of GHS-R1a and TM of σ_1R (in *orange*). (**B**) Time-evolution of the distance between the Cα atom of Glu58 (at Ras-domain) of α-subunit of Gi and Cα atom of Glu172 (key residue in the binding site) of the closer protomer of the σ_1R homotrimer (red arrow at *Figure 4.15*).

Cocaine increases colocalization of σ_1R and GHS-R1a

As cocaine binds σ_1R, which establishes direct interactions with ghrelin receptors as shown above, we hypothesized that cocaine affects ghrelin-mediated signals. First, we investigated the effect of cocaine in GHS-R1a expression. Immunocytochemical assays were

performed in cells expressing 0.75 µg cDNA for σ_1R-YFP and 1.66 µg cDNA for GHS-R1a-Rluc after cocaine addition (30 µM, 30 min). Figure 4.17A shows that plasma membrane expression of σ_1R increased when these cells were treated with a physiologically relevant dose of cocaine (Navarro *et al.*, 2010). A similar increase was observed when cells were incubated with the σ_1R agonist PRE-084 (100 nM, 30 min). The expression of GHS-R1a was not modified upon treatment with cocaine or PRE-084 but, interestingly, colocalization of σ_1R and GHS-R1a at the cell surface increased. Thus, cocaine and the σ_1R specific ligand PRE-084 are able to concomitantly affect co-expression of both receptors at the cell surface. Second, we evaluated the effect of cocaine and PRE-084 in the heteromerization of σ_1R and GHS-R1a. No differences in energy transfer recordings are observed in the absence of ligand or the presence of cocaine (30 µM) or PRE-084 (100 nM) in HEK-293T cells transfected with 0.075 µg cDNA for σ_1R-Rluc and 1.5 µg cDNA for GHS-R1a-GFP² (Figure 4.17B).

<u>Cocaine inhibits GHS-R1a signalling</u>

We first evaluated the effect of cocaine and PRE-084 on GHS-R1a-mediated signalling by measuring cAMP levels. HEK-293T cells endogenously express σ_1R but do not express ghrelin receptors (Navarro *et al.*, 2010). Moreover, it is known that low concentrations of GHS-R1b expression significantly increase GHS-R1a signalling (Navarro *et al.*, 2015). Thus, to analyse GHS-R1a signalling pathways in HEK-293T cells, we co-expressed GHS-R1a (1.66 µg cDNA) with low amounts of GHS-R1b (0.25 µg cDNA). Stimulation of these cells with ghrelin (100 nM) in the presence of forskolin (0.5 µM) significantly decreased cAMP levels (Figure 4.17C). This agrees with the previously reported coupling of ghrelin receptors with G_i

(www.guidetopharmacology.org) (Navarro *et al.*, 2016). The effect of the ghrelin peptide on forskolin-induced cAMP levels was completely blocked by pre-treatment with the GHS-R1a selective antagonist YIL-781 (2 µM) (Figure 4.17C). Interestingly, when cells were treated with cocaine (30 µM, 15 min) of the σ_1R agonist (PRE-084, 100 nM, 15 min) prior to ghrelin stimulation, the decrease in cAMP levels was avoided in a similar manner as the GHS-R1a selective antagonist YIL-781 (Figure 4.17C). This suggests that cocaine behaves as agonist of σ_1R and inhibits GHS-R1a signalling as efficiently as the GHS-R1a selective antagonist YIL-781 bound to the orthosteric binding site.

Measurement of ERK1/2 phosphorylation in HEK-293T cells transfected with GHS-R1a and GHS-R1b after stimulation with ghrelin shows an increase of 80% (Figure 4.17D). Similar to measurements of cAMP levels (Figure 4.17C), the GHS-R1a selective antagonist YIL-781 and the σ_1R agonists cocaine and PRE-084 inhibited ghrelin effects (Figure 4.17D). This indicates that cocaine and PRE-084 not only affects the α_i dependent pathway, but also the $\beta\gamma$ dependent signalling. We can, thus, conclude that the GHS-R1a selective antagonist YIL-781, cocaine and PRE-084 inhibited ghrelin effects, as measured by cAMP levels and ERK1/2 phosphorylation.

Figure 4.17 | Cocaine effects on ghrelin-mediated signalling. (**A**) HEK-293T cells transfected with cDNA for σ_1R-YFP, GHS-R1a-Rluc or both were treated with cocaine or PRE-084, then were monitored by the YFP fluorescence (*green*) or using a monoclonal anti-Rluc primary antibody and a cyanine-3-conjugated secondary antibody (*red*). Co-localization is shown in *yellow*. Nuclei were stained

with Hoechst (*blue*). Scale bar 10μm. (**B**) HEK-293T cells were transfected with cDNA for σ₁R-Rluc and GHS-R1a-GFP² and treated with cocaine, PRE-084 or vehicle for 30 min. Afterwards, the energy transfer signal was measured. Values are the mean ± S.E.M. of 7 different experiments. One-way ANOVA followed by Dunnett's *post-hoc* test did not show any significant effect of treatments versus control. **Panels C-D**: HEK-293T cells were transfected with GHS-R1a cDNA and GHS-R1b cDNA and treated with cocaine (*red*), the σ₁R agonist PRE-084 (*green*), the GHS-R1a antagonist YIL-781 (*blue*), or vehicle (*black*). Cells were then treated with ghrelin followed by forskolin only in cAMP accumulation assay (**C**). ERK1/2 phosphorylation (**D**) was analysed using an AlphaScreen® SureFire® kit (Perkin Elmer). Values are the mean ± S.E.M. from 8 to 11 different experiments. One-way ANOVA followed by Dunnett's *post-hoc* test showed a significant effect of treatments versus forskolin (cAMP assays, **C**) or control (pERK1/2 assays, **D**), ** $p < 0.01$ and *** $p < 0.001$, and a significant effect of treatments versus ghrelin, ## $p < 0.01$ and ### $p < 0.001$.
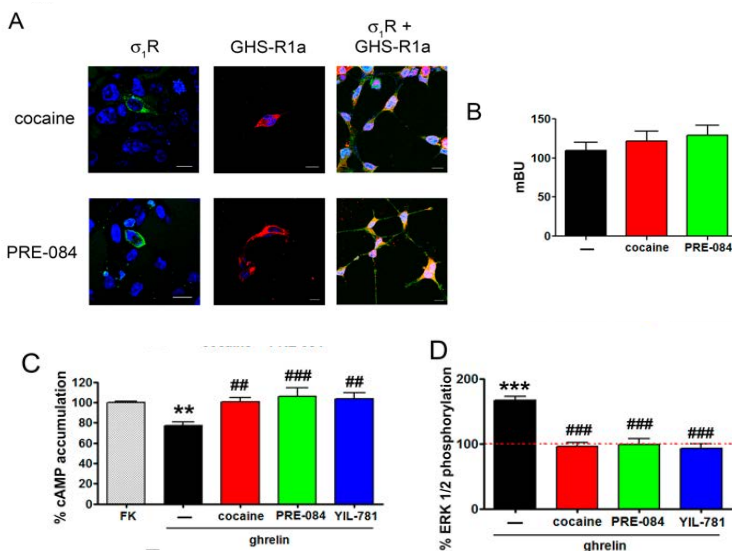
Cocaine inhibition of GHS-R1a signalling is mediated by σ_1R

To check whether cocaine inhibition of ghrelin-induced signalling was due to its interaction with σ_1R, HEK-293T cells expressing GHS-R1a (1.66 µg cDNA) and GHS-R1b (0.25 µg cDNA) were transfected with a siRNA designed to knock-down expression of σ_1R (3 µg siRNA). Cells incorporating siRNA responded to 100 nM of ghrelin in both, 0.5 µM forskolin-induced cAMP determination and ERK1/2 phosphorylation, with similar results to those in cells without siRNA (Figures 4.18A-B). However, in the presence of siRNA, cocaine (30 µM) or PRE-084 (100 nM) had no effect on ghrelin-induced signals while pre-treatment with YIL-781 (2 µM) blocked ghrelin-induced GHS-1A activation. These results show that cocaine effects over GHS-R1a receptor are mediated by σ_1R.

Disruption of the heteromeric complex between σ_1R/GHS-R1a by the TM1 interference peptide blocks the effect of cocaine on GHS-R1a function

As proposed above from data using TAT-fused synthetic peptides, the single TM helix of σ_1R likely interacts with TMs 1 and 2 of GHS-R1a. Accordingly, we can hypothesize that addition of the TM1 interference peptide would abolish the effect of cocaine on GHS-R1a function. Thus, HEK-293T cells expressing GHS-R1a (1.66 µg cDNA) were treated during 4h with 4 µM of TM1 (or TM7 as negative control). In agreement with our hypothesis, disruption of σ_1R-GHS-R1a heteromeric complex was achieved by TM1, but not TM7; in these experimental conditions in which TM1 peptide was present only the GHS-R1a selective antagonist YIL-781 (1µM) blocked ghrelin (100 nM) stimulation, whereas cocaine (30 µM) or PRE-084 (100 nM) did not display any effect in either cAMP levels

(Figures 4.18C-D). These results demonstrate that disruption of the GHS-R1a-σ_1R interaction using TM1, alters the cocaine effect on ghrelin receptors, thus reinforcing the idea that cocaine modulates GHS-R1a receptor function via σ_1R.
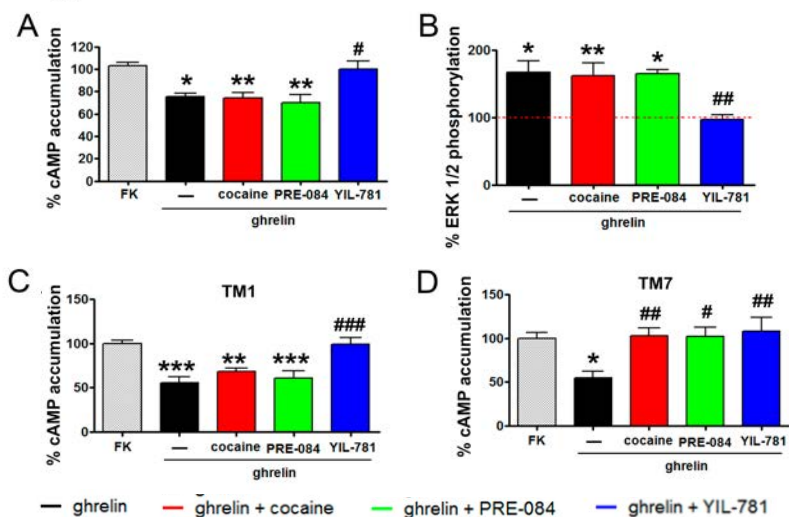

Figure 4.18 | Cocaine effects over GHs-R1a signalling depended on σ_1R expression. Panels A-B: HEK-293T cells were transfected with GHS-R1a, GHS-R1b and siRNA for σ_1R. These cells were treated with cocaine (*red*), the σ_1R agonist, PRE-084 (*green*), the GHS-R1a antagonist YIL-781 (*blue*), or vehicle (*black*) followed by ghrelin stimulation and forskolin in cAMP accumulation. Then, cAMP levels (**A**) and pERK1/2 (**B**) signals were recorded. Values are the mean ± S.E.M. from 8 to 11 different experiments. One-way ANOVA followed by Dunnett's *post-hoc* test showed a significant effect of treatments versus forskolin (cAMP assays, **A**) or control (pERK1/2 assays, **B**), * $p<0.05$, ** $p<0.01$, and a significant effect of treatments versus ghrelin, # $p<0.05$, ## $p<0.01$. **Panels C-D**: HEK-293T cells expressing GHS-R1a were treated for 4h with TM1 (**C**) or TM7 (**D**) TAT-peptides. Cells were subsequently treated with cocaine

129

(*red*), PRE-084 (*green*), YIL-781 (*blue*) or vehicle (*black*). Cells were then treated with ghrelin and forskolin only in cAMP experiments and cAMP levels were determined 15 min afterwards. Values are the mean ± S.E.M. of 6 different experiments. One-way ANOVA followed by Dunnett's *post-hoc* test showed a significant effect of treatments versus forskolin, * $p<0.05$, ** $p<0.01$ and *** $p<0.001$, and a significant effect of treatments versus ghrelin, # $p<0.05$, ## $p<0.01$ and ### $p<0.001$.

Next, we attempted to give insight into the he mechanism by which cocaine binds to σ_1R and blocks GHS-R1a function. Our structural model predicts that a single protomer of GHS-R1a cannot simultaneously bind σ_1R, via the TM1/2 interface, and G_i; accordingly, σ_1R may impede G_i binding and in consequence GHS-R1a function. However, this does not seem reasonable due to the possible formation of a GHS-R1a homodimer in which one protomer binds G_i and the second protomer σ_1R (Figure 4.15). Notably, this model and the subsequent MD simulation positions the cytoplasmic domain of one protomer of the σ_1R trimeric structure near the α-helical domain (αAH) of the α-subunit (Figure 4.15 and Figure 4.16B). It has been shown that the mechanism for receptor-catalysed nucleotide exchange in G proteins involves a large-scale opening of αAH, from the Ras domain, allowing GDP to freely dissociate (Dror *et al.*, 2015). This opening of αAH is not feasible in the presence of the σ_1R trimeric structure bound to TMs 1 and 2 of GHS-R1a. Modification of the GHS-R1a-σ_1R interaction, by inserting the TAT-fused TM1 peptide, would increase the distance between cytoplasmic domain of σ_1R and αAH, facilitating Gi function as found in our assays.

4.2.1.4. *Discussion*

The endogenous ligand of $\sigma_1 R$ remains unknown, however, synthetic agonists and antagonists are available. PRE-084 is a selective agonist due to its ability to dose-dependently dissociate $\sigma_1 R$ from a binding immunoglobulin protein/78kDa glucose-regulated protein (BiP/GPR-78) (Hayashi and Su, 2007). Despite not having a specific signalling machinery, $\sigma_1 R$ operates via translocation to the plasma membrane and via protein-protein-mediated modulation of cell responses upon agonist activation (Su *et al.*, 2016). This involves calcium signalling and ion activation of channels (Wu and Bowen, 2008), in addition to the regulation of GPCRs (Su *et al.*, 2016). For instance, it has been shown that $\sigma_1 R$ is involved in the negative control that glutamate N-methyl-d-aspartate acid receptors (NMDARs) exert on opioid anti-nociception (Rodriguez-Munoz *et al.*, 2015). Thus, $\sigma_1 R$ antagonists would enhance anti-nociception and reduce neuropathic pain induced by µ-opioid receptors. Another remarkable example is the extent control of $\sigma_1 R$ in the interaction between cannabinoid CB_1 and NMDAR receptors, whose failure might constitute a vulnerability factor for cannabis abuse, potentially precipitating schizophrenia (Sánchez-Blázquez *et al.*, 2014). Upon demonstrating that cocaine binds to $\sigma_1 R$ even at doses attained at recreational use, $\sigma_1 R$ is proposed to mediate locomotor activation (Menkel *et al.*, 1991; Barr *et al.*, 2015), seizures (Matsumoto *et al.*, 2001), drug sensitization (Ujike, Kuroda and Otsuki, 1996), and reward actions (Romieu, Martin-Fardon and Maurice, 2000; Romieu *et al.*, 2002). Reduction of $\sigma_1 R$ levels by injection of antisense nucleotides results in less convulsions (Matsumoto *et al.*, 2002), whereas the action of agonists or antagonists exacerbate or minimize, respectively, cocaine effects (Matsumoto *et al.*, 2002, 2003, 2004).

Although, it is known that activation of σ_1R by agonists is also involved in the appetitive properties of cocaine (Romieu *et al.*, 2002), the mechanism of action remains unknown. In this manuscript, we reveal for the first time that σ_1R mediate the hunger-suppressive action of cocaine by interacting with orexigenic ghrelin receptors.

The *nucleus accumbens*, one of the structures that conform the *striatum*, is part of the reward system. This system produces a pleasant sensation in front of food and other important actions for the individual survival (Kim and Hikosaka, 2015). It is important to note that in these mesolimbic regions, GHS-R1a is co-expressed with cocaine-sensitive σ_1R in neurons. From mechanistic and molecular point of views, this report highlights an interaction between σ_1R and GHS-R1a that is translated into a strong inhibition of ghrelin-induced GHS-R1a signalling, as measured by G protein dependent (cAMP accumulation) and independent (MAPK phosphorylation) signalling pathways. We have shown in transfected HEK-293T cells that pre-treatment with the σ_1R agonists cocaine or PRE-084 inhibits ghrelin-mediated signalling in a similar manner as the GHS-R1a antagonist YIL-781. This effect is mediated by σ_1R since, in σ_1R-RNAi treated cells, cocaine or PRE-084 had no effect on ghrelin-induced signals while YIL-781 maintained its effect.

Kotagle and collaborators have described the potent orexigenic neuropeptide Y (NPY) as a possible endogenous ligand for a subpopulation of σ_1Rs (Kotagale *et al.*, 2014), linking stimulation of sigma receptors with hunger. Our results show that cocaine binding to σ_1R counteracts the feeling of hunger. Part of this effect could also be due to the competition between cocaine and neuropeptide Y receptors to interact σ_1R.

The structure of σ_1R has traditionally considered to be formed by two TM helices. However, the recently released crystal structure of σ_1R has shown a single TM domain and a C-terminal tail having a buried ligand-binding site that arranges into homotrimers (Schmidt, et al., 2016). We have proposed, using TAT-fused synthetic peptides together with BiFC assays, that this single TM helix of σ_1R can be recognized by two different interacting interfaces of the 7TM bundle of GPCRs, either the interfaces formed by TMs 1 and 2 or TMs 5 and 6. Oligomerization of GPCRs via a particular interface might guide the interacting interface of σ_1R. In the case of σ_1R-GHS-R1a, due to the formation of the GHS-R1a-GHS-R1b heterotetramer via TM 4/5 and TM 5/6 interfaces, σ_1R can interact with GHS-R1a via the free TM 1/2 interface. We would like to speculate that, because σ_1R is a homotrimer, the two additional TM helices can bind two additional GHS-R1a-GHS-R1b heterotetramers, suggesting the possible existence of higher order complexes between GHS-R1a and σ_1R by the successive combination of these units. These clusters may form specialized machineries of σ_1R-mediated signalling.

4.2.1.5. *References*

Abraham, M. J. *et al.* (2015) 'Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers', *SoftwareX*, 1–2, pp. 19–25.

Barr, J. L. *et al.* (2015) 'Mechanisms of activation of nucleus accumbens neurons by cocaine via sigma-1 receptor-inositol 1,4,5-trisphosphate-transient receptor potential canonical channel pathways', *Cell Calcium*.

Berger, O., Edholm, O. and Jähnig, F. (1997) 'Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature.', *Biophys. J.*, 72(5), pp. 2002–2013.

Borroto-Escuela, D. O. *et al.* (2014) 'The G protein-coupled receptor heterodimer network (GPCR-HetNet) and its hub components.', *Int. J. Mol. Sci.*, 15(5), pp. 8570–8590.

Cassidy, R. M. and Tong, Q. (2017) 'Hunger and Satiety Gauge Reward Sensitivity.', *Front. Endocrinol. (Lausanne).*, 8, p. 104.

Chan, C.-B. and Cheng, C. H. K. (2004) 'Identification and functional characterization of two alternatively spliced growth hormone secretagogue receptor transcripts from the pituitary of black seabream Acanthopagrus schlegeli.', *Mol. Cell. Endocrinol.*, 214(1–2), pp. 81–95.

Chow, K. B. S. *et al.* (2012) 'The truncated ghrelin receptor polypeptide (GHS-R1b) is localized in the endoplasmic reticulum where it forms heterodimers with ghrelin receptors (GHS-R1a) to attenuate their cell surface expression.', *Mol. Cell. Endocrinol.* Elsevier Ireland Ltd, 348(1), pp. 247–54.

Conn, P. M. and Bowers, C. Y. (1996) 'A new receptor for growth hormone-release peptide.', *Science*, 273(5277), p. 923.

Corbera, J. *et al.* (2006) 'A Medicinal-Chemistry-Guided Approach to Selective and Druglike Sigma 1 Ligands', *ChemMedChem*, 1(1), pp. 140–154.

Cordomí, A., Caltabiano, G. and Pardo, L. (2012) 'Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters', *J. Chem. Theory Comput.*, 8(3), pp. 948–958.

Cordomí, A. *et al.* (2015) 'Structures for G-Protein-Coupled Receptor Tetramers in Complex with G Proteins', *Trends Biochem. Sci.* Elsevier Ltd, 40(10), pp. 548–551.

Damian, M. *et al.* (2015) 'Ghrelin receptor conformational dynamics regulate the transition from a preassembled to an active receptor:Gq complex.', *Proc. Natl. Acad. Sci. U. S. A.*, 112(5), pp. 1601–1606.

Dror, R. O. *et al.* (2015) 'Structural basis for nucleotide exchange in heterotrimeric G proteins', *Science (80-.).*, 348(6241), pp. 1361–1365.

Egloff, P. *et al.* (2014) 'Structure of signalling-competent neurotensin receptor 1 obtained by directed evolution in Escherichia coli.', *Proc. Natl. Acad. Sci. U. S. A.*, 111(6), pp. E655-62.

Geelissen, S. M. E. *et al.* (2003) 'Distribution and regulation of chicken growth hormone secretagogue receptor isoforms.', *Gen. Comp. Endocrinol.*, 134(2), pp. 167–174.

Hayashi, T. and Su, T.-P. (2007) 'Sigma-1 receptor chaperones at the ER-mitochondrion interface regulate Ca(2+) signalling and cell survival.', *Cell*, 131(3), pp. 596–610.

Hebert, T. E. *et al.* (1996) 'A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation.', *J. Biol. Chem.*, 271(27), pp. 16384–92.

Howick, K. *et al.* (2017) 'From Belly to Brain: Targeting the Ghrelin Receptor in Appetite and Food Intake Regulation', *Int. J. Mol. Sci.* Multidisciplinary Digital Publishing Institute, 18(2), p. 273.

Hradsky, J. *et al.* (2013) 'Super-resolution microscopy of the neuronal calcium-binding proteins Calneuron-1 and Caldendrin.', *Methods Mol. Biol.*, 963, pp. 147–169.

Kim, H. F. and Hikosaka, O. (2015) 'Parallel basal ganglia circuits for voluntary and automatic behaviour to reach rewards.', *Brain*, 138(Pt 7), pp. 1776–800.

Kotagale, N. R. *et al.* (2014) 'Involvement of hypothalamic neuropeptide Y in pentazocine induced suppression of food intake in rats.', *Neuropeptides*, 48(3), pp. 133–141.

Krumm, B. E. *et al.* (2015) 'Structural prerequisites for G-protein activation by the neurotensin receptor.', *Nat. Commun.*, 6, p. 7895.

Kutzner, C. *et al.* (2015) 'Best bang for your buck: GPU nodes for GROMACS biomolecular simulations', *J. Comput. Chem.*, 36(26), pp. 1990–2008.

Lever, J. R. *et al.* (2016) 'Cocaine occupancy of sigma1 receptors and dopamine transporters in mice', *Synapse*, 70(3), pp. 98–111.

Lindorff-Larsen, K. *et al.* (2010) 'Improved side-chain torsion potentials for the Amber ff99SB protein force field', *Proteins Struct. Funct. Bioinforma.*, 78(8), pp. 1950–1958.

Mary, S. *et al.* (2013) 'Heterodimerization with Its splice variant blocks the ghrelin receptor 1a in a non-signalling conformation: a study with a purified heterodimer assembled into lipid discs.', *J. Biol. Chem.*, 288(34), pp. 24656–24665.

Matsumoto, R. R. *et al.* (2001) 'Rimcazole analogs attenuate the convulsive effects of cocaine: correlation with binding to sigma receptors rather than dopamine transporters.', *Neuropharmacology*, 41(7), pp. 878–886.

Matsumoto, R. R. *et al.* (2002) 'Involvement of sigma receptors in the behavioral effects of cocaine: evidence from novel ligands and antisense oligodeoxynucleotides.', *Neuropharmacology*, 42(8), pp. 1043–55.

Matsumoto, R. R. *et al.* (2003) 'Sigma receptors: potential medications development target for anti-cocaine agents.', *Eur. J. Pharmacol.*, 469(1–3), pp. 1–12.

Matsumoto, R. R. *et al.* (2004) 'Novel analogs of the sigma receptor ligand BD1008 attenuate cocaine-induced toxicity in mice.', *Eur. J. Pharmacol.*, 492(1), pp. 21–26.

McCracken, K. A., Bowen, W. D. and Matsumoto, R. R. (1999) 'Novel sigma receptor ligands attenuate the locomotor stimulatory effects of cocaine.', *Eur. J. Pharmacol.*, 365(1), pp. 35–38.

Mei, J. and Pasternak, G. W. (2002) 'Sigma1 receptor modulation of opioid analgesia in the mouse.', *J. Pharmacol. Exp. Ther.*, 300(3), pp. 1070–1074.

Menkel, M. *et al.* (1991) 'Selective sigma ligands block stimulant effects of cocaine.', *Eur. J. Pharmacol.*, 201(2–3), pp. 251–252.

Moreno, E. *et al.* (2014) 'Cocaine disrupts histamine H3 receptor modulation of dopamine D1 receptor signalling: σ1-D1-H3 receptor complexes as key targets for reducing cocaine's effects.', *J. Neurosci.*, 34(10), pp. 3545–3558.

Navarro, G. *et al.* (2010) 'Direct involvement of sigma-1 receptors in the dopamine D1 receptor-mediated effects of cocaine.', *Proc. Natl. Acad. Sci. U. S. A.*, 107(43), pp. 18676–18681.

Navarro, G. *et al.* (2013) 'Cocaine inhibits dopamine D2 receptor signalling via sigma-1-D2 receptor heteromers.', *PLoS One*, 8(4), p. e61245.

Navarro, G. *et al.* (2015) 'Orexin-corticotropin-releasing factor receptor heteromers in the ventral tegmental area as targets for cocaine.', *J. Neurosci.*, 35(17), pp. 6639–6653.

Navarro, G. *et al.* (2016) 'A significant role of the truncated ghrelin receptor GHS-R1b in ghrelin-induced signalling in neurons', *J. Biol. Chem.*, 291(25), pp. 13048–13062.

Ng, G. Y. K. *et al.* (1996) 'Dopamine D2 Receptor Dimers and Receptor-Blocking Peptides', *Biochem. Biophys. Res. Commun.*, 227(1), pp. 200–204.

Price, D. J. and Brooks, C. L. (2004) 'A modified TIP3P water potential for simulation with Ewald summation', *J. Chem. Phys.*, 121(20), pp. 10096–10103.

Rasmussen, S. G. F. *et al.* (2011) 'Crystal structure of the β2 adrenergic receptor-Gs protein complex.', *Nature*, 477(7366), pp. 549–55.

Rodriguez-Munoz, M. *et al.* (2015) 'The sigma1 receptor engages the redox-regulated HINT1 protein to bring opioid analgesia under NMDA receptor negative control', *Antioxid Redox Signal*, 22(10), pp. 799–818.

Romieu, P. *et al.* (2002) 'Involvement of the sigma(1) receptor in cocaine-induced conditioned place preference: possible dependence on dopamine uptake blockade.', *Neuropsychopharmacology*, 26(4), pp. 444–455.

Romieu, P., Martin-Fardon, R. and Maurice, T. (2000) 'Involvement of the sigma1 receptor in the cocaine-induced conditioned place preference.', *Neuroreport*, 11(13), pp. 2885–8.

Sánchez-Blázquez, P. *et al.* (2014) 'The calcium-sensitive Sigma-1 receptor prevents cannabinoids from provoking glutamate NMDA receptor hypofunction: implications in antinociception and psychotic diseases.', *Int. J. Neuropsychopharmacol.*, 17(12), pp. 1943–55.

Schmidt, H. R. *et al.* (2016) 'Crystal structure of the human σ1 receptor.', *Nature*, 532(7600), pp. 527–530.

Skuza, G. (1999) 'Effect of sigma ligands on the cocaine-induced convulsions in mice.', *Pol. J. Pharmacol.*, 51(6), pp. 477–483.

Su, T.-P. *et al.* (2016) 'The Sigma-1 Receptor as a Pluripotent Modulator in Living Systems', *Trends Pharmacol. Sci.*, 37(4), pp. 262–278.

Sun, H. *et al.* (2016) 'Development of Novel Alkoxyisoxazoles as Sigma-1 Receptor Antagonists with Antinociceptive Efficacy.', *J. Med. Chem.*, 59(13), pp. 6329–6343.

Tesmer, J. J. G. *et al.* (1997) 'Structure of RGS4 bound to AlF4 - activated G(i alpha1): stabilization of the transition state for GTP hydrolysis.', *Cell*, 89(2), pp. 251–261.

Ujike, H., Kuroda, S. and Otsuki, S. (1996) 'sigma Receptor antagonists block the development of sensitization to cocaine.', *Eur. J. Pharmacol.*, 296(2), pp. 123–128.

Wu, Z. and Bowen, W. D. (2008) 'Role of sigma-1 receptor C-terminal segment in inositol 1,4,5-trisphosphate receptor activation: constitutive enhancement of calcium signalling in MCF-7 tumor cells.', *J. Biol. Chem.*, 283(42), pp. 28198–28215.

van Zundert, G. C. P. *et al.* (2016) 'The HADDOCK2.2 Web Server: User-Friendly Integrative Modelling of Biomolecular Complexes.', *J. Mol. Biol.*, 428(4), pp. 720–725.

PART 5.

CONCLUSIONS

5. CONCLUSIONS

In this Doctoral Thesis, we have shown how computational chemistry techniques, especially molecular dynamics (MD) simulations, can provide insight into the understanding of biological and chemical phenomena, and complemented with experimental results.

Overall, molecular dynamics applications in GPCRs have contributed to understand the structural determinants associated to their different allosteric mechanisms (bidirectional transmission of the signal): *activation mechanism*, *allosteric modulators regulation* and *oligomerization* with other GPCRs or additional proteins.

The main conclusions of this thesis are:

1. **Tyrotropin-release hormone receptor (TRHR).** The effect of the mutation I131[ICL2]T in TRHR, that reduce the affinity for the thyrotropin-release hormone (TRH) and cause central hypothyroidism (deficit in thyroid hormones) in homozygotes, is due to a direct disruption of the TRHR-Gq coupling.

2. **M_2 muscarinic acetylcholine receptor (M_2 mAChR).** M_2 mAChR deactivation starts with the rearrangement of the transmission switch, the subsequent opening of the extracellular portion of the receptor, and finally the closure of the intracellular part. The allosteric modulator LY2119620 restricts the conformations of Tyr177[ECL2] and Trp422[7.35]. This modulates the orientation of the Tyr426[7.39] and provide the strong positive cooperativity and the increase in affinity for orthosteric agonist iperoxo.

3. ***Ghrelin receptors (GHS-R1a and GHS-R1b) and sigma-1 receptor (σ_1R).*** The proposed quaternary structure of the complex between GHS-R1a, GHS-R1b and σ_1R and G proteins predicts close contact between the cytosolic domain of σ_1R and the G protein. This provides the framework for the inhibition of ghrelin signaling by cocaine, which cause anorexigenic effect.

PART 6.

LIST OF PUBLICATIONS

6. LIST OF PUBLICATIONS

PUBLICATIONS DIRECTLY DERIVED FROM THE CORE WORK OF THIS THESIS

García, M., González de Buitragro, J., **Jiménez-Rosés, M**., Pardo, L., Hinkle, P.M. and Moreno, J.C. (2017) 'Central Hypothyroidism Due to a TRHR Mutation Causing Impaired Ligand Affinity and Transactivation of Gq.', *J. Clin. Endocrinol. Metab.*, 102(7), pp. 2433–2442.

Jiménez-Rosés, M., Matsoukas, M.T., Caltabiano, G. and Cordomí, A. (2018) 'Ligand-Triggered Structural Changes in the M_2 Muscarinic Acetylcholine Receptor', *J. Chem. Inf. Model.*, 58(5), pp. 1074–1082.

Aguinaga, D., Medrano, M., Cordomí, A., **Jiménez-Rosés, M.**, Angelats, E., Casanovas, M., Vega-Quiroga, I., Canela, E.I., Petrovic, M., Gysling, K., Pardo, L., Franco, R. and Navarro, G. (2018) 'Cocaine blocks effects of hunger hormone, ghrelin, via interaction with neuronal sigma-1 receptors.', *Mol. Neurobiol.*, p. 1-15.
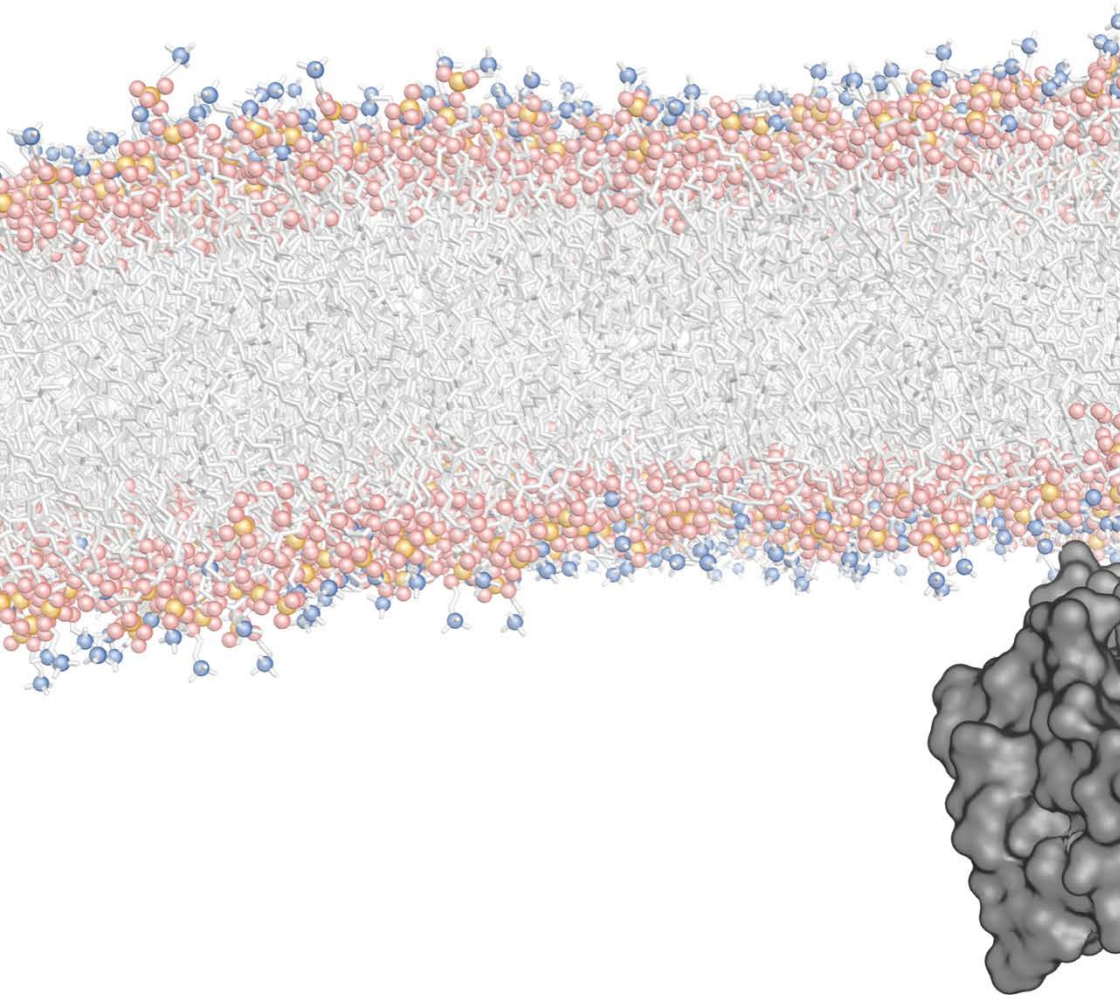
ADDITIONAL PUBLICATIONS

Marsango, S., Caltabiano, G., **Jiménez-Rosés, M.**, Millan, M.J., Pediani, J.D., Ward, R.J. and Milligan G. (2017) 'A Molecular Basis for Selective Antagonist Destabilization of Dopamine D_3 Receptor Quaternary Organization.', *Sci. Rep.*, 7(1), p. 1-17.

Srinivasan, S., Fernández-Sampedro, M.A., Ramon, E., **Jiménez-Rosés, M.**, Cordomí, A. and Garriga, P. (2018) 'Human Blue Cone Opsin Regeneration Involves Secondary Retinal Binding with Analog Specificity.', *Biophys. J.*, 114(6), pp. 1285–1294.

Pérez-Benito, L., Casajuana-Martin, N., **Jiménez-Rosés, M**., van Vlijmen, H. and Tresadern, G. (2019) 'Predicting activity cliffs with free energy perturbation.', *J. Chem. Theory Comput.*, p. 1–12.

Srinivasan, S., Jiménez-Rosés, M., Balke, J., Kim, T.Y., Cordomí, A., Alexiev, U. and Garriga, P. 'Efficient G-protein activation by the visual G protein-coupled receptor rhodopsin in the dark by the release of critical structural constraints.', **Manuscript under preparation**.