

# **Insight into G Protein-Coupled Receptors through Molecular Dynamics Simulations**

---

Name of Student:

Silvia Atanasio

Name of Supervisor:

Prof Christopher Reynolds

Dr Richard Strange

A thesis submitted for the degree of PhD Biological Sciences



Department of School of Life Sciences

University of Essex

October 2020





## Abstract

In recent decades, technological advancement has been increasing extremely rapidly and this allowed to open the way to wide computational approaches able to guide and support the experimental work of biological sciences. Currently, it is possible to model the behaviour of chemical systems and reach the atomistic scale. From this point of view, the ability to understand how a ligand (small molecule or peptide) interacts with its biological target and through which mechanism the allosteric and orthosteric binding sites crosstalk may lead to the development of the high efficacy therapeutic agents that may avoid negative side effects. Among the available computational techniques, molecular dynamics (MD) contributes significantly to the understanding of protein-drug recognition, providing detailed information on structural flexibility and conformational changes, the supervised molecular dynamics (SuMD) allows to understand the (un)binding paths differences, and metadynamics (MetaD) enhances the number of the states explored. These techniques were employed in this PhD thesis to study G protein-coupled receptors (GPCRs), one of the largest families of proteins in the mammalian genome. In particular two receptors have been analysed: free fatty acid receptor 1 (FFAR1) and calcitonin receptor (CTR), respectively belonging to class A and class B of GPCRs. The results highlight that allostereism is a crucial event for the activation mechanism whether exerted by a molecule (such as AP8 for FFAR1) or by an accessory protein (such as RAMP for CTR).

*The total word count of the thesis is: 41405*

*The word count of the thesis, excluding any appendices and references is: 30844*

## **Dedication**

Dedicated to my dear husband and best friend, Marco, for supporting and encouraging me throughout this important experience.

## Acknowledgements

First of all, I would like to express my sincere gratitude to the Peter Nicholls Scholarship, a UK studentship supported by a bequest from the estate of Professor Peter Nicholls, for allowing me the opportunity to commit to this PhD programme. Further, I would like to thank my supervisor, Professor Christopher Reynolds, for his excellent guidance, constructive criticisms, encouragement and recommendations that have enable me to successfully complete this PhD project. I am also thankful to Dr Giuseppe Deganutti for all the considerate advice, without his help this project would have not been the same! To conclude, I cannot forget to thank my family and friends for all their unconditional support in these very intense academic years.

## Table of Contents

<b>Abstract.....</b>	<b>III</b>
<b>Dedication .....</b>	<b>IV</b>
<b>Acknowledgements.....</b>	<b>V</b>
<b>Table of Contents .....</b>	<b>VI</b>
<b>List of Abbreviation .....</b>	<b>XI</b>
<b>Thesis structure .....</b>	<b>XIV</b>
<b>1. Literature review: An overview of G protein-coupled receptors (GPCRs).....</b>	<b>1</b>
1.1 Introduction .....	1
1.2 GPCRs: physiology .....	4
1.3 GPCRs: pathology.....	5
1.4 GPCRs: generic residue numbers.....	8
1.5 GPCRs: solved structures.....	12
1.6 GPCRs: structural architecture and common motifs.....	14
1.7 GPCRs: activation and the role of the G-protein and arrestins .....	18
1.8 Allostery and free fatty acid receptor 1 (FFAR1, formerly GPR40).....	25
1.9 GPCRs: experimental methods .....	29
1.10 Perspectives of my PhD project .....	31
1.11 Androglobin: a complex structure still to discover .....	32
<b>2. Methodology .....</b>	<b>35</b>

2.1 Protein fold recognition servers .....	35
2.2 Comparative protein structure modelling: Modeller .....	37
2.3 Molecular Dynamics (MD) .....	39
2.3.1 Molecular Mechanics: defining a force field and a system for MD simulations .....	39
2.3.2 Supervised molecular dynamics (SuMD) .....	43
2.3.3 Metadynamics (MetaD).....	46
2.3.4 System preparation for MD simulations .....	48
2.3.5 Ligand parametrisation belongs under Molecular Mechanics ...	51
2.3.6 System equilibration and MD settings belongs under Molecular Dynamics.....	52
2.3.7 MD analysis.....	54
2.3.8 General .....	54
<b>3. Allosteric effects in the free fatty acid receptor 1 (FFAR1) .....</b>	<b>55</b>
3.1 Abstract .....	55
3.2 Introduction .....	57
3.3 Materials and Methods .....	60
3.3.1 Sequence analysis.....	60
3.3.2 Molecular Dynamics (MD) simulations.....	62
<i>i) System preparation for MD simulations .....</i>	<i>62</i>
<i>ii) Ligand parametrisation .....</i>	<i>62</i>
3.4 Results and Discussion.....	64

3.4.1 Allosteric binding site A2: does it exist in GPCRs other than FFAR1? .....	64
3.4.2 Analysis of missing H8 in FFAR1 crystal structures .....	74
3.4.3 MD simulations of FFAR1: comparison between apo and holo form .....	81
3.5 Conclusion.....	89
<b>4. Addressing free fatty acid receptor 1 (FFAR1) activation using supervised molecular dynamics* .....</b>	<b>91</b>
4.1 Abstract .....	91
4.2 Introduction .....	92
4.3 Materials and Methods .....	97
4.3.1 System preparation for the MD .....	97
4.3.2 System equilibration and MD Settings.....	98
4.3.3 SuMD - binding.....	99
4.3.4 SuMD - unbinding.....	101
4.3.5 Analysis of the MD trajectories .....	104
4.4 Results and Discussion.....	105
4.4.1 Intermediate interactions between MK-8666 and ECL2 drive the FFAR1:MK-8666 complex formation .....	105
4.4.2 MK-8666 and AP8 are reciprocally stabilised during MD simulations .....	108
4.4.3 MK-8666 modifies the dynamics of TM6.....	111
4.4.4 AP8 influences MK-8666 unbinding from FFAR1.....	115

4.5 Conclusion.....	119
<b>5. Calcitonin receptor (CTR)* .....</b>	<b>123</b>
5.1 Abstract .....	124
5.2 Introduction .....	125
5.3 Materials and Methods .....	131
5.4 Results and Discussion.....	133
5.4.1 Geometric analysis: RMSD and RMSF .....	133
5.4.2 Geometric analysis: Angles, Distances, and Centroids.....	142
5.4.3 Contacts analysis .....	154
5.4.4 GBSA analysis .....	154
5.4.5 Effect of RAMP3 on the sCT:CTR complex .....	157
5.4.6 Effect of different CT on RAMP3:CTR complexes (AMY <sub>3</sub> receptor) .....	159
5.4.7 Effect of RAMP1 (AMY <sub>1</sub> ) or RAMP3 (AMY <sub>3</sub> ) on rAMY:CTR .....	161
5.4.8 Effect of different peptide residues in the key positions .....	162
5.5 Conclusion.....	166
<b>6. Concluding remarks.....</b>	<b>167</b>
6.1 Future research directions .....	167
6.2 Summary of scientific activities .....	169
6.2.1 Articles on peer-reviewed journals .....	169
6.2.2 Awards .....	170
6.2.3 Oral [O] and Poster [P] presentations during conferences .....	171



# PREFACE

---

<b>References .....</b>	<b>173</b>
<b>Appendices .....</b>	<b>191</b>
Appendix A. GPCRs solved structures .....	191
Appendix B. Molecular Dynamics (MD) simulation applied to Androglobin: a novel heme protein .....	199
Appendix C. Supplementary material of Chapter 5 .....	209

## List of Abbreviation

**2YB** – TAK-875

**6XQ** – Compound 1

**7OS** – AP8

**7TM** – Seven transmembrane

**A<sub>1</sub>AR** – Adenosine receptor A1

**Adgb** – Androglobin

**BLASTp** – Protein Basic Local Alignment Search Tool

**C5aR1** – C5a anaphylatoxin chemotactic receptor 1

**CLR** – Calcitonin receptor-like receptor

**CASP** – Critical Assessment of Structure Prediction

**cMD** – Classic Molecular Dynamics

**Cryo-EM** – Cryo-electron microscopy

**CT** – Calcitonin

**hCT** – human CT

**pCT** – porcine CT

**sCT** – salmon CT

**CTR** – Calcitonin receptor

**CVs** – Collective variables

**ECD** – Extracellular domain

**ECL** – Extracellular loop

**FFAR1** – Free fatty acid receptor 1

**GIP** – Gastric inhibitory polypeptide

**GLP-1** – Glucagon-like peptide 1

**GPCRs** – G protein-coupled receptors

**GPR40** – G protein-coupled receptor 40

**GPU** – Graphic Processing Unit

**H** – Helix

**hCGRP** – Human alpha-calcitonin gene-related peptide

**HTMD** – High-Throughput Molecular Dynamics

**ICL** – Intracellular or cytoplasmic loop

**Jnet** – Joint Network

**LCFAs** – Long-chain free fatty acids

**MAC** – Modelled active conformation

**MD** – Molecular Dynamics

**MetaD** – Metadynamics

**MK6** – MK-8666

**ML** – Machine learning

**MSM** – Markov State Modelling

**NAM** – Negative allosteric modulator

**NMR** – Nuclear magnetic resonance spectroscopy

**ns** – Nanosecond

**OPM** – Orientations of Proteins in Membranes

**PAM** – Positive allosteric modulator

**PDB ID** – Protein Data Bank identifier (a 4-character code)

**PME** – Particle Mesh Ewald

**POPC** – 1-palmitoyl-2-oleyl-sn-glycerol-3-phospho-choline

**PSIPRED** – PSI-blast based secondary structure PREDiction

**QM** – Quantum mechanics

**R** – Receptor inactive state

**R\*** – Receptor active state

**RAMP** – Receptor activity-modifying protein

**rAmy** – Rat amylin

**RESP** – Restrained electrostatic potential

**RMSD** – Root Mean Square Deviation

**RMSF** – Root Mean Square Fluctuation

**SuMD** – Supervised Molecular Dynamics

**T2DM** – Type 2 diabetes mellitus

**TCL** – Tool Command Language

**TM** – Transmembrane

**TMD** – Transmembrane domain

**VDW** – Van der Waals

**VMD** – Visual Molecular Dynamics

**WHO** – World Health Organization

**β<sub>2</sub>-AR** – Beta-2 adrenergic receptor

## Thesis structure

### Chapter 1: Literature review

An overview of G protein-coupled receptors (GPCRs) is presented to provide general background information on this superfamily of protein and introduce the purpose of the studies presented in this thesis.

### Chapter 2: Methodology

General principles on computational methodology are described in this section to provide an explanation for driving experimental designs presented in subsequent sections. Contrariwise, more detailed experimental procedures are given in each of the subsequent sections.

### Chapter 3: Allosteric effects in the free fatty acid receptor 1 (FFAR1)

A first study of a class A GPCR named free fatty acid receptor 1, is reported in this section. The aim of this research is to analyse the FFAR1 allosteric binding site and its effects respectively by means of static structural examination and molecular dynamics (MD) simulations, and to examine the conserved motifs with particular attention given to residues of helix 8 (H8).

## **Chapter 4: Addressing free fatty acid receptor 1 (FFAR1) activation using supervised molecular dynamics**

A second study of free fatty acid receptor 1 (FFAR1) is reported in this section. The previous analysis presented in Chapter 3 is here deepened with supervised molecular dynamics (SuMD) to deliver insights about the agonist (un)binding mechanism and the allosteric communications between the two experimentally determined FFAR1 binding sites.

## **Chapter 5: Calcitonin receptor (CTR)**

A study of a class B GPCR named calcitonin receptor (CTR) is reported in this section. This research project focuses on MD simulations of CTR in complex with three major distinct classes of the peptide hormone calcitonin (CT), such as salmon CT (sCT), human CT (hCT), and porcine CT (pCT), as well as rat amylin (rAmy) or human  $\alpha$ -CGRP (hCGRP) whose potency is enhanced by receptor activity-modifying protein (RAMP) interaction.

## **Chapter 6: Concluding remarks**

Limitations and recommendations for future directions are presented in this section. A summary of scientific activities is also reported here.



# 1. Literature review: An overview of G protein-coupled receptors (GPCRs)

## 1.1 Introduction

The superfamily of G protein-coupled receptors (GPCRs) is one of the largest families of proteins in the mammalian genome, which includes more than 800 members, sharing a common configuration of seven transmembrane  $\alpha$ -helices (Fredriksson *et al.*, 2003; Zhang *et al.*, 2015). These receptors are the most physiologically important membrane proteins because they recognise a multiplicity of ligands such as hormones, neurotransmitters and environmental stimuli, and are drug targets in the treatment of numerous diseases. Therefore, the study of GPCRs can have a very important impact on public health.

Two classification systems have been used to group this superfamily of proteins. The first system is designed to cover all GPCRs, in both vertebrate and invertebrate, and uses classes (or clans) A, B, C, D, E, and F, and subclasses are defined by roman number nomenclature (Attwood and Findlay, 1994; Kolakowski, 1994). The second system categorises most of human GPCRs into five main families by the GRAFS classification system (Figure 1.1): the Glutamate family (corresponded to class C) (15 members), the Rhodopsin family (corresponded to class A) (701 members), the Adhesion family (24 members), the Frizzled/taste family (corresponded to class F) (24 members), and the Secretin family (corresponded to class B, as does the Adhesion family) (15 members) (Fredriksson *et al.*, 2003).

The year 2012 was significant for the award of the Nobel Prize in Chemistry to Robert Lefkowitz and Brian Kobilka for ground-breaking discoveries that contributed

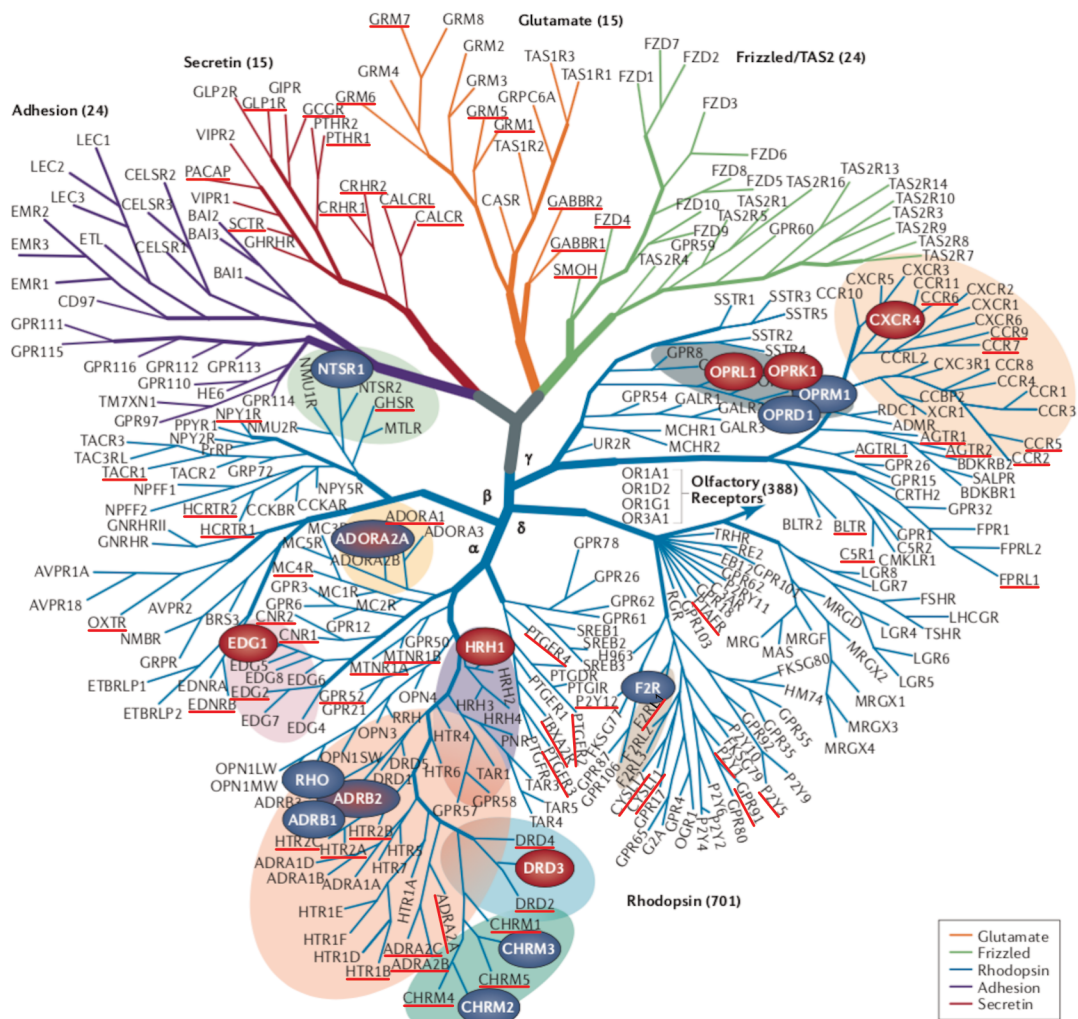


# CHAPTER 1

---

to our current understanding of the structure and the functioning of GPCRs, showing the scientific community that the whole family of GPCRs resemble each other and generally work in the same way. However, acquiring this knowledge has been challenging because these receptors have an innate conformational flexibility, which is essential for their activity under physiological conditions but, at the same time, complicates obtaining diffraction-quality crystals (Kobilka and Deupi, 2007). Therefore, the structure determination using traditional methods is very difficult and computational methods are required to address this problem.

In this chapter, the current knowledge of GPCRs is discussed; in particular, an overview of the solved structures and structural architecture is provided.

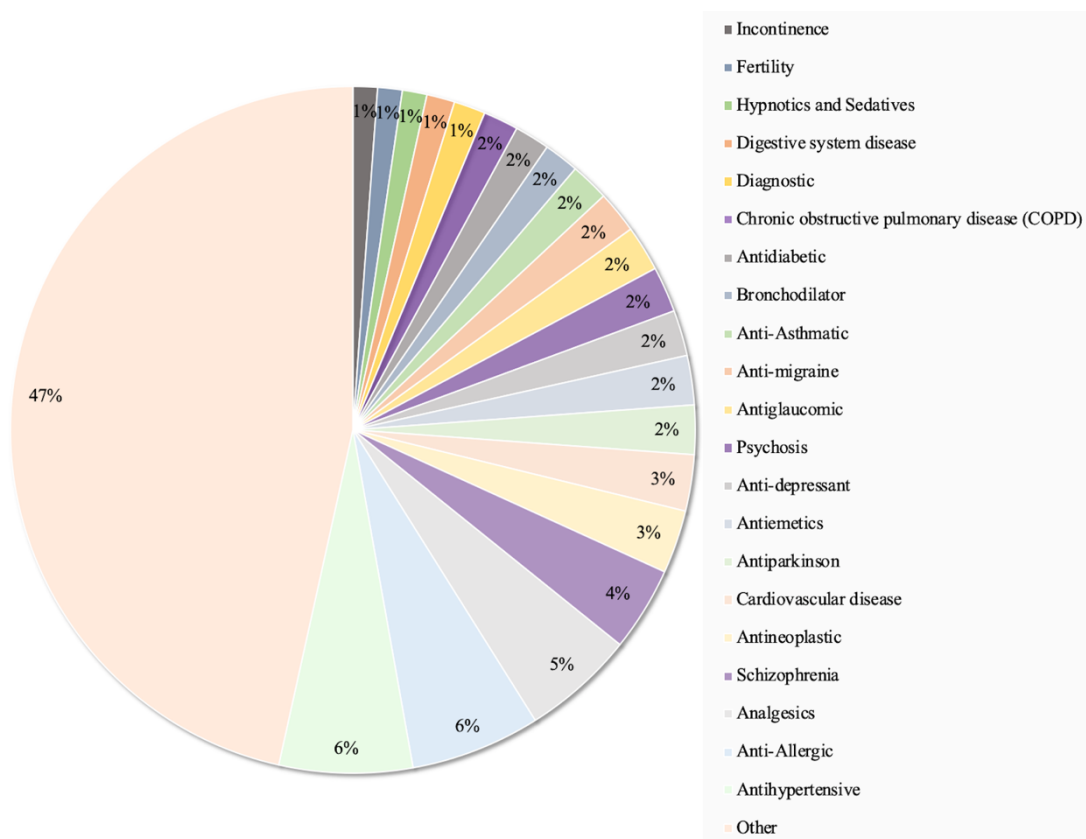


## 1.2 GPCRs: physiology

GPCRs have a central role in many physiological functions such as neurotransmission, metabolism regulation, cellular differentiation and growth, secretion, inflammatory and immune response, as well as vision, taste and smell (Li *et al.*, 2010; Filizola, 2014). GPCRs transmit their signal from extracellular binding of diverse ligands (*e.g.* ions, organic odorants, amines, peptides, proteins, lipids, nucleotides, photons, *etc.*) to modulate intracellular signalling molecules by the activation of intracellular proteins such as heterotrimeric G-proteins, arrestins, and kinases (Fredriksson *et al.*, 2003). The extracellular first messenger can be either endogenous or exogenous substance: the first category includes a large variety of molecules secreted by the organism, such as hormones and neurotransmitters, contrariwise, the second category comprehends environmental substances that are detected by the organism for example molecules responsible for odour or taste. Moreover, some GPCRs are stimulated by physical messengers, *i.e.* light photons (Filizola, 2014). The activation of the receptor by the extracellular binding results in a conformational change of the receptor itself, such that it couples with direct intracellular partners, namely heterotrimeric G-proteins, arrestins, and kinases, to transduce the stimulus into biochemical signals involving complex pathways (Pierce *et al.*, 2002).

### 1.3 GPCRs: pathology

GPCRs are involved in many diseases, such as central nervous system disorders (*i.e.* Alzheimer's disease and schizophrenia), inflammatory diseases, metabolic disorders (*i.e.* diabetes), cardiac diseases, monogenic diseases, cancer, and many other implications (Thomsen *et al.*, 2005) (Figure 1.2). Currently, ~34% of the drugs available on the market (Table 1.1) have GPCRs as their target and, because these receptors have druggable sites that are accessible at the cell surface, understanding the structure and dynamics of these proteins is a key factor for drug design (Thomsen *et al.*, 2005; Bakir and Sezerman, 2006; Garland, 2013). Moreover, a potential therapeutic advantage, such as increased spatial and temporal selectivity, can be achieved by modulation of GPCRs via allosteric sites which can alter the structure, dynamics, and function of the receptor (Christopoulos *et al.*, 2014) (discussed in section 1.8). Therefore, the comprehension of the structure of this superfamily of proteins in relation to not only orthosteric but also allosteric ligands allows for generating new more effective and selective drugs.



**Figure 1.2 | Disease indications associated with GPCRs.** The pie chart indicates the percentage of approved and distinct drugs targeting GPCRs (adapted figure from Isberg *et al.*, 2015) (<https://gpcrdb.org/drugs/drugstatistics>).

**Table 1.1 | New drugs for GPCRs approved by the FDA since 2014**

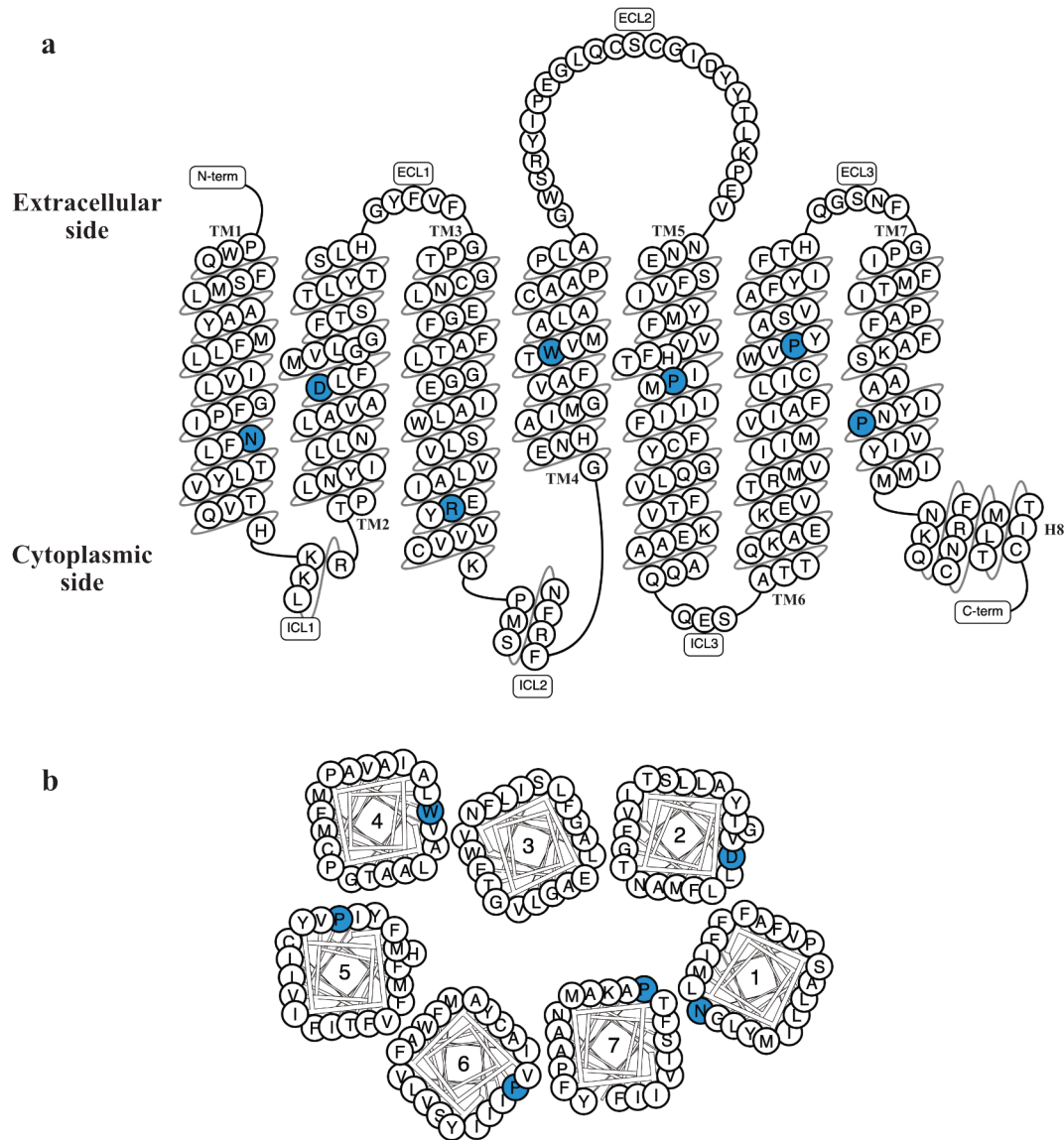
Substance	Brand name	Indications	Targets*	Year of approval
Albiglutide	Tanzeum	Antidiabetic	GLP-1R	2014
Alverine	Audmonal	Antispasmodics	5-HT1A	2014
Droxidopa	Northera	Orthostatic hypotension	$\beta_1$ -AR, $\beta_2$ -AR, $\beta_3$ -AR, $A_{2A}$ AR, $A_{2B}$ AR, $A_{2C}$ AR, and $A_{1A}$ AR, $A_{1B}$ AR, $A_{1D}$ AR	2014
Dulaglutide	Trulicity	Type 2 diabetes	GLP-1R	2014
Naloxegol	Movantik	Opioid-induced constipation	$\mu$ -OR	2014
Netupitant	Akynzeo	Nausea and/or vomiting	NK-1R	2014
Olodaterol	Striverdi respimat	Chronic obstructive pulmonary disease	$\beta_2$ -AR	2014
Suvorexant	Belsomra	Insomnia	Ox1R and Ox2R	2014
Tasimelteon	Hetlioz	Non-24-hour sleep-wake disorder	Mel-1A-R and Mel-1B-R	2014
Vorapaxar	Zontivity	Cardiovascular risk reduction	PAR1	2014
Aripiprazole lauroxil	Aristada	Schizophrenia	5-HT1A, 5-HT2A, and D2R	2015
Asfotase alfa	Strensiq	Hypophosphatasia	S1P <sub>1</sub>	2015
Brexipiprazole	Rexulti	Depression	5-HT1A, 5-HT2A, and D2R	2015
Cangrelor	Kengreal	Percutaneous coronary intervention	P2Y <sub>12</sub> R	2015
Cariprazine	Vraylar	Schizophrenia and bipolar disorder	D2R and D3R	2015
Eluxadoline	Viberzi	Irritable bowel syndrome	$\delta$ -OR and $\mu$ -OR	2015
Flibanserin	Addyi	Sexual dysfunction	5-HT1A and 5-HT2A	2015
Levodopa	Bidopal	Antiparkinson	D1R, D2R, D3R, D4R, and D5R	2015
Parathyroid hormone	Natpara	Anti-osteoporotic, Hypoparathyroidism, and Control hypocalcemia	PTH-1R and PTH-2R	2015
Rolapitant	Varubi	Nausea and/or vomiting	NK-1R	2015
Selexipag	Uptravi	Pulmonary hypertension	PI2R	2015
Sonidegib	Odomzo	Basal cell carcinoma	SMO	2015
Brexipiprazole	Rexulti	Anti-depressant	5-HT7R	2015
Buprenorphine	Subutex	Anti-depressant, Analgesics, and opioid addiction	k-OR and $\mu$ -OR	2016
Formoterol	Foradile	Chronic obstructive pulmonary disease (COPD)	$\beta_2$ -AR	2016
Lixisenatide	Adlyxin	Type 2 diabetes	GLP-1R	2016
Naltrexone	Nalorex	Addiction	$\delta$ -OR, k-OR, and $\mu$ -OR	2016
Oxycodone	Oxanest	Analgesics	$\delta$ -OR, k-OR, and $\mu$ -OR	2016
Pimavanserin	Nuplazid	Antiparkinson	5-HT2A	2016
Rupatadine	Rupafin	Anti-allergic	H1R and PAF-R	2016
Abaloparatide	Tymlos	Anti-osteoporotic	PTH-1R	2017
Etelcalcetide	Parsabiv	Parathyroid disorders	CASR	2017
Naldemedine	Symproic	Opioid-induced constipation	$\mu$ -OR	2017

(\*) Listed using the short name of the protein in UniProt (Adapted Table from Hauser *et al.*, 2017)

## 1.4 GPCRs: generic residue numbers

The peculiarity of GPCRs is based on sharing a common structural core of seven transmembrane (7TM) helices that readily facilitates the sequence alignment (by traditional methods or by structural alignment), to identify the corresponding residues, which are categorised with a generic residue number (Isberg *et al.*, 2015). There are different systems of GPCR residue numbering based on the type of class: the Ballesteros-Weinstein numbering scheme for class A (Ballesteros and Weinstein, 1995), the Wootten numbering scheme for class B (Wootten *et al.*, 2013), the Pin numbering scheme for class C (Pin, Galvez and Prèzeau, 2003), and lastly the Wang numbering scheme for class F (Wang *et al.*, 2014). These systems consist of two numbers separated by a dot, where the first number indexes the TM helix (1-7), while the second number denotes the residue position relative to the most conserved position, which is arbitrarily assigned the number 50. For example, the most conserved amino acid in TM7 is a proline whose identifier would be 7.50, *i.e.*, P<sup>7.50</sup>. A tyrosine residue located three amino acids after P<sup>7.50</sup> will be Y<sup>7.53</sup>. Moreover, the generic residue number can be related to the amino acid sequence of a particular GPCR protein, therefore, each identifier defined above can be followed by the numbering in the particular sequence as follows: P<sup>7.50(273)</sup>. In order to identify mutations, the wild-type identifier is followed by the mutant amino acid. For example, free fatty acid receptor 1 (FFAR1) shows a threonine instead of a tyrosine as residue 7.53 and this mutation is written in the following way: Y<sup>7.53(276)</sup>T. Figure 1.3 illustrates the residue relative to the most conserved position for each TM of GPCRs class A: N<sup>1.50</sup>: 98%, D<sup>2.50</sup>: 90%, R<sup>3.50</sup>: 95%, W<sup>4.50</sup>: 97%, P<sup>5.50</sup>: 78%, P<sup>6.50</sup>: 99%, P<sup>7.50</sup>: 88% (Isberg *et al.*, 2016). Furthermore, cross-class GPCR residue numbering is available in GPCRdb (<https://www.gpcrdb.org/>) and

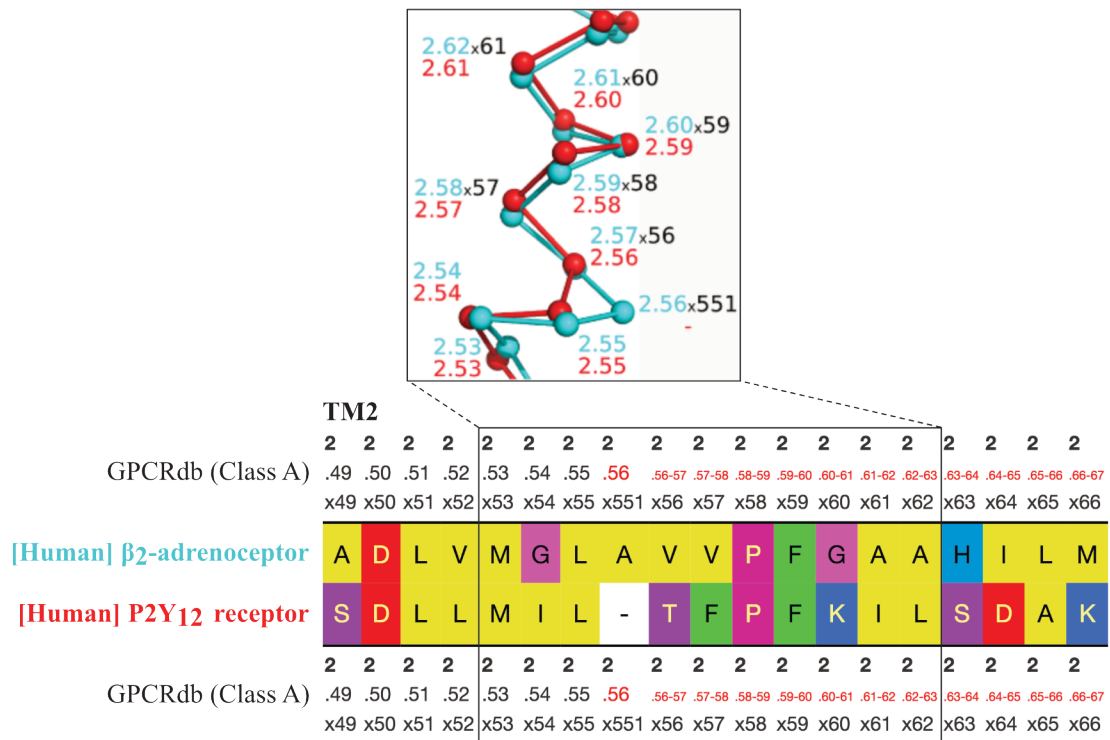
it is based on the Ballesteros-Weinstein numbering scheme specifying a class-specific number, for example Y<sup>7.53a.57b</sup> (Isberg *et al.*, 2015).



**Figure 1.3 | A two-dimensional model (a) and helix box plot (b) of Rhodopsin (class A).** The diagrams show the receptor residue topologies as seen from the side and top, respectively, and the key residues relative to the most conserved position are shown in blue circles (adapted figure from Isberg *et al.*, 2015).



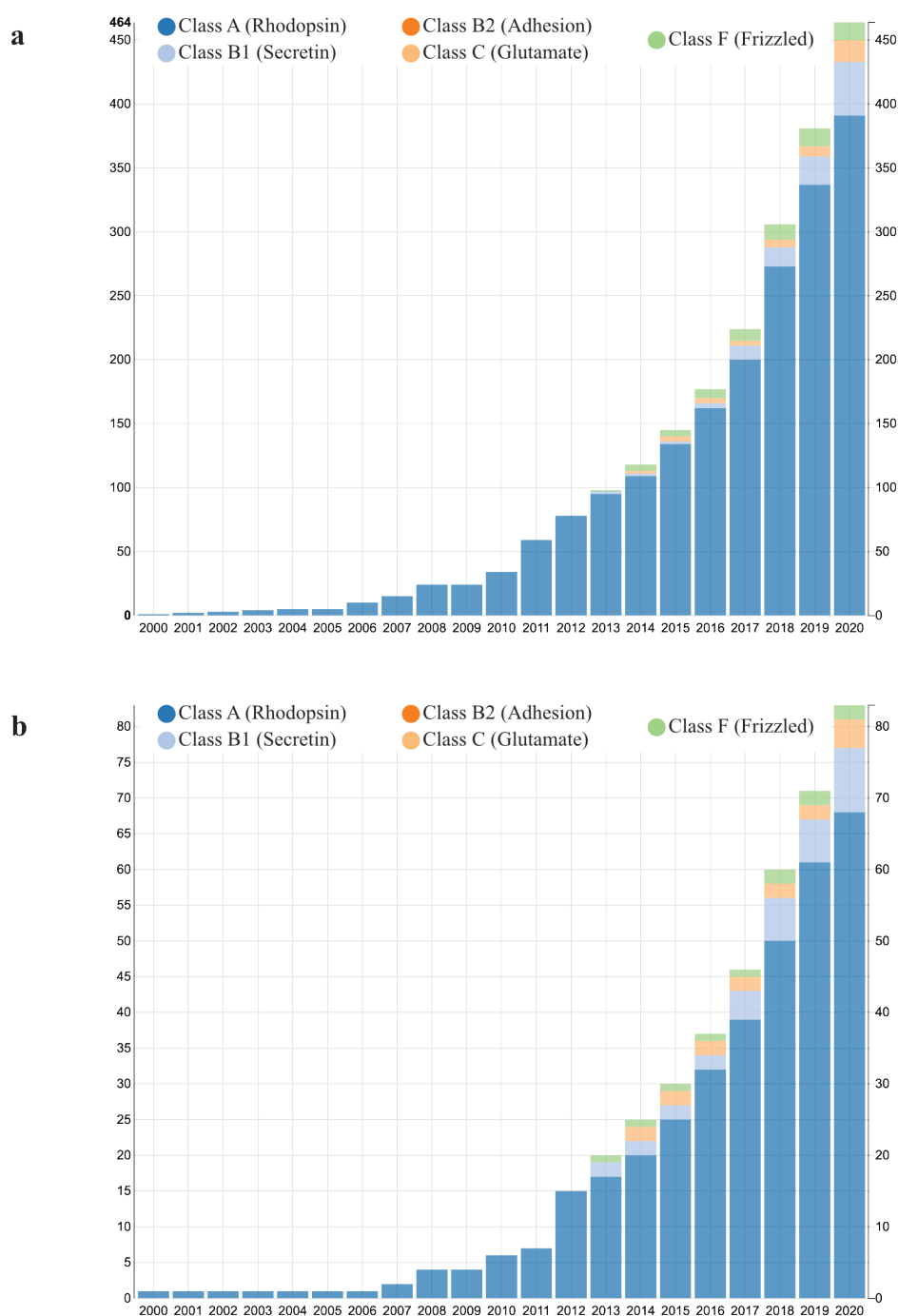
Lastly, a structure-based GPCRdb numbering scheme was developed based on crystal structures that corrects for structural distortions, helical bulges, and constrictions. The advantage of this scheme is that the residues aligned in sequence are those that align in structure (Isberg *et al.*, 2016). The GPCRdb numbering scheme consists of two numbers as above-mentioned systems, but these are separated by a unique separator ‘x’ instead of a dot (*e.g.* 2x56) and bulge residue is assigned the same number as the preceding residue followed by a 1 (*e.g.* for a bulge residue following position 55 is assigned the number 551). Furthermore, structure-based GPCRdb numbering scheme can be used together with one of the sequence-based schemes (*e.g.* 2.57x56) as shown in Figure 1.4 (which means that residue 2.57 is in the position normally occupied by residue 2.56, as shown in Figure 1.4).



**Figure 1.4 | Structure-based alignment between human  $\beta_2$ -adrenergic receptor (cyan) and human P2Y<sub>12</sub> receptor (red), both members of class A.** The amino acid A<sup>2.56(85)</sup> in  $\beta_2$ -adrenergic receptor is an extra residue that creates a bulge in TM2 and offsets in the sequence-based generic numbers when compared to the P2Y<sub>12</sub> receptor that lacks the bulge. This bulge residue is identified with the following numbering: 2.56x551 (adapted figure from Isberg *et al.*, 2015).

## 1.5 GPCRs: solved structures

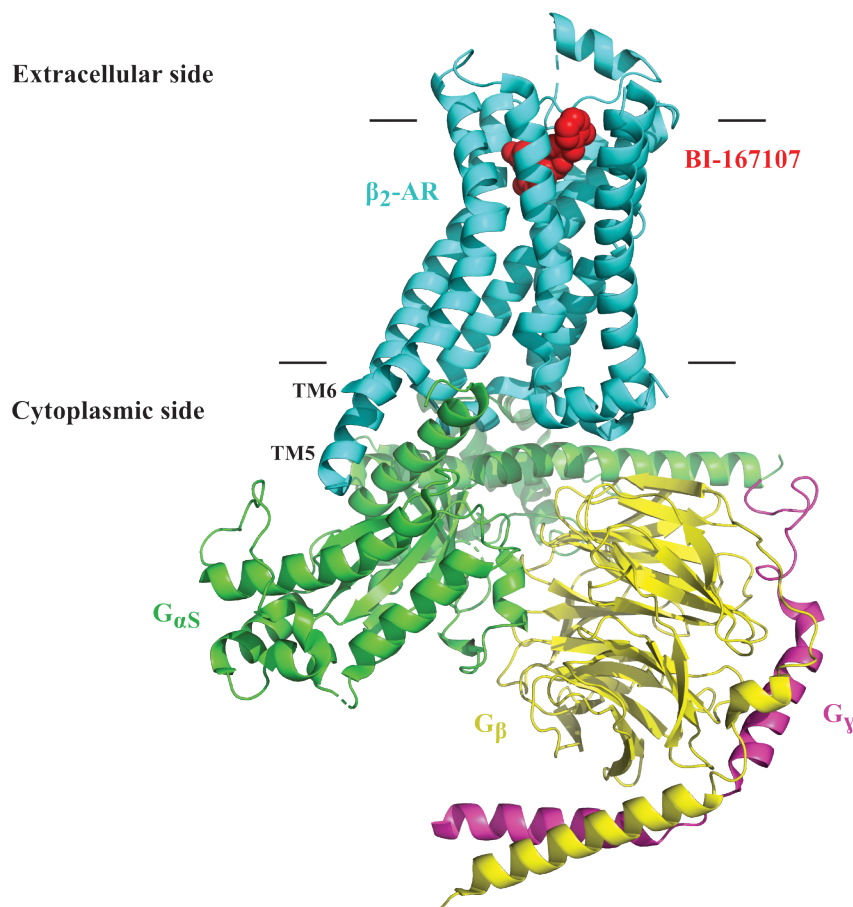
In recent years, an exponential growth in the number of GPCR crystal structures has been observed (Figure 1.5). Currently (August 2020), there are **464** solved structures of GPCRs (solved principally by X-ray diffraction and Cryo-electron microscopy) as follows (Figure 1.5a): 391 structures of the rhodopsin class (class A), 42 structures of secretin family (class B), 17 structures of glutamate family (class C), and 14 structures of frizzled family (class F), while there are not any solved structures for adhesion and taste family. Only **83** of these structures are unique crystallised receptors and the number for each class is respectively: 68, 9, 4, and 2 (Figure 1.5b) (Isberg *et al.*, 2016). A list of the currently available GPCRs solved structures is presented in the Appendix, section A.



**Figure 1.5 | GPCRs solved structures.** A plot of the number of GPCR crystal structures (a) and unique crystal structures (b) available each year updated until 26/08/2020 (<https://gpcrdb.org/structure/statistics>).

## **1.6 GPCRs: structural architecture and common motifs**

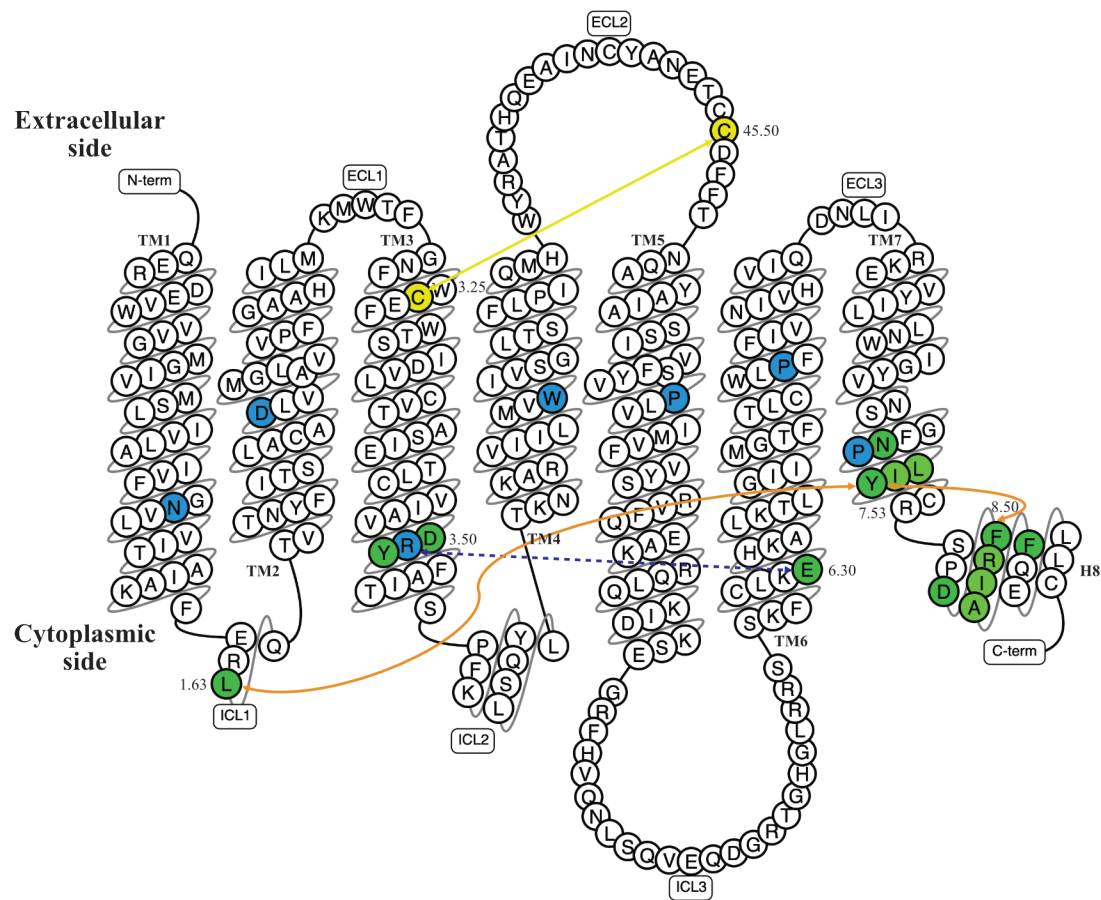
As mentioned previously, GPCRs share a common structural core of 7TM helices (Figure 1.3). These helices are connected by six loops, whereof three intracellular or cytoplasmic loops (ICL1-3) and three extracellular loops (ECL1-3). The N-terminus, located on the extracellular side, and ECLs are responsible for recognising and binding of diverse ligands (discussed in section 1.2). The extracellular signal is transduced through conformational changes to the intracellular regions where is located the C-terminus and, successively, these parts interact with cytosolic G-proteins and other downstream signalling effectors (Figure 1.6).



**Figure 1.6 | Active state crystal structure of  $\beta_2$ -adrenergic receptor (cyan) in complex with nucleotide-free  $G_s$  heterotrimer composed of  $G_{\alpha S}$  (green),  $G_{\beta}$  (yellow), and  $G_{\gamma}$  (purple) (PDB ID: 3SN6). The binding with the agonist (red spheres) induces conformational changes in the receptor that include a 14 Å outward movement at the cytoplasmic end of TM6 and an  $\alpha$ -helical extension of the cytoplasmic end of TM5 (Rasmussen *et al.*, 2011).**

Among the GPCRs, there are some common motifs that are revealed by the solved GPCR structures. Figure 1.7 is an example of  $\beta_2$ -adrenergic receptor where these motifs are shown. One of the most conserved motifs is the D/ERY<sup>3.51</sup> motif in TM3, which often makes a salt bridge with D/E<sup>6.30</sup> in TM6. This interaction between charged acidic (E<sup>6.30</sup>) and basic (R<sup>3.50</sup>) groups forms an “ionic lock” characteristic of the inactive conformation of GPCRs because it blocks the G-protein binding at the cytoplasmic region. Analogously, the receptor is locked in an inactive state by another important conserved motif: the NPxxY<sup>7.53</sup> motif in TM7, whose tyrosine (Y<sup>7.53a.57b</sup>) forms a hydrophobic cluster with leucine (L<sup>1.63</sup>) of the KKLH<sup>1.64</sup> motif in ICL1 and the phenylalanine (F<sup>8.50</sup>) of EFxxxL<sup>8.54</sup> motif in helix 8 (H8) (Vohra *et al.*, 2013), which consists of a 3-4 turn  $\alpha$ -helix contained in the C-terminus that runs parallel to the membrane.

In addition to the above-mentioned common motifs, some similarities are shown also in the ECLs. A highly conserved cysteine in TM3 (C<sup>3.25a.29b.29c.25f</sup>) forms a disulfide bond with a cysteine residue in ECL2 (C<sup>45.50</sup>) that plays an important role in stabilising the conformation of the ECL2 and shapes the entrance to the ligand-binding pocket (Zhang *et al.*, 2015).



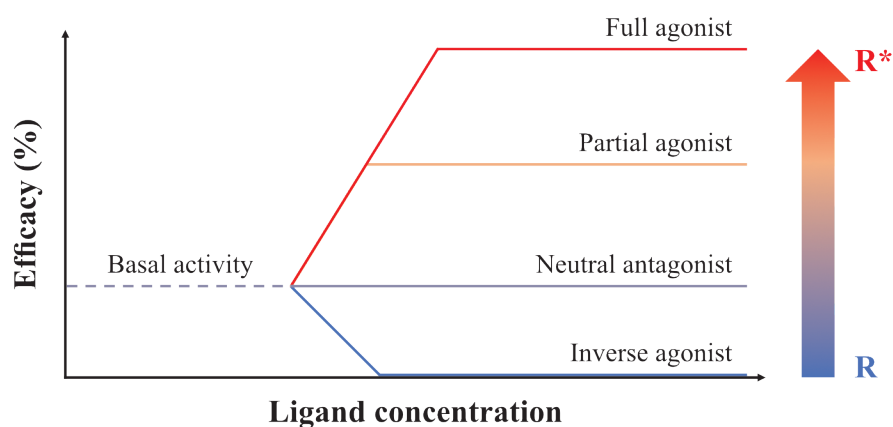
**Figure 1.7 | A two-dimensional model of  $\beta_2$ -adrenergic receptor (class A).** The diagram shows the key residues relative to the most conserved position in blue circles, the conserved motifs in green circles, and highly conserved cysteine in yellow circles. The following interactions are shown with arrows: disulfide bond between C<sup>3.25</sup> (TM3) and C<sup>45.50</sup> (ECL2) in yellow, the salt bridge between R<sup>3.50</sup> (TM3) and E<sup>6.30</sup> (TM6) in blue, and hydrophobic cluster between L<sup>1.63</sup> (ICL1), Y<sup>7.53</sup> (TM7), and F<sup>8.50</sup> (H8) in orange (adapted figure from Isberg *et al.*, 2015).



## 1.7 GPCRs: activation and the role of the G-protein and arrestins

The activation of GPCRs induced by ligand-binding, moves the equilibrium between the two conformations of GPCR, from the inactive state (R) to active state (R\*). At the base of the efficacy of ligands, reflected in their ability to alter the equilibrium between these two states, it is possible to define (Kobilka and Deupi, 2007) (Figure 1.8):

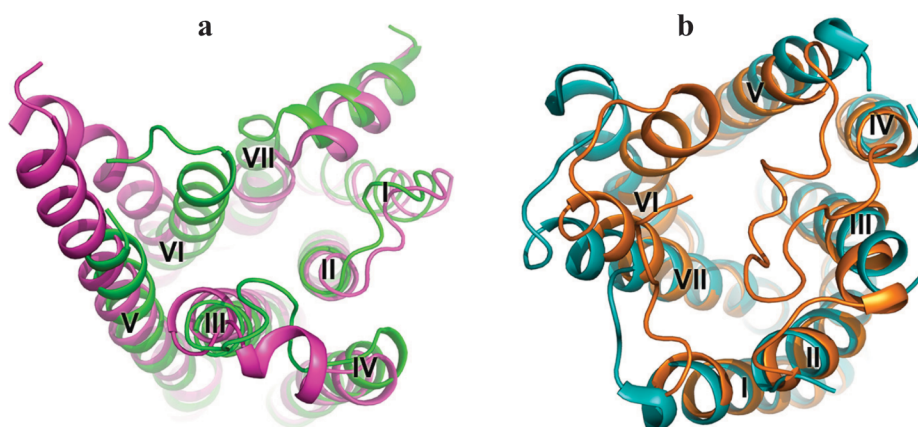
- *Full agonist*: ligand that binds to the receptor stabilising the active conformation and thus shifting the equilibrium to R\*;
- *Partial agonist*: ligand that has some affinity for both the conformations, therefore shifts the equilibrium towards R\* with less effect;
- *Inverse agonist*: ligand that stabilises the inactive conformation and thus shifting the equilibrium to R;
- *Neutral antagonist*: sits in the binding site and prevents the agonist from binding but does not shift the equilibrium.



**Figure 1.8 | Ligand-binding and activation state in GPCRs.** Full agonist (red line) and partial agonist (orange line) bind to the receptor leading to the active state R\*

enhancing the efficacy and promote signal transduction (respectively the first more than the second). While the inverse agonist (blue line) involves a reduction in basal activity (grey dashed line) stabilising the inactive conformation and thus shifting the equilibrium to R. The neutral antagonist (grey line) sits in the binding site and prevents the agonist from binding without altering  $R \leftrightarrow R^*$  equilibrium or directly modulate signalling.

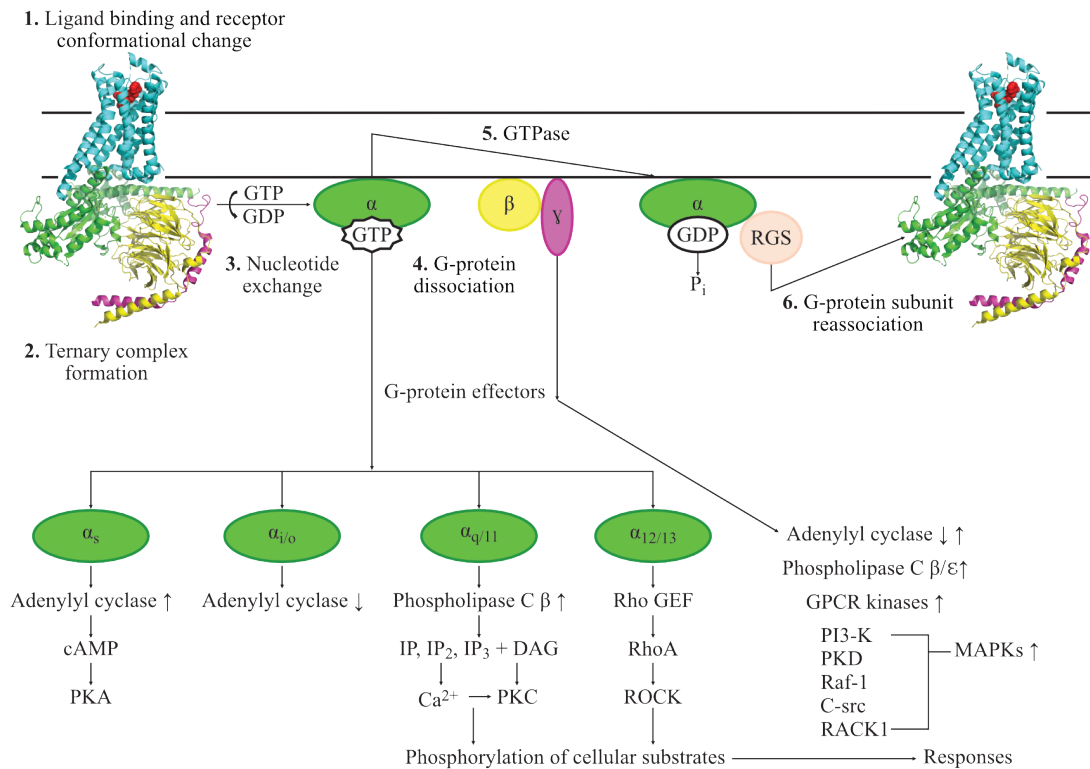
The activation mechanism is similar between GPCRs in spite of the considerable diversity. Some conserved motifs (mentioned in section 1.6) are defined as molecular micro-switches. A significant rearrangement of the 7TM domain is the key event of the receptor activation. Following, two examples of activation are reported. The first example is about  $\beta_2$ -AR; the activation of this receptor induces helical rearrangements including a small outward movement of TM5, a large outward movement of TM6, and an inward motion of TM7. This conformational change leads to an opening of the helical bundle at the cytoplasmic side and, therefore, facilitates the G-protein coupling and subsequent activation (Figure 1.9a) (Zhang *et al.*, 2015). The second example is about P2Y<sub>12</sub>R; in this case, the activation induces conformational changes in the extracellular part, where the extracellular tips of TM6 and TM7 in the active structure are shifted 5-10 Å towards the central axis of the 7TM domain (Figure 1.9b) (Zhang *et al.*, 2015). Therefore, in the receptor activation, both intracellular and extracellular parts have an important role.



**Figure 1.9 | Cartoon representation of the active/inactive structure of  $\beta_2$ -AR and P2Y<sub>12</sub>R.** The illustration shows on the left (a) the intracellular side of R\* (magenta) and R structure (green) of  $\beta_2$ -AR, while on the right (b) shows the extracellular side of R\* (orange) and the R (cyan) of P2Y<sub>12</sub>R (adapted figure from Zhang, Zhao and Wu, 2015).

As mentioned above, ligand binding to GPCRs induces conformational rearrangements that promote the coupling and the activation of G-protein and, therefore, signal transduction pathways and subsequent cellular responses are initiated. G-proteins are heterotrimeric proteins composed of three subunits, namely  $\alpha$ ,  $\beta$ , and  $\gamma$ . Their switching function depends on the ability of the G protein  $\alpha$ -subunit to cycle between an inactive GDP-bound conformation (in complex with  $\beta\gamma$ -subunits) and an active GTP-bound conformation (separated from  $\beta\gamma$  complex) (Figure 1.10) (Thomsen *et al.*, 2005; Oldham and Hamm, 2008). At the base of structural and functional similarity of the  $\alpha$  subunit, G-proteins are grouped into subfamilies, *i.e.* G<sub>s</sub>, G<sub>i/o</sub>, G<sub>q/11</sub>, and G<sub>12/13</sub> (Thomsen *et al.*, 2005). As shown in Figure 1.10, each type of G-protein acts with specific secondary messengers in the signalling pathways, thus determining the specificity of the signal. Moreover, GPCR signalling can be regulated

by desensitization and internalization, a process involving receptor phosphorylation (Thomsen *et al.*, 2005).



**Figure 1.10 | Activation/deactivation cycle of G-protein and GPCR signalling pathways.** The interaction between an agonist ligand and GPCR (1) induces conformational changes promoting its coupling with G-protein. The interaction with G-protein ternary complex with agonist-receptor complex (2) facilitates the nucleotide exchange of GTP for GDP in the  $\alpha$ -subunit (3). The  $\alpha$ -subunit activated with GTP dissociates from the  $\beta\gamma$  complex (4). Different types of  $\alpha$ -subunit of G-protein regulate the selective signal transduction pathways. The hydrolysis of GTP to GDP (5) by the intrinsic GTPase activity of the  $\alpha$ -subunit terminates the activity of the GTP-bound  $\alpha$ -subunit and the re-association of the GDP-bound  $\alpha$ -subunit with the  $\beta\gamma$  complex completes the cycle (6). G-protein signalling (RGS) proteins are accessory proteins that modulate the kinetics of the G-protein activation/deactivation cycle. [Abbreviation used: protein kinase A (PKA), protein kinase C (PKC), inositol phosphates (IP, IP<sub>2</sub>, and IP<sub>3</sub>), diacylglycerol (DAG)] (adapted figure from Thomsen, Frazer and Unett, 2005).

# CHAPTER 1

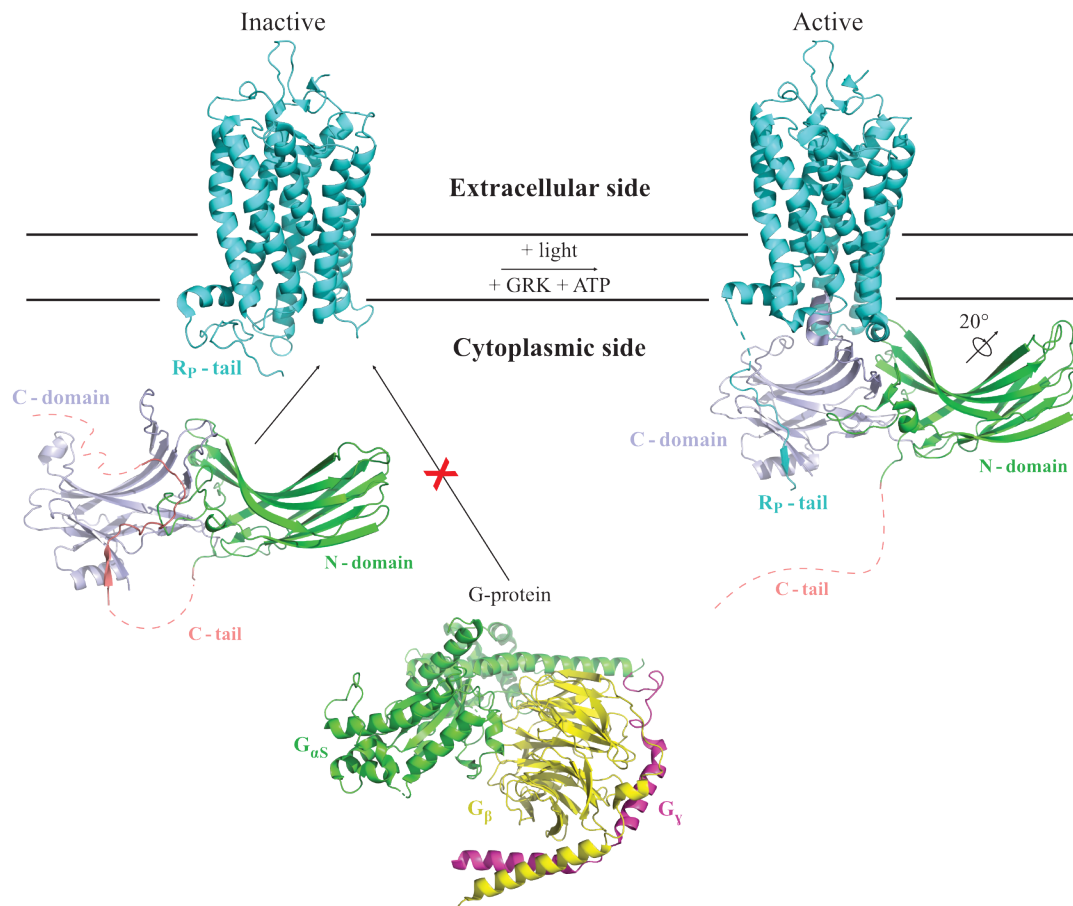
An important role is played by another family of protein *i.e.* arrestins, which block (arrest) the binding between the receptor and the G-protein by steric exclusion, as they interact with the cytoplasmic side of active phosphorylated GPCRs. In comparison with the multitude of GPCRs, the mammalian arrestin family includes only four members (Gurevich and Gurevich, 2006) such as arrestin-1, arrestin-2, arrestin-3, and arrestin-4 (Table 1.2). Arrestin-1 and -4 are defined as visual arrestin because of their selectively expression in the retina, while arrestin-2 and -3 are ubiquitously expressed.

**Table 1.2 | Arrestins and their functions**

Name	Alternative name	Expression	Known binding partners	Known functions
Arrestin-1	S-antigen, 48 kDa protein, visual arrestin, rod arrestin	Rod and cone photoreceptors	Rhodopsin; cone opsins; JNK3; Mdm2; NSF; AP2; ERK2; microtubules; parkin; calmodulin	Blocks rhodopsin cone opsin coupling to G proteins; mobilises Mdm2 to microtubules; supports synaptic function(s)
Arrestin-2	$\beta$ -arrestin, $\beta$ -arrestin1	Virtually every cell	Numerous non-visual GPCRs; trafficking proteins (clathrin, AP2, NSF); MAP kinases (ASK1, MKK4/7, JNK1/2/3, c-Raf1, MEK1, ERK1/2; p38); ubiquitin ligases (Mdm2, AIP4, parkin); calmodulin and many other signalling proteins	Blocks GPCR coupling to G proteins; promotes GPCR internalisation; scaffolds MAP kinase cascades; recruits ubiquitin ligases and deubiquitination enzymes to GPCRs; recruits Src family kinases to GPCRs; recruits Mdm2 and ERK2 to microtubules; scaffolds many other signalling cascades and localises them to GPCRs or other cellular compartments
Arrestin-3	$\beta$ -arrestin2, hTHY-ARRX	Virtually every cell	Numerous non-visual GPCRs; trafficking proteins (clathrin, AP2, NSF); MAP kinases (ASK1, MKK4/7, JNK1/2/3, c-Raf1, MEK1, ERK1/2; p38); ubiquitin ligases (Mdm2, AIP4, parkin); calmodulin and many other signalling proteins	Blocks GPCR coupling to G proteins; promotes GPCR internalisation; scaffolds MAP kinase cascades; recruits ubiquitin ligases and deubiquitination enzymes to GPCRs; recruits Src family kinases to GPCRs; recruits Mdm2 and ERK2 to microtubules; scaffolds many other signalling cascades and localises them to GPCRs or other cellular compartments
Arrestin-4	X- arrestin, cone arrestin	Cone photoreceptors	Cone opsins; JNK3; Mdm2; microtubules; parkin; calmodulin	Blocks cone opsin coupling to G proteins

(Adapted Table from Gurevich and Gurevich, 2014)

As shown in Table 1.2, arrestins have many biological functions due to their high flexibility that allows them to assume several conformations. Arrestins do not only block the interaction between GPCRs and the G-protein but also promote receptor internalisation, as well as, mediate several G-protein-independent signalling pathways (Latorraca *et al.*, 2018). Currently, there are only seven GPCR crystal structures binding arrestin (PDB ID: 4ZWJ, 5DGY, 5W0P, 6TKO, 6U1N, 6PWC, and 6UP7). Figure 1.11 shows the binding between the rhodopsin receptor and arrestin-1. The interaction mainly involves two distinct interfaces: (i) the receptor R<sub>P</sub>-tail binding within a positively charged trench in the N-domain of the arrestin-1 and (ii) the receptor core (TM helices and loops of the receptor) binding between the N-domain and C-domain (the central crest of the arrestin-1). This binding involves the displacement of entire C-tail, the binding of R<sub>P</sub>-tail in the lateral cleft of the C-domain, and the twist of the N-interdomain angle of 20° (Latorraca *et al.*, 2018). It is interesting that the active conformation of arrestin-1 is individually stabilised by either the receptor core or/and the R<sub>P</sub>-tail, even though they interact with different interfaces. Moreover, the binding of both receptor core and R<sub>P</sub>-tail stabilises further the active conformation of arrestin-1. This behaviour was reported by Latorraca *et al.*, 2018, where simulations of the rhodopsin-arrestin-1 complex crystal structure were done: (i) removing the R<sub>P</sub>-tail and leaving only the receptor core or (ii) removing the receptor core and leaving only the R<sub>P</sub>-tail. The simulations showed that, firstly, the removal of the R<sub>P</sub>-tail leads to an increased range of the N-interdomain twist angles, secondly, the binding of either part of the receptor in simulation considerably increases the fraction of time that arrestin-1 spends in active conformations and, lastly, the binding of the entire receptor has an even stronger effect. These computational predictions were also supported by the results obtained from site-directed fluorescence spectroscopy (Latorraca *et al.*, 2018).



**Figure 1.11 | Rhodopsin activation induced by arrestin-1 binding.** Rhodopsin (PDB ID: 5TE5, inactive; 5W0P, active) and arrestin-1 (PDB ID: 1CF1) are shown in cartoon representation, respectively in cyan and light blue (C-domain), green (N-domain), and pink (C-tail). Missing residues in the C-tail were added manually and are represented by a dashed line. The interaction between the receptor and G-protein is blocked by steric exclusion from arrestin-1 interaction that involves the receptor R<sub>P</sub>-tail and receptor core causing the displacement of entire C-tail, the binding of R<sub>P</sub>-tail in the lateral cleft of the C-domain, and the twist of the N-interdomain angle of 20°.

## 1.8 Allosterism and free fatty acid receptor 1 (FFAR1, formerly GPR40)

In recent years, studies of ligands that bind to the allosteric sites of GPCRs have been increasing because they have a therapeutic interest, given that allosteric binding site can in principle be more selective than orthosteric binding site because of greater sequence dissimilarity (the orthosteric binding site is usually very well conserved). Moreover, GPCR allosterism is relevant not only for drug discovery but also for structure-function analyses of GPCR action (May *et al.*, 2007).

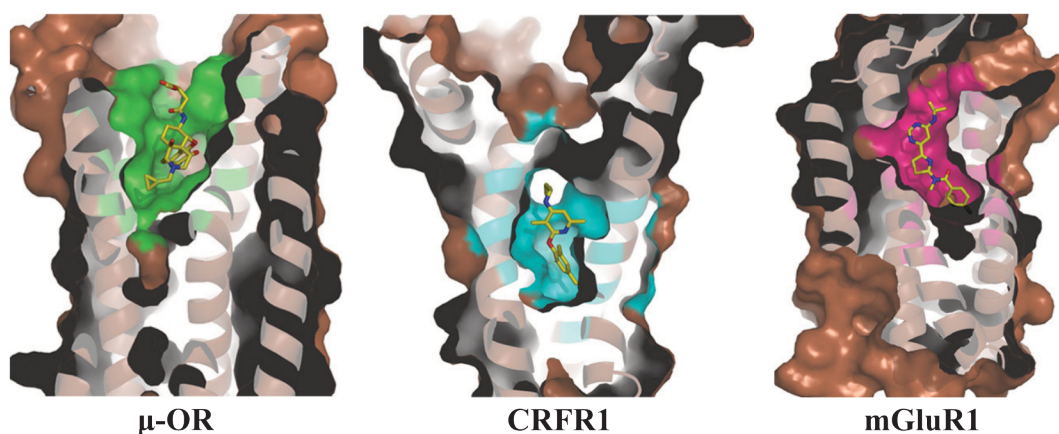
Before going into detail on this subject, it is necessary to make a brief reflection on the following terminology:

- *Orthosteric site*: the primary binding site on the receptor that is recognised by the endogenous agonist for that receptor (May *et al.*, 2007);
- *Allosteric site*: a binding site on the receptor that is topographically distinct from (does not exhibit any overlap with) the orthosteric site (May *et al.*, 2007);
- *Orthosteric ligand*: ligand that binds to the natural ligand-binding site on the receptor and thus directly competes with this natural ligand for receptor binding (Stevens *et al.*, 2013);
- *Allosteric ligand*: ligand that binds elsewhere from the orthosteric binding site and influences the functional properties of the receptor (Stevens *et al.*, 2013);
- *Positive allosteric modulator (PAM)*: also known as allosteric enhancer or potentiator, induces an amplification of the orthosteric agonist's effect, by enhancing orthosteric ligand affinity and/or agonist efficacy (May *et al.*, 2007);



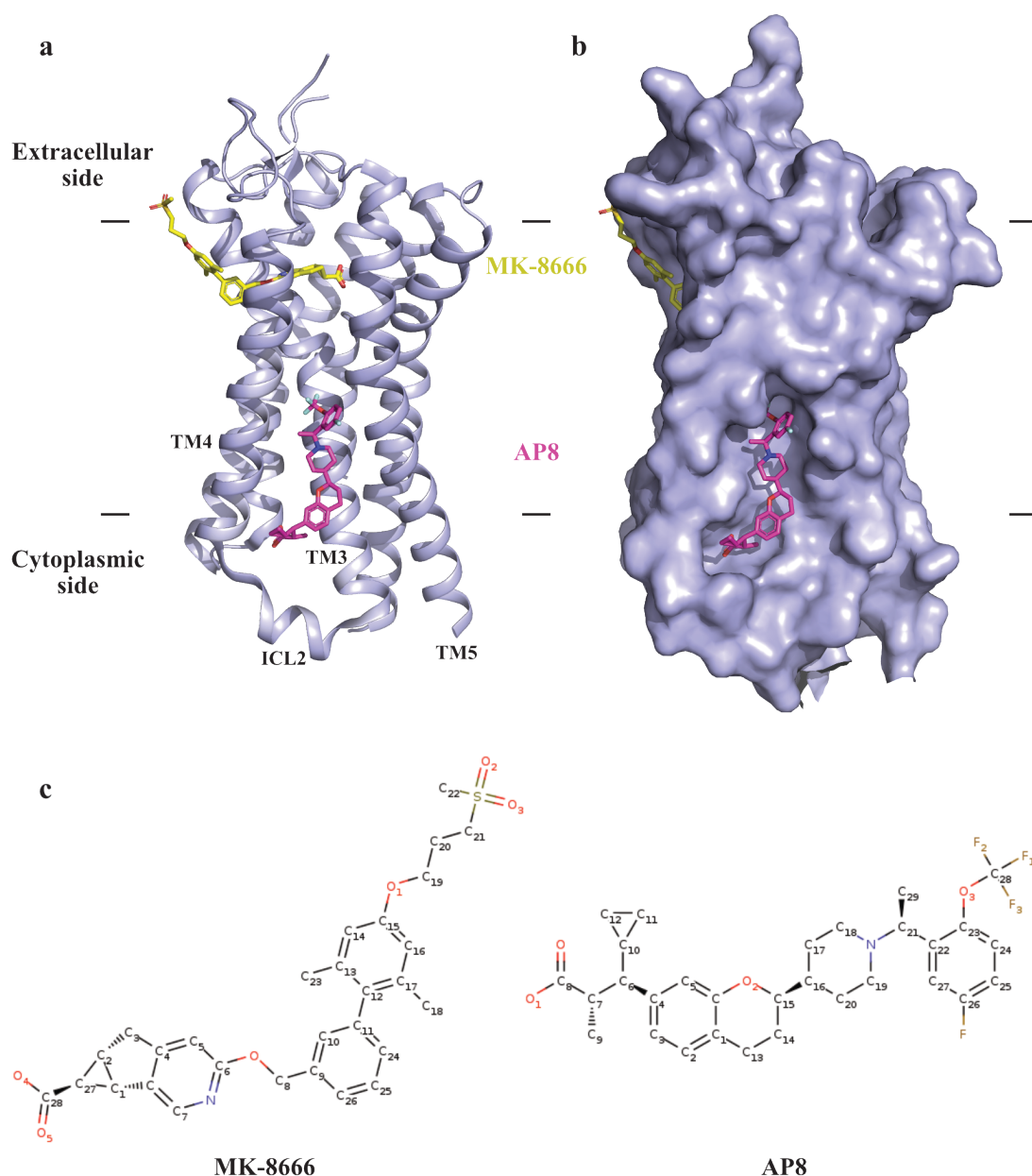
- *Negative allosteric modulator (NAM)*: also known as allosteric antagonist or inhibitor, induces a reduction of the orthosteric agonist's effect, by decreasing orthosteric ligand affinity and/or agonist efficacy (May *et al.*, 2007).

It is, therefore, possible to understand how allosteric ligands can provide a significant substance for the pharmacology of orthosteric ligands both ligand-, receptor-, and cell-dependent. *Ergo*, recognising the diversity of GPCR allosteric effects allows us to describe the molecular processes behind allosterism. The position of the orthosteric site is not similar among the GPCRs but depends on their class and, consequently, the position of the allosteric site is not similar among the GPCRs but depends on their class. For example, for class A, the orthosteric site is generally situated in the 7TM domain core. Contrariwise, as regards class C, the orthosteric site is positioned in the extracellular N-terminus (May *et al.*, 2007). For class B, the orthosteric site spans the extracellular N-terminus and the 7TM domain core. Figure 1.12 shows the orthosteric site for  $\mu$ -OR (class A) and an allosteric site for CRFR1 (class B) and mGluR1 (class C). The intracellular binding partners such as G proteins, are considered in some classifications as allosteric ligands because of their distance from the orthosteric site ( $\sim 30$  Å) (Stevens *et al.*, 2013).



**Figure 1.12 | Cartoon and surface representation of ligand-binding pockets of  $\mu$ -OR (class A, PDB ID: 4DKL), CRFR1 (class B, PDB ID: 4K5Y), and mGluR1 (class C, PDB ID: 4OR2).** Ligands are shown as yellow sticks and are respectively an antagonist for class A and B, and a negative allosteric modulator (NAM) for class C. The comparison shows that the orthosteric site is generally situated in the 7TM domain core for class A, while for class B and C the same position is an allosteric site (adapted figure from Zhang, Zhao and Wu, 2015).

Recently, research on FFAR1 has shown a different position for an allosteric site compared to other GPCRs (Lu *et al.*, 2017). Figure 1.13 shows a crystal structure of FFAR1 in complex with the partial agonist MK-8666 and the allosteric ligand AP8, a PAM. The binding cooperativity between these two ligands changes the receptor into a more active-like state. Further studies about the position of this allosteric site could reveal that far from being unique to this receptor, this position is general and could be relevant for drug design for a range of other receptors, for example the Beta-2 adrenergic receptor ( $\beta_2$ -AR) (Liu *et al.*, 2017) and C5a anaphylatoxin chemotactic receptor 1 (C5aR1) (Robertson *et al.*, 2018). Long-chain free fatty acids bind with FFAR1, activating the receptor such that this enhances glucose-dependent insulin secretion. Therefore, studies to better understand this receptor have an important role in the comprehension of treatment of type 2 diabetes mellitus.



**Figure 1.13 | Cartoon (a) and surface (b) representation of the FFAR1 in complex with the partial agonist MK-8666 (yellow) and the AP8 (purple). (c) Chemical structure of the MK-8666 and AP8. The binding with AP8 induces rearrangements of TM4 and TM5, and transition of the ICL2 into a short helix, therefore, the receptor assumes more active-like state.**

## 1.9 GPCRs: experimental methods

One of the greatest challenges of GPCRs is to obtain the native structure of these receptors due to their instability during the purification. GPCRs lose their native conformation and activity once moved from the cell membrane. To overcome this problem, the company Sosei Heptares which is an industry pioneer in GPCR structure-based drug design, developed the StaR® (Stabilised Receptor) technology. The StaR® technique generates a small number of point mutations in GPCR sequences to increase the stability of the receptors in a natural state without changing its pharmacology. The StaR GPCRs can be crystallised in multiple states (active, semi-active or inactive). Moreover, the ligands suitability to fit the receptor can be screened *in silico* and this information facilitates the precise design of drug candidates. Crystallizing proteins, however, is not an easily achievable process since that millions of identical copies of a protein have to be produced and aligned in the same orientation to solidify into a crystal. Contrary to X-ray diffraction, in the Nuclear magnetic resonance (NMR) spectroscopy crystals are not required, but this technique has limitations. Lately, with the advance of technologies, another technique has taken hold, namely Cryo-electron microscopy (Cryo-EM). The main advantage of such technique is that does not require crystals (Table 1.3). However, this technique is not new since it was developed in the 1970s with significant advances in 1993 (Baldwin, 1993) and 1997 (Baldwin *et al.*, 1997; Unger *et al.*, 1997) but its major breakthrough occurred only in 2013 when, the achievement of near-atomic resolution (Bai *et al.*, 2013; Li *et al.*, 2013) described as the ‘*resolution revolution*’ (Cheng, 2018) due to a series of technological and algorithmic advances, attracted wide attention to the approach. Moreover, in 2017 this technique was a protagonist of the Nobel Prize in Chemistry awarded to Jacques

# CHAPTER 1

Dubochet, Joachim Frank, and Richard Henderson "*for developing Cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution*" (Nobel Media, 2017).

**Table 1.3 | Comparison of X-ray diffraction, NMR, and Cryo-EM**

Method	Sample	Resolution	Advantages	Disadvantages	GPCR structures
X-ray diffraction	Crystallisable sample, soluble protein, membrane protein, ribosome, DNA/RNA, and protein complex	High	Well developed, high resolution, broad MW range, and easy for model building	Difficult for crystallisation and diffraction, solid structure preferred, and static crystalline state structure	393
NMR	Water soluble samples and sample with MW<35 KDa	-	3D structure in solution and good for dynamic study	Difficult for sample preparation	1
Cryo-EM	Large protein, membrane protein, ribosomes, protein complex, compounds, and sample with MW>50 KDa	Relatively low (>3.5 Å)	Structure in native state, small sample size, and easy sample preparation	Highly dependent on EM techniques and costly EM equipment	71

MW = molecular weights

## 1.10 Perspectives of my PhD project

The high flexibility of GPCRs is the reason behind the multiple activation pathways, which can be desirable or undesirable. The purpose of future projects is to review the current research of GPCRs from a dynamic point of view, ultimately in order to design drugs with only beneficial effects, avoiding the side effects deriving from parallel activation of alternative pathways. This type of design process is facilitated using molecular dynamics (MD) simulations, which is a computer technique that simulates the motion of the atoms in molecules. The MD algorithms introduces time as a new discrete variable in computational simulations, which allows consideration of the evolution of the systems according to the dynamics driven by the reciprocal interactions of their components. In other words, starting from an initial configuration on the energy landscape, it is possible to sample different conformational states and pathways from one local minimum of energy to another, passing through saddle points (transition states). Therefore, this system allows reconstructing binding events at a molecular scale (more details are discussed in section 2.3).

In this project, MD has been used to study in particular two GPCRs, free fatty acid receptor 1 (FFAR1; Rhodopsin family - class A) and calcitonin receptor (CTR; Secretin family - class B) in order to identify possible novel binding sites and improve our structural and functional understanding. I have explored GPCRs using traditional MD methods (Karplus and Petsko, 1990), supervised MD methods (Sabbadin and Moro, 2014), and metadynamics (Laio and Parrinello, 2002), since these latter techniques also offer the opportunity to explore conformational space more thoroughly than conventional MD (see section 2.3.2 and 2.3.3, respectively).

## 1.11 Androglobin: a complex structure still to discover

In recent years, there has been an increasing interest in developing a computational method able to predict the structure and the function of proteins from sequence information, without time and cost limitations (Li *et al.*, 2010). For this purpose, many fold recognition servers are freely usable (*e.g.* I-Tasser, HHpred, Phyre2, RaptorX, and *etc.* - discussed in section 2.1) and these tools are fast becoming a key instrument in the development of better model building (Taddese *et al.*, 2014). Moreover, a computer program named Modeller (section 2.2) is used to generate protein structure models obtained from the prediction tools, a relevant (multiple) sequence alignment, and a suitable template to evaluate the fold of the final model (Webb and Sali, 2014).

A secondary project about Androglobin is being conducted (see Appendix, section B), out of respect to the bequest of Peter Nicholls, whose main interest was in heme proteins. To date, only two articles were published about this interesting protein (Hoogewijs *et al.*, 2012; Huang *et al.*, 2014) and its structure is still unknown. Androglobin (Adgb), also known as Calpain-7-like protein (UniProtKB ID: Q8N7X0), is one of the eight globin types in vertebrates which include haemoglobin (Hb), myoglobin (Mb), neuroglobin (Ngb), cytoglobin (Cygb), globin X (GbX), globin Y (GbY), and globin E (GbE) (Burmester and Hankeln, 2014). Commonly, the globin fold and its general features are conserved among vertebrate globins. The main difference is related to the binding scheme of the iron atom of the heme in the deoxygenated state, which is penta-coordinated (with the sixth binding site of the  $\text{Fe}^{2+}$  being empty) in the Hb, Mb, and GbE, while is hexa-coordinated (with the distal amino acid of the protein chain bound to the  $\text{Fe}^{2+}$ ) in Ngb, Cygb, GbX, and Adgb. Instead, no

data is currently available on the coordination status of GbY (Burmester and Hankeln, 2014).

In humans, the Adgb gene is mainly expressed in the testis and experimental data suggested a role in reproduction, as shown an increase in expression of the gene in fertile compared to infertile males (Hoogewijs *et al.*, 2012).

The Adgb protein consists of 1667 amino acids in which were identified an N-terminal calpain-like domain, an internal globin domain, and an IQ calmodulin-binding motif (Hoogewijs *et al.*, 2012). A schematic representation of the Adgb domain organisation is reported in Figure B.1 (see Appendix, section B). However, compared to other heme proteins, the whole structure of Adgb is still unknown because of the instability of the protein during purification. Therefore, this study aimed to contribute to research into the three-dimensional protein structure of Adgb using fold recognition servers and Modeller to build a model. In particular, the analysis was done for the sequence fragments pre- and post- the globin part of Adgb, and for the globin domain. Preliminary results show an interesting model structure of the pre-internal globin domain consisting of a  $\beta$ -barrel structure. This information suggests new opportunities to model or purify and crystallise only a specific part of Adgb, which is a stable structural unit. Regarding to the globin domain, consisting of the typical 8 alpha-helical structure binding a heme group, and the IQ domain, which is unique to Adgb, several models were generated with a different structural position of the IQ domain. Two models were selected for molecular dynamics simulations to assess the difference in stability, enabling a putative model to be presented.



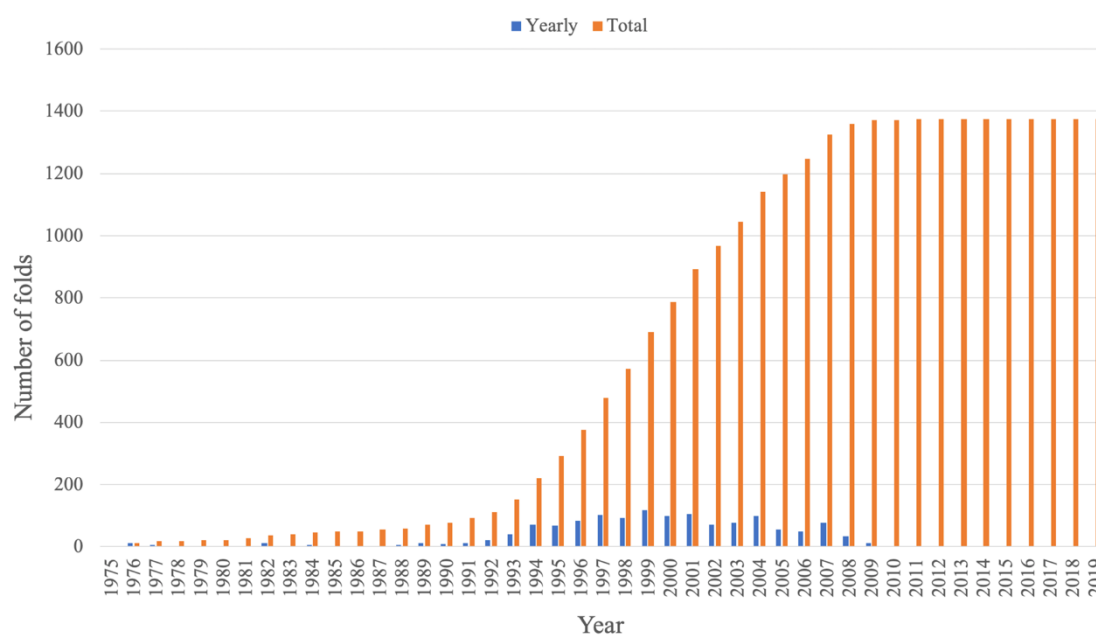


## 2. Methodology

### 2.1 Protein fold recognition servers

Fold recognition was used in the subsidiary project on Androglobin. Several fold recognition servers, which are freely available for protein structure and function prediction, were used in this project, such as I-TASSER (Roy *et al.*, 2010), Phyre2 (Bennett-Lovsey *et al.*, 2008), HHpred (Söding *et al.*, 2005), RaptorX (Källberg *et al.*, 2012), QUARK (Xu and Zhang, 2012), and IntFOLD (Roche *et al.*, 2011). The approach of those servers is based on the observation that the number of unique folds in nature seems to be restricted to around 1375 topologies since 2010 (as shown in Figure 2.1) (Berman *et al.*, 2000), thus considerably similar structures are adopted by many different remote homologue protein. This is consistent with the latest CASP (Critical Assessment of Structure Prediction) *i.e.* CASP13, in which it was noticed that the overall accuracy of models improved dramatically compared to the previous rounds (Kryshtafovych *et al.*, 2019).

Generally, the result of each server is a list of hits ordered by Z-score (the key metric) that is associated with a “high”, “medium” or “low” index of confidence, indicating the expected reliability of the result. Although the prediction is not necessarily correct, it could give an important indication of the structure and function of the protein. Therefore, in this project both the highest ranked hit and the top 10 hits were considered.



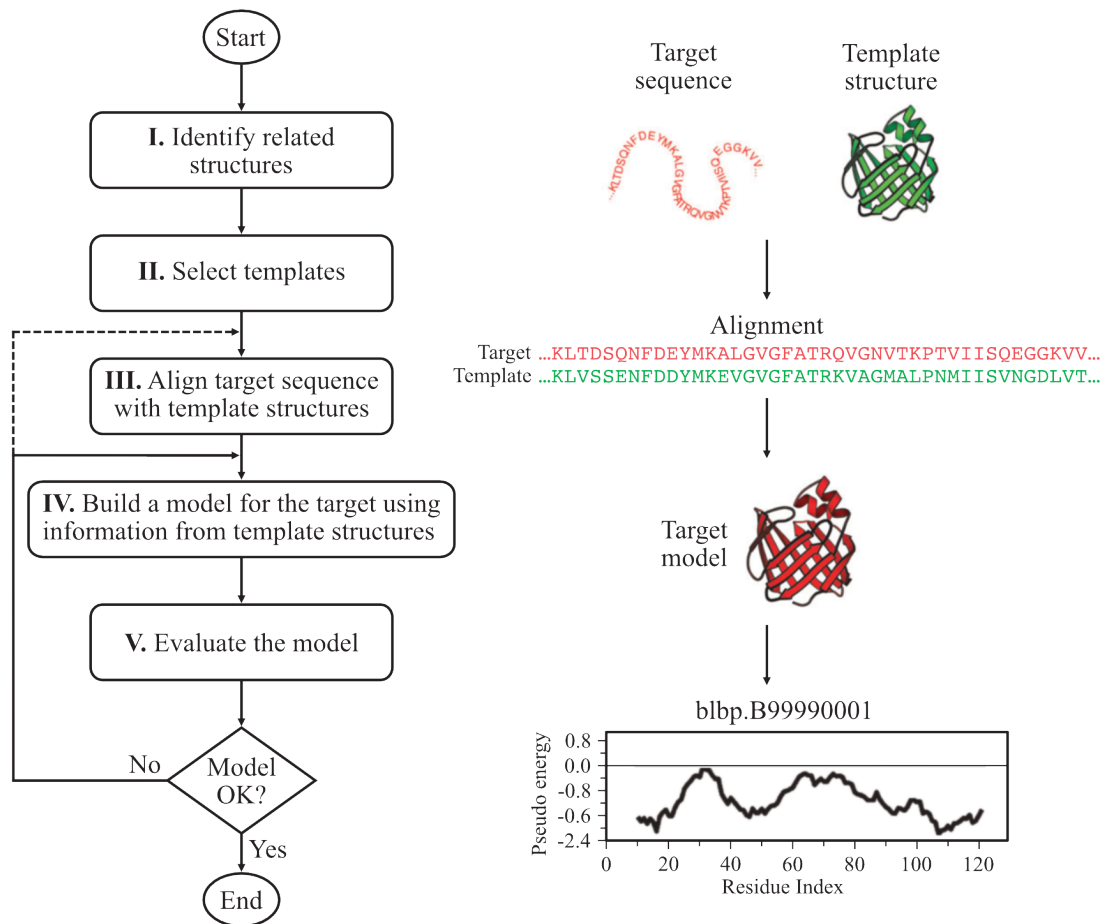
**Figure 2.1 | A plot of the growth of unique folds (topologies) per year defined by CATH (Dawson *et al.*, 2017).** The number of topologies available in the PDB has been growing exponentially since 1990 until 2010 where the graph reaches a plateau due to the absence of new folds. Data were taken from the RCSB PDB (<https://www.rcsb.org/>) (Berman *et al.*, 2000).

## 2.2 Comparative protein structure modelling: Modeller

The computer program Modeller 9.19 was used for homology modelling to generate protein structure models obtained from (i) the fold prediction tools, (ii) a relevant (multiple) sequence alignment, and (iii) a suitable template to evaluate the fold of the final model. Figure 2.2 shows the five steps of comparative modelling: *search for templates*, *selection of one or more templates*, *target-template alignment*, *model building*, and *model evaluation* (Webb and Sali, 2014). The template selection is a key step: it is important to obtain a high percentage of identical residues as well as a low number and short length of gaps in the alignment (*i.e.* higher overall sequence similarity) between the target and the template sequences. The program extracts the distance between each atom and dihedral angle restraints on the target from the template structure(s) to combine them with bond length and angle preferences following the general rules of protein structure (Webster, 2000). The sum of all these conditions constitutes the terms of the pseudo energy function. Based on this function, it is possible to choose the best model between those calculated for the same target by picking the model with the lowest value of the pseudo energy profile (Webb and Sali, 2014).

Protein secondary structure prediction servers such as JPred (Drozdetskiy *et al.*, 2015) and PSIPRED (acronym that stands for PSI-blast based secondary structure PREDiction) (Jones, 1999) were used with default parameters to further evaluate the model. Both JPred and PSIPRED use the artificial neural network machine learning methods in their algorithm in order to predict a protein's secondary structure (*i.e.* beta sheets, alpha helixes, and coils) from the primary sequence. Although JPred makes its predictions by using the JNet (Joint Network) algorithm (Cole *et al.*, 2008), while,

PSIPRED uses its own algorithm (PSIPRED algorithm) (Jones, 1999), they can both achieve a similar accuracy with a percentage accuracy higher than about 80% (Drozdetskiy *et al.*, 2015).



**Figure 2.2 | Comparative protein structure modelling.** The diagram shows the five steps of comparative modelling: *search for templates, selection of one or more templates, target-template alignment, model building, and model evaluation* (adapted figure from Webb and Sali, 2014).


## 2.3 Molecular Dynamics (MD)

Molecular dynamics (MD) is a computer simulation technique that describes the evolution of a chemical system over the time using the force field equation (discussed in section 2.3.1). It often plays an important role in drug discovery because it can reproduce very dynamic processes as molecular recognition and drug binding starting from the macromolecular structures (static models) produced by X-ray crystallography, NMR, Cryo-EM, and homology modelling. Therefore, the consideration of a macromolecule in constant motion rather than a single frozen structure allows one to understand better the process of the ligand-protein binding by exploring different conformations of the system.


### 2.3.1 Molecular Mechanics: defining a force field and a system for MD simulations

It is well known that protein flexibility plays an important role in ligand binding, but it is very complex to generate systems while taking into consideration the possible movement of each atom. To reduce this computational complexity, atomic motions have been approximated based on Newtonian physics. Therefore, an equation like the one reported in Figure 2.3 (2.1) estimates the forces acting on each of the system atoms.

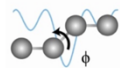
$$U = \underbrace{\sum_{\text{bonds}} K_r (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]}_{\text{Bonded}} + \underbrace{\sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]}_{\text{Non-bonded}} + \text{Improvers} \quad (2.1)$$




**I**




**II**



**III**



**IV**



**V**

**VI**

**Figure 2.3 | Force field equation.** The equation describes the atomic forces that govern molecular movement. This equation is constituted by the summation of bonded, non-bonded, and improper terms. Bonded forces (in green) arise due to interaction of covalently bound atoms, such as **(I)** oscillations about the equilibrium bond length (bonds), **(II)** oscillations of 3 atoms about an equilibrium bond angle (angles), and **(III)** torsional rotation of 4 atoms about a central bond (dihedrals). Otherwise, non-bonded forces (in blue) arise due to **(IV)** van der Waals interactions, modelled using the 6 - 12 Lennard-Jones potential, and **(V)** electrostatic interactions, modelled using Coulomb energy function. Finally, the last term **(VI, improvers, in red)** defines 4 atoms to maintain planarity for configurations such as the peptide bond (adapted figure from Durrant and McCammon, 2011); the term is similar to that for standard dihedrals but 2 of the atoms are not bonded.

The contributions of the various atomic forces that govern molecular dynamics, as described in the above equation, are parameterised to reach the optimal flexibility and lengths of the springs describing, firstly, bonds length and angles, secondly, the best partial atomic charges used for calculating electrostatic-interaction energies, and the proper van der Waals atomic radii. These parameters are obtained either from quantum mechanical calculations or by fitting to experimental data such as X-ray, NMR, Spectroscopy, *etc.* These parameters define a ‘force field’. Currently, many force fields are available, among the most common are AMBER and CHARMM, from which simulation software packages take their name. These force fields have been developed to provide a set of descriptions for common biological macromolecules such

as nucleic acids, proteins, lipids, *etc.* Once a suitable force field has been determined, (in our studies CHARMM), three important input files are needed to perform MD simulations such as parameter, topology, and coordinate files, each of them is in format *.prm*, *.psf*, and *.pdb*, respectively. Moreover, computer clusters that use many processors operating simultaneously in parallel are required; in our case these are the multiple cores of a GPU.

As discussed before, the chemical system during MD simulations follows the trajectories of  $N$  particles according to Newton's equation of motion:

$$F_i = m_i a_i \quad (2.2)$$

$$F_i = -\nabla U(r_i, \dots, r_N) \quad (2.3)$$

where  $U(r_i, \dots, r_N)$  is the potential energy function of  $N$  particles (2.1).

Given the position  $r$  and velocity  $v$  of  $N$  particles at time  $t$ , it is possible to integrate the Newton's equation according to Taylor's theorem and achieve the positions of the system components at the following time  $(t + \Delta t)$ :

$$v_i(t + \Delta t) = v_i(t) + \frac{F_i(t)}{m_i} \Delta t \quad (2.4)$$

$$r_i(t + \Delta t) = r_i(t) + v_i(t) \Delta t + \frac{F_i(t)}{2m_i} \Delta t^2 \quad (2.5)$$

Therefore, knowing the force field, it is possible to calculate the force  $F_i$  (2.3) that allows us to determine the acceleration  $a_i$  of the particles (2.2), to proceed with the calculation of velocity  $v_i$  (2.4), to finally calculate the position  $r_i$  (2.5).

In the very first step of a MD simulation, the assignation of the initial velocities is necessary. They are sampled from a Maxwell-Boltzman distribution:

$$P_{(v_{ix})} = \left( \frac{m_i}{2\pi k_B T} \right)^{1/2} \exp \left[ -\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T} \right] \quad (2.6)$$



where  $P_{(v_{ix})}$  is the probability that an atom  $i$  of mass  $m_i$  has a velocity  $v_{ix}$  in the direction  $x$  at a temperature  $T$ . This distribution is a Gaussian distribution that can be obtained using a random number generator (Leach, 2001).

Usually, MD simulations are performed with a time step of around 1-2 fs (which is a good compromise between accuracy in mathematical integration at biologically relevant time scales and computational cost) in the NVE ensemble, where the number of atoms (N), volume (V), and energy (E) are constant. On the contrary, the alternative NVT and NPT ensembles are employed to avoid the truncation errors that impede keeping the energy constant in long simulations. To keep the pressure (P) constant, the Berendsen barostat (Berendsen *et al.*, 1984) was employed, while for temperature (T), the Langevin thermostat (Loncharich *et al.*, 1992) was employed. Moreover, the M-SHAKE algorithm (Kräutler *et al.*, 2001) was used to constrain the bond lengths involving hydrogen atoms in order to reduce of the computational effort for obtaining a trajectory of a given length. In addition, to avoid unphysical edge effects during MD simulations, periodic boundary conditions are used which employ the use of a repeatable region (unit cell) to simulate a segment of molecules in a larger solution.

The introduction of Graphic Processing Unit (GPU) technology and dedicated MD codes during the last decade allowed MD simulations to reach the millisecond scale (Shaw, 2013). Nevertheless, even longer timescales (unreachable for this methodology) are sometimes necessary to explore enough sampling, in particular, to describe binding and unbinding events, because of the high energy barriers associated with transition states. In this regard, other methods are employed including Supervised Molecular Dynamics (SuMD) and Metadynamics (MetaD), as described below.

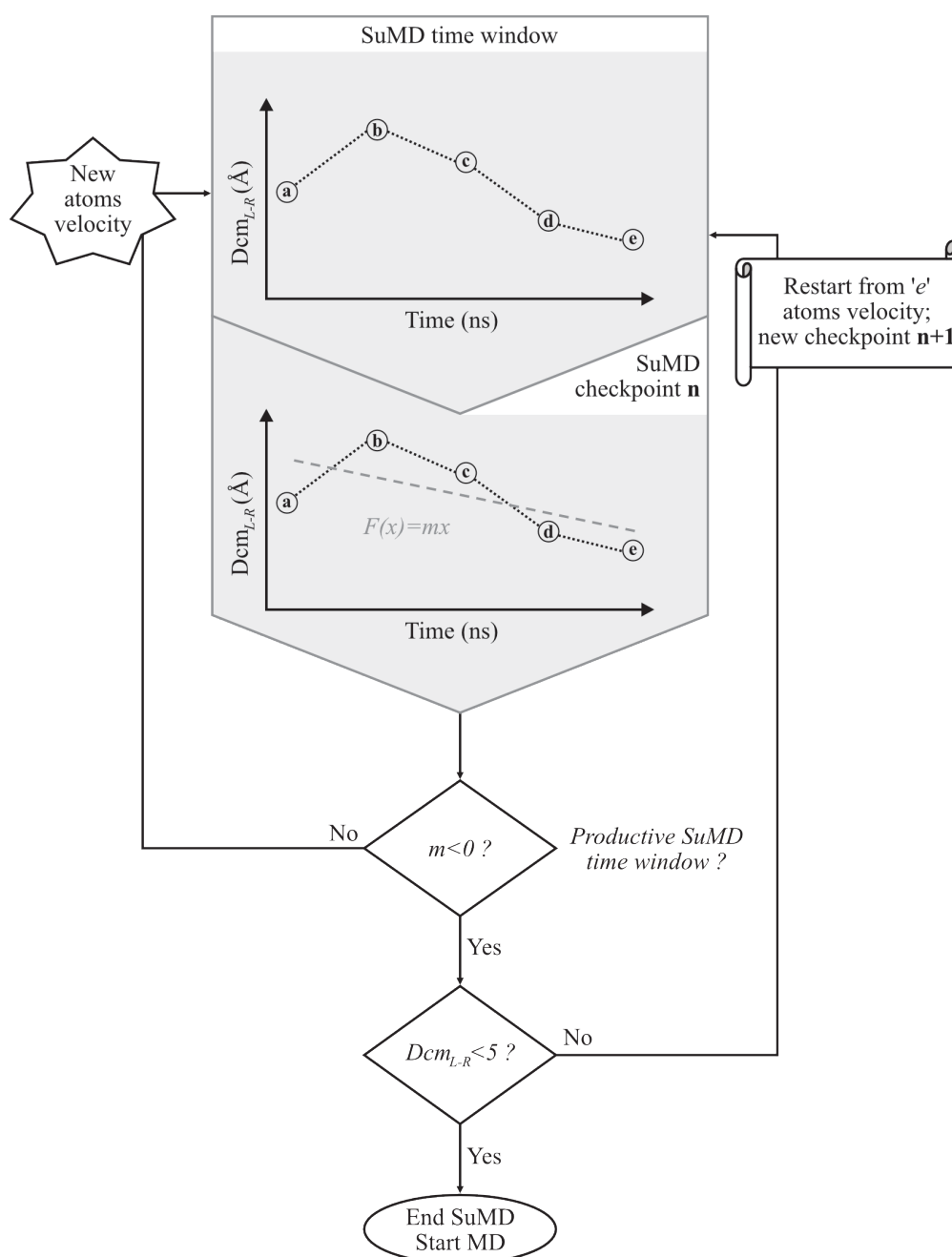
### 2.3.2 Supervised molecular dynamics (SuMD)

It is well known that the way a ligand recognises its receptor is a process that requires a long microsecond ( $\mu\text{s}$ ) time scale using a MD simulation. In order to overcome this time scale limitation, a new technique named Supervised Molecular Dynamics (SuMD) has been developed (Sabbadin and Moro, 2014). SuMD was successfully applied to GPCRs (Sabbadin and Moro, 2014; Degnutti *et al.*, 2018; Bissaro *et al.*, 2019; Atanasio *et al.*, 2020; Jakowiecki *et al.*, 2020). This technique can reduce the ligand-receptor recognition pathway up to nanosecond (ns) time scale, independently from the chemical structure of the ligand, its starting position, and its receptor binding affinity (Cuzzolin *et al.*, 2016). Moreover, it can provide details on transition states at the atomic level and on structural information of metastable intermediate states (meta-binding sites) that chronologically anticipate the orthosteric one, allowing the characterisation of multiple binding events (*i.e.* meta-binding, allosteric, and orthosteric sites). This approach is implemented as a command line tool written in Python, Tool Command Language (TCL), and bash programming languages (Sabbadin *et al.*, 2018). This is capable of monitoring the ligand-receptor distance focusing on the centre of masses of the ligand atoms and some key residues in the binding site of the receptor, using standard short MD simulations according to the algorithm summarised in Figure 2.4 (Sabbadin and Moro, 2014; Cuzzolin *et al.*, 2016). As shown, during a regular time interval of short MD simulations, a random number of ligand-receptor distance points ( $D_{\text{cmL-R}}$ ) is collected and a linear function  $f(x)=mx$  is plotted to interpolate the collected points. If the slope of the function is negative ( $m<0$ ), the ligand-receptor distance is reduced, and the simulation is restarted from the last produced set of coordinates. Otherwise, the previous step is repeated following

## CHAPTER 2

---

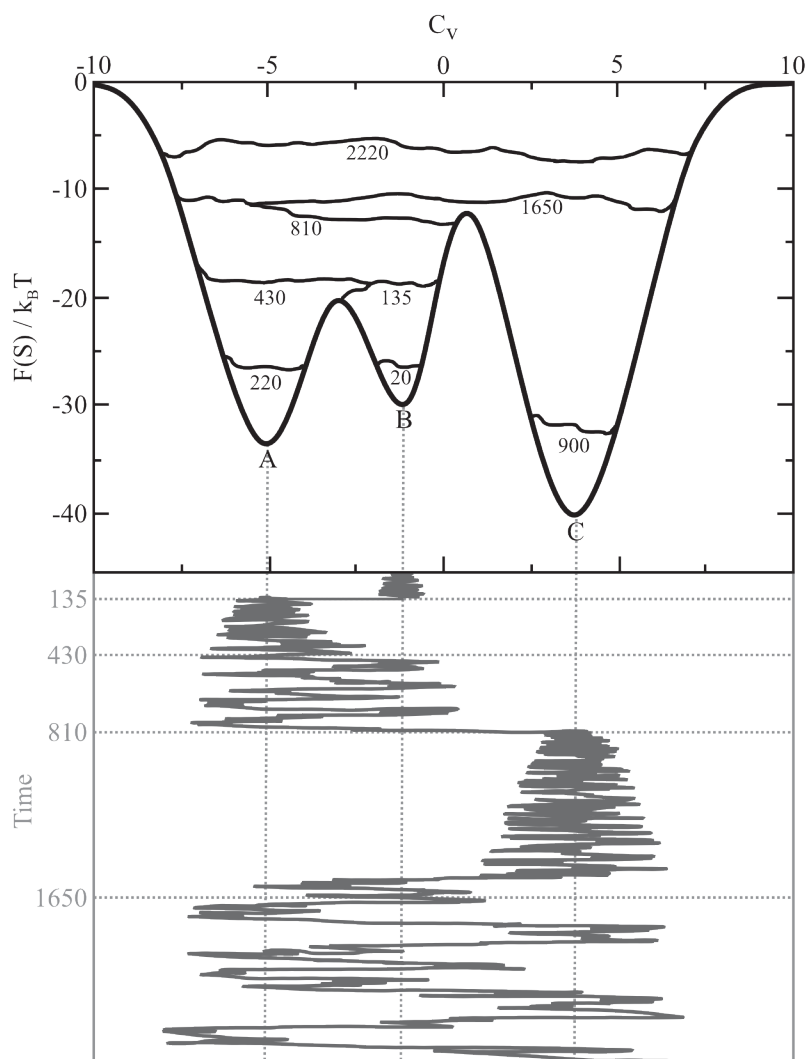
reassignment of the velocities of the atoms. The SuMD is repeated until the ligand-receptor distance is less than 5 Å. Thereafter, further classic MD steps are conducted.



**Figure 2.4 | The SuMD workflow.** A schematic representation of SuMD simulation steps. If the productive SuMD time window shows a shorter distance between ligand and receptor ( $D_{cm_{L-R}}$ ) compared to the previous step ( $m < 0$ ), the simulation is restarted from the last produced set of coordinates (checkpoint **n**); otherwise, the velocities of the atoms are reassigned by the repetition of the previous step (**new** productive SuMD time window). The SuMD cycle is repeated until  $D_{cm_{L-R}} < 5$  Å, when this distance has been reached, further classic MD steps are conducted.

### 2.3.3 Metadynamics (MetaD)

Metadynamics (MetaD) is a powerful method that helps to resolve the sampling problem, thus enhancing the number of the states explored. This technique simplifies the sampling by the introduction of an additional bias potential (in other words, a force) that operates on a selected number of degrees of freedom, usually named as collective variables (CVs) (Barducci *et al.*, 2011). Therefore, if an energy bias is added during the simulation along these CVs, the probability of observing the transition states increases, while the probability that the system will return to the previous already visited configuration decreases according to the energetic Gaussian function shape (Figure 2.5).



**Figure 2.5 | Example of a MetaD simulation that biases one CV. (Top)** Representation of the progressive filling of the potential energy minimums by means of the Gaussians deposited along the trajectory. (The numbers are explained in the next section). **(Bottom)** Time evolution of the collective variables during the simulation. The system, at the beginning of the simulation, is situated in the minimum B. The transition from the minimum B to the minimum A takes place after 135 energy depositions. Other 810 bias cycles are necessary to fill it and, after that, the system is allowed to explore the minimum C (adapted figure from Barducci *et al.*, 2011).

### 2.3.4 System preparation for MD simulations

The receptor systems were prepared for MD simulations of with the CHARMM36 force field. Setting up a membrane protein simulation system for the CHARMM36 force field (Figure 2.6) (Huang and MacKerell, 2013) usually requires a multistep insertion method that integrates both Python High-Throughput Molecular Dynamics (HTMD) (Doerr *et al.*, 2016) and TCL scripts. In general, five major working steps are required: *protein orientation*, *titration of ionisable groups*, *bilayer preparation*, *system solvation*, and *system neutralisation*. Below is reported the explanation of each step, using a FFAR1 structure (modelled active conformation, MAC) as an example.

*Protein orientation.* The first step for a membrane protein is its superimposition with a pre-existing bilayer of 1-palmitoyl-2-oleyl-sn-glycerol-3-phospho-choline (POPC) along the Cartesian z-axis using the Orientations of Proteins in Membranes (OPM) database (<https://opm.phar.umich.edu/>). This database is a store of membrane protein (*i.e.* transmembrane, monotopic, and peripheral proteins collected from the Protein Data Bank - PDB, <https://www.rcsb.org/>), whose correct spatial position in the lipid bilayer has been predicted and compared with experimental data (Lomize *et al.*, 2006). If an OPM file is not available for the specific structure under analysis, the database will suggest the most representative structure(s) of the protein in question. In this example, the MD system for FFAR1 structure (modelled active conformation, MAC) was prepared using 5TZY (another intermediate FFAR1 structure) as a representative structure for the coordinates (Figure 2.6a).

*Titration of ionisable groups.* The second step concerns the addition of the hydrogen atoms by means of PDB2PQR (Dolinsky *et al.*, 2004) and PROPKA (Olsson

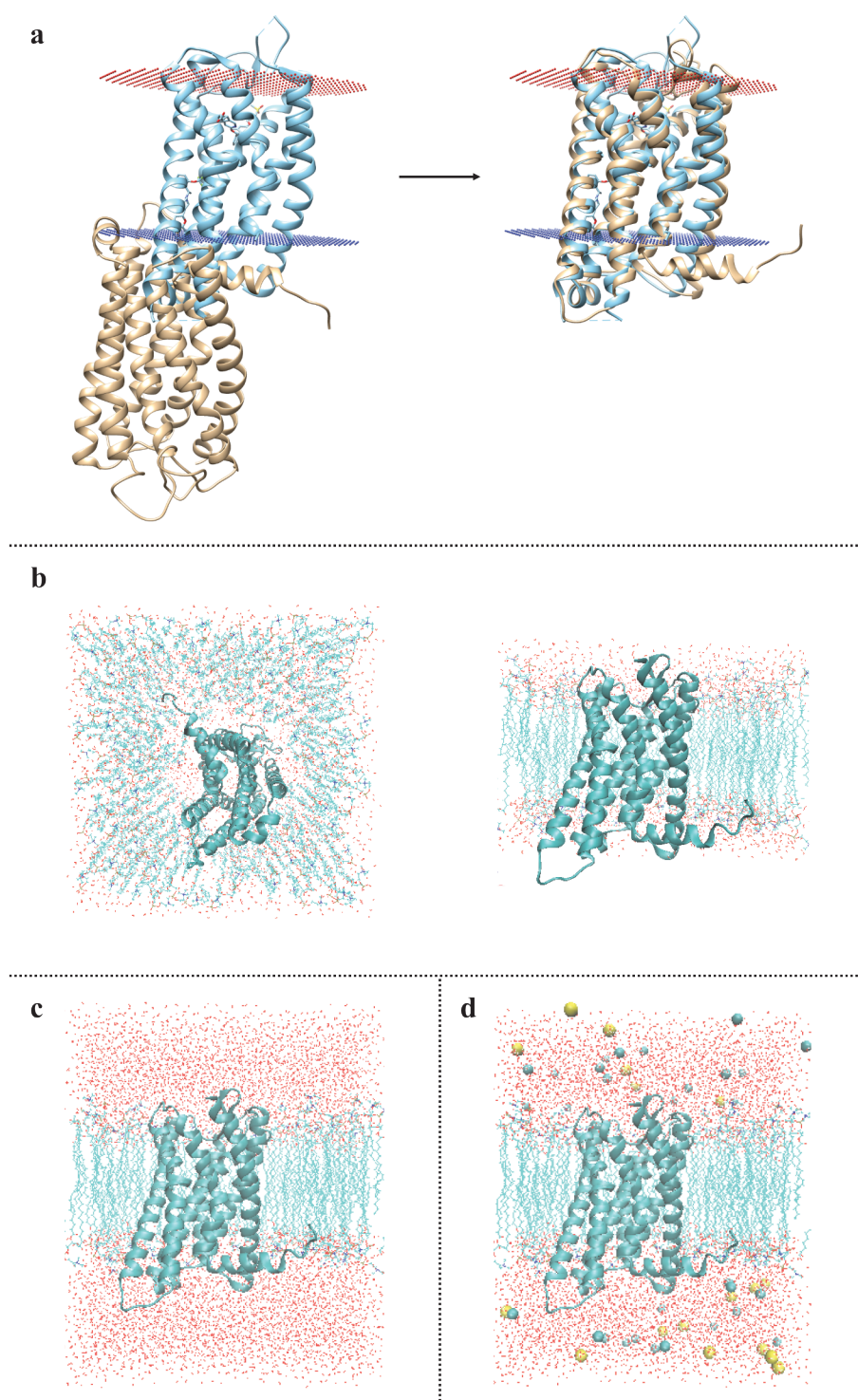
*et al.*, 2011) software (considering a simulated pH of 7.0). The protonation of titratable side chains was checked by visual inspection.

*Bilayer preparation.* The receptor was embedded in a square of 85 Å x 85 Å POPC bilayer using the VMD Membrane Builder plugin 1.1 (<http://www.ks.uiuc.edu/Research/vmd/plugins/membrane/>) through an insertion method (Sommer, 2013), using typically up to 200 lipids. Lipids overlapping the receptor transmembrane-domain bundle were removed on the basis of a simple distance cut-off (0.6 Å) between protein and lipid (Figure 2.6b).

*System solvation.* Solvation techniques use a geometrical rather than a thermodynamical approach to add water to a chemical system; here the TIP3P water model (Jorgensen *et al.*, 1983) is used. The main aspects of this step are to solvate cavities of proteins that do not easily exchange water molecules with the bulk solution and to have a sufficiently solvated starting structure. In this example, TIP3P water molecules (typically up to 20000) were added to the simulation box (*e.g.* 85 Å x 85 Å x 100 Å) by means of the VMD Solvate plugin v.1.5 (<http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/>) (Figure 2.6c).

*System neutralisation.* After solvation, it is important to neutralise the system; this was achieved by transmuted random water molecules into either KCl or NaCl, until the total system charge converged towards zero, for example by adding Na<sup>+</sup>/Cl<sup>-</sup> counter ions to a final concentration of 0.150 M by means of the VMD Autoionize plugin 1.3 (<http://www.ks.uiuc.edu/Research/vmd/plugins/autoionize/>) (Figure 2.6d).





**Figure 2.6 | System preparation for MD simulations. (a) *The protein orientation.*** The orientation process of the FFAR1 receptor (MAC) with OPM structure (PDB ID: 5TZY) was arranged using UCSF Chimera 1.12 program. Structures are shown in cartoon representation, respectively in gold and in cyan. On the left, the MAC and OPM structures in the original spatial position are shown, while, on the right, the MAC was

superimposed with the pre-existing bilayer of POPC of the OPM structure. **(b) *The bilayer preparation.*** the MAC view from the top (on the left) and from the front (on the right) after removal of overlapping lipid molecules. The protein is shown in cartoon representation while water and phospholipid bilayer are shown in line representation. **(c) *The system solvation.*** The water box creation (x,y,-z,+z) lies over and under the protein-phospholipid bilayer system. The MAC is shown in cartoon representation while water and phospholipid bilayer are shown in line representation. **(d) *The system neutralisation.*** The final simulation box, after complete solvation and neutralisation of the system. The MAC is shown in cartoon representation, water and phospholipids bilayer are shown in line representation, and Na<sup>+</sup> Cl<sup>-</sup> ions are shown in van der Waals (VDW) representation, respectively in cyan and yellow.

### 2.3.5 Ligand parametrisation belongs under Molecular Mechanics

The parametrisation of the ligand is a crucial part of the MD simulations since there is the possibility of wasting significant computational resources reaching misleading results (if inaccurate parameters are used). The ligands were parameterised in the CGenFF (Vanommeslaeghe and MacKerell, 2012; Vanommeslaeghe *et al.*, 2012) force field, which is designed to be combined with CHARMM36 (Huang and MacKerell, 2013), through a stepwise procedure that comprises (<https://cgenff.umaryland.edu>):

- 1) Retrieving a starting topology and parameters files from the web server ParamChem (Ghosh *et al.*, 2011) (the ligand coordinates have to be provided in mol2 format);
- 2) Evaluation of the penalties associated with each atomic partial charge and the parameters of the rotatable dihedrals of the ligand. A penalty score is affiliated automatically with each of possible substitution by ParamChem, as an approximation for the desired parameter, the one with the lowest total penalty

is chosen. Therefore, the penalty score is returned to the user as a measure for the accuracy of the approximation. If the penalties are below or close to the value of 50 it is possible to move to the last point of the procedure, otherwise, the next steps are necessary;

- 3) Computation of the new partial charges from scratch, on an optimised geometry of the ligand at a quantum mechanics (QM) level of theory (*i.e.* HF/6-31G\*), using the restrained electrostatic potential (RESP) approach (Bayly *et al.*, 1993);
- 4) Fragmentation of the ligand in order to reduce the number of atoms involved in the successive computation but keeping the chemical environment that defines the atom types of the dihedral that needs optimisation;
- 5) Dihedral optimisation by means of the Parameterise tool that comes with the HTMD (Doerr *et al.*, 2016) python suite (the fragment mol2 files are needed as input), or by means of the force field toolkit (ffTK) (Mayne *et al.*, 2013), implemented as a VMD plugin.

### 2.3.6 System equilibration and MD settings belongs under Molecular Dynamics

The *system equilibration* involves a MD session during which protein and ligand atoms are kept geometrically restrained by a harmonic force (first term of equation 2.1), whereas all the other atoms (*e.g.* water and lipid molecules) are free to diffuse and relax, exploring degrees of freedom and allowing a more stable and realistic chemical system. It is generally useful to keep in mind that at least 10 - 20 ns are required to equilibrate a phospholipid bilayer. Progress in equilibration can be judged

by monitoring the dimension of the simulation box, lipid properties, and total system energy.

The MD engine ACEMD (Harvey *et al.*, 2009) was employed for both the equilibration and productive simulations. Equilibration was achieved in isothermal-isobaric conditions (NPT) using the Berendsen barostat (Berendsen *et al.*, 1984) (target pressure 1 atm) and the Langevin thermostat (Loncharich *et al.*, 1992) (target temperature 300 K) with a low damping of  $1 \text{ ps}^{-1}$  and a positional restraints of  $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  on protein atoms. A four-stage procedure was performed employing an integration time step of 2 fs.

In the *first stage*, clashes between protein and lipid atoms were reduced through 2000 conjugate-gradient minimisation steps, then in the *second stage* a 2 ns long MD simulation was run with positional restraints of  $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  on protein and lipid phosphorus atoms. In the *third stage*, a 20 ns of MD simulation was performed retraining only the protein atoms and, finally, in the *fourth stage*, positional restraints were applied only to the protein backbone alpha carbons, for a further 30 ns.

Production trajectories were computed with an integration time step of 4 fs in the canonical ensemble (NVT) at 300 K, using a thermostat damping of  $0.1 \text{ ps}^{-1}$  and the M-SHAKE algorithm (Kräutler *et al.*, 2001) to constrain the bond lengths involving hydrogen atoms. The cut-off distance for electrostatic interactions was set at 9 Å, with a switching function applied beyond 7.5 Å. The Particle Mesh Ewald (PME) summation method (Essmann *et al.*, 1995) was employed to handle long-range Coulomb interactions with a mesh spacing to 1.0 Å.

### 2.3.7 MD analysis

Atomic contacts, hydrogen bonds, Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), distances between atoms, and dihedral angles were computed using VMD (Humphrey *et al.*, 1996) and PLUMED 2 (Tribello *et al.*, 2014). A contact was considered productive if the distance between two atoms was lower than 3.5 Å. For hydrogen bond detection, a strict donor-acceptor distance of 3 Å and an angle value of 160° were set as geometrical cut-offs when using VMD, and a more generous 3.5 Å and 150° when using a GetContacts python tool (<https://getcontacts.github.io/>).

The GBSA analysis were computed using the MMPBSA.py (Miller *et al.*, 2012) script, a program released with the open source AmberTools17 package (The Amber Molecular Dynamics Package, at <http://ambermd.org/>) after transforming the CHARMM *psf* topology files to an Amber *prmtop* format using ParmEd (ParmEd documentation at <http://parmed.github.io/ParmEd/html/index.html>).

The NAMD Energy calculations were calculated using NAMDenergy VMD plugin (<https://www.ks.uiuc.edu/Research/vmd/plugins/namdenergy/>).

### 2.3.8 General

Throughout the text, the Ballesteros-Weinstein GPCR universal numbering (Isberg *et al.*, 2015) is used alongside the normal residue numbers.

## 3. Allosteric effects in the free fatty acid receptor 1 (FFAR1)

### 3.1 Abstract

**Background:** The free fatty acid receptor 1 (FFAR1), also known as GPR40, is a G protein-coupled receptor (GPCR) that binds long-chain free fatty acids (LCFA) to enhance glucose-dependent insulin secretion. FFAR1 has an important role in the treatment of type 2 diabetes mellitus (T2DM). Therefore, studies based on a sequence and structure analysis of this receptor may help to develop new drugs for T2DM with improved metabolic profiles. **Aim:** The purpose of this research is: (i) to analyse the allosteric effect in FFAR1, whose binding site is located in a different position in FFAR1 compared to other GPCRs (Lu *et al.*, 2017), and to observe its possible existence in other receptors (I have worked also on the allosteric effect in Androglobin (Adgb) see Appendix, section B); (ii) to examine the conserved motifs, with particular attention given to residues of helix 8 (H8), which is a structurally conserved amphipathic helical motif in GPCRs that however is missing in all FFAR1 X-ray structures; (iii) to analyse the allosteric communication between binding sites from a dynamic point of view. **Methods:** Multiple sequence alignments, BLASTp search, and ConSurf analysis were conducted to analyse the allosteric binding site A2 across the class A GPCRs. FFAR1 has been computationally modelled with the addition of H8 and the time evolution of the apo- and the holo-protein system has been observed by means of molecular dynamics (MD) simulations (number of replicas: 3; time: 1  $\mu$ s; timestep: 4 fs). **Results and Conclusion:** According to a homology model of the Adenosine receptor A1 (A<sub>1</sub>AR), it has been proposed that this alternative binding site

A2 may exist in the A<sub>1</sub>AR, and by implication in other GPCRs. Moreover, the ConSurf results have shown that the residues implicated in the allosteric binding site A2 are generally very well conserved. MD simulations of FFAR1 have shown that the apo form is structurally more unstable than the holo form and this is in line with the absence of the crystal structure of FFAR1 apo form. Therefore, the ligand in allosteric binding site A2 may better stabilise the whole receptor structure including H8 that seems to keep part of its helicity although it remains very flexible during simulations due to the presence of several glycine residues in the helix. Moreover, MD simulations suggest that the presence of the allosteric ligand AP8 stabilises the partial agonist MK-8666 since the energy analysis of the latter shows a shift to lower energy values in presence of AP8. **Keywords:** G protein-coupled receptors (GPCRs); FFAR1; GPR40; Allosteric binding site A2; Molecular dynamics (MD) simulations.

### 3.2 Introduction

Free fatty acid receptor 1 (FFAR1 or GPR40) (Srivastava *et al.*, 2014) is a G protein-coupled receptor (GPCR) that binds long-chain free fatty acids (LCFA) to enhance glucose-dependent insulin secretion, directly and mediating a major part of the dietary triglyceride-induced secretion of the incretins Glucagon-like peptide 1 (GLP-1) and Gastric inhibitory polypeptide (GIP) (Ho *et al.*, 2018). Studies on this receptor have an important role in the comprehension of the treatment of type 2 diabetes mellitus (T2DM) that is the most common form of diabetes, accounting for around 90% of the total diabetes cases in the world (World Health Organization, 2016a). The prevalence of T2DM has risen dramatically in the past three decades and, as reported by the World Health Organization (WHO), in 2014 there were 422 million adults with diabetes and 1.6 million deaths were directly attributed to diabetes each year (World Health Organization, 2016b).

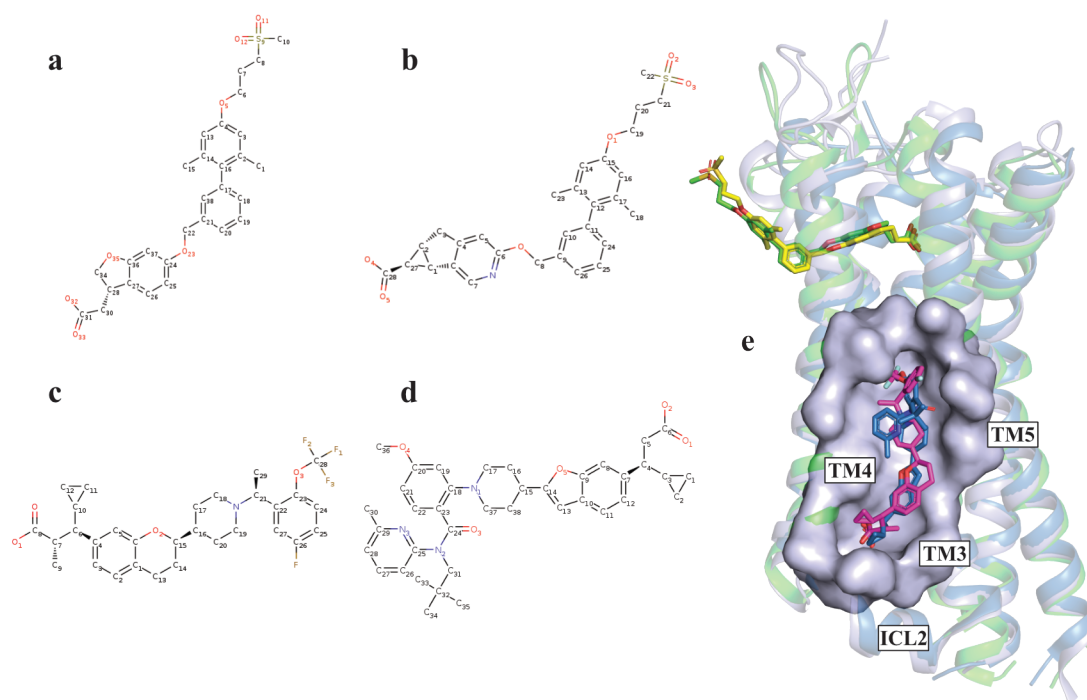
Currently, there are several clinical treatments of T2DM available on the market but many of them are associated with negative side effects (*i.e.* hypoglycaemia and weight gain) (Burant *et al.*, 2012). In this regard, studies of ligands that bind to allosteric sites instead of the orthosteric site of FFAR1 have drawn a considerable attention because the parallel activation of an alternative signalling pathway may avoid detrimental side effects. Two distinct allosteric binding sites are present on FFAR1: namely A1 (where ligands such as TAK-875 and MK-8666 bind) and A2 (where ligands such as AP8 and Compound 1 bind - Table 3.1 and Figure 3.1). The binding cooperativity between the partial agonist MK-8666 and the allosteric ligand AP8, changes the receptor into a more active-like state (Lu *et al.*, 2017). Further studies about



the position of the allosteric binding site A2 could reveal that it is not unique to this receptor, but it could be also relevant for other receptors drug design.

**Table 3.1 | Ligands engaged with allosteric binding sites A1 and A2 in FFAR1 structures.** The 3-character ligand names are taken from the relevant PDB files

FFAR1	Ligands in allosteric binding site A1	Ligands in allosteric binding site A2	References
4PHU	TAK-875 (2YB)	-	(Srivastava et al., 2014)
5TZR	MK-8666 (MK6)	-	(Lu et al., 2017)
5TZY	MK-8666 (MK6)	AP8 (7OS)	(Lu et al., 2017)
5KW2	-	Compound 1 (6XQ)	(Ho et al., 2018)



**Figure 3.1 | Allosteric binding sites A1 and A2 in FFAR1.** Chemical structure of: (a) TAK-875 (2YB), (b) MK-8666 (MK6), (c) AP8 (7OS), and (d) Compound 1(6XQ). (e) Structural superposition of FFAR1 structures (shown in cartoon representation) in complex with allosteric binding ligands (shown in stick representation): 4PHU in complex with TAK-875 (2YB) in A1 are in green, 5TZY is in light-blue in complex

with MK-8666 (MK6) in A1 shown in yellow and AP8 (7OS) in A2 shown in magenta and, lastly, 5KW2 in complex with Compound 1 (6XQ) in A2 are in sky-blue. The binding pocket A2 is shown as a surface representation.

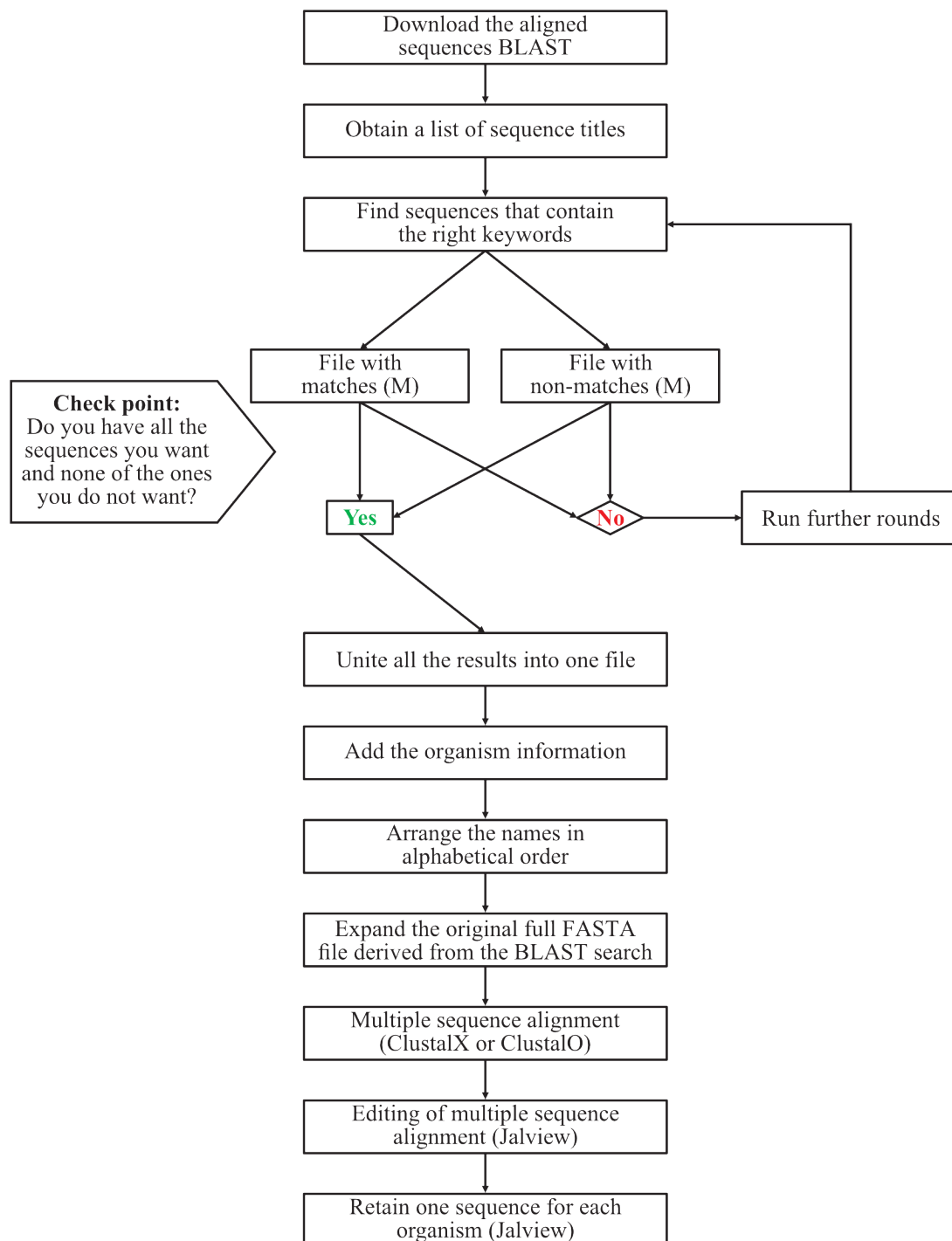
The second part of this project is based on the study of H8 in FFAR1 that is missing in all FFAR1 X-ray structures (PDB ID: 4PHU, 5TZR, 5TZY, and 5KW2). H8 generally consists of 3-4  $\alpha$ -helix turns and it is contained in the C-terminus that runs parallel to the membrane (Figure 1.3). The amphipathic structure of H8 is highly conserved in GPCRs, suggesting an important role for this helix in GPCR signalling, which is sustained by previous studies that described H8 involved in several cellular processes such as G-protein coupling (Ernst *et al.*, 2000), regulation of activation (Delos Santos *et al.*, 2006), receptor expression (Tetsuka *et al.*, 2004), and internalisation (Aratake *et al.*, 2012).

Finally, the last part of this project shows the MD simulations of the apo- and the holo- FFAR1 system.

### 3.3 Materials and Methods

#### 3.3.1 Sequence analysis

Multiple sequence alignments were generated to investigate possible evolutionary evidence for allosteric binding sites. Protein Basic Local Alignment Search Tool (BLASTp) (Altschul *et al.*, 1990) was used to search protein queries (UniProtKB entry name: FFAR1\_HUMAN, C5AR1\_HUMAN, and ADRB2\_HUMAN) against a protein sequence database. The BLASTp search settings used are shown as follows: (i) *Database*: non-redundant protein sequences (nr); (ii) *Algorithm*: blastp (protein-protein BLAST); (iii) *Max target sequences*: 20000; (iv) *Expect threshold*: 0.0001; (v) *Word size*: 3, and (vi) *Matrix*: BLOSUM62. The aligned sequences, obtained as BLASTp results, were gradually sifted by keywords in order to obtain the sequences truly related to each query, as specified in the annotation. The diagram below shows the workflow of the sifting BLAST searches by keywords (Figure 3.2). Multiple sequence alignment programs, such as ClustalX (Larkin *et al.*, 2007) and ClustalO (Sievers *et al.*, 2011), were used to align the sequences obtained at the end of the process and the final file was edited using Jalview to correct the automatic alignment (Waterhouse *et al.*, 2009). Lastly, the ConSurf web server (<https://consurf.tau.ac.il/>) (Ashkenazy *et al.*, 2010) was used to calculate the evolutionary conservation of amino acid positions from the multiple sequence alignment file for each initial query by means of the empirical Bayesian or Machine learning (ML) algorithms (Pupko *et al.*, 2002; Mayrose *et al.*, 2004).



**Figure 3.2 | The steps in sifting BLAST searches by keywords.**

### 3.3.2 Molecular Dynamics (MD) simulations

#### i) System preparation for MD simulations

In this study, the MD systems for FFAR1 structure (modelled active conformation, MAC) were prepared using 5TZY (another intermediate FFAR1 structure) as a representative structure, as explained in section 2.3.4. The disulfide bond between residues C79<sup>3.25</sup> and C170<sup>45.50</sup> was included. Table 3.2 summarises all the simulations performed.

**Table 3.2 | Summary of all the MD simulations performed on the FFAR1.** For each system, 3 replicas of a 1  $\mu$ s long MD simulation were performed

System	System details	Number of replicas	Total MD sampling time [ $\mu$ s]
apoFFAR1	<b>Receptor:</b> ClassA_ffar1_human_Active_5XSZ_2018-07-10_GPCRDB.pdb*	3	3
FFAR1:MK-8666	<b>Receptor:</b> ClassA_ffar1_human_Active_5XSZ_2018-07-10_GPCRDB.pdb* <b>Ligand:</b> MK-8666 (MK6)	3	3
FFAR1:AP8	<b>Receptor:</b> ClassA_ffar1_human_Active_5XSZ_2018-07-10_GPCRDB.pdb* <b>Ligand:</b> AP8 (7OS)	3	3
FFAR1:MK-8666:AP8	<b>Receptor:</b> ClassA_ffar1_human_Active_5XSZ_2018-07-10_GPCRDB.pdb* <b>Ligand 1:</b> MK-8666 (MK6) <b>Ligand 2:</b> AP8 (7OS)	3	3

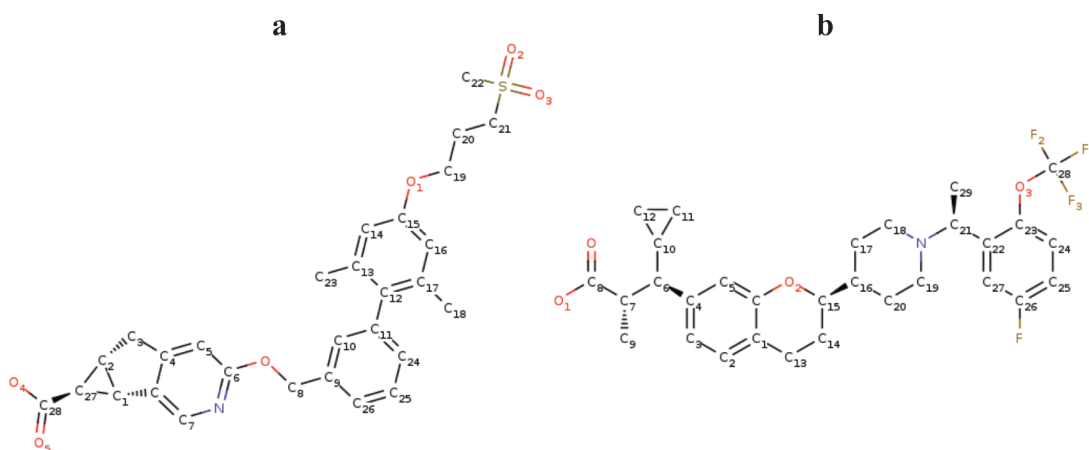
(\*) The FFAR1 structure modelled active conformation (MAC) was generated by GPCRdb (<https://www.gpcrdb.org/>) (Isberg *et al.*, 2016)

#### ii) Ligand parametrisation

The partial agonist agonist MK-8666 (MK6) (IUPAC name (5aR,6S,6aS)-3-({2',6'-dimethyl-4'-[3-(methylsulfonyl)propoxy][1,1'-biphenyl]-3-yl}methoxy)-5,5a,6,6a-tetrahydrocyclopropa[4,5]cyclopenta[1,2-c]pyridine-6-carboxylic acid) and the AP8 (7OS) (IUPAC name (2S,3R)-3-cyclopropyl-3-[(2R)-2-(1-((1S)-1-[5-fluoro-2-(trifluoromethoxy)phenyl]ethyl)piperidin-4-yl)-3,4-dihydro-2H-1-benzopyran-7-yl]-2-methylpropanoic acid) were parameterised in the CGenFF (Vanommeslaeghe

and MacKerell, 2012; Vanommeslaeghe *et al.*, 2012) force field, which is designed to be combined with the CHARMM36 force field (Huang and MacKerell, 2013), through a stepwise procedure as explained in section 2.3.5.

The ligands MK-8666 (Figure 3.3a) and AP8 (Figure 3.3b) were parameterised in the CGenFF, with the exception of the AP8 rotatable bonds C<sub>23</sub>-O<sub>3</sub> and the MK-8666 rotatable bond involving the carboxylic acid and the cyclopropyl ring that were optimised using HTMD and the ffTK, respectively.



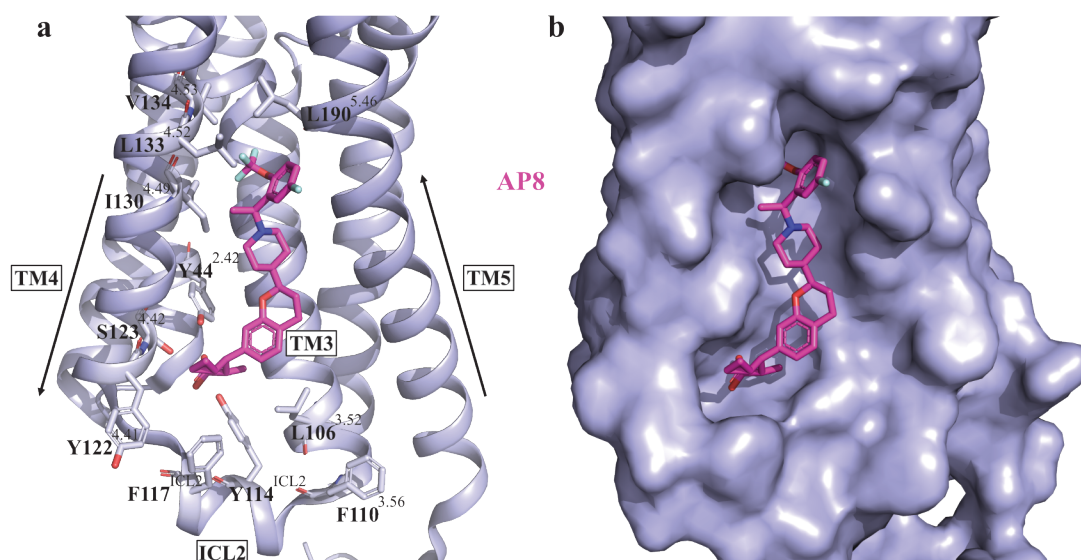
**Figure 3.3 | The chemical structure of the MK-8666 (MK6) (a) and AP8 (7OS) (b).**

The MD system equilibration and production stages are reported along with the analysis performed in section 2.3.6 and 2.3.7, respectively.

### 3.4 Results and Discussion

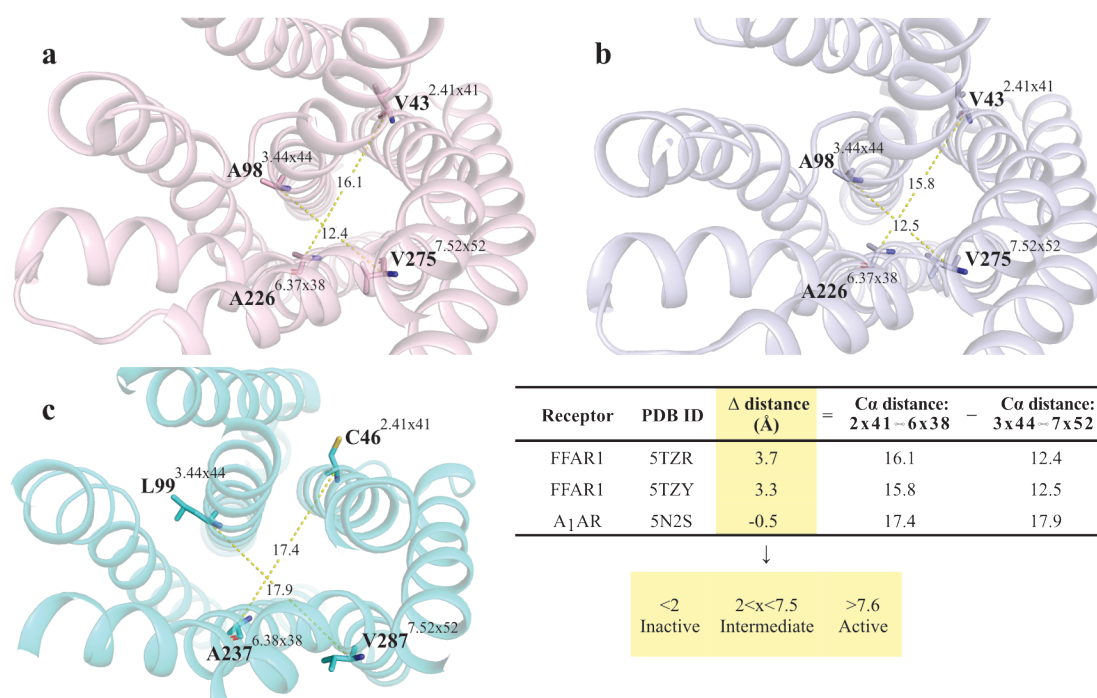
#### 3.4.1 Allosteric binding site A2: does it exist in GPCRs other than FFAR1?

Two inactive structures of the Adenosine receptor A1 (A<sub>1</sub>AR) (PDB ID: 5N2S, 5UEN) were analysed to evaluate the possibility that the allosteric binding site A2 is present in other GPCRs. It is known that the AP8 binding causes rearrangements of the transmembrane (TM) helices TM4 and TM5, and transition of the ICL2 into a short helix in FFAR1 (PDB ID: 5TZY), inducing the receptor to assume a more active-like state (Lu *et al.*, 2017) (Figure 3.4). Therefore, to compare intermediated-active FFAR1 with the inactive A<sub>1</sub>AR (Figure 3.5), a homology model of the A<sub>1</sub>AR was created without including TM4 and TM5 in the structural alignment. These two helices were deleted from 5TZY, whose structure was superimposed on that of A<sub>1</sub>AR (for both 5N2S and 5UEN structures) using two different programs such as UCSF Chimera 1.12 and Modeller 9.19. The final model for each structure (5N2Sm and 5UENm) was compared with the original 5TZY (Figure 3.6).



**Figure 3.4 | Cartoon (a) and surface (b) representation of the allosteric binding site of FFAR1 with AP8 (purple) (PDB ID: 5TZY).** AP8-binding site can be divided into three distinct regions: i) *hydrophobic pocket*: intersection of TM3, TM4, and TM5 residues, *i.e.* Ile130<sup>4.49</sup>, Leu133<sup>4.52</sup>, Val134<sup>4.53</sup>, and Leu190<sup>5.46</sup>; ii) *polar cavity*: key interactions include hydrogen bonds with the side chains of Tyr44<sup>2.42</sup> and Ser123<sup>4.42</sup>; iii) *intracellular environment*: hydrophobic and aromatic residues, *i.e.* Leu106<sup>3.52</sup>, Phe110<sup>3.56</sup>, Tyr114<sup>ICL2</sup>, Phe117<sup>ICL2</sup>, and Tyr122<sup>4.41</sup> (Lu *et al.*, 2017).



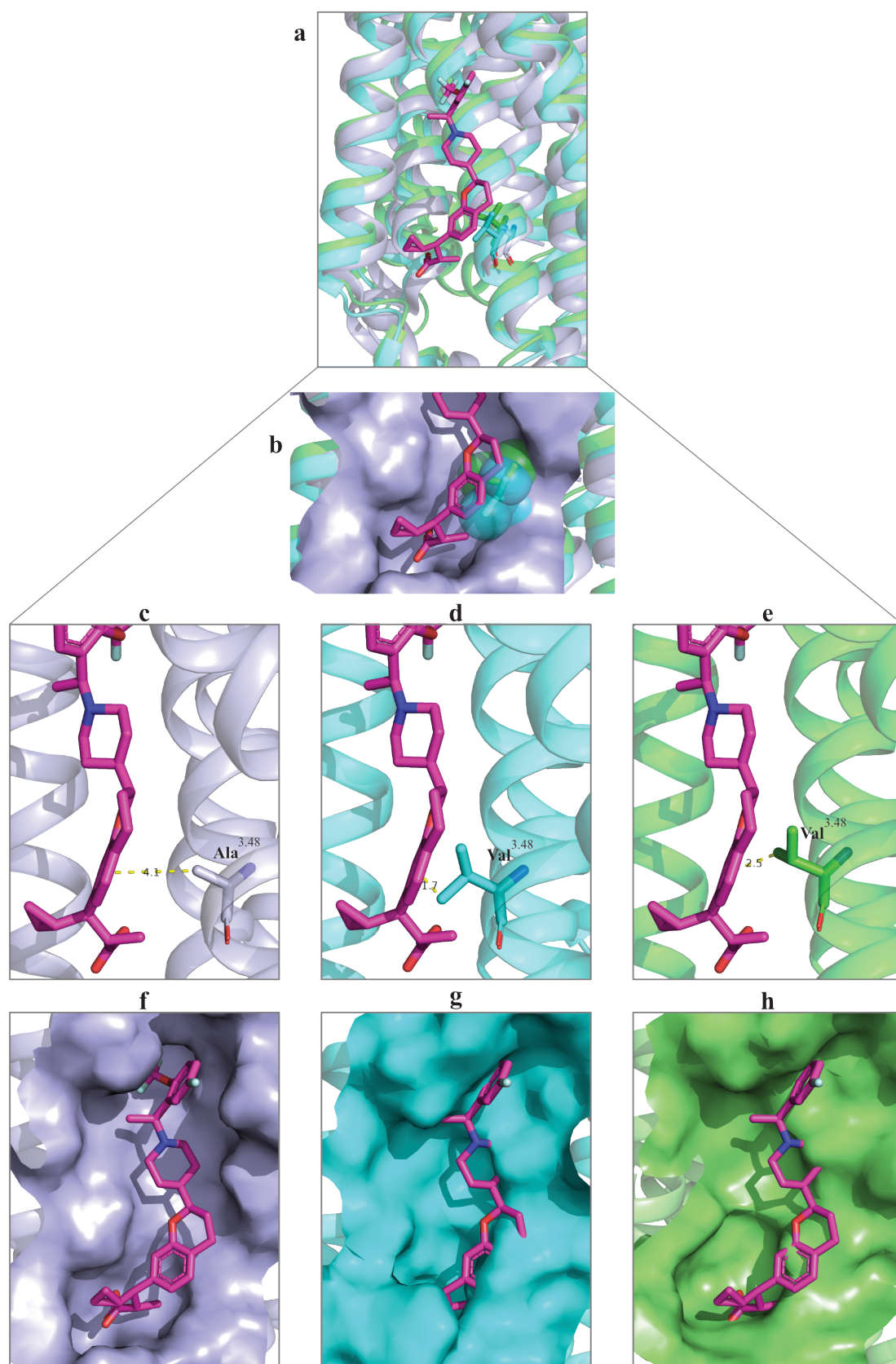


**Figure 3.5 | Analysis of the  $\Delta$  distance.** In this figure, all the receptors are shown in cartoon representation, respectively 5TZR (FFAR1) in light pink (a), 5TZY (FFAR1) in light blue (b), and 5N2S (A<sub>1</sub>AR) in cyan (c). Important residues are shown in stick representation. Two C $\alpha$ -distances between residues 2x41-6x38 and 3x44-7x52 has been analysed for each receptor (yellow dashed lines) to calculate the  $\Delta$  distance, determined by subtracting the 3x44-7x52 C $\alpha$ -distance from the 2x41-6x38 C $\alpha$ -distance, as shown in the Table. The value of this distance gives an indication of the state of the receptor as follow: <2, Inactive; 2<x<7.5, Intermediate; >7.6, Active (Isberg *et al.*, 2016).

An important residue that permits the allosteric agonist AP8 to bind is the one that precedes the D/ERY<sup>3.51</sup> motif in TM3 (GRY<sup>3.51</sup> motif in FFAR1), namely Ala<sup>3.48(102)</sup>, which is situated close to the middle region of the binding pocket. However, through the analysis conducted in the paper Lu *et al.*, 2017, it is interesting to note that the corresponding position 3.48 in the same receptor but from a different organism such as dog FFAR1 (UniProtKB entry name: E2RC57\_CANLF) is occupied by valine, and steric constraints on the binding of AP8 could be caused by its bulkier side chain (Lu *et al.*, 2017). Analogously with the dog FFAR1, the structural superposition of the

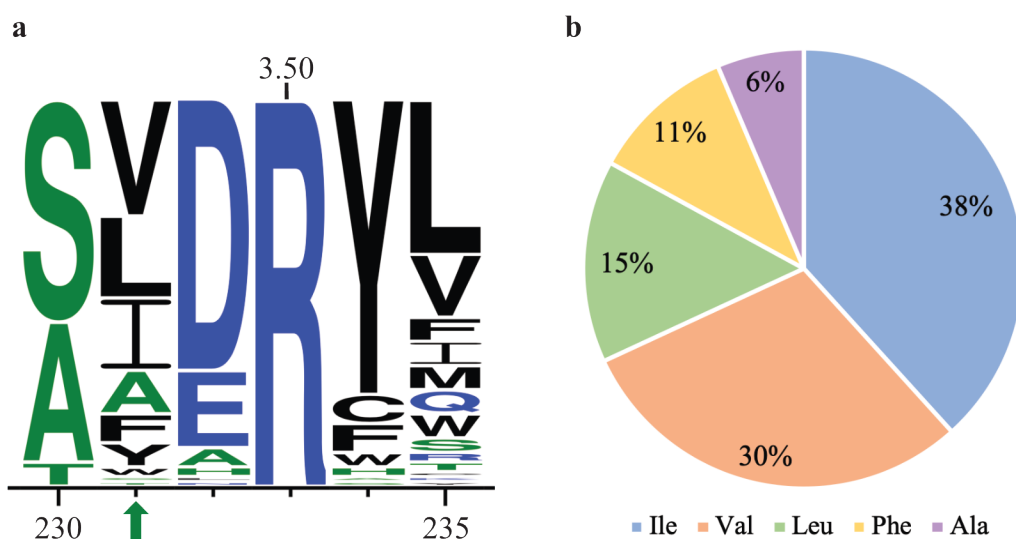
homology model of the A<sub>1</sub>AR with human FFAR1 shows a Val<sup>3.48(103)</sup> instead of Ala<sup>3.48(102)</sup> (Figure 3.6a,b). The distance between the side chain of the valine in A<sub>1</sub>AR and the ligand in the allosteric binding site A2 is shorter than the one in FFAR1. Figure 3.6c-e shows the following distances between the receptor and the ligand: 4.1 Å in 5TZY (light blue), 1.7 Å in 5N2Sm (cyan), and 2.5 Å in 5UENm (green). The longer valine side chain occupies the binding region of the AP8 aromatic ring, decreasing its binding affinity of about 10 fold (Lu *et al.*, 2017). However, the surface representation in Figure 3.6f-h shows the possible existence in the A<sub>1</sub>AR of an allosteric binding site analogue to the A2. A deeper analysis of this important position 3.48 was conducted in this project. Multiple sequence alignment of class A GPCRs has revealed a 30% of consensus for Val<sup>3.48</sup> followed mainly by Leu, Ile, Ala, and Phe residue (Figure 3.7a). Contrariwise, structure-based alignment of TM3 class A GPCRs, generated by GPCRdb (<https://www.gpcrdb.org/>) (Isberg *et al.*, 2016), has revealed a higher consensus (37%) for Ile<sup>3.48</sup> followed by Val (30%), Leu (15%), Phe (11%), and Ala (6%) (Figure 3.7b). Nevertheless, the two different alignments show the same top 3 residues (Ile, Leu, and Val) even if in a different order, indicating hydrophobic and aliphatic residues are privileged in this position.

In conclusion, the variant Ala<sup>3.48</sup>Val/Ile in the allosteric site could be overcome with a ligand optimisation, for example with the substitution of the rigid aromatic ring with a less rigid moiety (see Figure 3.6).



**Figure 3.6 | Analysis of allosteric binding site A2 and structural comparison between FFAR1 (PDB ID: 5TZY) and the homology model of the A<sub>1</sub>AR (5N2Sm,**

**5UENm**). In this figure, all the receptors are shown in cartoon representation, respectively 5TZY in light blue, 5N2Sm in cyan, and 5UENm in green. Residues in position 3.48 are shown in stick representation, same as the AP8 coloured in magenta. (a) Structural superposition of 5TZY, 5N2Sm, and 5UENm shows a valine in A<sub>1</sub>AR instead of an alanine in position 3.48 that causes steric constraints highlighted by spheres representation of these residues (b). Slices (c), (d), and (e) show the distance between Ala/Val<sup>3.48</sup> and the aromatic ring of AP8: 4.1 Å (in 5TZY), 1.7 Å (in 5N2Sm), and 2.5 Å (in 5UENm). In the slices (f), (g), and (h), the binding pocket A2 is shown as a surface representation.



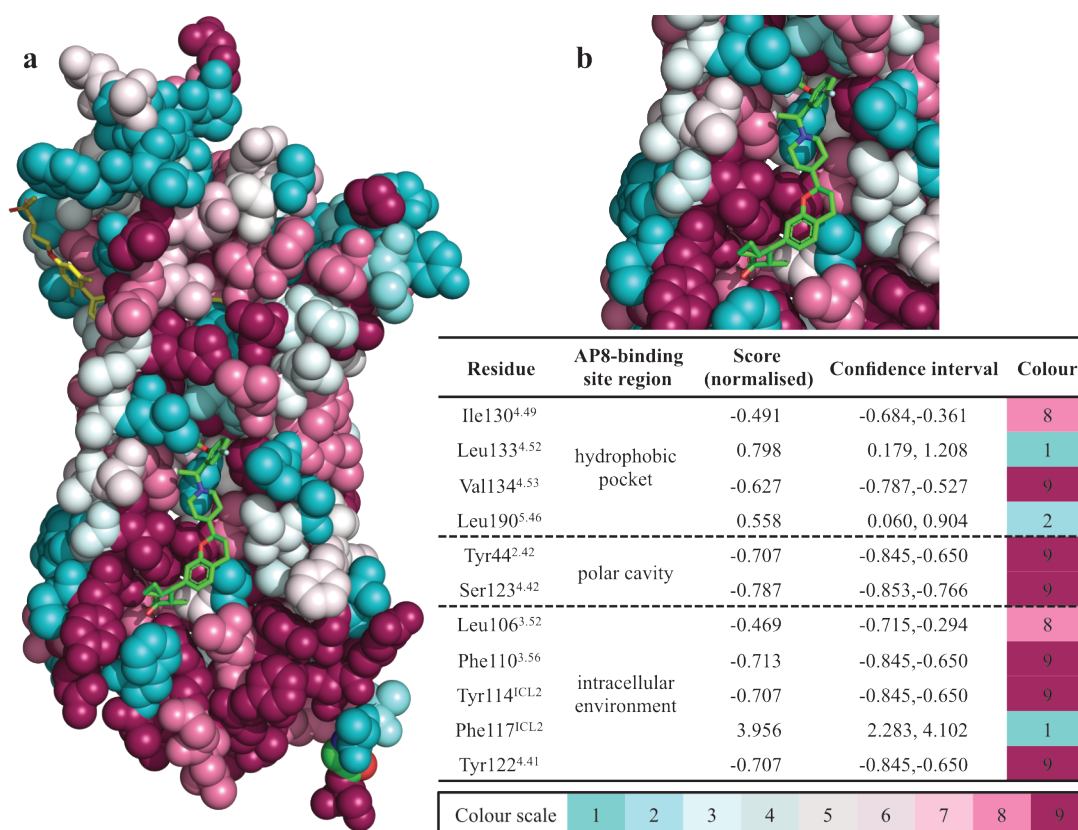
**Figure 3.7 | Analysis of position 3.48 in class A GPCRs.** (a) WebLogo representation (<http://weblogo.threeplusone.com/create.cgi>) (Crooks *et al.*, 2004) of multiple sequence alignment of class A GPCRs, generated by GPCRdb (<https://www.gpcrdb.org/>). The position 3.48 (indicated by an arrow) shows on the top a Val with 30% of consensus. (b) Pie chart of residue occurrence in position 3.48, obtained from a structure-based alignment of TM3 in class A GPCRs. The chart shows an Ile with 38% of consensus, followed by Val (30%), Leu (15%), Phe (11%), and Ala (6%).

Recently studies have revealed that the Beta-2 adrenergic receptor ( $\beta_2$ -AR) (Liu *et al.*, 2017) and C5a anaphylatoxin chemotactic receptor 1 (C5aR1) (Robertson *et al.*, 2018) contain an allosteric binding site A2, similar to that in FFAR1. A BLASTp search for these three receptors has been conducted in this project and the total number of hits obtained are as follow: 10555, FFAR1; 20000,  $\beta_2$ -AR; 20000, C5aR1. The aligned sequences were sifted according to a set of keywords, revealing that only the top 2% of the sequences are truly related to each query. A multiple sequence alignment of these related sequences was obtained using ClustalO (Sievers *et al.*, 2011) and edited using Jalview (Waterhouse *et al.*, 2009). Lastly, the evolutionary conservation of amino acid positions in these receptors was calculated using the ConSurf web server (<https://consurf.tau.ac.il/>) (Ashkenazy *et al.*, 2010). Figure 3.8 shows the output of FFAR1 ConSurf results using the multiple sequence alignment (obtained from the sifting BLAST searches by keywords) as input; contrariwise, Figure 3.9 shows the output of ConSurf results without using multiple sequence alignment as input. In this case only the FFAR1 sequence (FFAR1\_HUMAN) was submitted, and ConSurf obtained additional sequence for the multiple sequence alignment using HHpred (Söding *et al.*, 2005). Both searches used 5TZY as the know protein structure. The following parameters were calculated: (i) *Score*: the normalised conservation scores (the larger the value, the greater the mutation rate); (ii) *Confidence interval*: when using the Bayesian method for calculating rates, a confidence interval is assigned to each of the inferred evolutionary conservation scores, and (iii) *Colour*: the colour scale representing the conservation scores (9 - conserved, 1 - variable). The results reveal that the residues implicated in the AP8-binding site (shown in Figure 3.4) are generally very well conserved when the search is based on the multiple sequence alignment (Figure 3.8), *i.e.* based only on the FFAR1 family, while the same residues are more

variable in the second case where all suitable sequences from the HHpred search were included (Figure 3.9).

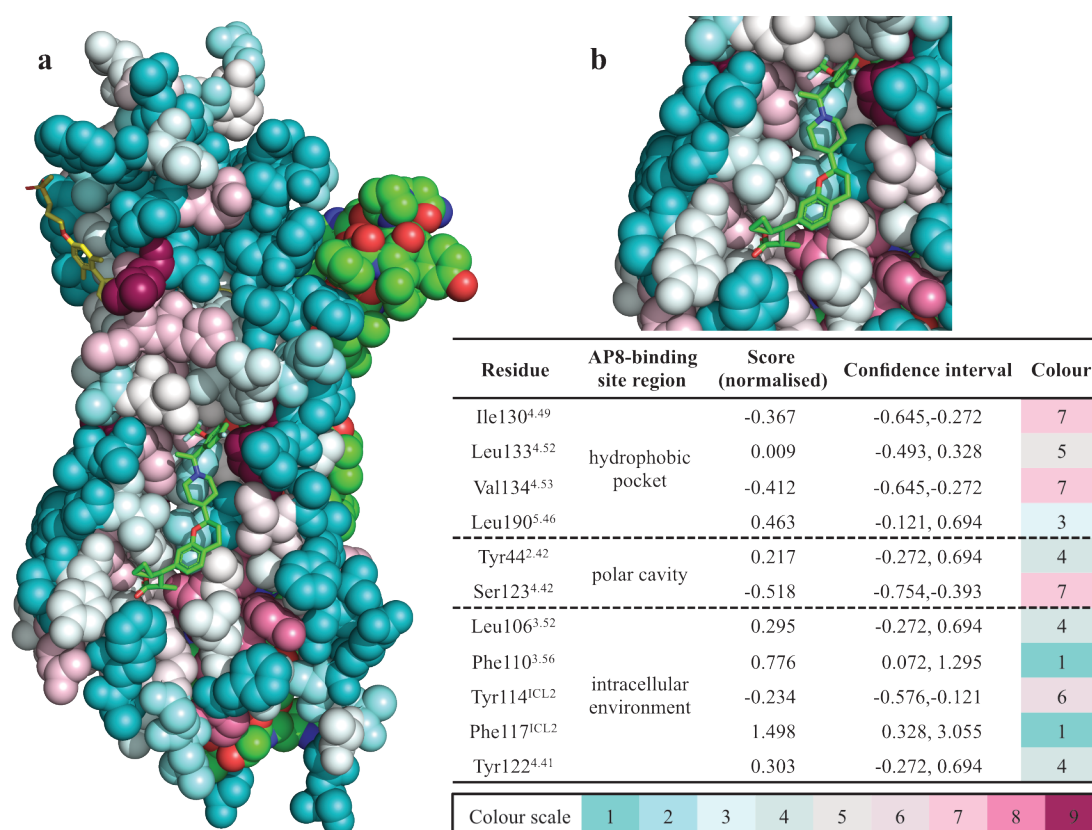
Preliminary results from our group highlight a similar conserved pattern also for the Beta-2 adrenergic receptor and for the adenosine A1 receptor.

Taken together, these results show that the binding site A2 exists not only in FFAR1 but also in  $\beta_2$ -AR, A<sub>1</sub>AR and by implication in other GPCRs. Furthermore, the use of the ConSurf web server with multiple sequence alignment as input has proven to be a powerful bioinformatics tool for estimating evolutionary conservation of amino acid positions in a protein.



**Figure 3.8 | Analysis of evolutionary conservation of amino acid positions in FFAR1 (PDB ID: 5TZY) calculated using ConSurf web server (<https://consurf.tau.ac.il/>) (Ashkenazy *et al.*, 2010) with a multiple sequence alignment as input.** The receptor is shown in spheres representation with a colour scale based on the conservation scores (9 - conserved, 1 - variable). AP8 (in allosteric bind site A2) and MK-8666 (in allosteric bind site A1) are shown in stick representation, respectively in green and yellow. **(a)** whole receptor and **(b)** detail of AP8-binding site show the high evolutionary conservation of residues implicated in the AP8 binding. The Table reports the (i) score computed using the empirical Bayesian or ML algorithms (Pupko *et al.*, 2002; Mayrose *et al.*, 2004), (ii) the confidence interval, and (iii) the colour for all these residues.





**Figure 3.9 | Analysis of evolutionary conservation of amino acid positions in FFAR1 (PDB ID: 5TZY) calculated using ConSurf web server (<https://consurf.tau.ac.il/>) (Ashkenazy *et al.*, 2010) without a multiple sequence alignment as input.** The receptor is shown in spheres representation with a colour scale based on the conservation scores (9 - conserved, 1 - variable). AP8 (in allosteric bind site A2) and MK-8666 (in allosteric bind site A1) are shown in stick representation, respectively in green and yellow. (a) whole receptor and (b) detail of AP8-binding site show a variable evolutionary conservation of residues implicated in the AP8 binding. The Table reports the (i) score computed using the empirical Bayesian or ML algorithms (Pupko *et al.*, 2002; Mayrose *et al.*, 2004), (ii) the confidence interval, and (iii) the colour for all these residues.



### 3.4.2 Analysis of missing H8 in FFAR1 crystal structures

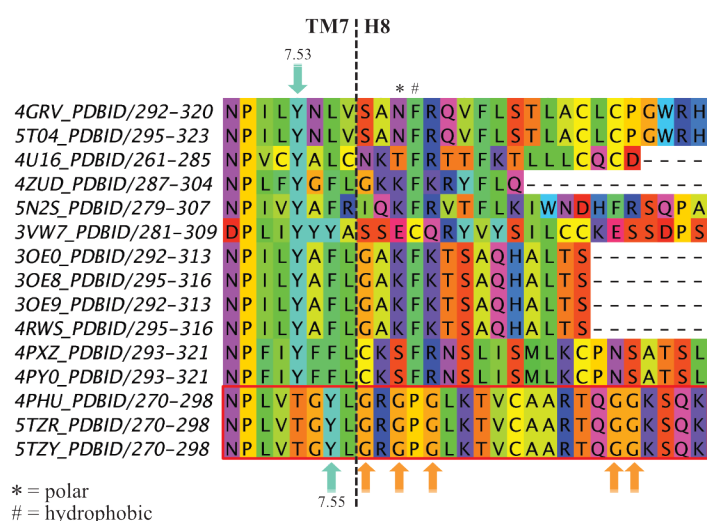
Sequences of class A GPCRs were analysed in order to understand the causes of missing H8 in FFAR1. H8 may play a role in receptor activation and may also interact with the G protein or other effectors; it may also play a role in biased signalling. For this reason, it is important to pay significant attention to its location. To date, 15 structures of class A GPCRs have been experimentally determined without the H8. Particular attention was given to three high conserved motifs of these receptors: the NPxxY<sup>7.53</sup> motif in TM7, whose tyrosine (Y<sup>7.53</sup>) forms a hydrophobic cluster with leucine (L<sup>1.63</sup>) of the KKLH<sup>1.64</sup> motif in ICL1 and the phenylalanine (F<sup>8.50</sup>) of the EFxxxL<sup>8.54</sup> motif in helix 8 (H8) (Vohra *et al.*, 2013), and also the positive residue at position 1.61 or 1.62 and the negative residues at position 8.49. It has been hypothesised that H8 loses its helical character when charged residues are located at position 8.49 and one of 1.61 or 1.62. The residues corresponding to the conserved motifs mentioned above are reported in Table 3.3 for every receptor with a missing H8.

With regard to FFAR1, Table 3.3 shows the lack of tyrosine in the NPxxY<sup>7.53</sup> motif in TM7. The hypothesis that the Y<sup>7.53</sup> is missing in FFAR1 has been analysed. Multiple sequence alignment of these sequences without H8 and all sequences of class A GPCRs was made by using ClustalO. The analysis of the results (shown in Figure 3.10) reveals an NPxxxxY<sup>7.55</sup> motif in TM7 for FFAR1 (two amino acids longer than the common motif). The question that has been raised is whether Y<sup>7.53</sup> is missing in FFAR1, it has moved forward two positions or has been replaced by T<sup>7.53</sup>.

**Table 3.3 | Conserved motifs of ICL1, TM7, and H8 in class A GPCRs structures with missing H8**

PDB ID	Receptor	KKLH <sup>1.64</sup> motif (ICL1)	NPxxY <sup>7.53</sup> motif (TM7)	EFxxxL <sup>8.54</sup> motif (H8)
5N2S	A <sub>1</sub> AR	QALR <sup>41</sup>	NPxxY <sup>288</sup>	KFxxxF <sup>299</sup>
4U16	M3R	KQLK <sup>99</sup>	NPxxY <sup>543</sup>	TFxxxF <sup>554</sup>
4GRV	NTR1	KSLQ <sup>96</sup>	NPxxY <sup>369</sup>	NFxxxF <sup>380</sup>
5T04	NTR1	KSLQ <sup>96</sup>	NPxxY <sup>369</sup>	NFxxxF <sup>380</sup>
3OE0	CXCR4	KKLR <sup>70</sup>	NPxxY <sup>302</sup>	KFxxxA <sup>313</sup>
3OE8	CXCR4	KKLR <sup>70</sup>	NPxxY <sup>302</sup>	KFxxxA <sup>313</sup>
3OE9	CXCR4	KKLR <sup>70</sup>	NPxxY <sup>302</sup>	KFxxxA <sup>313</sup>
4RWS	CXCR4	KKLR <sup>70</sup>	NPxxY <sup>302</sup>	KFxxxA <sup>313</sup>
4ZUD	AT1R	MKLK <sup>60</sup>	NPxxY <sup>302</sup>	KFxxxF <sup>313</sup>
4PXZ	P2Y <sub>12</sub> R	IRSK <sup>56</sup>	D*PxxY <sup>298</sup>	SFxxxL <sup>309</sup>
4PY0	P2Y <sub>12</sub> R	IRSK <sup>56</sup>	D*PxxY <sup>298</sup>	SFxxxL <sup>309</sup>
3VW7	PAR1	MKVK <sup>134</sup>	DPxxY <sup>371</sup>	ECxxxV <sup>382</sup>
4PHU	FFAR1	ARLR <sup>37</sup>	NPxxT <sup>276</sup>	GPxxxT <sup>287</sup>
5TZR	FFAR1	ARLR <sup>37</sup>	NPxxT <sup>276</sup>	GPxxxT <sup>287</sup>
5TZY	FFAR1	ARLR <sup>37</sup>	NPxxT <sup>276</sup>	GPxxxT <sup>287</sup>

(\*) Engineered mutation N294D

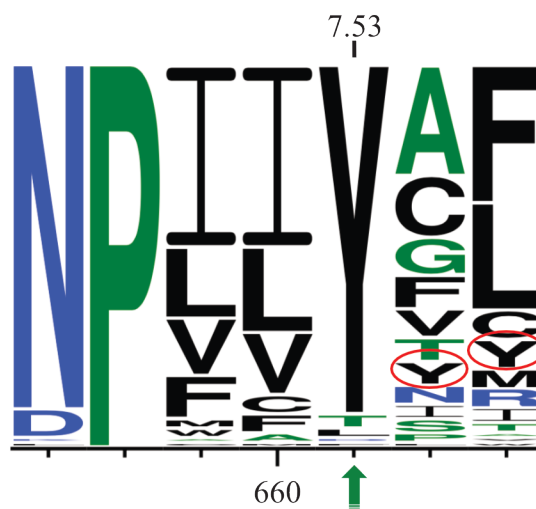


**Figure 3.10 | Detail of the multiple sequence alignment between class A GPCRs.**

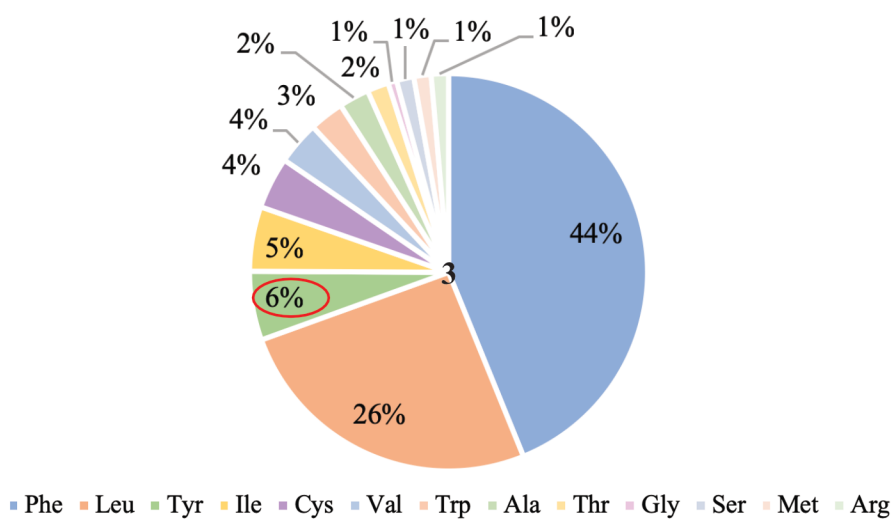
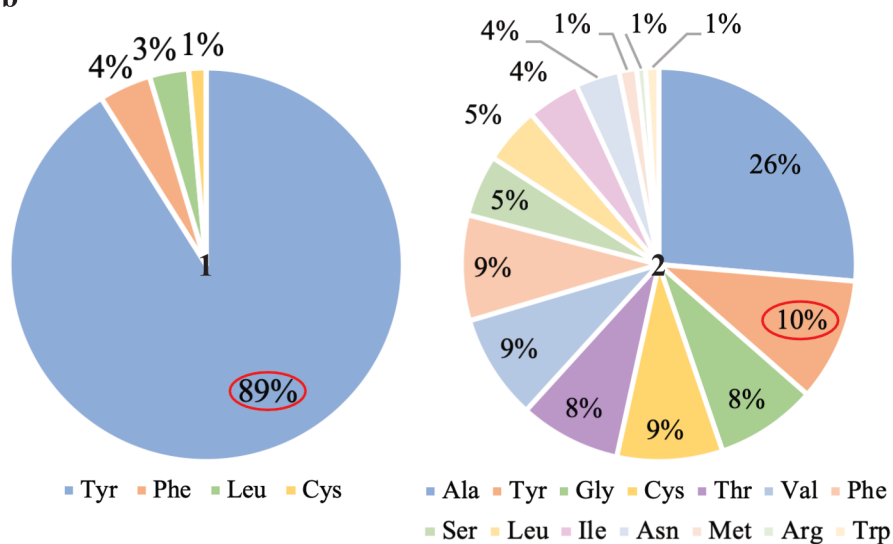
The arrow on the top indicates the tyrosine in position 7.53 in the NP<sup>7.50</sup>XXY motif, which is very high conserved across the alignment but not in the sequences of FFAR1 (highlighted in the red box), where the arrow on the bottom indicates another tyrosine in position, namely 7.55. The C-terminal segment of the transmembrane 7 (TM7) and the helix 8 (H8) of FFAR1 are characterised by numerous glycine residues (orange arrows). \* and # indicate respectively the position 8.49 and 8.50 normally filled by a polar and a hydrophobic residue in all class A GPCRs; contrariwise, FFAR1 shows a glycine and a proline residue. Residues are coloured in the Taylor scheme, as implemented in Jalview (Clamp *et al.*, 2004), where the colour reflects their biophysical properties as follows: blue, positive; red, negative or small polar; purple, polar; cyan, polar aromatic; green, large hydrophobic; yellow, small hydrophobic.

Multiple sequence alignment of class A GPCRs has revealed that, in addition to the Y<sup>7.53</sup> showing 91% of consensus (Figure 3.11a), other two tyrosine may be present in the next two positions: 7.54 and 7.55 as shown in Figure 3.11; as well as the structure-based alignment of TM7 class A GPCRs, generated by GPCRdb (<https://www.gpcrdb.org/>) (Isberg *et al.*, 2016), has revealed a tyrosine percentage identity as follows: 89% in position 7.53 (Figure 3.11,b1), 10% in position 7.54 (Figure 3.11,b2), and 6% in position 7.55 (Figure 3.11,b3). Therefore, I suggest the hypothesis that FFAR1 may not have the first tyrosine but only the third, in position 7.55. To understand if this tyrosine shifted by two positions (Y<sup>7.55</sup>) can form a hydrophobic cluster with L<sup>1.63</sup> and F<sup>8.50</sup>, a structure superposition of FFAR1 with the homology model of the A<sub>1</sub>AR (section 3.4.1) has been done.

a

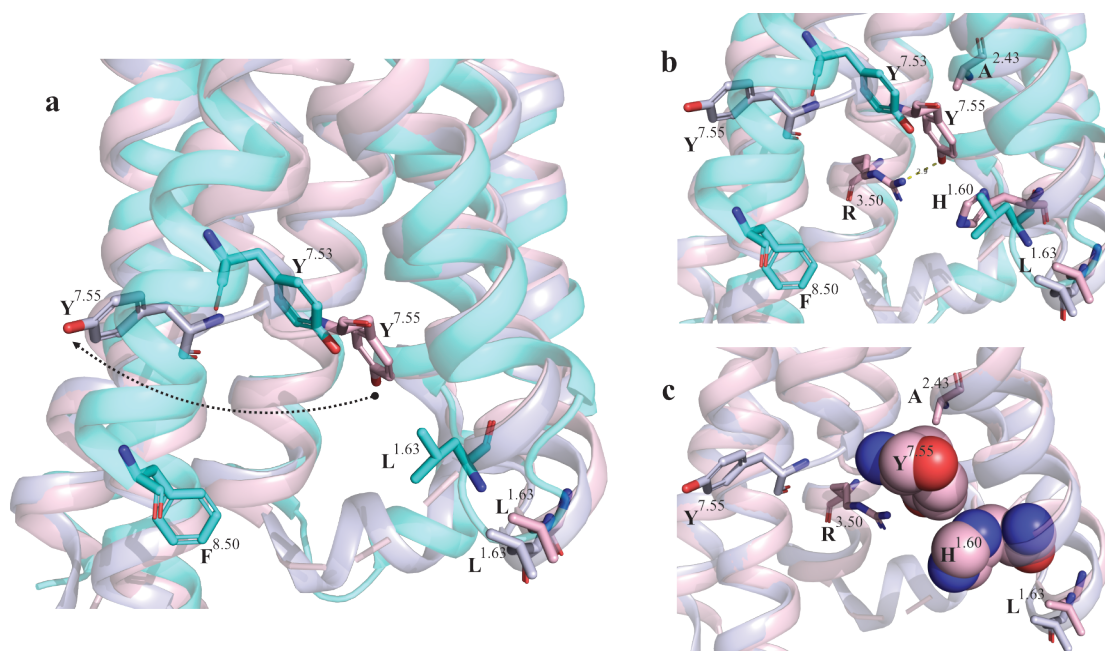


b



**Figure 3.11 | Analysis of tyrosine position in TM7 in class A GPCRs.** (a) WebLogo representation (<http://weblogo.threeplusone.com/create.cgi>) (Crooks *et al.*, 2004) of multiple sequence alignment of class A GPCRs. Position 7.53 (indicated by an arrow) is 91% conserved at Tyr. However, the Tyr may be present also in positions 7.54 and 7.55 (indicated by a red circle). (b) Pie charts of residues occurrence in positions: 7.53 (b1), 7.54 (b2), and 7.55 (b3). The data were obtained from a structure-based alignment of TM7 in class A GPCRs, generated by GPCRdb (<https://www.gpcrdb.org/>). The chart shows the percentage identity of tyrosine at each position: 89% (b1), 10% (b2), and 6% (b3).

The result of the structure superposition of FFAR1 with the homology model of the A<sub>1</sub>AR suggests that, although the tyrosine in FFAR1 is shifted by two positions in the sequence, the coordinates occupied by this tyrosine side chain are similar to those of the Y<sup>7.53</sup> of A<sub>1</sub>AR (Figure 3.12a) and, thus, it is possible for it to be involved in a hydrophobic cluster with H<sup>1.60</sup> and R<sup>3.50</sup> (not with L<sup>1.63</sup> and F<sup>8.50</sup> as in A<sub>1</sub>AR structure) (Figure 3.12b,c). In particular, Y<sup>7.55</sup> could form a hydrogen bond with R<sup>3.50</sup> and a  $\pi$ - $\pi$  stacking interaction with H<sup>1.60</sup>. Moreover, to support the hypothesis that this tyrosine Y<sup>7.55</sup> is part of the common motif in TM7, it moves outside of TM7 when the receptor assumes a more active-like state (Figure 3.12a).



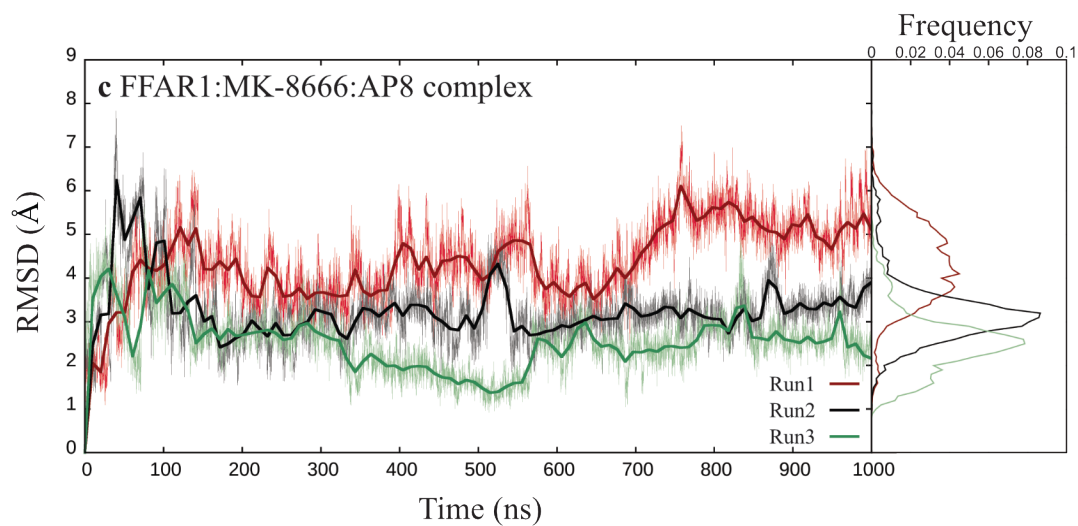
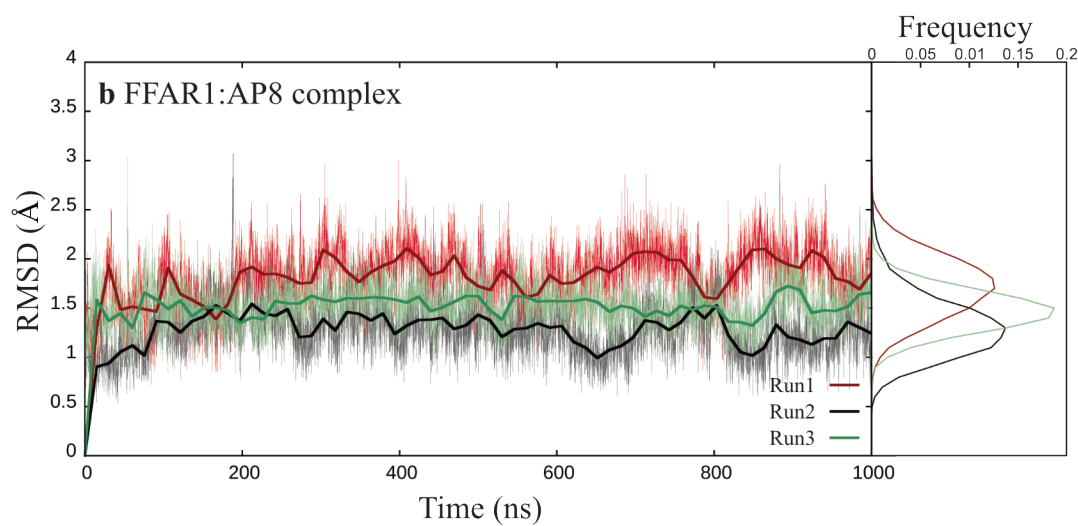
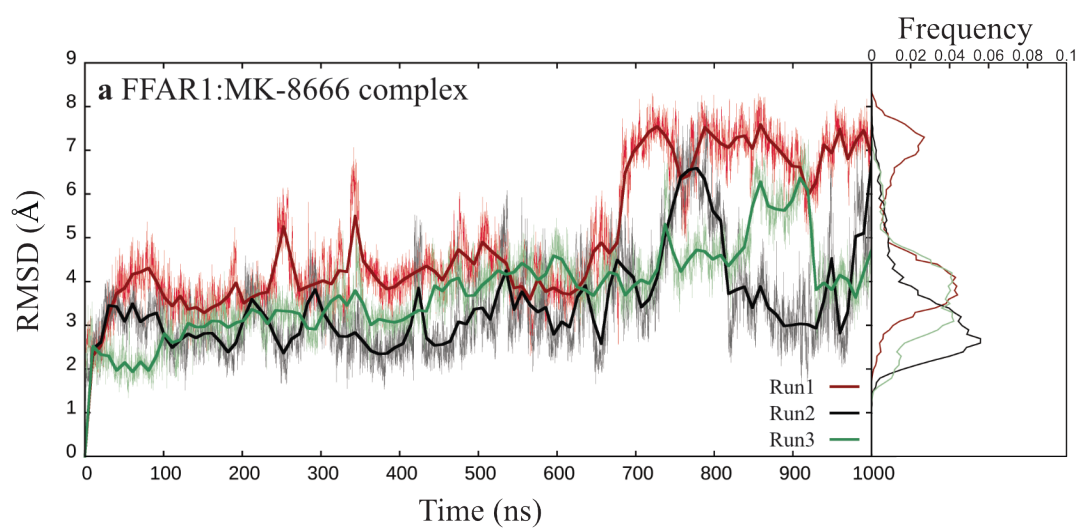
**Figure 3.12 | Analysis of NPxxY<sup>7.53</sup> motif - Comparison between FFAR1 and the homology model of the A<sub>1</sub>AR.** In this figure, all the receptors are shown in cartoon representation, respectively 5TZY (FFAR1) in light blue, 5TZR (FFAR1) in light pink, and 5N2Sm (A<sub>1</sub>AR) in cyan. Important residues are shown in stick representation. **(a)** Structure superposition of FFAR1 with A<sub>1</sub>AR shows a tyrosine in FFAR1 shifted by two positions in the sequence compared to the one in A<sub>1</sub>AR (Y<sup>7.53</sup> → Y<sup>7.55</sup>). The arrow indicates the outside movement of Y<sup>7.55</sup> from 5TZR to 5TZY structure when the receptor assumes a more active-like state. **(b)** The distance between Y<sup>7.55</sup> and R<sup>3.50</sup> in 5TZR is 2.9 Å thus a hydrogen bond is possible. **(c)** Structure superposition of FFAR1 with A<sub>1</sub>AR shows H<sup>1.60</sup> instead of L<sup>1.63</sup> (highlighted in spheres representation) that could form  $\pi$ - $\pi$  stacking interaction with Y<sup>7.55</sup>.

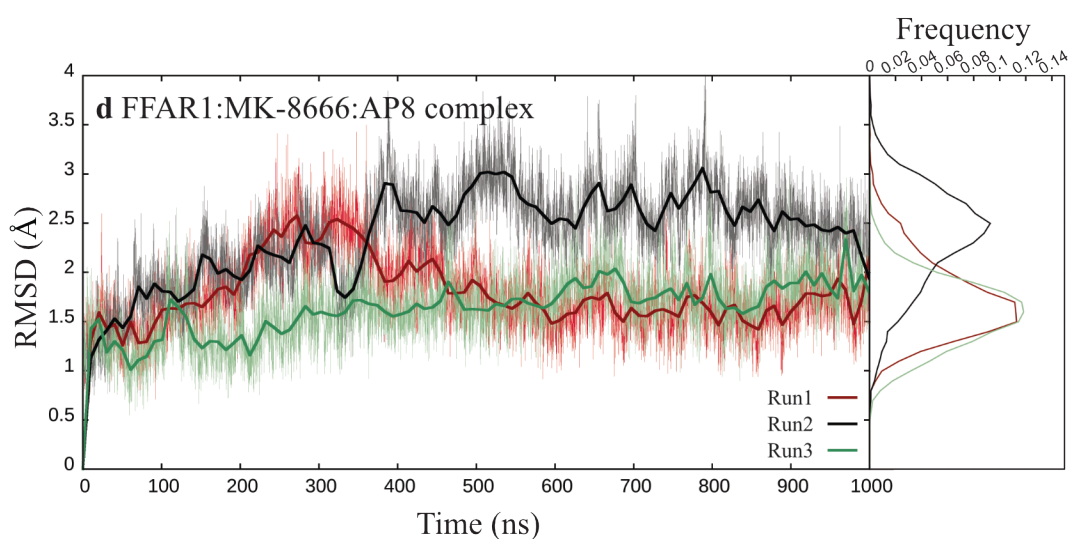
In conclusion, it is possible to assert that the common motif in TM7 is present also in FFAR1 (even if it is two residues longer) and that the tyrosine constituting this motif is involved in a hydrophobic cluster. Moreover, the missing H8 is probably not present in the FFAR1 X-ray structures due to its flexibility caused by the presence of several glycine residues in the helix (Figure 3.10) as well as due to the absence of the lipid bilayer during the purification.

### 3.4.3 MD simulations of FFAR1: comparison between apo and holo form

For each system reported in Table 3.2, MD simulations have been conducted and analysed to understand the possible effects on the ligand MK-8666 due to the presence of the allosteric ligand AP8. Figure 3.13 shows the RMSD analysis of both ligands (MK-8666 and AP8) calculated for each holo-FFAR1 system (summarised in Table 3.2). This analysis allows us to understand how much the ligand moves in the pocket. Therefore, the less it moves, the stronger are the interactions, the lower are the RMSD values, and consequently, the more stable/specific is the binding. From this analysis, it is possible to assert that the binding of MK-8666 is less stable in absence of AP8 (Figure 3.13a) than when AP8 is present in the allosteric binding site A2 (Figure 3.13c). In fact, the average RMSD of the ligand is higher in the first case (see Table 3.4) and, in the second part of the simulations (after 600 ns) the RMSD of MK-8666 tends to fluctuate more. Contrariwise, the binding of AP8 is stable in both systems (see Figure 3.13b,d).







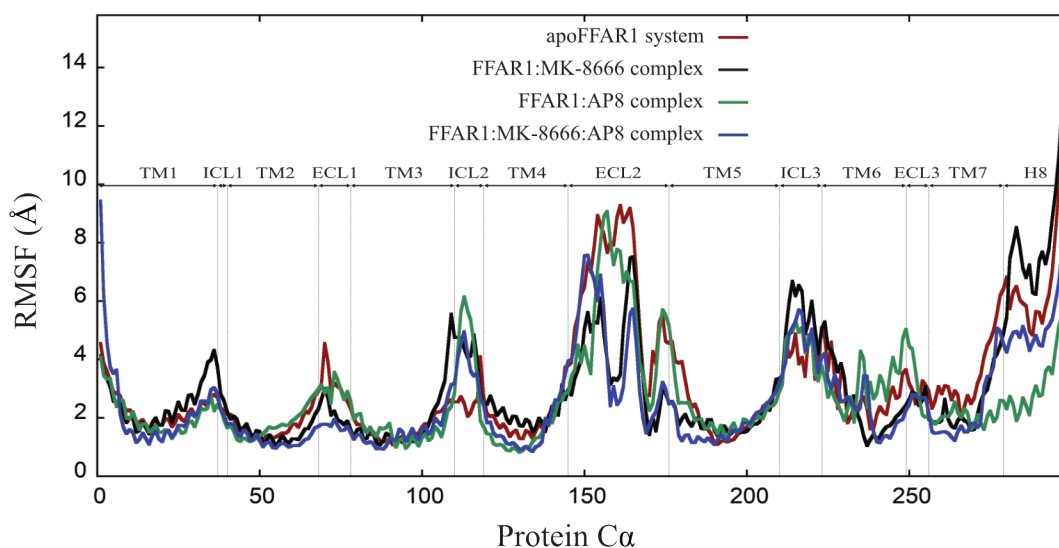
**Figure 3.13 | RMSD analysis of FFAR1 ligands.** RMSD plots and distributions of MK-8666 in the FFAR1:MK-8666 complex (a), RMSD of AP8 in the FFAR1:AP8 complex (b), RMSD of MK-8666 in the FFAR1:MK-8666:AP8 complex (c), and RMSD of AP8 in the FFAR1:MK-8666:AP8 complex (d) are shown with respect to the initial minimised structure for all the three MD simulations. The results suggest that the presence of AP8 in the allosteric binding site A2 stabilises the binding of MK-8666 in the allosteric binding site A1. The nomenclature is summarised in Table 3.2.

**Table 3.4 | Average RMSD and standard deviation (SD)**

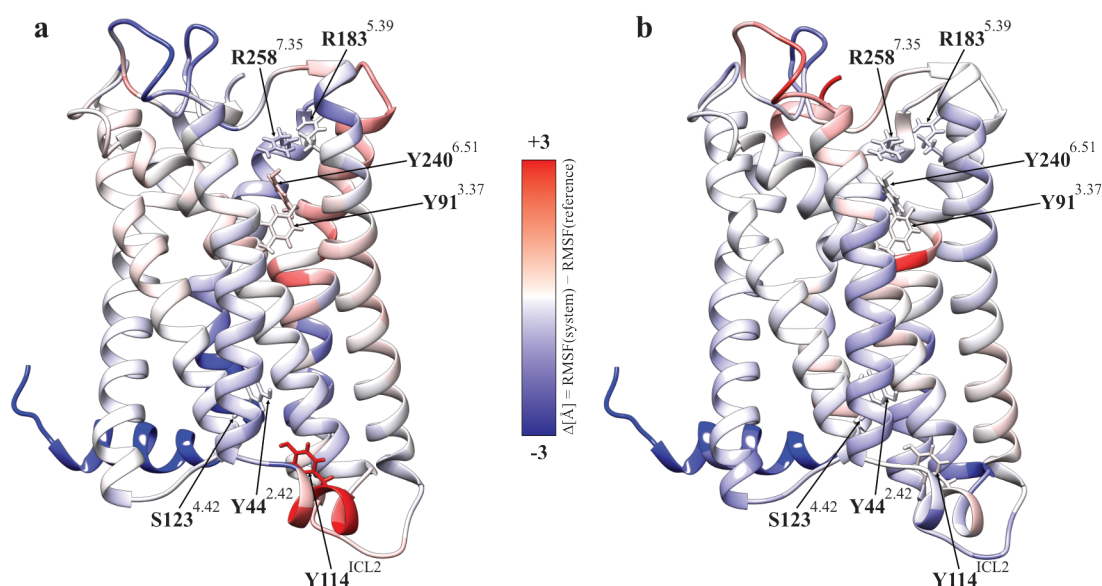
System	Ligand	RMSD $\pm$ SD of each replica [Å]		
		Run1	Run2	Run3
FFAR1:MK-8666	MK-8666	$4.98 \pm 1.56$	$3.43 \pm 1.12$	$3.78 \pm 1.03$
FFAR1:AP8	AP8	$1.77 \pm 0.31$	$1.37 \pm 0.31$	$1.53 \pm 0.21$
FFAR1:MK-8666:AP8	MK-8666	$4.37 \pm 0.92$	$3.26 \pm 0.75$	$2.55 \pm 0.69$
	AP8	$1.79 \pm 0.39$	$2.38 \pm 0.50$	$1.64 \pm 0.34$

The RMSF analysis conducted on the whole protein in the four systems (as previously reported in Table 3.2 - Figure 3.14) and the comparison between the systems (Figure 3.15a,b) allowed us to better understanding the influence that AP8 has on the apo and holo forms. In general, the presence of AP8 in concomitance with MK-8666 (FFAR1:MK-8666:AP8 complex, Figure 3.14; blue line) decreases the RMSF of the whole protein. In particular, there is a visible reduction in RMSF at the top of TM1, ECL1, the bottom TM4, and the top of TM5. In addition, stabilisation of H8 is

observable. This is in accordance with the colour variation from red-pink to blue in Figure 3.15. Therefore, when MK-8666 is in the allosteric binding site A1, the presence of AP8 in the allosteric binding site A2 decreases the flexibility of the bottom part of FFAR1 while the top residues in the TM4 located in the allosteric binding site A1 became more flexible. This is putatively due to an allosteric effect.



**Figure 3.14 | RMSF analysis of FFAR1 receptor.** RMSF plots from three MD simulations (Runs 1-3) of apoFFAR1 system (in red), FFAR1:MK-8666 complex (in black), FFAR1:AP8 complex (in green), and FFAR1:MK-8666:AP8 complex (in blue) were measured using the Cα of each residue. The presence of AP8 in concomitance with MK-8666 decreases the RMSF of the whole protein. In particular, it is visible a reduction of RMSF at the top of TM1, ECL1, the bottom TM4, the top of TM5, and H8.



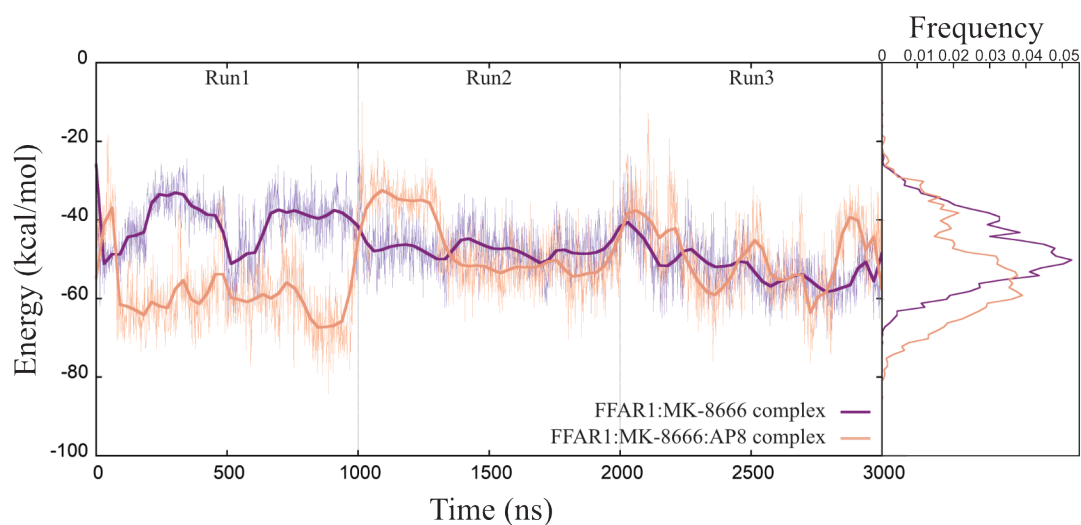
**Figure 3.15 | RMSF data difference comparison.** RMSF data difference between the apoFFAR1 (reference) and the FFAR1:AP8 (system) (a), and between the FFAR1:MK-8666 (reference) and the FFAR1:MK-8666:AP8 (system) (b). In this figure, the receptor is shown in cartoon representation, while important residues are shown in stick representation. The bar indicates the colour scale according to the RMSF difference ( $\Delta[\text{\AA}] = \text{RMSF}(\text{system}) - \text{RMSF}(\text{reference})$ ): negative values corresponding to blue and positive values to red.

The hydrogen bond occupancy analysis (Table 3.5) suggests that the hydrogen bonds between Tyr91<sup>3.37</sup> (an important residue located in the interface between the two allosteric binding sites A1 and A2) and Tyr240<sup>6.51</sup> or Asn241<sup>6.52</sup> increase in frequency by about 10% in the FFAR1:MK-8666 complex, compared to the apoFFAR1. Contrariwise, in the FFAR1:MK-8666:AP8 complex, Tyr91<sup>3.37</sup> makes a hydrogen bond with the ligand MK-8666 (therefore, only in presence of AP8) and decreases the hydrogen bond occupancy with Asn241<sup>6.52</sup> by about 15%.

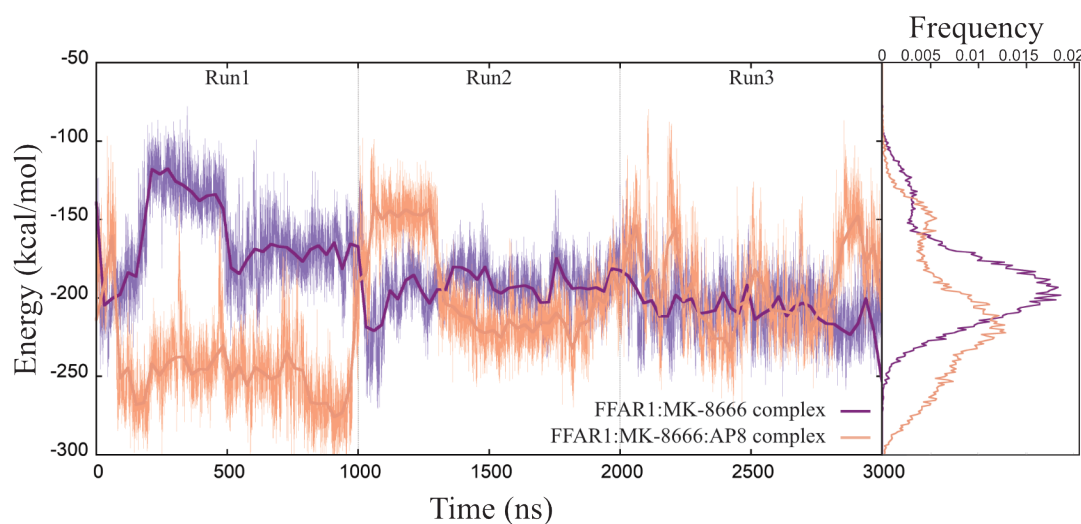
**Table 3.5 | Occupancy of the hydrogen bond between key residues of the allosteric binding site A1 (*i.e.* Tyr91<sup>3.37</sup>, Arg183<sup>5.39</sup>, Tyr240<sup>6.51</sup>, and Arg258<sup>7.35</sup>) and the whole protein**

System	Tyr91 <sup>3.37</sup>		Arg183 <sup>5.39</sup>		Tyr240 <sup>6.51</sup>		Arg258 <sup>7.35</sup>	
	Residue	%	Residue	%	Residue	%	Residue	%
apoFFAR1	Asn241 <sup>6.52</sup>	31.6	Glu145 <sup>4.64</sup>	61.8	His86 <sup>3.32</sup>	23.0	Glu172 <sup>45.52</sup>	45.2
	Tyr240 <sup>6.51</sup>	12.4	Glu172 <sup>45.52</sup>	45.2	Tyr91 <sup>3.37</sup>	12.4	Ser243 <sup>6.54</sup>	32.3
	Asn244 <sup>6.55</sup>	2.6	Asn244 <sup>6.55</sup>	5.3	Glu172 <sup>45.52</sup>	9.2	Asn244 <sup>6.55</sup>	26.5
			Tyr240 <sup>6.51</sup>	3.3	Asn244 <sup>6.55</sup>	3.5		
FFAR1:MK-8666					Arg183 <sup>5.39</sup>	3.3		
					Tyr12 <sup>1.39</sup>	3.0		
	Tyr240 <sup>6.51</sup>	23.6	MK-8666	82.1	MK-8666	61.2	MK-8666	85.0
	Asn241 <sup>6.52</sup>	21.5	Asp175 <sup>5.31</sup>	27.8	Tyr91 <sup>3.37</sup>	23.6	Glu172 <sup>45.52</sup>	52.0
FFAR1:AP8	Asn244 <sup>6.55</sup>	1.4	Glu172 <sup>45.52</sup>	26.1	Asn244 <sup>6.55</sup>	3.9	Asn244 <sup>6.55</sup>	21.2
			Ser247 <sup>6.58</sup>	11.4			Ser243 <sup>6.54</sup>	9.8
	Tyr240 <sup>6.51</sup>	20.0	Glu145 <sup>4.64</sup>	92.8	Tyr91 <sup>3.37</sup>	20.0	Asp175 <sup>5.31</sup>	9.3
	Ser187 <sup>5.43</sup>	16.4	Glu172 <sup>45.52</sup>	1.4	Asn244 <sup>6.55</sup>	8.9	Glu172 <sup>45.52</sup>	33.8
FFAR1:MK-8666:AP8	His137 <sup>4.56</sup>	2.1			His86 <sup>3.32</sup>	4.6	Asp175 <sup>5.31</sup>	29.0
					Ser58 <sup>2.56</sup>	2.6	Ser247 <sup>6.58</sup>	8.6
							Asn244 <sup>6.55</sup>	2.6
							Glu65 <sup>2.63</sup>	2.4
FFAR1:MK-8666:AP8	Tyr240 <sup>6.51</sup>	16.4	MK-8666	91.6	MK-8666	40.2	Glu172 <sup>45.52</sup>	62.2
	MK-8666	11.9	Glu172 <sup>45.52</sup>	15.2	Tyr91 <sup>3.37</sup>	16.4	MK-8666	61.6
	Asn241 <sup>6.52</sup>	7.5	Ser247 <sup>6.58</sup>	6.6	His86 <sup>3.32</sup>	6.0	Asn244 <sup>6.55</sup>	7.0
	Asn244 <sup>6.55</sup>	7.5	Asn244 <sup>6.55</sup>	1.6	Asn244 <sup>6.55</sup>	5.9	Ser243 <sup>6.54</sup>	5.8

According to the GBSA and the NAMD Energy analysis (computed as explained in section 2.3.7), the distribution of MK-8666 energy values of the FFAR1:MK-8666:AP8 complex (Figure 3.16 and Figure 3.17) shifts to lower energy values, suggesting that the binding of MK-8666 becomes more stable in presence of AP8.



**Figure 3.16 | MK-8666 GBSA analysis.** GBSA energy plot and distribution of MK-8666 in the FFAR1:MK-8666 complex (in purple) and FFAR1:MK-8666:AP8 complex (in pink) during MD simulations. The shift to lower energy values of the FFAR1:MK-8666:AP8 complex in the distribution plot indicates the increase of stability of MK-8666 binding in the presence of AP8.



**Figure 3.17 | MK-8666 NAMD Energy analysis.** NAMD Energy calculations operated on the two following systems: FFAR1:MK-8666 complex (in purple) and FFAR1:MK-8666:AP8 complex (in pink). The results show that the energy of MK-8666 is generally lower when AP8 is in the allosteric binding site A2. This determines the shift to lower energy values of the FFAR1:MK-8666:AP8 complex in the

distribution plot indicating the increase of stability of MK-8666 binding in the presence of AP8.

Following, the supporting movies (SM3.1-3) of the FFAR1 MD simulations are reported.

**Supporting movie SM3.1 | MD simulation of FFAR1 apo form available on the following link: <https://youtu.be/jRiqoS3It7A>.** The movie shows the FFAR1 apo form in cartoon representation and highlights the residues that are dynamically detected within 5 Å from H8 in stick representation during a MD simulation (MD simulation time 0-1 µs).

**Supporting movie SM3.2 | MD simulation of FFAR1:AP8 complex available on the following link: <https://youtu.be/0bhLDX9VRBo>.** The movie shows the FFAR1 holo form in cartoon representation and highlights the ligand AP8 (7OS) and residues that are dynamically detected within 5 Å from H8 in stick representation during a MD simulation (MD simulation time 0-1 µs).

**Supporting movie SM3.3 | MD simulation of FFAR1 holo form available on the following link: [https://youtu.be/sOMX9peLj\\_M](https://youtu.be/sOMX9peLj_M).** The movie shows the comparison between the FFAR1:MK-8666 complex (**left**) and FFAR1:MK-8666:AP8 complex (**right**). A different orientation of the residues implicated in the interaction with MK-8666 is detected. The backbone after the Tyr91<sup>3,37</sup> (located at the interface between the two binding sites) is less stable in the absence of AP8 and drives conformational changes of this residue able to destabilise the key electrostatic interactions between MK-8666 and Arg183<sup>5,39</sup>, Arg258<sup>7,35</sup>. Relevant residues and ligands are shown in stick representation while the receptor is shown in cartoon representation during MD simulations (MD simulation time 0-3 µs).



### 3.5 Conclusion

The sequence and structural analysis conducted in section 3.4.1 has revealed that the binding site A2 exists in other GPCRs and not only in FFAR1, indeed, residues implicated in the AP8-binding site are generally very well conserved. However, it is worth noting that the residue in position 3.48 is situated close to the middle region of the binding pocket A2 and, therefore, allows the AP8 binding in this region. The structural comparison of this position between FFAR1 and A<sub>1</sub>AR has shown an alanine and a valine residue, respectively. Experimental data showed that the mutation Ala<sup>3.48</sup>Val in FFAR1 decreases AP8 binding affinity of about 10 fold due to the longer valine side chain occupies the binding region of the AP8 aromatic ring (Lu *et al.*, 2017). However, sequence analysis of TM3 class A GPCRs has revealed Ile, Leu, and Val as top 3 residues, indicating hydrophobic and aliphatic residues are privileged in this position.

The sequence analysis conducted in section 3.4.2 on H8 has been performed before the availability of the A<sub>1</sub>AR active structure (PDB ID: 6D9H) (Draper-Joyce *et al.*, 2018) that broke the rules because of the presence of the H8 despite the potential positive repulsion between Arg<sup>12.51(41)</sup> in the ICL1 and Arg<sup>8.51(296)</sup> in the H8. This gave us the confidence to break the rules with FFAR1 which has Arg<sup>1.64(37)</sup> in the ICL1 and Lys<sup>8.53(286)</sup> in the H8. In FFAR1, the Tyr<sup>7.55(278)</sup> in the TM7 should be in position 276 where there is instead a Thr<sup>7.53(276)</sup>, and these residues should interact with Leu<sup>1.63(36)</sup> in the ICL1 and Phe<sup>8.54</sup> in the H8, that in this case is a Thr<sup>8.54(287)</sup>, to lock the receptor in an inactive state. However, because this potential positive repulsion between Arg<sup>1.64(37)</sup> in the ICL1 and Lys<sup>8.53(286)</sup> in the H8, this interaction seems to be missing in the X-ray structures. The question that has arisen is: with the loss of this interaction TM7-ICL1-



H8, what interaction is taking over? To try to answer this question I looked at the dynamic of the system with MD simulations (section 3.4.3) but no particular interactions have been identified for Thr276<sup>7.53</sup> and Y278<sup>7.55</sup>.

MD simulations of FFAR1 with AP8 in allosteric binding site A2 have revealed a general stabilisation of the whole receptor and, in particular, a more stable H8 that keeps part of its helicity. The glycine-rich flexible region situated at the end of TM7 and start of H8 (Gly277<sup>7.54</sup>, Gly280<sup>8.47</sup>, Gly282<sup>8.49</sup>, and Gly284<sup>8.51</sup>) allows a transverse movement of H8 while, after Gly295<sup>C-term</sup> and Gly296<sup>C-term</sup>, H8 loses completely its secondary structure showing a wider movement. This flexibility caused by the presence of several glycine residues in this region as well as the absence of the lipid bilayer may be the reason why H8 is missed in the FFAR1 X-ray structures. Moreover, in presence of AP8 an extra-movement of ICL2 is visible and this may allow the FFAR1 to interact with the G-protein in a different way compared to other class A GPCRs, but also may allow H8 to have a more defined structure and this is a clear example of the allosteric effect. Furthermore, during the simulations, it is possible to assert that the AP8-binding site appears and then disappears in the apo form while remains a valid pocket in FFAR1:MK-8666 complex.

In the next chapter, I will describe the suMD simulations of FFAR1 for both MK-8666 and AP8 to understand the binding paths differences. Considering the binding modes as for the X-ray crystal structures and the high lipophilicity characterising these ligands, a putative binding path from the membrane will be investigated.

## 4. Addressing free fatty acid receptor 1 (FFAR1) activation using supervised molecular dynamics\*

(\*) This chapter was published on 27/08/2020 in the *Journal of Computer-Aided Molecular Design*. doi: 10.1007/s10822-020-00338-6 (Atanasio *et al.*, 2020).

### 4.1 Abstract

**Background:** The free fatty acid receptor 1 (FFAR1, formerly GPR40), is a potential G protein-coupled receptor (GPCR) target for the treatment of type 2 diabetes mellitus (T2DM), as it enhances glucose-dependent insulin secretion upon activation by endogenous long-chain free fatty acids. The presence of two allosterically communicating binding sites and the lack of the conserved GPCR structural motifs challenge the general knowledge of its activation mechanism. To date, four X-ray crystal structures are available for computer-aided drug design. **Methods and Aim:** In this study, I employed molecular dynamics (MD) and supervised molecular dynamics (SuMD) to deliver insights into the (un)binding mechanism of the agonist MK-8666, and the allosteric communications between the two experimentally determined FFAR1 binding sites. **Results and Conclusion:** I found that FFAR1 extracellular loop 2 (ECL2) mediates the binding of the partial agonist MK-8666. Moreover, simulations showed that the agonists MK-8666 and AP8 are reciprocally stabilised and that AP8 influences MK-8666 unbinding from FFAR1. **Keywords:** G protein-coupled receptors (GPCRs); FFAR1; GPR40; Molecular dynamics (MD); Supervised Molecular dynamics (SuMD).

### 4.2 Introduction

The free fatty acid receptor 1 (FFAR1, formerly GPR40) is a G protein-coupled receptor (GPCR) naturally activated by saturated C<sub>12</sub>-C<sub>16</sub> or unsaturated C<sub>18</sub>-C<sub>20</sub> free fatty acids (Kotarsky *et al.*, 2003). These long chain free fatty acids (LCFAs) act as full agonists to enhance glucose-stimulated insulin secretion from pancreatic  $\beta$  cells (Itoh *et al.*, 2003), and secretion of the incretins Glucagon-like peptide 1 (GLP-1) and Gastric inhibitory polypeptide (GIP) from intestinal enteroendocrine cells (Edfalk *et al.*, 2008). The development of drugs activating FFAR1 represents a new potential therapeutic approach against type 2 diabetes mellitus (T2DM - counting for  $\sim 90\%$  of all diabetes cases (World Health Organization, 2016a)) as the activation of its alternative signalling (Watterson *et al.*, 2014; Li *et al.*, 2018) may avoid the negative side effects characterising the currently approved drugs, such as hypoglycaemia and weight gain (Burant *et al.*, 2012).

To date, the structure-based design of new compounds (Tikhonova and Poerio, 2015; Tikhonova, 2017) can exploit four crystallographic FFAR1 complexes (as reported in Table 3.1 of the previous chapter). In all these structures the receptor is captured in the inactive state, probably due to the stabilising mutations introduced and the insertion of lysozyme T4 into intracellular loop 3 (ICL3) (Ho *et al.*, 2018). Moreover, the glycine-rich C-terminal segment of transmembrane helix 7 (TM7) and helix 8 (H8) are not resolved. Even though FFAR1 is organised in the GPCR structural hallmark seven transmembrane domain (TMD), the conserved structural motifs NP<sup>7.50</sup>xxY (Figure 3.10), DR<sup>3.50</sup>Y, and P<sup>5.50</sup>I<sup>3.40</sup>F<sup>6.44</sup> as well as the “toggle switch” W<sup>6.48</sup> (Latorraca *et al.*, 2017; Weis and Kobilka, 2018), are not present.

Before any FFAR1 structural data were available (Table 3.1), a pharmacological body of evidence led to the hypothesis of up to three different binding sites (Lin *et al.*, 2012). In this scenario, full agonists would bind a different site to the partial agonists, with positive heterotropic cooperativity between the different binding sites. Indeed, the partial agonist fasiglifam (TAK-875, Figure 4.1, discontinued in phase III clinical trials due to liver toxicity) is able to amplify the agonistic activity of endogenous LCFA by binding to a largely internal FFAR1 allosteric site (Yabuki *et al.*, 2013). The first published FFAR1 structure (Srivastava *et al.*, 2014) located the bound TAK-875 enclosed between the top halves of TM3 and TM4 (Figure 4.1), with the carboxylic moiety pointing into a hydrophilic cavity partially buried from the extracellular bulk by extracellular loop 2 (ECL2). TAK-875 participates in an extended charge network involving R183<sup>5.39</sup>, R258<sup>7.35</sup>, E172<sup>ECL2</sup>, Y91<sup>3.37</sup>, S187<sup>5.43</sup>, N241<sup>6.52</sup>, and N244<sup>6.55</sup>. The disruption of the E145<sup>4.64</sup> - R183<sup>5.39</sup> and E172<sup>ECL2</sup> - R258<sup>7.35</sup> salt bridges upon TAK-875 binding have been proposed as contributing to receptor activation (Sum *et al.*, 2009; Lin *et al.*, 2012), and corroborated by mutagenesis data on Y91<sup>3.37</sup> and N244<sup>6.55</sup> (Yabuki *et al.*, 2013). In 2017 Lu J. *et al.* (Lu *et al.*, 2017) resolved FFAR1 in a ternary complex with the partial agonist MK-8666 and the agonist AP8 (Figure 4.1), revealing a second external binding site at the protein/lipid interface, which is responsible for the recognition of full agonists. While MK-8666 inserts in the same crevice and forms the same interaction pattern as TAK-875 (site A1 in Figure 4.1), AP8 is accommodated in a cleft shaped by TM3, TM4, and TM5 (site A2 in Figure 4.1), where I130<sup>4.49</sup>, L133<sup>4.52</sup>, V134<sup>4.53</sup>, and L190<sup>5.46</sup> form a hydrophobic pocket for the terminal trifluoromethoxyphenyl moiety of the ligand and Y44<sup>2.42</sup>, Y114<sup>I<sup>CL</sup>2</sup>, and S123<sup>4.42</sup> anchor the carboxylate through polar interactions (Figure 4.1). A direct comparison between the binary FFAR1:TAK-875 and the ternary FFAR1:MK-

8666:AP8 complexes shows TM4 and TM5 sliding approximately one-half of a helical turn in opposite directions to create a deeper cleft for AP8, which in turn stabilises ICL2 in a helix conformation. The subsequent crystal structure of the binary complex between FFAR1 and the full agonist Compound 1 (Figure 4.1) bound to site A2 sheds further light into the cross-talk between the different binding sites (Ho *et al.*, 2018); it shows a rearrangement of site A1 in which D175<sup>5.31</sup> inserts between R183<sup>5.39</sup> and R258<sup>7.35</sup> and moves ECL2 towards TM6 and TM7. This conformational rearrangement in the absence of partial agonists occupying site A2 prevents the formation of an additional proposed solvent exposed binding site between TM1 and TM7 (site A3 in Figure 4.1b) (Srivastava *et al.*, 2014; Lückmann *et al.*, 2019).

In this complex scenario, endogenous LCFAs are thought to bind to the site A2. An orthosteric binding site exposed to the membrane (site A2), responsible for full activation of the receptor, could have evolved along with a different structural mechanism of signal transduction, finely tuned by an allosteric communication network as intriguingly suggested by the absence of conserved class A structural motifs. Indeed, a chemical signal (the binding of LCFA) originating from the membrane, rather than from the extracellular environment (the case for almost all the other class A GPCRs) could underlie the FFAR1 divergence from the consensus structure, and the possibility of a 2:1 stoichiometry between LCFA and FFAR1 (with sites A1 and A2 both occupied simultaneously).

Here I address the FFAR1 activation first describing a putative binding mechanism for the formation of the binary FFAR1:MK-8666 complex using supervised molecular dynamics (SuMD) simulations. To date, few ligands accessing a GPCRs orthosteric site from the membrane have been addressed with atomistic unbiased simulations (Jakowiecki and Filipek, 2016; Stanley *et al.*, 2016). Then,

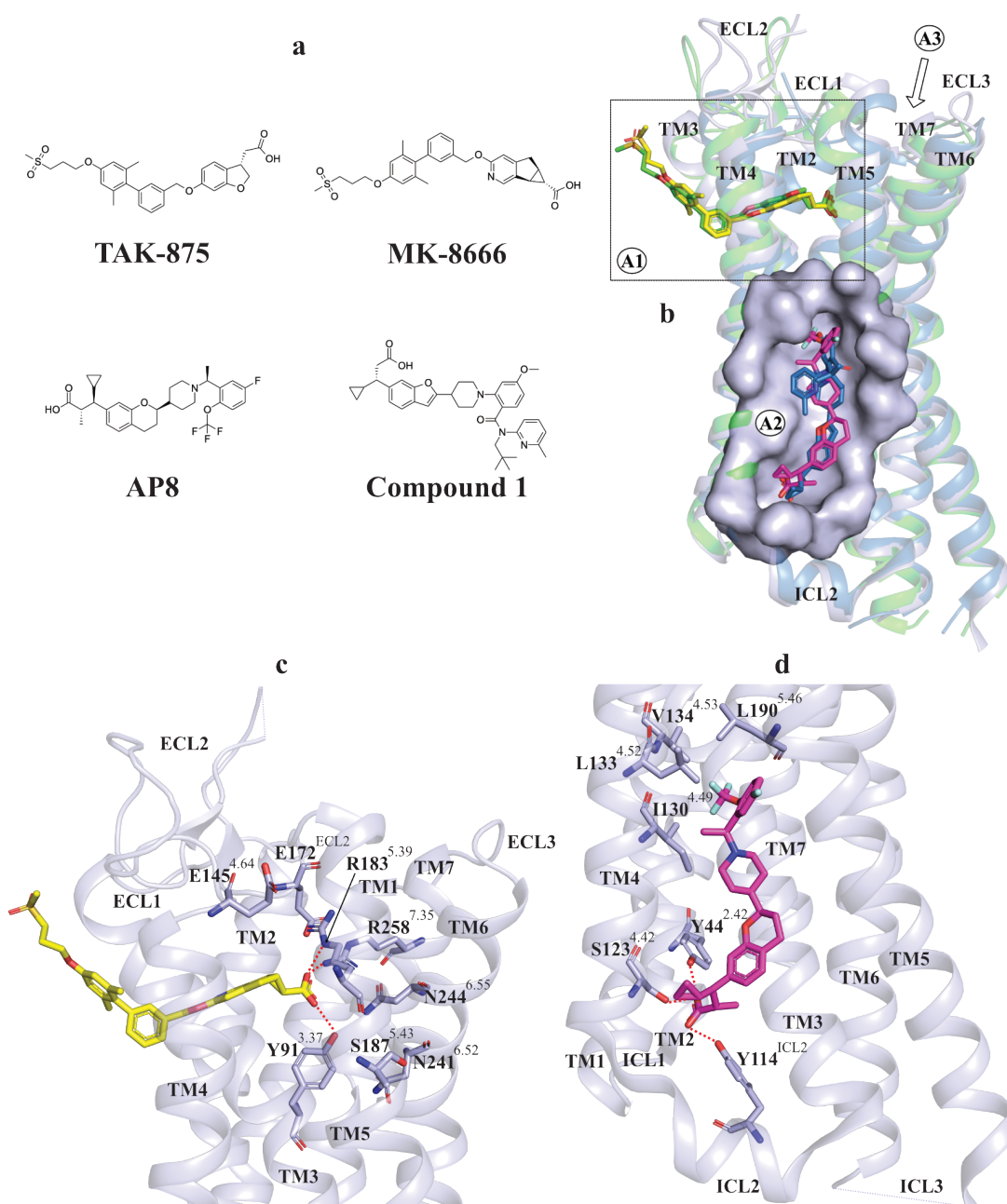
insights on possible interactions contributing to the allosteric communication between site A1 and A2 from equilibrium MD simulations performed on different FFAR1 complexes (Table 4.1) are presented. Finally, I applied a modified version of SuMD to simulate the MK-8666 unbiased unbinding from the binary FFAR1:MK-8666 and ternary FFAR1:MK-8666:AP8 complexes. Overall, this study produces mechanistic working hypothesis on the allosteric regulation and activation of FFAR1.

**Table 4.1 | Summary of the MD simulations performed.** Classic MD (cMD) and supervised MD (SuMD)

System	Total MD sampling time [ $\mu$ s] and number of replicas		
	cMD <sup>#</sup>	SuMD <sup>§</sup> - binding	SuMD <sup>§</sup> - unbinding
apoFFAR1	3 $\mu$ s (3 replicas)	-	-
FFAR1:MK-8666	3 $\mu$ s (3 replicas)	2.181 $\mu$ s (2 x SuMD replica/cMD replica)	1.148 $\mu$ s (3 replicas)
FFAR1:AP8	3 $\mu$ s (3 replicas)	-	-
FFAR1: MK-8666:AP8	3 $\mu$ s (3 replicas)	-	0.490 $\mu$ s (4 replicas)

(#) cMD: classic molecular dynamics

(§) SuMD: supervised molecular dynamics



**Figure 4.1 | Allosteric binding sites A1 and A2 in FFAR1.** (a) Chemical structure of the ligands co-crystallised with FFAR1; (b) binding sites of the different FFAR1 ligands: TAK-875 (green) and MK-8666 (yellow) bind to site A1, while the AP8 (magenta) and Compound 1 (blue) bind to site A2 (surface representation). The receptor is shown in cartoon representation (PDB ID: 4PHU in green, 5TZY in light-blue, and 5KW2 in sky-blue). Binding site details of (c) site A1 with MK-8666 (yellow) and (d) site A2 with AP8 (magenta). The receptor (PDB ID: 5TZY) is shown in cartoon representation, while important residues are shown in stick representation.

## 4.3 Materials and Methods

### 4.3.1 System preparation for the MD

Since FFAR1 is in the inactive state in all the available crystallographic structures (possibly due to the stabilising mutations and the insertion of T4 lysozyme into ICL3 (Ho *et al.*, 2018)), the intermediate active structure was retrieved from the GPCRdb website (<https://www.gpcrdb.org/>) (Isberg *et al.*, 2016). All the systems reported in Table 4.1 were prepared using a combination of High-Throughput Molecular Dynamics (HTMD) (Doerr *et al.*, 2016) and Tool Command Language (TCL) scripts. The disulfide bond between residues C79<sup>3.25</sup> and C170<sup>ECL2</sup> was included. The receptor orientation was determined from the Orientations of Proteins in Membranes (OPM) database entry 5TZY (<https://opm.phar.umich.edu/>) (Lomize *et al.*, 2006). The hydrogen atoms were added using PDB2PQR (Dolinsky *et al.*, 2004) and PROPKA (Olsson *et al.*, 2011) software (considering a simulated pH of 7.0) coupled with visual inspection. The receptor was embedded in a 85 Å x 85 Å 1-palmitoyl-2-oleyl-sn-glycerol-3-phospho-choline (POPC) bilayer using the Visual Molecular Dynamics (VMD) Membrane Builder plugin 1.1 (<http://www.ks.uiuc.edu/Research/vmd/plugins/membrane/>) through an insertion method (Sommer, 2013) with overlapping lipids removed (protein-lipid distance cut-off of 0.6 Å). The addition of the TIP3P water molecules to the simulation box (85 Å x 85 Å x 100 Å) was carried out using the VMD Solvate plugin v.1.5 (<http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/>). Finally, sodium and chloride ions were added to neutralise the system and mimic an ionic strength of 0.150 M using the VMD Autoionize plugin 1.3 (<http://www.ks.uiuc.edu/Research/vmd/plugins/autoionize/>).



### 4.3.2 System equilibration and MD Settings

The equilibration and productive simulations were performed using ACEMD (Harvey *et al.*, 2009), employing the CHARMM36 force field (Klauda *et al.*, 2010; Huang and MacKerell, 2013). The ligands AP8 and MK-8666 (Figure 4.1) were parameterised in the CGenFF force field (Vanommeslaeghe *et al.*, 2010), with the exception of the AP8 rotatable C-O bond linking the trifluoromethoxy group to the fluorobenzene ring, and the MK-8666 rotatable bond involving the carboxylic acid and the cyclopropyl ring. These dihedral terms were respectively optimised using HTMD parametrise (Doerr *et al.*, 2016) and the VMD plugin fTK (Mayne *et al.*, 2013), after fragmentation of the molecule.

The Langevin thermostat (Loncharich *et al.*, 1992) (target temperature: 300 K; low damping: 1 ps<sup>-1</sup>; positional restraints on protein atoms: 1 kcal mol<sup>-1</sup> Å<sup>-2</sup>) and the Berendsen barostat (Berendsen *et al.*, 1984) (target pressure: 1 atm) were used for the equilibration in isothermal-isobaric conditions (NPT) through a *four-stage* procedure employing an integration time step of 2 fs. The following *four-stage* procedure was used. *First stage*: clashes between protein and lipid atoms were reduced through 2000 conjugate-gradient minimisation steps; *Second stage*: a 2 ns long MD simulation was run with positional restraints of 1 kcal mol<sup>-1</sup> Å<sup>-2</sup> on protein and lipid phosphorus atoms; *Third stage*: a 20 ns of MD simulation was performed restraining only the protein atoms; *Fourth stage*: positional restraints were applied only to the protein backbone alpha carbons, for a further 30 ns.

Productive trajectories were computed with an integration time step of 4 fs in the canonical ensemble (NVT) at 300 K using a thermostat damping of 0.1 ps<sup>-1</sup>. The M-SHAKE algorithm (Kräutler *et al.*, 2001) was used to constrain the covalent bonds involving hydrogen atoms. A 9 Å cut-off distance was set for electrostatic interactions,

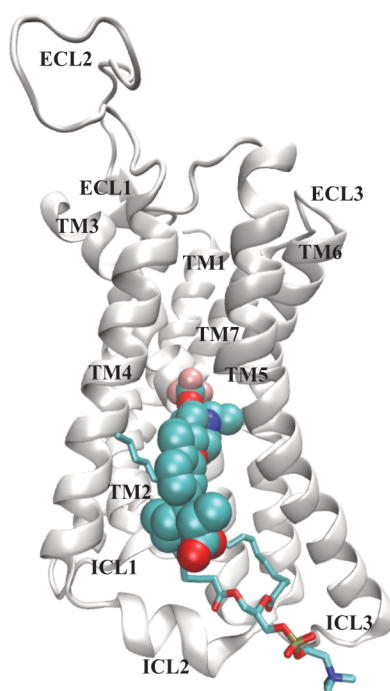
with a switching function applied beyond 7.5 Å. All simulations performed are summarised in Table 4.1. The particle mesh Ewald summation method (PME) (Essmann *et al.*, 1995) with a mesh spacing to 1.0 Å was employed to handle long-range Coulombic interactions.

### 4.3.3 SuMD - binding

The supervised MD (SuMD) (Sabbadin and Moro, 2014; Cuzzolin *et al.*, 2016) is an adaptive sampling method that uses a tabu-like algorithm to monitor the distance between the centres of masses (or the geometrical centres) of the ligand and the predicted binding site, during short classical MD simulations. After each simulation, the distances (collected at regular time intervals) are fitted to a linear function and if the resulting slope is negative then the next simulation step starts from the last set of coordinates and velocities produced, otherwise, the simulation is restarted by randomly assigning the atomic velocities. This approach allows simulating the formation of intermolecular complexes in timescales that are one or two orders of magnitude faster than the correspondent classical (unsupervised) MD simulations. Importantly, the sampling is gained without the introduction of any energetic bias.

MK-8666 was placed about 30 Å from its binding site, at the extracellular water/membrane interface, and the resulting system was prepared as reported in section 4.3.1. The distance between the MK-8666 (tetrahydrocyclopropa[4,5]cyclopenta[1,2-c]pyridine-6-carboxylic) moiety and the FFAR1 centroid computed on residues F87<sup>3.33</sup> and F142<sup>4.61</sup> (centroids distance) was supervised during successive 300 ns-long MD time windows. A single replica was run until the ligand reached the receptor, then, two different replicas were seeded (Replica 1 and Replica 2 in Supporting movie SM4.1).

AP8 was placed about 30 Å from its binding site, in the intracellular POPC layer, and the resulting system was prepared as reported in section 4.3.1. The distance between the AP8 and the FFAR1 centroid computed on residues A98<sup>3.44</sup> and A99<sup>3.45</sup> (centroids distance) was supervised during successive 300 ns-long MD time windows. No productive SuMD binding trajectories were sampled due to the presence of stable POPC residues at the FFAR1 site A2 (Figure 4.2).



**Figure 4.2 | AP8 SuMD binding to FFAR1 site A2 was hindered by phospholipids.** Stable POPC residues at the interface with site A2 prevented the full binding of AP8. The receptor is shown as a white cartoon, AP8 as van der Waals spheres, and POPC in stick representation.

#### 4.3.4 SuMD - unbinding

Recently, the SuMD approach has been successfully applied to the unbinding of GPCR ligands (Deganutti *et al.*, 2020). In this study, giving the extended electrostatic network characterising MK-8666 in the bound state, I have further modified and tested the protocol by supervising, besides the centroids distance, the number of water oxygen atoms that are within 4 Å from protein atoms involved in hydrogen bonds with the ligand (Figure 4.3). At the end of each MD time window, the simulation was considered productive (and the MD was restarted from the end) if the slopes of both the two plotted linear functions were positive, or the last centroids distance was longer than the one from the previous productive time window. Otherwise, the coordinates from the last productive MD time window were used and the velocities were reassigned. The general rationale for this water supervision is that the displacement of ligands strongly stabilised by ionic interactions should be facilitated if the hydration of protein hydrophilic spots interacting with the ligand is enhanced.

Differently from the original SuMD binding algorithm, the length ( $\Delta t$ ) of the short simulations performed increased along the unbinding pathway, according to the formula:

$$\Delta t = \Delta t_0 N t_i \quad (4.1)$$

$\Delta t_0$  is the duration of the very first MD time window and  $N t_i$  represents a factor that is picked from three user-defined values ( $N t_1$ ,  $N t_2$ , and  $N t_3$ ), according to the last ligand-protein distance detected. Three distance threshold values ( $D_1$ ,  $D_2$ , and  $D_3$ ) were set and the ligand-protein distance ( $r_L$ ) at the end of each MD run was compared to these

threshold values, allowing a decision on the value of the  $Nt_i$  factor according to the following conditions:

$$r_L \leq D_1 \rightarrow Nt_i = 1 \quad (4.2)$$

$$D_1 < r_L \leq D_2 \rightarrow Nt_i = Nt_1 \quad (4.3)$$

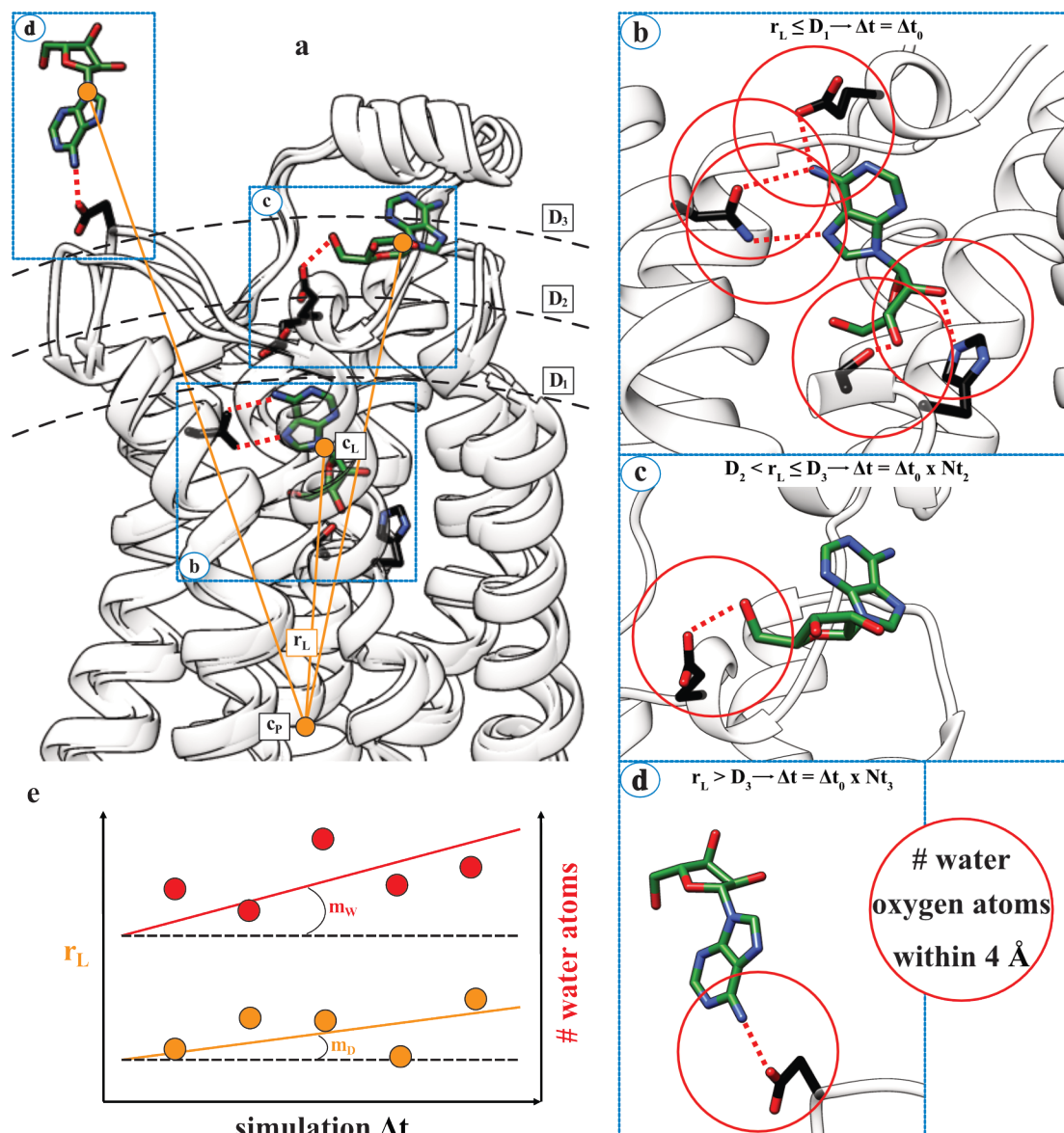
$$D_2 < r_L \leq D_3 \rightarrow Nt_i = Nt_2 \quad (4.4)$$

$$D_3 < r_L \rightarrow Nt_i = Nt_3 \quad (4.5)$$

For the FFAR1:MK-8666 unbinding, the initial time window length was 100 ps, with  $Nt_1$ ,  $Nt_2$ , and  $Nt_3$  set to 2, 4, and 8. Values of 10 Å, 14 Å, and 18 Å were used as  $D_1$ ,  $D_2$ , and  $D_3$  distances. The goal of increasing the simulation time window ( $\Delta t$  in equation 4.1) along the unbinding pathway is to facilitate the sampling of metastable states, which could otherwise be poorly visited.

The unbinding was iterated until no ligand-protein van der Waals contact was detected by means of the GetContacts scripts tools (<https://getcontacts.github.io>). The ligand-protein distance and the number of water oxygen atoms within 4 Å of protein donor/acceptor atoms were computed employing PLUMED 2 (Tribello *et al.*, 2014). After each productive MD time window, the GetContacts script (<https://getcontacts.github.io/index.html>) was employed to detect and update the protein atoms involved in hydrogen bonds with the ligand, considering a distance of 3.5 Å and an angle value of 120° as geometrical cut-offs. Notably, if no hydrogen bond between the ligand and the protein was present at the end of a productive MD time window, then protein atoms involved in water-mediated or van der Waals interactions were considered.

The FFAR1:MK-8666 and FFAR1:MK-8666:AP8 complexes were prepared for simulations as described in section 4.3.1.



**Figure 4.3 | Representation of the modified SuMD protocol for the unbinding of MK-8666.** (a) Three superimposed frames from a GPCR (white cartoon representation) - ligand (green stick representation) unbinding simulations are shown. The distance ( $r_L$ , orange line) between the ligand centroid ( $C_L$ ) and the protein centre ( $C_P$ ) is supervised during the protocol. Along the unbinding path (exemplified as states b, c, and d), the ligand experiences different hydrogen bonds; (b), (c), (d), the protein donor/acceptor atoms considered for the detection of water oxygen atoms within 4 Å are updated throughout the simulations. The length of each single short MD run is decided according to the ligand-protein distance ( $r_L$ ) and three cut-off distances ( $D_1$ ,

$D_2$ , and  $D_3$  - panel (a)); (e) after each short MD run,  $r_L$  and the number of water oxygen atoms are plotted: if the slopes ( $m_D$  and  $m_W$ ) of the interpolating linear functions are both positive, then the simulation is considered productive.

### 4.3.5 Analysis of the MD trajectories

Ligand-protein hydrogen bonds and van der Waals contacts were detected using the GetContacts scripts (<https://getcontacts.github.io/index.html>), setting a donor-acceptor distance of 3.3 Å and an angle value of 150° as geometrical cut-offs. Root mean square deviations (RMSD), root mean square fluctuations (RMSF), and dihedral angles were computed using VMD (Humphrey *et al.*, 1996). Distances between atoms were computed using PLUMED 2 (Tribello *et al.*, 2014) or VMD. The MMPBSA.py (Miller *et al.*, 2012) script, from the AmberTools17 suite (The Amber Molecular Dynamics Package, at <http://ambermd.org/>), was used to compute molecular mechanics energies combined with the generalised Born and surface area continuum solvation (MM/GBSA) method (Genheden and Ryde, 2015), after transforming the CHARMM *psf* topology files to an Amber *prmtop* format using ParmEd (ParmEd documentation at <http://parmed.github.io/ParmEd/html/index.html>). The Poisson-Boltzmann surface area (PBSA) was not employed to speed up the computation, *i.e.* the membrane was not implicitly modelled.

## 4.4 Results and Discussion

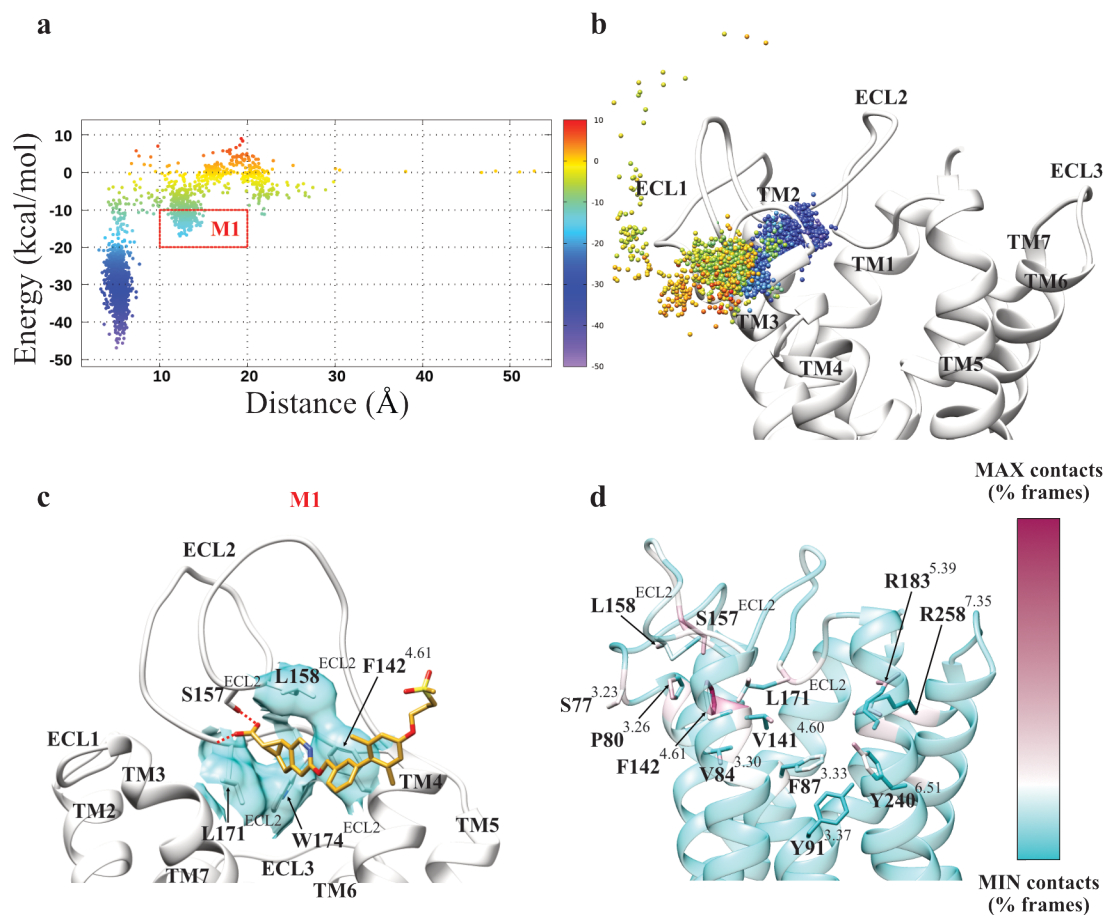
### 4.4.1 Intermediate interactions between MK-8666 and ECL2 drive the FFAR1:MK-8666 complex formation

The structural data indicates a probable MK-8666 binding route to site A1 between TM3 and TM4 (Figure 4.1), no further experiments have yet confirmed this. Indeed, an alternative path between TM4 and TM5 was not ruled out (Srivastava *et al.*, 2014). Overall, the SuMD simulations corroborate the first hypothesis, as a binding path between TM3 and TM4 (Figure 4.4) was sampled without any important steric barrier (differently from AP8 binding to site A2, Figure 4.2). From the unbound state, the ligand approached the FFAR1 extracellular vestibule from the bulk solvent, before moving to the membrane/bulk interface and establishing more extensive contacts with the receptor (Supporting movie SM4.1). Metastable states (M1 in the energy landscape reported in Figure 4.4) were characterised by MK-8666 carboxylate hydrogen bonds with FFAR1 residue S157<sup>ECL2</sup> and the backbone, while hydrophobic contacts were formed with L158<sup>ECL2</sup>, L171<sup>ECL2</sup>, W174<sup>ECL2</sup>, P80<sup>3.26</sup>, and F142<sup>4.61</sup> (Figure 4.4c,d). A conformational rearrangement of ECL2, with W174<sup>ECL2</sup> acting as a sort of gatekeeper, allowed MK-8666 deeper into the receptor through the crevice between TM3 and TM4 to reach site A1, where it formed electrostatic interactions with R183<sup>5.39</sup>, R258<sup>7.35</sup>, and Y240<sup>6.51</sup>. FFAR1 ECL2, which is rich in hydrophilic spots, could drive also the binding of LCFA from the membrane, stabilizing and channelling the polar head groups toward the final bound state. Interestingly, SuMD binding trajectories showed the disruption of an ionic lock between R183<sup>5.39</sup> and E145<sup>4.64</sup> upon binding of the partial agonist (Supporting movie SM4.1); this ionic lock is one of the proposed switches for FFAR1 partial activation (Sum *et al.*, 2009; Lin *et al.*, 2012; Srivastava *et al.*, 2014; Lu *et al.*, 2017). Mutagenesis studies pointed out the E145<sup>4.64</sup> - R183<sup>5.39</sup> and E172<sup>ECL2</sup> - R258<sup>7.35</sup>



salt bridges as possible alternative molecular switches involved in partial agonists binding to site A1 and the inactive/active transition of FFAR1. Our results corroborate this hypothesis, as the simulated association mechanism of the partial agonists MK-8666 showed the rupture of the E145<sup>4.64</sup> - R183<sup>5.39</sup> interaction, and the partial break of the E172<sup>ECL2</sup> - R258<sup>7.35</sup> ionic interaction, upon MK-8666 binding (Supporting movie SM4.1).

## FFAR1:MK-8666 complex



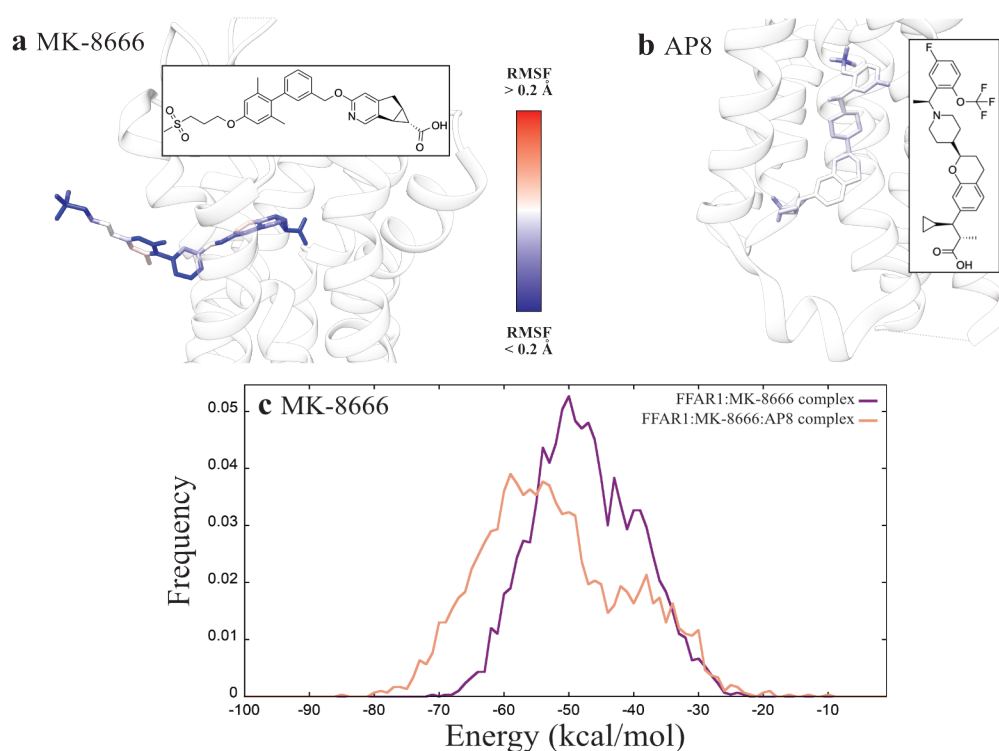
**Figure 4.4 | FFAR1:MK-8666 SuMD binding simulations.** (a) Energy landscape showing the FFAR1:MK-8666 interaction energy during the recognition; the x-axis denotes the distance between the ligand and the receptor centroids; (b) MK-8666 centroid positions coloured according to the interaction energy with FFAR1; (c) a representation of the FFAR1:MK-8666 interactions in the metastable states M1; (d) FFAR1 residues engaged during the simulated binding.

### 4.4.2 MK-8666 and AP8 are reciprocally stabilised during MD simulations

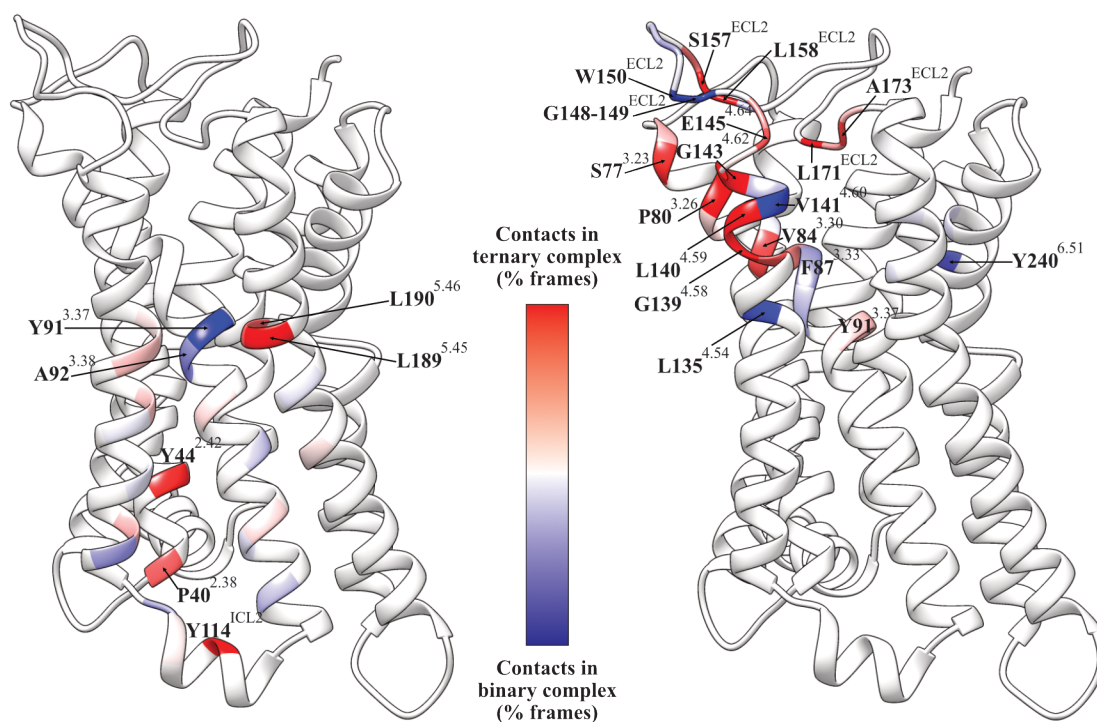
Simulations performed on the FFAR1 binary and ternary complexes (receptor bound to MK-8666 and AP8, Table 3.1, Table 4.1) highlighted several differences in the systems' dynamics. Interestingly, MK-8666 bound to site A1 gained stabilisation in the presence of AP8 (site A2), as shown by the shift of interaction energy towards lower values and the decrease of RMSF (Figure 4.5). This stabilisation could be ascribed to increasing the duration of contacts with FFAR1 residues S77<sup>3.23</sup>, P80<sup>3.26</sup>, V84<sup>3.30</sup>, located at the top of TM3, and L140<sup>4.59</sup> (Figure 4.6).

AP8, on the other hand, even though less influenced, was characterised by a RMSF decrease of the trifluoromethyl group (Figure 4.5b). This part of the molecule is in contact with the backbone of Y91<sup>3.37</sup>, which possibly stabilises the ligand in the ternary complex. A further comparison of the Root Mean Square Fluctuation (RMSF) between the different systems (Figure 3.14) indicates diminished flexibility of ECL2 and the top of TM4 upon binding of MK-8666, possibly influencing the site A2 and therefore AP8 stability. The AP8 interaction pattern in the ternary complex, compared to the FFAR1:AP8 binary one, was characterised by more contacts with L189<sup>5.45</sup>, L190<sup>5.46</sup>, Y144<sup>1CL2</sup>, Y44<sup>2.42</sup>, P40<sup>2.38</sup>, and fewer interactions with A92<sup>3.38</sup> and Y91<sup>3.37</sup>. Interestingly, in the ternary complex MK-8666 formed transient hydrogen bonds with the latter residue, losing interaction with Y240<sup>6.51</sup> (Table 4.2, Figure 4.6).

In our model, the energy stabilization experienced by MK-8666 (Figure 4.5) in the ternary complex is putatively due to increased contacts with the top of TM3 and, to less extent, with TM4 (Figure 4.6). AP8, instead, formed better interactions with residues located at TM2, along with a general reorganization of the contacts with TM3, TM4, and TM5 (Figure 4.6).



**Figure 4.5 | Ligand RMSF (a,b) and GBSA energy (c) distributions.** (a) MK-8666 RMSF comparison between FFAR1:MK-8666 and FFAR1:MK-8666:AP8 complexes; (b) AP8 RMSF comparison between FFAR1:AP8 and FFAR1:MK-8666:AP8 complexes; the RMSF values (a,b) were computed after superposing the MD trajectories on the initial ligand coordinates (blue atoms had lower mobility in the ternary complex, while red atoms had higher mobility). (c) MK-8666 GBSA energy comparison between FFAR1:MK-8666 (purple) and FFAR1:MK-8666:AP8 (pink) complexes. The shift toward lower energy values (of about 10 kcal/mol) of the FFAR1:MK-8666:AP8 complex in the GBSA distribution plot (c) indicates the increase in stability of MK-8666 in the presence of AP8.



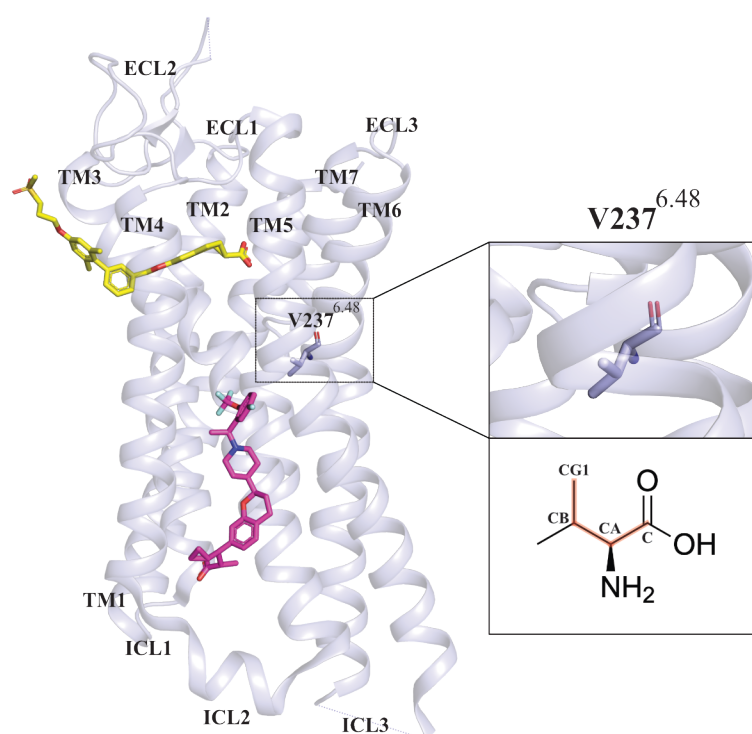
**Figure 4.6 | Protein-ligands contact differences between the FFAR1:MK-8666:AP8 ternary complex and the FFAR1:AP8 (left) and FFAR1:MK-8666 (right) binary complexes.** In this figure, the receptor is shown in cartoon representation. The colour indicates whether a given residue makes more contacts in the binary complex (blue) or more contacts in the ternary complex (red). The contact residues are P40<sup>2.38</sup>, Y44<sup>2.42</sup>, Y91<sup>3.37</sup>, A92<sup>3.38</sup>, Y114<sup>ICL2</sup>, L189<sup>5.45</sup>, and L190<sup>5.46</sup> on the left, and S77<sup>3.23</sup>, P80<sup>3.26</sup>, V84<sup>3.30</sup>, F87<sup>3.33</sup>, Y91<sup>3.37</sup>, L135<sup>4.54</sup>, G139<sup>4.58</sup>, L140<sup>4.59</sup>, V141<sup>4.60</sup>, G143<sup>4.62</sup>, E145<sup>4.64</sup>, G148<sup>ECL2</sup>, G149<sup>ECL2</sup>, W150<sup>ECL2</sup>, S157<sup>ECL2</sup>, L158<sup>ECL2</sup>, L171<sup>ECL2</sup>, A173<sup>ECL2</sup>, and Y240<sup>6.51</sup> on the right.

**Table 4.2 | MK-8666-FFAR1 hydrogen bonds formed in the site A1 in the binary and ternary complex with FFAR1**

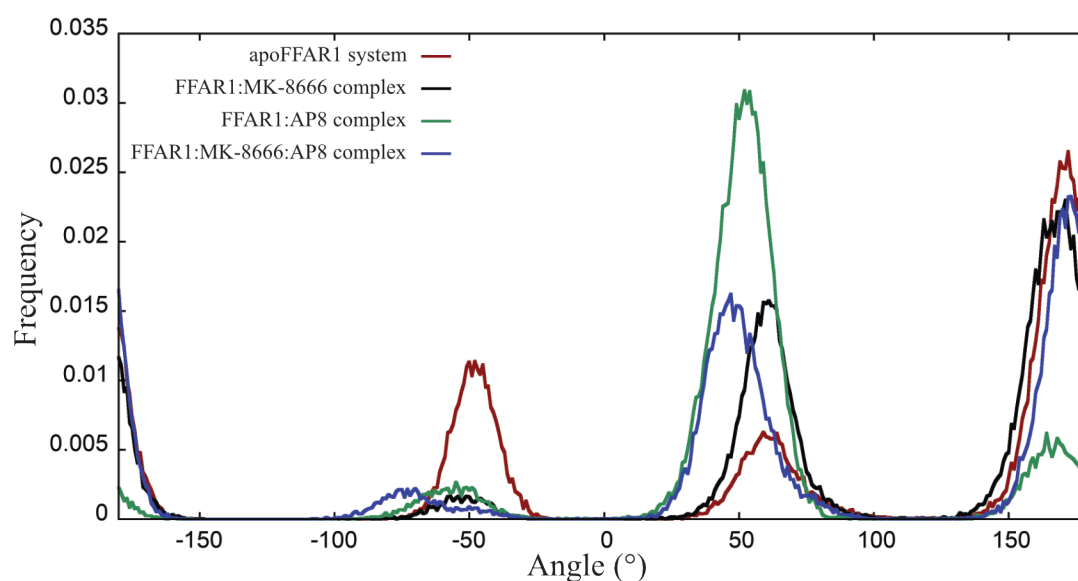
FFAR1 residue	FFAR1:MK-8666 complex (% MD frames)	FFAR1:MK-8666:AP8 complex (% MD frames)
R183 <sup>5.39</sup>	95.1	95.5
R258 <sup>7.35</sup>	89.4	88.8
Y240 <sup>6.51</sup>	70.4	56
N244 <sup>6.55</sup>	1.4	0.1
Y91 <sup>3.37</sup>	1.4	14.7

#### 4.4.3 MK-8666 modifies the dynamics of TM6

The main hallmark of the class A GPCR activation is the outward movement of TM6, which, in the resting position, is stabilised by the ionic lock between R<sup>3.50</sup> and E<sup>6.30</sup> that is common to many GPCRs (Latorraca *et al.*, 2017). In FFAR1 the basic residue K219<sup>6.30</sup> (part of the RRK motif) cannot take part in this inter-helix salt bridge with R104<sup>3.50</sup>. Moreover, TM6 in FFAR1 lacks the “toggle switch” residue W<sup>6.48</sup>, which strongly influences GPCR activation (Katritch *et al.*, 2014); position 6.48 is 81% tryptophan across all class A GPCRs. Position 6.48 in FFAR1 is occupied by V237<sup>6.48</sup>, which lies just above the A2 binding site (Figure 4.7). The comparison of the V237<sup>6.48</sup> dihedral angle distribution shows that the presence of AP8 bound to site A2 (binary complex FFAR1:AP8) changed the rotameric state of this residue (Figure 4.8) blocking a specific configuration, while the presence of MK-8666 favoured a bimodal distribution. However, this conformational state of V237<sup>6.48</sup> did not drive the overall flexibility of TM6. Indeed, as a general view, the presence of MK-8666 increased the FFAR1 TM6 flexibility (Figure 4.9). While the apo receptor and the binary FFAR1:AP8 complex have a similar dynamic in this region, the partial agonists bound to site A1 led to a wider TM6 tilt. It follows that the apoFFAR1 and the FFAR1:AP8 complex appeared constrained compared to the FFAR1:MK-8666 complex. In the presence of both ligands (FFAR1:MK-8666:AP8 complex) TM6 appeared more prone to outward movements, while in the FFAR1:MK-8666 complex TM6 experienced numerous closed conformations (Figure 4.9). This possibly facilitates the recruitment of intracellular effectors (G<sub>i/11</sub>, G<sub>s</sub> proteins, and  $\beta$ -arrestins 1 and 2) as the TM6 outward movement is the key feature shaping the binding site of the cognate proteins.

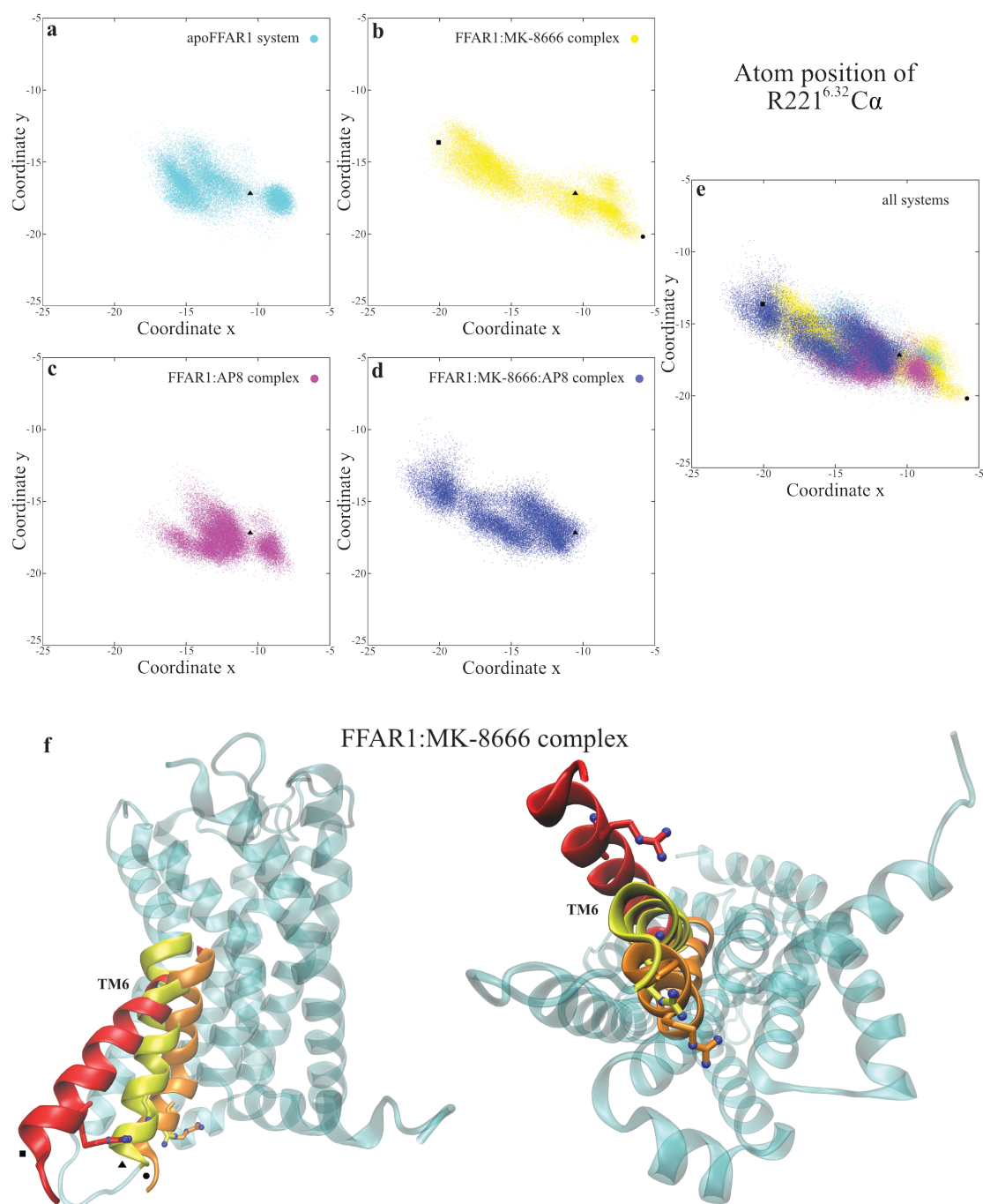


**Figure 4.7 | FFAR1 V237<sup>6.48</sup>.** While position 6.48 in all class A GPCRs is occupied by a tryptophan (which strongly influence the GPCR activation), FFAR1 bears a valine (V237). The dihedral angle of V237<sup>6.48</sup> taken under analysis includes atoms C, CA, CB, and CG1 of the residue. The receptor is shown in cartoon representation (PDB ID: 5TZY in light-blue) with ligands MK-8666 (yellow) and AP8 (magenta) shown in stick representation.



**Figure 4.8 | Dihedral angle of V237<sup>6.48</sup>.** Dihedral angles (atoms C, CA, CB, and CG1) of V237<sup>6.48</sup> from MD simulations of apoFFAR1 system (in red), FFAR1:MK-8666 complex (in black), FFAR1:AP8 complex (in green), and FFAR1:MK-8666:AP8 complex (in blue) were measured using VMD (Humphrey *et al.*, 1996). The main difference between these systems is that the V237<sup>6.48</sup> prefers an angle of about 50° in the FFAR1:AP8 complex while an angle of about 150° is preferred in all the other systems. Moreover, in the apoFFAR1 system, the same residue has the propensity to have an angle of about -50°.





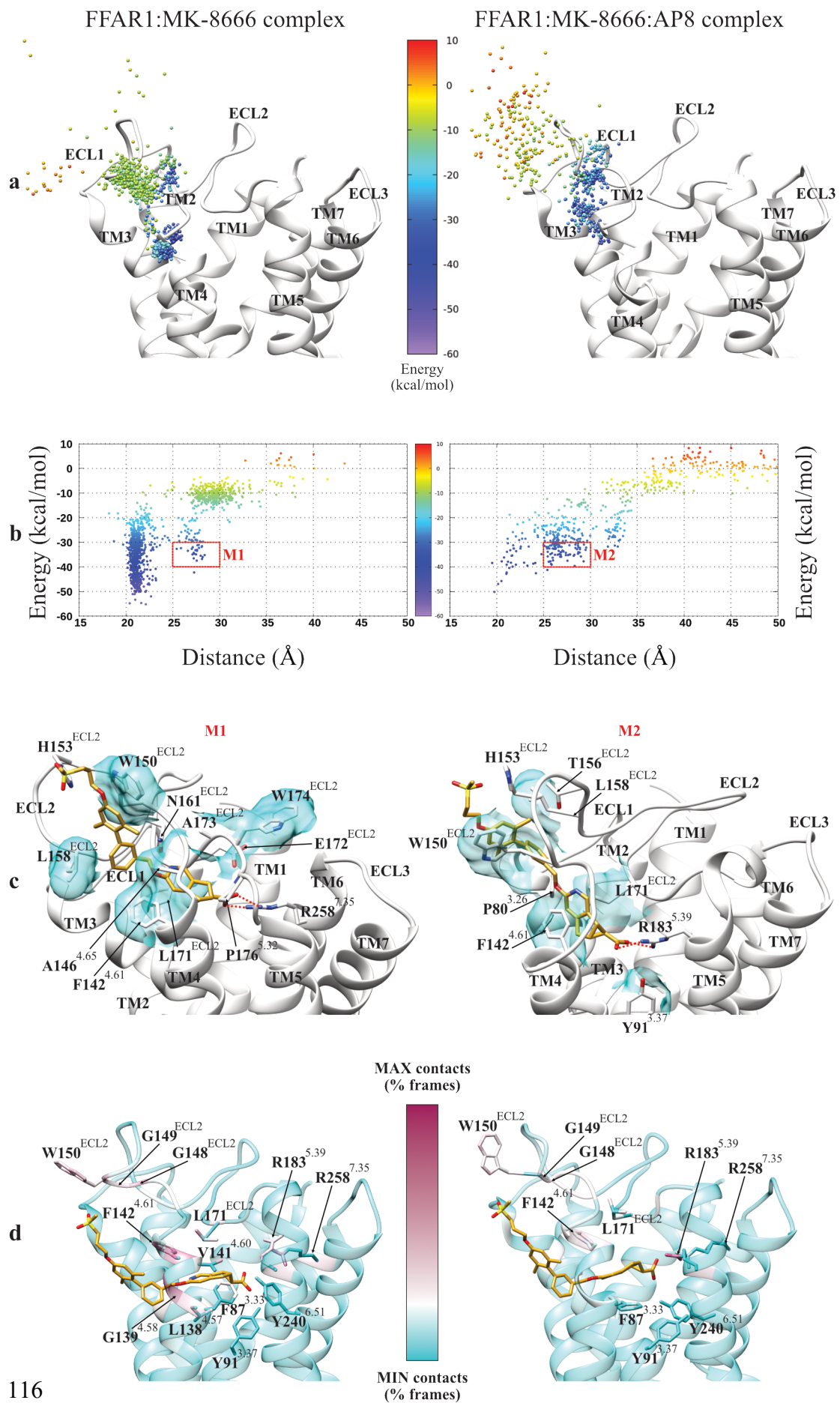
**Figure 4.9 | Atom positions of the  $\text{Ca}$  atom of  $\text{R221}^{6.32}$ .** The position of  $\text{R221}^{6.32}$  (shown in stick representation) was taken to represent the movement of TM6, which is implicated in activation. Plots (a) to (e) present the x and y coordinates of the  $\text{Ca}$  atom of  $\text{R221}^{6.32}$  respectively from (a) the apoFFAR1 (cyan), (b) the FFAR1:MK-8666 complex (yellow), (c) the FFAR1:AP8 complex (magenta), and (d) the FFAR1:MK-8666:AP8 ternary complex (blue). On the right, all systems are reported in the same plot (e). The triangle indicates the original PDB position of  $\text{R221}^{6.32}$  in each system,

while the circle and the square illustrate the position of R221<sup>6,32</sup> respectively in the closest and most open position of TM6 in the FFAR1:MK-8666 complex (**b,f**). TM6 is shown as a ribbon and coloured according to the three positions previously reported (the triangle indicates the original PDB position of R221<sup>6,32</sup>; the circle and the square show the position of R221<sup>6,32</sup> respectively in the closest and most open position of TM6 in the FFAR1:MK-8666 complex).

#### 4.4.4 AP8 influences MK-8666 unbinding from FFAR1

The putative FFAR1:MK-8666 dissociation mechanism (Supporting movie SM4.2), simulated using a modified version of SuMD, was sampled both in the presence and absence of AP8 bound to site A2. In both of these systems, the ligand left receptor site A1 following an unbinding pathway between TM3 and TM4 that resembles the binding pathway (Figure 4.10a). During the unbinding from the binary complex, MK-8666 made more contacts with residues F87<sup>3,33</sup>, V84<sup>3,30</sup>, L138<sup>4,57</sup>, G139<sup>4,58</sup>, V141<sup>4,60</sup>, F142<sup>4,61</sup>, and R258<sup>7,35</sup>, while in the ternary complex the partial agonist (MK-8666) engaged more residues located at the top of TM3 (A83<sup>3,29</sup>, V81<sup>3,27</sup>, P80<sup>3,26</sup>, and S77<sup>3,23</sup>) and on ECL2 (L144<sup>4,63</sup> and L158<sup>ECL2</sup>), as shown in Figure 4.10c and Figure 4.11. This shift in the barycentre of the interactions favoured the retaining of either one of the two electrostatic interactions with R183<sup>5,39</sup> or R258<sup>7,35</sup> during the early step of unbinding. In the absence of AP8, indeed, MK-8666 moved from the bound state straight to the metastable configuration M1 (Figure 4.10a,b), where it engaged R258<sup>7,35</sup>. In the ternary complex, instead, the ligand experienced several metastable states along the path characterised by the ionic interaction with R183<sup>5,39</sup> (metastable states M2 in Figure 4.10a,b). Numerous stable states scattered between the bound state and the final unbound state of the FFAR1:MK-8666:AP8 complex suggest an influence of the full agonist on the dissociation path from site A1.

# CHAPTER 4





Following, the supporting movies (SM4.1-2) of the FFAR1 SuMD simulations are reported.

**Supporting movie SM4.1 | FFAR1:MK-8666 SuMD binding simulations available on the following link: <https://youtu.be/KRBPpjrdBWQ>.** The video shows two SuMD replicas (diverging after 200 ns). The partial agonist MK-8666 (MK6 - van der Waals spheres) reaches the FFAR1 (cyan transparent ribbon) and makes interactions with ECL2 before deepening into the site receptor A1. During Replica 1, the ligand hydrogen bonds (red dashed lines) with R183<sup>5,39</sup> and displace E145<sup>4,64</sup> from the salt bridge. During Replica 2, MK-8666 forms an ionic interaction with R258<sup>7,35</sup>. The MK-8666 experimental bound conformation is reported as transparent grey van der Waals spheres.

**Supporting movie SM4.2 | FFAR1:MK-8666 SuMD unbinding simulations in the absence (left) and presence (right) of AP8 bound to site A2 available on the following link: [https://youtu.be/IdK\\_um2tfQc](https://youtu.be/IdK_um2tfQc).** The video shows the MK-8666 SuMD unbinding from the binary (left) and ternary complex (right). The partial agonist MK-8666 (MK6 - van der Waals spheres) leaves the FFAR1 (cyan transparent ribbon) site A1 following a route between TM3 and TM4 in both complexes. The hydrogen bonds (red dashed lines) are shown. The MK-8666 exit from the site A1 allows the E172<sup>ECL2</sup> - R258<sup>7,35</sup> salt bridge to reform.

## 4.5 Conclusion

The present computational work addresses the activation of FFAR1 by allosteric ligands. FFAR1, which is involved in glycaemic control, diverges from the other class A GPCRs as it presents several distinctive structural features. X-ray structures of FFAR1 unveiled the presence of two distinct binding sites, namely A1 and A2, responsible respectively for the binding of partial and full agonists, and in reciprocal allosteric communication, where site A2 is a novel site on the external lipid facing surface of the TM bundle. FFAR1 does not bear the conserved structural motifs NP<sup>7.50</sup>xxY, DR<sup>3.50</sup>Y, and P<sup>5.50</sup>I<sup>3.40</sup>F<sup>6.44</sup>, the highly conserved “toggle switch” W<sup>6.48</sup> or the ionic lock between R<sup>3.50</sup> and E<sup>6.30</sup> that is common to many class A GPCRs (Latorraca *et al.*, 2017; Weis and Kobilka, 2018). These characteristics, and the absence of a crystallised fully active reference state, suggest that the study of the FFAR1 activation mechanism is challenging. The pursuit of novel drugs able to tackle T2DM will increasingly consider FFAR1, as the structural knowledge of its activation and the allosteric mechanism is being unveiled by means of a plethora of different approaches. The X-ray crystal structures of FFAR1 (Srivastava *et al.*, 2014; Lu *et al.*, 2017; Ho *et al.*, 2018) paved the way for computational studies and rational structure-based drug design. This will likely lead to the development of new chemotypes capable to overcome the hepatotoxicity that prevented the approval of FFAR1 as therapeutics so far.

Our findings about the binding mechanism of TAK-875 to site A1 from the lipid/water interface highlighted the role of the extracellular vestibule, and ECL2 in particular, is in line with observations for the other GPCRs, where ECLs play a fundamental role in ligands binding (and functionality), providing the first interactions

and favourably orienting the incoming molecule prior to the orthosteric site (Peeters *et al.*, 2011; Wheatley *et al.*, 2012).

The allosteric cross-talking between FFAR1 sites A1 and A2 has created difficulties in classifying one of them as the orthosteric one. On one hand, site A1 is located within the TMD, consistent with the orthosteric site of other GPCRs, but trigger a partial activation, site A2, instead, is linked to a full activation but is located outside the TMD at the membrane interface as other GPCR allosteric sites characterised so far. Plausibly, in light of its exposure to the membrane, site A2 could be responsible for LCFA recognition under physiological conditions that are close to homeostasis, while the binding to the less (kinetically) accessible site A1 could take place when the local LCFA concentration increases. Intriguingly, the partial agonist TAK-875 is more effective in recruiting  $\beta$ -arrestins 1 and 2 than the endogenous ligands palmitate or oleate, which instead trigger coupling to the  $G_{q/11}$  path (Mancini *et al.*, 2015). Such deference in signalling profiles may underlie distinct metabolic regulation exerted by site A1 and site A2.

According to MD simulations of the binary FFAR1:AP8 and FFAR1:MK-8666 complexes and the ternary FFAR1:MK-8666:AP8 complex, TM3 and TM4 could be fundamental for the allosteric communication between binding sites. The simulations pointed out a possible role for Y91<sup>3,37</sup> at the interface between the sites A1 and A2 (Teng *et al.*, 2020), as the presence of AP8 slightly favoured the hydrogen bonding with the MK-8666 carboxylate, thereby stabilising the latter.

Focusing on other structural elements of FFAR1, TM6 showed variable degree of flexibility in the different complexes. As a general view, the presence of MK-8666 bound to site A1 increased TM6 mobility, with the greater flexibility occurring in the

ternary complex FFAR1:MK-8666:AP8. In the FFAR1 crystal structures, H8 was not resolved. The lack of structure is likely due to the numerous glycine residues in the primary sequence of the receptor. Simulations confirmed this instability, as the modelled H8 was unstable and unstructured in the many MD replicas. However, a slight tendency to keep helicity was detected when AP8 was bound to A2, probably due to a general stabilization of the intracellular side of the receptor.

Finally, SuMD unbinding of MK-8666 highlighted different mechanisms in the presence or absence of AP8. More precisely, the MK-8666 contacts with the top of TM3 were improved in the presence of AP8, and this may stabilise the ligand and hinder its dissociation, consistent with the slower dissociation rates measured for partial agonists in the ternary FFAR1 complexes (Lin *et al.*, 2012). For the first time, the SuMD protocol was extended to a second metric of the system, beside the distance between the ligand-protein centroids. This implementation could facilitate the unbiased simulation of the unbinding of ligands that, like MK-8666, are particularly stabilised by buried (or numerous) hydrogen bonds.

In conclusion, this work delivered computational insights on some important aspects of the poorly known FFAR1 activation and allostereism, providing a plethora of working hypotheses that I hope will be experimentally explored in future efforts.





## 5. Calcitonin receptor (CTR)\*

(\*) Part of the data in this chapter has been published in the following two papers:

Dal Maso, E. Glukhova, A. Zhu, Y. Garcia-Nafria, J. Tate, C. G. **Atanasio, S.** Reynolds, C. A. Ramírez-Aportela, E. Carazo, J.-M. Hick, C. A. Furness, S. G. B. Hay, D. L. Liang, Y.-L. Miller, L. J. Christopoulos, A. Wang, M.-W. Wootten, D. and Sexton, P. M. (2019). The Molecular Control of Calcitonin Receptor Signaling. *ACS pharmacology & translational science*, **2**(1), pp. 31–51. doi: 10.1021/acsptsci.8b00056.

Pham, V. Zhu, Y. Dal Maso, E. Reynolds, C. A. Deganutti, G. **Atanasio, S.** Hick, C. A. Yang, D. Christopoulos, A. Hay, D. L. Furness, S. G. B. Wang, M. Wootten, D. and Sexton, P. M. (2019). Deconvoluting the Molecular Control of Binding and Signaling at the Amylin 3 Receptor: RAMP3 Alters Signal Propagation through Extracellular Loops of the Calcitonin Receptor. *ACS pharmacology & translational science*. American Chemical Society, **2**(3), pp. 183–197. doi: 10.1021/acsptsci.9b00010.

### 5.1 Abstract

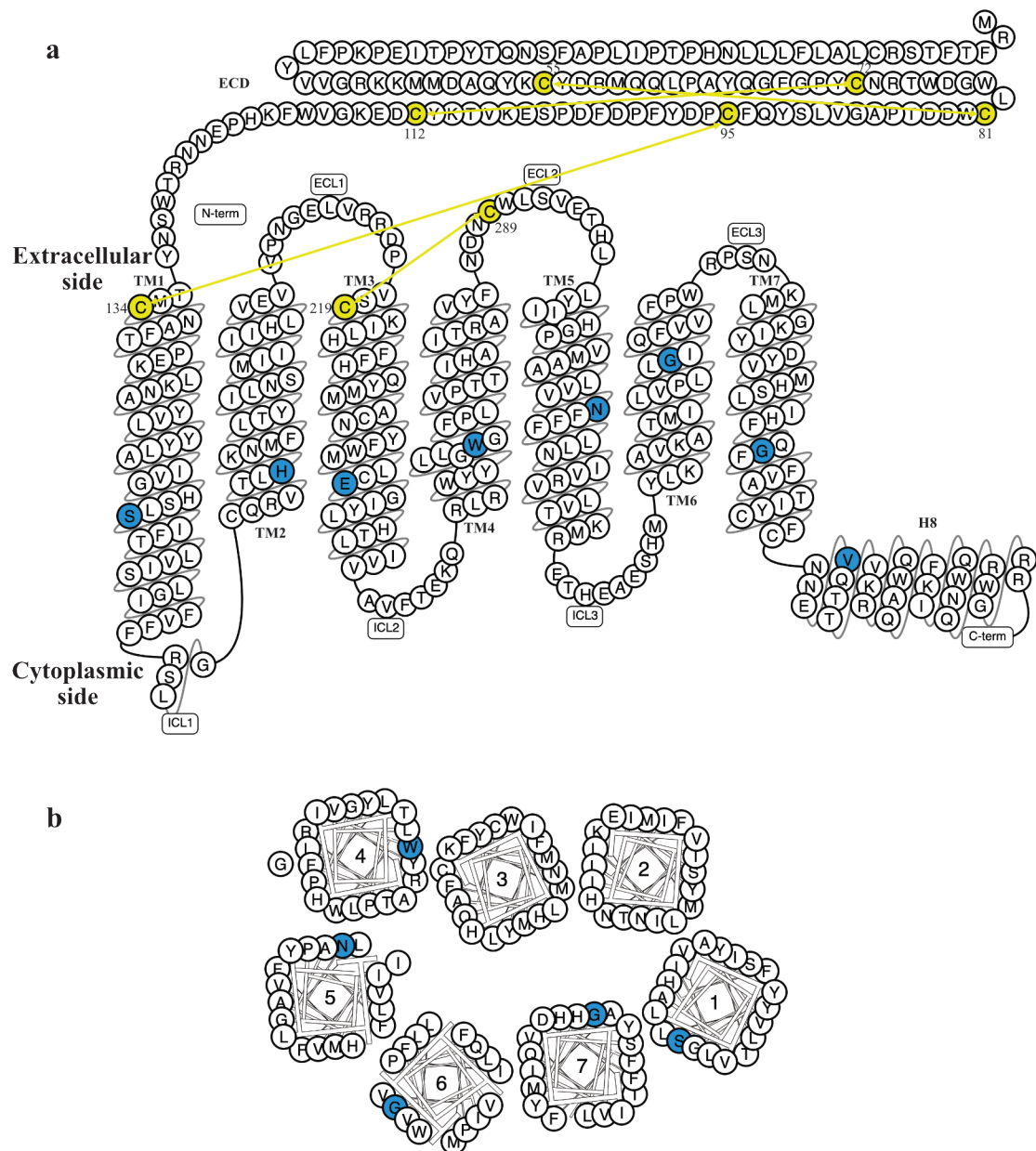
**Background:** The calcitonin receptor (CTR) is a potential G protein-coupled receptor (GPCR) target for the treatment of bone disorders including hypercalcemia of malignancy, osteoporosis, and Paget's disease, as it is highly expressed on osteoclasts. To date, only two structures are available for computer-aided drug design, namely PDB IDs: 6NIY and 5UZ7. **Methods and Aim:** In this study, I employed molecular dynamics (MD) to deliver insights into the CTR structure in complex with (i) the hormone calcitonin (CT) from different organisms such as human (hCT), porcine (pCT), and salmon (sCT), (ii) rat amylin (rAmy), and (iii) human alpha-calcitonin gene-related peptide (hCGRP). Moreover, the allosteric role of the receptor activity-modifying protein (RAMP) was investigated. **Results and Conclusion:** I found that RAMPs reduce the flexibility of the receptor ECD due to different hydrogen bonds with ECL2. Moreover, simulations showed that the stability of CT is higher in salmon than porcine and human. **Keywords:** G protein-coupled receptors (GPCRs); CTR; CT; Amy; hCGRP; RAMP; Molecular dynamics (MD).

## 5.2 Introduction

The calcitonin receptor (CTR) is a G protein-coupled receptor (GPCR) belonging to class B of the Secretin family. Although this receptor is principally expressed in tissues and cells including leucocytes, the brain, kidney, gastrointestinal tract, lung, and reproductive organs (Ostrovskaya *et al.*, 2017), it is best recognised for its expression in bone where it is highly expressed on osteoclasts. Therefore, as a consequence of its role in bone metabolism, studies on CTR have an important role in the comprehension of treatment of bone disorders including hypercalcemia of malignancy, osteoporosis, and Paget's disease (Ostrovskaya *et al.*, 2017).

As in all class B GPCRs, CTR has a two-domain structure which consists of an N-terminal extracellular domain (ECD) that in CTR is 131 amino acids long, and a transmembrane domain (TMD) that is the typical seven transmembrane (7TM) helical domain which distinguishes the GPCRs family (Figure 5.1). The whole structure is stabilised by four conserved disulfide bonds, three involving the ECD (C55-C81, C72-C112, and C95-C134<sup>1,25</sup>), and one connecting the top of TM3 and the ECL2 (C219<sup>3,29</sup>-C289<sup>45,50</sup>) (Figure 5.1). The class B GPCRs have a more open top half compared to the class A GPCRs to allow the binding of larger peptides that occur in two steps: firstly, a high-affinity interaction occurs between the C-terminal half of the peptide and the ECD; secondly, promoted by the first interaction, the N-terminal half of the peptide binds to the TMD, which is a low affinity interaction that is however responsible for receptor activation (Hoare, 2005). CTR interacts with different peptides such as the hormone calcitonin (CT), amylin (Amy), and human alpha-calcitonin gene-related peptide (human  $\alpha$ -CGRP or hCGRP). These peptides principally differ at their C-terminus. While CT ends with a Pro-amide that is critical for activity, amylin ends with

a Tyr-amide and similar to the Phe-amide of CGRP (Pioszak and Hay, 2020). A summary of the peptides is reported in Table 5.1 while sequence details are shown in Figure 5.2. Moreover, CTR can also interact with the receptor activity-modifying protein (RAMP) family that consist of RAMP1, RAMP2, and RAMP3. All three RAMPs have a similar basic structure characterised by a large ECD, a single TMD, and a short cytosolic domain. In this study I analysed CTR in complex with RAMP1 and RAMP3 (Table 5.2 and Figure 5.3). The sequence difference in the RAMP ECD leads the CTR peptide selectivity. A summary of CTR and the analogous calcitonin receptor-like receptor (CLR) in complex with the different RAMPs is reported in Table 5.3 to clarify the differences in terms of the affinity of each peptide for the different complexes. While CTR alone binds CT with a high-affinity, CTR in complex with RAMPs generally prefers to bind Amy. More specifically, the RAMP1:CTR complex (also known as  $AMY_1$  receptor) has high affinity for Amy and CGRP, while the RAMP2:CTR (also known as  $AMY_2$  receptor) and RAMP3:CTR (also known as  $AMY_3$  receptor) are high-affinity Amy receptors with lower affinity for related peptides (Gingell *et al.*, 2016).



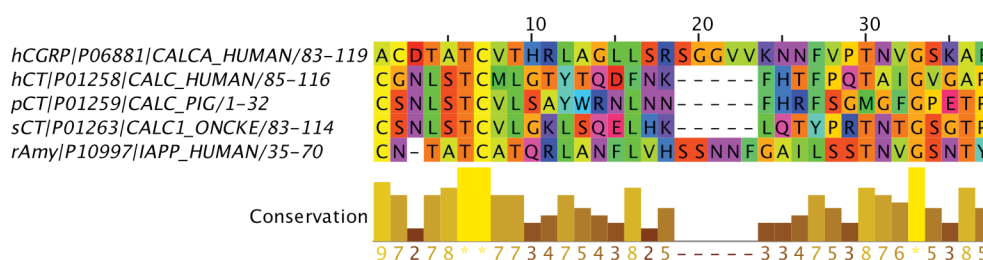
**Figure 5.1 | A two-dimensional model (a) and helix box plot (b) of CTR (class B).** The diagrams show the receptor residue topologies as seen from the side and top, respectively, and the key residues (*i.e.* the most conserved positions) are shown in blue circles. The yellow arrows show the disulfide bonds between C55<sup>ECD</sup>-C81<sup>ECD</sup> (N-term), C72<sup>ECD</sup>-C112<sup>ECD</sup> (N-term), C95<sup>ECD</sup> (N-term)-C134<sup>1.25</sup> (TM1), and C219<sup>3.29</sup> (TM3)-C289<sup>45.50</sup> (ECL2) (adapted figure from Isberg *et al.*, 2015).

### Table 5.1 | Summary of the peptides

Peptide name	Abbreviation	UniProtKB ID	Sequence
human $\alpha$ -CGRP	hCGRP	P06881	ACDTATCVTHRLAGLLSRSGGVVKNFVPTNVGSKAF (#83-119 $\rightarrow$ *1-37)
human CT	hCT	P01258	CGNLSTCMLGTYTQDFNKFHTFPQTAIGVGAP (#85-116 $\rightarrow$ *1-32)
porcine CT	pCT	P01259	CSNLSTCVLSAYWRNLNNFHRFSGMGFGPETP (*1-32)
salmon CT	sCT	P01263	CSNLSTCVLGKLSQELHLKLTYPRTNTGSGTP (#83-114 $\rightarrow$ *1-32)
rat Amylin	rAmy	P10997	KCNTATCATQRLANFLVHSSNNFGAILSSSTNVGSNTY (#35-70 $\rightarrow$ *1-37)

# original numbering

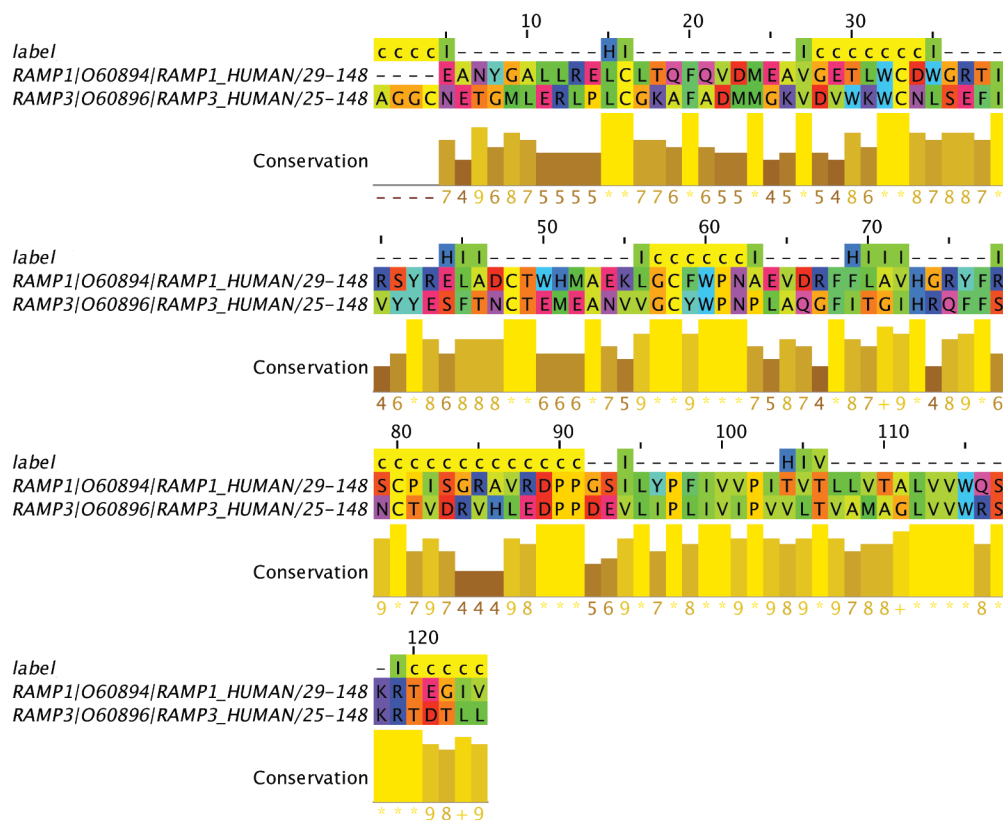
\* new numbering



**Figure 5.2 | Multiple sequence alignment between the peptides.** Residues are coloured in the Taylor scheme, as implemented in Jalview (Clamp *et al.*, 2004), where the colour reflects their biophysical properties as follows: blue, positive; red, negative or small polar; purple, polar; cyan, polar aromatic; green, large hydrophobic; yellow, small hydrophobic. The last line shows the conservation that is an automatically calculated quantitative alignment annotation in Jalview which measures the number of conserved physico-chemical properties conserved for each column of the alignment. Conserved columns are indicated by “\*” (score of 11 with default amino acid property grouping), and columns with mutations where all properties are conserved are marked with a “+” (score of 10, indicating all properties are conserved, not applicable here), then, the score is marked with a number from 9 to 0 (Jalview documentation at *Alignment Conservation Annotation* section).

**Table 5.2 | Summary of the RAMPs**

Protein name	Abbreviation	UniProtKB ID	Sequence
Receptor activity-modifying protein 1	RAMP1	O60894	EANYGALLRELCLTQFQVDMEAVGETLWCDWGRITIRSYRELA DCTWHMAEKLGCFWPNAEVDREFFLAVHGRYFRSCPISGRAVR DPPGSILYPFIVVPIVTLLVLTALVVWQSKRTEGIV (29-148)
Receptor activity-modifying protein 3	RAMP3	O60896	AGGCNETGMLERLPLCGKAFADMMGKVDVWKCWLSEFIVYY ESFTNCTEEMANVVGCVWPPLAQGFITGIHRQFFSNCTVDR VHLEDPPEVLIPLIVIPVVLTVAMAGLVVWRSKRTDTLL (25-148)



**Figure 5.3 | Multiple sequence alignment between RAMPs.** The first line represents the label where “c” stands for coil while HI-HIV are the four helices of the RAMPs structure (HI-HIII helices compose the ECD, whereas HIV helix constitutes the TMD). Residues are coloured in the Taylor scheme, as implemented in Jalview (Clamp *et al.*, 2004), where the colour reflects their biophysical properties as follows: blue, positive; red, negative or small polar; purple, polar; cyan, polar aromatic; green, large hydrophobic; yellow, small hydrophobic. The last line shows the conservation that is an automatically calculated quantitative alignment annotation in Jalview which measures the number of conserved physico-chemical properties conserved for each



column of the alignment. Conserved columns are indicated by “\*” (score of 11 with default amino acid property grouping), and columns with mutations where all properties are conserved are marked with a “+” (score of 10, indicating all properties are conserved), then, the score is marked with a number from 9 to 0 (Jalview documentation at *Alignment Conservation Annotation* section).

**Table 5.3 | A summary of the peptide affinity based on different GPCR-RAMP complex formation**

GPCR	RAMP	Peptide affinity	Receptor Name <sup>#</sup>
CTR	-	CT >> Amy/CGRP	-
CTR	1	Amy = CGRP > CT	AMY <sub>1</sub>
CTR	2	Amy > CGRP/CT	AMY <sub>2</sub>
CTR	3	Amy > CGRP/CT	AMY <sub>3</sub>
CLR	1	CGRP >> AM	CGRP
CLR	2	AM >> AM2	AM <sub>1</sub>
CLR	3	AM = AM2	AM <sub>2</sub>

<sup>#</sup> receptor nomenclature resulting from the association of individual RAMPs with either CTR or CLR

AM=Adrenomedullin

AM2=Adrenomedullin 2 or IMD=Intermedin

## 5.3 Materials and Methods

In this study, the receptor CTR structure with(out) RAMP1 or RAMP3, in complex with  $G_{(\alpha_{373-394})}$  and the respective peptide (hCGRP, hCT, pCT, sCT, and rAmy), was completed using Modeller (as explained in section 2.2): the missing loops were built from the cryo-EM structures of the CGRPR (which includes RAMP1, PDB ID: 6E3Y) and CTR (PDB ID: 6NIY), while the ECD was built from the adrenomedullin receptor (PDB ID: 4RWF). The best structure obtained from Modeller (*i.e.* the structure with the best discrete optimised protein energy (DOPE) score out of 1000 generated models) were used for the system preparation for MD simulations.

The MD systems for CTR structure were prepared using the calcitonin receptor- $G_s$  complex (PDB ID: 5UZ7) as a representative structure for the coordinates as explained in section 2.3.4 with the difference that a larger simulation box (106 Å x 106 Å x 136 Å) was used and a longer equilibration (*first stage*: 2500 conjugate-gradient minimisation steps; *second stage*: 2 ns; *third stage*: 33 ns; *fourth stage*: 35 ns) was performed. Table 5.4 summarises all the systems prepared and the simulations performed.

The MD system equilibration and production stages are reported along with the analysis performed in section 2.3.6 and 2.3.7, respectively.

The RAMPs influence on van der Waals contacts and hydrogen bonds was evaluated by computing the total numbers of contacts/hydrogen bonds between (i) peptide and CTR/RAMP and (ii) RAMP and CTR side chain during the simulations. In particular, the influence of the presence and absence of RAMP3 was analysed with the sCT peptide in the sCT:CTR: $G_\alpha$  (#1) and sCT:RAMP3:CTR: $G_\alpha$  (#5) system, while

the difference between RAMP1 and RAMP3 was analysed with rAmy peptide in the rAmy:RAMP1:CTR:G $\alpha$  (#6) and rAmy:RAMP3:CTR:G $\alpha$  (#7) system.

**Table 5.4 | Summary of the CTR MD simulations.** For each system, 3 replicas of a 1  $\mu$ s long MD simulation were performed. CTR denotes the part L40<sup>ECD</sup>-N414<sup>8.66</sup> of the receptor.  $\alpha$ 373-394 denotes the C terminal helix (R373 - L394) of the G protein  $\alpha$  subunit

System	Peptide	Complex	Disulfide bonds Cys-Cys between:			Number of replicas	Total MD sampling time [ $\mu$ s]
			CTR-CTR	CTR-RAMP	Peptide-Peptide		
#1	sCT <sup>d</sup>	CTR:G( $\alpha$ 373-394)	55-81 72-112 95-134 219-289	/	1-7	3	3
#2	hCGRP <sup>a</sup>	RAMP3:CTR:G( $\alpha$ 373-394)	55-81 72-112 95-134 219-289	28-82 40-72 57-104	2-7	3	3
#3	hCT <sup>b</sup>	RAMP3:CTR:G( $\alpha$ 373-394)	55-81 72-112 95-134 219-289	28-82 40-72 57-104	1-7	3	3
#4	pCT <sup>c</sup>	RAMP3:CTR:G( $\alpha$ 373-394)	55-81 72-112 95-134 219-289	28-82 40-72 57-104	1-7	3	3
#5	sCT <sup>d</sup>	RAMP3:CTR:G( $\alpha$ 373-394)	55-81 72-112 95-134 219-289	28-82 40-72 57-104	1-7	3	3
#6	rAmy <sup>e</sup>	RAMP1:CTR:G( $\alpha$ 373-394)	55-81 72-112 95-134 219-289	40-72 57-104	2-7	3	3
#7	rAmy <sup>e</sup>	RAMP3:CTR:G( $\alpha$ 373-394)	55-81 72-112 95-134 219-289	28-82 40-72 57-104	2-7	3	3

<sup>a</sup> human  $\alpha$ -CGRP

<sup>b</sup> human CT

<sup>c</sup> porcine CT

<sup>d</sup> salmon CT

<sup>e</sup> rat Amylin

## 5.4 Results and Discussion

### 5.4.1 Geometric analysis: RMSD and RMSF

The geometric analysis of the seven systems reported in Table 5.4 is described below and presented in Figures 5.4-5.8.

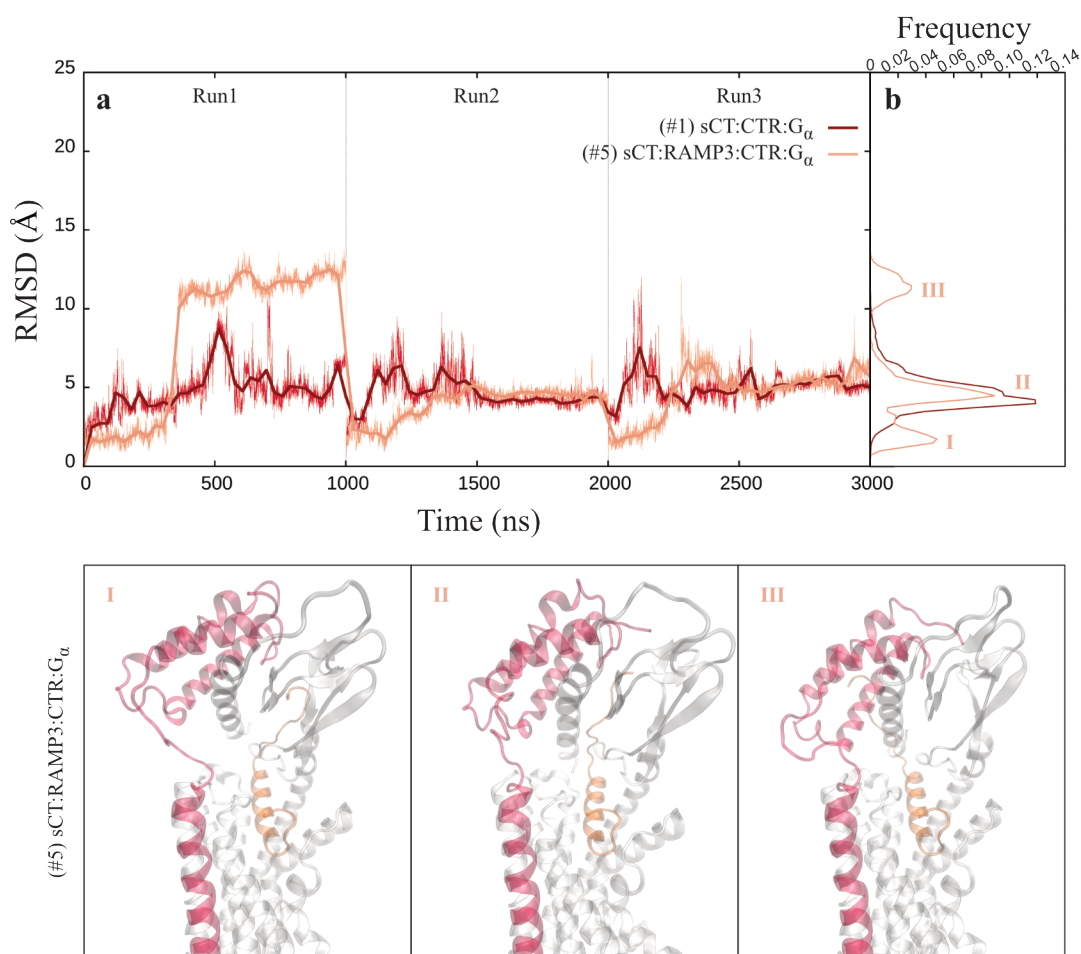
The RMSD comparison of sCT in the absence and presence of RAMP3, respectively in the sCT:CTR:Ga (#1; red) and sCT:RAMP3:CTR:Ga (#5; pink) is shown in Figure 5.4. The ECD of the receptor is the most flexible part during the simulations as shown in Figure 5.8a. The presence of RAMP3 affects the stability of the ECD of the receptor as expected since it applies a restraint to it. Therefore, RAMP3 seems to slow down the arrival at the plateau achieved in the second part of the simulations and this emerges as a peak at low peptide RMSD values (Figure 5.4bI). Different trend appearances during the Run1 where, after the initial slowdown of the changes in the RMSD value, the peptide in presence of the RAMP3 shows a conformational change at about 300 ns (Run1). The main difference in peptide conformation is the position of its C-terminal which switches from interacting with the ECD of the receptor to interacting with RAMP3, corresponding to peak III of Figure 5.4.

The comparison of all the systems in complex with CT belonging to the different organisms (Figure 5.5) shows that pCT gives the most variable bound peptide structures (RMSD ranges from 2 Å to 18 Å), therefore, pCT reaches higher values of RMSD compared to sCT and hCT. Two main conformations, corresponding to peaks I and II of the pCT RMSD system (Figure 5.5), denote once again a more flexible C-terminal peptide. While hCT and pCT have a similar influence on the RMSF of the receptor, the presence of sCT decreases the RMSF of the whole protein (Figure 5.8b).

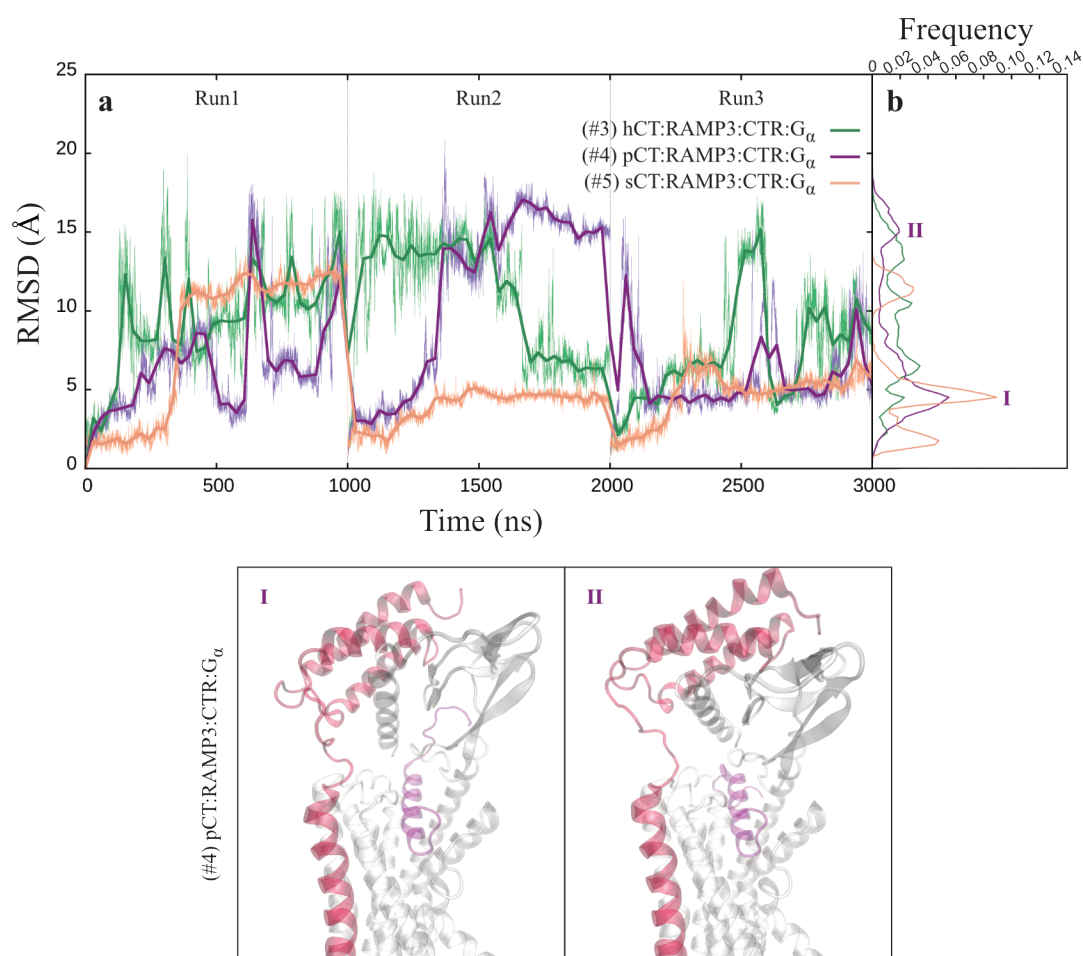
In particular, there is a visible reduction in the RMSF at the ECD, the top of TM1, the bottom of TM3, ICL2, the bottom of TM4, TM5, ICL3, TM6, ECL3, and the top of TM7.

In Figure 5.6, all systems composed of CTR in complex with RAMP3 are presented. This comparison once again highlights that sCT has lower RMSD values compared to the other peptides and this is in line with experimental data that suggest a higher affinity for the sCT. hCGRP follows sCT with low RMSD values showing a main peak of about 7 Å (Figure 5.6bI). The RMSF comparison of these systems is shown in Figure 5.8c. Noteworthy is the rAmy:RAMP3:CTR:Gα (#7) system that has very high RMSF values at the ECD and the lowest values at the TMD.

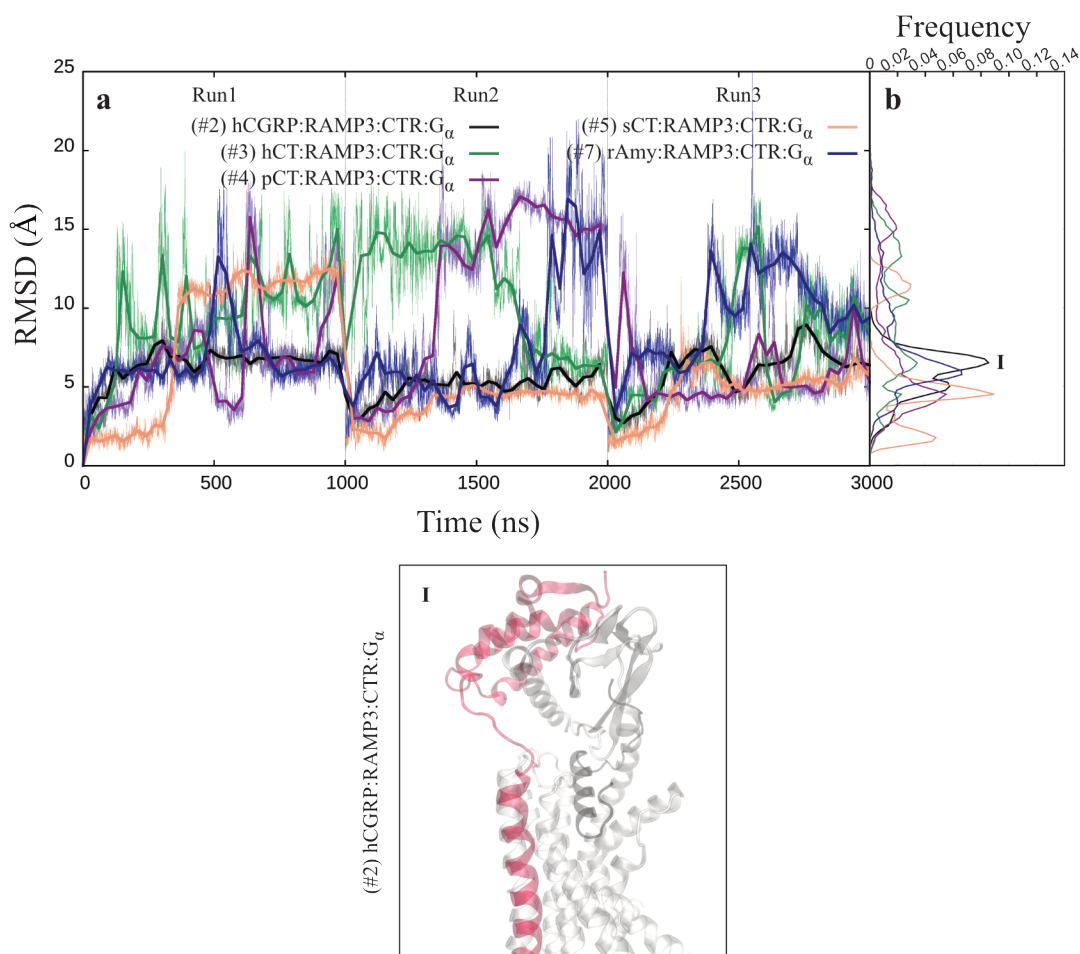
In Figure 5.7 CTR in complex with rAmy but a different RAMP is analysed. The effect of different RAMPs does not appear to significantly affect the stability of the peptide. rAmy:RAMP3:CTR:Gα (#7) system has a slightly more widespread RMSD, but the plot shows a similar trend for both systems with the two main maxima more or less aligned. On the contrary, the presence of RAMP1 seems to stabilise the whole ECD as shown in Figure 5.8d.



**Figure 5.4 | RMSD analysis.** RMSD plots (a) and distributions (b) from three MD simulations (Runs 1-3) of sCT peptide in the absence and presence of RAMP3, respectively in the sCT:CTR:G $\alpha$  (#1; red) and sCT:RAMP3:CTR:G $\alpha$  (#5; pink) systems. Panels I, II, and III show the system conformation corresponding to the three peaks obtained in the RMSD distribution (b) for sCT:RAMP3:CTR:G $\alpha$  (#5; pink) system. CTR, RAMP3, and sCT are shown in cartoon representation in white, red, and pink, respectively.

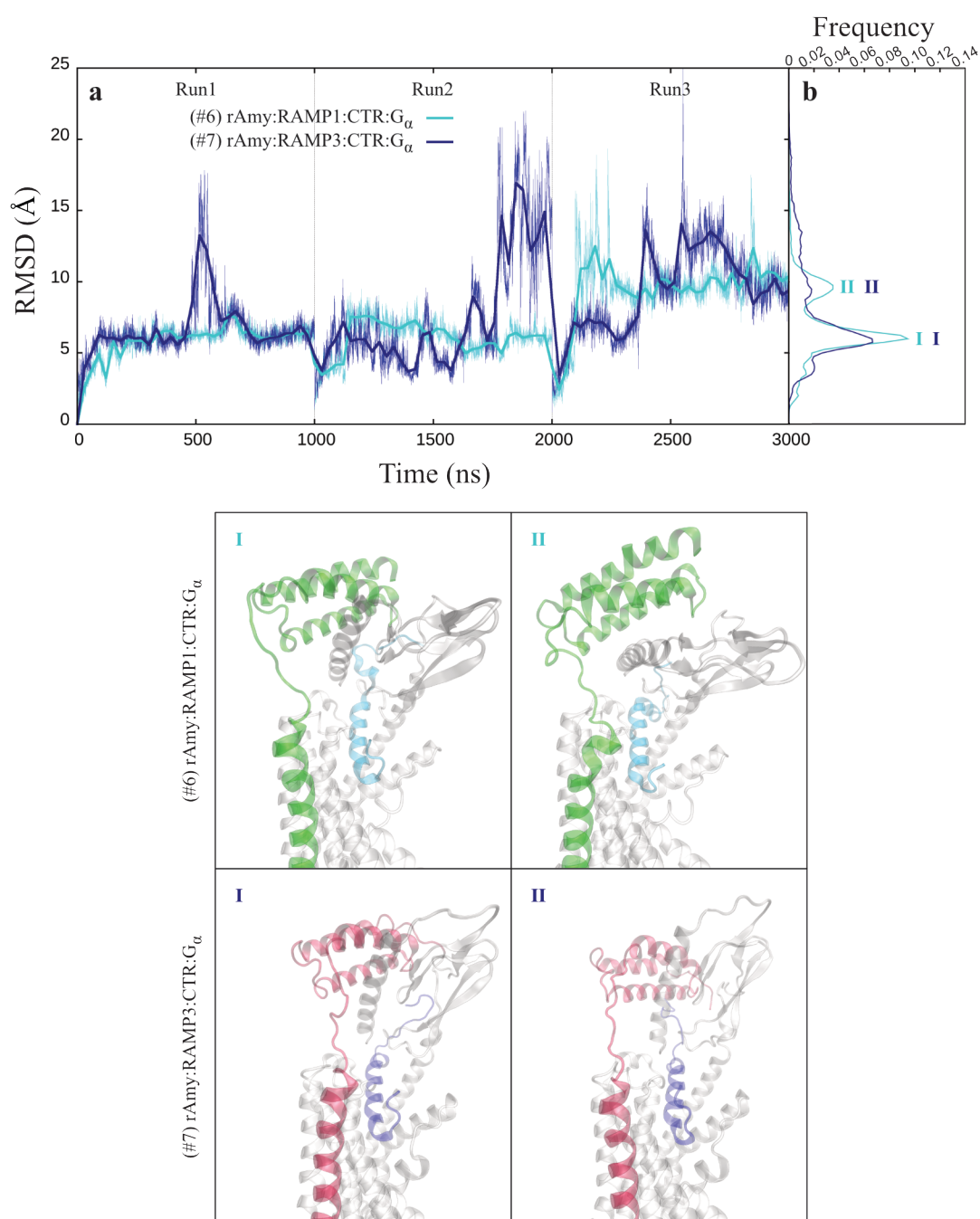


**Figure 5.5 | RMSD analysis.** RMSD plots (a) and distributions (b) from three MD simulations (Runs 1-3) of different organism CT peptides, *i.e.* human, porcine, and salmon, respectively in the hCT:RAMP3:CTR:G $\alpha$  (#3; green), pCT:RAMP3:CTR:G $\alpha$  (#4; purple), and sCT:RAMP3:CTR:G $\alpha$  (#5; pink) systems. Panels I and II show the system conformation corresponding to the two peaks obtained in the RMSD distribution (b) for pCT:RAMP3:CTR:G $\alpha$  (#4; purple) system. CTR, RAMP3, and pCT are shown in cartoon representation in white, red, and purple, respectively.



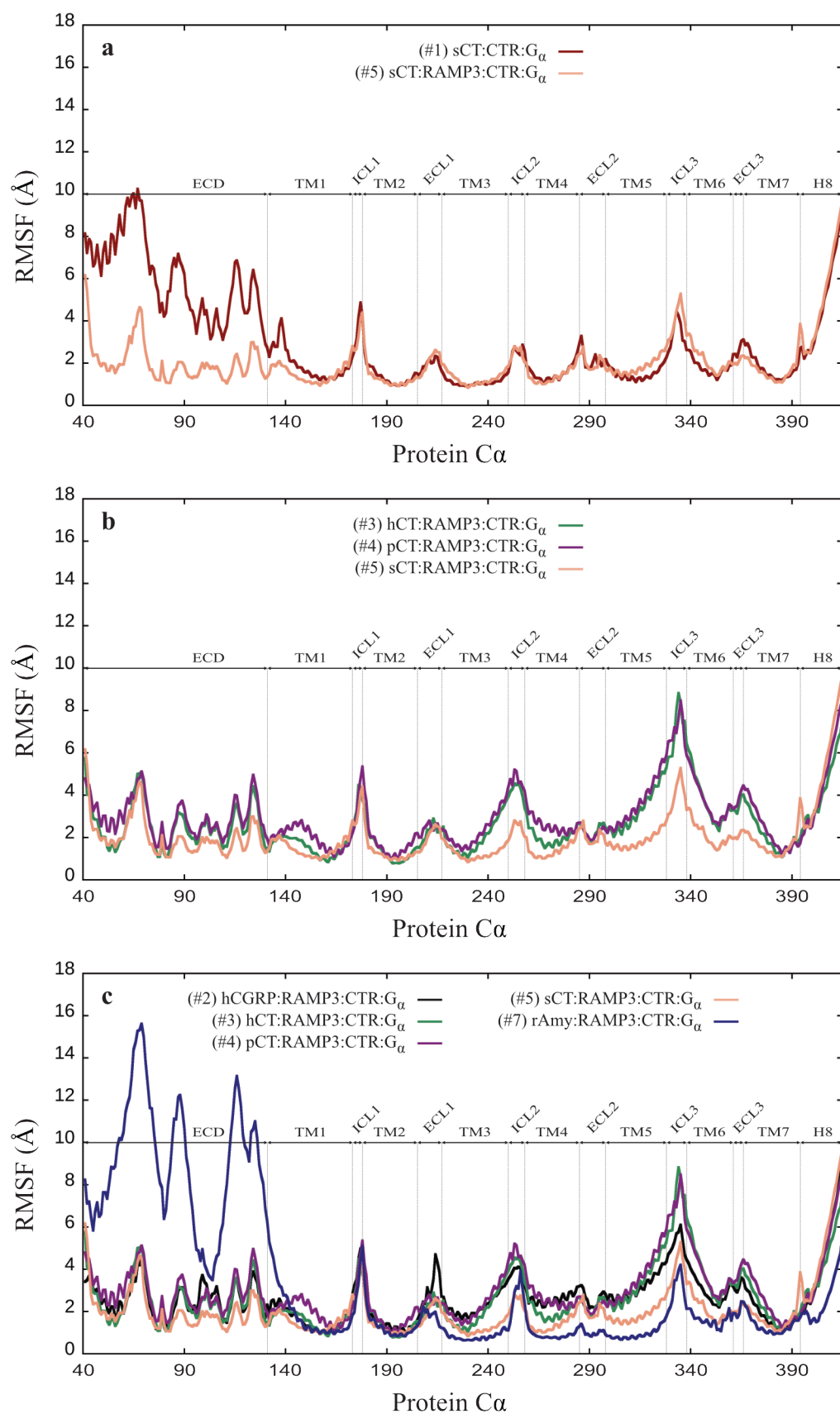
**Figure 5.6 | RMSD analysis.** RMSD plots (a) and distributions (b) from three MD simulations (Runs 1-3) of different peptides in complex with RAMP3, *i.e.* hCGRP:RAMP3:CTR:G $\alpha$  (#2; black), hCT:RAMP3:CTR:G $\alpha$  (#3; green), pCT:RAMP3:CTR:G $\alpha$  (#4; purple), sCT:RAMP3:CTR:G $\alpha$  (#5; pink), and rAmy:RAMP3:CTR:G $\alpha$  (#7; blue) systems. Panel I shows the system conformation corresponding to the main peak obtained in the RMSD distribution (b) for the hCGRP:RAMP3:CTR:G $\alpha$  (#2; black) system. CTR, RAMP3, and hCGRP are shown in cartoon representation in white, red, and black, respectively.

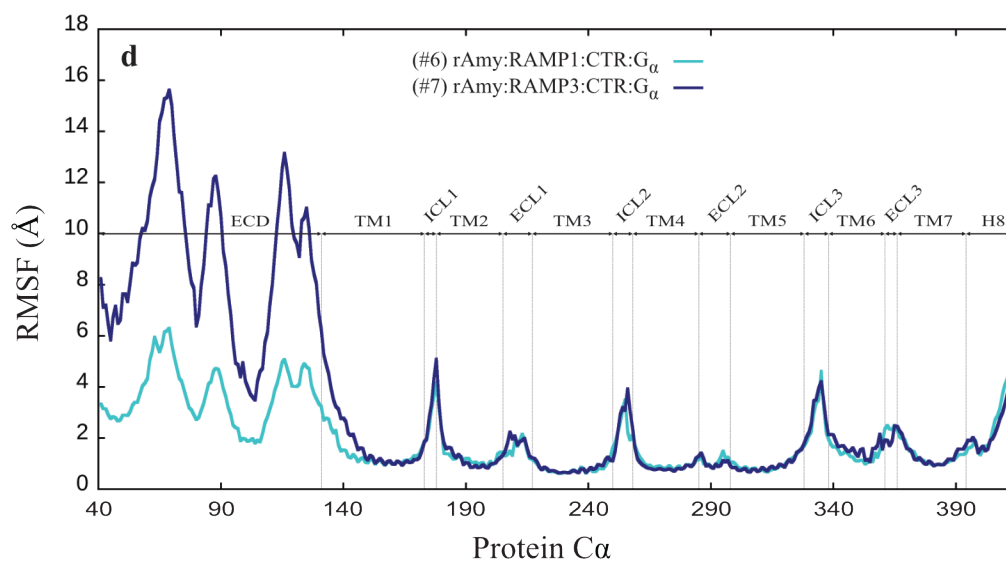




**Figure 5.7 | RMSD analysis.** RMSD plots (a) and distributions (b) from three MD simulations (Runs 1-3) of rAmy in complex with different RAMPs, *i.e.* RAMP1 and RAMP3, respectively in the rAmy:RAMP1:CTR:G $\alpha$  (#6; cyan) and rAmy:RAMP3:CTR:G $\alpha$  (#7; blue) systems. Panels I and II cyan-blue show the system conformation corresponding to the two peaks obtained in the RMSD distribution (b) for rAmy:RAMP1:CTR:G $\alpha$  (#6; cyan) and rAmy:RAMP3:CTR:G $\alpha$  (#7; blue) systems.

CTR, RAMP1, RAMP3, and rAmy are shown in cartoon representation in white, green, red, and cyan-blue, respectively.

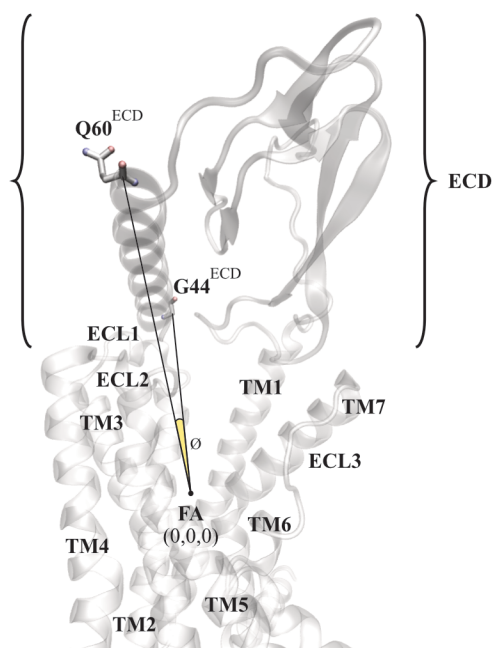




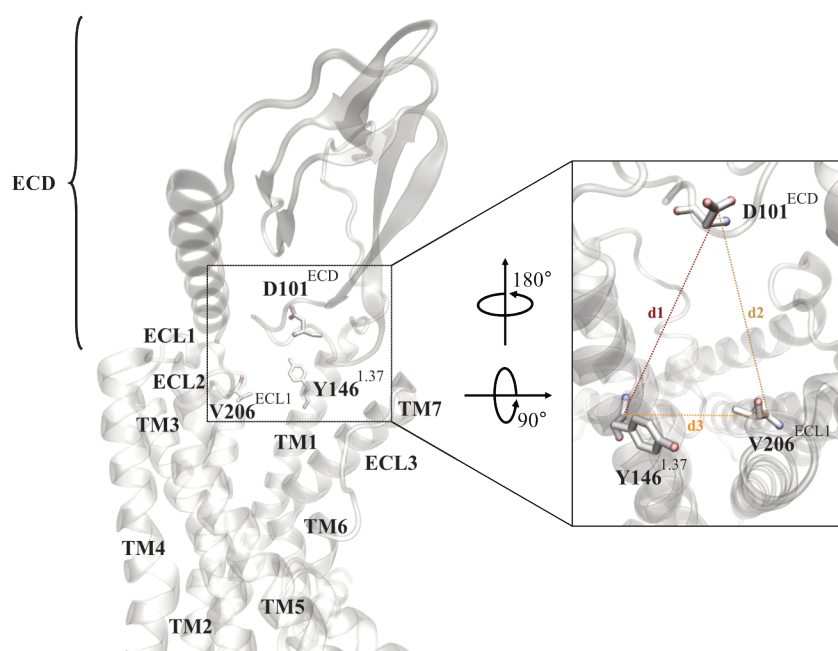
**Figure 5.8 | RMSF analysis.** RMSF plots of the receptor from three MD simulations (Runs 1-3) of each CTR complex summarised in Table 5.4. **(a)** RMSF comparison of CTR in complex with sCT in the absence and presence of RAMP3, respectively in the sCT:CTR:G $\alpha$  (#1; red) and sCT:RAMP3:CTR:G $\alpha$  (#5; pink) systems. **(b)** RMSF comparison of CTR in complex with RAMP3 and different organism CT peptides, *i.e.* human, porcine, and salmon, respectively in the hCT:RAMP3:CTR:G $\alpha$  (#3; green), pCT:RAMP3:CTR:G $\alpha$  (#4; purple), and sCT:RAMP3:CTR:G $\alpha$  (#5; pink) systems. **(c)** RMSF comparison of CTR in complex with RAMP3 and different peptides: hCGRP:RAMP3:CTR:G $\alpha$  (#2; black), hCT:RAMP3:CTR:G $\alpha$  (#3; green), pCT:RAMP3:CTR:G $\alpha$  (#4; purple), sCT:RAMP3:CTR:G $\alpha$  (#5; pink), and rAmy:RAMP3:CTR:G $\alpha$  (#7; blue) systems. **(d)** RMSF comparison of CTR in complex with rAmy and different RAMPs, *i.e.* RAMP1 and RAMP3, respectively in the rAmy:RAMP1:CTR:G $\alpha$  (#6; cyan) and rAmy:RAMP3:CTR:G $\alpha$  (#7; blue) systems.

### 5.4.2 Geometric analysis: Angles, Distances, and Centroids

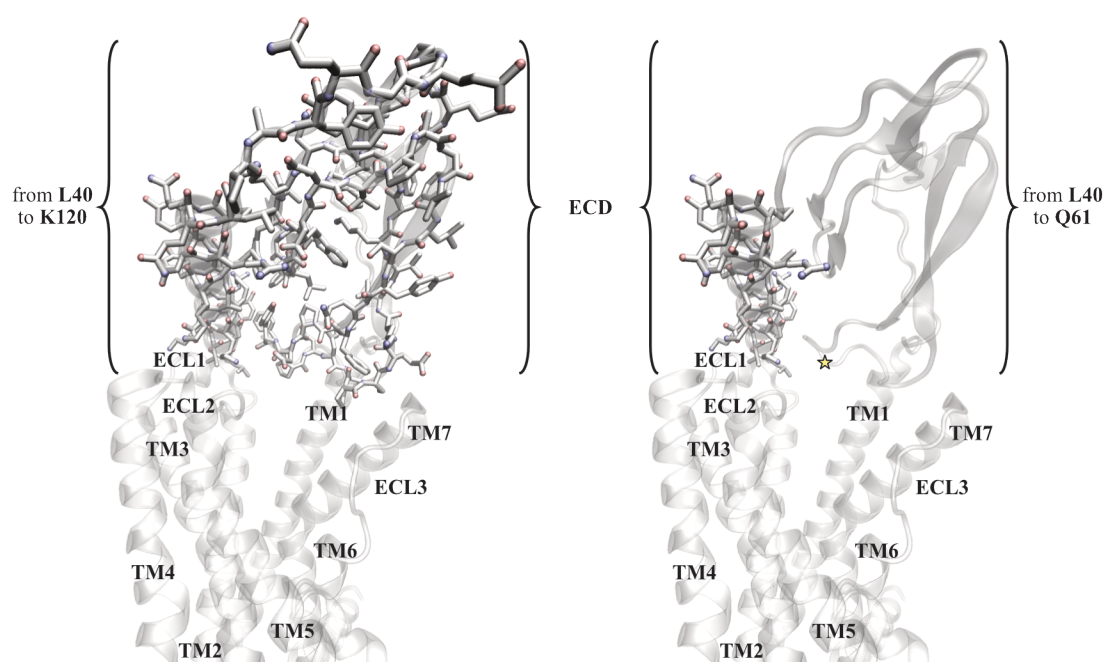
To understand the influence that different RAMPs and/or peptides have to the movement of the receptor ECD, several analyses were conducted. Firstly, the angle  $\emptyset$  between the C $\alpha$  atom of G44<sup>ECD</sup> (corresponding to the start of the ECD helix), a fixed Atom (FA, 0,0,0) (corresponding to the receptor centre of mass), and the C $\alpha$  atom of Q60<sup>ECD</sup> (corresponding to the end of the ECD helix) was analysed (Figure 5.9). Secondly, the distances between three residues, namely D101<sup>ECD</sup>, Y146<sup>1.37</sup>, and V206<sup>ECL1</sup> (d1=D101<sup>ECD</sup>-Y146<sup>1.37</sup>; d2=D101<sup>ECD</sup>-V206<sup>ECL1</sup>; and d3=Y146<sup>1.37</sup>-V206<sup>ECL1</sup>), were analysed (Figure 5.10). Lastly, the centroid analysis of the whole ECD (L40-K120), the ECD helix (L40-Q61), and the single residue P100 sitting in the loop that looks at the receptor cavity, were conducted (Figure 5.11).



**Figure 5.9 | Angle  $\emptyset$ .** The ECD movement during the MD simulations was analysed considering the variation of the angle ( $\emptyset$ ; yellow) between the C $\alpha$  atom of G44<sup>ECD</sup> (start of ECD helix), a fixed Atom (FA, 0,0,0) (receptor centre of mass), and the C $\alpha$  atom of Q60<sup>ECD</sup> (end of ECD helix). Important residues are shown in stick representation, while CTR is shown in cartoon representation.



**Figure 5.10 | Distances.** Three distances d1 (brown), d2 (gold), and d3 (orange) between D101<sup>ECD</sup>, Y146<sup>1.37</sup>, and V206<sup>ECL1</sup> (d1=D101<sup>ECD</sup>-Y146<sup>1.37</sup>; d2=D101<sup>ECD</sup>-V206<sup>ECL1</sup>; and d3=Y146<sup>1.37</sup>-V206<sup>ECL1</sup>) were calculated to analyse the ECD movement during MD simulations. Important residues are shown in stick representation, while CTR is shown in cartoon representation.



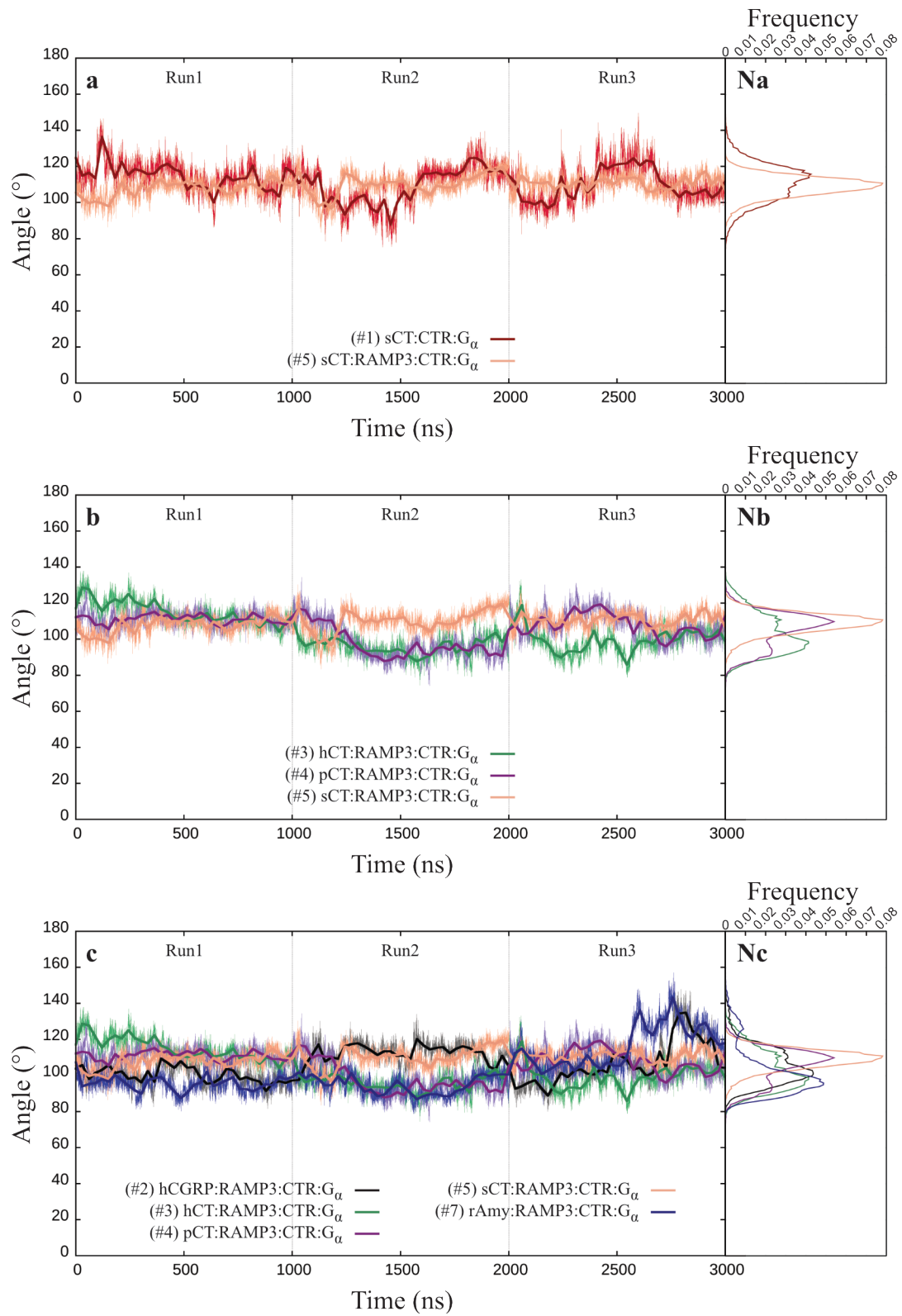
**Figure 5.11 | ECD selection for centroid analysis.** The ECD movement was analysed with three selections, *i.e.* (i) the whole ECD from L40 to K120, (ii) the helix of ECD from L40 to Q61, and (iii) the residue P100 (indicated by a yellow star). Important residues are shown in stick representation, while CTR is shown in cartoon representation.

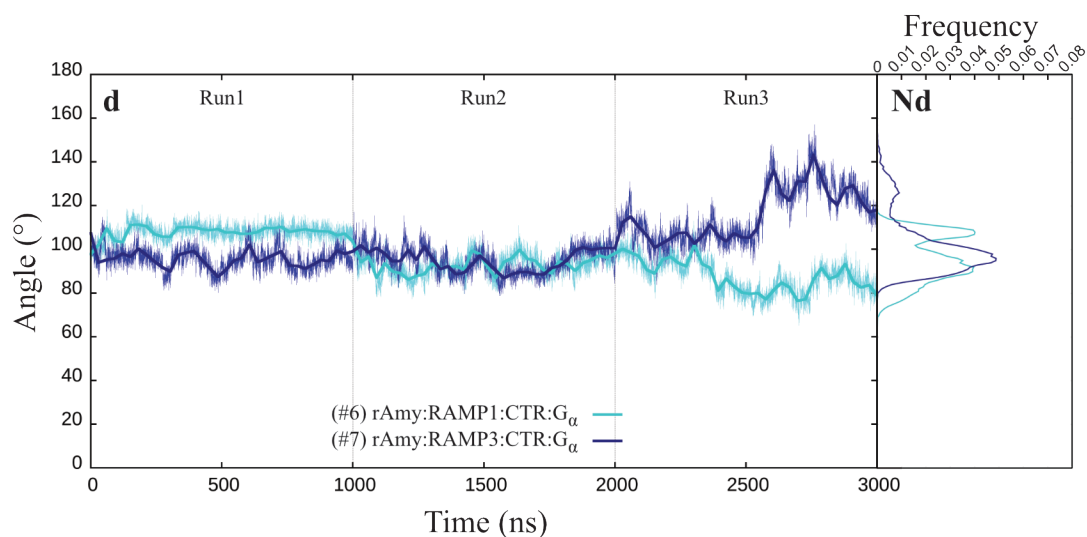


The analysis of the angle  $\emptyset$  (G44<sup>ECD</sup>-FA-Q60<sup>ECD</sup>; Figure 5.9 and Figure 5.12) shows that sCT:CTR:G $\alpha$  (#1; red) system has the most variable angle and this is consistent with the absence of the RAMP in the system (Figure 5.12a). Contrariwise, a very similar angle occurs for the three systems that have CT peptide (Figure 5.12b). Noteworthy is rAmy:RAMP3:CTR:G $\alpha$  (#7; blue) system where the angle changes from being more closed in Run1 to being more open in Run3 (Figure 5.12c,d). These high fluctuations lead to a higher RMSF plot (Figure 5.8c,d).

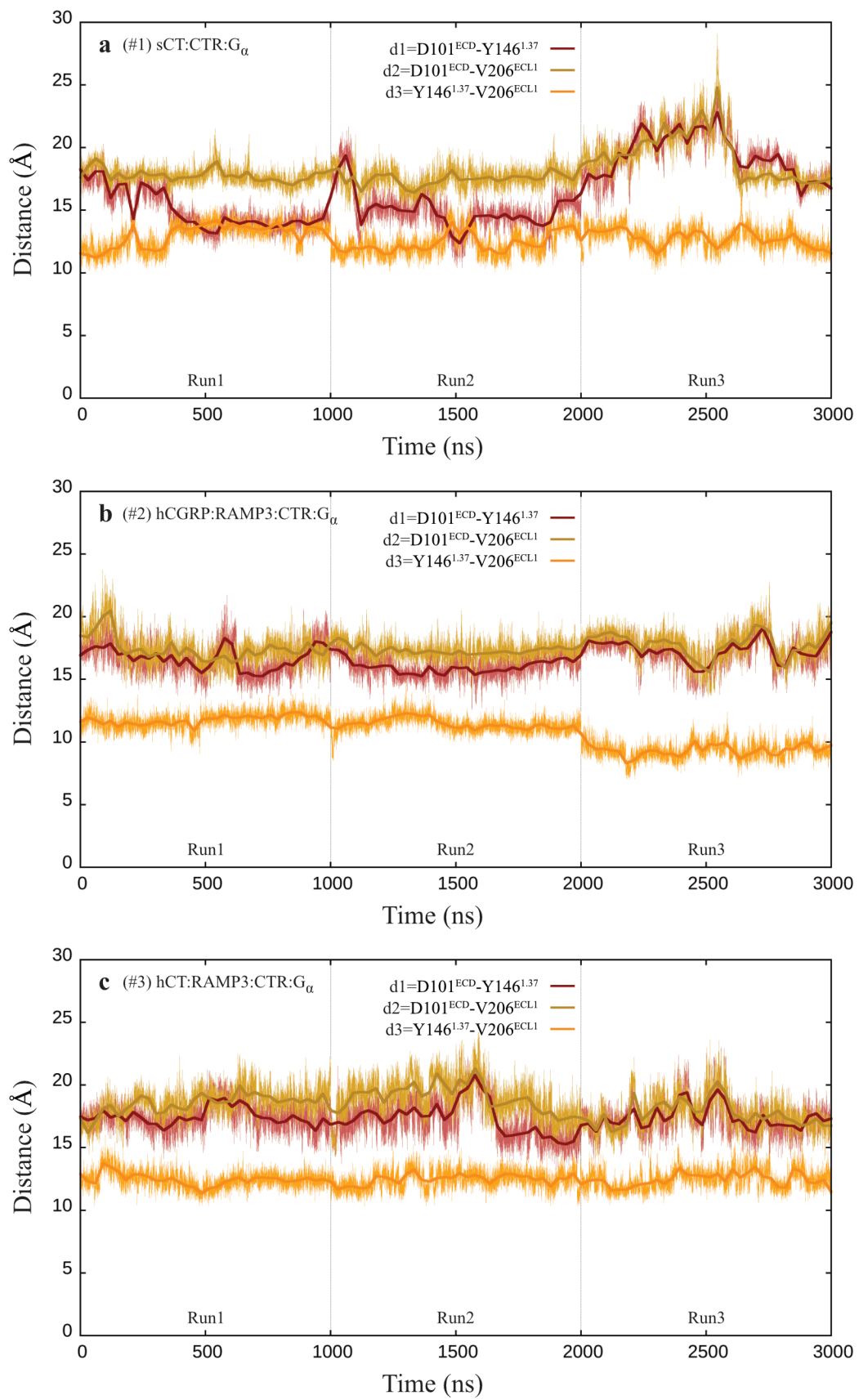
The analysis of the distances (d1=D101<sup>ECD</sup>-Y146<sup>1.37</sup>; d2=D101<sup>ECD</sup>-V206<sup>ECL1</sup>; and d3=Y146<sup>1.37</sup>-V206<sup>ECL1</sup>; Figure 5.10 and Figure 5.13) shows, in general, a similar trend for the distances d1 and d2, while d3 seems to be constant during the MD simulations. This leads to an increase in the distance of the ECD from the TMD. In particular, for the rAmy:RAMP3:CTR:G $\alpha$  (#7) system, d1 and d2 increase significantly during the Run1 with a very similar trend (Figure 5.13g), while the angle  $\emptyset$  in the corresponding simulation does not change significantly (Figure 5.12d). Therefore, a possible explanation for this event implies a rotation of the ECD where the triangular movement is accompanied by a swing in which there is no change in the angle  $\emptyset$  but a change in the distance d2 between D101<sup>ECD</sup> and V206<sup>ECL1</sup>.

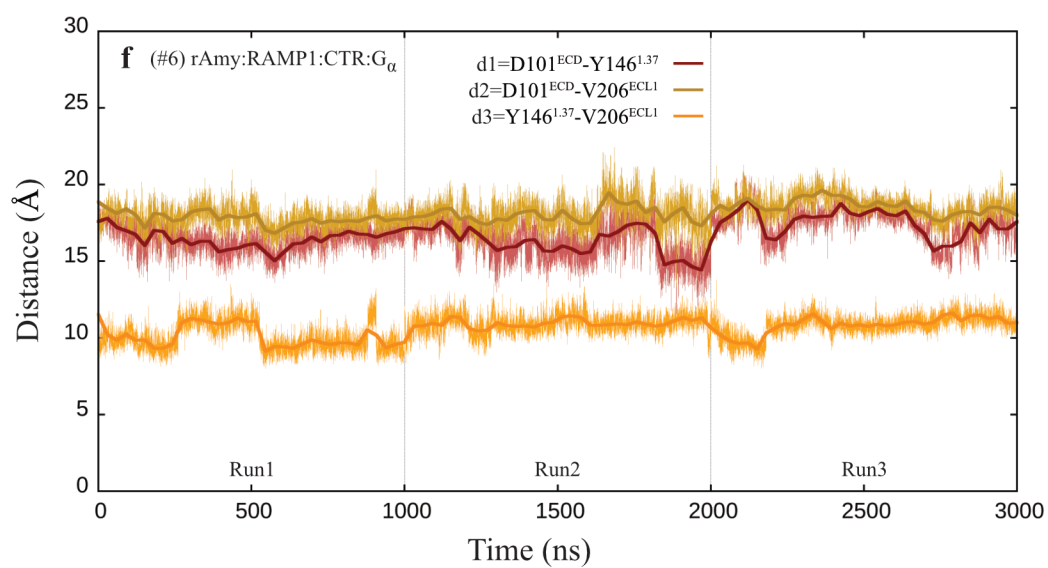
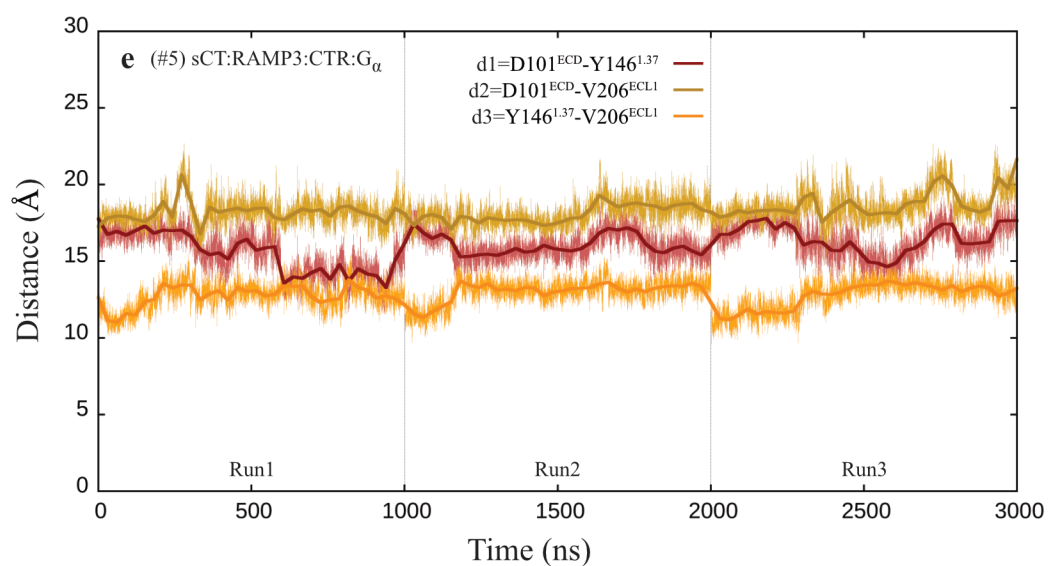
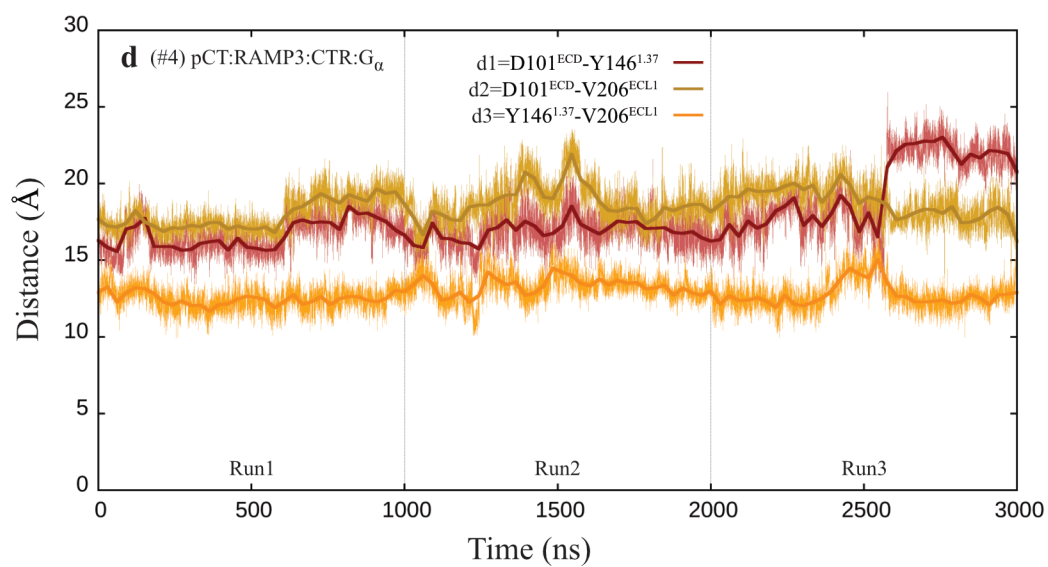
The final analysis concerns the centroids calculation for three selections such as (i) the whole ECD (L40-K120), (ii) the ECD helix (L40-Q61), and (iii) the single residue P100 (Figure 5.11 and Figure 5.14). This analysis allows to have a better idea of the ECD movement due to the presence of different peptides and the possible allosteric effect of RAMPs during the MD simulations.

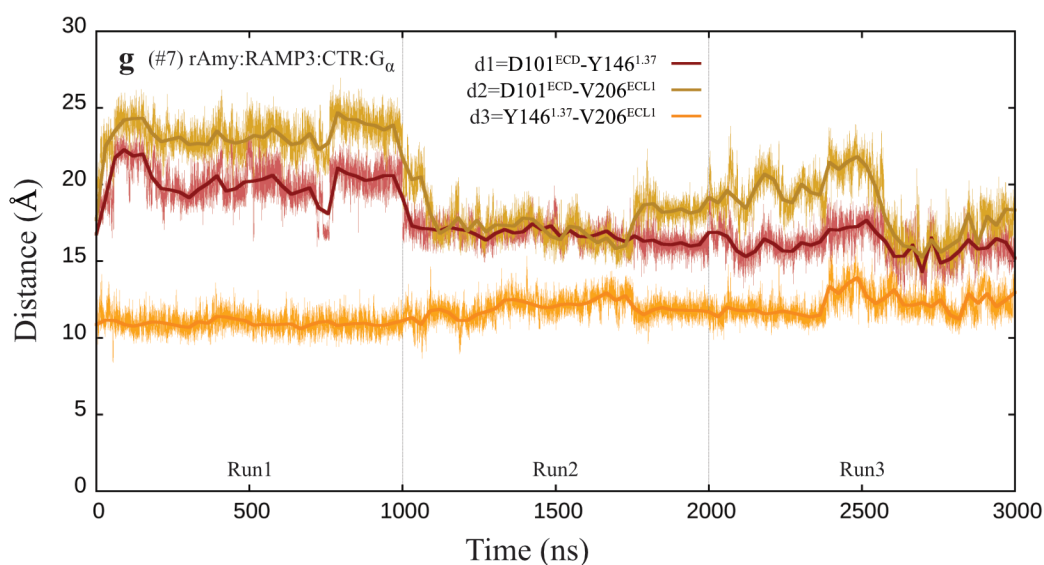




**Figure 5.12 | Analysis of angle  $\emptyset$  (G44<sup>ECD</sup>-FA-Q60<sup>ECD</sup>).** Angle  $\emptyset$  plots (a-d) and distributions (Na-Nd) from three MD simulations (Runs 1-3) of each CTR complex summarised in Table 5.4. (a) Angle  $\emptyset$  comparison of CTR in complex with sCT in the absence and presence of RAMP3, respectively in the sCT:CTR:G $\alpha$  (#1; red) and sCT:RAMP3:CTR:G $\alpha$  (#5; pink) systems. (b) Angle  $\emptyset$  comparison of CTR in complex with RAMP3 and different organism CT peptides, *i.e.* human, porcine, and salmon, respectively in the hCT:RAMP3:CTR:G $\alpha$  (#3; green), pCT:RAMP3:CTR:G $\alpha$  (#4; purple), and sCT:RAMP3:CTR:G $\alpha$  (#5; pink) systems. (c) Angle  $\emptyset$  comparison of CTR in complex with RAMP3 and different peptides: hCGRP:RAMP3:CTR:G $\alpha$  (#2; black), hCT:RAMP3:CTR:G $\alpha$  (#3; green), pCT:RAMP3:CTR:G $\alpha$  (#4; purple), sCT:RAMP3:CTR:G $\alpha$  (#5; pink), and rAmy:RAMP3:CTR:G $\alpha$  (#7; blue) systems. (d) Angle  $\emptyset$  comparison of CTR in complex with rAmy and different RAMPs, *i.e.* RAMP1 and RAMP3, respectively in the rAmy:RAMP1:CTR:G $\alpha$  (#6; cyan) and rAmy:RAMP3:CTR:G $\alpha$  (#7; blue) systems.



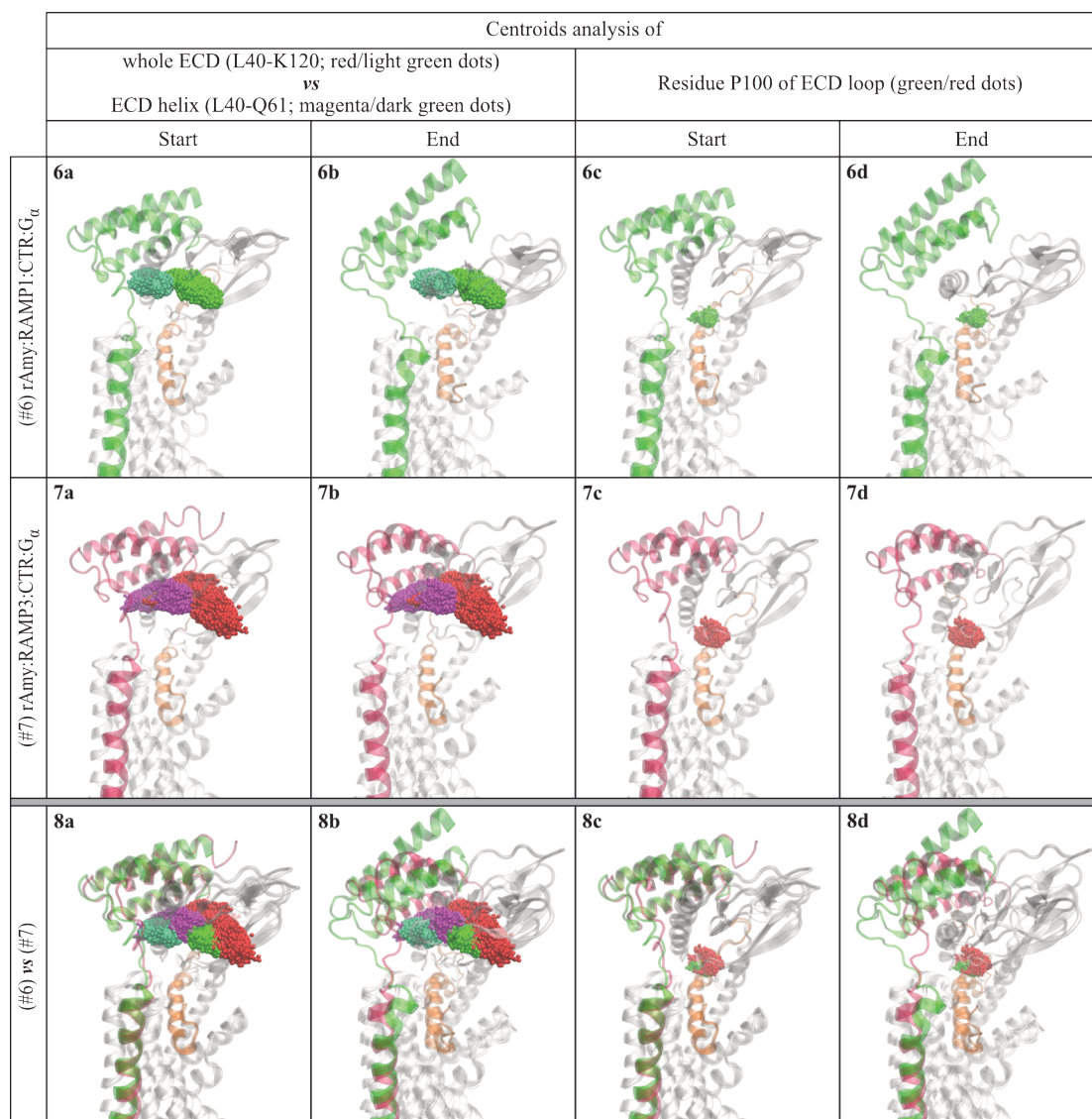




**Figure 5.13 | Analysis of the distances ( $d1=D101^{ECD}-Y146^{1.37}$ ;  $d2=D101^{ECD}-V206^{ECL1}$ ; and  $d3=Y146^{1.37}-V206^{ECL1}$ ).** The three distances d1, d2, and d3, in red, gold, and orange, respectively, were calculated from three MD simulations (Runs 1-3) of each CTR complex: **(a)** sCT:CTR:G $\alpha$  (#1), **(b)** hCGRP:RAMP3:CTR:G $\alpha$  (#2), **(c)** hCT:RAMP3:CTR:G $\alpha$  (#3), **(d)** pCT:RAMP3:CTR:G $\alpha$  (#4), **(e)** sCT:RAMP3:CTR:G $\alpha$  (#5), **(f)** rAmy:RAMP1:CTR:G $\alpha$  (#6), and **(g)** rAmy:RAMP3:CTR:G $\alpha$  (#7) systems.



Centroids analysis of				
whole ECD (L40-K120; red dots) vs ECD helix (L40-Q61; magenta dots)		Residue P100 of ECD loop (red dots)		
Start		End		
	Start	End	Start	End
#1) sCT:CTR:G <sub>u</sub>	1a	1b	1c	1d
#2) hCGRP:RAMP3:CTR:G <sub>u</sub>	2a	2b	2c	2d
#3) hCT:RAMP3:CTR:G <sub>u</sub>	3a	3b	3c	3d
#4) pCT:RAMP3:CTR:G <sub>u</sub>	4a	4b	4c	4d
#5) sCT:RAMP3:CTR:G <sub>u</sub>	5a	5b	5c	5d



**Figure 5.14 | Centroid analysis.** To investigate the receptor ECD movement, three selections for the centroid analysis were used, *i.e.* the whole ECD from L40 to K120 (column **a** and **b**; red/light green dots), the helix of ECD from L40 to Q61 (column **a** and **b**; magenta/dark green dots), and the residue P100 of ECD loop that looks at the receptor cavity where the peptide sits (column **c** and **d**; green/red dots). Column **a** shows the system conformation at the start of each MD simulation while column **b** shows the system conformation at the end of each MD simulation. The lines **1-7** present each CTR complex summarised in Table 5.4, while the last line (**8**) shows a comparison between rAmy:RAMP1:CTR:G<sub>α</sub> (#6), and rAmy:RAMP3:CTR:G<sub>α</sub> (#7) systems. CTR, RAMP1, RAMP3, and peptides are shown in cartoon representation in white, green, red, and orange, respectively.

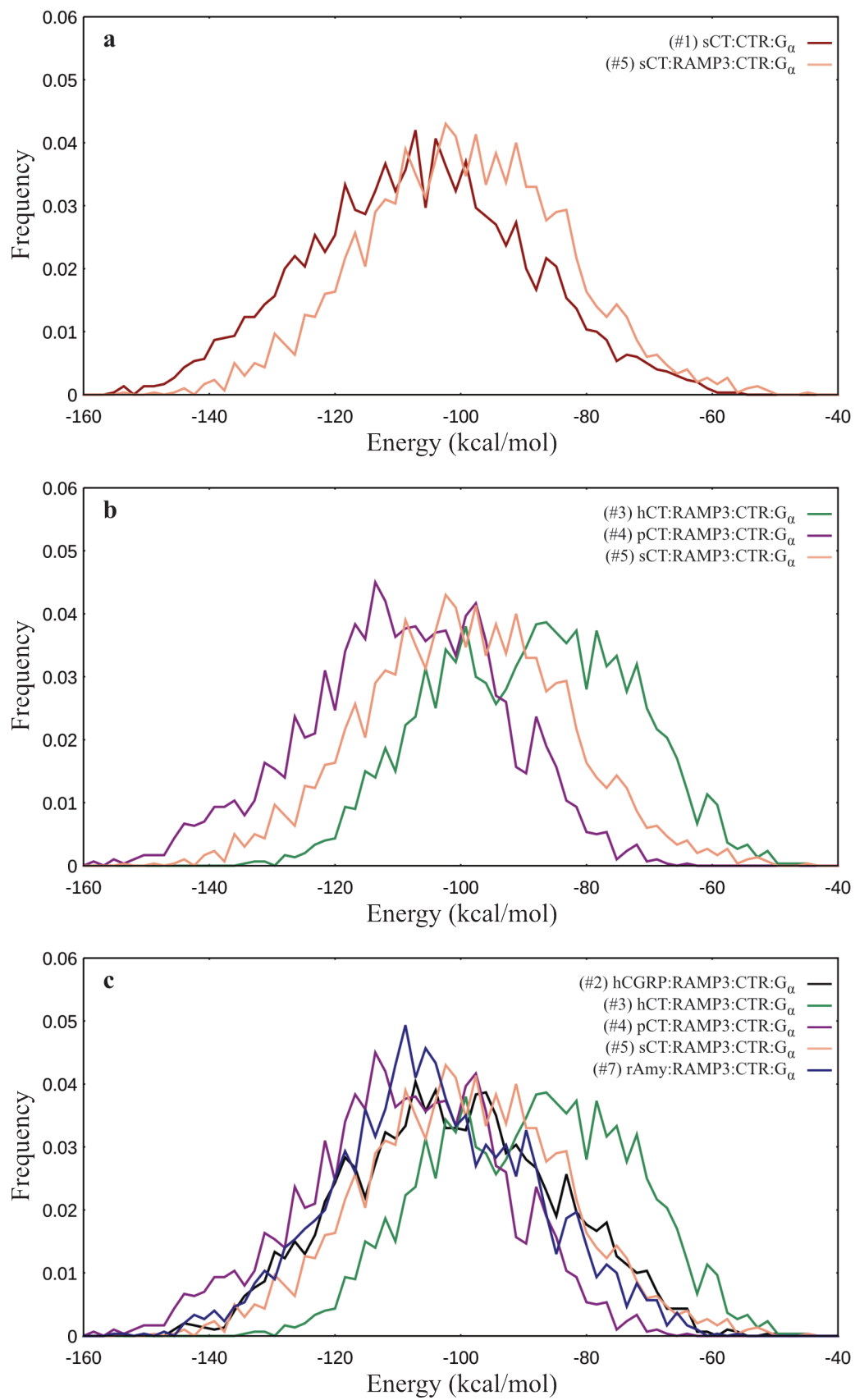


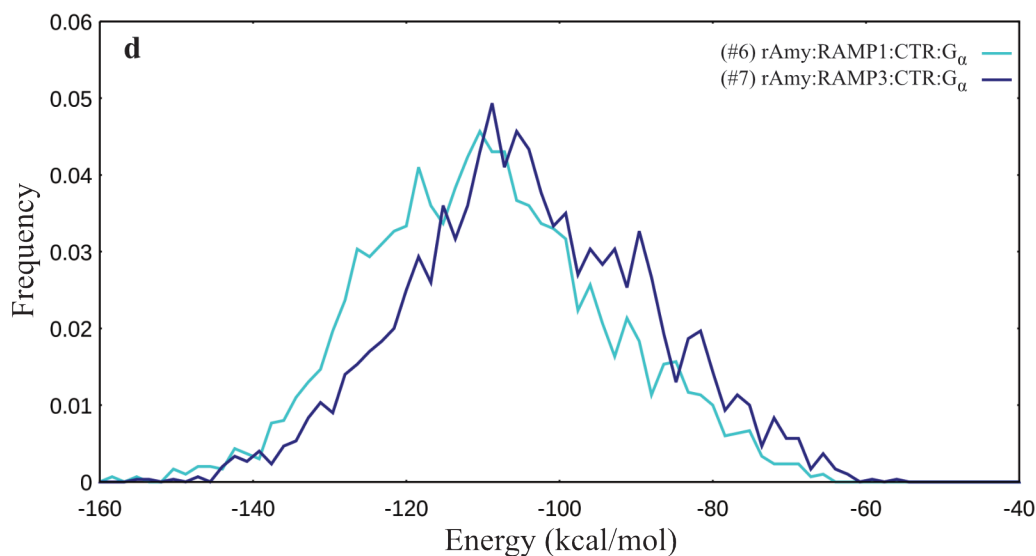
### 5.4.3 Contacts analysis

The van der Waals interactions (Table C.1-C.7) and the hydrogen bonds (Table C.8-C.14) occurred between (i) peptide and CTR or RAMP and (ii) RAMP and CTR, for each system during MD simulations are reported in the Appendix, section C. The threshold considered is respectively 30% for van der Waals interaction and 10% for hydrogen bonds.

### 5.4.4 GBSA analysis

Figure 5.15 shows the GBSA analysis conducted for each system. The comparison between sCT:CTR:Ga (#1; red) and sCT:RAMP3:CTR:Ga (#5; pink) shows better stability for the system without RAMP in the complex (Figure 5.15a). This is in line with experimental data that suggest a higher affinity for CT with CTR alone, while CTR in complex with RAMP loses affinity for CT to generally prefers the binding of no-related peptides such as Amy (Table 5.3). According to the MD simulations of all the systems in complex with CT belonging to the different organism (Figure 5.15b), the stability ranges from the most stable pCT:RAMP3:CTR:Ga (#4; purple) to sCT:RAMP3:CTR:Ga (#5; pink) to hCT:RAMP3:CTR:Ga (#3; green) systems. While hCGRP:RAMP3:CTR:Ga (#2; black) has a similar stability of sCT:RAMP3:CTR:Ga (#5; pink), rAmy:RAMP3:CTR:Ga (#7; blue) seems to be in line with pCT:RAMP3:CTR:Ga (#4; purple) (Figure 5.15c). Finally, the comparison of CTR in complex with rAmy but different RAMP (Figure 5.15d) shows a shift toward lower energy values (of about 10 kcal/mol) of the rAmy:RAMP1:CTR:Ga (#6; cyan) system in the GBSA distribution plot indicates the increase in stability of rAmy:CTR complex in the presence of RAMP1 instead of RAMP3.





**Figure 5.15 | GBSA analysis.** GBSA plots from three MD simulations (Runs 1-3) of each CTR complex summarised in Table 5.4. **(a)** GBSA comparison of CTR in complex with sCT in the absence and presence of RAMP3, respectively in the sCT:CTR:G $\alpha$  (#1; red) and sCT:RAMP3:CTR:G $\alpha$  (#5; pink) systems. **(b)** GBSA comparison of CTR in complex with RAMP3 and different organism CT peptides, *i.e.* human, porcine, and salmon, respectively in the hCT:RAMP3:CTR:G $\alpha$  (#3; green), pCT:RAMP3:CTR:G $\alpha$  (#4; purple), and sCT:RAMP3:CTR:G $\alpha$  (#5; pink) systems. **(c)** GBSA comparison of CTR in complex with RAMP3 and different peptides: hCGRP:RAMP3:CTR:G $\alpha$  (#2; black), hCT:RAMP3:CTR:G $\alpha$  (#3; green), pCT:RAMP3:CTR:G $\alpha$  (#4; purple), sCT:RAMP3:CTR:G $\alpha$  (#5; pink), and rAmy:RAMP3:CTR:G $\alpha$  (#7; blue) systems. **(d)** GBSA comparison of CTR in complex with rAmy and different RAMPs, *i.e.* RAMP1 and RAMP3, respectively in the rAmy:RAMP1:CTR:G $\alpha$  (#6; cyan) and rAmy:RAMP3:CTR:G $\alpha$  (#7; blue) systems.

### 5.4.5 Effect of RAMP3 on the sCT:CTR complex

According to the MD simulations of sCT:CTR:Gα (#1; red) and sCT:RAMP3:CTR:Gα (#5; pink), RAMP3 reduces the flexibility of the receptor ECD as shown in the RMSF (Figure 5.8a) and centroid analysis (Figure 5.14 - panel 1a,b and 5a,b). The observed movement of ECD in absence of the RAMP3 is more a swing movement (see start-end conformations in centroid analysis) than an open/close movement (there is not a huge difference in angle (Figure 5.12a) but it is still present as shown by spike more spread). RAMP3 stabilises and reduces the ECD movement through the following hydrogen bonds between the ECD and itself (Table C.12):

$$\text{Glu112}^{\text{RAMP3}} - \text{Lys54}^{\text{ECD}} = 31\%$$

$$\text{Glu112}^{\text{RAMP3}} - \text{Lys47}^{\text{ECD}} = 11\%$$

$$\text{Glu112}^{\text{RAMP3}} - \text{Lys46}^{\text{ECD}} = 11\%$$

Moreover, the presence of RAMP3 slightly increase the RMSF at TM5 in sCT:RAMP3:CTR:Gα (#5; pink) as shown in Figure 5.8a and this is maybe due to less hydrogen bonds between Thr6<sup>sCT</sup> and His302<sup>5.40</sup> that decreases from 90% (Table C.8) to 51% (Table C.12).

Regarding the peptide, sCT is less stable with RAMP3 as shown in Figure 5.4 where sCT unbinds from ECD (peak III) and slow down the reaching of the plateau (peak I). This higher stability in absence of RAMP3 is also confirmed by the GBSA analysis (Figure 5.15a) that shows a shift toward lower energy values (of about 10 kcal/mol) of the sCT:CTR:Gα (#1; red) system. This is maybe due to more or different hydrogen bonds with ECL2 that is consistent with the allosteric effect of RAMP3 exerted through ECL2 (Gingell *et al.*, 2016; Pham *et al.*, 2019). Indeed, with RAMP3 there are the following hydrogen bonds (Table C.12):

$$\text{Asp113}^{\text{RAMP3}} - \text{His296}^{\text{ECL2}} = 20\%$$

$$\text{Glu112}^{\text{RAMP3}} - \text{His296}^{\text{ECL2}} = 16\%$$

$$\text{Asp113}^{\text{RAMP3}} - \text{Thr295}^{\text{ECL2}} = 15\%$$

and, therefore, the peptide makes (Table C.12):

$$\text{Ser2}^{\text{sCT}} - \text{Glu294}^{\text{ECL2}} = 44\%$$

while without RAMP3 only two hydrogen bonds are formed (Table C.8):

$$\text{Lys11}^{\text{sCT}} - \text{Glu294}^{\text{ECL2}} = 19\%$$

$$\text{Gln14}^{\text{sCT}} - \text{Glu294}^{\text{ECL2}} = 12\%$$

Moreover, sCT makes more hydrogen bonds with the ECD in absence of RAMP3, as shown below.

With RAMP3 (Table C.12):

$$\text{Arg24}^{\text{sCT}} - \text{Asp101}^{\text{ECD}} = 54\%$$

$$\text{Thr25}^{\text{sCT}} - \text{Asp101}^{\text{ECD}} = 25\%$$

$$\text{Lys18}^{\text{sCT}} - \text{Asp97}^{\text{ECD}} = 11\%$$

Without RAMP3 (Table C.8):

$$\text{Arg24}^{\text{sCT}} - \text{Asp101}^{\text{ECD}} = 61\%$$

$$\text{Thr25}^{\text{sCT}} - \text{Asp101}^{\text{ECD}} = 49\%$$

$$\text{Lys18}^{\text{sCT}} - \text{Asp97}^{\text{ECD}} = 33\%$$

$$\text{Thr27}^{\text{sCT}} - \text{Asp101}^{\text{ECD}} = 30\%$$

$$\text{Lys11}^{\text{sCT}} - \text{Asp103}^{\text{ECD}} = 23\%$$

$$\text{Thr31}^{\text{sCT}} - \text{Asp77}^{\text{ECD}} = 17\%$$

$$\text{Lys18}^{\text{sCT}} - \text{Asp101}^{\text{ECD}} = 12\%$$

Taken together, these results show that the presence of RAMP3 influences both receptor and peptide, respectively, reducing the flexibility of the receptor ECD and the stability of the sCT binding. This is in accordance with the experimental data that suggest a higher affinity for CT with CTR alone (*ergo* without RAMP).

#### 5.4.6 Effect of different CT on RAMP3:CTR complexes (AMY<sub>3</sub> receptor)

According to the MD simulations of hCT:RAMP3:CTR:Gα (#3; green), pCT:RAMP3:CTR:Gα (#4; purple), and sCT:RAMP3:CTR:Gα (#5; pink) it is possible to assert that (Figure 5.5):

- sCT is the most stable peptide (sCT RMSD has one main peak (I) and two secondary peaks) and this is in line with sCT pK<sub>i</sub> on AMY<sub>3</sub> of 8.2 (The IUPHAR/BPS Guide to PHARMACOLOGY);
- pCT has one peak corresponding to peak I and one secondary peak (II);
- hCT is the most unstable and this is consistent with experiments on CTR alone (*ergo* without RAMP3) as show in the following Table (Dal Maso *et al.*, 2018).

	hCT pK <sub>i</sub>	sCT pK <sub>i</sub>	pCT pK <sub>i</sub>	sCT (8-32) pK <sub>i</sub>
WT	6.72 ± 0.06	9.87 ± 0.03	8.27 ± 0.07	9.70 ± 0.05

Therefore, I can hypothesise that the relative affinity (pK<sub>i</sub> ranking) on AMY<sub>3</sub> is the following: sCT > pCT > hCT.

From RMSF analysis (Figure 5.8b) it is possible to assert that the receptor RMSF value is generally lower when sCT is bound and this is consistent with sCT being the strongest binder. Moreover, when sCT is bound, the ECD is less mobile as shown in the RMSF (Figure 5.8b), the angle Ø (Figure 5.12b), and the centroid analysis (Figure 5.14 - panel 3a,b, 4a,b and 5a,b).

The GBSA analysis (Figure 5.15b) instead suggests pCT > sCT > hCT and this is in line with the lower quantity of hydrogen bonds between hCT and CTR, and numerous hydrogen bonds between pCT or sCT and the receptor.

Indeed, pCT has three hydrogen bonds with occupancy > 50 % (Table C.11):

$$\text{Trp13}^{\text{pCT}} - \text{His226}^{3.36} = 97\%$$

$$\text{Arg14}^{\text{pCT}} - \text{Glu294}^{\text{ECL2}} = 73\%$$

$$\text{Thr6}^{\text{pCT}} - \text{His302}^{5.40} = 50\%$$

while sCT has two (Table C.12):

$$\text{Arg24}^{\text{sCT}} - \text{Asp101}^{\text{ECD}} = 54\%$$

$$\text{Thr6}^{\text{sCT}} - \text{His302}^{5.40} = 51\%$$

and hCT just one (Table C.10):

$$\text{Thr6}^{\text{hCT}} - \text{His302}^{5.40} = 81\%$$

Moreover, sCT is the only one to have hydrogen bonds with RAMP3 though (Table C.12):

$$\text{Thr25}^{\text{sCT}} - \text{Glu67}^{\text{RAMP3}} = 20\%$$

$$\text{Thr27}^{\text{sCT}} - \text{Glu67}^{\text{RAMP3}} = 17\%$$

In conclusion, while it appears that the affinity of sCT >> hCT, the affinity between sCT and pCT is still not clear. Therefore, future studies are needed to understand better the difference between the peptides analysed in this chapter.

### 5.4.7 Effect of RAMP1 (AMY<sub>1</sub>) or RAMP3 (AMY<sub>3</sub>) on rAMY:CTR

According to the MD simulations of rAmy:RAMP1:CTR:Ga (#6; cyan) and rAmy:RAMP3:CTR:Ga (#7; blue), RAMP3 gives more ECD mobility as shown in the RMSF (Figure 5.8d) and centroid (Figure 5.14 - panel 6a,b and 7a,b) analysis. This is probably due to a swing movement rather than open/close transition (there is not a huge difference in angle, except for Run3 as shown in Figure 5.12d).

There are not any evident differences in rAMY stability (RMSD distributions are similar as shown in Figure 5.7), with RAMP1 slightly more stable in terms of RMSD (Figure 5.7) and GBSA (Figure 5.15d), possibly due to the following hydrogen bond (Table C.13):

$$\text{Arg11}^{\text{rAmy}} - \text{Asp103}^{\text{ECD}} = 100\%$$

which is not observed with RAMP3 (Table C.14). This, in turn, could be due to a different ECD orientation driven by the RAMP1 residue R102, which in RAMP3 is a serine, thus cannot protrude to hydrogen bonds with the ECD (Table C.13):

$$\text{Arg102}^{\text{RAMP1}} - \text{Asp57}^{\text{ECD}} = 84\%$$

$$\text{Arg102}^{\text{RAMP1}} - \text{Gln61}^{\text{ECD}} = 27\%$$

This is in line with the role that the RAMPs linker (region connecting the TM helix and the ECD of the RAMPs) has on the receptor ECD dynamics (Liang *et al.*, 2020); where the different RAMP linkers give different interactions with the ECD that implicate different ECD conformational changes and, therefore, the affinity changes (maybe due to changes in the binding pathway and hence kinetics).

In conclusion, it possible to assert that the presence of RAMP1 in complex with rAMY:CTR reduces the overall receptor ECD flexibility due to the key residue Arg102 situated in the linker region of RAMP1, which forms hydrogen bonds with the receptor



ECD. This allows the receptor to obtain a closer ECD conformation compared to the once with RAMP3 and this may increase the stability of the rAMY binding.

### 5.4.8 Effect of different peptide residues in the key positions

In the calcitonin class of peptide, there are important peptide residue positions, where variation may play a role in the selectivity of the ligands. The positions 11 and 18 are normally occupied by charged residues while the positions 8, 12, and 16 are normally filled by hydrophobic residues. Table 5.5 shows the residues present in these key positions for each peptide used in the MD simulations. Interestingly, hCT loses the charged residue in position 11 showing a polar residue, while pCT loses the charged residue in both positions 11 and 18 showing hydrophobic and polar residues, respectively. Contrariwise, the hydrophobic residue positions are very well conserved in all peptides (Table 5.5).

Differences in contacts at these positions are shown in Table 5.6 and Table 5.7. Although the hydrophobic residues at positions 8, 12, and 16 vary, visual analysis shows that the interaction between the peptide helix and its contact region on TM1 and TM7 did not vary during the simulations. This suggests that differences in peptide activity are not due to these hydrophobic residues.

It is interesting to ask whether the affinity differences are dependent on the variation in residues at position 11 and 18, or whether these residues can vary because they are not important. With regards to the affinity of calcitonin, sCT binds more strongly to CTR in the absence rather than the presence of the RAMP (RAMP3 in this case). Lys11 and Lys18 both make stronger hydrogen bonds (and contacts) to the receptor in the absence of the RAMP.

The affinity ranking of calcitonin is sCT > hCT > pCT. This correlates with the strength of hydrogen bonding to position 11, partly because sCT is the only CT peptide that is charged at this position. For position 18, the strength of the hydrogen bonds for sCT and hCT are similar, but position 18 is Asn in pCT and this probably contributes to the lower affinity of pCT.

In the presence of RAMP3, Amy binds to CTR more strongly than either CGRP or CT (Table 5.3); this again is reflected in the hydrogen bond strengths of Arg11 for Amy, compared to those of Arg11 of CGRP in that the main hydrogen bond to Asp373<sup>7,39</sup> is high at 50%. Position 18 for Amy is His, which is too short to hydrogen bond, so the hydrogen bond strengths at this position do not correlated with the affinity. Indeed, the hydrogen bond strengths of Arg18 interacting with Glu294<sup>ECL2</sup> is high (92%) and does not seem consistent with the trend. It should be noted that where there is an Arg at position 11 or 18, the hydrogen bonds seem to be stronger, but generally the residues at positions 11 and 18 spend a high proportion of their time interacting with the solvent. The Arg residue is longer than the common alternative, which is Lys, and this is probably a factor in Lys and other smaller residues forming fewer hydrogen bonds. Therefore, there does seem to be some link between the affinity and these positions, but there are clearly other factors.

**Table 5.5 | Important charged (positions 11 and 18) and hydrophobic (positions 8, 12, and 16) residues of peptides**

Peptide	Charged residue positions		Hydrophobic residue positions		
	11	18	8	12	16
hCGRP	R	R	V	L	L
hCT	T	K	M	Y	F
pCT	A	N	V	Y	L
sCT	K	K	V	L	L
rAmy	R	H	A	L	L

**Table 5.6 | Contacts (VDW) and hydrogen bonds (Hb) between positions 11 and 18 of the peptide, and CTR during MD simulations**

System	Charged residue positions													
	11							18						
	VDW				Hb			VDW				Hb		
	Peptide residue	CTR residue	C.o.#	Table	Peptide residue	CTR residue	Hb§	Table	Peptide residue	CTR residue	C.o.#	Table	Peptide residue	CTR residue
#1 sCT:CTR:Ga	Lys11	Val293 <sup>ECL2</sup>	34%	C.1	Lys11	Asp103 <sup>ECD</sup> Glu294 <sup>ECL2</sup> Asp373 <sup>7.39</sup>	23% 19% 13%	C.8	Lys18	Pro100 <sup>ECD</sup> Asp97 <sup>ECD</sup>	62% 39%	C.1	Lys18	Asp97 <sup>ECD</sup> Asp101 <sup>ECD</sup>
	Arg11	Asp373 <sup>7.39</sup>	40%	C.2	Arg11	Asp373 <sup>7.39</sup> Asp103 <sup>ECD</sup>	39% 28%	C.9	Arg18	Glu294 <sup>ECL2</sup> Pro100 <sup>ECD</sup> Leu40 <sup>ECD</sup> Ser292 <sup>ECL2</sup>	92% 58% 42% 31%	C.2	Arg18	Glu294 <sup>ECL2</sup>
	Thr11	Val293 <sup>ECL2</sup>	44%	C.3	Thr11	-	-	C.10	Lys18	Pro100 <sup>ECD</sup> Asp97 <sup>ECD</sup>	45% 30%	C.3	Lys18	Asp97 <sup>ECD</sup>
#3 hCT:RAMP3:CTR:Ga														
#4 pCT:RAMP3:CTR:Ga	Ala11	Val293 <sup>ECL2</sup>	49%	C.4	Ala11	-	-	C.11	Asn18	-	-	C.4	Asn18	-
#5 sCT:RAMP3:CTR:Ga	Lys11	-	-	C.5	Lys11	Asp373 <sup>7.39</sup>	18%	C.12	Lys18	Pro100 <sup>ECD</sup> Asp97 <sup>ECD</sup>	58% 43%	C.5	Lys18	Asp97 <sup>ECD</sup>
	Arg11	Asp103 <sup>ECD</sup> Ser105 <sup>ECD</sup>	100% 75%	C.6	Arg11	Asp103 <sup>ECD</sup> Ser105 <sup>ECD</sup>	100% 11%	C.13	His18	Asp97 <sup>ECD</sup> Pro100 <sup>ECD</sup> Phe99 <sup>ECD</sup> Phe102 <sup>ECD</sup>	88% 83% 80% 64%	C.6	His18	-
#6 rAmy:RAMP1:CTR:Ga														
#7 rAmy:RAMP3:CTR:Ga	Arg11	Asp373 <sup>7.39</sup> Val293 <sup>ECL2</sup>	51% 39%	C.7	Arg11	Asp373 <sup>7.39</sup> Glu294 <sup>ECL2</sup>	50% 13%	C.14	His18	Pro104 <sup>ECD</sup>	36%	C.7	His18	-

(#) Contact occupancy (%MD frames)

(§) Hydrogen Bond occupancy (%MD frames)

**Table 5.7 | Contacts (VDW) between positions 8, 12, and 16 of the peptide, and CTR during MD simulations**

Hydrophobic residue positions										
System	8			12			16			Table
	Peptide residue	CTR residue	Contact occupancy (%MD frames)	Peptide residue	CTR residue	Contact occupancy (%MD frames)	Peptide residue	CTR residue	Contact occupancy (%MD frames)	
( #1 ) sCT:CTR:Gα	Val8	His377 <sup>7,43</sup>	77%	Leu12	His377 <sup>7,43</sup>	73%	Leu16	Tyr149 <sup>1,40</sup>	77%	C.1
		Ile380 <sup>7,46</sup>	40%		Leu148 <sup>1,39</sup>	63%		Val206 <sup>ECLI</sup>	62%	
						Leu202 <sup>2,68</sup>		33%		
( #2 ) hCGRP:RAMP3:CTR:Gα	Val8	His377 <sup>7,43</sup>	94%	Leu12	His377 <sup>7,43</sup>	77%	Leu16	Val206 <sup>ECLI</sup>	78%	C.2
		Ile380 <sup>7,46</sup>	45%		Ala145 <sup>1,36</sup>	69%		Tyr149 <sup>1,40</sup>	64%	
					Leu148 <sup>1,39</sup>	56%		Ala145 <sup>1,36</sup>	44%	
( #3 ) hCT:RAMP3:CTR:Gα	Met8	Met376 <sup>7,42</sup>	63%	Tyr12	Tyr149 <sup>1,40</sup>	97%	Phe16	Tyr149 <sup>1,40</sup>	78%	C.3
		Tyr372 <sup>7,38</sup>	61%		Ala145 <sup>1,36</sup>	86%		Val206 <sup>ECLI</sup>	62%	
		His377 <sup>7,43</sup>	46%		His377 <sup>7,43</sup>	83%		Ala145 <sup>1,36</sup>	50%	
		Asp373 <sup>7,39</sup>	44%		Leu148 <sup>1,39</sup>	72%		Tyr146 <sup>1,37</sup>	44%	
		His377 <sup>7,43</sup>	53%		His377 <sup>7,43</sup>	95%		Val206 <sup>ECLI</sup>	64%	
( #4 ) pCT:RAMP3:CTR:Gα	Val8			Tyr12	Tyr149 <sup>1,40</sup>	92%	Leu16	Tyr149 <sup>1,40</sup>	33%	C.4
					Ala145 <sup>1,36</sup>	76%				
					Leu148 <sup>1,39</sup>	71%				
( #5 ) sCT:RAMP3:CTR:Gα	Val8	His377 <sup>7,43</sup>	97%	Leu12	His377 <sup>7,43</sup>	85%	Leu16	Tyr149 <sup>1,40</sup>	81%	C.5
		Ile380 <sup>7,46</sup>	54%		Leu148 <sup>1,39</sup>	63%		Val206 <sup>ECLI</sup>	62%	
		Met376 <sup>7,42</sup>	35%					Leu202 <sup>2,68</sup>	37%	
( #6 ) rAmy:RAMP1:CTR:Gα	Ala8	His377 <sup>7,43</sup>	68%	Leu12	Ala145 <sup>1,36</sup>	65%	Leu16	Val206 <sup>ECLI</sup>	75%	C.6
					Leu148 <sup>1,39</sup>	53%		Ala145 <sup>1,36</sup>	49%	
					His377 <sup>7,43</sup>	40%		Tyr149 <sup>1,40</sup>	45%	
( #7 ) rAmy:RAMP3:CTR:Gα	Ala8			Leu12			Leu16	Tyr146 <sup>1,37</sup>	35%	C.7
		His377 <sup>7,43</sup>	97%		His377 <sup>7,43</sup>	70%		Leu142 <sup>1,33</sup>	31%	
					Leu148 <sup>1,39</sup>	60%		Val206 <sup>ECLI</sup>	65%	
								Tyr149 <sup>1,40</sup>	63%	
								Ala145 <sup>1,36</sup>	46%	

### 5.5 Conclusion

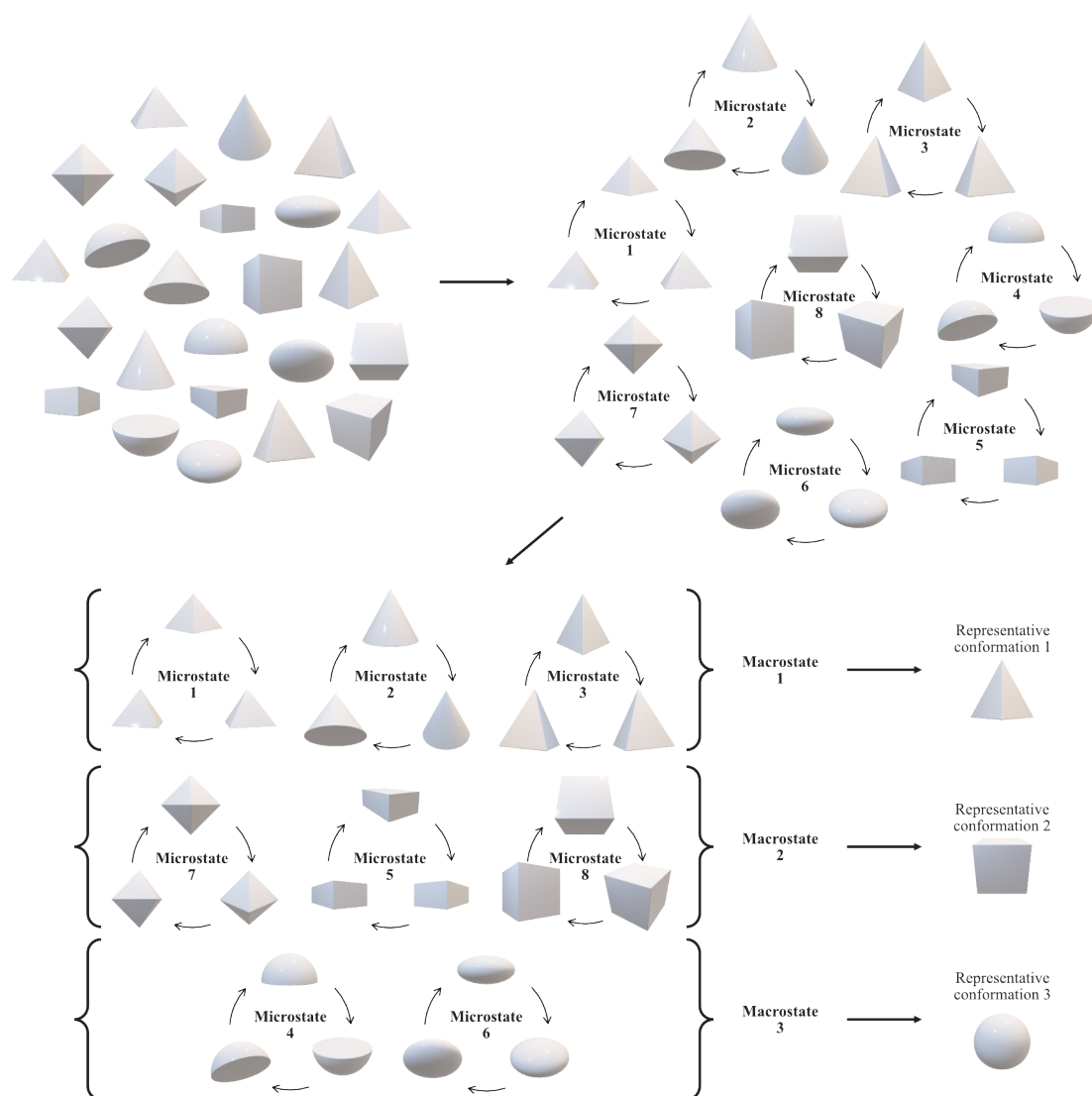
In this study the MD simulations of CTR in complex with different peptides and RAMPs has provided some insights into the dynamic of this class B GPCR receptor. In particular, the analysis of ECD movement has an important relevance since the structures currently available (PDB ID: 6NIY and 5UZ7) do not have information on the ECD structure which influences the very first step of the peptide binding. Moreover, the lack of charged residue in position 11 for hCT and in both positions 11 and 18 for pCT can explain why sCT binds more strongly compared to the other two peptides. Thus, charged residues on the peptides seem to be partially responsible for the differences in affinity. Moreover, because these residues can bias the interaction between the ECD, ECL2, and or the top of TM7, this too may affect affinity.

Future studies could investigate the binding and unbinding mechanism of each CTR peptide, employing SuMD since the peptide probably has to pass through a small opening where positions 11 and 18 may encounter steric interactions with the ECD, and more details may be provided on the peptide affinity and the allosteric effect due to the presence of different RAMPs.

## 6. Concluding remarks

### 6.1 Future research directions

The power of molecular dynamics to deal with complicate aspects of GPCRs (linked to their intrinsic dynamicity as macromolecular targets) has been used in this PhD thesis to study in particular two receptors: FFAR1 and CTR. Despite its power, MD simulation has a limitation in computer timescale (simulation run length restricted to the  $\mu$ s range), thus relevant conformational transitions may not be detectable (Pande *et al.*, 2010). To reach an experimentally relevant timescale, future projects could explore an alternative approach called Markov state modelling (MSM). The key idea behind the MSM is that many short simulations are run simultaneously and then joined together into a single statistical model. This inevitably means that MD simulations explore much more space, enabling the capture of rare events, compared to the run of a longer simulation (Kohlhoff *et al.*, 2014). In general, it is possible to describe MSM in three main steps (Figure 6.1): (i) grouping the data into microstates, which are small data sets based on structural similarity, (ii) grouping microstates together into macrostates (also called metastable states) based on kinetic similarity, (iii) extracting the most probable conformation from each macrostate determining the representative conformations (Bowman *et al.*, 2009). Therefore, only the present state determines the future dynamics while MD takes into consideration the prior states history (Suàrez, Adelman and Zuckerman, 2016).



**Figure 6.1 | Diagram of Markov state modelling (MSM) process.** In the first step, all the different shapes, representing different conformations that the protein may assume during the simulations, are grouped into microstates. Each microstate is based on a rapid interconversion between configurations, which is translated into a small RMSD. In the second step, microstates are grouped into macrostates (or metastable states) based on the kinetic similarity. Finally, the process extracts the most probable conformation from each macrostate determining the representative conformations.

## 6.2 Summary of scientific activities

### 6.2.1 Articles on peer-reviewed journals

Deganutti, G. **Atanasio, S.** Rujan, R.-M. Reynolds, C. A. (2020). Exploring Ligand Binding to Calcitonin Gene-Related Peptide Receptor. (*Under review*).

**Atanasio, S.** Deganutti, G. and Reynolds, C. A. (2020). Addressing free fatty acid receptor 1 (FFAR1) activation using supervised molecular dynamics. *Journal of computer-aided molecular design*. Springer International Publishing, **1**. doi: 10.1007/s10822-020-00338-6.

Pham, V. Zhu, Y. Dal Maso, E. Reynolds, C. A. Deganutti, G. **Atanasio, S.** Hick, C. A. Yang, D. Christopoulos, A. Hay, D. L. Furness, S. G. B. Wang, M. Wootten, D. and Sexton, P. M. (2019). Deconvoluting the Molecular Control of Binding and Signaling at the Amylin 3 Receptor: RAMP3 Alters Signal Propagation through Extracellular Loops of the Calcitonin Receptor. *ACS pharmacology & translational science*. American Chemical Society, **2**(3), pp. 183–197. doi: 10.1021/acsptsci.9b00010.

Dal Maso, E. Glukhova, A. Zhu, Y. Garcia-Nafria, J. Tate, C. G. **Atanasio, S.** Reynolds, C. A. Ramírez-Aportela, E. Carazo, J.-M. Hick, C. A. Furness, S. G. B. Hay, D. L. Liang, Y.-L. Miller, L. J. Christopoulos, A. Wang, M.-W. Wootten, D. and Sexton, P. M. (2019). The Molecular Control of Calcitonin Receptor Signaling. *ACS pharmacology & translational science*, **2**(1), pp. 31–51. doi: 10.1021/acsptsci.8b00056.



### 6.2.2 Awards

#### Oral Presentation Prize

Date: 24 April 2019

Event: *Three Minute Thesis (3MT) 2019* (Colchester, UK)

Title: G protein-coupled receptors (GPCRs): one family of proteins for regulating essential actions in the body

Award: University of Essex Audience's Choice 2019 (£100)

#### Poster Prize

Date: 26 September 2018

Event: *17<sup>th</sup> Annual Graduate Forum* (Colchester, UK)

Title: Molecular Dynamics Simulation applied to Androglobin: a novel heme protein

Award: Best Forum Poster 2018 (£30)

#### Peter Nicholls Research Scholarship

Date: October 2017

Description: This scholarship is generously supported by a bequest from the estate of Professor Peter Nicholls and it has a value of £12,500 per annum (£37,500 total) plus UK tuition fees.

### 6.2.3 Oral [O] and Poster [P] presentations during conferences

[O] **Atanasio, S.** Modelling studies: G Protein-Coupled Receptors (GPCRs). Oral session presented at: *18<sup>th</sup> Annual Graduate Forum*; 2019 Sep 25; Colchester, UK.

[P] Rujan, R.-M. **Atanasio, S.** Deganutti, G. Reynolds, C. A. G protein-coupled receptors as therapeutic targets for diabetes. Poster session presented at: *18<sup>th</sup> Annual Graduate Forum*; 2019 Sep 25; Colchester, UK.

[P] **Atanasio, S.** Rujan, R.-M. Deganutti, G. Reynolds, C. A. Allosteric effects in the Free fatty receptor 1 (FFAR1, formerly GPR40). Poster session presented at: *CCP5 Summer School*; 2019 Jul 10; Durham, UK.

[P] Rujan, R.-M. **Atanasio, S.** Deganutti, G. Reynolds, C. A. Identifying allosteric binding sites at the CGRP receptor. Poster session presented at: *CCP5 Summer School*; 2019 Jul 10; Durham, UK.

[O] **Atanasio, S.** G protein-coupled receptors (GPCRs): one family of proteins for regulating essential actions in the body. Oral session presented at: *Three Minute Thesis (3MT)*; 2019 Apr 24; Colchester, UK.

**Awarded the University of Essex Audience's Choice 2019**

[P] Altinay, C. **Atanasio, S.** Reynolds, C. A. Sequenced-based Approach to Allosteric Sites in G protein-coupled Receptors. Poster session presented at: *Networking event for Entrepreneurs in Residence and Industry Fellows*; 2018 Nov 23; London, UK.

[P] **Atanasio, S.** Deganutti, G. Reeder, B. J. Reynolds, C. A. Molecular Dynamics Simulation applied to Androglobin: a novel heme protein. Poster session presented at: *17<sup>th</sup> Annual Graduate Forum*; 2018 Sep 26; Colchester, UK.

**Awarded the Prize for the Best Forum Poster 2018**



## References

- Altschul, S. F. Gish, W. Miller, W. Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, **215**(3), pp. 403–10. doi: 10.1016/S0022-2836(05)80360-2.
- Aratake, Y. Okuno, T. Matsunobu, T. Saeki, K. Takayanagi, R. Furuya, S. and Yokomizo, T. (2012). Helix 8 of leukotriene B4 receptor 1 inhibits ligand-induced internalization. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, **26**(10), pp. 4068–78. doi: 10.1096/fj.12-212050.
- Ashkenazy, H. Erez, E. Martz, E. Pupko, T. and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic acids research*, **38**(Web Server issue), pp. W529-33. doi: 10.1093/nar/gkq399.
- Atanasio, S. Deganutti, G. and Reynolds, C. A. (2020). Addressing free fatty acid receptor 1 (FFAR1) activation using supervised molecular dynamics. *Journal of computer-aided molecular design*. Springer International Publishing, **1**. doi: 10.1007/s10822-020-00338-6.
- Attwood, T. K. and Findlay, J. B. (1994). Fingerprinting G-protein-coupled receptors. *Protein engineering*, **7**(2), pp. 195–203. doi: 10.1093/protein/7.2.195.
- Bai, X. Fernandez, I. S. McMullan, G. and Scheres, S. H. W. (2013). Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife*, **2**(2), pp. 2–13. doi: 10.7554/eLife.00461.
- Bakir, B. and Sezerman, O. U. (2006). Functional Classification of G-Protein Coupled Receptors, Based on Their Specific Ligand Coupling Patterns. In: Rothlauf, F. et al. (eds) *Applications of Evolutionary Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Computer Science), pp. 1–12. doi: 10.1007/11732242\_1.
- Baldwin, J. M. (1993). The probable arrangement of the helices in G protein-coupled receptors. *The EMBO journal*, **12**(4), pp. 1693–703. doi: 10.1002/j.1460-2075.1993.tb05814.x.

## REFERENCES

---

- Baldwin, J. M. Schertler, G. F. X. and Unger, V. M. (1997). An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *Journal of molecular biology*, **272**(1), pp. 144–64. doi: 10.1006/jmbi.1997.1240.
- Ballesteros, J. A. and Weinstein, H. (1995). [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In: *Methods in Neurosciences*, pp. 366–428. doi: 10.1016/S1043-9471(05)80049-7.
- Barducci, A. Bonomi, M. and Parrinello, M. (2011). Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **1**(5), pp. 826–843. doi: 10.1002/wcms.31.
- Bayly, C. I. Cieplak, P. Cornell, W. D. and Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *Journal of Physical Chemistry*, **97**(40), pp. 10269–10280. doi: 10.1021/j100142a004.
- Bennett-Lovsey, R. M. Herbert, A. D. Sternberg, M. J. E. and Kelley, L. A. (2008). Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins*, **70**(3), pp. 611–25. doi: 10.1002/prot.21688.
- Berendsen, H. J. C. Postma, J. P. M. van Gunsteren, W. F. DiNola, A. and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, **81**(8), pp. 3684–3690. doi: 10.1063/1.448118.
- Berman, H. M. Westbrook, J. Feng, Z. Gilliland, G. Bhat, T. N. Weissig, H. Shindyalov, I. N. and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic acids research*, **28**(1), pp. 235–42. URL <http://www.ncbi.nlm.nih.gov/pubmed/10592235>.
- Bissaro, M. Bolcato, G. Deganutti, G. Sturlese, M. and Moro, S. (2019). Revisiting the Allosteric Regulation of Sodium Cation on the Binding of Adenosine at the Human A2A Adenosine Receptor: Insights from Supervised Molecular Dynamics (SuMD) Simulations. *Molecules (Basel, Switzerland)*, **24**(15). doi: 10.3390/molecules24152752.

- Bowman, G. R. Huang, X. and Pande, V. S. (2009). Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods (San Diego, Calif.)*, **49**(2), pp. 197–201. doi: 10.1016/j.ymeth.2009.04.013.
- Burant, C. F. Viswanathan, P. Marcink, J. Cao, C. Vakilynejad, M. Xie, B. and Leifke, E. (2012). TAK-875 versus placebo or glimepiride in type 2 diabetes mellitus: a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet (London, England)*, **379**(9824), pp. 1403–11. doi: 10.1016/S0140-6736(11)61879-5.
- Burmester, T. and Hankeln, T. (2014). Function and evolution of vertebrate globins. *Acta Physiologica*, **211**(3), pp. 501–514. doi: 10.1111/apha.12312.
- Cheng, Y. (2018). Single-particle cryo-EM-How did it get here and where will it go. *Science (New York, N.Y.)*, **361**(6405), pp. 876–880. doi: 10.1126/science.aat4346.
- Christopoulos, A. Changeux, J.-P. Catterall, W. A. Fabbro, D. Burris, T. P. Cidlowski, J. A. Olsen, R. W. Peters, J. A. Neubig, R. R. Pin, J.-P. Sexton, P. M. Kenakin, T. P. Ehlert, F. J. Spedding, M. and Langmead, C. J. (2014). International Union of Basic and Clinical Pharmacology. XC. multisite pharmacology: recommendations for the nomenclature of receptor allosterism and allosteric ligands. *Pharmacological reviews*, **66**(4), pp. 918–47. doi: 10.1124/pr.114.008862.
- Clamp, M. Cuff, J. Searle, S. M. and Barton, G. J. (2004). The Jalview Java alignment editor. *Bioinformatics (Oxford, England)*, **20**(3), pp. 426–7. doi: 10.1093/bioinformatics/btg430.
- Cole, C. Barber, J. D. and Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic acids research*, **36**(Web Server issue), pp. W197–201. doi: 10.1093/nar/gkn238.
- Crooks, G. E. Hon, G. Chandonia, J.-M. and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome research*, **14**(6), pp. 1188–90. doi: 10.1101/gr.849004.
- Cuzzolin, A. Sturlese, M. Deganutti, G. Salmaso, V. Sabbadin, D. Ciancetta, A. and

- Moro, S. (2016). Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations. *Journal of chemical information and modeling*, **56**(4), pp. 687–705. doi: 10.1021/acs.jcim.5b00702.
- Dal Maso, E. Zhu, Y. Pham, V. Reynolds, C. A. Deganutti, G. Hick, C. A. Yang, D. Christopoulos, A. Hay, D. L. Wang, M.-W. Sexton, P. M. Furness, S. G. B. and Wootten, D. (2018). Extracellular loops 2 and 3 of the calcitonin receptor selectively modify agonist binding and efficacy. *Biochemical pharmacology*. Elsevier, **150**(February), pp. 214–244. doi: 10.1016/j.bcp.2018.02.005.
- Dawson, N. L. Lewis, T. E. Das, S. Lees, J. G. Lee, D. Ashford, P. Orengo, C. A. and Sillitoe, I. (2017). CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic acids research*, **45**(D1), pp. D289–D295. doi: 10.1093/nar/gkw1098.
- Deganutti, G. Moro, S. and Reynolds, C. A. (2020). A Supervised Molecular Dynamics Approach to Unbiased Ligand–Protein Unbinding. *Journal of Chemical Information and Modeling*, **60**(3), pp. 1804–1817. doi: 10.1021/acs.jcim.9b01094.
- Deganutti, G. Salmaso, V. and Moro, S. (2018). Could Adenosine Recognize its Receptors with a Stoichiometry Other than 1 : 1?. *Molecular informatics*, **37**(8), p. e1800009. doi: 10.1002/minf.201800009.
- Delos Santos, N. M. Gardner, L. A. White, S. W. and Bahouth, S. W. (2006). Characterization of the residues in helix 8 of the human beta1-adrenergic receptor that are involved in coupling the receptor to G proteins. *The Journal of biological chemistry*, **281**(18), pp. 12896–907. doi: 10.1074/jbc.M508500200.
- Doerr, S. Harvey, M. J. Noé, F. and De Fabritiis, G. (2016). HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *Journal of chemical theory and computation*, **12**(4), pp. 1845–52. doi: 10.1021/acs.jctc.6b00049.
- Dolinsky, T. J. Nielsen, J. E. McCammon, J. A. and Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic acids research*, **32**(Web Server issue), pp. W665–7. doi:

10.1093/nar/gkh381.

- Draper-Joyce, C. J. Khoshouei, M. Thal, D. M. Liang, Y.-L. Nguyen, A. T. N. Furness, S. G. B. Venugopal, H. Baltos, J.-A. Plitzko, J. M. Danev, R. Baumeister, W. May, L. T. Wootten, D. Sexton, P. M. Glukhova, A. and Christopoulos, A. (2018). Structure of the adenosine-bound human adenosine A1 receptor-Gi complex. *Nature*, **558**(7711), pp. 559–563. doi: 10.1038/s41586-018-0236-6.
- Drozdetskiy, A. Cole, C. Procter, J. and Barton, G. J. (2015). JPred4: a protein secondary structure prediction server. *Nucleic acids research*, **43**(W1), pp. W389-94. doi: 10.1093/nar/gkv332.
- Durrant, J. D. and McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC biology*, **9**(71), p. 71. doi: 10.1186/1741-7007-9-71.
- Edfalk, S. Steneberg, P. and Edlund, H. (2008). Gpr40 is expressed in enteroendocrine cells and mediates free fatty acid stimulation of incretin secretion. *Diabetes*, **57**(9), pp. 2280–7. doi: 10.2337/db08-0307.
- Ernst, O. P. Meyer, C. K. Marin, E. P. Henklein, P. Fu, W. Y. Sakmar, T. P. and Hofmann, K. P. (2000). Mutation of the fourth cytoplasmic loop of rhodopsin affects binding of transducin and peptides derived from the carboxyl-terminal sequences of transducin alpha and gamma subunits. *The Journal of biological chemistry*, **275**(3), pp. 1937–43. URL <http://www.ncbi.nlm.nih.gov/pubmed/10636895>.
- Essmann, U. Perera, L. Berkowitz, M. L. Darden, T. Lee, H. and Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, **103**(19), pp. 8577–8593. doi: 10.1063/1.470117.
- Filizola, M. (2014). *G Protein-Coupled Receptors - Modeling and Simulation*. Edited by M. Filizola. Dordrecht: Springer Netherlands (Advances in Experimental Medicine and Biology). doi: 10.1007/978-94-007-7423-0.
- Fredriksson, R. Lagerström, M. C. Lundin, L.-G. and Schiöth, H. B. (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Molecular pharmacology*, **63**(6), pp. 1256–72. doi: 10.1124/mol.63.6.1256.



## REFERENCES

---

- Garland, S. L. (2013). Are GPCRs Still a Source of New Targets?. *Journal of Biomolecular Screening*, **18**(9), pp. 947–966. doi: 10.1177/1087057113498418.
- Genheden, S. and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery*, **10**(5), pp. 449–61. doi: 10.1517/17460441.2015.1032936.
- Ghosh, J. Marru, S. Singh, N. Vanomesslaeghe, K. Fan, Y. and Pamidighantam, S. (2011). Molecular parameter optimization gateway (ParamChem). In: *Proceedings of the 2011 TeraGrid Conference on Extreme Digital Discovery - TG '11*. New York, New York, USA: ACM Press, p. 1. doi: 10.1145/2016741.2016779.
- Gingell, J. J. Simms, J. Barwell, J. Poyner, D. R. Watkins, H. A. Pioszak, A. A. Sexton, P. M. and Hay, D. L. (2016). An allosteric role for receptor activity-modifying proteins in defining GPCR pharmacology. *Cell discovery*. Nature Publishing Group, **2**, p. 16012. doi: 10.1038/celldisc.2016.12.
- Gurevich, E. V and Gurevich, V. V (2006). Arrestins: ubiquitous regulators of cellular signaling pathways. *Genome biology*, **7**(9), p. 236. doi: 10.1186/gb-2006-7-9-236.
- Gurevich, V. V. and Gurevich, E. V. (2014). Overview of different mechanisms of arrestin-mediated signaling. *Current protocols in pharmacology*, **67**, p. Unit 2.10.1-9. doi: 10.1002/0471141755.ph0210s67.
- Harvey, M. J. Giupponi, G. and Fabritiis, G. De (2009). ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *Journal of chemical theory and computation*, **5**(6), pp. 1632–9. doi: 10.1021/ct9000685.
- Hauser, A. S. Attwood, M. M. Rask-Andersen, M. Schiöth, H. B. and Gloriam, D. E. (2017). Trends in GPCR drug discovery: new agents, targets and indications. *Nature Reviews Drug Discovery*. Nature Publishing Group, **16**(12), pp. 829–842. doi: 10.1038/nrd.2017.178.
- Ho, J. D. *et al.* (2018). Structural basis for GPR40 allosteric agonism and incretin stimulation. *Nature Communications*, **9**(1), p. 1645. doi: 10.1038/s41467-017-

01240-w.

- Hoare, S. R. J. (2005). Mechanisms of peptide and nonpeptide ligand binding to Class B G-protein-coupled receptors. *Drug Discovery Today*, **10**(6), pp. 417–427. doi: 10.1016/S1359-6446(05)03370-2.
- Hoogewijs, D. Ebner, B. Germani, F. Hoffmann, F. G. Fabrizius, A. Moens, L. Burmester, T. Dewilde, S. Storz, J. F. Vinogradov, S. N. and Hankeln, T. (2012). Androglobin: A Chimeric Globin in Metazoans That Is Preferentially Expressed in Mammalian Testes. *Molecular Biology and Evolution*, **29**(4), pp. 1105–1114. doi: 10.1093/molbev/msr246.
- Huang, B. Lu, Y.-S. Li, X. Zhu, Z.-C. Li, K. Liu, J.-W. Zheng, J. and Hu, Z.-L. (2014). Androglobin knockdown inhibits growth of glioma cell lines. *International journal of clinical and experimental pathology*, **7**(5), pp. 2179–84. URL <http://www.ncbi.nlm.nih.gov/pubmed/24966926>.
- Huang, J. and MacKerell, A. D. (2013). CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *Journal of computational chemistry*, **34**(25), pp. 2135–45. doi: 10.1002/jcc.23354.
- Humphrey, W. Dalke, A. and Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, **14**(1), pp. 33–8, 27–8. doi: 10.1016/0263-7855(96)00018-5.
- Isberg, V. de Graaf, C. Bortolato, A. Cherezov, V. Katritch, V. Marshall, F. H. Mordalski, S. Pin, J.-P. Stevens, R. C. Vriend, G. and Gloriam, D. E. (2015). Generic GPCR residue numbers – aligning topology maps while minding the gaps. *Trends in Pharmacological Sciences*. Elsevier Ltd, **36**(1), pp. 22–31. doi: 10.1016/j.tips.2014.11.001.
- Isberg, V. Mordalski, S. Munk, C. Rataj, K. Harpsøe, K. Hauser, A. S. Vroiling, B. Bojarski, A. J. Vriend, G. and Gloriam, D. E. (2016). GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Research*, **44**(D1), pp. D356–D364. doi: 10.1093/nar/gkv1178.
- Itoh, Y. *et al.* (2003). Free fatty acids regulate insulin secretion from pancreatic beta cells through GPR40. *Nature*, **422**(6928), pp. 173–6. doi: 10.1038/nature01478.

## REFERENCES

---

- Jakowiecki, J. Orzeł, U. Chawananon, S. Miszt, P. and Filipek, S. (2020). The Hydrophobic Ligands Entry and Exit from the GPCR Binding Site-SMD and SuMD Simulations. *Molecules (Basel, Switzerland)*, **25**(8). doi: 10.3390/molecules25081930.
- Jakowiecki, J. and Filipek, S. (2016). Hydrophobic Ligand Entry and Exit Pathways of the CB1 Cannabinoid Receptor. *Journal of chemical information and modeling*, **56**(12), pp. 2457–2466. doi: 10.1021/acs.jcim.6b00499.
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology*, **292**(2), pp. 195–202. doi: 10.1006/jmbi.1999.3091.
- Jorgensen, W. L. Chandrasekhar, J. Madura, J. D. Impey, R. W. and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, **79**(2), pp. 926–935. doi: 10.1063/1.445869.
- Källberg, M. Wang, H. Wang, S. Peng, J. Wang, Z. Lu, H. and Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature protocols*, **7**(8), pp. 1511–22. doi: 10.1038/nprot.2012.085.
- Karplus, M. and Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature*, **347**(6294), pp. 631–9. doi: 10.1038/347631a0.
- Katritch, V. Fenalti, G. Abola, E. E. Roth, B. L. Cherezov, V. and Stevens, R. C. (2014). Allosteric sodium in class A GPCR signaling. *Trends in biochemical sciences*. Elsevier Ltd, **39**(5), pp. 233–44. doi: 10.1016/j.tibs.2014.03.002.
- Klauda, J. B. Venable, R. M. Freites, J. A. O'Connor, J. W. Tobias, D. J. Mondragon-Ramirez, C. Vorobyov, I. MacKerell, A. D. and Pastor, R. W. (2010). Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The journal of physical chemistry. B*, **114**(23), pp. 7830–43. doi: 10.1021/jp101759q.
- Kobilka, B. K. and Deupi, X. (2007). Conformational complexity of G-protein-coupled receptors. *Trends in pharmacological sciences*, **28**(8), pp. 397–406. doi: 10.1016/j.tips.2007.06.003.
- Kohlhoff, K. J. Shukla, D. Lawrenz, M. Bowman, G. R. Konerding, D. E. Belov, D.

- Altman, R. B. and Pande, V. S. (2014). Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nature chemistry*, **6**(1), pp. 15–21. doi: 10.1038/nchem.1821.
- Kolakowski, L. F. (1994). GCRDb: a G-protein-coupled receptor database. *Receptors & channels*, **2**(1), pp. 1–7. doi: 10.4236/ojgen.2012.24B003.
- Kotarsky, K. Nilsson, N. E. Flodgren, E. Owman, C. and Olde, B. (2003). A human cell surface receptor activated by free fatty acids and thiazolidinedione drugs. *Biochemical and biophysical research communications*, **301**(2), pp. 406–10. doi: 10.1016/s0006-291x(02)03064-4.
- Kräutler, V. Van Gunsteren, W. F. and Hünenberger, P. H. (2001). A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry*, **22**(5), pp. 501–508. doi: 10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V.
- Kryshtafovych, A. Schwede, T. Topf, M. Fidelis, K. and Moult, J. (2019). Critical assessment of methods of protein structure prediction (CASP)-Round XIII. *Proteins*, **86**(August), pp. 7–15. doi: 10.1002/prot.25823.
- Laio, A. and Parrinello, M. (2002). Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(20), pp. 12562–6. doi: 10.1073/pnas.202427399.
- Larkin, M. A. Blackshields, G. Brown, N. P. Chenna, R. McGettigan, P. A. McWilliam, H. Valentin, F. Wallace, I. M. Wilm, A. Lopez, R. Thompson, J. D. Gibson, T. J. and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, **23**(21), pp. 2947–8. doi: 10.1093/bioinformatics/btm404.
- Latorraca, N. R. Wang, J. K. Bauer, B. Townshend, R. J. L. Hollingsworth, S. A. Olivieri, J. E. Xu, H. E. Sommer, M. E. and Dror, R. O. (2018). Molecular mechanism of GPCR-mediated arrestin activation. *Nature*, **557**(7705), pp. 452–456. doi: 10.1038/s41586-018-0077-3.
- Latorraca, N. R. Venkatakrishnan, A. J. and Dror, R. O. (2017). GPCR Dynamics:

## REFERENCES

---

- Structures in Motion. *Chemical reviews*, **117**(1), pp. 139–155. doi: 10.1021/acs.chemrev.6b00177.
- Leach, A. (2001). *Molecular Modelling: Principles and Applications*. 2nd edition. Edited by Longman. Pearson Education.
- Li, X. Mooney, P. Zheng, S. Booth, C. R. Braunfeld, M. B. Gubbens, S. Agard, D. A. and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*, **10**(6), pp. 584–590. doi: 10.1038/nmeth.2472.
- Li, Z. Zhou, X. Dai, Z. and Zou, X. (2010). Classification of G-protein coupled receptors based on support vector machine with maximum relevance minimum redundancy and genetic algorithm. *BMC Bioinformatics*, **11**(1), p. 325. doi: 10.1186/1471-2105-11-325.
- Li, Z. Xu, X. Huang, W. and Qian, H. (2018). Free Fatty Acid Receptor 1 (FFAR1) as an Emerging Therapeutic Target for Type 2 Diabetes Mellitus: Recent Progress and Prevailing Challenges. *Medicinal research reviews*, **38**(2), pp. 381–425. doi: 10.1002/med.21441.
- Liang, Y.-L. Belousoff, M. J. Fletcher, M. M. Zhang, X. Khoshouei, M. Deganutti, G. Koole, C. Furness, S. G. B. Miller, L. J. Hay, D. L. Christopoulos, A. Reynolds, C. A. Danev, R. Wootten, D. and Sexton, P. M. (2020). Structure and Dynamics of Adrenomedullin Receptors AM<sub>1</sub> and AM<sub>2</sub> Reveal Key Mechanisms in the Control of Receptor Phenotype by Receptor Activity-Modifying Proteins. *ACS pharmacology & translational science*, **3**(2), pp. 263–284. doi: 10.1021/acsptsci.9b00080.
- Lin, D. C. H. Guo, Q. Luo, J. Zhang, J. Nguyen, K. Chen, M. Tran, T. Dransfield, P. J. Brown, S. P. Houze, J. Vimolratana, M. Jiao, X. Y. Wang, Y. Birdsall, N. J. M. and Swaminath, G. (2012). Identification and pharmacological characterization of multiple allosteric binding sites on the free fatty acid 1 receptor. *Molecular pharmacology*, **82**(5), pp. 843–59. doi: 10.1124/mol.112.079640.
- Liu, X. Ahn, S. Kahsai, A. W. Meng, K.-C. Latorraca, N. R. Pani, B. Venkatakrishnan, A. J. Masoudi, A. Weis, W. I. Dror, R. O. Chen, X. Lefkowitz, R. J. and Kobilka, B. K. (2017). Mechanism of intracellular allosteric  $\beta$ 2AR antagonist

- revealed by X-ray crystal structure. *Nature*, **548**(7668), pp. 480–484. doi: 10.1038/nature23652.
- Lomize, M. A. Lomize, A. L. Pogozheva, I. D. and Mosberg, H. I. (2006). OPM: orientations of proteins in membranes database. *Bioinformatics (Oxford, England)*, **22**(5), pp. 623–5. doi: 10.1093/bioinformatics/btk023.
- Loncharich, R. J. Brooks, B. R. and Pastor, R. W. (1992). Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers*, **32**(5), pp. 523–35. doi: 10.1002/bip.360320508.
- Lu, J. *et al.* (2017). Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. *Nature structural & molecular biology*. Nature Publishing Group, **24**(7), pp. 570–577. doi: 10.1038/nsmb.3417.
- Lückmann, M. Trauelsen, M. Bentsen, M. A. Nissen, T. A. D. Martins, J. Fallah, Z. Nygaard, M. M. Papaleo, E. Lindorff-Larsen, K. Schwartz, T. W. and Frimurer, T. M. (2019). Molecular dynamics-guided discovery of an ago-allosteric modulator for GPR40/FFAR1. *Proceedings of the National Academy of Sciences of the United States of America*, **116**(14), pp. 7123–7128. doi: 10.1073/pnas.1811066116.
- Mancini, A. D. Bertrand, G. Vivot, K. Carpentier, É. Tremblay, C. Ghislain, J. Bouvier, M. and Poitout, V. (2015).  $\beta$ -Arrestin Recruitment and Biased Agonism at Free Fatty Acid Receptor 1. *The Journal of biological chemistry*, **290**(34), pp. 21131–40. doi: 10.1074/jbc.M115.644450.
- May, L. T. Leach, K. Sexton, P. M. and Christopoulos, A. (2007). Allosteric modulation of G protein-coupled receptors. *Annual review of pharmacology and toxicology*, **47**(1), pp. 1–51. doi: 10.1146/annurev.pharmtox.47.120505.105159.
- Mayne, C. G. Saam, J. Schulten, K. Tajkhorshid, E. and Gumbart, J. C. (2013). Rapid parameterization of small molecules using the Force Field Toolkit. *Journal of computational chemistry*, **34**(32), pp. 2757–70. doi: 10.1002/jcc.23422.
- Mayrose, I. Graur, D. Ben-Tal, N. and Pupko, T. (2004). Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are

## REFERENCES

---

- superior. *Molecular biology and evolution*, **21**(9), pp. 1781–91. doi: 10.1093/molbev/msh194.
- Miller, B. R. McGee, T. D. Swails, J. M. Homeyer, N. Gohlke, H. and Roitberg, A. E. (2012). MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of chemical theory and computation*, **8**(9), pp. 3314–21. doi: 10.1021/ct300418h.
- Nobel Media (2017). *The 2017 Nobel Prize in Chemistry*. URL <https://www.nobelprize.org/>.
- Oldham, W. M. and Hamm, H. E. (2008). Heterotrimeric G protein activation by G-protein-coupled receptors. *Nature reviews. Molecular cell biology*, **9**(1), pp. 60–71. doi: 10.1038/nrm2299.
- Olsson, M. H. M. Søndergaard, C. R. Rostkowski, M. and Jensen, J. H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of chemical theory and computation*, **7**(2), pp. 525–37. doi: 10.1021/ct100578z.
- Ostrovskaya, A. Findlay, D. M. Sexton, P. M. and Furness, S. G. B. (2017). Calcitonin ☆. In: *Reference Module in Neuroscience and Biobehavioral Psychology*. Elsevier, pp. 220–230. doi: 10.1016/B978-0-12-809324-5.03223-5.
- Pande, V. S. Beauchamp, K. and Bowman, G. R. (2010). Everything you wanted to know about Markov State Models but were afraid to ask. *Methods (San Diego, Calif.)*, **52**(1), pp. 99–105. doi: 10.1016/j.ymeth.2010.06.002.
- Peeters, M. C. van Westen, G. J. P. Li, Q. and IJzerman, A. P. (2011). Importance of the extracellular loops in G protein-coupled receptors for ligand recognition and receptor activation. *Trends in pharmacological sciences*, **32**(1), pp. 35–42. doi: 10.1016/j.tips.2010.10.001.
- Pham, V. Zhu, Y. Dal Maso, E. Reynolds, C. A. Deganutti, G. Atanasio, S. Hick, C. A. Yang, D. Christopoulos, A. Hay, D. L. Furness, S. G. B. Wang, M. Wootten, D. and Sexton, P. M. (2019). Deconvoluting the Molecular Control of Binding and Signaling at the Amylin 3 Receptor: RAMP3 Alters Signal Propagation through Extracellular Loops of the Calcitonin Receptor. *ACS pharmacology &*

- translational science*. American Chemical Society, **2**(3), pp. 183–197. doi: 10.1021/acsptsci.9b00010.
- Pierce, K. L. Premont, R. T. and Lefkowitz, R. J. (2002). Signalling: Seven-transmembrane receptors. *Nature Reviews Molecular Cell Biology*, **3**(9), pp. 639–650. doi: 10.1038/nrm908.
- Pin, J.-P. Galvez, T. and Prézeau, L. (2003). Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. *Pharmacology & therapeutics*, **98**(3), pp. 325–54. URL <http://www.ncbi.nlm.nih.gov/pubmed/12782243>.
- Pioszak, A. A. and Hay, D. L. (2020). RAMPs as allosteric modulators of the calcitonin and calcitonin-like class B G protein-coupled receptors. *Advances in pharmacology (San Diego, Calif.)*. 1st edition. Elsevier Inc., **88**, pp. 115–141. doi: 10.1016/bs.apha.2020.01.001.
- Pupko, T. Bell, R. E. Mayrose, I. Glaser, F. and Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics (Oxford, England)*, **18 Suppl 1**(August 2002), pp. S71–7. doi: 10.1093/bioinformatics/18.suppl.
- Rasmussen, S. G. F. *et al.* (2011). Crystal structure of the  $\beta_2$  adrenergic receptor-Gs protein complex. *Nature*, **477**(7366), pp. 549–55. doi: 10.1038/nature10361.
- Robertson, N. Rappas, M. Doré, A. S. Brown, J. Bottegoni, G. Koglin, M. Cansfield, J. Jazayeri, A. Cooke, R. M. and Marshall, F. H. (2018). Structure of the complement C5a receptor bound to the extra-helical antagonist NDT9513727. *Nature*, **553**(7686), pp. 111–114. doi: 10.1038/nature25025.
- Roche, D. B. Buenavista, M. T. Tetchner, S. J. and McGuffin, L. J. (2011). The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. *Nucleic acids research*, **39**(Web Server issue), pp. W171–6. doi: 10.1093/nar/gkr184.
- Roy, A. Kucukural, A. and Zhang, Y. (2010). I-TASSER: a unified platform for



## REFERENCES

---

- automated protein structure and function prediction. *Nature protocols*, **5**(4), pp. 725–38. doi: 10.1038/nprot.2010.5.
- Sabbadin, D. Salmaso, V. Sturlese, M. and Moro, S. (2018). Supervised Molecular Dynamics (SuMD) Approaches in Drug Design. *Methods in molecular biology (Clifton, N.J.)*, **1824**(Md), pp. 287–298. doi: 10.1007/978-1-4939-8630-9\_17.
- Sabbadin, D. and Moro, S. (2014). Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale. *Journal of chemical information and modeling*, **54**(2), pp. 372–6. doi: 10.1021/ci400766b.
- Shaw, D. E. (2013). Millisecond-Long Molecular Dynamics Simulations of Proteins on a Special-Purpose Machine. *Biophysical Journal*. Biophysical Society, **104**(2), p. 45a. doi: 10.1016/j.bpj.2012.11.289.
- Sievers, F. Wilm, A. Dineen, D. Gibson, T. J. Karplus, K. Li, W. Lopez, R. McWilliam, H. Remmert, M. Söding, J. Thompson, J. D. and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, **7**, p. 539. doi: 10.1038/msb.2011.75.
- Söding, J. Biegert, A. and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research*, **33**(Web Server issue), pp. W244–8. doi: 10.1093/nar/gki408.
- Sommer, B. (2013). Membrane Packing Problems: A short Review on computational Membrane Modeling Methods and Tools. *Computational and structural biotechnology journal*, **5**(6), p. e201302014. doi: 10.5936/csbj.201302014.
- Srivastava, A. Yano, J. Hirozane, Y. Kefala, G. Gruswitz, F. Snell, G. Lane, W. Ivetac, A. Aertgeerts, K. Nguyen, J. Jennings, A. and Okada, K. (2014). High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature*, **513**(7516), pp. 124–7. doi: 10.1038/nature13494.
- Stanley, N. Pardo, L. and Fabritiis, G. De (2016). The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor. *Scientific reports*. Nature Publishing Group, **6**, p. 22639. doi: 10.1038/srep22639.

- Stevens, R. C. Cherezov, V. Katritch, V. Abagyan, R. Kuhn, P. Rosen, H. and Wüthrich, K. (2013). The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. *Nature reviews. Drug discovery*. Nature Publishing Group, **12**(1), pp. 25–34. doi: 10.1038/nrd3859.
- Suárez, E. Adelman, J. L. and Zuckerman, D. M. (2016). Accurate Estimation of Protein Folding and Unfolding Times: Beyond Markov State Models. *Journal of chemical theory and computation*, **12**(8), pp. 3473–81. doi: 10.1021/acs.jctc.6b00339.
- Sum, C. S. Tikhonova, I. G. Costanzi, S. and Gershengorn, M. C. (2009). Two arginine-glutamate ionic locks near the extracellular surface of FFAR1 gate receptor activation. *The Journal of biological chemistry*, **284**(6), pp. 3529–36. doi: 10.1074/jbc.M806987200.
- Taddese, B. Upton, G. J. G. Bailey, G. R. Jordan, S. R. D. Abdulla, N. Y. Reeves, P. J. and Reynolds, C. A. (2014). Do plants contain g protein-coupled receptors?. *Plant physiology*, **164**(1), pp. 287–307. doi: 10.1104/pp.113.228874.
- Teng, D. Chen, J. Li, D. Wu, Z. Li, W. Tang, Y. and Liu, G. (2020). Computational Insights into Molecular Activation and Positive Cooperative Mechanisms of FFAR1 Modulators. *Journal of Chemical Information and Modeling*, p. acs.jcim.0c00030. doi: 10.1021/acs.jcim.0c00030.
- Tetsuka, M. Saito, Y. Imai, K. Doi, H. and Maruyama, K. (2004). The basic residues in the membrane-proximal C-terminal tail of the rat melanin-concentrating hormone receptor 1 are required for receptor function. *Endocrinology*, **145**(8), pp. 3712–23. doi: 10.1210/en.2003-1638.
- Thomsen, W. Frazer, J. and Unett, D. (2005). Functional assays for screening GPCR targets. *Current opinion in biotechnology*, **16**(6), pp. 655–65. doi: 10.1016/j.copbio.2005.10.008.
- Tikhonova, I. G. (2017). Application of GPCR Structures for Modelling of Free Fatty Acid Receptors. *Handbook of experimental pharmacology*, **236**, pp. 57–77. doi: 10.1007/164\_2016\_52.
- Tikhonova, I. G. and Poerio, E. (2015). Free fatty acid receptors: structural models and

- p elucidation of ligand binding interactions.
- BMC structural biology*
- . BMC Structural Biology,
- 15**
- (1), p. 16. doi: 10.1186/s12900-015-0044-2.
- Tribello, G. A. Bonomi, M. Branduardi, D. Camilloni, C. and Bussi, G. (2014). PLUMED 2: New feathers for an old bird. *Computer Physics Communications*. Elsevier B.V., **185**(2), pp. 604–613. doi: 10.1016/j.cpc.2013.09.018.
- Unger, V. M. Hargrave, P. A. Baldwin, J. M. and Schertler, G. F. (1997). Arrangement of rhodopsin transmembrane alpha-helices. *Nature*, **389**(6647), pp. 203–6. doi: 10.1038/38316.
- Vanommeslaeghe, K. Hatcher, E. Acharya, C. Kundu, S. Zhong, S. Shim, J. Darian, E. Guvench, O. Lopes, P. Vorobyov, I. and Mackerell, A. D. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry*, **31**(4), pp. 671–90. doi: 10.1002/jcc.21367.
- Vanommeslaeghe, K. and MacKerell, A. D. (2012). Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *Journal of chemical information and modeling*, **52**(12), pp. 3144–54. doi: 10.1021/ci300363c.
- Vanommeslaeghe, K. Raman, E. P. and MacKerell, A. D. (2012). Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *Journal of chemical information and modeling*, **52**(12), pp. 3155–68. doi: 10.1021/ci3003649.
- Vohra, S. Taddese, B. Conner, A. C. Poyner, D. R. Hay, D. L. Barwell, J. Reeves, P. J. Upton, G. J. G. and Reynolds, C. a (2013). Similarity between class A and class B G-protein-coupled receptors exemplified through calcitonin gene-related peptide receptor modelling and mutagenesis studies. *Journal of the Royal Society, Interface*, **10**(79), p. 20120846. doi: 10.1098/rsif.2012.0846.
- Wang, C. Wu, H. Evron, T. Vardy, E. Han, G. W. Huang, X.-P. Hufeisen, S. J. Mangano, T. J. Urban, D. J. Katritch, V. Cherezov, V. Caron, M. G. Roth, B. L. and Stevens, R. C. (2014). Structural basis for Smoothed receptor modulation and chemoresistance to anticancer drugs. *Nature Communications*. Nature Publishing Group, **5**, p. 4355. doi: 10.1038/ncomms5355.

- Waterhouse, A. M. Procter, J. B. Martin, D. M. A. Clamp, M. and Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*, **25**(9), pp. 1189–91. doi: 10.1093/bioinformatics/btp033.
- Watterson, K. R. Hudson, B. D. Ulven, T. and Milligan, G. (2014). Treatment of Type 2 Diabetes by Free Fatty Acid Receptor Agonists. *Frontiers in Endocrinology*, **5**, p. 137. doi: 10.3389/fendo.2014.00137.
- Webb, B. and Sali, A. (2014). Comparative Protein Structure Modeling Using MODELLER. *Current protocols in bioinformatics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., **47**, pp. 5.6.1-32. doi: 10.1002/0471250953.bi0506s47.
- Webster, D. (2000). *Protein Structure Prediction*. New Jersey: Humana Press. doi: 10.1385/1592593682.
- Weis, W. I. and Kobilka, B. K. (2018). The Molecular Basis of G Protein–Coupled Receptor Activation. *Annual Review of Biochemistry*, **87**(1), pp. 897–919. doi: 10.1146/annurev-biochem-060614-033910.
- Wheatley, M. Wootten, D. Conner, M. T. Simms, J. Kendrick, R. Logan, R. T. Poyner, D. R. and Barwell, J. (2012). Lifting the lid on GPCRs: the role of extracellular loops. *British journal of pharmacology*, **165**(6), pp. 1688–1703. doi: 10.1111/j.1476-5381.2011.01629.x.
- Wootten, D. Simms, J. Miller, L. J. Christopoulos, A. and Sexton, P. M. (2013). Polar transmembrane interactions drive formation of ligand-specific and signal pathway-biased family B G protein-coupled receptor conformations. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(13), pp. 5211–6. doi: 10.1073/pnas.1221585110.
- World Health Organization (2016a). *Diabetes mellitus fact sheet*. URL <https://www.who.int/news-room/fact-sheets/detail/diabetes>.
- World Health Organization (2016b). *Global report on diabetes*. URL [https://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257\\_eng.pdf?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257_eng.pdf?sequence=1).
- Xu, D. and Zhang, Y. (2012). Ab initio protein structure assembly using continuous

## REFERENCES

---

- structure fragments and optimized knowledge-based force field. *Proteins*, **80**(7), pp. 1715–35. doi: 10.1002/prot.24065.
- Yabuki, C. Komatsu, H. Tsujihata, Y. Maeda, R. Ito, R. Matsuda-Nagasumi, K. Sakuma, K. Miyawaki, K. Kikuchi, N. Takeuchi, K. Habata, Y. and Mori, M. (2013). A novel antidiabetic drug, fasiglifam/TAK-875, acts as an ago-allosteric modulator of FFAR1. *PloS one*, **8**(10), p. e76280. doi: 10.1371/journal.pone.0076280.
- Zhang, D. Zhao, Q. and Wu, B. (2015). Structural Studies of G Protein-Coupled Receptors. *Molecules and cells*, **38**(10), pp. 836–42. doi: 10.14348/molcells.2015.0263.

## Appendices

### Appendix A. GPCRs solved structures

Most of the solved GPCR structures concern rhodopsin family (class A), divided in  $\alpha$ -rhodopsin,  $\beta$ -rhodopsin,  $\gamma$ -rhodopsin,  $\delta$ -rhodopsin, and other 7TM receptors.

About  $\alpha$ -rhodopsin, there are **thirty-two** GPCRs, including:

- **5-hydroxytryptamine receptor 1B** (5-HT1B) PDB ID: 4IAR, 4IAQ, 5V54, 6G79, 7C61;
- **5-hydroxytryptamine receptor 2A** (5-HT2A) PDB ID: 6A93, 6A94;
- **5-hydroxytryptamine receptor 2B** (5-HT2B) PDB ID: 4IB4, 4NC3, 5TUD, 5TVN, 6DRX, 6DRY, 6DRZ, 6DS0;
- **5-hydroxytryptamine receptor 2C** (5-HT2C) PDB ID: 6BQG, 6BQH;
- **Adenosine receptor A1** ( $A_1$ AR) PDB ID: 5N2S, 5UEN, 6D9H;
- **Adenosine receptor A2A** ( $A_{2A}$ AR) PDB ID: 2YDO, 2YDV, 3EML, 3PWH, 3QAK, 3REY, 3RFM, 3UZA, 3UZC, 3VG9, 3VGA, 4EIY, 4UG2, 4UHR, 5G53, 5IU4, 5IU7, 5IU8, 5IUA, 5IUB, 5JTB, 5K2A, 5K2B, 5K2C, 5K2D, 5MZJ, 5MZP, 5N2R, 5NLX, 5NM2, 5NM4, 5OLG, 5OLH, 5OLO, 5OLV, 5OLZ, 5OM1, 5OM4, 5UIG, 5UVI, 5VRA, 5WF5, 5WF6, 6AQF, 6GDG, 6GT3, 6JZH, 6MH8, 6PS7, 6S0Q;
- **Alpha-2A adrenergic receptor** ( $\alpha_{2A}$ -AR) PDB ID: 6KUX, 6KUY;
- **Alpha-2B adrenergic receptor** ( $\alpha_{2B}$ -AR) PDB ID: 6K41, 6K42;
- **Alpha-2C adrenergic receptor** ( $\alpha_{2C}$ -AR) PDB ID: 6KUW;

- **Beta-1 adrenergic receptor** ( $\beta_1$ -AR) PDB ID: 2VT4, 2Y00, 2Y01, 2Y02, 2Y03, 2Y04, 2YCW, 2YCY, 2YCX, 2Y CZ, 3ZPQ, 3ZPR, 4AMI, 4AMJ, 4BVN, 4GPO, 5A8E, 5F8U, 6H7J, 6H7L, 6H7M, 6H7N, 6H7O, 6IBL, 6TKO;
- **Beta-2 adrenergic receptor** ( $\beta_2$ -AR) PDB ID: 2R4R, 2R4S, 2RH1, 3D4S, 3KJ6, 3NY8, 3NY9, 3NYA, 3P0G, 3PDS, 3SN6, 4GBR, 4LDE, 4LDL, 4LDO, 4QKX, 5D5A, 5D5B, 5D6L, 5JQH, 5X7D, 6MXT, 6E67, 6N48, 6NI3, 6OBA, 6PRZ, 6PS0, 6PS1, 6PS2, 6PS3, 6PS4, 6PS5, 6PS6, 7BZ2;
- **Cannabinoid receptor 1** (CB1) PDB ID: 5TGZ, 5U09, 5XR8, 5XRA, 6KPG, 6KQI, 6N4B;
- **Cannabinoid receptor 2** (CB2) PDB ID: 5ZTY, 6KPC, 6KPF, 6PT0;
- **D(2) dopamine receptor** (D2R) PDB ID: 6CM4, 6LUQ, 6VMS;
- **D(3) dopamine receptor** (D3R) PDB ID: 3PBL;
- **D(4) dopamine receptor** (D4R) PDB ID: 5WIU, 5WIV, 6IQL;
- **G-protein coupled receptor 52** (GPR52) PDB ID: 6LI0, 6LI1, 6LI2, 6LI3;
- **Lysophosphatidic acid receptor 1** (LPA<sub>1</sub>) PDB ID: 4Z34, 4Z35, 4Z36;
- **Lysophosphatidic acid receptor 6A** (LPA<sub>6A</sub>) PDB ID: 5XSZ;
- **Melanocortin receptor 4** (MC4-R) PDB ID: 6W25;
- **Melatonin receptor type 1A** (MT<sub>1</sub>) PDB ID: 6ME2, 6ME3, 6ME4, 6ME5, 6PS8;
- **Melatonin receptor type 1B** (MT<sub>2</sub>) PDB ID: 6ME6, 6ME7, 6ME8, 6ME9;
- **Muscarinic acetylcholine receptor M1** (M1R) PDB ID: 5CXV, 6OIJ, 6WJC;
- **Muscarinic acetylcholine receptor M2** (M2R) PDB ID: 3UON, 4MQS, 4MQT, 5YC8, 5ZK3, 5ZK8, 5ZKB, 5ZKC, 6OIK, 6U1N;

- **Muscarinic acetylcholine receptor M3** (M3R) PDB ID: 4DAJ, 4U14, 4U15, 4U16, 5ZHP;
- **Muscarinic acetylcholine receptor M4** (M4R) PDB ID: 5DSG, 6KP6;
- **Muscarinic acetylcholine receptor M5** (M5R) PDB ID: 6OL9;
- **Prostaglandin E2 receptor EP3 subtype** (PE2R3) PDB ID: 6AK3, 6M9T, 5YHL, 5YWY;
- **Prostaglandin E2 receptor EP4 subtype** (PE2R4) PDB ID: 5YHL, 5YWY;
- **Rhodopsin** (RHO) PDB ID: 1F88, 1GZM, 1HZX, 1L9H, 1U19, 2G87, 2HPY, 2I35, 2I36, 2I37, 2J4Y, 2PED, 2X72, 2Z73, 2ZIY, 3AYM, 3AYN, 3C9L, 3C9M, 3CAP, 3DQB, 3OAX, 3PQR, 3PXO, 4A4M, 4BEY, 4BEZ, 4J4Q, 4PXF, 4WW3, 4X1H, 4ZWJ, 5DGY, 5DYS, 5EN0, 5TE3, 5TE5, 5W0P, 5WKT, 6CMO, 6FK6, 6FK7, 6FK8, 6FK9, 6FKA, 6FKB, 6FKC, 6FKD, 6FUF, 6NWE, 6OFJ, 6OY9, 6OYA, 6PEL, 6PGS, 6PH7, 6QNO;
- **Sphingosine 1-phosphate receptor 1** (S1P<sub>1</sub>) PDB ID: 3V2Y, 3V2W;
- **Thromboxane A2 receptor** (TXA<sub>2</sub>-R) PDB ID: 6IIU, 6IIV.

About  $\beta$ -rhodopsin, there are **eight** GPCRs, including:

- **Endothelin receptor type B** (ET-B) PDB ID: 5GLH, 5GLI, 5X93, 5XPR, 6IGK, 6IGL, 6K1Q, 6LRY;
- **Growth hormone secretagogue receptor type 1** (GHS-R) PDB ID: 6KO5;
- **Neuropeptide Y receptor type 1** (NPY1-R) PDB ID: 5ZBH, 5ZBQ;
- **Neurotensin receptor type 1** (NTR1) PDB ID: 3ZEV, 4BUO, 4BV0, 4BWB, 4GRV, 4XEE, 4XES, 5T04, 6OS9, 6OSA, 6PWC, 6UP7;
- **Orexin receptor type 1** (Ox1R) PDB ID: 4ZJ8, 4ZJC, 6TO7, 6TOD, 6TOS, 6TOT, 6TP3, 6TP4, 6TP6, 6TQ4, 6TQ6, 6TQ7, 6TQ9, 6V9S;



- **Orexin receptor type 2 (Ox2R)** PDB ID: 4S0V, 5WQC, 5WS3, 6TPG, 6TPJ, 6TPN;
- **Oxytocin receptor (OT-R)** PDB ID: 6TPK;
- **Tachykinin receptor 1** or **Neurokinin receptor 1 (NK1-R)** PDB ID: 6E59, 6HLL, 6HLO, 6HLP, 6J20, 6J21.

About γ-rhodopsin, there are **nineteen** GPCRs, including:

- **Apelin receptor (APLNR)** PDB ID: 5VBL, 6KNM;
- **C-C chemokine receptor type 2 (CCR2)** PDB ID: 5T1A, 6GPS, 6GPX;
- **C-C chemokine receptor type 5 (CCR5)** PDB ID: 4MBS, 5UIW, 6AKX, 6AKY, 6MEO, 6MET;
- **C-C chemokine receptor type 6 (CCR6)** PDB ID: 6WWZ;
- **C-C chemokine receptor type 7 (CCR7)** PDB ID: 6QZH;
- **C-C chemokine receptor type 9 (CCR9)** PDB ID: 5LWE;
- **C-X-C chemokine receptor type 1 (CXCR1)** PDB ID: 2LNL;
- **C-X-C chemokine receptor type 4 (CXCR4)** PDB ID: 3ODU, 3OE0, 3OE6, 3OE8, 3OE9, 4RWS;
- **C5a anaphylatoxin chemotactic receptor 1 (C5aR1)** PDB ID: 5O9H, 6C1Q, 6C1R;
- **Delta-type opioid receptor (δ-OR)** PDB ID: 4EJ4, 4N6H, 4RWA, 4RWD, 6PT2, 6PT3;
- **G-protein coupled receptor homolog US28 (HHRF3)** PDB ID: 4XT1, 4XT3, 5WB1, 5WB2;
- **Kappa-type opioid receptor (κ-OR)** PDB ID: 4DJH, 6B73, 6VI4;
- **Leukotriene receptor (BLT1)** PDB ID: 5X33;

- **Mu-type opioid receptor** ( $\mu$ -OR) PDB ID: 4DKL, 5C1M, 6DDE, 6DDF;
- **N-formyl peptide receptor 2** (FMLP-R-I) PDB ID: 6LW5, 6OMM;
- **Nociceptin receptor** (NOP) PDB ID: 4EA3, 5DHG, 5DHH;
- **Prostaglandin D2 receptor 2** (PD2R2) PDB ID: 6D26, 6D27;
- **Type-1 angiotensin II receptor** (AT1R) PDB ID: 4YAY, 4ZUD, 6DO1, 6OS0, 6OS1, 6OS2;
- **Type-2 angiotensin II receptor** (AT2R) PDB ID: 5UNF, 5UNG, 5UNH, 5XJM, 6JOD, 7C6A.

About  $\delta$ -rhodopsin, there are **nine** GPCRs, including:

- **Cysteinyl leukotriene receptor 1** (CysLTR1) PDB ID: 6RZ4, 6RZ5;
- **Cysteinyl leukotriene receptor 2** (CysLTR2) PDB ID: 6RZ6, 6RZ7, 6RZ8, 6RZ9;
- **Histamine H1 receptor** (H1R) PDB ID: 3RZE;
- **P2Y purinoceptor 1** (P2Y<sub>1</sub>R) PDB ID: 4XNV, 4XNW;
- **P2Y purinoceptor 12** (P2Y<sub>12</sub>R) PDB ID: 4NTJ, 4PXZ, 4PY0;
- **Platelet-activating factor receptors** (PAF-R) PDB ID: 5ZKP, 5ZKQ;
- **Proteinase-activated receptor 1** (PAR1) PDB ID: 3VW7;
- **Proteinase-activated receptor 2** (PAR2) PDB ID: 5NDD, 5NDZ, 5NJ6;
- **Succinate receptor 1** (SUCR1) PDB ID: 6IBB, 6RNK.

**One** from other 7TM receptors, including:

- **Free fatty acid receptor 1** (FFAR1) PDB ID: 4PHU, 5KW2, 5TZR, 5TZY.

**Seventy-three** solved GPCR structures concern classes B, C, and F.

**Nine** GPCRs of secretin family (class B), including:

- **Calcitonin gene-related peptide 1 (CGRP)** PDB ID: 6E3Y, 6UUN, 6UUS, 6UVA;
- **Calcitonin receptor (CTR)** PDB ID: 5UZ7, 6NIY;
- **Corticotropin releasing factor receptor 1 (CRFR1)** PDB ID: 4K5Y, 4Z9G, 6P9X, 6PB0;
- **Corticotropin releasing factor receptor 2 (CRFR2)** PDB ID: 6PB1;
- **Glucagon receptor (GLR)** PDB ID: 4L6R, 5EE7, 5VAI, 5XEZ, 5XF1, 5YQZ, 6LMK, 6LML, 6WHC, 6WPW;
- **Glucagon-like peptide 1 receptor (GLP-1R)** PDB ID: 5NX2, 5VEW, 5VEX, 6B3J, 6KJV, 6KK1, 6KK7, 6LN2, 6ORV, 6VCB, 7C2E;
- **Parathyroid hormone (PTH)** PDB ID: 6FJ3, 6NBF, 6NBH, 6NBI;
- **Pituitary adenylate cyclase-activating polypeptide type I receptor (PACAP-R1)** PDB ID: 6LPB, 6M1H, 6M1I, 6P9Y;
- **Secretin receptor (SCTR)** PDB ID: 6WI9, 6WZG.

**Four** GPCRs of glutamate family (class C), including:

- **Gamma-aminobutyric acid type B receptor subunit 1 (GABR1)** PDB ID: 6W2Y;
- **Gamma-aminobutyric acid type B receptor subunit 2 (GABR2)** PDB ID: 6UO8, 6UO9, 6UOA, 6VJM, 6W2X, 6WIV, 7C7Q, 7C7S;
- **Metabotropic glutamate receptor 1 (mGluR1)** PDB ID: 4OR2;
- **Metabotropic glutamate receptor 5 (mGluR5)** PDB ID: 4OO9, 5CGC, 5CGD, 6FFH, 6FFI, 6N51, 6N52.

**Two** GPCR of frizzled family (class F), including:

- **Frizzled-4 (Fz-4)** PDB ID: 6BD4;

- 
- **Smoothened homolog** (SMO) PDB ID: 4JKV, 4N4W, 4O9R, 4QIM, 4QIN, 5L7D, 5L7I, 5V56, 5V57, 6D32, 6D35, 6O3C, 6OT0.



## **Appendix B. Molecular Dynamics (MD) simulation applied to Androglobin: a novel heme protein**

Parallel to this project, a study on Androglobin (Adgb) has been done. There is an ongoing collaboration with Dr Brandon Reeder's group that aims to biochemically characterise the Adgb domains from the experimental point of view. Preliminary data about the calpain, heme, and IQ domains have been recently collected and the manuscript is in preparation.

Androglobin (Adgb), also known as Calpain-7-like protein (UniProtKB ID: Q8N7X0), is one of the eight globin types in vertebrates such as haemoglobin (Hb), myoglobin (Mb), neuroglobin (Ngb), cytoglobin (Cygb), globin X (GbX), globin Y (GbY), and globin E (GbE) (Burmester and Hankeln, 2014). While human Hb is one of the most studied proteins in science, Adgb is the least studied and understood proteins of the globin superfamily. To date, due to its structure instability during the purification process, only two articles have been published about this interesting protein (Hoogewijs *et al.*, 2012; Huang *et al.*, 2014) and its experimental structure is still unknown. The Adgb protein consists of 1667 amino acids in which were identified an N-terminal calpain-like domain, an internal, circularly permuted globin domain, and an IQ calmodulin-binding motif (Hoogewijs *et al.*, 2012) (Figure B.1).

# APPENDIX

10	20	30	40	50
MASKQTKKKE	VHRINSAHGS	DKSKDFYPFG	SNVQSGSTEQ	KKGKFPLWPE
60	70	80	90	100
WSEADINSEK	WDAGKGAKEK	<b>DKTGKSPVFH</b>	<b>FFEDPEGKIE</b>	<b>LPPSLKIYSW</b>
110	120	130	140	150
<b>KRPQDILFSQ</b>	<b>TPVVVKNEIT</b>	<b>FDLFSANEHL</b>	<b>LCSELMRWII</b>	<b>SEIYAVWKIF</b>
160	170	180	190	200
<b>NGGILSNYFK</b>	<b>GTSGEPELLP</b>	<b>WKPWEHIYSL</b>	<b>CKAVKGHMPL</b>	<b>FNSYGKYVVK</b>
210	220	230	240	250
<b>LYWMGCWRKI</b>	<b>TIDDFLPFDE</b>	<b>DNNLLLPATT</b>	<b>YEFELWPMLL</b>	<b>SKAI IKLANI</b>
260	270	280	290	300
<b>DIHVADRREL</b>	<b>GEFTVIHALT</b>	<b>GWLPEVISLH</b>	<b>PGYMDKVVWL</b>	<b>LKEILPEFKL</b>
310	320	330	340	350
<b>SDEASSESKI</b>	<b>AVLDSKLKEP</b>	<b>GKEGKEGKEI</b>	<b>KDGKEVKDVK</b>	<b>EFKPESSLTT</b>
360	370	380	390	400
<b>LKAPEKSDKV</b>	<b>PKEKADARDI</b>	<b>GKKRSKDGEK</b>	<b>EKFKFSLHGS</b>	<b>RPSSEVQYSV</b>
410	420	430	440	450
<b>QSLSDCSSAI</b>	<b>QTSBMVYAT</b>	FTPLYLFENK	IFSLEKMADS	AEKLREYGLS
460	470	480	490	500
HICSHPLVLT	RSRSCPLVAP	PKPPPLPPWK	LIRQKKETVI	TDEAQELIVK
510	520	530	540	550
KPERFLEISS	PFLNYRMTPF	TIPTEMHFVR	SLIKKGIPPG	SDLPSVSETD
560	570	580	590	600
ETATHSQTDL	SQITKATSQG	NTASQVILGK	GTDEQTDGFL	GDAHQSDGLN
610	620	630	640	650
LEREIVSQTT	ATQEKSQEEL	PTTNSVSKE	<u>IWLDFEDFCV</u>	<u>CFQNIYIFHK</u>
660	670	680	690	700
<u>PSSYCLNFQK</u>	<u>SEFKFSEERV</u>	<u>SYLFLVDSLK</u>	<u>PIELLVCFSA</u>	<u>LVRWGEYGAL</u>
710	720	730	740	750
<u>TKDSPPIEPG</u>	<u>LLTAETFSWK</u>	<u>SLKPGSLVLK</u>	<u>IHTYATKATV</u>	<u>VRLPVGRHML</u>
760	770	780	790	800
<u>LFNAYSPVGH</u>	<u>SIHICSMVSF</u>	<u>VIGDEHVVL</u>	<b>NFEPESCRFT</b>	<b>EQSLIMKAI</b>
810	820	830	840	850
<b>GNVIANFKDK</b>	<b>GKLSAALKDL</b>	<b>QTAHYVPFHH</b>	<b>DKELTAQHFR</b>	<b>VFHLSLWRLM</b>
860	870	880	890	900
<b>KKVQITKPPP</b>	<b>NFKFAFRAMV</b>	<b>LDLELLNSSL</b>	<b>EEVSLVEWLD</b>	<b>VKYCMPTSDK</b>
910	920	930	940	950
<b>EYSAEEVAAA</b>	<b>IKIQAMWRGT</b>	<b>YVRLLMKARI</b>	<b>PDTKEN ISVA</b>	<b>DTLQKVWAVL</b>
960	970	980	990	1000
<b>EMNLEQYAVS</b>	<b>LLRLMFKSKC</b>	<b>KSLESYPCYQ</b>	<b>DEETKIAFAD</b>	YTVTYQEQPP
1010	1020	1030	1040	1050
NSWFIVFRET	FLVHQDMILV	PKVYTTLPIC	ILHIVNNDTM	EQVPKVFQKV
1060	1070	1080	1090	1100
VPYLYTKNKK	GYTFVAEAF	GDTYVAASRW	KLRLIGSSAP	LPCLSRDSPC
1110	1120	1130	1140	1150
NSFAIKEIRD	YYIPNDKKIL	FRYSVKVLTP	QPATIQVRTS	KPDAFIKLQV
1160	1170	1180	1190	1200
LENEETMVSS	TGKGQAIIPA	FHFLKSEKGL	SSQSSKHILS	FHSASKKEQE
1210	1220	1230	1240	1250
VYVKKKAAQG	IQKSPKGRAV	SAIQDIGLPL	VEEETTSTPT	REDSSSTPLQ
1260	1270	1280	1290	1300
NYKYIIQCSV	LYNSWPLTES	QLTFVQALKD	LKKSNTKAYG	ERHEELINLG

1310	1320	1330	1340	1350
SPDSHTISEG	QKSSVTSKTT	RKGKEKSSEK	EKTAKEKQAP	RFEPQISTVH
1360	1370	1380	1390	1400
PQQEDPNKPY	WILRLVTEHN	ESELFVKKD	TERADEIRAM	KQAWETTEPG
1410	1420	1430	1440	1450
RAIKASQARL	HYLSGFIKKT	SDAESPPISE	SQTKPKEEVE	TAARGVKEPN
1460	1470	1480	1490	1500
SKNSAGSESK	EMTQTGSGSA	VWKKWQLTKG	LRDVAKSTSS	ESGGVSSPGK
1510	1520	1530	1540	1550
EEREQSTRKE	NIQTGPRTRS	PTILETSPRL	IRKALEFMDL	SQYVRKTDTD
1560	1570	1580	1590	1600
PLLQTDDELNQ	QQAMQKAE EI	HQFRQHRTRV	LSIRNIDQEE	RLKLKDEVLD
1610	1620	1630	1640	1650
MYKEMQDSLD	EARQKIFDIR	EEYRNKLLEA	EHLKLETLAA	QEAAMKLETE
1660				
KMTPAPDTQK	KKKGKKK			

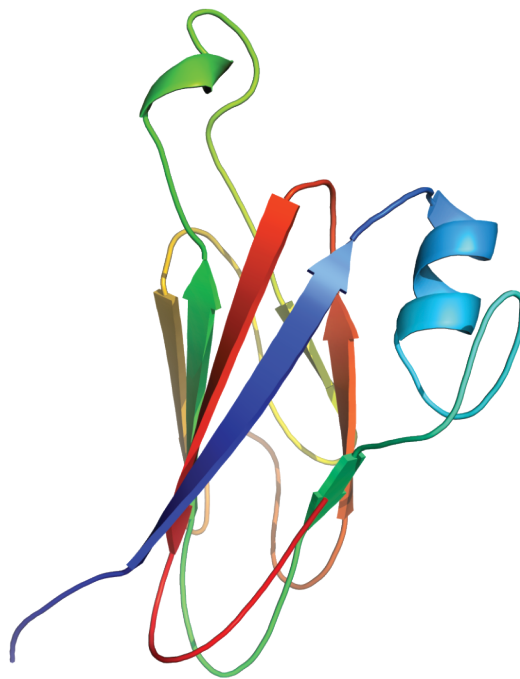
**Figure B.1 | Analysis of Adgb sequence.** The Adgb sequence (UniProtKB entry name: ADGB\_HUMAN) consists of 1667 amino acids where have been identified three main domains: (i) N-terminal calpain-like domain (residues 70-411) shown in green; (ii) an internal, circular permuted globin domain (residues 781-890, 937-985), shown in red, and (iii) an IQ calmodulin-binding motif (residues 891-936), shown in yellow. The sequence section of Adgb prior to the globin domain sequence is underlined and it has been studied in this project.

The aim of this study was to provide some insights on the three-dimensional protein structure of Adgb using an alignment method, molecular modelling, and molecular dynamics. These analyses were done for the sequence section of Adgb prior to the globin domain sequence (Figure B.1 - underlined section) and the globin domain sequence, consisted of (i) the typical 8 alpha-helical structure (helices A-H) binding a heme group (Figure B.1 - red section) and (ii) the IQ domain (Figure B.1 - yellow section), which is unique amongst heme proteins to Adgb.

With regard to the sequence sections of Adgb prior to the globin domain sequence, fold recognition servers (section 2.1) were used to do structure and function prediction of sequence sections of Adgb prior to the heme; strong hits obtained against

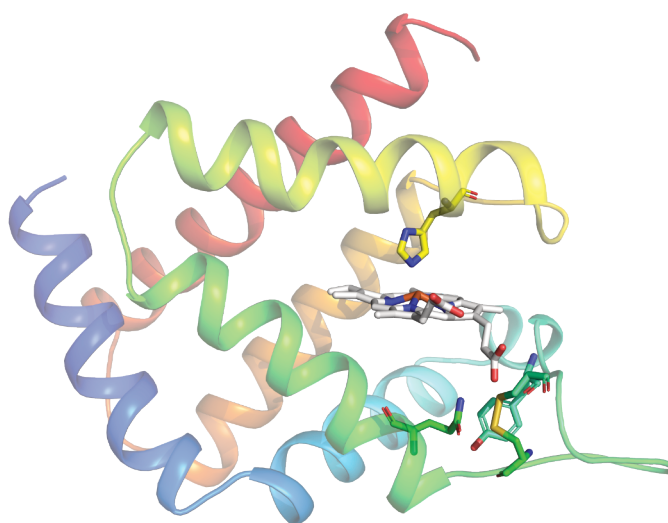


mu-calpain based on the key metric Z-score (associated with a “high”, “medium” or “low” interpretation index) were 3bow (chain A), 1kfx (chain L), and 1qxp (chain B). A model of the Adgb was built using Modeller (section 2.2) and preliminary results show an interesting model structure of the pre-internal globin domain consisting of a  $\beta$ -barrel structure (Figure B.2). According to the protein fold recognition server results, this domain is likely to be a calpain-like domain and putatively it appears that the Adgb structure comprises two of them. This information suggests new opportunities to model or purify and crystallise only a specific part of Adgb, which is a stable structural unit.



**Figure B.2 | Pre-internal globin domain of Adgb: a calpain-like domain, possibly presents twice in the Adgb structure.** The model of the pre-internal globin domain is shown in cartoon representation and consists of a  $\beta$ -barrel structure.

With regard to the globin domain sequence, several models were generated and three models (*Adgb0*, *Adgb1*, and *Adgb2*) (Figure B.3, B.4) were selected for MD simulations (Table B.1) to assess the difference in stability, enabling a putative model to be presented.



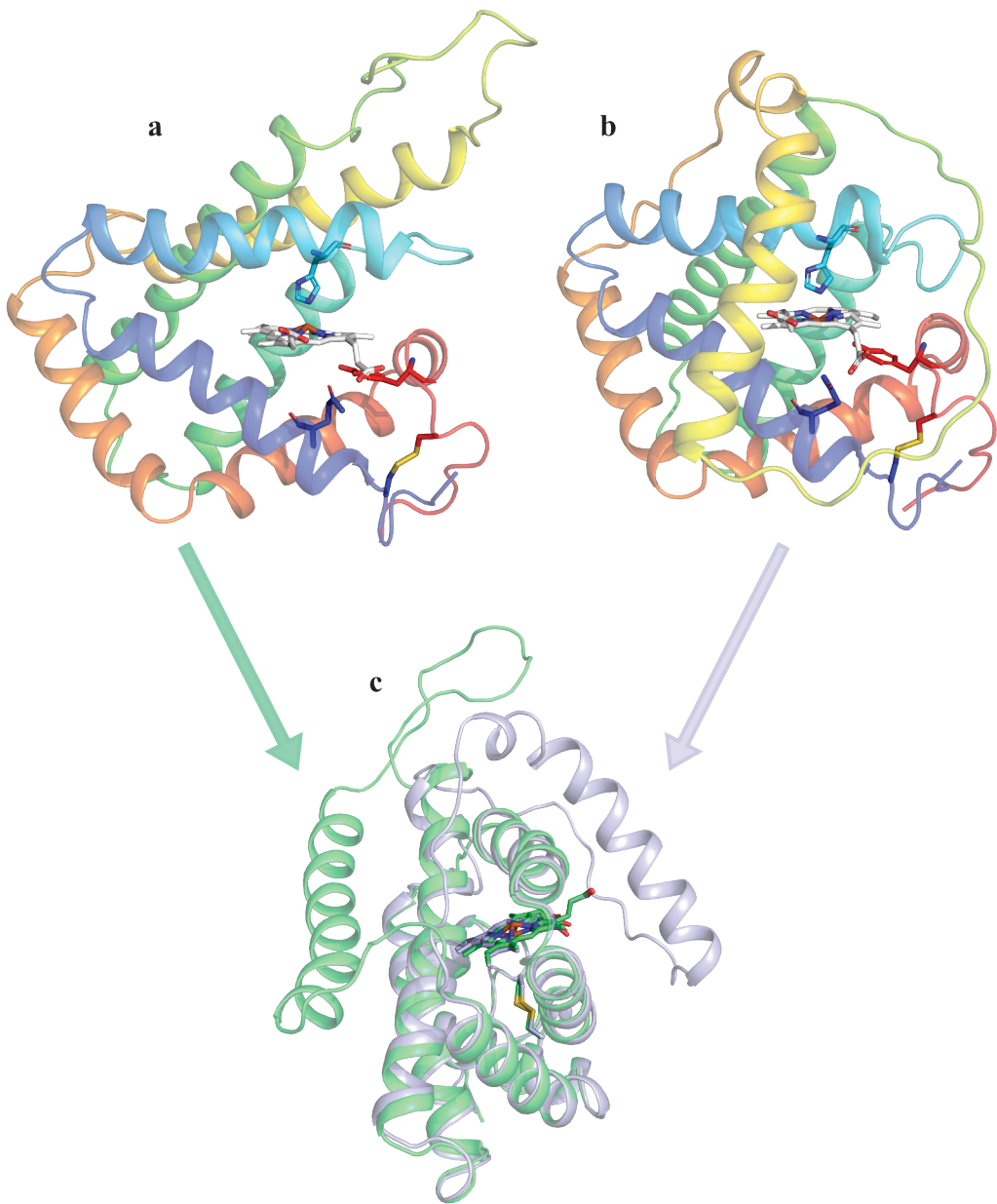
**Figure B.3 | Analysis of *Adgb0*.** The model *Adgb0* is shown in cartoon representation, while in stick representation are shown the 3 key amino acids, namely: Gln792 (in green), His824 (in yellow), and Tyr976 (in cyan) and the position of the disulfide bond between Cys787 and Cys978 (\*), forming a closed loop around the heme pocket.

**Table B.1 | Summary of all the MD simulations performed on the Androglobin models**

Model name	Model details	Number of replicas	Total MD sampling time [μs]
<i>Adgb0</i>	without IQ domain with disulfide bond (-S-S-) between Cys43(978)-Cys57(787)	3	1.5
<i>Adgb1</i>	with IQ domain behind the HEME group with disulfide bond (-S-S-) between Cys7(787)-Cys198(978)	3	1.5
	Extension of the most stable simulation	replica #2	1
<i>Adgb2</i>	with IQ domain front the HEME group with disulfide bond (-S-S-) between Cys7(787)-Cys198(978)	3	1.5
	Extension of the most stable simulation	replica #3	1

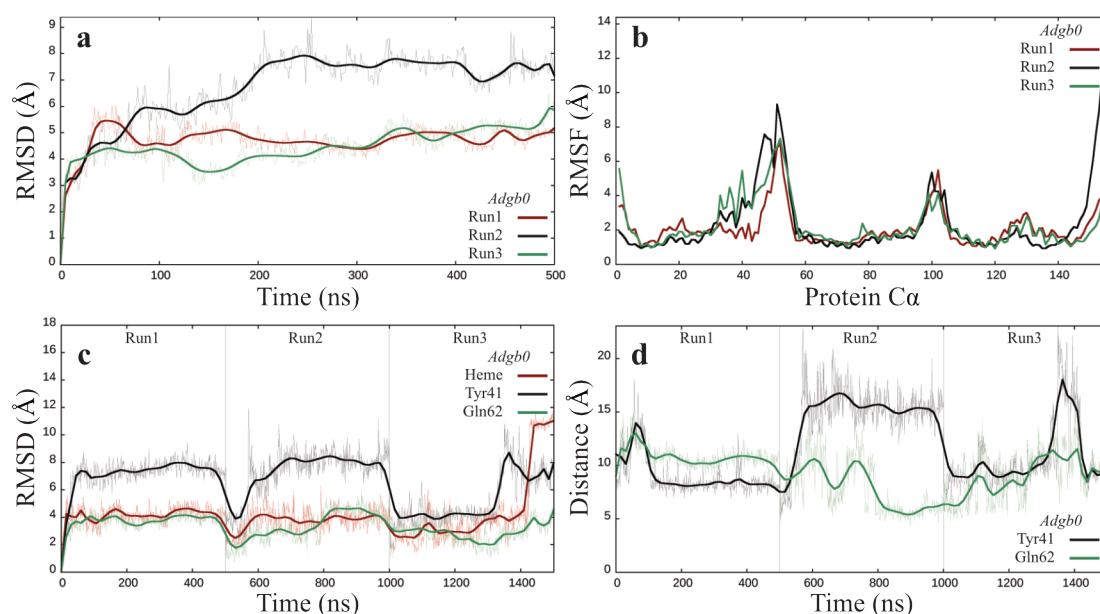
# APPENDIX

:	:	:	:	:
:	:	:	:	:
760	770	780	790	800
LFNAYSPVGH	SIHICSMVSF	VIGDEHVLP	NFEPESCRFT*	EQSLLIMKAI
810	820	830	840	850
GNVIANFKDK	GKLSAALKDL	QTAHYVPVFH	DKELTAQHFR	VFHLSLWRLM
860	870	880	890	900
KKVQITKPPP	NFKFAFRAMV	LDLELLNSSL	EEVSLVEWLD	VKYCMPTSDK
910	920	930	940	950
EYSAEEVAAA	IKIQAMWRGT	YVRLLMKARI	PDTKENISVA	DTLQKVWAVL
960	970	980	990	1000
EMNLEQYAVS	LLRLMFKSKC	KSLESYPCYQ*	DEETKLAFAFAD	YTVTYQEQQP
:	:	:	:	:
:	:	:	:	:



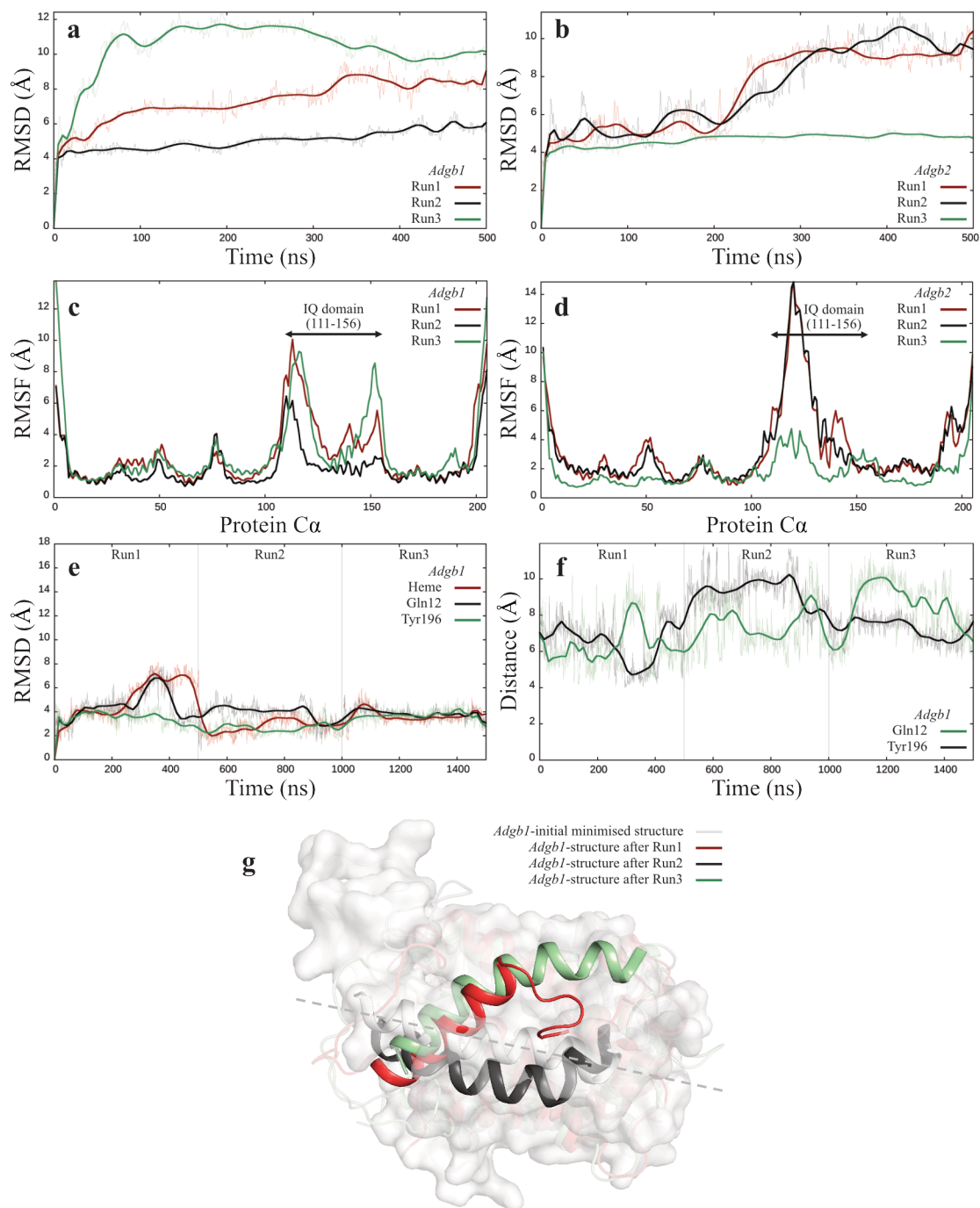
**Figure B.4 | Globin domain and IQ domain of Adgb.** The sequence above between the two arrows corresponds to the Adgb sequence of the globin and IQ domain. Two models *Adgb1* (a) and *Adgb2* (b) that have a different structural position of the IQ domain (c), were selected based on (i) the lowest DOPE score, (ii) the position of the 3 key amino acids, namely: Gln792 (in blue), His824 (in cyan), and Tyr976 (in red) and the position of the disulfide bond between Cys787 and Cys978 (\*), forming a closed loop around the heme pocket.

### Results of *Adgb0*



**Figure B.5 | Dynamic behaviour of *Adgb0*.** RMSD plot from three MD simulations of *Adgb0* (a) is shown with respect to the initial minimised structure for all the three MD simulations. RMSF plot from three MD simulations of *Adgb0* (b) was measured using the Cα of each residue. The RMSD plot for the heme and the key residues Gln62<sup>(792)</sup> and Tyr41<sup>(976)</sup> of *Adgb0* (c), and their distance from the heme iron (d), were calculated for all three MD simulations to investigate the heme pocket.

## Results of *Adgb1* and *Adgb2*



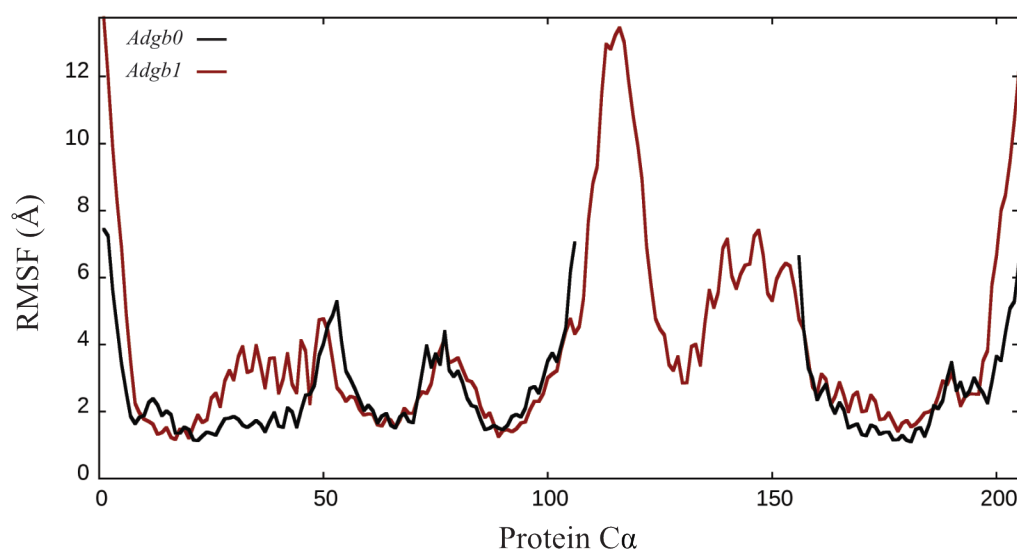
**Figure B.6 | Dynamic behaviour of the two models: *Adgb1* and *Adgb2*.** RMSD plots from three MD simulations of *Adgb1* (a) and *Adgb2* (b) are shown with respect to the initial minimised structure for all the three MD simulations. The results suggest *Adgb1* as the most stable model. RMSF plots from three MD simulations of *Adgb1* (c) and *Adgb2* (d) were measured using the Ca of each residue. The results show that the difference between the three runs is mainly due to the movement of the helix

corresponding to the IQ domain which rotates during the simulation trying to reach a more stable conformation (g). Table B.2 shows the occupancy of the hydrogen bond between residues of the IQ domain and the whole protein. The data indicate that Arg138<sup>(918)</sup> stabilises the structure interacting with the near Glutamic acid residues: Glu94<sup>(874)</sup>, Glu101<sup>(881)</sup>, and Glu102<sup>(882)</sup>. The RMSD plot for the heme and the key residues Gln12<sup>(792)</sup> and Tyr196<sup>(976)</sup> of *Adgb1* (e), and their distance from the heme iron (f), were calculated for all the three MD simulations to investigate the heme pocket. Tyr196<sup>(976)</sup> and the heme seem to have the same RMSD trend, suggesting a close relationship between them (in agreement with the results about distances).

**Table B.2 | Occupancy of the hydrogen bond between residues of the IQ domain and the whole protein.** In highlighting, the interaction between Arg138 of IQ domain and the near Glu94/101/102 that has high occupancy in all the 3 runs

Run1			Run2			Run3		
IQ domain	Protein	Occupancy	IQ domain	Protein	Occupancy	IQ domain	Protein	Occupancy
Arg149	Asp152	56.51%	Arg138	Glu101	108.54%	Arg138	Glu101	112.88%
Arg138	Glu101	51.88%	Lys120	Glu53	63.95%	Arg138	Glu94	105.80%
Arg138	Glu102	35.56%	Lys120	Asp51	50.07%	Arg68	Glu121	50.75%
Thr153	Asp92	34.12%	Arg149	Asp152	38.06%	Ser123	Glu126	39.96%
Lys112	Asp110	19.01%	Lys112	Glu102	29.10%	Tyr141	Asp92	38.61%
Ser123	Glu126	18.94%	Ser123	Glu126	25.88%	Lys112	Asp110	23.38%
Ser118	Glu53	18.06%	Gln134	Asn97	23.48%	Gln134	Glu94	20.79%
Lys154	Asp92	12.75%	Trp137	Glu94	20.80%	Lys189	Asp119	14.40%
Thr117	Glu53	10.59%	Lys154	Glu155	17.20%	Lys154	Asn81	10.61%

MD simulations of *Adgb1* and *Adgb2* suggested *Adgb1* as the most stable model because its structure gained a more stable conformation in the early part of the simulation and this agrees with our results from the RMSD analysis seen in Figure B.6a,b. Moreover, the position of IQ helix in *Adgb1* is more stable than the one in *Adgb2* as suggested by our results from RMSF analysis seen in Figure B.6c,d. The extension of the most stable runs of both models (Run2-*Adgb1* and Run3-*Adgb2*) has been done to confirm *Adgb1* as the best model and monitoring the final stable conformation of the IQ domain.



**Figure B.7 | RMSF analysis of the two models: *Adgb0* and *Adgb1*.** RMSF plots of *Adgb0* (in black) and *Adgb1* (in red) were measured using the Cα of each residue.

In conclusion, when the X-ray structure is not available, computational methods could give an important help to guide experimental work in the production of a mutant protein that could be successfully crystallised, as in this work where *Adgb* is a complex structure still to be discovered. Future works will include the analysis of other parts of this interesting multidomain protein.

## Appendix C. Supplementary material of Chapter 5

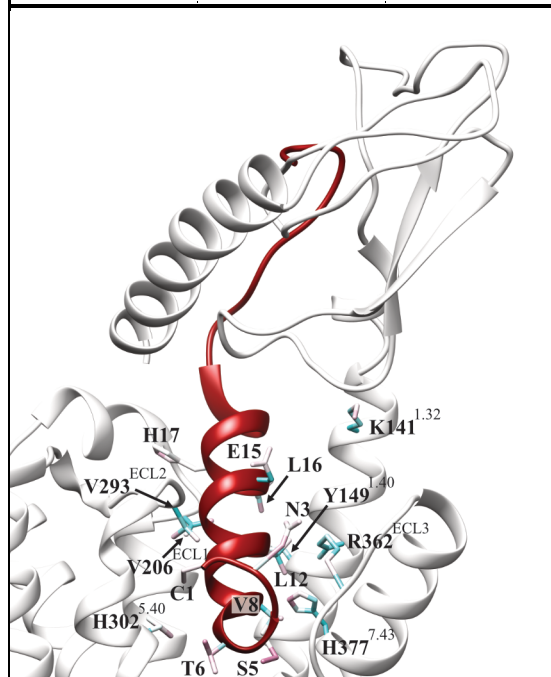
Table C1-14

**Table C.1 | Contacts between peptide and CTR during MD simulations of (#1) sCT:CTR:Ga system.** Contact persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)

(#1) sCT:CTR:Ga		
Peptide vs CTR		
sCT residue	CTR residue	Contact occupancy (%MD frames)
Thr6	His302 <sup>5.40</sup>	95%
Ser5	Phe356 <sup>6.53</sup>	91%
Glu15	Lys141 <sup>1.32</sup>	83%
Val8	His377 <sup>7.43</sup>	77%
Leu16	Tyr149 <sup>1.40</sup>	77%
Cys1	Val293 <sup>ECL2</sup>	77%
Ser13	Val206 <sup>ECL1</sup>	76%
His17	Val206 <sup>ECL1</sup>	75%
Asn3	Trp361 <sup>6.58</sup>	75%
Asn3	Arg362 <sup>ECL3</sup>	74%
Leu12	His377 <sup>7.43</sup>	73%
Arg24	Asp101 <sup>ECD</sup>	67%
Thr25	Asp101 <sup>ECD</sup>	66%
Leu12	Leu148 <sup>1.39</sup>	63%
Leu9	His381 <sup>7.47</sup>	63%
Lys18	Pro100 <sup>ECD</sup>	62%
Leu16	Val206 <sup>ECL1</sup>	62%
Leu19	Leu142 <sup>1.33</sup>	62%
Ser5	Pro360 <sup>6.57</sup>	60%
Gln20	Pro207 <sup>ECL1</sup>	54%
Gly10	Val293 <sup>ECL2</sup>	53%
His17	Val212 <sup>ECL1</sup>	52%
Gln14	Val293 <sup>ECL2</sup>	51%
Leu4	Pro360 <sup>6.57</sup>	49%
Cys1	Glu294 <sup>ECL2</sup>	49%
Thr27	Trp79 <sup>ECD</sup>	48%
Cys7	His302 <sup>5.40</sup>	45%



Ser5	Tyr372 <sup>7.38</sup>	44%
Cys1	Leu298 <sup>5.36</sup>	43%
Thr6	Met306 <sup>5.44</sup>	43%
Cys1	Tyr299 <sup>5.37</sup>	43%
Gln14	Leu291 <sup>ECL2</sup>	42%
Thr25	Phe99 <sup>ECD</sup>	42%
Asn26	Tyr41 <sup>ECD</sup>	40%
Val8	Ile380 <sup>7.46</sup>	40%
Lys18	Asp97 <sup>ECD</sup>	39%
Leu9	Ile198 <sup>2.64</sup>	37%
Ser5	Phe359 <sup>6.56</sup>	37%
Arg24	Pro100 <sup>ECD</sup>	36%
Leu19	Ala145 <sup>1.36</sup>	34%
Gln14	Ser292 <sup>ECL2</sup>	34%
Lys11	Val293 <sup>ECL2</sup>	34%
His17	Gly209 <sup>ECL1</sup>	33%
Leu16	Leu202 <sup>2.68</sup>	33%
Arg24	Tyr41 <sup>ECD</sup>	32%
Thr25	Tyr41 <sup>ECD</sup>	32%
Leu4	Met306 <sup>5.44</sup>	31%
Thr27	Asp101 <sup>ECD</sup>	30%
Gln20	Tyr146 <sup>1.37</sup>	30%
His17	Leu291 <sup>ECL2</sup>	30%

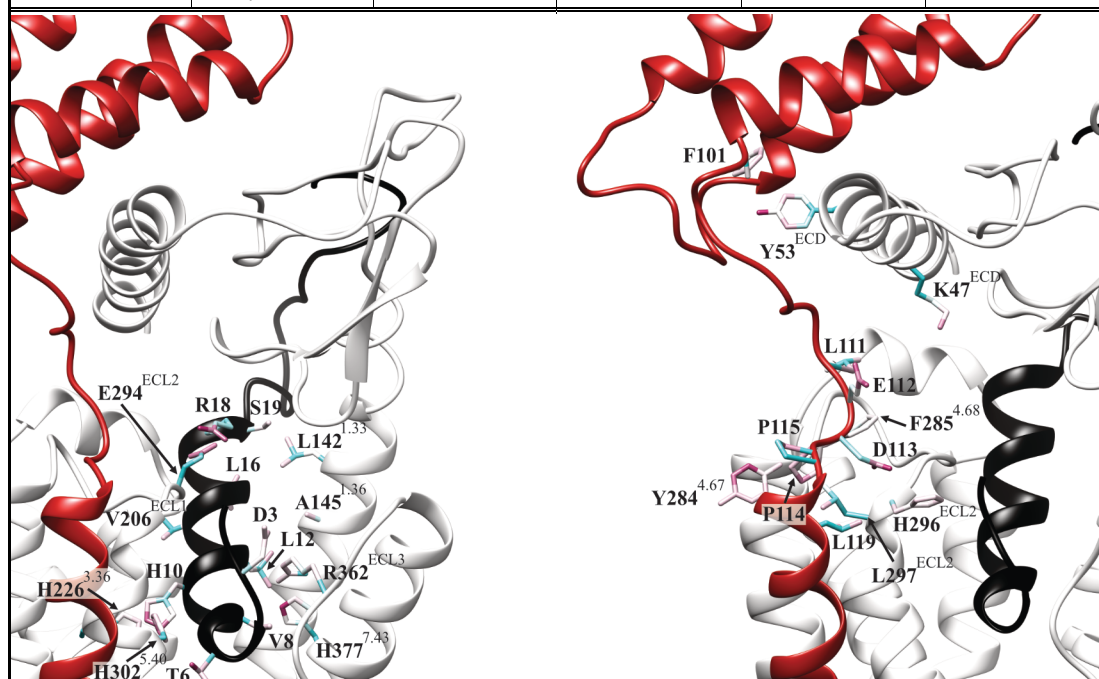


**Table C.2 | Contacts between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#2) hCGRP:RAMP3:CTR:Ga system.** Contact persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)

(#2) hCGRP:RAMP3:CTR:Ga					
Peptide vs CTR/RAMP			RAMP vs CTR		
hCGRP residue	CTR/RAMP residue	Contact occupancy (%MD frames)	RAMP residue	CTR residue	Contact occupancy (%MD frames)
Val8	His377 <sup>7.43</sup>	94%	Pro114	Tyr284 <sup>4.67</sup>	71%
Arg18	Glu294 <sup>ECL2</sup>	92%	Phe101	Tyr53 <sup>ECD</sup>	71%
Thr6	His302 <sup>5.40</sup>	88%	Glu112	Phe285 <sup>4.68</sup>	56%
Asp3	Arg362 <sup>ECL3</sup>	78%	Leu111	Tyr284 <sup>4.67</sup>	51%
Leu16	Val206 <sup>ECL1</sup>	78%	Leu119	His296 <sup>ECL2</sup>	47%
Leu12	His377 <sup>7.43</sup>	77%	Pro115	Tyr284 <sup>4.67</sup>	46%
His10	His226 <sup>3.36</sup>	74%	Glu112	Tyr284 <sup>4.67</sup>	44%
Leu12	Ala145 <sup>1.36</sup>	69%	Leu111	Phe285 <sup>4.68</sup>	43%
His10	Gln227 <sup>3.37</sup>	69%	Asp113	His296 <sup>ECL2</sup>	43%
Ser19	Leu142 <sup>1.33</sup>	69%	Pro114	Leu297 <sup>ECL2</sup>	41%
His10	Val293 <sup>ECL2</sup>	67%	Asp113	Tyr284 <sup>4.67</sup>	31%
Ala5	Phe356 <sup>6.53</sup>	66%	Glu112	Lys47 <sup>ECD</sup>	31%
Leu16	Tyr149 <sup>1.40</sup>	64%			
Arg18	Pro100 <sup>ECD</sup>	58%			
His10	His223 <sup>3.33</sup>	58%			
Thr9	His381 <sup>7.47</sup>	57%			
Leu12	Leu148 <sup>1.39</sup>	56%			
Ala13	Val206 <sup>ECL1</sup>	54%			
Ala5	Met376 <sup>7.42</sup>	54%			
Thr4	Pro360 <sup>6.57</sup>	53%			
Cys2	Val293 <sup>ECL2</sup>	52%			
Phe37	Ser129 <sup>ECD</sup>	52%			
Cys7	His302 <sup>5.40</sup>	50%			
Phe37	Trp128 <sup>ECD</sup>	47%			
Cys2	Leu298 <sup>5.36</sup>	45%			
Val8	Ile380 <sup>7.46</sup>	45%			
Thr6	Tyr234 <sup>3.44</sup>	45%			
Leu16	Ala145 <sup>1.36</sup>	44%			
Cys2	His302 <sup>5.40</sup>	43%			
Arg18	Leu40 <sup>ECD</sup>	42%			
Phe37	Trp79 <sup>ECD</sup>	42%			
Asp3	Trp361 <sup>6.58</sup>	42%			

# APPENDIX

Thr4	Met306 <sup>5.44</sup>	41%			
Arg11	Asp373 <sup>7.39</sup>	40%			
Phe37	Tyr131 <sup>ECD</sup>	37%			
Thr4	His302 <sup>5.40</sup>	37%			
Thr30	Asp101 <sup>ECD</sup>	36%			
His10	Leu291 <sup>ECL2</sup>	35%			
Leu15	Ala145 <sup>1.36</sup>	34%			
Ala5	Tyr372 <sup>7.38</sup>	34%			
Ala5	Ile380 <sup>7.46</sup>	32%			
Ser17	Leu291 <sup>ECL2</sup>	31%			
Arg18	Ser292 <sup>ECL2</sup>	31%			
Val23	Tyr146 <sup>1.37</sup>	31%			

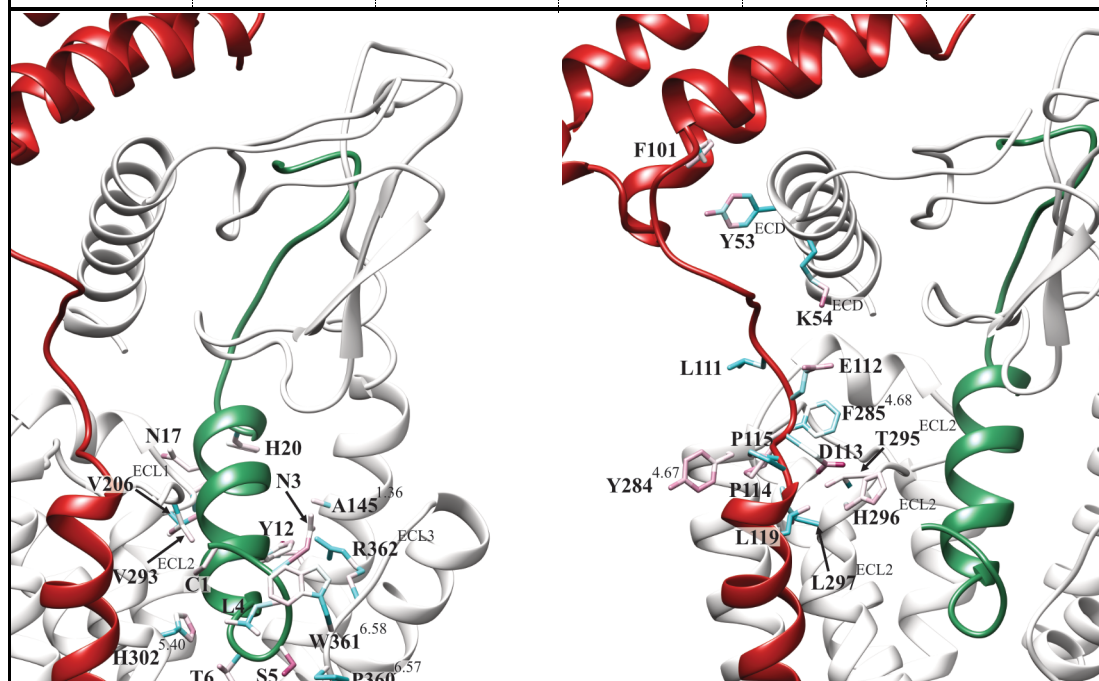


**Table C.3 | Contacts between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#3) hCT:RAMP3:CTR:Ga system.** Contact persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)

(#3) hCT:RAMP3:CTR:Ga					
Peptide vs CTR/RAMP			RAMP vs CTR		
hCT residue	CTR/RAMP residue	Contact occupancy (%MD frames)	RAMP residue	CTR residue	Contact occupancy (%MD frames)
Tyr12	Tyr149 <sup>1.40</sup>	97%	Phe101	Tyr53 <sup>ECD</sup>	78%
Asn17	Val206 <sup>ECL1</sup>	97%	Pro114	Tyr284 <sup>4.67</sup>	73%
Ser5	Phe356 <sup>6.53</sup>	95%	Leu119	His296 <sup>ECL2</sup>	62%
Thr6	His302 <sup>5.40</sup>	93%	Glu112	Phe285 <sup>4.68</sup>	55%
Asn3	Trp361 <sup>6.58</sup>	91%	Asp113	His296 <sup>ECL2</sup>	55%
His20	Pro207 <sup>ECL1</sup>	89%	Pro114	Leu297 <sup>ECL2</sup>	50%
Ser5	Pro360 <sup>6.57</sup>	88%	Leu111	Tyr284 <sup>4.67</sup>	48%
Asn3	Arg362 <sup>ECL3</sup>	88%	Pro115	Tyr284 <sup>4.67</sup>	43%
Tyr12	Ala145 <sup>1.36</sup>	86%	Glu112	Tyr284 <sup>4.67</sup>	39%
Cys1	Val293 <sup>ECL2</sup>	85%	Glu112	Lys54 <sup>ECD</sup>	36%
Leu4	Pro360 <sup>6.57</sup>	85%	Asp113	Thr295 <sup>ECL2</sup>	31%
Tyr12	His377 <sup>7.43</sup>	83%			
Phe19	Leu142 <sup>1.33</sup>	79%			
Phe16	Tyr149 <sup>1.40</sup>	78%			
Ser5	Phe359 <sup>6.56</sup>	75%			
Thr13	Val206 <sup>ECL1</sup>	75%			
Ser5	Tyr372 <sup>7.38</sup>	74%			
Tyr12	Leu148 <sup>1.39</sup>	72%			
Gly10	Val293 <sup>ECL2</sup>	69%			
Met8	Met376 <sup>7.42</sup>	63%			
Gln14	Leu291 <sup>ECL2</sup>	62%			
Phe16	Val206 <sup>ECL1</sup>	62%			
Cys1	Glu294 <sup>ECL2</sup>	61%			
Met8	Tyr372 <sup>7.38</sup>	61%			
Gln14	Ser292 <sup>ECL2</sup>	61%			
His20	Tyr146 <sup>1.37</sup>	56%			
Leu4	Trp361 <sup>6.58</sup>	56%			
Asn17	Asn208 <sup>ECL1</sup>	54%			
His20	Gly209 <sup>ECL1</sup>	51%			
Phe16	Ala145 <sup>1.36</sup>	50%			
Asp15	Lys141 <sup>1.32</sup>	48%			
Phe19	Thr138 <sup>1.29</sup>	47%			

# APPENDIX

Cys1	Leu298 <sup>5.36</sup>	46%			
Met8	His377 <sup>7.43</sup>	46%			
Thr21	Gly209 <sup>ECL1</sup>	45%			
Lys18	Pro100 <sup>ECD</sup>	45%			
Leu9	His381 <sup>7.47</sup>	44%			
Phe16	Tyr146 <sup>1.37</sup>	44%			
Met8	Asp373 <sup>7.39</sup>	44%			
Thr11	Val293 <sup>ECL2</sup>	44%			
Gln24	Tyr41 <sup>ECD</sup>	43%			
Asn17	Leu291 <sup>ECL2</sup>	42%			
Thr6	Met306 <sup>5.44</sup>	38%			
Cys7	His302 <sup>5.40</sup>	38%			
Phe19	Lys141 <sup>1.32</sup>	37%			
Cys7	Val293 <sup>ECL2</sup>	35%			
Gln14	Val293 <sup>ECL2</sup>	34%			
Phe22	Gly209 <sup>ECL1</sup>	34%			
Leu9	His226 <sup>3.36</sup>	33%			
Ser5	Met376 <sup>7.42</sup>	33%			
Cys1	Tyr299 <sup>5.37</sup>	31%			
Asn3	Tyr372 <sup>7.38</sup>	30%			
Lys18	Asp97 <sup>ECD</sup>	30%			

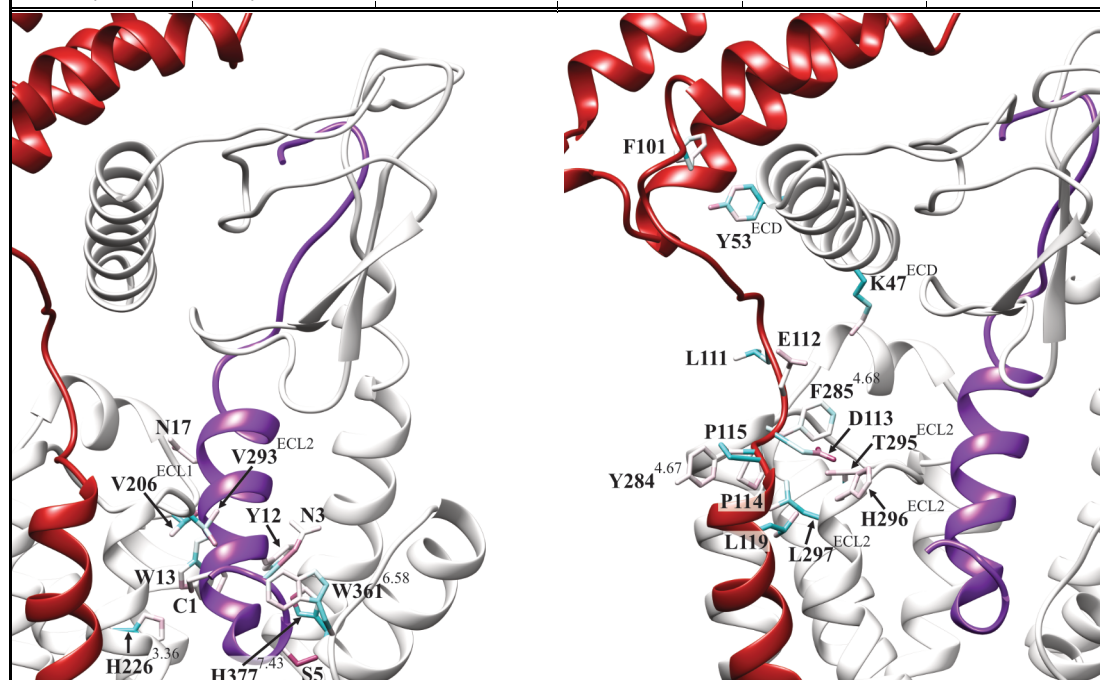


**Table C.4 | Contacts between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#4) pCT:RAMP3:CTR:Ga system.** Contact persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)

(#4) pCT:RAMP3:CTR:Ga					
Peptide vs CTR/RAMP			RAMP vs CTR		
pCT residue	CTR/RAMP residue	Contact occupancy (%MD frames)	RAMP residue	CTR residue	Contact occupancy (%MD frames)
Trp13	His226 <sup>3.36</sup>	99%	Phe101	Tyr53 <sup>ECD</sup>	75%
Ser5	Phe356 <sup>6.53</sup>	96%	Glu112	Phe285 <sup>4.68</sup>	74%
Cys1	Val293 <sup>ECL2</sup>	96%	Pro114	Tyr284 <sup>4.67</sup>	71%
Asn17	Val206 <sup>ECL1</sup>	96%	Asp113	His296 <sup>ECL2</sup>	65%
Asn3	Trp361 <sup>6.58</sup>	95%	Leu119	His296 <sup>ECL2</sup>	60%
Trp13	Val206 <sup>ECL1</sup>	95%	Asp113	Thr295 <sup>ECL2</sup>	50%
Tyr12	His377 <sup>7.43</sup>	95%	Glu112	Lys47 <sup>ECD</sup>	42%
Ser5	Pro360 <sup>6.57</sup>	93%	Pro114	Leu297 <sup>ECL2</sup>	41%
Asn3	Arg362 <sup>ECL3</sup>	93%	Pro115	Tyr284 <sup>4.67</sup>	37%
His20	Pro207 <sup>ECL1</sup>	92%	Leu111	Tyr284 <sup>4.67</sup>	36%
Tyr12	Tyr149 <sup>1.40</sup>	92%	Leu111	Phe285 <sup>4.68</sup>	35%
Thr6	His302 <sup>5.40</sup>	84%	Leu119	Leu297 <sup>ECL2</sup>	30%
Leu4	Pro360 <sup>6.57</sup>	83%			
Ser5	Tyr372 <sup>7.38</sup>	81%			
Cys7	His302 <sup>5.40</sup>	79%			
Phe19	Leu142 <sup>1.33</sup>	79%			
Tyr12	Ala145 <sup>1.36</sup>	76%			
Arg14	Leu291 <sup>ECL2</sup>	75%			
Ser10	Val293 <sup>ECL2</sup>	74%			
Arg14	Glu294 <sup>ECL2</sup>	73%			
Tyr12	Leu148 <sup>1.39</sup>	71%			
Asn17	Val212 <sup>ECL1</sup>	70%			
Trp13	Leu202 <sup>2.68</sup>	67%			
Trp13	His201 <sup>2.67</sup>	65%			
Leu16	Val206 <sup>ECL1</sup>	64%			
Ser5	Phe359 <sup>6.56</sup>	63%			
Arg21	Gly209 <sup>ECL1</sup>	60%			
Val8	His377 <sup>7.43</sup>	53%			
Arg14	Ser292 <sup>ECL2</sup>	51%			
Asn17	Leu291 <sup>ECL2</sup>	51%			
Leu9	His226 <sup>3.36</sup>	50%			
Ala11	Val293 <sup>ECL2</sup>	49%			

# APPENDIX

Thr6	Met306 <sup>5.44</sup>	47%			
Arg21	Leu40 <sup>ECD</sup>	47%			
Leu4	Trp361 <sup>6.58</sup>	47%			
Cys1	Leu298 <sup>5.36</sup>	46%			
Leu9	Met230 <sup>3.40</sup>	46%			
Cys1	Glu294 <sup>ECL2</sup>	45%			
Phe22	Pro100 <sup>ECD</sup>	44%			
Phe22	Leu40 <sup>ECD</sup>	44%			
Asn17	Asn208 <sup>ECL1</sup>	40%			
His20	Tyr146 <sup>1.37</sup>	40%			
Leu9	His381 <sup>7.47</sup>	39%			
Arg21	Val212 <sup>ECL1</sup>	38%			
Trp13	Ile198 <sup>2.64</sup>	36%			
Leu16	Tyr149 <sup>1.40</sup>	33%			
Phe19	Lys141 <sup>1.32</sup>	32%			
Ser23	Asp101 <sup>ECD</sup>	31%			
Ser5	Met376 <sup>7.42</sup>	31%			
Met25	Asp101 <sup>ECD</sup>	31%			
Arg21	Arg213 <sup>ECL1</sup>	31%			
Cys1	Tyr299 <sup>5.37</sup>	30%			



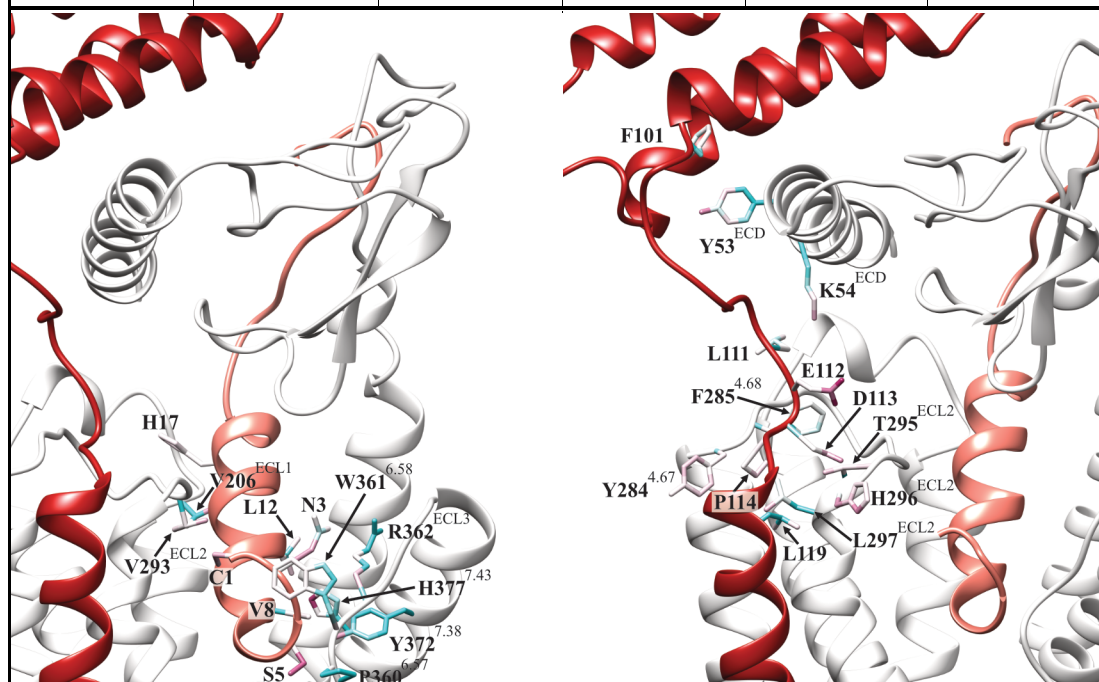
**Table C.5 | Contacts between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#5) sCT:RAMP3:CTR:Ga system.** Contact persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)

(#5) sCT:RAMP3:CTR:Ga					
Peptide vs CTR/RAMP			RAMP vs CTR		
sCT residue	CTR/RAMP residue	Contact occupancy (%MD frames)	RAMP residue	CTR residue	Contact occupancy (%MD frames)
Asn3	Arg362 <sup>ECL3</sup>	98%	Asp113	His296 <sup>ECL2</sup>	67%
Val8	His377 <sup>7.43</sup>	97%	Phe101	Tyr53 <sup>ECD</sup>	62%
Ser5	Phe356 <sup>6.53</sup>	97%	Leu111	Tyr284 <sup>4.67</sup>	61%
Asn3	Trp361 <sup>6.58</sup>	95%	Glu112	Phe285 <sup>4.68</sup>	56%
Ser5	Tyr372 <sup>7.38</sup>	92%	Pro114	Tyr284 <sup>4.67</sup>	50%
Cys1	Val293 <sup>ECL2</sup>	89%	Glu112	Tyr284 <sup>4.67</sup>	40%
Ser5	Pro360 <sup>6.57</sup>	86%	Pro114	Leu297 <sup>ECL2</sup>	36%
Leu12	His377 <sup>7.43</sup>	85%	Leu119	His296 <sup>ECL2</sup>	35%
His17	Val206 <sup>ECL1</sup>	85%	Glu112	His296 <sup>ECL2</sup>	33%
Leu16	Tyr149 <sup>1.40</sup>	81%	Glu112	Thr295 <sup>ECL2</sup>	32%
Arg24	Asp101 <sup>ECD</sup>	75%	Glu112	Lys54 <sup>ECD</sup>	31%
Ser13	Val206 <sup>ECL1</sup>	75%			
Leu4	Pro360 <sup>6.57</sup>	68%			
Thr6	His302 <sup>5.40</sup>	66%			
Leu12	Leu148 <sup>1.39</sup>	63%			
Leu16	Val206 <sup>ECL1</sup>	62%			
Leu19	Leu142 <sup>1.33</sup>	62%			
Gln14	Val293 <sup>ECL2</sup>	61%			
Gln20	Pro207 <sup>ECL1</sup>	59%			
Lys18	Pro100 <sup>ECD</sup>	58%			
Ser5	Phe359 <sup>6.56</sup>	56%			
Cys1	Glu294 <sup>ECL2</sup>	55%			
Val8	Ile380 <sup>7.46</sup>	54%			
Ser2	Glu294 <sup>ECL2</sup>	52%			
Thr6	Met306 <sup>5.44</sup>	50%			
Cys7	His302 <sup>5.40</sup>	48%			
Leu9	His226 <sup>3.36</sup>	48%			
Leu9	Ile198 <sup>2.64</sup>	46%			
His17	Leu40 <sup>ECD</sup>	46%			
His17	Leu291 <sup>ECL2</sup>	45%			
Leu4	His302 <sup>5.40</sup>	44%			
Thr25	Phe99 <sup>ECD</sup>	44%			



# APPENDIX

Gln14	Leu291 <sup>ECL2</sup>	44%			
Lys18	Asp97 <sup>ECD</sup>	43%			
Gly10	Val293 <sup>ECL2</sup>	43%			
Pro23	Tyr41 <sup>ECD</sup>	41%			
Glu15	Lys141 <sup>1.32</sup>	39%			
Leu16	Leu202 <sup>2.68</sup>	37%			
Asn26	Trp79 <sup>ECD</sup>	36%			
Gln20	Tyr146 <sup>1.37</sup>	36%			
Gln14	Ser292 <sup>ECL2</sup>	35%			
Val8	Met376 <sup>7.42</sup>	35%			
Thr27	Trp79 <sup>ECD</sup>	33%			
Thr6	Phe356 <sup>6.53</sup>	32%			
Leu19	Tyr146 <sup>1.37</sup>	31%			
Pro32	Trp128 <sup>ECD</sup>	31%			
Pro32	Ser129 <sup>ECD</sup>	30%			
Gly28	Trp79 <sup>ECD</sup>	30%			

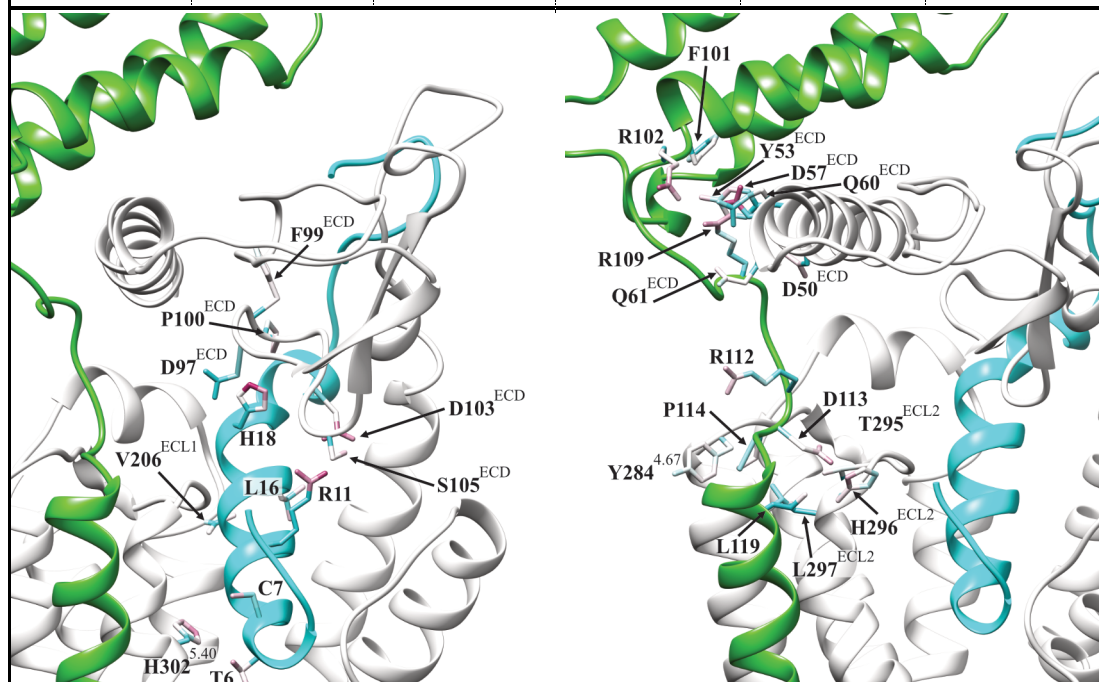


**Table C.6 | Contacts between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#6) rAmy:RAMP1:CTR:Ga system. Contact persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)**

<b>(#6) rAmy:RAMP1:CTR:Ga</b>					
<b>Peptide vs CTR/RAMP</b>			<b>RAMP vs CTR</b>		
<b>rAmy residue</b>	<b>CTR/RAMP residue</b>	<b>Contact occupancy (%MD frames)</b>	<b>RAMP residue</b>	<b>CTR residue</b>	<b>Contact occupancy (%MD frames)</b>
Arg11	Asp103 <sup>ECD</sup>	100%	Arg102	Asp57 <sup>ECD</sup>	84%
His18	Asp97 <sup>ECD</sup>	88%	Phe101	Tyr53 <sup>ECD</sup>	76%
His18	Pro100 <sup>ECD</sup>	83%	Arg109	Asp57 <sup>ECD</sup>	69%
Thr6	His302 <sup>5.40</sup>	82%	Arg102	Gln61 <sup>ECD</sup>	64%
His18	Phe99 <sup>ECD</sup>	80%	Pro114	Tyr284 <sup>4.67</sup>	56%
Leu16	Val206 <sup>ECL1</sup>	75%	Leu119	His296 <sup>ECL2</sup>	52%
Arg11	Ser105 <sup>ECD</sup>	75%	Asp113	His296 <sup>ECL2</sup>	44%
Cys7	His302 <sup>5.40</sup>	75%	Arg112	Asp50 <sup>ECD</sup>	44%
Gln10	His223 <sup>3.33</sup>	69%	Arg109	Tyr53 <sup>ECD</sup>	43%
Phe15	Lys141 <sup>1.32</sup>	69%	Arg102	Gln60 <sup>ECD</sup>	39%
Ala8	His377 <sup>7.43</sup>	68%	Arg112	Tyr284 <sup>4.67</sup>	36%
Ser19	Leu142 <sup>1.33</sup>	67%	Asp113	Thr295 <sup>ECL2</sup>	35%
Leu12	Ala145 <sup>1.36</sup>	65%	Leu119	Leu297 <sup>ECL2</sup>	31%
His18	Phe102 <sup>ECD</sup>	64%			
Phe15	Leu142 <sup>1.33</sup>	64%			
Asn14	Leu291 <sup>ECL2</sup>	63%			
Gln10	Val293 <sup>ECL2</sup>	59%			
Val17	Val205 <sup>2.71</sup>	54%			
Leu12	Leu148 <sup>1.39</sup>	53%			
Phe15	Asp103 <sup>ECD</sup>	52%			
Asn3	Phe356 <sup>6.53</sup>	52%			
Ala5	Phe356 <sup>6.53</sup>	49%			
Ala5	Tyr234 <sup>3.44</sup>	49%			
Leu16	Ala145 <sup>1.36</sup>	49%			
Ala5	Ile380 <sup>7.46</sup>	47%			
Ser19	Pro100 <sup>ECD</sup>	45%			
Leu16	Tyr149 <sup>1.40</sup>	45%			
Gln10	Leu291 <sup>ECL2</sup>	45%			
Val17	Val206 <sup>ECL1</sup>	44%			
Asn3	Pro360 <sup>6.57</sup>	44%			
Leu12	His377 <sup>7.43</sup>	40%			
Phe15	Ala145 <sup>1.36</sup>	39%			

# APPENDIX

Asn21	Leu40 <sup>ECD</sup>	39%			
Thr6	Met230 <sup>3.40</sup>	38%			
Asn21	Pro100 <sup>ECD</sup>	38%			
Lys1	Glu294 <sup>ECL2</sup>	37%			
Gln10	Gln227 <sup>3.37</sup>	37%			
Thr4	His302 <sup>5.40</sup>	36%			
Leu16	Tyr146 <sup>1.37</sup>	35%			
Cys7	Val293 <sup>ECL2</sup>	34%			
Val17	Leu211 <sup>ECL1</sup>	33%			
Thr4	Met306 <sup>5.44</sup>	33%			
Ala13	Leu291 <sup>ECL2</sup>	32%			
Asn22	Pro100 <sup>ECD</sup>	32%			
Asn35	Phe83 <sup>RAMP1</sup>	32%			
Asn22	Asp101 <sup>ECD</sup>	31%			
Thr4	Phe356 <sup>6.53</sup>	31%			
Leu16	Leu142 <sup>1.33</sup>	31%			
Ile26	Tyr41 <sup>ECD</sup>	31%			
Ser20	Gly209 <sup>ECL1</sup>	30%			
Tyr37	Cys82 <sup>RAMP1</sup>	30%			

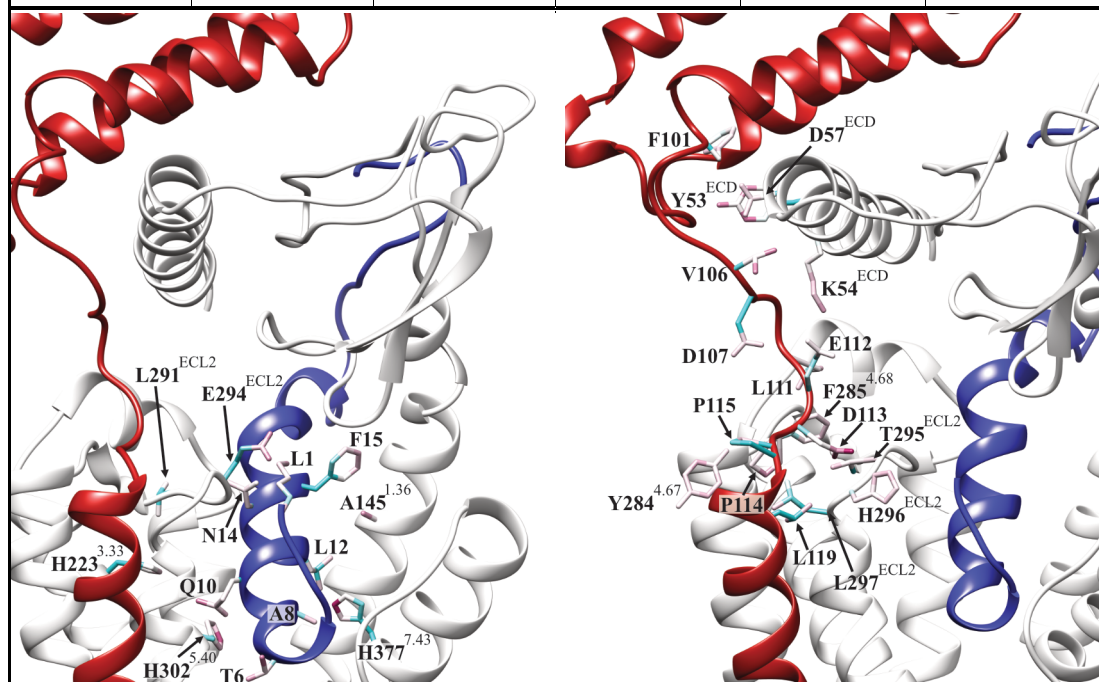


**Table C.7 | Contacts between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#7) rAmy:RAMP3:CTR:Ga system. Contact persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)**

(#7) rAmy:RAMP3:CTR:Ga					
Peptide vs CTR/RAMP			RAMP vs CTR		
rAmy residue	CTR/RAMP residue	Contact occupancy (%MD frames)	RAMP residue	CTR residue	Contact occupancy (%MD frames)
Ala8	His377 <sup>7.43</sup>	97%	Phe101	Tyr53 <sup>ECD</sup>	80%
Lys1	Glu294 <sup>ECL2</sup>	83%	Pro114	Tyr284 <sup>4.67</sup>	73%
Asn14	Leu291 <sup>ECL2</sup>	83%	Glu112	Phe285 <sup>4.68</sup>	69%
Phe15	Ala145 <sup>1.36</sup>	76%	Leu111	Phe285 <sup>4.68</sup>	61%
Thr6	His302 <sup>5.40</sup>	73%	Val106	Tyr53 <sup>ECD</sup>	59%
Gln10	His223 <sup>3.33</sup>	72%	Leu119	His296 <sup>ECL2</sup>	57%
Leu12	His377 <sup>7.43</sup>	70%	Asp113	His296 <sup>ECL2</sup>	52%
Thr6	Tyr234 <sup>3.44</sup>	67%	Val106	Lys54 <sup>ECD</sup>	46%
Ala5	Met376 <sup>7.42</sup>	66%	Pro114	Leu297 <sup>ECL2</sup>	40%
Phe15	Lys141 <sup>1.32</sup>	65%	Glu112	Tyr284 <sup>4.67</sup>	40%
Leu16	Val206 <sup>ECL1</sup>	65%	Asp107	Lys54 <sup>ECD</sup>	39%
Cys7	His302 <sup>5.40</sup>	63%	Pro115	Tyr284 <sup>4.67</sup>	39%
Leu16	Tyr149 <sup>1.40</sup>	63%	Leu119	Leu297 <sup>ECL2</sup>	37%
Ala5	Phe356 <sup>6.53</sup>	63%	Val106	Asp57 <sup>ECD</sup>	37%
Gln10	Val293 <sup>ECL2</sup>	60%	Asp113	Thr295 <sup>ECL2</sup>	35%
Leu12	Leu148 <sup>1.39</sup>	60%			
Cys2	Val293 <sup>ECL2</sup>	57%			
Asn3	Arg362 <sup>ECL3</sup>	57%			
Thr9	His226 <sup>3.36</sup>	55%			
Asn22	Pro100 <sup>ECD</sup>	52%			
Asn14	Val293 <sup>ECL2</sup>	51%			
Lys1	Val293 <sup>ECL2</sup>	51%			
Arg11	Asp373 <sup>7.39</sup>	51%			
Gln10	Gln227 <sup>3.37</sup>	50%			
Thr4	Met306 <sup>5.44</sup>	48%			
Cys2	Leu298 <sup>5.36</sup>	48%			
Asn22	Leu142 <sup>1.33</sup>	47%			
Phe23	Leu142 <sup>1.33</sup>	46%			
Leu16	Ala145 <sup>1.36</sup>	46%			
Gln10	His226 <sup>3.36</sup>	43%			
Ala5	Tyr372 <sup>7.38</sup>	43%			
Ser19	Leu142 <sup>1.33</sup>	42%			

# APPENDIX

Phe23	Tyr146 <sup>1.37</sup>	42%			
Thr4	His302 <sup>5.40</sup>	41%			
Ser29	Asp101 <sup>ECD</sup>	41%			
Thr30	Asp101 <sup>ECD</sup>	40%			
Val17	Leu291 <sup>ECL2</sup>	39%			
Arg11	Val293 <sup>ECL2</sup>	39%			
Ala13	Val206 <sup>ECL1</sup>	38%			
Ser20	Val206 <sup>ECL1</sup>	37%			
Asn3	Trp361 <sup>6.58</sup>	37%			
Val17	Val206 <sup>ECL1</sup>	37%			
Thr9	Ile380 <sup>7.46</sup>	36%			
His18	Pro104 <sup>ECD</sup>	36%			
Cys7	Val293 <sup>ECL2</sup>	32%			
Ala5	Ile380 <sup>7.46</sup>	31%			
Thr6	Val305 <sup>5.43</sup>	30%			
Ala13	His201 <sup>2.67</sup>	30%			



**Table C.8 | Hydrogen bonds between peptide and CTR during MD simulations of (#1) sCT:CTR:Ga system.** Hydrogen bond persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)

(#1) sCT:CTR:Ga		
Peptide vs CTR		
sCT residue	CTR residue	Hydrogen Bond occupancy (%MD frames)
Thr6	His302 <sup>5.40</sup>	90%
Glu15	Lys141 <sup>1.32</sup>	81%
Arg24	Asp101 <sup>ECD</sup>	61%
Thr25	Asp101 <sup>ECD</sup>	49%
Lys18	Asp97 <sup>ECD</sup>	33%
Thr27	Asp101 <sup>ECD</sup>	30%
Lys11	Asp103 <sup>ECD</sup>	23%
Gln20	Tyr146 <sup>1.37</sup>	21%
Lys11	Glu294 <sup>ECL2</sup>	19%
Thr31	Asp77 <sup>ECD</sup>	17%
Lys11	Asp373 <sup>7.39</sup>	13%
Gln14	Glu294 <sup>ECL2</sup>	12%
Lys18	Asp101 <sup>ECD</sup>	12%

**Table C.9 | Hydrogen bonds between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#2) hCGRP:RAMP3:CTR:Gα system.** Hydrogen bond persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)

(#2) hCGRP:RAMP3:CTR:Gα					
Peptide vs CTR/RAMP			RAMP vs CTR		
hCGRP residue	CTR/RAMP residue	Hydrogen Bond occupancy (%MD frames)	RAMP residue	CTR residue	Hydrogen Bond occupancy (%MD frames)
Arg18	Glu294 <sup>ECL2</sup>	92%	Glu112	Lys47 <sup>ECD</sup>	30%
Thr6	His302 <sup>5.40</sup>	77%	Arg108	Asp50 <sup>ECD</sup>	25%
His10	Gln227 <sup>3.37</sup>	46%	Glu112	Lys54 <sup>ECD</sup>	24%
Arg11	Asp373 <sup>7.39</sup>	39%	Asp113	Thr295 <sup>ECL2</sup>	23%
Asp3	Arg362 <sup>ECL3</sup>	32%	Asp113	His296 <sup>ECL2</sup>	20%
Thr30	Asp101 <sup>ECD</sup>	32%	Glu112	Lys46 <sup>ECD</sup>	10%
Arg11	Asp103 <sup>ECD</sup>	28%			
Lys35	Asp101 <sup>ECD</sup>	23%			
Thr9	His226 <sup>3.36</sup>	17%			
His10	His226 <sup>3.36</sup>	12%			



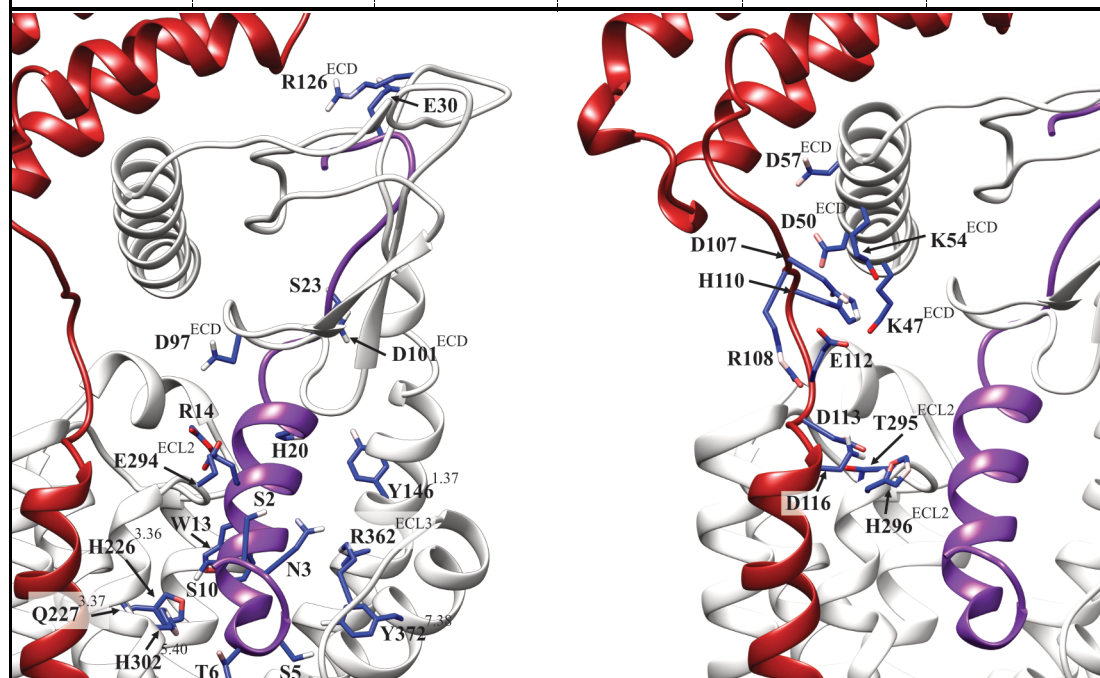
**Table C.10 | Hydrogen bonds between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#3) hCT:RAMP3:CTR:Ga system. Hydrogen bond persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)**

(#3) hCT:RAMP3:CTR:Ga					
Peptide vs CTR/RAMP			RAMP vs CTR		
hCT residue	CTR/RAMP residue	Hydrogen Bond occupancy (%MD frames)	RAMP residue	CTR residue	Hydrogen Bond occupancy (%MD frames)
Thr6	His302 <sup>5.40</sup>	81%	Glu112	Lys54 <sup>ECD</sup>	35%
Asp15	Lys141 <sup>1.32</sup>	46%	Asp113	Thr295 <sup>ECL2</sup>	30%
His20	Tyr146 <sup>1.37</sup>	46%	Asp113	His296 <sup>ECL2</sup>	28%
Lys18	Asp97 <sup>ECD</sup>	15%	Glu112	Lys46 <sup>ECD</sup>	17%
			Glu112	Lys47 <sup>ECD</sup>	16%
			Asp107	Lys54 <sup>ECD</sup>	11%



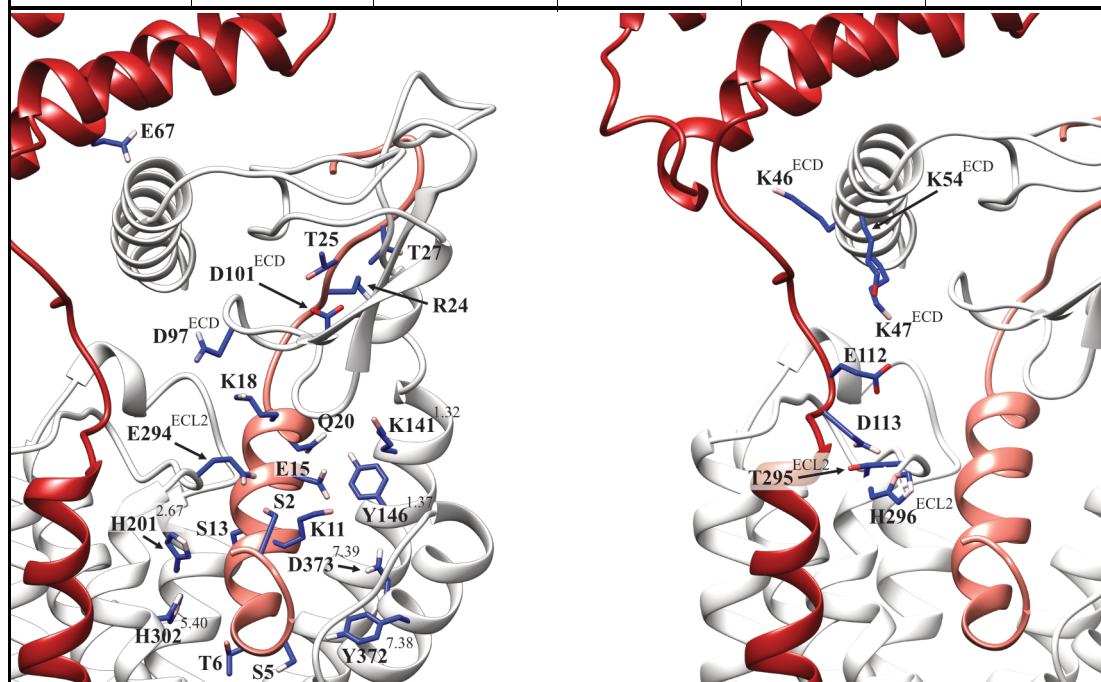
**Table C.11 | Hydrogen bonds between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#4) pCT:RAMP3:CTR:Ga system. Hydrogen bond persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)**

(#4) pCT:RAMP3:CTR:Ga					
Peptide vs CTR/RAMP			RAMP vs CTR		
pCT residue	CTR/RAMP residue	Hydrogen Bond occupancy (%MD frames)	RAMP residue	CTR residue	Hydrogen Bond occupancy (%MD frames)
Trp13	His226 <sup>3.36</sup>	97%	Asp113	Thr295 <sup>ECL2</sup>	48%
Arg14	Glu294 <sup>ECL2</sup>	73%	Glu112	Lys47 <sup>ECD</sup>	41%
Thr6	His302 <sup>5.40</sup>	50%	Asp113	His296 <sup>ECL2</sup>	39%
His20	Tyr146 <sup>1.37</sup>	31%	Glu112	Lys54 <sup>ECD</sup>	25%
Ser2	Glu294 <sup>ECL2</sup>	20%	Arg108	Asp50 <sup>ECD</sup>	18%
Ser23	Asp101 <sup>ECD</sup>	18%	Asp107	Lys54 <sup>ECD</sup>	15%
Glu30	Arg126 <sup>ECD</sup>	17%	Asp116	His296 <sup>ECL2</sup>	12%
Arg14	Asp97 <sup>ECD</sup>	17%	Arg108	Asp57 <sup>ECD</sup>	12%
Ser5	Tyr372 <sup>7.38</sup>	13%	His110	Asp50 <sup>ECD</sup>	11%
Asn3	Arg362 <sup>ECL3</sup>	12%			
Ser10	Gln227 <sup>3.37</sup>	12%			



**Table C.12 | Hydrogen bonds between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#5) sCT:RAMP3:CTR:Ga system. Hydrogen bond persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)**

(#5) sCT:RAMP3:CTR:Ga					
Peptide vs CTR/RAMP			RAMP vs CTR		
sCT residue	CTR/RAMP residue	Hydrogen Bond occupancy (%MD frames)	RAMP residue	CTR residue	Hydrogen Bond occupancy (%MD frames)
Arg24	Asp101 <sup>ECD</sup>	54%	Glu112	Thr295 <sup>ECL2</sup>	32%
Thr6	His302 <sup>5.40</sup>	51%	Glu112	Lys54 <sup>ECD</sup>	31%
Ser2	Glu294 <sup>ECL2</sup>	44%	Asp113	His296 <sup>ECL2</sup>	20%
Glu15	Lys141 <sup>1.32</sup>	37%	Glu112	His296 <sup>ECL2</sup>	16%
Gln20	Tyr146 <sup>1.37</sup>	28%	Asp113	Thr295 <sup>ECL2</sup>	15%
Thr25	Asp101 <sup>ECD</sup>	25%	Glu112	Lys47 <sup>ECD</sup>	11%
Ser13	His201 <sup>2.67</sup>	21%	Glu112	Lys46 <sup>ECD</sup>	11%
Thr25	Glu67 <sup>RAMP3</sup>	20%			
Lys11	Asp373 <sup>7.39</sup>	18%			
Thr27	Glu67 <sup>RAMP3</sup>	17%			
Ser5	Tyr372 <sup>7.38</sup>	17%			
Lys18	Asp97 <sup>ECD</sup>	11%			



**Table C.13 | Hydrogen bonds between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#6) rAmy:RAMP1:CTR:G $\alpha$  system.** Hydrogen bond persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)

(#6) rAmy:RAMP1:CTR:G $\alpha$					
Peptide vs CTR/RAMP			RAMP vs CTR		
rAmy residue	CTR/RAMP residue	Hydrogen Bond occupancy (%MD frames)	RAMP residue	CTR residue	Hydrogen Bond occupancy (%MD frames)
Arg11	Asp103 <sup>ECD</sup>	100%	Arg102	Asp57 <sup>ECD</sup>	84%
Thr6	His302 <sup>5.40</sup>	69%	Arg109	Asp57 <sup>ECD</sup>	69%
Lys1	Glu294 <sup>ECL2</sup>	17%	Arg112	Asp50 <sup>ECD</sup>	44%
Thr36	Asp101 <sup>ECD</sup>	16%	Asp113	Thr295 <sup>ECL2</sup>	32%
Gln10	His302 <sup>5.40</sup>	16%	Arg102	Gln61 <sup>ECD</sup>	27%
Asn22	Asp101 <sup>ECD</sup>	16%	Asp113	Lys54 <sup>ECD</sup>	12%
Ser34	Asp101 <sup>ECD</sup>	11%	Asp113	His296 <sup>ECL2</sup>	11%
Arg11	Ser105 <sup>ECD</sup>	11%			

**Table C.14 | Hydrogen bonds between (i) peptide and CTR or RAMP (ii) RAMP and CTR during MD simulations of (#7) rAmy:RAMP3:CTR:Ga system.**

Hydrogen bond persistency is quantified as the percentage of frames (over all the frames obtained by merging the different replicas)

(#7) rAmy:RAMP3:CTR:Ga					
Peptide vs CTR/RAMP			RAMP vs CTR		
rAmy residue	CTR/RAMP residue	Hydrogen Bond occupancy (%MD frames)	RAMP residue	CTR residue	Hydrogen Bond occupancy (%MD frames)
Thr6	His302 <sup>5.40</sup>	56%	Asp113	Thr295 <sup>ECL2</sup>	34%
Arg11	Asp373 <sup>7.39</sup>	50%	Asp113	His296 <sup>ECL2</sup>	31%
Lys1	Glu294 <sup>ECL2</sup>	45%	Asp107	Lys54 <sup>ECD</sup>	19%
Thr30	Asp101 <sup>ECD</sup>	38%	Asp107	Arg58 <sup>ECD</sup>	16%
Ser29	Asp101 <sup>ECD</sup>	26%	Asp116	His296 <sup>ECL2</sup>	15%
Thr9	His226 <sup>3.36</sup>	24%	Glu112	Arg58 <sup>ECD</sup>	13%
Gln10	His226 <sup>3.36</sup>	21%	Thr105	Asp57 <sup>ECD</sup>	11%
Gln10	Gln227 <sup>3.37</sup>	21%			
Gln10	His223 <sup>3.33</sup>	21%			
Asn3	Tyr299 <sup>5.37</sup>	17%			
Asn3	Arg362 <sup>ECL3</sup>	15%			
Gln10	His302 <sup>5.40</sup>	13%			
Arg11	Glu294 <sup>ECL2</sup>	13%			
Gln10	His201 <sup>2.67</sup>	13%			

