# An in silico study of the GCR1 role, mediated by Gibberellin GA1 and Abscisic Acid, to regulate GP in plant cells

**Pedro Miguel Hernández Acosta**
Thesis presented as a partial requirement for Doctorate in Engineering and Applied Sciences
**Andrés Jaramillo Botero**
Director
**Carlos Alberto Arango Mambuscay**
Co-Director


Facultad de Ingeniería y Ciencias
Pontificia Universidad Javeriana Cali
Cali, Colombia, Agosto 2023

# Contents

**Abstract**

GCR1 has been proposed as a plant analog to animal G-Protein coupled receptors (GPCR) that can promote or regulate several physiological processes by binding different phytohormones. The phytohormones Abscisic Acid (ABA) and Gibberellin A1 (GA1), for instance, have been shown to promote or regulate germination and flowering, root elongation, dormancy, and biotic and abiotic stresses, among others, which would put GCR1 at the heart of key signaling processes of agronomic importance. Unfortunately, this GPCR function has not been fully validated due to the lack of an X-Ray or cryo-EM 3D atomistic structure for GCR1.

Here we used the primary sequence data of GCR1 from a plant database and an *in silico* complete sampling method to screen 13 trillion possible packing of the 7 transmembrane helical domains of GCR1. From these, we down-select an ensemble of 25 configurations likely to be accessible to binding of the phytohormones Abscisic Acid (ABA) and Gibberellin A1 (GA1). We then predicted the extracellular binding sites for both phytohormones and the intracellular binding site for the G-$\alpha$ subunit, GPA1, of the plant G protein to GCR1, to determine the effect of each ligand on the GPA1 subunit and plausible activated and inactivated states of GPA1.

For the first time, we find and report significant changes in the GCR1 structure induced by the extracellular binding of GA1 and ABA. Comparing the equilibrated ABA–GCR1 and GA1–GCR1 complexes, we observed that helices in the GA1–GCR1 are offset to the right with respect to the apo-GCR1, while helices in the ABA–GCR1 are offset to the left with respect to apo-GCR1. This suggests that each ligand induces a different conformational change in the GCR1 structure. Significant changes are also observed in the intracellular loops for each structure. When we couple the GPA1 subunit to each complex, GA1-GCR1, and ABA-GCR1, once again, we find significant differences in the distances between the Ras and Helical domains, as well as structural changes that are driven by the presence/absence of salt bridges between the domains. This evidence provides essential insights into the potential for GDP-GTP exchange, induced by the phytohormones; hence contributing to support the role of GCR1 as a GPCR analog in plants.

The complex GA1-GCR1-GPA1 promotes the Ras-Helical distance increase, facilitating the GDP-GTP exchange and the subsequent activation of GPA1. In contrast, the complex ABA-GCR1-GPA1 promotes the Ras-Helical distance shortening related to the inactive form of GPA1. These

computational results agree with the experimental observations about the processes of germination and dormancy of seeds: germination is related to the active form of GPA1. Meanwhile, dormancy is associated with the inactive form of GPA1. Therefore, GA1 would promote the germination of seeds. ABA, on the other hand, would promote the dormancy of the seeds.

We employed metadynamics to produce the potential of mean force (PMF) for the complexes ABA-GCR1-GPA1-GDP-$Mg^{2+}$ and GA1-GCR1-GPA1-GDP- $Mg^{2+}$ as a function of the distance between the centers of mass of the Ras and Helical domains of GPA1. The PMF of the ABA and GA1 complexes are asymmetric double wells. The left well is located at $\approx 30 - 32\,\text{Å}$, the right well is located at $\approx 35 - 37\,\text{Å}$. The potential energy barriers, from inactive to active, are $\approx 100 - 120\,\text{kcal/mol}$ for ABA complex, and $\approx 3 - 5\,\text{kcal/mol}$. The left well of the ABA complex is lower in free energy than the right well by $\approx 60\,\text{kcal/mol}$, while the left well of the GA1 complex is higher in free energy than the right well by $\approx 14\,\text{kcal/mol}$. The free energy barrier between the wells is higher for the ABA complex than for the GA1 complex, by $\approx 100\,\text{kcal/mol}$, which leads to conclude that ABA maintains the protein GPA1 closed in its inactive conformation. At the same time, GA1 promotes the opening of GPA1 protein to its active conformation.

We compared our GCR1 structure with the ones of Tadesse and AlphaFold2. Our structure shows an ion lock between helices 2 and 4, similar to the characteristic of helices 2 and 6 of animal GPCRs. This ionic lock is not found in Tadesse and AlphaFold structures. We found the ionic lock from simulating physiological conditions using our MD simulation approach. At the secondary structure level, there are some similarities between our model and those of Tadesse and AlphFold2.

To enable experimental validation of the predicted ligand-GCR1 complexes, we identified a few mutations that would increase or weaken the binding interactions (stability of the ligands). Synthesizing and characterizing the corresponding mutants would contribute to establishing, once and for all, the physiological role of GCR1 in plants.

# Chapter 1

# Introduction

G protein-coupled receptors (GPCRs) are central trans-membrane proteins that detect molecules outside animal cells and activate cellular responses. Although there are no known GPCR proteins in plants, the transmembrane protein GCR1 has been proposed as a possible GPCR-analog protein in plant cells[1]. To date, there is no crystal structure of GCR1, and its tertiary structure is unknown. Although there have been efforts to produce a 3D-structure of GCR1, by in silico procedures, as by AlphaFold [2], this structure does not include the structural effects caused by the embedding of the protein in the cell membrane and the physiological environment, neither includes the molecular dynamics of the protein. The GEnSeMBLE methodology, developed by Goddard and co-workers[3], considers the physiological effects of the cell membrane and the physiological environment on protein. The GEnSeMBLE methodology has been used to predict successfully the 3D-structure of several GPCR proteins. In Chapters 6 and 7 of this dissertation/thesis, the GEnSeMBLE methodology has been used to elucidate the thermally equilibrated 3D-structure of GCR1.

Hormones are molecules that interact with receptor proteins to initiate the signaling pathway. Abscisic acid (ABA) and Gibberellins (GAs) are well-known phytohormones related to physiological responses such as plant growth, dormancy, stress response, and seed germination, among others [4, 5, 6, 7]. The first step of the plant signaling mechanism is the docking of the phytohormone to the binding site of the receptor protein. Biochemical and genetic evidence shows that GCR1 is involved in the G protein signaling mechanism [8, 9]. The binding between the hormones and receptors has been studied computationally using the method known as molecular docking. Molecular docking predicts the preferred orientation of the ligand (hormone)

in the binding site of the receptor protein. The docking of the ligand to the receptor induces conformational changes in the tertiary structure of the protein. These conformational changes are of central importance in the signaling mechanism. Chapters 8 and 9 of this dissertation/thesis study the use of molecular docking between the ligands, ABA and GA1, to GCR1 and the thermodynamic stability of these ligand-protein complexes.

In the animal signaling mechanism, the binding of the ligand to GPCRs triggers a series of structural changes, mainly in the TM3-TM6 ionic lock, that ultimately unlock the G-protein binding site[10]. It has been proposed that the TM2-TM4 interaction in GCR1 plays a similar role to the TM3-TM6 ionic lock of GPCRs [11]. In Chapter 6 of this dissertation/thesis, we show computational evidence that suggests that the regulation of the G-$\alpha$ subunit of the G-Protein, GPA1, is mediated by ABA and GA1. Depending on the ligand, ABA or GA1, GPA1 undergoes conformational changes measurable as the distance between the centers of mass of the Ras and Helical domains. We have found that the binding of GA1 to GCR1 induces the loss of salt bridges between the Ras and Helical domains of GPA1 which is associated to the activation of GPA1 by allowing the GDP/GTP exchange [12]. Contrary to GA1, the binding between ABA and GCR1, induces the formation of salt bridges between the Ras and Helical domains of GPA1 preventing the GDP/GTP exchange.

# Chapter 2

# Background

## 2.1 Bio-informatics

The gene gcr1 contains all the information to obtain the primary structure of the G-protein coupled receptor 1 (GCR1). The use of translation tools as Expasy (https://web.expasy.org/translate/) allows to obtain all the amino acid sequences of GCR1 from the nucleotide sequence (gcr1).

The use of sequence alignment, by the method known as Basic Local Alignment Search Tool (BLAST), allows us to compare the primary sequence of GCR1 (query protein) with a library of sequences, giving as output the sequences with greater resemblance to GCR1. BLAST resorts to heuristic methods for finding similar sequences by the process of seeding: finding short matches (words) between two sequences. In the first step, BLAST generates a list with all the 3-letter words (3-residues) in the query sequence; The list of words also includes similar words that are not part of the query protein sequence. A scan in the database looks up the words of the query protein list. A scoring system is applied to the matches between the words of the query protein and the database, BLAST employs the BLOSUM62 weighting scheme to calculate these scores and a threshold value to discard words with lower scores. Once the 3-letter word alignment has been concluded, BLAST extends the alignment in both directions. These extensions change the value of the score, and extensions that improve the score are included in the final output of BLAST.

## 2.2 Molecular mechanics and dynamics

The biological role of GPCRs in signal transmission can be studied at the atomic level by using Molecular Dynamics. The forces that drive the atoms of a bio-molecule are the result of the inter-atomic interactions. Molecular Mechanics (MM) provides computational and mathematical models that allow the calculation of these forces for all the atoms of a biochemical system. The use of the Molecular Dynamics (MD) and its techniques allow the time integration of the classical equations of motion of a biochemical system at constant pressure, temperature, and number of particles (NPT ensemble) [13].

The dynamical state at time $t$ of a molecular system of $N$ atoms is determined by the position and velocity vectors, $\boldsymbol{r}_\alpha(t)$ and $\dot{\boldsymbol{r}}_\alpha(t)$ respectively, with $\alpha = 1, 2, \ldots, N$. The positions and velocities are related by Newton's differential equations

$$\boldsymbol{F}_\alpha(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = m_\alpha \frac{d\dot{\boldsymbol{r}}_\alpha}{dt}, \qquad \alpha = 1, 2, \ldots, N, \tag{2.1}$$

with $m_\alpha$ as the mass of the $\alpha$-th atom, and $\boldsymbol{F}_\alpha$ as the force on the $\alpha$-atom. The force $\boldsymbol{F}_\alpha$, is obtained from the potential energy function,

$$\boldsymbol{F}_\alpha = -\left(\frac{\partial U}{\partial x_\alpha}, \frac{\partial U}{\partial y_\alpha}, \frac{\partial U}{\partial z_\alpha}\right), \tag{2.2}$$

with the function $U = U(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)$ as the potential energy function of the molecular system. The potential energy $U$ is conveniently written in terms of internal coordinates: bond lengths, $b_i$, bond angles $\theta_i$, and dihedral angles (proper and improper), $\phi_i$ [14, 13].

In molecular mechanics (MM) the potential energy $U$ includes bonded and non-bonded interactions,

$$U = U_{\text{bonded}} + U_{\text{non-bonded}}. \tag{2.3}$$

The bonded part of $U$ includes interaction between atoms that are linked through covalent bonds, bond angles, proper dihedral angles, and improper dihedral angles,

$$U_{\text{bonded}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}}, \tag{2.4}$$

with

$$U_{\text{bond}} = \sum_{\text{bonds } i} k_{b,i}(b_i - b_{0,i})^2, \tag{2.5}$$

$$U_{\text{angle}} = \sum_{\text{angles } i} k_{\theta,i}(\theta_i - \theta_{0,i})^2, \tag{2.6}$$

$$U_{\text{dihedral}} = \sum_{\text{dihedral } i} \begin{cases} k_{\phi,i}(1 + \cos(n_i\phi_i - \gamma_i)), & n_i \neq 0, \\ k_{\phi,i}(\phi_i - \gamma_i)^2, & n_i = 0. \end{cases} \tag{2.7}$$

The dihedral term has two possible cases: the proper dihedral angles, $n_i \neq 0$, and the improper dihedral, $n_i = 0$.

The non-bonded part $U_{\text{non-bonded}}$ includes electrostatic (Coulomb) and van der Waals interactions,

$$U_{\text{non-bond}} = U_{\text{Coulomb}} + U_{\text{vdW}}, \tag{2.8}$$

with

$$U_{\text{Coulomb}} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}, \tag{2.9}$$

$$U_{\text{vdW}} = \sum_i \sum_{j>i} 4\varepsilon_{ij}\left(s_{ij}^{12} - s_{ij}^{6}\right), \tag{2.10}$$

where $s_{ij} = \frac{\sigma_{ij}}{r_{ij}}$, $q_i$ and $q_j$ are the electric charge on atoms $i$ and $j$, respectively, and $r_{ij} = ||\boldsymbol{r}_i - \boldsymbol{r}_j||$. The constants $\varepsilon_0$, $\varepsilon_{ij}$ and $\sigma_{ij}$, are the electric permittivity of vacuum, and the Lennard-Jones parameters respectively [14, 13].

The molecular mechanics force field refers to the mathematical expressions (2.3)-(2.9) and their parameters. These are meant to track the true potential energy surface of a system with multiple degrees of freedom. Further details about the foundations of force fields and their applications can be found in [15]. In this work we use the AMBER and CHARMM force fields through the molecular dynamics program NAMD [16].

Physiological processes are given at constant temperature and pressure (310 K and 1 atm). Molecular dynamics controls these conditions using thermostats and barostats in the Nosé-Hoover Langevin piston pressure control method [16].

## 2.3 Metadynamics

The physiological response of a cell to external stimuli can be related to structural changes of proteins at the macro-molecular length scale. Confor-

mational states of proteins are described by collective variables (CV) such as distances or angles between structural domains. Examples of conformational changes of biochemical interest are the shift of an $\alpha$-helix in a GPCR or the distance between the Ras and Helical domains of GPA. The in silico MD study of these kinds of biochemical processes is difficult due to the inaccessibility of all the conformational states of interest in a practical simulation time, the initial state of the protein could never reach the state of interest after a propagation time of thousands of nanoseconds.

Metadynamics is a computer simulation method used to estimate state functions of systems with limited ergodicity due to energy barriers. [17] Conformational states that are difficult to reach by the system can be accessed by the addition of a bias potential to the molecular Hamiltonian, $H = T + V + V_{\mathrm{b}}$. The bias potential is formally given by

$$V_{\mathrm{b}}(\boldsymbol{S}) = \int_0^t \omega\delta(|\boldsymbol{S} - \boldsymbol{S}(s)|)ds, \tag{2.11}$$

with $\omega$ as the perturbation rate, $\delta$ as the Dirac $\delta$-function, and $\boldsymbol{S}(s)$ as the collective variable as function of time ($s$). This mathematical form of $V_{\mathrm{b}}$ guarantees to obtain the free energy $F$ as a function of the collective coordinates,

$$F(\boldsymbol{S}) = -\lim_{t\to\infty} V_{\mathrm{b}}(\boldsymbol{S}). \tag{2.12}$$

In practical terms, the bias potential must be approximated by a Gaussian function,

$$V_{\mathrm{b}}(\boldsymbol{S}) \approx \tau_{\mathrm{b}} \sum_{i=0}^{N_{\mathrm{b}}} \omega \exp\left(-\frac{1}{2}\left|\frac{\boldsymbol{S} - \boldsymbol{S}(i\tau_{\mathrm{b}})}{\boldsymbol{\sigma}}\right|^2\right), \tag{2.13}$$

with $N_{\mathrm{b}}$ as the number of perturbations hitting the system, $\tau_{\mathrm{b}} = \frac{t_{\mathrm{sim}}}{N_{\mathrm{b}}}$ as the time period of the Gaussian perturbation, and $\boldsymbol{\sigma}$ as the width of the perturbation along the collective variables.

Because of the insufficient sampling of MD simulations in systems with a sizeable number of degrees of freedom (many local minima separated by high-energy barriers), we additionally used metadynamics simulations to determine the free energy landscape for one of our systems, where the normal evolution of the system is biased by a history-dependent potential constructed as a sum of Gaussians centered along the trajectory followed by a suitably chosen set of collective variables [17].

To determinate the free energy profile using NAMD, we need, for instance, to define the reaction coordinate,$\epsilon$. NAMD method relies on the

continuous application of a dynamically adapted biasing force that compensates the current estimate of the free energy, thus virtually erasing the roughness of the free energy landscape as the system progresses along the reaction coordinate. [16] To reach this goal, the average force acting on, is evaluated from an unconstrained MD simulation:

## 2.4 State of the Art

### 2.4.1 Plant physiology

G–protein coupled receptors (GPCRs) are integral membrane receptor proteins. They interact intracellularly with heterotrimeric G–proteins, which are trimers composed by $\alpha$, $\beta$ and $\gamma$ subunits. GPCRs perceive extracellular stimuli by interacting with hormones, neurotransmitters, chemokines, lipid mediators, light, tastes and odorants. GPCRs convert the extracellular interactions into multiple intracellular signaling processes [18]. The structural flexibility associated with the pleiotropic functionality of GPCRs provides them with the capacity of activating multiple signaling cascades inside the cell, which affect numerous cellular behaviors [19]. This structural flexibility has led to propose that plant cells can promote or regulate several physiological processes by binding ligands such as hormones [20] to plant analog GPRCs, such as the G-protein coupled receptor 1, GCR1. Among these physiological processes are germination and flowering [20], root elongation [21, 22]. dormancy. and biotic and abiotic stress [9, 23] processes. Some of these processes have been associated to plant G–protein (GP) subunits, for example, the G$\alpha$ subunit, GPA1, has been associated with potentiation of seed germination [24], meanwhile Arabidopsis G$\beta$ subunit 1, AGB1, is central in root elongation control [21, 22], and flowers and fruits production [25]. GP subunits have also been implicated in abscisic acid (ABA) mediated processes, such as, dormancy, and biotic and abiotic stress response [26, 27, 28]. The GP complex is also related to growth control, cell proliferation, defense, movements of the stoma, channel regulation, sugar sensing and some hormonal responses [29].

It has been proposed that plants do not require any GPCR with guanine nucleotide exchange factor (GEF), as GCR1, to activate the G$\alpha$ subunit, because of the self-activating capacity of GPA1 [30]. However, authors such as Taddese propose that GCR1 is a GPCR homologue to class A and class B animal GPCRs [11], but no evidence has been found to support this function.

### 2.4.2 Plant stress signaling

Abscisic acid, ABA, and the components of the G-protein signaling pathway, promote plant dormancy [4, 20, 31], and are involved in abiotic [32, 28] and biotic regulation [9, 26, 33, 34]. ABA, and auxin mediated opening and closure stomata regulation, are crucial to reduce water loss and ensure plant survival, by maintaining hydration status during the drought. Evidence for negative crosstalk between auxin and ABA mediating stomatal opening supports processes mediated by GCR1 and GPA1, as mentioned by Cousson [35], where it is proposed that GCR1 positively interacts with GPA1 to activate guard cells. Aside from abiotic stress regulation, ABA and the G-protein signaling components are associated to biotic stress regulation. The transcriptome analyses on the *gcr1-5* mutant of *A. thaliana* showed that most of the differentially expressed genes, DEG, also belong to known G-protein regulated processes [9]. Therefore, the overlap between the genes/processes/responses of single and double mutants of GCR1 and GPA1 are the best genetic evidence in favor of their functional association [33], showing that, for the case of biotic and abiotic stress responses, DEGs such as DIN11, FMO1, MEE16, and PAD3 were reported as differentially regulated by both GPA1 and GCR1. Genes such as low temperature induced 78 (LTI78), plant defensin 2.5 (PDF2.5), ethylene response factor (ERF6), and several peroxidases and transcription factors are also involved in GPA1-GCR1 regulation. In general, the aforementioned genes are involved in abiotic stress responses to cold, heat, drought, salt etc., as well as in biotic stresses [33, 26, 26].

### 2.4.3 Membrane proteins, include GPCR

In plants, receptors like kinases, RLKs, and G-protein coupled receptor 1, GCR1, are the best-characterized receptor proteins. Because GPCRs are vital signaling molecules in diverse organisms, there is a growing interest in the identification of plant G protein-coupled receptors (GPCRs).

In addition to the crosstalk found regulating the opening and closing of stomata in guard cells, an important and interesting crosstalk between Brassinosteroids (BRs) and Gibberellins (GAs) receptors has been identified, not only promoting but also potentiating seed germination. BRs and GAs are growth-promoting phytohormones. Overlapping functions of BRs and GAs, support crosstalk between BR and GA signaling pathways in which both types of receptors, RLKs and GPCRs, seem to be participating. BRs are perceived by the plasma-membrane localized BR recep-

tor Brassinosteroid-Insensitive 1 (BRI1), a leucine-rich-repeat containing receptor-like kinase (LRR-RLK); GAs are perceived by the GA receptor gibberellin insensitive dwarf 1 (GID1) [36]. Ullah and coworkers propose that seed germination is regulated by many signals, since seeds bearing a GPA1 null mutation, *gpa1* mutants, are 100-fold less responsive to gibberellic acid, GA, and BR, but those seeds overexpressing GPA1 are at least a million-fold more responsive to GA, but yet still require GA for germination [24]. Similar to *gpa1, gcr1* seed mutants are less sensitive to GAs and BRs in promoting seed germination, while, seeds over-expressing GCR1, produce a loss of seed dormancy, suggesting that GCR1 may be a potential interaction node in GA- and BR-regulated seed germination, as is mentioned by Li and He [36]. Therefore, whether GPCRs exist in plants is a fundamental biological question.

### 2.4.4 GCR1 has all the structural characteristics found in animal GPCRs

Although the animal GPCR superfamily contains more than 1000 members, only until the late 1990s, two different research groups, Josefsson & Rask in 1997 and Plakidou-Dymock et al in 1998, independently identified a plant homolog of the GPCR superfamily, through a search in the database of expressed sequence tags (dbEST) [37, 38, 39]. These authors found several Arabidopsis sequences that predicted products with sequence similarity to GPCRs. In their work, they isolated the cDNA for the GCR1 gene and made a hydropathic analysis that indicated that GCR1 has a seven transmembrane-spanning domain. Analysis from membrane topology prediction algorithms support the thesis that GCR1 is indeed an analog of GPCRs [38, 40, 7, 8, 1]. Researchers from these groups, also found that GCR1 is a single-copy gene which shows 18–23% amino-acid identity (46–53% similarity) to, and good collinear alignment with, 7TM receptors from three different families. For example, GCR1 share a sequence identity with the Dictyostelium cAMP receptors [38], members of the Frizzled/Smoothened family of receptors, and also to the Methuselah-like proteins of Drosophila [40, 7]. Additional experiments to test whether GCR1 encodes a 7-TM receptor homolog revealed a structure characterized by an extracellular amino terminus, seven hydrophobic regions of 19–25 amino acids connected by alternating intracellular and extracellular hydrophilic loops, and a hydrophilic intracellular carboxyl terminus. The predicted protein sequence of GCR1 was determined computationally [41], and compared against the TMbase database [38]. A comparative sequence analysis between GCR1

and known sequences of GPCRs, concluded that the distribution of charged amino acids is consistent with the proposed structure of GCR1, with 24 of the 31 lysine or arginine residues predicted to be intracellular [38].

Plakidou-Dymock et al., and Josefsson-Rask also found a series of common characteristics among GCR1 and GPCRs, like the presence of cysteine, arginine, histidine, tryptophan, and aspartate, in highly conserved motifs on common sites of the GCR1 structure. This is very important not only to function but for the stability of GPCRs as well, since mutation of these are known to affect ligand binding and receptor expression. The authors also identified that analogous to several GPCRs, the putative GCR1 receptor has a potential N-linked glycosylation site [42], and multiple serines, and one threonine in the carboxyl terminus that may be sites of regulation [38, 7]. Last but not least, they aligned GCR1 against CAR1 and found that sequence similarity between these proteins extended across the whole of GCR1 and that the proteins had a similar arrangement of intracellular and extracellular loops and transmembrane domains [7].

After reviewing the structural characteristics of GCR1 that have led researchers to propose it as a GPCR analog in plants, we point to new evidence that involves GCR1 in the transduction mechanism through the GCR1-GP coupling, starting from its demonstrated participation in cell cycle modulation.

### 2.4.5 GCR1-mediated cell cycle modulation during seed germination and early seedling development.

Considering that GPA1 is self-activating [43, 44, 45] and that RGS1 is a direct regulator of GPA1 activity, and consequently an indirect regulator of the G$\beta\gamma$-dimer activity [46], and because GCR1 has been implicated in cell cycle regulation [47] of plants during seed dormancy and germination processes via ABA and GAs mediation, we review the potential involvement of GCR1 in the G-protein signaling pathway that regulates these two processes. We start by reviewing the existing body of work on specific overexpressors and (single, double, and triple) knock-out mutants of the genes *gcr1, gpa1, agb1*, and *rgs1* of seeds and early stages of plant development.

Although Plakidou-Dymock and co-workers [38] noted that the GCR1 gene was expressed at low levels, Colucci's group in 2002 was able to confirm the GCR1 expression in various organs of the plants [47, 48]. In their work, they also concluded that the level of expression of GCR1 varies among different organs, and through different stages of plant development. These observations led Colucci's group to conclude that GCR1 is a modulator of

the cell cycle and is expressed in a developmentally specific manner [47].

On the other hand, given that signaling through GPCRs is a highly conserved mechanism in animals and that GP along with GPCRs are key in the transduction of signals into the cytoplasm [49], different experiments have been carried out to evaluate the possible function of GCR1 in G-protein signaling in plant cells [48].

Based on the fact that GP is a key regulator of cell growth and differentiation in animal cells [50], and that over-expression in tobacco BY-2 cells of the Arabidopsis GPA1 gene leads to a premature advance of the cell cycle, while a GPA1 Arabidopsis null mutant (*gpa1*) had shown reduced cell division in its aerial parts [51], Colucci's group evaluated the possibility that the putative GCR1 could act as regulator of the cell cycle [47]. Thus, indirectly evaluating the mitotic index on synchronized transgenic Tobacco BY-2 cells over-expressing the GCR1 gene of *A. Thaliana* to show was consistently higher in GCR1-transformed cells than in wild type cells [48, 47].

Furthermore, Colluci and coworkers observed that not only the mitotic index was increased in GCR1-transformed BY-2 Cells but also in A. *thaliana* over-expressing the GCR1 gene, which stopped the dormancy while the expression of germination-associated genes was enhanced. One of the germination-associated genes found encodes the catalytic subunit of the serine/threonine phosphatase PP2A. These results suggested that GCR1 initiates germination processes in seeds that would otherwise be dormant [47].

The same group also proved that over-expression of GCR1 accelerates flowering, the production of siliques, and the expression of genes that control flowering such as LEAFY (LFY) compared to wild-type plants [47, 52]. The aforementioned processes are GA-activated. They propose, based on those results, that GCR1 transformants are more sensitive to GA and less sensitive to exogenous abscisic acid (ABA) than wt plants. This suggested that GA is a possible ligand to GCR1, that GCR1 plays a role in the regulation of the cell cycle, and that the developmental consequences of its over-expression are the result of a modulation of the cell cycle [47].

Other studies showed that over-expression of GPA1 led to excessive cell division in meristem and the initiation of additional meristems [51], which can be interpreted as if GPA1 were part of the signaling chain started by GCR1 [47].

In 2003, Apone and coworkers [48] provided support on the specific processes that indicate the cell cycle might be mediated by the interaction between GCR1 and GP. Independent experiments carried out on both the GCR1 and GPA1 over-expression showed an increase in the phosphatidylinositol specific phospholipase C (PI-PLC) activity, the inositol triphosphate

($IP_3$) and Diacylglycerol (DAG) levels, and DNA synthesis. Once they used PI-PLC specific inhibitors, the $IP_3$ production was suppressed and DNA synthesis stopped. This inhibitory effect was observed in wild-type cells as well. These results confirmed that both GCR1 as GPA1 would be involved in the same signaling pathway toward the effector activation, PI-PLC thus increasing the levels of the second messengers, $IP_3$ and DAG, and the subsequent thymidine incorporation required in DNA synthesis [48].

Another interesting finding made by Apone et al. [48] was that the activity of the kinases increased after the aphidicolin removal, and it was significantly higher in GCR1-over-expressing cells than in wild-type ones. As mentioned, there was an increase in the DAG levels, which might be related to an increase in kinases activity. It is known that in animal cells, DAG or phosphatidic acid, a metabolic product of DAG, produces the activation of several protein kinases with the subsequent activation of transcription factors [48, 53, 54]. This new evidence allowed the authors to confirm and expand the findings from prior work [47] carried out by Colucci et al. Consequently, Apone and collaborators suggest that GCR1 and GPA1 stimulate the same signal transduction pathway that can lead to DNA synthesis, and entry into the cell cycle through PI-PLC activation [48].

Colucci and collaborators showed the implications of GCR1 activating GPA1, regulation of cell growth, and cell differentiation through cell cycle activation [48, 47]. Consequently, Colucci and Apone groups propose that GCR1 activates GPA1 and both stimulate the same signal transduction pathway that leads to DNA synthesis, and entry into the cell cycle through PI-PLC activation. Since in 2002, it was still unknown that GPA1 was self-activating [43], it was incorrectly concluded that GPA1 had been activated by GCR1, just like in animal GPCRs.

To date, it is still unclear how GCR1 participates in the cell cycle modulation in plants, albeit additional evidence has emerged showing that GCR1 and GPA1 interact to promote the signal transduction pathway through the GP-GCR1 Coupling [20]. We describe those results next, in order to discuss the GCR1 role within this signal transduction mechanism.

### 2.4.6 GCR1-GPA1 physical interaction.

Given that until 2004 a direct interaction between GCR1 and GPA1 in plants had not been demonstrated, all of the studies were based on comparisons to animal model mechanisms in an attempt to explain the phenomena observed. Thus Pandey et al. [40] proposed that if GCR1 were to work as a GTP/GDP exchange factor (GEF), like a true GPCR, it should have a

physical interaction with GPA1 through some regions of the intracellular loops of GCR1. From this idea, Pandey et al. in 2004 were able to design novel strategies to demonstrate such interactions and the regions where these would take place. In an initial experiment, they used an in vitro binding assays from GPA1 fused with the GAL4 activation domain of yeast (*Saccharomyces cerevisiae*) (GPA1-GAD), and GCR1, followed by pull-down assays with anti-GAD antibodies. The results showed that Anti-GAD antibodies could precipitate GCR1 only in the presence of GPA1-GAD [40]. Next, they used a modified split-ubiquitin system to confirm the GCR1-GPA1 interaction and also showed that GPA1 requires a free GCR1 C-terminus for interaction. Through a series of modifications of this strategy, *i.e.*, using truncated constructs in split ubiquitin assays, the authors were able to identify GCR1 C-terminus, the presence of an intact ICL3 (intracellular loop 3) structure in GCR1, and some key amino acids between ICL2 and ICL3 that were also essential for its interaction with GPA1 [40]. In a later experiment from Pandey's group, they used *gcr1* mutant plants to further confirm the GCR1-GPA1 interaction. Using a variant of the first experiment that consisted in the insertion of full-length GCR1 cDNA fused with a FLAG epitope tag in gcr1 plants, the authors were able to detect GCR1 in the immunocomplex precipitated with anti-GPA1 antibodies, from GCR1:FLAG plants, but not in the gcr1-3 knockout proteins used as control [40].

From the confirmation that GCR1 and GPA1 proteins interact physically in plants, and the fact that ABA is one of the signals transduced by GPA1 [55, 56], Pandey and co-workers also carried out a series of experiments to evaluate the physiological response of GCR1 to ABA signaling. To do this, they evaluated and confirmed the presence of both GPA1 as a membrane-associated protein, and GCR1 as an integral protein of the cell membrane [40, 7, 57]. Because one of the classic responses regulated by ABA is the inhibition of root growth, and root protrusion is the paramount characteristic in seed germination, Pandey and co-workers demonstrated that this kind of inhibition was also present in gcr1 seedlings, which showed hypersensitivity to root growth inhibition by exogenous ABA, and also higher expression levels of a set of stress- and ABA-regulated genes, compared against wild-type plants. This demonstrated that gcr1 is implicated in seed dormancy [58] and gcr1 plants are also hypersensitive to ABA-mediated, root growth. An additional and interesting finding was the fact that *gcr1* mutant plants showed lower levels of GPA1 protein compared with wild-type plants [40].

In a similar study, Chen and collaborators [7] addressed the physiological function of GCR1, by using loss-of-function alleles of GCR1 (gcr1), GPA1 (gpa1), and AGB1 (agb1) plants. The study focused on evaluating the ge-

netic interaction among GCR1 and the G-protein subunits through the evaluation of leaf morphology and plant structure of single, double, and triple *gcr1, gpa1*, and *agb1* mutants. The author's proposal was that, if GCR1 were coupled with heterotrimeric GP, the *gcr1* plant mutants should share some or all phenotypes exhibited by loss-of-function of the G-protein subunit mutants. Nevertheless, gcr1 shows wild-type traits, while the *gpa1 and agb1* single mutants and the *gcr1-gpa1* and *agb1-gcr1* double and the *agb1-gcr1-gpa1* triple mutants showed a different phenotype. In consequence, the results of this study contradicted expectations regarding coupling between GCR1 and GP, and the notion that GCR1 might be a true GPCR [7].

On the other hand, based on accumulated findings showing that GCR1 and GPA1 proteins interact physically in plants [40], that GA and BR are associated with seed germination of GPA1 and *gpa1* over-expressing mutants, [47, 59] and from studies on germination and cell cycle regulation of GCR1 over-expressing plants [48, 47], Chen's group [7] identified both the GCR1-GPA1 genetic coupling, and the sensitiveness of single, double, and triple *gcr1, gpa1*, and *agb1* mutants seeds to GA and BR in seed germination. They observed that not only gpa1 but also gcr1 seeds have reduced sensitivities to GA and BR in seed germination, which provided evidence of the GCR1-GPA1 genetic coupling. Nevertheless, the reduced sensitivities to GA and BR by double and triple mutants, *gcr1-gpa1, agb1-gcr1,* and *agb1-gcr1-gpa1*, that showed additive or synergistic effects, seem to contradict the coupling between GCR1-GPA1 based on the idea that gcr1 and gpa1 or agb1 acted independently. On the other hand, one interesting finding was the fact that the sensitivities of the *gpa1-agb1* double mutant to GA and BR are the same as that of the *agb1* single mutant [7].

The aforementioned studies led Pandey and co-workers [58] (in 2006) to evaluate expression levels of GCR1, GPA1, and AGB1 genes during the germination and post-germination growth of Arabidopsis and to question if ABA could affect the expression of these genes. They concluded that the expression levels of these genes, during germination and post-germination growth, led to two different results depending on whether the seeds were stratified, or non-stratified. In the first case, only GCR1 and AGB1 genes were expressed early on. In the second case, GPA1 was detectable at earlier times. They repeated the experiments but this time using ABA-treated seeds during germination and post-germination growth. Comparing the results did not show any significant effect of ABA on the transcript level of these genes, moreover, ABA slightly increased the level of GPA1 protein at the post-germination stage [58].

In summary, while Chen and collaborators [7] showed that knockout

*gcr1, gpa1*, and *agb1* mutants display hypo sensitivity to GA and BR signaling, Pandey and co-workers [58] showed that single, double and triple *gcr1, gpa1*, and *agb1* mutants are hypersensitive to ABA and glucose (Glc). Nevertheless, the G-protein complex single mutants showed different germination rates when grown in the presence of ABA. The strongest hypersensitivity was observed for the *agb1* mutant. AGB1 regulates negatively to ABA signaling pathway. Interestingly, the *gcr1-gpa1* double mutant showed moderate hypersensitivity to ABA similar to the *gpa1* single mutants, which is in clear opposition to the results obtained by Chen's group [7]. This can only lead us to believe that GCR1, GPA1, and AGB1, all belong to the same signaling pathway activated by ABA and that the signaling pathway is not GA-mediated. In a similar result, ABA was found to be necessary for the Glc-induced inhibition of germination, both in wild-type and mutant seeds. Furthermore, it was also confirmed that the delay of germination by Glc requires endogenous ABA [7].

The role of each subunit of the GP is crucial to seed germination and root growth processes, and in order to relate these to the possible role of GCR1, we discuss in the next section the known role of each subunit, followed by that of the GP.

### 2.4.7 GCR1 implicated in the regulation of seed germination, root growth, and stomatal opening and closure during abiotic stress.

In 2006, Chen and co-workers [60] described that root growth rate combines: (i) the rate at which the root apical meristem (RAM) produces cell derivatives that are recruited into the distal root, and (ii) the rate at which these derivatives subsequently elongate. In an experiment, they saw that *agb1* and *gpa1* mutants had more and fewer lateral roots, respectively [60]. In their work, they indicated that the formation of a lateral root requires initiation of the division of one pericycle cell, and therefore, the number of lateral roots indicates the number of cell cycle entry events. Through the over-expression of GPA1 or AGB1 in *agb1* or *gpa1* mutant backgrounds, respectively, Chen and co-workers were able to determine the role of G-protein subunits in cell division in each cell type for both the rate of cell production in the primary RAM and the number of lateral roots. They also showed that increased root growth in both *agb1-2* single and *gpa1-4, agb1-2* double mutants was due to an increased cell production in the RAM and that the over-expression of AGB1 confers decreased cell production in the RAM. Thus, both results, loss- and gain-of-function in the wild-type background,

indicated that AGB1 is an attenuator of cell division in the primary root. On the other hand, no effect on primary root growth was observed when AGB1 was over-expressed in *gpa1* or *gpa1-agb1* mutant backgrounds, indicating that AGB1 action requires a functional GPA1. This suggests that the whole heterotrimeric complex would act by inhibiting cell division in the RAM. Through similar experiments, Chen and collaborators were able to demonstrate that activated GPA1 conferred an increase in cell production in the primary root in the presence of a functional AGB1 subunit and that AGB1 inhibits lateral root formation [60]. Therefore, the cell cycle, in each kind of cellular type, is modulated deferentially by GPA1, AGB1 ($G\beta\gamma$) dimer, and by GCR1.

The molecules ABA and Glucose (Glc) are of paramount importance in mediating 2 processes of the cell cycle regulation mechanism: (i) Deactivation of GTP-GPA1 subunit, and (ii) subsequent formation of the heterotrimer by promoting GDP-GPA1-AGB1 interaction. As was previously mentioned by Pandey's group [58], in the ABA medium, all mutant seedlings lacking the AGB1 subunit expressed very few lateral roots.

On the other hand, Pandey and co-workers had also evaluated the responsiveness of gcr1 stomata to ABA and S1P (Sphingosine-1-phosphate) in guard cells during drought and observed that gcr1 plants have an improved drought tolerance and lower rates of water loss than wt plants [40]. This was explained by considering that gcr1 guard cells were hypersensitive to both ABA and S1P in the inhibition of opening pre-closed stomata and the promotion of closure of pre-opened stomata, compared to wt plants. ABA promotes the inhibition of the opening of pre-closed stomata, through inhibition of inwardly rectifying $K^+$ channels. ABA also promotes the closure of opened stomata [61]. Given that gpa1 guard cells exhibited insensitivity to ABA with respect to the inhibition of opening of pre-closed stomata [55] and that GCR1 is expressed in guard cells of wild-type Arabidopsis plants, Pandey and co-workers conclude that in guard cells, GCR1 is a negative regulator of GPA1 in ABA and S1P signaling, and thus, GCR1 could work either as an S1P receptor or an ABA receptor, in guard cells [40], and from Chen's group results, maybe a possible negative regulator in the RAM [60] and seed germination [40, 58].

Although Pandey's group [58] in 2006 also noted an increase in the transcript level of some genes for the GPC mutants, each gene had its own distinctive expression profile. One special case was observed with the Rab18 gene, which was significantly more hypersensitive to ABA induction of gene expression in *agb1* mutants than the wild type, suggesting that not only GCR1 but also GPA1 and AGB1 are negative regulators of ABA induction

of gene expression during post-germination growth [58].

As we can see, GCR1 has been implicated in multiple ABA and GA-mediated processes. Since carotenoids have a similar structure to ABA, it is likely that GCR1 could be a blue light receptor.

### 2.4.8 Evidence of GCR1-GPA1 interaction in the blue light- and ABA-mediated synthesis of phenylpyruvate.

In 2006, Warpeha's group showed that the GCR1-GPA1 interaction has a specific role in the blue light-mediated synthesis of phenylpyruvate and then to phenylalanine (Phe), into the Shikimate metabolic pathway. This metabolic pathway uses two enzymatic activities, the chorismate mutase, which catalyzes the conversion of chorismate to prephenate, and the prephenate dehydratase 1 (PD1), which catalyzes the conversion of prephenate to phenylpyruvate. This pathway is activated by a single short pulse of blue light (BL), which triggers a signaling cascade that starts with the GPA1 activation mediated by GCR1 and the subsequent specific activation of PD1 to produce the accumulation of Phe. Although the physical interaction between GPA1 and PD1 had been proven using an in vitro pull-down procedure, and the fact that the BL activation of PD1 activity is lost in the *gpa1* mutant, the proposed mechanism was confirmed because neither Phe nor Tyr were accumulated after the BL irradiation of *pd1, gpa1*, and *gcr1* seedling single mutants [62]. Additional evidence suggests that ABA can act through the same pathway, i.e, ABA could act through GCR1, GPA1, and PD1 to form all or part of a signal transduction mechanism, which would prove that PD1 is an effector of GPA1 [62, 63].

Warpeha's group [62, 63] identified that in addition to the physical interaction between GPA1 and PD1, the presence of a single pulse of low-fluence BL induces the expression of Lhcb gene due to the interaction between GPA1 and PRN1, a cupin-fold superfamily member. These findings confirm that GCR1, GPA1, and PD1 or PRN1 form the whole or a part of a BL- or ABA-induced signal transduction mechanism. Given that ABA not only derives from a carotenoid but also has a cyclic head similar to that of the carotenoids that, along with BL, affects the production of Phe and their derivatives and that interacts with GCR1, the authors propose GCR1 could act as a single receptor for both, ABA and BL [62, 63].

After evaluating the above results, there are many doubts about the actual role of GCR1, and questions arise as to whether GCR1 is not a GPCR because all the information points to GCR1 performing a GPCR function.

### 2.4.9 What we know so far

To date, GCR1 has no proven ligand. Hence its role remains largely unknown. This has led some researchers, as Urano and co-workers in 2012 and 2013 [45, 44] to question the need for GPCRs in plants, given that the GP in plants is self-activating [43], *i.e*, G-$\alpha$ subunit would not need a guanine nucleotide exchange factor (GEF) like GCR1. Based on the premise that GCR1 would not be necessary to activate the GP, the fact that GCR1 has no known ligand, and the indirect evidence about GCR1-GP interaction, Urano's group proposed that RGS1 be the only transmembrane protein required to regulate GP activity. Thus, the GTPase activity, GAP, on the RGS1 protein promotes G-$\alpha$, GPA1, inactivation, and subsequent formation of the heterotrimer.

On the other hand, there is still sufficient biological evidence to confirm that GCR1 does function as a receptor and has homology to class A, class B, and class E GPCRs [40, 1]. A novel study to address these contradictory results involved helix alignment and fold recognition methods. Specifically, the approach by Taddese et al in 2013, combined transmembrane structure prediction and sequence analysis with fold recognition methods [1]. The results of this work indicate that GCR1 has strong homology with several classes of the GPCR families, being GCR1 the only candidate having a GPCR folding, by means of the analysis made with the servers I-TASSER, LOMETS, HHpred, FUGUE, and Phyre. Although in studies made by Urano and Jones in 2013 GCR1 shows limited homology to TM3 and TM4 of class E GPCRs [44], both, earlier and new reports, identified similarities to class A and class B GPCRs [38]. These similarities showed considerable homology to all eight helices of class A and class B GPCRs. Additionally, several motifs that were identified in GCR1 also were detected in class A, class B, and class E GPCRs, including the disulfide bond between TM3 and ECL2 [1]. The authors further confirmed the findings made by Pandey and Assmann in 2004, where they showed that GCR1 possesses motifs that would facilitate the GCR1-GPA1 interaction [40, 1]. These results were further proven through a variability analysis that showed that the GCR1 homologs have the same folding as class A and class B GPCRs. One additional conclusion of this study is that GCR1 homologs are generally closer to class A, class B, and class F GPCRs than these are to each other[1].

If the participation of GCR1 in the aforementioned processes has been identified through the results of genetic and biochemical experiments, as well as a study that manages to co-precipitate GCR1 with the G protein, GP, what kind of protein is GCR1? is it a receptor? or just a transmembrane

protein?

Current models of the signal transduction mechanism through the G protein complex, GPC, involve three different types of proteins: G protein-coupled receptors (GPCR), heterotrimeric G proteins (GP), and RGS (regulator of the G protein signaling) proteins. In animal cells, there are a variety of families of both GPCRs and GPs; each GPCR can activate one or more types of GPs and initiate a specific signaling pathway.

In plants, three important findings observed in the signal transduction mechanism through the GP-GCR1 coupling in Arabidopsis thaliana, provide evidence to propose a different model for this mechanism: (i) the existence of GCR1, a GPCR homolog in plants [37, 38] (ii) the self-activating capacity of GPA1 [64, 43, 45, 44] and (iii) the presence of RGS1, a different type of RGS protein [46, 60]. In particular, RGS1 is a hybrid protein that possesses a 7-TM domain and an RGS-box; the intracellular RGS-box has a well-known GTPasa activation, GAP, enzymatic activity, as Urano and Jones describe it [44].

So, do plants have GPCRs? Findings (ii) and (iii) have raised doubts about whether GCR1 is a true GPCR, inasmuch GCR1 does not have the nucleotide exchange function like animal GPCRs [44]. For this reason, Urano and Jones [44] propose RGS1 as the only regulatory protein of the G$\alpha$ subunit (and in consequence of the G$\beta\gamma$ dimer). These authors also mention that since loss-of-function mutations in RGS1 do not confer constitutive GPA1 signaling, the GPA1 regulatory process must be more complex [44].

This evidence raises a fundamental question: How can a single protein, such as RGS1, achieve G protein regulation? If after the hydrolysis of GTP to GDP on GPA1, mediated by RGS1, and the formation of the heterotrimer, RGS1 must release GP to continue its inactivating function on other active GPA1. Once released, the GP undergoes a fast self-activation since the GTP to GDP hydrolysis on GPA1 is the rate-limiting step, as demonstrated by Johnston [43]. Would the release of GP, followed by its fast self-activation, lead to a futile cycle?

In addition to the possibility of a futile cycle being generated through the RGS1-GP regulation, and considering that loss-of-function mutations in RGS1 do not lead to constitutive GPA1 signaling [43], we propose a synergistic interaction between GCR1 and RGS1. Specifically, it has been shown that RGS1 effectively inactivates GPA1 by promoting GTP hydrolysis. Once the GP is inactivated, its subsequent binding with the ABA-GCR1 complex promotes GP sequestration, thereby preventing GP self-activation and maintaining GP in an inactive state. On the other hand, alterations in environmental conditions can influence hormone concentrations, leading to

an increase in GA1 levels and a reduction in ABA concentrations in the extracellular region of plant cells. This fluctuation in GA1 concentration may result in competition between ABA and GA1 for the binding site on GCR1. When GA1 binds to GCR1, it could promote the release of the GP, allowing it to be self-activated or undergo GDP/GTP exchange and subsequent activation, mediated by the GA1-GCR1 complex.

# Chapter 3

# Problem statement

To date, no plant GPCR-homologue has been identified. The protein GCR1 is considered the best candidate to fill the role of GPCRs in plants [11, 65]. Although, as mentioned previously, there is a vast amount of biological evidence that involves GCR1 in several physiological processes, some of which can be mediated by promoting DNA synthesis [66], Urano [29] claims that since plant G proteins have a self-activating activity, they do not need a GPCR with guanine nucleotide exchange factor (GEF) to activate to G$\alpha$-subunit. To clarify this controversy, Taddese, in one of the most complete studies to date [11], carried out a structural bio-informatics study involving fold recognition methods to demonstrate that GCR1 has the full set of characteristics of a true GPCR, *i.e.* (i) variability comparisons provided evidence that GCR1 have the GPCR fold, (ii) helix-alignment of GCR1 to class A, class B, and class F GPCRs has shown that GCR1 is closer to class A and/or class B GPCRs than class A, class B, or class F GPCRs are to each other and, (iii) they identified motifs that are common to GCR1, class A, B, and E GPCRs.

Since there is genetic and biological evidence that involves GCR1 in regulatory processes mediated by signaling pathways through GP, such as dormancy, biotic and abiotic stresses regulation, germination and plant development, among others, it is of paramount importance to demonstrate or refute the GCR1 implication, and its specific role in such a process. Unfortunately, there are two significant obstacles in that way, where the biggest one is the lack of an X-Ray or cryo-EM 3D atomistic structure for GCR1, along with the fact that no one ligand has been proven to interact with it, becoming its role remains inconclusive, and led us to address the problem from a structure prediction.

To this end, we use the GEnSeMBLE approach [19], which integrates the homology approach to structure prediction with MD for sampling the potential energy surface of the GPCR. In combination, they offer a cost-effective path to GPCR activation and function analysis, as will be described throughout this paper. Goddard and coworkers [67] have developed and validated that the conformational flexibility associated with GPCRs can be targeted by the structure prediction method GEnSeMBLE, which samples and identifies low energy (physiologically important) conformations of a GPCR.

# Chapter 4

# Objectives

## 4.1 General objective

Using bioinformatics, molecular docking, and first-principles molecular dynamics to determine, at the structural and mechanistic level, the role of GCR1 as a GPCR analog in A. thaliana, the function of the GA1-GCR1 and ABA-GCR1 transmembrane complexes in the regulation of the G-protein signaling pathway, and their relationship with processes of germination, dormancy and stress in plant cells.

### 4.1.1 Specific objectives

- Predict the GCR1 3-D from the 1-D structure using first-principles modeling and simulations.
- Evaluate the preferential ligation to extracellular sites of GCR1 for ABA and GA1.
- Evaluate the preferential ligation to intracellular sites of the complex ligand-GCR1 with GP.
- Describing the GCR1 mechanisms involved in ABA- and GA-mediated signaling to regulate GP activation or inactivation.
- Contribute to existing state-of-the-art methods for predicting membrane protein structure in plants (THIS MIGHT INCLUDE CHANGES/ADAPTATIOS to GEnSeMBLE)

# Chapter 5

# Methodology

G-Protein coupled receptors (GPCRs) are transmembrane proteins embedded in the cellular membrane's lipid bilayer. These proteins have seven transmembrane (TM) $\alpha$-helices that are connected by three intracellular loops (ICL) and three extracellular loops (ECL), an initial N-terminus chain, and a final C-terminus chain. In animal or plant cells, the lipid bilayer environment of the plasma membrane provides specific boundary conditions for the structures of GPCRs and how they fold. Then, it is important to define a frame of reference (relative to the plasma membrane) to establish a set of coordinates that specifically describe the orientation of each of the seven $\alpha$-helices, in the GPCR TM bundle. Accordingly, a GPCR structure can then be characterized by the six orientation parameters of the seven $\alpha$-helices relative to the middle of the membrane. The central plane of the lipid bilayer, which corresponds to the $z = 0$ plane (the hydrophobic plane), and the normal (perpendicular) vector to this plane define an appropriate coordinate system to describe the GPCR. The hydrophobic center (HPC) of each TM helix is the helix's residue that crosses the hydrophobic plane. The origin of the coordinate system is given by the HPC of the third TM helix, with $x$ direction given by the vector from the origin to the HPC of the second TM helix. Due to the TM helices' structural rigidity, a helix's backbone can be modeled as a rigid body with a center of mass and a moment of inertia tensor. The cylindrical shape of an $\alpha$-helix approaches a prolate symmetric top with a principal (helical) axis given by the vector along the least moment of inertia.

Meanwhile, the ECLs, the ICLs, and the termini (N and C) are modeled as flexible chains. The conformational state of each TM helix is described by six coordinates: two coordinates for the position of the helix with respect

to the origin of coordinates ($x$ and $y$), one coordinate for the position of the helix center of mass with respect to the hydrophobic plane ($h$), and three Euler angles defining the orientation of the helical axis ($\theta$ and $\phi$) and an angle ($\eta$) defining the rotation of the helix above the helical axis. The ($\eta$) angle requires a unique definition of the helical axis to account for the reality of bent helices caused mainly by Proline residues in the middle of TM helices. These coordinates are exemplified in figure 5.2.

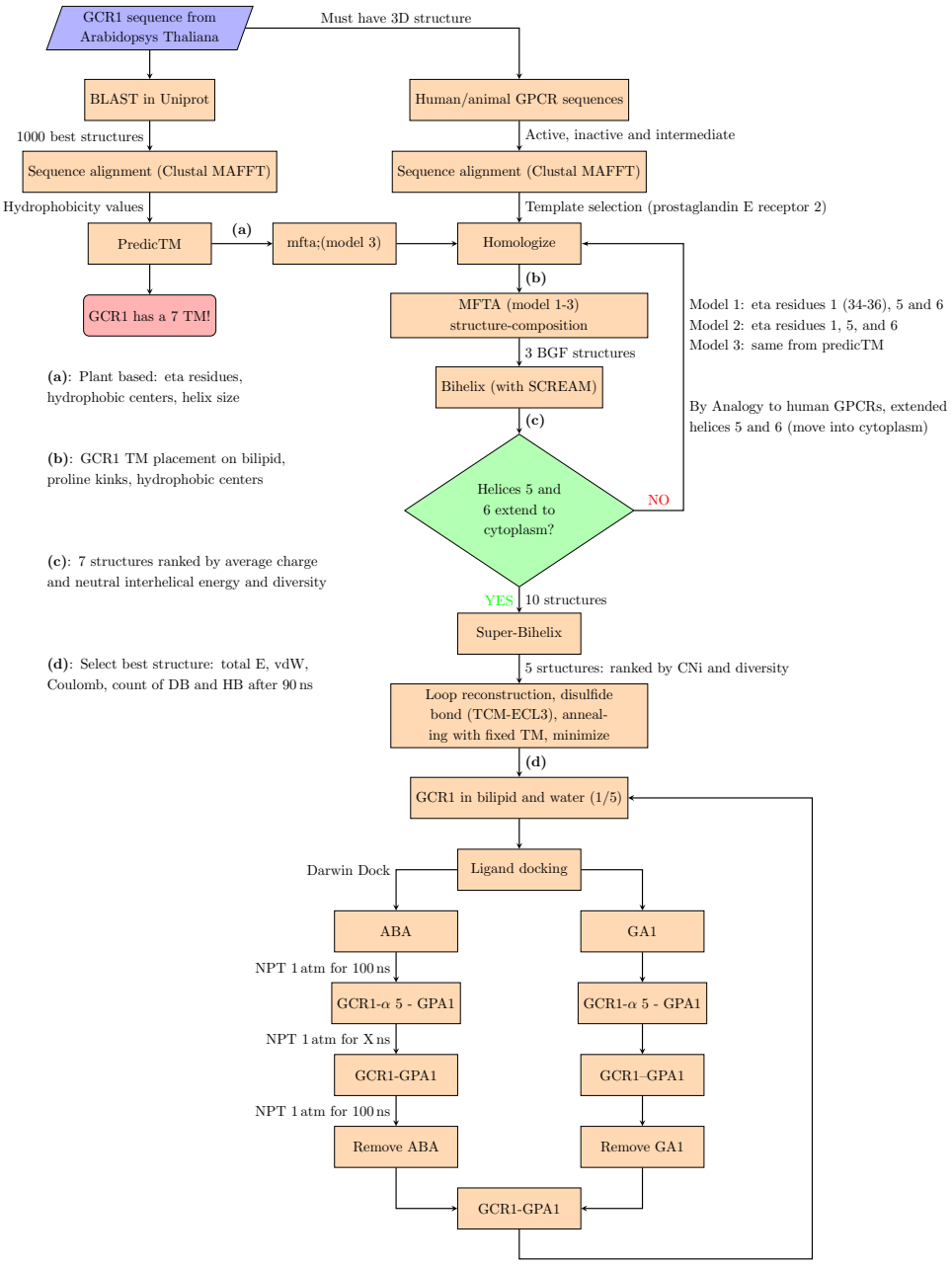


Figure 5.1: The GEnSeMBLE approach is depicted and combined with MD simulation as it was carried out during this research. Each step briefly describes the method used and its goal.

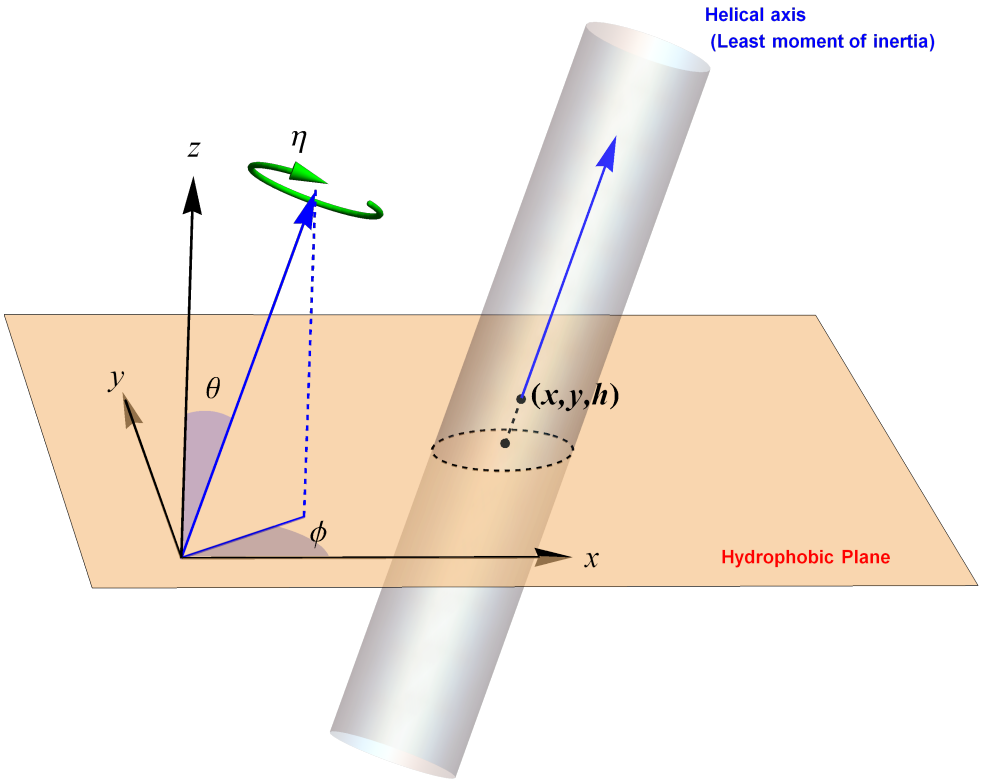


Figure 5.2: Coordinate system for an $\alpha$-helix. The hydrophobic plane, $z = 0$, is displayed in light orange color. The helix crosses the hydrophobic plane at the $(x, y)$ position, the distance from the hydrophobic plane to the hydrophobic center id given by $h$. The orientation of the helix id given by the blue vector. The orientation vector is specified by the angles $\theta$ and $\phi$. The rotation of the helix above the least moment of inertia gives the angle $\eta$.

The hydrophobic plane is the central plane of the lipid bilayer, and the hydrophobic center (HPC) residue of each helix is the residue that crosses

the hydrophobic plane. We define a Cartesian system of coordinates with origin at the HPC of TM3, the $x$ axis is defined from the origin to the HPC of TM2 [68], the hydrophobic plane defines the $z = 0$ plane, see figure 5.2. Each of the seven helices is modeled by the heavy atoms of the backbone with a moment of inertia. The use of the moment of inertia account for the bending of the helices caused mainly by Proline residues within the membrane region of the helix. The $x$ and $y$ positions of each helix are obtained from alignment with other GPCRs with known structures embedded in the lipid bilayer. The HPC of each helix, $h$, can be obtained in different ways: the residue at the raw mid of the helix, the residue at the mid of the cap helix, the residue at the mid area of the hydrophobicity profile of the helix, the residue at the buoyancy center of the helix, or by alignment with the template. The minor moment of inertia of each helix defines a vector, the direction of this vector is specified by two orientation angles, $\theta$ and $\phi$. Finally, the rotation of the helix on its own minor moment of inertia axis defines the angle $\eta$.

The helical interaction energy of class A GPCRs is approximated by 12 terms due to pairwise interaction between neighboring helices [69]. The interaction energy between two helices depends on six angles, $\theta$, $\phi$, and $\eta$ for each helix. A computational efficient method to obtain the minimum energy of the TM bundle has been designed by Goddard and collaborators [69, 68]. These methods, discretize the angles $\theta$, $\phi$, and $\eta$ optimizing a sampling space of about $10^{10}$ conformations, and use the SCREAM [70] method to optimize the position of the helices side chains.

The starting TM bundle is obtained from a homology GPCR. To find this homology GPCR it is necessary to identify the residues which belong to each helix from the primary structure of GCR1. This is done first by a BLAST search of similar sequences to the target, GCR1. These similar sequences are aligned to obtain an average hydrophobicity profile for GCR1, from which the seven TM residues are identified. A new alignment with animal GPCRs with known structure is needed to obtain a template structure that gives the starting TM bundle.

The possible templates are selected from https://gpcrdb.org/structure/, a data-base containing active, inactive, and/or intermediate X-ray structures, thus as from accurately predicted structures of animal and human GPCRs.

## 5.1 Sequence annotation

Analyses performed to identify and align homologous sequences were performed using the public BLAST service at NCBI (www.ncbi.nlm.nih.gov/BLAST).

## 5.2 Homologize procedure.

The Homologize helix procedure is performed in two steps. [71] In the first step, the sequence alignment, we carried out a condensed alignment of the target GCR1 to animal-GPCRs with known crystal/NMR or predicted structures. This step gives possible templates with the greatest homology to GCR1.

In the second step, homology helix shapes, we mutate the 3D structures of each helix from the templates to the amino acid sequence of GCR1. For this step, we downloaded from https://gpcrdb.org/structure/ the selected pdb files for each of the templates, and generated their corresponding GCR1-homologized bgf and mfta files. Since the pdb files from the template crystal structures downloaded do not provide absolute membrane orientation of GPCRs, we perform the alignment of each pdb template with the hBeta2_2rh1_opm.pdb, from a OPM database protein (Orientation of Proteins in Membrane database) followed by the homologyzed step, which mutates the 3D structures from the templates to the amino acid sequence of GCR1. This procedure also transfers the Proline kinks from the templates to the mutated 3D GCR1-structures. These proposed starting 3D GCR1-structures will be sampled in the BiHelix step. [69] Due to the fact that we are focused on the TM helices description, we deleted the intra and extracellular loops.

## 5.3 Conformational analysis

### 5.3.1 BiHelix sampling

The BiHelix method predicts the lowest pairwise energy conformation for neighbor helices. The Bihelix sampling is done on the $\eta$-angles, for all 12 nearest interacting helix pairs, resulting in $12 \times (12^2) = 1728$ energy conformations (using a 30° grid over full 360°). For each conformation tested, the side chain are optimized using the SCREAM method. [70] SCREAM uses a library of residue conformations with a conformational root mean square (CRMS) diversity between 0.4 Å to 1.6 Å, plus a Monte Carlo sampling which

uses full valence, and hydrogen bond and electrostatic interactions from an updated DREIDING force field (FF) [72] (referred to as DREIDING-III or D3FF), which is modified with a flat bottom van der Waals (vdW) potentials to reduce the penalty for contacts that are slightly too short, while retaining the normal attractive interactions at full strength. Finally, the two-helix pair, SCREAM optimized, is minimized for 10 steps using a conjugate gradient (CG) method. [69]

The 1728 energy conformations can be combined to estimate the energy of all possible $12^7$ conformations, which means we have an approximate energy for almost 36 million of plausible conformations. Each of the 1728 energies corresponds to a specific helix pair $i$-$j$, for a specific combination of $\eta_i$ and $\eta_j$ and is reported in the form of three energy components:

1. $e_{ij}^{inter}(\eta_i, \eta_j)$: This component describes the total interaction energy between the helices. It is calculated by subtracting the internal energy of individual helices from the total energy of the two interacting helices and captures sidechain-sidechain, sidechain-backbone, backbone-backbone interactions across the two helices.

2. $e_{ij}^{i,intra}(\eta_i, \eta_j)$: This component refers to the intra-helical energy of helix $i$ while it is interacting with helix $j$ in the $i$-$j$ pair.

3. $e_{ij}^{j,intra}(\eta_i, \eta_j)$: This component refers to the intra-helical energy of helix $j$ while it is interacting with helix $i$ in the $i$-$j$ pair.

For more detailed information refer to BiHelix article. [69]

The interhelical component of the energy is additive and can be summed to give the total interhelical energy of a seven-helix bundle. The intrahelical component of helix 3, for example, in the 2-3 helix pair will in general be different from that in the 3-4 helix pair. To accommodate this, the intrahelical energy of a helix in the seven-helix bundle is approximated as the average of that energy from all helix pairs involving that helix. Using this "mean-field" approximation, the energy of the ≈35 million conformations for the seven-helix bundle can be estimated. [69]

The resulting energies, obtained from single point energy calculations, after each modification of the rotational $\eta$-angle, are reordered to provide a new thermodynamic ranking of the structures from BiHelix, and the resulting lowest energy conformations can be analyzed in the context of the conformational space that was sampled, that, for this case is the $\eta$-angle for all the seven helices. This same principle is used for the next step, the SuperBiHelix sampling, but for that case, the conformational space sampled will be $\theta$- and $\phi$-angles, as can be seen in the next section.

### 5.3.2 SuperBiHelix sampling

During this sampling, the total subset of orientation parameters aforementioned were sampled over the starting template-based structures.

Since the BiHelix procedure optimizes the energy between helices by rigid rotations, [73] now, we have to optimize the tilts, $\theta$-angle, which is perpendicular to the $z$-axis of the membrane, and the azimuthal angle $\phi$ by rotations of 15° simultaneously with the rotation of $\eta$-angle. So, in SuperBiHelix procedure, the best combinations of rotations from BiHelix are followed by simultaneous changes in all three angles. We then output the 2000 best energy structures from this procedure, and similarly to the BiHelix case, the side for the SuperBiHelix step are also optimized using SCREAM, and the structure is minimized for 10 steps. [68] This energy ranking is more accurate than the obtained from BiHelix because all seven helices are present instead of just two, resulting in the ensemble of low-energy conformations most likely to play a role in the binding of ligands and activation of the GPCR.

## 5.4 Addition of extracellular and intracellular loops

We used the Maestro software [74] to perform a structural alignment of the 7TM bundle, obtained from SuperBiHelix, to the selected template. Then, we transferred the 3D loops from the selected template conformation to the target 7TM bundle. Next, we carried out 10 cycles of simulated annealing, keeping the 7TM region fixed, and followed this by a single minimization step, using the conjugated gradient method. From each cycle, the lowest energy structure was selected.

## 5.5 Thermodynamic analysis

### 5.5.1 Molecular dynamics simulations

The reconstructed GCR1 structures and their complexes with ligands in a periodic box were subject to full-solvent, full-lipid, and at physiological pH, to molecular dynamics simulation, using the program NAMD, which works with AMBER and CHARMM potential functions, parameters, and file formats. [16]

Previous to the equilibration step, we minimized the whole system by 10000 steps, in time steps of 1 fs.

The GCR1 equilibration was performed using a NPT ensemble, with a Langevin thermostat to keep the temperature at 310 K, and 1 atm pressure, in time steps of 1 fs.

## 5.6 Hormone docking and thermodynamic analysis

### 5.6.1 Docking ligands

We used the DarwinDock [75] complete sampling methodology to predict the binding site for the ligands. Here, we used just the top configuration from SuperBiHelix, where we protonated histidine, HSP-88(3.32), because its ability to form a salt bridge to the ligand. The DarwinDock procedure is as follows.

To find the optimal regions to bind the ligands on the equilibrated structure of GCR1 selected from SuperBiHelix, we replaced the six bulky nonpolar residues (I, L, V, F, Y, and W) of GCR1 with alanine (we call this step "alanization"). After this, we constructed $10 \times 10 \times 10$ boxes over the external surface of GCR1 and generated spheres that cover the empty region available for the ligand using the Sphgen procedure in the DOCK4.0 suite of programs [76, 77]. Then, we combined these spheres from several cubes for docking ligand poses. Next, we use Dock6 [76, 77] to generate 5000 poses (without energy evaluation), which we collected into Voronoi families based on RMSD less than 2.3 Å [78]. This procedure was repeated until the number of Voronoi families added was less than 5%. For the final binding site of ABA this led to 45000 poses in 2200 Voronoi families. For the final binding site of GA1 this led to 45000 poses in 2200 Voronoi families. The next step is to evaluate the interaction energy of each Voronoi family head with the GCR1. Then, for the best 10% we evaluated the interaction energy of all daughter poses with the GCR1, and selected the top 100 based on interhelical charged neutral energies, from which we selected the top 10 structures for further examination. For each one of these top 10 structures, we dealanized the protein side chains (converting back to the original I, L, V, F, Y, and W) and optimized the side chains using SCREAM.

The above procedure was done for each of the four rotamers of the ligand ABA and the two rotamers of the ligand GA1 generated using Maestro. Then, we combined the top 10 from each of the ABA and GA1 rotamers to select the best 10 overall for each ligand, as listed in Tables 6.15 and 6.18, respectively. The best ABA and GA1 docked structures based on cavity

and snapBE energies were selected for further steps, after pharmacophore analysis was performed on each of both top ten structures.

### 5.6.2 Molecular dynamics simulations

The newly constructed GCR1 structures with docked ligands were inserted into a periodic box containing the full-lipid bilayer with full-solvent at physiological pH. Then we carried out molecular dynamics (MD) simulation, using the NAMD program with AMBER force field [16].

Each GCR1 structure from the above steps (with and without docked ligands) was inserted into a periodic box containing the 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) lipid bilayer, with explicit water solvent and salt ($Na^+$ and $Cl^-$ ions) at physiological pH. (This step also balanced any net charge from the protein-ligand systems.) To eliminate steric clashes between these components, we energy minimized for 5000 conjugate gradient steps while keeping the protein fixed. The minimized systems were then equilibrated for 500 ps (with 1 fs steps) at 310 K, again keeping the protein fixed. To relax the side chains of the proteins, we next performed new minimizations and MD at 310 K, but this time keeping only the protein backbone fixed. Next, for all five systems, we carried out 60 ns of MD at 310 K while fixing only the 1,4-hydrogen bonds, angles, and dihedrals. Then, we performed MD with protein-ligand fixed for 60 ns, using time steps of 1 fs with an NPT ensemble at 310 K, and 1 atm of pressure. Finally, we removed the constraints between the protein and the ligand, and performed MD for 60 ns, using time steps of 1 fs with an NPT ensemble at 310 K and 1 atm of pressure, with a Nosé-Hoover Langevin piston pressure control method [16].

### 5.6.3 GCR1-G-protein coupling and thermodynamic analysis

We used the ionic lock K49-K115 found in GCR1. This ionic lock is the analog to the highly conserved ionic lock found in class A GPCRs, between D/E3.39 and H6.40, occurring with 94–98 % conservation among the human Opioid Receptors, hORs [79]. These authors observed that GPA1 is bound to the highly conserved ionic lock of the animal GPCR. In analogy, we bound the terminal carboxylate group on L383 of GPA1 to the ionic lock K49-K115 of GCR1.

The newly constructed full complexes ABA-GCR1-GPA1-GDP-$Mg^{2+}$ and GA1-GCR1-GPA1-GDP- $Mg^{2+}$ were inserted into a periodic box containing the full-lipid bilayer with full-solvent at physiological pH. Then we

carried out molecular dynamics (MD) simulation, using the NAMD program with AMBER force field [16].

Previous to the equilibration step, we minimized the whole system for 10,000 steps. In this minimization, we kept the proteins and ligands fixed while doing MD on the lipid and solvent to equilibrate them. Subsequently, the restrictions on the protein-ligand complex were removed, and again, a system-wide equilibration procedure was carried out. This equilibration procedure used 1 fs time steps and spanned 154 ns in the case of GA1 and 110 ns in the case of ABA. It was performed using an NPT ensemble, with a Nosé-Hoover Langevin piston pressure control method and a Langevin thermostat for physiological conditions: 310 K thermostat and a 1 atm pressure barostat [16].

These equilibrated systems were used as initial conditions of metadynamics simulations to obtain the free energy landscape for both ABA-GCR1-GPA1-GDP-$Mg^{2+}$ and GA1-GCR1-GPA1-GDP-$Mg^{2+}$.

### 5.6.4 Free energy landscape from Metadynamics simulations

The activation of GPA1 is promoted by the GDP/GTP exchange. The guanosine nucleotides, GDP or GTP, are located inside GPA1 between the Ras and Helical domains of this protein. The GDP/GTP exchange is facilitated by a spatial separation between the Ras and Helical domains of GPA1. Therefore we use the distance between the centers of mass of the Ras and Helical domains as the collective variable of the metadynamics simulations. We used metadynamics simulations to obtain the potential of mean force landscape (PMF) of the complexes, ABA-GCR1-GPA1-GDP-$Mg^{2+}$ and GA1-GCR1-GPA1-GDP-$Mg^{2+}$, as a function of the collective variable.

We used a set of metadyanmics parameters employed previously to simulate the Ras/Helical opening in animal GPA1. These parameters were a Gaussian width and height of of 1.0 Åand 0.01 kcal $mol^{-1}$ respectively, and a frequency for Gaussian hill deposition of 500. The minimum and maximum values that the collective variable will get during the run were set to 26 and 50 nm respectively. The final time of the metadynamics simulation were of 1.5 $\mu$s.

# Chapter 6

# Results and discussion

Since the identification of GCR1 as a true GPCR in plants is relevant from a biological and agronomic point of view, knowing the three-dimensional structure of GCR1 would provide important information to determine whether or not it is a true GPCR. In that way, Taddese and collaborators have used structural bioinformatics involving fold recognition methods to support the hypothesis of GCR1 as a true GPCR homolog in plants [11]. As mentioned in the methods section, Goddard and collaborators have developed efficient computational methods to determine the tertiary structure of animal GPCRs from the amino acid sequence [19, 68]. We have applied these methods to elucidate the tertiary structure of GCR1. Furthermore, the GPCR function of this 3D structure is tested by analyzing the conformational changes undergone by GCR1 and GPA1 due to the extracellular docking of GCR1 with the ligands GA1 and ABA.

## 6.1 GCR1 3D-structure prediction from sequence annotation

This section describes the use of the GEnSeMBLE methodology and full solvent MD simulations to obtain the tertiary structure of GCR1. The first step of this methodology is the prediction of the TM regions of GCR1 from its amino acid sequence and alignment with a series of similar plant proteins found by the BLAST search as described below.

### 6.1.1 Establishing protein transmembrane regions from primary sequence

The residue sequence of GCR1 for *Arabidopsis thaliana* [38] is available in the Uniprot database [80]. A BLAST search gives approximately one thousand proteins with a similar sequence to GCR1 down to 5%. The BLAST search returned a fasta file, O04714.fta, containing the sequence of GCR1 along with a series of similar protein sequences from the Uniprot database [80].

To identify the TM regions, we performed a multiple sequence alignment of those similar sequences in the O04714.fta file, using the MAFFT method and the E-INS-I algorithm [81], followed by substituting the amino acid residues with their corresponding hydrophobicity values, based on the Wimley-White scale [82], see figure 6.1. The raw helix regions are given by the residues with positive hydrophobicity values. These raw helix predictions are then "capped" by extending by 1 to 5 residues on both N and C-termini until a helix breaker residue (charged residues, prolines, and glycines) is found. These results are gathered in the file O04714.mfta shown in figure 6.2.

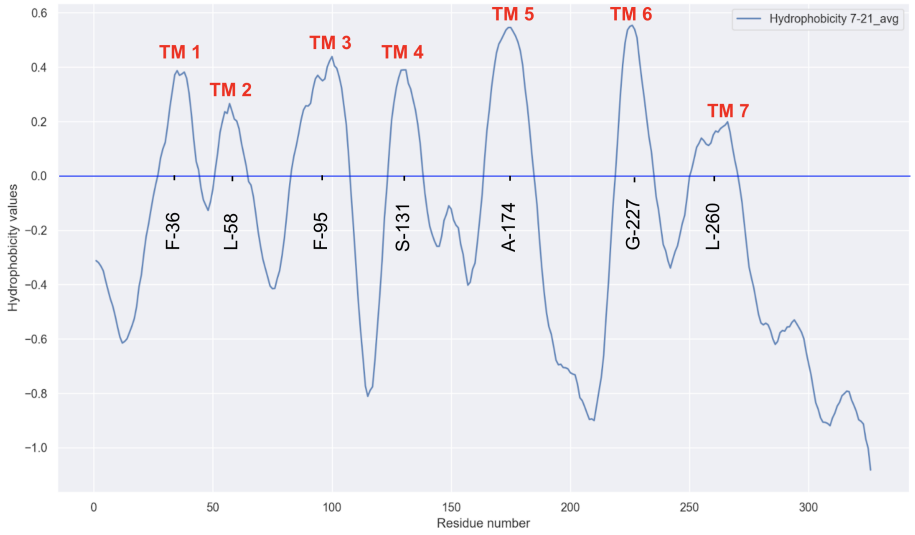


Figure 6.1: hydrophobicity profile based on the Wimley-White scale. Residues with hydrophobicity values above 0 are considered as no polar, and consequently, are the hydrophobic residues belonging to transmembrane regions, TM regions

```
>P1;sp|O04714|GCR1_ARATH G-protein coupled receptor 1
OS=Arabidopsis thaliana OX=3702 GN=GCR1 PE=1 SV=1
MSAVLTAGGGLTAGDRSIITAINTGASSLSFVGSAFIVLCYCLFKELRKFSFKLVFYLAL
SDMLCSFFLIVGDPSKGFICYAQGYTTHFFCVASFLWTTTIAFTLHRTVVKHKTDVEDLE
AMFHLYVWGTSLVVTVIRSFGNNHSHLGPWCWTQTGLKGKAVHFLTFYAPLWGAILYNGF
TYFQVIRMLRNARRMAVGMSDRVDQFDNRAELKVLNRWGYYPLILIGSWAFGTINRIHDF
IEPGHKIFWLSVLDVGTAALMGLFNSIAYGFNSSVRRAIHERLELFLPERLYRWLPSNFR
PKNHLILHQQQQQRSEMVSLKTEDQQ
*  1tm   27   44   25   44    X    X    X    X    F   36    A
*  2tm   51   64   51   68    X    X    X    X    D   62    A
*  3tm   83  107   77  111    X    X    X    X    T   87    A
*  4tm  124  138  120  141    X    X    X    X    W  128    A
*  5tm  164  184  163  187    X    X    X    X    P  170    A
*  6tm  219  235  217  236    X    X    X    X    Y  220    A
*  7tm  251  270  246  276    X    X    X    X    S  266    A
*  1hpc    35.50    34.50    35.00    36.07    X.00    X.00
*  2hpc    57.50    59.50    57.00    57.24    X.00    X.00
*  3hpc    95.00    94.00   100.00    96.89    X.00    X.00
*  4hpc   131.00   130.50   131.00   130.34    X.00    X.00
*  5hpc   174.00   175.00   175.00   174.07    X.00    X.00
*  6hpc   227.00   226.50   226.00   226.11    X.00    X.00
*  7hpc   260.50   261.00   266.00   261.49    X.00    X.00
```

Figure 6.2: Modified fasta file O04714.mfta. The file contains the primary sequence of GCR1, followed by seven rows (1tm-7tm) containing the seven TM domains with their corresponding $\eta$ residues. The next seven rows (1hpc-7hpc) give the hydrophobic centers obtained from the four possible definitions given below.

### Prediction of the 7-helix bundle by using the predicTM method

The predicTM method uses the BLAST output to make a sequence alignment. The final output of predicTM are the files O04714.data.csv and O04714.mfta, see Figure 6.2). The first file is used to plot the hydrophobicity profile, averaged over different lengths of neighboring residue windows (7-21 Moving window averages), to reduce the noise in the "raw hydrophobicity profile" for individual residues. Figure 6.1 depicts the seven TM regions. As mentioned above, a residue with averaged hydrophobicity value greater than zero is assigned to a TM helix, and makes part of the raw hydrophobic

regions.

The O04714.mfta file, shown in Figure 6.2, contains the amino acid sequence of GCR1, followed by 7 rows that describe the 7TM regions and their raw and capped helix lengths. The first column identifies the number of the helix, the second and third columns give the raw initial and final residues of the helices, the fourth and fifth columns are the initial and final residue of the capped helices, finally the tenth and eleventh columns give the amino acid abbreviation and the position of the most conserved, $\eta$, residue of each helix. This $\eta$ residue is used as the reference residue for the first angle sampling. The last 7 rows, from 1hpc to 7hpc, show the four different hydrophobic centers (HPC) of each helix calculated by four possible methods: raw-mid, cap-mid, peak, and area [3].

### 6.1.2 3D reconstruction of GCR1 protein via homology.

Since no plant GPCR has been identified, nor crystallized, we do not have candidate proteins to be used as templates to generate a 3D structure of GCR1 by homology. However, there are several animal and human crystallized GPCRs that could serve as templates to initiate the homology procedure on GCR1. To identify possible templates, among the animal and human GPCRs, is necessary to make an alignment against the available animal and human GPCRs. This second alignment employs animal GPCRs with known active/inactive or intermediate tertiary structures to obtain the shapes of the GCR1 helices, which preserve structural characteristics of the templates (helix bends, including the Proline kinks, among others). As a result of the aforementioned procedure, we selected the best four animal GPCR templates to produce the corresponding four O04714.mfta files. These files contain new helices lengths, $\eta$-residues, and hydrophobic centers for the TM regions of GCR1.

#### Alignment and homology with animal GPCRs.

As mentioned above, the GCR1 BLAST search returned one thousand plant protein amino acid sequences similar to GCR1 with no known tertiary structure. A second alignment with animal/human GPCRs, with known tertiary structure, was used to generate a group of GPCRs templates to obtain the mutated helix shape of GCR1. These templates are the result of selecting the GPCRs with the highest identity percentage of similarity and the highest average identity percentage for the seven helices. Table 6.1 shows the four selected GPCR templates: Human Prostaglandin D2 receptor 2,

PD2R2, in its inactive state; Human P2Y purinoceptor 1, human P2RY1, in its intermediate state; Beta-1 adrenergic receptor, wild turkey ADRB1, in its active state; and Prostaglandin E2 receptor EP3 subtype, human PE2R3, in its active state. These four GPCR templates are used as input for the homology helix shape procedure described below.

| Header | All | TM Avg | TM1 | TM2 | TM3 | TM4 | TM5 | TM6 | TM7 | State |
|---|---|---|---|---|---|---|---|---|---|---|
| sp\|O04714\|GCR1_ARATH | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |
| sp\|Q9Y5Y4\|PD2R2_HUMAN | 12.88 | 21.39 | 23.81 | 50 | 17.14 | 21.05 | 10.71 | 15 | 12 | Inactive |
| sp\|P24530\|EDNRB_HUMAN | 14.42 | 21.12 | 14.29 | 37.5 | 11.43 | 26.32 | 14.29 | 20 | 24 | Inactive |
| sp\|P25024\|CXCR1_HUMAN | 15.34 | 20.7 | 19.05 | 31.25 | 14.29 | 21.05 | 14.29 | 25 | 20 | Inactive |
| sp\|P21730\|C5AR1_HUMAN | 13.19 | 20.48 | 14.29 | 25 | 25.71 | 21.05 | 14.29 | 15 | 28 | Inactive |
| sp\|P61073\|CXCR4_HUMAN | 14.11 | 18.66 | 19.05 | 25 | 20 | 26.32 | 14.29 | 10 | 16 | Inactive |
| sp\|O43613\|OX1R_HUMAN | 13.5 | 18.13 | 14.29 | 25 | 14.29 | 26.32 | 25 | 10 | 12 | Inactive |
| sp\|P21453\|S1PR1_HUMAN | 13.19 | 18.08 | 14.29 | 37.5 | 17.14 | 15.79 | 17.86 | 0 | 24 | Inactive |
| sp\|P47900\|P2RY1_HUMAN | 12.27 | 17.54 | 9.52 | 37.5 | 25.71 | 15.79 | 14.29 | 0 | 20 | Intermediate |
| sp\|P25116\|PAR1_HUMAN | 12.27 | 17.41 | 14.29 | 31.25 | 22.86 | 15.79 | 10.71 | 15 | 12 | Intermediate |
| sp\|P25929\|NPY1R_HUMAN | 11.96 | 17.34 | 14.29 | 31.25 | 14.29 | 5.26 | 14.29 | 10 | 32 | |
| sp\|P07700\|ADRB1_MELGA | 14.11 | 17.27 | 9.52 | 18.75 | 17.14 | 21.05 | 21.43 | 5 | 28 | Active |
| sp\|P43115-2\|PE2R3_HUMAN | 13.19 | 17.13 | 4.76 | 31.25 | 20 | 21.05 | 17.86 | 5 | 20 | Intermediate |
| sp\|P43115-2\|PE2R3_HUMAN | 13.19 | 17.13 | 4.76 | 31.25 | 20 | 21.05 | 17.86 | 5 | 20 | Active |

Table 6.1: Animal and human GPCRs obtained after the condensed alignment. Gray and green color rows are the best-ranked GPCRs that are used as templates to generate the initial structures for GCR1 by the homologize procedure. In gray: human prostaglandin D2 receptor 2, PD2R2; human P2Y purinoceptor 1, P2RY1; Wild turkey Beta-1 adrenergic receptor, ADRB1; and in green the winner human prostaglandin E2 receptor PE3 subtype, PE2R3.

The coordinates of the four templates of Table 6.1 are not relative to a coordinate frame centered at the lipid bilayer. These templates need to be reoriented by using the coordinates of the protein hBeta2_2rh1_opm.pdb. The result of this orientation on the four templates is the helix positions on the hydrophobic plane, $x$ and $y$ coordinates, and orientations for each homologized helix bundle. The homology procedure with the GPCR templates produced four mfta and four bgf files, these files are displayed in Figures 10.1-10.4 of Appendix I.

| Helix | predicTM | PD2R2 | P2RY1 | ADRB1 | PE2R3 |
|---|---|---|---|---|---|
| TM1 | F36 | G25 | S34 | S34 | S34 |
| TM2 | D62 | D62 | D62 | D62 | D62 |
| TM3 | T87 | T87 | T87 | T87 | T87 |
| TM4 | W128 | W128 | W128 | W128 | W128 |
| TM5 | P170 | L157 | A174 | L157 | L157 |
| TM6 | Y220 | I224 | F240 | I224 | L223 |
| TM7 | S266 | M261 | S266 | S266 | S266 |

Table 6.2: Type and position of the $\eta$ residue for the five protein templates obtained from the alignment and homology procedures. The predicTM column is based on plant proteins; the columns PD2R2, P2RY1, ADRB1, and PE2R2, are based on animal/human GPCRs. Most of the $\eta$ residues of TM2, TM3, TM4 and TM6 are identical.

A direct observation of the mfta files obtained by predicTM, figure 6.2, and the ones obtained by GPCR homology, figures 10.1-10.4 of SI, shows differences in the positions of the $\eta$ residues and the hydrophobic centers, and the length of transmembrane regions. Table 6.2 displays the similarities and differences between the $\eta$ residues of the seven helices, for the four template-based structures of GCR1. It is interesting to see that the $\eta$ residue of helices TM2, TM3, TM4, and TM7, are practically identical. Conserved motives of class A GPCRs are also found in GCR1. The conserved motives, that can be seen in Table 6.7, are:

- The NSxxY motif in TM7 of GCR1 is present in as NPxxY in the templates of class A GPCRs [71, 79], therefore S266 was assigned as the 7.50 residue.
- All classes of GPCRs (except class B) have a polar residue in TM1 as the most conserved one. In the case of GCR1, the predicted $\eta$ residue is a F36, however, we chose a polar S34 as the 1.50 residue, consistent with the GPCR template
- For TM2, TM3, and TM4 predicTM and the homology models predicted D62, T87, and W128, respectively, as the most conserved residues, therefore these residues are assigned as the 2.50, 3.32 and 4.50, respectively for GCR1.
- Although, P170 was predicted as the most conserved residue by predicTM, but not for the homologized models in Table 6.2, P170 is assigned as P5.50 of GCR1. Assigning the P170 of GCR1 as the 5.50

residue aligns with a very conserved P residue found in most of the class A GPCRs and also this assignation produces an alignment between Y177 of GCR1 with Y5.57 (Y5.58) of class A GPCRs [71] [79].

- Finally, there are two possible assignments for the 6.50 residue of GCR1. In the first case, the P222 of GCR1 can be aligned with the highly conserved Proline of animal GPCRs. In the second case, the F231 of GCR1 can be aligned with an also highly conserved phenylalanine found in animal GPCRs [71] [79]. The second case has been suggested as the $\eta$ residue by Taddese [11].

The assignments of the $\eta$ residues of TM1-7 give rise to the three possible models: m1, m2, and m3 (predicTM), shown in table 6.3. These three models, for each of the 4 templates, were used as starting structures for the angular sampling of BiHelix[69].

| Helix | PredicTM (m3) | m1 | m2 | Taddesse |
|---|---|---|---|---|
| TM1 | F36 | S34 | S34 | S34 |
| TM2 | D62 | D62 | D62 | D62 |
| TM3 | T87 | T87 | T87 | T87 |
| TM4 | W128 | W128 | W128 | W128 |
| TM5 | P170 | P170 | P170 | A174 |
| TM6 | Y220 | P222 | F231 | F231 |
| TM7 | S266 | S266 | S266 | S266 |

Table 6.3: $\eta$ residues for the models m1, m2, and m3 (predicTM), after an alignment analysis against GPCRs as described above.

Due to the modifications on the $\eta$ residues for the four templates, it is necessary to carry out an additional homologize step. This new homologize produced four mfta files, for the corresponding template, for each model (m1, m2, and m3), in total there are twelve new mfta and twelve new bgf files. These mfta can be seen in Figures 10.5 to 10.7 of Appendix II.

Once the new shapes of the mutated helices were obtained (together with their corresponding lengths,$\eta$ and the hydrophobic residues), the pair-wise energies of the helices were calculated by sampling the $\eta$ angles, the 20 best energetically packed 7TM-bundle conformers were selected, as mentioned in the next BiHelix section.

### 6.1.3 Selection of transmembrane candidates by energy.

In a general view, this procedure can be seen in two steps called BiHelix and SuperBiHelix. (i) BiHelix involved optimizing the 7 helical orientations, through the sampling of the $\eta$ angle. Here we sampled rotations of each helix in increments of 30°. This led to $12^7$, about 35 million configurations. We then estimated the interaction energy for each configuration, by calculating independently the 12 pairwise interactions between helices (1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-1, 7-2, 7-3, 3-6, 3-5 and 4-2) [69]. We used SCREAM [70] to optimize the side chains for each case. Since in the BiHelix sampling, the helix pairwise energy components are additive [69], by calculating $12 \times (12^2) = 1728$ energies is possible to estimate the energies of all possible 35 million configurations. We then selected the best 1000 packing of 7TM conformers by energy, built the full 7-helix configuration and used SCREAM to find the optimum side chains. Each structure was then minimized in energy (10 steps) using a conjugate gradient method. We call this, the CombiHelix step. (ii) SuperBiHelix, starts using the results from CombiHelix, where we select the best structures based on energy and diversity, and optimize the tilts ($\theta$, $\phi$). So, in the second step, we evaluate the energy for the values of the angles $\theta$, $\phi$ and $\eta$. This led to a total of $(3 \times 5 \times 5)^7$, 13 trillion configurations. We then estimated the energy for all 13 trillion using pairwise interactions between the helices, selected the best 1000 (the SuperCombiHelix step), built the full 7-helix configuration, using SCREAM to find the optimum side chains, and finalized with a short (10 steps) energy minimization [68]. The best 25 structures were selected based on energy and diversity. The results of both samplings are shown below.

#### First BiHelix sampling

We employed the BiHelix sampling method [69] on the three models m1-3 of GCR1. This sampling produced a total of $10^{12}$ configurations for each model. A ranking based on charged and neutral interhelical energy, CInterH and NInterH respectively, and the $\eta$ angle diversity, was employed to sort and select the best 25 configurations for each model.

Tables 10.1-10.3 shown in Appendix III. summarize the BiHelix result for the three models m1,m2, and m3, for four homologized structures based on P2RY1, PE2R3, PD2fRho, and ADRB1 respectively.

To be able to select the best structures from models 1, 2, and 3, to sample in SuperBiHelix, we gather the 25 best structures obtained from BiHelix, for each model, and rearrange them, based on Energy and diversity of the

$\eta$-angles, as is shown in Table 6.4.

| ResBH | Eta | | | | | | | Source | Ind Energies | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ResBH | H1 | H2 | H3 | H4 | H5 | H6 | H7 | Source | CInterH | Order CinterH | NInterH | Order NinterH | Sum Order Cih-Nih |
| 1 | -90 | 30 | 30 | 60 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -654.037 | 1 | -380.406 | 7 | 8 |
| 2 | 90 | 0 | 30 | 180 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -636.980 | 3 | -380.442 | 6 | 9 |
| 3 | -120 | -90 | 30 | 180 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -632.983 | 4 | -377.307 | 11 | 15 |
| 4 | -120 | -90 | 30 | 60 | 0 | 0 | 0 | m1-GCR1-PE2R3 | -625.622 | 8 | -376.632 | 12 | 20 |
| 5 | -150 | 90 | 30 | -150 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -627.517 | 6 | -375.387 | 14 | 20 |
| 6 | 30 | 60 | 0 | 30 | 0 | 0 | 0 | m1-GCR1-PE2R3 | -618.004 | 17 | -384.654 | 4 | 21 |
| 7 | -90 | 60 | 0 | 30 | 0 | 0 | 0 | m1-GCR1-PE2R3 | -615.814 | 19 | -383.823 | 5 | 24 |
| 8 | -90 | 30 | 0 | 150 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -639.749 | 2 | -364.550 | 28 | 30 |
| 9 | 90 | 0 | 30 | 60 | 0 | 0 | 0 | m1-GCR1-PE2R3 | -614.432 | 21 | -377.621 | 10 | 31 |
| 10 | -90 | 30 | 0 | 30 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -622.425 | 13 | -373.946 | 18 | 31 |
| 11 | 150 | 30 | 0 | 150 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -625.243 | 9 | -366.698 | 23 | 32 |
| 12 | 30 | 60 | 0 | 150 | 0 | 0 | 0 | m1-GCR1-PE2R3 | -620.664 | 15 | -370.712 | 21 | 36 |
| 13 | -30 | 90 | 30 | 120 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -630.624 | 5 | -359.913 | 32 | 37 |
| 14 | -60 | 30 | 0 | 30 | 0 | 120 | 0 | m1-GCR1-PE2R3 | -620.750 | 14 | -364.940 | 27 | 41 |
| 15 | 90 | 0 | 30 | -120 | 0 | 0 | 0 | m1-GCR1-PE2R3 | -623.292 | 10 | -359.054 | 34 | 44 |
| 16 | 30 | 120 | 90 | -150 | 0 | 90 | -90 | m1-GCR1-PE2R3 | -619.971 | 16 | -360.323 | 31 | 47 |
| 17 | 90 | 0 | 30 | 30 | 0 | 0 | 0 | m1-GCR1-PE2R3 | -625.969 | 7 | -357.033 | 40 | 47 |
| 18 | 120 | -120 | 30 | 180 | 0 | 0 | 0 | m1-GCR1-PE2R3 | -614.214 | 22 | -365.047 | 26 | 48 |
| 19 | 120 | 0 | 30 | 90 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -569.016 | 51 | -399.658 | 1 | 52 |
| 20 | 30 | -120 | 30 | 60 | 0 | 150 | -60 | m1-GCR1-PE2R3 | -607.329 | 32 | -370.870 | 20 | 52 |
| 21 | -90 | 60 | 0 | 60 | 0 | 150 | 0 | m1-GCR1-PE2R3 | -612.804 | 23 | -363.045 | 30 | 53 |
| 22 | -120 | 60 | 30 | 150 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -558.579 | 52 | -396.710 | 2 | 54 |
| 23 | 180 | 30 | 0 | 90 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -623.241 | 11 | -356.224 | 43 | 54 |
| 24 | -120 | 60 | 30 | 120 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -547.543 | 53 | -390.312 | 3 | 56 |
| 25 | -90 | 30 | 0 | 150 | 0 | 60 | -30 | m2-GCR1-PE2R3 | -536.728 | 54 | -378.733 | 9 | 63 |

Table 6.4: 25 best energetically conformers obtained from BiHelix results, for each model m1, m2, and m3, gathered to select the best candidates based on energy and diversity

Finally a short list of 5 structures was obtained after choosing the best structures for each model by CInterH and NInterH, and diversity, as is shown in Table 6.5. These 5 structures were used as input for the SuperBiHelixSampling.

| ResBH | Eta | | | | | | | Source | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ResBH | H1 | H2 | H3 | H4 | H5 | H6 | H7 | Source | CInterH | NInterH | rCi | rNi |
| B1 | -90 | 30 | 30 | 60 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -654.036.800 | -380.405.600 | 1 | 6 |
| B2 | 90 | 0 | 30 | 180 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -636.980.100 | -380.441.600 | 2 | 3 |
| B3 | -120 | -90 | 30 | 180 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -632.983.200 | -377.306.700 | 3 | 5 |
| B5 | -150 | 90 | 30 | -150 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -627.517.100 | -375.386.800 | 15 | 11 |
| B22 | -120 | 60 | 30 | 150 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -558.579 | -396.710 | 2 | 54 |

Table 6.5: Selected structures from BiHelix results to be sampled in SuperBiHelix

Although model m1 was the only model that produced low-energy BiHelix conformers, based on the four templates selected from the condensed alignment, PD2, ADRB1, P2RY1, and PE2R3, as can be seen in Tables 10.1-10.3 of Appendix III, after combining all three models and down-select the best 25 energetically structures, the homologized GCR1 structure obtained from the PE2R3 template was the only one that produced the best energetic top 25 conformers, after sampling the $\eta$ angle. Then, we selected the five structures, from BiHelix-CombiHelix, shown in Table 6.5, which means that PE2R3 is by far the best template.

**First SuperBiHelix sampling.**

As mentioned above, we selected, from BiHelix-CombiHelix, the best five structures based on energy and diversity, for further optimization of tilts ($\theta$, $\phi$), to evaluate the energy of 13 trillion of configurations. This was done using pairwise interactions between the helices, selecting the best 1000 (the SuperCombiHelix step), building the full 7-helix configuration, using SCREAM to find the optimum side chains, and ending with a short energy minimization (10 steps) [68]. Table 6.6 shows the top 25 structures based on energy after sampling angles $\theta$, $\phi$, and $\eta$, of each helix. This sorted list is obtained by using the same energy and diversity criteria of BiHelix. The SuperBiHelix method[68] started from an approximated energy based on pairwise helix interactions, to generate multiple rotational combinations of $\theta$-, $\phi$-angles, followed by optimization of the side-chains using SCREAM[70], to obtain a new series of energy structures with a wider variability because the $\theta$-, $\phi$-, and of course $\eta$-angles sampling.

Results in table 6.6 show that all three models, m1-3, using PE2R3 as a template, were present in the top 25 best energetic structures after SuperBihelix sampling. Therefore, we selected this template to carry out the 2nd homologize step to continue with the prediction of GCR1 structure, as mentioned previously. In consequence, before completing the protein by the addition of loops, it is necessary to carry out a new sequence alignment. This new alignment is among PE2R3 template, GCR1, and a series of plant GCR1-like proteins found in the initial BLAST search, section 6.1.1. The goal of this new alignment is to find conserved residues, motifs, or conserved contacts present in both PE2R3 and GCR1, by means of the alignment of the proposed $\eta$ residues in GCR1 with those from the template. Once the $\eta$ residues have been aligned, the initial and final residues of the GCR1 helices are adjusted, extended or reduced, to the near-breaking helix residues in comparison with the helices of the PE2R3 template. Table 6.7 displays the

| | T | | | | | | | P | | | | | | | E | | | | | | | | Source | Ind Energies | | | | | I Ranks | | Comb Ranks |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H1 | H2 | H3 | H4 | H5 | H6 | H7 | Source | CInterH | O_CintH | NInterH | O_NintH | Sum C-H | rCi | rNi | rCNi |
| 1 | 0 | 0 | -10 | -10 | 10 | 0 | 0 | 0 | -30 | 15 | -30 | 0 | 0 | 0 | 30 | 15 | 15 | 30 | -15 | -15 | 15 | B2m3-GCR1-PE2R3 | -726.064 | 3 | -437.188 | 3 | 6 | 3 | 3 | 3 |
| 2 | -10 | 0 | 0 | -10 | 0 | -10 | -10 | 15 | 15 | 15 | 0 | -30 | -30 | -30 | 0 | 0 | 30 | 15 | 30 | 0 | 15 | B3m3-GCR1-PE2R3 | -718.050 | 8 | -462.102 | 1 | 9 | 8 | 1 | 4.5 |
| 3 | -10 | 0 | -10 | 0 | 0 | -10 | -10 | 15 | 0 | -30 | 15 | -30 | 30 | -30 | 0 | -15 | 15 | 15 | -15 | 30 | 30 | B2m3-GCR1-PE2R3 | -737.637 | 1 | -431.991 | 8 | 9 | 1 | 8 | 4.5 |
| 4 | -10 | 0 | -10 | 0 | 0 | -10 | -10 | 0 | -15 | 0 | -15 | -30 | 0 | 15 | 15 | 0 | 15 | 0 | 15 | 0 | 30 | B2m3-GCR1-PE2R3 | -720.470 | 7 | -435.396 | 5 | 12 | 7 | 5 | 6 |
| 5 | -10 | 0 | -10 | 0 | -10 | -10 | -10 | 15 | 0 | -15 | 0 | -30 | 30 | -30 | 0 | -15 | 0 | 30 | 30 | 0 | 30 | B2m3-GCR1-PE2R3 | -736.186 | 2 | -418.162 | 40 | 42 | 2 | 40 | 21 |
| 6 | -10 | 0 | -10 | -10 | 0 | -10 | -10 | 15 | -15 | 15 | -30 | -30 | 15 | -30 | 0 | 0 | 15 | 30 | 15 | -15 | 30 | B2m3-GCR1-PE2R3 | -699.915 | 16 | -419.690 | 31 | 47 | 16 | 31 | 23.5 |
| 7 | -10 | 0 | 0 | -10 | 0 | -10 | -10 | 15 | 15 | 15 | 15 | -30 | 0 | 15 | -15 | 0 | 30 | 15 | 30 | 0 | 30 | B3m3-GCR1-PE2R3 | -690.639 | 36 | -424.315 | 16 | 52 | 36 | 16 | 26 |
| 8 | -10 | 0 | 0 | -10 | 10 | -10 | -10 | 15 | 15 | 15 | 30 | -15 | -15 | 30 | -15 | -15 | 30 | 15 | 30 | 15 | 30 | B3m3-GCR1-PE2R3 | -690.948 | 35 | -421.975 | 21 | 56 | 35 | 21 | 28 |
| 9 | 0 | 0 | -10 | -10 | -10 | -10 | -10 | 0 | -15 | 0 | 0 | -30 | 30 | 30 | 0 | 15 | -15 | -30 | 15 | 15 | 15 | B1m1-GCR1-PE2R3 | -724.033 | 6 | -413.256 | 69 | 75 | 6 | 69 | 37.5 |
| 10 | 0 | 0 | -10 | 0 | -10 | -10 | 0 | 0 | -15 | 30 | 0 | -30 | 30 | 15 | 15 | 15 | 30 | 15 | 30 | 0 | 0 | B2m3-GCR1-PE2R3 | -703.424 | 11 | -413.440 | 67 | 78 | 11 | 67 | 39 |
| 11 | 0 | 0 | -10 | 0 | 0 | -10 | -10 | 0 | -15 | 30 | -30 | -30 | -30 | 0 | 0 | 30 | 0 | 30 | 0 | 0 | 0 | B2m3-GCR1-PE2R3 | -692.102 | 28 | -412.393 | 74 | 102 | 28 | 74 | 51 |
| 12 | -10 | 0 | 0 | -10 | 0 | -10 | -10 | 15 | 0 | 15 | 30 | -30 | 0 | 15 | 0 | 0 | 30 | 15 | 15 | 0 | 30 | B3m3-GCR1-PE2R3 | -689.018 | 44 | -414.479 | 63 | 107 | 44 | 63 | 53.5 |
| 13 | -10 | 0 | 0 | -10 | 0 | -10 | -10 | 15 | 15 | 15 | 0 | -30 | 0 | 15 | -15 | 0 | 30 | 15 | 30 | 0 | 30 | B3m3-GCR1-PE2R3 | -685.156 | 65 | -417.181 | 46 | 111 | 65 | 46 | 55.5 |
| 14 | 0 | -10 | -10 | 10 | -10 | -10 | -10 | -15 | 0 | 0 | -30 | -30 | 30 | 30 | 15 | 15 | 15 | 0 | 0 | 15 | 15 | B1m1-GCR1-PE2R3 | -725.169 | 4 | -406.492 | 118 | 122 | 4 | 118 | 61 |
| 15 | -10 | 0 | -10 | 0 | 10 | 0 | -10 | 15 | 0 | 15 | -30 | 0 | -15 | -30 | 0 | -15 | 15 | 0 | -15 | -15 | 30 | B2m3-GCR1-PE2R3 | -692.981 | 26 | -408.335 | 102 | 128 | 26 | 102 | 64 |
| 16 | -10 | 0 | -10 | 0 | 0 | -10 | -10 | 15 | -15 | -15 | 0 | -30 | 0 | 15 | 15 | 15 | 15 | 15 | 0 | 0 | 30 | B3m3-GCR1-PE2R3 | -678.330 | 102 | -420.156 | 27 | 129 | 102 | 27 | 64.5 |
| 17 | -10 | 0 | 0 | -10 | 0 | -10 | -10 | 15 | 15 | 15 | 15 | -30 | -30 | -30 | 0 | 0 | 30 | 15 | 30 | 0 | 15 | B3m3-GCR1-PE2R3 | -676.190 | 124 | -431.064 | 9 | 133 | 124 | 9 | 66.5 |
| 18 | -10 | -10 | -10 | 10 | -10 | -10 | -10 | -15 | 30 | -15 | -30 | 0 | 15 | 30 | 30 | 0 | 30 | -30 | 30 | 0 | 30 | B5m1-GCR1-PE2R3 | -700.570 | 14 | -405.963 | 120 | 134 | 14 | 120 | 67 |
| 19 | 0 | 0 | -10 | 0 | 10 | 0 | 0 | 0 | 0 | 30 | -30 | 0 | 0 | 0 | 15 | 15 | 0 | 15 | -15 | -15 | 30 | B2m3-GCR1-PE2R3 | -700.901 | 13 | -405.516 | 125 | 138 | 13 | 125 | 69 |
| 20 | -10 | -10 | -10 | 0 | -10 | -10 | -10 | -15 | 0 | 0 | 0 | -30 | 15 | 30 | 30 | 0 | 0 | -15 | 15 | 15 | 15 | B5m1-GCR1-PE2R3 | -685.049 | 66 | -409.921 | 95 | 161 | 66 | 95 | 80.5 |
| 21 | -10 | 0 | 0 | -10 | 0 | 0 | -10 | 0 | -30 | 15 | 30 | -15 | 0 | -15 | -30 | -30 | 0 | -15 | 15 | 15 | 30 | B1m1-GCR1-PE2R3 | -688.386 | 48 | -406.767 | 116 | 164 | 48 | 116 | 82 |
| 22 | -10 | -10 | -10 | 10 | -10 | -10 | -10 | -30 | 30 | -15 | -30 | -30 | 30 | 0 | 15 | 0 | 30 | -30 | 30 | 0 | 15 | B5m1-GCR1-PE2R3 | -691.353 | 32 | -403.857 | 135 | 167 | 32 | 135 | 83.5 |
| 23 | -10 | 0 | 0 | -10 | 0 | -10 | -10 | 15 | 15 | 15 | 0 | -30 | -30 | -30 | 15 | 0 | 30 | 15 | 30 | 0 | 15 | B3m3-GCR1-PE2R3 | -675.411 | 132 | -418.377 | 37 | 169 | 132 | 37 | 84.5 |
| 24 | -10 | -10 | -10 | 10 | -10 | -10 | -10 | -30 | 30 | -15 | -30 | -30 | 15 | 30 | -30 | 0 | 30 | -30 | 30 | 0 | 30 | B5m1-GCR1-PE2R3 | -678.456 | 100 | -412.514 | 73 | 173 | 100 | 73 | 86.5 |
| 25 | -10 | -10 | -10 | 10 | -10 | -10 | -10 | -30 | 30 | 0 | -30 | -30 | 0 | -15 | 0 | -30 | 0 | -30 | 0 | -15 | 0 | B22m2-GCR1-PE2R3 | -657.372 | 423 | -420.657 | 25 | 448 | 423 | 25 | 224 |

Table 6.6: Combined models SuperBiHelix results

final residues of GCR1 helices.

The final result is three new structures (with mfta files) for the three models m1-3 based on the PE2R3 template. The re-alignment and adjustment of the helix lengths require a new homologization process to generate new mfta and bgf files that consider the new helix lengths. Additionally, new BiHelix, and SuperBiHelix sampling and optimizations are needed since new interactions are present by the redefined helices.

### 6.1.4 Second Homologization

The selected template, the active Prostaglandine E Receptor 3, PE2R3 (P43115-2), has one of the best TM average similarity percentages, to address the Proline alignment issue in homology modeling, with 13.19% sequence identity (17.13% for TMD).

After unifying the helices' lengths for each helix, the only difference among the models, m1-3, are $\eta$-residues, for some helices as shown in Table 6.3. The homologized structures are used in a new BiHelix and SuperBiHelix sampling, as described in the next section.

The three new models, m1-3 respectively, are tabulated and shown in Appendix IV. The differences among the three models are only due to the selection of $\eta$ residues for helix 6.

#### Second BiHelix sampling

By using results from the new homologized structures based on PE2R3 template, we employed the BiHelix sampling method [69] on the three models m1-3 of GCR1. As mentioned before, first we optimized the 7 helical orientations ($\eta$). Consequently, we sampled rotations of each helix in increments of 30°. This led to $12^7$, about 35 million configurations. We then estimated the interaction energy for each configuration, by calculating independently the 12 pairwise interactions between helices (1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-1, 7-2, 7-3, 3-6, 3-5 and 4-2) [69]. Finally, we used SCREAM [70] to optimize the side chains for each case. This sampling produced a total of $10^{12}$ configurations for each model. A ranking based on charged and neutral interhelical energy, CInterH and NInterH respectively, the $\eta$ angle diversity, and conserved contacts present in both GCR1 and class A GPCRs is employed to sort and select the best 25 configurations for each model (Table 6.8).

The nomenclature used in Table 6.8 to describe each structure is as follows. The letter b next to the number indicates the row number where it appears in the column 1, bestBH. That is, the position among best energy

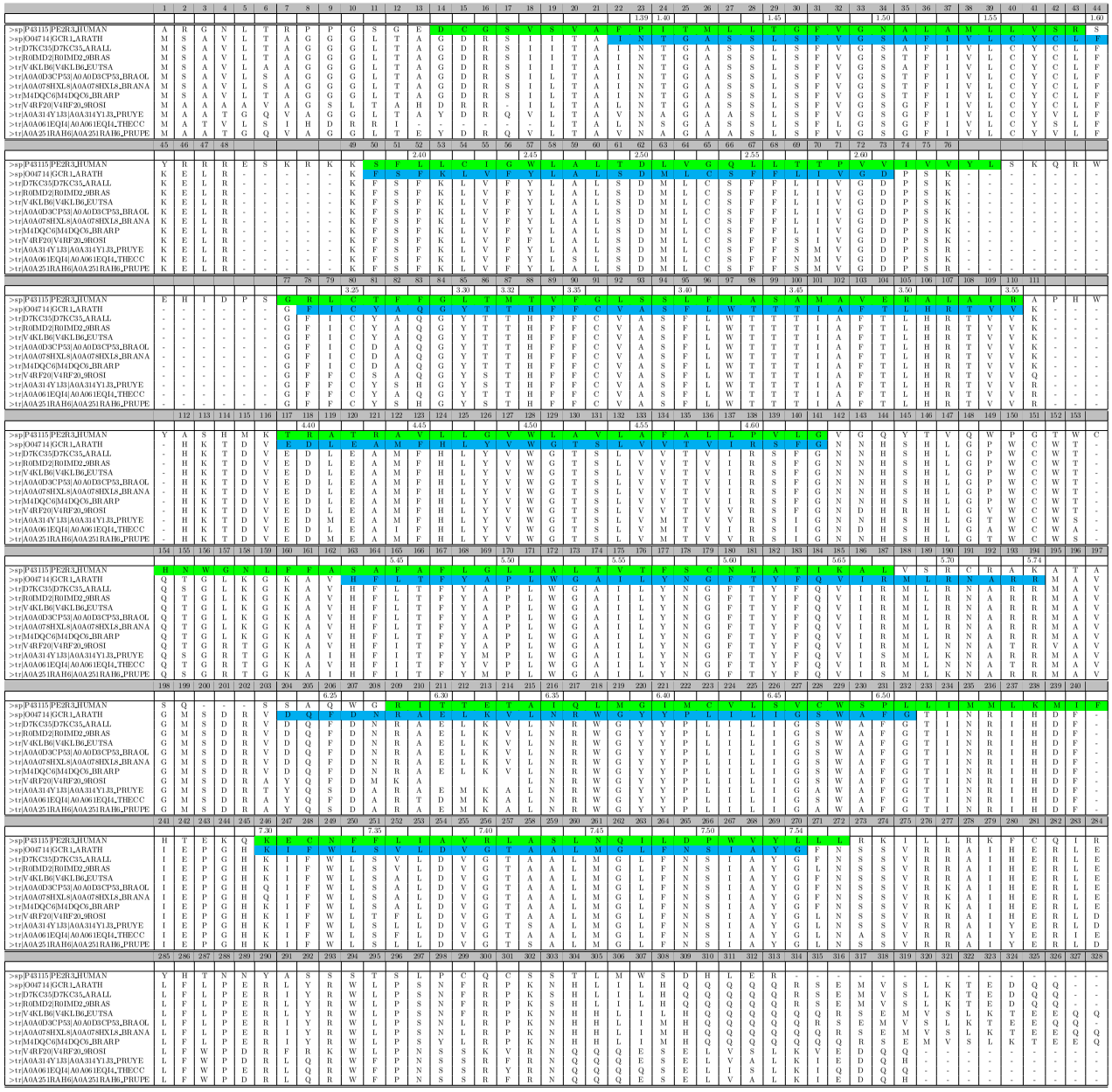

Table 6.7: Condensed alignment with animal-GPCRs and GCR1, analysis of TM and loops regions

| | Eta | | | | | | | Source | Ind. Energies | | | Ind. Ranks | | Comb. Ranks | | Contacts | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| bestBH | H1 | H2 | H3 | H4 | H5 | H6 | H7 | Source | CInterH | NInterH | E-Average | rCi | rNi | rCNi | RankNih | active | inactive | Class A Contacts |
| 1bBH | 0 | 30 | 0 | 0 | 0 | 120 | 30 | 1mGCR1-PE2R3 | -677.538 | -386.184 | -531.861 | 3 | 2 | 2.5 | 2 | 1 | 0 | 10 |
| 2bBH | -120 | -30 | 30 | 60 | 0 | 0 | 0 | 3mGCR1-PE2R3 | -665.730 | -390.911 | -528.321 | 10 | 1 | 5.5 | 1 | 1 | 2 | 19 |
| 3bBH | -30 | -90 | 30 | 150 | 0 | 30 | 0 | 2mGCR1-PE2R3 | -691.493 | -362.898 | -527.195 | 1 | 21 | 11 | 21 | 1 | 0 | 3 |
| 4bBH | -90 | 0 | 0 | -30 | 0 | 120 | 0 | 1mGCR1-PE2R3 | -683.068 | -368.933 | -526.000 | 2 | 12 | 7 | 12 | 1 | 0 | 14 |
| 5bBH | 90 | 0 | 30 | 60 | 0 | 0 | 0 | 3mGCR1-PE2R3 | -667.791 | -377.930 | -522.861 | 8 | 5 | 6.5 | 5 | 2 | 2 | 21 |
| 6bBH | -90 | 0 | 0 | 150 | 0 | 30 | 0 | 2mGCR1-PE2R3 | -670.945 | -374.361 | -522.653 | 5 | 8 | 6.5 | 8 | 2 | 0 | 7 |
| 7bBH | -90 | 0 | 30 | 60 | 0 | 0 | 0 | 3mGCR1-PE2R3 | -650.868 | -380.657 | -515.762 | 24 | 3 | 13.5 | 3 | 1 | 2 | 22 |
| 8bBH | -90 | 0 | 0 | 30 | 0 | 120 | 0 | 1mGCR1-PE2R3 | -659.378 | -367.969 | -513.673 | 13 | 13 | 13 | 13 | 2 | 0 | 9 |
| 9bBH | 180 | 0 | 30 | 60 | 0 | 30 | 30 | 2mGCR1-PE2R3 | -668.073 | -358.731 | -513.402 | 7 | 30 | 18.5 | 30 | 0 | 0 | 12 |
| 10bBH | 0 | 0 | 0 | 30 | 0 | 30 | 0 | 2mGCR1-PE2R3 | -647.608 | -377.472 | -512.540 | 29 | 6 | 17.5 | 6 | 2 | 0 | 8 |
| 11bBH | 0 | 0 | 30 | 60 | 0 | 30 | 30 | 2mGCR1-PE2R3 | -671.639 | -352.349 | -511.994 | 4 | 57 | 30.5 | 57 | 0 | 0 | 7 |
| 12bBH | 0 | 0 | 0 | -30 | 0 | 120 | 0 | 1mGCR1-PE2R3 | -662.190 | -360.454 | -511.322 | 12 | 25 | 18.5 | 25 | 1 | 0 | 14 |
| 13bBH | 0 | 0 | 0 | 30 | 0 | 120 | 0 | 1mGCR1-PE2R3 | -651.013 | -371.158 | -511.085 | 23 | 10 | 16.5 | 10 | 2 | 0 | 9 |
| 14bBH | -30 | 30 | 0 | 0 | 0 | 120 | 0 | 1mGCR1-PE2R3 | -643.416 | -375.863 | -509.640 | 38 | 7 | 22.5 | 7 | 2 | 0 | 10 |
| 15bBH | -90 | -30 | 30 | 60 | 0 | 30 | 30 | 2mGCR1-PE2R3 | -664.194 | -354.121 | -509.158 | 11 | 50 | 30.5 | 50 | 1 | 0 | 12 |
| 16bBH | 0 | 0 | 0 | -60 | 0 | 120 | 0 | 1mGCR1-PE2R3 | -669.093 | -346.288 | -507.690 | 6 | 102 | 54 | 102 | 1 | 0 | 10 |
| 17bBH | 0 | -30 | 0 | 30 | 0 | 120 | 0 | 1mGCR1-PE2R3 | -640.812 | -371.749 | -506.280 | 44 | 9 | 26.5 | 9 | 1 | 1 | 11 |
| 18bBH | -90 | -30 | 30 | 60 | 0 | 30 | 0 | 2mGCR1-PE2R3 | -666.769 | -341.450 | -504.110 | 9 | 132 | 70.5 | 132 | 1 | 0 | 11 |
| 19bBH | -30 | -90 | 90 | 150 | 30 | 30 | 0 | 2mGCR1-PE2R3 | -654.318 | -351.857 | -503.087 | 17 | 62 | 39.5 | 62 | 1 | 0 | 1 |
| 20bBH | -90 | -30 | 0 | 30 | 0 | 120 | 0 | 1mGCR1-PE2R3 | -641.551 | -364.275 | -502.913 | 43 | 19 | 31 | 19 | 1 | 0 | 13 |
| 21bBH | -30 | -30 | 0 | 30 | 0 | 120 | 0 | 1mGCR1-PE2R3 | -635.030 | -367.688 | -501.359 | 66 | 14 | 40 | 14 | 1 | 1 | 10 |
| 22bBH | -60 | -30 | 0 | 30 | 0 | 120 | -30 | 1mGCR1-PE2R3 | -634.842 | -367.512 | -501.177 | 68 | 15 | 41.5 | 15 | 1 | 0 | 9 |
| 23bBH | 0 | 0 | 30 | 60 | 0 | 0 | 0 | 3mGCR1-PE2R3 | -634.648 | -367.404 | -501.026 | 71 | 16 | 43.5 | 16 | 1 | 2 | 27 |
| 24bBH | -120 | 30 | 0 | 0 | 0 | 120 | 30 | 1mGCR1-PE2R3 | -634.751 | -367.218 | -500.984 | 70 | 17 | 43.5 | 17 | 1 | 0 | 11 |
| 25bBH | -30 | 0 | 0 | 30 | 0 | 120 | 0 | 1mGCR1-PE2R3 | -639.964 | -361.022 | -500.493 | 47 | 23 | 35 | 23 | 2 | 0 | 9 |

Table 6.8: BiHelix results from animal-plant comparative analysis of new homologized structures

structures, # BiHelix. The next seven columns, describe the $\eta$ angle value for helices from seven helices sampled, H1 to H7. The source, in column 8, indicates the model, m1-3, where the selected conformer is from, during the eta angle sampling, *i.e.,* we selected 1bBH whose source was model 1 of GCR1, which was built based on PE2R3 class-A GPCR template. Ind Energies, in the following columns, account for the charged and neutral interhelical energies of individual pairwise interactions. Comb Ranks, shows the rank for combined charged-neutral interhelical energies. And finally, in the last three columns there is a description of the conserved contacts found in the GCR1 TM bundle based on PE2R3 template, after sampling the $\eta$ angle. For instance, one of the most important contacts found in class A animal GPCRs is that between N1.50-D2.50-N7.49. Although we did not find these H-bonds in our BiHelix results, we do find a good orientation for these aminoacid residues, which can produce better distances among those residues in the selected structures, after sampling $\theta$- and $\phi$-tilts, see Figure 6.3, and Figures 10.8 - 10.14 of Appendix V for a detailed view of each structure. In addition to charged and neutral interhelical energies and diversity of the $\eta$-angle sampled, we also looked for other important characteristics such as, conserved contacts, that allowed us to select the best candidates for the SuperBiHelix sampling. The conserved contacts or interactions are shown using the Ballesteros-Weinstein, BW, numbering.

A short list of 7 structures was obtained, as is shown in Table 6.8. These 7 structures are used as input for the SuperBiHelix Sampling.

Figure 6.3. shows a superimposing picture of all 7 selected structures from BiHelix to be sampled in SuperBiHelix step.

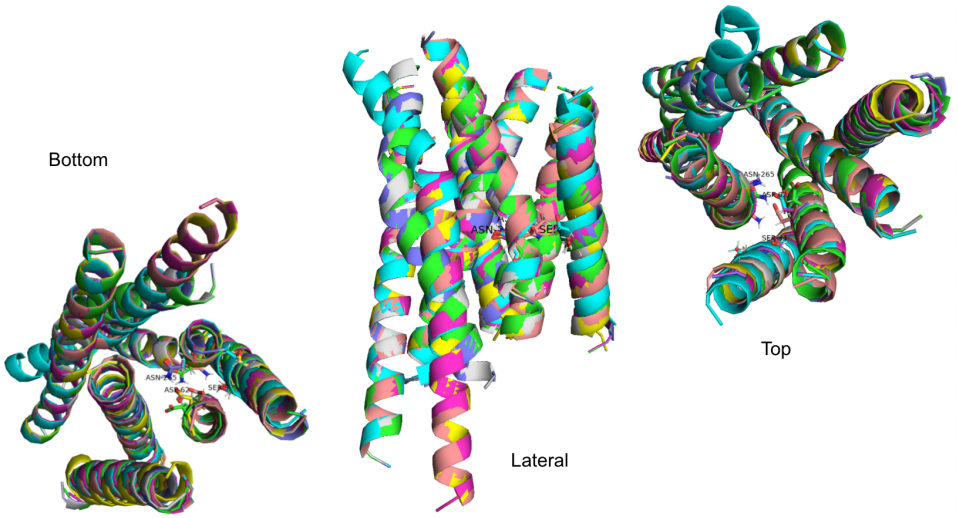


Figure 6.3: Bottom, lateral, and top views from all 7 candidate structures. The S1.50-D2.50-N7.49 positions are shown

### 6.1.5 2nd SuperBiHelix sampling

As we mentioned previously, from the latest BiHelix-CombiHelix, we obtained the seven best structures based on energy and diversity, and optimized tilts ($\theta$, $\phi$) in the SuperBiHelix sampling method [68]. Here, we evaluate the energy for several combinations of values for the $\theta$, $\phi$ and $\eta$ angles. This sampling led to a total of $(3 \times 5 \times 5)^7$, 13 trillion configurations. In consequence, we estimated the energy for all 13 trillion using pairwise interactions between the helices, selected the best 1000 (the SuperCombiHelix step), built the full 7-helix configuration, found the optimum side chains by using SCREAM and finalizing with a short (10 steps) energy minimization [68]. Table 6.9 shows the top 25 structures selected based on energy. This sorted list is obtained by using the same energy and diversity criteria as in the BiHelix sorting. Next, we performed a further analysis of the top 25 structures based on diversity among the $\eta$-, $\phi$- and $\theta$-angle values. Finally, we searched for possible conserved interactions found in both class A animal-GPCRs and in GCR1. The findings of this analysis are key elements to select the best 5 SuperBiHelix structures, and are shown in Tables 6.9 and 6.10.

A detailed analysis of the interactions between the helices of GCR1 shows

| | Theta | | | | | | | Phi | | | | | | | Eta | | | | | | | | Source | I Energies | | I Ranks | | Comb Ranks | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Best SBH | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H1 | H2 | H3 | H4 | H5 | H6 | H7 | BiHelix | CInterH | NIntraH | rCi | rNi | rCNi | active | inactive | Act-Inac | Class A Cont |
| 1SBiH | -10 | 0 | -10 | 0 | 10 | -10 | -10 | 15 | -15 | -30 | 30 | -30 | 30 | 15 | 30 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -787.343 | 218.851 | 3 | 2 | 2.5 | 2 | 0 | 2 | 3 |
| 2SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 30 | -15 | -30 | 30 | -30 | 30 | 30 | 30 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -783.271 | 243.569 | 4 | 3 | 3.5 | 2 | 0 | 2 | 3 |
| 3SBiH | -10 | 0 | -10 | 0 | 10 | -10 | -10 | 30 | -15 | -30 | 30 | -30 | 30 | 0 | 15 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -794.446 | 249.930 | 1 | 7 | 4.0 | 2 | 0 | 2 | 3 |
| 4SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 15 | -15 | -30 | 30 | -30 | 30 | 30 | 30 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -791.223 | 230.581 | 2 | 6 | 4.0 | 2 | 0 | 2 | 3 |
| 5SBiH | -10 | 0 | -10 | 0 | 0 | 0 | -10 | 30 | -15 | -15 | 15 | 0 | 15 | -30 | 30 | -15 | -15 | 0 | 0 | -30 | 0 | 15bBH | -772.064 | 185.382 | 10 | 1 | 5.5 | 1 | 0 | 1 | 6 |
| 6SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 15 | -15 | -30 | 30 | -30 | 30 | 30 | 15 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -778.551 | 197.227 | 7 | 9 | 8.0 | 2 | 0 | 2 | 3 |
| 7SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 15 | -15 | -30 | 30 | -30 | 30 | 15 | 30 | -15 | -30 | 30 | 30 | -30 | -30 | 15bBH | -769.898 | 193.981 | 12 | 5 | 8.5 | 2 | 0 | 2 | 4 |
| 8SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 0 | -15 | -30 | 30 | -30 | 30 | 0 | 30 | 0 | -30 | 30 | 30 | -30 | 30 | 15bBH | -779.959 | 248.009 | 6 | 13 | 9.5 | 0 | 0 | 0 | 5 |
| 9SBiH | -10 | 0 | -10 | 0 | 10 | -10 | -10 | 30 | -15 | -15 | 15 | -30 | 30 | -30 | 30 | -15 | -15 | 0 | 30 | -30 | 0 | 15bBH | -773.118 | 197.494 | 9 | 10 | 9.5 | 1 | 0 | 1 | 6 |
| 10SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 30 | -15 | -30 | 30 | -30 | 30 | 30 | 15 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -781.026 | 229.203 | 5 | 18 | 11.5 | 2 | 0 | 2 | 3 |
| 11SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 15 | -15 | -30 | 30 | -30 | 30 | 30 | -15 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -769.570 | 212.188 | 14 | 11 | 12.5 | 2 | 0 | 2 | 4 |
| 12SBiH | -10 | 0 | -10 | 0 | 10 | -10 | -10 | 15 | -15 | -30 | 30 | -30 | 30 | 0 | 30 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -775.106 | 242.619 | 8 | 26 | 17.0 | 0 | 0 | 0 | 3 |
| 13SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 30 | -15 | -15 | 30 | -30 | 30 | 30 | 30 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -769.255 | 230.495 | 15 | 21 | 18.0 | 2 | 0 | 2 | 3 |
| 14SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 0 | -15 | -30 | 30 | -30 | 30 | 30 | 30 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -766.408 | 220.944 | 21 | 20 | 20.5 | 2 | 0 | 2 | 3 |
| 15SBiH | 0 | -10 | -10 | 0 | 0 | -10 | -10 | -15 | 30 | 0 | -15 | -30 | 30 | 0 | 0 | 0 | 30 | 15 | 15 | 15 | 30 | 6bBH | -756.366 | 142.673 | 41 | 4 | 22.5 | 3 | 0 | 3 | 5 |
| 16SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 15 | -15 | -15 | 30 | -30 | 30 | 30 | 30 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -765.836 | 237.368 | 22 | 24 | 23.0 | 2 | 0 | 2 | 3 |
| 17SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 15 | -15 | -30 | 30 | -30 | 30 | 0 | 30 | 0 | -30 | 30 | 30 | -30 | 30 | 15bBH | -767.536 | 280.685 | 18 | 30 | 24.0 | 0 | 0 | 0 | 5 |
| 18SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 15 | -15 | -30 | 30 | -30 | 30 | 15 | 30 | 0 | -30 | 30 | 30 | -30 | 30 | 15bBH | -766.737 | 307.050 | 19 | 33 | 26.0 | 1 | 0 | 1 | 6 |
| 19SBiH | -10 | 0 | -10 | 0 | 10 | -10 | -10 | 15 | -15 | -30 | 30 | -30 | 30 | -30 | 30 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -769.736 | 210.493 | 13 | 48 | 30.5 | 0 | 0 | 0 | 3 |
| 20SBiH | 10 | -10 | -10 | 0 | 0 | -10 | -10 | -15 | 0 | -15 | 0 | -30 | 30 | 0 | -15 | 0 | 30 | 15 | 15 | 15 | 15 | 6bBH | -753.026 | 107.909 | 53 | 22 | 37.5 | 1 | 0 | 1 | 9 |
| 21SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 0 | -15 | -15 | 30 | -30 | 30 | 30 | -15 | -15 | -30 | 30 | 30 | -30 | 15 | 15bBH | -756.734 | 188.158 | 40 | 38 | 39.0 | 0 | 0 | 0 | 3 |
| 22SBiH | -10 | -10 | -10 | 0 | 0 | -10 | -10 | -15 | 15 | 15 | -15 | -30 | 30 | 0 | 30 | 15 | 15 | 15 | 15 | 15 | 0 | 6bBH | -752.404 | 163.283 | 57 | 23 | 40.0 | 2 | 0 | 2 | 5 |
| 23SBiH | -10 | -10 | -10 | 0 | 10 | -10 | -10 | -15 | -15 | 15 | -30 | -30 | 30 | 0 | -15 | 0 | 0 | 0 | 30 | -30 | 0 | 15bBH | -766.603 | 116.528 | 20 | 62 | 41.0 | 1 | 0 | 1 | 10 |
| 24SBiH | -10 | 0 | -10 | 0 | 10 | -10 | 0 | 15 | -15 | -30 | 30 | -30 | 30 | 15 | 30 | -15 | -30 | 30 | 30 | -30 | 30 | 15bBH | -771.299 | 308.988 | 11 | 74 | 42.5 | 1 | 0 | 1 | 3 |
| 25SBiH | 0 | -10 | -10 | 0 | 0 | -10 | -10 | -15 | 15 | 0 | -15 | -30 | 30 | 0 | 0 | 0 | 30 | 15 | 15 | 15 | 15 | 6bBH | -751.434 | 114.189 | 60 | 25 | 42.5 | 1 | 0 | 1 | 6 |

Table 6.9: SuperBiHelix results from animal-plant comparative analysis of new homologized structures

that there are HB and SB conserved interactions found in class A GPCRs. Examples of these findings are the conserved Hydrogen bonds network 1.50-2.50-7.49, and salt bridges between helices 2-3, 3-4, 2-4 and 5-6 [73], (B-W numbering [83] has been employed). These interactions are shown in Figures 10.15-10.24 and their corresponding tables of Appendix VI., and summarized in table 6.10, were useful for the selection of the best structures to be characterized via MD. As expected some phylogenetic traits, such as the existence of 7TM helices and interactions between amino acids belonging to each helix, are present in GCR1, which is consistent with the proposal of GCR1 serving as a GPCR. These findings were taken into account to determine the real lengths for TM 5 and 6.

We compared all top 25 structures and selected the best 5 candidates, shown in Table 6.9, based on the presence of stabilizing interhelical SBs and HBs interactions. Finally, the five full structures were constructed by adding loops using PE2R3 template homology, followed by 10 steps of simulated annealing and energy minimization. The resulting structures were used for full membrane full solvent MD simulations.

By comparing interhelical ionic and polar interactions in Table 6.10, for all five GCR1 proposed structures, with those found in some helices pairs of animal GPCRs, we found some similarities between both class A GPCRs and GCR1, such as those found particularly between helices 5-6, but also between H3-H4-H2 networks, added to 1-2-7 HB network [79]. On the other hand, although we did not find any 3-6 ionic lock, we did find a similar interaction between H2-H4. All these criteria, added to the charged and neutral interhelical energy, the variability in the $\eta$, $\phi$, and $\theta$ angles, and the active and conserved contacts obtained from the SuperBiHelix results, were used to select the five structures previously described. These will be used as input for thermodynamic analysis in the next section.

| Interaction | SBs | HBs | 1SBiH | 8SBiH | 15SBiH | 20SBiH | 23SBiH |
|---|---|---|---|---|---|---|---|
| H5 - H6 | R187(5.67) - E211(6.30) | | Y | Y | Y | Y | Y |
| | R190(5.70) - E211(6.30) | | N | Y | Y | Y | Y |
| | R190(5.70) - D207(6.26) | | Y | Y | Y | Y | Y |
| | R194(5.74) - D204(6.23) | | Y | Y | Y | Y | Y |
| | | N191(5.71) - R209(6.28) | N | N | Y | Y | N |
| | | R194(5.74) - N208(6.27) | N | N | N | Y | N |
| | | Q184(5.64) - N216(6.35) | N | N | Y | N | N |
| | | R187(5.67) - N208(6.27) | Y | Y | N | N | Y |
| Intra H5 | | R193(5.73) - R194(5.74) | Y | N | N | N | N |
| | | R190(5.70) - R194(5.74) | N | N | Y | Y | N |
| Intra H6 | R202(6.21) - D201(6.20) | | Y | Y | Y | Y | Y |
| | | Q205(6.24) - R202(6.21) | Y | Y | N | N | Y |
| H2 - H3 - H4 | K53(2.41) - D118(4.40) | | Y | Y | Y | Y | Y |
| | R107(3.52) - E120(4.42) | | Y | Y | N | N | Y |
| | R107(3.52) - D118(4.40) | | N | N | Y | Y | N |
| | R107(3.52) - E117(4.39) | | Y | Y | Y | Y | Y |
| | K53(2.41) - E117(4.39) | | N | N | N | N | Y |
| | | S61(2.49) - Y126(4.48) | N | N | N | Y | N |
| | | T100(3.45) - H124(4.46) | N | N | N | N | Y |
| 1.50-2.50-7.49 | | S34(1.50) - D62(2.50) | Y | N | N | N | N |
| | | S34(1.50) - N265(7.49) | Y | Y | N | N | N |
| | | D62(2.50) - N265(7.49) | Y | Y | Y | Y | Y |

Table 6.10: Salt Bridges, Hydrogen bonds, and conserved contacts from Animal-GPCRs present in selected SuperBiHelix GCR1 structures

## 6.2 Loops reconstruction and annealing

To be able to perform the molecular dynamics characterization of GCR1, we had to reassemble helices with loops, which were previously removed for BiHelix and SuperBiHelix sampling. Consequently, as a first step for GCR1 loop reconstruction, we used loops in PE2R3 as the template to locate and attach residues belonging to the proposed loops to the helices part of GCR1. This procedure was carried out with each of the five GCR1 structures selected from the SuperBiHelix sampling. An important finding that we had into account to locate correctly some loop regions, was the alignment between CYS80 in TM3, (3.25) and CYS151 in extracellular loop 2, the well-conserved disulfide bond ECL2 found in both GCR1 and animal GPCRs [84] [11]. The alignment between PE2R3 and the five SuperBiHelix structures, thus as the loop reconstruction and the reassembly with the corresponding helices was carried out using maestro software [74]. Finally, to ensure the best 3D conformation for the proposed loops, we carried out 10 simulated annealing cycles, keeping the helices fixed. We start each cycle at 50 K, using 50 K temperature intervals to increase the temperature

up to 600 K, as the highest temperature. The number of steps at each temperature was 100 steps, where each step ran for 1 fs, which means we ran 0.1 ps at each temperature. The Forcefield is dreiding-0.3.par. Once the ten annealing cycles of each of the mentioned structures have been completed, we select the 5 best energetically annealed structures for their subsequent minimization, see Table 6.11. Figure 6.4 shows an example of one selected annealed structure.

| | | | Annealing E (kcal/mol) | | |
|---|---|---|---|---|---|
| Cycle | 1SBiH | 8SBiH | 23SBiH | 20SBiH | 15SBiH |
| 1 | -2.64055 | -77.9442 | -162.693 | -155.336 | -3.94571 |
| 2 | -116.409 | -63.2221 | -210.722 | -152.765 | -123.092 |
| 3 | -144.018 | -70.7617 | -233.652 | -187.935 | -133.444 |
| 4 | -167.795 | -92.1374 | -225.566 | -211.252 | -154.445 |
| 5 | -146.396 | -80.5565 | -254.661 | -252.708 | -168.456 |
| 6 | -153.773 | -103.826 | -255.747 | -253.441 | -160.213 |
| 7 | -151.435 | -88.0709 | -282.483 | -197.296 | -173.511 |
| 8 | -167.714 | -131.251 | -285.503 | -245.707 | -167.049 |
| 9 | -156.242 | -143.947 | -246.771 | -246.202 | -216.490 |
| 10 | -158.664 | -147.910 | -294.057 | -191.266 | -218.399 |
| | | | Best Anneal_Cycle E | | |
| | 4_cycle = -167.795 | 10_cycle = -147.91 | 10_cycle = -294.057 | 6_cycle = -253.441 | 10_cycle = -218.399 |
| | | | Minimized Energy | | |
| Selected to MD | -758.577 | -736.703 | -778.169 | -745.654 | -814.37 |

Table 6.11: Simulated Annealing results for 5 best-selected SuperBiHelix structures

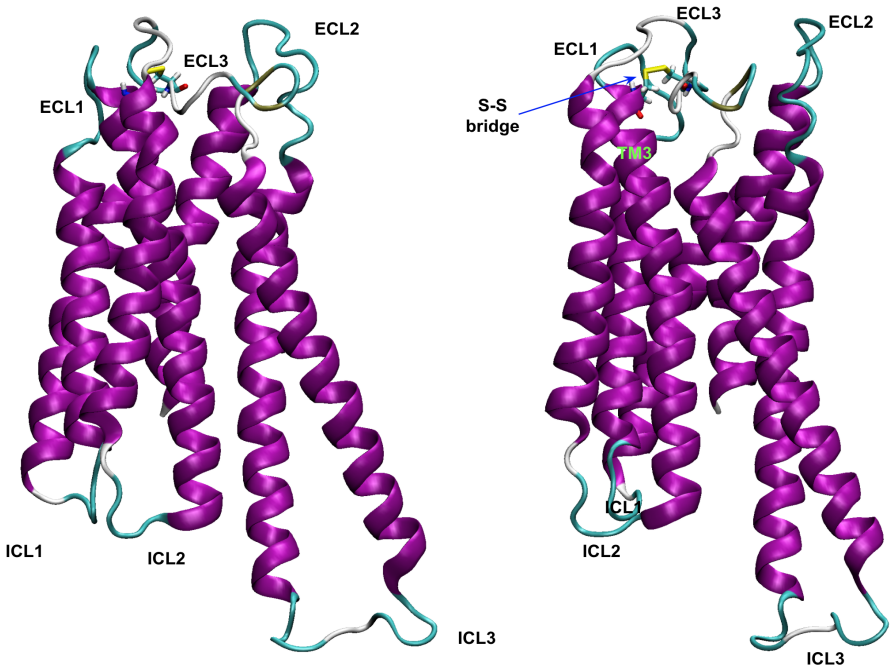


Figure 6.4: 1SBiH structure after 10 simulated annealing cycles and further minimization.

Figure 6.5 shows the best-annealed structures, by energy, and their corresponding minimized structures show small differences between them. The minimized structures were used in Molecular Dynamics simulations, after insertion in a lipid bilayer surrounded by water molecules and the inclusion of counter-ions to reflect the physiological cellular environment, as we will show in the next section.

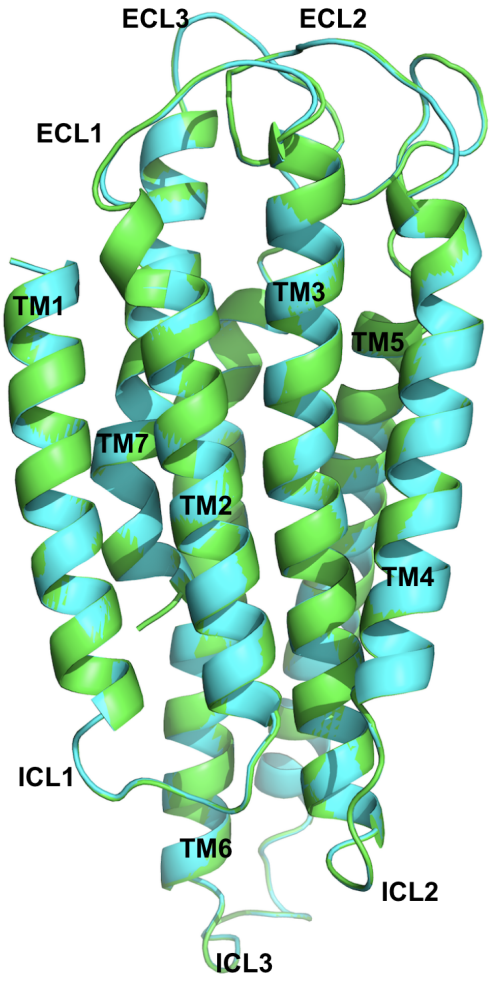


Figure 6.5: 1SBiH structure alignment of the fixed backbone annealed structure, in green, with their unfixed backbone minimized structure, in cyan. The RMSD to the reference structure is 0.179 for 3997 atoms aligned

## 6.3 Conformational analysis of the selected GCR1 structures in the physiological environment

As mentioned, the GCR1 candidate structures obtained from SuperBiHelix were reconstructed by adding the intracellular and extracellular loops. To carry out the MD studies in each reconstructed structure (1SBiH, 8SBiH,

15SBiH, 20SBiH and 23SBiH), it was necessary to simulate the membrane in the cellular environment and under physiological conditions, for which we inserted each of the five candidate structures in 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC), which represents the lipid bilayer of the cell membrane, followed by the addition of water molecules, TIP3, above and below the bilayer, as well as $Na^+$ and $Cl^-$ counterions to neutralize the entire system and simulate the physiological environment of the cell.

Once we had the protein into the membrane system, we minimized lipids, water, and ions, keeping the protein fixed, for 10000 steps of 1 fs each, to eliminate steric clashes between lipids, water, and ions. Next, we equilibrated for 500000 steps, of 1 fs each, at 310 K, lipids, water, and ions, again keeping the protein fixed. Once we had the lipids, water, and ions relaxed at 310 K, we carried out a new minimization and a new equilibration, at 310 K, but this time keeping the protein backbone fixed, *i.e.,* letting the protein side chains relax with lipids, waters, and ions. Finally, we simulated 60 ns of MD at 310 K keeping helices fixed (1,4-distances, angles, and dihedral values for Hydrogen bonds). In the final simulation processes, we removed any constraints used.

Figure 6.6 shows the values of electrostatic, Van-der-Waals and total energy for the five equilibrated structures. We use these comparative energies to select the most stable predicted structures for further analysis to allow us to select the best one to proceed with the docking with abscisic acid, ABA, and Gibberellin A1, GA1.

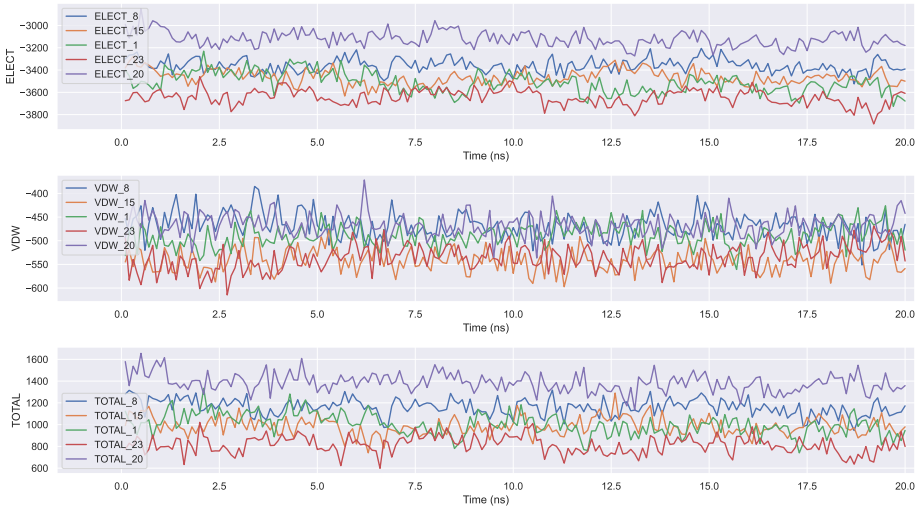


Figure 6.6: Electrostatic, vdW, and Total energies for the last 20 ns of MD simulations of 1SBiH in green, 23SBiH in red, 15SBiH in orange, 8SBiH in blue, and 20SBiH in purple, equilibrated at 310 K and at 1 atm.

As a result, we can see that structures 1SBiH in green, and 23SBiH in red, are on average the two best structures by total energy.

RMSD values are analyzed individually showing that all five structures converged around the last 2 ns of the total of 60 ns, as can be seen below. Convergence criteria was set to RMSD below 1.5 Å

Finally, we downloaded the trajectory files for each equilibrated structure and analyzed them one by one, to confirm that the interactions found in the previous SuperBiHelix step remained even after the MD simulation, or perhaps new hydrogen bonds, HBs, or salt bridges, SBs were formed.

Below we show a comparative analysis of structures 1SBiH and 23SBiH, as the two best final structures obtained, after completing the MD simulations of each structure, under NPT conditions (P = 1.04 atm and T = 310 K), and the analysis performed for the others is shown in Appendix VII.

### 6.3.1 Structural analysis of the equilibrated GCR1 candidate structure, 1SBiH

Figure 6.7, Panel A) shows the predicted GCR1 structure, 1SBiH, equilibrated at 310 K, and embedded in a POPC lipid bilayer. In addition, panel

B) shows the top view with the numbering of each helix, describing their corresponding orientations in the membrane. The RMSD shown in Figure 6.8 reports that the predicted GCR1 structure, 1SBiH, converged during the last 2 ns of the 60 ns MD simulation, under NPT, 310 K, and 1 atm conditions.

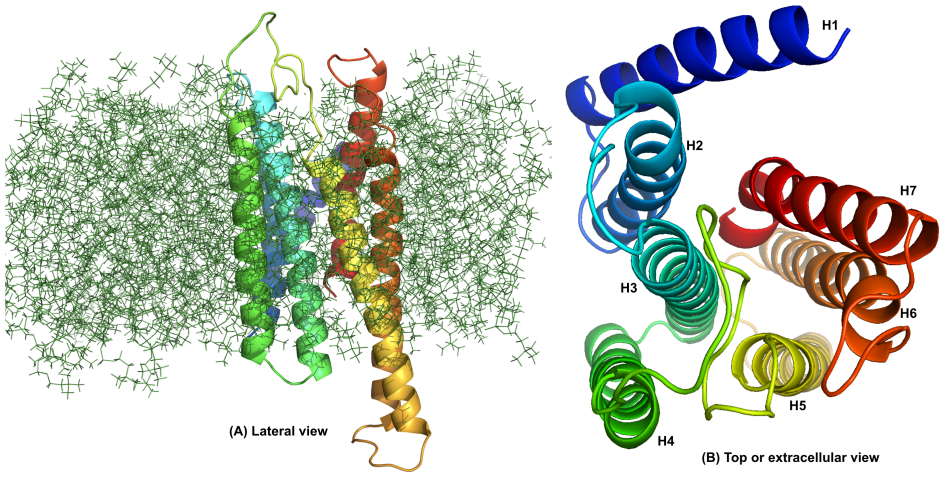


Figure 6.7: Equilibrated GCR1 structure, 1SBiH, embedded in a POPC lipid bilayer. Panel (A) shows a lateral view of the protein 1SBiH, and panel (B) shows the numbering of each helix and their spatial orientation.

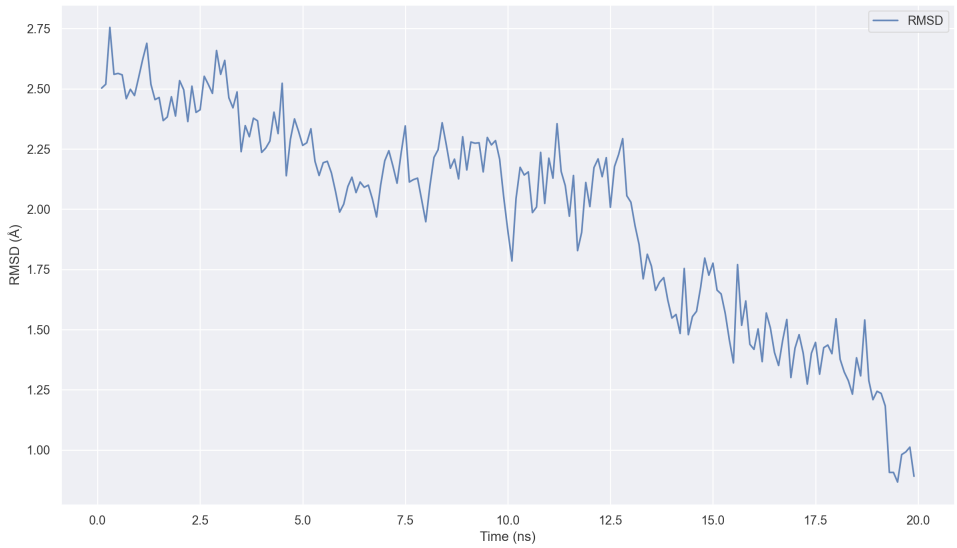


Figure 6.8: Equilibrated GCR1 structure, 1SBiH. Backbone RMSD for the last 20 ns of 60 ns of MD simulation for the predicted GCR1 structure, 1SBiH. Around last 2 ns are fully converged.

Figure 6.9 shows two SB networks found, in the boundary bilayer–cytoplasm, between D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74),E117(4.39)-R107(3.52), E120(4.42)-R107(3.52), D118(4.40)-K53(2.41), which are the same interactions found in the 7TM bundle obtained from 1SBiH SuperBiHelix. Interestingly, after the MD simulation, we observe that structure 1SBiH preserves all SBs found in the bundle structure from SuperBiHelix.

Figures 6.10 and 6.11, enable us to analyze SBs distances and their stabilities, for the complete last 20 ns of trajectory, besides identifying whether some of them are water-mediated SBs interactions.

Figure 6.10, shows 3 stable SBs between helices 5 and 6, D207(6.26)-R190(5.70), E211(6.30)-R187(5.67) and E211(6.30)-R190(5.70), and two stable SBs between between helices 3 and 4, E117(4.39)-R107(3.52) and E120(4.42)-R107(3.52).

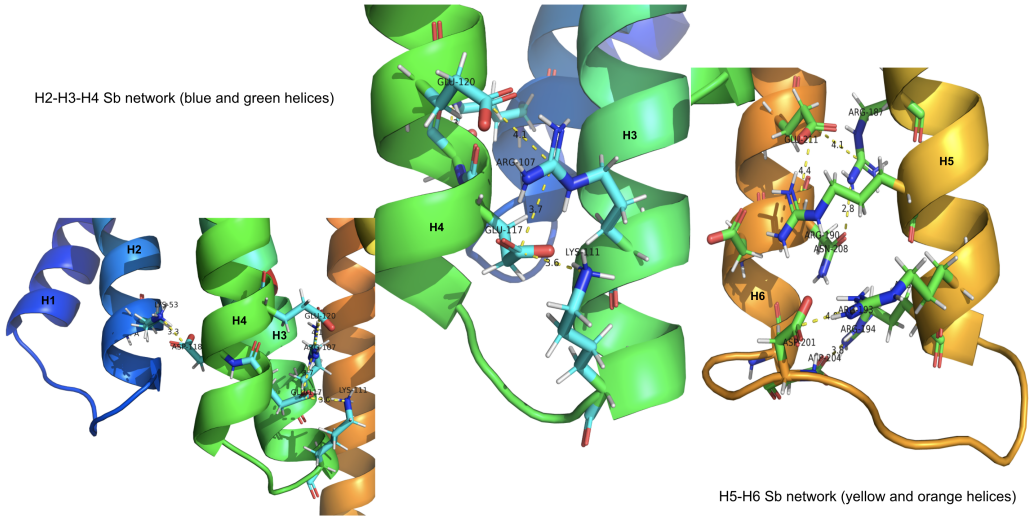


Figure 6.9: Equilibrated GCR1 structure, 1SBiH. A molecular view of the interhelical salt bridge networks found between helices 2-3-4 and 5-6 in the predicted GCR1 structure, 1SBiH.

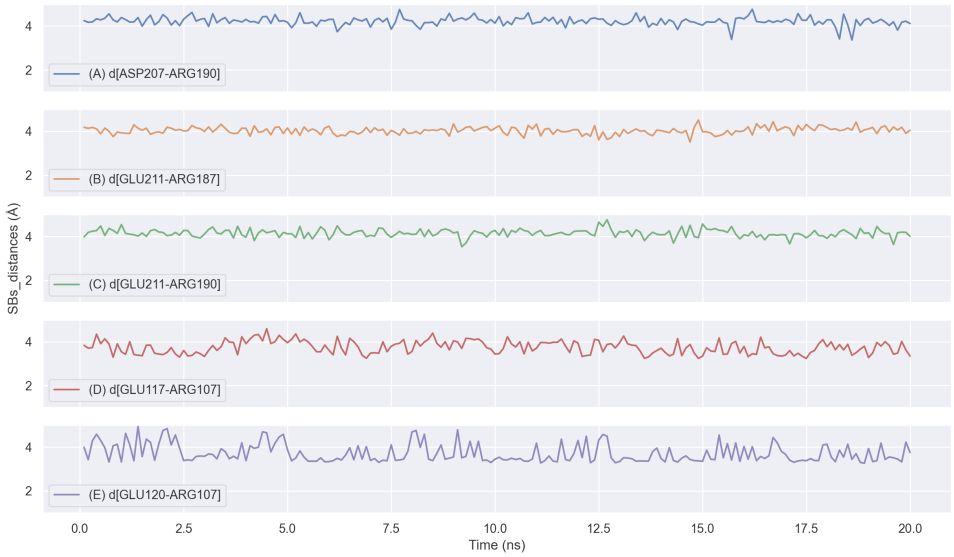


Figure 6.10: Equilibrated GCR1 structure, 1SBiH. Time evolution of the distances of stable interhelical salt bridges between residues for D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), E117(4.39)-R107(3.52), and, E120(4.42)-R107(3.52). Where $d[i-j]$, (Å) $i$ and $j$ of 1SBiH. We consider that stable SBs are less than 5 Å.

On the other hand, Figure 6.11 shows, in panels C) and D), two transient SBs which stabilize at the end part of the simulation. Panel D) shows an interaction between helix 5 and ICL3, where the electrostatic interaction with a highly flexible residue belonging to a highly flexible loop can explain this SB distance behavior. On the other hand, panels A), B), and E), show three SBs, that are stable during most of the simulation time. Panels A), B), and C) have as a particular trait, a LYS doing part of those interactions. LYS residues have a large side chain which can cause high mobility. Since panels A) to E), show good stability during most parts of the simulation, we considered those as stable SBs. However, we depicted them as unstable, because of the few fluctuations in their trajectory.

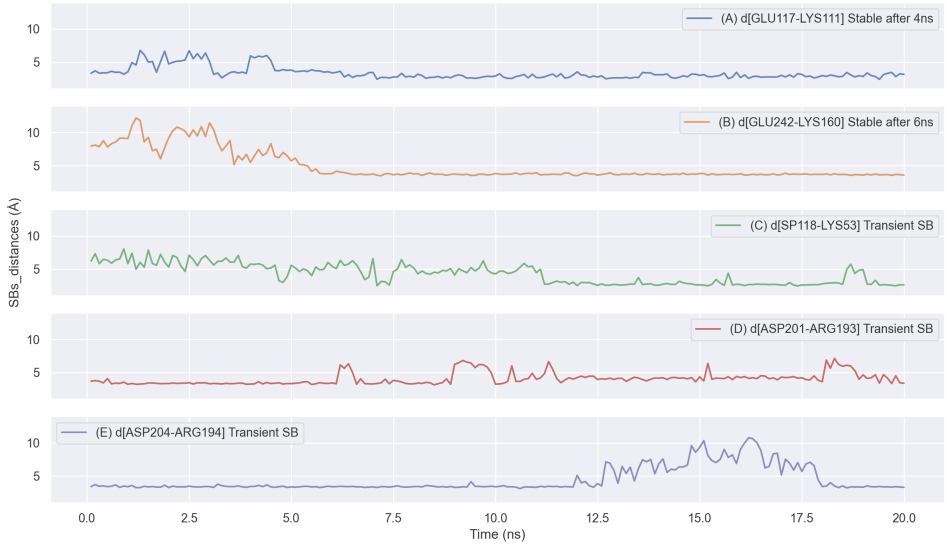


Figure 6.11: Equilibrated GCR1 structure, 1SBiH. Time evolution for the last 20 ns out of 60 ns of transient salt bridges. The salt bridge in (A) [E117(4.39)-K111(ICL2)] after 3.3 ns stabilizes out of unstable, one in (B) [E242(ECL3)-K160(ECL2)] after 5.5 ns stabilizes out of unstable, one in (C) [D118(4.40)-K53(2.41)] after 11.3 ns stabilizes out of unstable, however ones in (D) [D201(ICL3)-R193(5.73)] and (E) [D204(6.23)-R194(5.74)] are still marginal stable salt bridges.

Finally, we found one HB in the predicted GCR1 structure, 1SBiH, between N208(6.27)-R187(5.67), also part of the 5-6 SB network. This HB is stable during most of the simulation's time evolution as seen in Figure 6.12. By comparing the equilibrated structure, 1SBiH, with its corresponding 7TM bundle from SuperBiHelix, we observed that we have all previously found SBs, after finishing the MD simulation.

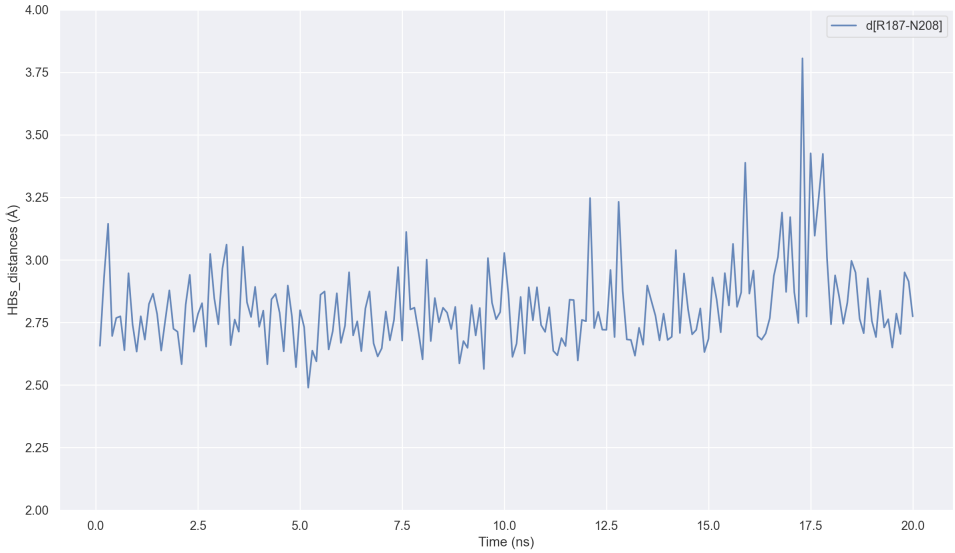


Figure 6.12: Equilibrated GCR1 structure, 1SBiH. Distance time evolution for the interhelical hydrogen bond found between N208(6.27)-R187(5.67).

### 6.3.2 Structural analysis of the equilibrated GCR1 candidate structure, 23SBiH

Figure 6.13, Panel (A) shows the lateral view of the equilibrated GCR1 structure, 23SBiH, at 310 K and a pressure of 1 atm with the numbering for each helix. Panel B) on the other hand, shows structure 23SBiH embedded the POPC lipid bilayer. In addition, Figure 6.14 shows the RMSD backbone, for the last 20 ns of the 40 ns MD simulation in the 23SBiH structure. The structure is fully equilibrated over the previous 4.5 ns.

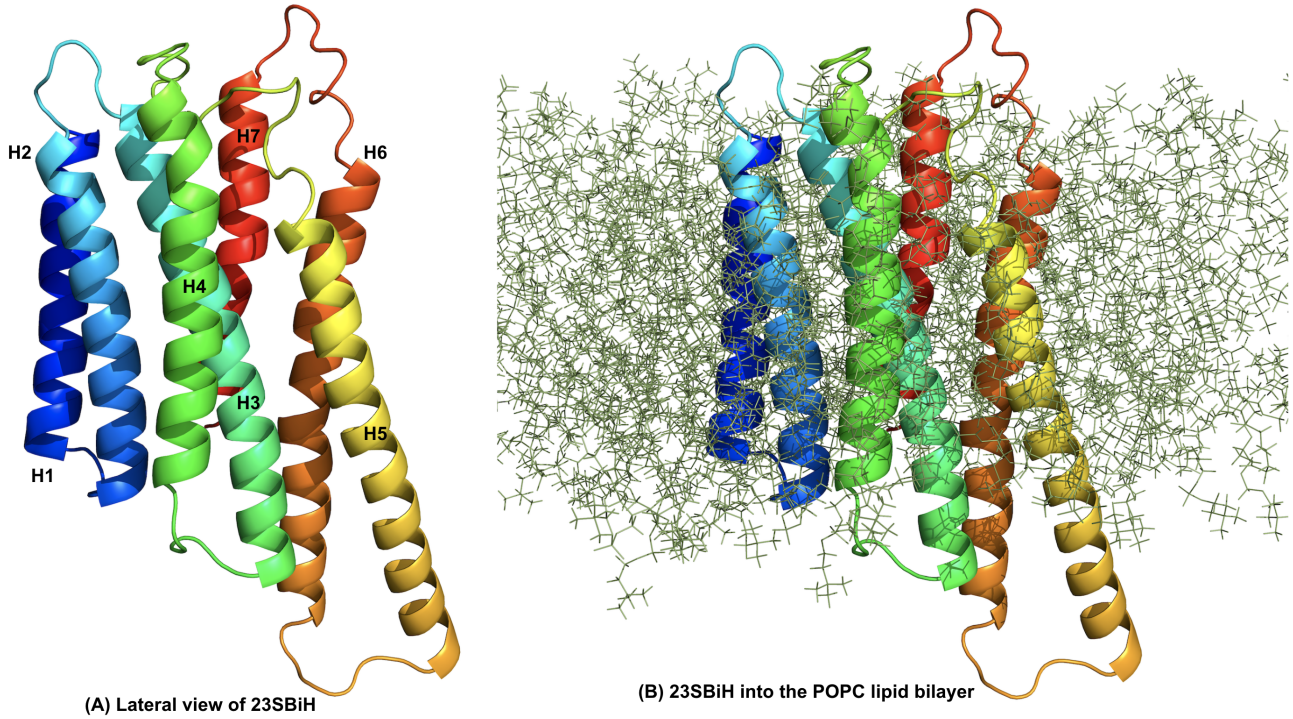


Figure 6.13: Equilibrated GCR1 structure, 23SBiH, at 310 K and 1 atm., embedded in the POPC lipid bilayer is shown in Panel (B). Panel (A) shows the side view with its respective color codes and numbers for each helix. (H1 = blue, H2 = cyan, H3 = teals, H4 = green, H5 = yellow, H6 = orange, and H7 = red).

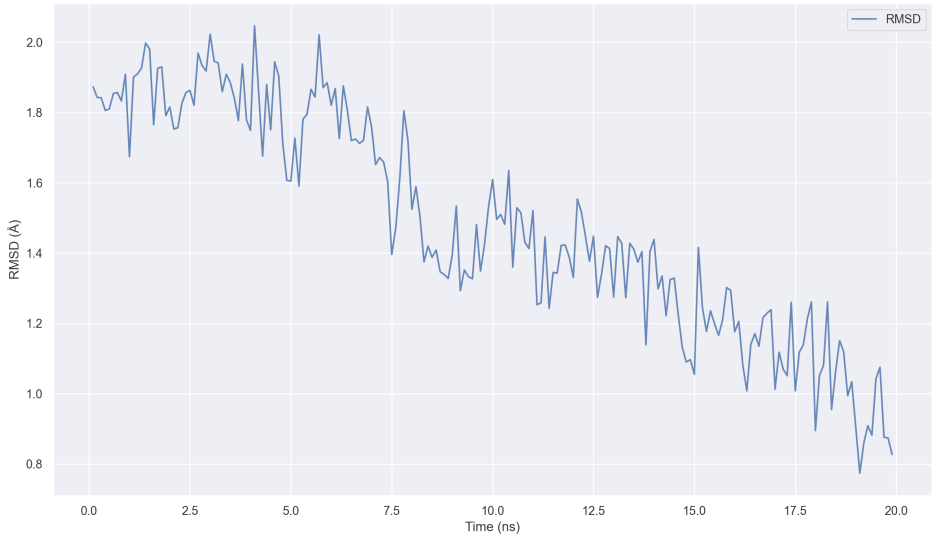


Figure 6.14: Equilibrated GCR1 structure, 23SBiH. Calculated backbone RMSD for the last 20 ns out of 40 ns for the predicted GCR1 structure, 20SBiH. Around the last 4.5 ns are fully converged.

Figure 6.15 for the 23SBiH structure displays in Panel A) two stable SBs networks, one of them between H2-H3-H4 structured by E117(4.39)-K53(2.41) and E120(4.42)-R107(3.52), and another between H5-H6, structured by E211(6.30)-R187(5.67) and E211(6.30)-R190(5.70). On the other hand, panel B) show an unstable SB present in ICL2 between D115 and K111, and finally, panel C) displays an unstable SB network between H5-H6, structured by D207(6.26)-R190(5.70), D204(6.23)-R190(5.70), and D204(6.23)-R194(5.74). The corresponding status as stable or unstable can be seen in Figures 6.16, and 6.17. We assume, for further analysis, that SBs in panels (A), (D), and (E) in Figure 6.17 can be considered stable since most of the simulation time was below 5 Å with a few exceptions.

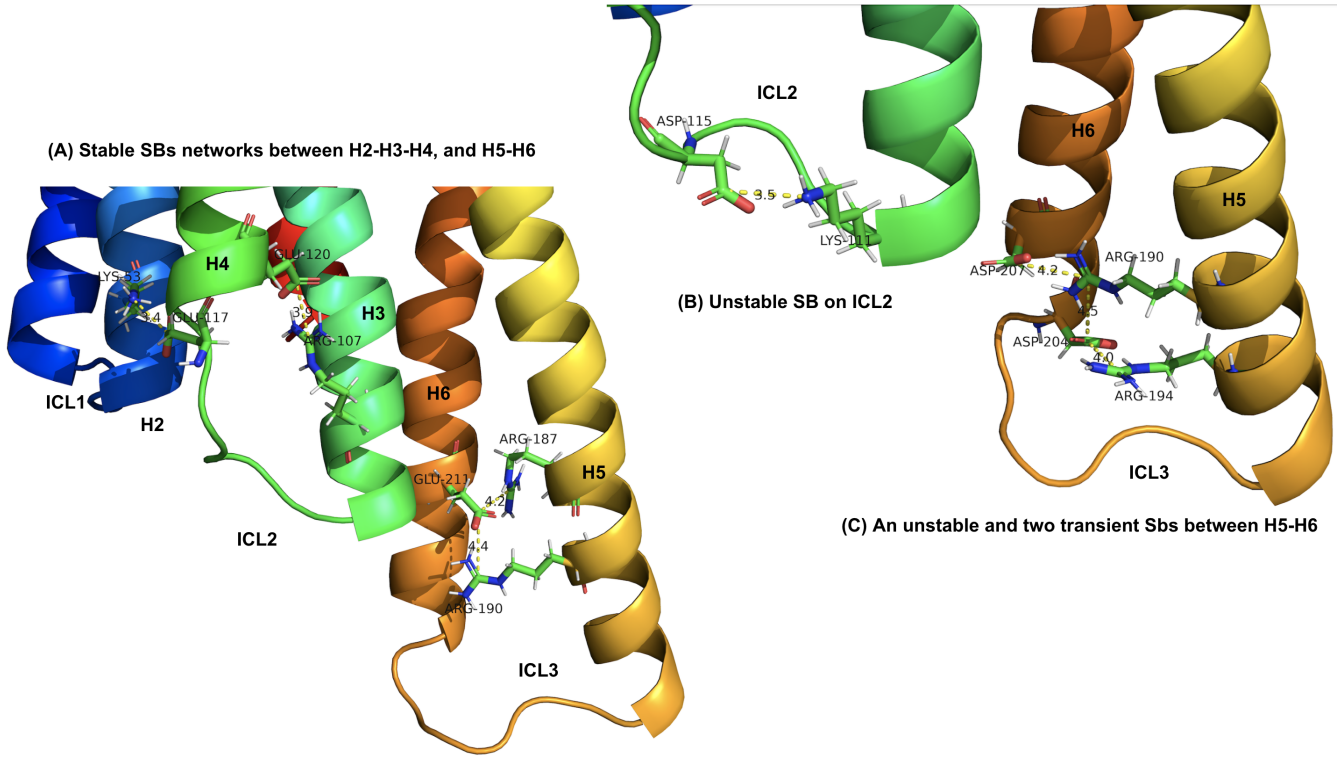


Figure 6.15: Equilibrated GCR1 structure, 23SBiH. Panel A) shows the molecular view of two stable SBs networks, one of them between H2-H3-H4 structured by E117(4.39)-K53(2.41) and E120(4.42)-R107(3.52), and another between H5-H6, structured by E211(6.30)-R187(5.67) and E211(6.30)-R190(5.70). Panel B) show an unstable SB present in ICL2 between D115 and K111, and finally, panel C) displays an unstable SB network between H5-H6, structured by D207(6.26)-R190(5.70), D204(6.23)-R190(5.70), and D204(6.23)-R194(5.74).

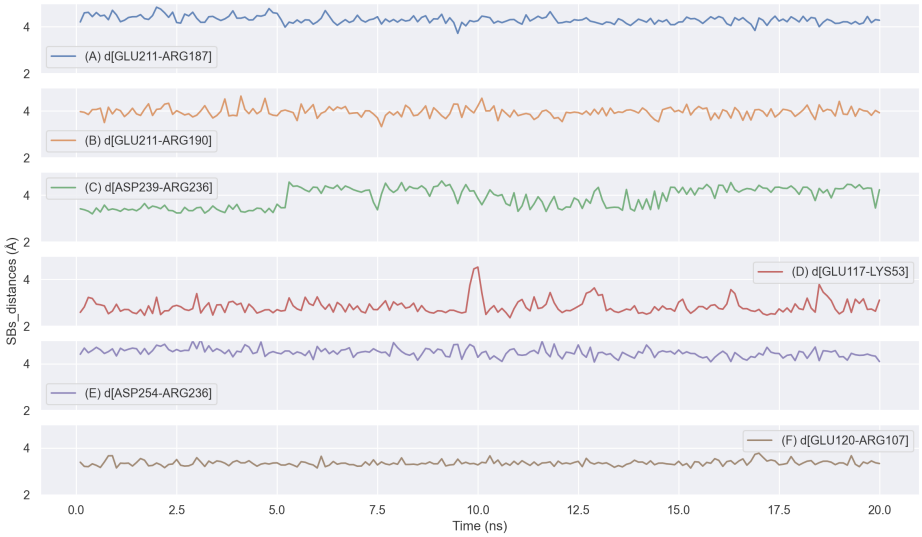


Figure 6.16: Equilibrated GCR1 structure, 23SBiH. Time evolution of stable SB for the last 20 ns of a total of 40 ns of MD simulation. (A) E211(6.30)-R187(5.67), (B) E211(6.30)-R190(5.70), (C) D239(EC3)-R236(EC3), (D) E117(4.39)-K53(2.41), (E) D254(7.38)-R236(EC3), and (F) E120(4.42)-R107(3.52). We consider that stable SBs are less than 5 Å.

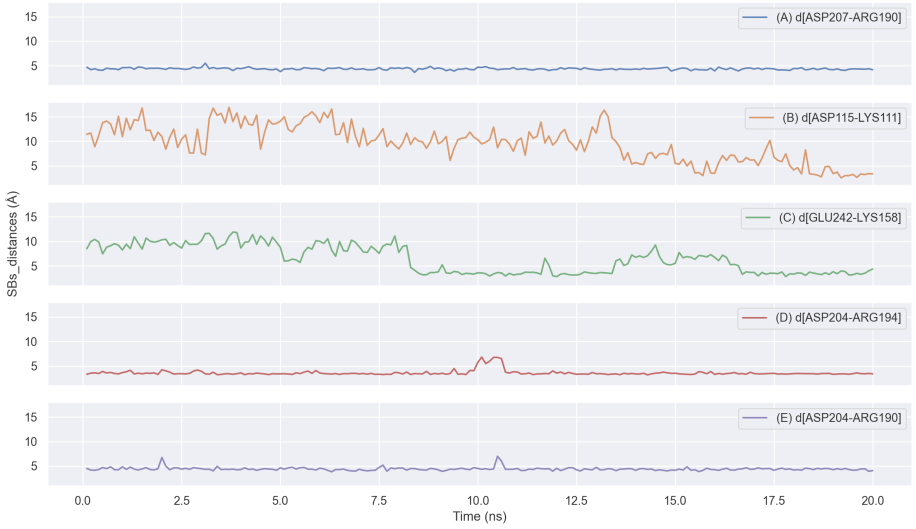


Figure 6.17: Equilibrated GCR1 structure, 23SBiH. Time evolution of transient salt bridges, for the last 20 ns out of a total of 40 ns of MD simulation. (A) D207(6.26)-R190(5.70), (B) D115(ICL2)-K111(ICL2), (C) E242(ECL3)-K158(ECL2), (D) D204(6.23)-R194(5.74), and (E) D204(6.23)-R190(5.70). We consider that stable SBs are less than 5 Å..

By comparing the equilibrated structure, 23SBiH with its corresponding 7TM bundle from SuperBiHelix, we observed that we lost the SB between K53(2.41)-D118(4.40), E117(4.39)-R107(3.52), but we gained the SB between D204(6.23)-R190(5.70), after finished the MD simulation.

Finally, we summarize in Table 6.12, the stable and transient salt bridges present in all five predicted and equilibrated structures for GCR1, named 1SBiH 23SBiH, 8SBiH, 15SBiH, and 20SBiH.

| | | 1SBiH | | 23SBiH | | 8SBiH | | 15SBiH | | 20SBiH | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| interacting regions | Salt bridges | Stable | Transient | Stable | Transient | Stable | Transient | Stable | Transient | Stable | Transient |
| TM4-TM2 | GLU117(4.39)-LYS53(2.41) | | | X | | | | | | | |
| | ASP118(4.40)-LYS53(2.41) | X | | | | | | | | | |
| TM4-TM3 | GLU117(4.39)-ARG107(3.52) | X | | | | X | | | X | | |
| | ASP118(4.40)-ARG107(3.52) | | | | | | | | X | X | |
| | GLU120(4.42)-ARG107(3.52) | X | | X | | X | | | | | |
| TM6-TM5 | ASP201(ICL3)-ARG193(5.73) | X | | | | | | | | | |
| | ASP204(6.23)-ARG194(5.74) | X | | X | | | | | | X | |
| | ASP204(6.23)-ARG190(5.70) | | | X | | | | X | | X | |
| | ASP207(6.26)-ARG190(5.70) | X | | X | | | | | | | X |
| | GLU211(6.30)-ARG187(5.67) | X | | X | | | | X | | | |
| | GLU211(6.30)-ARG190(5.70) | X | | X | | X | | X | | | |
| | GLU211(6.30)-ARG194(5.74) | | | | | X | | | | | |
| ICL1-ICL1 | GLU46(ICL1)-ARG48(ICL1) | | | | | | | | | | X |
| ICL2-ICL2 | ASP115(ICL2)-LYS111(ICL2) | | | | X | | | | | | |
| ECL3-ECL2 | ASP239(ECL3)-LYS158(ECL2) | | | | | | X | | | | |
| | GLU242(ECL3)-LYS158(ECL2) | | | X | | | X | | | | |
| | GLU242(ECL3)-LYS160(ECL2) | X | | | | | | | | | |
| ECL3-ECL3 | ASP239(ECL3)-ARG236(ECL3) | | | X | | X | | | | | |
| | GLU242(ECL3)-ARG236(ECL3) | | | | | | | | | | X |
| ICL1-TM6 | GLU46(ICL1)-ARG217(6.36) | | | | | | X | | | | |
| TM4-ICL2 | GLU117(4.39)-LYS111(ICL2) | X | | | | | X | | | | |
| TM7-ECL3 | ASP254(7.38)-ARG236(ECL3) | | | X | | | | | X | | |

Table 6.12: Stable and Transient salt bridges of SBiH models equilibrated of GCR1.

As a result, based on (i) A detailed inspection of each structure, (ii) followed by the comparison of interhelical polar and ionic interactions found in TM bundles from SuperBiHelix, versus those interactions found after equilibration of each equilibrated structure, and (iii) using Table 6.12 to compare among SBs found in each of the 5 new constructed GCR1 structures, along with (iv) the analysis of charged and neutral interhelical energy shown in Figure 6.6, we found that structures 1SBiH and 23SBiH are not only the two best energetic structures but also are the two structures with the bigger amount of stable and transient SB networks. However, the predicted GCR1 structure identified as 1SBiH, has more stable interhelical SBs than structure 23SBiH, and since the best interhelical SB networks may confer bigger folding stability, than loop-loop or loop-helical interactions, we selected the predicted structure 1SBiH, as the best candidate to be docked with both hormones ABA and GA1[85], as can be seen, below.

## 6.4 Regulation of GCR1 by phytohormones

To further examine its role as a true GPCR, we docked the GCR1 structure with both phytohormones, generating around 50 thousand poses for each ligand, in the binding site. We then used the cavity and snapBE energies to select the best-docked structures to run MD simulations. Finally, we analyzed the equilibrated structures.

### 6.4.1 ABA-GCR1 docking

We started downloading the 3D ABA structure from PubChem [86]. Next, using Maestro, we generated 4 charged conformers plus a neutral structure.

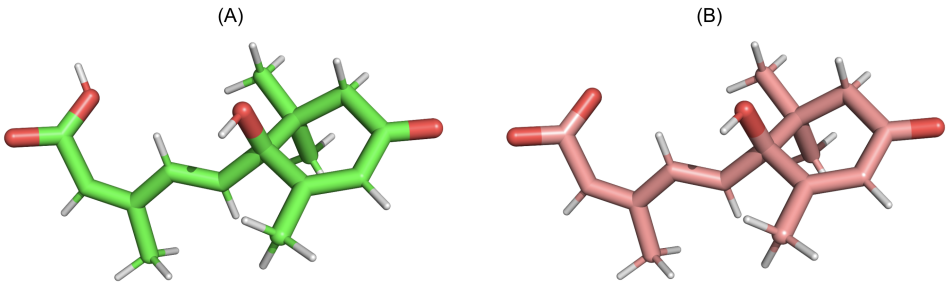


Figure 6.18: Panels (A) and (B) show the bgf files of neutral and charged.xray conformers for ABA, respectively

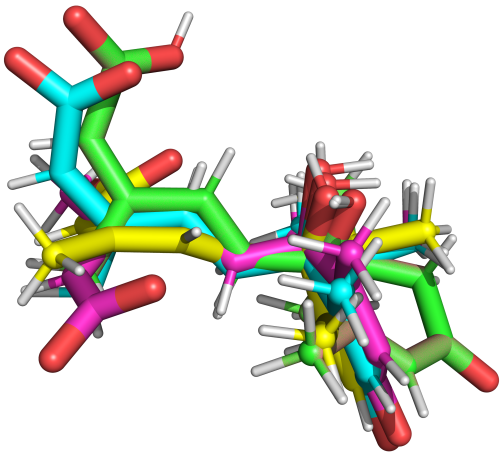

Figure 6.19: Superimposed bgf files of charged and neutral conformers for ABA. ABA_ntl (green) , ABA.no_0.bgf (blue), ABA.no_5.bgf (pink), ABA.no_11.bgf (yellow), and ABA.xray (salmon).

**Docking results for the HSP-88 screamed and alanized-GCR1 with ABA.**

Here, we also analyzed the top 10 ABA docked structures for the 10 HSP88-screamed GCR1 conformations. After testing more than fifty thousand poses of the four ABA conformers on each HSP88-screamed-GCR1 structure, we selected ten structures based on the cavity binding energy and snap binding energy. In addition, we used the number of salt bridges, hydrogen bonds, and hydrophobic contacts to select the best 5, yellow-highlighted structures shown in Table 6.13, as can be seen in Figures 10.43-10.53 of Appendix VIII. Only two conformers, ABA.no_0 and ABA.no_5, of the four docked ABA conformers, made part of the top ten structures.

Finally, using ligand interaction diagrams, which are shown at the final part of Appendix VIII., we identified a series of HBs interactions between carbonyl groups of ABA and amino acids HSP88 (3.33), R138 (4.60), H163 (5.43), F164 (5.44), and Y168 (5.48) on GCR1. Also, we identified one SB interaction between the ABA-carboxylate group and R138 (4.60), in two of the selected docked GCR1 structures. Tabla 6.14 summarizes types of interactions founded on the pharmacophore, for the top-ten docked structures.

**Dealanization, binding site annealing, and minimization of selected ABA Docked structures**

Next, we carried out the binding site annealing to the five selected and dealanized structures of GCR1, generating ten annealed structures of each docked structure. Finally, all 50 resulting structures were minimized. Table 6.15 shows the top 10 minimized structures selected, based on cavity binding energy and snapbe (snap binding) energy, for further analysis of interacting HBs and SBs in the pharmacophore, as can be seen in Figures 10.64 to 10.73, in Appendix IX.

| SCREAM | Pose | ucav | ucav_rank | Sum_Ranks | cpx_te | lig_te | ptn_te | snapbe | snapbe_rank | ucav_cou | ucav_hb | ucav_vdw |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| best_1 | *.ABA.no_0.rnd_1.c23860.h.c | -90.983 | 1 | 4 | -504.714 | 30.761 | -423.660 | -111.815 | 3 | -66.613 | -9.401 | -14.971 |
| best_5 | *.ABA.no_5.rnd_1.c31387.t.c | -90.794 | 4 | 5 | -490.847 | 39.845 | -417.558 | -113.134 | 1 | -62.021 | -5.994 | -22.779 |
| best_1 | *.ABA.no_0.rnd_1.c42248.th.c | -90.830 | 3 | 5 | -476.087 | 30.597 | -394.799 | -111.885 | 2 | -67.772 | -13.439 | -9.618 |
| best_10 | *.ABA.no_0.rnd_1.c24691.h.c | -89.857 | 5 | 11 | -490.965 | 30.545 | -409.974 | -111.536 | 6 | -65.023 | -9.886 | -14.942 |
| best_6 | *.ABA.no_0.rnd_1.c306.th.c | -88.691 | 11 | 19 | -471.789 | 30.651 | -391.487 | -110.953 | 8 | -69.230 | -15.260 | -4.204 |
| best_9 | *.ABA.no_0.rnd_1.c24494.h.c | -89.002 | 10 | 21 | -511.903 | 30.745 | -432.024 | -110.624 | 11 | -66.015 | -9.348 | -13.634 |
| best_3 | *.ABA.no_0.rnd_1.c40267.th.c | -88.360 | 15 | 27 | -487.378 | 30.590 | -407.614 | -110.354 | 12 | -65.744 | -13.555 | -9.058 |
| best_4 | *.ABA.no_0.rnd_1.c24315.h.c | -89.060 | 8 | 28 | -487.897 | 30.786 | -409.336 | -109.347 | 20 | -64.286 | -9.280 | -15.493 |
| best_9 | *.ABA.no_5.rnd_1.c20724.h.c | -88.599 | 12 | 29 | -498.936 | 39.274 | -428.791 | -109.419 | 17 | -63.246 | -5.042 | -20.311 |
| best_2 | *.ABA.no_5.rnd_1.c18183.ft.c | -87.952 | 20 | 30 | -487.608 | 39.531 | -416.453 | -110.686 | 10 | -63.362 | -5.021 | -19.568 |

Table 6.13: Docking results for the 10 screamed and alanized GCR1 structures with ABA. Table show the top 10 structures of combined 10 docking procedures, for each of the 10 screamed HPS-88 residues of GCR1 with 4 ABA conformers. We found only two ABA conformers in the top ten docked structures, ABA.no_0 and ABA.no_5. * = Alanized GCR1

| | | HB ABA - GCR1 | | | | | | SB ABA - GCR1 | |
|---|---|---|---|---|---|---|---|---|---|
| SCREAM | Pose | C=O - R138 | C=O - H163 | COO - H163 | C=O - F164 | COO - F164 | COO - Y168 | COO - HSP88 | COO - R138 |
| best_1 | *.ABA.no_0.rnd_1.c23860.h.c | X | | X | | | | X | |
| best_5 | *.ABA.no_5.rnd_1.c31387.t.c | | X | | X | | | X | |
| best_1 | *.ABA.no_0.rnd_1.c42248.th.c | | X | | X | | | X | X |
| best_10 | *.ABA.no_0.rnd_1.c24691.h.c | X | | X | | | | X | |
| best_6 | *.ABA.no_0.rnd_1.c306.th.c | X | | X | | X | | | |
| best_9 | *.ABA.no_0.rnd_1.c24494.h.c | X | | X | | | | X | |
| best_3 | *.ABA.no_0.rnd_1.c40267.th.c | X | | X | | X | | X | |
| best_4 | *.ABA.no_0.rnd_1.c24315.h.c | X | | X | | | | X | |
| best_9 | *.ABA.no_5.rnd_1.c20724.h.c | | X | | | | X | X | |
| best_2 | *.ABA.no_5.rnd_1.c18183.ft.c | | X | | | | X | X | X |

Table 6.14: Pharmacophore for ABA in alanized-GCR1 shows hydrogen bonds and salt bridges found between ABA and GCR1. * = Alanized GCR1. All measures are between C-C and C-N.

| Docking structure | COMPLEXMINIMIZE | ucav | ucav_rank | Sum_Rank | snapbe | snapbe_rank |
|---|---|---|---|---|---|---|
| b9_5_20724 | 9f.cm.bgf | -94,6350 | 1 | 2 | -139,6570 | 1 |
| b4_0_24315 | 8f.cm.bgf | -89,5630 | 4 | 7 | -130,8291 | 3 |
| b9_5_20724 | 8f.cm.bgf | -87,6640 | 7 | 9 | -130,9575 | 2 |
| b5_5_31387 | 10f.cm.bgf | -91,4310 | 2 | 11 | -129,7447 | 9 |
| b9_5_20724 | 5f.cm.bgf | -87,6430 | 8 | 13 | -130,6858 | 5 |
| b9_5_20724 | 1f.cm.bgf | -88,5250 | 5 | 13 | -129,7852 | 8 |
| b9_5_20724 | 4f.cm.bgf | -87,2660 | 10 | 14 | -130,7431 | 4 |
| b5_5_31387 | 9f.cm.bgf | -90,1360 | 3 | 15 | -128,6047 | 12 |
| b9_5_20724 | 7f.cm.bgf | -87,4220 | 9 | 16 | -129,8736 | 7 |
| b9_5_20724 | 3f.cm.bgf | -85,6010 | 16 | 22 | -130,5741 | 6 |

Table 6.15: Best top ten minimized structures, based on cavity binding energy and snapbe energy

Here we show a short list with specific interactions found in the pharmacophores (Appendix A) for each top ten structure. Table 6.16 summarizes a series of Hydrogen bonds located between the hydroxyl, carbonyl, and carboxylate groups of ABA with amino acids Y81 (3.26), R138 (4.60), S139 (4.61) ), V162(5.42), H163(5.43), and L165(5.45) of GCR1. Additionally, a salt bridge between the carboxylate group of ABA and HSP88 of GCR1.

After analyzing these results, we can conclude that structure b9_5_20724-9f.cm.bgf has the best cavity energy and position of ABA in the proposed binding site. However, we carried out a molecular dynamics simulation, and the results are shown in the next section. By comparing Table 6.14 with Table 6.16 we saw that the COO(ABA)-R138(GCR1) interaction is not present in any of the top ten of best ucav and snapbe energy structures, reflecting that the best position for a salt bridge is when HSP88 is close to the carboxylate group in ABA. Furthermore, the location of ABA also generates good and new stabilizing interactions of HBs.

| | | $H_b ABA-GCR1$ | | | | | | | | | $S_b ABA-GCR1$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Docking structure | COMPLEX MINIMIZE | OH - Y81 | OH - S139 | CO - V162 | CO - H163 | C=O - F164 | COO - L165 | COO - F164 | COO - H163 | CO - R138 | COO - HSP88 |
| b9_5_20724 | 9f.cm.bgf | X | X | | X | X | | | | | X |
| b4_0_24315 | 8f.cm.bgf | | | | | | X | X | X | X | X |
| b9_5_20724 | 8f.cm.bgf | | X | | X | X | | | | | X |
| b5_5_31387 | 10f.cm.bgf | | X | X | | | | | | | X |
| b9_5_20724 | 5f.cm.bgf | | | | X | X | | | | | X |
| b9_5_20724 | 1f.cm.bgf | | | | X | X | | | | | X |
| b9_5_20724 | 4f.cm.bgf | | | | X | X | | | | | X |
| b5_5_31387 | 9f.cm.bgf | | X | X | | | | | | | X |
| b9_5_20724 | 7f.cm.bgf | | X | | X | X | | | | | X |
| b9_5_20724 | 3f.cm.bgf | | | | X | X | | | | | X |

Table 6.16: Salt bridges and hydrogen bonds in the best top ten minimized structures, based on cavity binding energy and snapbe energy

### 6.4.2 GA1-GCR1 docking.

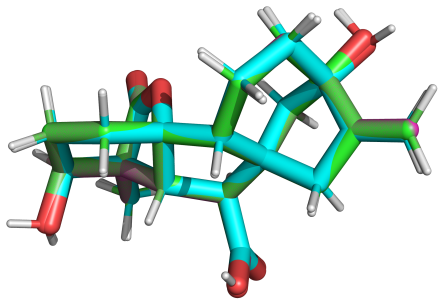

Figure 6.20: Superimposed bgf files of charged and neutral conformers for GA1, In green = GA1_ntl, blue = GA1.no_0.bgf, and pink = GA1.xray.bgf

**Docking results for alanized-GCR1 and GA1.**

Additionally, we analyzed the top 10 GA1 docked structures for the 10 HSP88-screamed GCR1 conformations. In a similar way described previously, we tested more than fifty thousand poses of the two GA1 conformers on each screamed-GCR1 structure and selected the top ten structures listed in table 6.17, with the best ucav and snapbe energies. The best docking sites were determined based on the cavity binding and snapbe energy. By last, through the pharmacophore analysis shown in Figures 10.84 to 10.94, along with Figures 10.74-10.80, shown in of Appendix X., we selected the best six, yellow-highlighted structures showed in Table 6.17, based on the number of salt bridges, hydrogen bonds, and hydrophobic contacts. Two GA1 conformers make part of the top ten, GA1.no_0 and GA1.xray. The selected structures were further annealed and minimized, and the results are shown later.

Last, the complete set of structures shown in table 6.17 containing Darwin Dock results were dealanized, followed by a binding site annealing and further minimization to be able to select the best structures. The selection was based on the cavity energy and binding site analysis. The selected structures were also further equilibrated, as seen in the next section.

| SCREAM | Pose | ucav | ucav_rank | Sum_Ranks | cpx_te | lig_te | ptn_te | snapbe | snapbe_rank | ucav_cou | ucav_HB | ucav_vdw |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| best_2 | *.GA1.xray.rnd_1.c32412.th.c | -85.910 | 3 | 18 | -408.265 | 71.730 | -363.869 | -116.126 | 15 | -67.335 | -11.595 | -6.979 |
| best_8 | *.GA1.xray.rnd_1.c32444.th.c | -84.601 | 6 | 22 | -454.215 | 71.853 | -410.490 | -115.578 | 16 | -66.941 | -11.542 | -6.117 |
| best_10 | *.GA1.xray.rnd_1.c845.th.c | -84.683 | 5 | 24 | -400.366 | 72.372 | -357.967 | -114.771 | 19 | -68.805 | -11.479 | -4.400 |
| best_10 | *.GA1.xray.rnd_1.c32809.h.c | -84.156 | 10 | 28 | -367.782 | 71.868 | -324.554 | -115.096 | 18 | -67.722 | -11.644 | -4.787 |
| best_10 | *.GA1.xray.rnd_1.c30864.th.c | -80.251 | 20 | 31 | -468.241 | 72.070 | -422.790 | -117.521 | 11 | -70.520 | -11.999 | 2.266 |
| best_3 | *.GA1.xray.rnd_1.c32054.th.c | -83.541 | 11 | 32 | -427.260 | 72.144 | -385.267 | -114.136 | 21 | -66.900 | -10.253 | -6.392 |
| best_1 | *.GA1.xray.rnd_1.c6914.h.c | -85.012 | 4 | 32 | -479.600 | 74.017 | -440.520 | -113.097 | 28 | -66.504 | -8.594 | -9.921 |
| best_9 | *.GA1.xray.rnd_1.c850.h.c | -82.335 | 14 | 37 | -462.923 | 72.704 | -421.819 | -113.808 | 23 | -68.936 | -11.876 | -1.523 |
| best_4 | *.GA1.no_0.rnd_1.c44628.h.c | -77.435 | 38 | 39 | -462.900 | 69.712 | -412.464 | -120.148 | 1 | -65.407 | -12.762 | 735 |
| best_4 | *.GA1.xray.rnd_1.c843.th.c | -82.652 | 13 | 39 | -456.933 | 72.624 | -416.046 | -113.511 | 26 | -69.614 | -11.864 | -1.172 |
| best_3 | *.xray.rnd_1.c18616.t.c | -86.390 | 1 | 45 | -429.865 | 72.291 | -392.203 | -109.953 | 44 | -65.991 | -5.728 | -14.666 |

Table 6.17: Docking results for alanized-GCR1 and GA1. Top 10 after combining all 10 dockings, for the 10 SCREAM GCR1 conformations with 2 GA1 conformers. * = Alanized GCR1

**Dealanization, binding site annealing, and minimization of selected GA1 Docked structures**

Table 6.18 shows the minimized top ten best structures. After pharmacophore analysis, based on best cavity and snapbe energies, and using Figures 10.95 - 10.97 and Figures 10.81 - 10.83, shown in Appendix XI., we selected the highlighted structures, which show good orientation of the ligand into the pharmacophore and good hydrogen bond, and salt bridges distances of GA1 inside the binding site.

| Docking structure | COMPLEXMINIMIZE | ucav | ucav_rank | Sum_Rank | total | snapbe | snapbe_rank |
|---|---|---|---|---|---|---|---|
| b10_x_845 | 9f.cm.bgf | -101,9210 | 3 | 9 | -594,4360 | -151,4978 | 6 |
| b10_x_845 | 8f.cm.bgf | -102,1290 | 2 | 10 | -604,8920 | -148,8565 | 8 |
| b10_x_845 | 4f.cm.bgf | -99,0650 | 10 | 11 | -588,1600 | -156,9573 | 1 |
| b10_x_845 | 7f.cm.bgf | -100,2140 | 8 | 17 | -605,1250 | -148,3489 | 9 |
| b10_x_845 | 5f.cm.bgf | -97,2550 | 16 | 18 | -599,7740 | -153,2524 | 2 |
| b8_x_32444 | 7f.cm.bgf | -99,0590 | 11 | 18 | -632,1780 | -151,4725 | 7 |
| b10_x_845 | 3f.cm.bgf | -97,6300 | 14 | 19 | -588,8120 | -151,5245 | 5 |
| b4_0_44628 | 2f.cm.bgf | -100,6240 | 6 | 25 | -648,7310 | -143,6290 | 19 |
| b4_0_44628 | 6f.cm.bgf | -102,7480 | 1 | 25 | -648,6000 | -141,2890 | 24 |
| b1_x_6914 | 10f.cm.bgf | -94,2440 | 24 | 27 | -664,9190 | -151,7133 | 3 |

Table 6.18: Top ten of minimized structures, based on cavity binding energy and snapbe energy. In yellow color, we show the selected structures which have a good orientation of the ligand into the pharmacophore and good hydrogen bond and salt bridge distances of GA1 inside the binding site.

After analyzing the energetic and pharmacophore results, we selected the three highlighted structures in Table 6.18 which account for both the best poses and have the best cavity energy and position of GA1 in the proposed binding site. We then carried out a molecular dynamics simulation on the selected structures, and the results are shown in the next section.

## 6.5 Conformational analysis of ligand docked GCR1-structures

To assess how docking with GA1 and ABA ligands might have affected the selected structure of GCR1 1SBiH, we inserted the annealed and minimized GA1- and ABA-docked-structures in a (1-palmitoyl-2-oleoyl-glycero-3-phosphocholine) (POPC) lipid bilayer, that reflects the cell membrane. Next, we added water, TIP3, above and below the bilayer, and finally, we

neutralized the whole system, adding Na and Cl ions to simulate the physiologic environment of the intra and extra-cellular regions of the cell.

The first step of the simulation consisted of 10000 steps of minimization, each 1 fs, on the lipids, water, and ions, keeping the protein fixed. This minimization aims to eliminate the steric clashes between lipids, water, ions, and amino acid residues. Next, we equilibrated by 500000 steps, of 1 fs each, at 310 K. Once we had the lipids, water, and ions relaxed, we carried out a new minimization, and again, we equilibrated at 310 K, keeping the protein backbone fixed, letting the protein side chains relax with lipids, waters, and ions. Finally, we equilibrated by 60 ns, keeping the helix fixed (1,4-Hydrogen bonds distances). Electrostatic, Van-der-Waals, and Total Energies values were analyzed to determine energy convergence and evaluate if the structure is equilibrated. Also, we analyzed the RMSD values for all processes as indicative of convergence. Finally, we downloaded the trajectory files and analyzed the whole structure seeking, out if the interactions found in previous processes remain, or maybe new Hydrogen Bonds, HBs, or Salt Bridges, SBs, might be found. Below we show the final structure of the MD simulations at NPT conditions (P = 1,04 atm and T = 310 K).

### 6.5.1 Selected GCR1-GA1 docked structures

Besides the interactions found during the time evolution for the equilibrated 1SBiH, selected as the best structure from SuperBihelix, where we found five stable salt bridges (D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), E117(4.39)-R107(3.52), and, E120(4.42)-R107(3.52)), we found three more stable salt bridges, which can be understood as product of a GA1 stabilizing interaction in the binding site that modify the helix interactions in the intracellular region of GCR1, for an example, see figure 6.22

#### GA1-b10x845-3fcm

Below, we show the analyzed parameters we used from MD simulations after equilibration, for selected docked structures, to select the 3fcm structure as the best candidate complex GA1-GCR1 complex to bind to GPA1. The RMSD for the equilibrated complexes, along with the SBs stability and energy figures, show that the complexes are fully converged, while the ligand interactions diagrams show the best pose for the ligand in the binding site, as shown below in Figures 6.21 – 6.24.

figure 6.21 show the final RMSD for the equilibrated structure.


Figure 6.21: Backbone RMSD for the last 20 ns out of 60 ns for GA1-b10x845-3c docked structure. Around last 2 ns are fully converged.

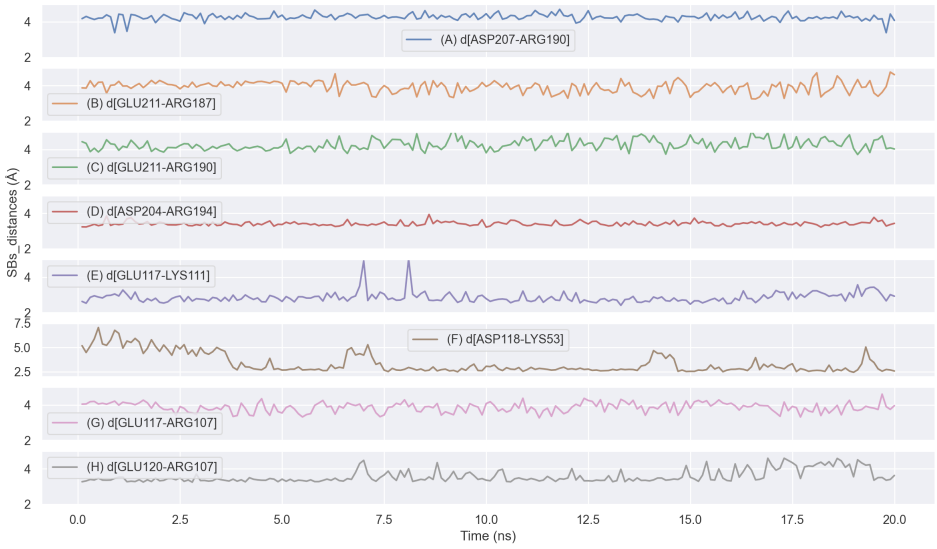


Figure 6.22: Time evolution for the equilibrated GA1-b10x845-3c salt bridges D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74), E117(4.39)-K111(ICL2), E117(4.39)-R107(3.52), D118(4.40)-K53(2.41), and E120(4.42)-R107(3.52). Where $d[i-j]$, (Å) is the distance of stable interhelical salt bridges between residues $i$ and $j$. We consider that stable SBs are less than 5 Å.

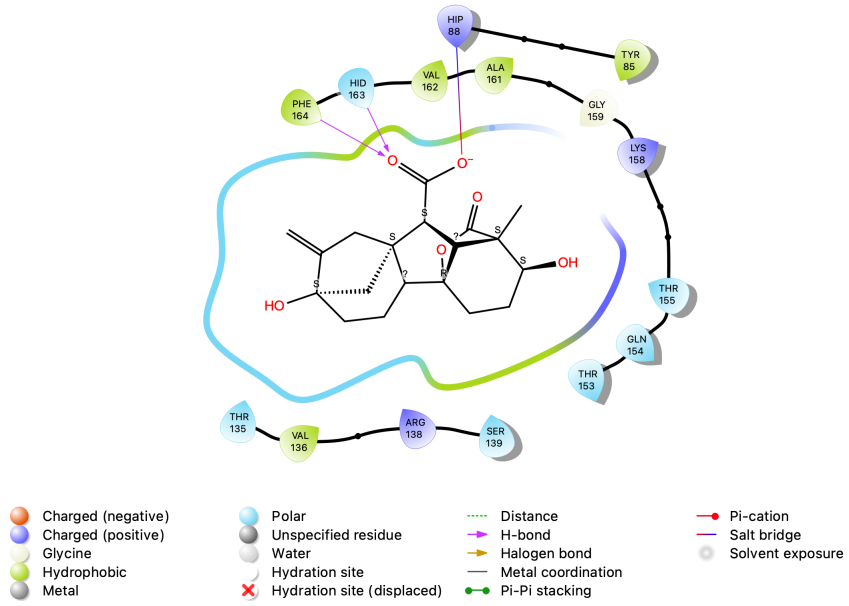


Figure 6.23: Shows the ligand interaction diagram, for the equilibrated and minimized structure, to the last structure obtained after 60 ns MD simulation, for the predicted interaction between GA1 and GCR1. The carboxylate group bears one salt bridge with HSP88, additionally two hydrogen bonds with a H-backbone of Phe164, and another with His163. A series of polar and hydrophobic interactions are shown with residues located at 4 Å away from GA1. For the sake of clarity, only polar hydrogen atoms are shown on the ligand.

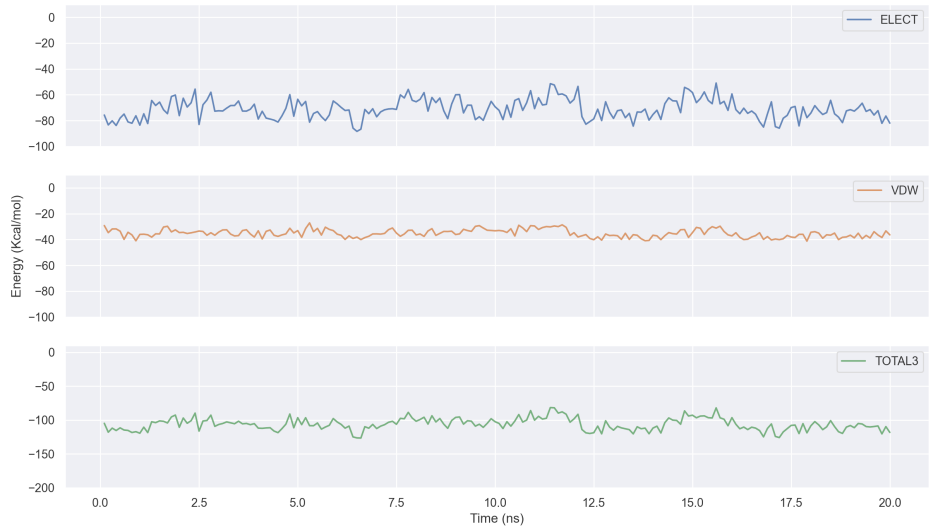


Figure 6.24: Shows the last 20 ns of a total of 60 ns molecular dynamics simulation, for the GA1-b10x845-3c structure. The electrostatic, Van der Waals, and total energies of the ligand-protein interaction reflect the stability of the complex during all 20 ns. The final electrostatic energy is around -80 Kcal/mol, while the VdW energy is close to -40 Kcal/mol. On the other hand, total energy, with a value close to -110 Kcal/mol, shows or reflects a higher contribution of the electrostatic interaction between the carboxylate group on GA1, and the residue HPS88 of GCR1.

In conclusion, we can see that structure b10x845-3fcm has the best total energy, with around -110 Kcal/mol, which is lower than those for the b10x845-5fcm and b1x6914-10fcm, shown in Appendix XII.. As we previously mentioned it is perhaps due to the electrostatic contribution of the SB interaction between the carboxylate group on GA1, and the residue HPS88 of GCR1 as can be seen in Figure 6.23 added to the number of HBs in each structure. On the other hand, structure b10x845-3fcm, has more stable interhelical salt bridges than the other equilibrated structures previously shown, which in turn can be associated with higher protein stability. Besides, since b10x845-3fcm structure has more SBs interactions than 1SBiH, the selected SuperBiHelix structure used to perform DarwinDock procedure, we could say that the ligand GA1 interaction inside the binding site promoted a conformational change in the intracellular region of those helices in GCR1.

### 6.5.2 Selected GCR1-ABA docked structures

Below, we show a comparative analysis of the ABA-GCR1 conformers obtained, from MD simulations which we used to select the best candidate ABA-GCR1 structure to bind to GPA1.

#### ABA-b9-5-20724-4c

In this case, the ABA-b9-5-20724-4c structure is the best candidate complex GA1-GCR1 complex to bind to GPA1. The RMSD for the equilibrated complexes, along with the SBs stability and energy figures, show that the complexes are fully converged, while the ligand interactions diagrams show the best pose for the ligand in the binding site, as shown bellow in Figures 6.25 – 6.29.

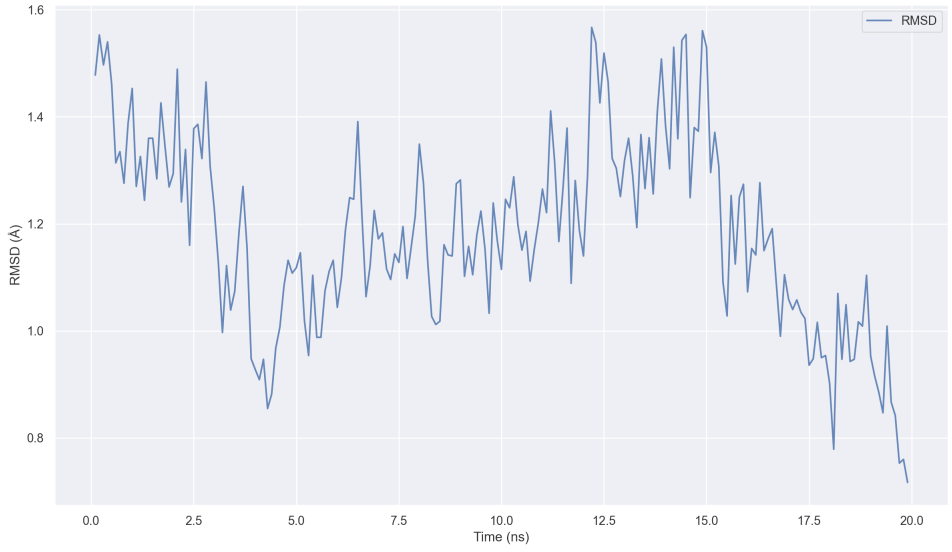


Figure 6.25: Backbone RMSD for the last 20 ns out of 60 ns for the equilibrated ABA-b9-5-20724-4c structure, which has the 1-4 HB distance fixed. Around the last 3 ns are fully converged.

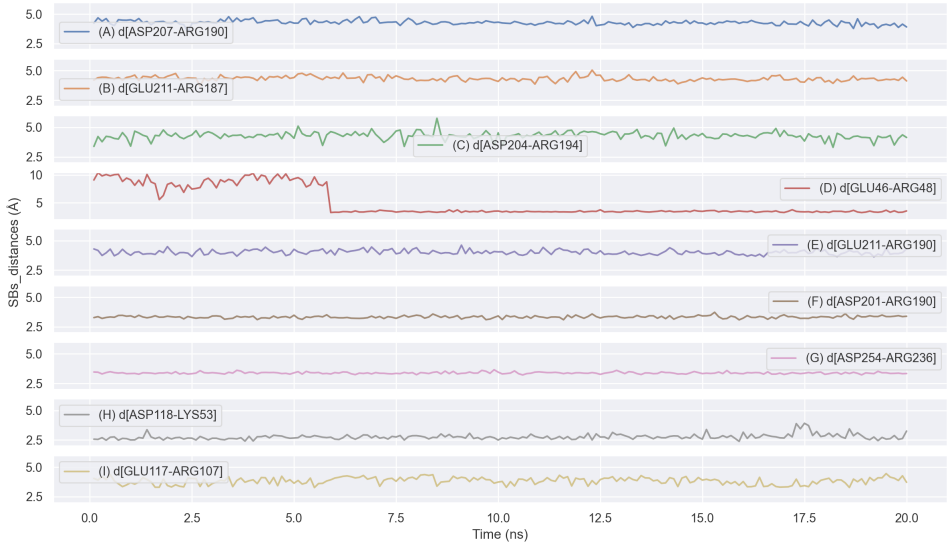


Figure 6.26: Time evolution for the equilibrated ABA-b9-5-20724-4c structure. Panels (A) to (I) show SBs found in the GCR1 region. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), D204(6.23)-R194(5.74), E46(ICL1)-R48(ICL1), E211(6.30)-R190(5.70), D201(ICL3)-R190(5.70), D254(7.38)-R236(ECL3), D118(4.40)-K53(2.41), and E117(4.39)-R107(3.52). We consider that stable SBs are less than 5 Å.

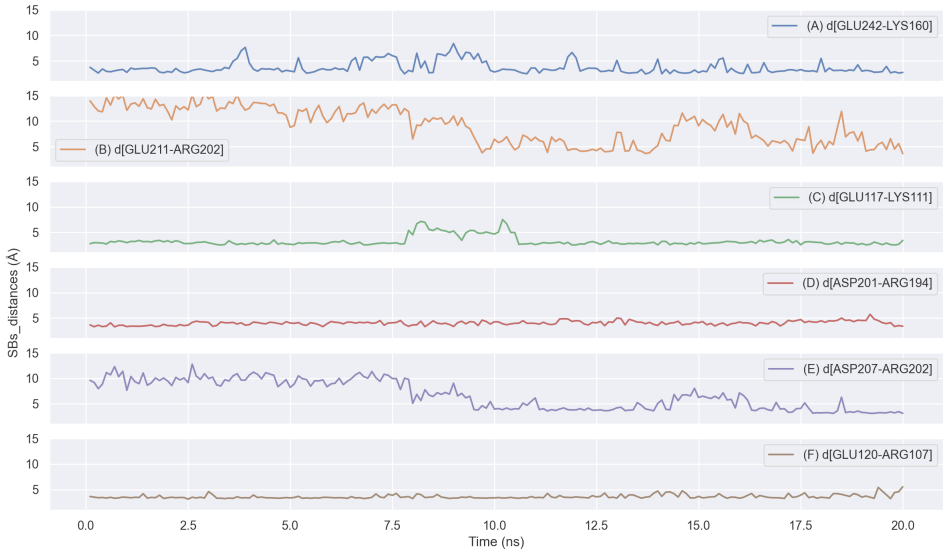


Figure 6.27: Time evolution for the equilibrated ABA-b9-5-20724-4c structure. Panels (A) to (F) show SBs found in the GCR1 region. Where $d[i-j]$, (Å) is the distance for transient interhelical salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: E242(ECL3)-K160(ECL2), E211(6.30)-R202(ICL3), E117(4.39)-K111(ICL2), D201(ICL3)-R194(5.74), D207(6.26)-R202(ICL3), and E120(4.42)-R107(3.52). We consider that the transient SBs break and form constantly during the time evolution of the dynamics but are around 4.5 Å.

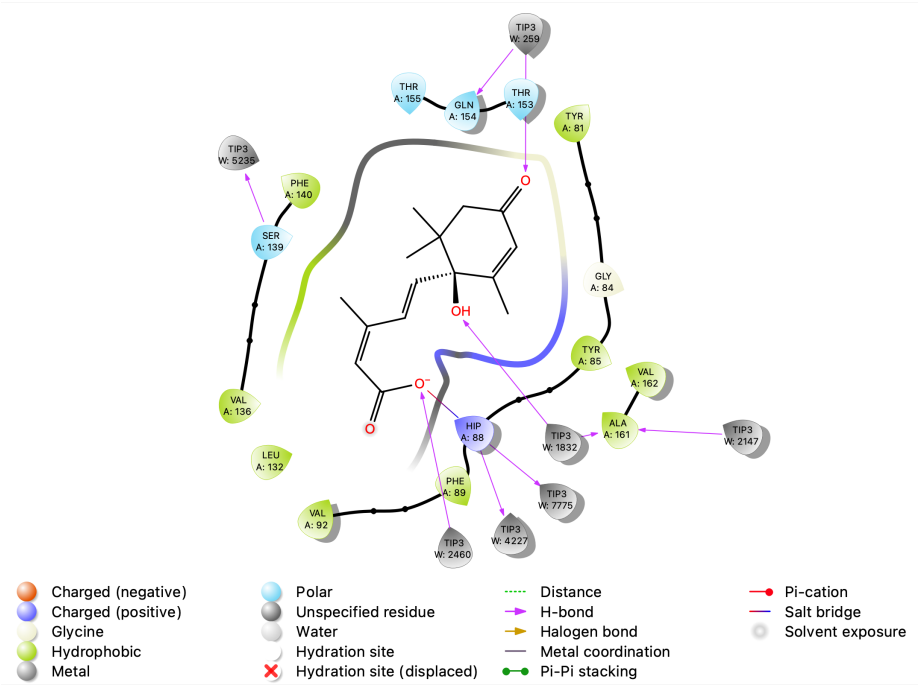


Figure 6.28: Ligand interaction diagram shows the last ABA-b9-5-20724-4c structure minimized and equilibrated in 60 ns of MD simulation. Here we can see a carboxylate group on ABA making part of a salt bridge with the HSP88-residue of GCR1 inside the predicted binding site. On the other hand, we can see two water-mediated hydrogen bonds, the first one is between the carbonyl group of ABA and GLN154, and the second one is between the hydroxyl group of ABA and ALA161. Likewise, a series of polar and hydrophobic interactions are shown with residues located at 4 Å away from ABA. For the sake of clarity, only polar hydrogen atoms are shown on the ligand.

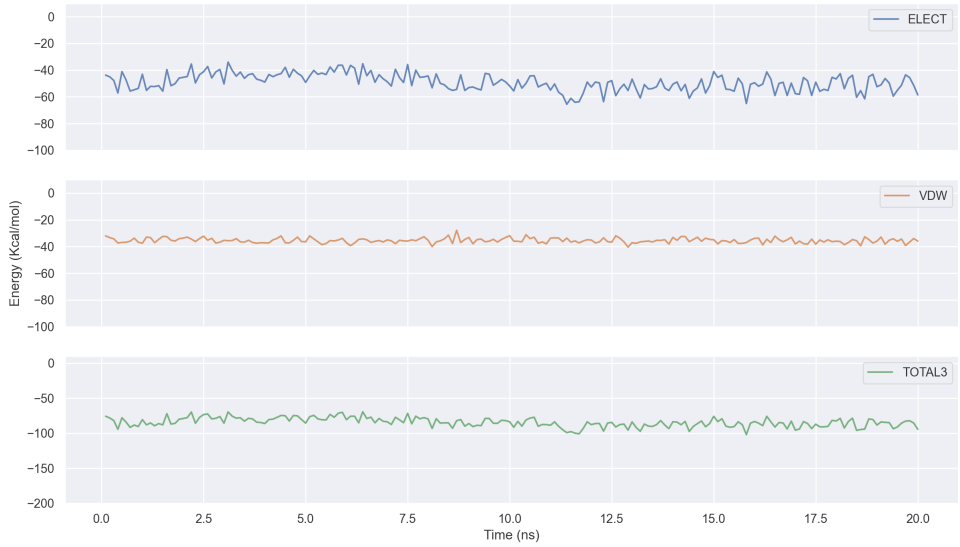


Figure 6.29: Ligand-protein energy. Electrostatic, Van der Waals, and total energies of the ligand-protein interaction reflect the stability of the complex during the last 20 ns of 60 ns molecular dynamics simulation for the ABA-b9-5-20724-4c structure.

Based on the energetic and structural analysis of the protein-ligand interaction performed in both, the binding site and inside the specific interaction between the carboxylate group in ABA with the HSP88 residue in GCR1, for the five equilibrated structures, four of them shown in Appendix XIII., we selected the ABA-b9-5-20724-4c structure as the best to bind to the G protein. Along with the above analysis, we looked over how those interactions could be related to a series of helix-helix, helix-loop, and loop-loop SBs found in all five structures, and we used this information as a predictor of which of the ligand poses inside the binding site, could generate the most stable conformation of GCR1.

From the results, we can see that structure named ABA-b9-5-20724-4c has the best electrostatic and Vdw energies for the specific COO-HSP interaction Figure 6.29, and also generates the highest amount of SBs in interhelical GCR1 region. On the other hand, the ABA-b9-5-20724-4c structure has two water-mediated HBs and one SB in the binding site region, as can be seen in Figures 6.26, 6.27, and 6.28, respectively.

### Comparative analysis of selected GCR1-ABA and GCR1-GA1 docked structures

Figure 6.30 shows a differential view of the interactions of both ligands inside the binding site. It is interesting to note that although both ligands have similar interactions inside the binding site, the carboxylate group on GA1 could also interact by a salt bridge with the HSD-162 residue, which we are modeling at the moment, data not included in this paper. We can also see that GA1 makes more polar interactions with residues in the binding site. These two different interacting ways of each ligand inside the binding site generate two different ways of interaction between GCR1 and GPA1 in the intracellular region. We are also modeling this at the moment, data not included in this paper.

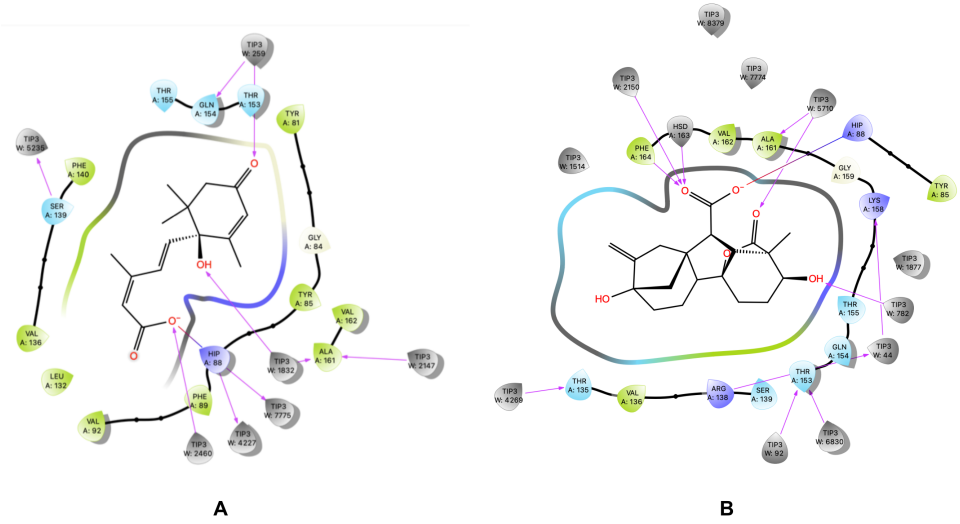


Figure 6.30: Ligand interaction diagram shows the last ABA-b9-5-20724-4c and GA1-b10-x-845-3C structures minimized and equilibrated in 60 ns of MD simulation. Here we can see a carboxylate group of ABA and GA1 making part of a salt bridge with the HSP88-residue of GCR1 inside the predicted binding site. On the other hand, we can see two different ways of water-mediated hydrogen bonds of each ligand into the binding site. Panel A shows one water-mediated HB between the carbonyl group of ABA and GLN-154, and a second one between the hydroxyl group of ABA and ALA161. On the other hand, GA1 makes one water-mediated HB between ALA 161 and the carbonyl group of the lactone but additionally makes two normal HBs, one with HSD-163 and another with PHE-164. Likewise, a series of polar and hydrophobic interactions are shown with residues located at 4 Å away from both ligands, GA1 and ABA. For clarity, only polar hydrogen atoms are shown on the ligand.

Figure 6.31 shows changes generated on the 1SBiH structures by binding GA1 and ABA to the proposed binding site. Comparing panels (A) and (B) allows to observe that helices in the cyan structure, corresponding to the GA1-GCR1 docked system, are offset to the right with respect to the undocked structure, while helices in the green structure, corresponding to the ABA-GCR1 docked structure, are offset to the left with respect to the undocked structure. This suggests that each ligand generates a different conformational change on the GCR1 structure, which can see better when

panels (C) and (D) of the same figure are observed in detailed. In these panels we can see a large changes in the orientations of some helices, for instance, H7 in the extracellular region of GCR1 as shown in panel (D), but also the intracellular regions of H1 and H2, as shown in panel (C), and H3, H4, H5, and H6 shown in panel (D). A large structural change can also be seen in each structure's intracellular loops.

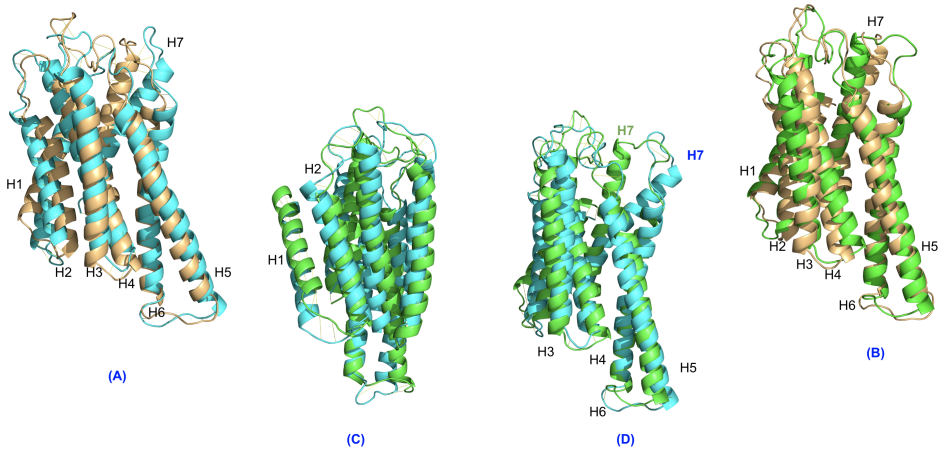


Figure 6.31: Changes generates on the 1SBiH structure by binding GA1 and ABA to the proposed binding site. Panel (A) shows a structural alignment between 1SBiH in orange and GA1-1SBiH docked structure in cyan. Panel (B) shows a structural alignment between 1SBiH in orange and ABA-1SBiH docked structure in green. Panels (C) and (D), show two different lateral views for the structural alignment between GA1-1SBiH docked structure in cyan and ABA-1SBiH docked structure in green.

## 6.6 Functional analysis of the phytohormones in the GCR1-GPA1 Signaling pathway

We docked GPA1 into the intracellular region of GCR1 in both complexes and carried out MD simulations. We found important differences in the interactions between the Ras and helical domains of GPA1, in each complex. For example, for the ABA complex, we found a series of SBs between the Ras and helical domains, mainly the one found between K288-D162. This SB was not found in the work of Jones and collaborators about the self-activation

of G-protein in plants [12]. In this work is suggested that the absence of this SB can increase solvent accessibility between Ras and helical domains, promoting the GDP/GTP exchange in GPA1. Our simulations show that this SB is formed in the ABA complex but is broken in the GA1 complex. We also found larger distances between the centers of mass of the Ras and Helical domains in the GA1 case than in the ABA case. These findings, shown in Appendix XIV. and Appendix XV, are supported by comparing the SBs distances found in both complexes. Finally, a pharmacophore view of the GDP between Ras and helical domains shows more interactions for the ABA case than for the GA1 case, which can be related with a pre-activated GPA1 mediated by the GA1-GCR1 interaction.

By comparing the equilibrium structures of the GA1-GCR1-GPA1-GDP-Mg and ABA-GCR1-GPA1-GDP-Mg, we found several differences. These differences suggest that the hormones ABA and GA1 promote antagonistic conformational changes. Figures 6.32 and 6.33 display several differences between salt bridges distances of both complexes, for instance, salt bridge D162-K288 in panel (B) represents a broken SB for the GA1-GCR1-GPA1 complex. In contrast, in panel (A), the D162-K288 distance shows a strong SB for the ABA-GCR1-GPA1 complex. On the other hand, although ARG190 is neither part of the Ras nor the Helical domains of GPA1, it is interesting to note that the SBs E254-R190 and D162-K288 are both formed in the ABA complex, meanwhile, they are both broken in the GA1 complex. This can be seen in panels (C) of Figures 6.32 (A) and (B). Panel (C) of the GA1 complex of figure 6.32(B) shows the distance of the SB E48-R190, instead of E254-R190, since there is no interaction E254-R190. These observations are corroborated by observing the SBs of panels (A) and (B) of Figure 6.33. These observations allow us to conjecture that the salt bridges E254-R190 and D162-K288 work together in a synergistic way to close the distance between Ras and Helical domains in the ABA complex in order to avoid the GDP/GTP exchange. Panel (A) of the Figure 6.34 shows that the Ras-Helical distance of the ABA complex remain small meanwhile panel (B) of the same Figure shows that he same distance grows larger for the GA1 complex, facilitating the GDP/GTP exchange and the subsequent GPA1 activation. The case of the ABA complex, where the Ras-Helical distance remains small, seems to be related to the synergistic interactions of the SBs E254-R190 and D162-K288.

The promoted differences, by binding of ABA and GA1 to GCR1, on the GPA1 conformational changes, inside of each complex, can be related as GCR1 interacting with GPA1 in the intracellular region, Figure 6.35. GCR1 has no salt bridges between helices 3 and 6, which is the place where the $\alpha$-

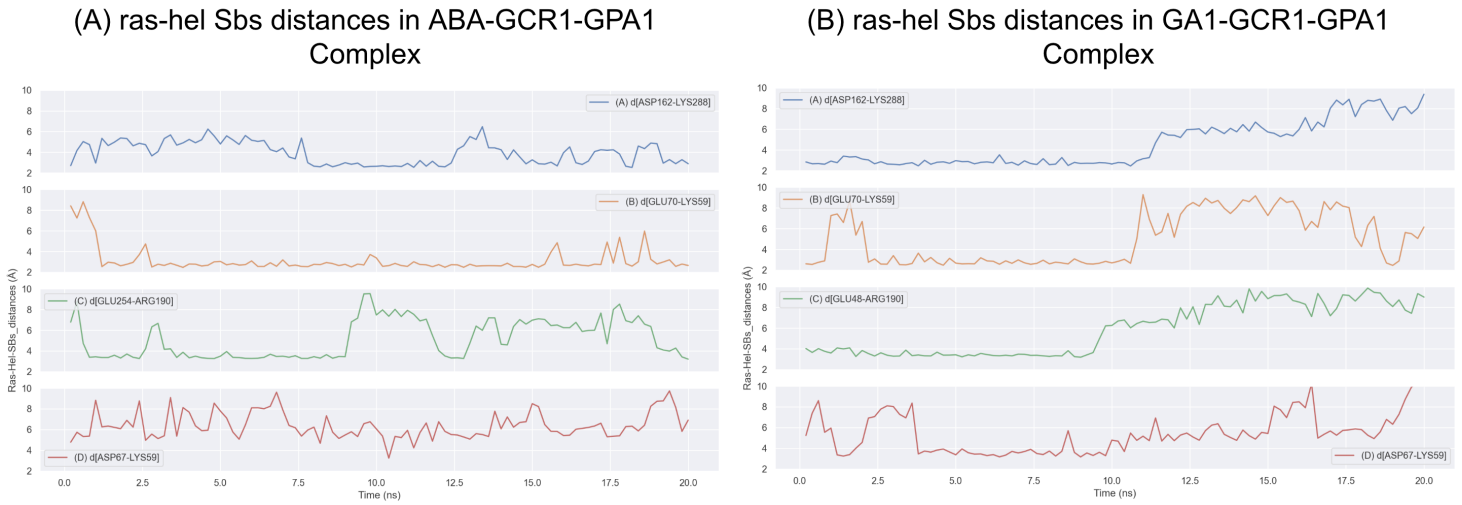


Figure 6.32: Panels (A) and (B) show the behavior of four electrostatic interactions between Ras and Helical domains of GPA1. Panel (A) shows three SBs interactions that stabilize at the end of 20 ns in the ABA-GCR1-GPA1 complex. Boxes (A), (B), and (C) in Panel (A), show distances for D162-K288, E70-K59, and E254-R190 interactions, respectively. In contrast, we can see in Panel (B), which describes the complex formed between GA1-GCR1-GPA1, that the D162-K288, E70-K59 interactions that show distances for stable SBs at the beginning of 20 ns, at the end of the trajectory are broken. In addition, R190 interacts with E46 but not with E254.

(A) ras-hel Sbs distances in ABA-GCR1-GPA1 Complex

(B) ras-hel Sbs distances in GA1-GCR1-GPA1 Complex

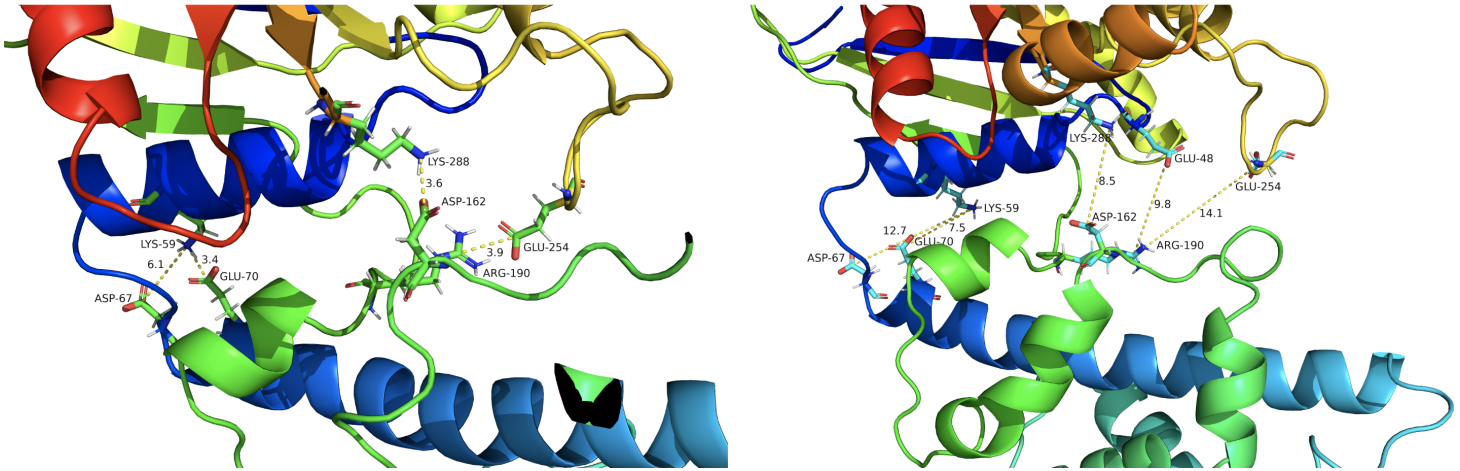


Figure 6.33: 3D view of the Ras-Helical interactions in the ABA-GCR1-GPA1 and GA1-GCR1-GPA1 complexes in panels (A) and (B), respectively. The location of the D162-K288 and E254-R190 interactions shows a possible effect of both interactions by keeping the Ras-Helical domains interacting and closing GPA1.

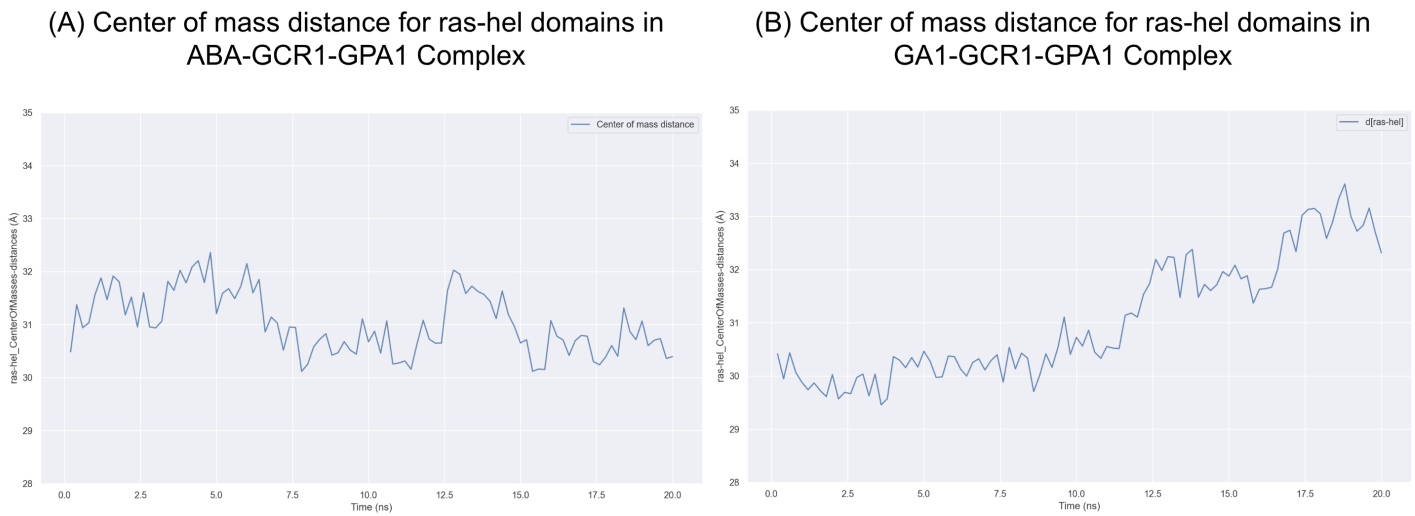


Figure 6.34: Ras-Helical center of mass distances in the ABA-GCR1-GPA1 and GA1-GCR1-GPA1 complexes in panels (A) and (B), respectively.

5-helix of G-$\alpha$ subunit binds to animal class-A GPCRs. Due to the absence of salt bridges between H3 and H6 in GCR1, we have decided to bind the residue L383, the last residue in GPA1 belonging to the $\alpha$-5-helix of GPA1, to the residue K49 of GCR1's helix 2; We have found a salt bridge between K49 (H2) and D115 (H4) on GCR1 which has the closer similitude to class-A animal GPCRs H3-H6 salt bridge. The interactions between L383, K49 and D115 can be seen as possible place to the interaction of GPA1 with GCR1, specifying the general ideas mentioned by Taddese and coworkers [11]. We found a big changes in the interhelical stable and transient netweorks of salt bridges in GCR1 for both complexes, as can be seen in Figures 6.36 and 6.37.

On the other hand, when we contrasted the binding site on GCR1 with ABA and GA1, we observed a bigger amount of polar interactions between GA1 and residues in the binding site than those found in the ABA case. Specifically, we found four HBs and an SB between GA1 and HSD-163(5.43), PHE-164(5.44), and THR-155(ECL2) and HSP88(3.33) respectively, while we found just one HB and one SB between ABA and GLN-154(ECL2) and HSP88(3.33) respectively as can be seen in figure 6.38. These findings can be interpreted not only as a stronger interaction between GA1 and GCR1 than between ABA and GCR1 but also as that GA1 is generating a different conformational change in GCR1 intracellular region, compared with that produced by the ABA-GCR1 interaction, which could explain an antagonistic behavior of those hormones after binding to GCR1. We supported the proposal for antagonistic roles of ABA and GA1 after analyzing Figures 6.32 and 6.33, where we can see that while, ABA interaction with GCR1 promotes the close up of the GPA1 through the formation of several SBs between Ras and Helical domains, which also close the distance Ras and Helical domains compared with the open up effect promoted by GA1, as can be seen in Figure 6.34. From a physiological view, this can be interpreted as GA1-GCR1 interaction promoting the GPA1 activation by allowing the open up of GPA1 and the subsequent interchange of GDP by GTP. In contrast, the ABA-GCR1 interaction avoids the GDP/GTP interchange inside the GPA1 by keeping both domains Ras-Hel closed up, after the inactivation of GPA1 promoted through the Arabidopsis regulator of G protein signaling-1 (At-RGS1) protein [30], as is also mentioned by Urano and collaborators [45, 44] to propose that plants do not require a GPCR with GEF activity to activate to the G-protein. The GPA1 activating GCR1 mediated proposed by us agrees with our hypothesis, where we stated that ABA-GCR1 interaction can generate the GPA1 sequestration to prevent the plausible GPA1 self-activation [30], while the GA1-GCR1 interaction can allow the GPA1

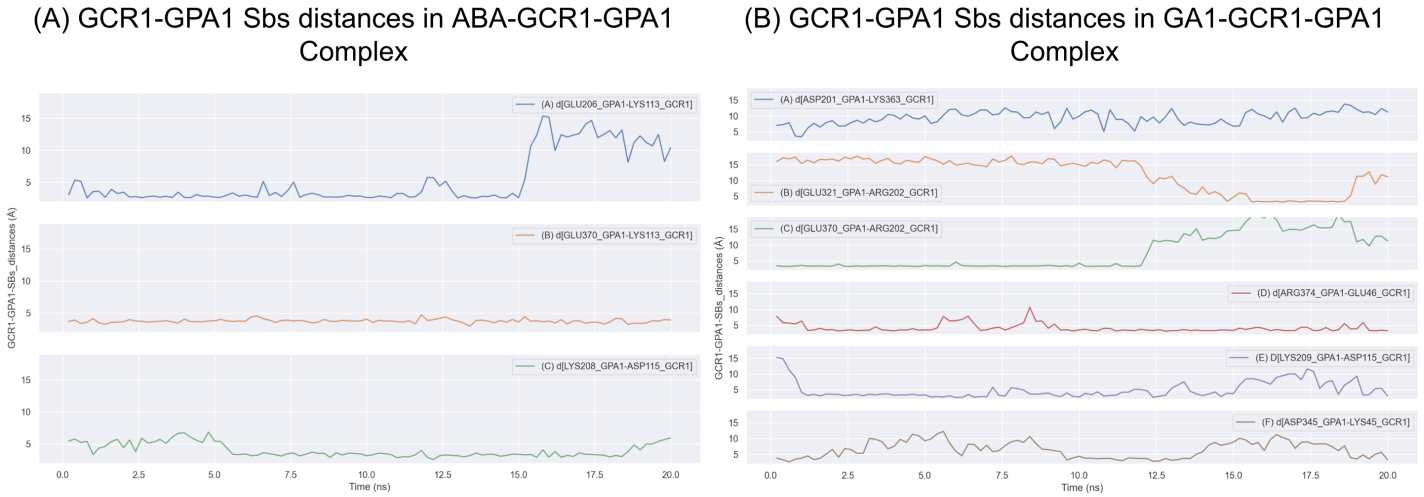


Figure 6.35: Panels (A) and (B) show two different ways of interactions between GCR1 and GPA1 in ABA-GCR1-GPA1 and GA1-GCR1-GPA1 complexes, respectively.

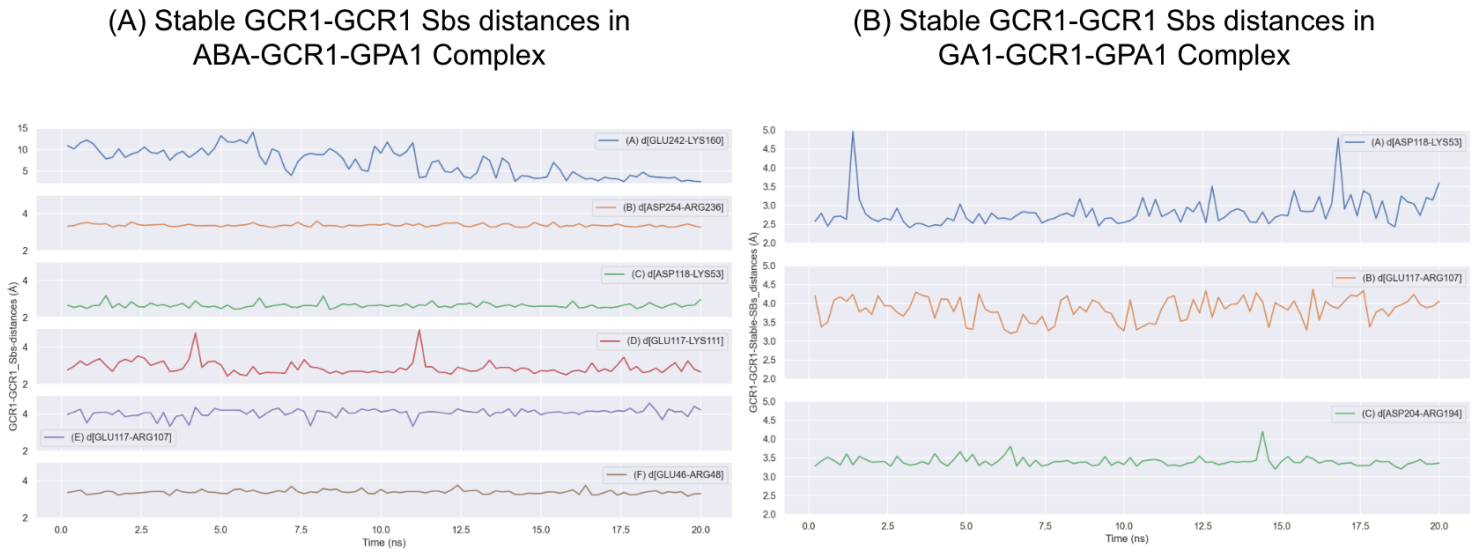


Figure 6.36: Panels (A) and (B) show two different interhelical GCR1-GCR1 stable SB networks in ABA-GCR1-GPA1 and GA1-GCR1-GPA1 complexes, respectively

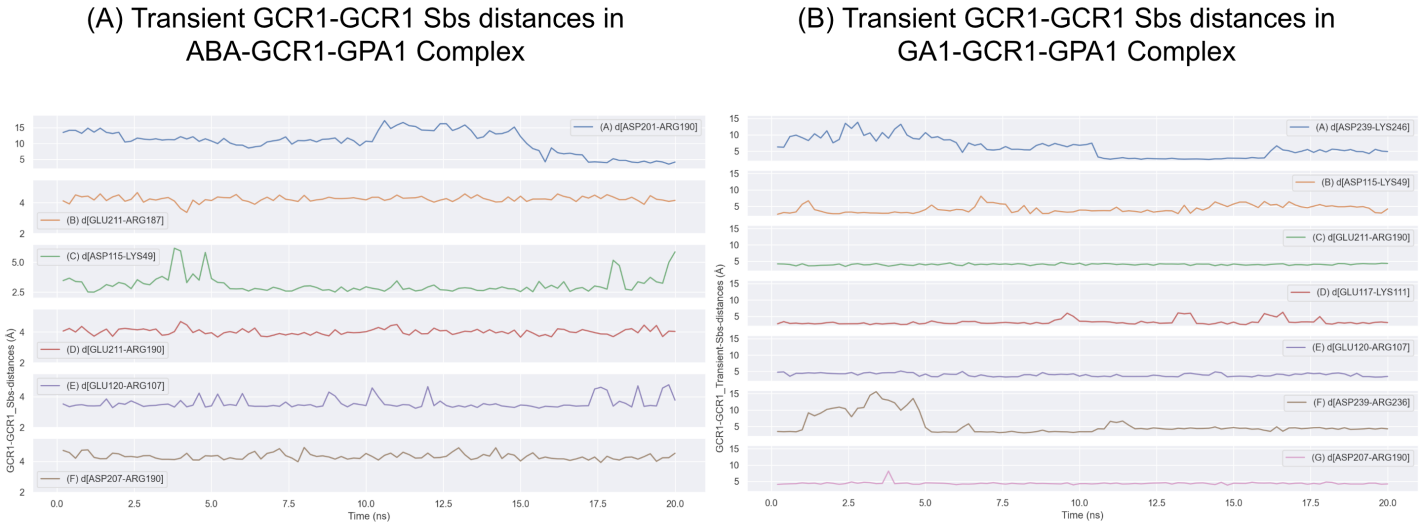


Figure 6.37: Panels (A) and (B) show two different interhelical GCR1-GCR1 transient SB networks in ABA-GCR1-GPA1 and GA1-GCR1-GPA1 complexes, respectively

activation mediated by the open-up of the GPA1 after braking of important inter Ras-Hel salt bridges[85], as is seen in Figures 6.33 and 6.34.

In addition, Figure 6.39 shows in panel (A) that there are more interactions between GDP and residues belonging to Ras and Helical domains in the ABA system than those found in the GA1 system. Although we can see some interaction between the Helical domain and GDP, which should not happen in an activated GPA1, this could show that GPA1 is in a pre-activated state, not in a fully activated state.

Finally, the energetic analysis shows interesting behaviors; for instance, in figure 6.40 we can see stronger binding electrostatic and total energy interactions between GA1 and GCR1-GPA1 than for ABA and GCR1-GCR1, but the vdW energy is similar in both cases. These findings are in agreement with the energy between Ras-Helical energy landscape shown in figure 6.41, where for the ABA case we see that the electrostatic and total components have better interactions between both domains, while we see worst Ras-Hel interaction, which can be suggesting that better ligand-GCR1 promotes opening up of GPA1, and poor ligand-GCR1 promotes better interactions between Ras and helical domains allowing both domains interact through several SBs, as shown in Figure 6.33.

This comprehensive presentation of results and analysis from my research enables me to propose a series of mutation experiments aimed at validating the predicted apo-GCR1, GA1-GCR1, and ABA-GCR1 complex structures, as well as the GPA1 regulation pathways mediated by GCR1. Additionally, I wanted in this study to include a comparison with other existing models, such as Taddese's and AlphaFold models, which will be discussed in the following two sections.

## 6.7 Proposed mutations to validate the predicted GCR1 Structure and the predicted ligand-GCR1 binding sites.

There are no cryoEM or x-ray structures for apo-GCR1 or for the activated structure bound to ABA or GA1. In order to provide a means to validate our predicted structure, we have identified (Table 6.19) mutations expected to decrease or increase stability in the apo-GCR1, ABA-GCR1, and GA1-GCR1 structures. Column 1 enumerates the interactions that are expected to stabilize the structures. Thus, we propose mutations to charged positive and negative residues respectively, that might increase the stability. On the other hand, mutating these polar residues to nonpolar ones should decrease

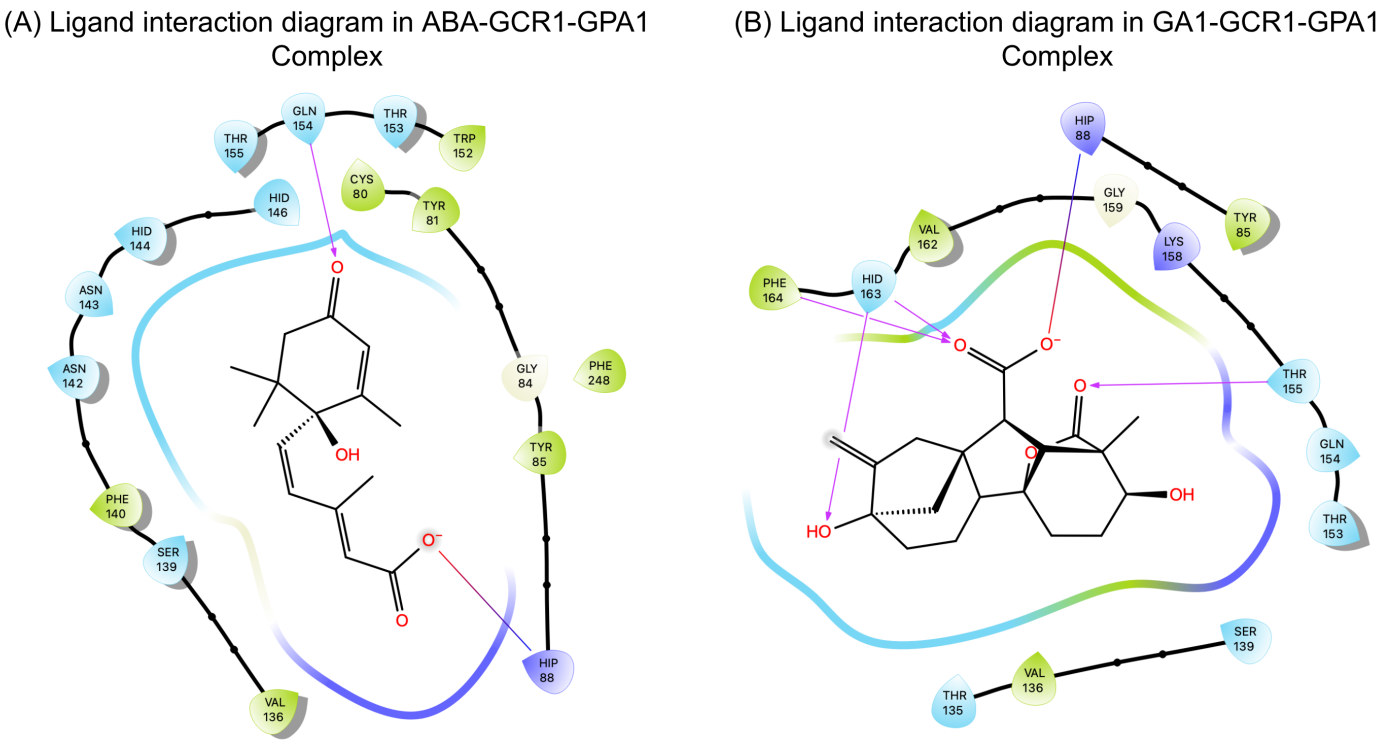


Figure 6.38: Panels (A) and (B) show interactions of ABA and GA1 inside the Binding site of GCR1 for both complexes, ABA-GCR1-GPA1 and GA1-GCR1-GPA1

(A) GDP-Ligand interaction diagram in ABA-GCR1-GPA1 Complex

(B) GDP-Ligand interaction diagram in GA1-GCR1-GPA1 Complex

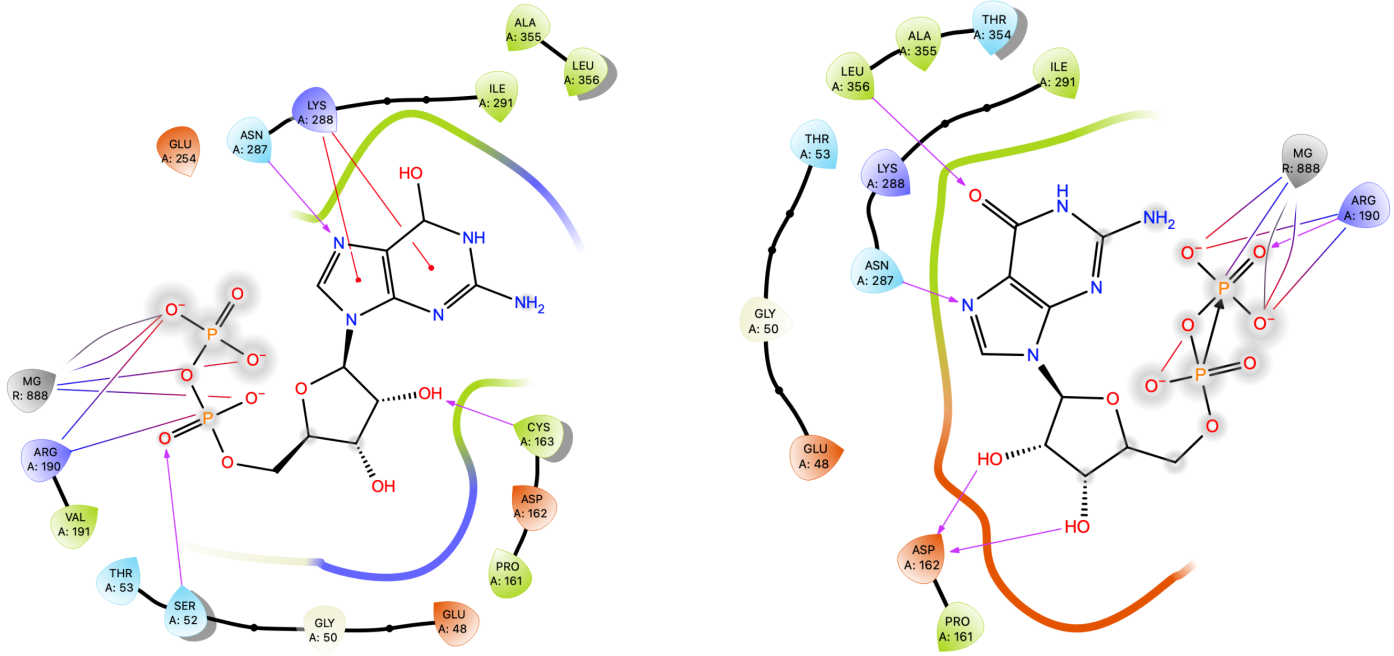


Figure 6.39: Panels (A) and (B) show interactions of GDP between Ras and Helical domains of GPA1 for both complexes, ABA-GCR1-GPA1 and GA1-GCR1-GPA1

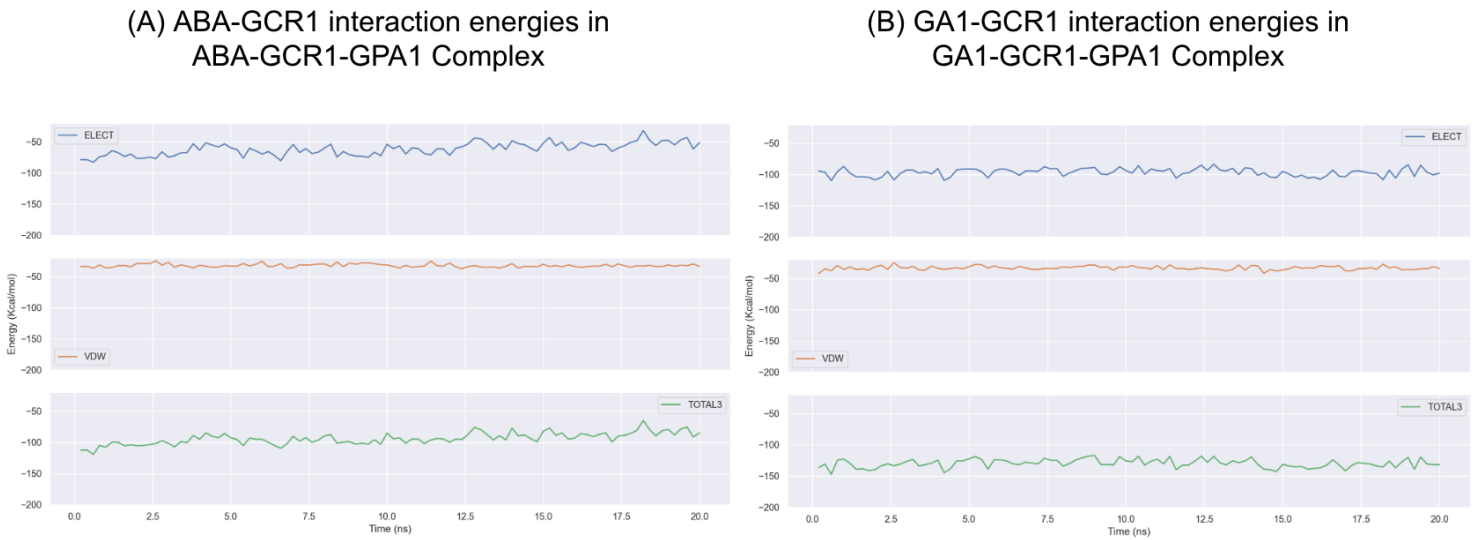


Figure 6.40: Comparative electrostatic, Van-der-Waals and total energies for the interaction between ABA-GCR1 and GA1-GCR1 in both complexes, ABA-GCR1-GPA1 and GA1-GCR1-GPA1.

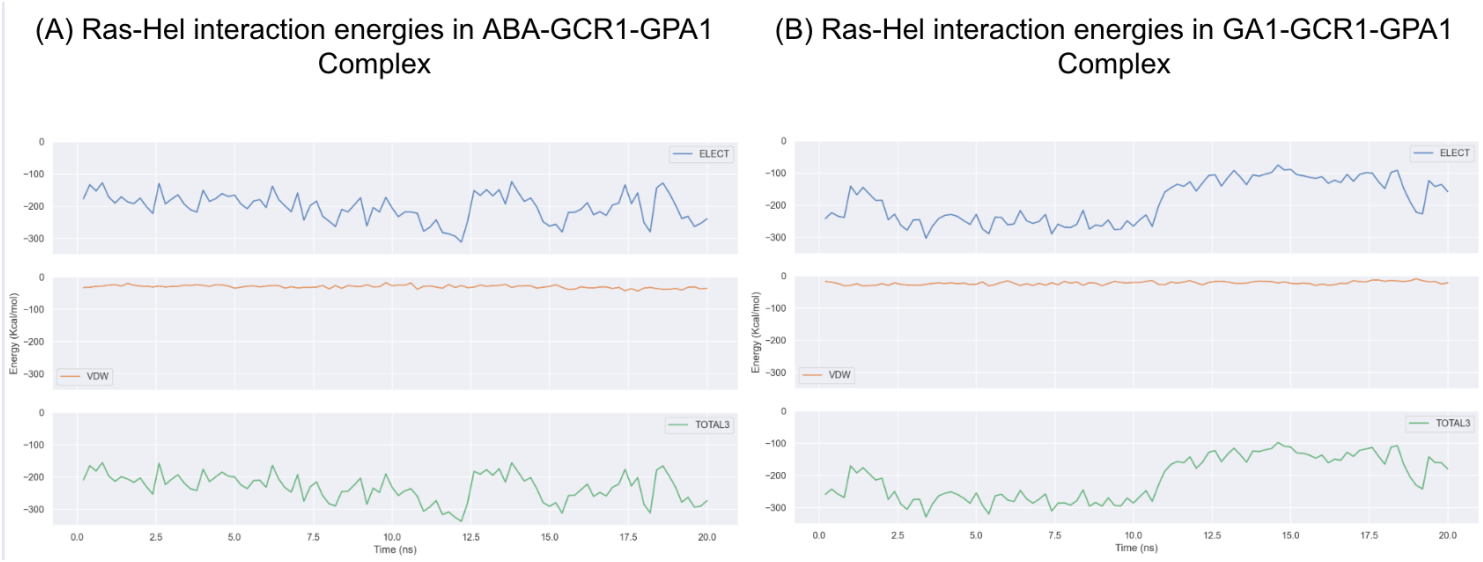


Figure 6.41: Comparative electrostatic, Van-der-Waals and total energies for the interaction between Ras-Helical domains of GPA1 in both complexes, ABA-GCR1-GPA1 and GA1-GCR1-GPA1.

the stability. We have identified the SBs in column 6 as stabilizing interactions, especially K53(2.41)-D118(4.40) and K49(ICL1)-D115(ICL2), which could be mutated to nonpolar residues such as GLY or ALA to produce a strong destabilization of the helix packaging in the GCR1, likely causing a loss of function of the protein. Table 6.12 showed a series of interhelical SBs between helices TM2-TM4, TM3-TM4, and TM5-TM6 that along with the two last HBs in column 1 of Table 6.19 generate a stabilized packing of those helices in the intracellular region of GCR1. Consequently, we propose that mutating these charged residues to non-polar residues such as GLY and PHE should strongly destabilize the GCR1 packing.

| H-bonds | BWN | GCR1 *d*, (Å) | GCR1-ABA *d*, (Å) | GCR1-GA1 *d*, (Å) | S-bridges | BWN |
|---|---|---|---|---|---|---|
| S34–D62 | 1.50–2.50 | 3,2 | 2,66 | 2,6 | K53–D118 | 2.41–4.40 |
| D62–N265 | 2.50–7.49 | 4 | 6,44 | | | |
| S34–S66 | 1.50–2.54 | | 2,86 | | K49–D115 | 2.37(ICL1)–4.37(ICL2) |
| Y41–S266 | 1.57–7.50 | 6,4 | 3,98 | | H163–D239 | 5.43–ICL3 |
| Y57–T100 | 2.45–3.45 | | 6,2 | | E46–K213 | ICL1–6.32 |
| S61–S94 | 2.49–3.39 | 4,17 | | | | |
| S61–Y269 | 2.49–7.53 | 2,94 | 5,66 | 4,3 | | |
| S61–N265 | 2.49–7.49 | | 4,7 | | | |
| Y81–N143 | 3.26–ECL2 | | 3.3 | | | |
| Y85–S139 | 3.30–4.61 | 5 | | | | |
| Y85–H144 | 3.30–ECL2 | | | 2,88 | | |
| T98–Y269 | 3.43–7.53 | 3,5 | 3,71 | 3,62 | | |
| T99–Y168 | 3.44–5.48 | | 2,81 | 2,85 | | |
| T99–Y221 | 3.44–6.40 | 4,17 | | | | |
| T104–D118 | 3.49–4.40 | | 4,5 | | | |
| T135–Y168 | 4.57–5.48 | 2,99 | | | | |
| T166–R236 | 5.46–ICL3 | 3,6 | | | | |
| T166–D254 | 5.46–7.38 | | 2,7 | | | |
| D254–R236 | 7.38–ICL3 | | 4 | | | |
| R187–N208 | 5.67–6.27 | | | | | |
| N192–N208 | 5.72–6.27 | | | | | |

Table 6.19: Predicted mutations on HB and SB interactions to validate or not the predicted GCR1 proposed structure. BWN=BW numbering, *d*=distance

In the same way, Table 6.20 enumerates the most important interaction found of ligands ABA and GA1 inside the binding site along with predicted deleterious and stabilizing mutations to validate or not the predicted ligand-GCR1 binding sites.

| Interaction between ABA–GCR1 | Interaction between GA1–GCR1 | Stabilizing mutation | Destabilizing mutation |
|---|---|---|---|
| SB HSP88(3.33)–ABA(COO) | SB HSP88(3.33)–GA1(COO) | PHE164→HSP | HSP88→PHE |
| HB HSD163(5.43)–ABA(COO) | HB HSD163(5.43)–GA1(COO) | HSD163→HSP | HSD163→PHE |
| HB PHE164(5.44)–ABA(COO) | HB PHE164(5.44)–GA1(COO) | PHE164→ARG | PHE164→PRO |
| HB GLN154(ECL2)–ABA(C=O) | HB GLN154(ECL2)–GA1(C=O) | GLN154→LYS | GLN154→ALA |
| HB SER139(4.61)–ABA(C=O) | HB SER139(4.61)–GA1(C=O) | SER139→ARG | SER139→GLY |
| HB TYR81(3.27)–ABA(OH) | HB TYR81(3.27)–GA1(OH) | TYR81→LYS | TYR81→ALA |

Table 6.20: predicted mutations on HB and SB interactions to validate or not the predicted ligand-GCR1 binding sites.

## 6.8 Free energy landscape from Meta-Dynamics simulations

We are employing metadynamics to produce the potential of mean force landscape (PMF) of the complexes ABA-GCR1-GPA1-GDP-$Mg^{2+}$, GA1-GCR1-GPA1-GDP-$Mg^{2+}$ and GPA1-GDP-$Mg^{2+}$. In these free energy landscapes, the abscissa is the collective coordinate made by measuring the distance between the centers of mass of the Ras and Helical domains, the ordinate is the free energy in kcal/mol. Partial results of these metadynamics calculations are displayed in Figure 6.42. The convergence of these calculations is verified by repeating the simulations with different initial bias potentials. The calculation of the smallest system, GA1-GCR1-GPA1-GDP-$Mg^{2+}$, has been repeated three times, meanwhile, the other two systems have been repeated twice.

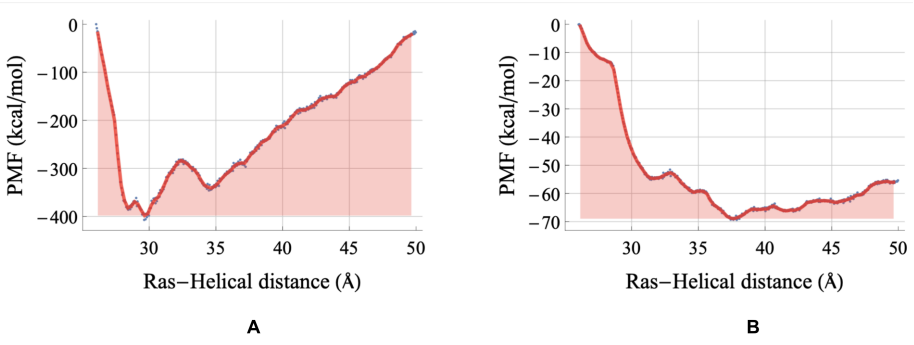


Figure 6.42: Panels (A) and (B) show the free energy as a function of Ras-Helical distance as the collective variable in ABA-GCR1-GPA1 and GA1-GCR1-GPA1 complexes

Figure 6.42 shows that all complexes display two free energy minima with a potential barrier between them. For both systems, the minimum at the smallest (largest) Ras-Helical distance corresponds to the inactive or closed (active or open) conformation of GPA1. For the three complexes, the inactive energy minima are located around 30 Å, and the active energy minima are located around 35 Å. The potential energy barriers from inactive to active are: ≈100 kcal/mol for the ABA complex, ≈60 kcal/mol for the GPA1 complex, and ≈10 kcal/mol for the GA1 complex. The lower the potential energy barrier, the more accessible the active conformation of GPA1, from its inactive conformation. Although the results are partial, they show agreement with our hypothesis that GA1 promotes the activation of GPA1 while ABA promotes the inactivation of GPA1.

Both GA1 and ABA engage in strong and stabilizing interactions within the GCR1 binding site, as demonstrated in Figure 6.43. Panel (A) displays the energetic landscape of ABA interactions with neighboring amino acids in the GCR1 binding site, while Panel (B) displays the energetic landscape for GA1 interactions with neighboring amino acids in the same site. These observations underscore the robust nature of the interactions between both hormones and the GCR1 receptor. Furthermore, each hormone induces distinct and characteristic conformational changes in the GPA1 structure within the ABA-GCR1-GPA1-GDP-$Mg^{2+}$ and GA1-GCR1-GPA1-GDP-$Mg^{2+}$ complexes. These conformational changes can be associated with the active or inactive states of GPA1. The ras-helical distance in GPA1 bound to each complex serves as an indicator of its activation state. Comparing the activated GTP-GPA1 structure reported by Jones et al. [46] with GPA1 bound to GA1-GCR1 and ABA-GCR1 complexes, reveals noticeable differences. For instance, in Panel (A) of Figure 6.44, the opening up of GPA1 corresponds to its activation state, enabling the GDP/GTP exchange. Conversely, in Panel (C), a closed conformation of GPA1 is observed, mediated by the ABA-GCR1 interaction, effectively enclosing the GDP-$Mg^{2+}$ and preventing the GDP/GTP exchange and subsequent inactivation. This observation is further supported by the presence of the D162-K288 salt bridge interaction, known to be associated with an inactive conformation of GPA1.

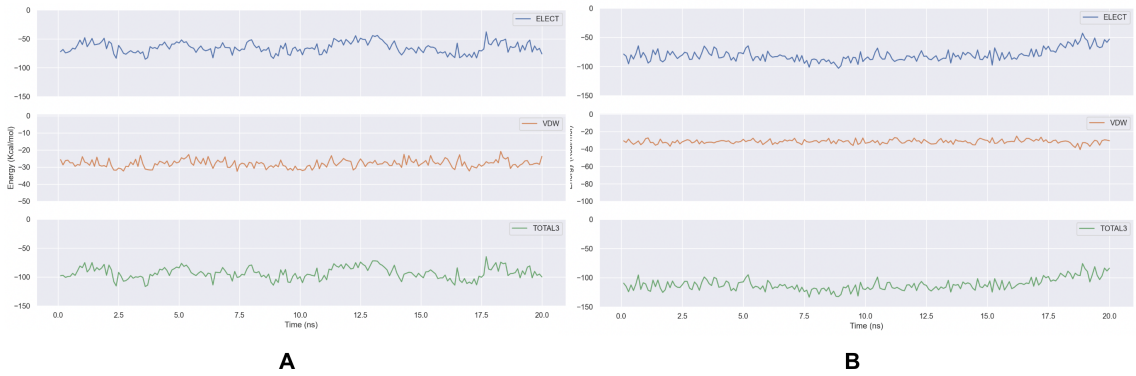


Figure 6.43: Energetic comparison of the interaction of ABA with amino acids of the binding site in panel (A) and GA1 with amino acids of the binding site in panel (B). Panels (A and B) depict the electrostatic, Van-der-Waals, and total energies for the last 20 ns of the MD simulation

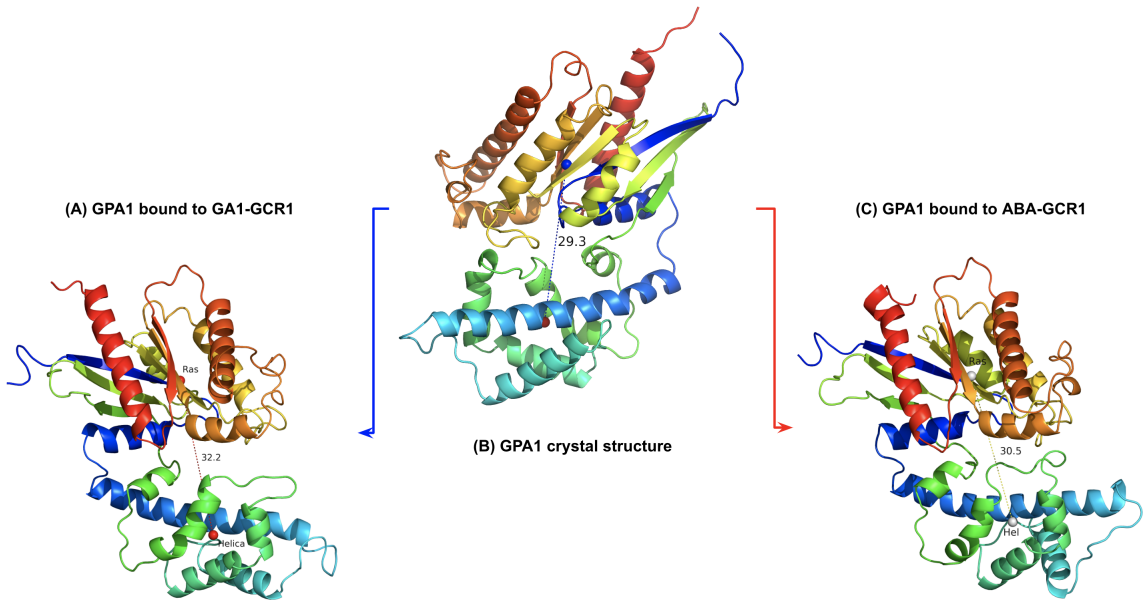


Figure 6.44: Center of masses comparison among panel (B) GPA1 crystal structure and GPA1 bound to GA1 in panel (A) and ABA in panel (C).

Finally, we compared our predicted tertiary structure of GCR1 with the 3D structures predicted by AlphaFold 2 and Taddese. We consider that our use of complete sampling to predict the structure followed by Molecular Dynamics lipid, solvent, and ion physiological environment, makes our structures more accurate. It would be interesting to insert the AlphaFold 2 and the Taddese structures into the full lipid and solvent and ion physiological environment followed by MD to see what differences there would be.

## 6.9 Comparative structural analysis between Tadesses's, AlphaFold 2, and GEnSeMBLE models

Taddese's study [1] utilized 5-fold recognition methods, which successfully matched GCR1 to the GPCR fold. The alignments of GCR1 homologs to class A, B, and F, GPCRs were clear and mutually consistent. Additionally, the helix-helix alignments involving GCR1 homologs exhibited a higher similarity than the well-established class A-class, B-class, and F-class GPCR alignments. The alignment identified 15 motifs that GCR1 shares with class A and class B GPCRs.

The artificial intelligence program AlphaFold 2 has predicted the tertiary structure of numerous proteins, including apo-GCR1 [2, 87, 88]. Figure 6.45 compares the AlphaFold 2 tertiary structure (residues 22 to residue 270) of apo-GCR1 in cyan with the GEnSeMBLE predicted structure in green. Notably, these predicted structures exhibit significant differences, as indicated by an RMSD of 5.027 Å (from PyMOL[89]).

In Table 6.21, it becomes evident that some transmembrane helices (TMDs) differ substantially between the three predictions. The most evident differences are given in helices 5 to 7. Helices 5 and 7 of Tadesse are noticeably shorter than the same helices of the other models. Helices 5 and 6 of AlphaFold are longer than the same helices of the other models. Helix 7 of GEnSeMBLE and AlphaFold display similar lengths.

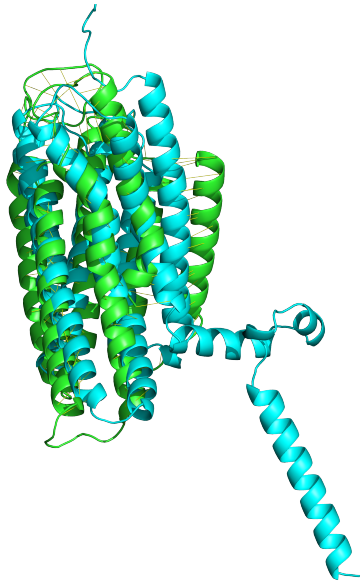

Figure 6.45: structural alignment of the two predicted proteins from residues 22 to residue 270 shows the similarity between the tertiary structure of GCR1 predicted by GEnSeMBLE (cyan) and by AlphaFold (green)

| Helix | GEnSeMBLE | Tadesse | AlphaFold 2 |
|---|---|---|---|
| 1 | I22-L43 | T20-L43 | L11-L43 |
| 2 | K49-V71 | F52-I70 | F50-V71 |
| 3 | I79-K111 | C80-H106 | G77 - V109 |
| 4 | E117-F140 | F123-F140 | E120-G141 |
| 5 | F164-M195 | H163-V185 | L157-A196 |
| 6 | D204-I234 | Q205-L215 | D201-I241 |
| 7 | H245-Y269 | A259-Y269 | F248-G270 |

Table 6.21: TM domains of for apo-GCR1 prediceted by GEnSeMBLE, Tadesse, and AlphaFold 2.

It has been observed [2, 87, 88] that the AlphaFold 1 predicted secondary structures are longer than observed experimentally. Although Alphafold 2 is considered to have solved the overcounting of interactions issues of AlphaFold 1, the predicted structure of GCR1 seems to show that AlphaFold 2 may still have problems. The intra and extracellular residues of the long

TMD 1, 5, and 6 of AlphaFold 2, which project well beyond the lipid bilayer, are exposed to polar and electrostatic interactions with water molecules and ions. We would expect strong electrostatic interactions that would affect the structural stability of the helices. Although GEnSeMBLE also predicts that TMD 5 and 6 project beyond the lipid bilayer, the use of Molecular Dynamics takes into account the polar and electrostatic interactions with water molecules and ions, producing more realistic lengths of the helices.

Is also interesting to notice that the highly conserved disulfide bridge C80(3.25)-C151(ECL2), was correctly predicted by all three methodologies. On the other hand, our predicted K49-D115 ion lock, which we consider a possible binding site for the GPA1 G-protein, was not identified by AlphaFold 2 nor by Taddese's model. In fact, Taddese predicted an ionic lock between K49 and E211, in the inactive structure.

# Chapter 7

# Summary and Conclusions

Using homology methods and first-principles molecular dynamics, we found that GCR1 displays the 7-TM domains typical of class-A animal GPCRs. The binding affinity of GCR1 to the phytohormones ABA and GA1 was obtained at the level of molecular docking. The resulting complexes ABA-GCR1 and GA1-GCR1 display evident energetic, structural, and conformational differences. Using the same methodologies described above, we modeled the complexes ABA-GCR1-GPA1 and GA1-GCR1-GPA1. We found that depending on the complex, ABA-GCR1-GPA1 or GA1-GCR1-GPA1, the protein GPA1 displays conformational differences measured as the distance between this protein's Ras and Helical domains. The complex GA1-GCR1-GPA1 promotes the Ras-Helical distance increase, facilitating the GDP/GTP exchange and the subsequent activation of GPA1, GPA1*. In contrast, the complex ABA-GCR1-GPA1 promotes the Ras-Helical distance shortening related to the inactive form of GPA1. These computational results agree with the experimental observations about the processes of germination and dormancy of seeds: germination is related to the active form of GPA1. Meanwhile, dormancy is related to the inactive form of GPA1. Therefore, GA1 would promote the germination of seeds. ABA, on the other hand, would promote the dormancy of the seeds. A deep and detailed analysis of the interaction between ABA and GA1 with the active site of GCR1 has allowed us to identify and propose extracellular binding domain mutations that would modify the affinity of the binding site of GCR1 to GA1 and ABA, and, ultimately, the response of GPA1.

Contrary to releasing GP for its self-activation, proposed by us in the hypothesis, our computational results show that the GP was not released away from GCR1 for self-activation, but the GA1-GCR1 binding alters the inter-

actions between GCR1 and GPA1. As a result, the Ras-Helical domains of GPA1 undergo modifications, causing GPA1 to open up and undergo GDP/GTP exchange, ultimately leading to the activation of GP. On the other hand, our calculations confirm our hypothesis with respect to ABA, since the ABA binding to GCR1 induced the formation of interactions between GCR1 and GP that halted the opening of the GP keeping GP in its inactive state.

We have adapted the GEnSeMBLE methodology for predicting the tertiary structure of the plant protein GCR1. This is the first predicted tertiary structure of this stress-regulating protein validated using first-principles-based methods. This involved performing a sequence alignment using available plant protein sequences to accurately capture plant proteins' unique characteristics, aligning the GCR1 with animal and human GPCRs, and accurately aligning to distantly related proteins. From this, we identified GPCRs with known 3D structures that could serve as templates to generate the best starting 3D structure using a homology-based model of GCR1. This model was then used to construct the best bundles of the seven transmembrane helices in GCR1.

The tertiary structure of GCR1 was predicted from its primary sequence using a methodology that has been well-validated for human GPCRs. This led to the following unique findings: (i) Motif NSxxY in GCR1-TM7 is conserved, this motif is present as NPxxY in class A GPCRs, this finding allowed us to assign the $\eta$ residue in helix 7. (ii) The residue P170, assigned as P5.50 of GCR1, is highly conserved, and produces an alignment between Y177 of GCR1 with Y5.57 (Y5.58) of class A GPCRs, confirming P170 as P5.50 in GCR1. (iii) The disulfide bond C80(3.25)-C151(ECL2) found in GCR1 is characteristic of class A GPCRs. (iv) Although the 3-6 ionic lock is absent in GCR1, which is highly conserved in animal GPCRs, we found a homologous site between helices 2-4, residues K49(2.37)-D115(4.37).

Using the predicted tertiary structure of GCR1, we studied the extracellular binding of potential ligands ABA and GA1. We found that HSP88(3.33) interacts via a salt bride with the carboxylate group on both ABA and GA1. Furthermore, we found a series of hydrogen bonds at the binding site between residues of GCR1 and polar groups for both ligands. Since the interactions in the binding site between GCR1 and both ligands are similar, we evaluated changes in the intracellular region of each ligand-GCR1 complex, finding differences as shifts in the helical structures. The binding site of the complex GA1-GCR1 has more polar interactions than the binding site of the complex ABA-GCR1. There are also important conformational differences in the intracellular region promoted by the interactions

between the GA1-GCR1 and ABA-GCR1 complexes. These structural differences generate different interactions between GCR1 and GPA1, which is related to the antagonistic behavior of those hormones after binding to GCR1.

To establish the possibility of a GDP-GTP exchange mechanism, we tracked the distance between the centers of mass of the domains Ras and Helical of GPA1 (involved in the G-protein regulation)[12] using metadynamics. Our MD simulations showed that on average, the Ras-Helical distance is larger for the complex GA1-GCR1-GPA1 than for the complex ABA-GCR1-GPA1. The larger Ras-Helical distance of GA1-GCR1-GPA1 would facilitate the GDP/GTP exchange, while the shorter Ras-Helical of ABA-GCR1-GPA1 would prevent this exchange. In more detail, we have found that the interaction of ABA with GCR1 ends up promoting the closure of GPA1, through the formation of several SBs between the Ras and helical domains, especially the one K288-D162, analog to the SB D150-K270 found in the animal G$\alpha i$ subunit and essential in keeping the GP inactive. An important aspect behind the relation between the complex ABA-GCR1 and the formation of SB K288-D162 is that Jones and collaborators did not find this SB. Hence, they proposed that the absence of this interaction is related to the self-activating activity of GPA1 [12]. Our findings demonstrate that GCR1 binds to ABA and GPA1 and that GCR1 promotes the inactivation of the GP through its interaction with ABA. On the other hand, the interaction GA1-GCR1 encourages the breaking of the SB K288-D162 and promotes the opening of GPA1 reflected in a more significant Ras-Helical distance and the absence of inter-domains SBs found in the ABA-GCR1-GPA1 complex. Although we found that the Ras-Helical distance for the GA1-GCR1-GPA1 system reached no more than 34 Å, we observed that this displacement is not a simple scissoring but also is a twisting between these domains which gives more room to promote a GDP-GTP exchange and ultimately relating the GA1-GCR1 interaction with the activation of GPA1.

Our metadynamics simulation results provides insightful potential of mean force (PMF) profiles for the ABA-GCR1-GPA1-GDP-$Mg^{2+}$ and GA1-GCR1-GPA1-GDP-$Mg^{2+}$ complexes, delineated by the interplay between the Ras and Helical domains of GPA1. These PMFs exhibit distinctive features as asymmetric double wells. Notably, the ABA complex manifests left and right wells at approximately $\approx 30 - 32, \text{Å}$ and $\approx 35 - 37, \text{Å}$, respectively, with the left well entailing lower free energy than its counterpart. In contrast, the GA1 complex showcases a reversed energetic profile, with the left well exhibiting higher free energy than the right well. Moreover,

the energy barrier demarcating the wells is more pronounced in the ABA complex compared to the GA1 complex. The comparative analysis of free energy landscapes unequivocally underscores ABA's role in maintaining the inactive conformation of the GPA1 protein, while GA1 actively promotes the transition of GPA1 into its active conformation.

These findings about GA1, ABA, GCR1, and GPA1, provide evidence to suggest that GCR1 is mediating agonist and antagonist mechanisms to activate or inactivate GPA1. Furthermore, we propose, based on the results of our simulations, that the inter-domain Ras-Hel SB between R190-E254 in GPA1 is a synergistic SB working together with K288-D162 to keep GPA1 inactive by binding of ABA in the extracellular region of GCR1. From the equilibrated complex, we propose residue mutations that would have a significant effect on the signaling pathway of the GP regulation, which would allow us to experimentally test (using gene editing tools) the predicted structure for GCR1 providing new evidence and insights that would improve our understanding of the role of GCR1 in plant physiology and its relationship with GPCRs.

# Chapter 8

# Outlook and future work

In this work, we identify a pocket in the GCR1 extracellular region where we bound ABA and GA1 to it. After equilibrating both complexes ABA-GCR1 and GA1-GCR1, in a physiological environment, we found two different conformational changes on GCR1, which we concluded as induced by the binding of ABA and GA1 to the same apo-GCR1. Then, I suggest that future doctoral research can use those findings to generate a series of mutations on the binding site using Crisper-Cas to confirm or reject our conclusions. In this way, the mutations to GCR1 proposed in tables 6.19 and **??** must be evaluated computationally and experimentally. The theoretical results of the simulations will provide helpful information that could support the signaling mechanism proposed in this dissertation.

A second doctoral research that can be derived from my work is based on the fact that I found an ionic lock between helices 2 and 4, similar to that described for animal GPCRs between helices 3 and 6, where I bound the terminal carboxylate group of GPA1. Although I did not find an ionic lock between H3 and H6, some Class A animal GPCRs also lack of this ionic lock, and the G-$\alpha$ subunit binds to H6 positive residue on the GPCR, which can be used to propose a different regulation mechanism for GPA1.

Additionally, it is important to study whether a synergistic work exists between GCR1 and RGS1 to regulate GPA1 or how is the relationship between GCR1 and RGS1 in the mechanism involved in regulating GPA1.

On the other hand, it has been proposed that GPCRs undergo pre-activation before binding with the ligands. The intracellular binding between the GPCR and the G protein induces this pre-activation. In the case of GCR1, we have the equilibrated complexes between GCR1, the ligands ABA and GA1, and the alpha subunit of the G protein (GPA1). By remov-

ing the ligands from these equilibrated structures, we expect to obtain the pre-activated GCR1 structure. We are currently working on this direction.

We are currently running metadynamics simulations on the GA1-GCR1-GPA1 and ABA-GCR1-GPA1. We expect to publish these results in order to validate our hypothesis about the mechanisms that regulate GPA1 mediated by the interactions of ABA-GCR1 and GA1-GCR1.

# Chapter 9

# Acknowledgement

I thank my doctoral director, professors Andrés Jaramillo-Botero from Pontificia Universidad Javeriana, Cali, and Carlos Arango from Universidad Icesi, Cali, their invaluable guidance, support, and encouragement throughout my graduate studies. Their expertise and insight have been instrumental in shaping my research and helping me develop as a scientist.

I would also like to thank Bill Goddard and Soo-Kyung Kim, who have played instrumental roles in shaping my research and providing invaluable guidance and feedback. Their insightful comments and suggestions have helped me refine my ideas and approach, and their support and encouragement have helped me stay focused and motivated during challenging times.

Finally, I would like to thank Goddard's group, who welcomed me into their research group at Caltech with open arms. Their willingness to share their knowledge and expertise, and their generosity in providing me with resources and support, have been instrumental in helping me navigate the challenges of graduate school.

I am deeply grateful to all of these individuals for their contributions to my academic and personal growth. I feel privileged to have had the opportunity to learn from and work with them.

I want to express my heartfelt appreciation to my wife for her unwavering love, support, and patience throughout my graduate studies. Her encouragement, understanding, and belief in me have been essential in enabling me to achieve my goals. Thank you, my love, for being my rock, my inspiration, and my partner in all of life's adventures. I could not have achieved any of this without you by my side.

I would like to express my special thanks to the Pontificia Universidad Javeriana campus in Cali for the tremendous support provided during my

doctoral studies. I would like to extend a very special thanks to the Deans of the Faculty of Health Sciences, Dr. Pedro Villamizar and Dr. Willy Paul Stangl Herrera, to the directors of the Department of Basic Health Sciences, Dr. Helberg Asencio and Dr. Freddy Moreno, and posthumously to the late first Director of the Department, Dr. Paula Margarita Hurtado. Her guidance and support were integral to completing my studies, and her passing is deeply felt.

This work has been partially funded by the OMICAS Program: In-silico Multiscale Optimization of Sustainable Agricultural Crops (Infrastructure and validation in Rice and Sugar Cane), sponsored within the Scientific Colombia Ecosystem, made up by the World Bank, Ministry of Science, Technology and Innovation (Minciencias), Icetex, Ministry of Education and Ministry of Industry and Tourism, Project ID: FP44842-217-2018.

# Chapter 10

# Appendices

## 10.1 Appendix I: MFTA files based on the plant alignment

return to main text

These four mfta files, Figures 10.1-10.4, contain the amino acid sequence of GCR1, followed by 7 rows, one per each helix of GCR1. The second and third columns of these rows contain the initial and final residue of each helix, the tenth and eleventh columns display the type and position of the $\eta$ residue. The position of the hydrophobic center residue of each helix is given by the second column of last seven rows of the files.

```
>P1;sp|O04714|GCR1_ARATH G-protein coupled receptor 1
OS=Arabidopsis thaliana OX=3702 GN=GCR1 PE=1 SV=1
MSAVLTAGGGLTAGDRSIITAINTGASSLSFVGSAFIVLCYCLFKELRKFSFKLVFYLAL
SDMLCSFFLIVGDPSKGFICYAQGYTTHFFCVASFLWTTTIAFTLHRTVVKHKTDVEDLE
AMFHLYVWGTSLVVTVIRSFGNNHSHLGPWCWTQTGLKGKAVHFLTFYAPLWGAILYNGF
TYFQVIRMLRNARRMAVGMSDRVDQFDNRAELKVLNRWGYYPLILIGSWAFGTINRIHDF
IEPGHKIFWLSVLDVGTAALMGLFNSIAYGFNSSVRRAIHERLELFLPERLYRWLPSNFR
PKNHLILHQQQQQRSEMVSLKTEDQQ
* 1tm     14   43   X X   X X   X X   G   25   A
* 2tm     53   79   X X   X X   X X   D   62   A
* 3tm     75  109   X X   X X   X X   T   87   A
* 4tm    122  143   X X   X X   X X   W  128   A
* 5tm    152  177   X X   X X   X X   L  157   A
* 6tm    213  245   X X   X X   X X   I  224   A
* 7tm    253  274   X X   X X   X X   M  261   A
* 1hpc     20.003   X.00   X.00   X.00   X.00   X.00
* 2hpc     63.75   X.00   X.00   X.00   X.00   X.00
* 3hpc     90.16   X.00   X.00   X.00   X.00   X.00
* 4hpc    135.19   X.00   X.00   X.00   X.00   X.00
* 5hpc    150.004   X.00   X.00   X.00   X.00   X.00
* 6hpc    221.67   X.00   X.00   X.00   X.00   X.00
* 7hpc    256.13   X.00   X.00   X.00   X.00   X.00
```

Figure 10.1: Human Prostaglandin D2 receptor 2, PD2R2 with 12.88% sequence identity (21.39% for TMD).

```
>P1;sp|O04714|GCR1_ARATH G-protein coupled receptor 1
OS=Arabidopsis thaliana OX=3702 GN=GCR1 PE=1 SV=1
MSAVLTAGGGLTAGDRSIITAINTGASSLSFVGSAFIVLCYCLFKELRKFSFKLVFYLAL
SDMLCSFFLIVGDPSKGFICYAQGYTTHFFCVASFLWTTTIAFTLHRTVVKHKTDVEDLE
AMFHLYVWGTSLVVTVIRSFGNNHSHLGPWCWTQTGLKGKAVHFLTFYAPLWGAILYNGF
TYFQVIRMLRNARRMAVGMSDRVDQFDNRAELKVLNRWGYYPLILIGSWAFGTINRIHDF
IEPGHKIFWLSVLDVGTAALMGLFNSIAYGFNSSVRRAIHERLELFLPERLYRWLPSNFR
PKNHLILHQQQQQRSEMVSLKTEDQQ
* 1tm     13   44   X X   X X   X X   S   34   A
* 2tm     50   77   X X   X X   X X   D   62   A
* 3tm     76  110   X X   X X   X X   T   87   A
* 4tm    117  142   X X   X X   X X   W  128   A
* 5tm    152  187   X X   X X   X X   A  174   A
* 6tm    211  245   X X   X X   X X   F  240   A
* 7tm    243  273   X X   X X   X X   S  266   A
* 1hpc     27.33   X.00   X.00   X.00   X.00   X.00
* 2hpc     64.57   X.00   X.00   X.00   X.00   X.00
* 3hpc     91.009   X.00   X.00   X.00   X.00   X.00
* 4hpc    130.55   X.00   X.00   X.00   X.00   X.00
* 5hpc    171.006   X.00   X.00   X.00   X.00   X.00
* 6hpc    237.71   X.00   X.00   X.00   X.00   X.00
* 7hpc    259.22   X.00   X.00   X.00   X.00   X.00
```

Figure 10.2: Human P2Y purinoceptor 1, P2RY1, with 12.27% sequence identity (17.54% for TMD).

```
>P1;sp|O04714|GCR1_ARATH G-protein coupled receptor 1
OS=Arabidopsis thaliana OX=3702 GN=GCR1 PE=1 SV=1
MSAVLTAGGGLTAGDRSIITAINTGASSLSFVGSAFIVLCYCLFKELRKFSFKLVFYLAL
SDMLCSFFLIVGDPSKGFICYAQGYTTHFFCVASFLWTTTIAFTLHRTVVKHKTDVEDLE
AMFHLYVWGTSLVVTVIRSFGNNHSHLGPWCWTQTGLKGKAVHFLTFYAPLWGAILYNGF
TYFQVIRMLRNARRMAVGMSDRVDQFDNRAELKVLNRWGYYPLILIGSWAFGTINRIHDF
IEPGHKIFWLSVLDVGTAALMGLFNSIAYGFNSSVRRAIHERLELFLPERLYRWLPSNFR
PKNHLILHQQQQQRSEMVSLKTEDQQ
* 1tm     17   43   X X   X X   X X   S   34   A
* 2tm     50   79   X X   X X   X X   D   62   A
* 3tm     76  110   X X   X X   X X   T   87   A
* 4tm    117  140   X X   X X   X X   W  128   A
* 5tm    156  194   X X   X X   X X   L  157   A
* 6tm    209  241   X X   X X   X X   I  224   A
* 7tm    251  270   X X   X X   X X   S  266   A
* 1hpc     26.74   X.00   X.00   X.00   X.00   X.00
* 2hpc     64.34   X.00   X.00   X.00   X.00   X.00
* 3hpc     91.57   X.00   X.00   X.00   X.00   X.00
* 4hpc    130.007   X.00   X.00   X.00   X.00   X.00
* 5hpc    154.45   X.00   X.00   X.00   X.00   X.00
* 6hpc    221.57   X.00   X.00   X.00   X.00   X.00
* 7hpc    259.75   X.00   X.00   X.00   X.00   X.00
```

Figure 10.3: Wild turkey Beta-1 adrenergic receptor, ADRB1, with 12.11% sequence identity (17.27% for TMD)

```
>P1;sp|O04714|GCR1_ARATH G-protein coupled receptor 1
OS=Arabidopsis thaliana OX=3702 GN=GCR1 PE=1 SV=1
MSAVLTAGGGLTAGDRSIITAINTGASSLSFVGSAFIVLCYCLFKELRKFSFKLVFYLAL
SDMLCSFFLIVGDPSKGFICYAQGYTTHFFCVASFLWTTTIAFTLHRTVVKHKTDVEDLE
AMFHLYVWGTSLVVTVIRSFGNNHSHLGPWCWTQTGLKGKAVHFLTFYAPLWGAILYNGF
TYFQVIRMLRNARRMAVGMSDRVDQFDNRAELKVLNRWGYYPLILIGSWAFGTINRIHDF
IEPGHKIFWLSVLDVGTAALMGLFNSIAYGFNSSVRRAIHERLELFLPERLYRWLPSNFR
PKNHLILHQQQQQRSEMVSLKTEDQQ
* 1tm     22   44   X X   X X   X X   S   34   A
* 2tm     50   79   X X   X X   X X   D   62   A
* 3tm     78  110   X X   X X   X X   T   87   A
* 4tm    117  139   X X   X X   X X   W  128   A
* 5tm    145  181   X X   X X   X X   L  157   A
* 6tm    198  232   X X   X X   X X   L  223   A
* 7tm    247  275   X X   X X   X X   S  266   A
* 1hpc     27.52   X.00   X.00   X.00   X.00   X.00
* 2hpc     64.92   X.00   X.00   X.00   X.00   X.00
* 3hpc     91.37   X.00   X.00   X.00   X.00   X.00
* 4hpc    130.36   X.00   X.00   X.00   X.00   X.00
* 5hpc    155.78   X.00   X.00   X.00   X.00   X.00
* 6hpc    221.42   X.00   X.00   X.00   X.00   X.00
* 7hpc    259.42   X.00   X.00   X.00   X.00   X.00
```

Figure 10.4: Human Prostaglandin E2 receptor EP3 subtype, PE2R3 with 13.19% sequence identity (17.13% for TMD)

## 10.2 Appendix II: MFTA files based on animal alignment

return to main text

Bellow we show a mfta file for each model based on the template P2RY1, Figures 10.5 to 10.7 where are the respective $\eta$ residues that characterize each model as m1, m2, and m3.

```
>P1;sp|O04714|GCR1_ARATH G-protein coupled receptor 1
OS=Arabidopsis thaliana OX=3702 GN=GCR1 PE=1 SV=1
MSAVLTAGGGLTAGDRSIITAINTGASSLSFVGSAFIVLCYCLFKELRKFSFKLVFYLAL
SDMLCSFFLIVGDPSKGFICYAQGYTTHFFCVASFLWTTTIAFTLHRTVVKHKTDVEDLE
AMFHLYVWGTSLVVTVIRSFGNNHSHLGPWCWTQTGLKGKAVHFLTFYAPLWGAILYNGF
TYFQVIRMLRNARRMAVGMSDRVDQFDNRAELKVLNRWGYYPLILIGSWAFGTINRIHDF
IEPGHKIFWLSVLDVGTAALMGLFNSIAYGFNSSVRRAIHERLELFLPERLYRWLPSNFR
PKNHLILHQQQQQRSEMVSLKTEDQQ
* 1tm     13   44   X X   X X   X X   S   34   A
* 2tm     50   69   X X   X X   X X   D   62   A
* 3tm     83  110   X X   X X   X X   T   87   A
* 4tm    117  142   X X   X X   X X   W  128   A
* 5tm    152  187   X X   X X   X X   P  170   A
* 6tm    202  236   X X   X X   X X   P  222   A
* 7tm    246  271   X X   X X   X X   S  266   A
* 1hpc     27.27   X.00   X.00   X.00   X.00   X.00
* 2hpc     64.62   X.00   X.00   X.00   X.00   X.00
* 3hpc     91.003   X.00   X.00   X.00   X.00   X.00
* 4hpc    130.54   X.00   X.00   X.00   X.00   X.00
* 5hpc    166.97   X.00   X.00   X.00   X.00   X.00
* 6hpc    219.65   X.00   X.00   X.00   X.00   X.00
* 7hpc    259.26   X.00   X.00   X.00   X.00   X.00
```

Figure 10.5: m1 for GCR1.mfta file based on template P2RY1

```
>P1;sp|O04714|GCR1_ARATH G-protein coupled receptor 1
OS=Arabidopsis thaliana OX=3702 GN=GCR1 PE=1 SV=1
MSAVLTAGGGLTAGDRSIITAINTGASSLSFVGSAFIVLCYCLFKELRKFSFKLVFYLAL
SDMLCSFFLIVGDPSKGFICYAQGYTTHFFCVASFLWTTTIAFTLHRTVVKHKTDVEDLE
AMFHLYVWGTSLVVTVIRSFGNNHSHLGPWCWTQTGLKGKAVHFLTFYAPLWGAILYNGF
TYFQVIRMLRNARRMAVGMSDRVDQFDNRAELKVLNRWGYYPLILIGSWAFGTINRIHDF
IEPGHKIFWLSVLDVGTAALMGLFNSIAYGFNSSVRRAIHERLELFLPERLYRWLPSNFR
PKNHLILHQQQQQRSEMVSLKTEDQQ
* 1tm     13   44   X X   X X   X X   S   34   A
* 2tm     50   69   X X   X X   X X   D   62   A
* 3tm     83  110   X X   X X   X X   T   87   A
* 4tm    117  142   X X   X X   X X   W  128   A
* 5tm    152  187   X X   X X   X X   P  170   A
* 6tm    211  236   X X   X X   X X   F  231   A
* 7tm    246  273   X X   X X   X X   A  268   A
* 1hpc     27.27   X.00   X.00   X.00   X.00   X.00
* 2hpc     64.62   X.00   X.00   X.00   X.00   X.00
* 3hpc     91.003   X.00   X.00   X.00   X.00   X.00
* 4hpc    130.54   X.00   X.00   X.00   X.00   X.00
* 5hpc    166.97   X.00   X.00   X.00   X.00   X.00
* 6hpc    228.65   X.00   X.00   X.00   X.00   X.00
* 7hpc    261.26   X.00   X.00   X.00   X.00   X.00
```

Figure 10.6: m2 for GCR1.mfta file based on template P2RY1

```
>P1;sp|O04714|GCR1_ARATH G-protein coupled receptor 1
OS=Arabidopsis thaliana OX=3702 GN=GCR1 PE=1 SV=1
MSAVLTAGGGLTAGDRSIITAINTGASSLSFVGSAFIVLCYCLFKELRKFSFKLVFYLAL
SDMLCSFFLIVGDPSKGFICYAQGYTTHFFCVASFLWTTTIAFTLHRTVVKHKTDVEDLE
AMFHLYVWGTSLVVTVIRSFGNNHSHLGPWCWTQTGLKGKAVHFLTFYAPLWGAILYNGF
TYFQVIRMLRNARRMAVGMSDRVDQFDNRAELKVLNRWGYYPLILIGSWAFGTINRIHDF
IEPGHKIFWLSVLDVGTAALMGLFNSIAYGFNSSVRRAIHERLELFLPERLYRWLPSNFR
PKNHLILHQQQQQRSEMVSLKTEDQQ
*  1tm   27   44   25   44    X    X    X    X    F   36    A
*  2tm   51   64   51   68    X    X    X    X    D   62    A
*  3tm   83  107   77  111    X    X    X    X    T   87    A
*  4tm  124  138  120  141    X    X    X    X    W  128    A
*  5tm  164  184  163  187    X    X    X    X    P  170    A
*  6tm  219  235  217  236    X    X    X    X    Y  220    A
*  7tm  251  270  246  276    X    X    X    X    S  266    A
*  1hpc    35.50    34.50    35.00    36.07    X.00    X.00
*  2hpc    57.50    59.50    57.00    57.24    X.00    X.00
*  3hpc    95.00    94.00   100.00    96.89    X.00    X.00
*  4hpc   131.00   130.50   131.00   130.34    X.00    X.00
*  5hpc   174.00   175.00   175.00   174.07    X.00    X.00
*  6hpc   227.00   226.50   226.00   226.11    X.00    X.00
*  7hpc   260.50   261.00   266.00   261.49    X.00    X.00
```

Figure 10.7: m3 for GCR1.mfta file based on template P2RY1

## 10.3 Appendix III: First BiHelix procedure, results using the four template-based homologized structures.

return to main text

Tables 10.1-10.3 summarize the BiHelix result for the three models m1,m2, and m3, for four homologized structures based on P2RY1, PE2R3, PD2fRho, and ADRB1 respectively.

| Strct | Eta | | | | | | | Source | Individual Energies | | Individual Ranks | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | H1 | H2 | H3 | H4 | H5 | H6 | H7 | Source | CInterH | NInterH | rCi | rNi |
| 1 | -90 | 30 | 0 | 150 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -639.749 | -364.550 | 3 | 43 |
| 2 | -90 | 30 | 30 | 60 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -654.037 | -380.406 | 1 | 6 |
| 3 | 180 | 30 | 0 | 90 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -623.241 | -356.225 | 21 | 85 |
| 4 | -90 | 30 | 0 | 30 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -622.425 | -373.946 | 24 | 17 |
| 5 | 30 | 90 | 30 | 180 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -605.271 | -357.846 | 86 | 73 |
| 6 | -60 | 30 | 0 | 30 | 0 | 150 | 0 | m1-GCR1-PE2R3 | -605.214 | -359.285 | 87 | 62 |
| 7 | 30 | -120 | 30 | 60 | 0 | 150 | -60 | m1-GCR1-PE2R3 | -607.329 | -370.870 | 76 | 22 |
| 8 | 180 | 30 | 0 | 0 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -604.027 | -356.990 | 94 | 78 |
| 9 | 150 | 30 | 0 | 150 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -625.243 | -366.698 | 16 | 34 |
| 10 | -60 | 30 | 0 | 30 | 0 | 120 | 0 | m1-GCR1-PE2R3 | -620.750 | -364.940 | 26 | 42 |
| 11 | -90 | 30 | 0 | 120 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -603.977 | -356.161 | 95 | 86 |
| 12 | -30 | 90 | 30 | 120 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -630.624 | -359.913 | 13 | 59 |
| 13 | -90 | 30 | 0 | -60 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -608.042 | -356.473 | 73 | 83 |
| 14 | -90 | 30 | 90 | 180 | 0 | 150 | 0 | m1-GCR1-PE2R3 | -622.692 | -347.281 | 22 | 191 |
| 15 | 180 | 30 | 0 | 30 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -605.636 | -353.658 | 83 | 113 |
| 16 | -150 | 90 | 30 | -150 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -627.517 | -375.387 | 15 | 11 |
| 17 | -90 | 30 | 0 | -60 | 0 | 150 | 0 | m1-GCR1-PE2R3 | -615.968 | -347.892 | 40 | 184 |
| 18 | -150 | 90 | 30 | 180 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -594.607 | -350.624 | 157 | 138 |
| 19 | -30 | 30 | 0 | 120 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -612.716 | -351.456 | 51 | 129 |
| 20 | -90 | 30 | 0 | 150 | 0 | 150 | 0 | m1-GCR1-PE2R3 | -608.399 | -349.048 | 71 | 168 |
| 21 | 180 | 30 | 0 | 120 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -610.530 | -343.652 | 58 | 242 |
| 22 | -30 | 90 | 0 | 120 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -590.075 | -358.950 | 202 | 65 |
| 23 | -90 | 60 | 0 | 60 | 0 | 150 | 0 | m1-GCR1-PE2R3 | -612.804 | -363.045 | 50 | 51 |
| 24 | 120 | 60 | 0 | 150 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -605.558 | -355.159 | 84 | 98 |
| 25 | -90 | 30 | 0 | 90 | 0 | 150 | -30 | m1-GCR1-PE2R3 | -615.200 | -351.156 | 43 | 132 |
| 295 | -60 | -90 | 60 | 30 | 60 | 60 | -30 | m1-GCR1-P2RY1 | -484.710 | -388.090 | 1817 | 1 |
| 337 | -30 | -120 | 60 | 30 | 60 | 60 | -60 | m1-GCR1-P2RY1 | -496.356 | -366.451 | 1616 | 37 |
| 517 | 0 | 0 | 30 | 0 | -150 | 120 | 60 | m1-GCR1-ADRB1 | -531.894 | -305.725 | 1064 | 1355 |
| 4896 | -150 | 30 | 180 | 90 | 90 | 60 | -60 | m1-GCR1-PD2 | -316.698 | -303.221 | 5851 | 1466 |

Table 10.1: Results for the 4 homologized GCR1 structures built from P2RY1, PE2R3, PD2fRho, and ADRB1 templates, proposed in Model 1

Similarly, in Tables 10.2 and 10.3 we summarize the BiHelix results for the models 2 and 3, based on P2RY1, PE2R3, PD2fRho, and ADRB1 receptors, respectively.

| | Eta | | | | | | | Source | I. Energies | | I. Ranks | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | H1 | H2 | H3 | H4 | H5 | H6 | H7 | Source | CInterH | NInterH | rCi | rNi |
| 1 | -90 | 30 | 0 | 150 | 0 | 60 | -30 | m2-GCR1-PE2R3 | -536.728 | -378.733 | 21 | 6 |
| 2 | -120 | 60 | 30 | 150 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -558.579 | -396.710 | 5 | 2 |
| 3 | 120 | 0 | 30 | 180 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -517.189 | -363.646 | 63 | 293 |
| 4 | -90 | 30 | 90 | 30 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -529.467 | -365.908 | 27 | 24 |
| 5 | 30 | 90 | 30 | 180 | 0 | 90 | -30 | m2-GCR1-PE2R3 | -511.410 | -366.624 | 80 | 23 |
| 6 | -90 | 30 | 90 | 150 | 0 | 90 | -30 | m2-GCR1-PE2R3 | -519.230 | -343.536 | 56 | 68 |
| 7 | 30 | 90 | 30 | -150 | 0 | 90 | -30 | m2-GCR1-PE2R3 | -510.942 | -374.623 | 82 | 13 |
| 8 | -120 | 60 | 30 | 120 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -547.543 | -390.312 | 10 | 3 |
| 9 | -90 | 30 | 0 | 30 | 0 | 60 | -30 | m2-GCR1-PE2R3 | -518.230 | -376.177 | 60 | 9 |
| 10 | 30 | 90 | 30 | 180 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -495.646 | -369.505 | 159 | 19 |
| 11 | -90 | 30 | 90 | 150 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -528.870 | -341.349 | 28 | 86 |
| 12 | -120 | 60 | 30 | 0 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -525.093 | -374.437 | 36 | 14 |
| 13 | -90 | 30 | 0 | 150 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -507.298 | -357.461 | 101 | 33 |
| 14 | 30 | -120 | 30 | 60 | 0 | 90 | -90 | m2-GCR1-PE2R3 | -498.036 | -371.952 | 142 | 16 |
| 15 | -90 | 30 | 0 | 150 | 0 | 90 | -30 | m2-GCR1-PE2R3 | -501.108 | -347.253 | 130 | 53 |
| 16 | 120 | 0 | 30 | 90 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -569.016 | -399.659 | 1 | 1 |
| 17 | 30 | 90 | 30 | -150 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -488.182 | -379.287 | 207 | 5 |
| 18 | -120 | 90 | 30 | -150 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -498.534 | -374.678 | 139 | 12 |
| 19 | -60 | 30 | 0 | 150 | 0 | 90 | 0 | m2-GCR1-PE2R3 | -501.670 | -343.439 | 126 | 71 |
| 20 | -90 | 30 | 0 | 120 | 0 | 60 | -30 | m2-GCR1-PE2R3 | -508.415 | -345.760 | 94 | 57 |
| 21 | -90 | 30 | 0 | 0 | 0 | 60 | -30 | m2-GCR1-PE2R3 | -498.958 | -355.453 | 137 | 38 |
| 22 | -90 | 30 | 90 | 30 | 0 | 90 | -30 | m2-GCR1-PE2R3 | -485.733 | -340.347 | 227 | 91 |
| 23 | -60 | 30 | 90 | 30 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -519.808 | -351.399 | 52 | 46 |
| 24 | 30 | 90 | 0 | 30 | 0 | 90 | -60 | m2-GCR1-PE2R3 | -484.940 | -358.729 | 238 | 32 |
| 25 | -90 | 30 | 0 | 30 | 0 | 90 | -30 | m2-GCR1-PE2R3 | -497.953 | -348.614 | 143 | 48 |

Table 10.2: Results for the 4 homologized GCR1 structures built from P2RY1, PE2R3, PD2fRho, and ADRB1 templates, proposed in Model 2

| | Eta | | | | | | | Source | Energies | | Ranks | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | H1 | H2 | H3 | H4 | H5 | H6 | H7 | Source | CInterH | NInterH | rCi | rNi |
| 1 | 120 | -120 | 30 | 180 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -614.214 | -365.047 | 18 | 10 |
| 2 | -120 | -90 | 30 | 60 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -625.622 | -376.632 | 8 | 6 |
| 3 | -90 | 60 | 0 | 30 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -615.814 | -383.823 | 16 | 2 |
| 4 | 90 | 0 | 30 | 180 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -636.980 | -380.442 | 2 | 3 |
| 5 | 90 | 0 | 30 | 60 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -614.432 | -377.621 | 17 | 4 |
| 6 | 90 | 0 | 30 | -120 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -623.292 | -359.054 | 9 | 14 |
| 7 | 30 | 60 | 0 | 30 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -618.004 | -384.654 | 13 | 1 |
| 8 | -120 | -90 | 30 | 180 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -632.983 | -377.307 | 3 | 5 |
| 9 | 30 | 120 | 90 | -150 | 0 | 90 | -90 | m3-GCR1-PE2R3 | -619.971 | -360.323 | 12 | 13 |
| 10 | 90 | 30 | 0 | 0 | 120 | 0 | 0 | m3-GCR1-PE2R3 | -610.298 | -358.572 | 24 | 17 |
| 11 | 120 | 90 | 30 | -150 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -601.126 | -346.314 | 40 | 40 |
| 12 | 30 | 60 | 0 | 150 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -620.664 | -370.712 | 11 | 8 |
| 13 | 90 | 30 | 0 | 120 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -593.452 | -354.713 | 53 | 23 |
| 14 | 30 | 120 | 30 | 120 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -587.337 | -346.305 | 69 | 41 |
| 15 | 90 | 30 | 0 | 150 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -599.767 | -349.144 | 42 | 32 |
| 16 | 90 | 30 | 0 | 0 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -580.716 | -358.727 | 92 | 15 |
| 17 | 120 | 90 | 30 | 180 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -589.737 | -347.733 | 64 | 34 |
| 18 | 30 | 120 | 30 | 120 | 180 | 0 | 0 | m3-GCR1-PE2R3 | -580.806 | -345.551 | 91 | 44 |
| 19 | 30 | 60 | 0 | 30 | 120 | 0 | 0 | m3-GCR1-PE2R3 | -609.287 | -350.223 | 27 | 29 |
| 20 | 30 | 120 | 0 | 90 | 0 | 90 | -60 | m3-GCR1-PE2R3 | -592.698 | -338.535 | 56 | 71 |
| 21 | 30 | 120 | 0 | 120 | 0 | 90 | -60 | m3-GCR1-PE2R3 | -579.360 | -337.537 | 99 | 78 |
| 22 | 30 | 120 | 0 | 120 | 0 | 90 | -90 | m3-GCR1-PE2R3 | -575.110 | -343.554 | 117 | 50 |
| 23 | 90 | 0 | 30 | 30 | 0 | 0 | 0 | m3-GCR1-PE2R3 | -625.969 | -357.033 | 7 | 19 |
| 24 | 90 | 30 | 0 | 150 | 0 | 90 | -60 | m3-GCR1-PE2R3 | -612.622 | -347.412 | 20 | 36 |
| 25 | 90 | 30 | 0 | 30 | 120 | 0 | 0 | m3-GCR1-PE2R3 | -610.291 | -348.330 | 25 | 33 |

Table 10.3: Results for the 4 homologized GCR1 structures built from P2RY1, PE2R3, PD2fRho, and ADRB1 templates, proposed in Model 3

## 10.4 Appendix IV: mfta files proposed based on models m1-m3.

return to main text

```
>P1;sp|O04714|GCR1_ARATH G-protein coupled receptor 1
OS=Arabidopsis thaliana OX=3702 GN=GCR1 PE=1 SV=1
MSAVLTAGGGLTAGDRSIITAINTGASSLSFVGSAFIVLCYCLFKELRKFSFKLVFYLAL
SDMLCSFFLIVGDPSKGFICYAQGYTTHFFCVASFLWTTTIAFTLHRTVVKHKTDVEDLE
AMFHLYVWGTSLVVTVIRSFGNNHSHLGPWCWTQTGLKGKAVHFLTFYAPLWGAILYNGF
TYFQVIRMLRNARRMAVGMSDRVDQFDNRAELKVLNRWGYYPLILIGSWAFGTINRIHDF
IEPGHKIFWLSVLDVGTAALMGLFNSIAYGFNSSVRRAIHERLELFLPERLYRWLPSNFR
PKNHLILHQQQQQRSEMVSLKTEDQQ
* 1tm     22   44   X X   X X   X X   S   34   A
* 2tm     50   73   X X   X X   X X   D   62   A
```

```
* 3tm     78  110   X X   X X   X X   T   87   A
* 4tm    117  141   X X   X X   X X   W  128   A
* 5tm    163  194   X X   X X   X X   P  170   A
* 6tm    204  232   X X   X X   X X   P  222   A
* 7tm    246  270   X X   X X   X X   S  266   A
* 1hpc     27.28   X.00   X.00   X.00   X.00   X.00
* 2hpc     65.009   X.00   X.00   X.00   X.00   X.00
* 3hpc     91.30   X.00   X.00   X.00   X.00   X.00
* 4hpc    130.40   X.00   X.00   X.00   X.00   X.00
* 5hpc    168.88   X.00   X.00   X.00   X.00   X.00
* 6hpc    220.37   X.00   X.00   X.00   X.00   X.00
* 7hpc    259.44   X.00   X.00   X.00   X.00   X.00

>P1;sp|O04714|GCR1_ARATH G-protein coupled receptor 1
OS=Arabidopsis thaliana OX=3702 GN=GCR1 PE=1 SV=1
MSAVLTAGGGLTAGDRSIITAINTGASSLSFVGSAFIVLCYCLFKELRKFSFKLVFYLAL
SDMLCSFFLIVGDPSKGFICYAQGYTTHFFCVASFLWTTTIAFTLHRTVVKHKTDVEDLE
AMFHLYVWGTSLVVTVIRSFGNNHSHLGPWCWTQTGLKGKAVHFLTFYAPLWGAILYNGF
TYFQVIRMLRNARRMAVGMSDRVDQFDNRAELKVLNRWGYYPLILIGSWAFGTINRIHDF
IEPGHKIFWLSVLDVGTAALMGLFNSIAYGFNSSVRRAIHERLELFLPERLYRWLPSNFR
PKNHLILHQQQQQRSEMVSLKTEDQQ
* 1tm     22   44   X X   X X   X X   S   34   A
* 2tm     50   73   X X   X X   X X   D   62   A
* 3tm     78  110   X X   X X   X X   T   87   A
* 4tm    117  141   X X   X X   X X   W  128   A
* 5tm    163  194   X X   X X   X X   P  170   A
* 6tm    204  232   X X   X X   X X   F  231   A
* 7tm    246  270   X X   X X   X X   S  266   A
* 1hpc     27.28   X.00   X.00   X.00   X.00   X.00
* 2hpc     65.009   X.00   X.00   X.00   X.00   X.00
* 3hpc     91.30   X.00   X.00   X.00   X.00   X.00
* 4hpc    130.40   X.00   X.00   X.00   X.00   X.00
* 5hpc    168.88   X.00   X.00   X.00   X.00   X.00
* 6hpc    229.37   X.00   X.00   X.00   X.00   X.00
* 7hpc    259.44   X.00   X.00   X.00   X.00   X.00

>P1;sp|O04714|GCR1_ARATH G-protein coupled receptor 1
OS=Arabidopsis thaliana OX=3702 GN=GCR1 PE=1 SV=1
MSAVLTAGGGLTAGDRSIITAINTGASSLSFVGSAFIVLCYCLFKELRKFSFKLVFYLAL
SDMLCSFFLIVGDPSKGFICYAQGYTTHFFCVASFLWTTTIAFTLHRTVVKHKTDVEDLE
```

```
AMFHLYVWGTSLVVTVIRSFGNNHSHLGPWCWTQTGLKGKAVHFLTFYAPLWGAILYNGF
TYFQVIRMLRNARRMAVGMSDRVDQFDNRAELKVLNRWGYYPLILIGSWAFGTINRIHDF
IEPGHKIFWLSVLDVGTAALMGLFNSIAYGFNSSVRRAIHERLELFLPERLYRWLPSNFR
PKNHLILHQQQQQRSEMVSLKTEDQQ
* 1tm     22   44   X X   X X   X X   F   36   A
* 2tm     50   73   X X   X X   X X   D   62   A
* 3tm     78  110   X X   X X   X X   T   87   A
* 4tm    117  141   X X   X X   X X   W  128   A
* 5tm    163  194   X X   X X   X X   P  170   A
* 6tm    204  232   X X   X X   X X   Y  220   A
* 7tm    246  270   X X   X X   X X   S  266   A
* 1hpc     29.28   X.00   X.00   X.00   X.00   X.00
* 2hpc     65.009   X.00   X.00   X.00   X.00   X.00
* 3hpc     91.30   X.00   X.00   X.00   X.00   X.00
* 4hpc    130.40   X.00   X.00   X.00   X.00   X.00
* 5hpc    168.88   X.00   X.00   X.00   X.00   X.00
* 6hpc    218.37   X.00   X.00   X.00   X.00   X.00
* 7hpc    259.44   X.00   X.00   X.00   X.00   X.00
```

## 10.5 Appendix V: Second BiHelix results.

return to main text

Figures 10.8 - 10.14 show how, although some amino acids, in the proposed GCR1 structure, have good interhelical orientations, these amino acids have not the required distances found in the conserved contacts in animal GPCRs.

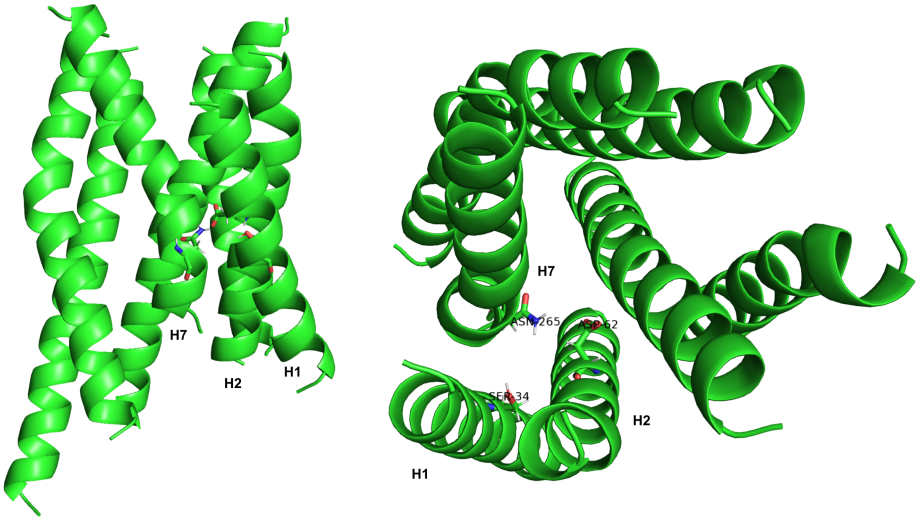


Figure 10.8: 1bBH.bgf is the best energy candidate, and also it holds 10 class A, and 1 active contact as shown in table 9.

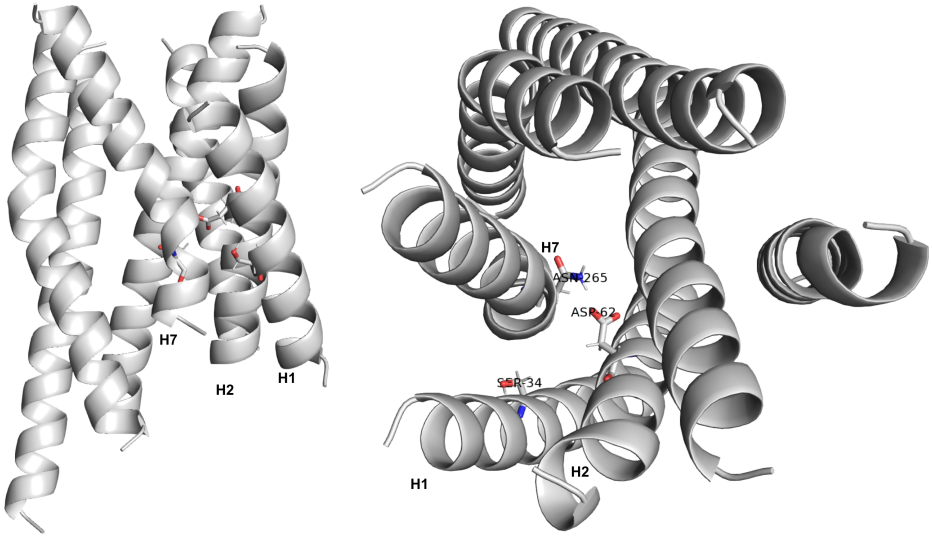


Figure 10.9: 4bBH.bgf has 14 class A contacts and 1 active contact.

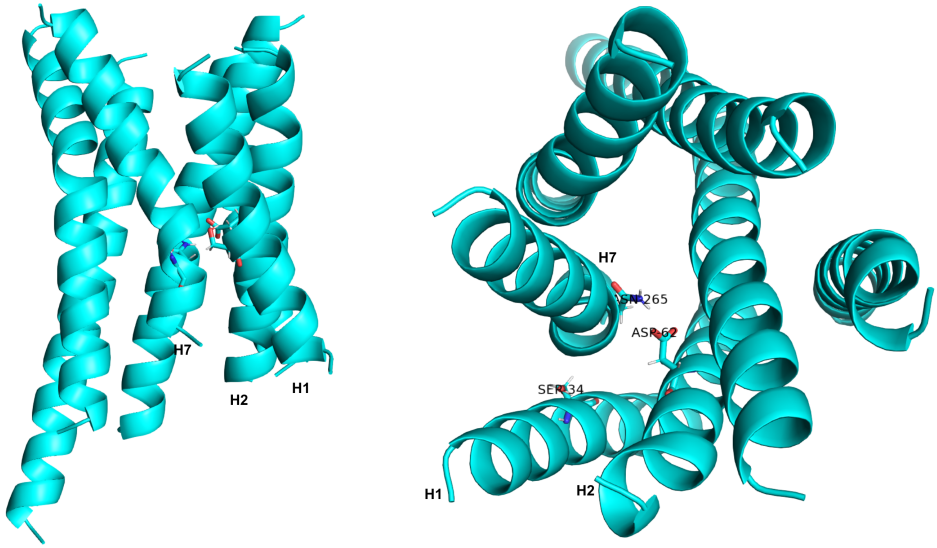


Figure 10.10: 5bBH.bgf has 21 class A contacts, 2 active, and 2 inactive contacts.

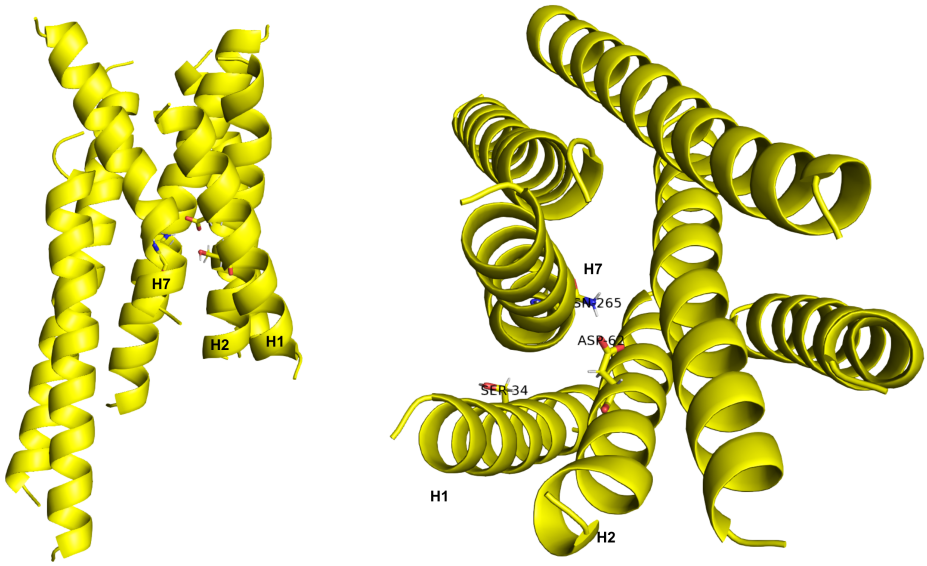


Figure 10.11: 6bBH.bgf has 7 class A contacts and 2 active contacts.

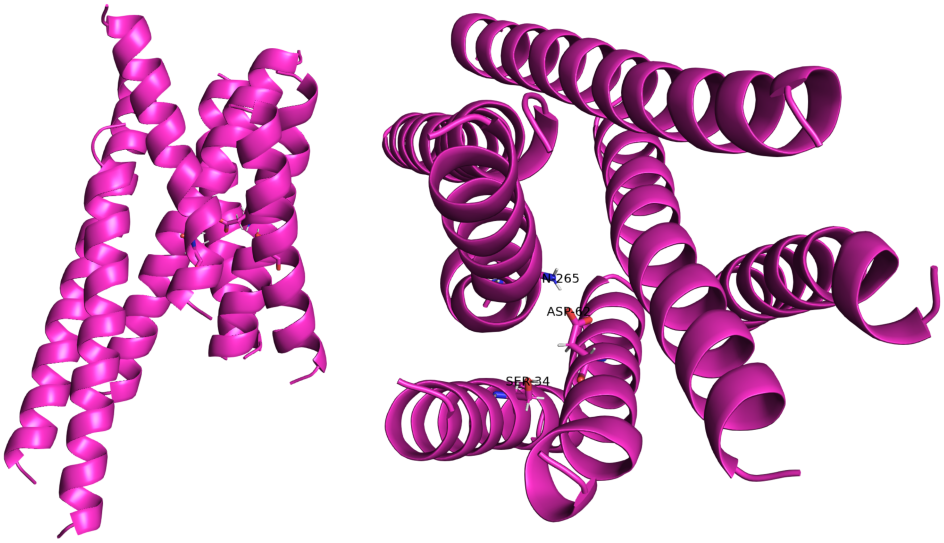


Figure 10.12: 10bBH.bgf has 8 class A contacts and 2 active contacts.

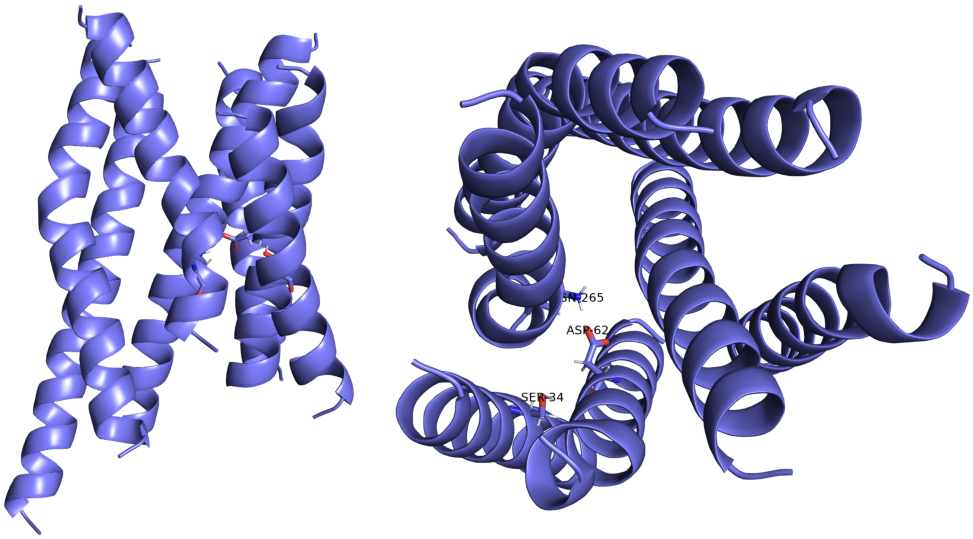


Figure 10.13: 12bBH.bgf has 14 class A contacts and 1 active contact.

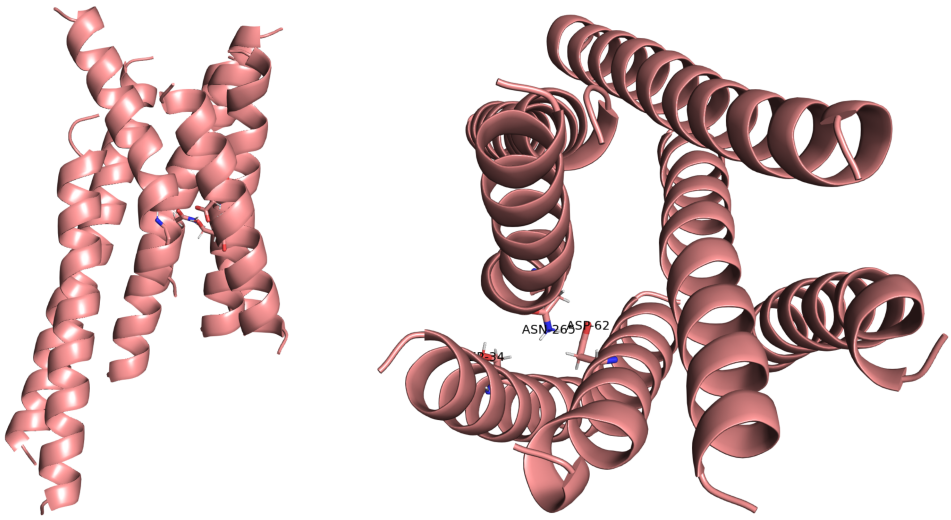


Figure 10.14: 15bBH.bgf has 12 class A contacts and 1 active contact.

## 10.6 Appendix VI: Second Super BiHelix results.

return to main text

1. 1SBiH

   1SBiH structure come from the 15th best BiHelix structure, 15bBH.

   Figures 10.15 and 10.16 show details of the 1SBiH structure. These figures show that the structure 1SBiH bears 2 active and 3 Class A contacts. Panel A of 10.15 shows an SB network between helices 2-3, 3-4, and 2-4, meanwhile, panel B displays the 1.50-2.50-7.49 HBs network [79]. In Figure 10.16 we show a network structured by a series of S-Bs and HBs, that stabilize the 5-6 helices package, in the intracellular regions.

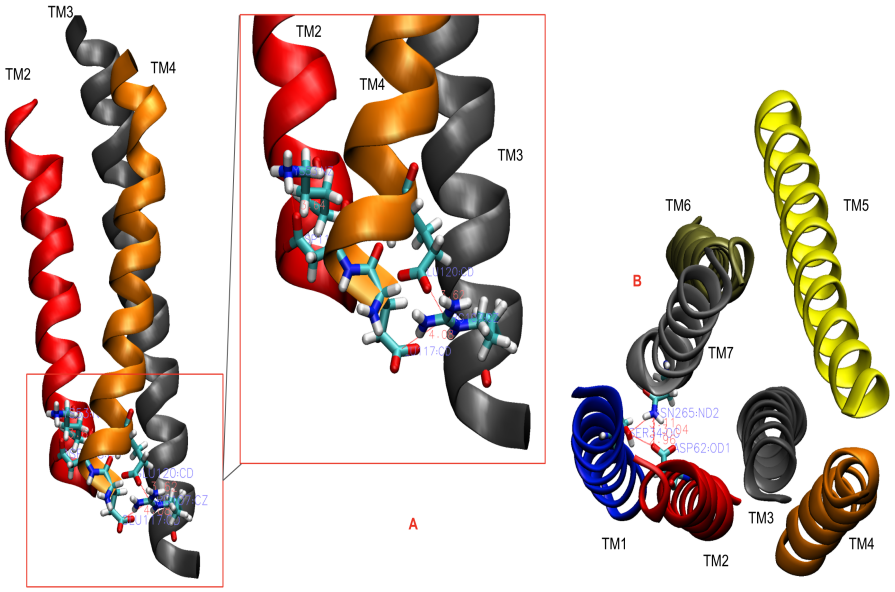


Figure 10.15: SBs and HBs in 1SBiH structure. Panel A) display the SB network K53(2.41)-D118(4.40), E120(4.42)-R107(3.52), and E117(4.39)-R107(3.52). Panel B) shows the conserved S34(1.50)-D62(2.50)-N(7.49) HB network, similarly found in most Class A GPCRs

In Figure 10.16, we show in Panel A) a network of interactions structured by a series of SBs and HBs in the intracellular region of helices 5-6. Panel B) shows the network of SB structured by D204(6.23)-R193(5.73), D207(6.26)-R190(5.70), and E211(6.30)-R187(5.67). Panel C) shows one interhelical HB between R187(5.67)-N208(6.27).

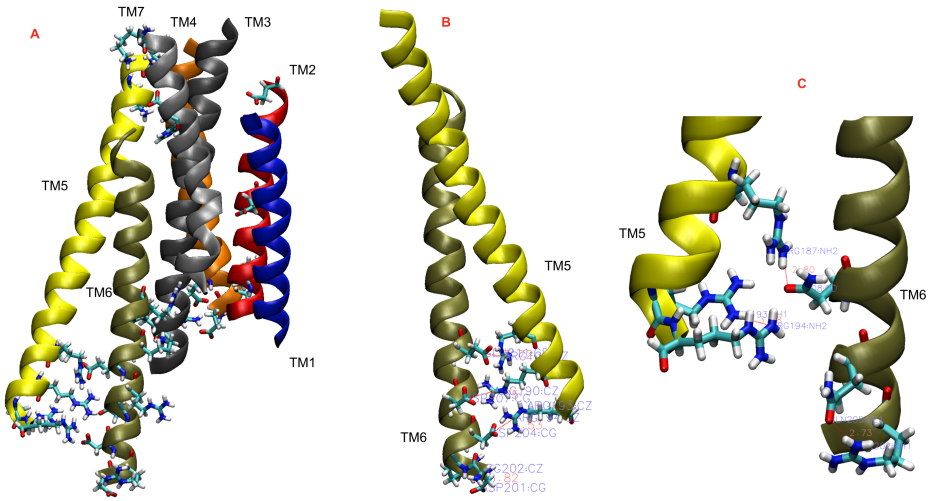


Figure 10.16: SBs and HBs in 1SBiH structure. Panel A) displays charged residues, especially near intracellular regions of some helices in the 7TM bundle. Panel B) shows the network structured by D204(6.23)-R193(5.73), D207(6.26)-R190(5.70), E211(6.30)-R190(5.70), and E211(6.30)-R187(5.67) SBs. Panel C) shows the interhelical R187(5.67)-N208(6.27) HB.

Table 10.4 contains the complete information about the SBs and the HBs networks showed in Figures 10.15 and 10.16

| Interaction | SBs | HBs | Distance/A |
|---|---|---|---|
| H5 - H6 | R187(5.67) - E211(6.30) | | 3.72 |
| | R190(5.70) - E211(6.30) | | 4.99 |
| | R190(5.70) - D207(6.26) | | 3.75 |
| | R194(5.74) - D204(6.23) | | 3.53 |
| | | R187(5.67) - N208(6.27) | 2.80 |
| Intra H5 | | R193(5.73) - R194(5.74) | 3.08 |
| Intra H6 | | Q205(6.24) - R202(6.21) | 2.73 |
| | R202(6.21) - D201(6.20) | | 3.82 |
| H2 - H3 - H4 | K53(2.41) - D118(4.40) | | 3.64 |
| | R107(3.52) - E120(4.42) | | 3.62 |
| | R107(3.52) - E117(4.39) | | 4.08 |
| 1.50-2.50-7.49 | | S34(1.50) - D62(2.50) | 2.96 |
| | | S34(1.50) - N265(7.49) | 3.11 |
| | | D62(2.50) - N265(7.49) | 3.04 |

Table 10.4: Distances for Salt Bridges, Hydrogen bonds, and conserved contacts found in both Animal-GPCRs and 1SBiH of GCR1.

2. 8SBiH

   8SBiH structure come from the 15th best BiHelix structure, 15bBH.

   Figures 10.17 and 10.18 show a series of SBs and HBs interactions in structure 8SBiH. Figure 10.17 shows a SB network between E120(4.42)-R107(3.52), E117(4.39)-R107(3.52), and K53(2.41)-D118(4.40). Figure 10.17, panel B) shows that although residues that make part of the conserved class A 1-2-7 HB network, have a good orientation to interact with each other, only one HB was found D62(2.50)-N265(7.49) 2.95 A. Another interaction between N265(7.49)-S34(1.50), has a distance of 3.30 A reflecting a weak HB. By last, interaction S34(1.50)-D62(2.50) has a distance of 6.59 A could be a water-mediated HB.

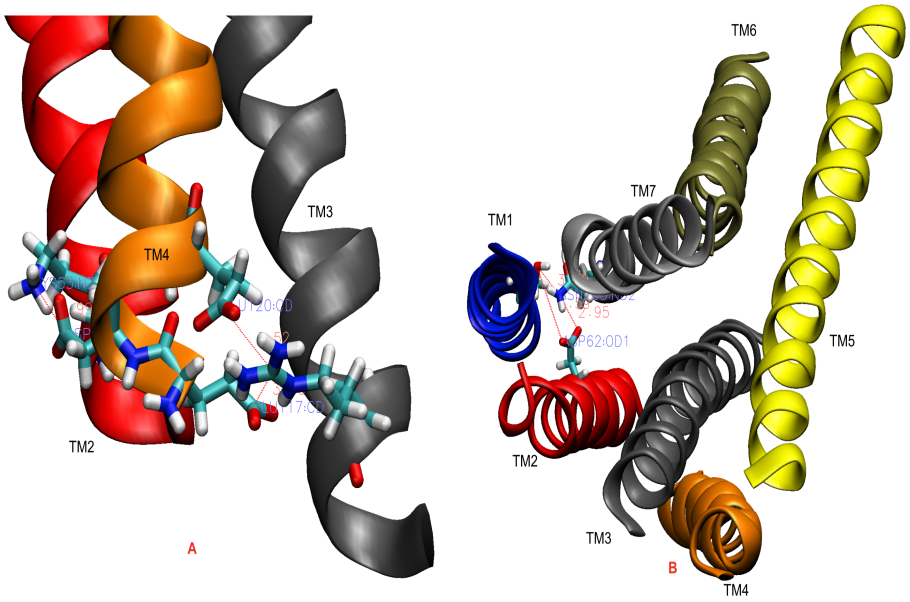


Figure 10.17: Structure 8SBiH shows in panel A) a three SBs network between helices 2-4 and 3-4. TM4 holding the three S_s, one of them with TM2, D118(4.40)-K53(2.41), and two with TM3, D120(4.42)-R107(3.52) and E117(4.39)-R107(3.52). Panel B) shows residues S34(1.50), D62(2.50) and N265(749) with similar location to the conserved class A 1.50-2.50-7.49 HB network, but with not good distances.

Figure 10.18, panel A) shows networks of SBs and HBs in the intracellular region of helices in structure 8SBiH. Panel B) shows the SB network structured by R187(5.67)-E211(6.30), R190(5.70)-E211(6.30), R190(5.70)-D207(6.26), and R194(5.74)-D204(6.23). Additionally, in panel B) is shown an intrahelical SB, R202(6.21)-D201(6.20). Finally, Panel C) shows two HBs, one intra H6 and another interhelical H5-H6, which added to the SB network, can confer high stability to the packing of those helices.

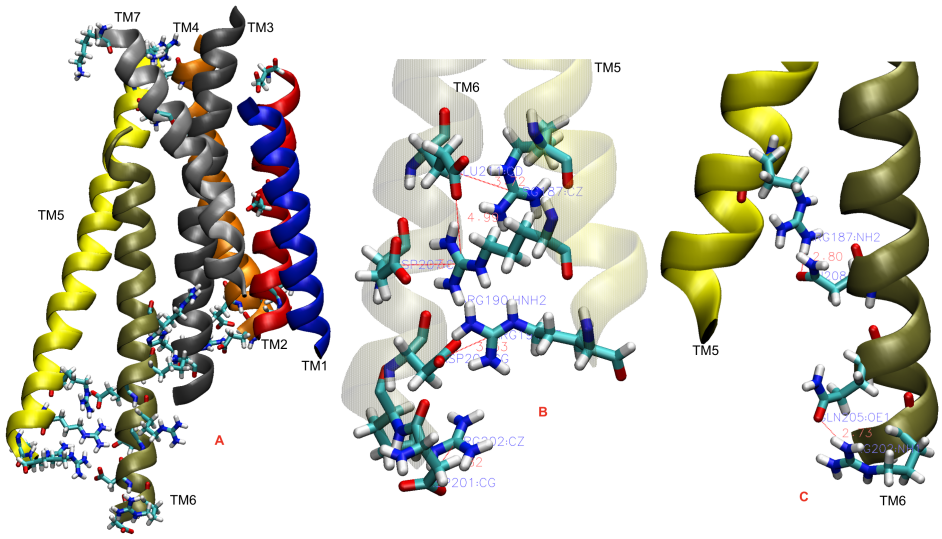


Figure 10.18: 8SBiH shows in A) the 7TM helices bundle with a series of SB and HB networks. Panel B) shows 4 interhelical H5-H6 SBs, R187(5.67)-E211(6.30), R190(5.70)-E211(6.30), R190(5.70)-D207(6.26), and R194(5.74)-D204(6.23), and one intrahelical H6. Panel C) we show a network of 2 Hydrogen bonds, R187(5.67)-N208(6.27) and Q205(6.24)-R202(6.21).

Table 10.5 contains the information about the SBs and HBs networks shown in Figures 12 and 13.

| Interaction | SBs | HBs | Distance/A |
|---|---|---|---|
| H5 - H6 | R187(5.67) - E211(6.30) | | 3.72 |
| | R190(5.70) - E211(6.30) | | 4.99 |
| | R190(5.70) - D207(6.26) | | 3.75 |
| | R194(5.74) - D204(6.23) | | 3.53 |
| | | R187(5.67) - N208(6.27) | 2.80 |
| Intra H6 | | Q205(6.24) - R202(6.21) | 2.73 |
| | R202(6.21) - D201(6.20) | | 4.25 |
| H2 - H3 - H4 | K53(2.41) - D118(4.40) | | 2.99 |
| | R107(3.52) - E120(4.42) | | 3.52 |
| | R107(3.52) - E117(4.39) | | 3.96 |
| 1.50-2.50-7.49 | | S34(1.50) - D62(2.50) | 6.59 |
| | | S34(1.50) - N265(7.49) | 3.30 |
| | | D62(2.50) - N265(7.49) | 2.95 |

Table 10.5: Distances for Salt Bridges, Hydrogen bonds. Only two of three conserved 1-2-7 contacts found in Animal-GPCRs were found in 8SBiH.

3. 15SBiH

   15SBiH structure come from the 6th best BiHelix structure, 6bBH.

   Figures 10.19 and 10.20 show a series of SB and HB interactions in structure 15SBiH. Figure 10.19, Panel A) shows the SB network K53(2.41)-D118(4.40), R107(3.52)-D118(4.40), and R107(3.52)-E117(4.39). Panel B) shows the D62(2.50)-N265(7.49) HB 3.11 A. The other two interactions S34(1.50)-D62(2.50) and N265(7.49)-S34(1.50) which are conserved in animal GPCRs are far away from each other in structure 15SBiH, 7.68 Å and 8.59 Å, respectively, although they have a good location between helices.

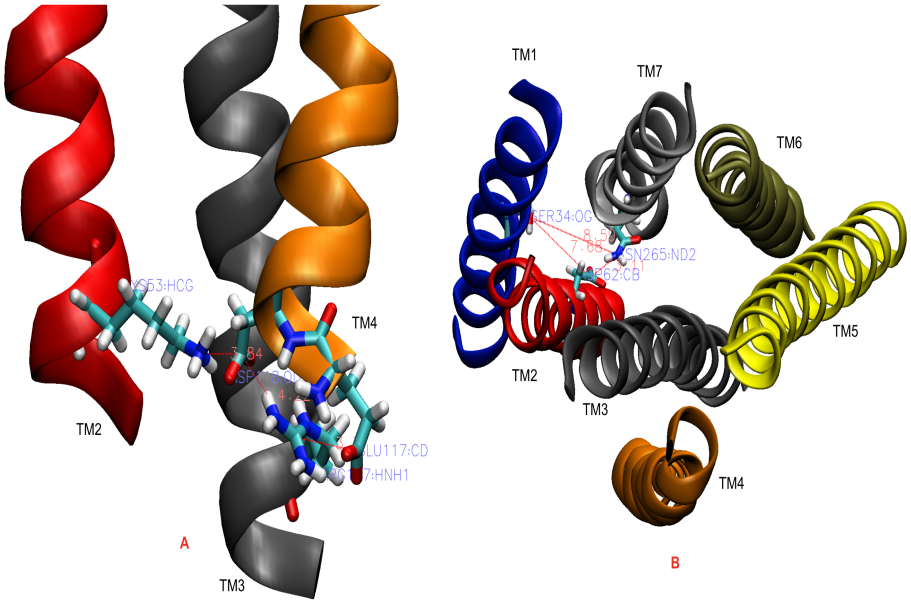


Figure 10.19: Structure 15SBiH. Panel A) a network of SBs between helices with TM4 holding three SBs one of them with TM2, K53(2.41)-D118(4.40), and two with TM3, R107(3.52)-D118(4.40), and R107(3.52)-E117(4.39). Panel B) shows those residues that make part of the conserved 1.50-2.50-7.49 HB in animal GPCRs, but only one HB, D62-N265, is found in 15SbiH structure.

Figure 10.20 shows a network of SBs and HBs, that we propose as stabilizing of the H5-H6 packing: R187(5.67)-E211(6.30), R190(5.70)-E211(6.30), R190(5.70)-D207(6.26) and R194(5.74)-D204(6.23), and N191(5.71)-R209(6.28) and Q184(5.64)-N216(6.35), respectively.

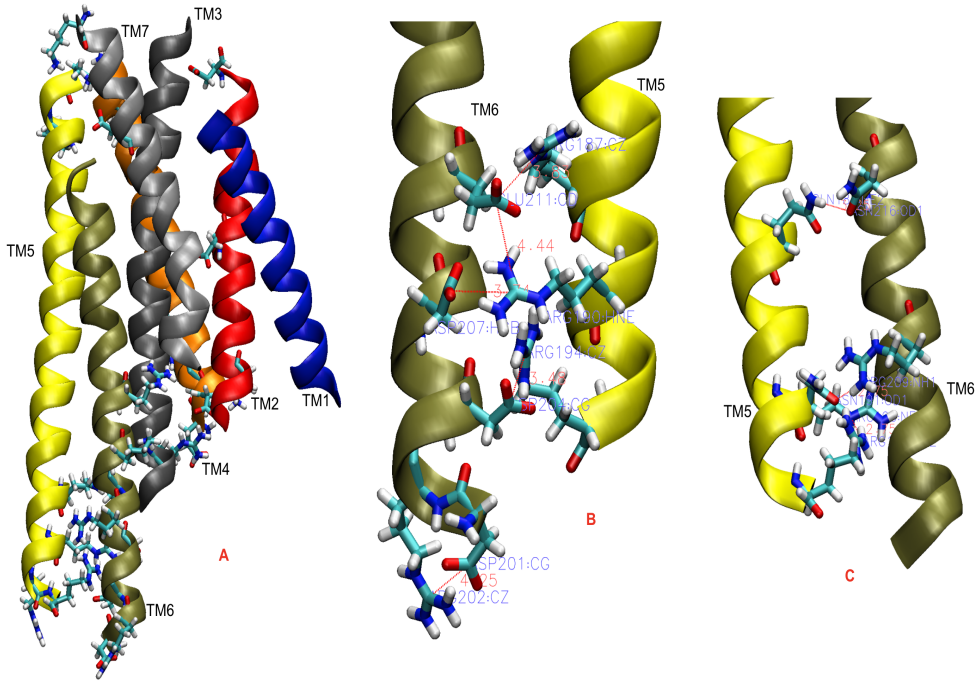


Figure 10.20: 15SBiH structure. Panel A) shows a series of interactions in the 7TM bundle. Panel B) shows four interhelical SBs R187(5.67)-E211(6.30), R190(5.70)-E211(6.30), R190(5.70)-D207(6.26), and R194(5.74)-D204(6.23) and one is intrahelical H6, R202(6.21)-D201(6.20). Panel C) shows 3 Hydrogen bonds, two of them are interhelical HBs, N191(5.71)-R209(6.28) and Q184(5.64)-N216(6.35), which added to the SB network can stabilize the packing of those helices.

Table 10.6 contains the information about the SBs and HBs networks showed in Figures 10.19 and 10.20.

| Interaction | SBs | HBs | Distance/A |
|---|---|---|---|
| H5 - H6 | R187(5.67) - E211(6.30) | | 3.83 |
| | R190(5.70) - E211(6.30) | | 4.44 |
| | R190(5.70) - D207(6.26) | | 3.74 |
| | R194(5.74) - D204(6.23) | | 3.48 |
| | | N191(5.71) - R209(6.28) | 2.75 |
| | | Q184(5.64) - N216(6.35) | 3.25 |
| Intra H5 | | R190(5.70) - R194(5.74) | 275 |
| Intra H6 | R202(6.21) - D201(6.20) | | 4.25 |
| H2 - H3 - H4 | K53(2.41) - D118(4.40) | | 3.84 |
| | R107(3.52) - D118(4.40) | | 4.22 |
| | R107(3.52) - E117(4.39) | | 3.77 |
| 1.50-2.50-7.49 | | S34(1.50) - D62(2.50) | 7.68 |
| | | S34(1.50) - N265(7.49) | 8.59 |
| | | D62(2.50) - N265(7.49) | 3.11 |

Table 10.6: Distances for Salt Bridges, Hydrogen bonds, and conserved contacts found in both Animal-GPCRs and 15SBiH of GCR1. R202-D201 is an intra 6-helical SB

4. 20SBiH

   20SBiH structure come from the 6th best BiHelix structure, 6bBH.

   Figures 10.21 and 10.22 show a series of SB and HB interactions in structure 20SBiH. Figure 10.21, panel A), shows a SB-HB network of 3 SB and 1 HB, K53(2.41)-D118(4.40), R107(3.52)-D118(4.40) and R107(3.52)-E117(4.39), and S61(2.49)-Y126(4.48), respectively. In panel B), we found the D62(2.50)-N265(7.49) HB but not the conserved animal 1-2-7 HB network, with a distance of 3.13 Å, resembling a weak HB between on those two residues.

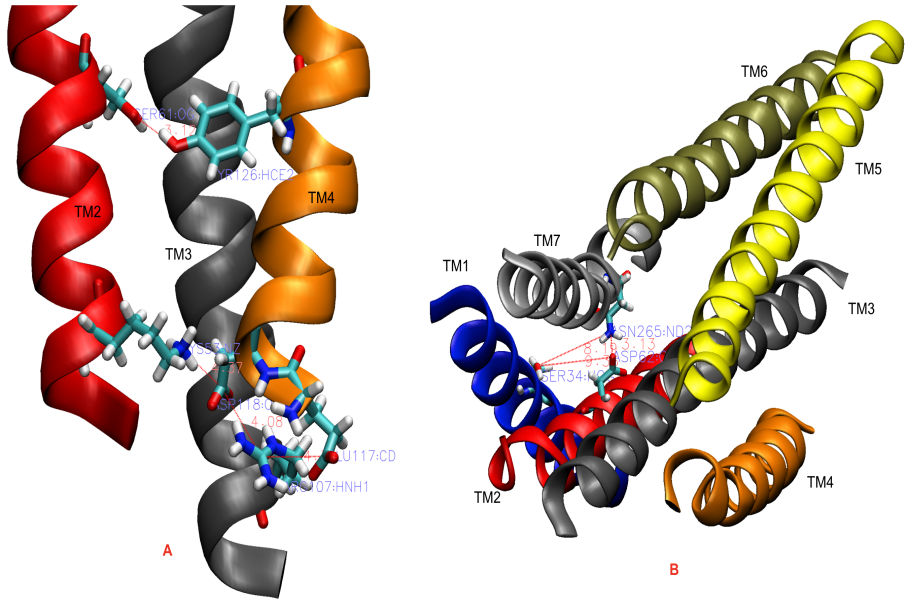


Figure 10.21: Structure 20SBiH. Panel A) shows a network of SBs and HBs between H2-H3 and H3-H4. TM4 holding three SBs, one of them with TM2, K53(2.41)-D118(4.40), two more with TM3, R107(3.52)-D118(4.40) and R107(3.52)-E117(4.39), additionally, TM4 has a HB with TM2 S61(2.49)-Y126(4.48). On the other hand, Panel B) shows the residues that make part of the conserved class A 1.50-2.50-7.49 HB network, although for the 20SBiH only one HB is found, D62(2.50)-N265(7.49).

Figure 10.22, shows a series of HB and SB interactions between R187(5.67)-E211(6.30), R190(5.70)-E211(6.30), R190(5.70)-D207(6.26) and R194(5.74)-D204(6.23), and R194(5.74)-N208(6.27) and N191(5.71)-R209(6.28), panels B) and C) respectively. We propose that those interactions keep the H5-H6 packing, in the intracellular regions.

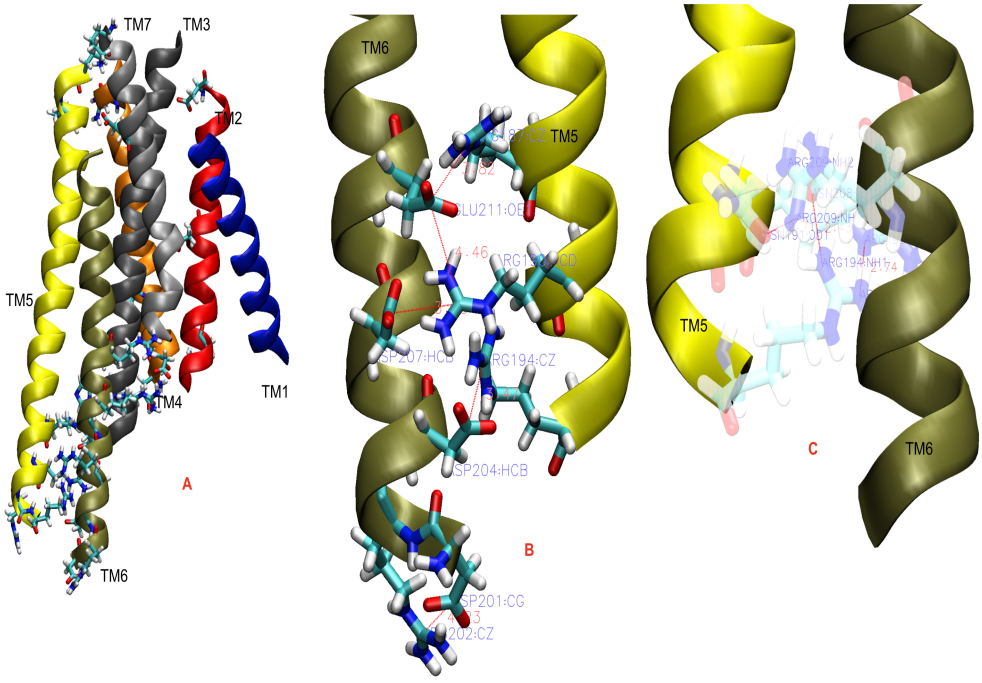


Figure 10.22: 20SBiH shows in A) the 7TM helices bundle. 5 salt bridges are found between helices 5 and 6, four of them are interhelical SBs (R187-E211; R190-E211; R190-D207; R194-D204), and one is helix6-intrahelical (R202-D201), as can be seen in B). On the other hand, in C) we show a network of 4 Hydrogen bonds, (R194-N208; R209-N191; R190-R194; Q184-N216) which added to the SB network, should confer high stability to those helices.

Table 10.7 contains the information about the SBs and HBs networks showed in Figures 10.21 and 10.22

| Interaction | SBs | HBs | Distance/A |
|---|---|---|---|
| H5 - H6 | R187(5.67) - E211(6.30) | | 3.82 |
| | R190(5.70) - E211(6.30) | | 4.46 |
| | R190(5.70) - D207(6.26) | | 3.73 |
| | R194(5.74) - D204(6.23) | | 3.47 |
| | | R194(5.74) - N208(6.27) | 2.43 |
| | | N191(5.71) - R209(6.28) | 2.71 |
| Intra H5 | | R190(5.70) - R194(5.74) | 2.74 |
| Intra H6 | R202(6.21) - D201(6.20) | | 4.23 |
| H2 - H3 - H4 | K53(2.41) - D118(4.40) | | 4.37 |
| | R107(3.52) - D118(4.40) | | 4.08 |
| | R107(3.52) - E117(4.39) | | 4.31 |
| | | S61(2.49) - Y126(4.48) | 3.12 |
| 1.50-2.50-7.49 | | S34(1.50) - D62(2.50) | 9.35 |
| | | S34(1.50) - N265(7.49) | 8.19 |
| | | D62(2.50) - N265(7.49) | 3.13 |

Table 10.7: Distances for salt bridges, hydrogen bonds, and conserved contacts from Animal-GPCRs present in 20SBiH OF GCR1

5. 23SBiH

   23SBiH structure come from the 15th best BiHelix structure, 15bBH.

   Figures 10.23 and 10.24 show a series of SBs and HBs interactions in structure 23SBiH. In Figure 10.23, Panels A) and B) show both SBs and HB making a network that keeps helices 2-3-4 together, with 4 SB and 1 HB, K53(2.41)-D118(4.40), R107(3.52)-E120(4.42), R107(3.52)-E117(4.39) and K53(2.41)-E117(4.39), and T100(3.45)-H124(4.46), respectively. Panel C) shows the residues that make part of class A conserved 1-2-7 HB network but for structure 23SBiH no one good interaction was found. The better interaction D62(2.50) and N265(7.49) has a distance 3.23 Å that resembles a very weak HB.

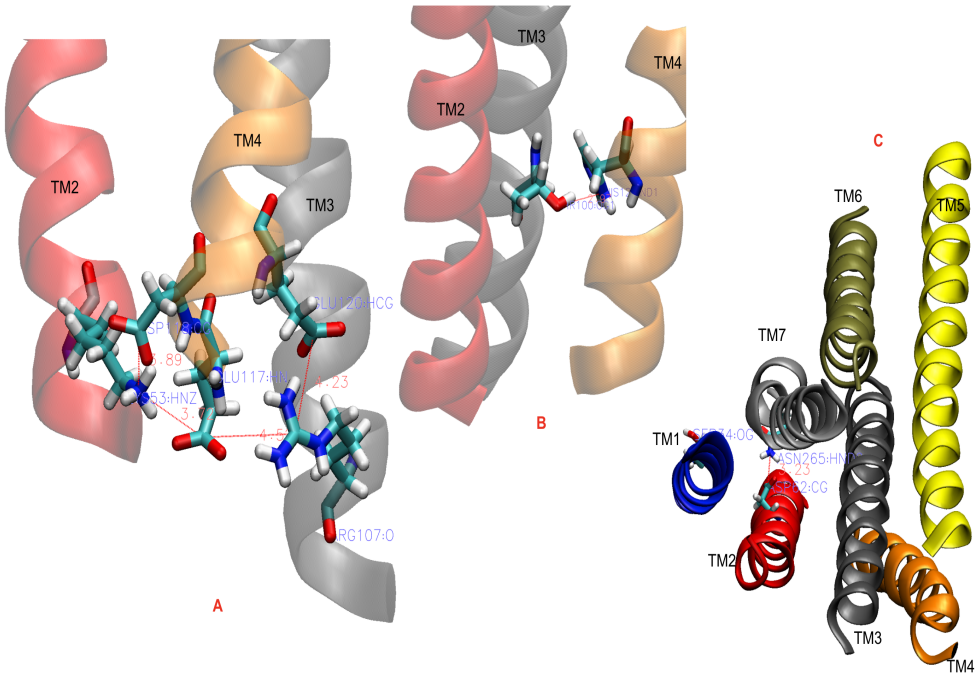


Figure 10.23: Structure 23SBiH. Panel A) shows a network of 4 SBs between helices H2-H4 and H3-H4. TM4 makes part of all four SBs. Two SBs with TM2, K53(2.41)-D118(4.40) and K53(2.41)-E117(4.39. Two more with TM3, R107(3.52)-E117(4.39) and R107(3.52)-E120(4.42). Panel B) shows the HB T100(3.45)-H124(4.46). Panel C) shows residues that make part of the conserved animal GPCR 1.50-2.50-7.49 HB network, but for the 23SBiH structure, we found a weak D62-N265 interaction of 3.23 A.

Figure 10.24, Panel A) shows an SB-HB interaction network, keeping the 5-6 helices packing close together in the intracellular regions. Figure 10.24, Panel B) specifically shows a SB network between R187(5.67)-E211(6.30), R190(5.70)-E211(6.30), R190(5.70)-D207(6.26) and R194(5.74)-D204(6.23), and the interhelical R187(5.67)-N208(6.27) HB in Panel C).

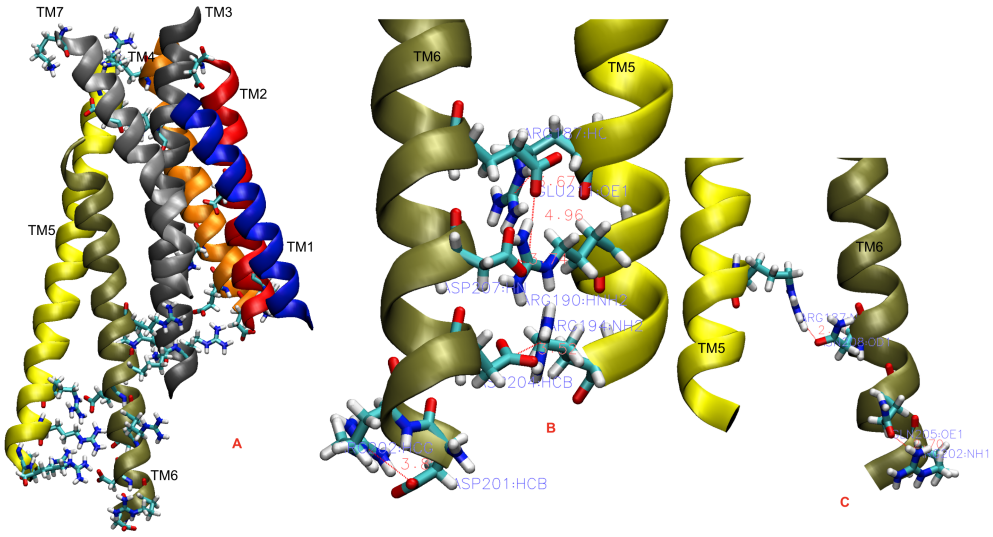


Figure 10.24: 23SBiH structure shows in A) the 7TM helices bundle. 5 salt bridges are found between helices 5 and 6, four of them are interhelical SBs, and one is helix6-intrahelical, as can be seen in B). By last, in C) we show 2 Hydrogen bonds, which added to the SB network, hold the intracellular region of those helices.

Table 10.8 contains the information about the SBs and HBs networks showed in Figures 10.23 and 10.24

| Interaction | SBs | HBs | Distance/A |
|---|---|---|---|
| H5 - H6 | R187(5.67) - E211(6.30) | | 3.67 |
| | R190(5.70) - E211(6.30) | | 4.96 |
| | R190(5.70) - D207(6.26) | | 3.74 |
| | R194(5.74) - D204(6.23) | | 3.52 |
| | | R187(5.67) - N208(6.27) | 2.81 |
| Intra H6 | | Q205(6.24) - R202(6.21) | 2.70 |
| | R202(6.21) - D201(6.20) | | 3.80 |
| H2 - H3 - H4 | K53(2.41) - D118(4.40) | | 3.89 |
| | R107(3.52) - E120(4.42) | | 4.23 |
| | R107(3.52) - E117(4.39) | | 4.52 |
| | K53(2.41) - E117(4.39) | | 3.72 |
| | | T100(3.45) - H124(4.46) | 2.97 |
| 1.50-2.50-7.49 | | S34(1.50) - D62(2.50) | 9.40 |
| | | S34(1.50) - N265(7.49) | 8.30 |
| | | D62(2.50) - N265(7.49) | 3.23 |

Table 10.8: Distances for salt bridges, hydrogen bonds, and conserved contacts from Animal-GPCRs present in 23SBiH GCR1

## 10.7 Appendix VII: Structural analysis of the equilibrated GCR1 candidate structures.

return to main text

### 10.7.1 Structural analysis of the equilibrated GCR1 candidate structure, 8SBiH

Figure 10.25 shows the predicted GCR1 structure, 8SBiH, equilibrated at 310 K. Panel A) displays the lateral view with the numbering of each helix. Panel B) shows the same structure embedded in a POPC lipid bilayer. Figure 10.26 displays the RMSD, where we can see that structure 8SBiH converges during the last 2ns of a total of 80ns of MD simulation, under NPT, 310K, and 1 atm conditions.

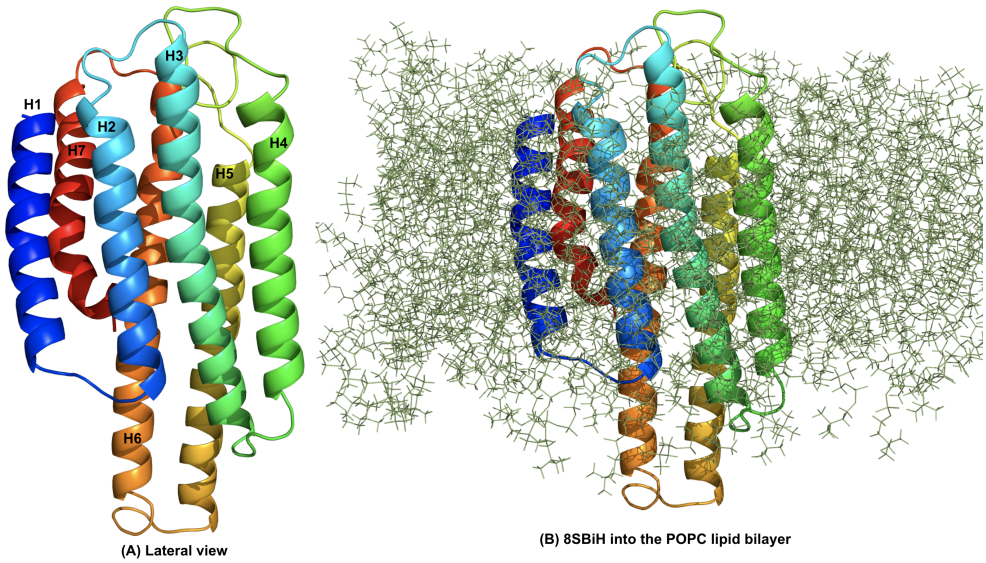


Figure 10.25: Equilibrated GCR1 structure, 8SBiH. Panel (A) shows lateral view of the protein with the helical numbering. Panel (B) shows protein embedded in a POPC lipid bilayer.

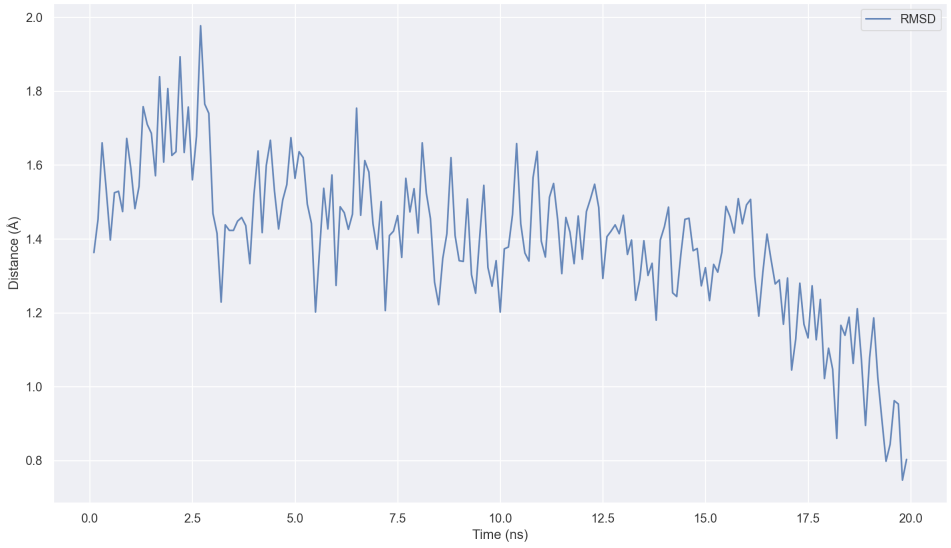


Figure 10.26: Equilibrated GCR1 structure, 8SBiH. Backbone RMSD for the last 20 ns out of 80 ns for the predicted GCR1 structure, 8SBiH. Around last 2 ns are fully converged.

Figure **??**, panel A) shows a SB netwerok structured by E117(4.39)-R107(3.52) and E120(4.42)-R107(3.52). Additionally, we observe three HBs, the first one in panel B) between R194(5.74)-N208(6.27). The second one, in panel C) between S34(1.50)-N265(7.49), which is highly conserved in class A GPCRs, and finally, a third one in D) between S61(2.49)-S94(3.39). The one between R194(5.74)-N208(6.27), makes also part of the H5-H6 SB network in panel C) of Figure 10.28, which is structured by R190(5.70)-E211(6.30) and R194(5.74)-E211(6.30). On the other hand, panel A) displays a SB between K111(ICL2)-E117(4.39), panel B) shows a SB between R217(6.36)-E46, and panel D) shows a second network structured by two SBs, D239(ECL3)-K158(ECL2) and E242(ECL3)-K158(ECL2).

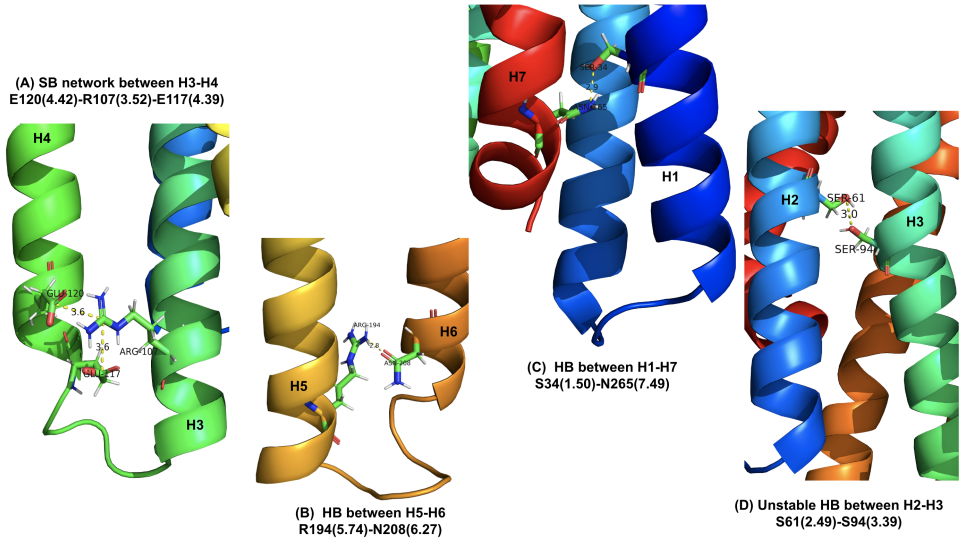


Figure 10.27: Equilibrated GCR1 structure, 8SBiH. Panel (A) shows the molecular view of the interhelical salt bridge network, structured by E117(4.39)-R107(3.52) and E120(4.42)-R107(3.52). Panels (B, C, and D) show three HBs R194(5.74)-N208(6.27), S34(1.50)-N265(7.49), and S61(2.49)-S94(3.39), respectively.

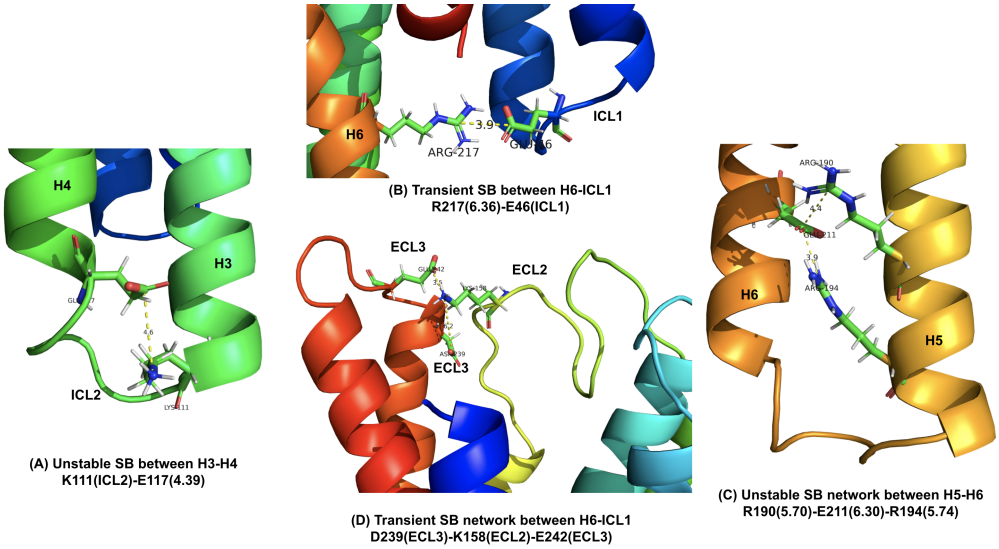


Figure 10.28: Equilibrated GCR1 structure, 8SBiH. Molecular view of unstable and transient interhelical salt bridges. Panel A) displays a SB between K111(ICL2)-E117(4.39). Panel B) shows a SB between R217(6.36)-E46, and panels (C and D) show two networks, the one structured by R190(5.70)-E211(6.30) and R194(5.74)-E211(6.30), and another by D239(ECL3)-K158(ECL2) and E242(ECL3)-K158(ECL2), respectively.

By comparing the equilibrated 8SBiH with its corresponding bundle from SuperBiHelix, we found a series of differences between the interhelical SBs, for instance, SB E211(6.30)-R194(5.74) is new, but E211(6.30)-R187(5.67), D207(6.26)-R190(5.70), D204(6.23)-R194(5.74), and D118(4.40)-K53(2.41) were lost, as can be confirmed in Figures 10.29 and 10.30.

Figures 10.29 and 10.30 show the time evolution, for the last 20 ns out of 80 ns in the MD simulation, for the stable SBs shown in Figure 10.27, and also for those shown in Figure 10.28. Panels (A and B) in Figure 10.30 shows two SBs stable enough during the time simulation, whereby, SBs E211(6.30)-R194(5.74) and E211(6.30)-R190(5.70) are considered as stable for further analysis. On the other hand, panels (C to F) represent transient SBs.

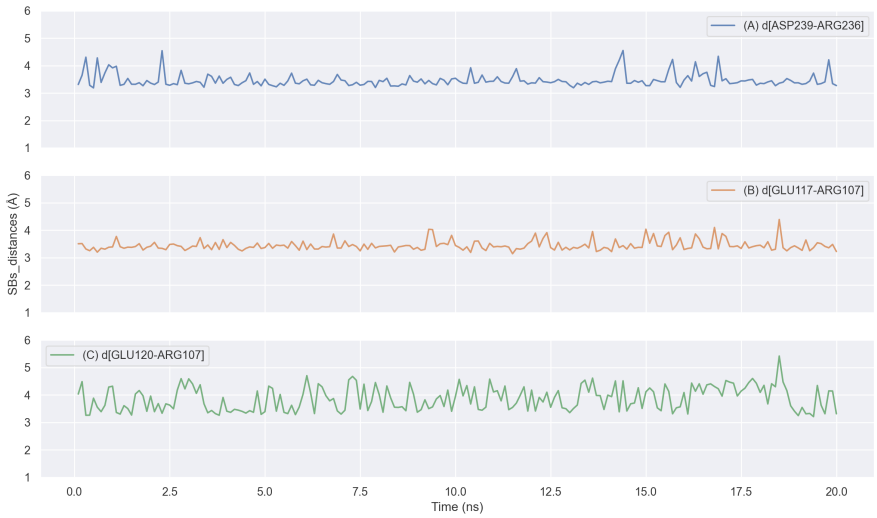


Figure 10.29: Equilibrated GCR1 structure, 8SBiH. Time evolution of distances for D239(ECL3)-R236(ECL3) E117(4.39)-R107(3.52), and E120(4.42)-R107(3.52). We consider that stable SBs are less than 5 Å.


Figure 10.30: Equilibrated GCR1 structure, 8SBiH. Time evolution for the last 20 ns out of 80 ns of transient salt bridges. The salt bridge in (A) E211(6.30)-R194(5.74) after 5.5 ns stabilizes out of unstable, one in (B) E211(6.30)-R190(5.70) is stable most of the simulation time, becoming unstable during near 1 ns, around 15.5 and 16.5 ns, then stabilizes out of unstable, one in (C) E46(ICL1)-R217(6.36) after 15 ns unstabilizes out of stable but stabilizing at the end of the time, however ones in (D) D239(ECL3)-K158(ECL2), (E) D242(ECL3)-K158(ECL2), and (F) E117(4.39)-K111(ICL2) are still marginally stable salt bridges.

We show in Figure 10.31 the time evolution for the three HBs interactions shown in panels (B, C, and D) of Figure 10.27. Besides, Figure 10.31 Panel (A) shows the HB between [R194(5.74)-N208(6.27)] that make part of the SB network described in Figure 10.28 panel (C). Panel (B) in Figure 10.31, shows the time evolution for one of the most conserved HBs in class-A animal-GPCRs, which can be seen that is stable for all the simulation time.

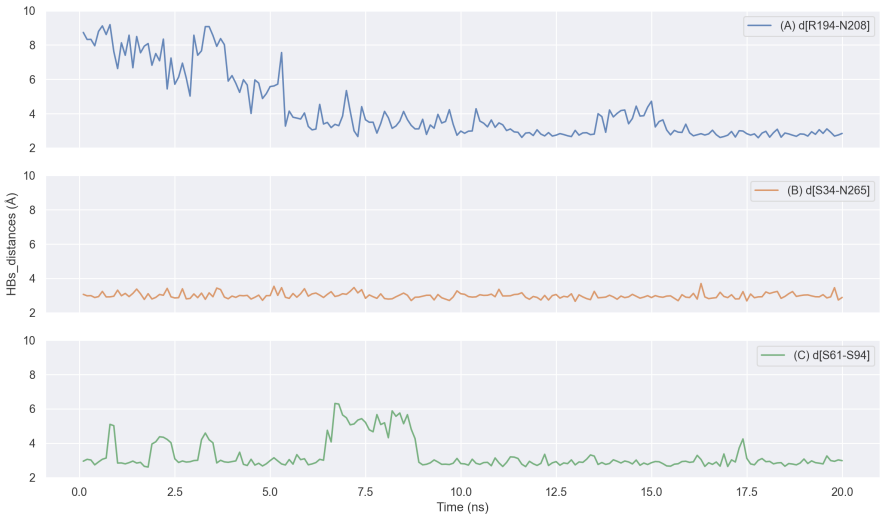


Figure 10.31: Predicted GCR1 structure, 8SBiH. Distance time evolution for the interhelical hydrogen bonds. Panel (A) R194(5.74)-N208(6.27), panel (B) S34(1.50)-N265(7.49), panel (C) S61(2.49)-S94(3.39)

### 10.7.2 Structural analysis of the equilibrated GCR1 candidate structure, 15SBiH

Figure 10.32, panels (A) and (C) show the top and lateral views for the predicted and equilibrated GCR1 structure, 15SBiH, with helical numbering and coloring codes to identify each helix on the structure. Finally, panel (B) shows the GCR1 structure, equlibrated at 310 K, embedded into a POPC lipid bilayer. In addition, Figure 10.33 shows that structure 15SBiH converged during the last part of the 25 ns of a total of 60 ns of MD simulation. Although, we can see that the RMSD value decreases, it returns to the average value obtained during the last 10 ns of the simulation.

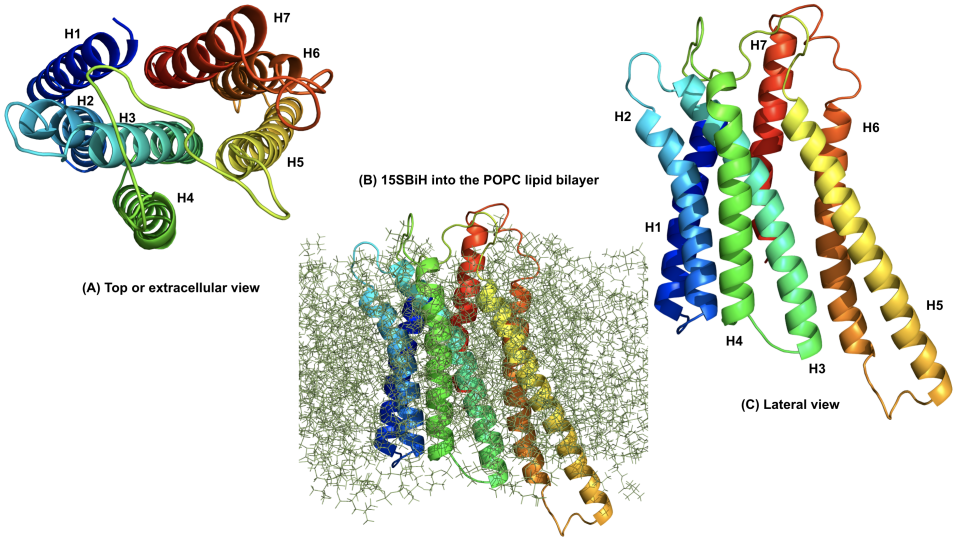


Figure 10.32: Equilibrated GCR1 structure, 15SBiH, embedded in a POPC lipid bilayer. Panels (A) and (C) show a top and a lateral views of the protein with the helical numbering. Panel (B) shows the equilibrated protein embedded in a POPC lipid bilayer.

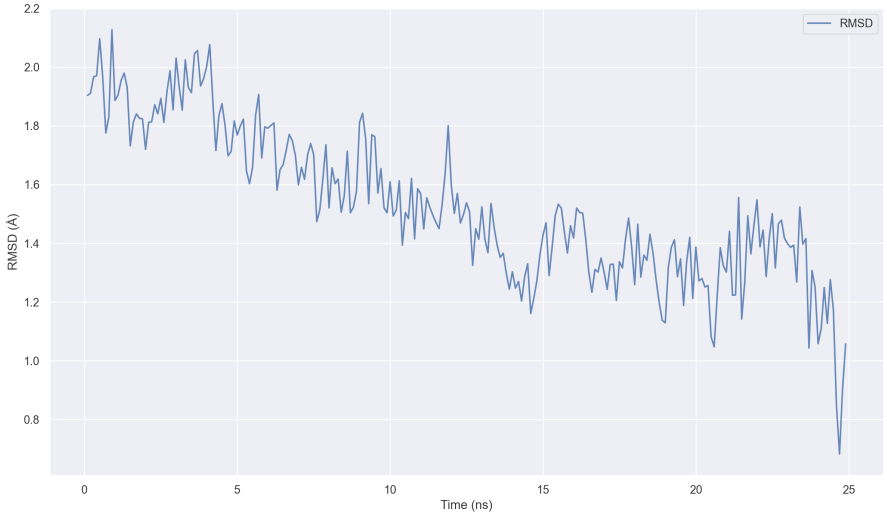


Figure 10.33: Equilibrated GCR1 structure, 15SBiH. Backbone RMSD for the last 25 ns of 60 ns of the MD simulation. Around last 10 ns are fully converged.

Figure 10.34 shows two types of salt bridges, stable and transient. Panel (A) shows stable SBs between E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), and D204(6.23)-R190(5.70). Panels (B) and (C) show transient SBs between E118(4.40)-R107(3.52) and E117(4.39)-R107(3.52), and D254(7.38) and R236(ECL3), respectively. The status of stable or transient can be confirmed by Figures 10.35, and 10.36, respectively. By comparing the equilibrated structures with their corresponding bundle structure from SuperBiHelix, we can see some differences between the interhelical SBs found in both. We gain a new SBs, D204(6.23)-R190(5.70), but we lost three SBs, D204(6.23)-R194(5.74), E207(6.26)-R190(5.70) and K53(2.41)-D118(4.40). These changes may be related to the movement restrictions imposed by the addition of loops.

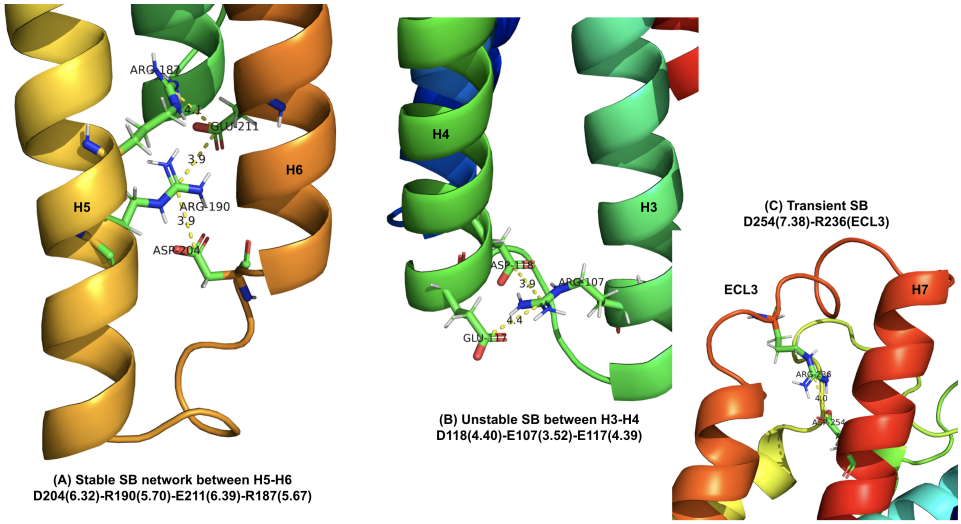


Figure 10.34: Equilibrated GCR1 structure, 15SBiH. Panel (A) shows the molecular view of the interhelical H5-H6 salt bridge network, structured by E211(6.39)-R187(5.67), E211(6.39)-R190(5.70), and D204(6.32)-R190(5.70). Panel (B) shows the unstable SB network between E118(4.40)-R107(3.52) and E117(4.39)-R107(3.52). Panel (C) show a transient SB between D254(7.38) and R236(ECL3).

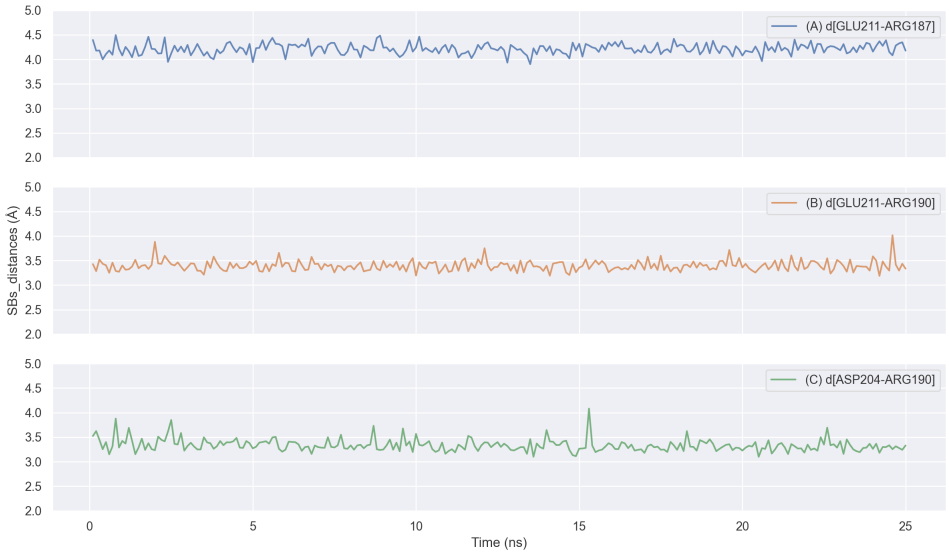


Figure 10.35: Equilibrated GCR1 structure, 15SBiH. Time evolution of the distances for E211(6.39)-R187(5.67), E211(6.39)-R190(5.70), and D204(6.32)-R190(5.70). We consider that stable SBs are less than 5 Å.

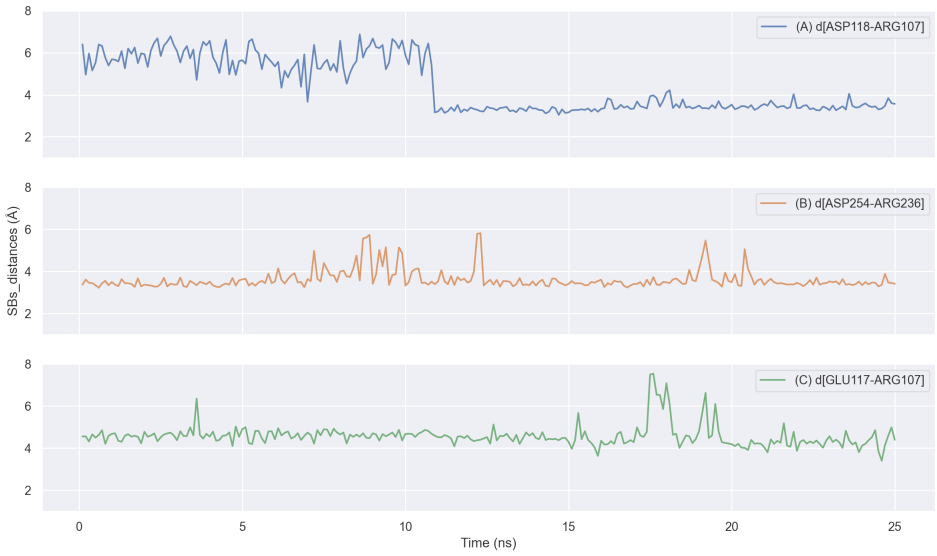


Figure 10.36: Equilibrated GCR1 structure, 15SBiH. Time evolution of transient salt bridges E118(4.40)-R107(3.52), E117(4.39)-R107(3.52) panels (A y C), and D254(7.38)-R236(ECL3) Panel (B).

Finally, we also found three transient HBs shown in Figure 10.37

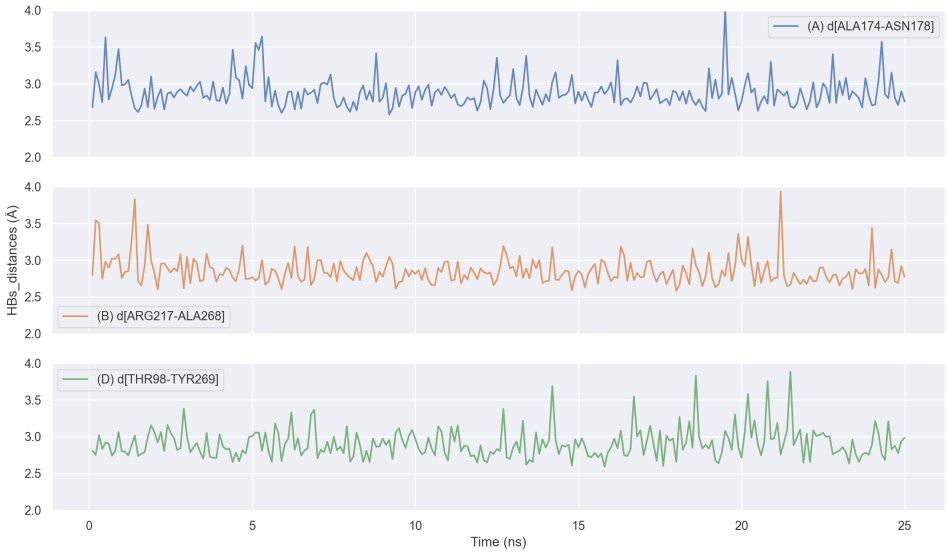


Figure 10.37: Equilibrated GCR1 structure, 15SBiH. Distance time evolution for the interhelical hydrogen bonds. Panel (A) A174(5.54)-N178(5.58), panel (B) R217(6.36)-A268(7.52), and panel (C) T98(3.43)-Y269(7.53)

### 10.7.3 Structural analysis of the equilibrated GCR1 candidate structure, 20SBiH

Figure 10.38, panel (A), shows the predicted and equilibrated GCR1 structure, 20SBiH, embedded in a POPC lipid bilayer. Panel (B), shows the lateral view, with the numbering for each helix. Additionally, Figure 10.39 shows the RMSD, for the last 20 ns out of 60 ns of the MD simulation at 310 K and a pressure of 1 atm. 20SBiH structure is fully equilibrated over the last 4.5 ns.

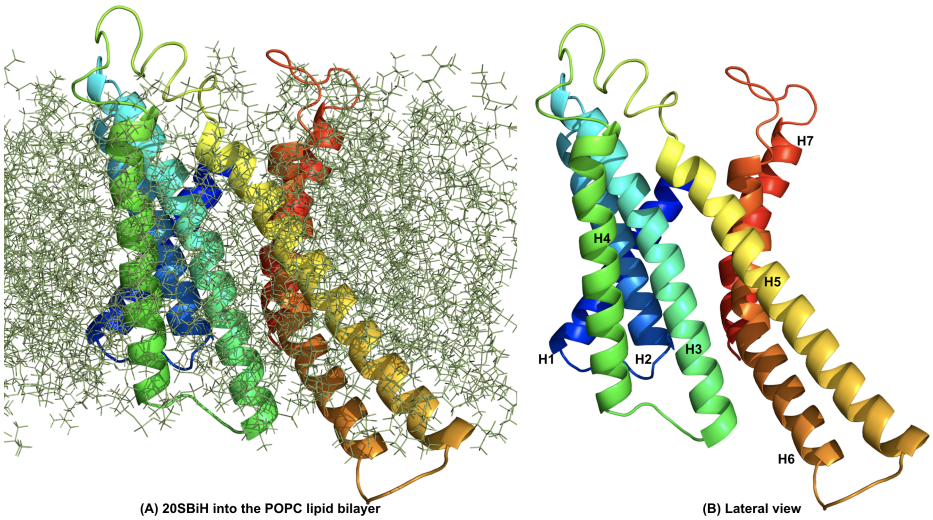


Figure 10.38: Equilibrated GCR1 structure, 20SBiH, embedded in the POPC lipid bilayer is shown in Panel (A). Panel (B) shows the side view of the 20SBiH structure with its respective color code and numbers for each helix. H1 = blue, H2 = cyan, H3 = teals, H4 = green, H5 = yellow, H6 = orange, and H7 = red.

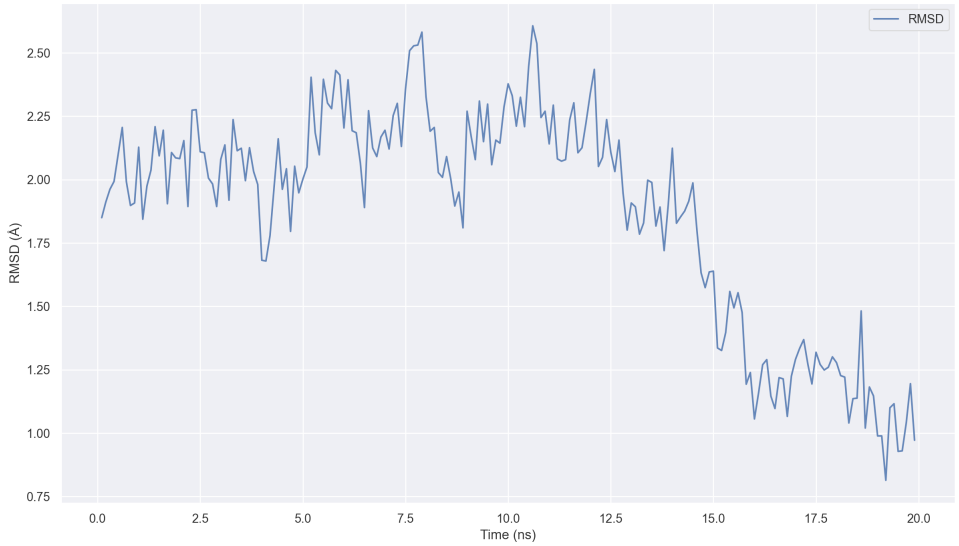


Figure 10.39: Equilibrated GCR1 structure, 20SBiH. Calculated backbone RMSD for the last 20 ns out of 60 ns for the predicted GCR1 structure, 20SBiH. It can see that around the last 4.5 ns are fully converged.

Figures 10.40, show interhelical SBs between helices E118(4.40)-R107(3.52), in panel A) while in panel B) we see a network between D207(6.26)-R190(5.70) SB, D204(6.23)-R190(5.70) and D204(6.23)-R194(5.74). Trajectories shown in Figures 10.41, and 10.42 for the SBs shown in Figures 10.40, enable us to determine their corresponding stabilities. Then, we consider that both shown in Figure 10.41, and one in panel (A) plus one in panel (D) of Figure 10.42 are stable enough during most part of the simulation time. Therefore, we concluded that we found four stable and three transient SBs in the equilibrated structure. These two structures were initially considered transient because they sometimes exceeded 5 Å. By comparing the equilibrated structure 20SBiH with its corresponding bundle structure from SuperBiHelix, we can see that we lost several interhelical structures, such us E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), E117(4.39)-R107(3.52), D118(4.40)-R107(3.52), and D118(4.40)-K53(2.41).

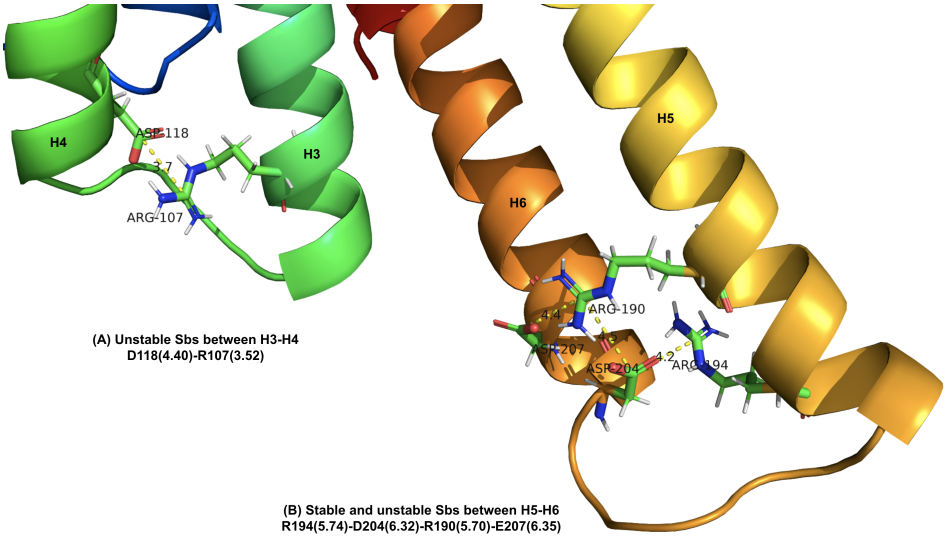


Figure 10.40: Equilibrated GCR1 structure, 20SBiH. Panel (A) shows the molecular view of the unstable interhelical H3-H4 SB, E118(4.40)-R107(3.52). Panel (B) shows an unstable D207(6.26)-R190(5.70) SB, and two stable D204(6.23)-R190(5.70) and D204(6.23)-R194(5.74) SBs, making part of the H5-H6 network.

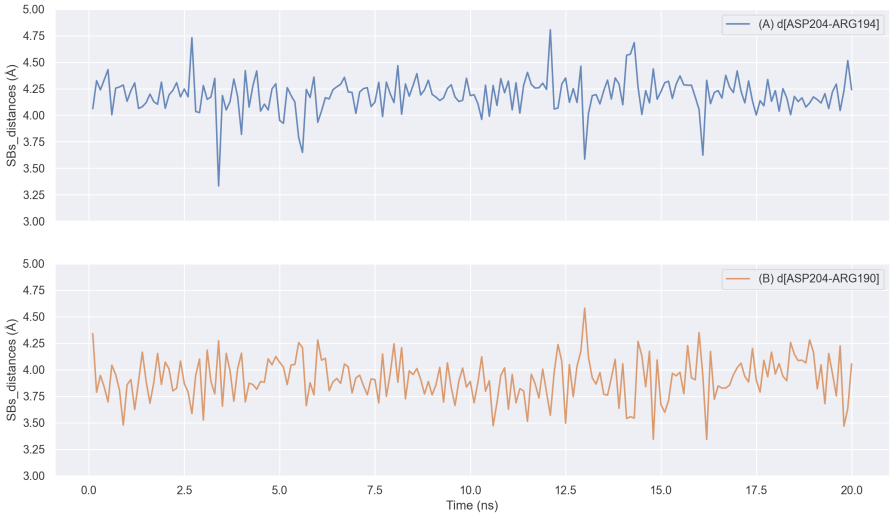


Figure 10.41: Equilibrated GCR1 structure, 20SBiH. Time evolution of the distances for D204(6.32)-R194(5.74) and D204(6.32)-R190(5.70), for the last 20 ns out of 60 ns NPT MD simulation. We consider that stable SBs are less than 5 Å.

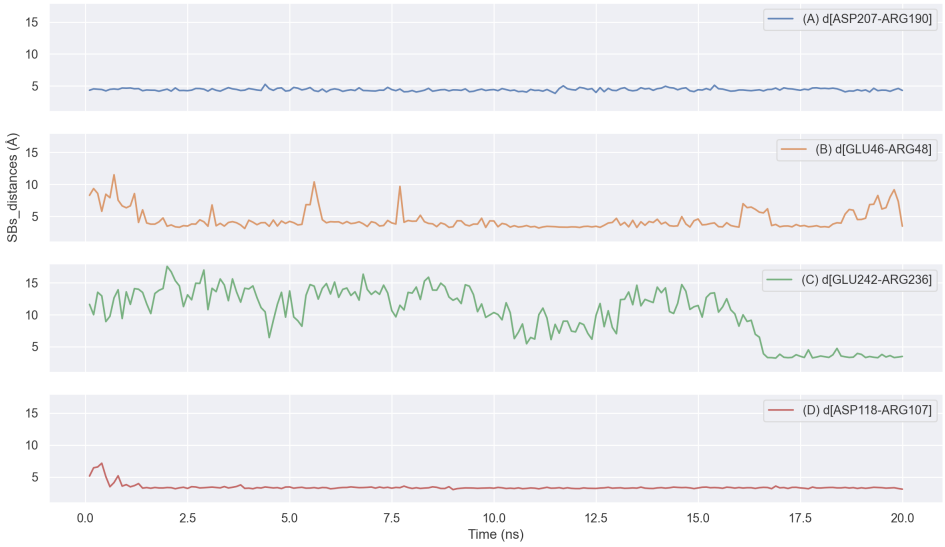


Figure 10.42: Equilibrated GCR1 structure, 20SBiH. Time evolution of transient and unstable salt bridges, for the last 20 ns out of 60 ns NPT MD simulation. Panels (A) D207(6.35)-R190(5.70) and (B) E46(ICL1)-R48(ICL1) show transient Sbs. Although panel (A) seems to be stable during all 20 ns out of 60 ns, it sometimes goes up 5 Å which allows us to think that it shows a transient behavior. The one in (C) E242(ECL3)-R236(ECL3) stabilizes after being unstable for 16.5 ns of dynamics. By last, the one in (D) D118(4.40)-R107(3.52) is unstable at the beginning but is stable during most part of the simulation time.

## 10.8 Appendix VIII: Docking results for the HSP-88 screamed and alanized-GCR1 with ABA.

return to main text

1. b1-ABAno_0.c23860hc

   As can be seen in Figures 10.43 and 10.44, the ligand interacts with residues in the binding site shows that, so, in addition to a series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found a salt bridge and two hydrogen bonds, HSP88-COO, and HIS163-COO and ARG138-CO, respectively.

This figure shows the raw docking results, the distances are measured between charged N and O for the sb, and for HB interactions.

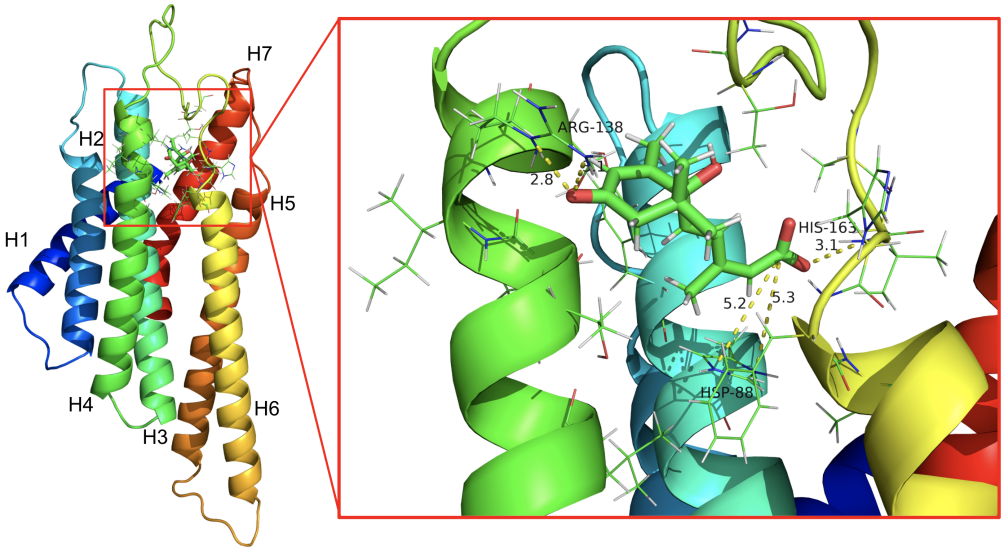


Figure 10.43: GCR1 3D lateral view of the pharmacophore shows distances for the salt bridge and two hydrogen bonds, HSP88-COO, HIS163-COO and ARG138-CO. The distances for the SB are measured between the carbon atom bearing charged O and the N-charged atom, and between heavy atoms for HB interactions.

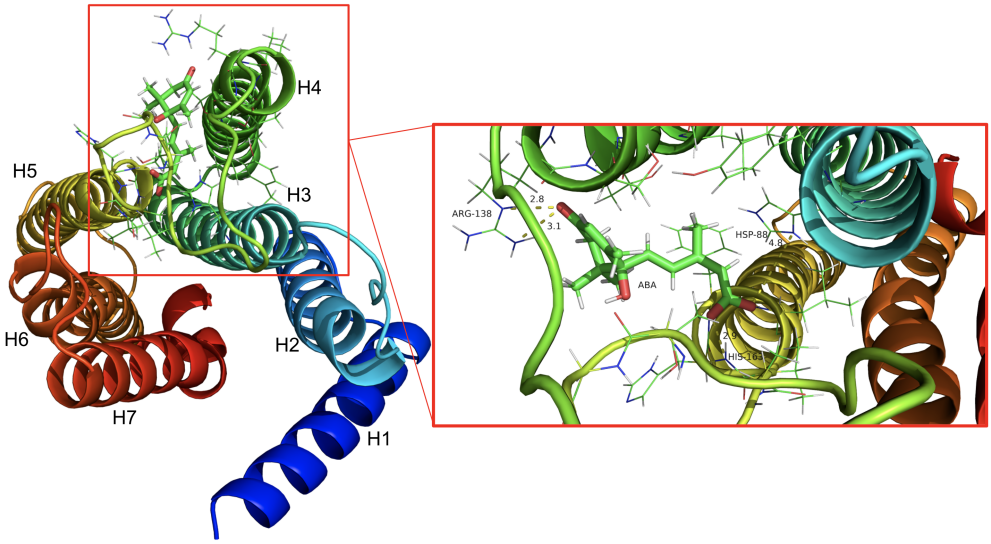


Figure 10.44: GCR1 3D upper view of the pharmacophore shows the location of the ligand in the region located between the helices H3, H4, and H5.

2. b5-ABAno_5.c31387tc

   Although we also found two HBs making part of the interaction between GCR1 and ABA, these are different from those previously found in b1-ABAno_0.c23860hc complex, as can be seen in Figure 10.45.

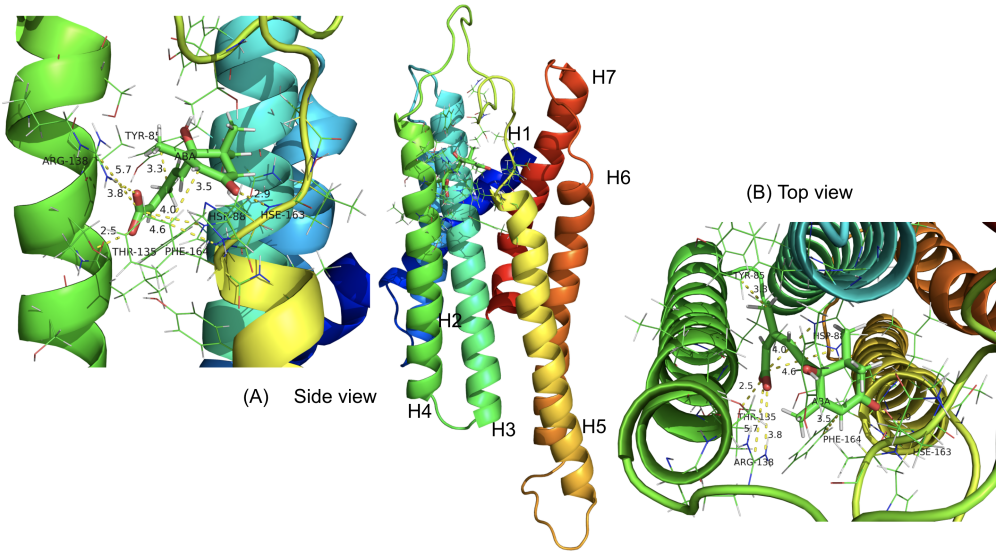


Figure 10.45: GCR1 3D lateral and upper views of the pharmacophore show both, the distances for the salt bridge and for two hydrogen bonds, HSP88-COO, HIS163-CO and PHE164-CO, respectively, also the location of the ligand among helices H3, H4, and H5. The distances for the SB are measured between the carbon atom bearing charged O and the N-charged atom, and between heavy atoms for HB interactions.

3. b1-ABAno_0.c42248thcs

   We found two HBs and two SBs. The SBs are using the same COO group of ABA but two different residues of GCR1, ARG138, and HSP88.

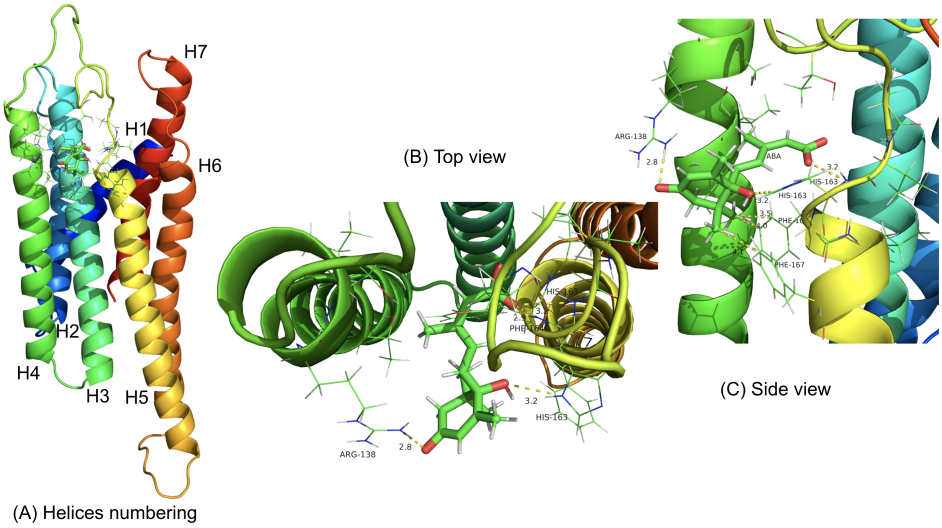


Figure 10.46: GCR1 3D lateral and upper views of the pharmacophore show both, the distances for salt bridges and hydrogen bonds, HSP88-COO and ARG138-COO, HIS163-CO and PHE164-CO, respectively. Also the ligand location among helices H3, H4, and H5. The distances for the SB are measured between the carbon atom bearing charged O and the N-charged atom, and between heavy atoms for HB interactions.

4. b10-ABA.no_0.c24691.h.c

   Similar to the first structure analyzed, this structure has one SB and one HB between the COO group of the ligand with two different residues of GCR1, HSP88 and HIS163, respectively. The other HB is between the CO group of ABA with ARG138.

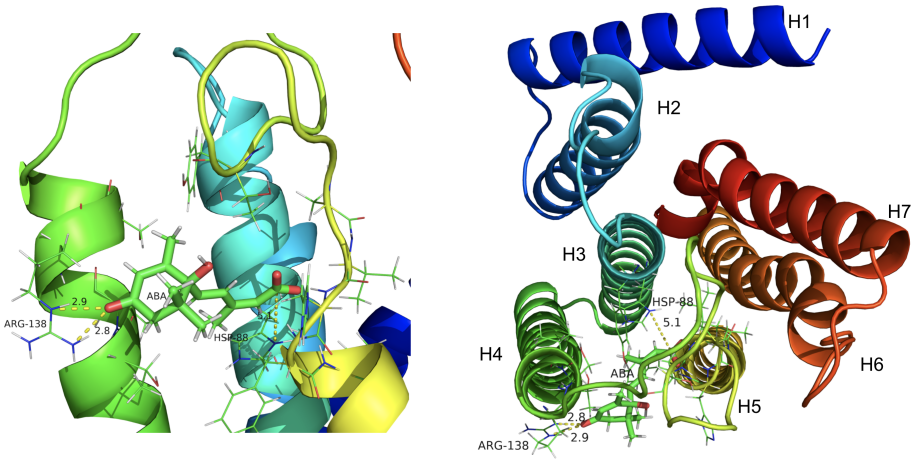


Figure 10.47: GCR1 3D lateral and upper views of the pharmacophore show distances for the salt bridge and two hydrogen bonds, HSP88-COO, HIS163-COO, and ARG138-CO. the distances for the SB are measured between the carbon atom bearing charged O and the N-charged atom, and between heavy atoms for HB interactions.

5. b6-ABA.no_0.c306th

   Different from the previous ones, this structure does not have any salt bridge as part of the pharmacophore, it has three HBs interactions, two of which are with the COO group on ABA and the other with the CO group. The participating residues are always the same, HIS163, PHE164, and ARG138.

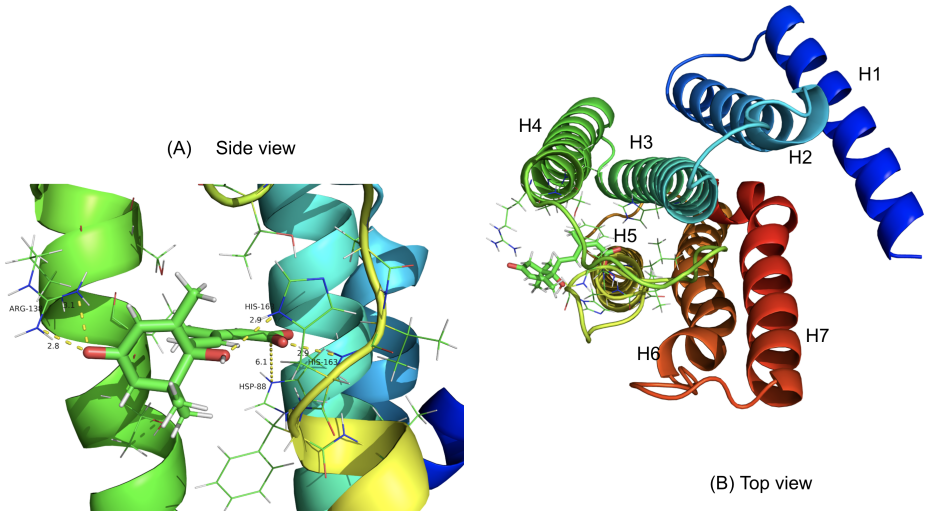


Figure 10.48: GCR1 3D lateral and upper views of the pharmacophore show distances for three hydrogen bonds, HIS163-COO, PHE164-COO, and ARG138-COO, respectively, also the ligand location among helices H3, H4, and H5. The distances are measured between heavy atoms for HB interactions.

6. b9-ABA.no_0.c24494hc

   Again, we found a structure that has one SB and one HB between the COO group of the ligand with two different residues of GCR1, HSP88, and HIS163, respectively. The other HB is between the CO group of ABA with ARG138.

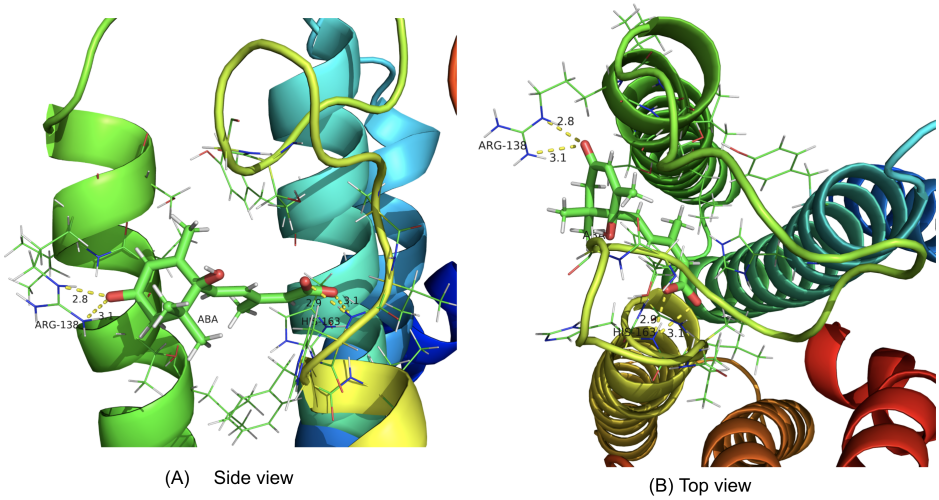


Figure 10.49: GCR1 3D lateral and upper views of the pharmacophore show distances for the salt bridge and two hydrogen bonds, HSP88-COO, HIS163-COO, and ARG138-CO. the distances for the SB are measured between the carbon atom bearing charged O and the N-charged atom, and between heavy atoms for HB interactions.

7. b3-ABA.no_0c40267thc

   In this case, we also found three HBs added to the SB between HSP88 and the COO group of ABA, but the three HBs, although are with the same residues HIS163 and PHE164, at this time the HBs are with the COO group of ABA. The last HB is between ARG138 and the CO group of ABA.

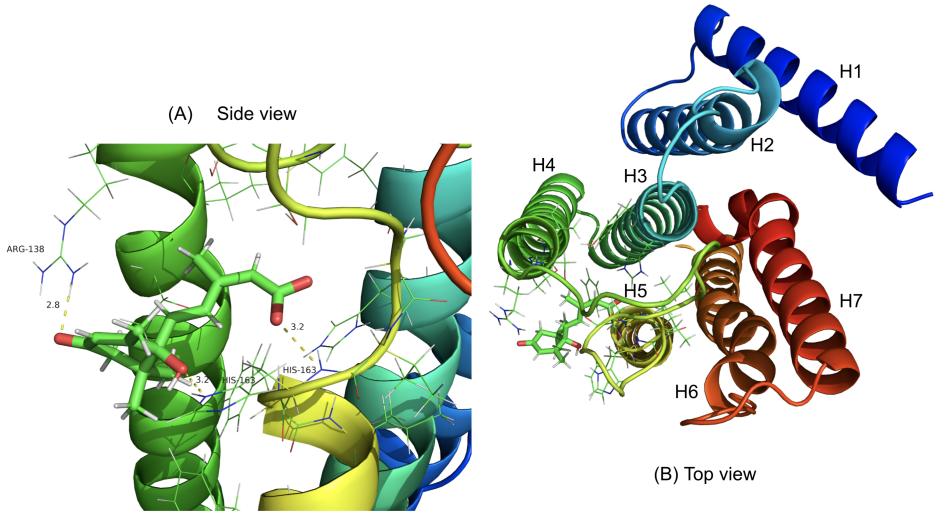


Figure 10.50: GCR1 3D lateral and upper views of the pharmacophore show both, the distances for the one salt bridge and for three hydrogen bonds, HSP88-COO, HIS163-COO, PHE164-COO, and ARG138-CO, respectively, besides, the ligand location among helices H3, H4, and H5 of GCR1. The distances for the SB are measured between the carbon atom bearing charged O and the N-charged atom, and between heavy atoms for HB interactions.

8. b4-ABA.no_0c24315hc

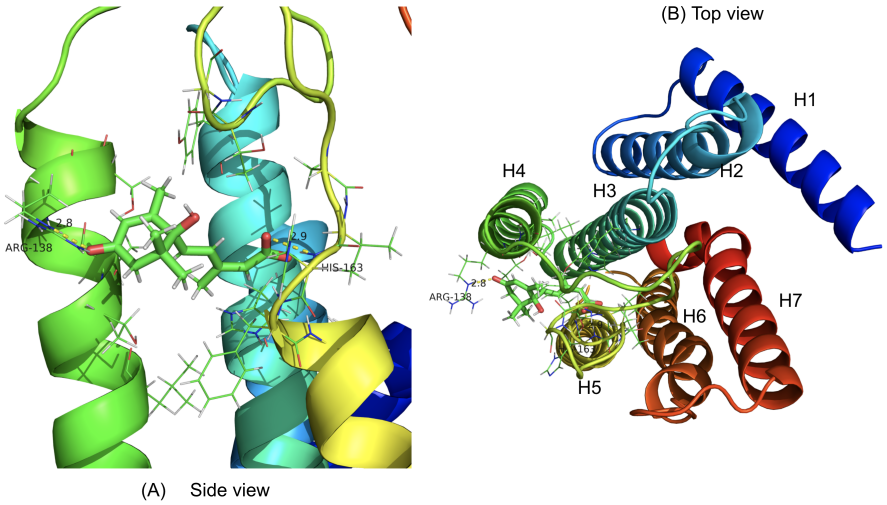


Figure 10.51: GCR1 3D lateral and upper views of the pharmacophore show distances for the salt bridge and two hydrogen bonds, HSP88-COO, HIS163-COO, and ARG138-CO, respectively. the distances for the SB are measured between the carbon atom bearing charged O and the N-charged atom, and between heavy atoms for HB interactions.

9. b9-ABA.no_5c20724hc

   Although this structure also has one SB and two HBs, like previous ones, one of the HB uses a different residue of GCR1 to interact with the charged O of the COO group of ABA, TYR168-COO.

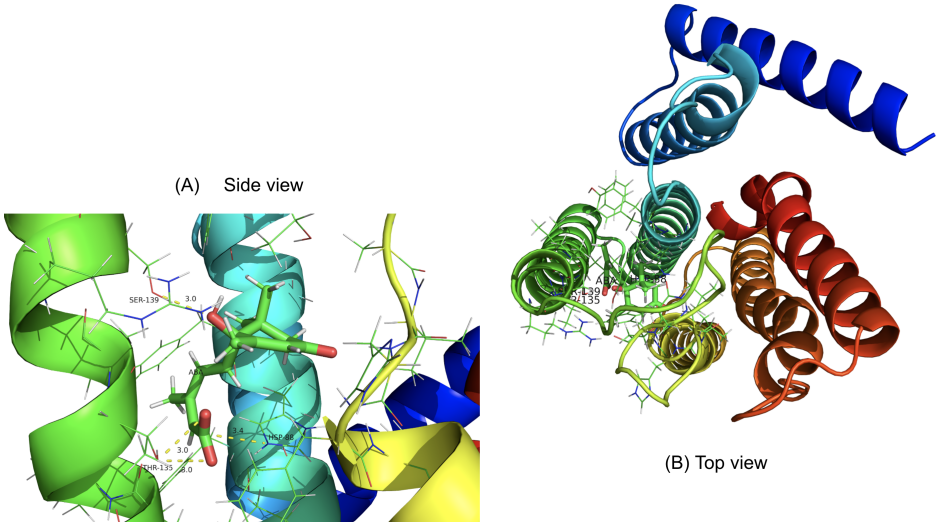


Figure 10.52: GCR1 3D lateral and upper views of the pharmacophore show distances for the salt bridge and two hydrogen bonds, HSP88-COO, HIS163-CO, and TYR168-COO, respectively. the distances for the SB are measured between the carbon atom bearing the charged O and the N-charged atom, and between heavy atoms for HB interactions.

10. b2-ABA.no_5c18183ftc

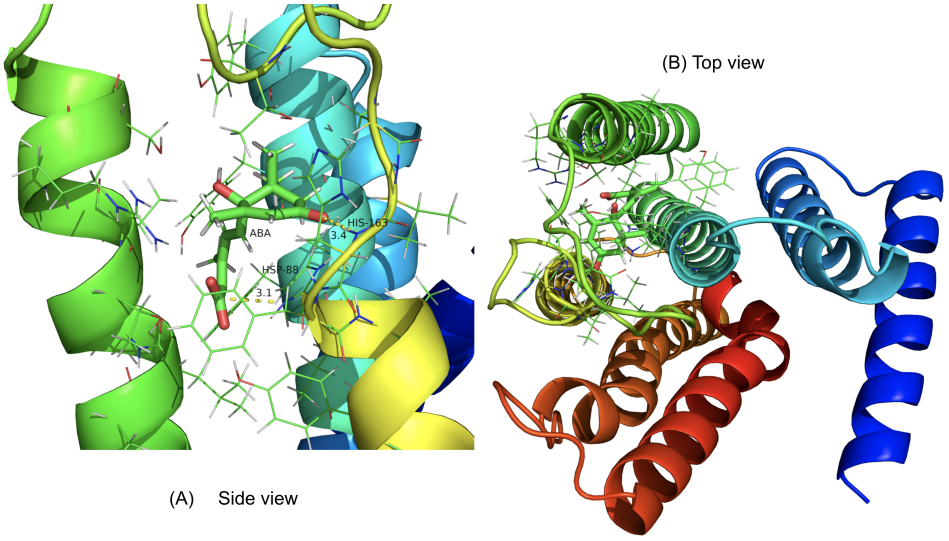


Figure 10.53: GCR1 3D lateral and upper views of the pharmacophore show distances for one salt bridge and one hydrogen bond. The distances for the SB are measured between the carbon atom bearing the charged O and the N-charged atom, and between heavy atoms for HB interactions.

1. b1-ABAno_0.c23860hc

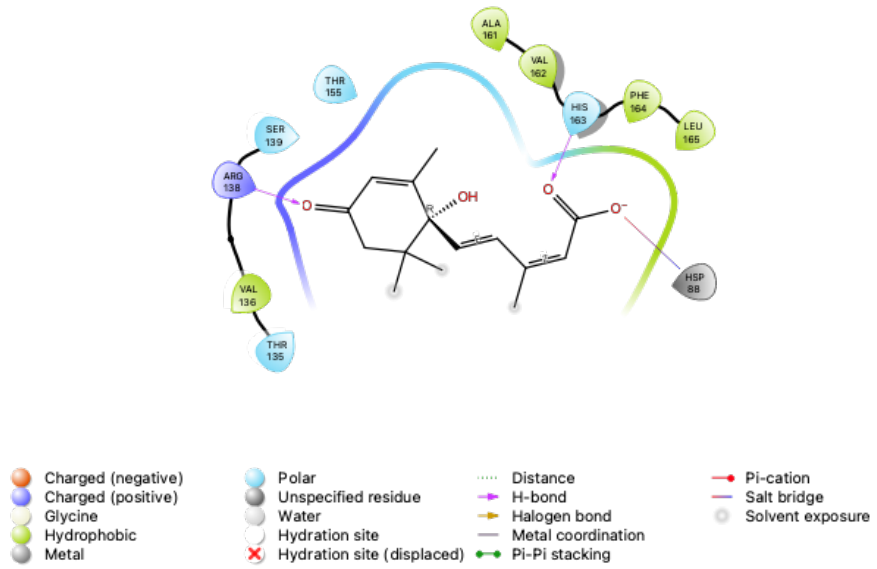


Figure 10.54: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found a salt bridge and two hydrogen bonds, HSP88-COO, and HIS163-COO and ARG138-CO, respectively. This figure shows the raw docking results, the distances are measured between charged N and O for the sb, and between heavy atoms for HB interactions.

2. b5-ABAno_5.c31387tc

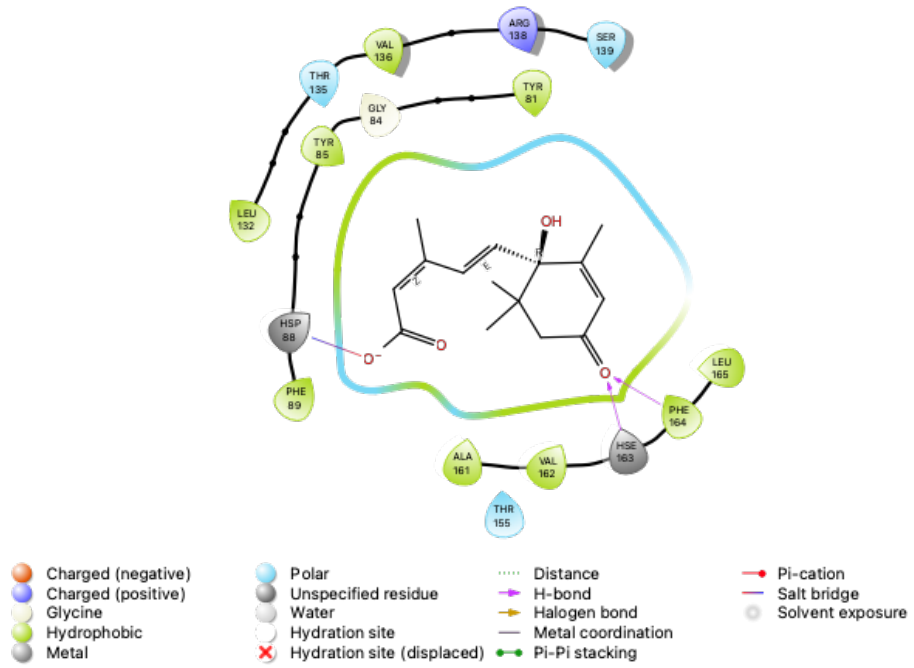


Figure 10.55: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found a salt bridge and two hydrogen bonds, HSP88-COO, and HIS163-CO and PHE164-CO, respectively. This figure shows the raw docking results, the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

3. b1-ABAno_0.c42248thcs

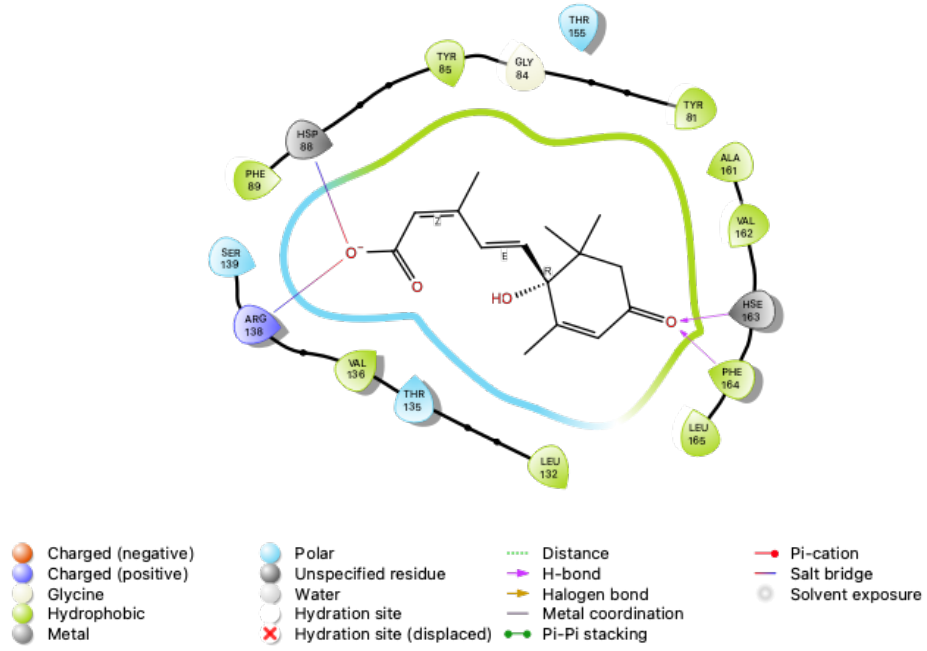


Figure 10.56: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found two salt bridges, HSP88-COO and ARG138-COO, and also two hydrogen bonds, HIS163-CO and PHE164-CO. This figure shows the raw docking results, the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

4. b10-ABA.no_0.c24691.h.c

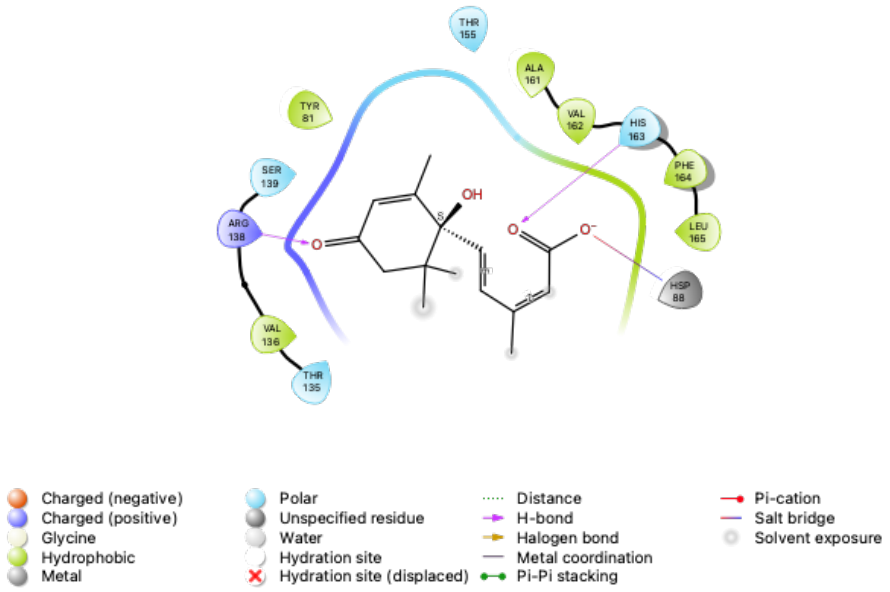


Figure 10.57: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found a salt bridge and two hydrogen bonds, HSP88-COO, and HIS163-COO and ARG138-CO, respectively. This figure shows the raw docking results, the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

5. b6-ABA.no_0.c306th

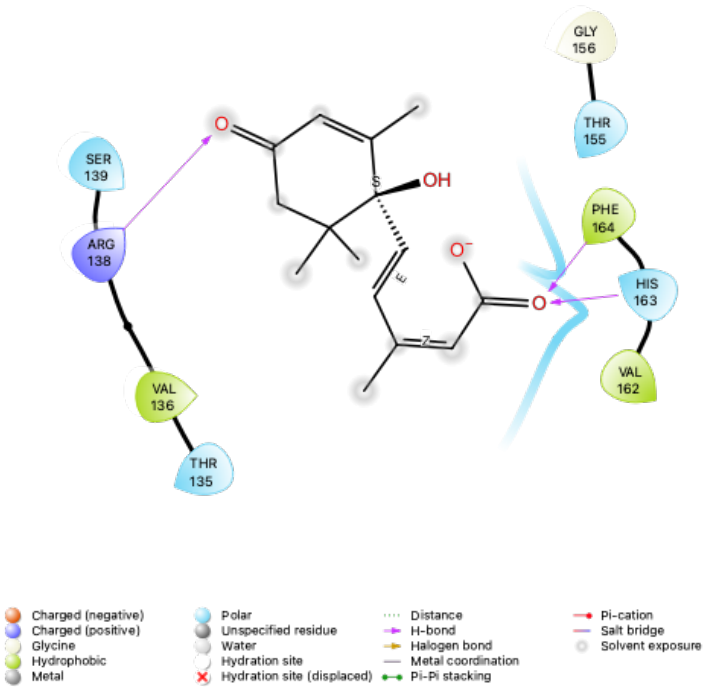


Figure 10.58: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, in this case, we only found three hydrogen bonds, HIS163-COO, PHE164-COO, and ARG138-COO respectively. This figure shows the raw docking results, the distances are measured between heavy atoms for the HB interactions

6. b9-ABA.no_0.c24494hc

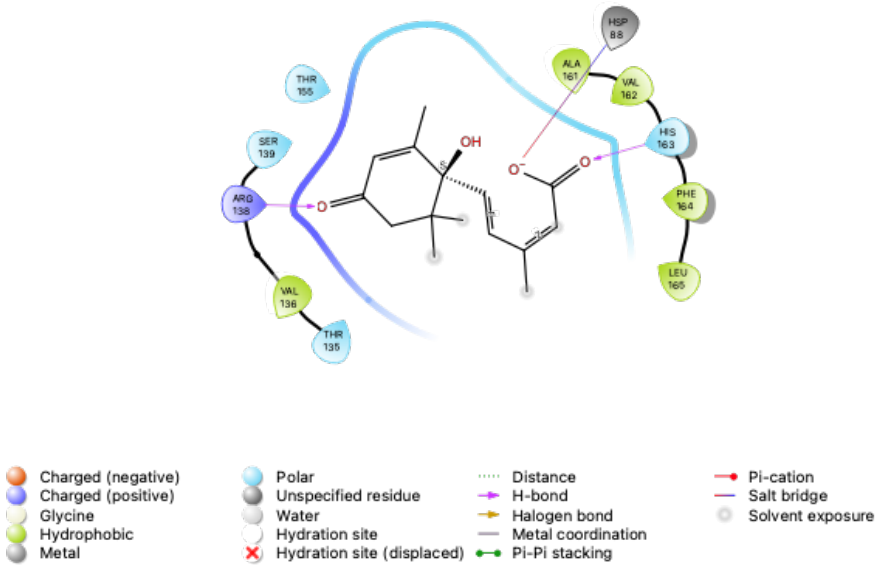


Figure 10.59: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found a salt bridge and two hydrogen bonds, HSP88-COO, and HIS163-COO and ARG138-CO, respectively. This figure shows the raw docking results, the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

7. b3-ABA.no_0c40267thc

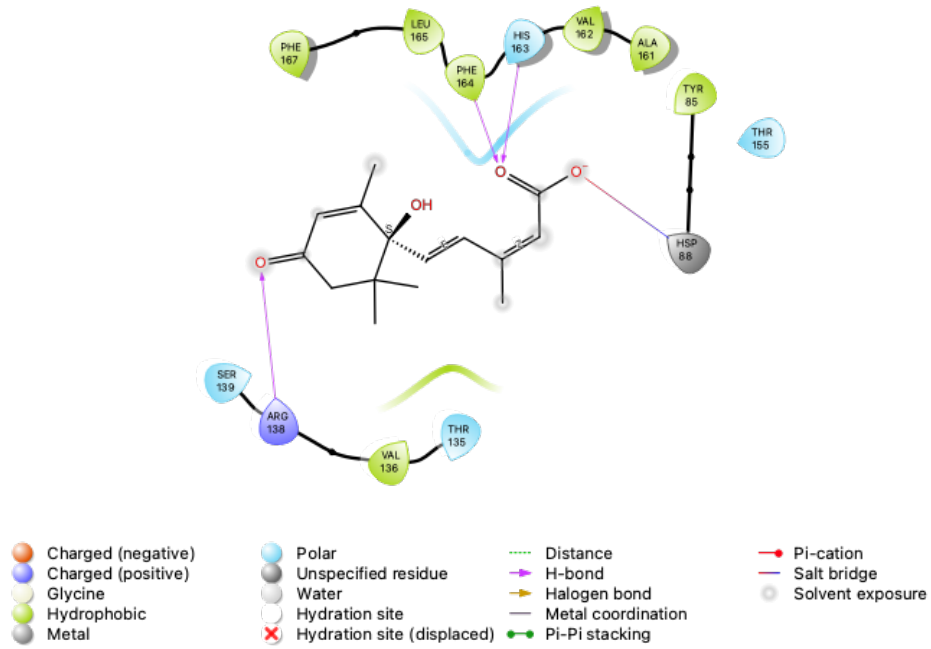


Figure 10.60: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found a salt bridge and three hydrogen bonds, HSP88-COO, and HIS163-COO, PHE164-COO, and ARG138-CO, respectively. This figure shows the raw docking results, the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions.

8. b4-ABA.no_0c24315hc

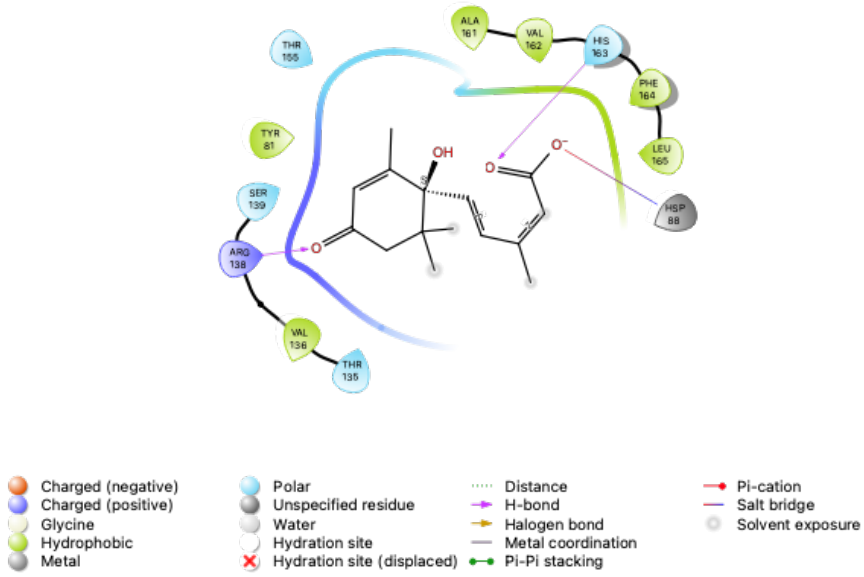


Figure 10.61: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found a salt bridge and two hydrogen bonds, HSP88-COO, and HIS163-COO and ARG138-CO, respectively. This figure shows the raw docking results, the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

9. b9-ABA.no_5c20724hc

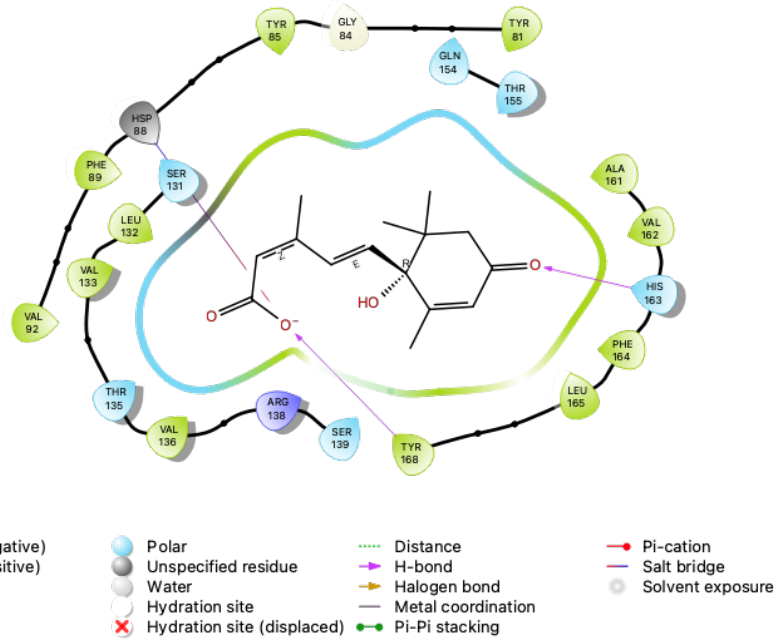


Figure 10.62: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found a salt bridge and two hydrogen bonds, HSP88-COO, and HIS163-CO and TYR-COO, respectively. This figure shows the raw docking results, the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

10. b2-ABA.no_5c18183ftc

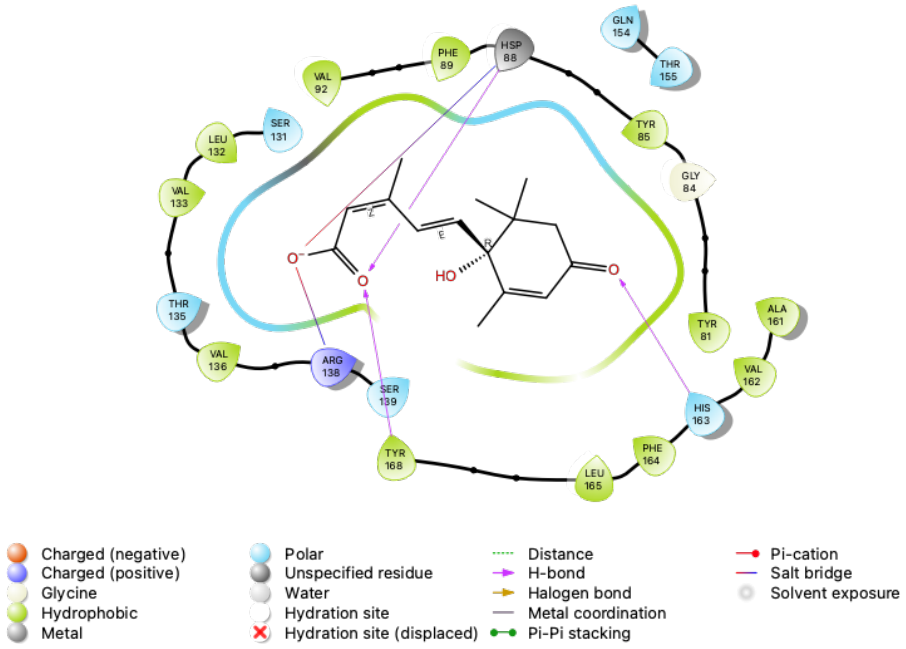


Figure 10.63: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found two salt bridges and three hydrogen bonds. So HSP88 residue has one SB and one HB interaction with the -COO group of ABA. On the other hand, another SB is found, ARG138-COO, and additionally, other two HBs which are between, HIS163-CO, TYR168-COO, respectively. This figure shows the raw docking results, the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

## 10.9 Appendix IX: Dealanization, binding site annealing, and minimization of selected ABA Docked structures.

return to main text

(a) b9-5-20724-9f.cm

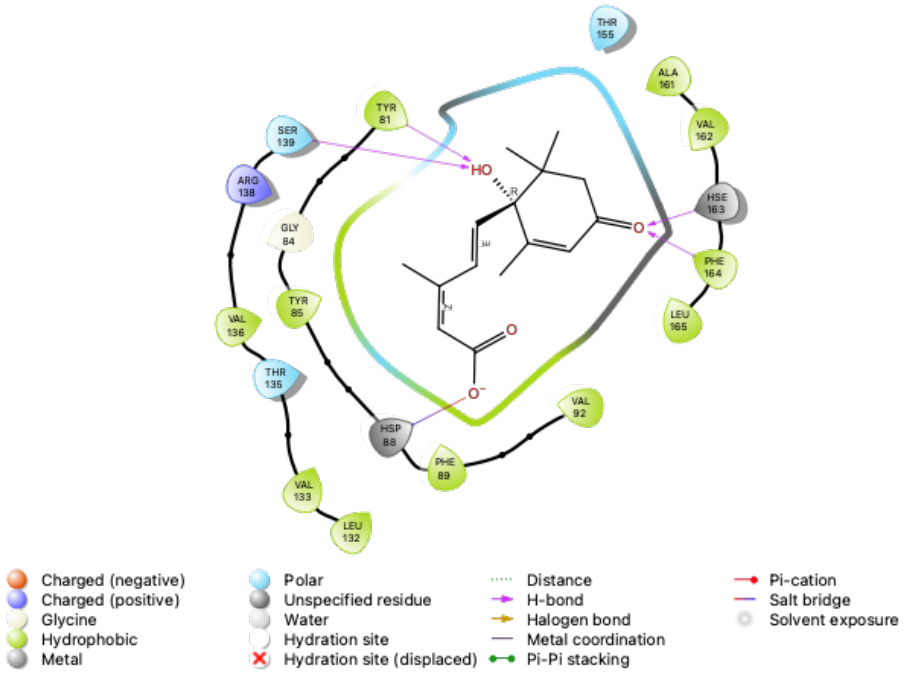


Figure 10.64: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found one salt bridge and four hydrogen bonds involving, besides the COO and CO groups, the hydroxyl group of ABA. So, the HSP88 interacts with the COO group of ABA through a SB, and the HBs are between, SER139-OH, TYR81-OH, HIS163-CO, and PHE164-CO. This figure shows the annealed and minimized structures obtained after the docking step. The distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

(b) b9_5_20724-8f.cm

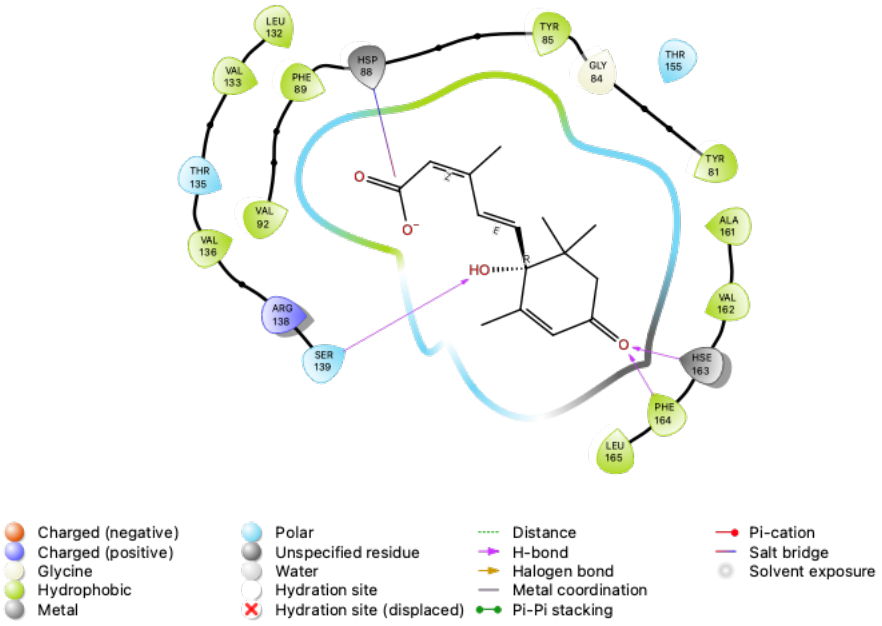


Figure 10.65: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found one salt bridge and three hydrogen bonds involving, besides the COO and CO groups, the hydroxyl group of ABA. So, the HSP88 interacts with the COO group of ABA through an SB, and the HBs are between, SER139-OH, HIS163-CO, and PHE164-CO. This figure shows the annealed and minimized structures obtained after the docking step. The distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

(c) b9_5_20724-5f.cm

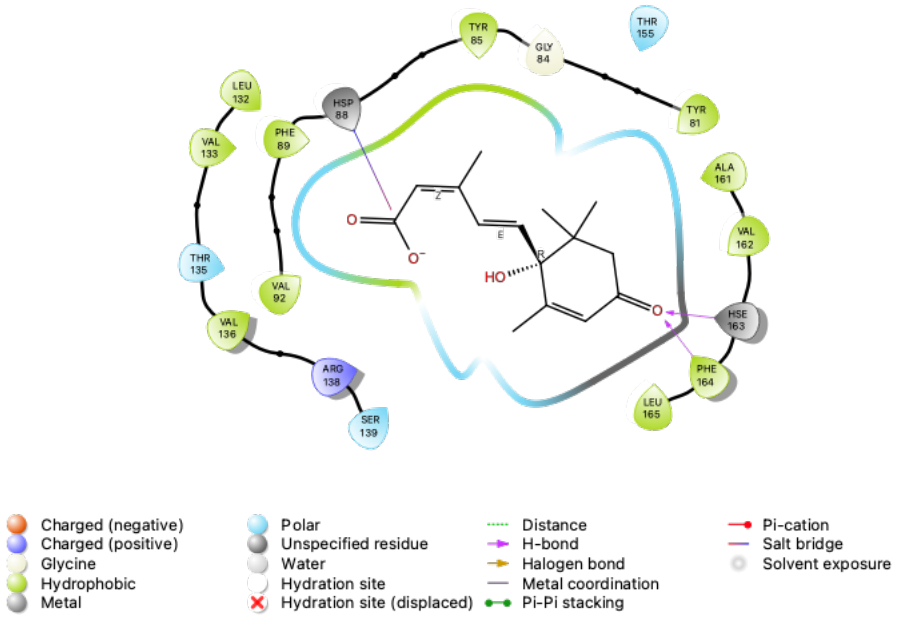


Figure 10.66: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found one salt bridge and two hydrogen bonds. So, the HSP88 interacts with the COO group of ABA through an SB, and the HBs are between HIS163-CO and PHE164-CO. This figure shows the annealed and minimized structures obtained after the docking step. The distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

(d) b9_5_20724-1f.cm

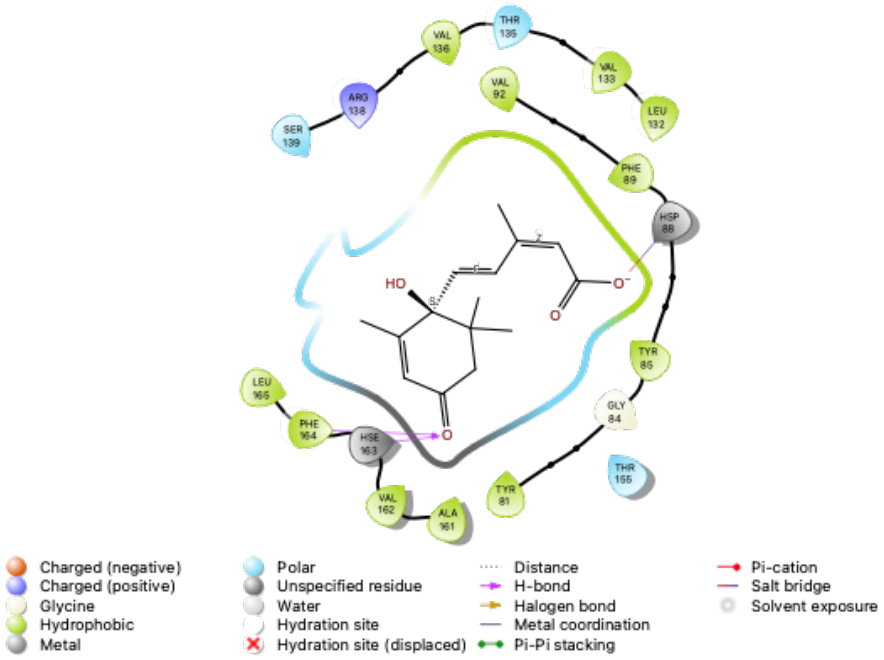


Figure 10.67: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found one salt bridge and two hydrogen bonds. So, the HSP88 interacts with the COO group of ABA through an SB, and the HBs are between HIS163-CO and PHE164-CO. This figure shows the annealed and minimized structures obtained after the docking step. The distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

(e) b9_5_20724-4f.cm

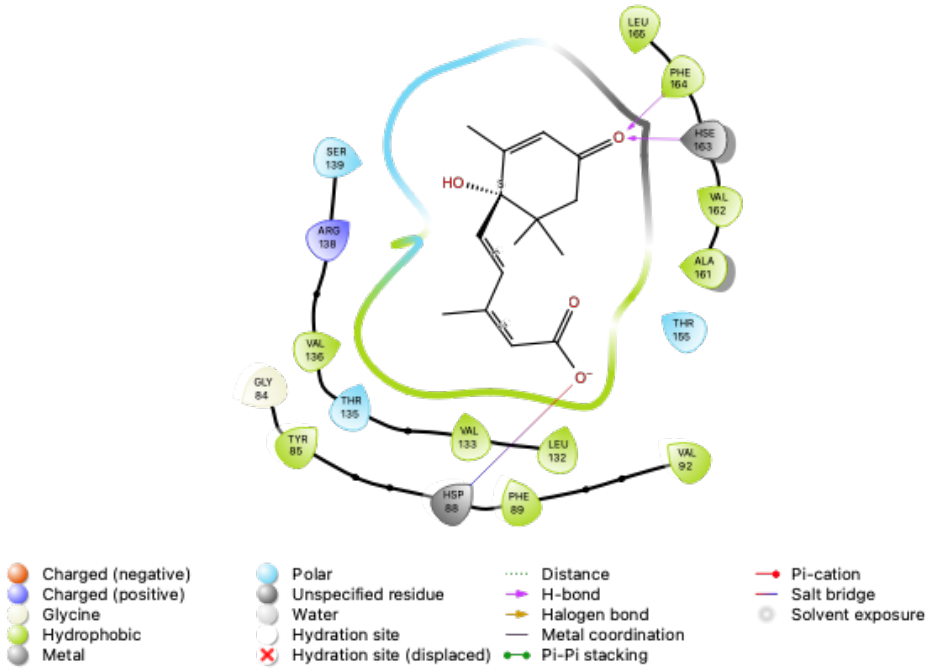


Figure 10.68: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found one salt bridge and two hydrogen bonds. So, the HSP88 interacts with the COO group of ABA through an SB, and the HBs are between HIS163-CO and PHE164-CO. This figure shows the annealed and minimized structures obtained after the docking step. The distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

(f) b9_5_20724-7f.cm

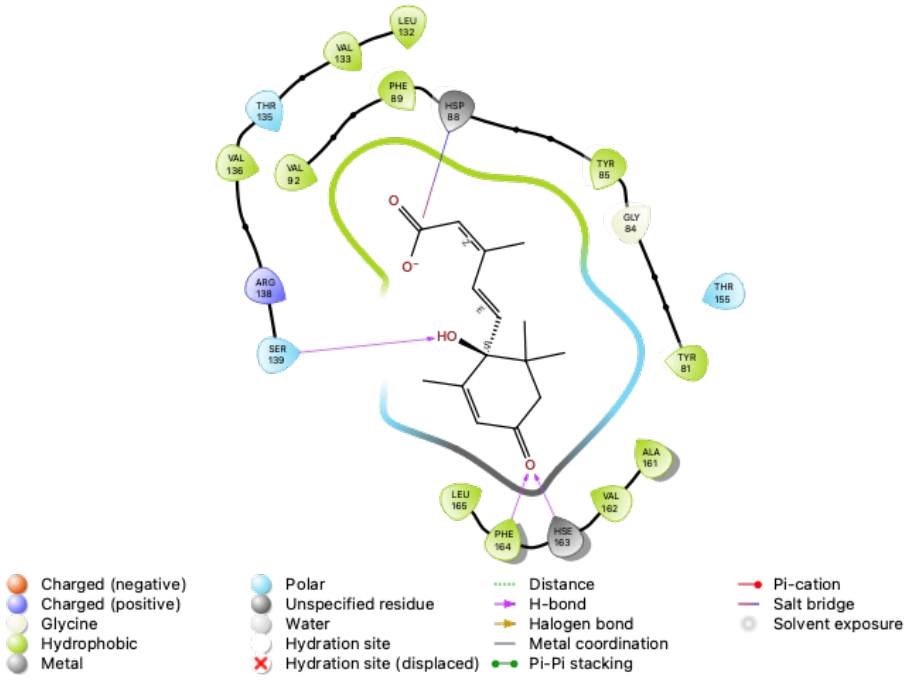


Figure 10.69: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found one salt bridge and three hydrogen bonds involving, besides the COO and CO groups, the hydroxyl group of ABA. So, the HSP88 interacts with the COO group of ABA through an SB, and the HBs are between, SER139-OH, HIS163-CO, and PHE164-CO. This figure shows the annealed and minimized structures obtained after the docking step. The distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

(g) b9_5_20724-3f.cm

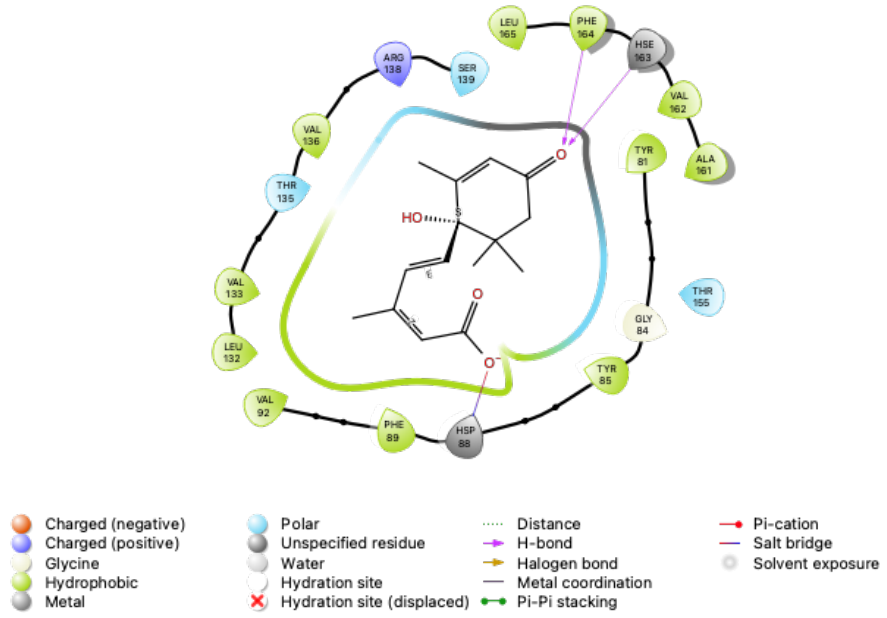


Figure 10.70: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found one salt bridge and two hydrogen bonds. So, the HSP88 interacts with the COO group of ABA through an SB, and the HBs are between HIS163-CO and PHE164-CO. This figure shows the annealed and minimized structures obtained after the docking step. The distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

(h) b4_0_24315-8f.cm

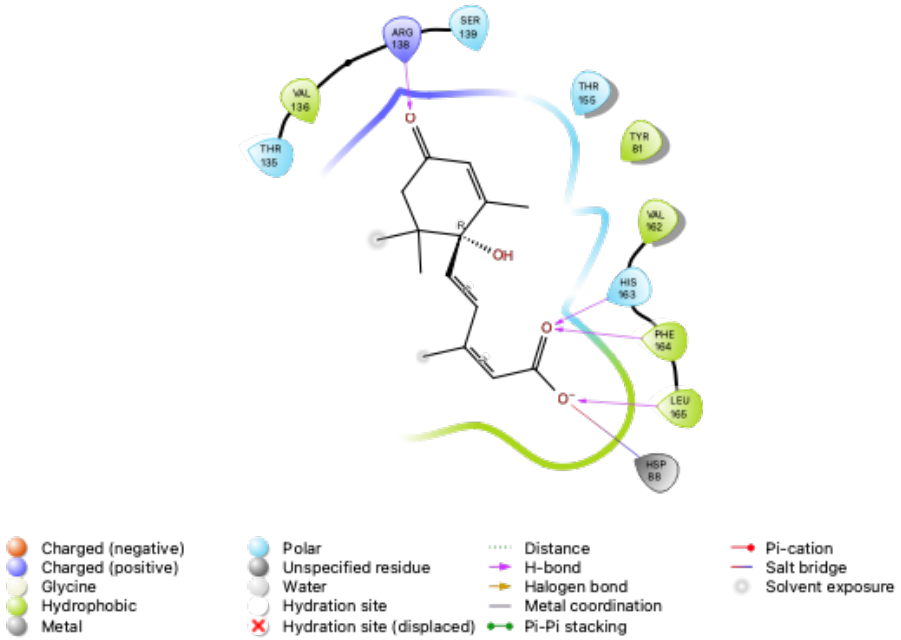


Figure 10.71: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found one salt bridge and four hydrogen bonds. So, the HSP88 interacts with the COO group of ABA through an SB, and the HBs are between ARG138-CO, HIS163-COO, PHE164-COO, and LEU165-COO. This figure shows the annealed and minimized structures obtained after the docking step. The distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

(i) b5_5_31387-10f.cm

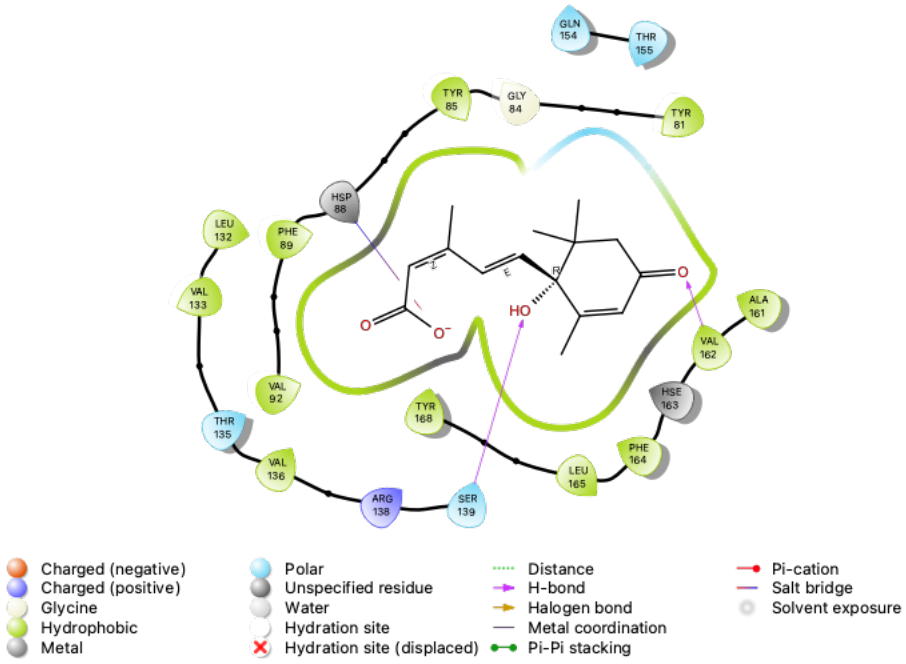


Figure 10.72: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found one salt bridge and two hydrogen bonds. So, the HSP88 interacts with the COO group of ABA through an SB, and the HBs are between SER139-OH and VAL162-CO. This figure shows the annealed and minimized structures obtained after the docking step. The distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

(j) b5_5_31387-9f.cm

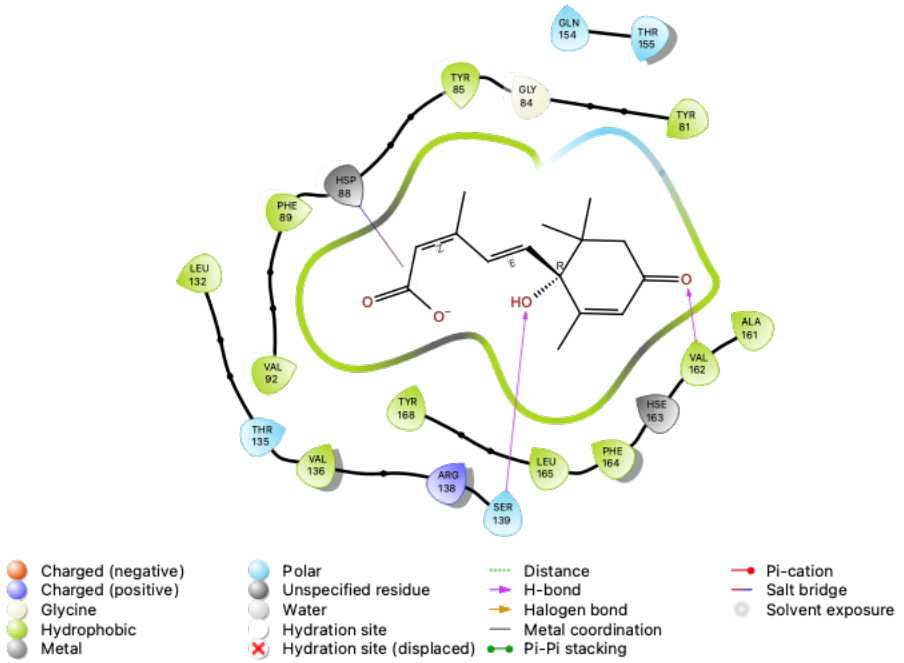


Figure 10.73: The ligand interaction diagram of the pharmacophore shows that, in addition to the series of polar and hydrophobic interactions between the amino acid residues of GCR1 and ABA, we also found one salt bridge and two hydrogen bonds. So, the HSP88 interacts with the COO group of ABA through an SB, and the HBs are between SER139-OH and VAL162-CO. This figure shows the annealed and minimized structures obtained after the docking step. The distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

## 10.10 Appendix X: Docking results for the HSP-88 screamed and alanized-GCR1 with GA1.

return to main text

1. b2-GA1.xray.c32412thc

   Like the ABA-docked GCR1 structures above, for the GA1 case, we also found a good location of GA1 in the binding site of GCR1, which besides shows a similar SB interaction between the GA1 carboxylate

group with the HSP88 in GCR1. On the other hand, something interesting to highlight is the fact that most of the GCR1 residues that interact with ABA, through HBs, also interact with GA1, as can be seen below.

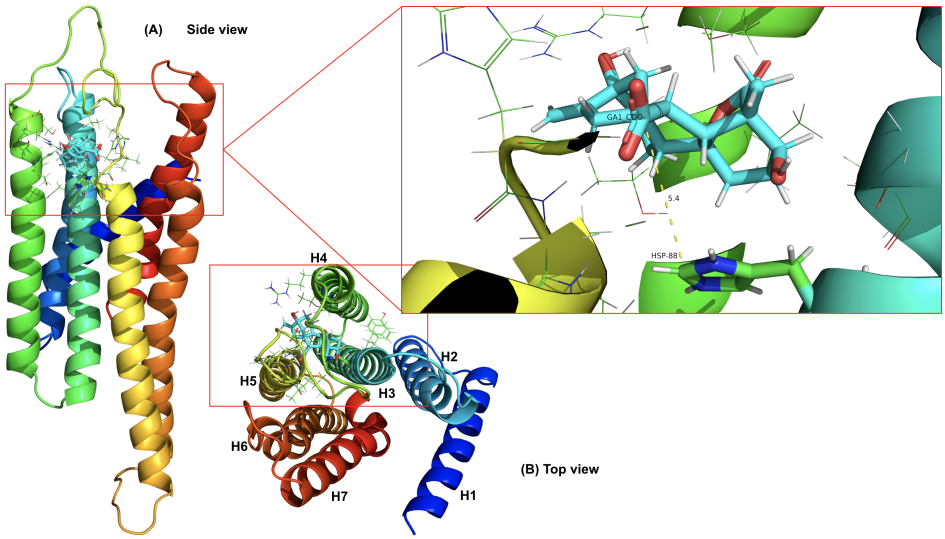


Figure 10.74: GCR1 3D, lateral and upper, views of the pharmacophore show both distances between the HSP88 GCR1-residue and -COO in GA1, and location of the ligand among helices H3, H4, and H5. Distances between HSP88 residue and -COO are measured from the charged N atom, or from the C atom bound to the charged N atom on HSP88 to the C atom, bound to the charged O atom in GA1. On the other hand, distances for HBs are measured between heavy interacting atoms.

2. b8-GA1.xray.c32444thc

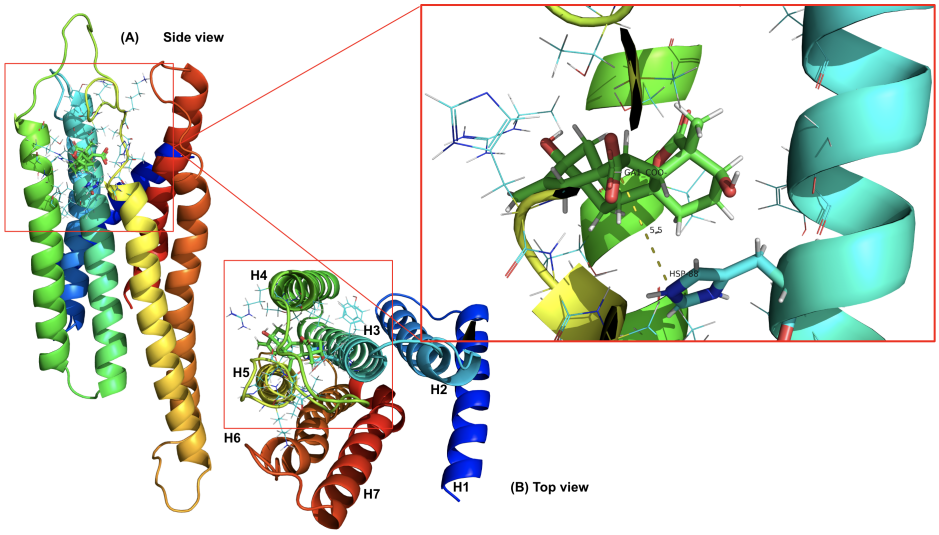


Figure 10.75: GCR1 3D, lateral and upper, views of the pharmacophore show both distances between the HSP88 GCR1-residue and -COO in GA1, and location of the ligand among helices H3, H4, and H5. Distances between HSP88 residue and -COO are measured from the charged N atom, or the C atom bound to the charged N atom on HSP88 to the C atom, bound to the charged O atom in GA1. On the other hand, distances for HBs are measured between heavy interacting atoms.

3. b10-GA1.xray.rnd_1.c845.thc

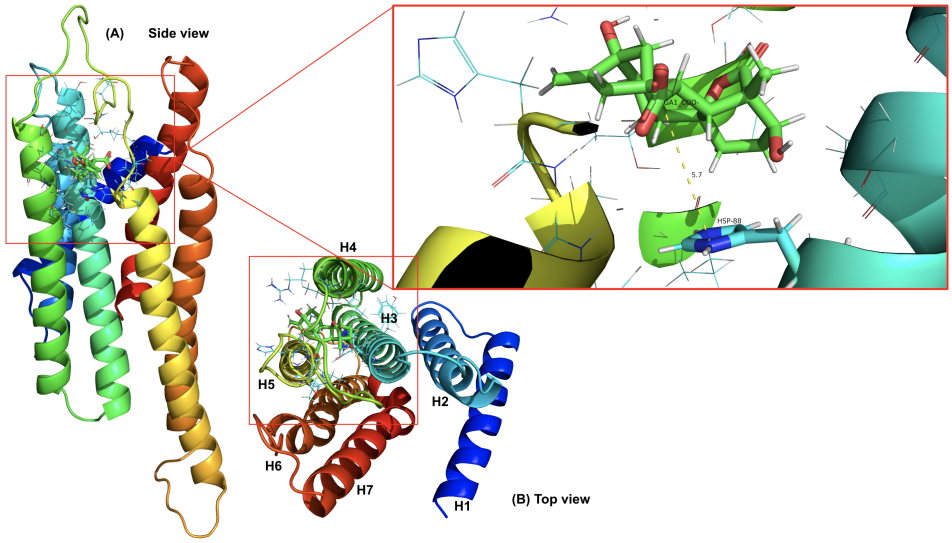


Figure 10.76: GCR1 3D, lateral and upper, views of the pharmacophore show both distances between the HSP88 GCR1-residue and -COO in GA1, and location of the ligand among helices H3, H4, and H5. Distances between HSP88 residue and -COO are measured from the charged N atom, or from the C atom bound to the charged N atom on HSP88 to the C atom, bound to the charged O atom in GA1. On the other hand, distances for HBs are measured between heavy interacting atoms.

4. b10-GA1.xray.rnd_1.c32809.hc

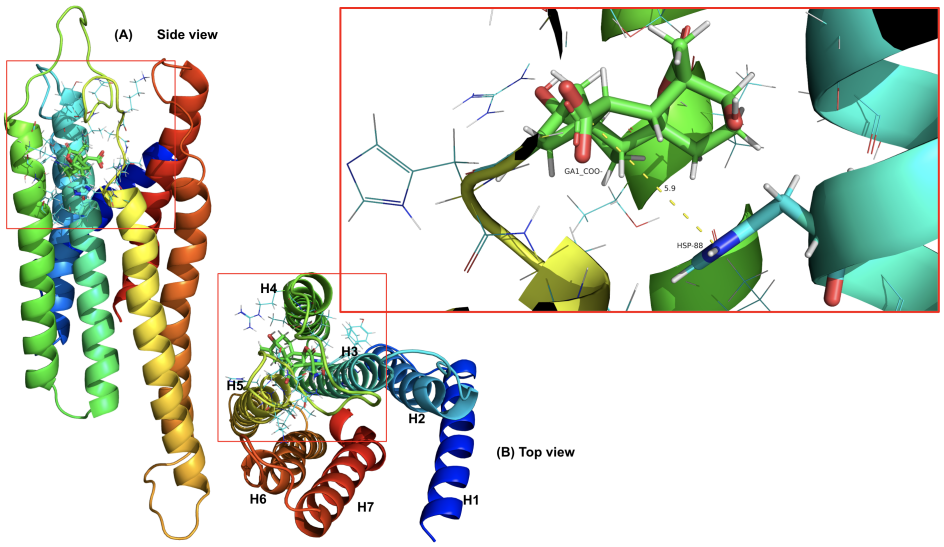


Figure 10.77: GCR1 3D, lateral and upper, views of the pharmacophore show both distances between the HSP88 GCR1-residue and -COO in GA1, and location of the ligand among helices H3, H4, and H5. Distances between HSP88 residue and -COO are measured from the charged N atom or the C atom bound to the charged N atom on HSP88 to the C atom, bound to the charged O atom in GA1. On the other hand, distances for HBs are measured between heavy interacting atoms.

5. b10-GA1.xray.rnd_1.c30864.thc
6. b3-GA1.xray.rnd_1.c32054.thc
7. b1-GA1.xray.rnd_1.c6914.hc

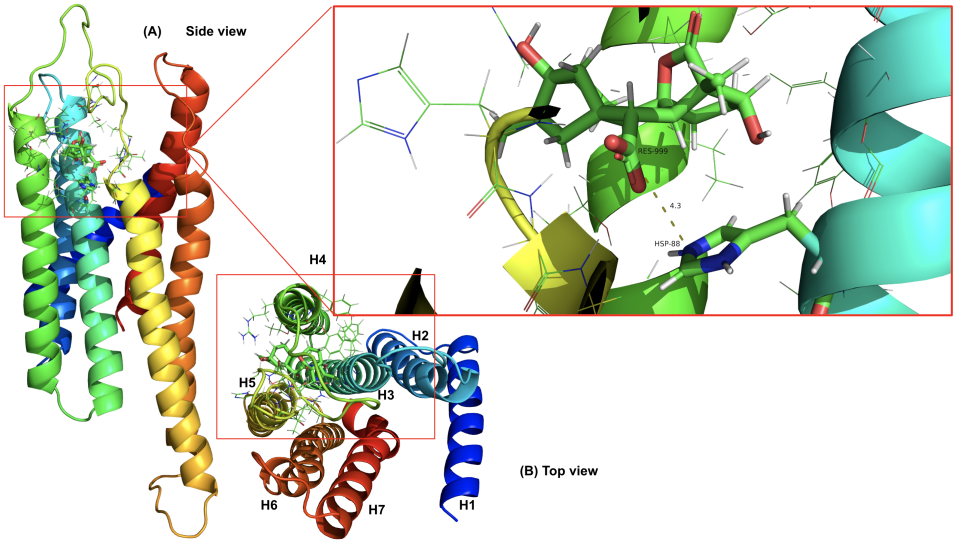


Figure 10.78: GCR1 3D, lateral and upper, views of the pharmacophore show both distances between the HSP88 GCR1-residue and -COO in GA1, and location of the ligand among helices H3, H4, and H5. Distances between HSP88 residue and -COO are measured from the charged N atom or the C atom bound to the charged N atom on HSP88 to the C atom bound to the charged O atom in GA1. On the other hand, distances for HBs are measured between heavy interacting atoms.

8. b9-GA1.xray.rnd_1.c850hc

9. b4-GA1.no_0.rnd_1.c44628hc

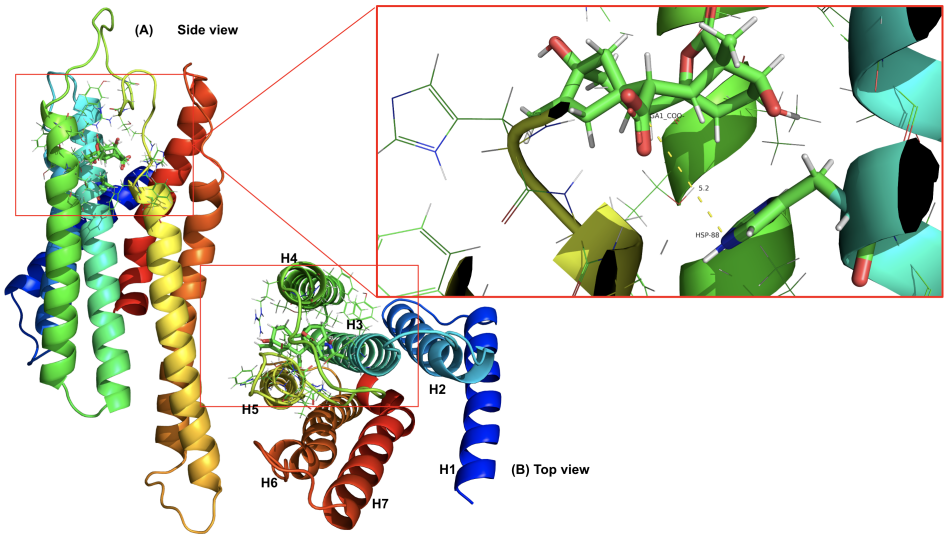


Figure 10.79: GCR1 3D, lateral and upper, views of the pharmacophore show both distances between the HSP88 GCR1-residue and -COO in GA1, and location of the ligand among helices H3, H4, and H5. Distances between HSP88 residue and -COO are measured from the charged N atom or the C atom bound to the charged N atom on HSP88 to the C atom bound to the charged O atom in GA1. On the other hand, distances for HBs are measured between heavy interacting atoms.

10. b4-GA1.xray.rnd_1.c843thc

11. b3-GA1.xray.rnd_1.c18616tc

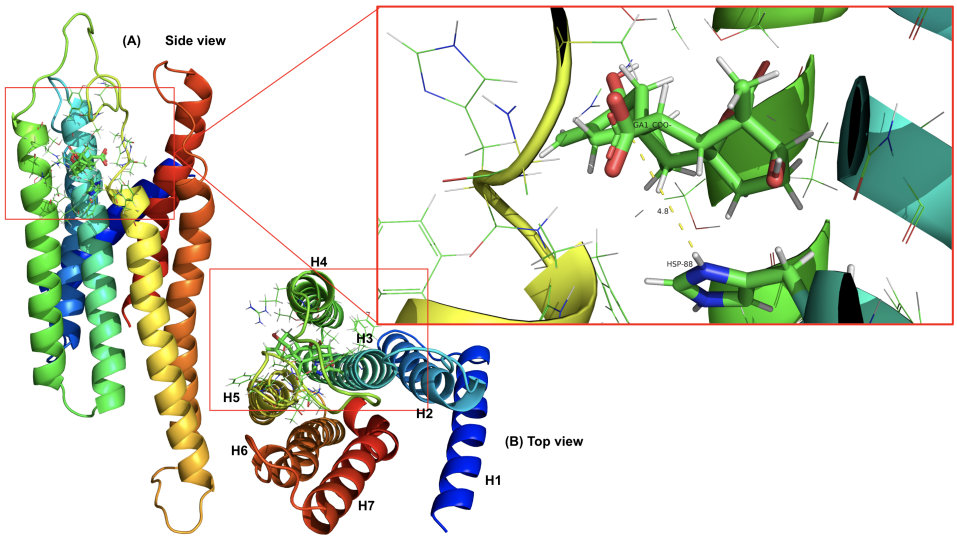


Figure 10.80: GCR1 3D, lateral and upper, views of the pharmacophore show both distances between the HSP88-GCR1 residue and -COO in GA1, and location of the ligand among helices H3, H4, and H5. Distances between HSP88 residue and -COO are measured from the charged N atom or the C atom bound to the charged N atom on HSP88 to the C atom bound to the charged O atom in GA1. On the other hand, distances for HBs are measured between heavy interacting atoms.

## 10.11 Appendix XI: Dealanization, binding site annealing, and minimization of selected GA1 Docked structures.

return to main text

1. b10_x_845-5fcm

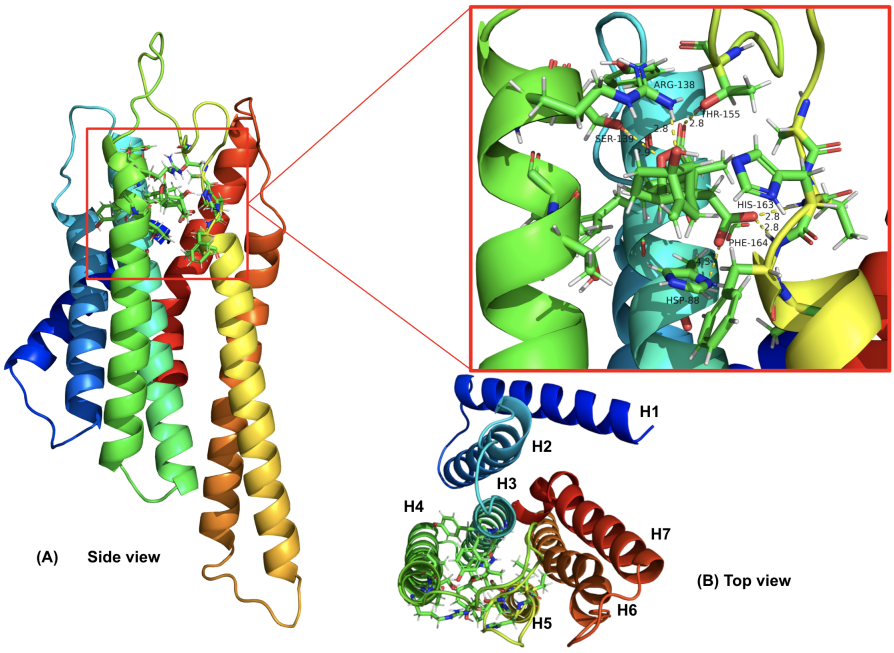


Figure 10.81: Minimized-annealed GCR1-3D GA1-docked structure, b10-x-845-5c. lateral and upper views of the pharmacophore show both SB interaction distances between the HSP88-GCR1 residue and -COO in GA1, aside from distances for HBs between HIS163-COO, PHE164-COO, ARG138-OH, SER139-O-lac, and TYR155-COO-lac. Additionally, we show the location of the ligand among helices H3, H4, and H5. Distances between HSP88 residue and -COO are measured from the C atom bound to the charged N atom on HSP88 to the C atom bound to the charged O atom in GA1. On the other hand, distances for HBs are measured between heavy interacting atoms.

2. b10_x_845-3fcm

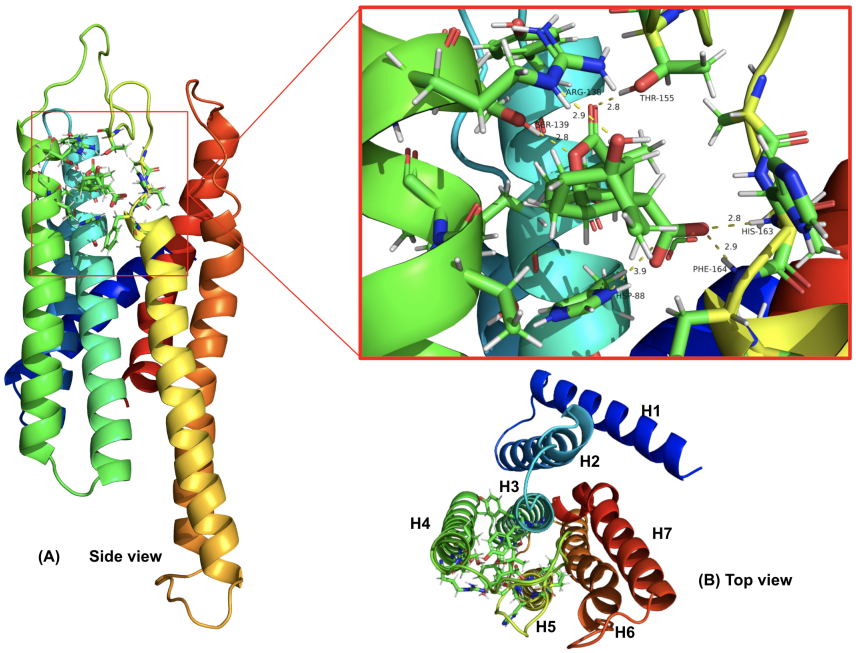


Figure 10.82: Minimized-annealed GCR1-3D GA1-docked structure, b10-x-845-3c. lateral and upper views of the pharmacophore show both SB interaction distances between the HSP88-GCR1 residue and -COO in GA1, aside from distances for HBs between HIS163-COO, PHE164-COO, ARG138-OH, SER139-O-lac, and TYR155-COO-lac. Additionally, we show the location of the ligand among helices H3, H4, and H5. Distances between HSP88 residue and -COO are measured from the C atom bound to the charged N atom on HSP88 to the C atom bound to the charged O atom in GA1. On the other hand, distances for HBs are measured between heavy interacting atoms.

3. b1_x_6914-10fcm

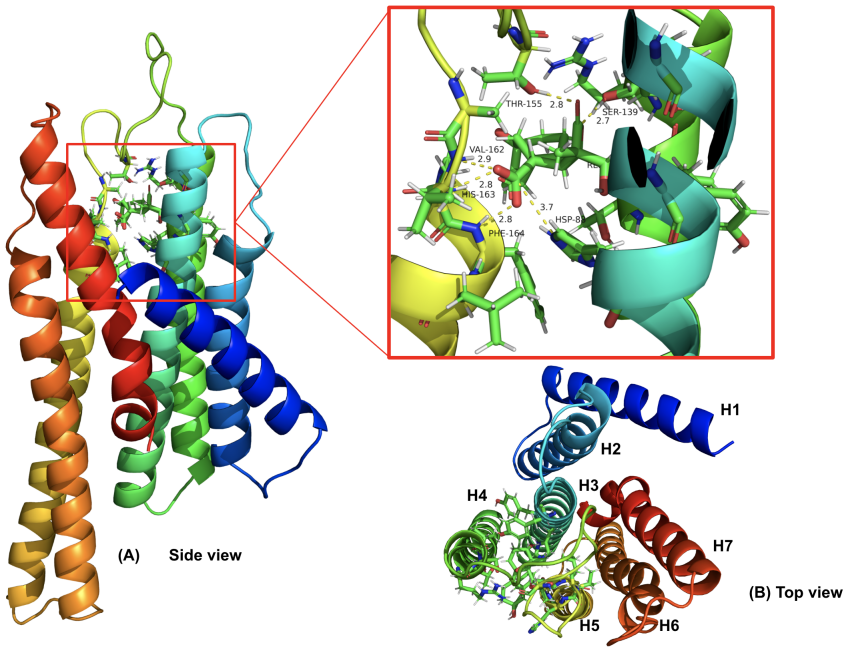


Figure 10.83: Minimized-annealed GCR1-3D GA1-docked structure, b1_x_6914-10c. lateral and upper views of the pharmacophore show both SB interaction distance between the HSP88-GCR1 residue and -COO in GA1, aside from distances for HBs between VAL162-COO, HIS163-COO, PHE164-COO, SER139-O-lac, and TYR155-COO-lac. Additionally, we show the location of the ligand among helices H3, H4, and H5. Distances between HSP88 residue and -COO are measured from the C atom bound to the charged N atom on HSP88 to the C atom bound to the charged O atom in GA1. On the other hand, distances for HBs are measured between heavy interacting atoms.

1. b2-GA1.xray.c32412thc

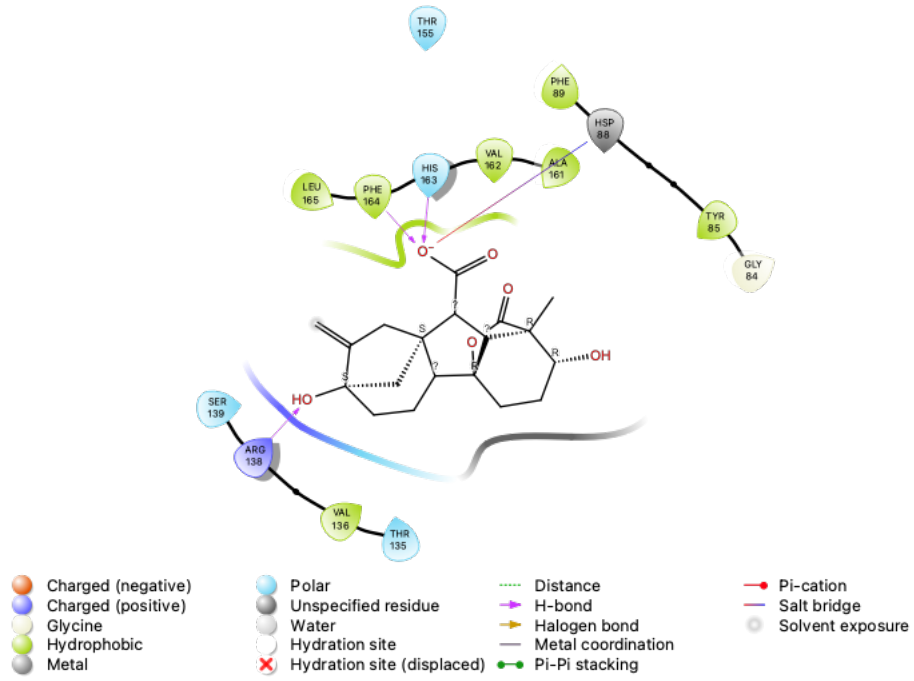


Figure 10.84: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found one salt bridge and three hydrogen bonds. So, the HSP88 interacts with the carboxylate -COO group of GA1 through a SB, and the HBs are between HIS163-COO, PHE164-COO, and ARG138-OH. This figure shows the raw docking results; the distances are measured between charged N and O for the SB and between heavy atoms for HB interactions

2. b8-GA1.xray.c32444thc

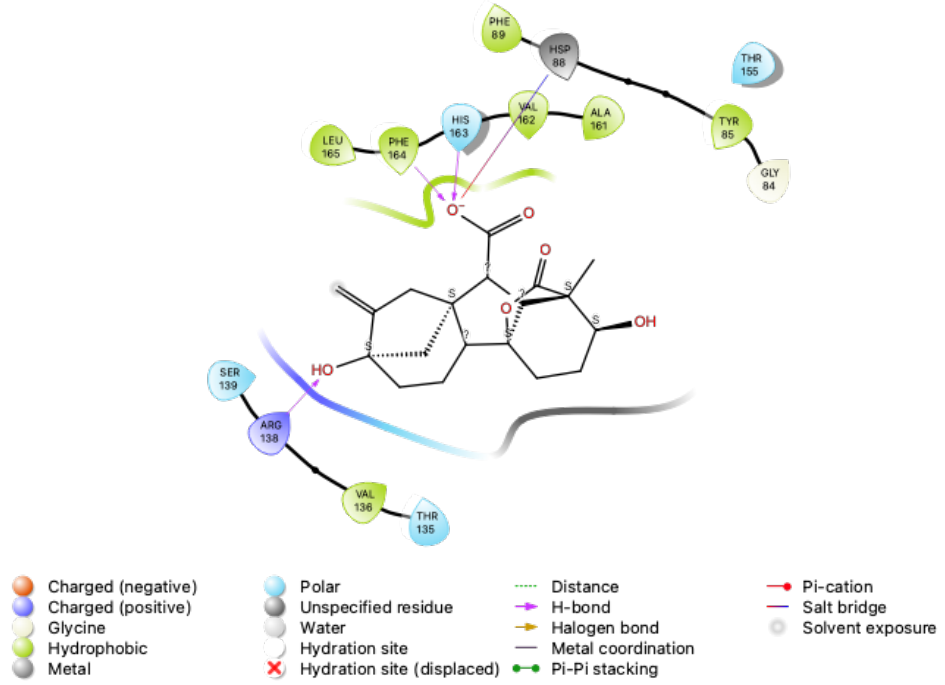


Figure 10.85: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found one salt bridge and three hydrogen bonds. So, the HSP88 interacts with the carboxylate -COO group of GA1 through an SB, and the HBs are between HIS163-COO, PHE164-COO, and ARG138-OH. This figure shows the raw docking results; the distances are measured between charged N and O for the SB and between heavy atoms for HB interactions

3. b10-GA1.xray.rnd_1.c845.thc

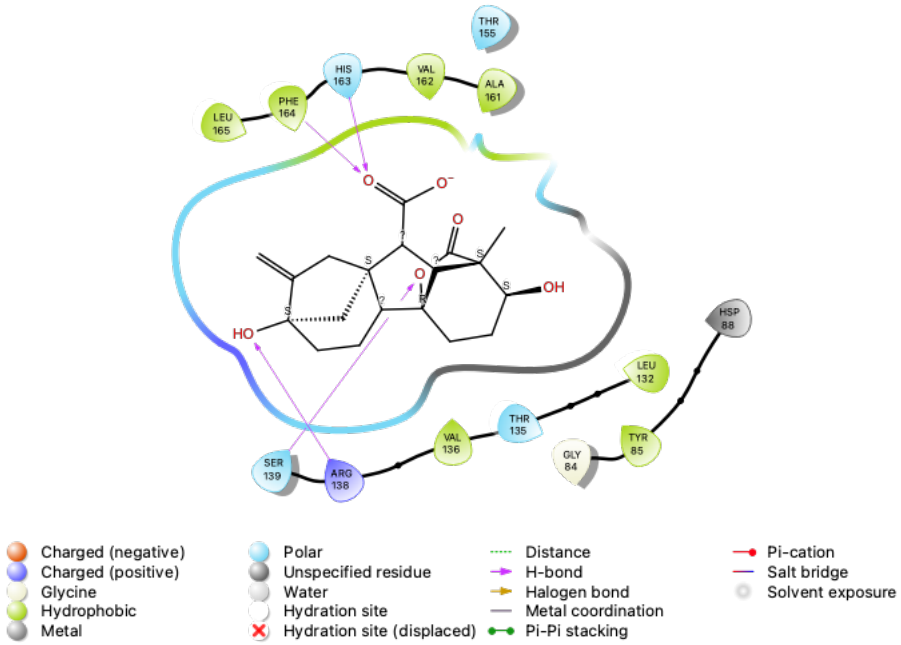


Figure 10.86: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found four hydrogen bonds, but no one salt bridge. The HBs are between HIS163-COO, PHE164-COO, ARG138-OH, and finally, HB between the oxygen in the lactone group of GA1 with SER139. This figure shows the raw docking results; the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

4. b10-GA1.xray.rnd_1.c32809.hc

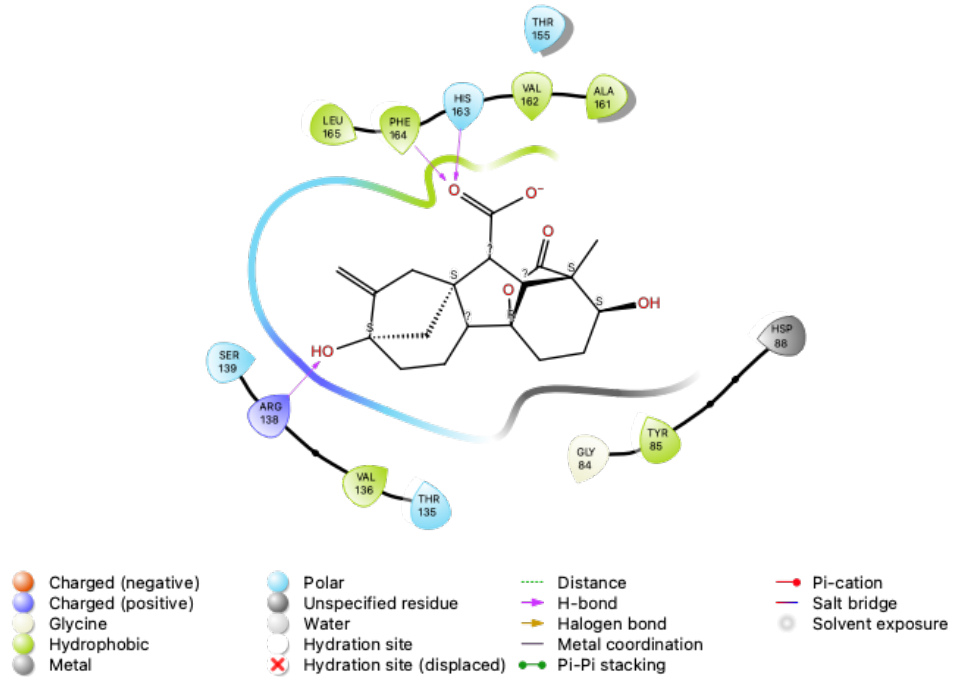


Figure 10.87: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found three hydrogen bonds, but no one salt bridge. The HBs are between HIS163-COO, PHE164-COO, and ARG138-OH. This figure shows the raw docking results, the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

5. b10-GA1.xray.rnd_1.c30864.thc

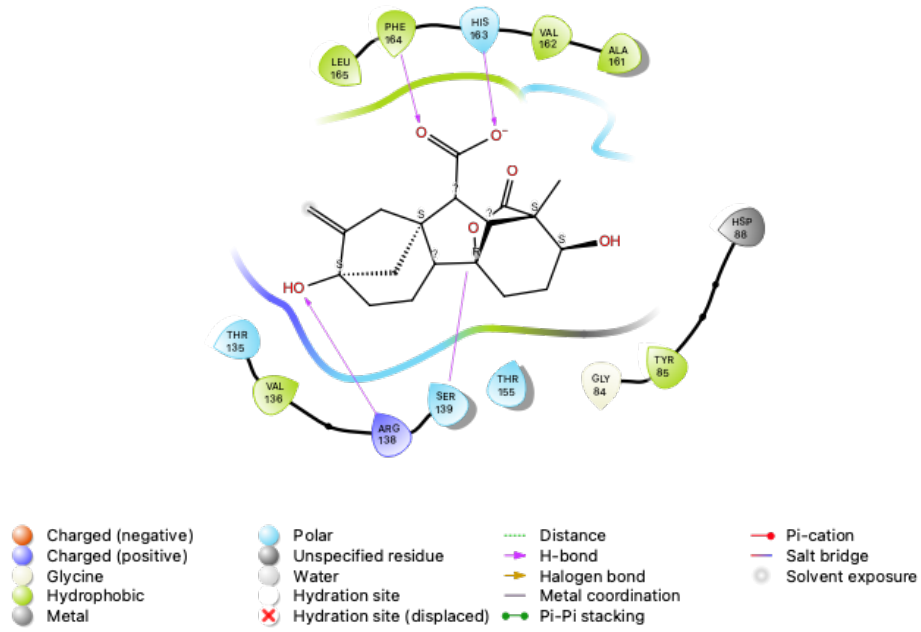


Figure 10.88: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found four hydrogen bonds, but no one salt bridge. The HBs are between HIS163-COO, PHE164-COO, ARG138-OH, and SER139-O-lac. This figure shows the raw docking results; the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

6. b3-GA1.xray.rnd_1.c32054.thc

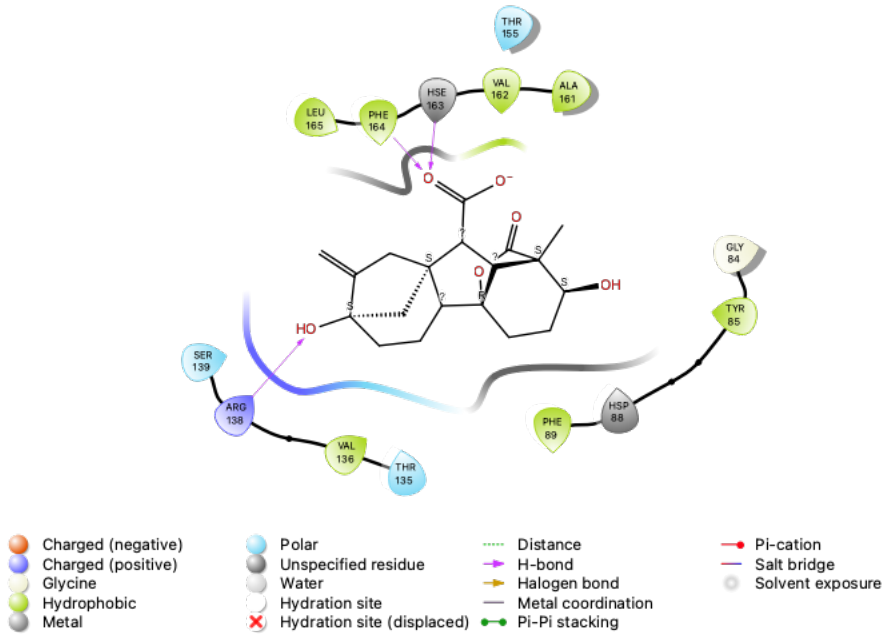


Figure 10.89: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found three hydrogen bonds, but no one salt bridge. The HBs are between HIS163-COO, PHE164-COO, and ARG138-OH. This figure shows the raw docking results; the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

7. b1-GA1.xray.rnd_1.c6914.hc

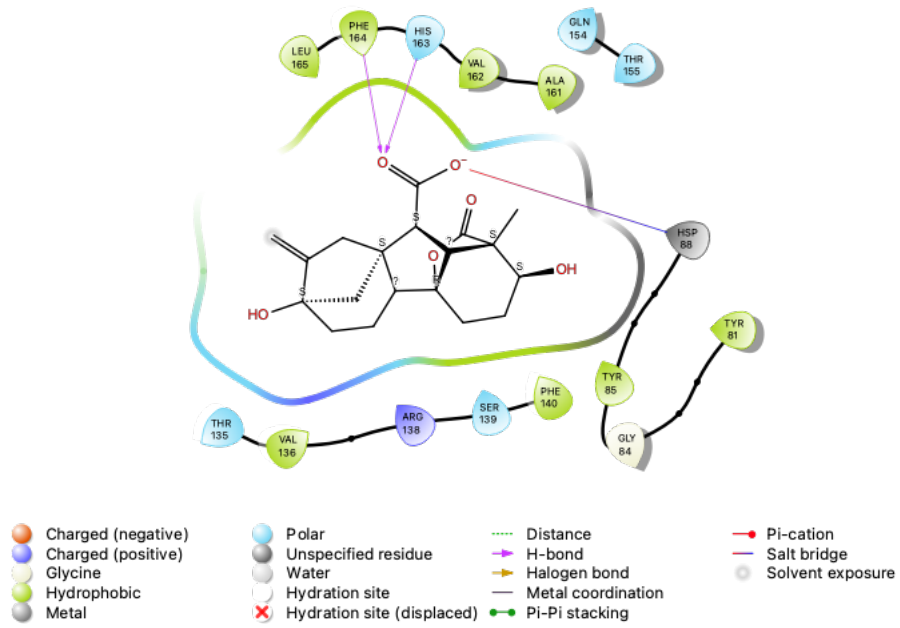


Figure 10.90: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found two hydrogen bonds and one salt bridge between GCR1-HSP88 residue and GA1-COO. The HBs are between HIS163-COO and PHE164-COO. This figure shows the raw docking results; the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

8. b9-GA1.xray.rnd_1.c850hc

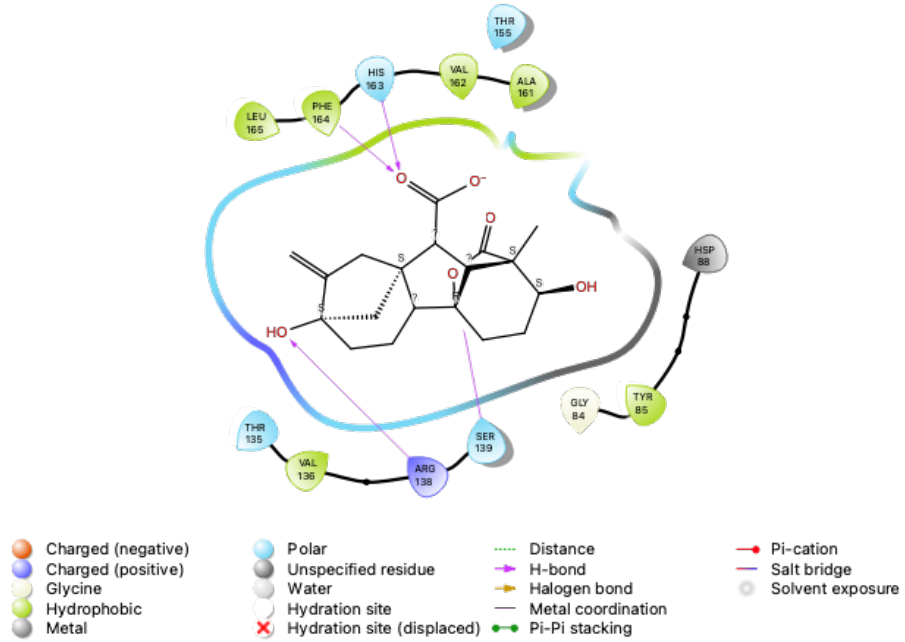


Figure 10.91: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found four hydrogen bonds, but no one salt bridge. The HBs are between HIS163-COO, PHE164-COO, ARG138-OH, and SER139-O-lac. This figure shows the raw docking results; the distances are measured between charged N and O for the SB and between heavy atoms for HB interactions

9. b4-GA1.no_0.rnd_1.c44628hc

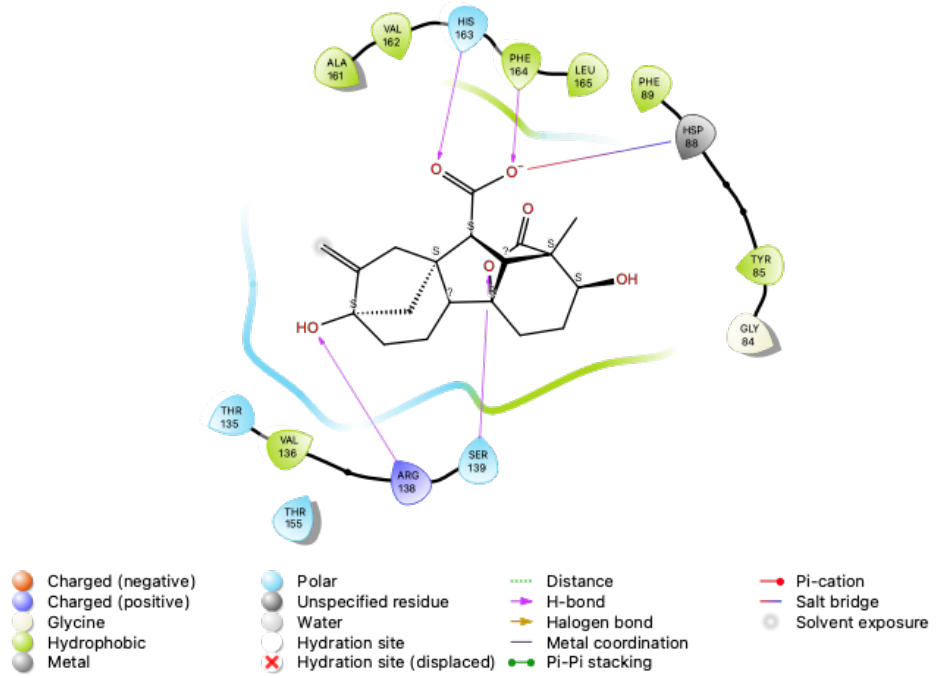


Figure 10.92: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found four hydrogen bonds and one salt bridge between GCR1-HSP88 residue and GA1-COO. The HBs are between HIS163-COO, PHE164-COO, ARG138-OH, and SER139-O-lac. This figure shows the raw docking results; the distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

10. b4-GA1.xray.rnd_1.c843thc

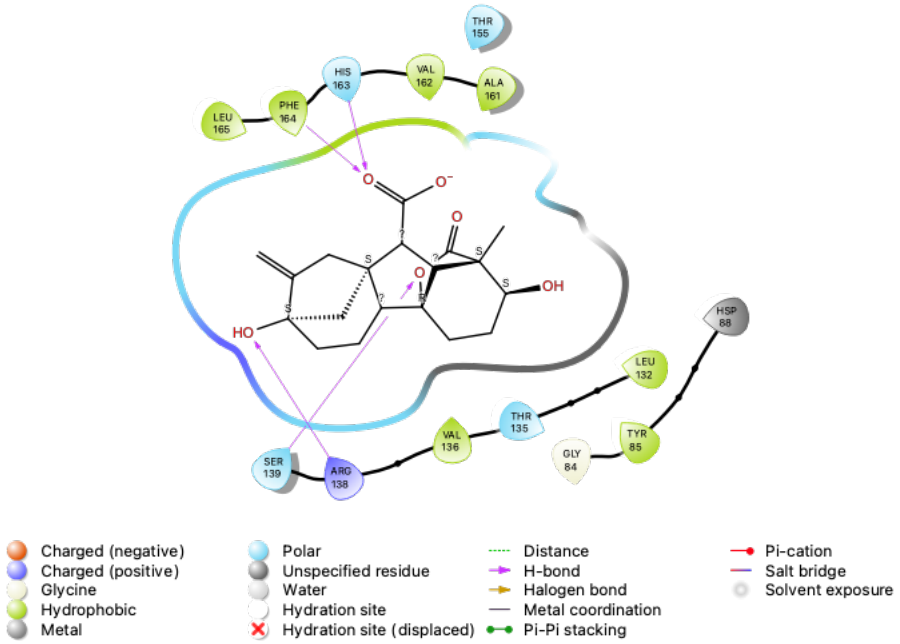


Figure 10.93: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found four hydrogen bonds, but no one salt bridge. The HBs are between HIS163-COO, PHE164-COO, ARG138-OH, and SER139-O-lac. This figure shows the raw docking results. The distances are measured between charged N and O for the SB, and between heavy atoms for HB interactions

11. b3-GA1.xray.rnd_1.c18616tc

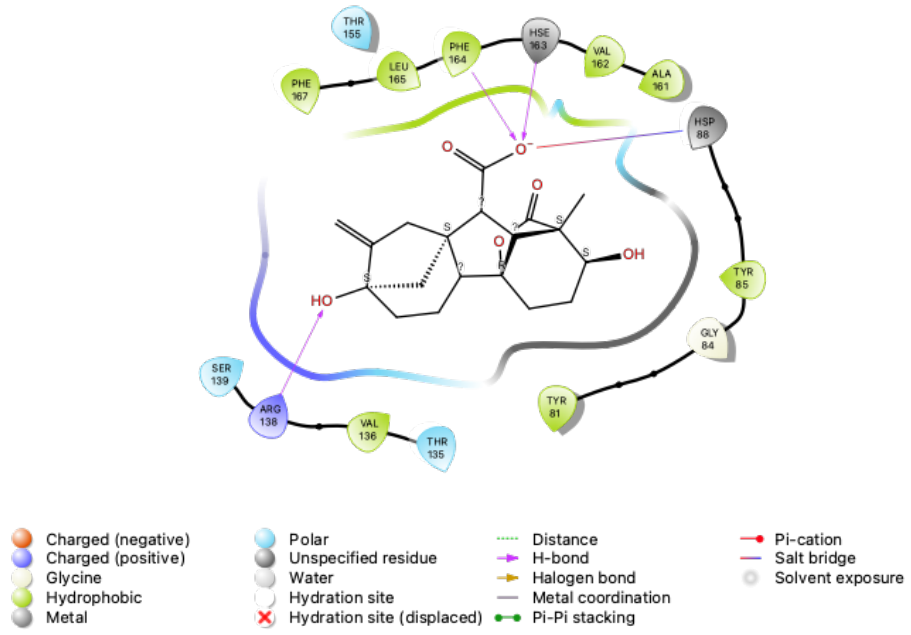


Figure 10.94: The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found three hydrogen bonds and one salt bridge between GCR1-HSP88 residue and GA1-COO. The HBs are between HIS163-COO, PHE164-COO, and ARG138-OH. This figure shows the raw docking results; the distances are measured between charged N and O for the SB and between heavy atoms for HB interactions

(a) b10_x_845-5fcm

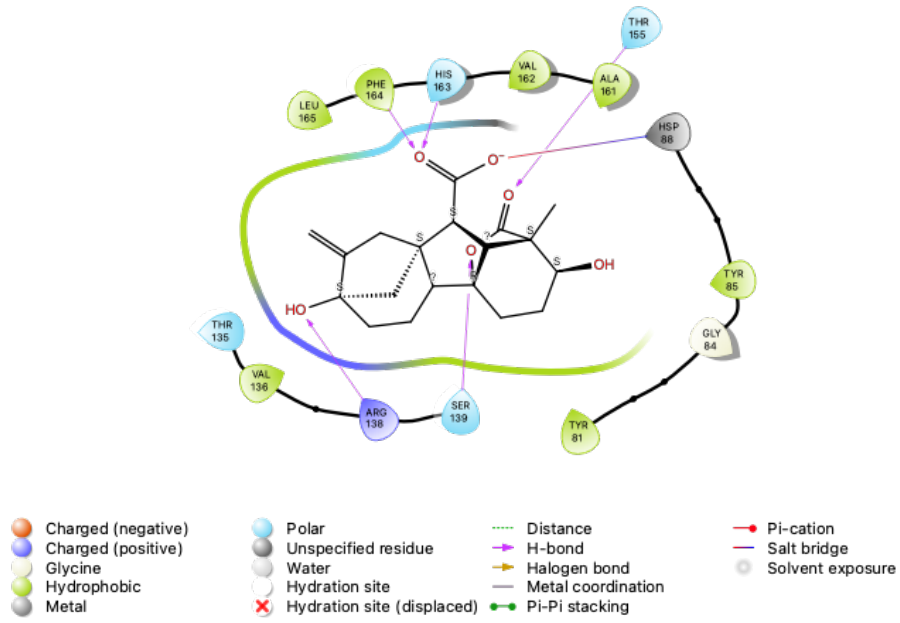


Figure 10.95: This figure shows the annealed and minimized results for the GA1 docked-GCR1 structure, b10-x-845-5c. The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found five hydrogen bonds and one salt bridge. The SB is between GCR1-HSP88 residue and GA1-COO. The HBs are between HIS163-COO, PHE164-COO, ARG138-OH, SER139-O-lac, and TYR155-COO-lac. Distances are measured between charged N and O for the SB and between heavy atoms for HB interactions

(b) b10_x_845-3fcm

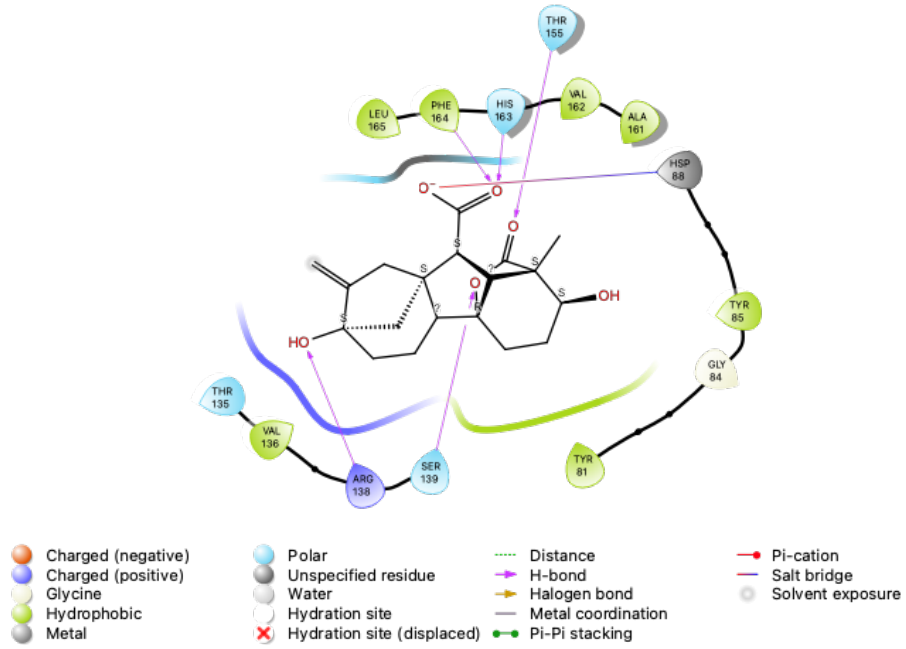


Figure 10.96: This figure shows the annealed and minimized results for the GA1 docked-GCR1 structure, b10-x-845-3c. The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found five hydrogen bonds and one salt bridge. The SB is between GCR1-HSP88 residue and GA1-COO. The HBs are between HIS163-COO, PHE164-COO, ARG138-OH, SER139-O-lac, and TYR155-COO-lac. Distances are measured between charged N and O for the SB and between heavy atoms for HB interactions

(c) b1_x_6914-10fcm

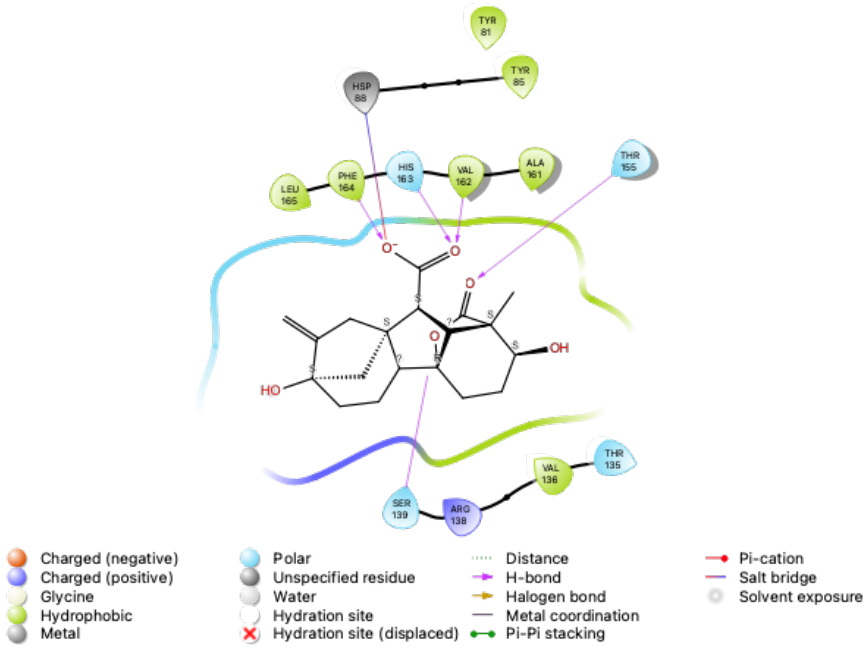


Figure 10.97: This figure shows the annealed and minimized results for the GA1 docked-GCR1 structure, b1_x_6914-10c. The ligand interaction diagram of the pharmacophore shows that, in addition to a series of polar and hydrophobic interactions between residues belonging to the predicted binding site of GCR1 with GA1, we also found five hydrogen bonds and one salt bridge. The SB is between GCR1-HSP88 residue and GA1-COO. The HBs are between VAL162-COO, HIS163-COO, PHE164-COO, SER139-O-lac, and TYR155-COO-lac. Distances are measured between charged N and O for the SB and between heavy atoms for HB interactions

## 10.12 Appendix XII: Selection of GCR1-GA1 docked structures.

return to main text

**GA1-b10x845-5fcm**

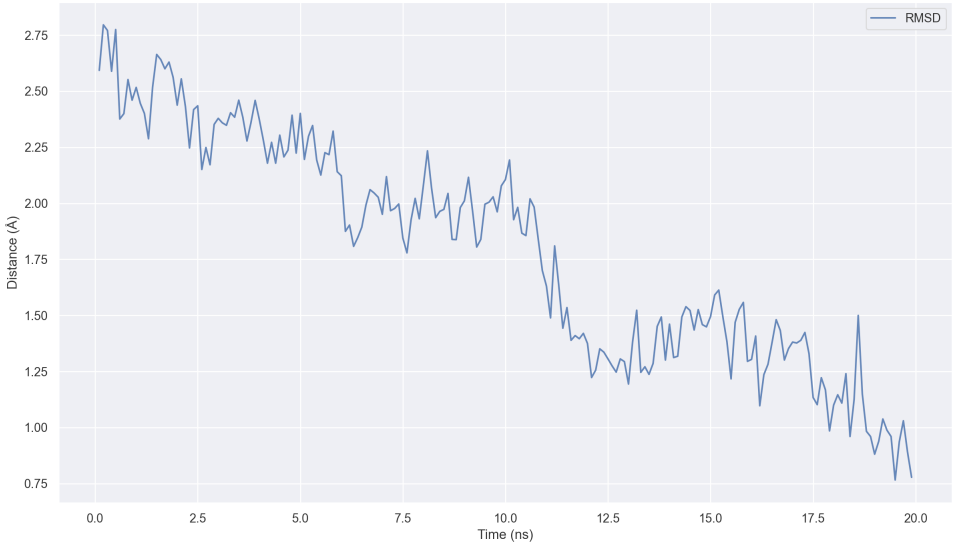


Figure 10.98: Backbone RMSD for the last 20 ns out of 40 ns for the GA1-b10x845-5fcm equilibrated structure with 1-4 HB fixed. Around the last 2 ns are fully converged.

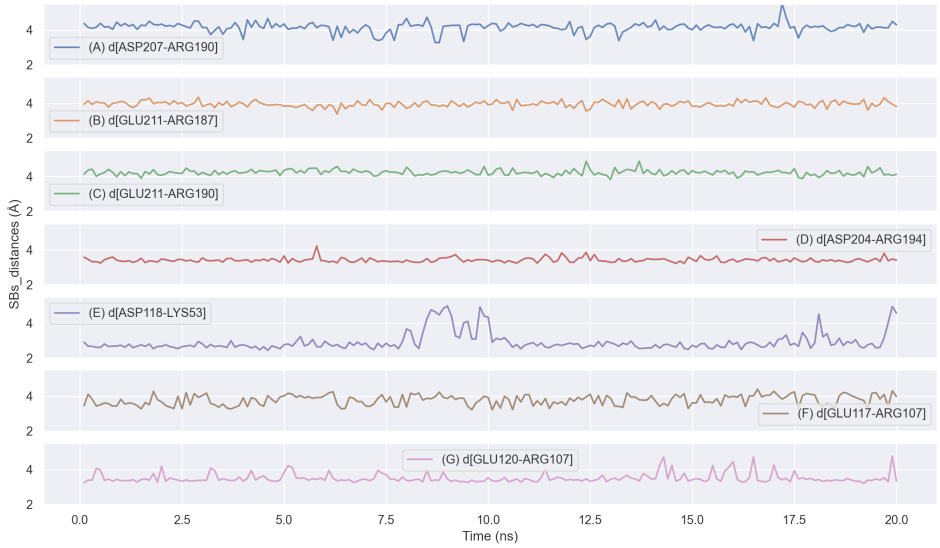


Figure 10.99: Time evolution for the equilibrated GA1-b10x845-5fcm. This snapshot shows the stable salt bridges distances for the last 20 ns of a total of 40 ns MD simulation D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74), D118(4.40)-K53(2.41), E117(4.39)-R107(3.52), and E120(4.42)-R107(3.52). Where $d[i-j]$, (Å) is the distance of stable interhelical salt bridges between residues $i$ and $j$. We consider that stable SBs are less than 5 Å.


Figure 10.100: Time evolution for the equilibrated GA1-b10x845-5fcm. This snapshot shows the transient salt bridges distances for the last 20 ns of a total of 40 ns MD simulation D201(ICL3)-R190(5.70), D201(ICL3)-R193(5.73), and E117(4.39)-K111(ICL2). Where $d[i-j]$, (Å) is the distance of stable interhelical salt bridges between residues $i$ and $j$. We consider that stable SBs are less than 5 Å.

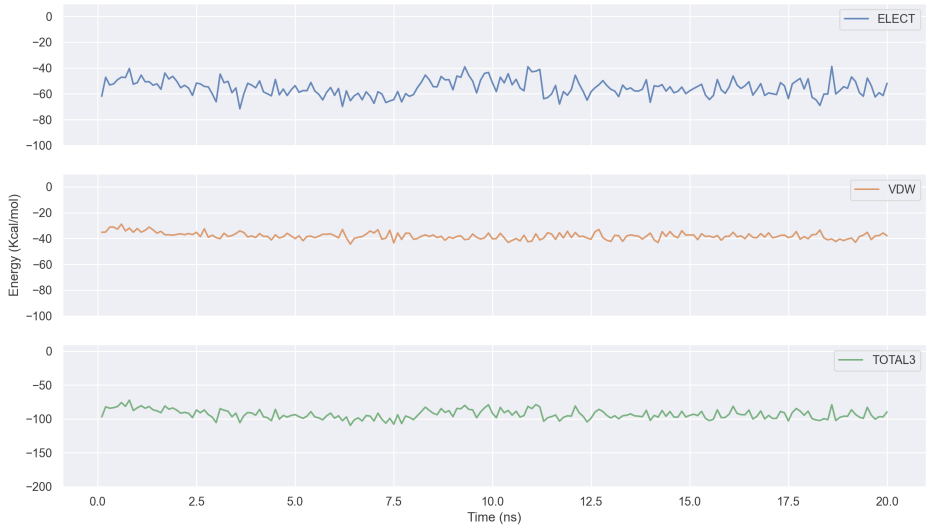


Figure 10.101: Shows the last 20 ns of 40 ns molecular dynamics simulation, for the GA1-b10x845-5c structure. The electrostatic, Van der Waals and total energies of the ligand-protein interaction reflect the stability of the complex during the 20 ns. The final electrostatic energy is around -40 Kcal/mol and -60 Kcal/mol, while the VdW energy is close to -40 Kcal/mol. On the other hand, the total energy around -90 Kcal/mol, which is a little bit higher to that for the previous structure, is perhaps due to the lack of the electrostatic contribution of the SB interaction between the carboxylate group on GA1, and the residue HPS88 of GCR1.

**GA1-b1x6914-10fcm**


Figure 10.102: Backbone RMSD for the last 20 ns out of 40 ns for the GA1-b1x6914-10c equilibrated structure with 1-4 HB fixed. Around the last 3 ns are fully converged.

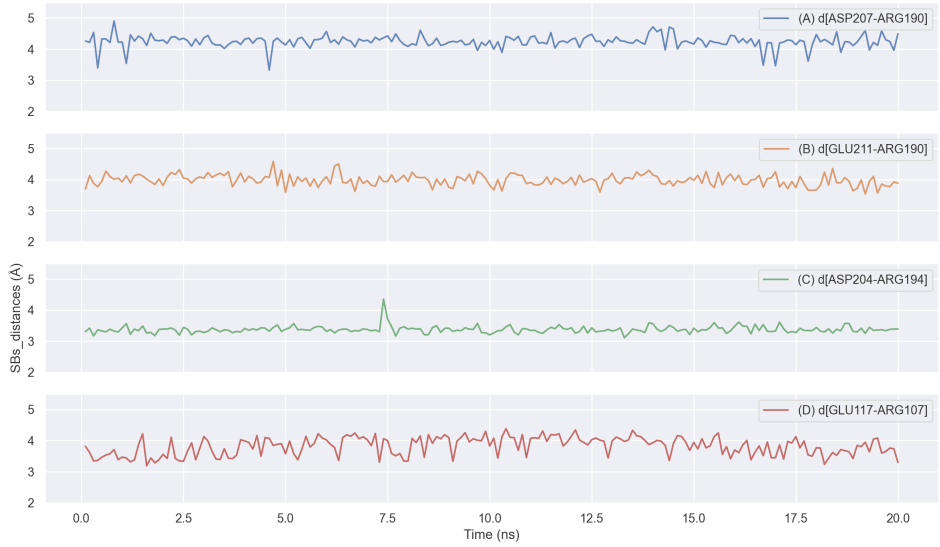


Figure 10.103: Time evolution for the equilibrated GA1-b1x6914-10c. This snapshot shows the stable salt bridges distances for the last 20 ns of a total of 40ns MD simulations D207(6.26)-R190(5.70), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74), E117(4.39)-R107(3.52). Where $d[i-j]$, (Å) is the distance of stable interhelical salt bridges between residues $i$ and $j$. We consider that stable SBs are less than 5 Å.

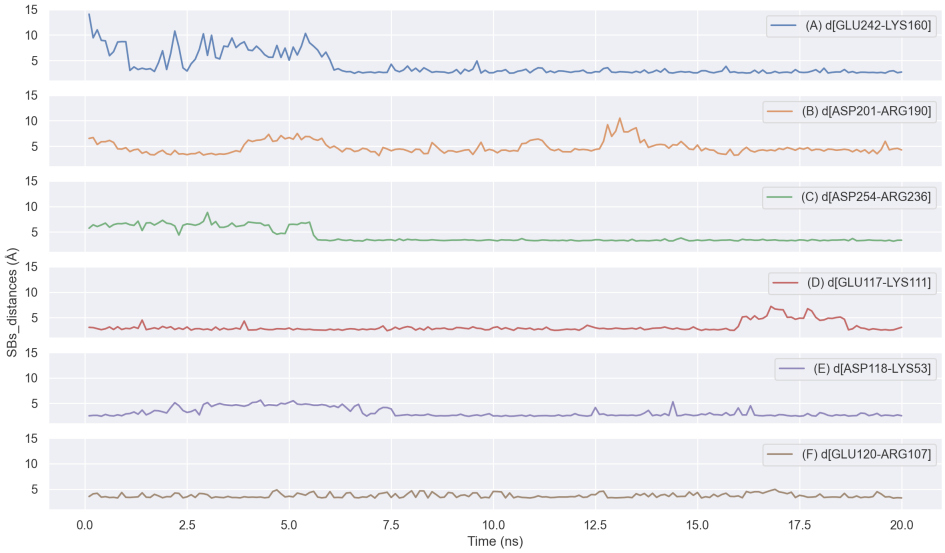


Figure 10.104: Time evolution for the equilibrated GA1-b1x6914-10c. This snapshot shows the transient salt bridges distances for the last 20 ns of a total of 40ns MD simulation E242(ECL2)-K160(ECL2), D201(ICL3)-R190(5.70), D254(7.38)-R236(ECL3), E117(4.39)-K111(ICL2), D118(4.40)-K53(2.41), and E120(4.42)-R107(3.52). Where $d[i-j]$, (Å) is the distance of stable interhelical salt bridges between residues $i$ and $j$ of. We consider that stable SBs are less than 5 Å.

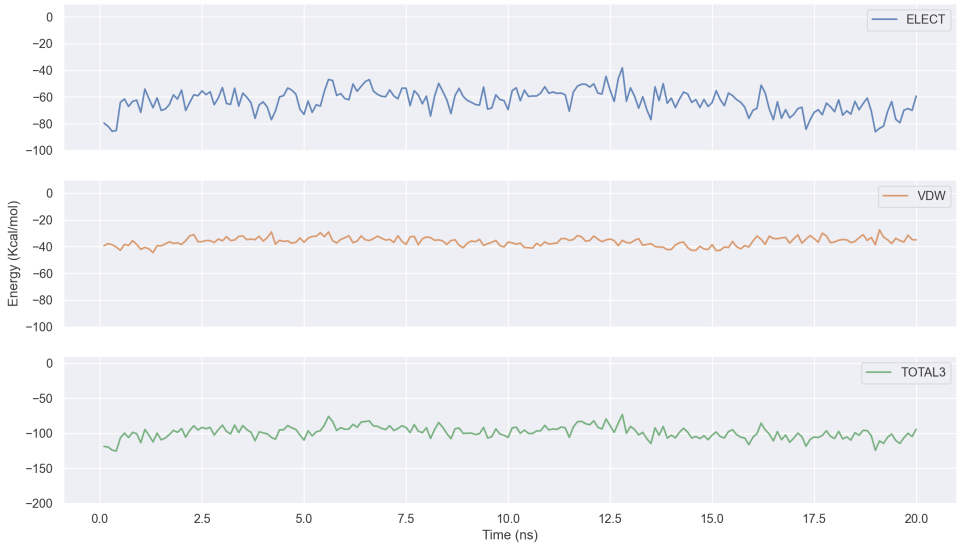


Figure 10.105: Shows the last 20 ns of 40 ns molecular dynamics simulation, for the GA1-b1x6914-10c structure. The electrostatic, Van der Waals and total energies of the ligand-protein interaction reflect the stability of the complex during the last 20 ns. As can be seen, the last 3 ns show that the electrostatic energy is between -60 Kcal/mol and -80 Kcal/mol, while the VdW energy is close to -40 Kcal/mol. On the other hand, the total energy is around -100 Kcal/mol, which is a little bit lower than the previous structure but higher than the first structure, which also might be due to the lack of the electrostatic contribution of the SB interaction between the carboxylate group on GA1, and the residue HPS88 of GCR1.

## 10.13 Appendix XIII: Selection of GCR1-ABA docked structures.

return to main text

**ABA-b9-5-20724-1c**

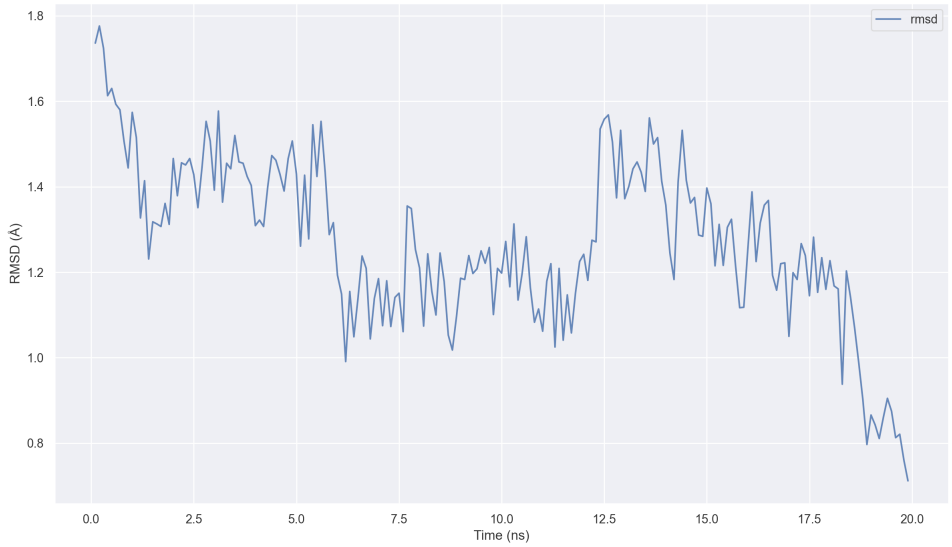


Figure 10.106: Backbone RMSD for the last 20 ns out of 60 ns for the ABA-b9-5-20724-1c equilibrated structure with 1-4 HB fixed. Around the last 2 ns are fully converged.

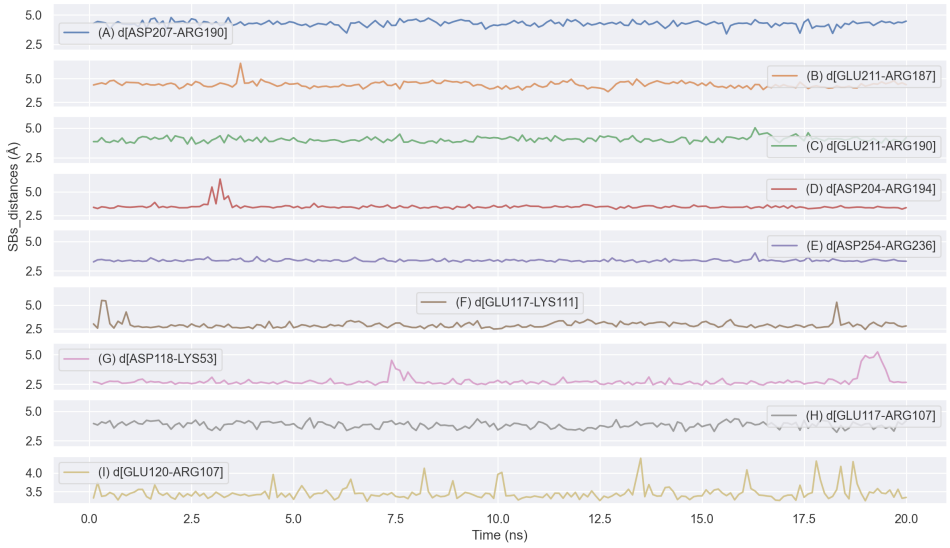


Figure 10.107: Time evolution for the equilibrated ABA-b9-5-20724-1c structure. Panels (A) to (I) show SBs found in the GCR1 region. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74), D254(7.38)-R236(ECL3), E117(4.39)-K111(ICL2), D118(4.40)-K53(2.41), E117(4.39)-R107(3.52), and E120(4.42)-R107(3.52). We consider that stable SBs are less than 5 Å.

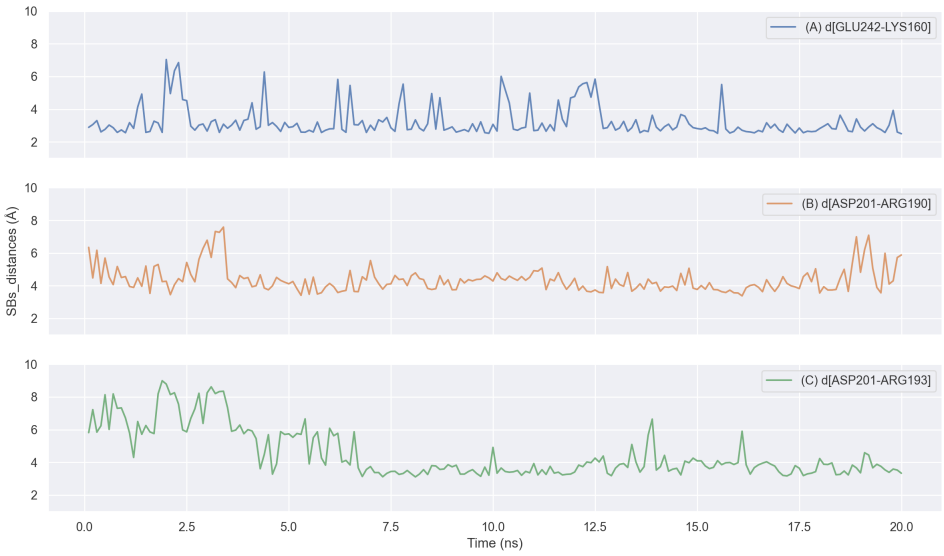


Figure 10.108: Time evolution for the equilibrated ABA-b9-5-20724-1c structure. Panels (A) to (C) show SBs found into the GCR1 region. Where $d[i-j]$, (Å) is the distance for transient interhelical salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: E242(ECL3)-K160(ECL2), D201(ICL3)-R190(5.70), and D201(ICL3)-R193(5.73). We consider that the transient SBs break and form constantly during the time evolution of the dynamics but are around 4.5 Å.

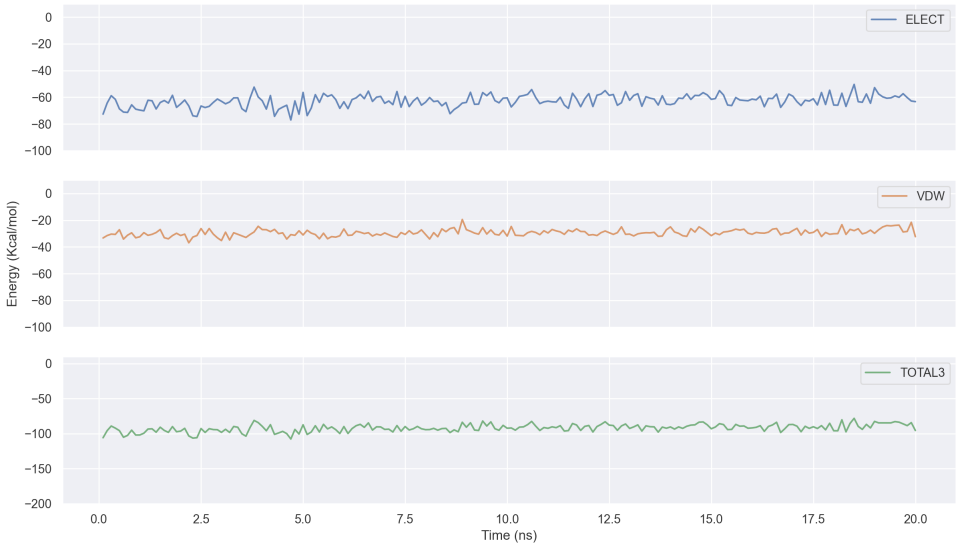


Figure 10.109: Ligand-protein energy. Electrostatic, Van der Waals and total energies for the ligand-protein interaction reflect the stability of the complex during the last 20 ns of 60,ns molecular dynamics simulation, for the ABA-b9-5-20724-1c structure.

**ABA-b9-5-20724-9c**

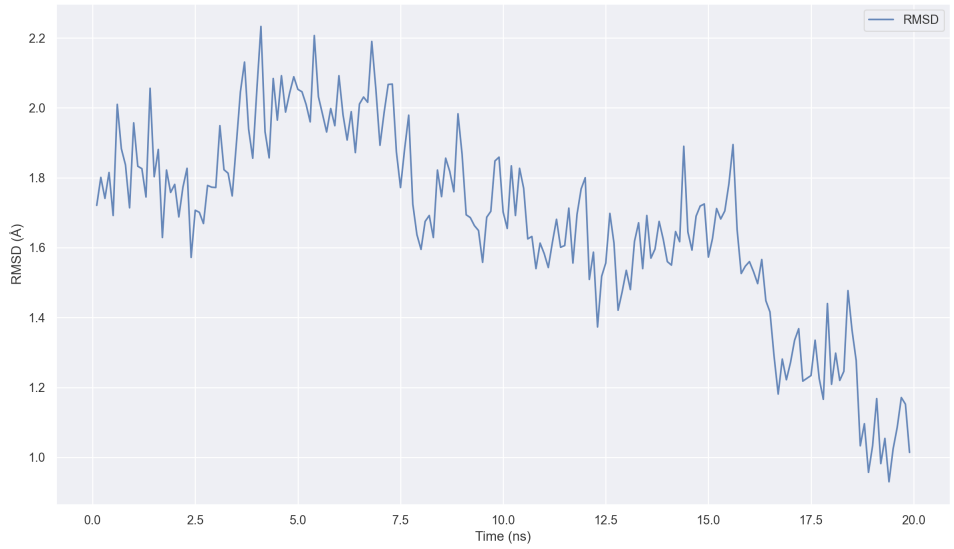


Figure 10.110: Backbone RMSD for the last 20 ns out of 60 ns for the ABA-b9-5-20724-9c equilibrated structure with 1-4 HB fixed. Around the last 2 ns are fully converged.

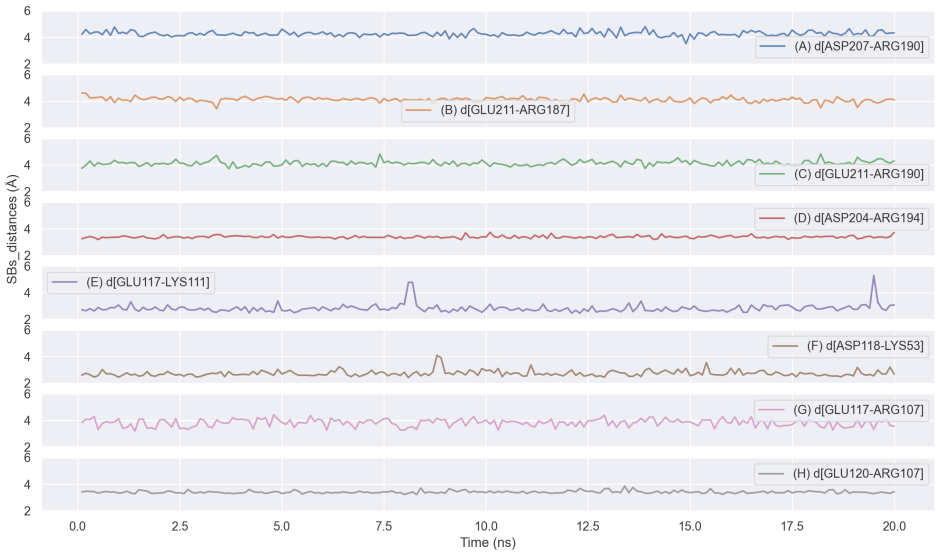


Figure 10.111: Time evolution for the equilibrated ABA-b9-5-20724-9c structure. Panels (A) to (H) show SBs found in the GCR1 region. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74), E117(4.39)-K111(ICL2), D118(4.40)-K53(2.41), E117(4.39)-R107(3.52), and E120(4.42)-R107(3.52). We consider that stable SBs are less than 5 Å.


Figure 10.112: Time evolution for the equilibrated ABA-b9-5-20724-9c structure. Panels (A) to (D) show SBs found in the GCR1 region. Where $d[i-j]$, (Å) is the distance for transient interhelical salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: E242(ECL3)-K160(ECL2), D254(7.38)-HSP88(3.33), D201(ICL3)-R190(5.70), and D201(ICL3)-R193(5.73). We consider that the transient SBs break and form constantly during the time evolution of the dynamics but are around 4.5 Å.

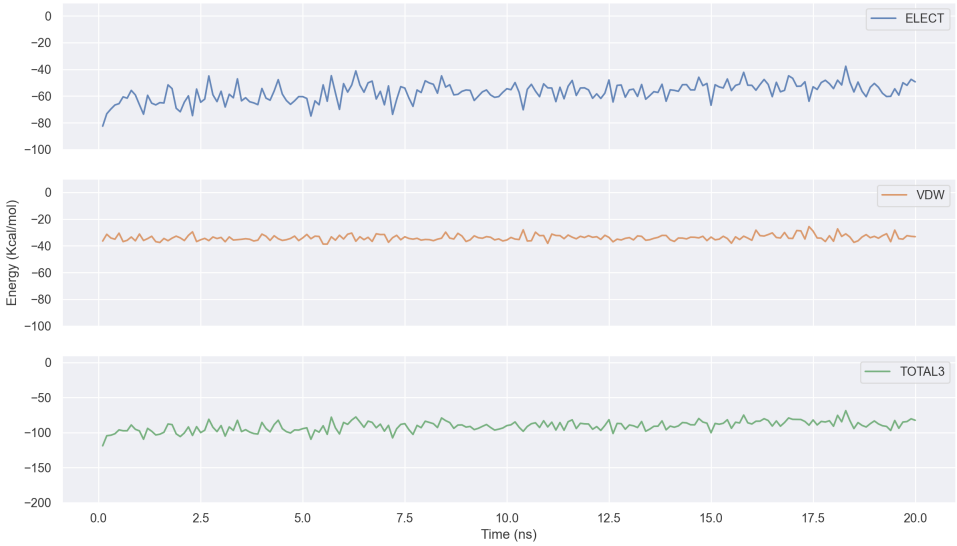


Figure 10.113: Ligand-protein energy. Electrostatic, Van der Waals, and total energies for the ligand-protein interaction reflect the stability of the complex during the last 20 ns of 60,ns molecular dynamics simulation, for the ABA-b9-5-20724-9c structure.

**ABA-b4-0-24315-8c**

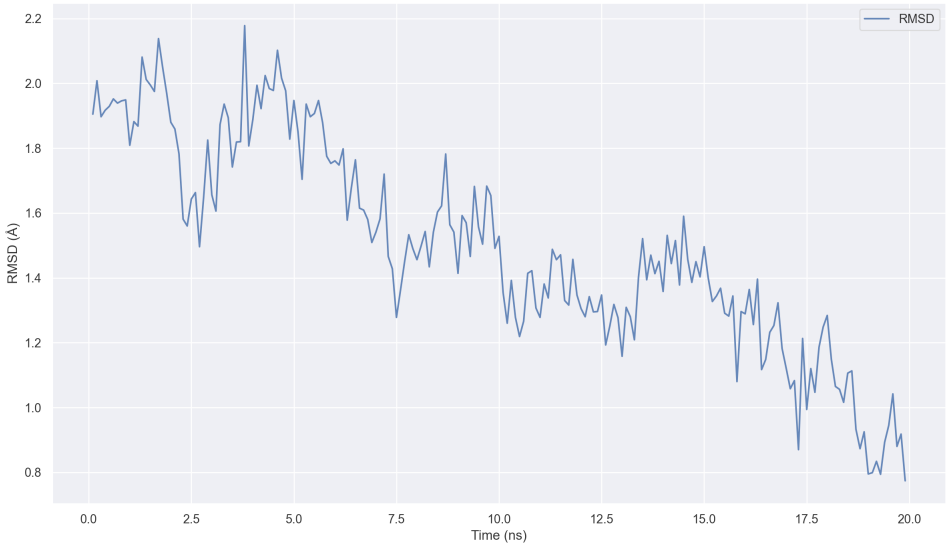


Figure 10.114: Backbone RMSD for the last 20 ns out of 40 ns for the ABA-b4-0-24315-8c equilibrated structure with 1-4 HB fixed. Around the last 2 ns are fully converged.

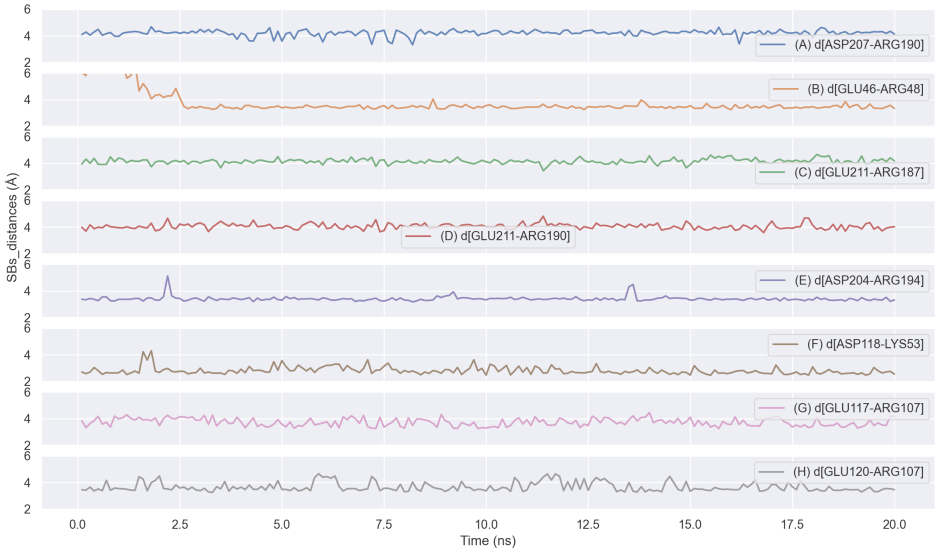


Figure 10.115: Time evolution for the equilibrated ABA-b4-0-24315-8c structure. Panels (A) to (H) show SBs found in the GCR1 region. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 40 ns MD simulation. B-W numbering is used to describe each residue: D207(6.26)-R190(5.70), E46(ICL1)-R48(ICL1), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74), D118(4.40)-K53(2.41), E117(4.39)-R107(3.52), and E120(4.42)-R107(3.52). We consider that stable SBs are less than 5 Å.

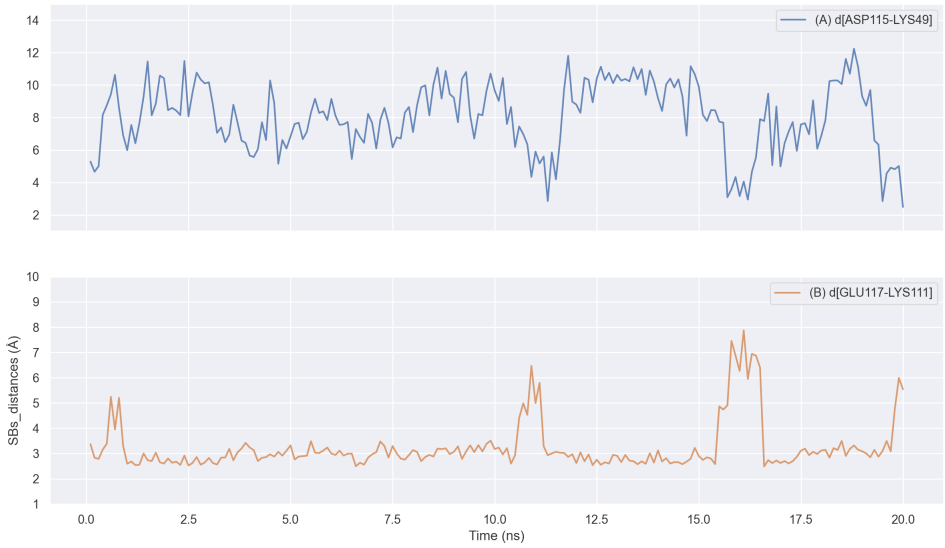


Figure 10.116: Time evolution for the equilibrated ABA-b4-0-24315-8c structure. Panels (A) and (B) show SBs found in the GCR1 region. Where $d[i-j]$, (Å) is the distance for transient interhelical salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 40 ns MD simulation. B-W numbering is used to describe each residue: D115(ICL2)-K49(ICL1) and E117(4.39)-K111(ICL2). We consider that the transient SBs break and form constantly during the time evolution of the dynamics but are around 4.5 Å.

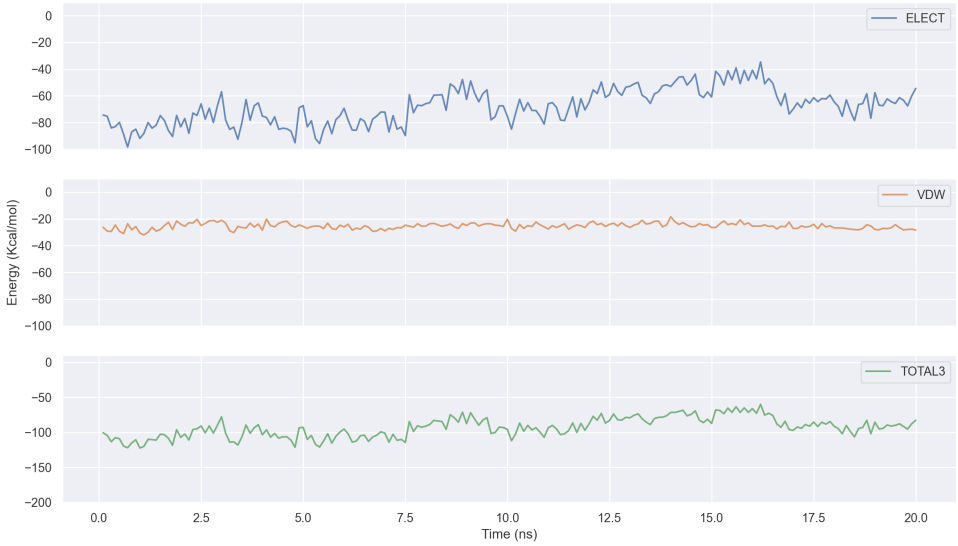


Figure 10.117: Ligand-protein energy. Electrostatic, Van der Waals, and total energies for the ligand-protein interaction reflect the stability of the complex during the last 20 ns of 40 ns molecular dynamics simulation, for the ABA-b4-0-24315-8c structure.

**ABA-b5-5-31387-10c**

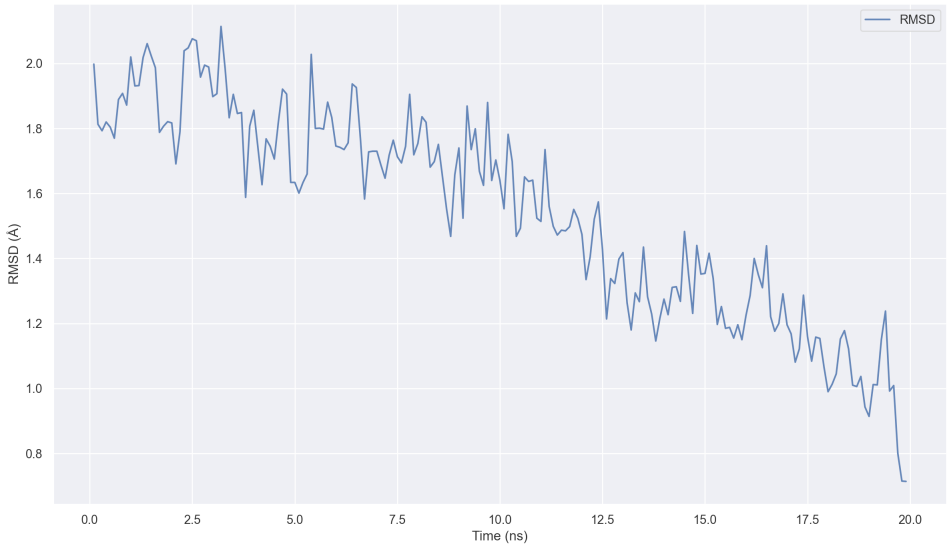


Figure 10.118: Backbone RMSD for the last 20 ns out of 60 ns for the ABA-b5-5-31387-10c equilibrated structure with 1-4 HB fixed. Around last 3 ns are fully converged.

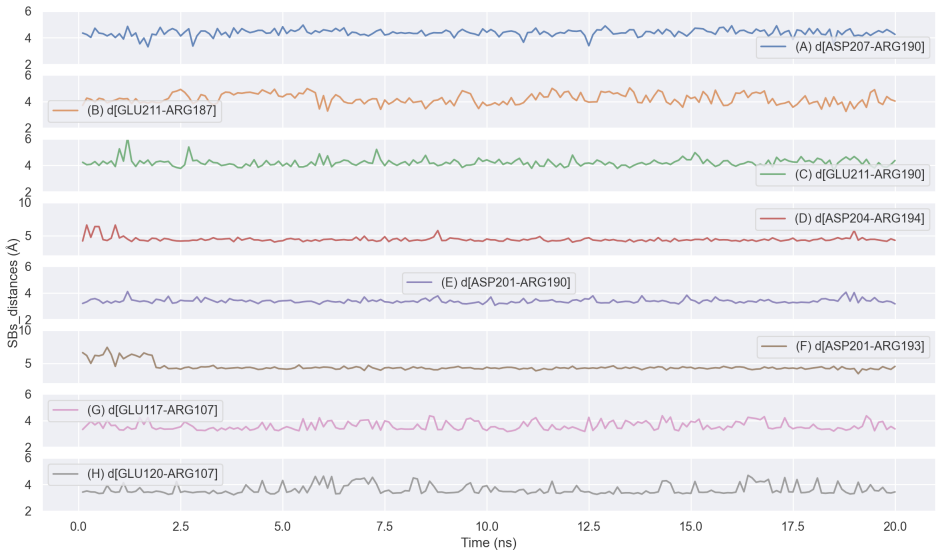


Figure 10.119: Time evolution for the equilibrated ABA-b5-5-31387-10c structure. Panels (A) to (I) show SBs found in the GCR1 region. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74), D201(ICL3)-R190(5.70), D201(ICL3)-R193(5.73), E117(4.39)-R107(3.52), and E120(4.42)-R107(3.52). We consider that stable SBs are less than 5 Å.

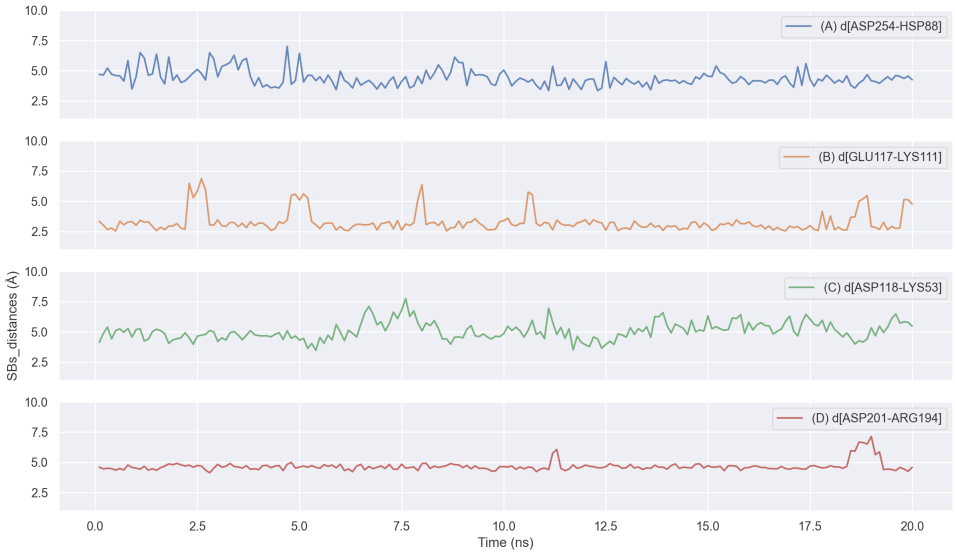


Figure 10.120: Time evolution for the equilibrated ABA-b5-5-31387-10c structure. Panels (A) to (D) show SBs found in the GCR1 region. Where $d[i-j]$, (Å) is the distance for transient interhelical salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: D254(7.38)-HSP88(3.33), E117(4.39)-K111(ICL2), D118(4.40)-K53(2.41), and D201(ICL3)-R194(5.74). We consider that the transient SBs break and form constantly during the time evolution of the dynamics but are around 4.5 Å.

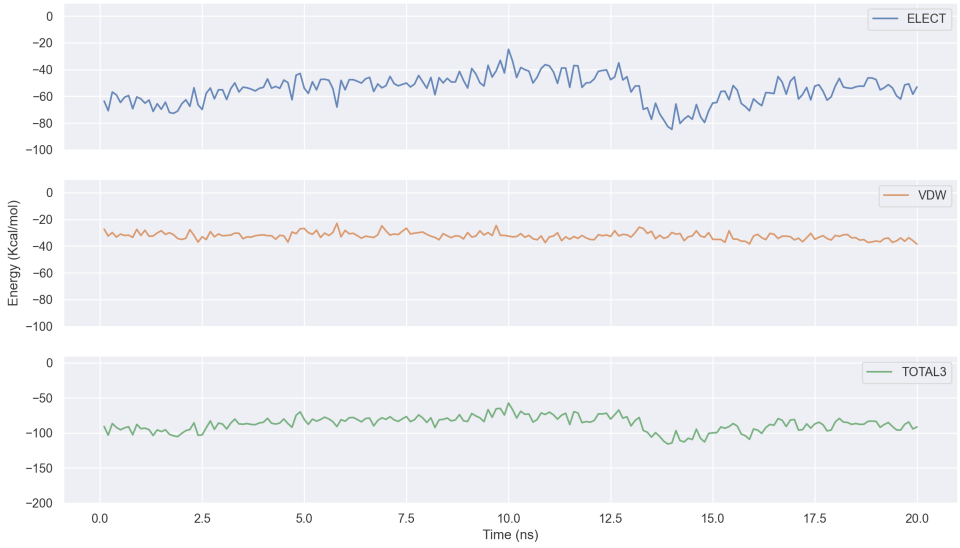


Figure 10.121: Ligand-protein energy. Electrostatic, Van der Waals, and total energies for the ligand-protein interaction reflect the stability of the complex during the last 20 ns of 60 ns molecular dynamics simulation, for the ABA-b5-5-31387-10c structure.

## 10.14 Appendix XIV: Proposed GCR1-GPA1 Signaling mechanism, GA1-mediated

return to main text

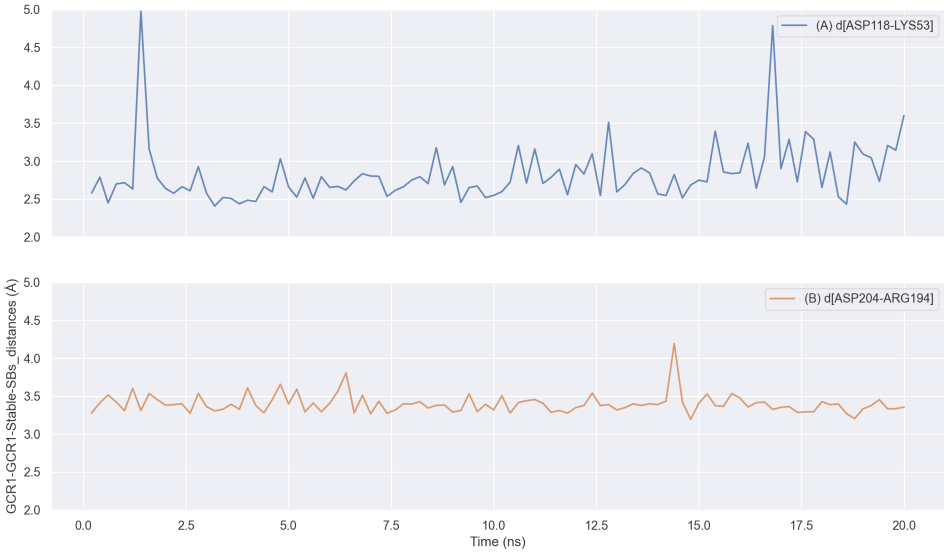


Figure 10.122: GCR1-GCR1 SBs in GA1-GCR1-GPA1. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: D204(6.23)-R194(5.74), D118(4.40)-K53(2.41). We consider that stable SBs are less than 4 Å.

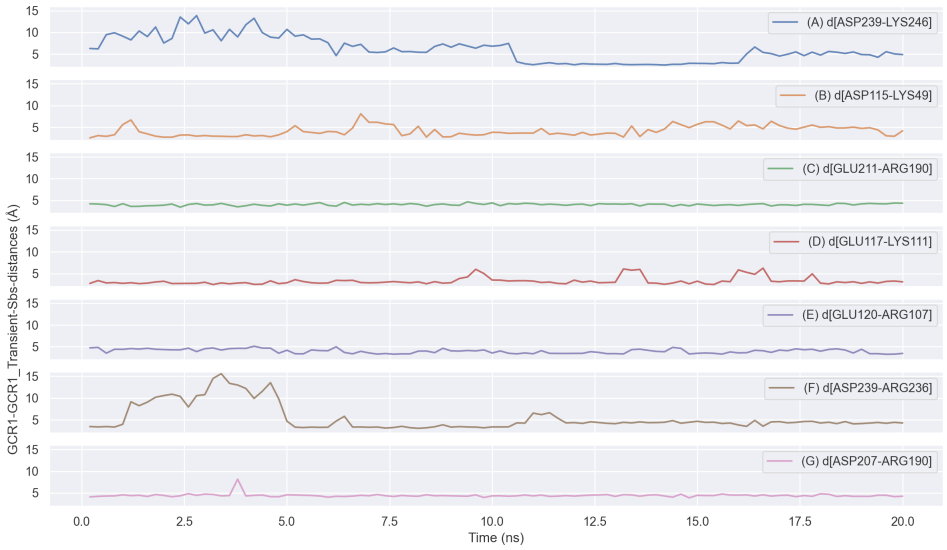


Figure 10.123: GCR1-GCR1 stable SBs in GA1-GCR1-GPA1. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74), E117(4.39)-K111(ICL2), D118(4.40)-K53(2.41), E117(4.39)-R107(3.52), and E120(4.42)-R107(3.52). We consider that stable SBs are less than 5 Å.


Figure 10.124: GCR1-GPA1 SBs in GA1-GCR1-GPA1. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74), E117(4.39)-K111(ICL2), D118(4.40)-K53(2.41), E117(4.39)-R107(3.52), and E120(4.42)-R107(3.52). We consider that stable SBs are less than 5 Å.

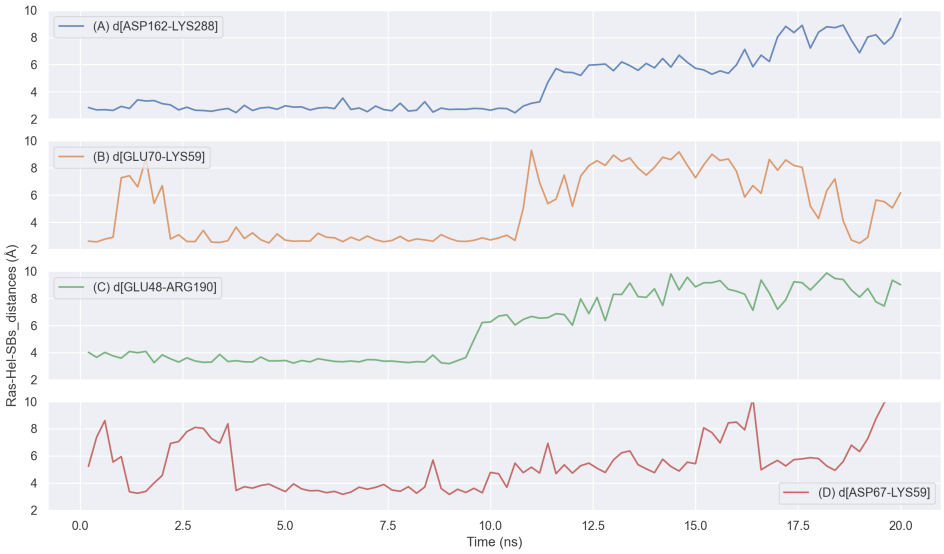


Figure 10.125: Time evolution for Ras-Hel SBs in GA1-GCR1-GPA1. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74), E117(4.39)-K111(ICL2), D118(4.40)-K53(2.41), E117(4.39)-R107(3.52), and E120(4.42)-R107(3.52). We consider that stable SBs are less than 5 Å.

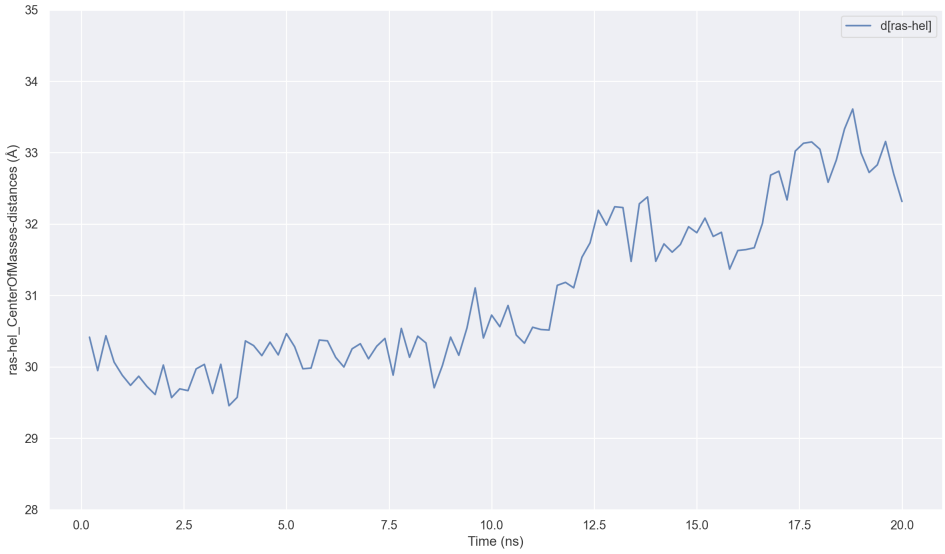


Figure 10.126: Time evolution for Center of mass of Ras-Hel SBs in GA1-GCR1-GPA1. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. B-W numbering is used to describe each residue: D207(6.26)-R190(5.70), E211(6.30)-R187(5.67), E211(6.30)-R190(5.70), D204(6.23)-R194(5.74), E117(4.39)-K111(ICL2), D118(4.40)-K53(2.41), E117(4.39)-R107(3.52), and E120(4.42)-R107(3.52). We consider that stable SBs are less than 5 Å.

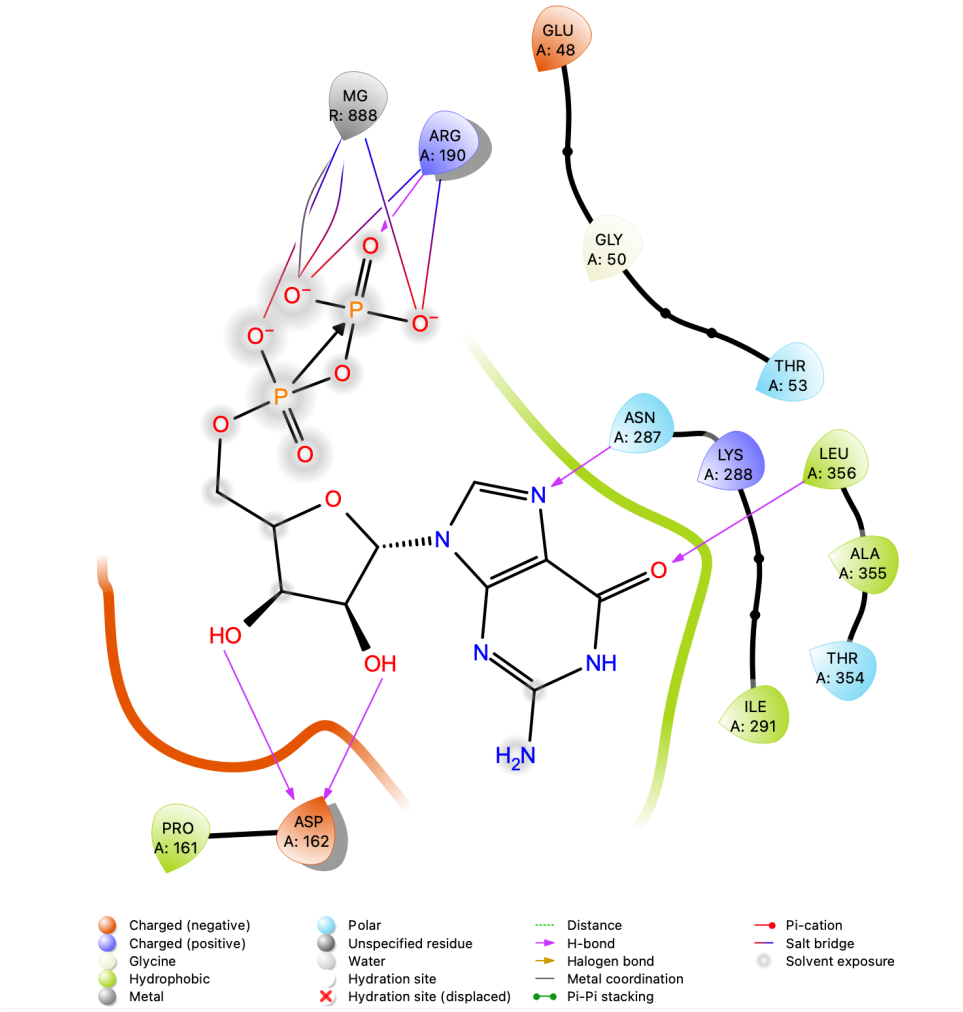


Figure 10.127: Ligand interaction diagram shows the last GDP-Mg ligand interaction in the GA1-GCR1-GPA1 structure minimized and equilibrated in 73 ns of MD simulation. Likewise, a series of polar and hydrophobic interactions are shown with residues located at 4 Å away from ABA. For the sake of clarity, only polar hydrogen atoms are shown on the ligand.

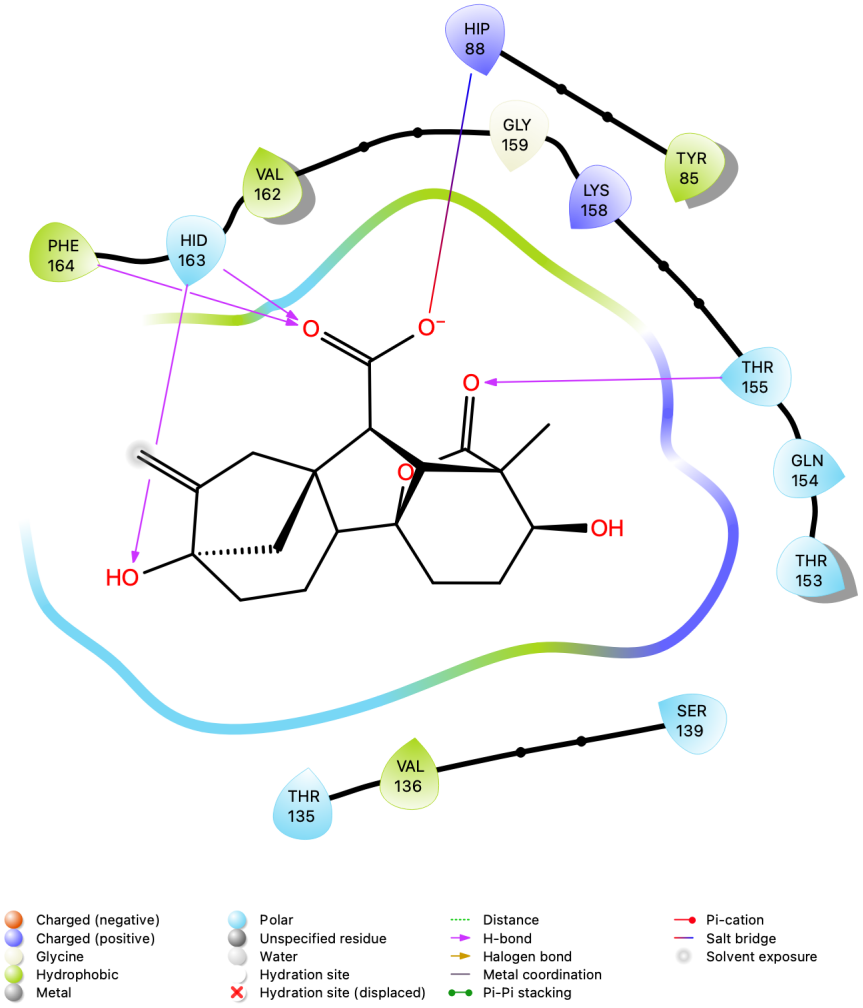


Figure 10.128: Ligand interaction diagram shows the last GDP-Mg ligand interaction in the GA1-GCR11-GPA1 structure minimized and equilibrated in 73 ns of MD simulation. Here we can see a carboxylate group on GA1 making part of a salt bridge with the HSP88-residue of GCR1 inside the predicted binding site. Likewise, a series of polar and hydrophobic interactions are shown with residues located at 4 Å away from ABA. For the sake of clarity, only polar hydrogen atoms are shown on the ligand.

## 10.15 Appendix XV: Proposed GCR1-GPA1 Signaling mechanism, ABA-mediated

return to main text


Figure 10.129: Time evolution for Ras-Hel SBs in ABA-GCR1-GPA1. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. We consider that stable SBs are less than 4 Å.

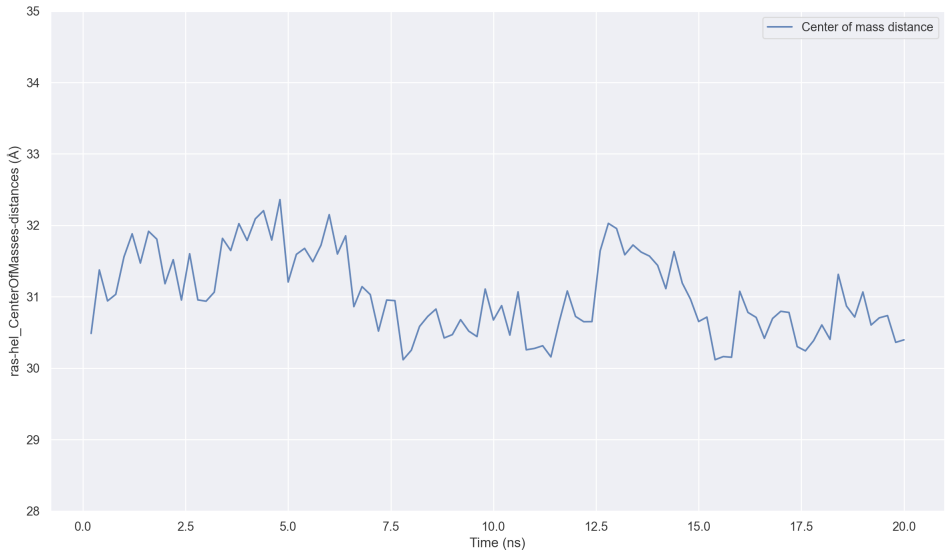


Figure 10.130: Time evolution for Center of mass of Ras-Hel SBs in ABA-GCR1-GPA1. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation.

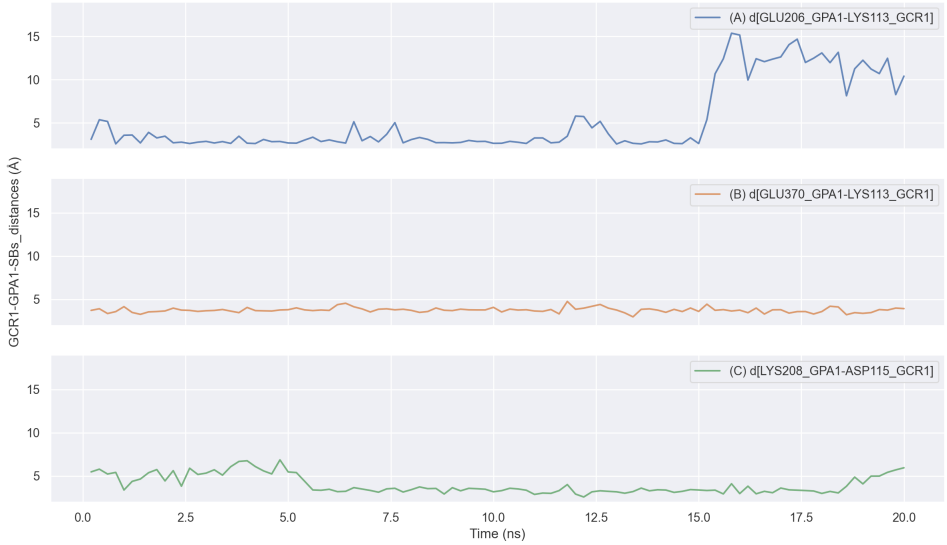


Figure 10.131: Time evolution for GCR1-GPA1 SBs in ABA-GCR1-GPA1. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. We consider that stable SBs are less than 4 Å.

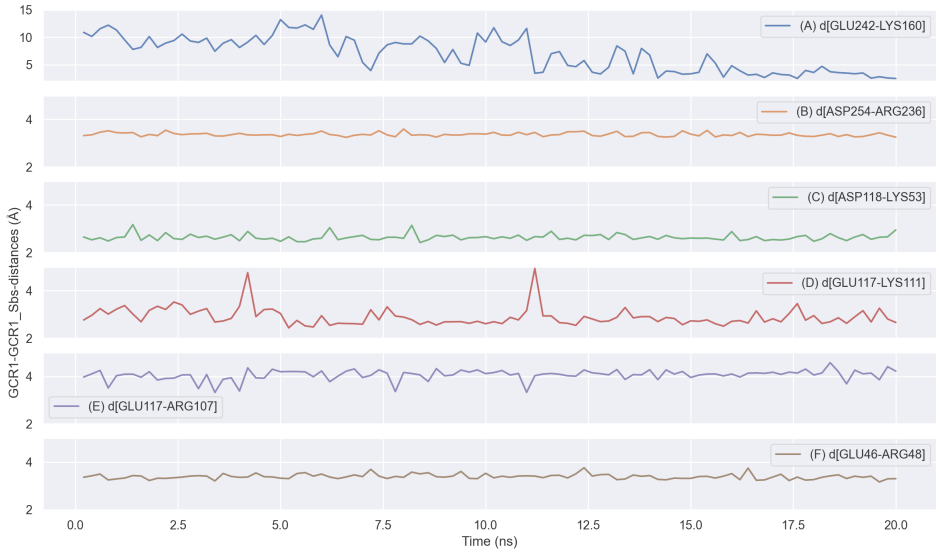


Figure 10.132: Time evolution for GCR1-GCR1 SBs in ABA-GCR1-GPA1. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. We consider that stable SBs are less than 4 Å.

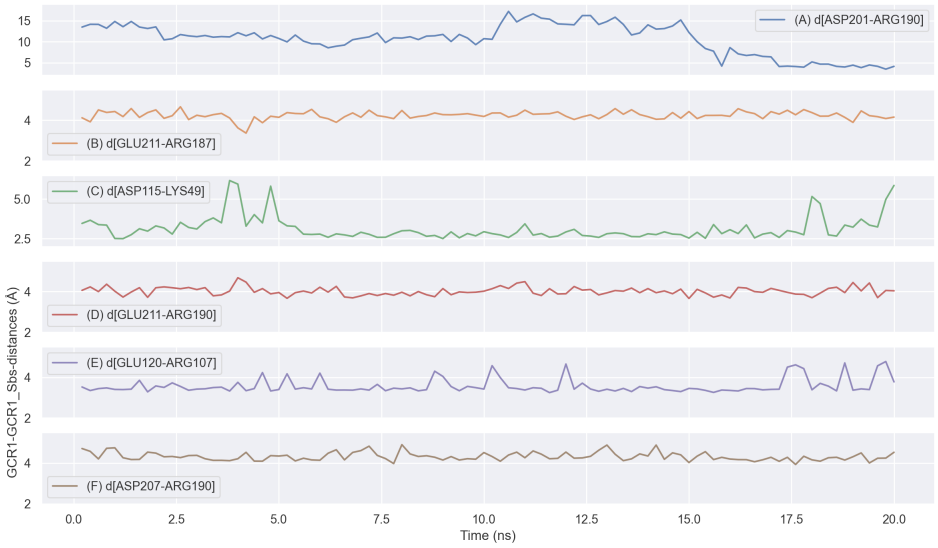


Figure 10.133: Time evolution for GCR1-GCR1 SBs in ABA-GCR1-GPA1. Where $d[i-j]$, (Å) is the distance for stable salt bridges between residues $i$ and $j$ during the last 20 ns of a total of 60 ns MD simulation. We consider that stable SBs are less than 4 Å.

# Bibliography


[1] B Taddese, G J G Upton, G R Bailey, S R D Jordan, N Y Abdulla, P J Reeves, and C A Reynolds. Do plants contain G Protein-Coupled receptors? *Plant Physiol.*, 164(1):287–307, 2014.

[2] Alphafold protein structure database: G-protein coupled receptor 1. http:/https://alphafold.ebi.ac.uk/entry/O04714. Accessed: 2022-08-01.

[3] William A. Goddard, Soo Kyung Kim, Youyong Li, Bartosz Trzaskowski, Adam R. Griffith, and Ravinder Abrol. Predicted 3D structures for adenosine receptors bound to ligands: Comparison to the crystal structure. *Journal of Structural Biology*, 170(1):10–20, 2010.

[4] Ruth R. Finkelstein, Srinivas S. L. Gampala, and Christopher D. Rock. Abscisic Acid Signaling in Seeds and Seedlings. *The Plant Cell*, 14(suppl 1):S15–S45, 2017.

[5] Zhen-Ming Pei, Kazuyuki Kuchitsu, John M Ward, Martin Schwarz, and Julian I Schroeder. Differential abscisic acid regulation of guard cell slow anion channels in arabidopsis Wild-Type and abi1 and abi2 mutants. *Plant Cell*, 9(3):409, 2007.

[6] S M Assmann. G protein signaling in the regulation of rice seed germination. *Sci. Signal.*, 2005(310):12, 2005.

[7] J-G Chen. GCR1 can act independently of heterotrimeric G-Protein in response to brassinosteroids and gibberellins in arabidopsis seed germination. *Plant Physiol.*, 2004.

[8] Richard Hooley. Plant hormone perception and action: a role for g–protein signal transduction? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1374):1425–1430, 1998.

[9] Navjyoti Chakraborty, Priyanka Sharma, Kostya Kanyuka, Ravi R. Pathak, Devapriya Choudhury, Richard A. Hooley, and Nandula Raghuram. Transcriptome analysis of arabidopsis GCR1 mutant reveals its roles in stress, hormones, secondary metabolism and phosphate starvation. *PLoS ONE*, 10(2):1–17, 2015.

[10] Amirhossein Mafi, Soo Kyung Kim, and William A. Goddard. The mechanism for ligand activation of the GPCR-G protein complex. *Proceedings of the National Academy of Sciences of the United States of America*, 119(18), 2022.

[11] B. Taddese, G. J. G. Upton, G. R. Bailey, S. R. D. Jordan, N. Y. Abdulla, P. J. Reeves, and C. A. Reynolds. Do Plants Contain G Protein-Coupled Receptors? *Plant Physiology*, 164(1):287–307, 2014.

[12] Janice C. Jones, Jeffrey W. Duffy, Mischa Machius, Brenda R.S. Temple, Henrik G. Dohlman, and Alan M. Jones. The crystal structure of a self-activating g protein $\alpha$ subunit reveals its distinct mechanism of signal initiation. *Science Signaling*, 4(159):1–8, 2011.

[13] Andrew Leach. *Molecular Modelling: Principles and Applications*. 2001.

[14] Mark Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. 2010.

[15] Andres Jaramillo-Botero, Robert Nielsen, Ravi Abrol, Julius Su, Tod Pascal, Jonathan Mueller, and William A. Goddard III. First-principles-based multiscale, multiparadigm molecular mechanics and dynamics methods for describing complex chemical processes. *Top Curr Chem*, 307:1–42, 2012.

[16] James C Phillips, Rosemary Braun, W E I Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, and Henri Poincare. Scalable Molecular Dynamics with NAMD. 2005.

[17] Alessandro Laio and Francesco L Gervasio. Metadynamics : a method to simulate rare events and reconstruct the free energy in biophysics , chemistry and material science. 2008.

[18] Martin Rodbell and Alfred North Whitehead. Signal Transduction: Evolution of. (December):338–345, 1994.

[19] Ravinder Abrol, Soo Kyung Kim, Jenelle K. Bray, Bartosz Trzaskowski, and William A. Goddard. Conformational ensemble view of G protein-coupled receptors and the effect of mutations and ligand binding. *Methods in Enzymology*, 520:31–48, 2013.

[20] Gabriella Colucci, Fabio Apone, Nicole Alyeshmerni, Derek Chalmers, and Maarten J Chrispeels. GCR1, the putative Arabidopsis G protein-coupled receptor gene is cell cycle-regulated, and its overexpression abolishes seed dormancy and shortens time to flowering. *Proceedings of the National Academy of Sciences*, 99(7):4736–4741, 2002.

[21] Fang Liu, Ziriu Bian, Zhenhua Jia, Qian Zhao, and Shuishan Song. The GCR1 and GPA1 Participate in Promotion of Arabidopsis Primary Root Elongation Induced by N -Acyl-Homoserine Lactones, the Bacterial Quorum-Sensing Signals. *Molecular Plant-Microbe Interactions*, 25(5):677–683, 2012.

[22] Laetitia Perfus-Barbeoch, Alan M. Jones, and Sarah M. Assmann. Plant heterotrimeric G protein function: Insights from Arabidopsis and rice mutants. *Current Opinion in Plant Biology*, 7(6):719–731, 2004.

[23] Navjyoti Chakraborty, Navneet Singh, Kanwaljeet Kaur, and Nandula Raghuram. G-protein Signaling Components GCR1 and GPA1 Mediate Responses to Multiple Abiotic Stresses in Arabidopsis. *Frontiers in Plant Science*, 6(November):1–10, 2015.

[24] Ullah H; Chen JG; Wang S; Jones AM. Role of a Heterotrimeric G Protein in Regulation of Arabidopsis Seed Germination. *Plant Physiology*, 129(2):897–907, 2002.

[25] H. Ullah. The beta-Subunit of the Arabidopsis G Protein Negatively Regulates Auxin-Induced Cell Division and Affects Multiple Developmental Processes. *the Plant Cell Online*, 15(2):393–409, 2003.

[26] Navjyoti Chakraborty, Priyanka Sharma, Kostya Kanyuka, Ravi Ramesh Pathak, Devapriya Choudhury, Richard Hooley, and Nandula Raghuram. G-protein $\alpha$-subunit (GPA1) regulates stress, nitrate and phosphate response, flavonoid biosynthesis, fruit/seed development and substantially shares GCR1 regulation in A. thaliana. *Plant Molecular Biology*, 89(6):559–576, 2015.

[27] Ting-Ying Wu and Daisuke Urano. Genetic and Systematic Approaches Toward G Protein-Coupled Abiotic Stress Signaling in Plants. *Frontiers in Plant Science*, 9(September), 2018.

[28] Girish Mishra, Wenhua Zhang, Fan Deng, Jian Zhao, and Xuemin Wang. A bifurcating pathway directs abscisic acid effects on stomatal closure and opening in Arabidopsis. *Science*, 312(5771):264–266, 2006.

[29] D. Urano, J.-G. Chen, J. R. Botella, and A. M. Jones. Heterotrimeric G protein signalling in the plant kingdom. *Open Biology*, 3(3):120186–120186, 2013.

[30] Christopher A. Johnston, J Philip Taylor, Yajun Gao, Adam J Kimple, Jeffrey C Grigston, Jin-Gui Chen, David P Siderovski, Alan M. Jones, and Francis S Willard. GTPase acceleration as the rate-limiting step in Arabidopsis G protein-coupled sugar signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 104(44):17317–17322, 2007.

[31] Sarah M. Assmann. G protein signaling in the regulation of arabidopsis seed germination. *Science's STKE*, 2005(308):cm11–cm11, 2005.

[32] Sylvie Coursol, Liu Min Fan, Hervé Le Stunff, Sarah Splegel, Simon Gilroy, and Sarah M. Assman. Sphingolipid signalling in Arabidopsis guard cells involves heterotrimeric G proteins. *Nature*, 423(6940):651–654, 2003.

[33] Navjyoti Chakraborty, Kostya Kanyuka, Dinesh Kumar Jaiswal, and Abhineet Kumar. GCR1 and GPA1 coupling regulates nitrate , cell wall , immunity and light responses in Arabidopsis. (June 2018):1–17, 2019.

[34] Alejandro C. Colaneri and Alan M. Jones. The wiring diagram for plant G signaling, 2014.

[35] A. Cousson. Indolyl-3-butyric acid-induced Arabidopsis stomatal opening mediated by 3,5-cyclic guanosine-monophosphate. *Plant Physiology and Biochemistry*, 48(12):977–986, 2010.

[36] Qian Feng Li and Jun Xian He. Mechanisms of signaling crosstalk between brassinosteroids and gibberellins. *Plant Signaling and Behavior*, 8(7), 2013.

[37] Lars-Goran Josefsson and Lars Rask. Cloning of a putative g-protein-coupled receptor from. *DNA Seq.*, 420:415–420, 1997.

[38] Stella Plakidou-Dymock, David Dymock, and Richard Hooley. A higher plant seven-transmembrane receptor that influences sensitivity to cytokinins. *Curr. Biol.*, 8(6):315–324, 1998.

[39] Mark S Boguski, Todd M J Lowe, and Carolyn M Tolstoshev. dbEST — database for "expressed sequence tags". *Nat. Genet.*, 4(4):332–333, 1993.

[40] S Pandey. The arabidopsis putative G Protein-Coupled receptor GCR1 interacts with the G protein subunit GPA1 and regulates abscisic acid signaling. *Plant Cell*, 2004.

[41] András Fiser, Richard Kinh Gian Do, et al. Modeling of loops in protein structures. *Protein science*, 9(9):1753–1773, 2000.

[42] C D Strader, T M Fong, M R Tota, D Underwood, and R A Dixon. Structure and function of G protein-coupled receptors. *Annu. Rev. Biochem.*, 63:101–132, 1994.

[43] Christopher A Johnston, J Philip Taylor, Yajun Gao, Adam J Kimple, Jeffrey C Grigston, Jin-Gui Chen, David P Siderovski, Alan M Jones, and Francis S Willard. GTPase acceleration as the rate-limiting step in arabidopsis G protein-coupled sugar signaling. *Proc. Natl. Acad. Sci. U. S. A.*, 104(44):17317–17322, 2007.

[44] D Urano, J-G Chen, J R Botella, and A M Jones. Heterotrimeric G protein signalling in the plant kingdom. *Open Biol.*, 3(3):120186–120186, 2013.

[45] Daisuke Urano, Janice C Jones, Hao Wang, Melissa Matthews, William Bradford, Jeffrey L Bennetzen, and Alan M Jones. G protein activation without a GEF in the plant kingdom. *PLoS Genet.*, 8(6), 2012.

[46] Janice C Jones, Brenda R S Temple, Alan M Jones, and Henrik G Dohlman. Functional reconstitution of an atypical G protein heterotrimer and regulator of G protein signaling protein (RGS1) from arabidopsis thaliana. *J. Biol. Chem.*, 286(15):13143–13150, 2011.

[47] Gabriella Colucci, Fabio Apone, Nicole Alyeshmerni, Derek Chalmers, and Maarten J Chrispeels. GCR1, the putative arabidopsis G protein-coupled receptor gene is cell cycle-regulated, and its overexpression

abolishes seed dormancy and shortens time to flowering. *Proceedings of the National Academy of Sciences*, 99(7):4736–4741, 2002.

[48] Fabio Apone, Nicole Alyeshmerni, Kathryn Wiens, Derek Chalmers, Maarten J Chrispeels, and Gabriella Colucci. The G-Protein-Coupled receptor GCR1 regulates DNA synthesis through activation of Phosphatidylinositol-Specific phospholipase C. *Plant Physiol.*, 2003.

[49] Michael P Strathmann and Narasimhan Gautam. Diversity of G proteins. *Science*, 252(1971):802–808, 1991.

[50] Stefan Offermanns. In vivo functions of heterotrimeric g-proteins: studies in g$\alpha$-deficient mice. *Oncogene*, 20(13):1635, 2001.

[51] Hemayet Ullah, Jin-Gui Chen, Jeff C Young, Kyung-Hoan Im, Michael R Sussman, and Alan M Jones. Modulation of cell proliferation by heterotrimeric g protein in arabidopsis. *Science*, 292(5524):2066–2069, 2001.

[52] S M Assmann. Plant G proteins, phytohormones, and plasticity: Three questions and a speculation. *Sci. Signal.*, 2004(264):20, 2004.

[53] Teun Munnik. Phosphatidic acid: an emerging plant lipid second messenger. *Trends in plant science*, 6(5):227–233, 2001.

[54] C S D'Santos, J H Clarke, and N Divecha. Phospholipid signalling in the nucleus. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1436(1-2):201–232, 1998.

[55] Xi-Qing Wang, Hemayet Ullah, Alan M Jones, and Sarah M Assmann. G protein regulation of ion channels and abscisic acid signaling in arabidopsis guard cells. *Science*, 292(5524):2070–2072, 2001.

[56] Sylvie Coursol, Liu Min Fan, Hervé Le Stunff, Sarah Splegel, Simon Gilroy, and Sarah M Assman. Sphingolipid signalling in arabidopsis guard cells involves heterotrimeric G proteins. *Nature*, 423(6940):651–654, 2003.

[57] Tania Vivienne Humphrey and José Ramón Botella. Re-evaluation of the cytokinin receptor role of the arabidopsis gene GCR1. *J. Plant Physiol.*, 158(5):645–653, 2001.

[58] Sona Pandey, Jin-Gui Chen, Alan M Jones, and Sarah M Assmann. G-protein complex mutants are hypersensitive to abscisic acid regulation of germination and postgermination development. *Plant Physiol.*, 141(May):243–256, 2006.

[59] Ullah H; Chen JG; Wang S; Jones AM. Role of a heterotrimeric G protein in regulation of arabidopsis seed germination. *Plant Physiol.*, 129(2):897–907, 2002.

[60] J-G Chen. Differential roles of arabidopsis heterotrimeric G-Protein subunits in modulating cell division in roots. *Plant Physiol.*, 141(3):887–897, 2006.

[61] Nancy A Eckardt, Hyung-Taeg Cho, Robyn M Perrin, and Matthew R Willmann. Plant biology 2001, 2001.

[62] Katherine Mary Warpeha, Syed Salman Lateef, Yevgeniya Lapik, Marybeth Anderson, Bao-Shiang Lee, and Lon Seth Kaufman. G-protein-coupled receptor 1, g-protein g$\alpha$-subunit 1, and prephenate dehydratase 1 are required for blue light-induced production of phenylalanine in etiolated arabidopsis. *Plant physiology*, 140(3):844–855, 2006.

[63] K M Warpeha, S Upadhyay, J Yeh, J Adamiak, S I Hawkins, Y R Lapik, M B Anderson, and L S Kaufman. The GCR1, GPA1, PRN1, NF-Y signal chain mediates both blue light and abscisic acid responses in arabidopsis. *Plant Physiol.*, 143(4):1590–1600, 2007.

[64] Alan M Jones. G-protein-coupled signaling in arabidopsis. *Curr. Opin. Plant Biol.*, 5(5):402–407, 2002.

[65] S. Vohra, S. V. Chintapalli, C. J.R. Illingworth, P. J. Reeves, P. M. Mullineaux, H. S.X. Clark, M. K. Dean, G. J.G. Upton, and C. A. Reynolds. Computational studies of Family A and Family B GPCRs. *Biochemical Society Transactions*, 35(4):749–754, 2007.

[66] Fabio Apone, Nicole Alyeshmerni, Kathryn Wiens, Derek Chalmers, Maarten J Chrispeels, and Gabriella Colucci. The G-Protein-Coupled Receptor GCR1 Regulates DNA Synthesis through Activation of Phosphatidylinositol-Specific Phospholipase C. 2003.

[67] Pyeonghwa Jeong, Soo Kyung Kim, Quanjie Li, Su jin Oh, Seonil Son, Guangju Chen, Hongwei Tan, Siwon Kim, Jong Hyun Park, Ki Duk Park, Yeo Ok Kim, Myung Ha Yoon, Yong Chul Kim, and William A.

Goddard. Discovery of Novel Biased Opioid Receptor Ligands through Structure-Based Pharmacophore Virtual Screening and Experiment. *ChemMedChem*, 14(20):1783–1794, 2019.

[68] J. K. Bray, R. Abrol, W. A. Goddard, B. Trzaskowski, and C. E. Scott. SuperBiHelix method for predicting the pleiotropic ensemble of G-protein-coupled receptor conformations. *Proceedings of the National Academy of Sciences*, 111(1):E72–E78, 2014.

[69] Ravinder Abrol, Jenelle K. Bray, and William A. Goddard. Bihelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. *Proteins: Structure, Function and Bioinformatics*, 80(2):505–518, 2012.

[70] Victor Wai Tak Kam and William A. Goddard. Flat-bottom strategy for improved accuracy in protein side-chain placements. *Journal of Chemical Theory and Computation*, 4(12):2160–2169, 2008.

[71] Vaclav Cvicek, William A. Goddard, and Ravinder Abrol. *Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications*, volume 12. 2016.

[72] Stephen L. Mayo, Barry D. Olafson, and William A. Goddard. DREIDING: A generic force field for molecular simulations. *Journal of Physical Chemistry*, 94(26):8897–8909, 1990.

[73] Vishnu Shankar, William A. Goddard, Soo Kyung Kim, Ravinder Abrol, and Fan Liu. The 3D Structure of Human DP Prostaglandin G-Protein-Coupled Receptor Bound to Cyclopentanoindole Antagonist, Predicted Using the DuplexBiHelix Modification of the GEnSeMBLE Method. *Journal of Chemical Theory and Computation*, 14(3):1624–1642, mar 2018.

[74] Schrodinger. Maestro.

[75] Adam Reid Griffith. Darwindock gag - dock : Methods and applications for sma ll molecule. 2017, 2017. DARWINDOCK amp; GAG-DOCK: METHODS AND APPLICATIONS FOR SMALL MOLECULE DOCKING.

[76] Elaine C. Meng, Brian K. Shoichet, and Irwin D. Kuntz. Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry*, 13(4):505–524, 1992.

[77] T J A Ewing and I D Kuntz. Critical evaluation of search algorithms used in automated molecular docking. *Comput Appl Biosci*, 18:1175–1189, 1997.

[78] Soo Kyung Kim, Peter Fristrup, Ravinder Abrol, and William A. Goddard. Structure-based prediction of subtype selectivity of histamine H 3 receptor selective antagonists in clinical trials. *Journal of Chemical Information and Modeling*, 51(12):3262–3274, 2011.

[79] Soo Kyung Kim and William A. Goddard. Predicted 3D structures of olfactory receptors with details of odorant binding to OR1G1. *Journal of Computer-Aided Molecular Design*, 28(12):1175–1190, 2014.

[80] Uniprot. Gcr1 - g-protein coupled receptor 1 - arabidopsis thaliana, 1999.

[81] Kazutaka Katoh, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14):3059–3066, 07 2002.

[82] William C. Wimley and Stephen H. White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural Biology*, 3(10):842–848, 1996.

[83] D. B. Bylund, D. C. Eikenberg, J. P. Hieble, S. Z. Langer, R. J. Lefkowitz, K. P. Minneman, P. B. Molinoff, R. R. Ruffolo, and U. Trendelenburg. IV. International union of pharmacology nomenclature of adrenoceptors. *Pharmacological Reviews*, 46(2):121–136, 1994.

[84] Youyong Li, Fangqiang Zhu, Nagarajan Vaidehi, William A. Goddard, Felix Sheinerman, Stephan Reiling, Isabelle Morize, Lan Mu, Keith Harris, Ali Ardati, and Abdelazize Laoui. Prediction of the 3D structure and dynamics of human DP G-protein coupled receptor bound to an agonist and an antagonist. *Journal of the American Chemical Society*, 129(35):10720–10731, 2007.

[85] Pedro M. Hernández, Carlos A. Arango, Soo-Kyung Kim, Andres Jaramillo-Botero, and William A. III Goddard. Predicted three-dimensional structure of the gcr1 putative gpcr in arabidopsis thaliana and its binding to abscisic acid and gibberellin a1. *Journal of Agricultural and Food Chemistry*, 71:5770–5782, 2023. doi: 10.1021/acs.jafc.2c06846.

[86] Abscisic acid. https://pubchem.ncbi.nlm.nih.gov/compound/%28%2B%29-abscisic%20acid. Accessed: 2022-08-01.

[87] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A.A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

[88] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Zídek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, Michael Figurnov, Andrew Cowie, Nicole Hobbs, Pushmeet Kohli, Gerard Kleywegt, Ewan Birney, Demis Hassabis, and Sameer Velankar. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1):D439–D444, 2022.

[89] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.