Unveiling the activation mechanism of the GnRH Receptor

Elpiniki Paspali

Thesis submitted to the University of Strathclyde for the degree of Doctor of Philosophy

Department of Chemical & Process Engineering

University of Strathclyde

2024

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Elpiniki Paspali Date: 02/04/2024

Acknowledgments

This section will be long, and I am glad to say so. I would like to express my deepest gratitude to Dr Paul Mulheran and Dr Valerie Ferro, my research supervisors, for their exceptional mentorship, advice, and support. Their profound insights, patience, and encouragement have been pivotal in my growth as a researcher. Their continuous support through academic and health challenges I faced throughout my PhD has been an unwavering source of strength and reassurance, for which I am profoundly grateful.

My deepest thanks go to my family for their unconditional support and belief in my abilities. Their love, encouragement, and sacrifices have been the bedrock of my academic journey and personal growth. To my mum: words are not enough to express my love for you, nothing is. You sacrificed yourself to raise us, and I know you would do it all over again. I owe everything that I am, have achieved, and will be to you. To my siblings: my life would have been incomplete without you. I cannot imagine walking this earth without the two of you. Being your sister has been the greatest privilege of my life. To my grandparents: Thank you for raising me and keeping me safe. To my aunt: your life experience has been an example to me. Thank you for everything you have done for me. To my father: Thank you for your love, support, and all the important life lessons. Despite the challenges, you all carried me through three academic degrees, and I love you all with all my heart.

My deepest thanks to my dear friends Vasiliki and Doxie. You have truly witnessed all the struggles I went through and helped me keep going no matter what. You supported me and never stopped believing in me. I could never have wished for better and truer friends than you two. Thank you for everything.

To all my teachers and professors, starting from high school to now, you have taught me everything I know. Each one of you has been a unique example to me. You made me love, understand, and pursue science, and for that, I am grateful.

To my fellow PhD students, each one of you has helped me in a unique way. Magda, you were always there to help whenever I needed. Ben, your kindness, and support meant the world to me. Haider, you are a true inspiration to the world.

I am indebted to the University of Strathclyde for the financial support, which made this research possible. Their commitment to advancing knowledge is commendable. Additionally, I would like to express my gratitude to the ARCHIE-WeSt Supercomputer administration staff and especially Karina, for their patience and continuous troubleshooting throughout my PhD studies.

This work would not have been possible without the collective contributions and encouragement of all those mentioned and many others whose support has been a source of strength throughout my doctoral studies.

Abstract

G-protein coupled receptors (GPCRs) are indispensable signalling molecules that orchestrate a myriad of physiological processes, rendering them pivotal therapeutic targets. Remarkably, approximately 35% of approved drugs modulate GPCR activity, underscoring their profound significance. Nonetheless, understanding the mechanisms underpinning GPCR activation remains a challenge, particularly in the context of the GnRH1 receptor (GnRH1R). This receptor is dedicated to mediating the effects of gonadotropin-releasing hormone (GnRH) and the production of fertility hormones, exhibits unusual structural characteristics, such as the absence of a C-terminal helix and variations in what are otherwise highly conserved motifs amongst the GPCR superfamily. Unveiling the GnRH binding that triggers GnRH1R activation is imperative, and characterisation of the GnRH1R active conformation will address a critical knowledge gap.

The research methodology employed a combination of computational docking and molecular dynamic (MD) simulations. Docking simulations were performed using the Rosetta software and tailored protocols for membrane proteins and flexible peptide ligands. The docking process resulted into thousands of potential GnRH binding modes which were clustered and filtered based on criteria such as energy metrics and contacts with experimentally important receptor residues. The binding mode elimination process yielded two promising binding modes representing the native-like GnRH-GnRH1R complex. MD simulations of the selected best ranking binding modes were performed to investigate their ability to activate the GnRH1R.

Receptor activation was monitored and evaluated by the increase of the distance between two key transmembrane helices (TM3 and TM6) - a widely accepted characteristic of GPCR activation. One of the predicted binding modes induced activation after 1.0 μs of simulation and displayed an increased TM3-TM6 distance (~13 Å) compared to the inactive GnRH1R (~8 Å). GnRH binding displayed high stability and involved several π-π interactions, especially through tryptophan 3 (W3) and tyrosine 5 (Y5) of GnRH and the CWxPY motif of the receptor. The crucial residue R8 of GnRH was found to form cation-π interactions with $W280^{6.48}$ of the CWxPY motif and in addition, it mediated interaction to the G-protein binding pocket (DRS motif) and the DPxxY motif through interactions with residues of TM3 and water mediated hydrogen bonds respectfully. The resulting active GnRH1R conformation exhibited an open G-protein binding pocket, an enlarged ligand binding pocket and key rearrangements of the DPxxY motif indicative of activation. Communication between GnRH1R and lipid molecules was also observed.

The significance of this work transcends the confines of GnRH1R, contributing essential knowledge to the broader field of GPCR biology and computational studies. The identified binding mechanism and proposed interhelical communication networks offer valuable insights for drug design targeting GnRH1R, with implications for developing more precise therapeutics for reproductive system-related disorders. Finally, the details of the discovered communication pathways initiated by GnRH pave the way for a deeper understanding of GPCR activation.

Table of Contents

List of Figures

List of Tables

Abbreviations

1

1-palmitoyl-2-oleoylphosphatidylcholine
POPC 43

3

3-10 helices
G 94

5

5-(pi) helices
I 94

A

alpha helices
H 94

B

Backbone
BB 45
bends
I 94
beta-strands
B 94
binding energy
ΔG binding 13, 40, 76, 154
Brunger-Brooks-Karplus
BBK 26

C

$C279^{6.47}$-$W280^{6.48}$-$T281^{6.49}$-$P282^{6.50}$-$Y283^{6.51}$
CWxPY 7
Canonical ensemble
NVT 24
Cluster Centre of Mass
CoM 72
confidence score indicator
pLDDT 34

D

$D138^{3.49}$- $R139^{3.50}$-$S140^{3.51}$
DRS 7
$D138^{3.49}$- $R139^{3.50}$-$Y140^{3.51}$
DRY 7
$D319^{7.49}$-$P320^{7.50}$-$L321^{7.51}$-$I322^{7.52}$-$Y323^{7.53}$
DPxxY 7
Define Secondary Structure of Proteins
DSSP 92, 93
diacylglycerol
DAG 2

E

Effect Size
ES 59
endoplasmic reticulum
ER 2
extended strands
E 94
extracellular loops
ECLs 2

F

family-wise error rate
FWER 40, 59
Fast Fourier Transform
FFT 27
follicle-stimulating hormone
FSH 1

G

Gonadotrophin releasing 1 hormone
GnRH 1
Gonadotrophin releasing 1 hormone receptor
..... 1
G-protein coupled receptors
GPCRs ..... v, 1
G-proteins
GPs 1
guanosine diphosphate
GDP 4

H

half-maximal effective concentration
EC_{50} 9
Hydrogen bonds
H-bonds 48, 78, 92, 173, 174

I

inositol 1,4,5-triphosphate
IP_3 2

Amino acid Abbreviations

Amino acid name	3-letter code	1-letter code	Properties (at pH 7)
Alanine	Ala	A	Neutral
Arginine	Arg	R	Positive
Asparagine	Asn	N	Neutral
Aspartate	Asp	D	Negative
Cysteine	Cys	C	Neutral
Glutamine	Gln	Q	Neutral
Glutamate	Glu	E	Negative
Glycine	Gly	G	Neutral
Histidine	His	H	Positive 10%, Neutral 90%
Isoleucine	Ile	I	Neutral
Leucine	Leu	L	Neutral
Lysine	Lys	K	Positive
Methionine	Met	M	Neutral
Phenylalanine	Phe	F	Neutral
Proline	Pro	P	Neutral
Serine	Ser	S	Neutral
Threonine	Thr	T	Neutral
Tryptophan	Trp	W	Neutral
Tyrosine	Tyr	Y	Neutral
Valine	Val	V	Neutral

Chapter 1 - Introduction

1.1 Synopsis

G-protein coupled receptors (GPCRs) are vital signalling molecules involved in various physiological processes and are currently targeted by 35% of approved drugs[1]. This thesis is dedicated to the investigation of the specific interactions between the GPCR-Gonadotrophin releasing 1 hormone receptor (GnRH1R) and its native agonist GnRH with a specific focus on elucidating the molecular events that lead to receptor activation. To achieve this, a combination of computational docking and molecular dynamics (MD) simulations were employed. Despite their importance, GnRH receptors have been relatively underexplored by computational means. The behaviour and influence of the unique structural characteristics of GnRH1R, including the lack of a C-terminal helix, as well as differences to several conserved motifs amongst GPCRs, make it an interesting candidate for understanding agonist activation of similar GPCRs. Additionally, the GnRH binding mode that induces activation will be beneficial for the development of new and more effective agonists with improved therapeutic applications due to the novel insights of this research.

In this chapter, the properties of GPCRs will be reviewed, focussing on their signalling pathways, active/inactive conformation characteristics, and functionality of critical motifs. The contrasting structure of the GnRH1R is presented next, highlighting the current gaps in understanding its active conformation. Known aspects of the binding between GnRH1R and GnRH is reviewed, emphasising the impact of mutagenesis studies in revealing key interaction sites. Finally, the aims and specific objectives of the thesis are detailed, addressing the knowledge gaps that could significantly impact the current understanding of the activation of similar GPCRs and to the development of new, more effective therapeutics targeting the GnRH1R.

1.2 Background

GPCRs serve as pivotal orchestrators in cellular signalling, effectively transducing extracellular stimuli into intracellular responses through selective coupling to G-proteins (GP)[2], [3]. The GnRH1R is classified within Class A of the GPCR superfamily[4]. The primary function of the GnRH1R resides in the regulation of the reproductive system, where it catalyses the synthesis of luteinising hormone (LH) and follicle-stimulating hormone (FSH) within the pituitary gland[4]. This process commences upon the binding of the GnRH hormone which is released from GnRH neurons in a pulsative manner[5], [6], [7], [8]. Expression of GnRH1R predominantly localises to the surface of gonadotrophic cells within the pituitary gland. To date it is believed there are

two forms of GnRH present in mammals - GnRH1 and GnRH2 - that correspond to their cognate receptors GnRH1R and GnRH2R, respectively.

GnRH1: pE1-H2-W3-S4-Y5-G6-L7-R8-P9-G10-NH_2

GnRH2: pE1-H2-W3-S4-H5-G6-W7-Y8-P9-G10-NH_2

GnRH1R can be activated by binding both GnRH isoforms, although it has higher affinity for GnRH1[9]. In contrast, GnRH2R displays selectivity for GnRH2[10]. Thus, GnRH1R initiates different signalling pathways when activated by GnRH1 and GnRH2, resulting in different physiological effects[11]. The GnRH2 and GnRH2R genes are present in humans, but the GnRH2R gene displays coding errors that often hinder full-length protein production, resulting in five or seven transmembrane helices[12]. Consequently, this thesis focusses on the GnRH1-GnRH1R complex, and henceforth GnRH refers to GnRH1 unless otherwise specified.

Class A GPCRs are characterised by seven transmembrane (TM) helices, three extracellular loops (ECLs), three intracellular loops (ICLs) and an additional C-terminal domain (H8). Diverging from the typical structural features observed in most Class A GPCRs, the GnRH1R is characterised by the absence of the cytosolic H8 domain. Consequently, GnRH1R couples to G_q/G_{11} or G_i proteins, thereby initiating a signalling cascade that primarily stimulates phospholipase C (PLC) rather than adenylate cyclase (AC)[13], [14], [15], [16]. PLC functions by hydrolysing phosphatidylinositol-4,5-diphosphate ($P1P_2$) into inositol 1,4,5-triphosphate (IP_3) and diacylglycerol (DAG), resulting in an upsurge in protein kinase C (PKC) activity [13], [14], [15], [16]. Here, IP3 binds to specific receptors on the endoplasmic reticulum (ER) membrane, thereby instigating the efflux of calcium ions from the ER internal reservoir into the cytoplasm. The subsequent elevation in cytoplasmic calcium concentration acts as a signal for various cellular processes (Figure 1.1).

Furthermore, PKC initiates a cascade of events culminating in the activation of the Rapidly Accelerated Fibrosarcoma (RAF) kinases within the mitogen-activated protein kinase (MAPK) pathway. Subsequently, RAF phosphorylates and activates downstream kinases, resulting in the activation of MAPKs such as ERK1/2 [13], [14], [15], [16]. These activated MAPKs assume a pivotal role in the regulation of cellular growth and proliferation, thus underscoring the significance of GnRH1R in the modulation of these processes [13], [14], [15], [16].

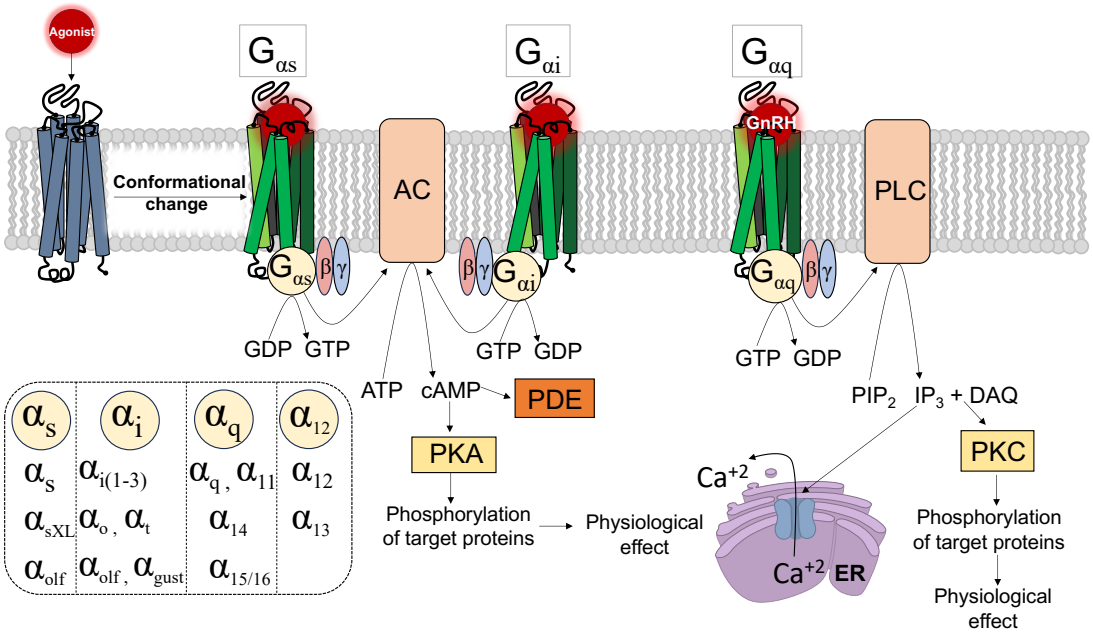


Figure 1.1: Signalling cascade induced by agonist binding in GPCRs. The image depicts three different signalling cascades induced by Gαs, Gαi, and Gαq. The GnRH1R initiates the $G\alpha_{q/11}$ or G_i cascades as shown at the right end and centre of the figure.

Consequently, the expression of GnRH1R has been documented in various tissues, including pituitary adenomas, ovary, breast, testis, prostate, and granulosa-luteal cells. Furthermore, GnRH1R exhibits overexpression in cancers that affect gonadal steroid-dependant organs, including 86% prostate cancers, 80% of endometrial and ovarian adenocarcinomas, and 50% of breast cancer cases[17], [18], [19], [20], [21], [22]. In such cases, direct GnRH1R inhibition has been administrated for treatment. GnRH derivatives have found extensive use in the treatment of reproductive dysfunctions and sex-hormone dependent diseases (Table 1.1).

Table 1.1: Marketed GnRH Agonists and Antagonists targeting the GnRH1R. All data retrieved from PubChem (https://pubchem.ncbi.nlm.nih.gov).

Name (Brand name | PubChem ID	**Medical Application**	**Action and Structure**
GnRH	Native ligand	Stimulation of GnRH1R pGlu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH_2
Buserelin (Superfact | 50225)	Endometriosis, breast cancer, prostate cancer, endometrial hyperplasia, uterine fibroids, female Fertility,	**Agonist** H-Pyr-His-Trp-Ser-Tyr-D-Ser(tBu)-Leu-Arg-Pro-NHEt
Goserelin (Zoladex | 5311128)	Endometriosis, endometrial hyperplasia, breast cancer, prostate cancer, female fertility, uterine fibroids, uterine haemorrhage	**Agonist** H-Pyr-His-Trp-Ser-Tyr-D-Ser(tBu)-Leu-Arg-Pro-NHNHCONH_2
Leuprorelin (Lupron | 657181)	Endometriosis, breast cancer, menorrhagia, precocious puberty, prostate cancer, uterine fibroids	**Agonist** H-Pyr-His-Trp-Ser-Tyr-D-Leu-Leu-Arg-Pro-NHEt

Gonadorelin (Factrel | 638793)	Cryptorchidism, hypogonadotrophic hypogonadism, delayed puberty, veterinary medicine (assisted reproduction)	**Agonist** H-Pyr-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH_2
Nafarelin (Synarel | 25077405)	Endometriosis, precocious puberty	**Agonist** H-Pyr-His-Trp-Ser-Tyr-D-2Nal-Leu-Arg-Pro-Gly-NH_2
Triptorelin (Decapeptyl | 25074470)	Female infertility, Endometriosis, precocious puberty, breast cancer, uterine fibroids, prostate cancer	**Agonist** H-Pyr-His-Trp-Ser-Tyr-D-Trp-Leu-Arg-Pro-Gly-NH_2
Ganirelix (Orgalutran, Antagon | 16130957)	Ovulation, assisted reproduction, endometriosis, control,	**Antagonist** Ac-D-2Nal-D-Phe(4-Cl)-D-3Pal-Ser-Tyr-D-hArg(Et,Et)-Leu-hArg(Et,Et)-Pro-D-Ala-NH_2
Elagonix | 11250647	Heavy menstrual bleeding, uterine fibroids, endometriosis, heavy uterine bleeding, folliculogenesis,	**Antagonist** Nonpeptide
Fertirelin (Ovalyse | 188304)	Veterinary medicine (assisted reproduction)	**Agonist** H-Pyr-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-NHEt

1.3 General mechanism of Class A GPCR activation

GPCRs, including serotonin, opioid, adenosine, chemokine, adrenergic, cannabinoid, muscarinic, and dopamine receptors, have been observed to couple GPs after the disruption of the ionic lock between conserved residues $R^{3.50}$ and $E^{6.30}$ (Ballesteros-Weinstein numbering scheme[23] X.YY, where X is the TM number and YY is the residue number according to the most conserved residue 50, e.g., $E^{6.30}$ is located on TM6 and position 30 before the most conserved residue of the helix 6.50) of the inactive receptor. This disruption arises due to the interference exerted by the GP-α5 subunit, which promotes the establishment of a pre-coupled GPCR-GP complex (Figure 1.2). Upon the binding of an agonist, the pre-coupled complex undergoes full activation. The activation process is finalised by the exposure of guanosine diphosphate (GDP) of the Gα subunit which facilitates the initiation of signalling pathways[24]. The GP coupling of GPCRs is guided by a selectivity 'barcode' unique to each of the 16 Gα proteins (Figure 1.1), where different receptors can interpret the same barcode utilising distinct regions and conformations[25].

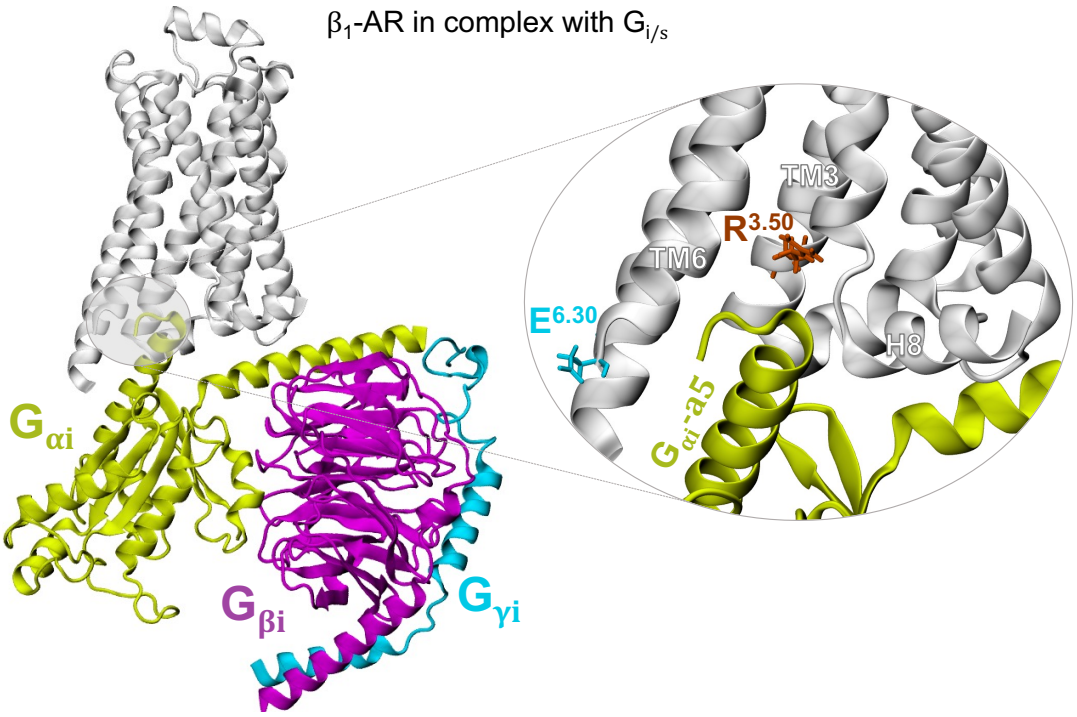


Figure 1.2: Example of GPCR-GP coupling. The β_1-AR in complex with the heterotrimeric $G_{i/s}$ chimera protein (PDBID:7S0G). The ionic lock between TM3-TM6 dissociates for GP interaction.

The disruption of the ionic lock between R3.50 and E6.30 is deemed critical for GP coupling and subsequent activation. The salt bridge (SB) existing between R3.50, a component of the conserved DRY motif, and E6.30 plays a pivotal role in establishing a tightly linked connection between TM3 and TM6 (Figure 1.3). To elaborate further, receptors known for coupling with G_q/G_{11}, such as the M_1-Muscarinic Receptor-G_{11}, M_3-Muscarinic Receptor-Gq, α_2A-Adrenergic Receptor-G_q, 5-HT_2c-Serotonin Receptor-G_q, and 5-HT_2A-Serotonin Receptor-G_q, exhibit post-GP coupling increase of the distance between R3.50 and E6.30 of approximately 10, 7, 8, 8, and 7 Å, respectively[24]. The disruption of the TM3-TM6 ionic lock following GP coupling, results in the opening of the cytoplasmic region of GPCRs leading to an outward movement of the cytosolic end of TM6[2], [24], [25], [26], [27], [28], [29]. Therefore, a substantial outward kink of the cytosolic region of TM6 in Class A receptors is observed, spanning a range of 7-13 Å, along with an average outward displacement of ~2.0 Å at the extracellular end and a substantial rotation (greater than 23°) [27] (Figure 1.3). However, the E6.30 position is not highly conserved across GPCRs and cannot provide a universal mechanism for stabilising inactive conformations[30]. Although TM distance difference varies across GPCRs, the accepted activation distances between TM3-TM6 are ~6Å increase upon activation and TM3-TM7 ~3Å decrease upon activation.

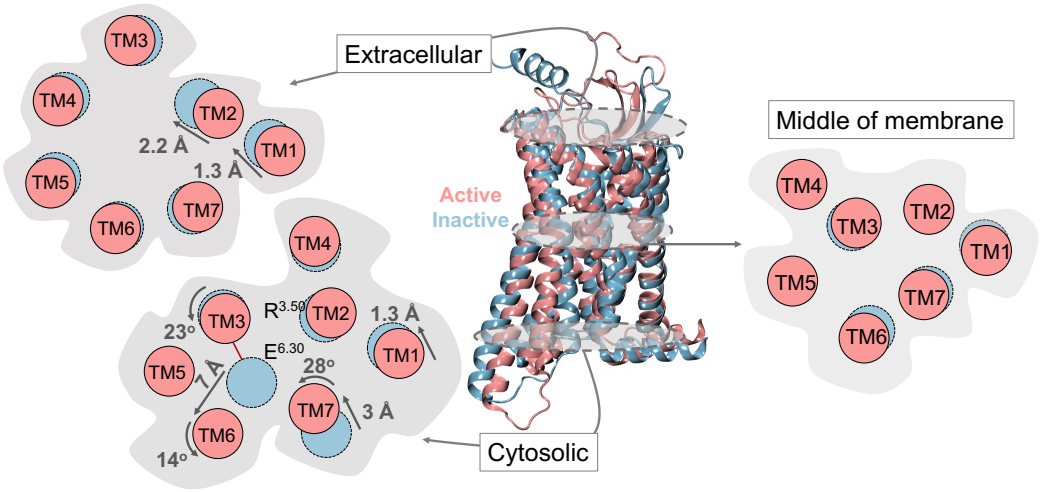


Figure 1.3: Comparison of TM distances and rotations in the active (pink) and inactive (blue) GPCR conformations. The figure was recreated using the GPCRdb website[31] (https://gpcrdb.org/structure_comparison/comparative_analysis#). Data were collected from 10 inactive and active Class A GPCRs that couple G_q/G_{11} and G_i (UniProtID: NTR1, OPRM, CNR1, AA1R, ACM2, 5HT1B, ACM1, OPRK, GHSR, 5HT5A). The figure represents the active (pink) and inactive (blue) adenosine receptor (A_1R).

Furthermore, it has been demonstrated that GP selectivity is primarily guided by the intracellular loop 3 (ICL3) of GPCRs[28]. Despite the absence of strict conservation, even among closely related GPCRs, ICL3 plays a pivotal role in the autoregulation of receptor activity by orchestrating a dynamic conformational equilibrium between states that either obstruct or expose the receptor's GP-binding site.

Following GP binding, ICL3 undergoes an outward movement, effectively enhancing receptor activity[28]. Moreover, the length of ICL3 varies across different receptor. GPCRs possessing long ICL3 domains, such as β_2AR, M_1R, CB_1R, V_1AR, β_1AR, and D_1R, rely on this region to maintain GP selectivity, unlike receptors with shorter ICL3 domains, exemplified by A_1R[28].

Additionally, in the case of Class A GPCRs, there is a growing body of evidence suggesting that the conformation of TM7, particularly the conserved NPxxY motif (where x represents any residue), can influence the efficiency of arrestin coupling and activation, thereby contributing to the observed signalling bias among agonists[32], [33], [34], [35].

1.4 Unique characteristics of the GnRH1R

In contrast to most Class A GPCRs, the GnRH1R exhibits several distinctive features and variations (Table 1.2). The most prominent difference is the lack of the cytoplasmic C-terminal H8 helix which is commonly utilised for arrestin recruitment. Additionally, the experimentally determined inactive

GnRH1R lacks the highly conserved D138$^{3.49}$- R139$^{3.50}$-Y140$^{3.51}$ (DRY) motif typically found in other GPCRs, instead featuring the D138$^{3.49}$- R139$^{3.50}$-S140$^{3.51}$ (DRS) motif.

Table 1.2: Motifs and conserved residues comparison between Class A GPCRs and the GnRH1R.

Conserved GPCR motif	**Function in GPCRs**	**GnRH1R motif**	**Function in GnRH1R**
D$^{3.49}$-R$^{3.50}$-Y$^{3.51}$ **DRY**	Ionic lock-GP interaction site [36], [37]	D138$^{3.49}$-R139$^{3.50}$-S140$^{3.51}$ **DRS**	Structural and activation of cellular signalling [38], [39].
C$^{6.47}$-W$^{6.48}$-x-P$^{6.50}$-Y$^{6.51}$ **CWxPY**	-Conformation-independent conserved interhelical network[40], [41]. -Conserved water-mediated polar network[42]. -Forms an exaggerated kink that opens the G-protein binding pocket when TM6 rotates[43].	C279$^{6.47}$-W280$^{6.48}$-T281$^{6.49}$-P282$^{6.50}$-Y283$^{6.51}$ **CWxPY**	Structural and ligand binding affinity[40], [41].
N$^{7.49}$-P$^{7.50}$-x-x-Y$^{7.53}$ **NPxxY**	-Conformation-independent conserved interhelical network[40], [41]. -Conserved water-mediated polar network[42]. -Forms conformation-specific interhelical interactions[44].	D319$^{7.49}$-P320$^{7.50}$-L321$^{7.51}$-I322$^{7.52}$-Y323$^{7.53}$ **DPxxY**	Structural, possible Na^+ counter ion, activation of cellular signalling[45], [46], [47], [48]
P$^{5.50}$-I$^{3.40}$-F$^{6.44}$ **PIF**	Facilitates movement of the cytoplasmic TM6 region. Part of the transmission switch[42], [49], [50], [51]	P223$^{5.50}$-A129$^{3.40}$-F276$^{6.44}$ **PAF**	Unknown function [52]
Y$^{5.58}$ (96% of Class A)	Important for receptor activation [42], [49], [50], [51]	N231$^{5.58}$ Polar interaction with S136$^{3.47}$	Tight packing of TM5 with TM3 and TM6 in inactive GnRH1R[52]
Cytoplasmic tail-Helix 8	Present	Absent	Y323$^{7.53}$ contacts F56$^{1.53}$ and W63$^{1.60}$ in TM1 (due to absence of H8 in inactive GnRH1R) [52].
X$^{6.40}$	Short hydrophobic residues.	F272$^{6.40}$ is highly conserved in tailless mammalian GnRH1R.	Important to activation: function unknown [52].
D$^{2.50}$	Allosteric Na^+ binding or activation[40], [41], [42], [53].	N87$^{2.50}$ -Direct polar interaction with D319$^{7.49}$ [52].	-Potential involvement in receptor activation and the interhelical mediated network[52].

Additionally, the widely observed R$^{3.50}$-E$^{6.30}$ ionic lock present in most inactive GPCRs, including Rhodopsin, is notably absent in the GnRH1R. Instead of E$^{6.30}$, the GnRH1R possesses R262$^{6.30}$, which obviously cannot form an ionic lock with R139$^{3.50}$. This non-conserved position in the GnRH1R, leads to the formation of a SB between D138$^{3.49}$ and R139$^{3.50}$ and a polar interaction

between R139$^{3.50}$ and T265$^{6.33}$ (Figure 1.4). Therefore, activation characteristics for the GnRH1R are determined by the TM3-TM6 R139$^{3.50}$ and T265$^{6.33}$ distance instead.

Furthermore, the highly conserved NPxxY motif found in most Class A GPCRs is altered in the GnRH1R, where it becomes DPxxY, and D319$^{7.49}$ replaces the conserved N$^{7.49}$. This difference results in a unique negatively charged motif in TM7 that could potentially serve as a sodium-binding site. Additionally, the GnRH1R possesses an extended ICL3, comprising 15 amino acids, further distinguishing it from other Class A GPCR like Rhodopsin.

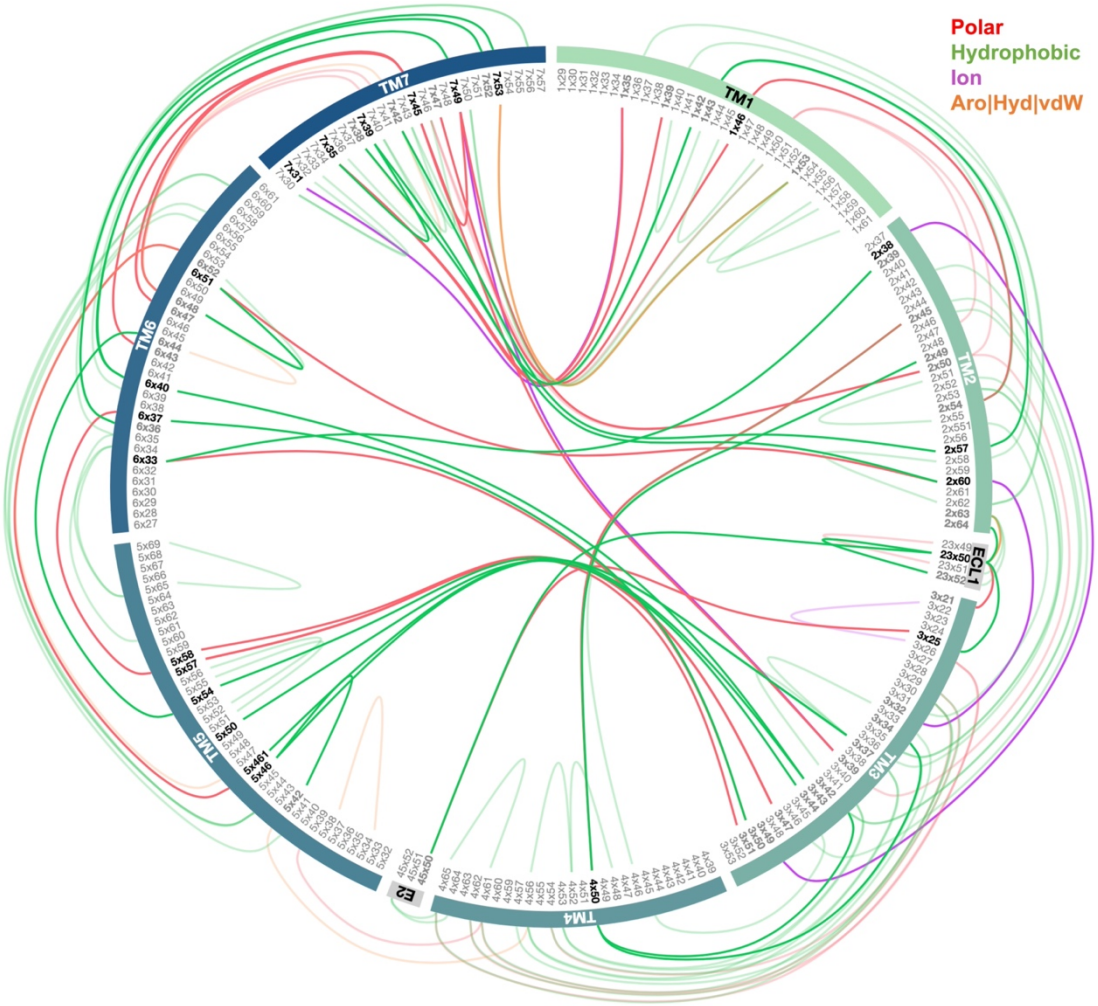


Figure 1.4: Flare plot of residue interactions in the inactive GnRH1R. The most important interhelical interactions are highlighted. Flare plot created using the GPCRdb website[31] (UniProtID: GnRHR)

1.5 GnRH binding to GnRH1R

The development of novel therapeutics targeting the GnRH1R critically depends on a foundational premise: the characterisation of its active conformational state and the corresponding GnRH binding mode. Currently, the inactive crystal structure of the GnRH1R in complex with the antagonist

elagonix is available[52]; however, the essential active conformation remains elusive. Experimental mutagenesis studies have played a vital role in identifying key residues within the GnRH1R that influence GnRH binding (Table 1.3).

Table 1.3: Experimentally defined GnRH-GnRH1R binding pocket.

Receptor mutant	**Signal Transduction IP assays**		**Reference**
GnRH1R	**GnRH**	**Elagonix**	
Mutation	**EC_{50} (nM)**	**IC_{50} (nM)**	
WT	3.9±0.44	8.1	[54]
L23A	4.9±0.85	5.8±2.3†	[54]
Q174A	21±2.8	17	[54]
F178A	76±30	9.3†	[54]
W280F	ND		[54]
Y283F Y283A	7.3±3.1† ND	0.32†	[54]
Y284F	7.6±1.5	2.8	[54]
L286A	69±8.6		[54]
H306A	55±25	51	[54]
Y323F Y323A	21±5.1 ND		[54]
D302N	12.6±1.8*		[55]
Y290A Y290A	 2.7±0.40†	1262±113**	[56] [54]
R38A	989-fold reduction in binding affinity		[57]
D98A	ND		[54]
K121Q	ND		[54]
N87D	ND		[54]
M125A	ND		[54]
P128K	ND		[54]
N321Y	ND		[54]
F272V	ND		[54]
N87D	ND		[54]
D319N	ND		[54]

*Wild Type (WT):EC_{50}=0.29±0.007, p<0.001, **WT: IC_{50}=3.8±0.14, †: value multiplied by 10^3. ND means that the EC_{50} and IC_{50} values could not be calculated based on the experimental data [54]. IP accumulation was measured after stimulation with GnRH ligands for 2h [54]. EC_{50} values are expressed as means ± SEM (n=3) at 3 times independently experiment repeats with similar results [54].

These studies have demonstrated a residue's impact on GnRH binding or GnRH1R function by an increase of the EC_{50} (half-maximal effective concentration). In this context, the EC_{50} signifies the concentration of a substance required to elicit a response halfway between the baseline and the maximum, thus providing insight into the receptor's potency in inducing a biological effect. Conversely, the IC_{50} represents the concentration necessary to inhibit a biological process or response by 50%, reflecting the receptor's

inhibitory potency. Interaction between the wild type (WT) receptor residue and GnRH is evident when the EC_{50} values increase, whereas interaction of the antagonist elagonix (present in the crystal structure) with receptor residues is evident by the increase of IC_{50} values increase (Table 1.3, Figure 1.5B). Although these mutations prove valuable, it is unknown whether residues affect receptor function and expression by direct GnRH binding or because they are essential for the receptor's structural integrity.

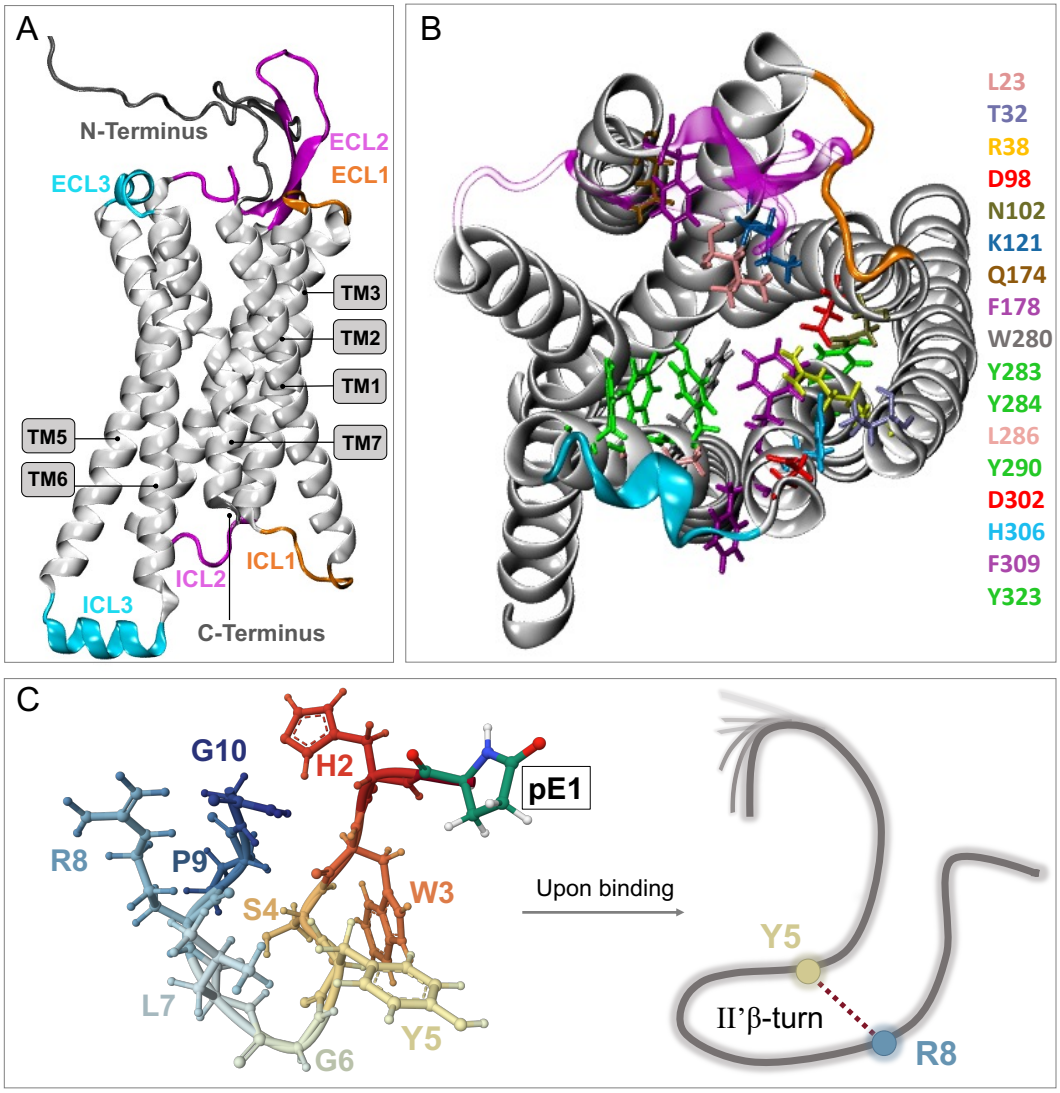


Figure 1.5: A) Structural characteristics of GnRH1R depicting the 7TM helices, Termini, ECLs and ICLs (coloured) (PDBID:7BR3). B) Experimentally defined GnRH binding pocket (Table 1.3) (several GnRH-binding associated residues are not displayed for clarity). C) Structural characteristics of GnRH where pE1 atoms are showing in green for carbon, red for oxygen, blue for nitrogen and white for hydrogen (PDBID:1YY1).

The GnRH hormone consists of ten amino acids with the sequence: pE1-H2-W3-S4-Y5-G6-L7-R8-P9-G10. The N-terminal residues, including pE1, H2, and W3, have been associated receptor specificity[4]. On the other hand, the

C-terminal residues, particularly R8, are essential for establishing high-affinity binding to the GnRH1R. Although GnRH can adopt various conformations due to its structural flexibility, the prevalent conformation involves a II'β-turn with opposed termini[58]. This specific conformation is characterised by the presence of an intermolecular H-bond between the hydrogen atom of the backbone nitrogen of Y5 (i) and the backbone oxygen atom of R8 ($i+3$) (Figure 1.5C). Additionally, this conformation is defined by the trans conformation of the nitrogen in G6 and oxygen in L7.
Although mutations have offered insights into the roles of specific receptor residues in GnRH binding, the exact nature of these interactions remains speculative. Previous computational docking studies have aimed to elucidate the binding characteristics and proposed several interactions [4].

1.5.1 D302$^{7.31}$-R8

The proposed interaction between D302$^{7.31}$ and the unique residue R8 in mammalian GnRH, was suggested due to the distinctive nature of R8. Receptor specificity for R8 through D302$^{7.31}$ was proposed[59] and the mutation of D302$^{7.31}$ resulted in a decrease in GnRH binding affinity when mutated to uncharged amino acids, while mutations to charged amino acids had no observable effects[59], [60]. The proposed mechanism involved the formation of a salt bridge (SB) between R8 and D302$^{7.31}$, leading to a high-affinity GnRH conformation. However, it was concluded that the binding affinity of GnRH is not exclusively dependent on the R8-D302$^{7.31}$ interaction[60].

1.5.2 F308$^{7.37}$-W3 assisted by H306$^{7.35}$

The proposed interaction between F308$^{7.37}$ and W3, mediated by H305$^{7.35}$ in the mouse GnRH receptor, was suggested by a computational study and evaluated through mutations. The mutation of the H305$^{7.35}$ residue in the mouse GnRH receptor, particularly with nonpolar amino acids, resulted in a decrease in GnRH binding affinity, while mutations with polar amino acids had no discernible effect[61]. It was hypothesised that H305$^{7.35}$ forms an intrahelical contact with F308$^{7.37}$, consequently driving F308$^{7.37}$ to engage in π-π interactions with W3 of GnRH[61]. In the human GnRH1R investigated herein, H305$^{7.35}$ of the mouse receptor corresponds to H306$^{7.35}$ in the human receptor. Mutation of both F308$^{7.37}$ and H306$^{7.35}$ resulted in reduced GnRH binding affinity by significant increase in EC_{50} values[52].

1.5.3 Y290$^{6.58}$-Y5

The proposed interaction involving Y290$^{6.58}$ and Y5, as indicated by computational studies, gained significance due to the observed reduction in

GnRH binding affinity upon the mutation of Y290$^{6.58}$[52], [62]. The suggested functional role of Y290$^{6.58}$ was its potential function as a coupling agent for agonists, initiating the rotation of TM6.

1.5.4 Y283$^{6.51}$ and F309$^{7.38}$-W3

The Y283$^{6.51}$ residue of the CWxPY motif is located in TM6 and showed a decreased GnRH-binding affinity upon mutation[52], [63]. Its importance was suggested to explore GnRH binding similarly to ETB binding to the endothelin and apelin receptor deep in the intrahelical bundle[64], [65]. Experimental studies suggested involvement of W3-F309$^{7.38}$[61] in GnRH binding by the disruption of the Y283$^{6.51}$ -F309$^{7.38}$ intrahelical contact[66].

1.5.5 R38$^{1.35}$-G10

Computational studies suggested interactions of R38$^{1.35}$ with G10 of GnRH due their proximity in the computational model. Mutation of R38$^{1.35}$ resulted in a reduced GnRH-binding affinity[67] although the proposed interaction was not confirmed. The decreased binding affinity was justified as a broken intrahelical contact that assists GnRH binging and is lost after the R38$^{1.35}$ mutation. Additionally, the R38$^{1.35}$:D98$^{2.60}$ SB was suggested in the inactive receptor conformation which is disrupted by GnRH binding in the active conformation.

1.5.6 W280$^{6.48}$-W3

Experimental studies have shown that mutations of W280$^{6.48}$ significantly affect GnRH binding and cellular signalling[52], [63]. The proposed interaction from computational studies of W280$^{6.48}$ and W3 occur through aryl-aryl interactions[63], [68], [69].

1.5.7 K121$^{3.32}$-E1/H2/W3

The inactive receptor conformation was suggested to possess two salt bridges between E90$^{2.53}$:K121$^{3.32}$ and R38$^{1.35}$:D98$^{2.60}$ [70], [71]. Upon activation, GnRH binding is thought to disturb the E90$^{2.53}$:K121$^{3.32}$ SB and cause the rotation of K121$^{3.32}$ towards the formation of the D98:K121$^{3.32}$ SB in the active conformation[72]. However, the existence of the E90$^{2.53}$:K121$^{3.32}$ SB was disproved by later studies[73]. Computational models previously indicated interactions of K121$^{3.32}$ with GnRH residues E1, H2, and W3. Mutation studies of K121$^{3.32}$ revealed reduced GnRH binding affinity and signal transduction[52], while antagonist binding remained unaffected[63], [74]. Despite these findings, the precise role of K121$^{3.32}$ in GnRH binding remains ambiguous, warranting further investigation to conclusively determine its functional significance in receptor activation and GnRH interaction.

1.5.8 $L23^{N\text{-}ter}$

The $L23^{N\text{-}ter}$ residue of GnRH1R N-terminus increased receptor sensitivity values, EC_{50}, upon mutation[52]. However, the specific involvement of $L23^{N\text{-}ter}$ to GnRH binding has not been elucidated.

1.5.9 $T32^{1.29}$

Mutation of $T32^{1.29}$I naturally occurs in the human GnRH1R gene in patients with idiopathic hypogonadotrophic hypogonadism.[75] This mutation reduces GnRH1R function primarily by reducing GnRH binding affinity[75]. A GnRH-$T32^{1.29}$ interaction has not currently been suggested by experimental or computational studies.

1.5.10 $N102^{2.64}$-G10

The mutation of $N102^{2.64}$ has been shown to induce a 225-fold loss in GnRH potency[76]. Interaction of $N102^{2.64}$ with G10 was proposed by prior computational docking studies[11].

1.5.11 $D98^{2.60}$-E1 or H2

Mutation of $D98^{2.60}$ resulted in a reduction in GnRH binding affinity and the H2-$D98^{2.60}$ interaction was proposed[77]. Interestingly, it has been suggested that $D98^{2.60}$ forms a SB with $K121^{3.32}$ upon GnRH binding and receptor activation. Contrastingly, the $R38^{1.35}$:$D98^{2.60}$ SB is formed in the inactive receptor conformation[72].

1.6 Aims and Objectives

The characterisation of the active conformation of GPCRs is crucial for advancing the current understanding of the activation process and for the development of improved therapeutics. This thesis focusses on uncovering the details of the GnRH binding mode that induces GnRH1R activation and the characterisation of the GnRH1R active conformation.

Aim 1: The prediction of 'close-to-native' binding modes employing computational docking simulations with Rosetta.
Objectives:

1. Conduct large-scale docking poses from a several starting positions around the binding pocket to ensure unbiased binding.
2. Implement an energy- and conformation- biased clustering protocol to categorise binding poses into clusters.
3. Analyse the binding interface based on binding energy, total energy, and solvent-accessible area to eliminate clusters with low populations and unfavourable bindings.

4. Perform contact analysis to identify clusters with experimentally meaningful GnRH-GnRH1R contacts.
5. Apply multiple criteria for binding pose selection to identify the most energetically favourable binding modes with the highest number of experimentally validated contacts.
6. Select the two best scoring structures for further investigation through MD simulations in search of the active GnRH1R conformation and GnRH binding.

Aim 2: The characterisation of the GnRH1R active conformation and GnRH binding using MD simulations.
Objectives:

1. Conduct MD simulations for all selected binding modes to assess the stability of the binding mode and receptor activation.
2. Analyse TM3-TM6 distances to determine whether the binding mode induced receptor activation.
3. Perform trajectory analysis to characterise in depth the active binding and receptor conformation.
4. Investigate the behaviour of unique motifs as well as their communication and function.
5. Monitor the conformational transition of the GnRH1R from the inactive to the active state.
6. Generate communication maps to elucidate the signalling process from GnRH to GnRH1R.
7. Discuss how the active binding mode found supports experimental evidence regarding GnRH binding and how it contrasts with previously suggested interactions from computational docking studies.

1.7 Thesis overview

Chapter 2 describes the fundamental theoretical background of Molecular Dynamics (MD) and computational docking simulations, laying the groundwork for subsequent research.

Chapter 3 provides a detailed guide of the protocols, methods, and analysis techniques employed to achieve the research objectives. It offers technical and practical guidance for computational docking and MD simulations, ensuring complete transparency in the methodology used for this thesis.

Chapter 4 demonstrates the results from computational docking simulations of the GnRH-GnRH1R complex. The selected binding modes, chosen from thousands of structures, highlight unique characteristics and form the basis for MD simulations exploring GnRH-binding and GnRH1R activation.

Chapter 5 analyses the evolution of GnRH binding modes as selected by computational docking simulations through MD simulations. Extensive details of the binding mode are revealed, offering new insights into the binding process.

Chapter 6 is dedicated to the active conformation of GnRH1R and its characteristics. This chapter employs long-scale MD simulations to uncover various unique structural changes during the conformational transition from inactive to active GnRH1R states. In this chapter the novel activation process of GnRH1R is revealed and insights into the functionality of several key residues and motifs are discussed.

Chapter 2 - Theory

2.1 Molecular Dynamics

Complex chemical molecules such as proteins, nucleic acids and polymers consist of a network of atoms bonded in a specific way that create life at a molecular level[78]. For example, proteins consist of a combination of the 20 amino acids, folded in a 3D structure consistent with their biological function[79]. Given the intricate structural and functional nature of biomolecules, computational methods have emerged as valuable tools for in-depth analysis. They enable the exploration of behaviours and functions that are currently beyond the reach of conventional experimental techniques[80]. With faster and more powerful computational modelling, it becomes possible to delve into complex biological systems, gaining in depth understanding[81].

Molecular dynamics (MD) is one of the first computational methods to arise and become a valuable tool in the fields of physics, chemistry, and biology[82]. This computational technique is used to simulate the behaviour of atoms and molecules over time, offering insights into their collective behaviour[82]. A wide range of phenomena can be investigated, such as the binding process of ligands to proteins and subsequent conformational changes[83].

Atomic representation

Experimentally solved protein structures are most commonly used as the initial structure for MD simulations[84]. A biological system in MD is represented as a network of atoms where each atom has a precise position and velocity. Each atom is represented as a point object with mass and various interatomic interactions such as covalent bonds represented as elastic springs[85]. To explore the motion of the network of atoms in a biomolecule, MD implements Newton's equations of motion. A biological system composed of *N* atoms has a given geometry:

$$\vec{R} = \{\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_N\} \quad [2\text{-}1]$$

where the overall system can be expressed as a set of distinct atomic vectors $\vec{r} = (x, y, z)$. For example, a water molecule is comprised of three atoms, one oxygen atom, $\vec{r}_1$, and two hydrogen atoms $\vec{r}_2$ and $\vec{r}_3$ (Figure 2.1). The vectors $\vec{r}_1$, $\vec{r}_2$, and $\vec{r}_3$ describe the position of each atom in space.

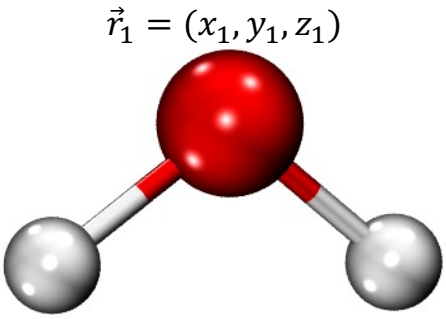


Figure 2.1: Water molecule containing three atoms. R, is the set of vectors that consists of one oxygen atom $\vec{r}_1 = (x_1, y_1, z_1)$ and two hydrogen atoms $\vec{r}_2 = (x_2, y_2, z_2)$ and $\vec{r}_3 = (x_3, y_3, z_3)$.

Successive configurations of the system are obtained by implementing Newton's second law of motion[86] as follows:

$$\vec{F} = m\vec{a} \qquad [2\text{-}2]$$

where $\vec{F}$ is the force acting on an atom, $\vec{a}$ is the acceleration and m is the mass of the atom. In MD, Newton's law [2-2] is a general principle that relates force to acceleration and governs the behaviour of atoms by guiding their movements. The acceleration of an atom is the second derivative of its position $\vec{r}_i$, and from [2-2]:

$$\overrightarrow{F_i}(\vec{R}) = m_i\vec{a}_i = m_i\frac{d^2\vec{r}_i(t)}{dt^2} = -\nabla_i \mathrm{U}(\vec{R}) \qquad [2\text{-}3]$$

where $\overrightarrow{F_i}(\vec{R})$ is the force acting on atom i, which in general is a function of the coordinates of all atoms in the system. It is evaluated by finding the negative gradient with respect to $\vec{r}_i$ of the potential energy $\mathrm{U}(\vec{R})$ of the system. MD aims to solve Newton's equation and find the position of all atoms in the system at a given time, $\vec{r}_i(t)$, where $i = 1,2 \ldots N$. For a system composed of more than two atoms, the above equations cannot be solved analytically. To overcome this problem, a numerical approach is implemented. This approach proposes that integration can be divided into a finite time steps, Δt. Molecular motions are usually in the range of femtoseconds. Therefor from [2-3]:

$$\vec{a}_i(t) = -\left(\frac{1}{m_i}\right)\nabla_i \mathrm{U}(\vec{R}) \qquad [2\text{-}4]$$

Given the position $\vec{r}_i(t)$ and velocity $\vec{v}_i(t)$, the acceleration $\vec{a}_i(t)$ can be calculated and the position and velocity at a future time $(t + \Delta t)$ can be predicted. The application of this method for multiple steps results in a trajectory that describes the position and velocities of the total number of atoms present in the system over time. The most commonly implemented integration method is the Velocity Verlet algorithm that ustilises two 3rd order Taylor expansions[87] to approximate the position and dynamic properties of

a molecule at a future time $t + \Delta t$ and at a past time $t - \Delta t$ based on its position, velocity and acceleration at a current time t:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{1}{2}\vec{a}_i(t)\Delta t^2 \quad [2\text{-}5]$$

while the velocity is updated every $\Delta t + {}^{\Delta t}/_{2}$ steps:

$$\vec{v}_i\left(t + \frac{\Delta t}{2}\right) = \vec{v}_i(t) + \frac{1}{2}\vec{a}_i(t)\Delta t \quad [2\text{-}6]$$

Therefore, the acceleration at the subsequent step at time $t + \Delta t$ is:

$$\vec{a}_i(t + \Delta t) = -\left(\frac{1}{m_i}\right)\nabla U(\vec{r}_i(t + \Delta t)) \quad [2\text{-}7]$$

and the velocity in the next step at time $t + \Delta t$ is:

$$\vec{v}_i(t + \Delta t) = \vec{v}_i\left(t + \frac{\Delta t}{2}\right) + \frac{1}{2}\vec{a}_i(t + \Delta t)\Delta t \quad [2\text{-}8]$$

2.2 Force field

A force field is a combination of equations used to calculate the potential energy of a system of atoms that can include both bonded and non-bonded interactions. Therefore, the incorporation of a force field is a way to describe interactions between atoms accurately and thus calculate the forces at play[88]. Various force fields are suitable for MD simulations, such as CHARMM[89], Amber[90], and GROMOS[91]. In MD, all force fields use the same equation to calculate the total potential energy of a system and produce similar results[92] as follows:

$$U_{total} = U_{bonded} + U_{nonbonded} \quad [2\text{-}9]$$

where the total potential energy of the system (U_{total}) is the sum of the bonded (U_{bonded}) and non-bonded ($U_{nonbonded}$) potential energies of all atoms.

Each MD software package has a recommended force field that complements its algorithm. The CHARMM[89] force field and Nanoscale Molecular Dynamics (NAMD) package were selected for this research due to their compatibility, accuracy, and efficiency.

The CHARMM[89] force field is an all-atom force field that consists of parameters for a wide range of chemical groups present in biological systems. The components of the CHARMM[89] additive force field calculate the potential energy of bonded and non-bonded interactions. The bonded potential accounts for bond stretching, angle torsions and proper/improper dihedral torsions, while the non-bonded potential accounts for van der Waals and electrostatic interactions.

2.2.1 Bonded interactions

Bonded interactions consist of covalently bonded atoms. In CHARMM, the harmonic oscillator approximation is used to describe covalent bonds. Therefore, for the duration of the MD simulations covalent bonds cannot be formed or broken. The potential energy of bonded interactions include:

- 2-atom spring bond potential
- 3-atom spring angle potential
- 4-atom torsional angle potential: proper and improper dihedral torsion potential

Bond Potential

A bond between two atoms is represented as a spring that in isolation yields a harmonic vibrational motion between an i, j atom pair:

$$U_{bonds}(r_{i,j}) = \sum_{bonds} K_b(r_{i,j} - r_0)^2 \qquad [2\text{-}10]$$

where $r_{i,j} = \|\vec{r}_j - \vec{r}_i\|$ is the distance between the atoms, r_0 is the equilibrium distance and K_b is the spring constant.

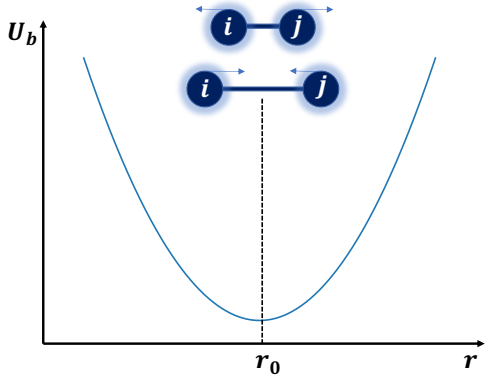


Figure 2.2: The potential curve and representation of bond stretching. The potential energy, U_b, increases when the atoms i and j are in minimum and maximum distances and is minimal at the optimum distance r_0.

The potential energy (U_b) associated with bond stretching tends to be minimised when the bond is at its equilibrium length (r_0) reflecting the natural tendency of systems to achieve a state of lower energy for increased stability and balance. Deviations from this equilibrium leads to an increased potential energy.

Angle Potential

An angle is formed between three atoms (i, j, k) and is described by the angular harmonic potential:

$$U_{angles}(\theta) = \sum_{angles} K_\theta(\theta - \theta_0)^2 \qquad [2\text{-}11]$$

where θ is the angle in radians between the vectors:

$$r_{ij} = \vec{r}_j - \vec{r}_i \; and \; r_{kj} = \vec{r}_j - \vec{r}_k \qquad [2\text{-}12]$$

θ_0 is the equilibrium angle and K_θ is the spring constant.

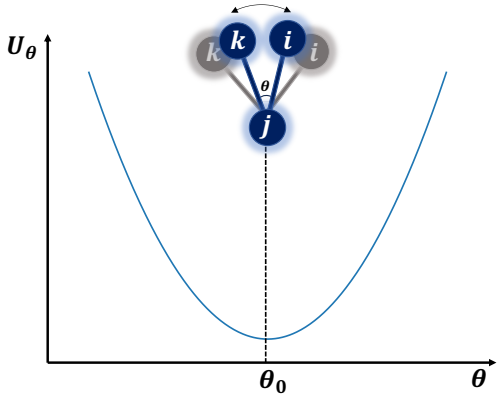


Figure 2.3: The potential curve and representation of angle bending motions. The potential energy, U_θ, increases at minimum and maximum angles and is minimised at the equilibrium angle, θ_0.

Proper dihedral torsion

A set of four atoms i, j, k, l can rotate around a bond axis. The potential energy is the result of the rotation of one group of atoms relative to the other around this axis. The potential energy in a proper dihedral angle is defined through:

$$U_{torsion}(\varphi) = \sum_{proper\ dihedral} K_n\big(1 + cos(n\varphi - \delta_n)\big) \quad if \quad n > 0 \qquad [2\text{-}13]$$

Here ϕ is the angle in radians between the i, j, k plane and the j, k, l-plane (Figure 2.4). The integer constant n and δ_n are the dihedral multiplicity and phase. The dihedral angle, φ, can range between 0 and 2π radians. Multiple minima and maxima indicate different stable conformations and rotational barriers respectfully (Figure 2.4).

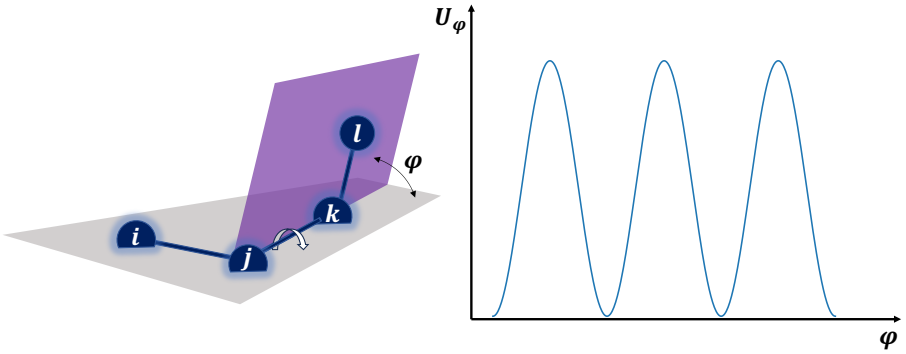


Figure 2.4: Proper dihedral torsion. Left: visual representation of the dihedral torsion around the central $j - k$ bond. Right: The potential curve of a proper dihedral torsion.

Improper dihedral

An improper dihedral angle is an "out-of-plane" dihedral and describes the orientation of four atoms (i, j, k, l) in a molecule. In an improper dihedral the angle is expressed by the three vectors formed by atoms in two different planes (Figure 2.5):

$$U_{improper\ dihedral}(\vec{R}) = \sum_{impropers} K_{\varphi}(\varphi - \varphi_0)^2 \quad [2\text{-}14]$$

Where φ is the angle in radians between the two planes creates by the j, i, k and i, k, l planes, φ_0 is the equilibrium angle and K_{φ} is the multiplicative constant.

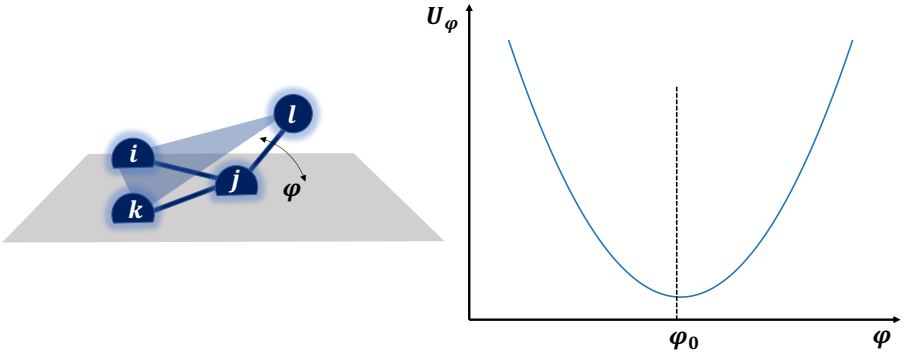


Figure 2.5: Improper dihedral torsion. Left: visual representation of the improper dihedral torsion formed between 2 planes (grey, blue) and 4 atoms: j, i, k and i, k, l. Right: The potential curve of improper dihedral angles.

The potential curve of improper dihedral torsions is a characteristic harmonic oscillator. When $\varphi = \varphi_0$ the potential energy is at a minimum and when it deviates from φ_0 the energy increases symmetrically in a quadratic fashion. In general, bonded interactions contribute to the potential energy as follows:

$$U_{bonds} > U_{angles} > U_{dihedrals} > U_{improper\ dihedrals}$$

This ranking aligns with the general understanding that bond stretching tends to have the most influence on the potential energy as a more direct and strong force, followed by angles and then dihedrals, which involve more subtle vibrational and rotational effects.

2.2.1 Nonbonded interactions

The nonbonded interactions involve atom pairs that are not covalently bonded. The two primary nonbonded interactions included in a force field are the van der Waals and electrostatic interactions[93]. In the calculation of nonbonded potentials, 1-4 interactions are excluded as interactions involving atoms separated by exactly three bonds are already taken into consideration in the bonded potential terms. To prevent the duplication of these interactions

in the energy calculations, 1-4 interactions are treated separately using force field parameters specifically designed for such interactions (see Chapter 3 - 3.8). This approach ensures accurate energy calculations without double-counting interactions in MD simulations.

Van der Waals

The van der Waals (vdW) interactions[94] are short-range repulsive and attractive forces that occur due to electron density fluctuations of molecules (dipoles) (Figure 2.6). VdW forces include:

- Keesom force[95]: the force between two permanent dipoles.
- Debye force[96]: the force between a permanent dipole and a corresponding induced dipole.
- London dispersion force[97]: the force between two instantaneously induced dipoles.

A permanent dipole is formed due to a stable and lasting separation of electric charges and the formation of two oppositely charged poles ($\delta^+$and δ^-). Typical examples of molecules with permanent dipoles include water (H_2O) and ammonia (NH_3) where the high difference in electronegativity between oxygen-hydrogen and nitrogen-hydrogen atoms cause a permanent separation of charge that lead to polar covalent bonds. Permanent dipoles can form an induce dipole in another molecule when in proximity.

The potential curve of vdW interactions described by the Lennard-Jones potential[98] resembles a Morse curve[99]. The potential energy strongly increases in minimum atom-atom distances, minimises in optimum atom-atom distances and approximates zero in maximum atom-atom distances[99] (Figure 2.6).

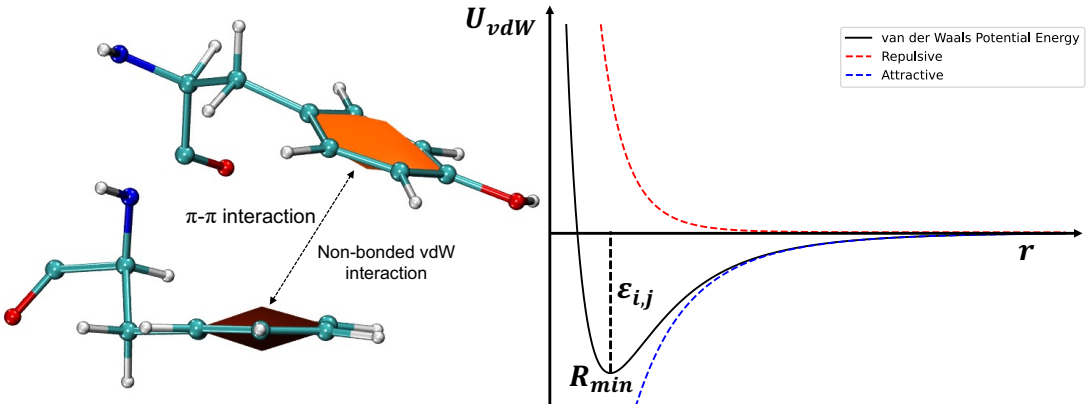


Figure 2.6: VdW interactions. Left: visual representation of vdw interactions present in a molecule. Right: The potential curve of vdW interactions. The repulsive and attractive term of the potential is shown in red and blue colours, respectively.

The Lennard-Jones potential[98] is commonly used to describe both attractive and repulsive vdW forces:

$$U_{vdW} = \varepsilon_{i,j}\left[\left(\frac{R_{min}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min}}{r_{ij}}\right)^{6}\right] \quad [2\text{-}15]$$

where $\varepsilon_{i,j}$ is the well depth, R_{min} is the radius at which the potential is minimised and r_{ij} is the distance between atoms i and j. The first term of the equation represents the short-range repulsive forces, and the second term represents the longer-range attractive forces.

Electrostatics

Electrostatic interactions are expressed with Coulomb's law[100]:

$$U_{el} = \sum_{i>j} \frac{q_i q_j}{4\pi D r_{ij}} \quad [2\text{-}16]$$

where q_i and q_j are the charges of atoms i and j, D is the dielectric constant of the medium and r_{ij} is the distance of atoms i and j. The potential energy curve of electrostatic interactions forms a steep rise at short distances due to strong atom repulsion (Figure 2.7).

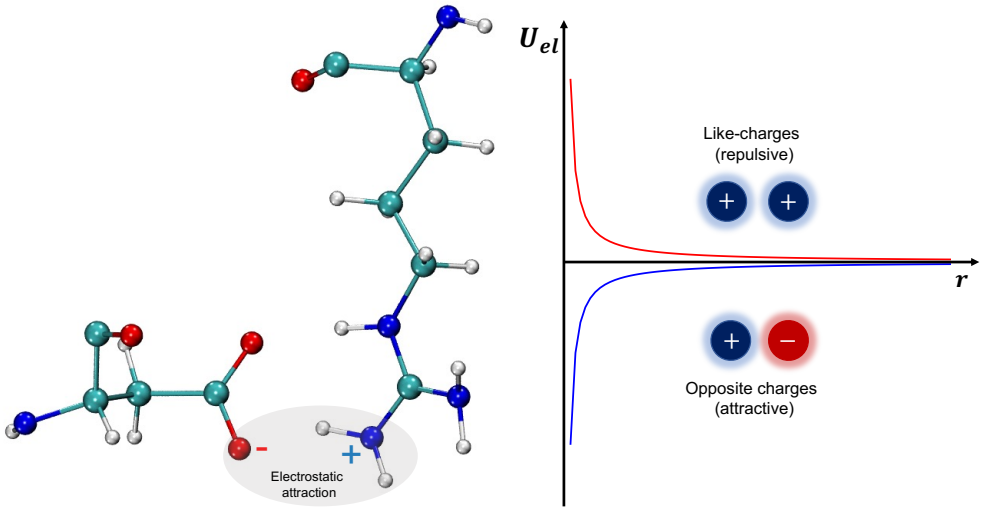


Figure 2.7: Electrostatic interactions. Left: two molecules experiencing electrostatic attraction. Right: The electrostatic potential curve of attractive (blue) and repulsive (red) interactions.

The typical workflow in an MD simulation is given below (Figure 2.8). The initial configuration of the desired system is provided, and the algorithm calculates the potential energy according to the force field. In a second step, the forces and subsequently the acceleration of all atoms in the system is computed. According to the Velocity Verlet algorithm [2-5]-[2-8] the software assigns new positions and velocities to all atoms which results in a new configuration. These steps are repeated for the desired amount of time and the formation of the MD trajectory.

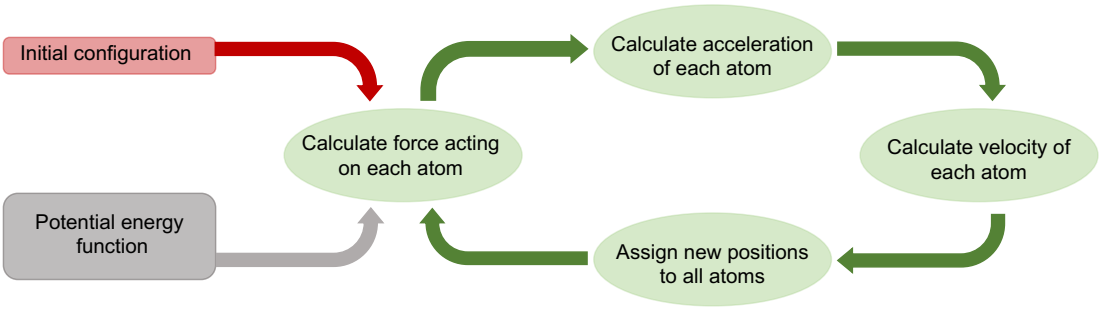


Figure 2.8: Diagram depicting the MD pipeline for the calculation of atom position, velocity and acceleration that correspond to an MD trajectory. Modifications of this workflow to accommodate constant pressure and temperature are described below.

2.3 NAMD

NAMD[101] is a widely used software package for the application of MD simulations. NAMD is optimum for the simulation of large biomolecular systems due to its efficiency and scalability in using parallel computing resources including CPUs and GPU accelerated systems. Therefore, NAMD[101] was chosen as the software package to conduct MD simulations for this research.

2.3.1 NAMD workflow

The first step in obtaining a NAMD trajectory is the preparation of the system of interest. The second step includes the force field choice (CHARMM in this research) (Figure 2.9). In the next step, the user specifies parameters and conditions for the production of the resulting MD simulations of the system.

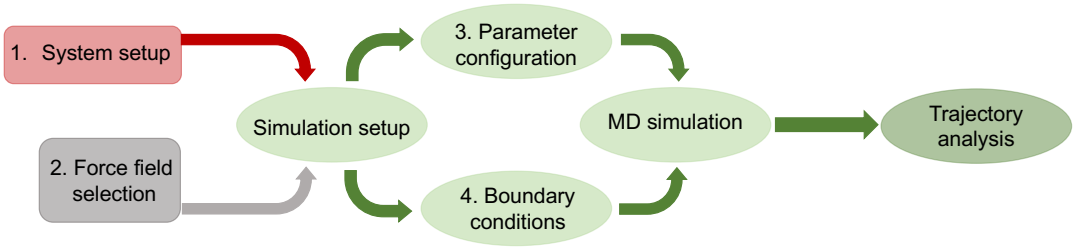


Figure 2.9: General workflow of a NAMD simulation.

2.3.2 Ensembles

In the third step of the NAMD workflow (Figure 2.9), parameters such as temperature, pressure, time step and simulation length are specified. Here the user must choose the ensemble of the MD simulation. There are three main ensembles[102] for MD simulations:

1. Microcanonical ensemble (NVE)[103]: the number of atoms (N), the system volume (V) and the total energy (E) are conserved.
2. Canonical ensemble (NVT)[104]: the number of atoms (N), the system volume (V) and the temperature (T) are conserved.

3. Isothermal-isobaric ensemble (NPT): the number of atoms (N), the system pressure (P) and the temperature (T) are conserved.

NAMD uses a thermostat to control the temperature and a barostat to control the pressure. These components help achieve and maintain the desired thermodynamic conditions during the simulation.

2.3.3 Temperature control

A thermostat is employed to achieve temperature control in the NPT and NVT ensembles by velocity adjustment of the atoms in the simulation. The thermostat monitors the temperature of the system and compares is with the user specified temperature (target temperature). If the monitored and target temperature do not agree, the thermostat rescales or adjusts atom velocities to maintain the desired temperature. Common thermostat algorithms are the Berendsen[105], Andersen[106], Nosé-Hoover[107] and Langevin. The Langevin thermostat was selected for this research as it is a stochastic method that balances accuracy and computational speed, commonly used with NAMD.

Temperature is expressed as a function of the momenta of all atoms. In a system containing N atoms then the temperature is estimated from the average of the kinetic energy:

$$\langle \frac{1}{2} m_i \vec{v}_i^2 \rangle = \frac{1}{2} N_f K_B T \qquad [2\text{-}17]$$

where N_f is the number of degrees of freedom in the system containing N atoms and K_B is the Boltzmann constant $\left(1.38 \cdot 10^{-23} \frac{m^2 kg}{s^2 K}\right)$. In MD simulations, temperature is depended on time as atom velocities change over time. Therefore:

$$T(t) = \sum_{i=1}^{N} \frac{m_i \vec{v}_i^2(t)}{K_B N_f} \qquad [2\text{-}18]$$

2.3.4 Pressure control

A barostat is employed in the NPT ensemble of MD simulations to ensure pressure control while allowing the simulation cell to expand or contract to maintain constant pressure. For this research, the Nosé-Hoover Langevin piston barostat was used as it is recommended for MD simulations with the NAMD package.

2.3.5 Langevin Dynamics

In Langevin Dynamics[108], an atom is subjected to both deterministic forces (driven by potential energy gradients) and stochastic forces (influenced by thermal fluctuations). Stochastic forces consist of a dissipative force and a

fluctuating force. Newtown's equations of motion with the addition of the dissipative and fluctuating force form the Langevin equation:

$$m_i \frac{d\vec{v}_i}{dt} = \vec{F}_i - m_i \gamma \vec{v}_i + \sqrt{2 m_i K_B T \gamma}\, \vec{R}_i(t) \qquad [2\text{-}19]$$

Here m_i and $\vec{v}_i$ are the mass and velocity of the atom i, $\vec{F}_i$ is the total force acting on the atom i (Langevin forces excluded), γ is the friction coefficient, K_B is Boltzmann's constant, T is the desired temperature and $\vec{R}_i(t)$ is the random force acting on atom i at time t derived from a normal Gaussian distribution. The term $-m_i \gamma \vec{v}_i$ in [2-19] represents the dissipative force, which expresses the frictional interactions between the atom and the surrounding medium (usually water). The dissipative force expresses the tendency of an atom in motion to lose energy and reduce its velocity due to interactions with its environment. Additionally, the term $\sqrt{2 m_i K_B T \gamma}\, \vec{R}_i(t)$ represents the fluctuating force and consists of the stochastic or random component of the forces in the system. The fluctuating force accounts for the thermal fluctuations in the system by monitoring the random and unpredictable motions of the atoms due to their thermal energy. Moreover, this term ensures that at a given temperature, the atom contributes to the dynamic behaviour of the system and undergoes Brownian motion.

Together, the dissipative and fluctuating force in Langevin dynamics[108] ensure balance between deterministic and stochastic motion. The deterministic forces guide the atoms' trajectory according to the potential energy landscape, while the stochastic forces provide randomness and ensure that the system explores various configurations consistent with the Boltzmann distribution at a given temperature. To compute the velocities of atoms in the presence of a Langevin thermostat[103], the Brunger-Brooks-Karplus[109] (BBK) method is employed. The BBK algorithm combines deterministic, stochastic, and frictional forces to update the velocities of the atoms in the system. The BBK algorithm is an extension of the Velocity Verlet algorithm [2-5]-[2-8], and updates the velocity every half timestep $(t + \Delta t/2)$ as follows:

$$\vec{v_i}(t + \Delta t/2) = \vec{v_i}(t) + \frac{\vec{F}_i(t)}{2m_i}\Delta t - \left(\frac{\gamma}{m_i}\right)\vec{v_i}(t)\Delta t + \vec{R}_i(t)\sqrt{\frac{2K_B T\gamma}{\Delta t\, m_i}} \qquad [2\text{-}20]$$

Additionally, the position of all atoms is updated every $t + \Delta t$ as follows:

$$\vec{x_i}(t + \Delta t) = \vec{x_i}(t) + \vec{v_i}(t + \Delta t/2)\Delta t \qquad [2\text{-}21]$$

2.3.6 Periodic Boundary conditions

In biological systems, molecules exist in a vast environment that is virtually infinite on the molecular scale. Periodic Boundary Conditions (PBC) are employed to mimic this phenomenon in MD simulations. PBC allows the

simulation of a small representative portion of a larger system by using an infinitely repeating lattice. More specifically, the simulation box is defined by three cell vectors representing the shape and size of the unit cell which is replicated to form the periodic arrangement. The contents of the simulation box are copied in all directions to form the lattice of periodic images. When one atom exits the unit cell from one side, it is replaced by a copy (periodic image) entering the cell on the opposite side (Figure 2.10). PBC are valuable for the calculation of vdW interactions as they ensure correct treatment by including interactions with periodic images.

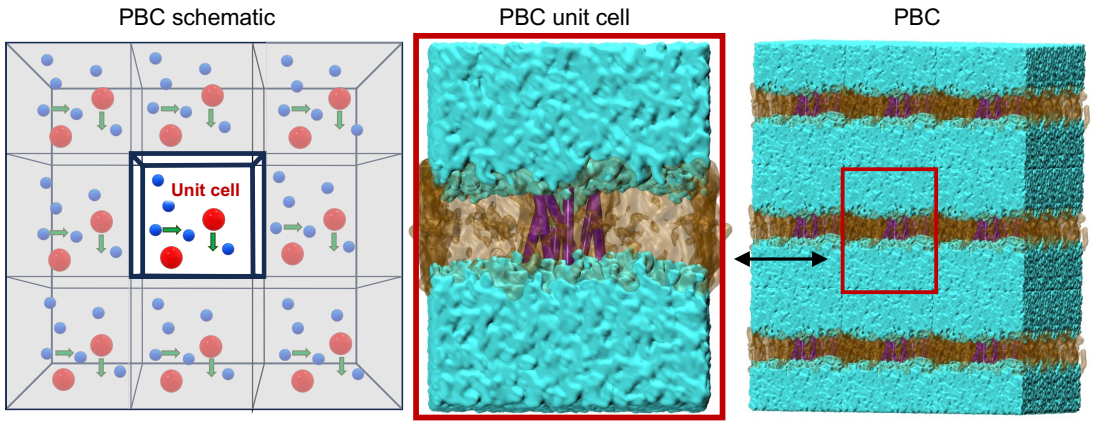


Figure 2.10: Periodic boundary conditions. Left: Schematic representation of a periodic cell. Each atom in the system is replicated periodically. Middle: A single unit cell. Right: Periodic unit cells.

With PCB, atoms at the edge of the box experience the same forces and interactions with the atoms in the bulk. Additionally, PBC allows the conservation of the total number of atoms, momentum, and energy in the system as well as statistical accuracy of the ensemble. There are several types of PBC cell geometries, such as cubic box, octahedron, dodecahedron and more. For this research, the cubic box geometry in MD simulations was employed (Figure 2.10).

2.3.7 Particle Mesh Ewald

The Particle Mesh Ewald (PME)[110], [111] method is widely used in MD simulations to compute long-range electrostatic interactions in periodic systems. The Ewald method[112] involves the summation of short and long-range electrostatic contributions. For short-range interactions, direct summations occur within a user specified cutoff distance. The long-range electrostatic interactions are calculated in reciprocal space using the Fast Fourier Transform (FFT) algorithm[112], [113]. The PME extension of the Ewald method, allows the direct summation of both short and long-range electrostatic interactions without the need for cutoff distances. Long-range electrostatic interactions are computed by employing a 3D grid (charge grid).

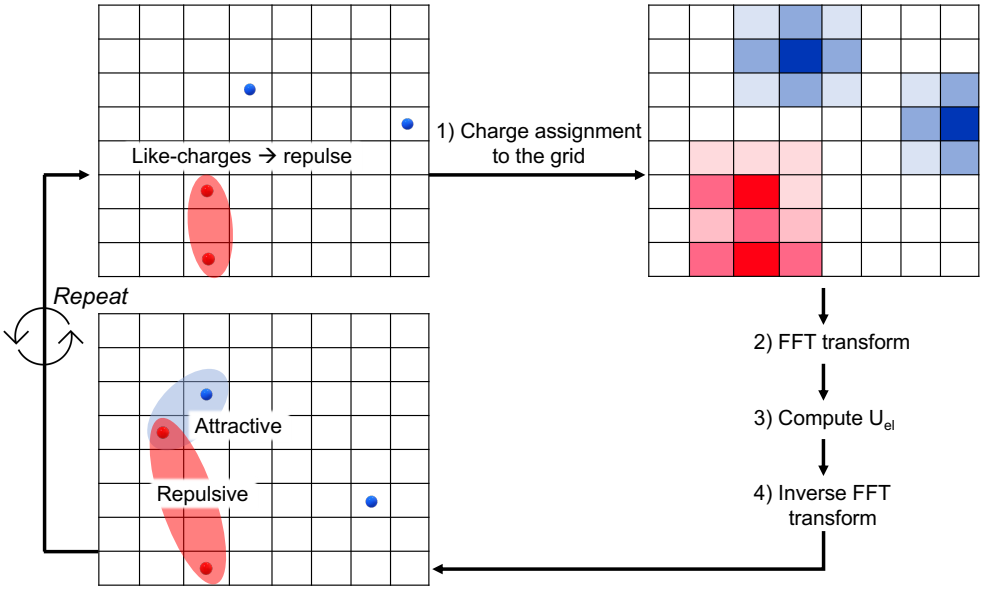


Figure 2.11: PME method pipeline for the calculation of long-range interactions.

Here, the charges of individual atoms are assigned to the grid using interpolation techniques, which involve spreading the charge densities from the atoms to the grid points (Figure 2.11, step 1). In the second step, a FFT is performed on the charge grid to convert the charge distribution from real space to reciprocal space. This enables the efficient evaluation of long-range electrostatics (Figure 2.11, step 2). In the third step, the electrostatic potential (U_{el}) is computed using the FFT-transformed charge distribution and the reciprocal lattice vectors (Figure 2.11, step 3). The last step converts the charges from reciprocal to real space with an inverse FFT transformation and provides the long-range electrostatic contribution (Figure 2.11, step 4). Here the long and short-range electrostatics are combined to obtain the total electrostatic potential of the system. This process is repeated throughout the length of the simulation.

2.4 Computational Docking

Computational docking is a technique that aims to predict the energetically favourable binding modes of a ligand-receptor complex. The Rosetta commons suite[114] was implemented to perform computational docking simulations. Rosetta predicts a binding pose and its associated energy metrics based on a) a force field, b) empirical scoring functions and c) knowledge-based potential energies. The Rosetta scoring function seeks to approximate the bonded and non-bonded potential energies of a ligand-receptor complex as well as free energy contributions.

2.4.1 Rosetta

Rosetta v3.2[114] is a Monte Carlo (MC) based docking algorithm used to identify the native-like ligand-receptor binding pose. The algorithm starts from either a random initial distance between the docking partners or a user specified distance from the receptor's binding pocket (if known). Thereafter, docking is performed in two distinct stages: 1) low resolution docking and 2) high resolution local refinement and minimisation. In stage 1, the protein sidechains (size and charge) are represented in a centroid mode to save computational time (Figure 2.12A). In stage 2, the protein is represented by all atoms in which sidechain orientation between the docking partners is optimised at the expense of computational time (Figure 2.12B).

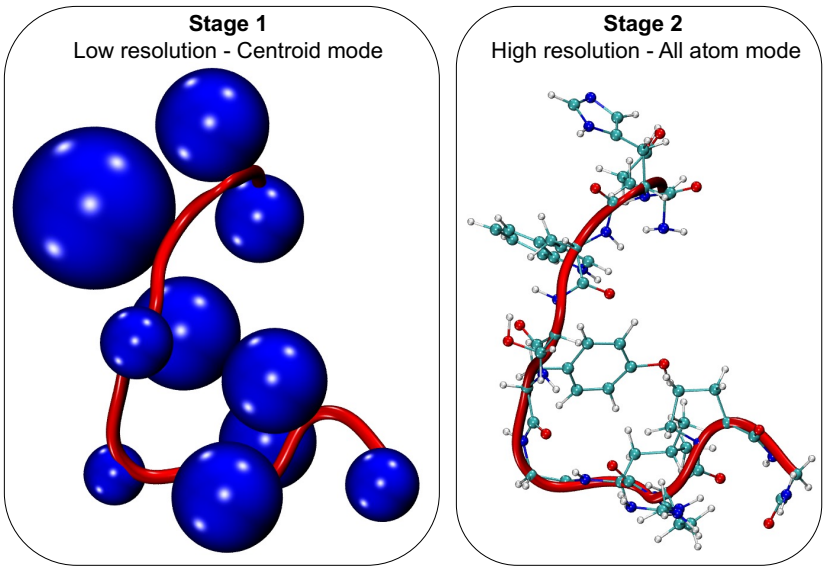


Figure 2.12: Docking stages in Rosetta. Left: Stage 1:Low resolution, atom representation in centroid mode. Sidechains are represented as spheres of representative size and charge. Right: Stage 2: High resolution-all atom mode.

Stage 1: Low resolution docking

Docking initiates with the low resolution stage where the ligand is roughly placed close the binding site to sample possible docking partner-orientations under a simplified scoring function. An initial perturbation of the ligand by ~3 Å and rotation of 8° is performed to create new conformations. The most stable conformations according to the Metropolis Monte Carlo (MMC) method pass to the high resolution stage (Figure 2.12B). Under the MMC acceptance criterion:

$$\Delta E = E_{trial\ conformation} - E_{initial\ conformation} \qquad [2\text{-}22]$$

If $\Delta E \leq 0$, the acceptance probability $P_{accept} = 1$ and the conformation is accepted. If $\Delta E > 0$ then the acceptance probability decreases exponentially with increasing energy difference following the Boltzmann factor $e^{-\Delta E/kT} < 1$. In this case, Rosetta repeats the low resolution stage 500 times and if the trial

conformation continues to have positive ΔE then the conformation is rejected, and Rosetta proceeds with the next trial conformation (Figure 2.13). The acceptance probability ensures that only low energy conformations are accepted. In this stage the potential minima of a conformation are located quickly and efficiently and only the lowest energy conformations pass to the next stage.

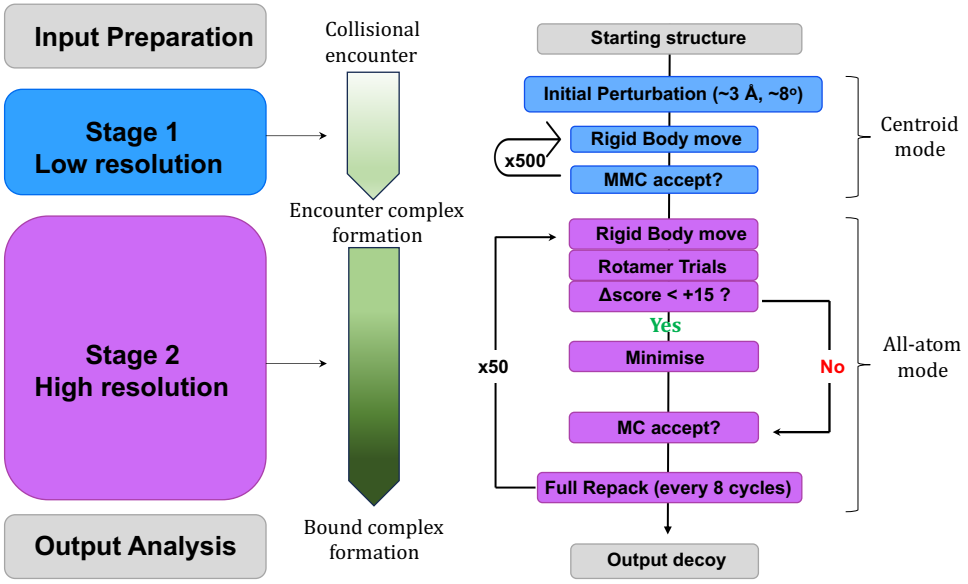


Figure 2.13: Diagram of the docking protocol.

Stage 2: High resolution docking

The successful conformations from Stage 1 undergo high resolution local refinement and minimisation in Stage 2. Here, all atoms of the complex are represented (Figure 2.12B). The position of the docking partners found in Stage 1 is optimised in rotamer trials where the algorithm explores and evaluates different sidechain-backbone orientation. Rotamer trials take into consideration experimentally sidechain orientations and the lowest energy rotamer is selected for the final conformation. The scoring of rotamers $\Delta score$ represent the entropy change due to the selection of a particular rotamer and:

- If $\Delta_{score} < +15$, then the sidechain conformation is relatively constrained or ordered with low entropy.
- If $\Delta_{score} > +15$, then Rosetta performs the MMC method to accept or discard the trial conformation.

Generally, lower Δ_{score} imply more favourable and ordered rotamers that result in stable conformations within a protein. The trial conformations with $\Delta_{score} < +15$ undergo minimisation and repacking. Here, Rosetta applies the energy scoring function[115] considering bonded and nonbonded interactions, solvation effects and other factors to score sidechain conformation. The positions of the sidechains in the lowest scoring conformations are updated to the selected energetically favourable

conformations. This step is repeated for the whole ligand-receptor complex. The resulting trial conformations that successfully pass all steps undergo high resolution refinement for additional 50 times (Figure 2.13). The resulting docking poses include the most favourable lowest ligand-receptor conformations.

Rosetta energy function

Rosetta uses the energy function REF2015[115] which approximates the energy of a binding pose. This energy is computed as a linear combination of energy terms E_i as functions of geometric degrees of freedom, Θ. Each term is scaled by a weight (w_i) according to their chemical identity as follows:

$$\Delta E_{total} = \sum_i w_i E_i(\Theta_i, aa_i) \quad [2\text{-}23]$$

Table 2.1: Summary table of energy terms used in REF15[115] for proteins scoring.

Term	Description	Weight	Units
Fa_atr	Attractive energy between two atoms on different residues separated by a distance d	1.0	kcal/mol
Fa_rep	Repulsive energy between two atoms on different residues separated by a distance d	0.55	kcal/mol
fa_intra_rep	Repulsive energy between two atoms on same residues separated by a distance d	0.005	kcal/mol
fa_sol	Gaussian exclusion implicit solvation energy between protein atoms in different residues	1.0	kcal/mol
lk_ball_wtd	Orientation-dependent solvation of polar atoms assuming ideal water geometry	1.0	kcal/mol
fa_intra_sol	Gaussian exclusion implicit solvation energy between protein atoms in the same residue	1.0	kcal/mol
Fa_elec	Energy of interaction between two nonbonded charged atoms separated by a distance d	1.0	kcal/mol
hbond_lr_bb	Energy of short-range hydrogen bonds	1.0	kcal/mol
hbond_sr_bb	Energy of long-range hydrogen bonds	1.0	kcal/mol
hbond_bb_sc	Energy of backbone−side-chain hydrogen bonds	1.0	kcal/mol
hbond_sc	Energy of side-chain−side-chain hydrogen bonds	1.0	kcal/mol
dslf_fa13	Energy of disulphide bridges	1.25	kcal/mol
rama_prepro	Probability of backbone φ, ψ angles given the amino acid type	(0.45 kcal/mol)/kT	kT
p_aa_pp	Probability of amino acid identity given backbone φ, ψ angles	(0.4 kcal/mol)/kT	kT
fa_dun	Probability that a chosen rotamer is native-like given backbone φ, ψ angles	(0.7 kcal/mol)/kT	kT
Omega	Backbone-dependent penalty for cis ω dihedrals that deviate from 0° and trans ω dihedrals that deviate from 180°	(0.6 kcal/mol)/AU	AU
pro_close	Penalty for an open proline ring and proline ω bonding energy	(1.25 kcal/mol)/AU	AU
Ref	Reference energies for amino acid types	(1.0 kcal/mol)/AU	AU

The complete scoring function of Rosetta are presented in detail in reference [115]. All MD and computational docking simulations were performed using the ARCHIE-WeSt supercomputer (https://www.archie-west.ac.uk/for-academia/acknowledging-archie/).

Chapter 3 - Methodology

3.1 Rationale

The methodology employed in this thesis is fundamental for the generation and analysis of MD and computational docking simulations. This chapter details the technical aspects of the protocols and techniques used to ensure the reliability and validity of the obtained results. Beyond the mere exposition of methodologies, this chapter serves as a guide, aiding researchers with similar interest in applying these protocols effectively offering scientific transparency. The decision-making processes embedded in this chapter offer a nuanced understanding of why specific tools, servers, or algorithms were selected for this research.

3.2 Aims

1. Crystal structures of GPCRs are often reported as engineered complexes with foreign proteins and missing extracellular loops (ECLs) and intracellular loops (ICLs). However, the initial model is critical to obtain robust results that represent the molecule of interest with accuracy. This chapter reveals the decision-making process of selecting the best predicted complete model of the GnRH1R and GnRH, which were used as the basis for all subsequent computations.
2. The establishment and report of a step-by-step protocol for the production and analysis of GnRH-GnRH1R docking poses using Rosetta. In search of the native-like binding mode, thousands of docking poses were produced, selected, and analysed based on several elimination steps and techniques. Incorporation of experimental data in the binding mode selection process reinforces the biological relevancy of the selected representative poses. The screening process resulted in the selection of two close-to-native binding poses for further investigation with MD simulations.
3. The careful preparation of the GnRH-GnRH-membrane systems using established protocols to ensure biologically representative systems through MD simulations.
4. Strategic analysis of the MD simulations to uncover various GnRH1R conformational transitions and structural characteristics. This includes a detailed description of analysis protocols used for the MD simulations for all selected systems, including several plugins, customised code for data analysis and visual representation of the results designed to provide clarity and understanding of the results.

3.3 The GnRH1R model structure

The inactive crystal structure of GnRH1R is available at the Protein Data Bank (PBD) website (https://www.rcsb.org) with PDBID: 7BR3. Due to the challenging crystallisation process of a membrane protein, the GnRH1R crystal structure was reported in absence of several sequences such as the ECLs, ICLs, and N-terminus (Figure 3.1, left).

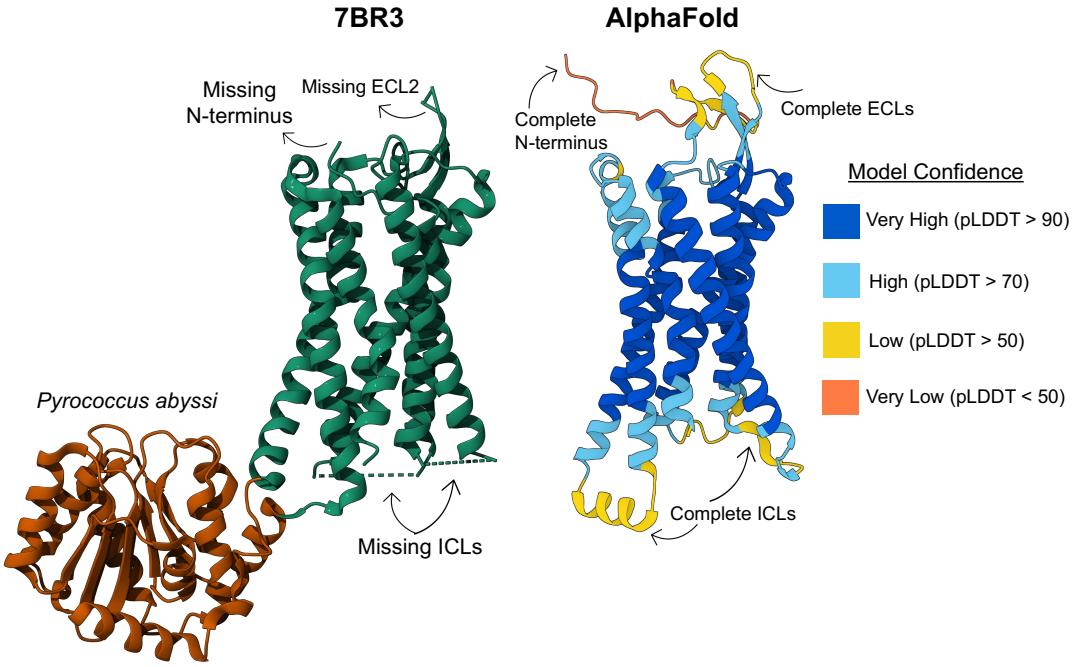


Figure 3.1: The crystal structure of the GnRH1R with PDBID:7BR3 on the left, in comparison with the AlphaFold entry: P306968 predicted model of the wild-type 7BR3 on the right. The AlphaFold model includes the complete structure of 7BR3 where all the missing sequences were modelled and the *Pyrococcus abyssi* moiety is removed. The confidence level of the predicted model is depicted on the right with the confidence score indicator: pLDDT.

To ensure the most accurate representation of the wild-type GnRH1R, the AlphaFold-predicted model was employed (Figure 3.1, right). The AlphaFold model serves as a reliable alternative to the published crystal structure, offering the most accurate wild-type representation of the complete GnRH1R structure, thereby enhancing the overall understanding of its molecular architecture.

The confidence level associated with the predictive AlphaFold model is visually represented on the right side of the Figure 3.1, with the confidence score indicator, (pLDDT). This score serves as a measure of the reliability and certainty associated with guessed sequences. Given the very low confidence score (pLDDT $<$ 50) of the predicted N-terminus due to the absence of pertinent structural data, a judicious decision was made to exclude the initial eleven N-terminal residues (Figure 3.1, AlphaFold: orange colour) from consideration in computational docking simulations due its strong

interference with the docking process. This deliberate choice not only mitigates interference with GnRH binding but also preserves the structural integrity of the disulphide bridge between C14 and C200 in the GnRH1R. In MD simulations where the N-terminus was included, a strong interference with GnRH impeded the binding dynamics of the peptide. For this reason, the N-terminus was omitted from MD systems where GnRH was left to roam the extracellular space for unconstrained binding. However, the N-terminus was included for the MD simulations of Rosetta docked structures and the Apo-GnRH1R (free) systems.

3.4 The GnRH model structure

Lastly, the crystal structure of the GnRH peptide was acquired from the PDB, identified by the PDBID: 1YY1. The primary structure of the GnRH peptide is:

$$pE1 - H2 - W3 - S4 - Y5 - G6 - L7 - R8 - P9 - G10 - NH_2$$

The native GnRH features pyroglutamic acid (PCA) as its N-terminal residue, results from the cyclisation of glutamic acid. While the pyro ring is important to the GnRH conformation, computational exploration of this feature proved not feasible within the confines of this thesis.

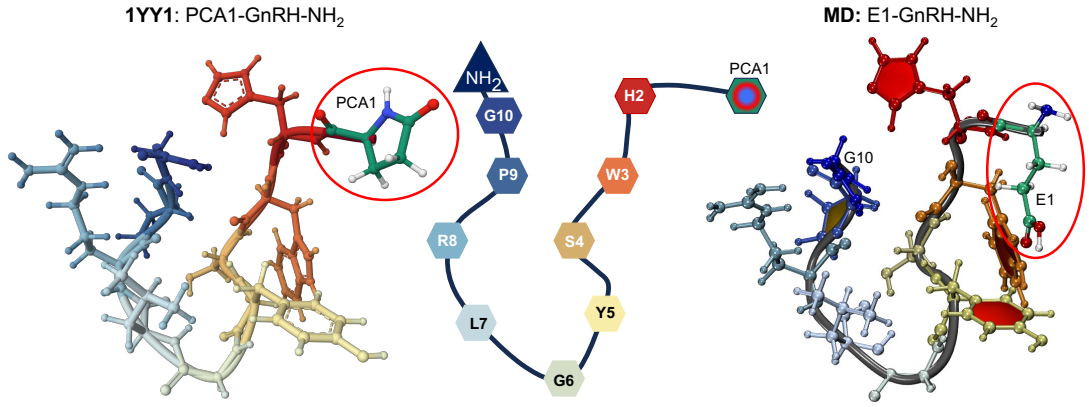


Figure 3.2: 3-dimentional structure of GnRH. Left: PDBID:1YY1 entry for the mammalian GnRH featuring the PCA N-terminal ring. Middle: Schematic representation of 1YY1 denoting the primary structure. Right: E1-GnRH used for MD and docking simulations in this thesis. For the PCA/E1 residue: carbon atoms in green, oxygen atoms in red, nitrogen atoms in blue and hydrogen atoms in white colours.

To preserve as many native GnRH characteristics as possible, the non-canonical amino acid, PCA, was substituted with a protonated glutamic acid (E1) (Figure 3.2). This substitution was chosen to eliminate the negative charge from the N-terminus, aligning with the neutral charge of the PCA ring. It is essential to acknowledge that the chemical properties of E1 deviates from that of PCA and consequently, does not accurately represent the native state of GnRH. Therefore, any interactions reported in this thesis concerning E1

should not be acknowledged as native interactions. Notably, in experimental GnRH binding assays following GnRH1R mutation, the synthetic cysteine-GnRH is commonly used. To approximate native structural interactions of GnRH, the decision was made to employ E1 at the N-terminus instead of cysteine or the complete omission of the N-terminal residue.

In summary, the AlphaFold predicted structure for GnRH1R and E1-GnRH-NH_2 were utilised to investigate their binding mode using computational docking simulations and subsequent MD simulations were conducted to explore the active and inactive conformations of the GnRH1R (Figure 3.1,Figure 3.2).

3.5 Computational Docking

Computational docking simulations were conducted using the Rosetta Commons suite (https://www.rosettacommons.org), specifically employing the FlexPepDock application (https://www.rosettacommons.org/docs/latest/application_documentation/docking/flex-pep-dock).

The overall aim is to simulate the flexible backbone of GnRH and enhance the probability of identifying a binding mode closely resembling native binding. Experimentally proven GnRH interacting residues in terms of EC_{50} and IC_{50} (Table 1.3, Figure 1.5) were taken into consideration in the binding mode selection process.

The docking protocol consisted of a total of 27,000 docking poses to thoroughly explore the conformational space. To accommodate the software requirement of reading structures starting from residue number 1, the initial structure was renumbered, aligning N12 with N1. Additionally, contiguous numbering of GnRH and GnRH1R residues was ensured. Consequently, the first-last residue for GnRH1R was numbered 1-317, while for GnRH 318-327.

3.6 Docking step-by-step guide

Docking simulations, incorporating high flexibility of the GnRH backbone and GnRH1R sidechains, aimed to increase the accuracy and reliability of the predictions regarding the binding conformation between GnRH and GnRH1R. The intensive sampling of docking poses allowed for the exploration of potential binding modes, facilitating the identification of the most biologically relevant conformations.

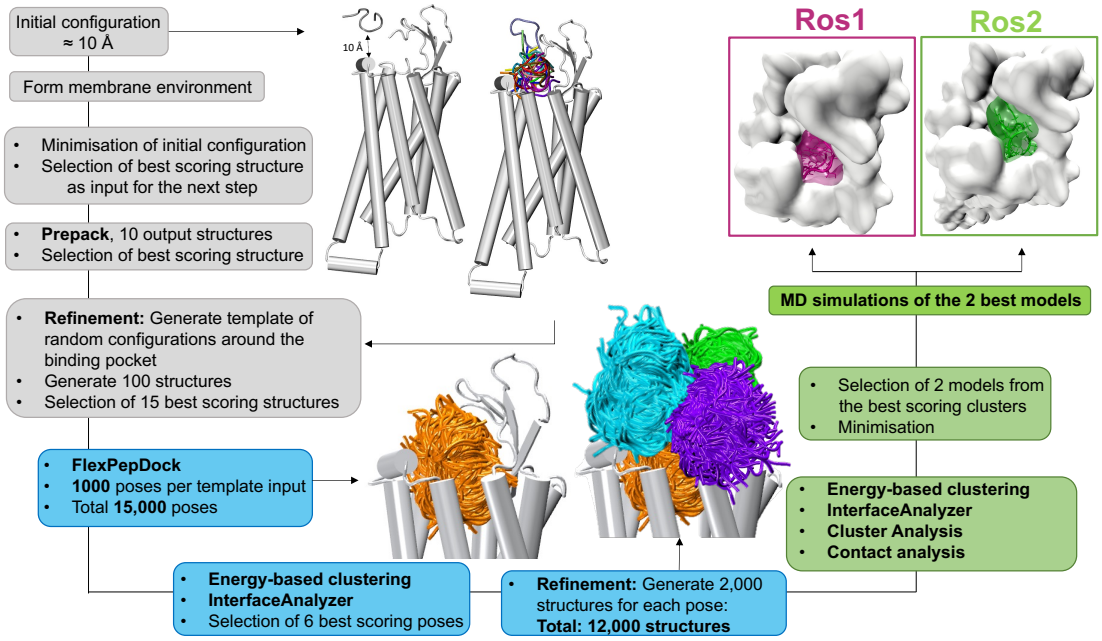


Figure 3.3: GnRH to GnRH1R docking pipeline: Stage 1 - Structure preparation and template formation (grey); Stage 2 - Initial docking and final refinement structure selection (blue); Stage 3 - Binding mode selection *via* cluster, contact, and hydrogen bond analysis (green). Optimal binding poses meeting criteria were chosen for MD simulations.

3.6.1 Input preparation

The step-by-step process for setting up the GnRH-GnRH1R system in a membrane environment is outlined as follows:

1. Initial configuration: The GnRH-GnRH1R starting configuration was established with the two entities positioned at 10 Å distance.
2. Membrane prediction: The membrane environment surrounding the GnRH1R was predicted using the `mp_span_from_pdb` application (https://www.rosettacommons.org/docs/latest/application_documentation/membrane_proteins/RosettaMP-App-MPSpanFromPDB) [116], [117] implemented with Rosetta using the following command:

```
/path/to/Rosetta/main/source/bin/mp_span_from_pdb.static.linuxgccrelease -in:file:s 10A_start_AP.pdb
```

This command executes the `mp_span_from_pdb` application and predicts the membrane environment around the GnRH1R based on the input structure provided (10A_start_AP.pdb). The successful prediction of the membrane environment around the GnRH1R is a crucial step for simulating the system in a biologically relevant context, especially when considering the influence of the membrane on protein conformation and interactions. The process for minimising, prepacking, and refining models using FlexPepDock in the context of the GnRH-GnRH1R system is detailed in the Appendix: Minimisation and Prepacking sections.

3.6.2 Formation of initial template

A total of 100 models were created and the 15 best scoring models were chosen as initial inputs for further refinement[118]. The template formation protocol is described in the Appendix: Template formation section. Here scores are derived from the scoring file as created using the protocol. The scoring file contains energy and statistical metrics for each model. This protocol refines 100 models using FlexPepDock, and the resulting 15 best-scoring models (yielded the lowest total energy scores) were selected as templates for further refinement. The use of multiple templates allows the exploration of various configurations around the binding pocket. Additionally, multiple template structures ensures that docking is not biased to a specific orientation[118]. This series of steps involving minimisation, prepacking, refinement, and template formation are crucial for generating accurate models of the GnRH-GnRH1R system, particularly in the context of flexible peptide docking and incorporation of the membrane environment.

3.6.3 Docking protocol

The provided script (Appendix, Docking protocol) outlines the execution of FlexPepDock for further refinement of the 15 template models. The script iterates over the specified list of input PDB files, performing 1,000 FlexPepDock simulations for each template model. The FlexPepDocking protocol is outlined in the Appendix: FlexPepDocking section. Here, each template input is used to produce 1,000 models. This process was repeated for all 15 template models to create a total of 15,000 models.

The six best scoring poses offering conformational variety around the binding pocket were selected to represent the second template. For this step, additional 2,000 docking poses were produced per template for a total of 12,000 poses by repeating the previous step. The second template consisted of the best scoring poses screened from a pool of 15,000 poses. This process not only limits the need for extensive amounts of docking poses but also allows for a more effective refinement of the best scoring configurations.

3.6.4 Energy-based clustering

The resulting 12,000 models were clustered based on Total Score measured in Rosetta Energy Units (REU) and Root Mean Square Deviation (RSMD) using the Rosetta energy-based clustering application[119] (https://www.rosettacommons.org/docs/latest/application_documentation/analysis/energy_based_clustering_application). This protocol is described in the Appendix: Clustering section.

The clustering step is crucial for identifying distinct energy and conformational clusters within the ensemble of models, helping to understand the diversity of the generated structures, and selection the best representative

clusters for further analysis. The energy-based clustering by Rosetta was used due to its reliability regardless of the number of structures and speed. Additionally, energy and RMSD biased clustering was highly desirable in this case, as it facilitates the search for the native-like binding mode more effectively compared to solely RMSD-biased clustering. The application implements an incremental clustering approach where entries are ordered by energy. More specifically, it selects the lowest-energy structure from the unclustered list as the centre of the current cluster and removes it from the unclustered list using a "cookie-cutter" algorithm. The application thereafter constructs an RMSD vector between the current cluster centre and all remaining structures in the unclustered list. The RMSD is based on the Cartesian coordinates. As a next step, the application selects all structures with RMSD values below the specified cutoff (1 Å) and groups the structures while also removing them from the unclustered list. This process is repeated until there are no structures remaining in the unclustered list. The application is fully deterministic, ensuring repeated runs on the same database produce the same output. The output files of this protocol are generated under the naming convention:

c.<cluster_number><struct_number>.pdb.

For example:

- c.1.1.pdb belongs to Cluster 1, pose 1.
- c.1.2.pdb belongs to Cluster 1, pose 2.
- c.2.1.pdb belongs to Cluster 2, pose 1.
- c.2.2.pdb belongs to Cluster 2, pose 2.

3.6.5 Interface analysis

The InterfaceAnalyzer application implemented with Rosetta, (https://www.rosettacommons.org/docs/latest/application_documentation/analysis/interface-analyzer) is a valuable tool for in-depth analysis the protein-peptide interface of the generated poses as it provides insights into binding energies, buried interface surface areas, and other important metrics[120]. The InterfaceAnalyzer protocol is provided in the Appendix: InteraceAnalyzer section.

The protocol calls a task operator to identify the interface and sequentially the mover to analyse the interface:

1) The RestrictToInterfaceVector Task operator. The Task operator is necessary for the identification of the interface based on the given criteria (https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/TaskOperations/taskoperations_pages/RestrictToInterfaceVectorOperation).

2) The InterfaceAnalyzerMover. This mover is responsible for the in-depth analysis of the found interface. (https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/Movers/movers_pages/analysis/InterfaceAnalyzerMover).

The output file contains metrics for all input structures (calculation of each metric detailed in Chapter 2 - 2.4.1).

The close-to-native binding mode is most likely to exist in a highly populated cluster with low Total Score and binding energy. Additionally, experimentally supported residue contacts (Table 1.3) were taken into consideration for the selection of the representative binding poses.

3.6.6 Cluster statistical analysis

Clusters were analysed based on metrics calculated with the InterfaceAnalyzer protocol. In this case, the Total Score (REU) represents the total energy of the complex. The ΔG binding (REU) represents the binding energy and is calculated after separation of the docking partners' chains and a repacking stage. Finally, the Solvent Accessible Solvent Area (ΔSASA) (Å^2) represents the solvent accessible surface area buried at the interface of the complex. These three metrics were used to analyse all clusters and select the best clusters to represent the system. Statistical analysis of the Total Score (REU), ΔG binding (REU), and ΔSASA (Å^2) interface metrics of the clusters was conducted using:

1. Distribution normality evaluation using the Kolmogorov-Smirnov statistical test[121], in conjunction with histograms and Quantile-Quantile plots (Q-Q) (Appendix, QQ_plots.py)
2. Mann-Whitney double-sided U tests[122], were conducted to eliminate statistically similar clusters, ultimately selecting the most representative clusters (Appendix, Best_clusters_pvalues.py).
3. Effect size calculations to determine the actual difference of the metrics distributions[123].

The Mann-Whitney tests incorporated a significance threshold of $p=0.05$, where the null hypothesis assumes no significant difference between the analysed metrics within the clusters. Subsequently, the resulting p-values underwent adjustment using the Bonferroni correction[124], [125]. This correction method accounts for multiple comparisons, ensuring a more stringent control over the family-wise error rate and reducing the likelihood of observing significant results by chance when conducting numerous tests. The results were visualised using the Matplotlib (https://matplotlib.org) and Seaborn (https://seaborn.pydata.org) libraries *via* customised Python scripts in the form of violin plots. Violin plots are a combination of traditional

histograms and kernel density plots and provide valuable insights of the distributions.

3.6.7 Cluster contact analysis

In addition to cluster statistical analysis, contact heatmaps were constructed to unravel the specific pair interactions between GnRH and GnRH1R within each cluster. These contact maps serve as detailed visualisations of the interacting residues, offering crucial insights into the key molecular contacts that govern the diverse binding modes observed in the clusters. The contact maps were constructed using customised Python scripts (Appendix: Find_contacts.py, Contact_map.py). Prior to contact analysis, the structure files were renumbered to represent the true residue numbers rather than the Rosetta required numbering version using customised Python scripts (Appendix, Renumber.py)

Parameters & advantages of contact analysis

- Contacts were calculated based on the Cβ-Cβ distances of GnRH-GnRH1R residues (Cα for glycine)[118], [120].
- The calculation involved only consideration of residue pairs within a 5 Å cutoff distance using customised Python scripts (Appendix: Find_contacts.py).
- Contact maps were presented in the form of heatmaps where different colours indicate population levels and signify the number of structures within the clusters that form the specific contact (Appendix, Contact_map.py).
- Contact maps for each cluster allows for a comparative analysis of interactions in different clusters.
- Variations in contact patterns across clusters provide valuable information about conformational diversity.

3.6.8 Hydrogen bond analysis

Similarly to cluster contact analysis, hydrogen bond (H-bond) analysis was performed for the successful clusters. H-bond analysis was performed using customised Python scripts with specifications of polar atom search (N, O with H), and 3.5 Å distance and 120^{o} angle cutoffs[126] (Appendix, Hbonds.py).

3.6.9 Binding mode selection for MD simulations

The overall goal of this process was to select the binding poses that best represent the native binding mode. Elimination rounds were implemented in various stages following clustering as detailed in Table 3.1.

Table 3.1: Elimination criteria in various stages of the analysis. Population: number of poses in the cluster. General contacts: any GnRH1R residue. Important contacts: experimentally shown GnRH1R binding pocket residues (Table 1.3, Figure 1.5).

Elimination criteria	Stage 1 (Clustering results)	Stage 2 (Cluster analysis)	Stage 3 (Contact analysis)	Stage 4 (Candidate selection)	Selection
Population	< 50	$< 1{,}000$	-	-	-
Total Score	-	$> -620\ REU$	-	-	-
ΔG Binding	-	$> 0\ REU$	-	$> -20\ REU$	Lowest
General contacts	-	-	< 1	-	-
Important contacts	-	-	-	< 1	Highest

Therefore, binding poses that satisfied the elimination criteria at any stage were discarded from further analysis. At the final stage of elimination, the poses were ranked based on ΔG binding (REU) and a higher number of important contacts (Table 1.3, Figure 1.5). Consequently, the successful candidate poses for MD simulations represent the GnRH-GnRH1R complexes with the lowest energy and the highest number of important contacts, signifying their potential as biologically relevant binding modes. Candidate selection and ranking was performed using customised Python scripts (Appendix, Rank_contacts.py, Candidate_selection.py, Candidate_rank.py).

3.7 MD simulations

For the investigation of both the active and inactive conformations of the GnRH1R, four distinct systems were constructed using the CHARMM-GUI webserver (https://charmm-gui.org). The active conformation simulations were designed to incorporate the GnRH peptide. A series of simulations was performed: one where the GnRH was allowed to circulate freely within the extracellular space (GnRH-GnRH1R, and two binding modes selected from computational docking simulations (ROS-1, ROS-2). The inactive conformation simulations were performed in the absence of the GnRH (Apo-GnRH1R). The preparation of these systems adhered to the established protocol for membrane proteins as outlined by previous methodologies[127], [128] (Figure 3.4). The system preparation protocol ensures a consistent and validated framework for the simulations, facilitating examination of the dynamic behaviour of GnRH1R. The resultant output files of this process encompass all requisite inputs essential for the execution of NAMD

simulations generated by CHARMM-GUI. The membrane environment consisted of 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) lipids.

While the use of POPC membranes is a standard approach in computational studies due to their well-characterised properties and ability to provide a stable environment, it is important to acknowledge that POPC membranes may not fully replicate the complexity of native cellular membranes. Native membranes are composed of a diverse mixture of lipids, proteins, and cholesterol, which can influence the behaviour and dynamics of embedded proteins in ways that are not entirely captured by POPC alone. However, POPC provides a simplified yet realistic model that balances computational efficiency with a reasonable approximation of a biological membrane's physical properties, making it a widely accepted choice for membrane protein simulations.

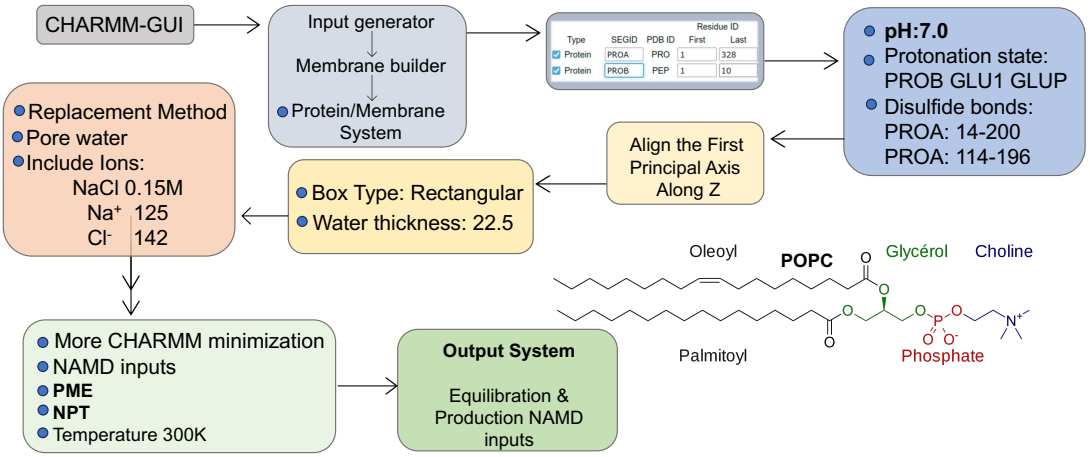


Figure 3.4: Input preparation pipeline using CHARMM-GUI. The depicted process concerns the GnRH-GnRH1R system. The Rosetta docked systems and the Apo-GnRH1R were prepared with the same main parameters and the absence of GnRH in the initial input file in the latter. All systems contained 150 POPC lipids per leaflet.

Following the procedure outlined in Figure 3.4, the output structures underwent a two-step post-processing stage involving the application of two scripts: Ctrbox.tcl and Fix_protein.tcl (provided in Appendix). The initial script (Ctrbox.tcl) was used to centre the GnRH1R-POPC system in the centre of the water box while the latter (Fix_protein.tcl) was used for the mobilisation of all atoms but those of GnRH-GnRH1R-POPC. The output files generated from Ctrbox.tcl, were used as representatives input coordinates and topology files, respectively. Lastly, the fix_protein.pdb was used in the first step of equilibration only, to immobilise all but water and ion atoms in the simulation.

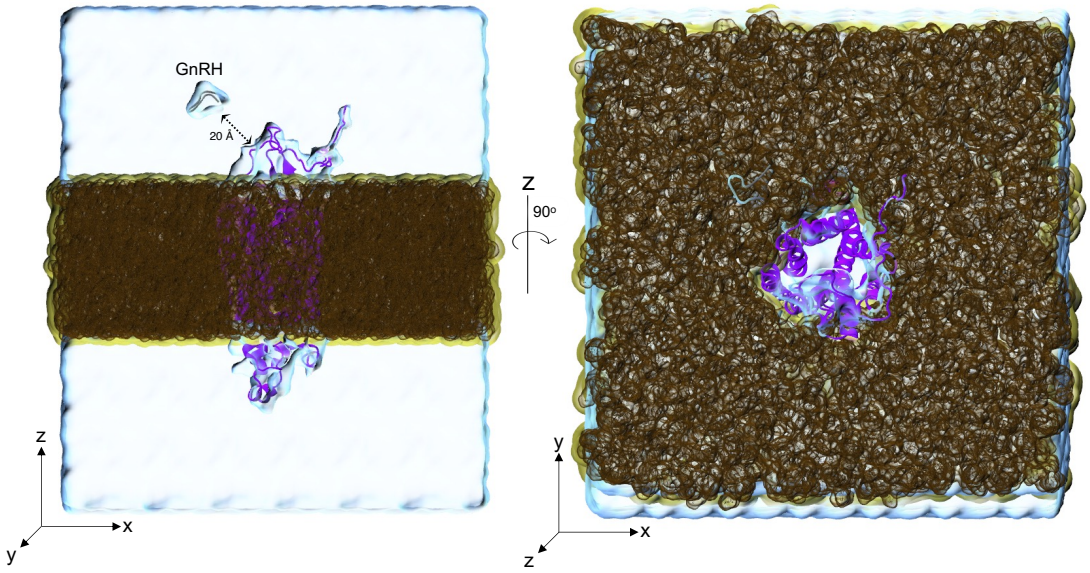


Figure 3.5: Representation of the GnRH-GnRH1R system for equilibration and subsequent production of MD simulations where GnRH is not docked. Left Panel: Depicts the side view of the system, featuring the GnRH1R (purple), POPC bilayer (brown), GnRH (grey) positioned 20 Å away from GnRH1R, and the surrounding water environment (cyan). Ions have been omitted for clarity. Right Panel: Illustrates the top view, highlighting the orientation of the GnRH1R within the lipid bilayer and emphasising the interhelical pore occupied exclusively by water molecules.

3.7.1 Systems Studied

The accurate representation of the native behaviour of a 7-TM receptor necessitates careful simulation. A total of four distinct systems were prepared for MD simulations, as outlined in Table 3.2. To bolster the accuracy of the findings, two independent replicas were performed for each system. This approach, involving multiple replicas for each system, was adopted to ensure the reproducibility and reliability of the results, thereby enhancing the scientific validity and credibility of the simulations.

Table 3.2: Details of the five systems used and the MD simulation details.

System	GnRH-GnRH1R	Ros-1	Ros-2	Apo-GnRH1R
Replicas	2	2	2	2
Time (μs)	1.1	1.1	1.1	1.1

The GnRH-GnRH1R system was studied to observe GnRH-induced-GnRH1R activation. In these simulations, the GnRH peptide was initially positioned 20 Å from GnRH1R and allowed to freely circulate. Conversely, the ROS-1 and ROS-2, systems represent the most optimally scored docking poses, derived from extensive FlexPepDock simulations conducted using the Rosetta software suite.

Finally, the Apo-GnRH1R system characterises the inactive conformation, as unveiled by MD simulations conducted in the absence of the activating GnRH peptide. The disparity between active and inactive conformations is delineated through an in-depth analysis of the structural and molecular differences observed in the two states. This comparative analysis aims to unravel the nuanced differences that underlie the conformational states of GnRH1R.

3.8 System Equilibration and Production of MD simulations

To achieve proper and stable MD simulations, eight equilibration trajectories were conducted with gradually reducing restraints. Following equilibration, the production simulation run was executed without any imposed restraints. The duration of the simulations for all studied systems was determined from the system that induced activation which yielded 1.1 μs of simulation time (Table 3.3).

Table 3.3: Detailed parameters for equilibration with slowly releasing restraints and production trajectories with no restraints.

Equilibration Step	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Force Constants for Harmonic Restraints	**Max**	**Max**	**Max**	**High**	**High**	**Moderate**	**Low**	**Min**
BB	10	10	10	5.0	2.5	1.0	0.5	0.1
SC	5.0	5.0	5.0	2.5	1.0	0.5	0.1	0.0
wforce	2.5	2.5	2.5	2.5	1.0	0.5	0.1	0.0
tforce	2.5	2.5	2.5	2.5	1.0	0.5	0.1	0.0
mforce	2.5	2.5	2.5	2.5	1.0	0.5	0.1	0.0
ion	10.0	10.0	10.0	0.0	0.0	0.0	0.0	0.0
fcis	250	250	250	100	50	50	25	0.0
fc2	250	250	250	100	50	50	25	0.0
Time (ps)	100	300	125	125	125	250	250	250
Ensemble	NPT	NVT	NPT	NPT	NPT	NPT	NPT	NPT
Timestep	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Production run	Ensemble: NPT Timestep: 2.0 Restraints: N/A Time: 1.1 μs							

BB: Backbone, **SC**: Sidechain, **wforce**: force constant to keep water molecules away from the hydrophobic core, **tforce**: force constant to keep the lipid tail below +/-%, **mforce**: force constant to keep the lipid head groups close to target values, **ion**: force constant applied on all ions in the system. **fcis** and **fc2**: lipid dihedral restraints were applied to keep the cis double bond and c2 chirality. Force constant values are presented in (Kcal/mol(Å)^2). (Appendix, inputs: D0.1.inp to D0.8.inp for equilibration and D1.production.inp for production runs).

Each equilibration cycle was conducted over a specified timeframe, employing subsequent cycles of NPT-NVT-NPT at a timestep of 1.0 ps. Step 1

of equilibration includes water-ion equilibration only followed by slow temperature increase and equilibration of the entire system in the NVT ensemble at the second step (Table 3.3).

Equilibration steps 3-8 apply slow restraint release in the NPT ensemble. Following the equilibration phase, the production runs were conducted in the NPT ensemble with a timestep of 2.0 ps and no restraints using the leap-frog algorithm. The Nosé-Hoover Langevin-piston and the Langevin temperature coupling were employed to control pressure and temperature in the NPT ensemble (Table 3.3).

The necessary input files to perform all equilibration and production of MD simulations are provided in the Appendix section (MD Equilibration and Production input files).

3.9 MD Simulation Analysis

Trajectory visualisation was performed with the Visual Molecular Dynamics software (VMD). Trajectory analysis was then conducted utilising VMD plugins alongside customised Python scripts tailored for specific analyses (Appendix, MD analysis). The incorporation of these tools allowed for the examination and characterisation of the dynamics and conformational changes exhibited by the simulated systems. Furthermore, all visual representations of the simulated systems depicted in the figures of this thesis were conducted using VMD and the Tachyon renderer in combination with the GPLS render mode. Assembly and schematic representation figures were constructed in Microsoft Power Point. ChemDraw 22.2.0 was used to construct chemical structures.

3.9.1 Salt bridge analysis

Generally, charge residues within a protein can form a special type of electrostatic interactions between oppositely charged amino acid sidechains called salt bridges. The charged amino acids that constitute a salt bridge are lysine, arginine with aspartic acid, glutamic acid (Figure 3.6F). Salt Bridges (SBs) within the simulated systems were calculated utilising the VMD plugin 'salt bridges' under a 3.5 Å N-O cutoff distance. Plots of the SB evolution during simulation of all systems were performed using customised Python scripts (Appendix, SaltBridge.py). The script not only provides a clear visual representation that facilitates the tracking and monitoring of active/inactive SBs throughout the simulation but also provides the absolute time of duration of each SB. This approach contributes to a nuanced understanding of the role of SBs in the conformational dynamics of the simulated systems.

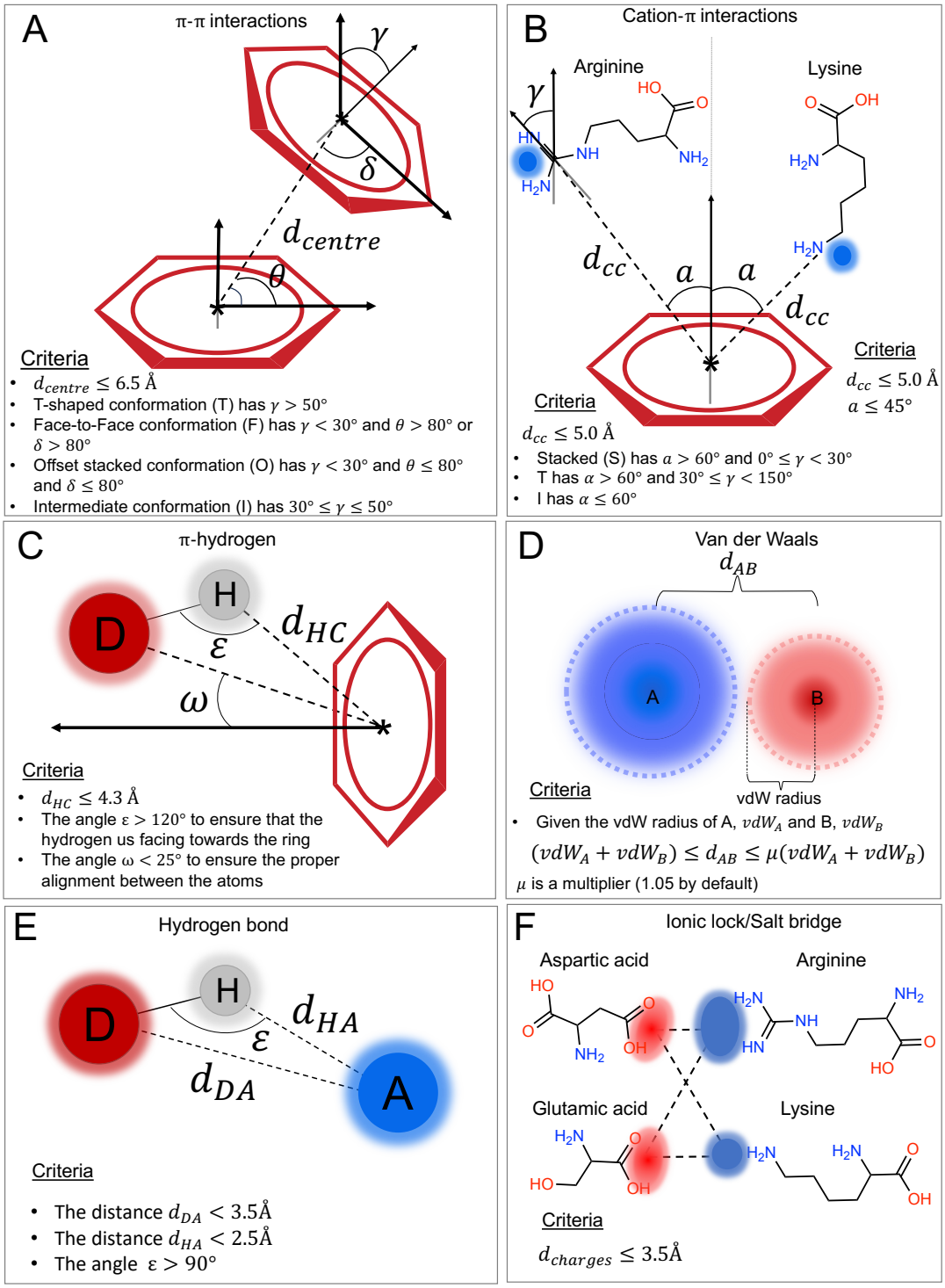


Figure 3.6: Examples on interactions calculated using RING. A) π-π, B) Cation-π interactions, C) π-hydrogen interactions, D) vdW interactions, E) hydrogen bonds and F) Ionic locks (SBs). In the schematics, the letter D corresponds to electromagnetic (nitrogen, oxygen) donor atoms and A to electromagnetic acceptor atoms.

3.9.2 Hydrogen bond analysis

Hydrogen bonds (H-bonds) within the simulated systems were analysed using the VMD plugin 'Hydrogen Bonds'. This method included the calculation of H-bonds for all components in the systems, including those formed between GnRH1R and the POPC membrane and GnRH. Parameters of 3.5 Å cutoff distance and 20° Donor-Hydrogen-Acceptor angle cutoff were utilised. H-bonds were assessed based on their occupancy throughout the simulations. The investigation was particularly focused on H-bonds formed between GnRH and GnRH1R residues, warranting an in-depth analysis of the molecular interactions. The reporting order of Hydrogen bonding pairs adhered to the standard Donor-Acceptor convention (Figure 3.6E).

3.9.3 Analysis of interactions

In addition to salt bridge and hydrogen bond, π-π stacking, cation-π, π-hydrogen, and vdW interactions were calculated using RING webserver (https://ring.biocomputingup.it). For this analysis, the strict cutoff thresholds option was applied (Figure 3.6).

3.9.4 Root Mean Square Deviation calculations (RMSD)

All RMSD calculations were executed using the VMD plugin 'RMSD Trajectory Tool'. The procedure involved backbone alignment of the GnRH1R to the equilibrated structure as reference (Frame 0). RMSD was plotted using customised Python scripts (Appendix, RMSD.py). To ensure that subsequent trajectory analysis focused on the backbone equilibrated GnRH1R, the time required for backbone relaxation was excluded from analysis of all systems. Specifically, the first 150-250 ns of the trajectories were discarded based on the above criteria. The RMSD was calculated based on the equation:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}\left|\vec{r}_{i,traj} - \vec{r}_{i,ref}\right|^2}{N}} \quad \textbf{[3-1]}$$

where N is the number of backbone atoms in the GnRH1R, $\vec{r}_{i,traj}$ and $\vec{r}_{i,ref}$ is the position of the $i-th$ atom in the trajectory and reference frame, respectively.

3.9.5 Root Mean Square Fluctuation calculations (RMSF)

RMSF calculations were performed with customised Python scripts and the mdtraj library (Appendix, RMSF.py, Plot_RMSF.py). The calculations were conducted only for backbone atoms. This approach allowed for the assessment of the fluctuations exhibited by backbone residues, enabling the identification of unique conformation dependent differences. RMSF and RMSD calculations follow the same principle, where the only difference is that

RMSF calculations divide the difference between the reference and trajectory frame per residue in the molecule.

3.9.6 Principal Component Analysis (PCA)

PCA is a powerful statistical method employed in MD simulations to extract essential information regarding the dominant motions and fluctuations within a complex system. PCA is a dimensionality reduction method used to simplify a large data set into meaningful patterns or trends.

PCA was applied to the extracted GnRH coordinates from the MD simulation. PCA is a dimensionality reduction technique that identifies the principal components (PCs) or directions of maximum variance in the data. In this analysis, PCA is performed with two components to reduce the dimensionality of the peptide coordinates.

Kernel Density Estimation (KDE) is a statistical technique used to estimate the probability density function of a continuous random variable. In this step, a 2D KDE is created using the PCA-transformed coordinates (PCA 1 and PCA 2) of the peptide. This KDE represents the density of points in the reduced 2D space. The KDE is evaluated on a grid of points that spans the range of PCA 1 and PCA 2 values from the data. This step creates a density map in the reduced 2D space, where each point on the grid represents the estimated population density of data points in the original MD trajectory.

Finally, results of PCA are represented in a density plot as heatmaps. The x-axis and y-axis represent the values of PCA 1 and PCA 2, respectively. The colour intensity at each point on the plot represents the population density, with higher intensity indicating a higher density of data points in that region of the 2D space. Application of PCA, offer great insights into the GnRH binding. PCA was conducted through customised python scripts (Appendix, PCA.py) and the mdtraj, scikit-learn and SciPy libraries.

3.9.7 Ramachandran analysis

Ramachandran plots were generated for all systems using the customised Python scripts and the MDanalysis library (Appendix, Ramachadran.py). The Ramachandran plot is a pivotal tool in structural biology and focuses on the Phi (φ) and Psi (ψ) dihedral angles of residues. The Phi angle describes the rotation around the Cα-N bond whereas the Psi angle describes the rotation around the Cα-C=O bond. The plot provides the distribution of these angles in the protein structure and categorises conformations into regions where amino acids are likely to adopt energetically favourable configurations. These regions are often referred to as “allowed regions” and represent the dominant conformations that proteins typically adopt. Areas with partial allowance for flexibility and regions indicating “disallowed” or uncommon conformations are also identified on the plot.

Deviations from the allowed regions may suggest structural irregularities, such as steric clashes or incorrect assignments in the model. In this analysis, the overall conformational quality of GnRH and GnRH1R were assessed.

Conclusions

This chapter presents the framework for the investigation of the GnRH-GnRH1R complex through computational docking and MD simulations. Detailed protocols and application of each method were described alongside instructions for more challenging techniques. The methodology described here is applied to the following chapters where the GnRH-GnRH1R system is investigated in detail.

Key Points:

1) Rosetta docking explores binding modes: Computational docking simulations facilitated the search of the close-to-native GnRH-GnRH1R binding modes. Energy and RMSD clustering of thousands of binding modes, in addition to statistical analysis of cluster energy metrics and contact analysis, allowed the selection of binding modes through numerous eliminations stages.
2) Docking evolution through MD simulations: The selection of optimal docking modes paved the way for MD simulations, allowing the validation the predicted binding modes and their effect on the GnRH1R conformation.
3) MD simulations unveil conformational dynamics: MD simulations provided invaluable insights into the dynamic transitions between the active and inactive states of GnRH1R.
4) Conformation characterisation: The application of a range of analysis techniques, including Salt Bridges, Hydrogen Bonds, RMSD/RMSF calculations and PCA enabled in-depth characterisation of both active and inactive conformations.

Chapter 4 - Computational docking

4.1 Rationale

The identification of the binding mode between GnRH and GnRH1R is crucial for understanding the functionality of key residues that contribute to receptor activation.

Prior computational studies have proposed various GnRH-GnRH1R interactions such as a SB between R8 of GnRH and D302$^{7.31}$ of the GnRH1R. However subsequent studies indicated this interaction is not essential for activation. To mitigate bias from pre-existing computational studies and assumptions, completely unrestricted computational docking simulations were implemented herein, with binding mode selection based solely on experimental mutagenesis data.

This chapter aims to identify the native-like binding mode of the GnRH-GnRH1R complex utilising computational docking simulations with Rosetta. An extensive series of 27,000 dockings were conducted employing the FlexPepDocking protocol. The resulting docking poses were clustered using the energy-based clustering protocol integrated with Rosetta. Through cluster analysis, successive structure elimination rounds and contact analysis, two binding modes were identified for further examination through MD simulations. The selection of these successful models was grounded in a dual-criteria approach: first, the energy metrics of the docking poses, and second, the presence of important contacts as delineated by mutation experiments that characterise the GnRH1R binding pocket.

The outcomes of this chapter apply a methodology for the prediction and selection of the native-like binding mode of the GnRH-GnRH1R complex. This approach not only enhances current understanding of the molecular interactions within this complex but also sets a precedent for similar studies focused on receptor-ligand interactions.

4.2 Aims

The primary aim of this chapter is to ascertain the native-like binding modes of the GnRH-GnRH1R complex. The successfully identified binding modes will serve as the initial structures for MD simulations, intended to investigate the active conformation of the GnRH1R (presented in Chapter 5 and Chapter 6). More specifically, this chapter aims to:

1. Assess and compare the binding characteristics of GnRH-GnRH1R complexes through Rosetta FlexPepDock simulations to identify the close-to-native binding mode.

- Conduct extensive computational docking simulations to model the interaction between GnRH and GnRH1R.
- Perform a series of 27,000 FlexPepDock dockings to capture a wide array of potential binding modes, thereby ensuring a thorough exploration of the binding pocket.

2. Cluster and analysis of the dockings.
    - Implement energy-based clustering protocols to organise the docking results into meaningful categories (clusters).
    - Evaluate and compare the clusters using statistical tests, including Mann-Whitney U tests and effect size calculations, to discern significant differences in cluster energy metrics.
3. Perform contact analysis to uncover specific binding characteristics and determine the biological relevance of the identified structures within the clusters.
4. Application of elimination rounds to narrow down the pool of potential binding modes.
5. Select and validate potential close-to-native binding modes.
    - Select the promising binding modes for further investigation based on specific criteria including energy metrics and the presence of key contacts as identified from mutation experiments.

By achieving these objectives, this chapter aims to distinguish biologically meaningful binding modes from false positives and contribute to the development of more accurate peptide docking strategies. The successful binding modes will undergo further examination through MD simulations.

4.3 Introduction to computational docking

Computational docking has a pivotal role in molecular biology and drug discovery, offering valuable insights into biomolecular interactions between proteins and ligands[129], [130], [131], [132]. It is instrumental in predicting binding modes and affinity, contributing to the understanding of molecular recognition, and the acceleration of drug development[133].

In this research, computational docking is a key element in the search of the native-like binding mode between GnRH and GnRH1R. The choice of Rosetta commons docking suite is motivated by its reputation for accuracy in predicting protein-ligand interactions[134], [135]. Rosetta utilises a sophisticated energy-based approach to explore the conformational space of biomolecular complexes, facilitating the prediction of favourable binding modes[136], [137], [138].

By combining computational docking with MD simulations, this research aims to achieve a comprehensive and accurate investigation of the GnRH-GnRH1R interaction and activation process[139], [140]. Computational docking complements this by predicting energetically favourable binding poses, while MD simulations provide dynamic insights of the binding mode over time. Integrating these approaches allows for a more holistic understanding of the GnRH-GnRH1R interaction, encompassing both static and dynamic aspects. The generation of 27,000 docking poses adds depth, enabling a thorough exploration of conformational space and increasing the chances of identifying the native-like binding mode[140], [141].

4.4 Methodology

The detailed protocols for the preparation, production, and selection of binding modes are described in Chapter 3 -3.6 and Appendix of the thesis. Briefly, two GnRH-GnRH1R binding modes were selected though a pool of 27,000 docking poses generated with FlexPepDocking, namely ROS-1 and ROS-2. The docking poses originated from two groups of template structures: one comprised of 15 GnRH positions, generating 15,000 poses (1,000/template), and the other of 6 GnRH positions, producing 12,000 poses (2,000/template). The aim of these templates was to ensure unbiased docking, avoiding any undue preference to specific interactions around the binding pocket. The second template was constructed from the six best scoring models from the first template based on binding energy, ensuring the likelihood of favourable dockings. The resulting poses were clustered based on energy and RMSD and were subjected through various elimination stages to ensure selection of biologically relevant binding modes. Elimination rounds were guided through cluster populations, statistical analysis of energy metrics, contact analysis, the presence of important contacts and binding energy, (Chapter 3 -3.6.4-3.6.9). The binding modes selected for MD simulations were chosen based on the highest number of important contacts and lowest binding energies. The GnRH centre of mass in regards to the GnRH1R binding pocket for each cluster and the ranked binding modes were plotted using customised Python scripts (Appendix, Cluster_CoM.py, Candidate_rank.py)

4.5 Results

4.5.1 Cluster representation

The 12,000 final docking poses underwent energy-based clustering and yielded a total of 84 clusters, with populations ranging from 2,022 to 47 structures. From these clusters, only those with populations greater than 50 were chosen for subsequent analysis. The 17 selected clusters, with populations ranging from 2,022 to 52, underwent cluster analysis, and further

exclusion of structures with a Total Score greater than -620 REU and a ΔG binding greater than 0 REU was performed Violin plots for the Total Score, ΔG binding, and ΔSASA of the interface were constructed for these clusters (Figure 4.1).

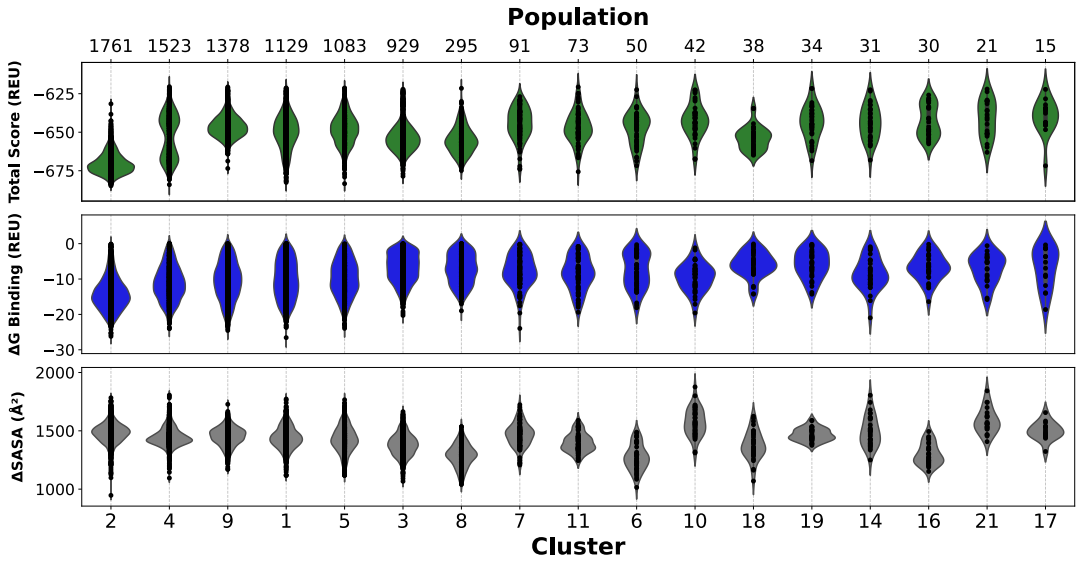


Figure 4.1: Energy metrics of the 17 most populated clusters: Total Rosetta score (REU) (green), ΔG Binding (REU) (blue), and ΔSASA (Å²) of the binding interface (grey) distributions in violin plots. Cluster identifiers and populations are provided on the lower and upper x-axes.

Table 4.1: Cluster population before and after applying elimination criteria. Structures with Total Score values > -620 REU and ΔG binding > 0 were eliminated.

Cluster name	Population before elimination	Population after elimination	Structures eliminated
2	2022	1761	261
4	1809	1523	286
9	1787	1378	409
1	1763	1129	634
5	1528	1083	445
3	1573	929	644
8	507	295	212
7	190	91	99
11	109	73	36
6	81	50	31
10	73	42	31
18	63	38	25
14	66	34	32
19	107	31	76
16	67	30	37
21	54	21	33
17	52	15	37

Clustering was conducted based on Total Score metrics, with clusters consisting of conformations sharing similar RMSD values. Due to the stochastic nature of Rosetta, it is possible for a cluster to include conformations that are structurally similar but exhibit a wide range of Total Scores. In such situations, priority is given to the clusters with the lowest Total Scores, as they are more indicative of a physically plausible state. Higher-energy clusters may be trapped in local minima, which a physical receptor would typically easily overcome.

In the subsequent phase of elimination, clusters with populations of fewer than 1,000 structures were excluded from further consideration. Consequently, clusters 2, 4, 9, 1, and 5, with populations of 1,761, 1,523, 1,378, 1,129, and 1,083 respectively, were taken into consideration (Figure 4.1,Table 4.1). However, despite the significant size of Cluster 9, its low ΔG Binding values did not result in a corresponding reduction in Total Score (Figure 4.1). Conversely, Cluster 9 displayed the highest Total Score values at approximately -650 REU, whereas the remaining clusters had values closer to -680 REU. Since all clusters exhibited similar lower ΔG Binding values, it would be expected that favourable conformations within Cluster 9 would also have a lower Total Score. This inconsistency suggests that the conformations within Cluster 9, despite their abundance, may not represent energetically favourable states. Therefore, Cluster 9 was considered unreliable for containing biologically relevant conformations and was excluded from further analysis. The final clusters 2, 4, 1, and 5 were then subjected to a statistical analysis.

4.5.2 Statistical analysis of clusters

Prior to conducting statistical analysis on the chosen clusters, evaluation of the data distributions was deemed necessary for the application of appropriate statistical tests. As a result, the Kolmogorov-Smirnov (K-S) statistical test, in conjunction with histograms and Quantile-Quantile (Q-Q) plots of each metric, was carried out to gain an understanding of the data distributions for all selected clusters (Figure 4.2). These tests offer insights into whether the data deviate or adhere to a normal Gaussian distribution. Histograms provide a visual representation of the frequency distribution of each variable and the overall shape of the distributions. Q-Q plots allow for a comparison of the distribution of metrics with a theoretical normal distribution. The K-S test determines whether the distribution follows normality or not.

$$D_n = sup_x|F_n(x) - F(x)| \qquad [4\text{-}1]$$

where D_n represents the K-S statistic, $F_n(x)$, is the empirical distribution function of the sample, and $F(x)$ is the cumulative distribution function of the

compared distribution (in this case, the normal distribution). The K-S statistic, along with the p-value, is employed to assess the hypothesis that the data in each case conform to a normal distribution. In this context, if $p > 0.05$, then the data is considered to follow a normal distribution; otherwise, it is considered not to follow a normal distribution. The outcomes of this analysis offer insights into whether parametric or nonparametric statistical tests can be applied to the cluster metrics, as parametric tests can only be applied to normal distributions. The p-value of the K-S test was computed as an approximation for large data sets in the following manner:

$$p \approx 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2D^2n} \quad [4\text{-}2]$$

where 1 represents the initial point of the calculation and signifies the total probability prior to subtracting the calculated probability of the observed test statistic. The term $(-1)^{k-1}$ ensures the convergence of the series for both positive and negative values, corresponding to even and odd k values, respectively.

The K-S statistic, denoted as D, stands for the maximum distance between the empirical cumulative distribution function of the sample and the theoretical distribution. Multiplying D by the sample size, n, holds significance as larger sample sizes enhance the ability of the test to detect disparities between the sample distribution and the theoretical distribution. As the K-S test demonstrates that half of the metric distributions conform to normality, while the other half deviate from normality (Figure 4.2,Table 4.2), the choice of non-parametric statistical tests for further analysis was made.

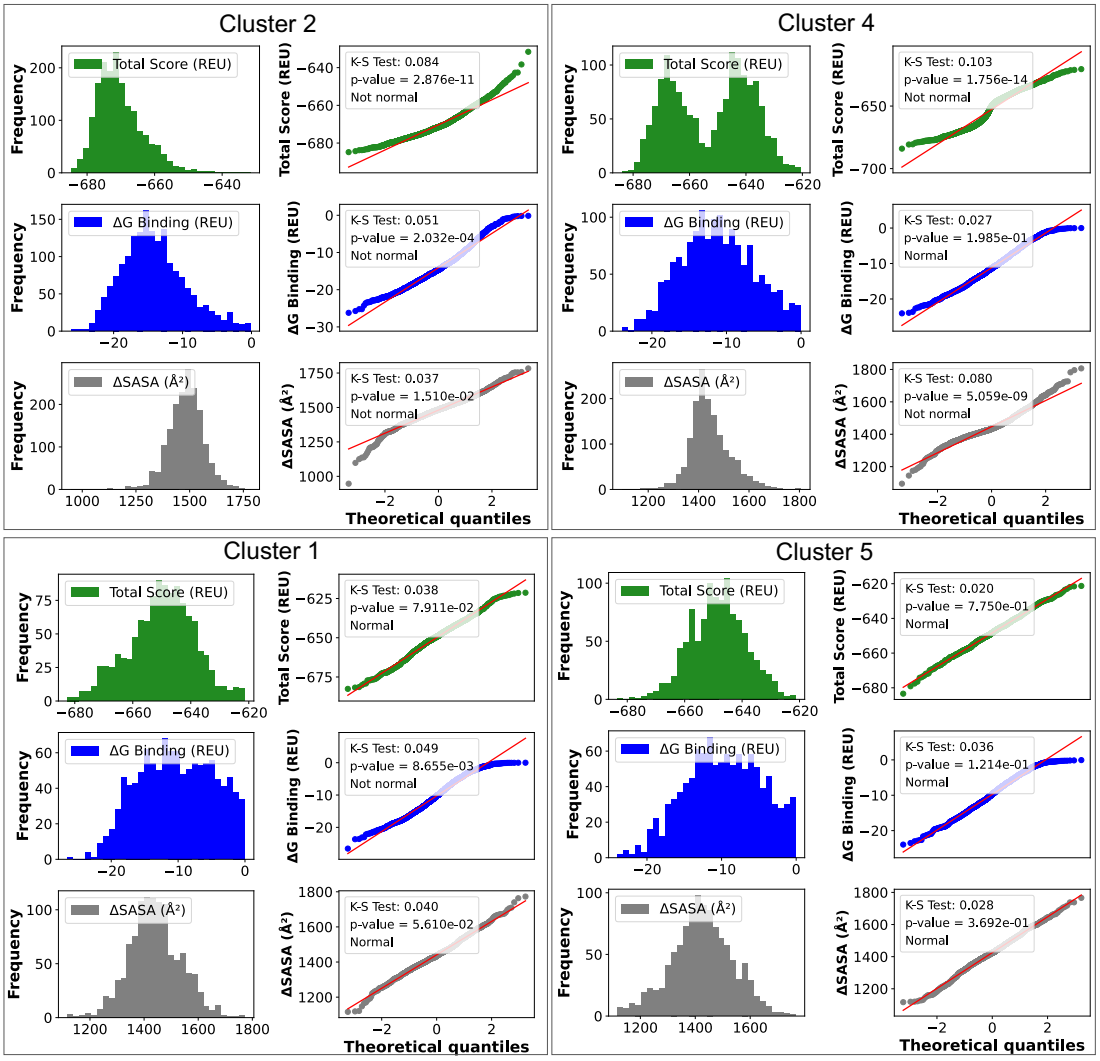


Figure 4.2: Histograms (left) and Q-Q plots (right) for Total Score (REU) (green), ΔG Binding (REU) (blue), and ΔSASA (Å²) (grey) in Clusters 2, 4, 1, and 5. Legend box includes p-values against the theoretical Gaussian distribution (red line) and the K-S coefficient. Distribution normality test results are also shown in each legend.

Table 4.2: K-S test results for all tested metrics in Clusters 2, 4, 1, and 5.

Cluster	Total Score (REU)	ΔG Binding	ΔSASA (Å²)
Cluster 2	Not normal	Not normal	Not normal
Cluster 4	Not normal	Normal	Not normal
Cluster 1	Normal	Not normal	Normal
Cluster 5	Normal	Normal	Normal

This preference arises from the inherent flexibility of non-parametric approaches in accommodating mixed distribution types, their robustness against data peculiarities such as skewness and outliers, and their independence from the stringent assumptions that underlie parametric tests. Especially in the upcoming cluster statistical analysis, the objective is to identify and categorise inherent structures within multidimensional data. This

enhances the validity and generalisability of the findings and embodies a methodologically sound approach in situations characterised by heterogeneous data distributions.

4.5.3 Cluster statistical analysis

The selected clusters were presented in the form of violin plots, which offer a valuable means of visualising distinct behavioural patterns for each evaluated metric. Non-parametric double-sided Mann-Whitney U tests were employed to evaluate each pairwise combination of the four clusters (Figure 4.3). The null hypothesis assumed was no significant difference between the distributions of the two clusters in comparison. The test involves ranking all the data points from both groups together and then analysing the sum of ranks in each group. A significant result (p-value < 0.05) indicates that the likelihood of observing the difference in ranks by random chance is low, suggesting a genuine difference of the data distributions. Conversely, p-values > 0.05 suggest similar values for the groups. The U values of the Mann-Whitney double-sided test are calculated as follows:

$$U_X = n_X \cdot n_Y + \frac{n_X(n_X + 1)}{2} - R_X \quad [4\text{-}3]$$

where U_X represents the Mann-Whitney U statistic for sample X, n_X is the number of observations in sample X, n_Y is the number of observations in sample Y, and R_X is the sum of the ranks of the observations in sample X when all observations from both samples X and Y are ranked together. The product $n_X \cdot n_Y$ represents the total number of possible pairings between an observation in sample X and an observation in sample Y. Each pairing contributes to the overall ranking comparison. The term $n_X(n_X + 1)/2$ calculates the sum of the ranks that would be assigned to sample X based on the formula for the sum of the first n natural numbers, which is $n_X(n_X + 1)/2$. The subtraction of R_X from this term accounts for the actual ranks obtained by sample X. Essentially U quantifies the extent of rank dominance of one sample over the other. A small U value indicates that most observations in sample X tend to have lower ranks (are smaller) compared to those in sample Y. Conversely, a large U value suggests that observations in sample X tend to have higher ranks (are larger). The Mann-Whitney U test employs this statistic to determine whether there is a statistically significant difference between the two independent samples in terms of their rank ordering. The U_Y statistic is calculated seperatly in a similar way using the modified version of [4-3] for sample Y and the smaller U value is used for the p-value calculation. In cases with large sample sizes, the U statistic approximates a normal distribution, and the z-score is computed as follows:

$$z = \frac{U - \mu_U}{\sigma_U} \quad [4\text{-}4]$$

where, $\mu_U = n_1 n_2/2$ is the mean and $\sigma_U = \sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}$, is the standard deviation of the U distribution.

The p-value is computed based on the minimum U statistic and it indicates whether the observed difference in rank ordering between the two samples is statistically significant. For a double-sided test, the p-value is computed using the z-score as follows:

$$p = 2 \cdot P(Z > |z|) \quad [4\text{-}5]$$

This calculation places particular emphasis on both tails of the distribution by determining the probability of a z-score greater than the absolute value of the calculated z-score. The $P(Z > |z|)$ part of the equation involves the standard normal random distribution (Z) and represents the probability (P) that a standard normal random variable Z is greater or equal to the absolute value of the test statistic $|z|$. The multiplication by 2 is performed to facilitate the double-sided test.

Furthermore, to account for multiple comparisons, the Bonferroni correction was applied to adjust the p-values obtained from the pairwise Mann-Whitney U tests. The Bonferroni correction is a method used to control the family-wise error rate (FWER), which represents the probability of making at least one Type I error (false positive) among all the conducted tests. The Bonferroni correction is calculated as follows:

$$p - value(Bonferroni) = p - value \cdot N \quad [4\text{-}6]$$

where N is the number of comparison and is calculated as:

$$N = \frac{N_c \cdot (N_c - 1)}{2} \quad [4\text{-}7]$$

where N_c is the number of clusters being compared, and the division by 2 is used due to the double-sided comparison. This correction is employed to reduce the likelihood of obtaining a significant result by chance when conducting multiple tests. It is essential to note that while p-values indicate the presence or absence of an effect, they do not quantify the size of the effect. Therefore, additional statistical measures, such as Effect Size (ES), are reported to measure the magnitude of differences observed between the clusters. In the context of the Mann-Whitney U test, ES is a metric that calculates the magnitude of difference between two groups. It is computed using the formula:

$$ES = 1 - \frac{2U}{n_1 n_2} \quad [4\text{-}8]$$

where U is the Mann-Whitney U statistic, and n_1 and n_2 are the sample sizes of the two groups. This ES value signifies the extent of overlap between the

distributions of the two groups, offering a standardised measure of the magnitude of the difference. An ES value nearing 0 indicates a substantial overlap, suggesting a small or negligible effect (similar distributions). Conversely, a value closer to 1 implies less overlap, signifying a larger effect (different distributions).

In this analysis, the Total Score, ΔG binding, and ΔSASA of the interface among the most populated clusters were compared. The objective of this analysis is to uncover similarities and differences in these metrics among the clusters, leading to a deeper understanding of their representative binding modes. Specifically, all clusters exhibited statistically significant differences in their Total Score values, as indicated by p-values < 0.05 (Figure 4.3). The most significant differences were observed in the comparisons of Cluster 2 with the three other clusters, whereas smaller differences were observed between pairs of Clusters 4, 1, and 5. This disparity is also evident in the significant difference in the median Total Score value, with Cluster 2 at approximately -670 REU and Clusters 4, 1, and 5 at approximately -650 REU. This difference is also reflected by the large ES values. Conversely, although statistically different, Clusters 4, 1, and 5 displayed more subtle differences in the Total Score, with small or moderate ES values. Intriguingly, Cluster 4 was composed of two distinct groups of energy 'subclusters', yielding Total Scores of approximately -640 and approximately -665 REU (Figure 4.3)

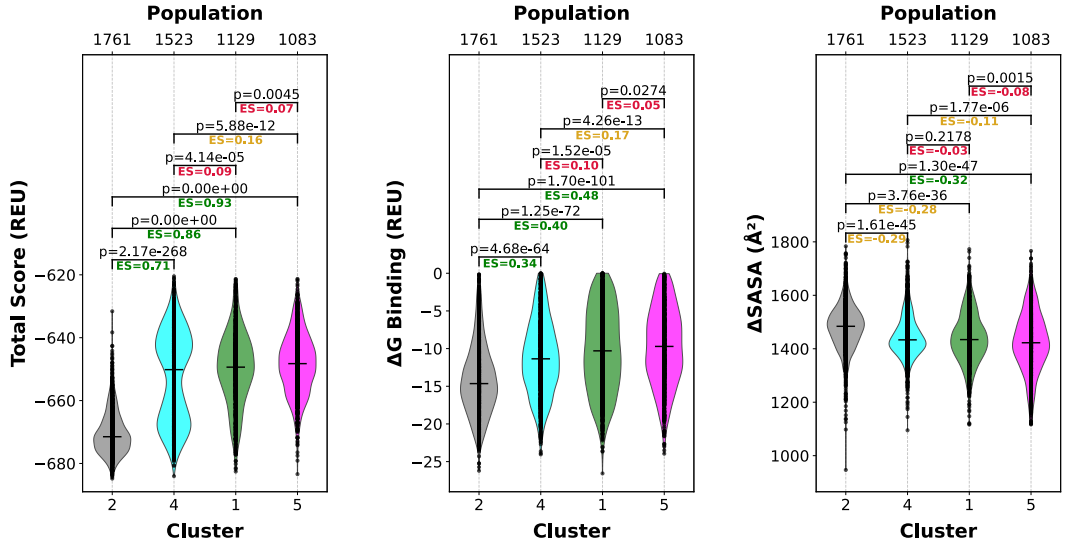


Figure 4.3: Chosen Clusters for in-depth cluster analysis: A) Total Score (REU), B) ΔG of binding (REU), and C) ΔSASA buried at the interface of selected clusters. Population of each cluster post-elimination is shown on the upper x-axis. P-values calculated using double-sided Mann-Whitney U test with Bonferroni correction. ES is indicated as large (green), moderate (yellow), and small (red) effects.

In terms of ΔG binding, Cluster 2 exhibits a significant median value difference of approximately -15 REU, while Clusters 4, 1, and 5 have values ranging from -12 to -10 REU. All clusters exhibit significant differences in

terms of ΔG binding except for the comparison of Clusters 1 and 5 ($p > 0.05$) (Figure 4.3). This suggests that Clusters 1 and 5 include similar binding poses in terms of energy. However, as the ΔSASA comparison is significantly different ($p < 0.05$), these clusters must exhibit significant differences in their conformations. The decrease in the ΔSASA of Cluster 5 compared to Cluster 1, indicated by the small negative decrease (-0.08) in ES, suggests that conformations within Cluster 5 induce a stronger conformational change than those in Cluster 1. However, both clusters are similar in terms of ΔSASA and ΔG binding compared to the other compared clusters. Furthermore, small and moderate ES values between Clusters 4-1 and 4-5 indicate that the dual population of Cluster 4 (as indicated in the Total Score) yields conformations with similar ΔG binding to either Cluster 1 or 5, while more similar to those of Cluster 1, as indicated by the small values of their ΔSASA ES (-0.03) (Figure 4.3).

It is intriguing that despite the extremely low values of Cluster 2 in both Total Score and ΔG binding, the median ΔSASA values are increased compared to the rest of the clusters. This anomaly may serve as an indication that GnRH conformations within Cluster 2 result in highly favourable binding energies due to possible interference with the N-terminus, possibly indicating a false-positive docking pool[142]. Since a decrease in ΔSASA values typically indicates a significant conformational change, the overall increase observed in Cluster 2 may suggest the existence of biologically irrelevant conformations. The lowest ΔSASA values are observed in conformations within Cluster 5, followed by Clusters 1 and 4. Further analysis of the specific contacts between GnRH and GnRH1R are essential to gain a deeper understanding of the nature of these differences (Figure 4.3)

4.5.4 Contact & Hydrogen Bond analysis

Contact analysis was conducted for the selected clusters to extract specific characteristics of the various binding modes. In this analysis, only structures that exhibited at least one contact between GnRH and GnRH1R were considered (Figure 4.4). Contacts were identified based on a 5 Å cutoff distance between Cβ carbons of the peptide-receptor complex (Cα for glycine).

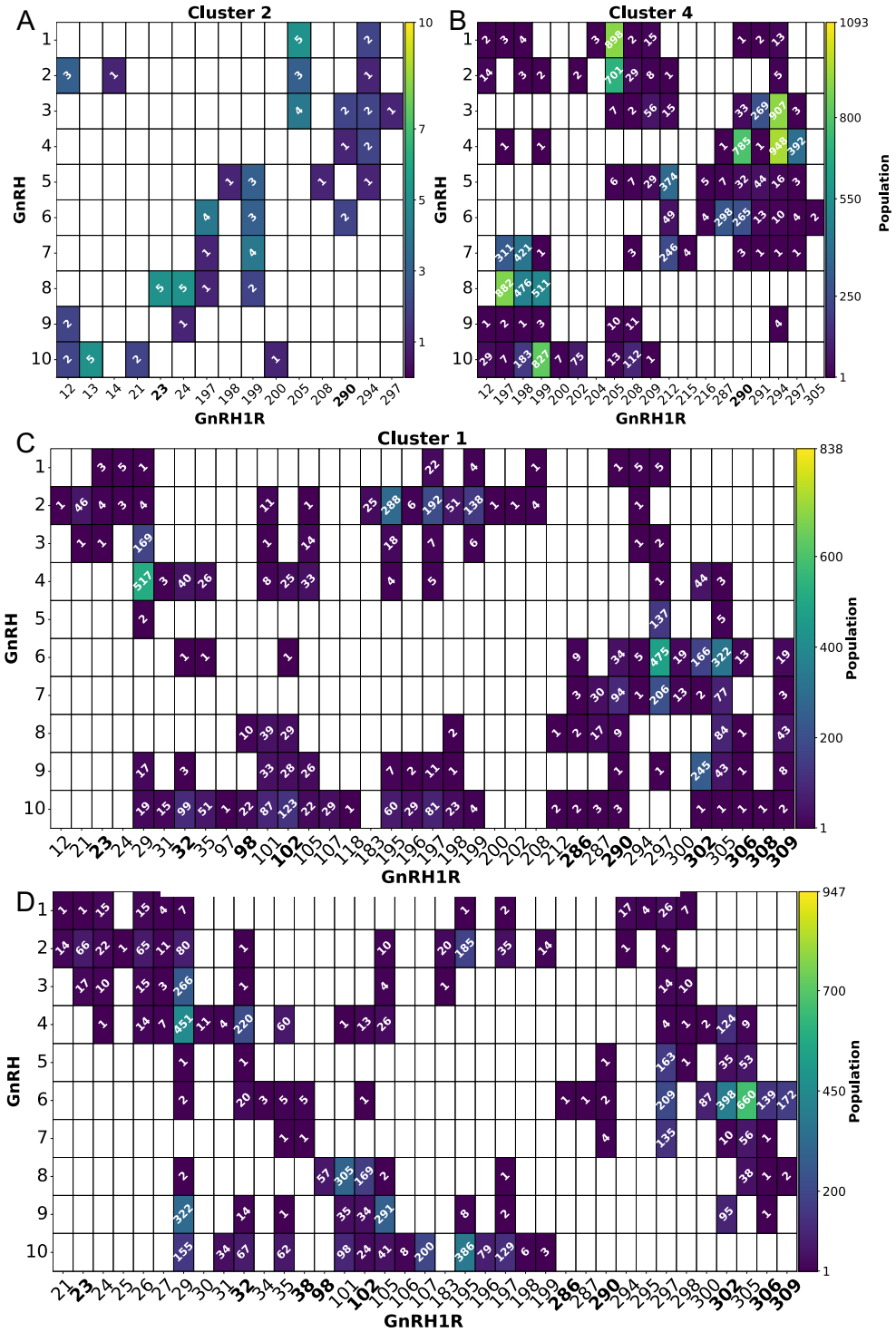


Figure 4.4: Contact analysis of Clusters 2 (A), 4 (B), 1 (C), and 5 (D) in the form of heatmaps. The Population bar represents the number of structures that display the specific contact in a color-coded manner: high (yellow) and low (blue) Population. In bold font, experimentally proven important GnRH-interacting residues are highlighted.

This elimination criterion led to a substantial reduction in the cluster populations, as numerous structures did not display any contacts and were thus excluded from further consideration (Table 4.3).

Table 4.3: Cluster population before and after the second elimination round, applying criteria requiring at least one contact.

Cluster number	Population before 2nd elimination	Population after 2nd elimination	Structures eliminated
2	1761	10	1751
4	1523	1093	430
1	1129	838	291
5	1083	947	136

It is noteworthy that the majority of population within Cluster 2 was excluded based on the second elimination criterion (Figure 4.4A). This further supports previous indications that the binding modes of Cluster 2 may be part of the false-positive dockings. As only 10 structures of this cluster formed at least one contact with the receptor, and notably, only one pose showed two important contacts with residues L23 and Y290$^{6.58}$, this cluster was deemed unsuitable for further analysis.

Similarly, despite the high population numbers in Cluster 4, only the important contact Y290$^{6.58}$-GnRH was identified (Figure 4.4B). In contrast, Clusters 1 and 5 exhibit a substantial number of important contacts and a significant population (Figure 4.4C, D). This implies that the docking configurations within these clusters have a strong likelihood of including the close-to-native GnRH and GnRH1R.

As clusters 2 and 4 did not present high numbers or important contacts, hydrogen bond analysis was conducted only for clusters 1 and 5.(Figure 4.5). The aim of the contact and H-bond analysis was to firstly eliminate poses that do not present any meaningful contacts, and additionally perform an in-depth analysis of the binding interface of the clusters. Special emphasis is given to contacts and H-bonds formed between GnRH and the GnRH1R binding site.

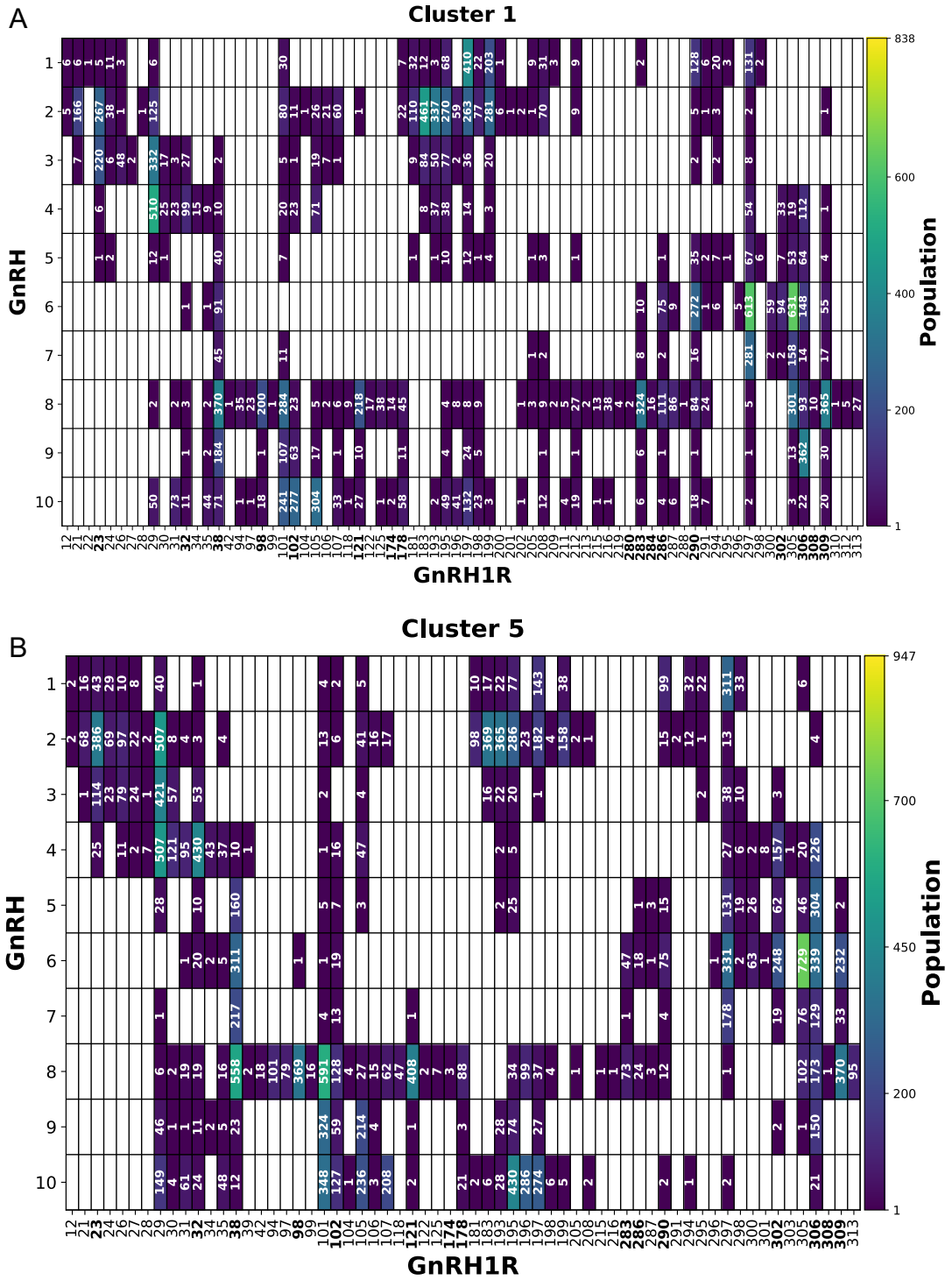


Figure 4.5: H-bond analysis of Clusters 1, and 5 in the form of heat-maps. The Population bar represents the number of structures that display the specific contact in a color-coded manner: high (yellow) and low (blue) Population. In bold font, experimentally proven important GnRH-interacting residues are highlighted.

D302$^{7.31}$

The proposed interaction between the negatively charged aspartic acid, D302$^{7.31}$ with the positively charged arginine, R8 in mammalian GnRH, was suggested due to the distinctive nature of R8. Receptor specificity for R8 through D302$^{7.31}$ was proposed[59] and the mutation of D302$^{7.31}$ resulted in a decrease in GnRH binding affinity when mutated to uncharged amino acids, while mutations to charged amino acids had no observable effects[59], [60]. The proposed mechanism involved the formation of a SB between R8 and D302$^{7.31}$, leading to a high-affinity GnRH conformation. However, it was concluded that the binding affinity of GnRH is not exclusively dependent on the R8-D302$^{7.31}$ interaction[60]. D302$^{7.31}$ in Cluster 1 exhibits moderate contact frequencies, particularly with P9 at 29.24% (p.245) (where p. stands for population) and G6 at 19.81% (p.166) (Figure 4.4C). The hydrogen bonding with G6 and S4 is also noteworthy, at 11.2% (p.94) and 3.9% (p.33) respectively (Figure 4.5A).

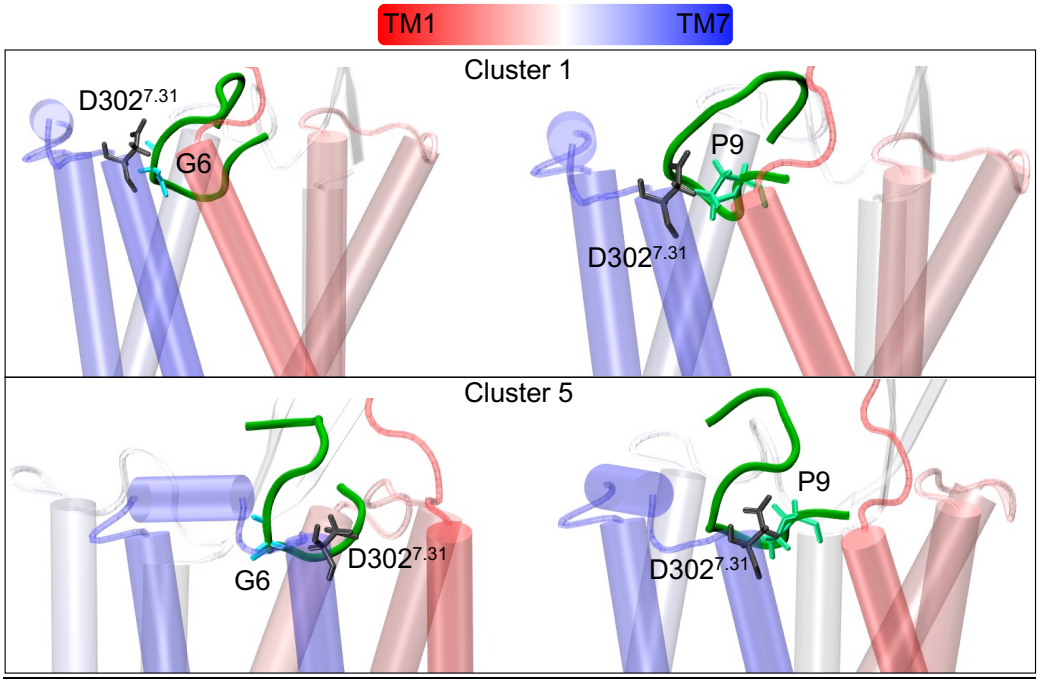


Figure 4.6: Dominant contacts of residue D302$^{7.31}$ in clusters 1 (upper panel) and 5 (lower panel).

In Cluster 5, D302$^{7.31}$ displays a different activity pattern, showing a high contact frequency with G6 at 42.0% (p.398) (Figure 4.4D). The contacts with S4 and P9 at 13.1% (p.124) and 10.0% (p.95) respectively, along with moderate hydrogen bonding frequencies with Y5 at 6.5% (p.62) and S4 at 16.6% (p.157), further highlight a versatile role of D302$^{7.31}$ in the GnRH-GnRH1R interface (Figure 4.4D, Figure 4.5B). Although the R8-D302$^{7.31}$ contact or H-bond did not form in any clusters, a SB may be possible and observable through subsequent

MD simulations. However, these results indicate a different binding mechanism for the R8 residue of GnRH than previously proposed.

F308$^{7.37}$ assisted by H306$^{7.35}$

The proposed interaction between F308$^{7.37}$ and W3, mediated by H305$^{7.35}$ in the mouse GnRH receptor, was suggested from a computational study and assessed through experimental studies. The mutation of the H305$^{7.35}$ residue in the mouse GnRH receptor, particularly with nonpolar amino acids, resulted in a decrease in GnRH binding affinity, while mutations with polar amino acids had no discernible effect[61]. It was hypothesised that H305$^{7.35}$ forms an intrahelical contact with F308$^{7.37}$, consequently driving F308$^{7.37}$ to engage in π-π interactions with W3 in GnRH[61]. In the human GnRH1R investigated in this research, H305$^{7.35}$ of the mouse receptor corresponds to H306$^{7.35}$ in the human receptor. Mutation of both F308$^{7.37}$ and H306$^{7.35}$ resulted in reduced GnRH binding affinity and significant increase in EC_{50} values[52]. In Cluster 1, interactions of F308$^{7.37}$ are negligible, as contacts with G10 at 0.12% (p.1) (Figure 4.4C) and 0.1% (p.1) in Cluster 5 are present (Figure 4.4D).

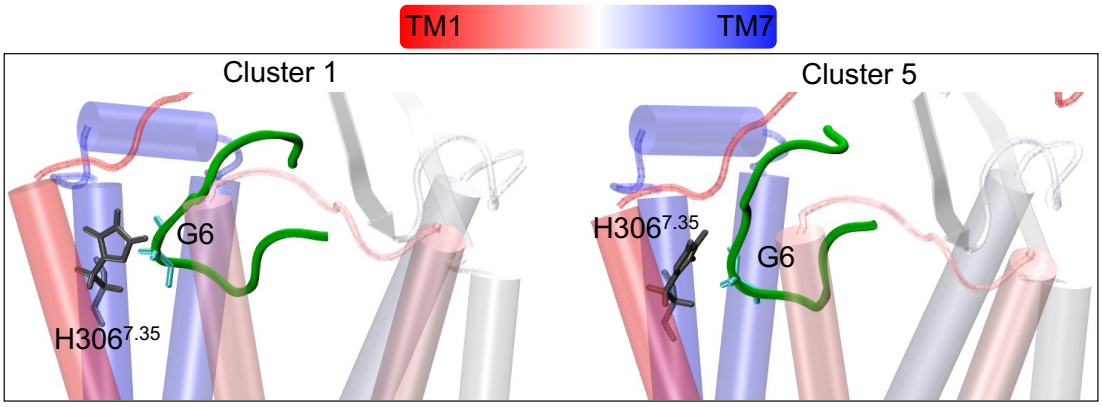


Figure 4.7: Dominant contacts of residue H306$^{7.35}$ in clusters 1 (left panel) and 5 (right panel).

In Cluster 1, H306$^{7.35}$ shows a preferred interaction pattern through H-bonds particularly with P9 at 43.2% (p.362), R8 at 11.1% (p.93), G6 at 17.7% (p.148) and S4 at 13.4% (p.112) suggesting a potential involvement in GnRH binding (Figure 4.5A). The hydrogen bonding frequencies of these residues with H306$^{7.35}$ increases significantly in Cluster 5 with G6 reaching frequencies of 35.8% (p.339), Y5 at 32.1% (p.304) S4 at 23.9% (p.226), L7 at 13.6% (p.129) and R8 at 18.3% (p.150) (Figure 4.5B). Notably the P9-H306$^{7.35}$ H-bond frequency is reduced to 15.8% (p.150) in Cluster 5 and the additional H2-H306$^{7.35}$ H-bond formed at minimal frequency of 0.4% (p.4). Notable contacts with G6 at 5.25% (p.44) in Cluster 1 and at 14.7% (p.139) in Cluster 5 are also observed (Figure 4.4C, D). The significant role of the H306$^{7.35}$-F308$^{7.37}$ will be further investigated through MD simulations to further understand the roles of F308$^{7.37}$ and H306$^{7.35}$ in GnRH binding.

Y290$^{6.58}$

The proposed interaction involving Y290$^{6.58}$ and Y5, as suggested by computational studies, gained significance due to the observed reduction in GnRH binding affinity upon the mutation of Y290$^{6.58}$[52], [62]. The suggested functional role of Y290$^{6.58}$ was its potential function as a coupling agent for agonists, initiating the rotation of TM6. Y290$^{6.58}$ in Cluster 1 predominantly engages with L7, showing a contact frequency of 11.22% (p.94) (Figure 4.4C). However, H-bond interactions are markedly higher, with the most notable being with G6 at 32.5% (p.272), E1 at 15.3% (p.128) and R8 at 10% (p.84) (Figure 4.5A). Less frequent H-bonds form with Y5 at 4.2% (p.35) and G10 at 2.1% (p.18). Cluster 5 reveals reduced contact and H-bond interactions with GnRH, with the most notable being the G6 at 7.9% (p.75) and E1 at 10.5% (p.99) H-bonds (Figure 4.5B), whereas minimal H-bond interaction forms with Y5 and H2. The interaction dynamics of Y290$^{6.58}$ across the clusters indicates a preference for hydrogen bonding.

Y283$^{6.51}$ and F309$^{7.38}$

The Y283$^{6.51}$ residue of GnRH1R located in TM6 decreased GnRH-binding affinity upon mutation[52], [63]. This residue participates in ligand binding of endothelin and apelin receptors and it was suggested that this residue may have a similar behaviour in the GnRH1R[64], [65]. Experimental studies suggested involvement of W3-F309$^{7.38}$[61] in GnRH binding by the disruption of the Y283$^{6.51}$-F309$^{7.38}$ intrahelical contact[66]. Y283$^{6.51}$ displays a selective interaction pattern across the clusters. In Clusters 1 and 5, Y283$^{6.51}$ does not form any contacts with GnRH. However, Y283$^{6.51}$ in Cluster 1 shows a pronounced increase in H-bond frequency with R8 at a remarkable 38.7% (p.324) (Figure 4.5A). In Cluster 5, the interactions of Y283$^{6.51}$ are limited to H-bonds with R8 at 7.7% (p.73) and less prominently with L7 and G6 at 0.1% (p.1) and 5.0% (p.47) respectively (Figure 4.5D).

Additionally, F309$^{7.38}$ exhibits a diverse range of interactions across the studied clusters. In Cluster 1, the contact with GnRH residues is notably low, where the most notable contacts are with R8 at 5.13% (p.43) and G6 at 18.2% (p.172) in Cluster 5 (Figure 4.4C, D). However, F309$^{7.38}$ engages strongly through H-bonds with R8 at 43.6% (p.365) in Cluster 1 and 39.1% (p.370) in Cluster 5 (Figure 4.5). Additionally, notable H-bonds form between G6 and F309$^{7.38}$ at a 24.5% (p.232) frequency in Cluster 5. The interaction pattern of F309$^{7.38}$ is indicative of its selective yet substantial involvement in the GnRH binding function, with a particular affinity for R8. The proposed W3-Y283$^{6.51}$ interaction was not observed in any cluster, however, the interaction pattern of the Y283$^{6.51}$ and F309$^{7.38}$ will be further studied through MD simulations.

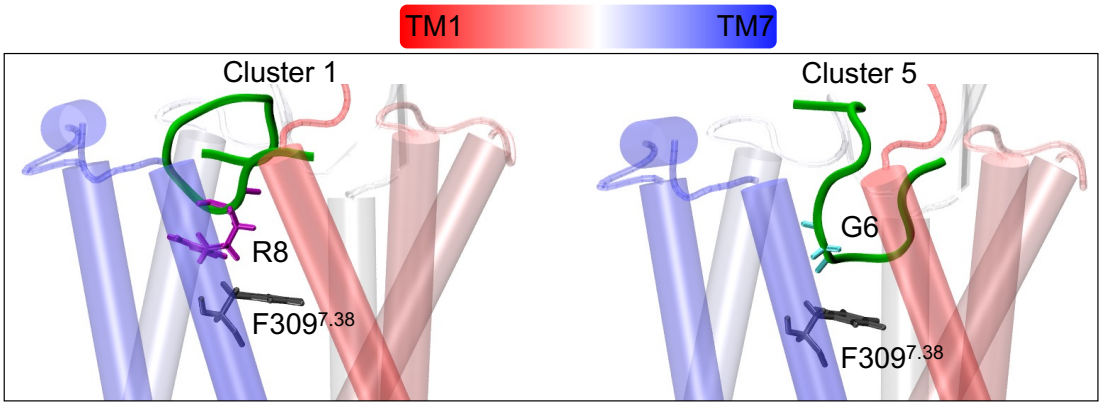


Figure 4.8: Dominant contacts of residue F309$^{7.38}$ in clusters 1 (left panel) and 5 (right panel).

R38$^{1.35}$

Prior computational studies suggested interactions of R38$^{1.35}$ with G10 of GnRH due their proximity in the computational model. Mutation of R38$^{1.35}$ resulted in a reduced GnRH-binding affinity[67] although the proposed interaction was not confirmed. The decreased binding affinity was justified as a broken intrahelical contact that assists GnRH binging and is lost after the R38$^{1.35}$ mutation. Additionally, the R38$^{1.35}$:D98$^{2.60}$ SB was suggested in the inactive receptor conformation which is disrupted by GnRH binding in the active conformation [67]. R38$^{1.35}$ does not show any significant contacts within any clusters, however, high levels of H-bonds form within Cluster 1 and 5 (Figure 4.5). In Cluster 1, R38$^{1.35}$ becomes notably active, establishing H-bonds primarily with R8 at 44.2% (p.370), P9 at 22% (p.184), G6 at 10.9% (p.91), G10 at 8.5% (p.71) and Y5 at 4.8% (p.40) frequencies (Figure 4.5A). The interaction pattern of R38$^{1.35}$ deviates in Cluster 5 with a strong H-bond frequency with R8 at 58.9% (p.558), G6 at 32,8% (p.311), L7 at 22.9% (p.217) and Y5 at 16.9% (p.160) (Figure 4.5B). The proposed interaction between G10 and R38$^{1.35}$ was observed through moderate frequency H-bonds in Cluster 1, however, the overall interactions of R38$^{1.35}$ with GnRH and its conformational role will be further observed and analysed by MD simulations.

W280$^{6.48}$

The role of the W280$^{6.48}$ residue, a key component of the CWxPY motif in TM6, warrants careful consideration. Experimental studies have shown that mutations in W280$^{6.48}$ have a significantly affected GnRH binding affinity and cellular signalling[52], [63]. However, the GnRH1R was rescued by pharmacophores after W280$^{2.48}$ mutation which indicated that this residue is important for the conformation but does not participate in ligand binding. The proposed interaction of W280$^{6.48}$ and W3 occurs through aryl-aryl

interactions[63], [68], [69]. In this study, the interaction profile of $W280^{6.48}$ across various clusters in minimum. Cluster 1 exhibits low-frequency H-bond interactions between $W280^{6.48}$ and R8, observable in only 2 poses (Figure 4.5A). This modest interaction level contrasts with the absence of contacts or H-bonds in other clusters. Despite the lack of high frequency contacts or H-bonds between $W280^{6.48}$ and GnRH within the clusters, its functionality as part of the CWxPY motif will be further investigated through MD simulations.

$K121^{3.32}$

$K121^{3.32}$ is a conserved residue across the GPCR family and in the GnRH1R the formation of a SB with $E90^{2.53}$ has been previously suggested[70], [71]. The inactive receptor conformation is suggested to possess two SBs between $E90^{2.53}$:$K121^{3.32}$ and $D38^{1.35}$:$N98^{2.60}$. Upon activation, GnRH binding is suggested to disrupt the $E90^{2.53}$:$K121^{3.32}$ SB and cause the rotation of $K121^{3.32}$ towards the formation of the $N98^{2.60}$:$K121^{3.32}$ SB in the active conformation[72]. However, the $E90^{2.53}$:$K121^{3.32}$ SB was later disproved in MD simulations of the Apo receptor[73]. Computational models previously indicated interactions of $K121^{3.32}$ with GnRH residues E1, H2, and W3. Mutation studies of $K121^{3.32}$ revealed reduced GnRH binding affinity and signal transduction[52], while antagonist binding remained unaffected[63], [74]. Despite these findings, the precise role of $K121^{3.32}$ in GnRH binding remains ambiguous, warranting further investigation to conclusively determine its functional significance in receptor activation and GnRH interaction. $K121^{3.32}$ shows a distinctive pattern of hydrogen bonding across different Clusters 1 and 5. Cluster 1 reveals a robust profile for $K121^{3.32}$, with a high frequency of H-bonds with R8 at 26.0% (p.218) (Figure 4.5C). The interaction with G10, although lower, is still considerable at 3.2% (p.27) (Figure 4.5C). In Cluster 5, $K121^{3.32}$ engages in a high frequency of hydrogen bonding with R8 at 43.1% (p.408) (Figure 4.5D). MD simulations will provide a deeper understanding of the $K121^{3.32}$ functionality in GnRH1R activation and any association to GnRH binding with greater certainty.

$L23^{N\text{-}ter}$

The $L23^{N\text{-}ter}$ residue of GnRH1R N-terminus increased receptor sensitivity values, EC_{50}, upon mutation[52]. However, the specific involvement of $L23^{N\text{-}ter}$ to GnRH binding has not been elucidated. $L23^{N\text{-}ter}$ displays negligible contact interactions in Clusters 1 and 5. However, significant H-bond interaction occurs between E1 and H2 of Cluster 1 with frequencies of 31.9% (p.267) and 26.3% (p.220) (Figure 4.5C). Interestingly, H-bonds with H2 appear to dominate interactions with L23 in Cluster 5, with a striking frequency of 40.8% (p.386) and W3 at 12% (p.114) (Figure 4.5D). This increased activity, especially the pronounced contact with H2, indicates the

L23 $^{N\text{-ter}}$ potential involvement in stabilising the complex configuration. A potential function of this residue may be to 'cap' GnRH in the binding pocket and reenforce the binding conformation. Further MD simulation will uncover the role of the N-terminus and ECL2 in assisting or participating in GnRH binding.

T32$^{1.29}$

Mutation of T32$^{1.29}$I naturally occurs in the human GnRH1R gene in patients with idiopathic hypogonadotropic hypogonadism[75]. This mutation reduces GnRH1R function primarily by reducing GnRH binding affinity[75]. In Cluster 1, residue T32$^{1.29}$ primarily engages with G10 with a contact frequency of 11.81% (p.99) (Figure 4.4C). H-bond interactions are also observed, with a modest frequency of 1.3% (p.11) with G10. Additionally, T32$^{1.29}$ establishes contacts and H-bonds with S4 at significant frequencies of 4.77% (p.40) and 11.8% (p.99) (Figure 4.5C).

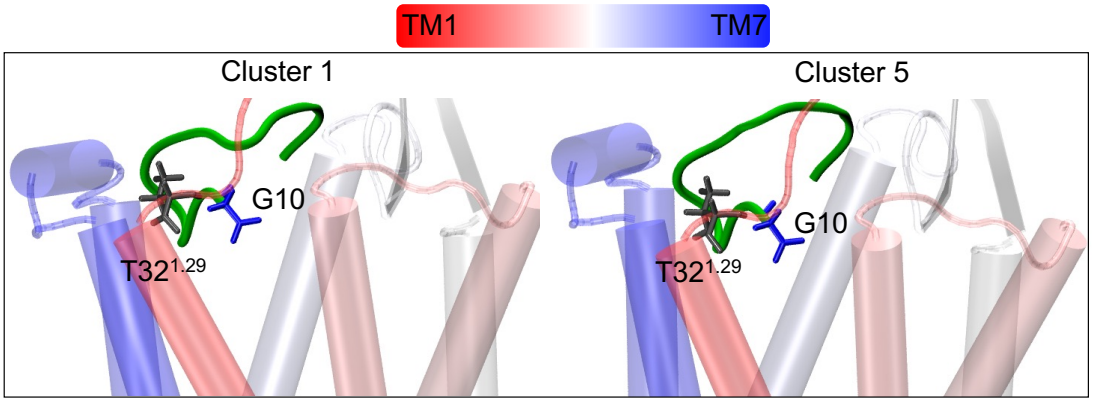


Figure 4.9: Dominant contacts of residue T32$^{1.29}$ in clusters 1 (left panel) and 5 (right panel).

Transitioning to Cluster 5, T32$^{1.29}$ exhibits a drastic change in its interaction profile, where it engages extensively in H-bonds with S4 at an exceptionally high frequency of 45.4% (p.430) (Figure 4.5C, D). Contacts with S4 remain robust at 23.2% (p.220), further emphasising the importance of the S4-T32$^{1.29}$ interaction. The presence of contacts with other residues such as G10 at 7.1% (p.67) and H-bonds at 2.5% (p.24) also indicates the versatile interaction network of T32$^{1.29}$.

N102$^{2.64}$

The mutation of N102$^{2.64}$ has been shown to induce a 225-fold loss in GnRH potency[76]. Interaction of N102$^{2.64}$ with G10 was proposed by prior computational studies[11]. In Cluster 1, N102$^{2.64}$ exhibits substantial contact interactions with G10, as evidenced by a contact frequency of 14.68% (p.123) (Figure 4.4C).

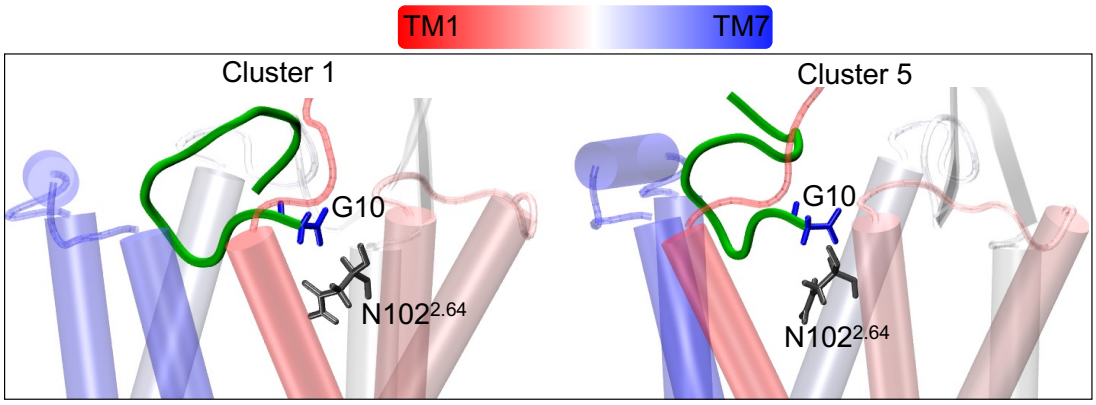


Figure 4.10: Dominant contacts of residue N102$^{2.64}$ in clusters 1 (left panel) and 5 (right panel).

The H-bond interactions are more prominent, with N102$^{2.64}$ forming H-bonds with G10 at a 27.1% (p.227) (Figure 4.5C). In Cluster 5, N102$^{2.64}$ maintains minimal contact interactions with G10, with a contact frequency of 2.5% (p.24) and a higher H-bond frequency of 13.4% (p.127). Contacts with R8 are also notably frequent, at 17.8% (p.169) and H-bond frequency of 13.5% (p.128) (Figure 4.4D, Figure 4.5D).

D98$^{2.60}$

Mutation of D98$^{2.60}$ resulted in a reduction in GnRH binding affinity and the H2-D98$^{2.60}$ interaction was proposed[77]. Interestingly, it has been suggested that D98$^{2.60}$ forms a SB with K121$^{3.32}$ upon GnRH binding and receptor activation. Contrastingly, the R38$^{1.35}$:D98$^{2.60}$ SB is formed in the inactive receptor conformation[72]. In Cluster 1, D98$^{2.60}$ presents a moderate interaction with G10 and R8, with a contact frequency of 2.63% (p.22), 1.2% (p.10) (Figure 4.4C). The H-bond interactions, however, are more prominent, with an interaction with R8 at 23.9% (p.200) and G10 at 2.1% (p.18) suggesting a specific and dominant role in the GnRH hydrogen bonding network especially through R8 interactions (Figure 4.5C).

In Cluster 5, the interaction landscape for D98$^{2.60}$ changes markedly, with R8 dominating the interactions *via* H-bonds at a high frequency of 39.0% (p.369) (Figure 4.5C). Contacts with R8 are also notable, dominating the interactions with a frequency of 6.0% (p.84).

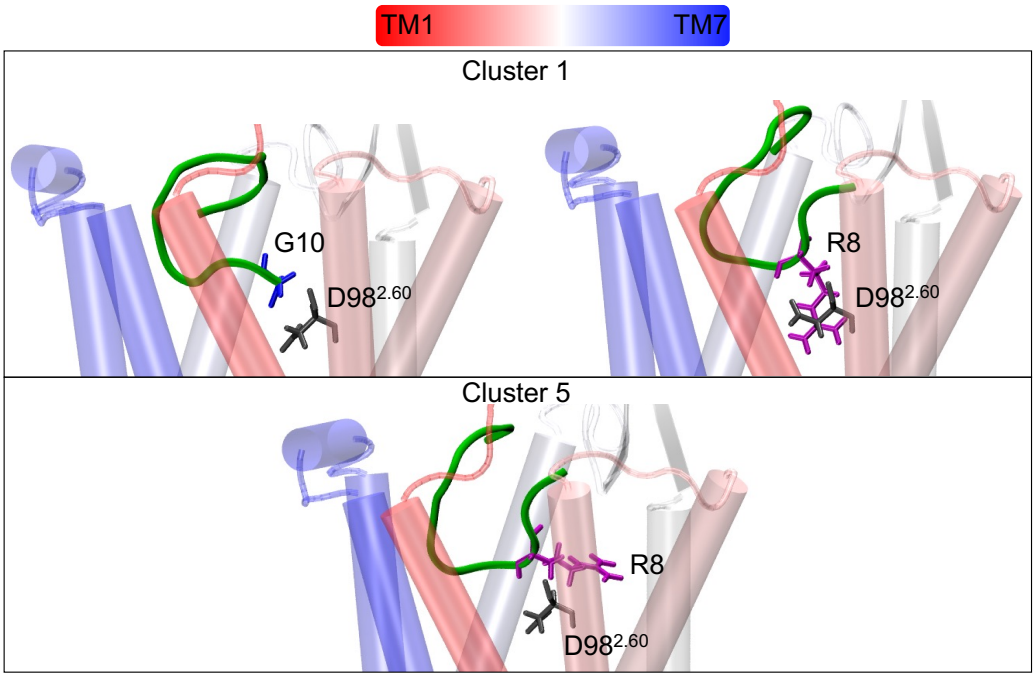


Figure 4.11: Dominant contacts of residue D98$^{2.60}$ in clusters 1 (upper panel) and 5 (lower panel).

The contact profile of D98$^{2.60}$ in Cluster 5 is primarily focused on R8, indicating a specificity in interaction. The proposed H2-D98$^{2.60}$ interaction was not observed in any clusters; however, MD simulations will clarify the interaction pattern and SB network of D98$^{2.60}$.

4.5.5 Cluster Centre of Mass (CoM)

The GnRH-CoM of the clusters is significantly different with respect to the binding pocket. GnRH-CoM yields towards TM4-TM5 in Cluster 4, ECL3-TM5-TM6-TM7-TM1 in Clusters 1 and TM1-TM6-TM7 in Cluster 5, where poses in Cluster 1 appear to adopt similar positions to both Clusters 4 and 5 (Figure 4.12).

The significant population of all clusters suggests that all 3 GnRH-CoM configurations are possible, however, the positional similarities, number of important contacts and H-bonds of Clusters 1 and 5 suggest that the close-to-native GnRH binding configuration may exist is those clusters.

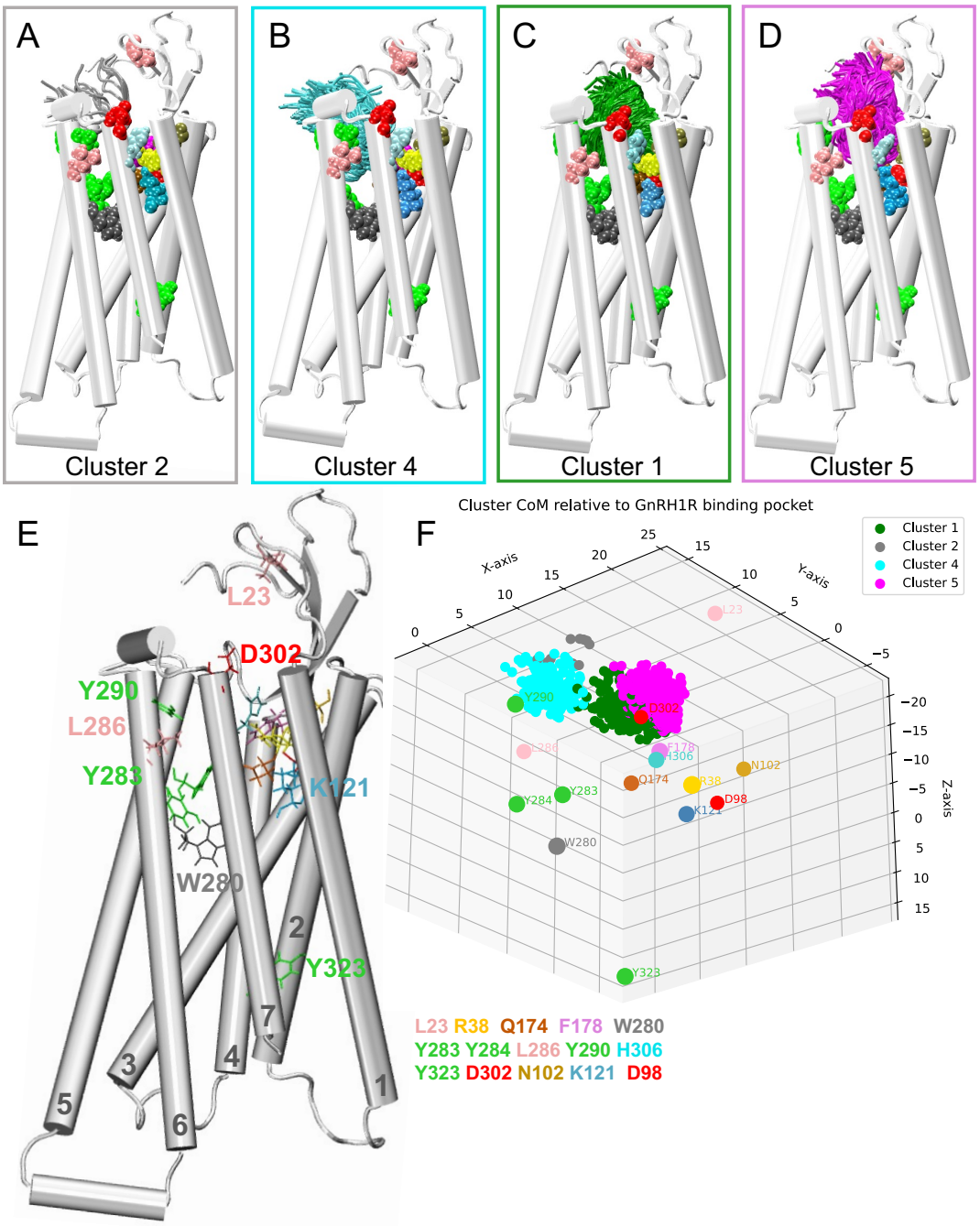


Figure 4.12: Visual representation of Clusters 2 (A), 4 (B), 1 (C), and 5 (D) in relation to the experimentally identified GnRH1R important residues shown in vdW representation. E) Side view representation of the GnR1RH binding pocket residues. F) 3D plot illustrating the CoM of GnRH in each cluster relative to the positions of binding pocket residues.

4.5.6 Candidate selection

In the final stage of structure elimination, binding modes from clusters were selected based on the following criteria:

1. ΔG binding < -20 REU
2. Important contacts > 1

As clusters 1 and 4 did not shown more than one important contact they were eliminated and discarded from further analysis. Subsequently, poses from clusters 1 and 5 were ranked based on the lowest ΔG Binding energies and highest numbers of important contacts (Figure 4.13).

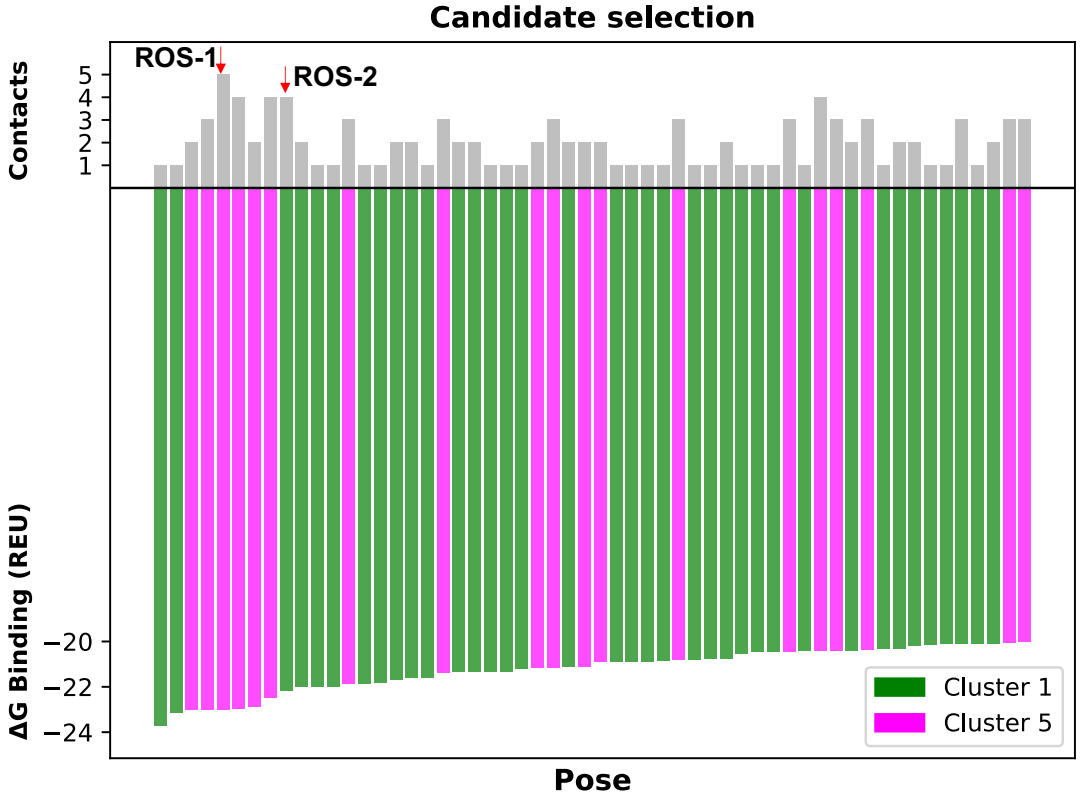


Figure 4.13: Application of the last stage elimination criteria for candidate pose selection for MD simulation noted as ROS-1 and ROS-2.

The final selection included one structure from each cluster that presented both the lowest ΔG Binding and highest number of important contacts in their representative clusters. This resulted in the selection of ROS-1 and ROS-2 binding modes from the parent clusters 5 and 1 respectively (Figure 4.13, Figure 4.14). ROS-1 displayed a total of 5 important contacts and ΔG Binding of –23.012 REU and ROS-2 displayed 4 important contacts and ΔG Binding of -22.159 REU (Figure 4.13).

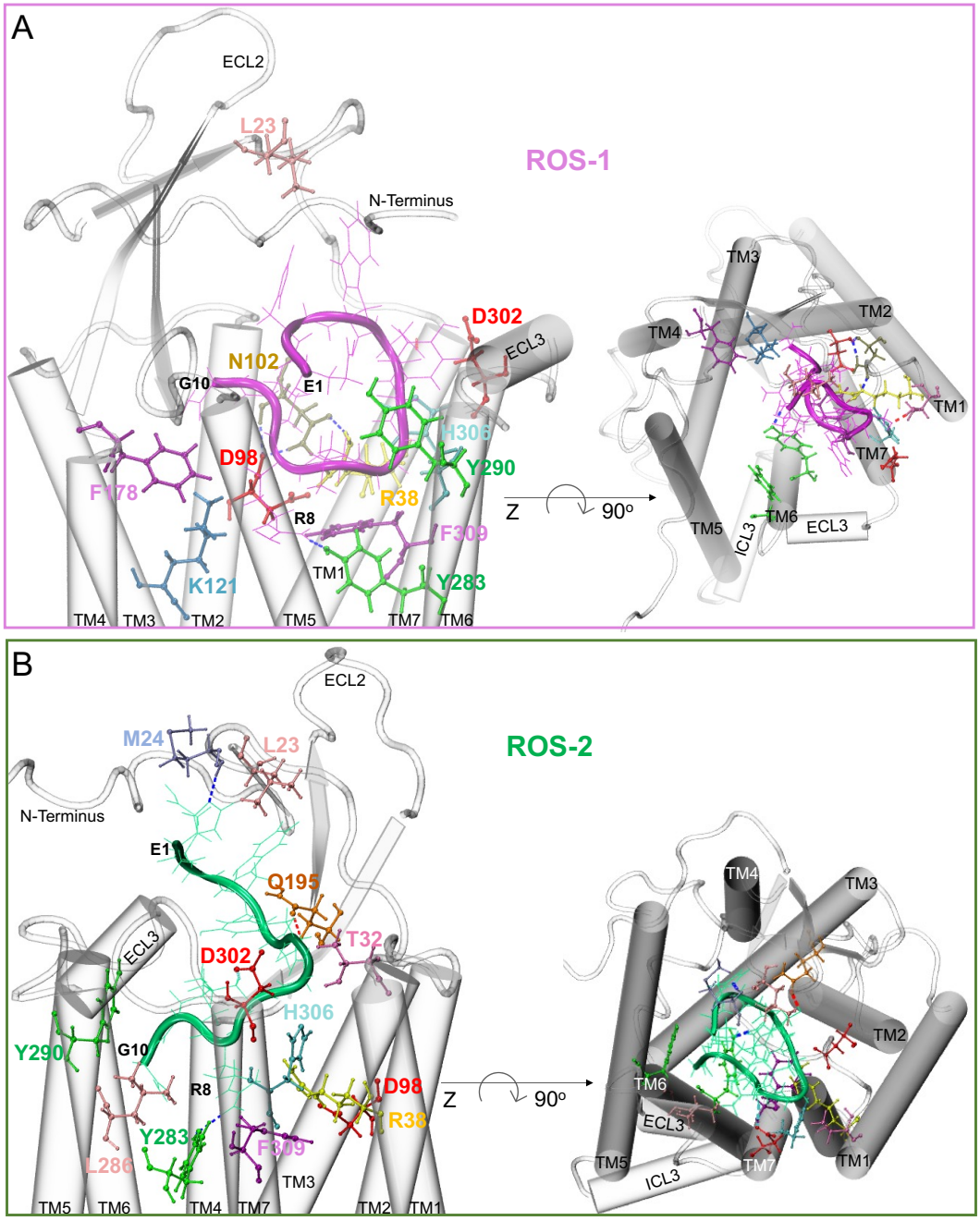


Figure 4.14: Visual depiction of ROS-1 (A) and ROS-2 (B) along with their significant contacts or residues within a 5 Å distance, and the presence of H-bonds is indicated in blue dashed lines.

4.6 Conclusions

In this Chapter, the analysis of GnRH-GnRH1R docking simulations using the Rosetta FlexPepDock protocol and energy-based clustering were presented and analysed. Key energy and structural metrics, including Total Score, ΔG binding, and ΔSASA, across selected clusters were examined. Mann-Whitney U tests and effect size calculations were used to assess the

statistical significance and magnitude of differences between the metrics of each cluster. Notably, Cluster 2 displayed substantially lower binding energy values than the rest of the clusters but lacked biological relevance as revealed by the minimal contacts between GnRH and GnRH1R. On the contrary, Clusters 1 and 5 exhibited both significant contacts and population, suggesting potential native-like binding modes. Cluster 4 exhibited a high population in general contacts between GnRH-GnRH1R however, displayed only one high frequency important contact with $Y290^{6.58}$.

The results of cluster contacts of the GnRH1R-GnRH complex, show interactions with key receptor residues including $F308^{7.37}$, $H306^{7.35}$, $Y290^{6.58}$, $Y283^{6.51}$, $F309^{7.38}$, $R38^{1.38}$, $L23^{N\text{-}ter}$, and $K121^{3.32}$. Cluster 5 emerged as potentially representative of the native-like binding mode, characterised by pronounced interactions, particularly for $F309^{7.38}$ and $R38^{1.38}$ whilst $Y290^{6.58}$ and $Y283^{6.51}$ predominantly engaged through H-bonds. The interaction of $K121^{3.32}$ with R8, through hydrogen bonding, challenges traditional understandings of GnRH binding, hinting at a more complex mechanism. A possible GnRH binding mechanism may be the formation of the R8:$E90^{2.53}$ or R8:$D98^{2.60}$ SB or a cation-π interaction with $W280^{6.48}$ of the CWxPY motif.

Results of this chapter reveal the complexity of GPCR-ligand interactions and highlight the need for further in-depth studies. Therefore, further investigation of the close-to-native representative binding poses, namely ROS-1 and ROS-2 through MD simulations are necessary, to unravel the full spectrum of GnRH1R-GnRH binding dynamics and receptor activation.

Chapter 5 - GnRH binding through MD simulation

5.1 Rationale

In this chapter, the Rosetta-derived binding modes of the GnRH-GnRH1R complex were examined using MD simulations to identify the native binding mode that induces GnRH1R activation. The two top-ranked docked complexes were selected for MD simulations at a microsecond timescale, quantifying activation based on the hallmark increase in transmembrane helix 3-6 (TM3-TM6) distance.

Following MD simulations, the complete network of intermolecular interactions underlying the native-like active binding mode was successfully described.

Defining the physiological GnRH-bound state that induces receptor activation will facilitate improved design of both agonists and antagonists targeting this therapeutically valuable GPCR, especially for hormone-dependent disorders and cancers.

5.2 Aims and Objectives

1) To perform atomistic MD simulations of the top-ranked ROS-1 and ROS-2 binding modes to assess their ability to induce GnRH activation.
    - Set up and perform unbiased MD simulations of the top-docked GnRH-GnRH1R complexes.
    - Calculate the TM3-TM6 distances over time to determine receptor activation.
2) To characterise the GnRH-GnRH1R interactions of the active binding in detail.
    - Employ several techniques and methodologies including customised code, VMD plugins and various webservers to compute complicated intermolecular interactions including π-π, cation-π, hydrogen bonding and vdW.
    - Analyse the network of interactions between GnRH and GnRH1R whilst justifying each interaction based on experimental mutagenesis data.
3) To investigate the conformational landscape of GnRH during active binding.
    - Perform principal component analysis (PCA) to sample the major GnRH conformations during active binding.
4) To perform critical analysis of previously suggested GnRH-GnRH1R interactions against those observed herein.

- Justify disagreements with prior computational studies and provide evidence to support the accuracy of the active-GnRH binding mode identified in this work.

5.3 Methods

5.3.1 MD Simulations

The selected GnRH-GnRH1R binding modes from the docking simulations were subjected to microsecond-timescale MD simulations (of duration 1.1 μs) to impartially assess their stability and capacity to activate the receptor. Additional MD simulations were performed on an unbound GnRH-GnRH1R system to investigate the possibility of GnRH binding and receptor activation naturally. Simulations were also conducted for the GnRH1R in absence of GnRH (Apo-GnRH1R) to develop an equilibrated model of the inactive receptor state and ascertain the propensity for activation in absence of GnRH. All systems were constructed as outlined in Chapter 3-3.7. Two fully independent replica simulations were initiated for each system. Furthermore, for the ROS-1 binding mode that induced activation, three "dependent" replica simulations were performed. The depended replicas initiated from time equal to 0.85 μs of the original ROS-1 simulation. At this timepoint, the receptor is still in its inactive conformation with TM3-TM6 distance of ~8 Å. This allows the statistical reliability of the activation process to be assessed using reasonable computational resources.

5.3.2 Analysis of Receptor Activation

To distinguish between inactive and active GnRH1R conformations, TM3-TM6 distances were calculated over the trajectories based on the distance between R139$^{3.50}$ and T265$^{6.33}$. For reference, the inactive crystal structure of the GnRH1R has TM3-TM6 distance of 7.9 Å, while active-like GnRH1R homology models predicted by AlphaFold show a distance of 11 Å. TM3-TM6 distances were calculated using MDTraj and custom Python scripts (Appendix, Tm3_6_7_distance.py, tm3_tm6_all_plot.py). Further analyses focused exclusively on the activated GnRH1R portion of the trajectories between 1.0 and 1.1 μs, which yielded 100 ns of activated GnRH1R simulation time. Hydrogen bonding was quantified with a 3.5 Å N-O distance and 20° angle cutoff using the VMD plugin 'Hydrogen bonds'. Complete intermolecular and intramolecular interaction profiles (π-π, vdW, ionic, π-hydrogen bonds, and hydrogen bond interactions) were generated using the RING webserver (https://ring.biocomputingup.it) over the active conformation timeframe (final 100 ns, 100 frames) and plotted using customised Python scripts (Appendix, Ring_interaction_map.py).

5.3.3 GnRH Conformational Analysis

The conformations assumed by GnRH over the activating trajectory were examined by principal component analysis (PCA) of the peptide coordinates. PCA was conducted using Scikit-Learn and SciPy on the final 100 ns to identify any distinct GnRH conformations during active binding. PCA and the corresponding variance explained were calculated and plotted using custom Python scripts (Appendix, PCA.py).

5.4 Background

Previous *in silico* studies have utilised molecular docking simulations to model the interaction between GnRH and earlier homology models of the GnRH1R, given the lack of a crystal structure at the time. Guided by experimental mutagenesis data, these studies proposed potential binding modes and interactions between GnRH and GnRH1R. One such interaction is a salt bridge (SB) between the arginine at position 8 of GnRH (R8) and aspartic acid 302 ($D302^{7.31}$) of GnRH1R, was hypothesised and supported by the observation that mutation of $D302^{7.31}$ decreased binding affinity[59], [60]. However, subsequent studies concluded that this ionic interaction was not requisite for receptor activation as conformationally constrained GnRH peptides were not affected by the presence of absence of R8 or $D302^{7.31}$ or both[66]. This finding is noteworthy due to the unconventional nature of the observed interaction. Typically, an arginine at position 8 is deemed indispensable for receptor activation induced by all GnRH peptide agonists. Consequently, one would anticipate that a mutation affecting the receptor residue interacting with R8 would result in the abolishment of both binding and signalling.

While the salt bridge scenario clarifies why GnRH1 and its agonists necessitate R8 for binding GnRH1R *via* a $D302^{7.31}$ SB, it fails to address how GnRH2, featuring a tyrosine at position 8, can bind to both GnRH1R and GnRH2R, which contain aspartic acid and proline at position $302^{7.31}$, respectively. This observation introduces an additional layer of complexity to the binding mechanism, suggesting a more intricate process than initially hypothesised. Moreover, it is essential to emphasise that the mutation of $D302^{7.31}$ does not lead to the abolition of GnRH binding or signalling[52], [59]. This observation necessitates re-evaluation of the initially proposed binding mechanism.

Subsequent experimental and computational studies have brought to light inconsistencies in the prior mutagenesis data concerning the role of $E90^{2.53}$ in GnRH1R. Notably, the same group which suggested the R8-$D302^{7.31}$ SB performed mutation of $E90^{2.53}$, revealed no discernible effect on receptor function, cell surface expression, or GnRH binding[59]. This was later disproved as other groups[70] demonstrated that the $E90^{2.53}$ mutation not only

inactivates the receptor but is also observed in patients with hypogonadotrophic hypogonadism, leading to low production of follicle-stimulating hormone (FSH) and luteinising hormone (LH). Significantly, both E90$^{2.53}$ and D302$^{7.31}$ are negatively charged residues capable of forming a SB with R8. However, due to the initial evidence suggesting a R8-D302$^{7.31}$ SB, all subsequent docking studies have favoured this interaction[70], and the potential for an R8-E90$^{2.53}$ SB or an entirely different binding mode was neither proposed nor investigated. Additionally, the R8-D302$^{7.51}$ interaction was not predicted in any of the clusters investigated in Chapter 4. On the contrary, R8 showed interaction with residues located deeper in the intrahelical bundle of the GnRH1R. The inconsistencies revealed by prior research for this interaction and the lack of contacts between R8 and D302$^{7.51}$ in any of the examined clusters in this research, calls for a re-evaluation of the current understanding and points towards a different binding mechanism of GnRH with its receptor.

To circumvent potential biases stemming from previous computational studies, which have favoured specific interactions, this research employed completely unbiased docking simulations. These simulations were conducted without making assumptions about the binding site or potential interactions between the receptor and GnRH. While this approach may reduce the efficacy of native-like binding mode predictions, it facilitated an unbiased exploration of the GnRH1R binding pocket. The outcome of this unbiased docking led to the identification of two distinct binding modes, named ROS-1 and ROS-2, which were subjected to further in-depth investigation through MD simulations.

5.5 Results

5.5.1 Binding mode evolution though MD simulations

Notably, the MD simulations revealed dynamic behaviour in the binding modes, particularly with the ROS-1 configuration exhibiting inward movements, while ROS-2 maintained its initially predicted binding mode (Figure 5.1). ROS-1 primarily interacted with the aromatic residues of TM6 through GnRH residues 3-8. In contrast, ROS-2 assumed an orientation opposite to ROS-1, positioning its C-terminus proximal to TM6 and TM7. This unique orientation of the binding mode facilitated an examination of the previously proposed R8-D302$^{7.51}$ interaction from earlier computational studies[59], [60].

However, it is noteworthy that prior studies had suggested an interaction between the GnRH N-terminus and residues of both TM6-TM3 and TM6-TM7 [59], [60]. This would imply the positioning the GnRH N-terminus simultaneously at TM3 and TM7 - a conformation that appears implausible.

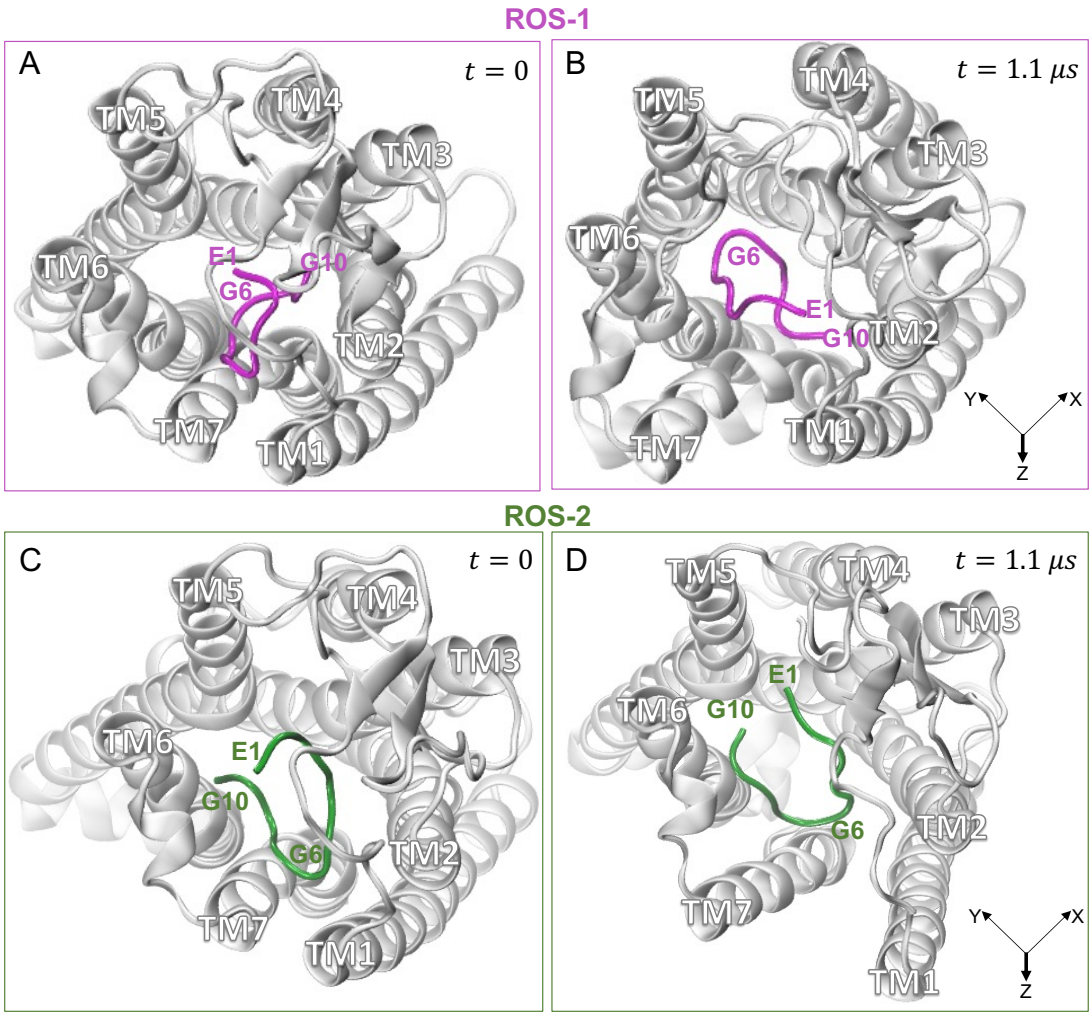


Figure 5.1: Evolution of ROS-1 and ROS-2 binding modes through MD simulations. A) and B) depict the initial and final (1.1 µs) binding modes of ROS-1, respectively. Similarly, C) and D) illustrate the initial and final (1.1 µs) binding modes of ROS-2. The N-terminus, C-terminus, and the central G6 residue of GnRH are denoted.

5.5.2 GnRH1R activation

GPCR activation has been extensively characterised by increases in the distance between TM3-TM6 coupled with decreases between TM3 and TM7, as outlined in the Chapter 1. These coordinated conformational changes, observed across various Class A GPCRs, signify the transition from the inactive to active state. Therefore, the GnRH binding mode responsible for inducing receptor activation is anticipated to evoke similar conformational changes in the GnRH1R.

As anticipated, the TM3-TM6 distances in the Apo-GnRH1R (GnRH free) system remained consistently inactive throughout the simulation, maintaining approximately 8 Å (Figure 5.2D). Similarly, the undocked- GnRH-GnRH1R systems also remained in an inactive state for the duration of the simulation. Notably, the second replica of this system displayed the formation

of an R8-D302$^{7.31}$ SB without triggering receptor activation (Figure 5.2C, Appendix: Figure 8.1).

The two selected binding modes, ROS-1 and ROS-2, along with the Apo-GnRH1R and the undocked GnRH-GnRH1R systems, were subjected to MD simulations. For each system, two independent simulation replicas were performed. However, activation was observed only in the ROS-1 system, occurring after 1.0 μs of simulation. To save computational resources and time, two additional simulations were initiated from the original ROS-1 trajectory, starting at 0.85 μs. At this point in the original simulation, the receptor was still in its inactive conformation, making it an ideal time to restart and validate the activation process in these new simulations.

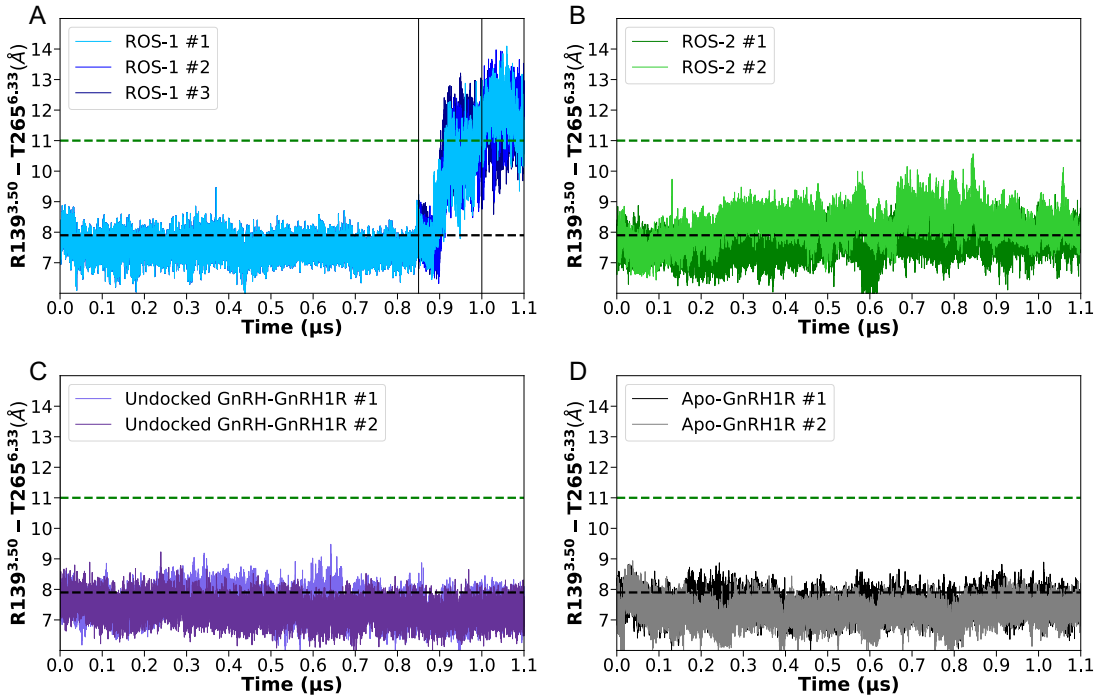


Figure 5.2: Time evolution of the TM3-TM6 distances measured by R139$^{3.50}$-T265$^{6.33}$ Cα atoms. A) Dynamics of ROS-1 activation, showcasing TM3-TM6 distance stabilisation at ~1.0 μs. Analysis focused on the activated receptor during the 1.0-1.1 μs timeframe, with replica initiation at 0.85 μs. B) ROS-2 replicas. C) The undocked GnRH-GnRH1R systems. D) The Apo-GnRH1R system. Green line: active-like AlphaFold predicted TM3-TM6 distance; Black line: inactive GnRH1R crystal TM3-TM6 distance.

This observation provides additional evidence challenging the notion that the previously proposed R8-D302$^{7.31}$ ionic interaction is a requisite for activation and thus does not represent the native binding mode.

Upon comparing the top-ranked Rosetta docked binding modes (ROS-1 and ROS-2), it became evident that only ROS-1 induced receptor activation (Figure 5.2A, B). Notably, within the two independent replicas of ROS-1, activation occurred solely in the second replica, attributed to the complete

rearrangement of the initial binding mode in the first replica. In contrast, the three dependent replicas of the second ROS-1 replica successfully reproduced the activation process (Figure 5.2A). This indicates that among the various binding modes explored during MD simulations, including two undocked GnRH states, two ROS-2 replicas, and two different binding modes in ROS-1 replica 1 and 2, only the latter binding mode in ROS-1 replica 2 triggered activation.

This observation strongly suggests that ROS-1 (second replica) is close to the native binding mode, as it effectively induced TM3-TM6 conformational changes and established robust interactions with the conserved CWxPY aromatic motif of TM6. Moreover, ROS-1 interacted with various residues experimentally shown to be crucial for GnRH binding, with their mutation leading to the abolition of GnRH binding. Consequently, ROS-1 demonstrated a unique capacity to activate GnRH1R, as supported by the increase of the TM3-TM6 distance. Hence, the ROS-1 binding mode serves as the primary focus for all subsequent analyses presented in this study.

5.6 Active binding mode

5.6.1 Active binding characteristics

The active binding mode demonstrates prominent interactions facilitated by a network of aromatic residues, establishing crucial interactions between the N-terminus of GnRH and the aromatic residues of TM6-TM7 in the GnRH1R (Figure 5.3). Specifically, T-shaped π-π stacking interactions was observed between W3 of GnRH and the aromatic residues Y290$^{6.58}$ and H306$^{7.35}$ of GnRH1R, complemented by an irregular (I)-shaped π-π stacking interaction with F309$^{7.38}$ (Figure 5.3B, E). Furthermore, W3 of GnRH engages in intramolecular T-shaped π-π stacking interactions with Y5 (Figure 5.3B). Subsequently, Y5 forms additional π-π stacking interactions with Y290$^{6.58}$, Y283$^{6.51}$, and F309$^{6.58}$, accompanied by vdW and hydrogen bonding interactions with L286$^{6.54}$ (Figure 5.3B, F). It is noteworthy that the mutation of these receptor residues leads to either the complete abolition or a significant reduction in GnRH binding (Chapter 1-1.5, Table 1.3).

Significantly, the aliphatic residues W3 and Y5 of GnRH establish simultaneous interactions with multiple GnRH1R residues spanning both TM6 and TM7 helices, rather than interacting independently. This network of aliphatic interactions between GnRH and the conserved CWxPY residues, as well as F309$^{7.35}$ and H306$^{7.35}$ of GnRH1R, results in a robust and stable interaction throughout the entire trajectory. Importantly, these interactions play a pivotal role in activating the receptor, underlining their significance in the overall mechanism of GnRH1R activation.

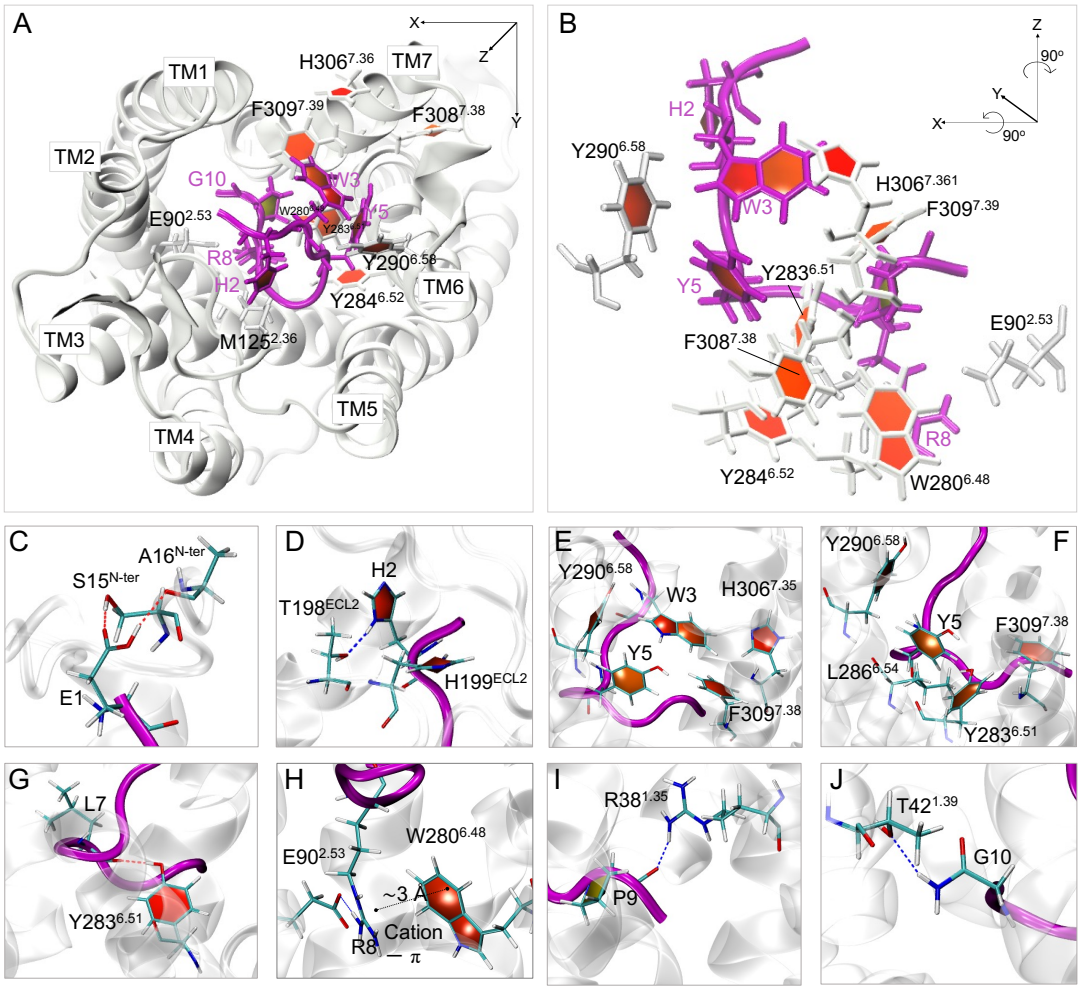


Figure 5.3: Illustration of the ROS-1 activating binding mode. A) Overview of the binding pose seen from the extracellular space and B) isolated view. C-J) Highlighted residues involved in the binding mode, featuring GnRH1R (white) and GnRH (magenta). Snapshots captured at 1.1 μs.

Furthermore, the critical GnRH residue, R8, establishes favourable cation-π interactions with the conserved W280$^{6.48}$ residue of the CWxPY motif (Figure 5.3H), deviating from the previously proposed SB with D302$^{7.31}$ [59], [60]. In addition to this interaction, low-occupancy H-bonds and low duration SB form between R8 and E90$^{2.53}$ as well as vdW and H-bond interactions with

M125$^{3.36}$ (Figure 5.4). Both E90$^{2.53}$ and M125$^{3.36}$ exhibit significant impacts on GnRH binding and receptor function after mutations[52], [73].

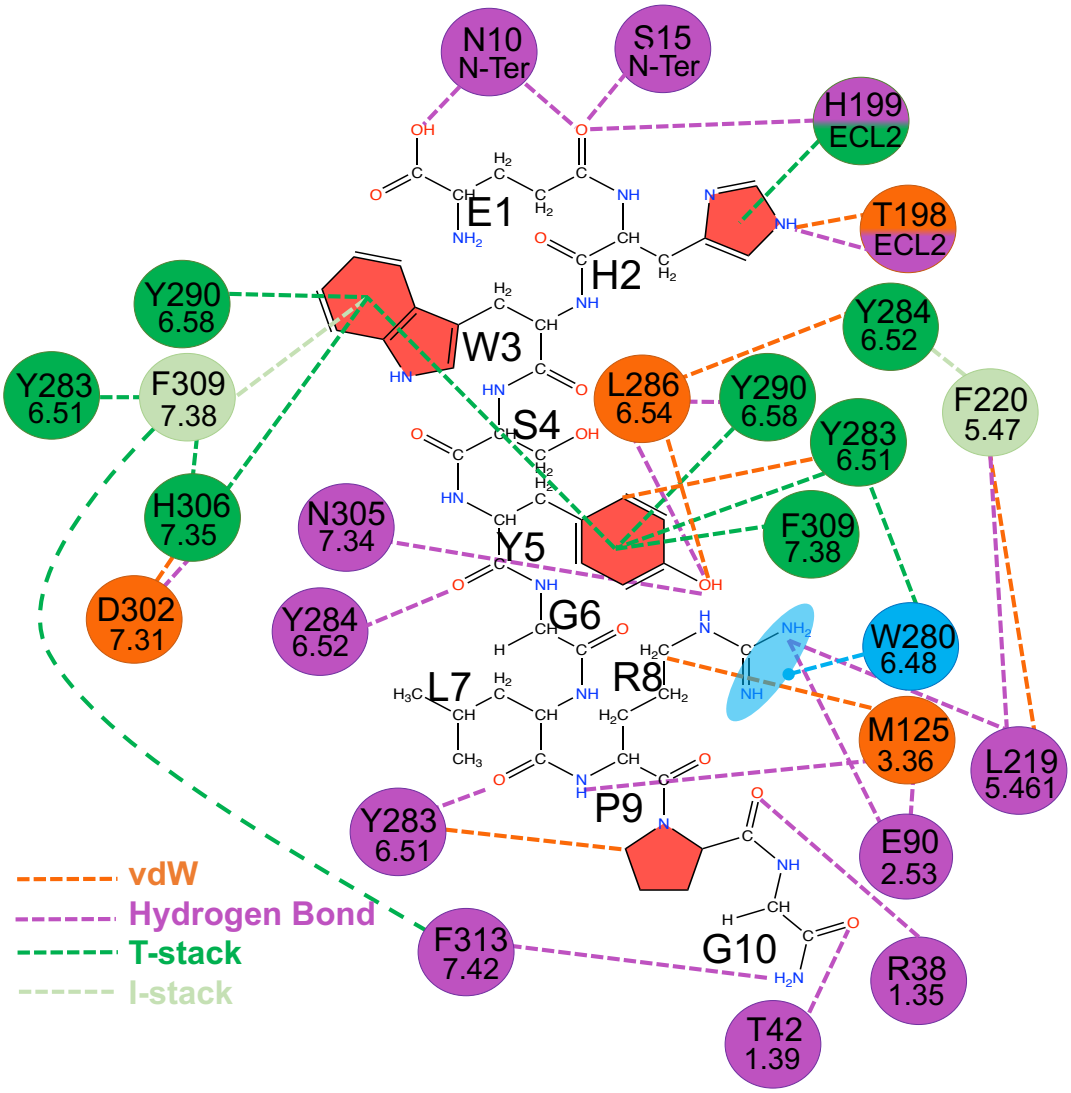


Figure 5.4: Network of GnRH and GnRH1R interactions. GnRH is as its chemical 2D structure, while GnRH1R residues are depicted as circles, with each colour indicating its dominant interactions.

The non-native terminal E1 residue of the modelled GnRH forms H-bonds with N-terminal residues N10$^{N\text{-}ter}$, S15 $^{N\text{-}ter}$, and A16 $^{N\text{-}ter}$ (Figure 5.3C, Figure 5.4). This suggests the potential role of the receptor's N-terminus in GnRH binding, challenging previous studies that proposed that the N-terminus co-occupies the orthosteric binding pocket and interacts with agonists and antagonists, but does not participate in GnRH binding[52]. It is important to note, however, that the N-terminus interactions observed in this research are mediated through E1 rather than the native pyroglutamate ring, raising considerations about their biological relevance. Additionally, H2 forms T-shaped π-π stacking and hydrogen bond interactions with H199^{ECL2} and

T198^{ECL2} of ECL2 (Figure 5.3D, Figure 5.4). Finally, the C-terminal P9 forms H-bonds with R38$^{1.35}$ and engages in vdW interactions with Y283$^{6.51}$ of the CWxPY motif (Figure 5.3I, Figure 5.4) and G10 forms H-bonds with T42$^{1.39}$ (Figure 5.3J, Figure 5.4) Notably, Y283$^{6.51}$ also forms direct H-bonds with the backbone oxygen of L7, either directly or through a bridging water molecule (Figure 5.3G, Figure 5.4).

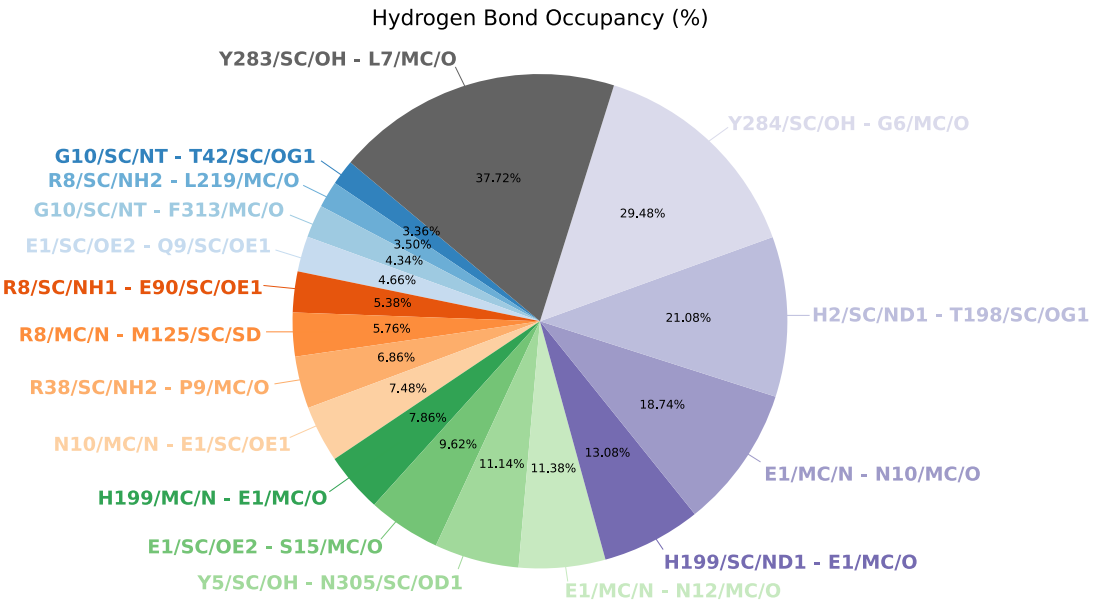


Figure 5.5: H-bond occupancy between GnRH and GnRH1R in the final 100 ns of activation (1.0-1.1 μs). H-bonds with occupancy < 3% are omitted for clarity. Residue order follows the Donor-Acceptor format, with MC indicating main chain and SC indicating side chain. Atoms participating in the H-bonds are listed at the end of each label.

5.6.2 Critical analysis GnRH-GnRH1R interactions

Previous computational studies strongly suggested the existence of a SB between D302$^{7.31}$ and R8 in GnRH1R[59], [60]. However, even though D302$^{7.31}$, located on ECL3 at the beginning of TM7, became a focal point of investigation due to the decreased binding affinity observed upon its mutations, it was later concluded that the R8-D302$^{7.31}$ interaction is not essential for receptor activation[66]. Subsequent computational studies favoured the R8-D302$^{7.31}$ interaction[70]. In the simulations conducted in this research, R8 was found deep within the interhelical bundle of the receptor, directly interacting with the CWxPY motif of GnRH through a cation-π interaction with W280$^{6.48}$. Importantly, W280$^{6.48}$ is highly conserved not only across GnRH receptors but also throughout the entire GPCR superfamily. However, expression of the W280$^{6.48}$ mutant GnRH1R was pharmacoperone rescued and the mutant receptors displayed unchanged ligand-binding affinity and signalling[143]. This suggests that GnRH does not directly contact W280$^{6.48}$. However, GnRH persistently interacts though cation-π interactions

with W280$^{6.48}$ and activates the receptor in the simulations. The unchanged binding after pharmacophore recovery may be attributed to the change of interactions of R8-W280$^{6.48}$ with R8-E90$^{2.53}$ through a SB (Figure 6.6).

The findings of this study provide clarification on why the previously proposed R8-D302$^{7.31}$ interaction is not a prerequisite for activation. Additionally, it sheds light on how GnRH2, with a tyrosine at position 8, can both bind GnRH1R and GnRH2R. In GnRH2, where R8 is substituted with Y8, an aliphatic residue capable of forming either direct π-π stacking interactions with W280$^{6.48}$ or cation-π interactions with the neighbouring K121$^{3.32}$, serves as a plausible explanation for the binding of all GnRH isoforms. Additionally, it is important to note that the 7.31 position is not conserved across GnRH receptors, indicating that the SB interaction cannot provide a universal mechanism for GnRH isoform binding and activation.

A plausible explanation for the decreased binding affinity observed in the D302$^{7.31}$ mutated GnRH1R may be attributed to the H-bond and vdW interactions of D302$^{7.31}$ with GnRH-interacting H306$^{7.35}$ through I-stacking π-π interactions with W3 in the MD simulations. Disruption of the D302$^{7.31}$-H306$^{7.35}$ interactions could potentially destabilise the GnRH-binding pocket, assisted by H306$^{7.35}$ and F309$^{7.38}$, which would explain the decrease in binding affinity and the non-necessity of the D302$^{7.31}$-R8 interaction.

The conserved residue K121$^{3.32}$ in GnRH1R has been previously suggested to interact with the GnRH peptide, particularly through pG1 and H2. Additionally, early computational models proposed that K121$^{3.32}$ forms a SB with D98$^{2.60}$ in the active conformation. In a recent MD study involving a non-peptide-GnRH1R complex, K121$^{3.32}$ was shown to engage with non-peptide agonists through cation-π interactions, and the expected K121$^{3.32}$-R98$^{2.60}$ SB did not form[144]. Moreover, mutation of R98$^{2.60}$ did not affect the binding of the non-peptide agonist [144]. However, the results of this thesis indicate that K121$^{3.32}$ does not actively participate in GnRH binding, despite the consistent observation of the K121$^{3.32}$-R98$^{2.60}$ SB throughout the trajectories (Chapter 6 -6.5.3, Figure 6.6). It is important to note that the non-native structure of E1 in this model might influence conclusions regarding interactions between K121$^{3.32}$ and pG1/H2.

The findings of this study suggest that interactions involving K121$^{3.32}$ are not deemed essential for activation, as GnRH1R was observed to activate without participation of K121$^{3.32}$ in binding and in the presence of the K121$^{3.32}$-R98$^{2.60}$ SB. Nevertheless, mutation of the conserved K121$^{3.32}$ residue results in the abolition of GnRH binding or receptor function. This suggests that K121$^{3.32}$ is likely to be crucial for maintaining receptor conformation and for shaping its binding pocket to accommodate a diverse array of ligands rather than its direct participation in GnRH binding.

Prior computational studies have proposed that the key residue W280$^{6.48}$ of the CWxPY motif interacts with W3 through π-π stacking interactions[68]. However, this research demonstrates that W280$^{6.48}$ engages in cation-π interactions with the unique residue R8 of GnRH. The role of this residue in transmitting signal will be explored further in Chapter 6.

Additionally, prior studies have suggested the interaction of F309$^{7.38}$ with W3[61], [66], potentially disrupting the Y283$^{6.51}$-F309$^{7.38}$ interaction[61]. This research confirms the presence of this interaction, as W3 consistently interacts with F309$^{7.38}$ through a network of π-π stacking interactions. Specifically, I-stacking interactions occur between F309$^{7.38}$ and W3, while T-stacking interactions occur between F309$^{7.38}$ and Y283$^{6.51}$. Thus, this study confirms the direct interaction between W3 and F309$^{7.38}$, while also supporting the stability of the Y283$^{6.51}$-F309$^{7.38}$ intrahelical contact.

Previous computational studies proposed that F308$^{7.37}$ engages in π-π interactions with H306$^{7.35}$, thereby prompting F308$^{7.37}$ to participate in π-π interactions with W3[61]. However, in this study, W3 directly forms T-stacking π-π interactions with H306$^{7.35}$, while F308$^{7.37}$ is oriented towards the membrane environment. Although F308$^{7.37}$ is undoubtedly a crucial residue for the functionality of GnRH1R, it does not directly partake in GnRH binding. Instead, it aids in the π-π interactions of H306$^{7.35}$ and has a potential role in the communication of TM7 with the membrane environment.

The conserved residue Y290$^{6.58}$ of GnRH1R has been identified as a crucial residue in agonist coupling, and prior computational studies suggested an interaction between Y290$^{6.58}$ and Y5[62]. This research validates the existence of this interaction, revealing that Y290$^{6.58}$ engages in direct T-stacking π-π interactions with Y5. Furthermore, several other essential residues interact with Y5, including T-stacking π-π interactions with F309$^{7.38}$ and the Y283$^{6.51}$ residue of the CWxPY motif. Additionally, Y290$^{6.58}$ forms π-aryl interactions with L286$^{6.54}$, mutation of which resulted in a decrease in binding affinity. Moreover, L286$^{6.54}$ engages in vdW and H-bond interactions with the alcohol group of Y5 in the MD trajectories.

Finally, R38$^{1.35}$ has been proposed to interact with G10 according to prior computational studies[67], and this interaction is confirmed by the present research. In the MD trajectories, R38$^{1.35}$ forms H-bond interactions with the backbone oxygen atom of P9, as well as vdW interactions with the N-terminal group of G10.

Experimental studies have revealed that the mutation of the N-terminal residue L23$^{N\text{-}ter}$ leads to a decrease in the binding affinity of GnRH[52]. However, prior research has suggested that N-terminal residues of GnRH1R do not directly participate in GnRH binding [52]. In contrast to this prevailing notion, the findings of this research propose that the N-terminus is indeed involved in GnRH binding, providing a potential explanation for the observed

decrease in binding affinity upon mutation of the N-terminal residue L23$^{N\text{-}ter}$. However, it was not observed that this specific residue directly interacts with GnRH. Instead, N10$^{N\text{-}ter}$, S15$^{N\text{-}ter}$, and A16$^{N\text{-}ter}$ were identified to form H-bonds with E1 of GnRH. It is crucial to note that these interactions, while plausible, may not fully represent the native behaviour of pG1 in GnRH.

Considering these observations, this research suggests two potential scenarios: a) N-terminal residues such as pG1 and H2 are involved in N-terminus binding, elucidating why mutations affecting the N-terminus impact GnRH binding, as observed in this study; and b) pG1 and H2 of GnRH tilts backwards and position towards TM2, where they may form cation-π or H-bond interactions with K121^{3,32}, as previously proposed in computational studies[4]. While the first scenario aligns with the activating binding mode observed in this research, the second scenario has been suggested by computational studies cannot be dismissed, especially since a recent non-peptide agonist was found to form cation-π interactions with K121^{3,32} in MD trajectories [144].

5.6.3 Conformational stability of active binding mode

To evaluate the stability and dynamics of the active binding mode during MD simulations, PCA was employed for the specific timeframe corresponding to the active state of the receptor, spanning 100 ns (Figure 5.6).

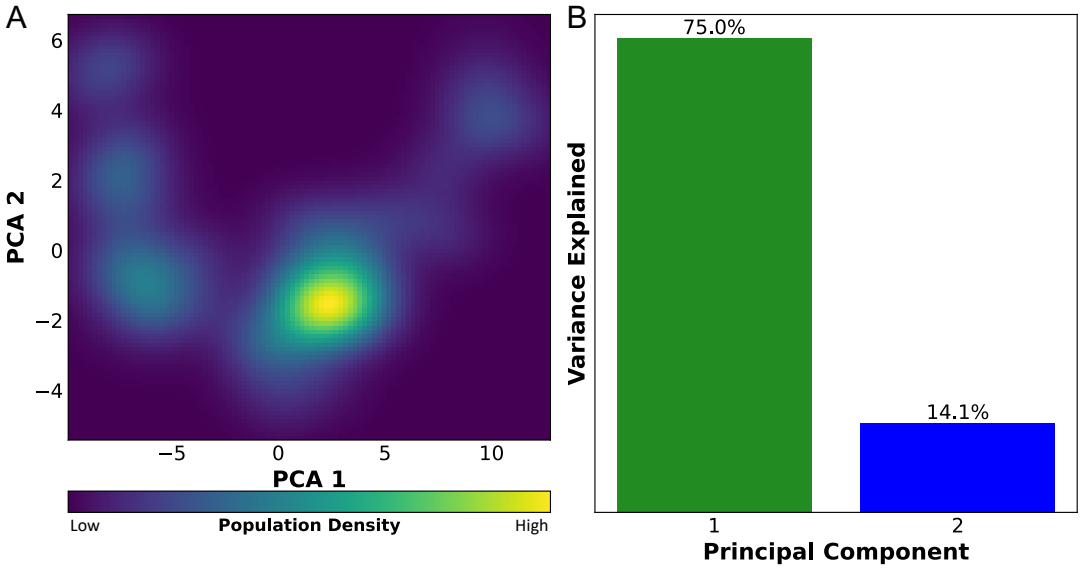


Figure 5.6: PCA of active binding mode (1.0-1.1 μs). A) Identification of a single high-population cluster, indicating a persistent GnRH binding conformation. B) Overview of the PCA variance explained.

The PCA results revealed a distinctive cluster capturing the binding mode, suggesting that the activating binding mode maintains a consistent conformation throughout the activation period, characterised by the interactions described earlier.

An in-depth examination of the variance explained by PCA highlighted that PCA1 accounted for a substantial 75% of the observed conformational variance, emphasising its pivotal role in describing the primary structural fluctuations within the active binding mode. Complementing this, PCA2 contributed 14.1% to the overall variance, offering additional perspectives on secondary structural dynamics within the system.

5.7 Conclusions

Investigation of the activation and binding mechanism of GnRH1R through a multi-faceted approach of MD simulations and structural analyses has revealed nuanced insights into the processes governing receptor activation and GnRH binding. The key findings of this chapter can be summarised as follows:

1. Rethinking R8 function and interactions:

Contrary to prior assumptions, the critical residue R8 of GnRH was found to engage in cation-π interactions with the conserved $W280^{6.48}$ of the CWxPY motif, challenging the previously proposed SB with $D302^{7.31}$.

2. Aromatic residue network:

A robust network of aromatic residues, including $Y290^{6.58}$, $Y283^{6.51}$, $F309^{7.38}$, and $H306^{7.36}$, plays a pivotal role in stabilising the binding mode and facilitating GnRH1R activation through π-π stacking interactions with W3 and Y5 of GnRH.

3. N-Terminus involvement:

Contrary to previous research, the N-terminal residues, particularly $N10^{N\text{-}ter}$, $S15^{N\text{-}ter}$, and $A16^{N\text{-}er}$, were found to form H-bonds with E1 of GnRH, indicating their direct involvement in binding. The study proposes two scenarios where N-terminal residues participate in N-terminus binding or form interactions with TM2, contributing to the overall understanding of the GnRH binding process.

4. Binding mode stability:

The activating binding mode exhibited remarkable stability during the activated portion of GnRH1R in the MD simulations, as evidenced by PCA, with PCA1 capturing a significant 75% of the conformational variance. This stability underscores the reliability of the identified activating binding mode and its persistence throughout the activation period.

5. Implications for drug design:

The refined understanding of the GnRH binding mechanism, particularly the unexpected interactions involving R8 and the aromatic residue network, provides valuable insights for drug design targeting this receptor. These findings can guide the development of more precise and effective therapeutics for reproductive system-related disorders.

Chapter 6 - GnRH1R activation

6.1 Rationale

The chapter focuses on elucidating the mechanism underlying GnRH1R activation. Through detailed computational analyses, the research aims to unravel the dynamic interactions and communication networks within the receptor, shedding light on the conformational changes and signalling events.

6.2 Aims and Objectives

1. To investigate communication networks: Explore and characterise the communication networks within GnRH1R activated by GnRH, with a specific focus on the communication pathways initiated by key residues such as R8.
2. To analyse structural dynamics: Examine the structural dynamics of GnRH1R during activation.
3. To compare the active and inactive GnRH1R conformations: Uncover conformation dependent differences between the two conformational states.
4. To uncover lipid-mediated communication: Explore potential lipid-mediated communication within GnRH1R, identifying residues participating in lipid interactions and their role in signal transmission and receptor activation.

6.3 Methodology

MD simulations were performed using the methodology detailed in Chapter 3, complemented by the following analyses:

- Ramachandran calculations were conducted for the trajectories using Python scripts (Appendix, Ramachadran.py).
- Backbone RMSD calculations were performed using the VMD plugin 'RMSD Trajectory Tool' and plotted using customised Pythons scripts (Appendix, RMSD.py).
- Backbone RMSF calculations were performed using customised Python scripts (Appendix, RMSF.py, Plot_RMSF.py) and comparison with experimental B-factors were performed using the published inactive crystal structure of the GnRH1R (PDBID:7BR3)[52] and customised Python scripts (Appendix, B_factor_rmsf.py).
- Pearson's correlation coefficient and the associated p-values were calculated using 'pearsonr' from the 'scipy.stats' Python library.

- Secondary structure evaluation was performed using the Define Secondary Structure of Proteins (DSSP) methodology through customised Python scripts (Appendix, DSSP.py).
- Salt bridge analysis was conducted using the VMD plugin 'Salt bridges' using a 3.5 Å N-O distance cutoff and the duration of each salt bridge was calculated and plotted through custom Python scripts (Appendix, SaltBridge.py).
- Hydrogen bonds between GnRH1R residues and POPC molecules were calculated using the VMD plugin 'Hydrogen bonds' and plotted using custom Python scripts (Appendix, Lipid_GnRH1R_hbonds.py). Hydrogen bonds and distances between specific receptor residues and lipids were calculated plotted using custom Python scripts (Appendix, F272_POPC_find_distance.py, Plot_F272_POPC_distance.py, R240_hbonds.py).
- Interactions including: π-π stacking, π-hydrogen, cation-π, vdW and hydrogen bonds were calculated using RING (https://ring.biocomputingup.it) and strict cutoffs (Chapter 3- 3.9.3, Figure 3.6).

Results

6.4 Structural quality of GnRH and GnRH1R

6.4.1 Ramachandran dihedral angles

Analysis of Ramachandran dihedral angles was performed for GnRH and GnRH1R during MD simulations for both ROS-1 and Apo-GnRH1R systems (Figure 6.1). Ramachandran analysis serves to assess the stability and conformational quality of the proteins during simulation. The Ramachandran plot, a graphical representation of the phi-psi space, delineates regions corresponding to 'allowed' and 'disallowed' conformations based on steric hindrance and clash considerations. Energetically favourable conformations reside in allowed regions, while disallowed regions signify sterically unfavourable interactions.

In the ROS-1 system, the GnRH residues consistently populate allowed regions throughout the simulation, underscoring the peptide's stability during simulation (Figure 6.1A). Conversely, for the GnRH1R in the ROS-1 system, while the majority of the analysed frames reside in allowed regions, there is a minor frame population in disallowed regions (Figure 6.1B).

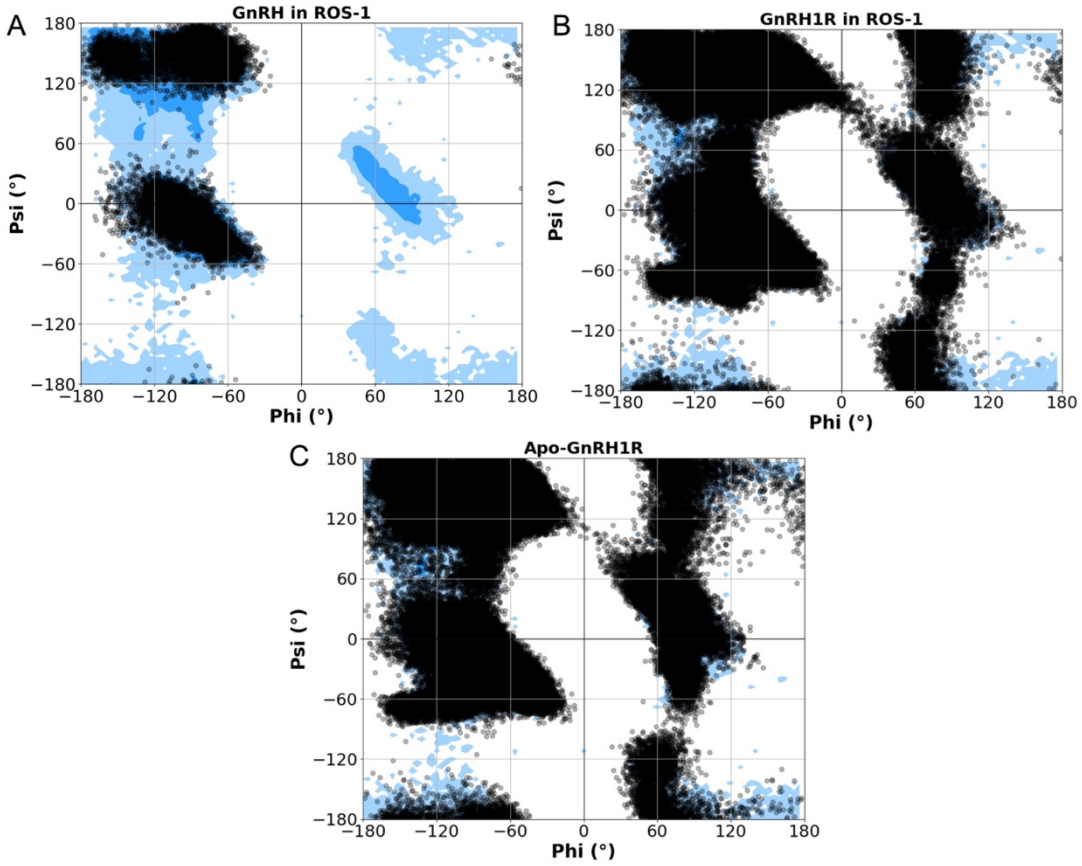


Figure 6.1: Ramachandran plots for A) GnRH residues and B) GnRH1R residues in ROS-1 simulation, and C) GnRH1R in the Apo System. The blue colour indicates allowed regions, while black points represent all residues across 2,500 frames at the 1.0-1.1 μs timeframe.

A similar pattern is observed in the Apo-GnRH1R system, albeit with fewer frames participating in disallowed regions compared to the ROS-1 GnRH1R (Figure 6.1C). This observation implies that the receptor, particularly in its inactive conformation, adopts a more favourable and stable conformation than when in its active state, a result aligned with expectations. Despite occasional frame populations in disallowed regions, the predominant localisation of frames within allowed regions in both systems signifies the overall high-quality conformation and stability of all proteins in the simulations.

6.4.2 Secondary structure analysis

Examination of secondary structures was conducted to assess the structural integrity of the receptor in the simulated systems (Figure 6.2). The analysis focuses on the percentage distribution of secondary structure elements using the Define Secondary Structure of Proteins (DSSP) methodology. This included the identification of irregular elements, beta-

strands (B), extended strands (E), 3-10 helices (G), alpha helices (H), 5-(pi) helices (I), bends (S) and turns (T) in the GnRH1R. Maintaining the structural integrity of the receptor during simulations is crucial, as the resulting structure forms the basis for reliable conclusions.

The comparative assessment of the active GnRH1R in ROS-1 and Apo-GnRH1R revealed a commendable preservation of structural integrity in both simulation systems. However, subtle differences were observed, such as an increased percentage of alpha helices, bends, and extended strands in the Apo-GnRH1R, suggesting a slightly more structured inactive conformation compared to the active ROS-1 receptor. This observation is supported by a decrease in the percentage of irregular elements and loops in the DSSP analysis for the Apo-GnRH1R.

Significantly, a notable increase in pi-helices was observed in the active ROS-1, registering at 4.11%, in contrast to the 1.8% presence in the Apo-GnRH1R. Pi-helices, a distinctive helical conformation, though less common than alpha helices and 3/10 helices, showed a pronounced increase in the active ROS-1 configuration. Finally, hydrogen bonded turns also appear increased in the active conformation yielding a 4.2% in comparison with the 3.8% in the inactive conformation (Figure 6.2).

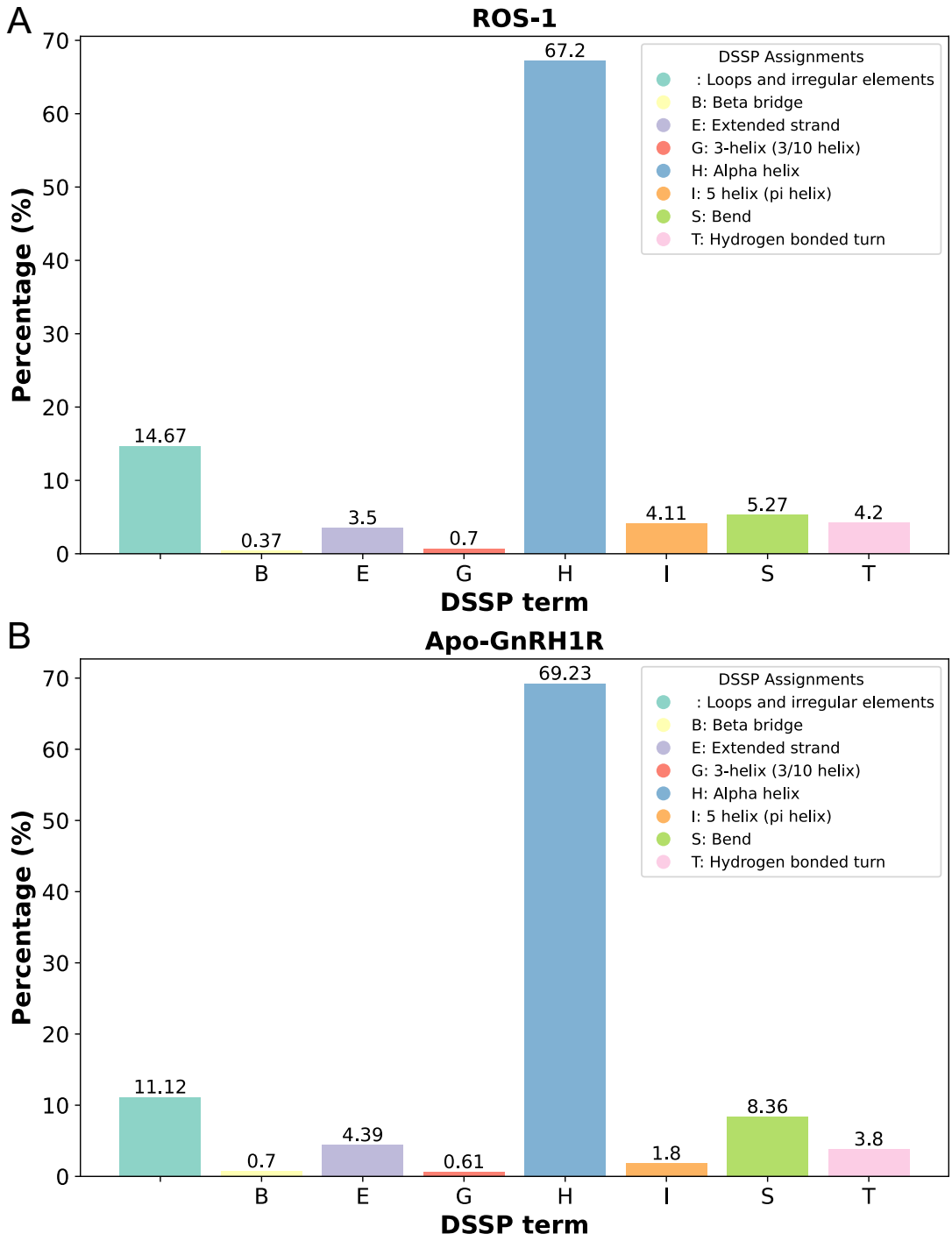


Figure 6.2: DSSP analysis for GnRH1R in A) ROS-1, and B) Apo-GnRH1R simulations. DSSP was conducted for the last 50 ns of the trajectories (1.05-1.1 μs).

Overall, the active and inactive states of the receptor present minor differences in terms of secondary structure elements and both retain their original characteristics and integrity to a high degree.

6.5 Trajectory analysis

6.5.1 RMSD calculations

To assess the time required for the backbone equilibration of GnRH1R, RMSD calculations were conducted throughout the simulation period for both ROS-1 and Apo-GnRH1R systems (Figure 6.3). The 7TM bundle of the receptor demonstrated swift equilibration within 100-200 ns, displaying an RMSD difference of approximately 2 Å from the initial system (reference) for both systems. Notably, the N-terminus, particularly the 100% AlphaFold-predicted residues 1-17 (Chapter 3 - Figure 3.1B), emerged as the most unstable sequence.

For the ROS-1 system, this poorly predicted sequence reached equilibration around 500 ns, with an RMSD of approximately 6 Å from its initial configuration (Figure 6.3A). Conversely, the 18-32 residues of the N-terminus, reported with the crystal structure, exhibited rearrangement but equilibrated promptly with an RMSD of approximately 3 Å from the initial position. In the Apo-GnRH1R system, a comparable pattern was observed for the 7TM bundle and the 18-33 N-terminal residues (Figure 6.3B).

A distinctive contrast between the ROS-1 and Apo-GnRH1R systems was evident in the behaviour of the first 17 residues of the N-terminus. While this sequence achieved equilibration after approximately 500 ns in ROS-1, equilibration was not realised in the Apo-GnRH1R. This observation suggests that residues 1-17 of the N-terminus necessitate the presence of a ligand to 'lock' into a specific position and conformation. In the absence of a ligand, this sequence remains in a fluctuating conformation, indicating the dependence of its stability on the ligand interaction.

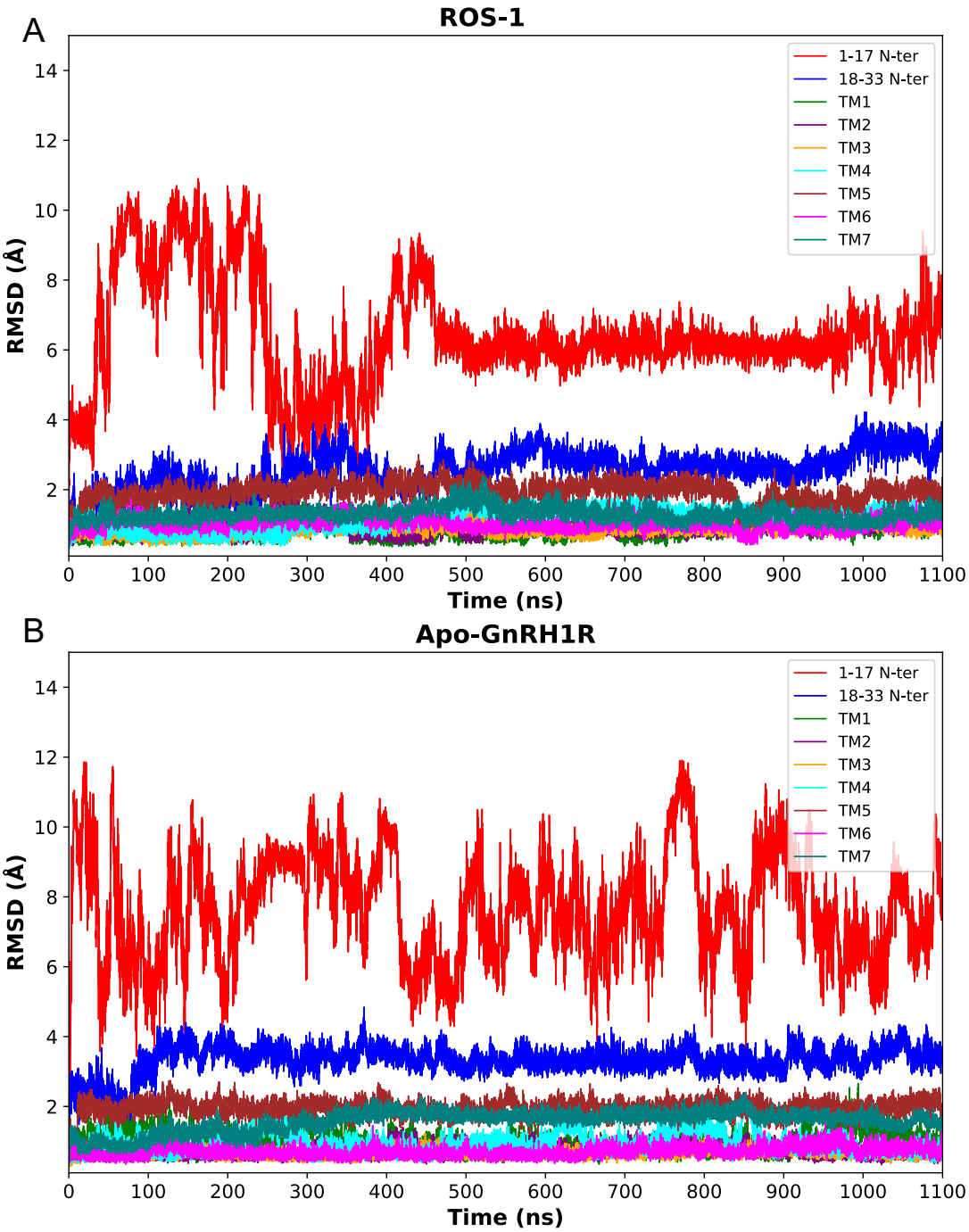


Figure 6.3: Time evolution of RMSD values for A) ROS-1 and B) Apo-GnRH1R systems.

PCA was employed to further investigate the behaviour of the initial seventeen residues of the N-terminus in both the ROS-1 GnRH1R and Apo-GnRH1R systems (Figure 6.4). This analysis was conducted over the timeframe of 0.5-1.1 μs, as backbone equilibration of the 1-17 N-terminal residues is observed to occur at 0.5 μs for the ROS-1 GnRH1R system.

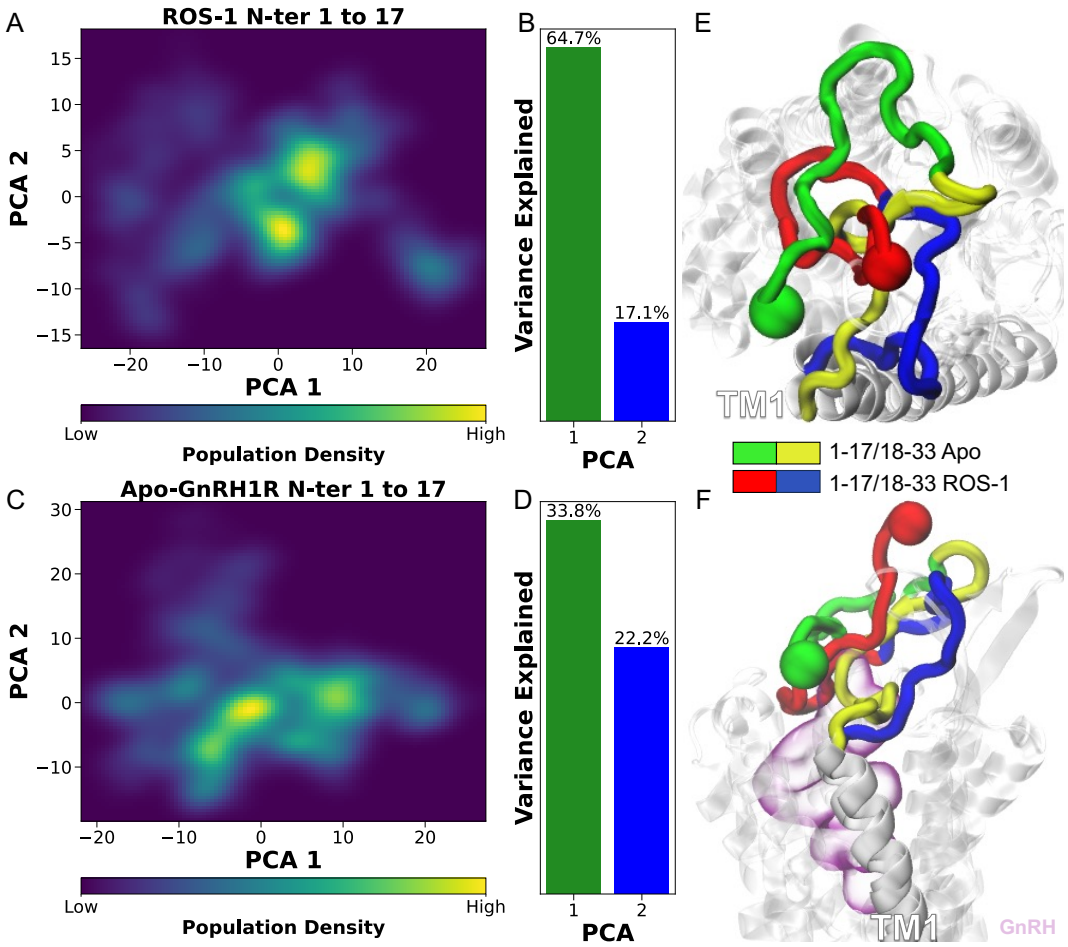


Figure 6.4: PCA of N-Terminus residues 1 to 17 and variance explained for A), B) ROS-1 and C), D) Apo-GnRH1R systems. N-Terminus is visualised from the extracellular space in E) and side view in F). In Apo-GnRH, the 1-17 and 18-33 N-terminal residues are depicted in green and yellow, respectively, while in ROS-1, they are represented in red and blue colours. GnRH is displayed in pink in F) for context and has been omitted from E) for clarity. PCA was conducted for the timeframe of 0.5-1.1 μs in both systems.

As anticipated, the ROS-1 N-terminal exhibited a markedly superior degree of stability, evidenced by the presence of two distinct clusters of the 1-17 N-terminus, each representing a unique conformation (Figure 6.4A). In contrast, the Apo-GnRH1R systems displayed a more heterogeneous behaviour, characterised by a multitude of clusters, with a high degree of population dispersion across the PCA space (Figure 6.4C). This finding is indicative of a greater degree of conformational flexibility and structural heterogeneity within the Apo-GnRH1R system, potentially attributable to the absence of the GnRH.

Quantitatively, the variance explained by the principal components accounted for a substantial 64.7% and 17.1% for PCA1 and PCA2, respectively, in the case of the ROS-1 system (Figure 6.4B). This underscores the ability of

the PCA1 and PCA2 components to capture a significant portion of the conformational variability within the ROS-1 GnRH1R system. However, the same principal components failed to encapsulate the conformational dynamics of the N-terminus in the Apo-GnRH1R system, accounting for a mere 33.8% and 22.2% of the variance for PCA1 and PCA2, respectively (Figure 6.4D). This suggests that the conformational landscape of the Apo-GnRH1R system is characterised by a greater degree of complexity, necessitating the inclusion of additional principal components to adequately capture its structural heterogeneity.

6.5.2 RMSF calculations

RMSF calculations were performed for the active and inactive GnRH1R conformations to analyse the flexibility and dynamic behaviour of different regions within the receptor (Figure 6.5). The comparison of the active and inactive states of the receptor provides insights into the conformational changes associated with receptor activation and inactivation.

Comparison of the RMSF values, revealed that the active conformation is less flexible and less prone to conformational changes than the inactive conformation which agrees with similar studies on the active and inactive conformations of the cannabinoid receptor [145]. Comparison of the B-factors of the inactive crystal structure of the GnRH1R with the RMSF of the inactive conformation during MD simulations (normalised values) reveals a positive correlation in terms of Pearson's coefficient (0.49) (Figure 6.5A). This suggests that the flexibility of the Apo-GnRH1R observed in MD simulation agrees with the flexibility observed in the experimental crystal structure and provides supporting evidence for the credibility of the RMSF results obtained from the simulations (Figure 6.5B).

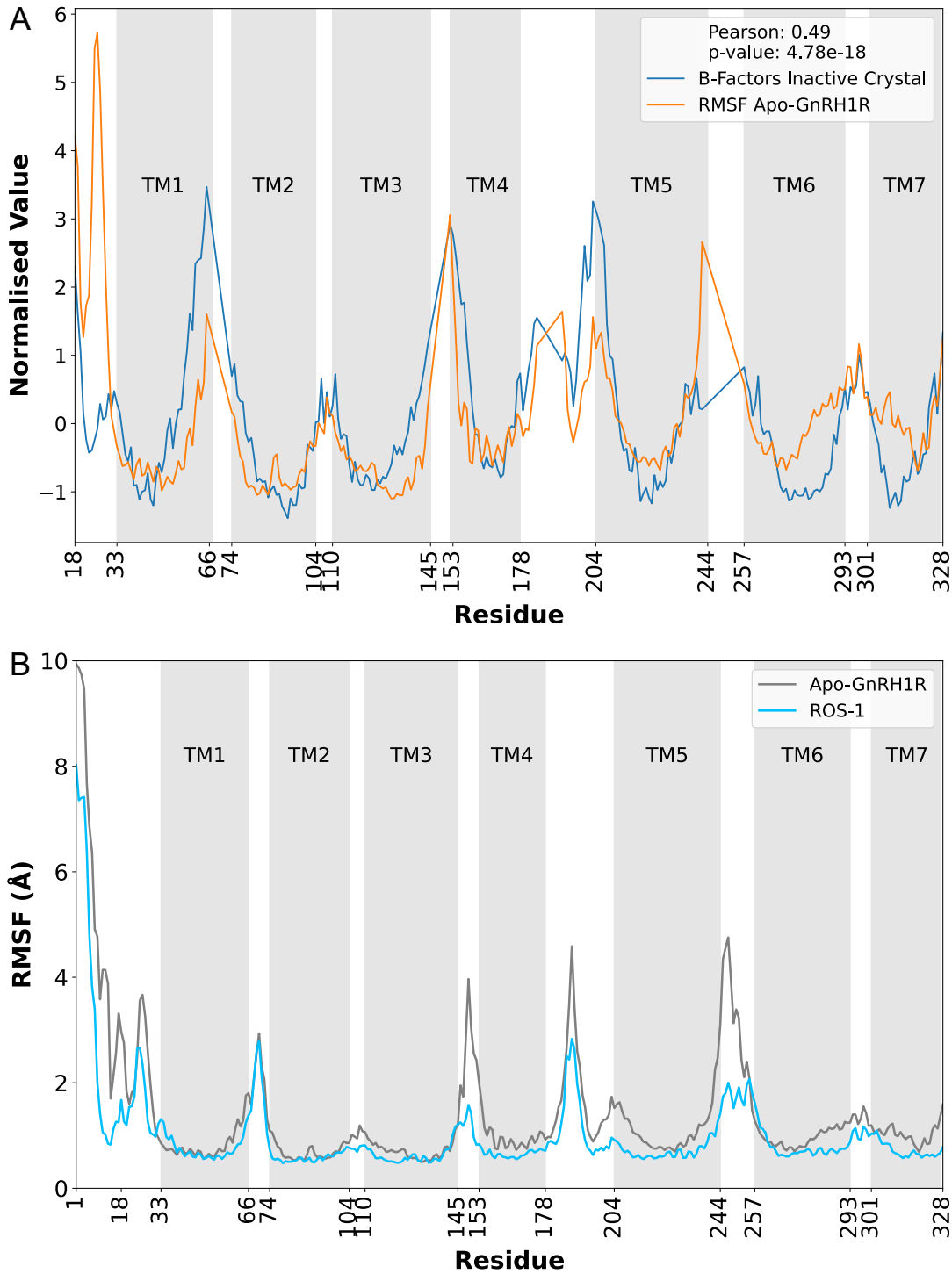


Figure 6.5: Analysis of RMSF. A) Normalised B-factors of the GnRH1R crystal structure (blue) and RMSF of the Apo-GnRH1R system (orange) with Pearson's correlation coefficient and associated p-value presented in the legend. B) RMSF comparison between the Apo-GnRH1R (grey) in the inactive conformation and ROS-1 (blue) in the active conformation.

6.5.3 Salt bridge analysis

Salt bridge analysis was undertaken throughout the simulations for both the active and inactive conformations of GnRH1R, aiming to unravel the

dynamics of ionic interactions and their dependence on the receptor's conformation (Figure 6.6).

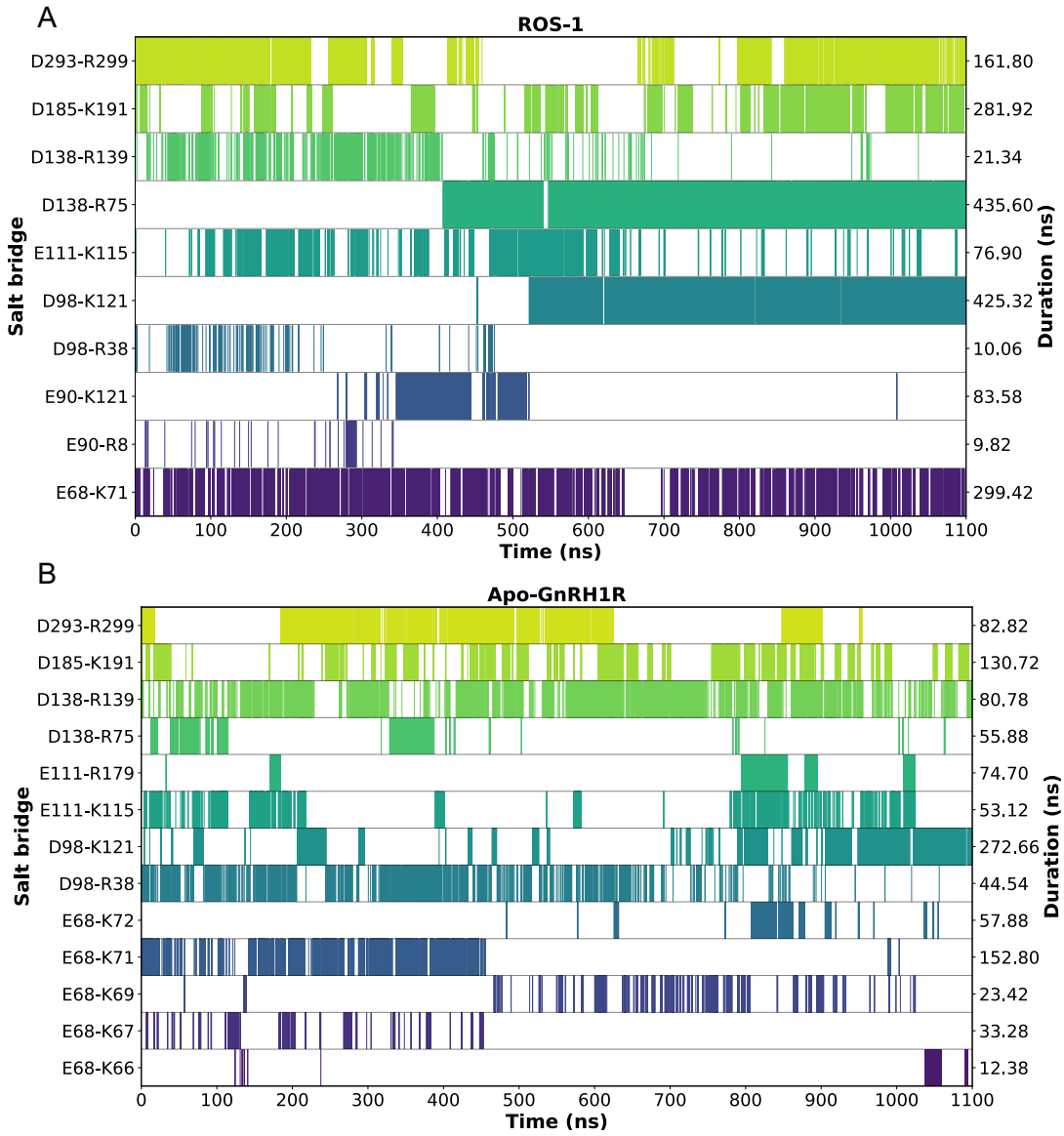


Figure 6.6: Analysis of salt bridge formation and evolution in A) ROS-1 and B) Apo-GnRH1R systems throughout the simulation.

In the ROS-1 system, a total of 10 ionic locks were identified, with the most notable SB observed between D98$^{2.60}$ and K121$^{3.32}$, lasting for an impressive 435.60 ns. Remarkably, this ionic lock is established subsequent to the dissociation of both D98$^{2.60}$-R38$^{1.35}$ and E90$^{2.53}$-K121$^{3.32}$ ionic locks, aligning with prior findings that support the conformational dependency of the D98$^{2.60}$-K121$^{3.32}$ SB [52], [73], indicating its association with the active conformation (Figure 6.6A).

In contrast, the D98$^{2.60}$-K121$^{3.32}$ SB in the Apo-GnRH1R system is less prominent but sporadically formed, particularly beyond 700 ns, with a duration of 272.66 ns (Figure 6.6B). The D98$^{2.60}$-R38$^{1.35}$ SB in the Apo-GnRH1R

exerts a more substantial influence, persisting for a total duration of 44.54 ns compared to 10 ns in ROS-1, with the residues in close and consistent interaction throughout the simulation. In agreement with prior research, the E90$^{2.53}$-K121$^{3.32}$ SB did not form in the Apo-GnRH1R system[73]. The second most significant ionic lock in the active ROS-1 system is the D138$^{3.49}$-R75$^{2.38}$ SB, lasting 435.60 ns. This SB gains intensity after 400 ns, while in the Apo-GnRH1R, the D138$^{3.49}$-R139$^{3.50}$ ionic lock remains stable throughout the simulation, with the D138$^{3.49}$-R75$^{2.38}$ SB forming briefly at the beginning and around 400 ns. This conformation-dependent observation suggests that the R75$^{2.38}$-D138$^{3.49}$ SB may contribute to orienting TM3 closer to TM2 and stabilising the active conformation of TM3.

The third most notable ionic lock forms between E68^{ICL1}-K71$^{12.48(ICL1)}$ in ICL1, with a duration of 299.42 ns and a stable presence throughout the 1.1 μs of simulation in ROS-1. In the Apo-GnRH1R, this ionic lock is observed for half the duration of ROS-1 and completely dissociates after 450 ns.

Of particular interest is the residue K121$^{3.32}$ which exhibits intriguing dynamics. In the Apo-GnRH1R, K121$^{3.32}$ forms only one SB with D98$^{2.60}$, while in the presence of GnRH, K121$^{3.32}$ undergoes rearrangement, forming ionic locks with E90$^{2.53}$ before stabilising with D98$^{2.60}$. This highlights the crucial role of K121$^{3.32}$ in the plasticity and reorganisation of the binding pocket to accommodate diverse ligands. Although K121$^{3.32}$ does not directly interact with GnRH in ROS-1, it plays a prominent role in stabilising the extracellular portions of TM3 and TM2, forming the outer wall of the binding pocket.

6.6 Active receptor characteristics

6.6.1 TM3-TM6 distance

The prevailing activation mechanism observed across Class A GPCRs involves an increase in the TM3-TM6 distance and the disruption of the ionic lock between residues typically found in positions 3.50 and 6.30, often featuring arginine and glutamic acid respectfully[146]. However, in GnRH1R, the conserved arginine in position 3.50, R139, forms a polar interaction with T265 in position 6.33 instead. Consequently, monitoring of TM3-TM6 distances based on these residues revealed activation at approximately 1.0 μs in the ROS-1 system. To establish a baseline, the inactive TM3-TM6 distances of the crystal structure and AlphaFold-predicted active models were compared for both conformational states (Figure 6.7C-D).

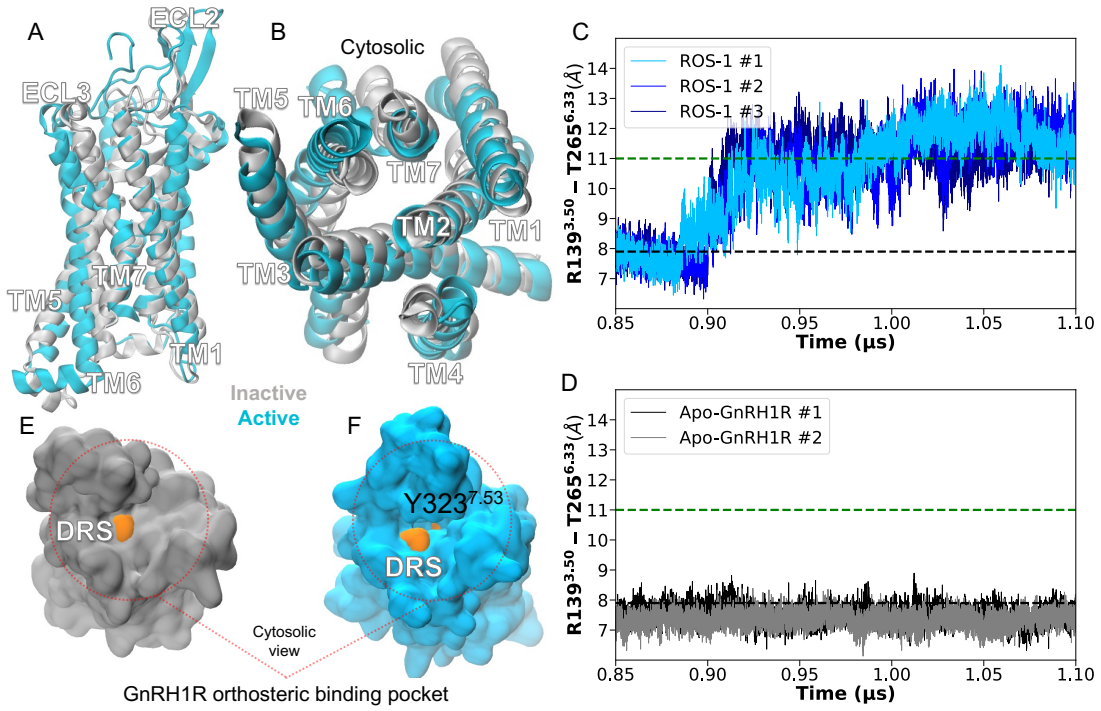


Figure 6.7: GnRH1R activation dynamics. A) and B) Side and cytosolic views of GnRH1R, illustrating the conformational change of TM3. Traces of TM3-TM6 distances over time for C) the active ROS-1 and D) the inactive Apo-GnRH1R systems. E) and F) Represent the cytosolic site of the receptor in its E) inactive Apo-GnRH1R and F) active ROS-1 conformations. The DRS is depicted in orange, illustrating a closed Gα pocket in the inactive conformation and an exposed GnRH1R orthosteric binding pocket where both DRS and $Y323^{7.53}$ are visible in the active conformation.

The distance analysis unveiled an ~5 Å increase in the distance between TM3 and TM6 during activation, a consistent observation with distances reported in other GPCRs. Intriguingly, the active conformation of GnRH1R through MD simulations displayed a TM3-TM6 distance approximately 1 Å higher than the AlphaFold-predicted active TM3-TM6 distance (Figure 6.7C). This suggests that the predicted structure may lack full equilibration, a finding also noted in a previous study where the active-predicted structure inactivated during MD simulation with a non-peptide agonist[144]. Notably, an evident outward movement of TM6 was observed in the active ROS-1 conformation, stabilising at ~ 12.5 Å with respect to TM3, while the inactive conformation maintained ~ 7.5 Å distance from TM3 (Figure 6.7A-D). Additionally, a slight inward movement of TM7, consistent with the general behaviour of TM7 in active Class A GPCRs, was observed (Figure 6.7B).

Furthermore, a slight outward movement of the cytosolic end of TM3 was noted in the active conformation (Figure 6.7B). Notably, a lateral rearrangement of the cytosolic end of TM4 occurred, shifting away from TM3

and towards TM2, while the cytosolic ends of TM1 and TM2 and TM5 remained unchanged (Figure 6.7B).

As anticipated, the activation of GnRH1R leads to the opening of the orthosteric pocket, encompassing the DRS motif and cytosolic TM7, while it remains closed and occupied by TM6 in the inactive conformation (Figure 6.7E, F).

The conserved residue P282$^{6.50}$ of the CWxPY motif serves as a pivotal point for the outwards bending of TM6 upon activation, a phenomenon observed in numerous GPCRs[147]. It is believed to function as a hinge for the bending due to the absence of H-bonds between its backbone nitrogen atom and the carbonyl group of the residue one helical turn above[147]. The function of P282$^{6.50}$ in GnRH1R is underscored by the findings of this research as well, wherein the outwards movement of TM6 initiates following this conserved residue (Figure 6.8).

Throughout the simulation, the ICLs of GnRH1R maintain stability, except for ICL3. Notably, ICL3 exhibits an outwards turn and becomes a parallel extension of TM5 and TM6 during GnRH1R activation. The substantial rearrangement of ICL3 aligns with previously studies observing outward movements of ICL3[28]. The reorganisation of ICL3 appears to resemble conformations observed in the C-terminal Helix 8 of other GPCRs, an element absent in GnRH1R. Thus, ICL3 may potentially have a role in enlarging the interacting area for G-protein coupling and potentially β-arrestin coupling in the absence of a Helix 8 tail.

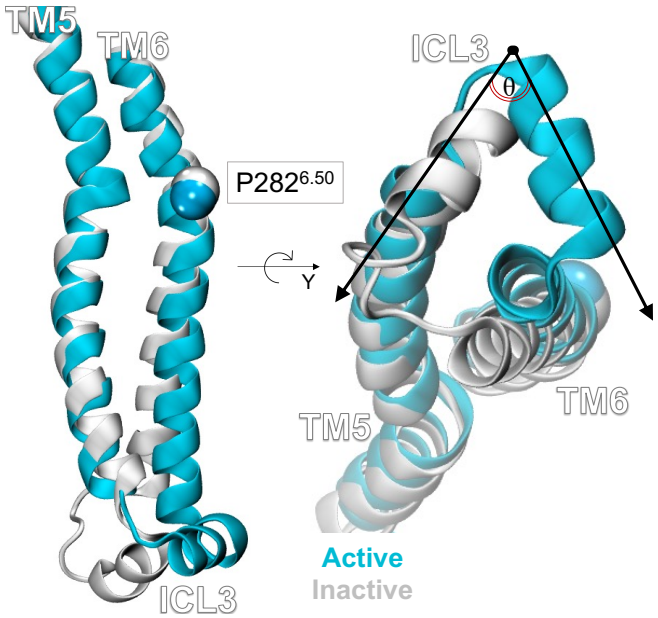


Figure 6.8: Conformation of ICL3 and TM6 kink following P282$^{6.50}$ in the active conformation (blue). ICL3 and TM6 conformation of the inactive conformation is presented in white colour.

6.6.2 CWxPY – DPxxY motifs and water mediated communication

GnRH1R exhibits a negatively charged DPxxY motif, deviating from the conserved polar NPxxY motif found in other GPCRs. While the widely conserved CWxPY motif in GPCRs is known to interact with ligands, this study affirms that GnRH engages with all residues of the CWxPY motif. Notably, the interaction involves cation-π interactions between R8 and W280$^{6.48}$, as well as various π-π stacking interactions, including those between W3, Y5, and Y283$^{6.51}$, among others (Chapter 5 - Figure 5.3).

An intriguing and novel pattern emerges, revealing communication between the CWxPY motif, GnRH, and the DPxxY motif through water-mediated H-bonds (Figure 6.9). Specifically, R8 forms H-bonds either directly with E90$^{2.53}$ or through a common water molecule. This water molecule also engages in interactions with N87$^{2.50}$ through H-bonds, supporting the proposed role of N87$^{2.50}$ in water-mediated networks observed in other GPCRs[40], [41], [52], [53]. In turn, N87$^{2.50}$ establishes water-mediated H-bonds with either D319$^{7.49}$ or Y323$^{7.53}$ of the DPxxY motif, concluding the communication between GnRH-CWxPY and DPxxY.

This observation elucidates the necessity for arginine in position 8 in GnRH-peptide agonist binding and GnRH1R activation. Beyond its favourable cation-π interactions with W280$^{6.48}$, the long and charged side chain of R8 participates in the water-mediated network and signal transmission between TM6, TM2, and TM7.

Notably, whilst the active GnRH1R conformation forms a water mediated network from R8 of GnRH to the CWxPY, N87$^{2.50}$ and to the DPxxY, the inactive conformation possesses a Na^+ atom close to D319$^{7.49}$ of the DPxxY motif.

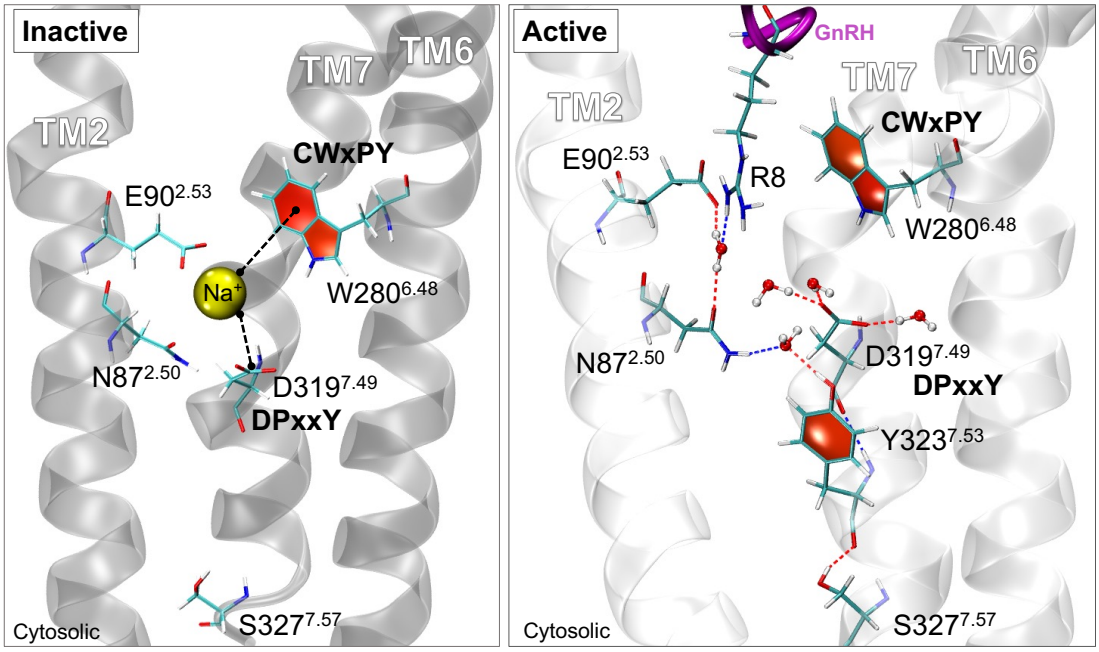


Figure 6.9: Left: Sodium (Na^+) binding pocket at the DPxxY motif. Right: Water mediated network between GnRH/CWxPY and $E90^{2.53}$/$N87^{2.50}$ and DPxxY motif.

Monitoring the distance between $D319^{7.49}$ and the sodium atom in the Apo-GnRH1R simulation, it is evident that the sodium atom inserts to the interhelical space early in the simulation and remains in ~2.5 Å distance from the negatively charge $D319^{7.49}$ (Figure 6.10B). Additionally, the same sodium atom retains an ~7 Å distance from the aliphatic ring of $W280^{6.48}$ in the CWxPY motif. This observation confirms the DPxxY motif of the GnRH1R as a sodium binding site and it additionally shows that the binding of sodium is a characteristic of the inactive conformation, whilst the positively charged R8 of GnRH stabilises the charged interactions in the active conformation.

Additionally, the sodium ion is not capable of mediating signal to $N87^{2.50}$ whilst R8 mediates communicates through water with $N87^{2.50}$ and $E90^{2.53}$ as well as through vdW interactions with $M125^{3.36}$ which in turn communicates with $A129^{3.40}$ of the PAF motif and the G-protein binding pocket (DRS motif) (Figure 6.15). Calculation of the hydrogen bonds formed between water molecules and $N87^{2.50}$ in the active ROS-1 and inactive Apo-GnRH1R simulations revealed that $N87^{2.50}$ forms one less hydrogen bond with water in the inactive conformation than it does in the active (Figure 6.10A). This difference is potentially attributed to the hydrogen bond mediated between R8 of GnRH and $N87^{2.50}$ in the active GnRH1R.

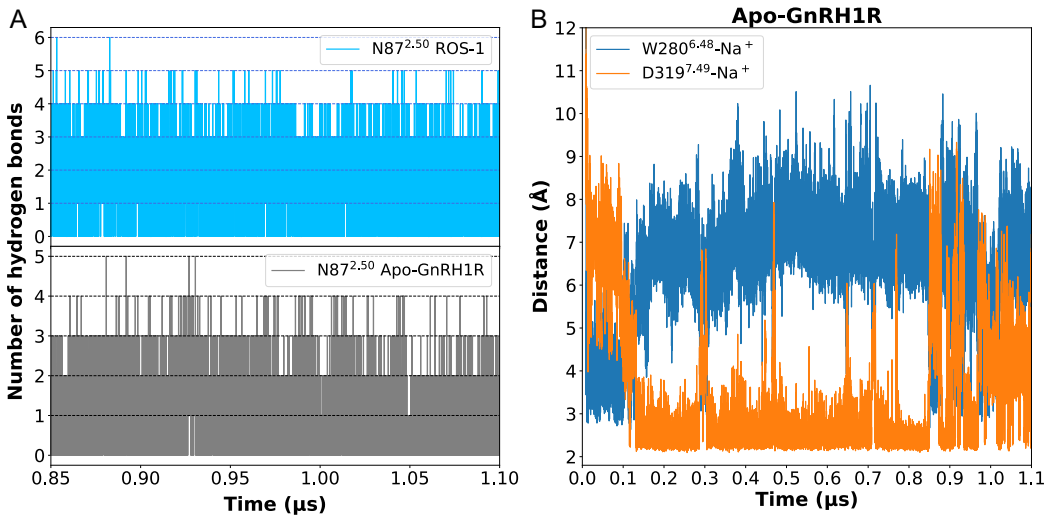


Figure 6.10: A) Hydrogen bonds between N87$^{2.50}$ and water molecules in the ROS-1 and Apo-GnRH1R simulations. B) Distance of W280$^{6.48}$ and D319$^{7.49}$ with the sodium atom.

6.6.3 TM1-TM7 communication upon activation

The absence of a C-terminal H8 helix in GnRH1R results in the tight packing of TM1 and TM7 as evident by the interactions of cytosolic residues F56$^{1.53}$ and W63$^{1.60}$ with Y323$^{7.53}$ of the DPxxY motif in the inactive conformation[52]. However, as TM7 experiences slight inward movement in the active conformation, the π-π interactions among the triad Y323$^{7.53}$-F56$^{1.53}$-W63$^{1.60}$ break, while the Y323$^{7.53}$-F56$^{1.53}$ interaction remains stable (Figure 6.11, Figure 6.12). Upon activation, the rotation of Y323$^{7.53}$ and the rearrangement of its aromatic residue towards the centre of the intrahelical bundle aligns with the observed behaviour of this residue in other active GPCRs[44], [146].

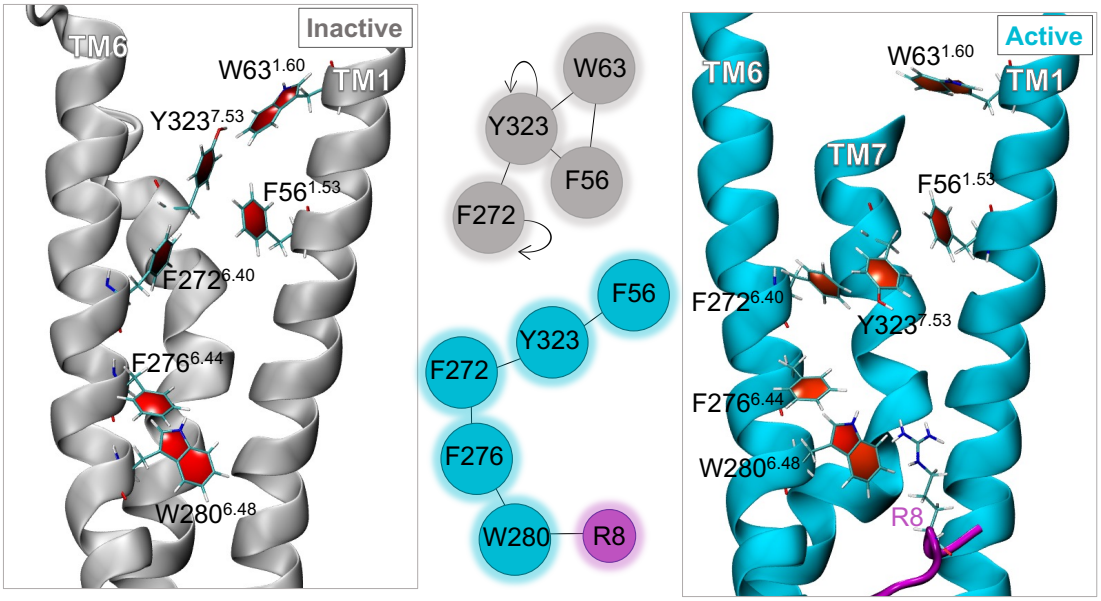


Figure 6.11: Communication pathway between TM1-TM7-TM6 in the inactive (grey) and active (blue) conformations. In the inactive conformation, $Y323^{7.53}$ of the DPxxY interacts with TM1 residues $F56^{1.53}$ and $W63^{1.60}$. In the active conformation (blue), the pathway to CWxPY is open, where $Y323^{7.53}$ communicates with $W280^{6.48}$/R8 through $F272^{6.40}$ and $F276^{6.40}$ of the $P^{5.50}A^{3.40}F^{6.40}$ motif. The contact map for π-π interactions of the inactive GnRH1R is presented in the Appendix, Figure 8.4 and for the active GnRH1R in Figure 6.12.

In the active conformation, $Y323^{7.53}$ forms stable π-π interactions with $F272^{6.40}$. It is noteworthy that GnRH1R belongs to the minority of Class A GPCRs that possess a large aromatic residue in position 6.40 instead of a small hydrophobic [52]. The functionality of $F272^{6.40}$ is currently unknown, although this residue has been revealed to be important for activation as shown by mutagenesis studies[52]. Here, the function of $F272^{6.40}$ is shown to be mediating interactions between $Y323^{7.53}$ (DPxxY motif), $F276^{6.44}$ (PAF motif) and subsequently $W280^{6.48}$ (CWxPY motif) in the active conformation (Figure 6.11, Figure 6.12). This observation explains why mutation of $F272^{6.40}$ affects receptor activation and its association with GnRH1R activation by previous studies[52].

The unique $F272^{6.40}$ residue of the GnRH1R is located one helical turn from $F276^{6.44}$, and these residues establish π-π interactions that persist throughout the simulation (Figure 6.12). Additionally, $F276^{6.44}$ establishes stable π-π interactions with $W280^{6.48}$ of the CWxPY motif (Figure 6.12). Notably, $W280^{6.48}$ engages in cation-π interactions with R8 of GnRH, highlighting the intricate network of aromatic interactions involved in the active conformation of GnRH1R.

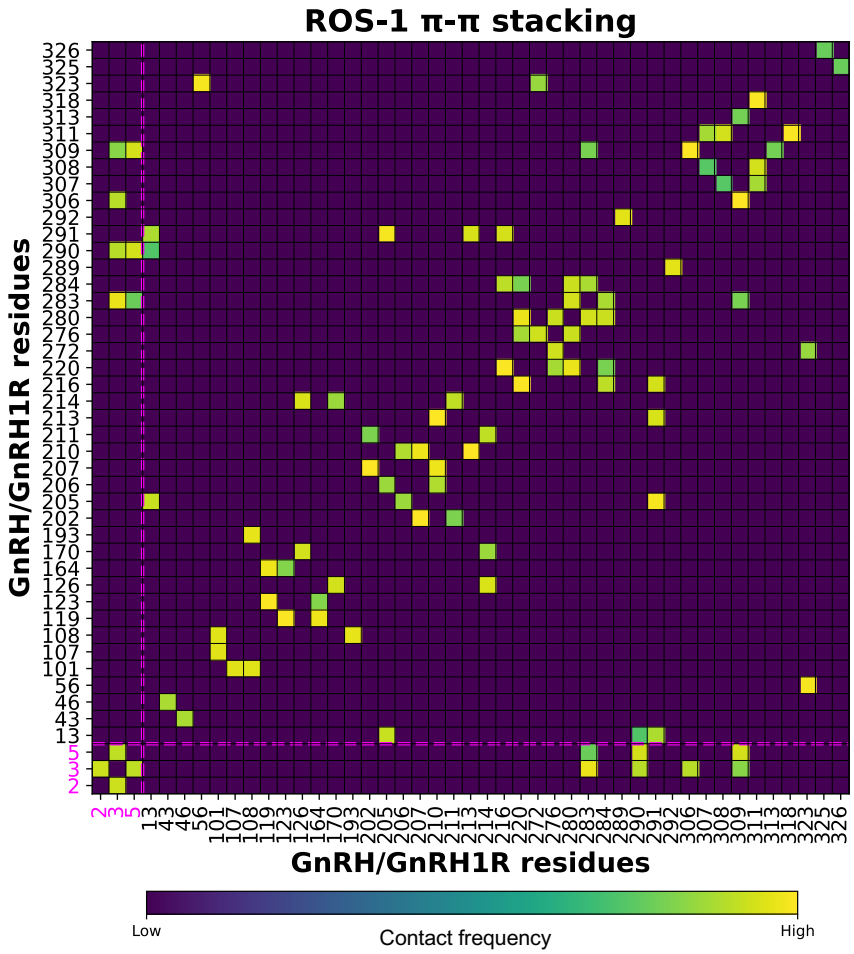


Figure 6.12: π-π stacking interaction analysis in the ROS-1 system throughout the activated simulation. Pink numbers and lines represent GnRH residues and its intramolecular or intermolecular interactions. Calculations conducted using RING (https://ring.biocomputingup.it).

During the activation process, the focus of the Y323$^{7.53}$ residue from the DPxxY motif shifts from TM1 to TM6. The interaction between W63$^{1.60}$, F56$^{1.53}$, and Y323$^{7.53}$ is lost, and a new interaction between F56$^{1.53}$, Y323$^{7.53}$, and F272$^{6.40}$ is formed. Communication among the GnRH-CWxPY-DPxxY motifs is established through R8-W280$^{6.48}$ and transmitted through F272$^{6.40}$-F276$^{6.44}$ to Y323$^{7.53}$ at the cytosolic end of TM7.

6.6.4 PAF motif and membrane communication

The PIF motif in GPCRs, while not highly conserved, has been proposed to play a role in facilitating the outward movement of the cytosolic end of TM6. The GnRH1R features the P$^{5.50}$A$^{3.40}$F$^{6.44}$ motif, comprising P223$^{5.50}$-A129$^{3.40}$-F276$^{6.44}$. Upon comparing the PAF motif between the active and inactive conformations, no significant differences were observed, except for a slight rotation of F276$^{6.44}$ towards TM5 and the membrane environment in the active conformation (Figure 6.13A).

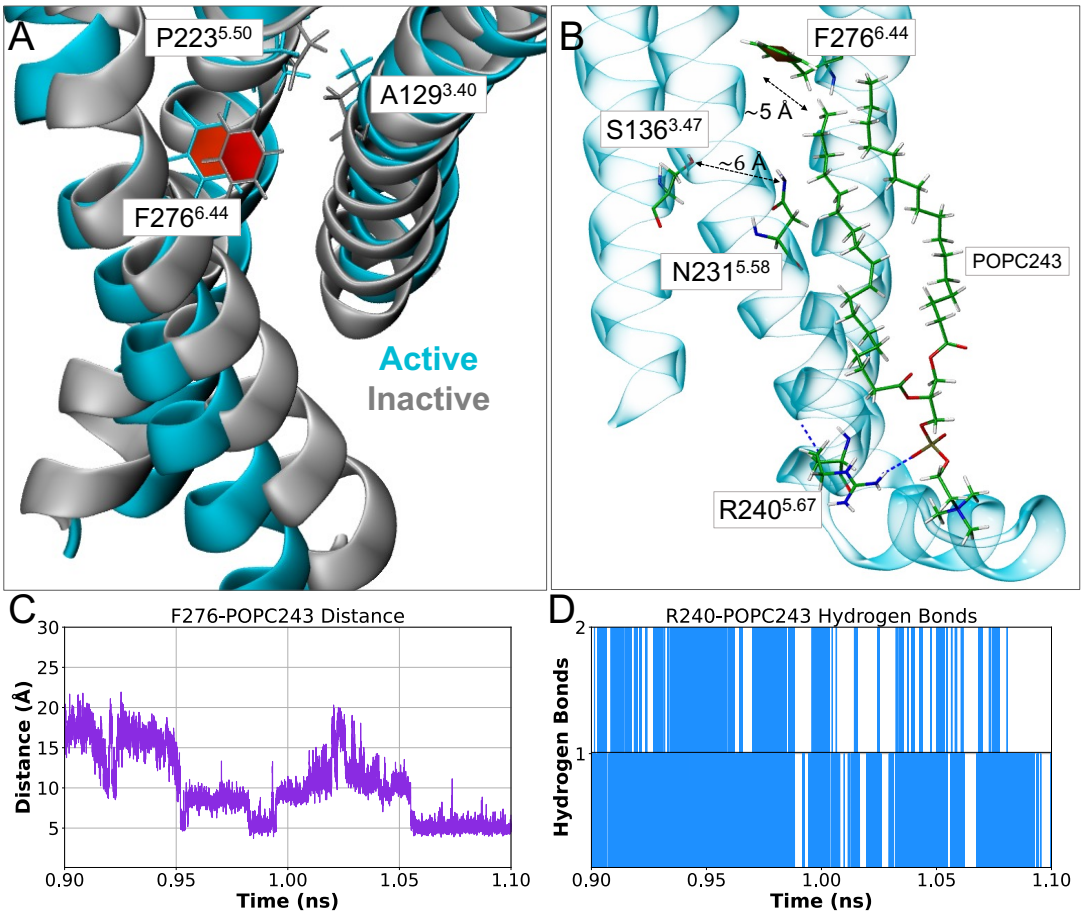


Figure 6.13: A) Comparison of the PAF motif in the inactive Apo-GnRH1R (grey) and active ROS-1 (blue). B) Communication of F276$^{6.44}$ with R240$^{5.67}$ through a common lipid molecule (POPC243). C) Distance between F276$^{6.44}$ and the aryl tail of POPC243. D) Total number of H-bonds formed between R240$^{5.67}$ and POPC243 over time.

Notably, F276$^{6.44}$ was observed within 5 Å of a lipid tail after activation, which in turn interacts consistently with R240$^{5.67}$ at the cytosolic end of TM5 through H-bonds (Figure 6.13B,C). This proximity suggests a potential role for F276$^{6.44}$ in mediating signal through the membrane and stabilising the cytosolic movement of ICL3 and TM6. Additionally, the same lipid molecule was found close to N231$^{5.58}$. In 96% of Class A GPCRs, the residue in position 5.58 is a tyrosine and it has been shown to be important for receptor activation (Chapter 1-1.4). However, the GnRH1R consists of a polar asparagine in position 5.58. In the inactive GnRH1R, N231$^{5.58}$ forms a tight polar interaction with S136$^{3.47}$ which stabilises the interhelical distance of TM3 and TM5 to 4 Å[52]. In the ROS-1 active GnRH1R conformation, N231$^{5.58}$ was found at a 6 Å distance from S136$^{3.47}$ and additionally in close proximity with the same lipid that forms vdW interactions with F276$^{6.44}$ of the PAF motif and hydrogen bonds with R240$^{5.67}$ of ICL3(Figure 6.13B,D). This result suggests that the

unique N231$^{5.58}$ residue of the GnRH1R may be involved in receptor activation though membrane communication.

6.6.5 Membrane communication

In computational studies of GPCRs, the interactions between the receptor and its surrounding lipid environment play a crucial role in determining the protein's structure, dynamics, and function. Among the various lipid types, POPC membranes are commonly used due to their well-defined properties and the stability they offer during MD simulations. While POPC lipids provide a simplified representation of the more complex native cellular membranes, they nonetheless serve as a practical model for studying GPCR-lipid interactions. The use of POPC allows for a balanced investigation of protein behaviour within a membrane environment without the need to account for the high complexity of native membranes, which are composed of multiple lipid species, cholesterol, and membrane-associated proteins. Although POPC membranes lack the heterogeneity of native systems, they offer a reasonable approximation for examining fundamental aspects of lipid-protein interactions.

Interactions between specific amino acid residues of the protein and the surrounding POPC lipids can significantly influence the receptor's conformational states and activity. These interactions involve the lipid head groups, acyl chains, and glycerol backbone, which may form transient or stable contacts with residues located in the transmembrane helices and extracellular domains of the GPCR. Understanding these interactions is essential for deciphering how the lipid environment modulates receptor function and how POPC lipids, in particular, contribute to maintaining the structural integrity of GPCRs in simulations.

To investigate the role of lipids in GnRH1R activation, lipid interactions were monitored for the duration of activation, with a focus on H-bonds (Figure 6.14). Notably, it is observed that the head polar groups of POPC lipids form high-occupancy H-bonds with polar residues located at the ECLs and ICLs of the receptor. The network of H-bond interactions between the membrane and GnRH1R are mediated mostly through serine, arginine, and lysine residues.

On the extracellular side of the receptor, high communication through H-bonds is observed *via* TM5 and ECL3 residues. Particularly, S301$^{7.30}$, located just after ECL3, and R299^{ECL3} demonstrate significant participation in these interactions. Additionally, residues on the extracellular side of TM5 and ECL2, such as Q204$^{5.32}$, W206$^{5.34}$, H207$^{5.35}$, and S203^{ECL2}, also exhibit high occupancy H-bonds (Figure 6.14).

Lipid-receptor H-bonds occur at the cytosolic side of TM7, involving Y325$^{7.55}$ and K233$^{5.60}$. Additionally, Y325$^{7.55}$, located one residue after Y323$^{7.53}$ of the DPxxY motif, participates actively in hydrogen bonding with lipids.

Meanwhile, K233$^{5.60}$ at the cytosolic end of TM6 also contributes to lipid interactions. The ICL3 also participated in receptor-lipid H-bonds *via* R240$^{5.67}$, while ICL2 communicates through R145$^{3.56}$ (Figure 6.14).

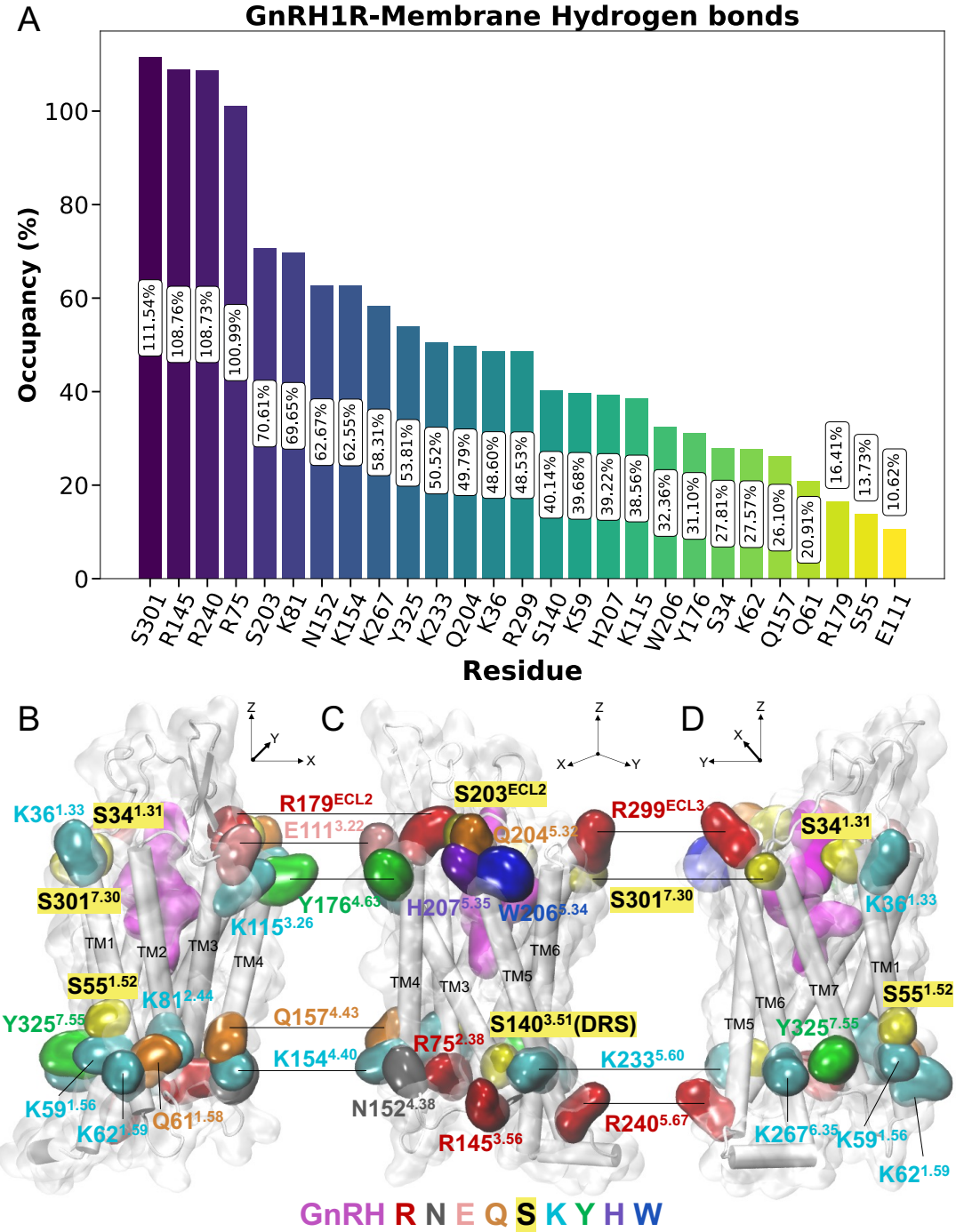


Figure 6.14: Analysis of GnRH1R-Lipid H-bond occupancy in the active ROS-1 System. A) Overall H-bond occupancy between GnRH1R and lipids. B), C), and D) Illustration of GnRH1R residues forming H-bonds with lipids in different orientations around the principal z-axis, represented with surface representation. Note that the occupancy values in graph A exceed 100% because hydrogen bonds between the receptor and lipids are treated as whole residues rather than individual atoms. For example, S301 formed multiple hydrogen bonds with different oxygen

atoms of the lipid. To simplify the graph, these individual atom-atom H-bonds were summed, leading to occupancy values greater than 100%.

Notably, the S140$^{3.51}$ residue (which is unique to GnRH1R) of the DRS motif presents high occupancy H-bonds with lipids, suggesting its potential involvement in receptor internalisation, through ICL2 especially in the absence of a C-terminal H8 helix. Furthermore, R75$^{2.38}$ at the cytosolic end of TM2 forms high occupancy H-bonds with lipids while retaining an ionic lock with D138$^{3.49}$ of the DRS motif. The participation of the DRS residue S140$^{3.51}$ and DRS-interacting residue R75$^{2.38}$ in lipid interactions, highlights their importance in GnRH1R function and potentially internalisation processes (Figure 6.14). Receptor-lipid interactions are also observed at the cytosolic end of TM1, where S55$^{1.52}$, K59$^{1.56}$, and K62$^{1.59}$ form moderate H-bonds with lipids. Similarly, the cytosolic end of TM2 presents lipid H-bonds with Q61$^{1.58}$ and K81$^{2.44}$.

6.7 Summary

The signalling mechanism and communication networks

This research identifies two major communication pathways following GnRH binding and receptor activation (Figure 6.15). GnRH binding is characterised by a cation-π interaction between R8 and W280$^{6.48}$ of the CWxPY motif and is re-enforced with various other interactions, including W3 and Y5 π-π stacking interactions with key residues of such as Y290$^{6.58}$, F309$^{7.38}$ and H306$^{7.35}$ and Y283$^{6.51}$.

R8 of GnRH plays a secondary role beyond establishing a stable cation-π interaction with W280$^{6.48}$. It serves as a mediator for signal transmission to the DPxxY motif and the DRS motif. The communication between R8 and the DRS motif is facilitated through M125$^{3.36}$. More specifically, R8 forms vdW interactions M125$^{3.36}$ which also engages in vdW interactions and H-bonds with A129$^{3.40}$ of the PAF motif. A129$^{3.40}$, in turn, forms H-bonds with V133$^{3.44}$, which is one helical turn from the DRS motif. Hydrogen bonding between V134$^{3.45}$ and D138$^{3.49}$ of the DRS motif is established. Notably, D138$^{3.49}$ forms a stable ionic lock with R75$^{2.38}$ of TM2, reinforcing tight packing of the cytosolic ends of TM2 and TM3. Additionally, the DRS motif interacts with various residues of ICL2 and predominantly serine, threonine, and isoleucine residues -commonly associated with receptor internalisation.

The second identified communication network, initiated with R8, involves a water-mediated pathway connecting the CWxPY and the DPxxY motif in GnRH1R. In this pathway, R8 forms H-bonds either directly with E90$^{2.53}$ or through a common water molecule. This water molecule, in turn, contacts N87$^{2.50}$ and D319$^{7.49}$/Y323$^{7.53}$ of the DPxxY motif. Notably, the DPxxY motif is confirmed herein as a sodium binding pocket where the binding of the ion is conformationally dependent of the inactive state whilst R8 replaces its

charge and further mediates interactions with N87$^{2.50}$ and the DPxxY motif through water mediated H-bonds in the active conformation.

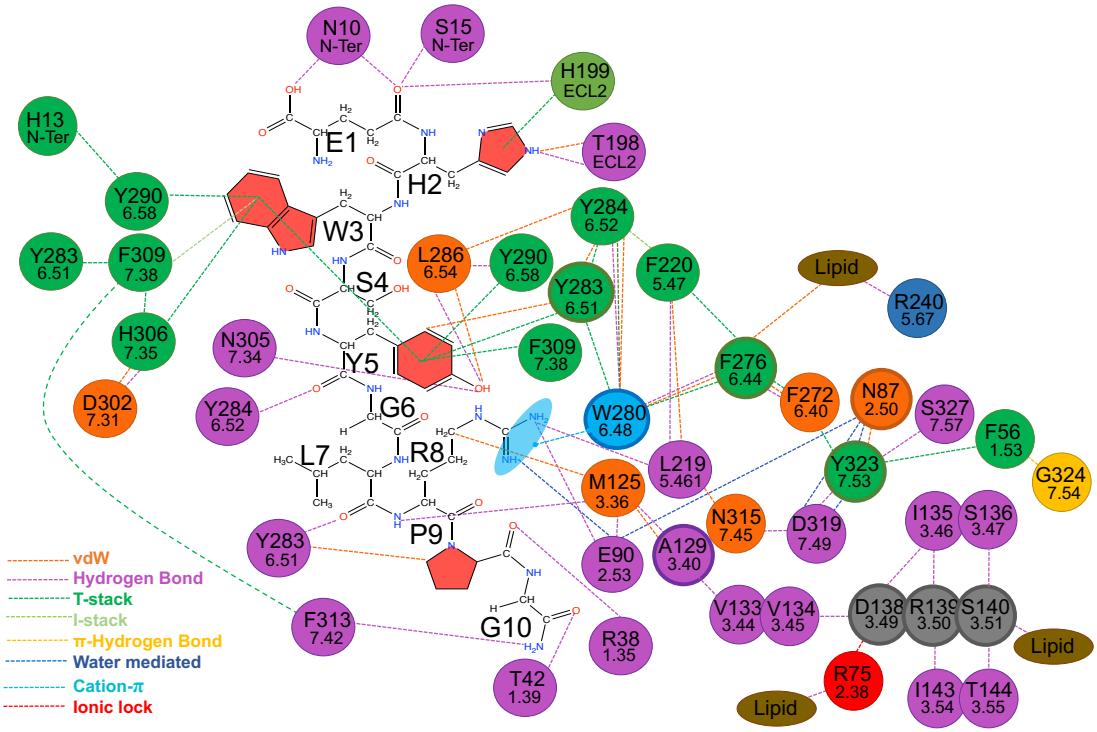


Figure 6.15: Total GnRH-GnRH1R interactions and allosteric networks in the GnRH1R. Contact maps of π-π stacking interactions, cation-π, hydrogen-π are presented in Figure 6.12, and Appendix: Figure 8.2 and Figure 8.3. Ionic locks (SBs) are presented in Figure 6.6. Water mediated interactions are presented in Figure 6.9 and lipid-GnRH1R hydrogen bonds are presented in Figure 6.14. Intermolecular hydrogen bonds and vdW interactions specified in the figure were extracted directly from raw RING data files due to their magnitude.

Furthermore, W280$^{6.48}$ of the CWxPY, which interacts with R8, engages in π-π stacking interactions with F276$^{6.44}$ of the PAF motif (-interacts with the DRS through A129$^{3.40}$). F276$^{6.44}$, in turn, contacts F272$^{6.40}$, and subsequently, Y323$^{7.53}$ of the DPxxY motif and cytosolic F56$^{1.53}$ of TM1. The rotamer change of Y323$^{7.53}$ repositions the residue at the centre of the 7TM bundle, making it available for the water-mediated channel.

In addition to the water-mediated pathway membrane communication is potentially facilitated by F276$^{6.44}$-lipid vdW interactions. This communication is reinforced through R240$^{5.67}$-lipid H-bonds involving the same lipid molecule. Several other residues, most notably S140$^{3.51}$ of the DRS motif and R75$^{2.38}$-DRS interacting residues, also form high occupancy H-bond interactions with lipids.

Table 6.1: Comparative summary between the GPCR/GnRH1R characteristics mentioned in the literature and those found in this study for the GnRH1R.

Residue	*Function*	*Literature*	*This Study*
R38$^{1.35}$	Interaction with G10 of GnRH (computational study). Forms intrahelical contacts that assist GnRH binding.		
	R38$^{1.35}$ formed H-bonds with the backbone oxygen of P9 during simulation. However, G10 is neighbours P9 and therefore is highly likely to interact also		
N87$^{2.50}$	Receptor functionality: part of the conserved water-mediated network.		
	N87$^{2.50}$ was found to participate in the water mediated network (Figure 6.9) involving E90$^{2.53}$- N87$^{2.50}$-DPxxY in the active conformation. Due to the binding of a Na^{+} atom from the DPxxY in the inactive conformation, N87$^{2.50}$ was found to form one less H-bond with water in the inactive conformation.		
D98$^{2.60}$	Structural or GnRH binding. Interaction with H2 (computational study). Forms conformation-independent ionic locks with K121$^{3.32}$. Adjusts the binding pocket based on the ligand.		
	D98$^{2.60}$ did not participate in GnRH binding but showed conformation dependent SB behaviour. The active conformation formed the D98:K121 SB whereas the inactive conformation formed the D98:R38 SB (Figure 6.6) .		
K121$^{3.32}$	Interacts with non-peptide agonists. Suggested interactions with GnRH residues pG1, H2, and W3 (computational studies). Regulates the conformation of the extracellular side of TM3 by forming a salt bridge with D98$^{2.60}$. The salt bridge is dependent on the nature of the ligand.		
	Forms a SB with D98$^{2.60}$ in the active conformation. Important for the plasticity of the binding pocket, stabilises the extracellular portions of TM2 and TM3 and forms the outer wall of the binding pocket.		
M125$^{3.36}$	Suggested to form the lower wall of the orthosteric pocket in antagonists and blocks access to the toggle switch area (CWxPY).		
	M125$^{3.36}$ formed H-bonds and vdW with R8 of GnRH. Communicates with the PAF motif through vdW and H-bonds with A129$^{3.40}$, which is located two helical turns below the DRS.		
Q174$^{4.61}$	Participates in antagonist binding.		
	Q174$^{4.61}$ did not participate in GnRH binding.		
F178$^{4.65}$	Participates in antagonist binding.		
	F178$^{4.65}$ did not participate in GnRH binding.		
W280$^{6.48}$	Part of the toggle switch (CWxPY), this residue has been shown to directly contact ligands in various GPCRs. Mutation abolishes signalling response. Suggested to interact with W3 through $\pi-\pi$ interactions.		
	W280$^{6.48}$ forms cation-π interactions with R8 and π- π interactions with the F276$^{6.44}$, F272$^{6.40}$ and the Y323$^{7.53}$ in the active conformation. This network is closed as in the inactive conformation.		
Y283$^{6.51}$	Belongs to the CWxPY motif. Crucial for ligand binding and activation. Suggested to assist in GnRH binding between W3 and Y283$^{6.51}$ upon disruption of the Y283$^{6.51}$-F309$^{7.38}$ intrahelical contact (computational studies).		
	Forms π-π interactions with Y5 and F309$^{7.38}$. Belongs to the group of π-π interactions consisting of W3, Y283$^{6.51}$, Y290$^{6.58}$, H306$^{7.35}$ and F309$^{7.38}$.		
L286$^{6.54}$	Interacts with antagonists.		
	L286$^{6.54}$ formed H-bonds and vdW interactions with the alcohol group of Y5. Additionally, it formed H-bonds and vdW interactions with Y290$^{6.58}$ and Y284$^{6.53}$ respectfully.		

Y290$^{6.58}$	Suggested interaction with Y5 of GnRH.
	Y290$^{6.58}$ was found to form π-π interactions with W3 and Y5 and was part of the π-π network including Y283$^{6.51}$, Y290$^{6.58}$, H306$^{7.35}$ and F309$^{7.38}$.
D302$^{7.31}$	Suggested interaction with R8 through a salt bridge.
	D302$^{7.31}$ did not participate in GnRH binding. It was found to support the interaction of H306$^{7.35}$ with W3 and F309$^{7.38}$ through vdW and H-bonds with H306$^{7.35}$.
H306$^{7.35}$	Suggested to form intrahelical contact with F308$^{7.37}$, which consequently forms π−π interactions with W3 of GnRH.
	H306$^{7.35}$ was found to interact with W3 of GnRH through π-π interactions. The orientation of F308$^{7.37}$ was towards the membrane environment where it could assist the stability of the binding pocket through hydrophobic interactions with the membrane.
NPxxY	Participates in the conformation-independent conserved interhelical network, in the water-mediated polar network and forms conformation-specific interhelical interactions.
	(DPxxY in the GnRH1R) In the inactive conformation, Y323$^{7.53}$ of the DPxxY interacts with TM1 residues F56$^{1.53}$ and W63$^{1.60}$. In the active conformation (blue), the pathway to CWxPY is open, where Y323$^{7.53}$ communicates with W280$^{6.48}$/R8 through F272$^{6.40}$ and F276$^{6.40}$ of the P$^{5.50}$A$^{3.40}$F$^{6.40}$ motif. D319$^{7.49}$ of the DPxxY acts as a sodium binding residue in the inactive conformation and participates in the water-mediated network between GnRH/CWxPY and E90$^{2.53}$/N87$^{2.50}$ and DPxxY.
DRY	G-protein interaction site. R$^{3.50}$ of the DRY motif forms an ionic lock with E$^{6.30}$ in numerous GPCRs. An increase of the distance between these residues (>5 Å) commonly indicates the conformational state of the GPCR. The corresponding residues in the GnRH1R are R139$^{3.50}$ and T265$^{6.33}$ which form a polar interaction in the inactive conformation that stabilises the distance between TM3 and TM6 at ~7.5 Å.
	(DRS in the GnRH1R) The distance between R139$^{3.50}$ and T265$^{6.33}$ in the active conformation increases at ~13 Å. The DRS and PAF motifs interact through M125$^{3.36}$ which interacts directly with R8. S140$^{3.51}$ of the DRS interacts with lipids through H-bonds. D138$^{3.49}$ of the DRS forms a SB with R75$^{2.38}$ which also interacts with a lipid through H-bonds.
F276$^{6.44}$ and N231$^{5.58}$	N231$^{5.58}$ is associated with receptor activation.
	F276$^{6.44}$ of the PAF rotates towards TM5 and membrane environment in the active conformation where it forms vdW interactions with a lipid molecule. The same lipid molecule acts as a mediator as it also interacts with N231$^{5.58}$. The distance between N231$^{5.58}$ and S136$^{3.47}$ in the inactive conformation was ~4 Å whilst increased at ~6 Å in the active conformation. Additionally, F276$^{6.44}$ was observed within 5 Å of a lipid tail which also interacts with R240$^{5.67}$ though H-bonds in the active conformation.
F272$^{6.40}$	In most GPCRs the residue in position 6.40 is a short hydrophobic amino acid. The GnRH1R has a phenylalanine in this position, and it is highly conserved in tailless mammalian GnRH receptors. F272$^{6.40}$ has been found important for activation, but its exact function was not discovered.
	Mediates interactions between Y323$^{7.53}$ (DPxxY motif), F276$^{6.44}$ (PAF motif) and subsequently W280$^{6.48}$ (CWxPY motif) through π-π interactions in the active conformation.

These detailed interactions highlight the complexity of the communication network initiated by GnRH and provide insights into the

intricate molecular mechanisms governing signal transmission and conformational changes in GnRH1R during activation.

The research conducted in this study provides valuable novel insights into the activation mechanism of the GnRH1R. The identified allosteric networks, H-bond interactions, and lipid-mediated communication pathways contribute to the understanding of the structural dynamics and functional implications of GnRH1R activation. These findings pave the way for further research and may inform future drug development efforts targeting this crucial receptor involved in reproductive processes.

Chapter 7 - Discussion

7.1 Summary of results

GPCRs are master regulators of various physiological processes and are highly targeted by therapeutic drugs. However, GPCR activation by agonists has been challenging. This thesis explored the activation process of the GnRH1 receptor induced by its native ligand GnRH. The binding of GnRH to the GnRH1R triggers receptor activation, which signals the production of fertility hormones (LH and FSH). The primary goal of this research was to uncover the active conformation of the GnRH1R and its complementary GnRH binding mode.

To accomplish this, the research methodology employed computational docking simulations with Rosetta. A total of 27,000 docking simulations were performed, and a protocol designed to select the best scoring and biologically relevant binding modes was developed and employed. The binding mode selection protocol included elimination rounds based on cluster population, energy metrics and experimentally proved GnRH-GnRH1R contacts. The overall goal of the docking and binding mode elimination protocols was the successful selection of two binding modes that provide the best possible starting configuration of the complex.

To avoid the excessive production of binding modes in search for the native-like binding mode, a strategic methodology that developed the formation of templates was employed. Templates not only included energetically favourable binding modes but allowed extensive exploration of the binding pocket due to their distinct binding conformations. This approach allowed for unbiased docking simulation and the output binding modes offered a wide range of configurations around the GnRH1R binding pocket as evident in the contact analysis of the different clusters. Subsequent MD simulations of the selected native-like binding modes was performed to equilibrate the Rosetta predicted configurations and to observe GnRH1R activation.

MD simulations for the two best scoring binding modes ROS-1 and ROS-2, as well as simulations of the undocked GnRH1R systems, reveal that the docking position is invaluable for the successful activation of the receptor. Even though the Rosetta predicted ROS-1 binding mode did equilibrate to a more stable conformation, the binding mode selection process proved successful as it did provide GnRH-GnRH1R configuration that resulted in activation.

The ROS-1 binding mode induced GnRH1R activation after 1.0 μs of simulation time, yielding a ~12.5 Å TM3-TM6 distance compared to the inactive Apo-GnRH1R with a TM3-TM6 distance of ~7.5 Å. GnRH binding predominantly occurred through π-π stacking and H-bond interactions. More specifically, Y5 of GnRH formed π-π interactions with $Y283^{6.51}$ of the

$C^{6.47}W^{6.48}xP^{6.50}Y^{6.51}$ motif, $F309^{7.38}$ and $Y290^{6.58}$, and H-bond interaction with $L286^{6.54}$. Additionally, Y5 formed intramolecular π-π stacking interactions with W3. In turn, W3 contacted $Y290^{6.58}$, $F309^{7.38}$ and $H306^{7.35}$ through π-π interactions.

The GnRH residue R8, which is unique to mammalian GnRH, formed dominant cation-π interactions with $W280^{6.48}$ of the CWxPY, as well as sporadic ionic and H-bond interactions with neighbouring $E90^{2.53}$ and H-bonds with $M125^{3.36}$. The pivotal R8 residue of GnRH was found to have a more intricate function beyond binding to receptor residues. It was observed to mediate the water network to the $D^{7.49}P^{7.50}xxY^{7.53}$ motif through common water-mediated H-bonds with $N87^{2.50}$. A conformation dependent difference was observed in the case of the Apo-GnRH1R, where the DPxxY motif was found to bind a sodium atom throughout the simulation and $N87^{2.50}$ formed one less H-bond with water molecules.

Additionally, signal transmission to the G-protein interacting site was facilitated through vdW interactions of R8 with $M125^{3.36}$, which in turn formed H-bond and vdW interactions with $A129^{3.40}$ of the $P^{5.50}A^{3.40}F^{6.44}$ motif. In the inactive conformation of the GnRH1R, TM1 and TM7 are tightly packed through the interactions between $Y323^{7.53}$ (DPxxY motif) with $F56^{1.53}$ and $W63^{1.60}$. On the contrary, the active conformation of the GnRH1R was characterised by an open G-protein orthosteric pocket where the $Y323^{7.53}$ residue rotated towards the centre of the intrahelical bundle of the receptor, ceasing interactions with $W63^{1.60}$ and forming π-π interactions with the unique residue $F272^{6.40}$. In turn, $F272^{6.40}$ formed π-π interactions with $F276^{6.44}$ of the PAF motif and established vdW interactions with a lipid molecule, which also communicated with the $R240^{5.67}$ at the cytosolic side of TM5 through H-bonds. The receptor was also found to communicate with lipid molecules through H-bonds, especially through the non-conserved residue $S140^{3.51}$ of the $D^{3.49}R^{3.50}S^{3.51}$ (G-protein binding site).

The conformation of the N-terminus of the GnRH1R has been shown herein to be dependant of the presence of a GnRH. More specifically, in absence of GnRH, the N-terminus of the receptor did not equilibrate in a stable conformation, whereas in the activating ROS-1 system, the N-terminus retained a stable conformation after 0.5 μs. This suggests that the N-terminus of GnRH1R fluctuates in the inactive conformation while allowing for an exposed ligand binding pocket, whilst it equilibrated into a stable conformation which 'caps' GnRH inside the binding pocket in the active conformation. The behaviour of the N-terminus in the GnRH1R could apply to other GPCRs and their ligands.

7.2 Conclusions and Future work

The discovery of active GPCR conformations is valuable for the acceleration of structure-based drug design and development as it aids traditional trial-and-error methods. Additionally, receptor-agonist binding modes that induce activation enhance current understanding on the specific signalling pathways and their regulation. In this thesis, the GnRH binding mode that induces GnRH1R activation was unveiled, and the active conformation of the receptor was characterised in detail and compared to the inactive conformation. These results will facilitate drug development targeting the GnRH1R and benefit patients suffering from fertility hormone related cancers, such as prostate, ovarian, cervical and breast cancers and disorders such as endometriosis, hypogonatotrophic hypogonadism and more.

The significance of this work transcends the confines of GnRH1R, contributing essential knowledge to the broader field of GPCRs. The identified binding mechanism and proposed communication networks offer valuable insights for a deeper understanding of GPCR activation. The GnRH1R is classified within the Class A of GPCRs but has differences in conserved motifs such a DPxxY instead of the conserved NPxxY, DRS instead of DRY and more. Additionally, the lack of the cytosolic H8 helix make the active conformation of the GnRH1R revealed here an example for the functionality of similar GPCRs.

Future studies could involve further exploration of the identified binding mode through experimental mutagenesis studies to validate (or contradict) and expand upon the findings presented herein. Additionally, receptor internalisation in absence of the H8 helix would be highly beneficial. Furthermore, GnRH1Rs are expressed in numerous tissues and therefore display high G-protein variability. Hence, structural studies of the receptor in complex with different effectors will not only provide concrete proof of the active conformation but aid better understanding of the mechanisms of GnRH1R activation. Finally, the outcomes of this research hold the potential to catalyse the development of novel and more efficacious therapeutics for HPG axis regulation through the GnRH1R.

References

[1] K. Sriram and P. A. Insel, "G protein-coupled receptors as targets for approved drugs: How many targets and how many drugs?," *Mol Pharmacol*, vol. 93, no. 4, pp. 251–258, Apr. 2018, doi: 10.1124/MOL.117.111062/-/DC1.

[2] L. Cheng *et al.*, "Structure, function and drug discovery of GPCR signaling," *Molecular Biomedicine*, vol. 4, no. 1. Springer, pp. 1–39, Dec. 04, 2023. doi: 10.1186/s43556-023-00156-w.

[3] N. Tuteja, "Signaling through G protein coupled receptors," *Plant Signaling and Behavior*, vol. 4, no. 10. pp. 942–947, 2009. doi: 10.4161/psb.4.10.9530.

[4] R. P. Millar, "GnRHs and GnRH receptors," *Anim Reprod Sci*, vol. 88, no. 1-2 SPEC. ISS., pp. 5–28, 2005, doi: 10.1016/j.anireprosci.2005.05.032.

[5] L. Wildt *et al.*, "Frequency and amplitude of gonadotropin-releasing hormone stimulation and gonadotropin secretion in the rhesus monkey," *Endocrinology*, vol. 109, no. 2, pp. 376–385, 1981, doi: 10.1210/endo-109-2-376.

[6] S. M. Moenter, R. M. Brand, A. R. Midgley, and F. J. Karsch, "Dynamics of gonadotropin-releasing hormone release during a pulse," *Endocrinology*, vol. 130, no. 1, pp. 503–510, 1992, doi: 10.1210/endo.130.1.1727719.

[7] H. A. Ferris and M. A. Shupnik, "Mechanisms for pulsatile regulation of the gonadotropin subunit genes by GNRH1," *Biol Reprod*, vol. 74, no. 6, pp. 993–998, 2006, doi: 10.1095/BIOLREPROD.105.049049.

[8] J. Lehnert and A. Khadra, "How pulsatile kisspeptin stimulation and GnRH autocrine feedback can drive GnRH Secretion: A modeling investigation," *Endocrinology*, vol. 160, no. 5, pp. 1289–1306, May 2019, doi: 10.1210/en.2018-00947.

[9] R. P. Millar, "Gnrh II and type II GnRH receptors," *Trends in Endocrinology and Metabolism*, vol. 14, no. 1. Elsevier, pp. 35–43, Jan. 01, 2003. doi: 10.1016/S1043-2760(02)00016-4.

[10] J. D. Neill, L. W. Duck, J. C. Sellers, and L. C. Musgrove, "A gonadotropin-releasing hormone (GnRH) receptor specific for GnRH II in primates," *Biochem Biophys Res Commun*, vol. 282, no. 4, pp. 1012–1018, Apr. 2001, doi: 10.1006/bbrc.2001.4678.

[11] R. P. Millar, Z. L. Lu, A. J. Pawson, C. A. Flanagan, K. Morgan, and S. R. Maudsley, "Gonadotropin-releasing hormone receptors," *Endocrine Reviews*, vol. 25, no. 2. pp. 235–275, 2004. doi: 10.1210/er.2003-0002.

[12] J. D. Neill, L. C. Musgrove, and L. W. Duck, "Newly recognized GnRH receptors: Function and relative role," *Trends in Endocrinology and Metabolism*, vol. 15, no. 8, pp. 383–392, Oct. 2004, doi: 10.1016/j.tem.2004.08.005.

[13] R. M. Perrett and C. A. McArdle, "Molecular mechanisms of gonadotropin-releasing hormone signaling: Integrating cyclic nucleotides into the network," *Front Endocrinol (Lausanne)*, vol. 4, no. NOV, p. 65072, Nov. 2013, doi: 10.3389/fendo.2013.00180.

[14] S. P. Bliss, A. M. Navratil, J. Xie, and M. S. Roberson, "GnRH signaling, the gonadotrope and endocrine control of fertility," *Front Neuroendocrinol*, vol. 31, no. 3, pp. 322–340, Jul. 2010, doi: 10.1016/J.YFRNE.2010.04.002.

[15] X. Lin and P. M. Conn, "Transcriptional Activation of Gonadotropin-Releasing Hormone (GnRH) Receptor Gene by GnRH: Involvement of Multiple Signal Transduction Pathways*This study was supported by NIH Grants HD-19899, RR-00163 and HD-18185.," *Endocrinology*, vol. 140, no. 1, pp. 358–364, Jan. 1999, doi: 10.1210/ENDO.140.1.6452.

[16] S. Kraus, Z. Naor, and R. Seger, "Intracellular signaling pathways mediated by the gonadotropin-releasing hormone (GnRH) receptor," *Archives of Medical Research*, vol. 32, no. 6. Arch Med Res, pp. 499–509, 2001. doi: 10.1016/S0188-4409(01)00331-9.

[17] M. P. Garrido, A. Hernandez, M. Vega, E. Araya, and C. Romero, "Conventional and new proposals of GnRH therapy for ovarian, breast, and prostatic cancers," *Front Endocrinol (Lausanne)*, vol. 14, no. March, pp. 1–10, 2023, doi: 10.3389/fendo.2023.1143261.

[18] "Increased Incidence of Luteinizing Hormone-releasing Hormone Receptor Gene Messenger RNA Expression in Hormone-refractory Human Prostate Cancers | Clinical Cancer Research | American Association for Cancer Research." Accessed: Feb. 06, 2024. [Online]. Available: https://aacrjournals.org/clincancerres/article/7/8/2340/199972/Increased-Incidence-of-Luteinizing-Hormone

[19] C. Gründker *et al.*, "Antitumor effects of the cytotoxic luteinizing hormone-releasing hormone analog AN-152 on human endometrial and ovarian cancers xenografted into nude mice," *Am J Obstet Gynecol*, vol. 187, no. 3, pp. 528–537, Sep. 2002, doi: 10.1067/mob.2002.124278.

[20] J. Engel, G. Emons, J. Pinski, and A. V. Schally, "AEZS-108: a targeted cytotoxic analog of LHRH for the treatment of cancers positive for LHRH receptors," *Expert Opin Investig Drugs*, vol. 21, no. 6, pp. 891–899, Jun. 2012, doi: 10.1517/13543784.2012.685128.

[21] C. Gründker, J. Ernst, M. Reutter, B. M. Ghadimi, and G. Emons, "Effective targeted chemotherapy using AEZS-108 (AN-152) for LHRH receptor-positive pancreatic cancers," *Oncol Rep*, vol. 26, no. 3, pp. 629–635, Sep. 2011, doi: 10.3892/OR.2011.1340/HTML.

[22] A. Nagy and A. V. Schally, "Targeting of Cytotoxic Luteinizing Hormone-Releasing Hormone Analogs to Breast, Ovarian, Endometrial,

and Prostate Cancers," *Biol Reprod*, vol. 73, no. 5, pp. 851–859, Nov. 2005, doi: 10.1095/BIOLREPROD.105.043489.

[23] J. A. Ballesteros and H. Weinstein, "[19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors," *Methods in Neurosciences*, vol. 25, no. C, pp. 366–428, Jan. 1995, doi: 10.1016/S1043-9471(05)80049-7.

[24] A. Mafi, S. K. Kim, and W. A. Goddard, "The mechanism for ligand activation of the GPCR-G protein complex," *Proc Natl Acad Sci U S A*, vol. 119, no. 18, May 2022, doi: 10.1073/pnas.2110085119.

[25] D. Wootten, A. Christopoulos, M. Marti-Solano, M. M. Babu, and P. M. Sexton, "Mechanisms of signalling and biased agonism in G protein-coupled receptors," *Nat Rev Mol Cell Biol*, vol. 19, no. 10, pp. 638–653, 2018, doi: 10.1038/s41580-018-0049-3.

[26] D. Hilger, "The role of structural dynamics in GPCR-mediated signaling," *FEBS Journal*, vol. 288, no. 8. John Wiley & Sons, Ltd, pp. 2461–2489, Apr. 01, 2021. doi: 10.1111/febs.15841.

[27] A. S. Hauser *et al.*, "GPCR activation mechanisms across classes and macro/microscales," *Nature Structural & Molecular Biology 2021 28:11*, vol. 28, no. 11, pp. 879–888, Nov. 2021, doi: 10.1038/s41594-021-00674-7.

[28] F. Sadler, N. Ma, M. Ritt, Y. Sharma, N. Vaidehi, and S. Sivaramakrishnan, "Autoregulation of GPCR signalling through the third intracellular loop," *Nature*, vol. 615, no. 7953, pp. 734–741, 2023, doi: 10.1038/s41586-023-05789-z.

[29] K. O. Alegre *et al.*, "Structural basis and mechanism of activation of two different families of G proteins by the same GPCR," *Nat Struct Mol Biol*, vol. 28, no. 11, pp. 936–944, 2021, doi: 10.1038/s41594-021-00679-2.

[30] J. A. Ballesteros *et al.*, "Activation of the β2-Adrenergic Receptor Involves Disruption of an Ionic Lock between the Cytoplasmic Ends of Transmembrane Segments 3 and 6," *Journal of Biological Chemistry*, vol. 276, no. 31, pp. 29171–29177, Aug. 2001, doi: 10.1074/jbc.M103747200.

[31] G. Pándy-Szekeres *et al.*, "The G protein database, GproteinDb," *Nucleic Acids Res*, vol. 50, no. D1, pp. D518–D525, Jan. 2022, doi: 10.1093/nar/gkab852.

[32] A. M. Schönegge *et al.*, "Evolutionary action and structural basis of the allosteric switch controlling β2AR functional selectivity," *Nat Commun*, vol. 8, no. 1, pp. 1–12, Dec. 2017, doi: 10.1038/s41467-017-02257-x.

[33] Y. Peng *et al.*, "5-HT2C Receptor Structures Reveal the Structural Basis of GPCR Polypharmacology," *Cell*, vol. 172, no. 4, pp. 719-730.e14, Feb. 2018, doi: 10.1016/J.CELL.2018.01.001.

[34] D. Wacker *et al.*, "Structural Features for Functional Selectivity at Serotonin Receptors," *Science*, vol. 340, no. 6132, p. 615, May 2013, doi: 10.1126/SCIENCE.1232808.

[35] N. R. Latorraca, A. J. Venkatakrishnan, and R. O. Dror, "GPCR dynamics: Structures in motion," *Chem Rev*, vol. 117, no. 1, pp. 139–155, Jan. 2017, doi: 10.1021/ACS.CHEMREV.6B00177/ASSET/IMAGES/LARGE/CR-2016-001779_0012.JPEG.

[36] B. Carpenter, R. Nehmé, T. Warne, A. G. W. Leslie, and C. G. Tate, "Structure of the adenosine A2A receptor bound to an engineered G protein," *Nature 2016 536:7614*, vol. 536, no. 7614, pp. 104–107, Jul. 2016, doi: 10.1038/nature18966.

[37] J. A. Goncalves *et al.*, "Highly conserved tyrosine stabilizes the active state of rhodopsin," *Proc Natl Acad Sci U S A*, vol. 107, no. 46, pp. 19861–19866, Nov. 2010, doi: 10.1073/PNAS.1009405107/SUPPL_FILE/PNAS.1009405107_SI.PDF.

[38] A. K. Topaloglu *et al.*, "Molecular genetic analysis of normosmic hypogonadotropic hypogonadism in a Turkish population: identification and detailed functional characterization of a novel mutation in the gonadotropin-releasing hormone receptor gene," *Neuroendocrinology*, vol. 84, no. 5, pp. 301–308, Mar. 2006, doi: 10.1159/000098147.

[39] J. Ballesteros *et al.*, "Functional microdomains in G-protein-coupled receptors: The conserved arginine-cage motif in the gonadotropin-releasing hormone receptor," *Journal of Biological Chemistry*, vol. 273, no. 17, pp. 10445–10453, 1998, doi: 10.1074/jbc.273.17.10445.

[40] A. J. Venkatakrishnan, X. Deupi, G. Lebon, C. G. Tate, G. F. Schertler, and M. Madan Babu, "Molecular signatures of G-protein-coupled receptors," *Nature*, vol. 494, no. 7436. pp. 185–194, 2013. doi: 10.1038/nature11896.

[41] V. Cvicek, W. A. Goddard, and R. Abrol, *Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications*, vol. 12, no. 3. 2016. doi: 10.1371/journal.pcbi.1004805.

[42] W. Huang *et al.*, "Structural insights into μ-opioid receptor activation," *Nature*, vol. 524, no. 7565, pp. 315–321, 2015, doi: 10.1038/nature14886.

[43] J. Standfuss *et al.*, "The structural basis of agonist-induced activation in constitutively active rhodopsin," *Nature 2011 471:7340*, vol. 471, no. 7340, pp. 656–660, Mar. 2011, doi: 10.1038/nature09795.

[44] A. J. Venkatakrishnan *et al.*, "Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region," *Nature 2016*

536:7617, vol. 536, no. 7617, pp. 484–487, Aug. 2016, doi: 10.1038/nature19107.

[45] W. Zhou *et al.*, "A reciprocal mutation supports helix 2 and helix 7 proximity in the gonadotropin-releasing hormone receptor.," *Mol Pharmacol*, vol. 45, no. 2, 1994.

[46] C. A. Flanagan *et al.*, "The functional microdomain in transmembrane helices 2 and 7 regulates expression, activation, and coupling pathways of the gonadotropin-releasing hormone receptor," *Journal of Biological Chemistry*, vol. 274, no. 41, pp. 28880–28886, Oct. 1999, doi: 10.1074/jbc.274.41.28880.

[47] J. A. Tello, C. L. Newton, J. Bouligand, A. Guiochon-Mantel, R. P. Millar, and J. Young, "Congenital hypogonadotropic hypogonadism due to GNRH receptor mutations in three brothers reveal sites affecting conformation and coupling," *PLoS One*, vol. 7, no. 6, p. e38456, Jun. 2012, doi: 10.1371/journal.pone.0038456.

[48] K. K. Arora, Z. Cheng, and K. J. Catt, "Dependence of agonist activation on an aromatic moiety in the DPLIY motif of the gonadotropin-releasing hormone receptor.," *Molecular Endocrinology*, vol. 10, no. 8, pp. 979–986, Aug. 1996, doi: 10.1210/mend.10.8.8843414.

[49] B. Trzaskowski, D. Latek, S. Yuan, U. Ghoshdastider, A. Debinski, and S. Filipek, "Action of Molecular Switches in GPCRs - Theoretical and Experimental Studies," *Curr Med Chem*, vol. 19, no. 8, pp. 1090–1109, Feb. 2012, doi: 10.2174/092986712799320556.

[50] X. Deupi and J. Standfuss, "Structural insights into agonist-induced activation of G-protein-coupled receptors," *Current Opinion in Structural Biology*, vol. 21, no. 4. Elsevier Current Trends, pp. 541–551, Aug. 01, 2011. doi: 10.1016/j.sbi.2011.06.002.

[51] X. Deupi, "Relevance of rhodopsin studies for GPCR activation," *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 1837, no. 5, pp. 674–682, May 2014, doi: 10.1016/J.BBABIO.2013.09.002.

[52] W. Yan *et al.*, "Structure of the human gonadotropin-releasing hormone receptor GnRH1R reveals an unusual ligand binding mode," *Nat Commun*, vol. 11, no. 1, pp. 1–10, 2020, doi: 10.1038/s41467-020-19109-w.

[53] V. Katritch, G. Fenalti, E. E. Abola, B. L. Roth, V. Cherezov, and R. C. Stevens, "Allosteric sodium in class A GPCR signaling," *Trends Biochem Sci*, vol. 39, no. 5, pp. 233–244, May 2014, doi: 10.1016/J.TIBS.2014.03.002.

[54] W. Yan *et al.*, "Structure of the human gonadotropin-releasing hormone receptor GnRH1R reveals an unusual ligand binding mode," *Nat Commun*, vol. 11, no. 1, pp. 1–10, 2020, doi: 10.1038/s41467-020-19109-w.

[55] B. J. Fromme, A. A. Katz, R. W. Roeske, R. P. Millar, and C. A. Flanagan, "Role of aspartate7.32(302) of the human gonadotropin-releasing hormone receptor in stabilizing a high-affinity ligand conformation.,"

Mol Pharmacol, vol. 60, no. 6, pp. 1280–1287, Dec. 2001, doi: 10.1124/mol.60.6.1280.

[56] M. Coetsee, R. P. Millar, C. A. Flanagan, and Z.-L. Lu, "Identification of Tyr(290(6.58)) of the human gonadotropin-releasing hormone (GnRH) receptor as a contact residue for both GnRH I and GnRH II: importance for high-affinity binding and receptor activation.," *Biochemistry*, vol. 47, no. 39, pp. 10305–10313, Sep. 2008, doi: 10.1021/bi800911z.

[57] A. J. Stewart, R. Sellar, D. J. Wilson, R. P. Millar, and Z.-L. Lu, "Identification of a novel ligand binding residue Arg38(1.35) in the human gonadotropin-releasing hormone receptor.," *Mol Pharmacol*, vol. 73, no. 1, pp. 75–81, Jan. 2008, doi: 10.1124/mol.107.040816.

[58] K. Kochman, "Evolution of gonadotropin-releasing hormone (GnRH) structure and its receptor," *J Anim Feed Sci*, vol. 21, no. 1, pp. 3–30, 2012, doi: 10.22358/jafs/66031/2012.

[59] C. A. Flanagan *et al.*, "Glutamate 301 of the mouse gonadotropin-releasing hormone receptor confers specificity for arginine 8 of mammalian gonadotropin-releasing hormone," *Journal of Biological Chemistry*, vol. 269, no. 36, pp. 22636–22641, 1994, doi: 10.1016/s0021-9258(17)31693-9.

[60] B. J. Fromme, A. A. Katz, R. W. Roeske, R. P. Millar, and C. A. Flanagan, "Role of aspartate7.32(302) of the human gonadotropin-releasing hormone receptor in stabilizing a high-affinity ligand conformation.," *Mol Pharmacol*, vol. 60, no. 6, pp. 1280–1287, Dec. 2001, doi: 10.1124/mol.60.6.1280.

[61] N. M. I. Mayevu *et al.*, "Histidine7.36(305) in the conserved peptide receptor activation domain of the gonadotropin releasing hormone receptor couples peptide binding and receptor activation," *Mol Cell Endocrinol*, vol. 402, pp. 95–106, 2015, doi: 10.1016/j.mce.2015.01.008.

[62] M. Coetsee, R. P. Millar, C. A. Flanagan, and Z.-L. Lu, "Identification of Tyr(290(6.58)) of the human gonadotropin-releasing hormone (GnRH) receptor as a contact residue for both GnRH I and GnRH II: importance for high-affinity binding and receptor activation.," *Biochemistry*, vol. 47, no. 39, pp. 10305–10313, Sep. 2008, doi: 10.1021/bi800911z.

[63] S. F. Betz, G. J. Reinhart, F. M. Lio, C. Chen, and R. S. Struthers, "Overlapping, nonidentical binding sites of different classes of nonpeptide antagonists for the human gonadotropin-releasing hormone receptor," *J Med Chem*, vol. 49, no. 2, pp. 637–647, 2006, doi: 10.1021/jm0506928.

[64] W. Shihoya *et al.*, "Activation mechanism of endothelin ET B receptor by endothelin-1," *Nature*, vol. 537, no. 7620, pp. 363–368, 2016, doi: 10.1038/nature19319.

[65] J. F. White *et al.*, "Structure of the agonist-bound neurotensin receptor," *Nature*, vol. 490, no. 7421, pp. 508–513, 2012, doi: 10.1038/nature11558.
[66] C. A. Flanagan and A. Manilall, "Gonadotropin-releasing hormone (GnRH) receptor structure and GnRH binding," *Front Endocrinol (Lausanne)*, vol. 8, no. OCT, pp. 1–14, 2017, doi: 10.3389/fendo.2017.00274.
[67] A. J. Stewart, R. Sellar, D. J. Wilson, R. P. Millar, and Z.-L. Lu, "Identification of a novel ligand binding residue Arg38(1.35) in the human gonadotropin-releasing hormone receptor.," *Mol Pharmacol*, vol. 73, no. 1, pp. 75–81, Jan. 2008, doi: 10.1124/mol.107.040816.
[68] J. A. Söderhäll, E. E. Polymeropoulos, K. Paulini, E. Günther, and R. Kühne, "Antagonist and agonist binding models of the human gonadotropin-releasing hormone receptor.," *Biochem Biophys Res Commun*, vol. 333, no. 2, pp. 568–582, Jul. 2005, doi: 10.1016/j.bbrc.2005.05.142.
[69] S. Hövelmann, S. H. Hoffmann, R. Kühne, T. Ter Laak, H. Reiländer, and T. Beckers, "Impact of aromatic residues within transmembrane helix 6 of the human gonadotropin-releasing hormone receptor upon agonist and antagonist binding," *Biochemistry*, vol. 41, no. 4, pp. 1129–1136, 2002, doi: 10.1021/bi0113162.
[70] S. H. Hoffmann, T. Ter Laak, R. Kühne, H. Reiländer, and T. Beckers, "Residues within transmembrane helices 2 and 5 of the human gonadotropin-releasing hormone receptor contribute to agonist and antagonist binding," *Molecular Endocrinology*, vol. 14, no. 7, pp. 1099–1115, 2000, doi: 10.1210/mend.14.7.0483.
[71] R. Forfar and Z. L. Lu, "Role of the transmembrane domain 4/extracellular loop 2 junction of the human gonadotropin-releasing hormone receptor in ligand binding and receptor conformational selection," *Journal of Biological Chemistry*, vol. 286, no. 40, pp. 34617–34626, 2011, doi: 10.1074/jbc.M111.240341.
[72] J. A. Janovick, I. D. Pogozheva, H. I. Mosberg, and P. M. Conn, "Salt bridges overlapping the gonadotropin-releasing hormone receptor agonist binding site reveal a coincidence detector for G protein-coupled receptor activation," *Journal of Pharmacology and Experimental Therapeutics*, vol. 338, no. 2, pp. 430–442, 2011, doi: 10.1124/jpet.111.180869.
[73] A. Manilall, B. A. Stander, M. T. Madziva, R. P. Millar, and C. A. Flanagan, "Glu 2.53(90) of the GnRH receptor is part of the conserved G protein-coupled receptor structure and does not form a salt-bridge with Lys 3.32(121)," 2018, doi: 10.1016/j.mce.2018.11.009.
[74] W. Zhou *et al.*, "A locus of the gonadotropin-releasing hormone receptor that differentiates agonist and antagonist binding sites," *Journal of*

Biological Chemistry, vol. 270, no. 32, pp. 18853–18857, 1995, doi: 10.1074/jbc.270.32.18853.

[75] G. Y. Bédécarrats *et al.*, "Four naturally occurring mutations in the human GnRH receptor affect ligand binding and receptor function," *Mol Cell Endocrinol*, vol. 205, no. 1–2, pp. 51–64, Jul. 2003, doi: 10.1016/S0303-7207(03)00201-6.

[76] J. S. Davidson, C. A. McArdle, P. Davies, R. Elario, C. A. Flanagan, and R. P. Millar, "Asn102 of the Gonadotropin-releasing Hormone Receptor Is a Critical Determinant of Potency for Agonists Containing C-terminal Glycinamide*," *Journal of Biological Chemistry*, vol. 271, no. 26, pp. 15510–15514, 1996, doi: https://doi.org/10.1074/jbc.271.26.15510.

[77] C. A. Flanagan *et al.*, "Multiple interactions of the Asp(2.61(98)) side chain of the gonadotropin-releasing hormone receptor contribute differentially to ligand interaction," *Biochemistry*, vol. 39, no. 28, pp. 8133–8141, Jul. 2000, doi: 10.1021/bi000085g.

[78] L. P. Datta, S. Manchineella, and T. Govindaraju, "Biomolecules-derived biomaterials," *Biomaterials*, vol. 230, no. November 2019, p. 119633, 2020, doi: 10.1016/j.biomaterials.2019.119633.

[79] N.V. Bhagavan and H. Chung-Eun, "Amino Acids," in *Essentials of Medical Biochemistry*, vol. 96, no. 2005, Academic Press, 2015, pp. 21–29.

[80] S. Poblete, S. Pantano, K. I. Okazaki, Z. Liang, K. Kremer, and A. B. Poma, "Editorial: Recent advances in computational modelling of biomolecular complexes," *Frontiers in Chemistry*, vol. 11. pp. 1–3, 2023. doi: 10.3389/fchem.2023.1200409.

[81] P. Lecca, A. Re, A. Ihekwaba, I. Mura, and T.-P. Nguyen, "Overview of Biological Network Inference and Modeling of Dynamics," in *Computational Systems Biology*, 2016.

[82] T. Schlick and S. Portillo-Ledesma, "Biomolecular modeling thrives in the age of technology," *Nature Computational Science*, vol. 1, no. 5. pp. 321–331, 2021. doi: 10.1038/s43588-021-00060-9.

[83] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí, "Molecular dynamics simulations: Advances and applications," *Advances and Applications in Bioinformatics and Chemistry*, vol. 8, no. 1. pp. 37–47, 2015. doi: 10.2147/AABC.S70333.

[84] J. M. Rondeau and H. Schreuder, *Protein Crystallography and Drug Discovery*. Elsevier Ltd, 2015.

[85] G. Gallavotti, "Introductory Article: Classical Mechanics," in *Encyclopedia of Mathematical Physics: Five-Volume Set*, Academic Press, 2004, pp. 1–33. doi: 10.1016/B0-12-512666-2/00318-7.

[86] I. Newton, "Newton's Principia: the mathematical principles of natural philosophy 1642-1727," in *New-York: Daniel Adee*, 1846.

[87] L. Verlet, "Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules," *Physical Review*, vol. 159, no. 1, pp. 98–103, 1967, doi: 10.1103/PhysRev.159.98.
[88] M. González, "Force fields and molecular dynamics simulations," *EDP Sciences*, vol. 72. pp. 169–200, 2011.
[89] K. Vanommeslaeghe *et al.*, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," *J Comput Chem*, vol. 31, no. 4, pp. 671–690, 2010, doi: 10.1002/jcc.21367.
[90] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general Amber force field," *J Comput Chem*, vol. 25, no. 9, pp. 1157–1174, 2004, doi: 10.1002/jcc.20035.
[91] N. Schmid *et al.*, "Definition and testing of the GROMOS force-field versions 54A7 and 54B7," *European Biophysics Journal*, vol. 40, no. 7, pp. 843–856, 2011, doi: 10.1007/s00249-011-0700-9.
[92] D. Petrov and B. Zagrovic, "Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments?," *PLoS Comput Biol*, vol. 10, no. 5, p. 1003638, 2014, doi: 10.1371/journal.pcbi.1003638.
[93] L. Fabbrizzi, "Beyond the Molecule: Intermolecular Forces from Gas Liquefaction to X–H···π Hydrogen Bonds," *ChemPlusChem*, vol. 87, no. 1. pp. 1–23, 2022.
[94] J. Hermann, R. A. DiStasio, and A. Tkatchenko, "First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications," *Chem Rev*, vol. 117, no. 6, pp. 4714–4758, 2017, doi: 10.1021/acs.chemrev.6b00446.
[95] W. H. Keesom, "The second virial coefficient for rigid spherical molecules whose mutual attraction is equivalent to that of a quadruplet placed at its center," *Proceedings of the Royal Netherlands Academy of Arts and Sciences*, vol. 18, pp. 636–646, 1915.
[96] P. Debye, "Van der Waals Cohesive Forces," *Phys*, vol. 21, pp. 178–187, 1920.
[97] F. London, "The General Theory of Molecular Forces," *Transactions of the Faraday Society*, vol. 33, pp. 8–26, 1937.
[98] J. E. Lennard-Jones, "No Title," *Proc Phys Soc*, vol. 43. pp. 461–482, 1931.
[99] X. Wang, S. Ramírez-hinestrosa, J. Dobnikar, and D. Frenkel, "The Lennard-Jones potential : when (not) to use it.," *Physical Chemistry Chemical Physics*, no. 19, pp. 1–11, 2020, doi: 10.1039/cXCP00000x/.
[100] G. B. Arfken, D. F. Griffing, D. C. Kelly, and J. Priest, "ELECTRIC FIELD AND GAUSS' LAW," in *International Edition University Physics*, Academic Press, 1984, pp. 490–513. doi: 10.1016/b978-0-12-059858-8.50032-7.

[101] J. C. Phillips *et al.*, "Scalable molecular dynamics with NAMD," *Parallel Science and Engineering Applications: The Charm++ Approach*, vol. 26, no. 16, pp. 60–76, 2016, doi: 10.1201/b16251-15.
[102] P. Taylor, G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, "Explicit reversible integrators for extended systems dynamics," *Mol Phys*, vol. 87, no. 5, pp. 1117–1157, 1996.
[103] D. J. Tobias, G. J. Martyna, and M. L. Klein, "Molecular Dynamics Simulations of a Protein in the Canonical Ensemble," *Journal of Physical Chemistry*, vol. 97, no. 12959–12966, 1993.
[104] K. Tu, D. J. Tobias, and M. L. Klein, "Constant pressure and temperature molecular dynamics simulation of a fully hydrated liquid crystal phase dipalmitoylphosphatidylcholine bilayer," *Biophys J*, vol. 69, no. 6, pp. 2558–2562, 1995, doi: 10.1016/S0006-3495(95)80126-8.
[105] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *J Chem Phys*, vol. 81, no. 8, pp. 3684–3690, 1984.
[106] K. Zhou and B. Liu, "Control techniques of molecular dynamics simulation," *Molecular Dynamics Simulation*, pp. 67–96, 2022, doi: 10.1016/b978-0-12-816419-8.00008-8.
[107] S. Nosè, "A unified formulation of the constant temperature molecular dynamics methods," *J Chem Phys*, vol. 81, no. 1, pp. 511–519, 1984.
[108] M. Sachs, B. Leimkuhler, and V. Danos, "Langevin dynamics with variable coefficients and nonconservative forces: From stationary states to numerical methods," *Entropy*, vol. 19, no. 12, pp. 1–28, 2017.
[109] C. L. Brooks, A. Brünger, and M. Karplus, "Active site dynamics in protein molecules: A stochastic boundary molecular-dynamics approach," *Biopolymers*, vol. 24, no. 5, pp. 843–865, 1985, doi: 10.1002/bip.360240509.
[110] B. C. B. Symons and P. L. A. Popelier, "Flexible multipole moments in smooth particle mesh Ewald," *Journal of Chemical Physics*, vol. 156, no. 244107, pp. 1–10, 2022.
[111] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems," *J Chem Phys*, vol. 98, no. 12, pp. 10089–10092, 1993, doi: 10.1063/1.464397.
[112] P. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale," *Ann Phys*, vol. 369, no. 3, pp. 253–287, 1921.
[113] J. W. Cooley, "The re-discovery of the fast Fourier transform algorithm," *Mikrochim Acta*, vol. 93, no. 1–6, pp. 33–45, 1987, doi: 10.1007/BF01201681.
[114] A. Leaver-Fay *et al.*, "Rosetta3: An object-oriented software suite for the simulation and design of macromolecules," *Methods Enzymol*, vol. 487, pp. 545–574, 2011, doi: 10.1016/B978-0-12-381270-4.00019-6.

[115] R. F. Alford *et al.*, "The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design," *J Chem Theory Comput*, vol. 13, no. 6, pp. 3031–3048, 2017, doi: 10.1021/acs.jctc.7b00125.
[116] J. K. Leman, B. K. Mueller, and J. J. Gray, "Expanding the toolkit for membrane protein modeling in Rosetta," *Bioinformatics*, vol. 33, no. 5, pp. 754–756, 2017, doi: 10.1093/BIOINFORMATICS/BTW716.
[117] R. F. Alford *et al.*, "An Integrated Framework Advancing Membrane Protein Modeling and Design," *PLoS Comput Biol*, vol. 11, no. 9, pp. 1–23, 2015, doi: 10.1371/journal.pcbi.1004398.
[118] R. E. Jefferson *et al.*, "Computational design of dynamic receptor—peptide signaling complexes applied to chemotaxis," *Nat Commun*, vol. 14, no. 1, 2023, doi: 10.1038/s41467-023-38491-9.
[119] P. Hosseinzadeh *et al.*, "Comprehensive computational design of ordered peptide macrocycles," *Science*, vol. 358, no. 6369, pp. 1461–1466, Dec. 2017, doi: 10.1126/SCIENCE.AAP7577.
[120] M. Matic, P. Miglionico, M. Tatsumi, A. Inoue, and F. Raimondi, "GPCRome-wide analysis of G-protein-coupling diversity using a computational biology approach," *Nat Commun*, vol. 14, no. 1, 2023, doi: 10.1038/s41467-023-40045-y.
[121] F. Antoneli, F. M. Passos, L. R. Lopes, and M. R. S. Briones, "A Kolmogorov-Smirnov test for the molecular clock based on Bayesian ensembles of phylogenies," *PLoS One*, vol. 13, no. 1, Jan. 2018, doi: 10.1371/JOURNAL.PONE.0190826.
[122] M. P. Fay and M. A. Proschan, "Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules," *Stat Surv*, vol. 4, p. 1, 2010, doi: 10.1214/09-SS051.
[123] G. M. Sullivan and R. Feinn, "Using Effect Size—or Why the P Value Is Not Enough," *J Grad Med Educ*, vol. 4, no. 3, p. 279, Sep. 2012, doi: 10.4300/JGME-D-12-00156.1.
[124] M. Jafari and N. Ansari-Pour, "Why, When and How to Adjust Your P Values?," *Cell Journal (Yakhteh)*, vol. 20, no. 4, p. 604, 2019, doi: 10.22074/CELLJ.2019.5992.
[125] V. Vickerstaff, R. Z. Omar, and G. Ambler, "Methods to adjust for multiple comparisons in the analysis and sample size calculation of randomised controlled trials with multiple primary outcomes," *BMC Med Res Methodol*, vol. 19, no. 1, pp. 1–13, Jun. 2019, doi: 10.1186/S12874-019-0754-4/TABLES/5.
[126] K. Pern Tan, K. Singh, A. Hazra, and M. Madhusudhan, "Peptide bond planarity constrains hydrogen bond geometry and influences secondary structure conformations," 2021, doi: 10.1016/j.crstbi.2020.11.002.

[127] S. Jo, J. B. Lim, J. B. Klauda, and W. Im, "CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes," *Biophys J*, vol. 97, no. 1, pp. 50–58, 2009, doi: 10.1016/j.bpj.2009.04.013.

[128] S. Jo, T. Kim, and W. Im, "Automated builder and database of protein/membrane complexes for molecular dynamics simulations," *PLoS One*, vol. 2, no. 9, 2007, doi: 10.1371/journal.pone.0000880.

[129] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular docking: a powerful approach for structure-based drug discovery," *Curr Comput Aided Drug Des*, vol. 7, no. 2, pp. 146–157, Nov. 2011, doi: 10.2174/157340911795677602.

[130] Romano T. Kroemer, "Structure-Based Drug Design: Docking and Scoring," *Curr Protein Pept Sci*, vol. 8, no. 4, pp. 312–328, Jul. 2007, doi: 10.2174/138920307781369382.

[131] A. A. T. Naqvi, T. Mohammad, G. M. Hasan, and Md. I. Hassan, "Advancements in Docking and Molecular Dynamics Simulations Towards Ligand-receptor Interactions and Structure-function Relationships," *Curr Top Med Chem*, vol. 18, no. 20, pp. 1755–1768, Jan. 2019, doi: 10.2174/1568026618666181025114157.

[132] M. Zivkovic, M. Zlatanovic, N. Zlatanovic, M. Golubović, and A. M. Veselinović, "The Application of the Combination of Monte Carlo Optimization Method based QSAR Modeling and Molecular Docking in Drug Design and Development," *Mini-Reviews in Medicinal Chemistry*, vol. 20, no. 14, pp. 1389–1402, Feb. 2020, doi: 10.2174/1389557520666200212111428.

[133] G. D. Geromichalos, "Importance of molecular computer modeling in anticancer drug development," *Journal of B.U.ON.*, vol. 12, no. SUPPL. 1, Sep. 2007.

[134] J. K. Leman *et al.*, "Macromolecular modeling and design in Rosetta: recent methods and frameworks," *Nat Methods*, vol. 17, no. 7, p. 665, Jul. 2020, doi: 10.1038/S41592-020-0848-2.

[135] K. W. Kaufmann, G. H. Lemmon, S. L. Deluca, J. H. Sheehan, and J. Meiler, "Practically useful: What the R osetta protein modeling suite can do for you," *Biochemistry*, vol. 49, no. 14, pp. 2987–2998, Apr. 2010, doi: 10.1021/BI902153G/SUPPL_FILE/BI902153G_SI_003.ZIP.

[136] S. A. Combs *et al.*, "Small-molecule ligand docking into comparative models with Rosetta," *Nat Protoc*, vol. 8, no. 7, p. 1277, 2013, doi: 10.1038/NPROT.2013.074.

[137] C. Wang, P. Bradley, and D. Baker, "Protein-Protein Docking with Backbone Flexibility," *J Mol Biol*, vol. 373, no. 2, pp. 503–519, Oct. 2007, doi: 10.1016/J.JMB.2007.07.050.

[138] B. Kuhlman and P. Bradley, "Advances in protein structure prediction and design," *Nat Rev Mol Cell Biol*, doi: 10.1038/s41580-019-0163-x.

[139] P. Renault, M. Louet, J. Marie, G. Labesse, and N. Floquet, "Molecular Dynamics Simulations of the Allosteric Modulation of the Adenosine A 2 a Receptor by a Mini-G Protein," *Sci Rep*, vol. 9, no. 1, pp. 1–12, 2019, doi: 10.1038/s41598-019-41980-x.

[140] A. E. Badaczewska-Dawid, S. Kmiecik, and M. Koliński, "Docking of peptides to GPCRs using a combination of CABS-dock with FlexPepDock refinement," *Brief Bioinform*, vol. 22, no. 3, pp. 1–9, 2021, doi: 10.1093/bib/bbaa109.

[141] A. Khramushin *et al.*, "Modeling beta-sheet peptide-protein interactions: Rosetta FlexPepDock in CAPRI rounds 38-45," *Proteins: Structure, Function and Bioinformatics*, vol. 88, no. 8, pp. 1037–1049, 2020, doi: 10.1002/prot.25871.

[142] B. J. Bender *et al.*, "A practical guide to large-scale docking," *Nat Protoc*, vol. 16, no. 10, pp. 4799–4832, 2021, doi: 10.1038/s41596-021-00597-z.

[143] S. F. Betz, G. J. Reinhart, F. M. Lio, C. Chen, and R. S. Struthers, "Overlapping, nonidentical binding sites of different classes of nonpeptide antagonists for the human gonadotropin-releasing hormone receptor," *J Med Chem*, vol. 49, no. 2, pp. 637–647, Jan. 2006, doi: 10.1021/jm0506928.

[144] X. Yang *et al.*, "Discovery of Small Molecule Agonist of Gonadotropin-Releasing Hormone Receptor (GnRH1R)," *J Chem Inf Model*, vol. 62, no. 20, pp. 5009–5022, Oct. 2022, doi: 10.1021/acs.jcim.2c00639.

[145] J. Hu *et al.*, "Difference and Influence of Inactive and Active States of Cannabinoid Receptor Subtype CB2: From Conformation to Drug Discovery," *J Chem Inf Model*, vol. 56, no. 6, pp. 1152–1163, Jun. 2016, doi: 10.1021/acs.jcim.5b00739.

[146] Q. Zhou *et al.*, "Common activation mechanism of class a GPCRs," *Elife*, vol. 8, pp. 1–31, 2019, doi: 10.7554/eLife.50279.

[147] W. I. Weis and B. K. Kobilka, "The Molecular Basis of G Protein–Coupled Receptor Activation," *Annu Rev Biochem*, vol. 87, p. 897, Jun. 2018, doi: 10.1146/ANNUREV-BIOCHEM-060614-033910.

8.1 Rosetta docking protocol

Minimisation

Minimisation:

```
/path/to/Rosetta/main/source/bin/FlexPepDocking.static.linuxgccrel
ease \
-database /path/to/Rosetta/main/database/ \
-s 7BR3_PEP_start.pdb \
-in:file:spanfile 7BR3_PEP.span \
-membrane:Membed_init \
-membrane:Mhbond_depth \
-score:weights membrane_highres \
-ex1 \
-ex2aro \
-flexPepDocking:flexpep_score_only \
-scorefile min.score.sc \
-nstruct 10 \
-out:suffix _min
```

Prepacking

Prepacking:

```
/path/to/Rosetta/main/source/bin/FlexPepDocking.static.linuxgccrel
ease \
-in:file:s 7BR3_PEP_start_min_0003.pdb \
-in:file:spanfile 7BR3_PEP.span \
-membrane:Membed_init \
-membrane::Mhbond_depth \
-score:weights membrane_highres  \
-scorefile ppk.sc \
-out:no_nstruct_label \
-out:prefix ppk. \
-flexPepDocking:flexpep_prepack \
-flexPepDocking:flexpep_score_only \
-ex1 \
-ex2aro \
-use_input_sc \
-unboundrot native.pdb \
-nstruct 10
```

Template formation

```
/path/to/Rosetta/main/source/bin/FlexPepDocking.static.linuxgccrel
ease @options_template > template.log
```

Where the `options_template` contents:

```
-database /path/to/Rosetta/main/database/
-in:file:s ppk.min.7BR3_PEP.pdb
-in:file:spanfile 7BR3_PEP.span
-membrane:Membed_init
-membrane:Mhbond_depth
-score:weights membrane_highres
-scorefile score.sc
```

```
-flexPepDocking:lowres_preoptimize
-flexPepDocking:pep_refine
-ex1
-ex2aro
-nstruct 100
-use_input_sc
-out:prefix tem.
```

FlexPepDocking

```
#!/bin/bash
#set the full path to the Rosetta database
DATABASE_PATH="//path/to/Rosetta/main/database/"
# Specify the options file
OPTIONS_FILE="options_flex.inp"
#list of specific input PDB files
INPUT_FILES=("tem_0060.pdb")
#iterate over the list of input PDB files
for pdb_file in "${INPUT_FILES[@]}"
do
    #set the input PDB file
    INPUT_PDB="-s $pdb_file"
    #set the output prefix based on the input file name
    OUTPUT_PREFIX="-out:prefix flex_$(basename $pdb_file .pdb)"
    #run the FlexPepDocking command
/path/to/Rosetta/main/source/bin/FlexPepDocking.static.linuxgccrel
ease $INPUT_PDB $OUTPUT_PREFIX @$OPTIONS_FILE > dock_60.log
Done
```

Where the `options_template` consists of:

```
-in:file:spanfile 7BR3_PEP.span
-membrane:Membed_init
-membrane:Mhbond_depth
-score:weights membrane_highres
-scorefile score.sc
-flexPepDocking:lowres_preoptimize
-flexPepDocking:pep_refine
-ex1
-ex2aro
-nstruct 1000
-use_input_sc
```

Clustering

```
/path/to/Rosetta/main/source/bin/energy_based_clustering.static.li
nuxgccrelease @make_clusters.options.inp > cluster.log
```

Where the `make_clusters.options.inp` consists of:

```
-in:file:l pdb_list #specifies the list of PDB files to be
clustered.
-in:file:fullatom #indicates that full-atom structures are used
for clustering.
```

```
-cluster:energy_based_clustering:cluster_radius 1.0 #sets the
clustering radius to 1.0 Å.
-cluster:energy_based_clustering:limit_structures_per_cluster 0
#no limit on the number of structures per cluster.
-cluster:energy_based_clustering:cluster_by bb_cartesian
#clustering based on backbone Cartesian coordinates.
-cluster:energy_based_clustering:use_CB false #disables the use of
Cβ atoms in clustering.
-cluster:energy_based_clustering:cyclic false #specifies non-
cyclic clustering
-cluster:energy_based_clustering:cluster_cyclic_permutations false
# disables clustering of cyclic permutations.
-cluster:energy_based_clustering:perform_ABOXYZ_bin_analysis true
#enables analysis of ABOXYZ bin during clustering (Ramachandran).
```

InteraceAnalyzer

```
/path/to/Rosetta/main/source/bin/rosetta_scripts.static.linuxgccrele
ase @interfaceAnalyzer.options -in:file:s *c.*pdb >
interfaceAnalyzer.log
```

Where the `interfaceAnalyzer.options` contains:

```
-parser:protocol InterfaceAnalyzer.xml
-use_input_sc
-ex1
-ex2aro
-out:file:scorefile cluster_InterfaceAnalyzer.csv
-out:no_nstruct_label
-out:file:score_only
-in:file:spanfile 7BR3_PEP.span
-membrane:Membed_init
-membrane:Mhbond_depth
-score:weights membrane_highres.wts
```

Where the `InterfaceAnalyzer.xml` contains:

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
        <ScoreFunction name="membrane_highres"
weights="membrane_highres.wts" />
    </SCOREFXNS>
    <TASKOPERATIONS>
        <RestrictToInterfaceVector name="rtiv" chain1_num="1"
chain2_num="2" CB_dist_cutoff="10.0" nearby_atom_cutoff="5.5"
vector_angle_cutoff="75" vector_dist_cutoff="9.0" />
    </TASKOPERATIONS>
    <MOVERS>
        <InterfaceAnalyzerMover name="iface_analyzer"
scorefxn="membrane_highres" interface="A_P" packstat="0"
pack_input="0" pack_separated="1" tracer="0" />
    </MOVERS>
    <FILTERS>
        <AverageDegree name="avg_degree" threshold="0"
distance_threshold="10" task_operations="rtiv" />
        <Rmsd name="rmsd" superimpose="1" threshold="2500" />
    </FILTERS>
    <APPLY_TO_POSE>
```

```
    </APPLY_TO_POSE>
    <PROTOCOLS>
        <Add mover="iface_analyzer" />
        <Add filter="avg_degree" />
        <Add filter="rmsd" />
    </PROTOCOLS>
    <OUTPUT scorefxn="membrane_highres" />
</ROSETTASCRIPTS>
```

8.2 Rosetta analysis

Total_clusters.py

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

def load_data(file_path, column_index, condition_column_name,
condition_value, total_score_condition):
    data = pd.read_csv(file_path, delim_whitespace=True)

    #apply the conditions to filter data
    filtered_data = data[(data[condition_column_name] <
condition_value) & (data.iloc[:, 1] < total_score_condition)]

    return filtered_data.iloc[:, column_index]
file_names = [
    'c2.txt',
    'c4.txt',
    'c9.txt',
    'c1.txt',
    'c5.txt',
    'c3.txt',
    'c8.txt',
    'c7.txt',
    'c11.txt',
    'c6.txt',
    'c10.txt',
    'c18.txt',
    'c19.txt',
    'c14.txt',
    'c16.txt',
    'c21.txt',
    'c17.txt',
]

labels = [name.split('.')[0][1:] for name in file_names]
fig, (ax1, ax2, ax3) = plt.subplots(3, 1, figsize=(20, 10),
sharex=True, gridspec_kw={'hspace': 0.05})
#plot the data for each subplot with the conditions
def plot_data(ax, column_index, ylabel, is_first_plot=False,
is_last_plot=False, label_text='', condition_column_name=None,
condition_value=None, total_score_condition=None, label_pad=None):
    combined_data = pd.DataFrame({label: load_data(file_name,
column_index, condition_column_name, condition_value,
total_score_condition) for file_name, label in zip(file_names,
labels)})
```

```
    combined_data.columns = [label.replace('.txt', '') for label
in combined_data.columns]
    if is_first_plot:
        palette = ['forestgreen'] * len(combined_data.columns)
    elif is_last_plot:
        palette = ['grey'] * len(combined_data.columns)
    else:
        palette = ['blue'] * len(combined_data.columns)
    sns.violinplot(data=combined_data, palette=palette, ax=ax)
    sns.stripplot(data=combined_data, color='black', jitter=False,
marker='o', ax=ax)
    ax.set_ylabel(ylabel, fontsize=14, fontweight='bold',
labelpad=5)
    ax.tick_params(axis='both', which='major', labelsize=14)
    if is_first_plot:
        upper_data_points = combined_data.count()
        upper_labels = upper_data_points.tolist()
        upper_ax = ax.twiny()
        upper_ax.set_xlim(ax.get_xlim())
        upper_ax.set_xticks([tick for tick in ax.get_xticks()])
        upper_ax.set_xticklabels([f"{label}" for label in
upper_labels], fontsize=14, ha='center', rotation=0)
        upper_ax.set_xlabel("Population", fontsize=16,
fontweight='bold', labelpad=14)
        upper_ax.set_xlim(ax.get_xlim())
    if not is_last_plot:
        ax.tick_params(axis='x', which='both', bottom=False,
top=False, labelbottom=False)
    for tick in ax.get_xticklabels():
        tick.set_rotation(0)
    ax.xaxis.grid(True, linestyle='--', alpha=0.7)
    ax.set_axisbelow(True)
    ax.text(0.999, 0.009, label_text, transform=ax.transAxes,
fontsize=18, fontweight='bold', ha='right', va='bottom')
    if label_pad is not None and not is_first_plot:
        ax.yaxis.labelpad = label_pad
plot_data(ax1, 1, "Total Score (REU)", is_first_plot=True,
is_last_plot=False, condition_column_name='dG_separated',
condition_value=0, total_score_condition=-620)
plot_data(ax2, 6, "ΔG Binding (REU)", is_first_plot=False,
is_last_plot=False, condition_column_name='dG_separated',
condition_value=0, total_score_condition=-620, label_pad=10)
plot_data(ax3, 9, "ΔSASA (Å²)", is_first_plot=False,
is_last_plot=True, condition_column_name='dG_separated',
condition_value=0, total_score_condition=-620)
ax3.set_xlabel("Cluster", fontsize=16, fontweight='bold',
labelpad=5)
plt.tight_layout()
plt.show()
```

QQ_plots.py

To find if the data distributions is normal Gaussian or not.

```
import pandas as pd
import scipy.stats as stats
import matplotlib.pyplot as plt
import numpy as np
```

```
from matplotlib.lines import Line2D

data = pd.read_csv("c4.txt", delim_whitespace=True, header=None,
skiprows=1)
filtered_data = data[(data.iloc[:, 1] < -620) & (data.iloc[:, 6] <
0)]

columns_info = {
    1: ("Total Score (REU)", 'forestgreen', 16),
    6: ("ΔG Binding (REU)", 'blue', 16),
    9: ("ΔSASA (Å²)", 'grey', 16)
}

num_rows = len(columns_info)
num_cols = 2
fig, axs = plt.subplots(num_rows, num_cols, figsize=(10, 5 *
num_rows))
axs = axs.ravel()

for i, (column_index, (ylabel, color, fontsize)) in
enumerate(columns_info.items(), start=0):
    column_data = filtered_data.iloc[:,
column_index].dropna().astype(float)

    #histogram
    axs[2*i].hist(column_data, bins=30, color=color, label=ylabel)
    axs[2*i].set_ylabel('Frequency', fontsize=18,
fontweight='bold')
    axs[2*i].legend(loc='upper left', fontsize=16)
    axs[2*i].tick_params(axis='both', labelsize=16)
    axs[2*i].set_xlabel('')

    #Q-Q plot
    (osm, osr), (slope, intercept, r) =
stats.probplot(column_data, dist="norm", plot=None)
    axs[2*i+1].plot(osm, osr, 'o', color=color, label=ylabel)
    axs[2*i+1].plot(osm, slope*osm + intercept, 'r-',
label='Gaussian Distribution')
    axs[2*i+1].set_ylabel(ylabel, fontsize=16, fontweight='bold')
    axs[2*i+1].tick_params(axis='both', labelsize=16)  tick

    if i < num_rows - 1:
        axs[2*i+1].set_xlabel('')
        if i == 0 or i == 1:
            axs[2*i+1].tick_params(bottom=False,
labelbottom=False)

    else:
        axs[2*i+1].set_xlabel('Theoretical quantiles',
fontsize=18, fontweight='bold')

    #Kolmogorov-Smirnov test
    ks_stat, ks_p_value = stats.kstest((column_data -
np.mean(column_data)) / np.std(column_data, ddof=1), 'norm')
    normality = "Normal" if ks_p_value > 0.05 else "Not normal"
    p_value_display = f"{ks_p_value:.3e}" if ks_p_value < 0.001
else f"{ks_p_value:.3f}"
```

```
    text_legend_elements = [Line2D([0], [0], color='w', label=f'K-
S Test: {ks_stat:.3f}'),
                            Line2D([0], [0], color='w', label=f'p-
value = {ks_p_value:.3e}'),
                            Line2D([0], [0], color='w',
label=normality)]
    handles, labels = axs[2*i+1].get_legend_handles_labels()
    handles.extend(text_legend_elements)
    labels.extend([label.get_label() for label in
text_legend_elements])
    axs[2*i+1].legend(handles=handles[2:], labels=labels[2:],
loc='upper left', fontsize=14, handlelength=0, handletextpad=0)
    axs[2*i].set_aspect('auto')
    axs[2*i+1].set_aspect('auto')
fig.suptitle('Cluster 4', fontsize=20, fontweight='bold', y=0.99,
x=0.5)
plt.tight_layout(pad=4)
plt.subplots_adjust(top=0.95)
plt.show()
```

Best_clusters_pvalues.py

```
import pandas as pd
import matplotlib.pyplot as plt
from scipy.stats import mannwhitneyu
from itertools import combinations

def load_data(file_path, column_index, condition_column_name,
condition_value, total_score_condition):
    data = pd.read_csv(file_path, delim_whitespace=True)
    filtered_data = data[(data[condition_column_name] <
condition_value) & (data.iloc[:, 1] < total_score_condition)]
    return filtered_data.iloc[:, column_index]

def mann_whitney_test_with_effect(data1, data2):
    u_statistic, p_value = mannwhitneyu(data1, data2,
alternative='two-sided')
    n1 = len(data1)
    n2 = len(data2)
    effect_size = 1 - (2 * u_statistic) / (n1 * n2)
    return p_value, effect_size

def calculate_p_values_and_effects(file_names, cluster_names,
column_index, condition_column_name, condition_value,
total_score_condition):
    results = {}
    num_comparisons = len(file_names) * (len(file_names) - 1) / 2
#total number of pairwise comparisons
    for i in range(len(file_names)):
        for j in range(i + 1, len(file_names)):
            cluster1 = load_data(file_names[i], column_index,
condition_column_name, condition_value, total_score_condition)
            cluster2 = load_data(file_names[j], column_index,
condition_column_name, condition_value, total_score_condition)
            p_value, effect_size =
mann_whitney_test_with_effect(cluster1, cluster2)
```

```
            #apply Bonferroni correction
            corrected_p_value = p_value * num_comparisons
            if corrected_p_value > 1:
                corrected_p_value = 1.0

            key = f"{cluster_names[i]} vs {cluster_names[j]}"
            results[key] = {'p_value': corrected_p_value,
'effect_size': effect_size}
    return results

def add_p_value_brackets(ax, x1, x2, y, p_value, effect_size,
bracket_height=0.02, text_y_offset=0.1, p_fontsize=14,
es_fontsize=12, es_offset=0.1):
    bracket_height *= (ax.get_ylim()[1] - ax.get_ylim()[0])
    text_y_offset *= bracket_height
    ax.plot([x1, x1], [y, y + bracket_height], color='black',
lw=1.5)
    ax.plot([x1, x2], [y + bracket_height, y + bracket_height],
color='black', lw=1.5)
    ax.plot([x2, x2], [y, y + bracket_height], color='black',
lw=1.5)

    midpoint = (x1 + x2) / 2

    p_text = f"p={p_value:.2e}" if p_value < 0.001 else
f"p={p_value:.4f}"
    ax.text(midpoint, y + bracket_height + text_y_offset, p_text,
ha='center', va='bottom', color='black', fontsize=p_fontsize)
    abs_effect_size = abs(effect_size)
    es_color = 'green' if abs_effect_size >= 0.3 else 'goldenrod'
if abs_effect_size >= 0.1 else 'crimson'
    es_text = f"ES={effect_size:.2f}"
    ax.text(midpoint, y + bracket_height - (text_y_offset +
es_offset), es_text, ha='center', va='top', color=es_color,
fontsize=es_fontsize, fontweight='bold')

file_names = ['c2.txt', 'c4.txt', 'c1.txt', 'c5.txt']
cluster_names = ['2', '4', '1', '5']
violin_width = 6
tick_positions = [5, 15, 25, 35]

fig, axes = plt.subplots(1, 3, figsize=(20, 10), sharex=True,
sharey=False, gridspec_kw={'wspace': 0.5})

specified_y_coordinates_plot1 = [-618, -608, -598, -588, -578, -
568]
specified_y_coordinates_plot2 = [1, 6, 10, 14, 18, 22]
specified_y_coordinates_plot3 = [1800, 1910, 2020, 2130, 2240,
2350]

for ax, column_index, ylabel, specified_y_coordinates in zip(axes,
[1, 6, 9], ["Total Score (REU)", "ΔG Binding (REU)", "ΔSASA
(Å²)"], [specified_y_coordinates_plot1,
specified_y_coordinates_plot2, specified_y_coordinates_plot3]):
    for file_name, label, position in zip(file_names,
cluster_names, tick_positions):
```

```
        data = load_data(file_name, column_index, 'dG_separated',
0, -620)
        parts = ax.violinplot([data], positions=[position],
widths=violin_width, showmedians=True, showextrema=False)
        color = {'2': 'grey', '4': 'cyan', '1': 'forestgreen',
'5': 'magenta'}.get(label, 'black')
        for pc in parts['bodies']:
            pc.set_facecolor(color)
            pc.set_edgecolor('black')
            pc.set_alpha(0.7)
        ax.scatter([position] * len(data), data, color='black',
alpha=0.7, s=10)
        parts['cmedians'].set_edgecolor('black')
    ax.set_ylabel(ylabel, fontsize=20, fontweight='bold',
labelpad=5)
    ax.set_xticks(tick_positions)
    ax.set_xticklabels([int(cluster) for cluster in
cluster_names], fontsize=16)
    ax.tick_params(axis='both', which='major', labelsize=16)
    ax.xaxis.grid(True, linestyle='--', alpha=0.7)
    ax.set_axisbelow(True)
    if ylabel == "Total Score (REU)":
        ax.set_ylim(-689, -550)
    elif ylabel == "ΔG Binding (REU)":
        ax.set_ylim(-29, 29)
    elif ylabel == "ΔSASA (Å²)":
        ax.set_ylim(865, 2500)
    upper_ax = ax.twiny()
    upper_ax.set_xlim(ax.get_xlim())
    upper_ax.set_xticks(tick_positions)
    upper_ax.set_xticklabels([f"{load_data(file_name,
column_index, 'dG_separated', 0, -620).count()}" for file_name in
file_names], fontsize=16, ha='center', rotation=0)
    upper_ax.set_xlabel("Population", fontsize=20,
fontweight='bold', labelpad=14)
    results = calculate_p_values_and_effects(file_names,
cluster_names, column_index, 'dG_separated', 0, -620)
    for ((cluster1, cluster2), y_coord) in
zip(combinations(cluster_names, 2), specified_y_coordinates):
        key = f"{cluster1} vs {cluster2}"
        if key in results:
            p_value = results[key]['p_value']
            effect_size = results[key]['effect_size']
            x1, x2 =
tick_positions[cluster_names.index(cluster1)],
tick_positions[cluster_names.index(cluster2)]
            add_p_value_brackets(ax, x1, x2, y_coord, p_value,
effect_size)
    ax.set_xlabel('Cluster', fontsize=20, fontweight='bold',
labelpad=10)
axes[1].set_yticks([0, -5, -10, -15, -20, -25])
axes[0].set_yticks([-620, -640, -660, -680])
axes[2].set_yticks([1800, 1600, 1400, 1200, 1000])
plt.tight_layout()
plt.show()
```

Renumber.py

```
#renumbers all pdb files to original numbering and make a new
directory for them.
import os
from Bio import PDB

def renumber_pdb(input_file, output_file, chain_a_start,
chain_p_start):
    parser = PDB.PDBParser(QUIET=True)
    structure = parser.get_structure('struct', input_file)

    #step 1: assign temporary unique identifiers
    temp_id = 10000
    for model in structure:
        for chain in model:
            for residue in chain.get_unpacked_list():
                if residue.id[0] == ' ':
                    residue.id = (' ', temp_id, ' ')
                    temp_id += 1

    #step 2: assign final residue numbers
    for model in structure:
        for chain in model:
            if chain.id == 'A':
                #renumber Chain A
                residue_num = chain_a_start
                for residue in chain.get_unpacked_list():
                    residue.id = (' ', residue_num, ' ')
                    residue_num += 1
            elif chain.id == 'P':
                #renumber Chain P
                residue_num = chain_p_start
                for residue in chain.get_unpacked_list():
                    residue.id = (' ', residue_num, ' ')
                    residue_num += 1

    io = PDB.PDBIO()
    io.set_structure(structure)
    io.save(output_file)
#parameters
source_directory = '.'
destination_directory = 'renumber'
if not os.path.exists(destination_directory):
    os.makedirs(destination_directory)

for file in os.listdir(source_directory):
    if file.endswith(".pdb") and "c." in file:
        input_file = os.path.join(source_directory, file)
        output_file = os.path.join(destination_directory, file)
        renumber_pdb(input_file, output_file, chain_a_start=12,
chain_p_start=1)
print("Renumbering complete.")
```

Find_contacts.py

```
import os
from Bio import PDB
```

```
def find_contacts(pdb_file, chain_gpcr, chain_peptide,
distance_threshold=5.0):
    parser = PDB.PDBParser(QUIET=True)
    structure = parser.get_structure('struct', pdb_file)
    #function to choose the right atom for distance calculation
    def get_representative_atom(residue):
        if residue.get_resname() == 'GLY':
            return residue['CA'] if 'CA' in residue else None
        else:
            return residue['CB'] if 'CB' in residue else
residue.get('CA', None)
    #extract representative atoms from GPCR and peptide
    rep_atoms_gpcr = []
    rep_atoms_peptide = []
    for model in structure:
        for chain in model:
            if chain.id == chain_gpcr:
                for residue in chain.get_unpacked_list():
                    if residue.id[0] == ' ':
                        atom = get_representative_atom(residue)
                        if atom:
                            rep_atoms_gpcr.append(atom)
            elif chain.id == chain_peptide:
                for residue in chain.get_unpacked_list():
                    if residue.id[0] == ' ':
                        atom = get_representative_atom(residue)
                        if atom:
                            rep_atoms_peptide.append(atom)
    #check if there are atoms to work with
    if not rep_atoms_gpcr or not rep_atoms_peptide:
        print(f"No valid atoms found in {pdb_file} for contact
analysis.")
        return []
    # finding contacts
    ns = PDB.NeighborSearch(rep_atoms_gpcr)
    contacts = []
    for atom in rep_atoms_peptide:
        close_atoms = ns.search(atom.get_coord(),
distance_threshold)
        for close_atom in close_atoms:
            contacts.append((atom, close_atom))
    return contacts
source_directory = '.'
output_filename = 'contacts_c5.txt'
chain_gpcr = 'A'  #chain ID of the GPCR
chain_peptide = 'P'  #chain ID of the peptide
with open(output_filename, 'w') as output_file:
    for file in os.listdir(source_directory):
        if file.endswith(".pdb") and "c.5." in file:
            pdb_file = os.path.join(source_directory, file)
            contacts = find_contacts(pdb_file, chain_gpcr,
chain_peptide)
            output_file.write(f'Contacts in {file}:\n')
            for atom1, atom2 in contacts:
                res1 = atom1.get_parent()
                res2 = atom2.get_parent()
```

```
                output_file.write(f"{res1.get_resname()}
{res1.id[1]} {atom1.get_name()} - {res2.get_resname()}
{res2.id[1]} {atom2.get_name()}\n")
            output_file.write("\n")
print("Contact analysis complete. Results are in
'contacts_c5.txt'")
```

Best_fontacts.py

```
import os
import shutil

#target residue numbers
target_residues = {23, 38, 178, 174, 280, 283, 284, 286, 290, 306,
323, 302, 309, 308, 305, 102}
def find_and_copy_relevant_pdbs(contact_files, source_dir,
dest_dir):
    if not os.path.exists(dest_dir):
        os.makedirs(dest_dir)
    relevant_pdbs = set()
    for contact_file in contact_files:
        with open(contact_file, 'r') as file:
            current_pdb = None
            for line in file:
                if line.startswith("Contacts in"):
                    #extracting the filename without adding extra
'.pdb'
                    current_pdb = line.split()[2].rstrip(':')
                    if not current_pdb.endswith(".pdb"):
                        current_pdb += ".pdb"
                elif current_pdb and any(str(residue) in line for
residue in target_residues):
                    relevant_pdbs.add(current_pdb)
    for pdb in relevant_pdbs:
        source_path = os.path.join(source_dir, pdb)
        dest_path = os.path.join(dest_dir, pdb)
        if os.path.exists(source_path):
            shutil.copy(source_path, dest_path)
        else:
            print(f"File not found: {source_path}")
    return relevant_pdbs
#parameters
contact_files = ['contacts_c2.txt', 'contacts_c4.txt',
'contacts_c1.txt', 'contacts_c5.txt']
source_directory = '.'
destination_directory = 'best_contacts'
#process the files
relevant_pdbs = find_and_copy_relevant_pdbs(contact_files,
source_directory, destination_directory)
print(f"Copied {len(relevant_pdbs)} PDB files to
{destination_directory}")
```

Contact_map.py

Plots the best clusters scanned through literature indicated contacts.

```
from collections import defaultdict
```

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib as mpl
from mpl_toolkits.axes_grid1 import make_axes_locatable
from matplotlib.cm import ScalarMappable

def extract_residue_number(part):
    try:
        return int(''.join(filter(str.isdigit, part)))
    except ValueError:
        return None
contact_files = ['contacts_c2.txt', 'contacts_c4.txt',
'contacts_c1.txt', 'contacts_c5.txt']
titles = {'contacts_c2.txt': 'Cluster 2', 'contacts_c4.txt':
'Cluster 4',
          'contacts_c1.txt': 'Cluster 1', 'contacts_c5.txt':
'Cluster 5'}
bold_residues = {23, 32, 38, 121, 174, 178, 280, 283, 284, 286,
290, 306, 323, 302, 102, 308, 309, 98}
title_colors = {
    'contacts_c2.txt': 'black',
    'contacts_c4.txt': 'black',
    'contacts_c1.txt': 'black',
    'contacts_c5.txt': 'black'
}

for cluster_file in contact_files:
    contact_frequency = defaultdict(lambda: defaultdict(int))
    total_pdb_files = 0
    peptide_residues_in_contact = set()
    receptor_residues_in_contact = set()

    with open(cluster_file, 'r') as file:
        for line in file:
            if line.startswith("Contacts in"):
                total_pdb_files += 1
                current_pdb = line.strip().split()[2]
                continue
            parts = line.split('-')
            if len(parts) == 2:
                peptide_residue = extract_residue_number(parts[0])
                receptor_residue =
extract_residue_number(parts[1])
                if peptide_residue is not None and
receptor_residue is not None:

peptide_residues_in_contact.add(peptide_residue)

receptor_residues_in_contact.add(receptor_residue)

contact_frequency[peptide_residue][receptor_residue] += 1
    peptide_residues_sorted = sorted(peptide_residues_in_contact)
    receptor_residues_sorted =
sorted(receptor_residues_in_contact)
    matrix = np.zeros((len(peptide_residues_sorted),
len(receptor_residues_sorted)))
```

```
    for i, pep_res in enumerate(peptide_residues_sorted):
        for j, rec_res in enumerate(receptor_residues_sorted):
            matrix[i, j] = contact_frequency[pep_res][rec_res]

    frequency_ticks = {
        'contacts_c2.txt': [1, 3, 5, 7, 10],
        'contacts_c4.txt': [1, 250, 550, 800, 1093],
        'contacts_c1.txt': [1, 200, 400, 600, 838],
        'contacts_c5.txt': [1, 200, 450, 700, 947]
    }
    custom_cmap =
mpl.colors.ListedColormap(sns.color_palette("viridis",
as_cmap=True)(np.linspace(0, 1, 256)))
    custom_cmap.set_under(color='white')
    fig = plt.figure(figsize=(10, 10))
    ax = fig.add_subplot(111)
    heatmap = sns.heatmap(matrix, ax=ax, cmap=custom_cmap,
annot=False, cbar=False, linewidths=.5, linecolor='black',
                          vmin=0, vmax=total_pdb_files,
mask=matrix==0)
    ax.set_title(titles[cluster_file], fontweight='bold',
fontsize=18)
    ax.set_ylabel('GnRH', fontweight='bold', fontsize=18)
    ax.set_xlabel('GnRH1R', fontweight='bold', fontsize=18)
    ax.set_xticks(np.arange(len(receptor_residues_sorted)) + 0.5)
    ax.set_yticks(np.arange(len(peptide_residues_sorted)) + 0.5)
    xtick_labels = [f'$\\mathbf{{{residue}}}$' if residue in
bold_residues else str(residue) for residue in
receptor_residues_sorted]
    ax.set_xticklabels(xtick_labels, rotation=90, fontsize=14)
    ax.set_yticklabels(peptide_residues_sorted, fontsize=14)
    cluster_title = titles[cluster_file]
    ax.set_title(cluster_title, fontweight='bold', fontsize=18,
color=title_colors[cluster_file])

    #annotate boxes with frequency values
    for i in range(len(peptide_residues_sorted)):
        for j in range(len(receptor_residues_sorted)):
            value =
contact_frequency[peptide_residues_sorted[i]][receptor_residues_so
rted[j]]
            if value > 0:
                ax.text(j + 0.5, i + 0.5, str(value), ha='center',
va='center', color='white', fontsize=10, fontweight='bold',
rotation=90)
    ax.spines['top'].set_visible(False)
    ax.spines['right'].set_visible(True)
    ax.spines['bottom'].set_visible(True)
    ax.spines['left'].set_visible(False)
    max_freq = total_pdb_files
    sm = ScalarMappable(cmap=custom_cmap,
norm=plt.Normalize(vmin=0, vmax=max_freq))
    sm.set_array([])
    divider = make_axes_locatable(ax)
    cax = divider.append_axes("right", size="2.5%", pad=0.04)
    cbar = plt.colorbar(sm, cax=cax,
ticks=frequency_ticks[cluster_file])
```

```
    cbar.set_label('Population', fontweight='bold', fontsize=14)
    cbar.ax.tick_params(labelsize=12)
    fig.savefig(f'{cluster_file.replace(".txt", "_heatmap.png")}')
    plt.close(fig)
```

Hbonds.py

```
import os
import numpy as np
import matplotlib.pyplot as plt
from Bio.PDB import PDBParser, NeighborSearch
import glob
from mpl_toolkits.axes_grid1 import make_axes_locatable

parser = PDBParser(QUIET=True)
def is_hbond_donor(atom):
    return 'H' in atom.get_id()
def is_hbond_acceptor(atom):
    return atom.element in ('O', 'N')
def find_hbonds_between_chains(chain_donor, chain_acceptor):
    ns = NeighborSearch(list(chain_acceptor.get_atoms()))
    hbonds = set()
    for residue in chain_donor:
        for donor_atom in residue:
            if is_hbond_donor(donor_atom):
                acceptor_atoms = ns.search(donor_atom.coord, 3.5)
                for acceptor_atom in acceptor_atoms:
                    if is_hbond_acceptor(acceptor_atom):
                        hbonds.add((residue.get_id()[1],
acceptor_atom.get_parent().get_id()[1]))
    return list(hbonds)

def plot_hbond_density(hbonds, title, ax, title_color, max_freq,
bold_residues, fontweight='bold', fontsize=20):
    if not hbonds:
        print(f"No hydrogen bonds found for {title}")
        return

    num_ticks = 5
    cbar_ticks = {
        'Cluster 2': [1, 3, 5, 7, 10],
        'Cluster 4': [1, 250, 550, 800, 1093],
        'Cluster 1': [1, 200, 400, 600, 838],
        'Cluster 5': [1, 200, 450, 700, 947]
    }
    donors, acceptors = zip(*hbonds)
    donor_residues = sorted(set(donors))
    acceptor_residues = sorted(set(acceptors))
    hbond_freq = {}
    for d, a in hbonds:
        hbond_freq[(d, a)] = hbond_freq.get((d, a), 0) + 1

    matrix = np.full((len(acceptor_residues),
len(donor_residues)), np.nan)
    for (d, a), freq in hbond_freq.items():
        matrix[acceptor_residues.index(a),
donor_residues.index(d)] = freq
```

```
    acceptor_residues = acceptor_residues[::-1]
    matrix = matrix[::-1, :]
    cmap = plt.cm.viridis.copy()
    cmap.set_bad(color='white')
    cax_plot = ax.matshow(matrix, cmap=cmap, origin='lower',
vmin=1, vmax=max_freq, aspect='auto')
    ax.set_xticks(range(len(donor_residues)))
    ax.set_yticks(range(len(acceptor_residues)))
    x_tick_labels = [f'{x:.0f}' for x in donor_residues]

    for i, label in enumerate(x_tick_labels):
        if int(label) in bold_residues:

ax.xaxis.get_major_ticks()[i].label1.set_fontweight('bold')

    ax.set_xticklabels(x_tick_labels, fontsize=10)
    ax.set_yticklabels([f'{y:.0f}' for y in acceptor_residues],
fontsize=10)
    ax.tick_params(axis='x', which='major', length=4,
labelrotation=90)
    ax.tick_params(axis='y', which='major', length=4)
    ax.xaxis.set_ticks_position('bottom')
    ax.xaxis.set_label_position('bottom')
    ax.tick_params(axis='x', which='minor', bottom=False,
length=6, labelrotation=90)
    ax.set_ylabel('GnRH', fontweight='bold', fontsize=18,
labelpad=5)
    ax.yaxis.set_label_position('left')
    ax.tick_params(axis='y', which='minor', left=False, length=4)
    ax.set_xticks(np.arange(-.5, len(donor_residues), 1),
minor=True)
    ax.set_yticks(np.arange(-.5, len(acceptor_residues), 1),
minor=True)
    ax.grid(which='minor', color='black', linestyle='-',
linewidth=1)
    divider = make_axes_locatable(ax)
    cax = divider.append_axes("right", size="2.5%", pad=0.05)
    cbar_ticks_plot = cbar_ticks.get(title, [1, 2, 3, 4, 5])
    cbar = plt.colorbar(cax_plot, cax=cax, aspect=40, pad=0.04,
ticks=cbar_ticks_plot)
    cbar.set_label('Population', fontweight='bold', fontsize=14)
    cbar.set_ticklabels([str(int(tick)) for tick in
cbar_ticks_plot])
    ax.set_xlabel('GnRH1R', fontweight='bold', fontsize=18,
labelpad=5)
    ax.set_title(title, fontsize=fontsize, fontweight=fontweight,
color=title_color, y=0.99, pad=20)

    #annotate
    for i in range(len(acceptor_residues)):
        for j in range(len(donor_residues)):
            freq = matrix[i, j]
            if not np.isnan(freq) and freq >= 1:
                ax.text(j, i, f'{int(freq)}', va='center',
ha='center', color='white', rotation=90, fontweight='bold',
fontsize=8)
```

```
def calculate_max_frequency(pdb_files):
    return len(pdb_files)

pdb_files_path = '.'
cluster_paths = ['*c.2.*pdb', '*c.4.*pdb', '*c.1.*pdb',
'*c.5.*pdb']
cluster_titles = ['Cluster 2', 'Cluster 4', 'Cluster 1', 'Cluster
5']

title_colors = {
    'Cluster 2': 'black',
    'Cluster 4': 'black',
    'Cluster 1': 'black',
    'Cluster 5': 'black'
}
bold_residues = {23, 32, 38, 121, 174, 178, 280, 283, 284, 286,
290, 306, 323, 302, 102, 308, 309, 98}

for cluster_path, cluster_title in zip(cluster_paths,
cluster_titles):
    pdb_files = glob.glob(os.path.join(pdb_files_path,
cluster_path))
    cluster_hbonds = []

    for pdb_file in pdb_files:
        structure = parser.get_structure('cluster', pdb_file)
        chain_A = structure[0]['A']
        chain_P = structure[0]['P']
        hbonds = find_hbonds_between_chains(chain_A, chain_P)
        cluster_hbonds.extend(hbonds)
    title_color = title_colors.get(cluster_title, 'black')
    max_freq = calculate_max_frequency(pdb_files)
    fig, ax = plt.subplots(figsize=(10, 10))
    plot_hbond_density(cluster_hbonds, cluster_title, ax,
title_color, max_freq, bold_residues, fontweight='bold',
fontsize=18)
    plt.show()
```

Cluster_CoM.py

```
import os
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

def get_center_of_mass_from_pdb(pdb_file, chain_id):
    with open(pdb_file, 'r') as file:
        lines = file.readlines()
    coordinates = []
    for line in lines:
        if line.startswith('ATOM') and line[21] == chain_id:
            x = float(line[30:38].strip())
            y = float(line[38:46].strip())
            z = float(line[46:54].strip())
            coordinates.append((x, y, z))
    if not coordinates:
```

```
        raise ValueError(f"No coordinates found for chain
{chain_id} in {pdb_file}")
    center_of_mass = np.mean(coordinates, axis=0)
    return center_of_mass

def get_binding_pocket_coords_and_sizes(pdb_file, chain_id,
pocket_residues):
    with open(pdb_file, 'r') as file:
        lines = file.readlines()
    pocket_coords = {}
    for residue_number in pocket_residues:
        residue_atoms = []
        for line in lines:
            if line.startswith('ATOM') and line[21] == chain_id
and int(line[22:26].strip()) == residue_number:
                x = float(line[30:38].strip())
                y = float(line[38:46].strip())
                z = float(line[46:54].strip())
                residue_atoms.append((x, y, z))
        if residue_atoms:
            pocket_coords[residue_number] =
np.array(residue_atoms)
    return pocket_coords

def calculate_residue_sizes(pocket_coords):
    residue_sizes = {}
    for residue_number, coords in pocket_coords.items():
        size = np.std(coords, axis=0).mean()
        residue_sizes[residue_number] = size * 100
    return residue_sizes

def plot_binding_pocket(ax, binding_pocket_coords, residue_sizes,
pocket_residues_info):
    for residue_number, coords in binding_pocket_coords.items():
        info = next((item for item in pocket_residues_info if
item[0] == residue_number), None)
        if info:
            _, color, label = info
            size = residue_sizes[residue_number]
            center = coords.mean(axis=0)
            ax.scatter(*center, color=color, marker='o', s=size)
            ax.text(*center, '  ' + label, color=color,
fontsize=9)

def plot_clusters(ax, cluster_files, label, color):
    first_file = True
    for cluster_file in cluster_files:
        cluster_coords = get_center_of_mass_from_pdb(cluster_file,
'P')
        if cluster_coords is not None:
            if first_file:
                ax.scatter(*cluster_coords, marker='o',
label=label, color=color, s=50)
                first_file = False
            else:
                ax.scatter(*cluster_coords, marker='o',
color=color, s=50)
```

```
directory =
'/users/nkb19202/ROSETTA_af/flexpepdock/template/cluster_2/cl_tem/
cluster/best_clusters/renumber/best_contacts/candidates'
cluster_files = {}
for cluster in [1, 2, 4, 5]:
    cluster_files[cluster] = [os.path.join(directory, file) for
file in os.listdir(directory) if file.endswith('.pdb') and
f'c.{cluster}.' in file]

pocket_residues_info = [
    (23, 'pink', 'L23'),
    (38, 'gold', 'R38'),
    (174, 'chocolate', 'Q174'),
    (178, 'violet', 'F178'),
    (280, 'gray', 'W280'),
    (283, 'limegreen', 'Y283'),
    (284, 'limegreen', 'Y284'),
    (286, 'pink', 'L286'),
    (290, 'limegreen', 'Y290'),
    (306, 'mediumturquoise', 'H306'),
    (323, 'limegreen', 'Y323'),
    (302, 'red', 'D302'),
    (102, 'goldenrod', 'N102'),
    (121, 'steelblue', 'K121'),
    (98, 'red', 'D98'),
]

pocket_residue_numbers = [residue_info[0] for residue_info in
pocket_residues_info]

representative_pdb = '/path/to/file/c.5.314.pdb'
binding_pocket_coords =
get_binding_pocket_coords_and_sizes(representative_pdb, 'A',
pocket_residue_numbers)
residue_sizes = calculate_residue_sizes(binding_pocket_coords)

fig = plt.figure(figsize=(12, 8))
ax = fig.add_subplot(111, projection='3d')
cluster_colors = ['green', 'grey', 'cyan', 'magenta']
cluster_labels = ['Cluster 1', 'Cluster 2', 'Cluster 4', 'Cluster
5']
for idx, (cluster, files) in enumerate(cluster_files.items()):
    plot_clusters(ax, files, cluster_labels[idx],
cluster_colors[idx])
plot_binding_pocket(ax, binding_pocket_coords, residue_sizes,
pocket_residues_info)
ax.set_xlabel('X-axis')
ax.set_ylabel('Y-axis')
ax.set_zlabel('Z-axis')
ax.set_title('Cluster CoM relative to GnRH1R binding pocket')
ax.view_init(elev=-205, azim=230)
ax.legend(loc='lower center', bbox_to_anchor=(1, 0.85))
plt.show()
```

Rank_contacts.py

```
import os
import shutil

#target residue numbers
target_residues = {23, 38, 178, 174, 280, 283, 284, 286, 290, 306,
323, 302, 309, 308, 305, 102}
contact_files = ['contacts_c2.txt', 'contacts_c4.txt',
'contacts_c1.txt', 'contacts_c5.txt']
pdb_contact_count = {}

#process contact files
for file_name in contact_files:
    with open(file_name, 'r') as file:
        current_pdb = ''
        for line in file:
            if line.startswith("Contacts in"):
                current_pdb = line.split()[2].strip(':')
            else:
                parts = line.split()
                if parts and int(parts[5]) in target_residues:
                    pdb_contact_count[current_pdb] =
pdb_contact_count.get(current_pdb, 0) + 1
#sort PDB files by the number of contacts
sorted_pdbs = sorted(pdb_contact_count.items(), key=lambda x:
x[1], reverse=True)
#create candidates directory
candidates_dir = 'candidates'
os.makedirs(candidates_dir, exist_ok=True)
#move top PDB files to candidates directory and write to a ranking
file
ranking_file = 'pdb_ranking.txt'
with open(ranking_file, 'w') as rank_file:
    for pdb, count in sorted_pdbs:
        shutil.copy(pdb, os.path.join(candidates_dir, pdb))
        rank_file.write(f"{pdb}: {count} contacts\n")
print(f"PDB files ranked and copied to {candidates_dir}. Ranking
saved in {ranking_file}.")
```

Candidate_selection.py

```
import pandas as pd
def read_energy_file(filename):
    df_energy = pd.read_csv(filename, sep='\s+', skiprows=2)
    #extract the 'dG_separated' (index 6) and 'description' (index
44) columns
    df_energy = df_energy.iloc[:, [6, 44]]
    #rename the columns for clarity
    df_energy.columns = ['dG_separated', 'file_name']
    return df_energy
def read_ranking_file(filename):
    #read the ranking file
    with open(filename, 'r') as file:
        lines = file.readlines()
    #extract file names and number of contacts
    data = []
    for line in lines:
```

```
        parts = line.split(':')
        file_name = parts[0].strip('.pdb')  #remove .pdb extension
        contacts = int(parts[1].split()[0])
        data.append({'file_name': file_name, 'contacts':
contacts})
    return pd.DataFrame(data)
#read and process the files
df_energy = read_energy_file('30_15_gr_seperated.txt')
df_ranking = read_ranking_file('pdb_ranking.txt')
#merge the dataframes on the file names
df_merged = pd.merge(df_energy, df_ranking, on='file_name',
how='inner')
#sort by dG_separated (ascending) and contacts (descending)
df_sorted = df_merged.sort_values(by=['dG_separated', 'contacts'],
ascending=[True, False])
#output the top structures
top_structures = df_sorted.head(100)
#write the results to an output file
output_file = 'top_ranked_candidates.txt'
top_structures.to_csv(output_file, index=False, sep='\t')
print(f"Top structures have been saved to {output_file}.")
```

The output file provides ranking of all selected structures based on best scoring binding energy and most contacts.

```
dG_separated file_name   contacts
-23.713      c.1.209       1
-23.128      c.1.211       1
-23.017      c.5.239       2
-23.014       c.5.50       3
-23.012      c.5.203       5
-22.959       c.5.3        4
-22.85       c.5.362       2
-22.469       c.5.81       4
…
```

GnRH_position.py

```
import os
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

def get_center_of_mass_from_pdb(pdb_file, chain_id):
    with open(pdb_file, 'r') as file:
        lines = file.readlines()
    coordinates = []
    for line in lines:
        if line.startswith('ATOM') and line[21] == chain_id:
            x = float(line[30:38].strip())
            y = float(line[38:46].strip())
            z = float(line[46:54].strip())
            coordinates.append((x, y, z))
    if not coordinates:
```

```
        raise ValueError(f"No coordinates found for chain
{chain_id} in {pdb_file}")
    center_of_mass = np.mean(coordinates, axis=0)
    return center_of_mass
def get_binding_pocket_coords_and_sizes(pdb_file, chain_id,
pocket_residues):
    with open(pdb_file, 'r') as file:
        lines = file.readlines()
    pocket_coords = {}
    for residue_number in pocket_residues:
        residue_atoms = []
        for line in lines:
            if line.startswith('ATOM') and line[21] == chain_id
and int(line[22:26].strip()) == residue_number:
                x = float(line[30:38].strip())
                y = float(line[38:46].strip())
                z = float(line[46:54].strip())
                residue_atoms.append((x, y, z))
        if residue_atoms:
            pocket_coords[residue_number] =
np.array(residue_atoms)
    return pocket_coords
def calculate_residue_sizes(pocket_coords):
    residue_sizes = {}
    for residue_number, coords in pocket_coords.items():
        size = np.std(coords, axis=0).mean()
        residue_sizes[residue_number] = size * 100
    return residue_sizes
def plot_binding_pocket(ax, binding_pocket_coords, residue_sizes,
pocket_residues_info):
    for residue_number, coords in binding_pocket_coords.items():
        info = next((item for item in pocket_residues_info if
item[0] == residue_number), None)
        if info:
            _, color, label = info
            size = residue_sizes[residue_number]
            center = coords.mean(axis=0)
            ax.scatter(*center, color=color, marker='o', s=size)
            ax.text(*center, '  ' + label, color=color,
fontsize=9)
def plot_clusters(ax, cluster_files, label, color):
    first_file = True
    for cluster_file in cluster_files:
        cluster_coords = get_center_of_mass_from_pdb(cluster_file,
'P')
        if cluster_coords is not None:
            if first_file:
                ax.scatter(*cluster_coords, marker='o',
label=label, color=color, s=50)
                first_file = False
            else:
                ax.scatter(*cluster_coords, marker='o',
color=color, s=50)
directory =
'/users/nkb19202/ROSETTA_af/flexpepdock/template/cluster_2/cl_tem/
cluster/best_clusters/renumber/best_contacts/candidates'
cluster_files = {}
```

```
for cluster in [1, 2, 4, 5]:
    cluster_files[cluster] = [os.path.join(directory, file) for
file in os.listdir(directory) if file.endswith('.pdb') and
f'c.{cluster}.' in file]

pocket_residues_info = [
    (23, 'pink', 'L23'),
    (38, 'gold', 'R38'),
    (174, 'chocolate', 'Q174'),
    (178, 'violet', 'F178'),
    (280, 'gray', 'W280'),
    (283, 'limegreen', 'Y283'),
    (284, 'limegreen', 'Y284'),
    (286, 'pink', 'L286'),
    (290, 'limegreen', 'Y290'),
    (306, 'mediumturquoise', 'H306'),
    (323, 'limegreen', 'Y323'),
    (302, 'red', 'D302'),
    (102, 'goldenrod', 'N102'),
    (121, 'skyblue', 'K121'),
    (98, 'red', 'D98'),
]
pocket_residue_numbers = [residue_info[0] for residue_info in
pocket_residues_info]
representative_pdb = '/path/to/file/c.5.314.pdb'
binding_pocket_coords =
get_binding_pocket_coords_and_sizes(representative_pdb, 'A',
pocket_residue_numbers)
residue_sizes = calculate_residue_sizes(binding_pocket_coords)
fig = plt.figure(figsize=(12, 8))
ax = fig.add_subplot(111, projection='3d')
cluster_colors = ['green', 'grey', 'cyan', 'magenta']
cluster_labels = ['Cluster 1', 'Cluster 2', 'Cluster 4', 'Cluster
5']
for idx, (cluster, files) in enumerate(cluster_files.items()):
    plot_clusters(ax, files, cluster_labels[idx],
cluster_colors[idx])
plot_binding_pocket(ax, binding_pocket_coords, residue_sizes,
pocket_residues_info)
ax.set_xlabel('X-axis')
ax.set_ylabel('Y-axis')
ax.set_zlabel('Z-axis')
ax.set_title('Cluster COM relative to GnRH1R binding pocket')
ax.view_init(elev=-205, azim=230)
ax.legend(loc='lower center', bbox_to_anchor=(1, 0.85))
plt.show()
```

Candidate_rank.py

```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches

file_path = '/path/to/file/2top_ranked_candidates.txt'
data = pd.read_csv(file_path, sep='\t')

fig, ax = plt.subplots()
```

```
def assign_color(x):
    if 'c.1.' in x:
        return 'green'
    elif 'c.4.' in x:
        return 'cyan'
    elif 'c.5.' in x:
        return 'magenta'
    else:
        return 'gray'

colors = data['file_name'].apply(assign_color)

ax.bar(data['file_name'], data['dG_separated'], color=colors,
alpha=0.7)

ax.bar(data['file_name'], data['contacts'], color='gray',
alpha=0.5)

ax.set_xlabel('Pose', fontsize=12, fontweight='bold')
ax.set_ylabel('ΔG Binding (REU)
Contacts', fontsize=10, fontweight='bold')
plt.title('Candidate selection', fontsize=12, fontweight='bold')
legend_labels = ['Cluster 1', 'Cluster 4', 'Cluster 5']
legend_colors = ['green', 'cyan', 'magenta']
legend_patches = [mpatches.Patch(color=color, label=label) for
color, label in zip(legend_colors, legend_labels)]
plt.xticks([])
plt.yticks([-24, -23, -22, -21, -20, 1, 2, 3, 4, 5])
ax.axhline(0, color='black', linewidth=1)
plt.legend(handles=legend_patches, loc='lower right', fontsize=8)
plt.tight_layout()
plt.show()
```

8.3 MD analysis

SaltBridge.py

```
import matplotlib.pyplot as plt
import numpy as np
import re
import seaborn as sns

markersize = 48

data_files = [
    'saltbr-GLU68_chainP_segnamePROA-
LYS71_chainP_segnamePROA.dat',
    'saltbr-GLU90_chainP_segnamePROA-ARG8_chainP_segnamePROB.dat',
    'saltbr-GLU90_chainP_segnamePROA-
LYS121_chainP_segnamePROA.dat',
    'saltbr-ASP98_chainP_segnamePROA-
ARG38_chainP_segnamePROA.dat',
    'saltbr-ASP98_chainP_segnamePROA-
LYS121_chainP_segnamePROA.dat',
    'saltbr-GLU111_chainP_segnamePROA-
LYS115_chainP_segnamePROA.dat',
```

```
    'saltbr-ASP138_chainP_segnamePROA-
ARG75_chainP_segnamePROA.dat',
    'saltbr-ASP138_chainP_segnamePROA-
ARG139_chainP_segnamePROA.dat',
    'saltbr-ASP185_chainP_segnamePROA-
LYS191_chainP_segnamePROA.dat',
    'saltbr-ASP293_chainP_segnamePROA-
ARG299_chainP_segnamePROA.dat',
]

fig, ax1 = plt.subplots(figsize=(10, 8))
ax2 = ax1.twinx()

all_x_values = []
all_y_values = []
durations = []

color_iterator = iter(sns.color_palette("viridis",
len(data_files)))

labels = []

for i, data_file in enumerate(data_files):
    with open(data_file, 'r') as file:
        data = [line.split() for line in file.readlines()]
        frames, distances = zip(*[(int(frame), float(distance))
for frame, distance in data])

    filtered_data = [(frame * 2 / 100, i * markersize + markersize
/ 2) for frame, distance in zip(frames, distances) if distance <=
3.51]

    duration_frames = len(filtered_data)
    duration_ns = duration_frames * 2 / 100

    all_x_values.extend([x for x, _ in filtered_data])
    all_y_values.extend([y for _, y in filtered_data])
    durations.append(duration_ns)

    match = re.match(r'^saltbr-([A-Z]+)(\d+)_chain([A-
Z]+)_segname[A-Z]+-([A-Z]+)(\d+)_chain([A-Z]+)_segname[A-
Z]+.dat$', data_file)
    if match:
        amino_acid1 = match.group(1).replace('GLU',
'E').replace('ARG', 'R').replace('ASP', 'D').replace('LYS', 'K')
        number1 = match.group(2)
        chain1 = match.group(3)
        amino_acid2 = match.group(4).replace('GLU',
'E').replace('ARG', 'R').replace('ASP', 'D').replace('LYS', 'K')
        number2 = match.group(5)
        chain2 = match.group(6)

        label = f"{amino_acid1[0]}{number1}-
{amino_acid2[0]}{number2}"
        labels.append(label)

        color = next(color_iterator)
```

```
        ax1.plot([x for x, _ in filtered_data], [y for _, y in
filtered_data], '|', markersize=markersize, color=color,
label=label)
    else:
        print(f"Skipping invalid filename: {data_file}")

ax1.set_xlabel('Time (ns)', fontsize=20, fontweight='bold')
ax1.set_ylabel('Salt bridge', color='black', fontsize=20,
fontweight='bold')
ax1.set_title('ROS-1', fontsize=20, fontweight='bold')
yticks = np.arange(markersize / 2, len(data_files) * markersize,
markersize)
ax1.set_yticks(yticks)
ax1.set_yticklabels(labels, fontsize=18)
ax1.set_ylim(min(all_y_values) - markersize / 2, max(all_y_values)
+ markersize / 2)
for ytick in yticks:
    ax1.axhline(y=ytick + markersize / 2, color='black',
linestyle='-', linewidth=0.5, alpha=0.5)

x_max = max(all_x_values)
ax1.set_xlim(0, x_max)
xticks = np.arange(0, 1101, 100)
ax1.set_xticks(xticks)
ax1.set_xticklabels(xticks, fontsize=18)
yticks_right = yticks
yticklabels_right = [f'{duration:.2f}' for duration in durations]
ax2.set_yticks(yticks_right)
ax2.set_yticklabels(yticklabels_right, fontsize=18)
ax2.set_ylabel('Duration (ns)', color='black', fontsize=20,
fontweight='bold')
ax2.set_ylim(0, len(data_files) * markersize)
plt.tight_layout()
plt.show()
```

HydrogenBonds_pie.py

```
import matplotlib.pyplot as plt

data_file = "D20_D21_HBONDShbonds-details_sorted.dat"
labels = []
occupancies = []

def custom_autopct(pct):
    total = sum(occupancies)
    occupancy = round(pct * total / 100.0, 2)
    return f'{occupancy:.2f}%'

with open(data_file, 'r') as file:
    next(file)  #skip the header
    for line in file:
        line = line.strip().split('\t')
        donor = line[0]
        acceptor = line[1]
        occupancy = float(line[2].rstrip('%'))  #remove % and
convert to float
        if occupancy > 3.0:
```

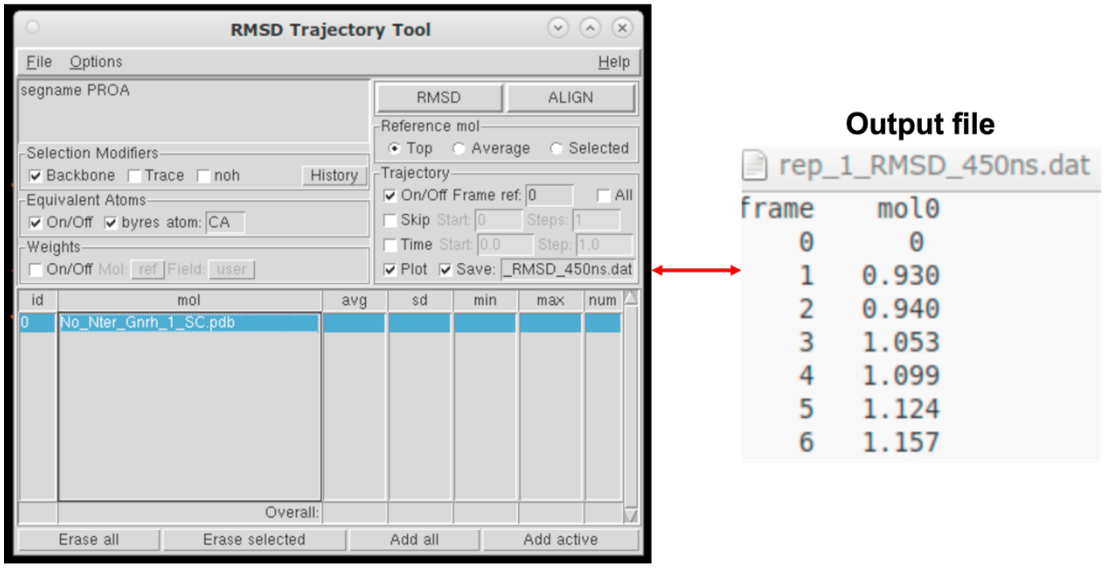
```
            label = f"{donor}-{acceptor}"
            labels.append(label)
            occupancies.append(occupancy)

fig, ax = plt.subplots(figsize=(8, 8))
colors = plt.cm.tab20c(range(len(occupancies)))
patches, texts, autotexts = ax.pie(occupancies, labels=labels,
startangle=140, colors=colors, autopct=custom_autopct)

for text, color in zip(texts, colors):
    text.set_color(color)
    text.set_fontsize(16)
    text.set_fontweight('bold')
ax.set_title('Hydrogen Bond Occupancy (%)', pad=20, fontsize=18)
plt.axis('equal')
plt.show()
```

RMSD.py

For RMSD data extraction, the VMD plugin 'RMSD Trajectory Tool' was utilised as follows:



The above process calculates the RMSD of the whole GnRH1R sequence (segname PROA selection). To retrieve RMSD values for each TM7, the selections:

- segname PROA and resid 33 to 66 (for TM1)
- segname PROA and resid 74 to 104 (for TM2)
- segname PROA and resid 110 to 145 (for TM3)
- segname PROA and resid 153 to 178 (for TM4)
- segname PROA and resid 204 to 244 (for TM5)
- segname PROA and resid 257 to 293 (for TM6)
- segname PROA and resid 301 to 328 (for TM7)
- segname PROA and resid 1 to 17 (for N-ter)

– segname PROA and resid 18 to 32 (for N-ter)
–

The resulting RMSD data file is plotted using the customised python script for each system:

```
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib.ticker import MultipleLocator

file_names = [
    'rmsd_1to18_D1_D22.dat', 'rmsd_18to33_D1_D22.dat',
'TM1_rmsd.dat',
    'TM2_rmsd.dat', 'TM3_rmsd.dat', 'TM4_rmsd.dat',
    'TM5_rmsd.dat', 'TM6_rmsd.dat', 'TM7_rmsd.dat'
]

plt.figure(figsize=(10, 6))
colors = ['red', 'blue']
custom_colors = ['green', 'purple', 'orange', 'cyan', 'brown',
'magenta', 'teal']

max_frame = 0

for i, file_name in enumerate(file_names):
    df = pd.read_csv(file_name, delim_whitespace=True)
    df['frame'] /= 50
    if i == 0:
        label = '1-17 N-ter'
    elif i == 1:
        label = '18-33 N-ter'
    else:
        label = file_name.split('_')[0]
    if i < len(colors):
        plt.plot(df['frame'], df['mol0'], label=label,
linewidth=1, color=colors[i])
    else:
        plt.plot(df['frame'], df['mol0'], label=label,
linewidth=1, color=custom_colors[i-len(colors)])
    max_frame = max(max_frame, max(df['frame']))

plt.xlim([0, 1100])
plt.ylim([0.1, 15])
plt.xlabel('Time (ns)', fontsize=14, fontweight='bold')
plt.ylabel('RMSD (Å)', fontsize=14, fontweight='bold')
plt.title('ROS-1 #2', fontsize=16, fontweight='bold')
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.gca().xaxis.set_major_locator(MultipleLocator(100))
plt.legend(loc='upper right')
plt.grid(False)
plt.show()
```

RMSF.py

```
import MDAnalysis as mda
```

```
from MDAnalysis.analysis import rms, align
import matplotlib.pyplot as plt
import warnings
import numpy as np

warnings.filterwarnings('ignore')

adk_topology = 'ROS_1_SC.pdb'
adk_trajectory = 'D13.dcd'

u = mda.Universe(adk_topology, adk_trajectory)
protein_proa = u.select_atoms('protein and name CA and (resid
1:328)')
average = align.AverageStructure(u, select='protein and name CA
and (resid 1:328)', ref_frame=0).run()
ref = average.universe
aligner = align.AlignTraj(u, ref, select='protein and name CA and
(resid 1:328)', in_memory=True).run()
c_alphas = protein_proa
R = rms.RMSF(c_alphas).run()
data = np.column_stack((c_alphas.resids, R.rmsf))
np.savetxt('rmsf_data.txt', data, header='Residue_number RMSF',
fmt='%d %.4f') #save data only

plt.plot(c_alphas.resids, R.rmsf)
plt.xlabel('Residue number', fontsize=14, weight='bold')
plt.ylabel('RMSF ($\AA$)', fontsize=14, weight='bold')
plt.yticks(fontsize=14)
plt.xticks(fontsize=14)
plt.xlim(12, 328)
plt.xticks([12, 33, 67, 74, 93, 110, 145, 152, 177, 204, 244, 257,
293, 302, 328], rotation=90)
midpoints = [(33 + 67) / 2, (74 + 93) / 2, (110 + 145) / 2, (152 +
177) / 2, (204 + 244) / 2, (257 + 293) / 2, (302 + 327) / 2]

for midpoint, label in zip(midpoints, ['TM1', 'TM2', 'TM3', 'TM4',
'TM5', 'TM6', 'TM7']):
    plt.text(midpoint, 4.5, label, ha='center', va='center',
color='black', fontsize=14, weight='bold')

plt.axvspan(33, 67, zorder=0, alpha=0.2, color='grey',
label='TM1')
plt.axvspan(74, 93, zorder=0, alpha=0.2, color='grey',
label='TM2')
plt.axvspan(110, 145, zorder=0, alpha=0.2, color='grey',
label='TM3')
plt.axvspan(152, 177, zorder=0, alpha=0.2, color='grey',
label='TM4')
plt.axvspan(204, 244, zorder=0, alpha=0.2, color='grey',
label='TM5')
plt.axvspan(257, 293, zorder=0, alpha=0.2, color='grey',
label='TM6')
plt.axvspan(302, 327, zorder=0, alpha=0.2, color='grey',
label='TM7')
plt.show()
```

Plot_RMSF.py

```
import matplotlib.pyplot as plt

fig = plt.figure(figsize=(10, 10))

data_files = [
    {"file": "rmsf_inactive1_50ns.txt", "label": "Inactive
GnRH1R", "color": "black"},
    {"file": "rmsf_1ROS2.txt", "label": "ROS-2 (1)", "color":
"red"},
    {"file": "rmsf_2ROS2.txt", "label": "ROS-2 (2)", "color":
"green"},
]
legend_handles = []
for data_info in data_files:
    with open(data_info["file"], "r") as file:
        lines = file.readlines()
    x_values = []
    y_values = []
    for line in lines:
        parts = line.split()
        if len(parts) == 2:
            x_values.append(int(parts[0]))
            y_values.append(float(parts[1]))
    plt.plot(x_values, y_values, label=data_info["label"],
color=data_info["color"])
    legend_handles.append(plt.Line2D([0], [0],
color=data_info["color"], label=data_info["label"]))
plt.xlabel('Residue number', fontsize=14, weight='bold')
plt.ylabel('RMSF ($\AA$)', fontsize=14, weight='bold')
plt.yticks(fontsize=14)
plt.xticks(fontsize=14)
plt.xlim(1, 328)
plt.xticks([33, 66, 74, 104, 110, 145, 153, 178, 204, 244, 257,
293, 301, 327], rotation=90)
plt.ylim(0, 10)
midpoints = [(33 + 66) / 2, (74 + 104) / 2, (110 + 145) / 2, (153
+ 178) / 2, (204 + 244) / 2, (257 + 293) / 2, (301 + 327) / 2]
top_y_coordinate = 8.0
for midpoint, label in zip(midpoints, ['TM1', 'TM2', 'TM3', 'TM4',
'TM5', 'TM6', 'TM7']):
    plt.text(midpoint, top_y_coordinate, label, ha='center',
va='bottom', color='black', fontsize=14, weight='bold')

plt.axvspan(33, 66, zorder=0, alpha=0.2, color='grey',
label='TM1')
plt.axvspan(74, 104, zorder=0, alpha=0.2, color='grey',
label='TM2')
plt.axvspan(110, 145, zorder=0, alpha=0.2, color='grey',
label='TM3')
plt.axvspan(153, 178, zorder=0, alpha=0.2, color='grey',
label='TM4')
plt.axvspan(204, 244, zorder=0, alpha=0.2, color='grey',
label='TM5')
plt.axvspan(257, 293, zorder=0, alpha=0.2, color='grey',
label='TM6')
```

```
plt.axvspan(301, 327, zorder=0, alpha=0.2, color='grey',
label='TM7')
plt.legend(handles=legend_handles, fontsize=12)
plt.show()
```

Ramachadran.py

```
import MDAnalysis as mda
from MDAnalysis.analysis.dihedrals import Ramachandran
import matplotlib.pyplot as plt

psf_file = 'ROS_1_SC.psf'
dcd_file = 'D22.dcd'
u = mda.Universe(psf_file, dcd_file)
selection = u.select_atoms("segid PROA")
R = Ramachandran(selection).run()
fig, ax = plt.subplots(figsize=(10, 8))
R.plot(ax=ax, color='k', marker='o', alpha=0.3, ref=True)
ax.set_xlabel('Phi (°)', fontsize=24, weight='bold')
ax.set_ylabel('Psi (°)', fontsize=24, weight='bold')
ax.set_title('GnRH1R in ROS-1', fontsize=22, weight='bold')
ax.tick_params(axis='both', which='major', labelsize=22)
plt.grid(True)
plt.tight_layout()
plt.show()
```

Tm3_6_7_distance.py

```
import MDAnalysis as mda
import numpy as np
import matplotlib.pyplot as plt

dcd_files = ['D18.dcd', 'D19.dcd', 'D20.dcd', 'D21.dcd',
'D22.dcd'] #D1.dcd to D22.dcd was used = 1.1 µs in total

psf_file = 'ROS_1_SC.psf'
u = mda.Universe(psf_file)

#define selections for CA atoms of residues 139 (TM3), 265 (TM6),
and 323 (TM7) in segname PROA
tm3_selection = u.select_atoms("segid PROA and resid 139 and name
CA")
tm6_selection = u.select_atoms("segid PROA and resid 265 and name
CA")
tm7_selection = u.select_atoms("segid PROA and resid 323 and name
CA")

tm3_tm6_distances = []
tm3_tm7_distances = []
time_values = []

previous_time = 0  #initialise time from previous trajectory
for dcd_file in dcd_files:
    u.load_new(dcd_file)

    for ts in u.trajectory:
        #calculate distances between CA atoms of TM3 and TM6
```

```
        distance_tm3_tm6 = np.linalg.norm(tm3_selection.positions
- tm6_selection.positions)
        tm3_tm6_distances.append(distance_tm3_tm6)
        time_values.append(ts.time + previous_time)
    previous_time = time_values[-1]
time_values = np.array(time_values)
tm3_tm6_distances = np.array(tm3_tm6_distances)

output_file = "tm3_tm6_distances.txt"
np.savetxt(output_file, np.column_stack((time_values,
tm3_tm6_distances)), fmt='%10.5f', header='Time (ps)\tTM3-TM6
Distance (angstrom)', comments='')

# Plot
plt.plot(time_values, tm3_tm6_distances, color='b', linestyle='-')
plt.axhline(y=8, color='k', linestyle='--')
plt.xlabel('Time (ps)')
plt.ylabel('TM3-TM6 Distance (angstrom)')
plt.title('TM3-TM6 Distance over Time')
plt.grid(True)
plt.show()
```

tm3_tm6_all_plot.py

```
import pandas as pd
import matplotlib.pyplot as plt

csv_file_path = '/path/to/file/distances.csv'
df = pd.read_csv(csv_file_path, skiprows=2)

plt.figure(figsize=(40, 20))

# ROS-1
plt.subplot(2, 2, 1)
line1, = plt.plot(df.iloc[:, 0], df.iloc[:, 3], label='ROS-1 #3',
color='cadetblue')  # Plot column 0:1 first
line2, = plt.plot(df.iloc[:, 0], df.iloc[:, 2], label='ROS-1 #2',
color='darkblue')  # Plot column 2 next
line3, = plt.plot(df.iloc[:, 0], df.iloc[:, 1], label='ROS-1 #1',
color='deepskyblue')  # Plot column 3 last
plt.axhline(y=7.9, color='black', linestyle='--', linewidth=6)  #
Line at y=8 is inactive crystal
plt.axhline(y=11, color='green', linestyle='--', linewidth=6)  #
Line at y=11 is active AF
plt.xlim(0, 1100)
plt.ylim(6, 15)
plt.xticks(range(0, 1101, 100), fontsize=36)
plt.yticks(range(7, 15), fontsize=30)
plt.tick_params(axis='both', which='major', labelsize=36)
plt.axvline(x=1000, color='black', linestyle='-', linewidth=2)
plt.axvline(x=850, color='black', linestyle='-', linewidth=2)
plt.xlabel('Time (ns)', fontsize=40, fontweight='bold')
plt.ylabel('$\mathbf{R139^{3.50}-T265^{6.33}}(Å)$', fontsize=40,
fontweight='bold')
plt.legend(handles=[line3, line2, line1], labels=['ROS-1 #1',
'ROS-1 #2', 'ROS-1 #3'], loc='upper left', fontsize=36)
```

```
# ROS-2
plt.subplot(2, 2, 2)
plt.plot(df.iloc[:, 0], df.iloc[:, 4], label='ROS-2 #1',
color='g')
plt.plot(df.iloc[:, 0], df.iloc[:, 5], label='ROS-2 #2',
color='limegreen')
plt.axhline(y=7.9, color='black', linestyle='--', linewidth=6)  #
Line at y=8 is inactive crystal
plt.axhline(y=11, color='green', linestyle='--', linewidth=6)  #
Line at y=11 is active AF
plt.xlim(0, 1100)
plt.ylim(6, 15)
plt.xticks(range(0, 1101, 100), fontsize=36)
plt.yticks(range(7, 15), fontsize=30)
plt.tick_params(axis='both', which='major', labelsize=36)
plt.xlabel('Time (ns)', fontsize=40, fontweight='bold')
plt.ylabel('$\mathbf{R139^{3.50}-T265^{6.33}}(Å)$', fontsize=40,
fontweight='bold')
plt.legend(loc='upper left', fontsize=36)

# GnRH-Gnrh1r
plt.subplot(2, 2, 3)
plt.plot(df.iloc[:, 0], df.iloc[:, 6], label='Undocked GnRH-GnRH1R
#1', color='darkslateblue')
plt.plot(df.iloc[:, 0], df.iloc[:, 7], label='Undocked GnRH-GnRH1R
#2', color='darkviolet')
plt.axhline(y=7.9, color='black', linestyle='--', linewidth=6)  #
Line at y=8 is inactive crystal
plt.axhline(y=11, color='green', linestyle='--', linewidth=6)  #
Line at y=11 is active AF
plt.xlim(0, 1100)
plt.ylim(6, 15)
plt.xticks(range(0, 1101, 100), fontsize=36)
plt.yticks(range(7, 15), fontsize=36)
plt.tick_params(axis='both', which='major', labelsize=36)
plt.axvline(x=0, color='black', linestyle='-', linewidth=2)
plt.xlabel('Time (ns)', fontsize=40, fontweight='bold')
plt.ylabel('$\mathbf{R139^{3.50}-T265^{6.33}}(Å)$', fontsize=40,
fontweight='bold')
plt.legend(loc='upper left', fontsize=36)

# free
plt.subplot(2, 2, 4)
plt.plot(df.iloc[:, 0], df.iloc[:, 8], label='Apo-GnRH1R #1',
color='k')
plt.plot(df.iloc[:, 0], df.iloc[:, 9], label='Apo-GnRH1R #2',
color='grey')
plt.axhline(y=7.9, color='black', linestyle='--', linewidth=6)  #
Line at y=8 is inactive crystal
plt.axhline(y=11, color='green', linestyle='--', linewidth=6)  #
Line at y=11 is active AF
plt.xlim(0, 1100)
plt.ylim(6, 15)
plt.xticks(range(0, 1101, 100), fontsize=36)
plt.yticks(range(7, 15), fontsize=36)
```

```
plt.tick_params(axis='both', which='major', labelsize=36)
plt.xlabel('Time (ns)', fontsize=40, fontweight='bold')
plt.ylabel('$\mathbf{R139^{3.50}-T265^{6.33}}(Å)$', fontsize=40,
fontweight='bold')
plt.legend(loc='upper left', fontsize=36)

plt.tight_layout()
plt.subplots_adjust(hspace=0.2)

plt.savefig('/path/to/file/tms_multisubplots.pdf')
plt.show()
```

B_factor_rmsf.py

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import pearsonr

rmsf_file_path = '/path/to/file/rmsf_inactive1_50ns.txt' #md rmsf
of inactive
rmsf_data = np.loadtxt(rmsf_file_path, usecols=(0, 1))
pdb_file_path = '/path/to/file/7br3_clean.pdb' #b-factor pdb

#extract B-factor values from PDB for each residue
def extract_b_factors(pdb_file):
    b_factors = {}
    with open(pdb_file, 'r') as file:
        for line in file:
            if line.startswith('ATOM'):
                residue_id = int(line[22:26].strip())
                b_factor = float(line[60:66].strip())
                if residue_id not in b_factors:
                    b_factors[residue_id] = []
                b_factors[residue_id].append(b_factor)
    return b_factors
b_factors_per_residue = extract_b_factors(pdb_file_path)

#align RMSF data with residue numbers from the PDB file
aligned_rmsf_data = np.zeros((len(b_factors_per_residue), 2))
for i, residue_id in enumerate(b_factors_per_residue.keys()):
    aligned_rmsf_data[i, 0] = residue_id
    aligned_rmsf_data[i, 1] = rmsf_data[rmsf_data[:, 0] ==
residue_id, 1]

#normalize both RMSF and B-factor values
normalized_rmsf = (aligned_rmsf_data[:, 1] -
np.mean(aligned_rmsf_data[:, 1])) / np.std(aligned_rmsf_data[:,
1])
normalized_b_factors = np.array([np.mean(b_factors) for b_factors
in b_factors_per_residue.values()])
normalized_b_factors = (normalized_b_factors -
np.mean(normalized_b_factors)) / np.std(normalized_b_factors)

#plot
plt.figure(figsize=(12, 8))
plt.plot(aligned_rmsf_data[:, 0], normalized_b_factors, label='B-
Factors Inactive Crystal')
```

```
plt.plot(aligned_rmsf_data[:, 0], normalized_rmsf, label='RMSF
Apo-GnRH1R')
plt.xlabel('Residue', fontsize=22, fontweight='bold', labelpad=10)
plt.ylabel('Normalised Value', fontsize=22, fontweight='bold')
plt.xlim(18, 328)

#quantitative comparison (Pearson correlation coefficient)
correlation_coefficient, p_value = pearsonr(normalized_b_factors,
normalized_rmsf)
correlation_coefficient_rounded = round(correlation_coefficient,
2)
p_value_scientific = f"{p_value:.2e}"

legend_text = f'Pearson: {correlation_coefficient_rounded}\np-
value: {p_value_scientific}'
plt.legend(title=legend_text, loc='upper right', fontsize=16,
title_fontsize=16)
plt.xticks([18, 33, 66, 74, 104, 110, 145, 153, 178, 204, 244,
257, 293, 301, 328], rotation=90, fontsize=20)
midpoints = [(33 + 66) / 2, (74 + 104) / 2, (110 + 145) / 2, (153
+ 178) / 2,
             (204 + 244) / 2,(257 + 293) / 2,(301 +328) /2]

for midpoint,label in
zip(midpoints,['TM1','TM2','TM3','TM4','TM5','TM6','TM7']):
    plt.text(midpoint,max(normalized_b_factors),label,
             ha='center',va='center',color='black',fontsize=18)
plt.yticks(fontsize=18)
plt.axvspan(33 ,67,zorder=0,alpha=0.2,color='grey',label='TM1')
plt.axvspan(74 ,93,zorder=0,alpha=0.2,color='grey',label='TM2')
plt.axvspan(110 ,145,zorder=0,alpha=0.2,color='grey',label='TM3')
plt.axvspan(152 ,177,zorder=0,alpha=0.2,color='grey',label='TM4')
plt.axvspan(204 ,244,zorder=0,alpha=0.2,color='grey',label='TM5')
plt.axvspan(257 ,293,zorder=0,alpha=0.2,color='grey',label='TM6')
plt.axvspan(302 ,327,zorder=0,alpha=0.2,color='grey',label='TM7')
plt.tight_layout()
plt.show()
```

DSSP.py

```
import MDAnalysis as mda
import mdtraj as md
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

u = mda.Universe("NTER_FREE_SC.psf", "D26.dcd")
protein_PROA = u.select_atoms("segid PROA")
traj = md.load("D26.dcd", top="NTER_FREE_SC.psf", stride=500)
protein_indices = protein_PROA.resids - 1
dssp = md.compute_dssp(traj, simplified=False)
protein_dssp = dssp[:, protein_indices]
structure_types, structure_counts = np.unique(protein_dssp,
return_counts=True)
total_residues = len(protein_indices) * protein_dssp.shape[0]
structure_percentages = {s: np.sum(protein_dssp == s) /
total_residues * 100 for s in structure_types}
```

```
colors = sns.color_palette("Set3", n_colors=len(structure_types))
plt.rcParams.update({'axes.labelsize': 16, 'xtick.labelsize': 14,
'ytick.labelsize': 14, 'legend.fontsize': 10})
plt.figure(figsize=(10, 6))
bars = plt.bar(structure_percentages.keys(),
structure_percentages.values(), color=colors)
plt.xlabel('DSSP term', weight='bold')
plt.ylabel('Percentage (%)', weight='bold')
plt.title('Apo-GnRH1R', weight='bold', fontsize=16)

legend_labels = {'H': 'Alpha helix',
                 'B': 'Beta bridge',
                 'E': 'Extended strand',
                 'G': '3-helix (3/10 helix)',
                 'I': '5 helix (pi helix)',
                 'T': 'Hydrogen bonded turn',
                 'S': 'Bend',
                 ' ': 'Loops and irregular elements'}

legend_handles = [plt.Line2D([0], [0], marker='o', color='w',
markerfacecolor=colors[i], markersize=10) for i in
range(len(structure_types))]
legend_texts = [f'{key}: {legend_labels[key]}' for key in
structure_types]
plt.legend(legend_handles, legend_texts, title='DSSP Assignments',
loc='upper right')
for bar in bars:
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width() / 2, yval, round(yval,
2), ha='center', va='bottom', fontsize=12)
plt.show()
```

PCA.py

```
import mdtraj as md
import numpy as np
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from scipy.stats import gaussian_kde
from mpl_toolkits.axes_grid1 import make_axes_locatable
from collections import Counter
from sklearn.cluster import DBSCAN

traj_files = ['D10.dcd', 'D11.dcd', 'D12.dcd', 'D13.dcd',
'D14.dcd', 'D15.dcd', 'D16.dcd', 'D17.dcd', 'D18.dcd', 'D19.dcd',
'D20.dcd', 'D21.dcd', 'D22.dcd']
traj = md.load(traj_files, top='ROS_1_SC.pdb')

#specify residue range for the peptide or N-ter
peptide_residues = range(1, 17) #1,17 and seganame PROA was used
for PCA of receptor N-terminus (D10-D22 was t>500ns. segname PROB
and resid 1, 10 was used for GnRH PCA, D20-D22 was activation
trajectories.
selection_string = f'segname PROA and (residue
{peptide_residues[0]} to {peptide_residues[-1]})'
```

```
peptide_traj =
traj.atom_slice(traj.topology.select(selection_string))

#perform PCA
peptide_coords = peptide_traj.xyz.reshape(peptide_traj.n_frames, -
1)
pca = PCA(n_components=2)
peptide_pcs = pca.fit_transform(peptide_coords)

#create a 2D KDE
x = peptide_pcs[:, 0]
y = peptide_pcs[:, 1]
kde = gaussian_kde(np.vstack([x, y]))
x_grid, y_grid = np.mgrid[x.min():x.max():100j,
y.min():y.max():100j]
z = kde(np.vstack([x_grid.ravel(), y_grid.ravel()]))

#################### comment for gpcr residues //// only uncomment
for segname PROB selection (peptide) ####################
#perform DBSCAN clustering
#dbscan = DBSCAN(eps=0.1, min_samples=3)
#cluster_labels = dbscan.fit_predict(peptide_pcs)

#count the number of frames in each cluster
#cluster_counter = Counter(cluster_labels)

#sort clusters based on population
#sorted_clusters = sorted(cluster_counter.items(), key=lambda x:
x[1], reverse=True)

#print frame ranges of the three most populated clusters
#for i in range(3):
   # label, count = sorted_clusters[i]
   # frames = np.where(cluster_labels == label)[0]
   # print(f'Cluster {label}: Frame range [{frames.min()},
{frames.max()}], Number of frames: {count}')
#####################################################################
#####################################################
#plot
fig, axs = plt.subplots(1, 2, figsize=(12, 10),
gridspec_kw={'width_ratios': [6, 2]})  # Adjusted figsize and
width_ratios
#density Plot
im = axs[0].pcolormesh(x_grid, y_grid, z.reshape(x_grid.shape),
shading='auto', cmap='viridis')
axs[0].set_xlabel('PCA 1', fontsize=30, fontweight='bold')
axs[0].set_ylabel('PCA 2', fontsize=30, fontweight='bold')
axs[0].set_title('ROS-1 N-ter 1 to 17', fontsize=28,
fontweight='bold')
axs[0].tick_params(axis='both', which='both', length=8,
labelsize=22)
divider = make_axes_locatable(axs[0])
cax = divider.append_axes("bottom", size="5%", pad=1)
cbar = plt.colorbar(im, cax=cax, orientation='horizontal')
cbar_ticks = [z.min(), z.max()]
cbar.set_ticks(cbar_ticks)
cbar.set_ticklabels(['Low', 'High'])
```

```
cbar.ax.set_xlabel('Population Density', fontsize=24,
fontweight='bold')
cbar.ax.tick_params(axis='both', which='both', labelsize=22)

#scree plot (Variance Explained)
variance_explained = pca.explained_variance_ratio_
bars = axs[1].bar(range(1, len(variance_explained) + 1),
variance_explained, align='center', color=['forestgreen', 'blue'])
axs[1].set_xlabel('PCA', fontsize=30, fontweight='bold')
axs[1].set_ylabel('Variance Explained', fontsize=30,
fontweight='bold')
axs[1].set_xticks([1, 2])
axs[1].set_yticks([])
axs[1].tick_params(axis='both', which='both', length=8,
labelsize=22)

for i, bar in enumerate(bars):
    height = bar.get_height()
    axs[1].text(bar.get_x() + bar.get_width()/2., height,
'{:.1%}'.format(variance_explained[i]), ha='center', va='bottom',
fontsize=22)
plt.tight_layout()
plt.show()
```

R240_hbonds.py

```
import matplotlib.pyplot as plt
import numpy as np
from matplotlib.ticker import FormatStrFormatter, MultipleLocator

data = np.loadtxt('R240_hbonds.dat')  # Replace 'R240_hbonds.dat'
with the actual filename
time = (data[:, 0] / 50.0 + 900.0) / 1000  #divide the time column
by 50 and convert to ns
num_hydrogen_bonds = data[:, 1]

#plot
plt.figure(figsize=(10, 6))
plt.plot(time, num_hydrogen_bonds, linestyle='-', markersize=3,
color='dodgerblue')
plt.axhline(y=1.01, color='black', linestyle='-', linewidth=1)  #
Black line at y=1
plt.xlabel('Time (μs)', fontsize=36, fontweight='bold')
plt.ylabel('Hydrogen Bonds', fontsize=36, fontweight='bold')
plt.title('R240-POPC243 Hydrogen Bonds', fontsize=36)
plt.grid(True)
plt.xlim(0.9, 1.1)
plt.ylim(0.1, 2)
plt.yticks([1, 2], fontsize=30)
plt.xticks(fontsize=30)
plt.gca().xaxis.set_major_locator(MultipleLocator(0.05))  # Set x-
axis ticks every 0.05
plt.gca().xaxis.set_major_formatter(FormatStrFormatter('%.2f'))
plt.tight_layout()
plt.savefig('hbonds_240_POPC.pdf')
plt.show()
```

F272_POPC_find_distance.py

```
import MDAnalysis as mda
import numpy as np
import matplotlib.pyplot as plt

#intermediate to active
trajectory_files = ["D19.dcd", "D20.dcd", "D21.dcd", "D22.dcd"]
protein_residue = "resname PHE and resid 276"
POPC_residue = "resname POPC and resid 243 and name C218"
distances = []
for traj_file in trajectory_files:
    u = mda.Universe("ROS_1_SC.pdb", traj_file)
    protein_atoms = u.select_atoms(protein_residue)
    POPC_atoms = u.select_atoms(POPC_residue)
    for ts in u.trajectory:
        distance = np.linalg.norm(protein_atoms.center_of_mass() -
POPC_atoms.positions)
        distances.append(distance)
time_axis = np.arange(0, len(distances)) / 50.0 #convert to ns

#plot
plt.figure(figsize=(10, 6))
plt.plot(time_axis, distances, label='Distance Protein Residue 276
- POPC Residue 243 C212', linestyle='-', linewidth=1,
color='deepskyblue')
plt.xlabel('Time (ns)', fontsize=16, fontweight='bold')
plt.ylabel('Distance (Å)', fontsize=16, fontweight='bold')
plt.title('F276 CoM and Lipid aryl group distance', fontsize=16)
plt.xlim(0, 100)
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)
plt.grid(True)
np.savetxt('distance_data.txt', np.column_stack((time_axis,
distances)), header='Time (ns) Distance (Å)', fmt='%1.3f %1.3f')
plt.show()
```

Plot_F272_POPC_distance.py

```
import matplotlib.pyplot as plt
import numpy as np
from matplotlib.ticker import FormatStrFormatter, MultipleLocator

data = np.loadtxt('distance_F272_lipid_2ROS1.txt')
time_data = (data[:, 0] + 900) / 1000
distance_data = data[:, 1]

#plot
plt.figure(figsize=(10, 6))
plt.plot(time_data, distance_data, linestyle='-',
color='blueviolet', label='Distance Data')
plt.xlabel('Time (µs)', fontsize=36, fontweight='bold')
plt.ylabel('Distance (Å)', fontsize=36, fontweight='bold')
plt.title('F276-POPC243 Distance', fontsize=36)
plt.xticks(fontsize=30)
plt.yticks(fontsize=30)
plt.gca().xaxis.set_major_formatter(FormatStrFormatter('%.2f'))
plt.gca().xaxis.set_major_locator(MultipleLocator(0.05))
```

```
plt.xlim(0.9, 1.1)
plt.ylim(0.1, 30)
plt.grid(True)
plt.tight_layout()
plt.savefig('distance_F276_POPC.pdf')
plt.show()
```

Lipid_GnRH1R_hbonds.py

```
import os
import matplotlib.pyplot as plt
import numpy as np

output_file_path = '/path/to/directory/sorted_results.txt'
with open(output_file_path, 'r') as output_file:
    lines = output_file.readlines()
x_labels = [line.split('_')[0] for line in lines]
y_values = [float(line.split(': ')[1].strip()) for line in lines]
filtered_indices = [i for i, value in enumerate(y_values) if value
> 10 and not x_labels[i].startswith('POPC')]
filtered_x_labels = [x_labels[i] for i in filtered_indices]
filtered_y_values = [y_values[i] for i in filtered_indices]
colors = plt.cm.viridis(np.linspace(0, 1, len(filtered_x_labels)))

fig, ax = plt.subplots(figsize=(18, 12))
bars = ax.bar(filtered_x_labels, filtered_y_values, color=colors)
plt.xlabel('Residue', fontsize=32, fontweight='bold')
plt.ylabel('Occupancy (%)', fontsize=28, fontweight='bold')
plt.title('GnRH1R-Membrane Hydrogen bonds', fontsize=32,
fontweight='bold')
plt.xticks(rotation=65, ha='center', fontsize=26)
plt.yticks(fontsize=28)

for bar, value, label in zip(bars, filtered_y_values,
filtered_x_labels):
    height = bar.get_height()
    width = bar.get_width()
    x = bar.get_x() + width / 2

    if label in ['R179', 'S55', 'E111']:
        ax.text(x, height + 10, f'{value:.2f}%', ha='center',
va='center',
                fontsize=18, color='black',
bbox=dict(facecolor='white', edgecolor='black',
boxstyle='round,pad=0.3'),
                rotation=90)
    else:
        ax.text(x, height / 2, f'{value:.2f}%', ha='center',
va='center',
                fontsize=18, color='black',
bbox=dict(facecolor='white', edgecolor='black',
boxstyle='round,pad=0.3'),
                rotation=90)

ax.tick_params(axis='both', which='both', direction='out',
length=10, width=2)
plt.savefig('/path/to/file/memb_hbonds.pdf', bbox_inches='tight')
```

```
plt.show()
import os
import matplotlib.pyplot as plt
import numpy as np

output_file_path = '/path/to/file/sorted_results.txt'
with open(output_file_path, 'r') as output_file:
    lines = output_file.readlines()
x_labels = [line.split('_')[0] for line in lines]
y_values = [float(line.split(': ')[1].strip()) for line in lines]
filtered_indices = [i for i, value in enumerate(y_values) if value
> 10 and not x_labels[i].startswith('POPC')]
filtered_x_labels = [x_labels[i] for i in filtered_indices]
filtered_y_values = [y_values[i] for i in filtered_indices]
colors = plt.cm.viridis(np.linspace(0, 1, len(filtered_x_labels)))

fig, ax = plt.subplots(figsize=(18, 12))
bars = ax.bar(filtered_x_labels, filtered_y_values, color=colors)
plt.xlabel('Residue', fontsize=32, fontweight='bold')
plt.ylabel('Occupancy (%)', fontsize=28, fontweight='bold')
plt.title('GnRH1R-Membrane Hydrogen bonds', fontsize=32,
fontweight='bold')
plt.xticks(rotation=65, ha='center', fontsize=26)
plt.yticks(fontsize=28)

for bar, value, label in zip(bars, filtered_y_values,
filtered_x_labels):
    height = bar.get_height()
    width = bar.get_width()
    x = bar.get_x() + width / 2

    if label in ['R179', 'S55', 'E111']:
        ax.text(x, height + 10, f'{value:.2f}%', ha='center',
va='center',
                fontsize=18, color='black',
bbox=dict(facecolor='white', edgecolor='black',
boxstyle='round,pad=0.3'),
                rotation=90)
    else:
        ax.text(x, height / 2, f'{value:.2f}%', ha='center',
va='center',
                fontsize=18, color='black',
bbox=dict(facecolor='white', edgecolor='black',
boxstyle='round,pad=0.3'),
                rotation=90)

ax.tick_params(axis='both', which='both', direction='out',
length=10, width=2)
plt.savefig('/path/to/file/memb_hbonds.pdf', bbox_inches='tight')
plt.show()
```

Ring_interaction_map.py

```
import numpy as np
import matplotlib.pyplot as plt

#example for pipistack interactions
```

```
data = []
with open("pipistack.txt", "r") as file:
    for line in file:
        parts = line.strip().split("\t")
        residue1 = int(parts[0].split(":")[1])
        residue2 = int(parts[2].split(":")[1])
        distance = float(parts[3])
        data.append((residue1, residue2, distance))

residues = sorted(set(residue for pair in data for residue in
pair[:2]))
contact_map = np.zeros((len(residues), len(residues)))

for pair in data:
    i = residues.index(pair[0])
    j = residues.index(pair[1])
    contact_map[i, j] = pair[2]
    contact_map[j, i] = pair[2]
plt.figure(figsize=(10, 8))
img = plt.imshow(contact_map, cmap='viridis',
interpolation='nearest', vmin=0, vmax=np.max(contact_map),
extent=[0, len(residues), 0, len(residues)], origin='lower',
aspect='auto')
plt.colorbar()
tick_positions = np.arange(len(residues)) + 0.5
plt.xticks(ticks=tick_positions, labels=residues,
rotation='vertical', fontsize=12)
plt.yticks(ticks=tick_positions, labels=residues, fontsize=12)
magenta_ticks = [2, 3, 5]
for tick in magenta_ticks:
    plt.xticks()[1][residues.index(tick)].set_color('magenta')
    plt.yticks()[1][residues.index(tick)].set_color('magenta')
plt.axhline(y=np.where(np.array(residues) == 5)[0][0] + 1,
color='magenta', linestyle='--', linewidth=2)
plt.axvline(x=np.where(np.array(residues) == 5)[0][0] + 1,
color='magenta', linestyle='--', linewidth=2)

for i in range(len(residues) + 1):
    plt.axhline(y=i, color='black', linestyle='-', linewidth=0.5)
    plt.axvline(x=i, color='black', linestyle='-', linewidth=0.5)
plt.title("ROS-1 π-π stacking", fontsize=18, fontweight='bold')
plt.xlabel("GnRH/GnRH1R residues", fontsize=16, fontweight='bold')
plt.ylabel("GnRH/GnRH1R residues", fontsize=16, fontweight='bold')
plt.gca().set_facecolor('white')
plt.show()
```

Ion_D319_W280_distance.py

```
import MDAnalysis as mda

trajectory_files = ["D1.dcd", "D2.dcd", "D3.dcd", "D4.dcd",
"D5.dcd", "D6.dcd", "D7.dcd", "D8.dcd", "D9.dcd", "D10.dcd",
"D11.dcd", "D12.dcd", "D13.dcd", "D14.dcd", "D15.dcd", "D16.dcd",
"D17.dcd", "D18.dcd", "D19.dcd", "D20.dcd", "D21.dcd", "D22.dcd",
"D23.dcd", "D24.dcd", "D25.dcd", "D26.dcd", "D27.dcd"]

topology_file = "NTER_FREE_SC.psf"
```

```
u = mda.Universe(topology_file)
ion = u.select_atoms("(resid 88 and name SOD)")
#residue = u.select_atoms("(resid 319 and name OD2)")
residue = u.select_atoms("(resid 280 and name HE1)")
output_file = open("ion_W280_APO_distances.txt", "w")

for trajectory_file in trajectory_files:
    u.load_new(trajectory_file)

    for ts in u.trajectory:
        distance = mda.lib.distances.distance_array(ion.positions,
residue.positions)[0][0]
        output_file.write(f"{distance:.3f}\n")
output_file.close()
```

Plot_ion_distance.py

```
import matplotlib.pyplot as plt
import numpy as np

file_path = "/path/to/file/d318_W280_ion__distance.csv"
with open(file_path, 'r') as file:
    lines = file.readlines()
data = []
for line in lines[2:]:
    parts = line.split(',')
    time_ns = float(parts[0]) / 1000  #convert to µs
    distance_d319 = float(parts[1])
    distance_w280 = float(parts[2])
    data.append((time_ns, distance_d319, distance_w280))

x = [entry[0] for entry in data]
y_d319 = [entry[1] for entry in data]
y_w280 = [entry[2] for entry in data]

plt.figure(figsize=(18, 12))
plt.plot(x, y_w280, label='W280-Na$^+$')
plt.plot(x, y_d319, label='D319-Na$^+$')
plt.xlabel('Time (µs)', fontsize=38, fontweight='bold')
plt.ylabel('Distance (Å)', fontsize=38, fontweight='bold')
plt.title('Apo-GnRH1R', fontsize=40, fontweight='bold')
plt.xticks(np.arange(0, 1.2, 0.1), fontsize=36)
plt.yticks(np.arange(2, 13, 1), fontsize=36)
plt.legend(fontsize=33)
plt.ylim(1.8, 12)
plt.xlim(0, 1.1)
plt.savefig('ion_D319_w280_distance.pdf')
plt.show()
```

8.4 MD system preparation

Ctrbox.tcl

```
# centres the waterbox regarding the geometry

proc ctrbox { in_psf in_pdb out_pfx } {

    resetpsf
```

```
    readpsf $in_psf
    coordpdb $in_pdb

    mol load psf $in_psf pdb $in_pdb

    set all [ atomselect top all ]
    set minmax [ measure minmax $all ]

    foreach {min max} $minmax { break }
    foreach {xmin ymin zmin} $min { break }
    foreach {xmax ymax zmax} $max { break }

    set mx [ expr $xmin + abs(($xmin - $xmax) / 2) ]
    set my [ expr $ymin + abs(($ymin - $ymax) / 2) ]
    set mz [ expr $zmin + abs(($zmin - $zmax) / 2) ]

    $all moveby [ vecsub {0 0 0} [list $mx $my $mz] ]

    foreach atom [$all get {segid resid name x y z}] {
        foreach {segid resid name x y z} $atom { break }
        coord $segid $resid $name [list $x $y $z]
    }

    writepsf $out_pfx.psf
    writepdb $out_pfx.pdb

    set minmax [ measure minmax $all ]
    foreach {min max} $minmax { break }
    foreach {xmin ymin zmin} $min { break }

      set vec1 [ expr 2 * abs($xmin) + 0.5 ]
      set vec2 [ expr 2 * abs($ymin) + 0.5 ]
      set vec3 [ expr 2 * abs($zmin) + 0.5 ]

        set fp [ open "$out_pfx.pdb" a+ ]
          puts $fp "REMARK cellBasisVector1 $vec1   0.0 0.0"
      puts $fp "REMARK cellBasisVector2     0.0 $vec2   0.0"
      puts $fp "REMARK cellBasisVector3     0.0 0.0 $vec3"
      puts $fp "REMARK cellOrigin       0.0 0.0 0.0"

        close $fp

    mol delete top
}
```

Fix_protein.tcl

```
set all [ atomselect top all ]
set fix [ atomselect top "all not water" ]
$all set occupancy 0
$fix set occupancy 1
$all writepdb fix_protein.pdb
```

8.5 MD Equilibration and Production input files

D0.1.inp

```
#For water only minimisation. First open step5 .pdb and .psf to do
ctrbox.tcl and fix_protein.tcl.
#Do not change PME info in this script or add new lines. Only use
.SC.pdb and .SC.psf for input structures, see below.
#Open the fix_protein.pdb change occupancy of IONS from 1.00 0.00
to 0.00 0.00 to make them mobile. This input uses
#NPT and constrains to minimise water + ions only.

structure               No_Nter_Gnrh_1_SC.psf
coordinates             No_Nter_Gnrh_1_SC.pdb

set temp                300;
set outputname          D0.1;

# read system values written by CHARMM (need to convert uppercases
to lowercases)
exec tr "\[:upper:\]" "\[:lower:\]" < ../step5_assembly.str | sed
-e "s/ =//g" > step5_input.str
source                  step5_input.str

temperature             $temp;

outputName              $outputname; # base name for output from
this run NAMD writes two files at the end, final coord and vel in
the format of first-dyn.coor and first-dyn.vel
firsttimestep           0; # last step of previous run
restartfreq             1000;  # 1000 steps = every 2ps
dcdfreq                 5000;
dcdUnitCell             yes;    # the file will contain unit cell
info in the style of charmm dcd files. if yes, the dcd files will
contain unit cell information in the style of charmm DCD files.
xstFreq                 5000;  # XSTFreq: control how often the
extended systen configuration will be appended to the XST file
outputEnergies          125;   # 125 steps = every 0.25ps. number
of timesteps between each energy output of NAMD
outputTiming            1000;  # The number of timesteps between
each timing output shows time per step and time to completion

# Force-Field Parameters
paraTypeCharmm          on; # We're using charmm type parameter
file(s) multiple definitions may be used but only one file per
definition
parameters              toppar/par_all36m_prot.prm
parameters              toppar/par_all36_na.prm
parameters              toppar/par_all36_carb.prm
parameters              toppar/par_all36_lipid.prm
parameters              toppar/par_all36_cgenff.prm
parameters              toppar/par_interface.prm
parameters              toppar/toppar_all36_moreions.str
parameters              toppar/toppar_all36_nano_lig.str
parameters              toppar/toppar_all36_nano_lig_patch.str
parameters              toppar/toppar_all36_synthetic_polymer.str
parameters
toppar/toppar_all36_synthetic_polymer_patch.str
```

```
parameters              toppar/toppar_all36_polymer_solvent.str
parameters              toppar/toppar_water_ions.str
parameters              toppar/toppar_dum_noble_gases.str
parameters              toppar/toppar_ions_won.str
parameters              toppar/cam.str
parameters              toppar/toppar_all36_prot_arg0.str
parameters
toppar/toppar_all36_prot_c36m_d_aminoacids.str
parameters
toppar/toppar_all36_prot_fluoro_alkanes.str
parameters              toppar/toppar_all36_prot_heme.str
parameters              toppar/toppar_all36_prot_na_combined.str
parameters              toppar/toppar_all36_prot_retinol.str
parameters              toppar/toppar_all36_prot_model.str
parameters              toppar/toppar_all36_prot_modify_res.str
parameters              toppar/toppar_all36_na_nad_ppi.str
parameters              toppar/toppar_all36_na_rna_modified.str
parameters              toppar/toppar_all36_lipid_sphingo.str
parameters              toppar/toppar_all36_lipid_archaeal.str
parameters              toppar/toppar_all36_lipid_bacterial.str
parameters              toppar/toppar_all36_lipid_cardiolipin.str
parameters              toppar/toppar_all36_lipid_cholesterol.str
parameters              toppar/toppar_all36_lipid_dag.str
parameters              toppar/toppar_all36_lipid_inositol.str
parameters              toppar/toppar_all36_lipid_lnp.str
parameters              toppar/toppar_all36_lipid_lps.str
parameters
toppar/toppar_all36_lipid_mycobacterial.str
parameters
toppar/toppar_all36_lipid_miscellaneous.str
parameters              toppar/toppar_all36_lipid_model.str
parameters              toppar/toppar_all36_lipid_prot.str
parameters              toppar/toppar_all36_lipid_tag.str
parameters              toppar/toppar_all36_lipid_yeast.str
parameters              toppar/toppar_all36_lipid_hmmm.str
parameters              toppar/toppar_all36_lipid_detergent.str
parameters              toppar/toppar_all36_lipid_ether.str
parameters              toppar/toppar_all36_lipid_oxidized.str
parameters              toppar/toppar_all36_carb_glycolipid.str
parameters              toppar/toppar_all36_carb_glycopeptide.str
parameters              toppar/toppar_all36_carb_imlab.str
parameters              toppar/toppar_all36_label_spin.str
parameters              toppar/toppar_all36_label_fluorophore.str

# Nonbonded Parameters
exclude                 scaled1-4  # non-bonded exclusion policy
to use "none,1-2,1-3,1-4,or scaled1-4"# 1-2: all atoms pairs that
are bonded are going to be ignored # 1-3: 3 consecutively bonded
are excluded #scaled1-4: include all the 1-3, and modified 1-4
interactions


1-4scaling              1.0  # electrostatic scaled by 1-4scaling
factor 1.0# vdW special 1-4 parameters in charmm parameter file.

switching               on # switching is used for vdW forces.
```

```
vdwForceSwitching       on;  # New option for force-based
switching of vdW # if both switching and vdwForceSwitching are on
CHARMM force

cutoff                  12.0;
switchdist              8.0;  # cutoff - 2. # switchdist - where
you start to switch # cutoff - where you stop accounting for
nonbond interactions.
pairlistdist            14.0;  # stores the all the pairs with in
the distance it should be larger # than cutoff( + 2.)
stepspercycle           20;   # 20 redo pairlists every ten steps
pairlistsPerCycle       2;   # 2 is the default # cycle represents
the number of steps between atom reassignments # this means every
20/2=10 steps the pairlist will be updated

# Integrator Parameters
timestep                1.0;                # fs/step
rigidBonds              water; # Bound constraint all bonds
involving H are fixed in length
nonbondedFreq           1;    # nonbonded forces every step
fullElectFrequency      1;   # PME every step

# Constant Temperature Control ONLY DURING EQUILB
reassignFreq            500;   # reassignFreq:  use this to
reassign velocity every 500 steps
reassignTemp            $temp;

# Periodic Boundary conditions. Need this since for a start.
if { $boxtype == "hexa" } {
   set b [expr {$a / 2 * sqrt(3)}]
   set d [expr {$a / 2}]
   set wrapnearst on
} else {
   set wrapnearst off
   set d 0.0
}

cellBasisVector1     $a   0.0   0.0;        # vector to the next
image
cellBasisVector2     $d    $b   0.0;
cellBasisVector3    0.0   0.0    $c;
cellOrigin          0.0   0.0 $zcen;        # the *center* of the
cell

wrapWater               on;                 # wrap water to
central cell
wrapAll                 on;                 # wrap other molecules
too
wrapNearest             $wrapnearst;  # use for non-rectangular
cells (wrap to the nearest image)

# PME (for full-system periodic electrostatics)
PME                     yes;
PMEInterpOrder          6;   # interpolation order (spline order 6
in charmm)
PMEGridSpacing          1.0;  # maximum PME grid space / used to
calculate grid size
```

```
# Pressure and volume control
useGroupPressure        yes;  # use a hydrogen-group based pseudo-
molecular viral to calcualte pressure and # has less fluctuation,
is needed for rigid bonds (rigidBonds/SHAKE)
useFlexibleCell         yes;  # yes for anisotropic system like
membrane
useConstantRatio        yes;  # keeps the ratio of the unit cell
in the x-y plane constant A=B

langevin                on
langevinDamping         5.0
langevinTemp            $temp
langevinHydrogen        off

# constant pressure
langevinPiston          on
langevinPistonTarget    1.01325
langevinPistonPeriod    50.0
langevinPistonDecay     25.0
langevinPistonTemp      $temp

constraints             on
consexp                 2
consref                 restraints/prot_posres.ref
conskfile               restraints/prot_posres.ref
conskcol                B
constraintScaling       10.0

# planar restraint
exec sed -e "s/Constant \$fc/Constant 5/g" step5_input.colvar.str
> restraints/$outputname.col
colvars                 on
colvarsConfig           restraints/$outputname.col

# dihedral restraint
exec sed -e "s/\$FC/500/g" restraints/dihe.txt >
restraints/$outputname.dihe
extraBonds              yes
extraBondsFile          restraints/$outputname.dihe

fixedAtoms      on
fixedAtomsFile  fix_protein.pdb
fixedAtomsCol   O

minimize 1000
run 100000
```

Comments are given only for this script. The same explanations apply to rest of the inputs.

D0.2.inp

```
structure               No_Nter_Gnrh_1_SC.psf
coordinates             No_Nter_Gnrh_1_SC.pdb

set outputname          D0.2;
```

```
source                  step5_input.str

set inputname           D0.1;
outputname              $outputname;
binCoordinates          $inputname.coor;
extendedSystem          $inputname.xsc;

source                  step5_input.str
restartfreq             1000;
dcdfreq                 5000;
dcdUnitCell             yes;
xstFreq                 5000;
outputEnergies          125;
outputTiming            1000;
paraTypeCharmm          on;
parameters              toppar/par_all36m_prot.prm
parameters              toppar/par_all36_na.prm
parameters              toppar/par_all36_carb.prm
parameters              toppar/par_all36_lipid.prm
parameters              toppar/par_all36_cgenff.prm
parameters              toppar/par_interface.prm
parameters              toppar/toppar_all36_moreions.str
parameters              toppar/toppar_all36_nano_lig.str
parameters              toppar/toppar_all36_nano_lig_patch.str
parameters              toppar/toppar_all36_synthetic_polymer.str
parameters
toppar/toppar_all36_synthetic_polymer_patch.str
parameters              toppar/toppar_all36_polymer_solvent.str
parameters              toppar/toppar_water_ions.str
parameters              toppar/toppar_dum_noble_gases.str
parameters              toppar/toppar_ions_won.str
parameters              toppar/cam.str
parameters              toppar/toppar_all36_prot_arg0.str
parameters
toppar/toppar_all36_prot_c36m_d_aminoacids.str
parameters
toppar/toppar_all36_prot_fluoro_alkanes.str
parameters              toppar/toppar_all36_prot_heme.str
parameters              toppar/toppar_all36_prot_na_combined.str
parameters              toppar/toppar_all36_prot_retinol.str
parameters              toppar/toppar_all36_prot_model.str
parameters              toppar/toppar_all36_prot_modify_res.str
parameters              toppar/toppar_all36_na_nad_ppi.str
parameters              toppar/toppar_all36_na_rna_modified.str
parameters              toppar/toppar_all36_lipid_sphingo.str
parameters              toppar/toppar_all36_lipid_archaeal.str
parameters              toppar/toppar_all36_lipid_bacterial.str
parameters              toppar/toppar_all36_lipid_cardiolipin.str
parameters              toppar/toppar_all36_lipid_cholesterol.str
parameters              toppar/toppar_all36_lipid_dag.str
parameters              toppar/toppar_all36_lipid_inositol.str
parameters              toppar/toppar_all36_lipid_lnp.str
parameters              toppar/toppar_all36_lipid_lps.str
parameters
toppar/toppar_all36_lipid_mycobacterial.str
parameters
toppar/toppar_all36_lipid_miscellaneous.str
```

```
parameters              toppar/toppar_all36_lipid_model.str
parameters              toppar/toppar_all36_lipid_prot.str
parameters              toppar/toppar_all36_lipid_tag.str
parameters              toppar/toppar_all36_lipid_yeast.str
parameters              toppar/toppar_all36_lipid_hmmm.str
parameters              toppar/toppar_all36_lipid_detergent.str
parameters              toppar/toppar_all36_lipid_ether.str
parameters              toppar/toppar_all36_lipid_oxidized.str
parameters              toppar/toppar_all36_carb_glycolipid.str
parameters              toppar/toppar_all36_carb_glycopeptide.str
parameters              toppar/toppar_all36_carb_imlab.str
parameters              toppar/toppar_all36_label_spin.str
parameters              toppar/toppar_all36_label_fluorophore.str

exclude                 scaled1-4
1-4scaling              1.0
switching               on
vdwForceSwitching       on;
cutoff                  12.0;
switchdist              8.0;
pairlistdist            14.0;
stepspercycle           20;
pairlistsPerCycle       2;
timestep                1.0;
rigidBonds              water;
nonbondedFreq           1;
fullElectFrequency      1;

if { $boxtype == "hexa" } {
   set b [expr {$a / 2 * sqrt(3)}]
   set d [expr {$a / 2}]
   set wrapnearst on
} else {
   set wrapnearst off
   set d 0.0
}

cellBasisVector1     $a   0.0   0.0;
cellBasisVector2     $d    $b   0.0;
cellBasisVector3    0.0   0.0    $c;
cellOrigin          0.0   0.0 $zcen;

wrapWater               on;
wrapAll                 on;
wrapNearest             $wrapnearst;
PME                     yes;
PMEInterpOrder          6;
PMEGridSpacing          1.0;
useGroupPressure        yes;
useFlexibleCell         yes;
useConstantRatio        yes;

temperature     0

reassignFreq        1000
reassignIncr        10
reassignHold        300
```

```
constraints             on
consexp                 2
consref                 restraints/prot_posres.ref
conskfile               restraints/prot_posres.ref
conskcol                B
constraintScaling       10.0

# planar restraint
exec sed -e "s/Constant \$fc/Constant 5/g" step5_input.colvar.str
> restraints/$outputname.col
colvars                 on
colvarsConfig           restraints/$outputname.col

# dihedral restraint
exec sed -e "s/\$FC/500/g" restraints/dihe.txt >
restraints/$outputname.dihe
extraBonds              yes
extraBondsFile          restraints/$outputname.dihe

minimize            10000
run                 300000
```

D0.3.inp

```
structure               No_Nter_Gnrh_1_SC.psf
coordinates             No_Nter_Gnrh_1_SC.pdb

set temp                300;
set outputname          D0.3;
source                  step5_input.str

set inputname           D0.2;
outputname              $outputname;
binCoordinates          $inputname.coor;
extendedSystem          $inputname.xsc;

source                  step5_input.str

temperature             $temp;
restartfreq             1000;
dcdfreq                 5000;
dcdUnitCell             yes;
xstFreq                 5000;
outputEnergies          125;
outputTiming            1000;
paraTypeCharmm          on;
parameters              toppar/par_all36m_prot.prm
parameters              toppar/par_all36_na.prm
parameters              toppar/par_all36_carb.prm
parameters              toppar/par_all36_lipid.prm
parameters              toppar/par_all36_cgenff.prm
parameters              toppar/par_interface.prm
parameters              toppar/toppar_all36_moreions.str
parameters              toppar/toppar_all36_nano_lig.str
parameters              toppar/toppar_all36_nano_lig_patch.str
parameters              toppar/toppar_all36_synthetic_polymer.str
```

```
parameters
toppar/toppar_all36_synthetic_polymer_patch.str
parameters              toppar/toppar_all36_polymer_solvent.str
parameters              toppar/toppar_water_ions.str
parameters              toppar/toppar_dum_noble_gases.str
parameters              toppar/toppar_ions_won.str
parameters              toppar/cam.str
parameters              toppar/toppar_all36_prot_arg0.str
parameters
toppar/toppar_all36_prot_c36m_d_aminoacids.str
parameters
toppar/toppar_all36_prot_fluoro_alkanes.str
parameters              toppar/toppar_all36_prot_heme.str
parameters              toppar/toppar_all36_prot_na_combined.str
parameters              toppar/toppar_all36_prot_retinol.str
parameters              toppar/toppar_all36_prot_model.str
parameters              toppar/toppar_all36_prot_modify_res.str
parameters              toppar/toppar_all36_na_nad_ppi.str
parameters              toppar/toppar_all36_na_rna_modified.str
parameters              toppar/toppar_all36_lipid_sphingo.str
parameters              toppar/toppar_all36_lipid_archaeal.str
parameters              toppar/toppar_all36_lipid_bacterial.str
parameters              toppar/toppar_all36_lipid_cardiolipin.str
parameters              toppar/toppar_all36_lipid_cholesterol.str
parameters              toppar/toppar_all36_lipid_dag.str
parameters              toppar/toppar_all36_lipid_inositol.str
parameters              toppar/toppar_all36_lipid_lnp.str
parameters              toppar/toppar_all36_lipid_lps.str
parameters
toppar/toppar_all36_lipid_mycobacterial.str
parameters
toppar/toppar_all36_lipid_miscellaneous.str
parameters              toppar/toppar_all36_lipid_model.str
parameters              toppar/toppar_all36_lipid_prot.str
parameters              toppar/toppar_all36_lipid_tag.str
parameters              toppar/toppar_all36_lipid_yeast.str
parameters              toppar/toppar_all36_lipid_hmmm.str
parameters              toppar/toppar_all36_lipid_detergent.str
parameters              toppar/toppar_all36_lipid_ether.str
parameters              toppar/toppar_all36_lipid_oxidized.str
parameters              toppar/toppar_all36_carb_glycolipid.str
parameters              toppar/toppar_all36_carb_glycopeptide.str
parameters              toppar/toppar_all36_carb_imlab.str
parameters              toppar/toppar_all36_label_spin.str
parameters              toppar/toppar_all36_label_fluorophore.str

exclude                 scaled1-4
1-4scaling              1.0
switching               on
vdwForceSwitching       on;
cutoff                  12.0;
switchdist              8.0;
pairlistdist            14.0;

stepspercycle           20;
pairlistsPerCycle       2;
timestep                1.0;
```

```
rigidBonds              water;
nonbondedFreq           1;
fullElectFrequency      1;
reassignFreq            500;
reassignTemp            $temp;

if { $boxtype == "hexa" } {
   set b [expr {$a / 2 * sqrt(3)}]
   set d [expr {$a / 2}]
   set wrapnearst on
} else {
   set wrapnearst off
   set d 0.0
}

cellBasisVector1     $a   0.0   0.0;
cellBasisVector2     $d    $b   0.0;
cellBasisVector3    0.0   0.0    $c;
cellOrigin          0.0   0.0 $zcen;

wrapWater               on;
wrapAll                 on;
wrapNearest             $wrapnearst;
PME                     yes;
PMEInterpOrder          6;
PMEGridSpacing          1.0;
useGroupPressure        yes;
useFlexibleCell         yes;
useConstantRatio        yes;

langevin                on
langevinDamping         5.0
langevinTemp            $temp
langevinHydrogen        off

# constant pressure
langevinPiston          on
langevinPistonTarget    1.01325
langevinPistonPeriod    50.0
langevinPistonDecay     25.0
langevinPistonTemp      $temp

constraints             on
consexp                 2
consref                 restraints/prot_posres.ref
conskfile               restraints/prot_posres.ref
conskcol                B
constraintScaling       10.0

# planar restraint
exec sed -e "s/Constant \$fc/Constant 5/g" step5_input.colvar.str
> restraints/$outputname.col
colvars                 on
colvarsConfig           restraints/$outputname.col

# dihedral restraint
```

```
exec sed -e "s/\$FC/500/g" restraints/dihe.txt >
restraints/$outputname.dihe
extraBonds              yes
extraBondsFile          restraints/$outputname.dihe

minimize                10000

numsteps                90000000
run                     125000
```

D0.4.inp

```
structure               No_Nter_Gnrh_1_SC.psf
coordinates             No_Nter_Gnrh_1_SC.pdb

set temp                300;
set outputname          D0.4;

source                  step5_input.str
set inputname           D0.3;
outputname              $outputname;
binCoordinates          $inputname.coor;
binVelocities           $inputname.vel;
extendedSystem          $inputname.xsc;

restartfreq             1000;
dcdfreq                 5000;
dcdUnitCell             yes;
xstFreq                 5000;
outputEnergies          125;
outputTiming            1000;
paraTypeCharmm          on;
parameters              toppar/par_all36m_prot.prm
parameters              toppar/par_all36_na.prm
parameters              toppar/par_all36_carb.prm
parameters              toppar/par_all36_lipid.prm
parameters              toppar/par_all36_cgenff.prm
parameters              toppar/par_interface.prm
parameters              toppar/toppar_all36_moreions.str
parameters              toppar/toppar_all36_nano_lig.str
parameters              toppar/toppar_all36_nano_lig_patch.str
parameters              toppar/toppar_all36_synthetic_polymer.str
parameters
toppar/toppar_all36_synthetic_polymer_patch.str
parameters              toppar/toppar_all36_polymer_solvent.str
parameters              toppar/toppar_water_ions.str
parameters              toppar/toppar_dum_noble_gases.str
parameters              toppar/toppar_ions_won.str
parameters              toppar/cam.str
parameters              toppar/toppar_all36_prot_arg0.str
parameters
toppar/toppar_all36_prot_c36m_d_aminoacids.str
parameters
toppar/toppar_all36_prot_fluoro_alkanes.str
parameters              toppar/toppar_all36_prot_heme.str
parameters              toppar/toppar_all36_prot_na_combined.str
parameters              toppar/toppar_all36_prot_retinol.str
```

```
parameters              toppar/toppar_all36_prot_model.str
parameters              toppar/toppar_all36_prot_modify_res.str
parameters              toppar/toppar_all36_na_nad_ppi.str
parameters              toppar/toppar_all36_na_rna_modified.str
parameters              toppar/toppar_all36_lipid_sphingo.str
parameters              toppar/toppar_all36_lipid_archaeal.str
parameters              toppar/toppar_all36_lipid_bacterial.str
parameters              toppar/toppar_all36_lipid_cardiolipin.str
parameters              toppar/toppar_all36_lipid_cholesterol.str
parameters              toppar/toppar_all36_lipid_dag.str
parameters              toppar/toppar_all36_lipid_inositol.str
parameters              toppar/toppar_all36_lipid_lnp.str
parameters              toppar/toppar_all36_lipid_lps.str
parameters
toppar/toppar_all36_lipid_mycobacterial.str
parameters
toppar/toppar_all36_lipid_miscellaneous.str
parameters              toppar/toppar_all36_lipid_model.str
parameters              toppar/toppar_all36_lipid_prot.str
parameters              toppar/toppar_all36_lipid_tag.str
parameters              toppar/toppar_all36_lipid_yeast.str
parameters              toppar/toppar_all36_lipid_hmmm.str
parameters              toppar/toppar_all36_lipid_detergent.str
parameters              toppar/toppar_all36_lipid_ether.str
parameters              toppar/toppar_all36_lipid_oxidized.str
parameters              toppar/toppar_all36_carb_glycolipid.str
parameters              toppar/toppar_all36_carb_glycopeptide.str
parameters              toppar/toppar_all36_carb_imlab.str
parameters              toppar/toppar_all36_label_spin.str
parameters              toppar/toppar_all36_label_fluorophore.str

exclude                 scaled1-4
1-4scaling              1.0
switching               on
vdwForceSwitching       on;
cutoff                  12.0
switchdist              8.0;
pairlistdist            14.0;
stepspercycle           20;
pairlistsPerCycle       2;
timestep                1.0;
rigidBonds              water;
nonbondedFreq           1;
fullElectFrequency      1;
reassignFreq            500;
reassignTemp            $temp;

if { $boxtype == "hexa" } {
   set wrapnearst on
} else {
   set wrapnearst off
}

wrapWater               on;
wrapAll                 on;
wrapNearest             $wrapnearst;
PME                     yes;
```

```
PMEInterpOrder          6;
PMEGridSpacing          1.0;
useGroupPressure        yes;
useFlexibleCell         yes;
useConstantRatio        yes;

langevin                on
langevinDamping         5.0
langevinTemp            $temp
langevinHydrogen        off

# constant pressure
langevinPiston          on
langevinPistonTarget    1.01325
langevinPistonPeriod    50.0
langevinPistonDecay     25.0
langevinPistonTemp      $temp

constraints             on
consexp                 2
consref                 restraints/prot_posres.ref
conskfile               restraints/prot_posres.ref
conskcol                B
constraintScaling       5.0

# planar restraint
exec sed -e "s/Constant \$fc/Constant 5/g" step5_input.colvar.str
> restraints/$outputname.col
colvars                 on
colvarsConfig           restraints/$outputname.col

# dihedral restraint
exec sed -e "s/\$FC/200/g" restraints/dihe.txt >
restraints/$outputname.dihe
extraBonds              yes
extraBondsFile          restraints/$outputname.dihe

numsteps                90000000
run                     125000
```

D0.5.inp

```
structure               No_Nter_Gnrh_1_SC.psf
coordinates             No_Nter_Gnrh_1_SC.pdb

set temp                300;
set outputname          D0.5;

source                  step5_input.str
set inputname           D0.4;
outputname              $outputname;
binCoordinates          $inputname.coor;
binVelocities           $inputname.vel;
extendedSystem          $inputname.xsc;

restartfreq             1000;
dcdfreq                 5000;
```

```
dcdUnitCell             yes;
xstFreq                 5000;
outputEnergies          125;
outputTiming            1000;
paraTypeCharmm          on;
parameters              toppar/par_all36m_prot.prm
parameters              toppar/par_all36_na.prm
parameters              toppar/par_all36_carb.prm
parameters              toppar/par_all36_lipid.prm
parameters              toppar/par_all36_cgenff.prm
parameters              toppar/par_interface.prm
parameters              toppar/toppar_all36_moreions.str
parameters              toppar/toppar_all36_nano_lig.str
parameters              toppar/toppar_all36_nano_lig_patch.str
parameters              toppar/toppar_all36_synthetic_polymer.str
parameters
toppar/toppar_all36_synthetic_polymer_patch.str
parameters              toppar/toppar_all36_polymer_solvent.str
parameters              toppar/toppar_water_ions.str
parameters              toppar/toppar_dum_noble_gases.str
parameters              toppar/toppar_ions_won.str
parameters              toppar/cam.str
parameters              toppar/toppar_all36_prot_arg0.str
parameters
toppar/toppar_all36_prot_c36m_d_aminoacids.str
parameters
toppar/toppar_all36_prot_fluoro_alkanes.str
parameters              toppar/toppar_all36_prot_heme.str
parameters              toppar/toppar_all36_prot_na_combined.str
parameters              toppar/toppar_all36_prot_retinol.str
parameters              toppar/toppar_all36_prot_model.str
parameters              toppar/toppar_all36_prot_modify_res.str
parameters              toppar/toppar_all36_na_nad_ppi.str
parameters              toppar/toppar_all36_na_rna_modified.str
parameters              toppar/toppar_all36_lipid_sphingo.str
parameters              toppar/toppar_all36_lipid_archaeal.str
parameters              toppar/toppar_all36_lipid_bacterial.str
parameters              toppar/toppar_all36_lipid_cardiolipin.str
parameters              toppar/toppar_all36_lipid_cholesterol.str
parameters              toppar/toppar_all36_lipid_dag.str
parameters              toppar/toppar_all36_lipid_inositol.str
parameters              toppar/toppar_all36_lipid_lnp.str
parameters              toppar/toppar_all36_lipid_lps.str
parameters
toppar/toppar_all36_lipid_mycobacterial.str
parameters
toppar/toppar_all36_lipid_miscellaneous.str
parameters              toppar/toppar_all36_lipid_model.str
parameters              toppar/toppar_all36_lipid_prot.str
parameters              toppar/toppar_all36_lipid_tag.str
parameters              toppar/toppar_all36_lipid_yeast.str
parameters              toppar/toppar_all36_lipid_hmmm.str
parameters              toppar/toppar_all36_lipid_detergent.str
parameters              toppar/toppar_all36_lipid_ether.str
parameters              toppar/toppar_all36_lipid_oxidized.str
parameters              toppar/toppar_all36_carb_glycolipid.str
parameters              toppar/toppar_all36_carb_glycopeptide.str
```

```
parameters              toppar/toppar_all36_carb_imlab.str
parameters              toppar/toppar_all36_label_spin.str
parameters              toppar/toppar_all36_label_fluorophore.str

exclude                 scaled1-4
1-4scaling              1.0
switching               on
vdwForceSwitching       on;
cutoff                  12.0;
switchdist              8.0;
pairlistdist            14.0;
stepspercycle           20;
pairlistsPerCycle       2;
timestep                1.0;
rigidBonds              water;
nonbondedFreq           1;
fullElectFrequency      1;
reassignFreq            500;
reassignTemp            $temp;

if { $boxtype == "hexa" } {
   set wrapnearst on
} else {
   set wrapnearst off
}

wrapWater               on;
wrapAll                 on;
wrapNearest             $wrapnearst;
PME                     yes;
PMEInterpOrder          6;
PMEGridSpacing          1.0;
useGroupPressure        yes;
useFlexibleCell         yes;
useConstantRatio        yes;

langevin                on
langevinDamping         5.0
langevinTemp            $temp
langevinHydrogen        off

# constant pressure
langevinPiston          on
langevinPistonTarget    1.01325
langevinPistonPeriod    50.0
langevinPistonDecay     25.0
langevinPistonTemp      $temp

constraints             on
consexp                 2
consref                 restraints/prot_posres.ref
conskfile               restraints/prot_posres.ref
conskcol                B
constraintScaling       2.5

# planar restraint
```

```
exec sed -e "s/Constant \$fc/Constant 2/g" step5_input.colvar.str
> restraints/$outputname.col
colvars                 on
colvarsConfig           restraints/$outputname.col

# dihedral restraint
exec sed -e "s/\$FC/100/g" restraints/dihe.txt >
restraints/$outputname.dihe
extraBonds              yes
extraBondsFile          restraints/$outputname.dihe

numsteps                90000000
run                     125000
```

D0.6.inp

```
structure               No_Nter_Gnrh_1_SC.psf
coordinates             No_Nter_Gnrh_1_SC.pdb

set temp                300;
set outputname          D0.6;

source                  step5_input.str
set inputname           D0.5;
outputname              $outputname;
binCoordinates          $inputname.coor;
binVelocities           $inputname.vel;
extendedSystem          $inputname.xsc;

restartfreq             1000;
dcdfreq                 5000;
dcdUnitCell             yes;
xstFreq                 5000;
outputEnergies          125;
outputTiming            1000;
paraTypeCharmm          on;
parameters              toppar/par_all36m_prot.prm
parameters              toppar/par_all36_na.prm
parameters              toppar/par_all36_carb.prm
parameters              toppar/par_all36_lipid.prm
parameters              toppar/par_all36_cgenff.prm
parameters              toppar/par_interface.prm
parameters              toppar/toppar_all36_moreions.str
parameters              toppar/toppar_all36_nano_lig.str
parameters              toppar/toppar_all36_nano_lig_patch.str
parameters              toppar/toppar_all36_synthetic_polymer.str
parameters
toppar/toppar_all36_synthetic_polymer_patch.str
parameters              toppar/toppar_all36_polymer_solvent.str
parameters              toppar/toppar_water_ions.str
parameters              toppar/toppar_dum_noble_gases.str
parameters              toppar/toppar_ions_won.str
parameters              toppar/cam.str
parameters              toppar/toppar_all36_prot_arg0.str
parameters
toppar/toppar_all36_prot_c36m_d_aminoacids.str
```

```
parameters
toppar/toppar_all36_prot_fluoro_alkanes.str
parameters              toppar/toppar_all36_prot_heme.str
parameters              toppar/toppar_all36_prot_na_combined.str
parameters              toppar/toppar_all36_prot_retinol.str
parameters              toppar/toppar_all36_prot_model.str
parameters              toppar/toppar_all36_prot_modify_res.str
parameters              toppar/toppar_all36_na_nad_ppi.str
parameters              toppar/toppar_all36_na_rna_modified.str
parameters              toppar/toppar_all36_lipid_sphingo.str
parameters              toppar/toppar_all36_lipid_archaeal.str
parameters              toppar/toppar_all36_lipid_bacterial.str
parameters              toppar/toppar_all36_lipid_cardiolipin.str
parameters              toppar/toppar_all36_lipid_cholesterol.str
parameters              toppar/toppar_all36_lipid_dag.str
parameters              toppar/toppar_all36_lipid_inositol.str
parameters              toppar/toppar_all36_lipid_lnp.str
parameters              toppar/toppar_all36_lipid_lps.str
parameters
toppar/toppar_all36_lipid_mycobacterial.str
parameters
toppar/toppar_all36_lipid_miscellaneous.str
parameters              toppar/toppar_all36_lipid_model.str
parameters              toppar/toppar_all36_lipid_prot.str
parameters              toppar/toppar_all36_lipid_tag.str
parameters              toppar/toppar_all36_lipid_yeast.str
parameters              toppar/toppar_all36_lipid_hmmm.str
parameters              toppar/toppar_all36_lipid_detergent.str
parameters              toppar/toppar_all36_lipid_ether.str
parameters              toppar/toppar_all36_lipid_oxidized.str
parameters              toppar/toppar_all36_carb_glycolipid.str
parameters              toppar/toppar_all36_carb_glycopeptide.str
parameters              toppar/toppar_all36_carb_imlab.str
parameters              toppar/toppar_all36_label_spin.str
parameters              toppar/toppar_all36_label_fluorophore.str

exclude                 scaled1-4
1-4scaling              1.0
switching               on
vdwForceSwitching       on;
cutoff                  12.0;
switchdist              8.0;
pairlistdist            14.0;
stepspercycle           20;
pairlistsPerCycle       2;
timestep                1.0;
rigidBonds              water;
nonbondedFreq           1;
fullElectFrequency      1;
reassignFreq            500;
reassignTemp            $temp;

if { $boxtype == "hexa" } {
   set wrapnearst on
} else {
   set wrapnearst off
}
```

```
wrapWater               on;
wrapAll                 on;
wrapNearest             $wrapnearst;
PME                     yes;
PMEInterpOrder          6;
PMEGridSpacing          1.0;
useGroupPressure        yes;
useFlexibleCell         yes;
useConstantRatio        yes;

langevin                on
langevinDamping         5.0
langevinTemp            $temp
langevinHydrogen        off

# constant pressure
langevinPiston          on
langevinPistonTarget    1.01325
langevinPistonPeriod    50.0
langevinPistonDecay     25.0
langevinPistonTemp      $temp

constraints             on
consexp                 2
consref                 restraints/prot_posres.ref
conskfile               restraints/prot_posres.ref
conskcol                B
constraintScaling       1.0

# planar restraint
exec sed -e "s/Constant \$fc/Constant 1/g" step5_input.colvar.str
> restraints/$outputname.col
colvars                 on
colvarsConfig           restraints/$outputname.col

# dihedral restraint
exec sed -e "s/\$FC/100/g" restraints/dihe.txt >
restraints/$outputname.dihe
extraBonds              yes
extraBondsFile          restraints/$outputname.dihe

numsteps                90000000
run                     250000
```

D0.7.inp

```
structure               No_Nter_Gnrh_1_SC.psf
coordinates             No_Nter_Gnrh_1_SC.pdb

set temp                300;
set outputname          D0.7;

source                  step5_input.str
set inputname           D0.6;
outputname              $outputname;
binCoordinates          $inputname.coor;
```

```
binVelocities           $inputname.vel;
extendedSystem          $inputname.xsc;

restartfreq             1000;
dcdfreq                 5000;
dcdUnitCell             yes;
xstFreq                 5000;
outputEnergies          125;
outputTiming            1000;
paraTypeCharmm          on;
parameters              toppar/par_all36m_prot.prm
parameters              toppar/par_all36_na.prm
parameters              toppar/par_all36_carb.prm
parameters              toppar/par_all36_lipid.prm
parameters              toppar/par_all36_cgenff.prm
parameters              toppar/par_interface.prm
parameters              toppar/toppar_all36_moreions.str
parameters              toppar/toppar_all36_nano_lig.str
parameters              toppar/toppar_all36_nano_lig_patch.str
parameters              toppar/toppar_all36_synthetic_polymer.str
parameters
toppar/toppar_all36_synthetic_polymer_patch.str
parameters              toppar/toppar_all36_polymer_solvent.str
parameters              toppar/toppar_water_ions.str
parameters              toppar/toppar_dum_noble_gases.str
parameters              toppar/toppar_ions_won.str
parameters              toppar/cam.str
parameters              toppar/toppar_all36_prot_arg0.str
parameters
toppar/toppar_all36_prot_c36m_d_aminoacids.str
parameters
toppar/toppar_all36_prot_fluoro_alkanes.str
parameters              toppar/toppar_all36_prot_heme.str
parameters              toppar/toppar_all36_prot_na_combined.str
parameters              toppar/toppar_all36_prot_retinol.str
parameters              toppar/toppar_all36_prot_model.str
parameters              toppar/toppar_all36_prot_modify_res.str
parameters              toppar/toppar_all36_na_nad_ppi.str
parameters              toppar/toppar_all36_na_rna_modified.str
parameters              toppar/toppar_all36_lipid_sphingo.str
parameters              toppar/toppar_all36_lipid_archaeal.str
parameters              toppar/toppar_all36_lipid_bacterial.str
parameters              toppar/toppar_all36_lipid_cardiolipin.str
parameters              toppar/toppar_all36_lipid_cholesterol.str
parameters              toppar/toppar_all36_lipid_dag.str
parameters              toppar/toppar_all36_lipid_inositol.str
parameters              toppar/toppar_all36_lipid_lnp.str
parameters              toppar/toppar_all36_lipid_lps.str
parameters
toppar/toppar_all36_lipid_mycobacterial.str
parameters
toppar/toppar_all36_lipid_miscellaneous.str
parameters              toppar/toppar_all36_lipid_model.str
parameters              toppar/toppar_all36_lipid_prot.str
parameters              toppar/toppar_all36_lipid_tag.str
parameters              toppar/toppar_all36_lipid_yeast.str
parameters              toppar/toppar_all36_lipid_hmmm.str
```

```
parameters              toppar/toppar_all36_lipid_detergent.str
parameters              toppar/toppar_all36_lipid_ether.str
parameters              toppar/toppar_all36_lipid_oxidized.str
parameters              toppar/toppar_all36_carb_glycolipid.str
parameters              toppar/toppar_all36_carb_glycopeptide.str
parameters              toppar/toppar_all36_carb_imlab.str
parameters              toppar/toppar_all36_label_spin.str
parameters              toppar/toppar_all36_label_fluorophore.str

exclude                 scaled1-4
1-4scaling              1.0
switching               on
vdwForceSwitching       on;
cutoff                  12.0;
switchdist              8.0;
pairlistdist            14.0;
stepspercycle           20;
pairlistsPerCycle       2;
timestep                1.0;
rigidBonds              water;
nonbondedFreq           1;
fullElectFrequency      1;
reassignFreq            500;
reassignTemp            $temp;

if { $boxtype == "hexa" } {
   set wrapnearst on
} else {
   set wrapnearst off
}

wrapWater               on;
wrapAll                 on;
wrapNearest             $wrapnearst;
PME                     yes;
PMEInterpOrder          6;
PMEGridSpacing          1.0;
useGroupPressure        yes;
useFlexibleCell         yes;
useConstantRatio        yes;

langevin                on
langevinDamping         5.0
langevinTemp            $temp
langevinHydrogen        off

# constant pressure
langevinPiston          on
langevinPistonTarget    1.01325
langevinPistonPeriod    50.0
langevinPistonDecay     25.0
langevinPistonTemp      $temp

constraints             on
consexp                 2
consref                 restraints/prot_posres.ref
conskfile               restraints/prot_posres.ref
```

```
conskcol                B
constraintScaling       0.5

# planar restraint
exec sed -e "s/Constant \$fc/Constant 0.2/g"
step5_input.colvar.str > restraints/$outputname.col
colvars                 on
colvarsConfig           restraints/$outputname.col

# dihedral restraint
exec sed -e "s/\$FC/50/g" restraints/dihe.txt >
restraints/$outputname.dihe
extraBonds              yes
extraBondsFile          restraints/$outputname.dihe

numsteps                90000000
run                     250000
```

D0.8.inp

```
structure               No_Nter_Gnrh_1_SC.psf
coordinates             No_Nter_Gnrh_1_SC.pdb

set temp                300;
set outputname          D0.8;

source                  step5_input.str
set inputname           D0.7;
outputname              $outputname;
binCoordinates          $inputname.coor;
binVelocities           $inputname.vel;
extendedSystem          $inputname.xsc;

restartfreq             1000;
dcdfreq                 5000;
dcdUnitCell             yes;
xstFreq                 5000;
outputEnergies          125;
outputTiming            1000;
paraTypeCharmm          on;
parameters              toppar/par_all36m_prot.prm
parameters              toppar/par_all36_na.prm
parameters              toppar/par_all36_carb.prm
parameters              toppar/par_all36_lipid.prm
parameters              toppar/par_all36_cgenff.prm
parameters              toppar/par_interface.prm
parameters              toppar/toppar_all36_moreions.str
parameters              toppar/toppar_all36_nano_lig.str
parameters              toppar/toppar_all36_nano_lig_patch.str
parameters              toppar/toppar_all36_synthetic_polymer.str
parameters
toppar/toppar_all36_synthetic_polymer_patch.str
parameters              toppar/toppar_all36_polymer_solvent.str
parameters              toppar/toppar_water_ions.str
parameters              toppar/toppar_dum_noble_gases.str
parameters              toppar/toppar_ions_won.str
parameters              toppar/cam.str
```

```
parameters              toppar/toppar_all36_prot_arg0.str
parameters
toppar/toppar_all36_prot_c36m_d_aminoacids.str
parameters
toppar/toppar_all36_prot_fluoro_alkanes.str
parameters              toppar/toppar_all36_prot_heme.str
parameters              toppar/toppar_all36_prot_na_combined.str
parameters              toppar/toppar_all36_prot_retinol.str
parameters              toppar/toppar_all36_prot_model.str
parameters              toppar/toppar_all36_prot_modify_res.str
parameters              toppar/toppar_all36_na_nad_ppi.str
parameters              toppar/toppar_all36_na_rna_modified.str
parameters              toppar/toppar_all36_lipid_sphingo.str
parameters              toppar/toppar_all36_lipid_archaeal.str
parameters              toppar/toppar_all36_lipid_bacterial.str
parameters              toppar/toppar_all36_lipid_cardiolipin.str
parameters              toppar/toppar_all36_lipid_cholesterol.str
parameters              toppar/toppar_all36_lipid_dag.str
parameters              toppar/toppar_all36_lipid_inositol.str
parameters              toppar/toppar_all36_lipid_lnp.str
parameters              toppar/toppar_all36_lipid_lps.str
parameters
toppar/toppar_all36_lipid_mycobacterial.str
parameters
toppar/toppar_all36_lipid_miscellaneous.str
parameters              toppar/toppar_all36_lipid_model.str
parameters              toppar/toppar_all36_lipid_prot.str
parameters              toppar/toppar_all36_lipid_tag.str
parameters              toppar/toppar_all36_lipid_yeast.str
parameters              toppar/toppar_all36_lipid_hmmm.str
parameters              toppar/toppar_all36_lipid_detergent.str
parameters              toppar/toppar_all36_lipid_ether.str
parameters              toppar/toppar_all36_lipid_oxidized.str
parameters              toppar/toppar_all36_carb_glycolipid.str
parameters              toppar/toppar_all36_carb_glycopeptide.str
parameters              toppar/toppar_all36_carb_imlab.str
parameters              toppar/toppar_all36_label_spin.str
parameters              toppar/toppar_all36_label_fluorophore.str

exclude                 scaled1-4
1-4scaling              1.0
switching               on
vdwForceSwitching       on;
cutoff                  12.0;
switchdist              8.0;
pairlistdist            14.0;
stepspercycle           20;
pairlistsPerCycle       2;
timestep                1.0;
rigidBonds              water;
nonbondedFreq           1;
fullElectFrequency      1;

reassignFreq            500;
reassignTemp            $temp;

if { $boxtype == "hexa" } {
```

```
   set wrapnearst on
} else {
   set wrapnearst off
}

wrapWater               on;
wrapAll                 on;
wrapNearest             $wrapnearst;
)
PME                     yes;
PMEInterpOrder          6;
PMEGridSpacing          1.0;
useGroupPressure        yes;
useFlexibleCell         yes;
useConstantRatio        yes;

langevin                on
langevinDamping         5.0
langevinTemp            $temp
langevinHydrogen        off

langevinPiston          on
langevinPistonTarget    1.01325
langevinPistonPeriod    50.0
langevinPistonDecay     25.0
langevinPistonTemp      $temp

constraints             on
consexp                 2
consref                 restraints/prot_posres.ref
conskfile                restraints/prot_posres.ref
conskcol                B
constraintScaling       0.1

exec sed -e "s/Constant \$fc/Constant 0/g" step5_input.colvar.str
> restraints/$outputname.col
colvars                 on
colvarsConfig           restraints/$outputname.col

exec sed -e "s/\$FC/0/g" restraints/dihe.txt >
restraints/$outputname.dihe
extraBonds              yes
extraBondsFile          restraints/$outputname.dihe

numsteps                90000000
run                     250000
```

D1.production.inp

```
#Production run in NPT.
structure               No_Nter_Gnrh_1_SC.psf
coordinates             No_Nter_Gnrh_1_SC.pdb

set temp                300;
outputName              D1;

set inputname           D0.8;
```

```
binCoordinates          $inputname.coor;
binVelocities           $inputname.vel;
extendedSystem          $inputname.xsc;

dcdfreq                 1000;
dcdUnitCell             yes;
xstFreq                 5000;
outputEnergies          100;
outputTiming            100;
restartfreq             100000;
paraTypeCharmm          on;
parameters              toppar/par_all36m_prot.prm
parameters              toppar/par_all36_na.prm
parameters              toppar/par_all36_carb.prm
parameters              toppar/par_all36_lipid.prm
parameters              toppar/par_all36_cgenff.prm
parameters              toppar/par_interface.prm
parameters              toppar/toppar_all36_moreions.str
parameters              toppar/toppar_all36_nano_lig.str
parameters              toppar/toppar_all36_nano_lig_patch.str
parameters              toppar/toppar_all36_synthetic_polymer.str
parameters
toppar/toppar_all36_synthetic_polymer_patch.str
parameters              toppar/toppar_all36_polymer_solvent.str
parameters              toppar/toppar_water_ions.str
parameters              toppar/toppar_dum_noble_gases.str
parameters              toppar/toppar_ions_won.str
parameters              toppar/cam.str
parameters              toppar/toppar_all36_prot_arg0.str
parameters
toppar/toppar_all36_prot_c36m_d_aminoacids.str
parameters
toppar/toppar_all36_prot_fluoro_alkanes.str
parameters              toppar/toppar_all36_prot_heme.str
parameters              toppar/toppar_all36_prot_na_combined.str
parameters              toppar/toppar_all36_prot_retinol.str
parameters              toppar/toppar_all36_prot_model.str
parameters              toppar/toppar_all36_prot_modify_res.str
parameters              toppar/toppar_all36_na_nad_ppi.str
parameters              toppar/toppar_all36_na_rna_modified.str
parameters              toppar/toppar_all36_lipid_sphingo.str
parameters              toppar/toppar_all36_lipid_archaeal.str
parameters              toppar/toppar_all36_lipid_bacterial.str
parameters              toppar/toppar_all36_lipid_cardiolipin.str
parameters              toppar/toppar_all36_lipid_cholesterol.str
parameters              toppar/toppar_all36_lipid_dag.str
parameters              toppar/toppar_all36_lipid_inositol.str
parameters              toppar/toppar_all36_lipid_lnp.str
parameters              toppar/toppar_all36_lipid_lps.str
parameters
toppar/toppar_all36_lipid_mycobacterial.str
parameters
toppar/toppar_all36_lipid_miscellaneous.str
parameters              toppar/toppar_all36_lipid_model.str
parameters              toppar/toppar_all36_lipid_prot.str
parameters              toppar/toppar_all36_lipid_tag.str
parameters              toppar/toppar_all36_lipid_yeast.str
```

```
parameters              toppar/toppar_all36_lipid_hmmm.str
parameters              toppar/toppar_all36_lipid_detergent.str
parameters              toppar/toppar_all36_lipid_ether.str
parameters              toppar/toppar_all36_lipid_oxidized.str
parameters              toppar/toppar_all36_carb_glycolipid.str
parameters              toppar/toppar_all36_carb_glycopeptide.str
parameters              toppar/toppar_all36_carb_imlab.str
parameters              toppar/toppar_all36_label_spin.str
parameters              toppar/toppar_all36_label_fluorophore.str
source                  step5_input.str

exclude                 scaled1-4
1-4scaling              1.0
switching               on
vdwForceSwitching       on;
cutoff                  12.0;
switchdist              8.0;
pairlistdist            14.0;
stepspercycle           20;
pairlistsPerCycle       2;
timestep                2.0;
rigidBonds              water;
nonbondedFreq           1;
fullElectFrequency      1;

wrapWater               on;
wrapAll                 on;
if { $boxtype == "hexa" } {
wrapNearest             on;
} else {
wrapNearest             off;
}
PME                     yes;
PMEInterpOrder          6;
PMEGridSpacing          1.0;
useGroupPressure        yes;
useFlexibleCell         yes
useConstantRatio        yes;

langevinPiston          on;
langevinPistonTarget    1.01325;
langevinPistonPeriod    50.0;
langevinPistonDecay     25.0;
langevinPistonTemp      $temp;

langevin                on;
langevinDamping         5.0;
langevinTemp            $temp;
langevinHydrogen        off;

run                     25000000; # 50ns time(ns)= timestep x run
/1000000
```

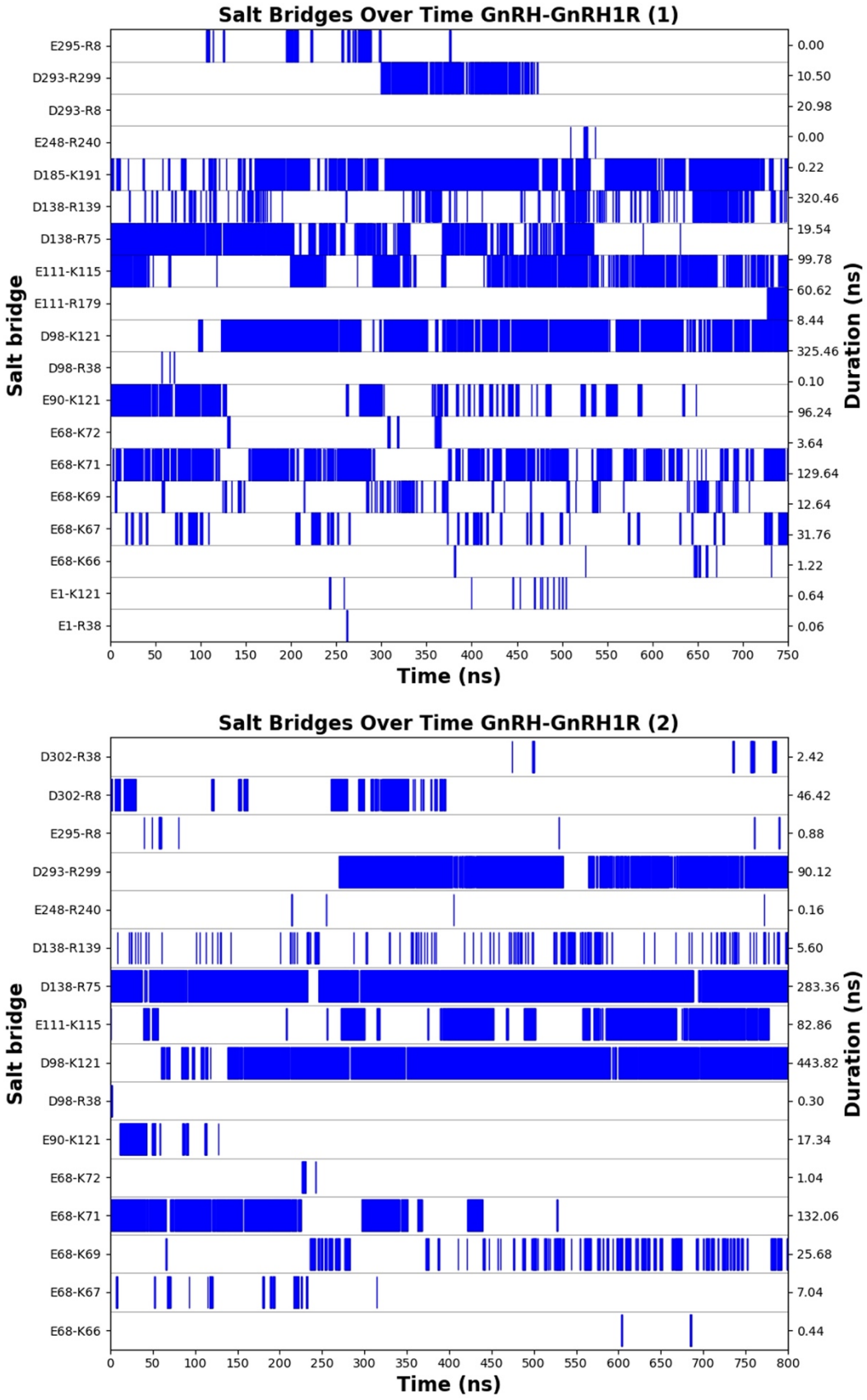


Figure 8.1: Salt bridges over time for the GnRH-GnRH1R systems

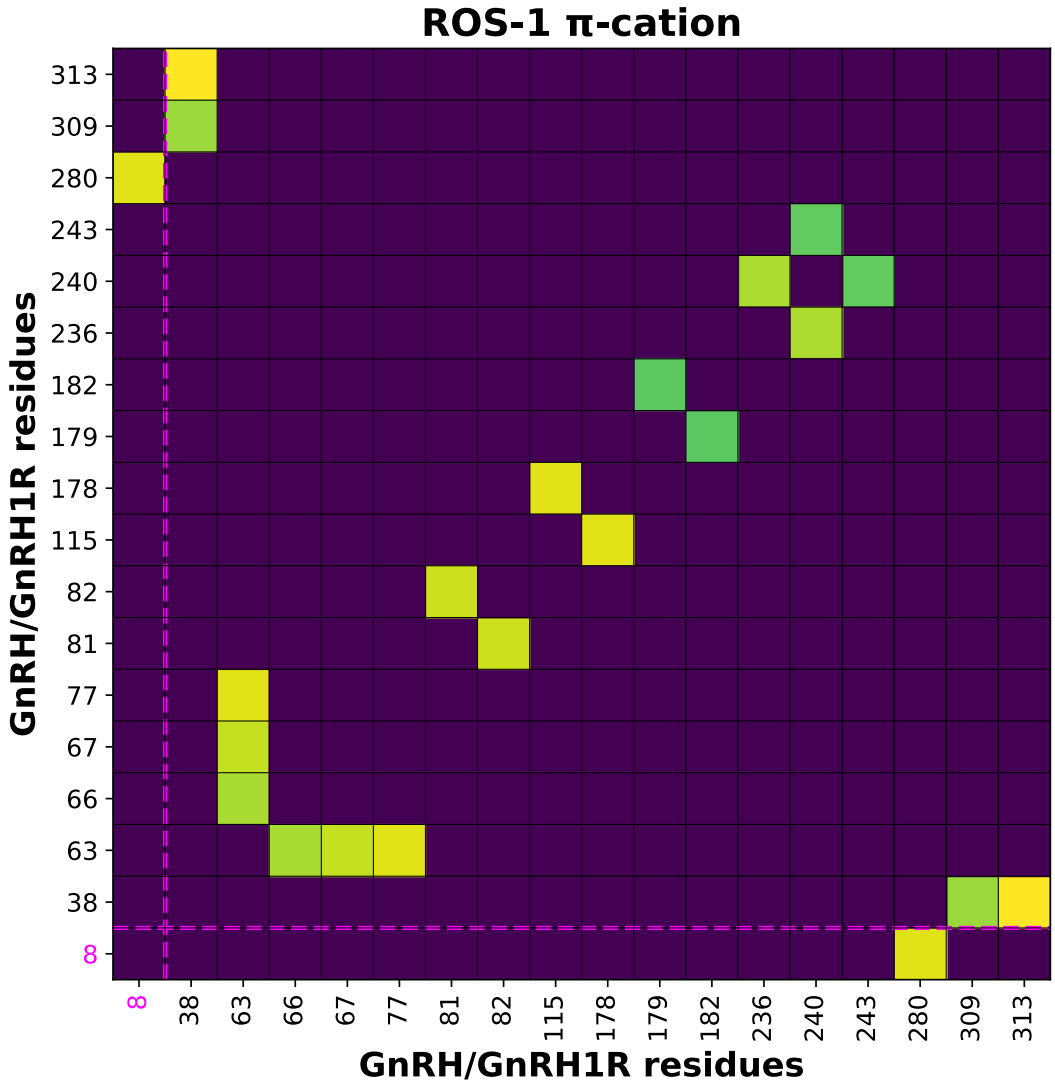


Figure 8.2: Cation-π interactions of the ROS-1 system. Intramolecular and intermolecular interactions of GnRH are showing with magenta indicators. Colour scheme follows: Blue: low and yellow: high frequency.

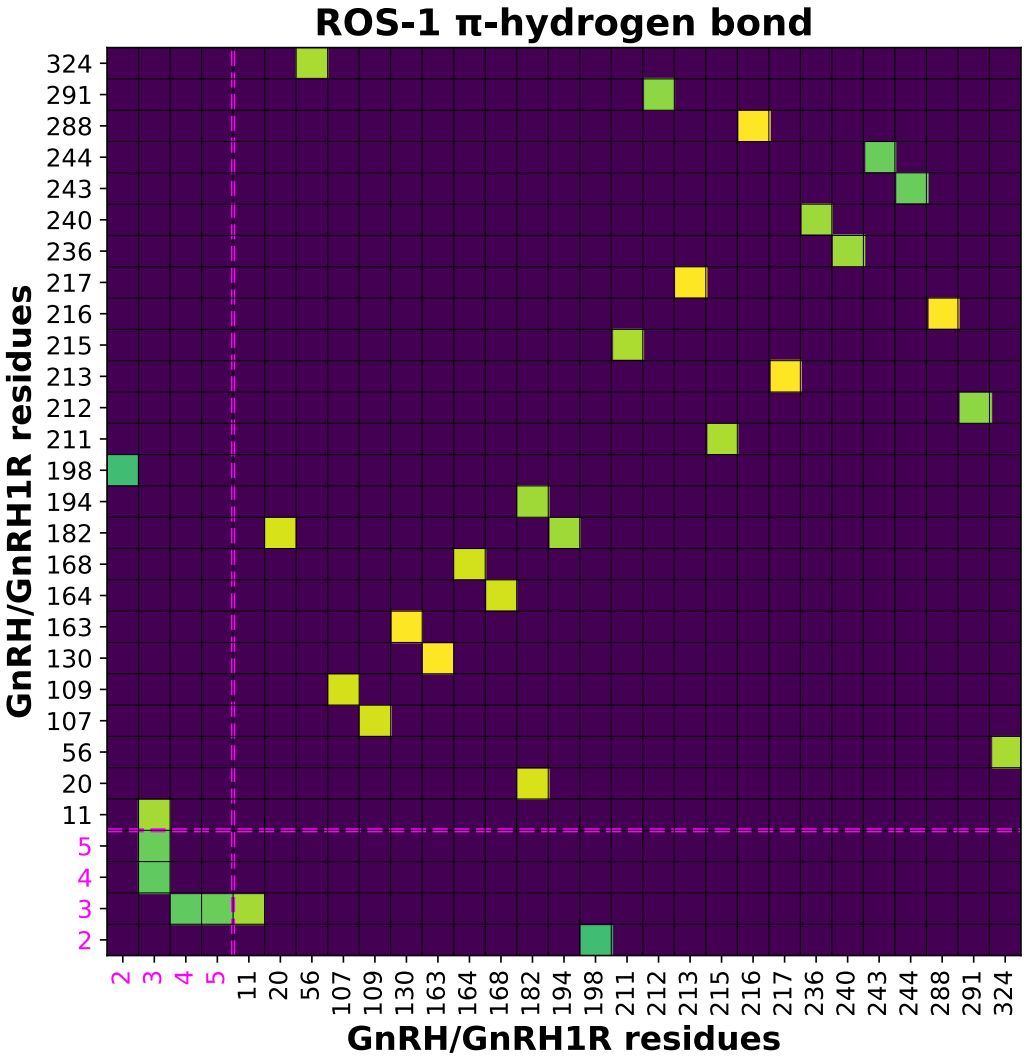


Figure 8.3: Hydrogen-π interactions of the ROS-1 system. Intramolecular and intermolecular interactions of GnRH are showing with magenta indicators. Colour scheme follows: Blue: low and yellow: high frequency.

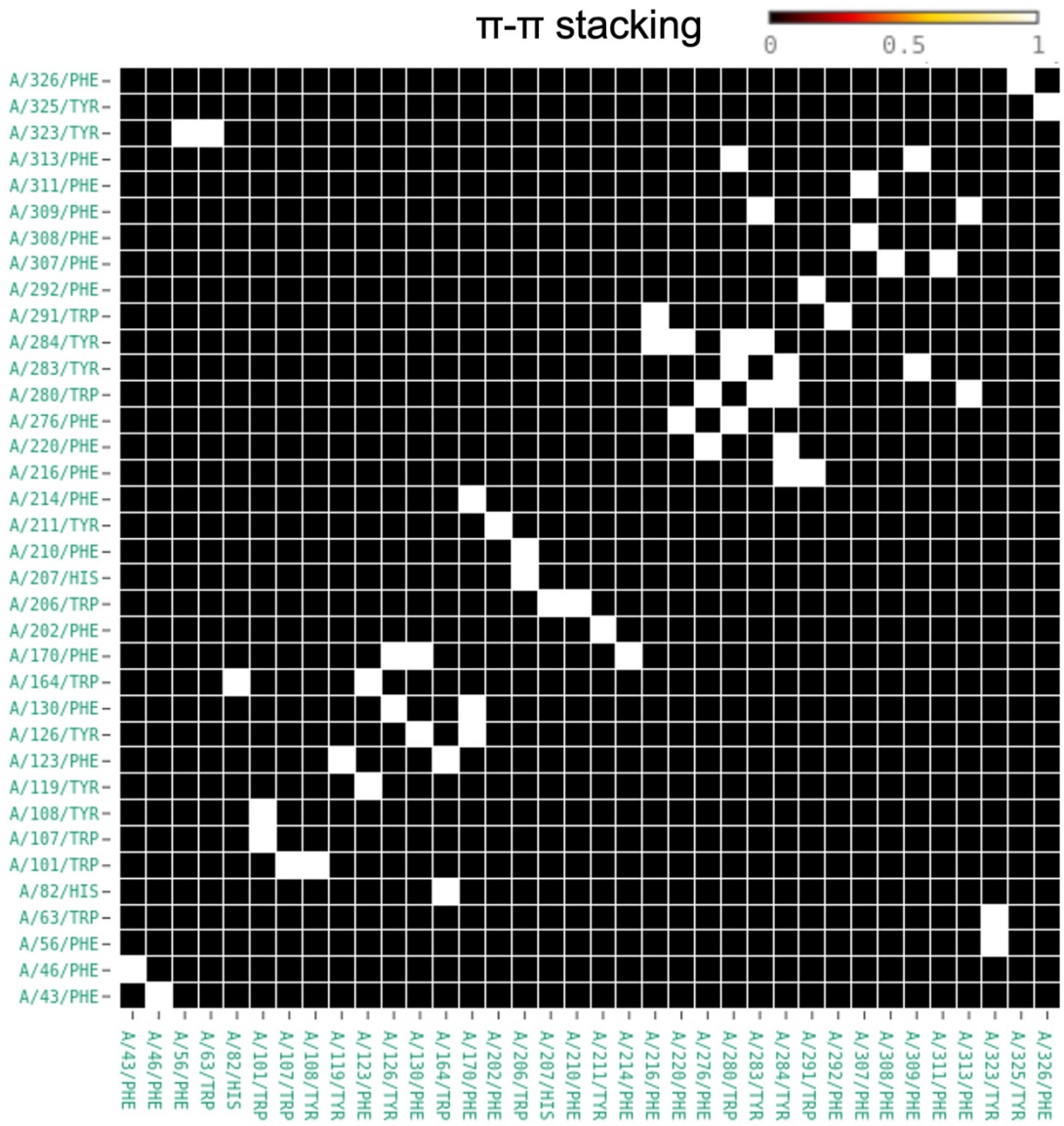


Figure 8.4: π-π stacking interactions of the inactive GnRH1R.