UNIVERSIDADE D
COIMBRA

Daniel José Vasconcelos Silvério

DECODING PARTNER SPECIFICITY IN OPIOID RECEPTOR FAMILY

Dissertação no âmbito do Mestrado em Bioquímica orientada pela Professora Doutora Irina de Sousa Moreira e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Janeiro de 2021

Daniel José Vasconcelos Silvério

Orientação: Professora Doutora Irina de Sousa Moreira

Decoding partner specificity in opioid receptor family

Departamento de Ciências da Vida

Janeiro de 2021

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Table of contents

Resumo

Foi realizada uma caracterização extensiva da família dos recetores de opióides para criar novo conhecimento acerca das propriedades farmacológicas e fisiológicas destes alvos de fármacos importantes. Foi realizado modelação por homologia usando estruturas do recetor opióide do tipo- κ (humano), do recetor de neurotensina 1 e recetor muscarínico M2 para gerar complexos confiáveis de recetores de opióides ligados a proteína-G ou arrestina, empregando cinco scores diferentes para selecionar o melhor modelo de homologia em cada condição. Foram realizadas simulações de Dinâmica Molecular (equilibração) de forma a relaxar os melhores modelos. As estruturas relaxadas foram alinhadas com os modelos de parceiros para poder formar os complexos. Após a formação dos complexos foi aplicada uma ampla variedade de métodos computacionais para avaliar e providenciar uma descrição detalhada das interfaces de interação de todos os membros da família de recetores de opióides [μ (MOR), δ (DOR), κ (KOR), nociceptina (NOP) com os seus parceiros de ligação correspondentes (ARRs: ARR2, ARR3; proteína-G: G_{i1}, G_{i2}, G_{i3}, G_o, G_{ob}, G_z, G_q, G_{11}, G_{12}, G_{14}, G_{15}, $G_{s(sh)}$, $G_{s(lo)}$)]. Esta descrição inclui os seguintes parâmetros estruturais: distâncias inter-hélice, distâncias electroestáticas, resíduos que interagem, percentagens de interação dos resíduos, ligações de hidrogénio, pontes salinas, área de superfície acessível ao solvente, número de átomos à superfície e enterrados. Além disso, análise dinâmica, no âmbito da Análise de Modo Normal, foi também executada para avaliar dois parâmetros dinâmicos: mudanças de flexibilidade e mudanças no fold em flutuação média. A construção e análise destes 57 modelos, envolvendo recetores de opióides, representa uma nova e excitante análise de grandes dados dos determinantes estruturais na interface dos complexos e constituí um passo seguinte na compreensão da especificidade funcional da família de recetores de opióides.

Palavras-chave: Recetores de opióides; Proteinas-G; Arrestinas; Modulação por homologia; Simulações de Dinâmica Molecular

Abstract

An extensive characterization of the opioid receptor family was carried out to create new knowledge about the physiological and pharmacological properties of these important drug targets. Homology modelling was performed using κ-type opioid receptor (human), neurotensin receptor 1 and muscarinic M2 receptor structures to generate reliable complexes of opioid receptor bound to either G-protein or arrestin, employing five different scores to select the best homology model in each condition. Molecular Dynamic simulations (equilibration) were performed in order to relax the best models. The relaxed structures were aligned with the partner models in order to form the complexes. After the complex formation a wide range of computational methods were applied to assess and provide a detailed description of the interaction interfaces of all members of the opioid receptor family [μ (MOR), δ (DOR), κ (KOR), nociceptin (NOP) with their corresponding binding partners (ARRs: ARR2, ARR3; G-protein: G_{i1}, G_{i2}, G_{i3}, G_o, G_{ob}, G_z, G_q, G_{11}, G_{12}, G_{14}, G_{15}, $G_{s(sh)}$, $G_{s(lo)}$)]. This description includes the following structural parameters: inter-helical distances, electrostatic distances, interacting residues, residue interaction percentages, hydrogen bonds, salt bridges, solvent accessible surface area, number of surface and buried atoms. Moreover, dynamic analysis, under the scope of Normal Mode Analysis, was also performed to evaluate two dynamical aspects of the complexes: flexibility changes and average fluctuation fold changes. The construction and analysis of these 57 models, involving opioid receptors, represents a novel and exciting big data analysis of the complexes interface structural determinants and constitutes a further step into the understanding of opioid receptor family functional specificity.

Keywords: Opioid receptors; G-proteins; Arrestins; Homology modelling; Molecular Dynamic simulations

Acknowledgements

6 years after the beginning of my journey at the University of Coimbra, I want to thank my parents for all their resilience and perseverance during my difficult moments but also for their never-ending support. They were my guide and help me achieving all my accomplishments until now.

I am also deeply grateful that Irina Moreira, my advisor, allowed me to enter in her research group where I learnt so many things about computational biology that enriched my scientific knowledge. All the group members, and specially Carlos Barreto, António Gomes and Salete Baptista, I want to thank your availability and patience in teaching me all the necessary tools to do this work.

The friends that I met in Coimbra, particularly Gonçalo Afonso, Mário Coutinho, Ivo Machado, Luísa Abreu and Pedro Filipe, that since the first day accompanied me.

Not forgetting also my friends in Sertã: Rubén Farinha, André Ferreira, Pedro Neves, Rafael Fernandes, Filipe Gaspar, André Pinto and Ricardo António. They are my truly friends since I was very young.

Table Index

Figure Index

Equation Index

List of abbreviations

GPCR: G-protein coupled receptors

TM: Transmembrane α-helices

ECL: Extracellular loops

ICL: Intracellular loops

MP: Membrane proteins

GAIN: GPCR autoproteolysis-inducing domain

β2AR: β-2-adrenergic receptor

A_{2A}AR: A_{2A} adenosine receptor

GDP: Guanosine-triphosphate

GTP: Guanosine-diphosphate

AC: Adenylyl cyclase

PLC: Phospholipase C

GRK: G protein-coupled receptor kinases

ERK 1/2: Extracellular signal-regulated kinases 1/2

EGFR: Epidermal growth factor receptor

MMP: Matrix metalloprotease

EGF: Epidermal growth factor

OR: Opioid receptor

MOR: μ opioid receptor

DOR: δ opioid receptor

KOR: κ opioid receptor

NOP: Nociceptin opioid receptor

VGCC: Voltage-gated calcium channels

3D: Three-dimensional

NMR: Nuclear magnetic resonance

MSA: Multiple sequence alignments

ProSA: Protein Structure Analysis

MC: Monte-Carlo

MD: Molecular Dynamics

TIP: Transferable intermolecular potential

φ: *phi*

Ψ: *psi*

CHARMM: Chemistry at Harvard Molecular Mechanics

AMBER: Assisted Model Building with Energy Refinement

GROMOS: Groningen Molecular Simulation

OPLS: Optimized Potential for Liquid Simulation

RESP: Restrained Electrostatic Potential

QM: Quantum mechanics

MM: Molecular mechanics

DFT: Density Functional Theory

NMA: Normal mode analysis

ENM: Elastic network model

M2R: Muscarinic M2 receptor

NTSR1: Neurotensin receptor 1

DOPE: Discrete Optimized Protein Energy

molpdf: molecules probability density function

ProQ: Protein Quality

RMSD: Root Mean Square Deviation

CHL: Cholesterol

POPC: 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine

M1R: Muscarinic M1 receptor

HADDOCK: High Ambiguity Driven protein protein DOCKing

HB: Hydrogen bonds

SB: Salt bridges

SASA: Solvent accessible surface area

PIPSA: Protein Interaction Property Similarity Analysis

UHBD: University of Houston Brownian Dynamics

Å: Angstroms

COCOMAPS: BioCOmplexes Contact MAPS

BC: Bhattacharyya coefficient

MS: Multidimensional scaling

PDB: Protein Data Bank

V2R: V2 vasopressin receptor

5-HTR_{2A}: 5-hydroxytryptamine receptor 2A

1. Introduction

1.1 G-protein coupled receptors

Hundreds of G-protein coupled receptors **(**GPCR) were discovered in human organism. These receptors are embedded in the cellular membrane and all of them have similar structures, containing seven transmembrane α-helices (TM). These TM are connected through three extracellular loops (ECL) and three intracellular loops (ICL). Finally, one α-helix is located in the C-terminal (designated by H8).[1] There are several GPCR families, based on sequence similarity analysis: the rhodopsin receptors belong to the class A (which is the biggest GPCR group), the secretin receptors belong to the class B1, the adhesion ones are the class B2 the glutamate receptors belong to the class C, and finally, the last one is the class F (which includes the frizzled and smoothened receptors). These receptors are involved in several cellular signalling processes, and are a major target of approved drugs.[1,2] Thus, the understanding of the structural features and molecular mechanisms involving GPCR are of utmost interest in drug discovery and development pipeline.[1] Membrane proteins (MP) are particularly difficult to isolate and crystalize, although in the last few years, with the improvement of several experimental techniques, there was an increase of GPCR structures deposited in public databases.[3] Nowadays, and according with GPCRdb statistics, there are about 500 GPCR structures and 75 GPCR-G-protein complexes available, including structures of the same type of receptor or complex (updated in 30/12/2020).[4,5] The majority of the GPCR-partner complexes were solved using cryo-electron microscopy (cryo-EM).[3]

Despite the GPCR possess a typical structure, enunciated in the beginning, some differences can be spotted particularly between different GPCR, mainly in the extracellular regions while the TM and intracellular regions maintain structural conservation between different subfamilies.[6–9] Figure 1 shows some of these differences between various GPCR classes.

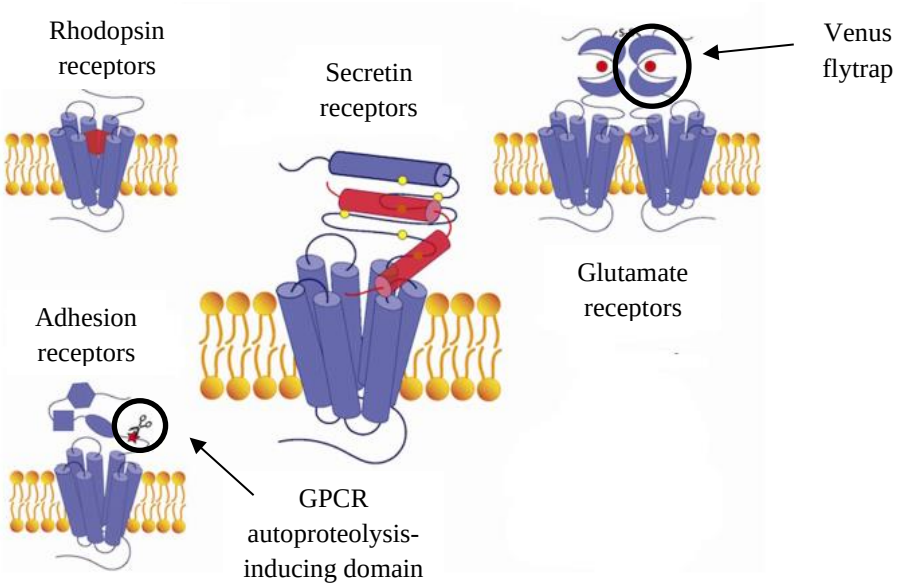


Figure 1: Different GPCR classes and specific structural differences between them, including some functional domains that are highlighted in this figure. Adapted from: Culhane, Kelly J. *et al. Transmembrane signal transduction by peptide hormones via family B G protein-coupled receptors* (2015). doi: 10.3389/fphar.2015.00264

The N-terminus (extracellular region), for instance, differs significantly between different GPCR subfamilies in terms of length, sequence and shape.[10–12] It is also relevant to note that the class A GPCR has, commonly, a short N-terminus (with few exceptions) whereas the other subfamilies (particularly the class B2) have long ones, however it is reported as being a cysteine-rich region, independent of the receptor class.[9,11] They can be unstructured, with no recognizable shape (class A GPCR), or they can also be globular.[12] There is also a wide spectrum of functional domains in the N-terminus and in the extracellular region, many of them are subfamily specific. The class A has very few domains, like the leucin rich repeat region (present in the thyrotropin receptor and others). This is in part due to fact that the N-terminus is very short, as previously discussed. The class B1 receptors share, between them, the hormone binding domain. The conserved proteolytic domain or, in other words, the GPCR autoproteolysis-inducing domain (GAIN) makes part of the class B2 receptors and it has an important role in protein cleavage.[8,13] The glutamate receptors have a conserved region which is designated as Venus flytrap motif due to shape it creates to accommodate the ligands during the coupling process. Other features are conserved in almost all GPCR, like the disulphide bridge between the ECL1 and the ECL2.[13] Other conserved disulphide bridge is the one connecting the TM3 and the ECL2, this link stabilizes the extracellular region even after receptor activation.[11]

The ligand binding pocket (in class A GPCR) also behaves differently according to the ligands it binds. Some pockets close after the coupling process (like the rhodopsin receptor), others remain open which is linked to the ligand hydrophobicity.[10] The pocket deepness is also a differentiating factor, the β-2-adrenergic receptor (β2AR) has a relatively deep pocket comparing with the A_{2A} adenosine receptor (A_{2A}AR), for example.[12] Despite the extracellular area has some conserved regions, particularly subfamily-specific, the TM and the intracellular region is generally more conserved in GPCR.[8] The ionic lock is one of these conserved contacts that is present in almost all GPCR. This is a salt bridge that connects the conserved E/DRY motif (in the TM3) with the $E^{6.30}$ (using the Ballesteros-Weinstein numbering)[14] and is essential to stabilize the GPCR when it is inactive.[8,15,16] There are other conserved motifs in the TM region like the NPxxY (in TM7) and the WxP (in TM5). The notable exceptions are the ICL3 and the C-terminus that can vary extraordinarily in length between different GPCR.[10,12]

1.2 Classic signalling and desensitization pathways

The classic signalling and desensitization pathways are well described in Figure 2.

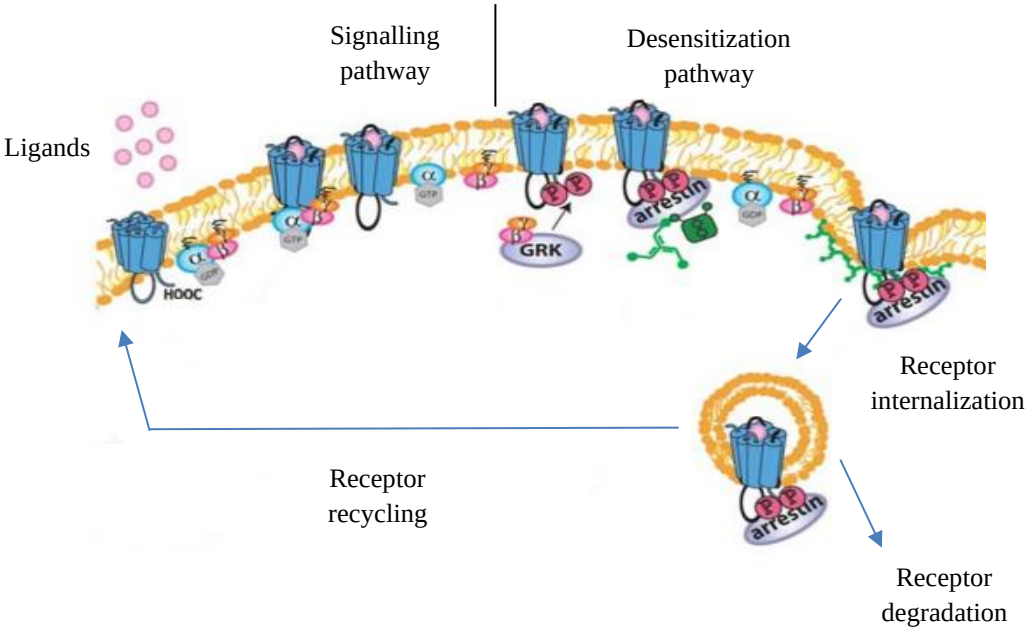


Figure 2: Classic GPCR signalling and desensitization mechanisms. Adapted from: Billington, Charlotte K. *et al*. *Signaling and regulation of G protein-coupled receptors in airway smooth muscle* (2003). doi: 10.1186/1465-9921-4-2

In order to trigger a signalling cascade is necessary that an agonist binds to the receptor in an area designated by ligand-binding pocket, leading to a conformational change in the receptor structure (activation) opening a pocket on the cytoplasmic side. The structure activation allows the coupling of heterotrimeric G-proteins or arrestins. The classic perspective states that either G-proteins activate signalling cascades or arrestins activate the desensitization pathway.[17] The G-proteins are molecules divided into two main subunits, α and βγ subunits; before receptor activation, the Gα subunit has a guanosine-diphosphate (GDP) molecule attached to it (inactive state), that is exchanged by a guanosine-triphosphate (GTP) after receptor activation.[18] Upon ligand activation the GPCR suffer conformational changes, in particular at the following TM segments: TM3, TM5, TM6 and TM7, whereas the other TM apparently do not move, significantly, from their inactive state positions.[11] In particular, the TM5 and TM7 inward and the TM6 outward movements are both critical to open the receptor intracellular binding pocket in order to accommodate the α5-helix of the G-protein.[6,19] This is accompanied by the cleavage of the ionic lock and the formation of other conserved contact occurs between the NPxxY motif and the $Y^{5.58}$, this is normally denominated as the water lock (hydrogen bond), and has a stabilizing role on the active state structure.[8,15]

Several complexes have shown specific GPCR subdomains that have a preponderance to interact with G-proteins. In general, the TM3, TM5, TM6 and the ICL interact directly with the G-proteins.[3]

As the G-protein couples to the receptor and induces a switch on the Gα subunit promoting the exchange of a GDP by a GTP, this switch activates the G-protein and leads to the dissociation of the two main subunits, each of one affecting different effector proteins (second messengers).[18]

It is important to take into account that exists four different G-protein subfamilies (Gs, Gi/o, Gq/11 and G12/13) and they are divided based on the α-subunit sequence similarity resulting in a total of 21 different G-proteins.[20] Different Gα subunits can affect different effectors. There are several known different effectors including adenylyl cyclase (AC) or phospholipase C (PLC).[18,20] As consequence each G-protein subfamily has a unique interaction with certain effectors, for example, Gi/o subfamily inhibits AC activity whereas Gs subfamily stimulates AC[18,21]. The AC activity modulation interferes directly with the cAMP intracellular levels thanks to its enzymatic activity. This enzyme is capable of transforming ATP in cAMP. The Gq/11 subfamily activates PLC[18]. The G12/13 stimulates GTPase activity.[22]

Arrestins are another family of molecules that play a major role in the GPCR molecular mechanisms. They are directly involved in the desensitization of GPCR, which is a major protective mechanism to defend the cells from overstimulation. There are four different arrestins: arrestins 1 and 4 named as visual arrestins (as they present only in the eye), and arrestins 2 and 3, considered non-visual arrestins.[23] These monomers are divided in two domains (N-domain and C-domain), they are separated by an area filled with loops containing hydrophilic residues, denominated by polar core. This region, in conjunction with the three-element region, keeps the inactive state arrestin structure cohesive.[24–26] To initiate this pathway there is an initial phosphorylation of the receptor (normally the binding occurs in the ICL and the C-terminus regions)[27–29], which is mediated by G protein-coupled receptor kinases (GRK), allowing the coupling between the arrestin and the receptor. This leads to polar core disruption and the arrestin C-terminus release.[26] Arrestins prevent the coupling of the G-protein with the receptor, inhibiting the signalling cascade, mediated by the G-protein.[23] Then the receptor is internalized by clathrin-mediated endocytosis and can follow two different paths: being recycled to the cellular membrane, to be used again, or be degraded by lysosomes.[30]

However, it is important to point out that many other independent signalling pathways were associated with the coupling of arrestins to GPCR, and these systems are key to understand GPCR function.[23]

1.3 Non-classic GPCR signalling pathways

Several new signalling pathways have been discovered that shaped our understanding of GPCR. Initially, it was thought that only G-proteins could couple with GPCR and activate signalling cascades, modulating several cellular responses as result.[31] However, in recent years it was discovered that specific agonists can induce different signalling pathways by promoting, preferentially, the coupling of certain transducers (G-proteins, arrestins or GRK) with the GPCR.[32] This phenomenon is named biased activation.[31,32] The activation mechanism can be explained by different receptor conformational changes induced by this biased ligands.[31] Evidence shows that the biased activation mechanism is more complex since the coupling of a biased ligand or an allosteric modulator it is not necessary to trigger the receptor biased activation. There are other ways to trigger this phenomenon such as coupling with specific molecules (e.g. intrabodies) from the intracellular side of the receptor.[33]

Even though GPCR are membrane receptors, it was demonstrated that they can activate signalling cascades intracellularly via internalization mechanisms.[34] Arrestins induce the clathrin-mediated endocytosis of GPCR allowing the receptor internalization, the specific mechanisms underlying GPCR intracellular activation depends on the specific GPCR class, for example, receptors belonging to the class A form a complex with arrestin promoting the internalization via endocytosis (forming a clathrin-coated pit) and then they activate effector proteins [extracellular signal-regulated kinases 1/2 (ERK 1/2)]. Unlike the class A, the receptors belonging to the class B will bind with arrestin and form an endosome before the posterior activation of effectors.[31] This difference can be explained by the different stability of the complexes.[34] Other intracellular activation pathways involve: only arrestins to signal ERK 1/2 without the need to form an endosome containing the complex; using G-proteins (coupled with the receptor) inside the cytoplasm to signal ERK 1/2 via cAMP and other signalling pathways independent of internalization.[31] In the classic signalling pathway GPCR monomers signal G-proteins without any cooperation from other receptors, however this is being put in question by new evidence showing that GPCR can cooperate together forming GPCR oligomers. Under these circumstances some GPCR signal through different mechanisms comparing with GPCR monomers, for example, they can use different G-proteins or use arrestin instead of G-protein.[31]

Under specific agonist activation some GPCR can also activate other membrane receptors (e.g.; epidermal growth factor receptor (EGFR))[31,35], this mechanism is called transactivation and until recently the majority of GPCR capable to induce EGFR activation belong to the class A. There are two ways to activate it: GPCR ligand-dependent or independent EGFR transactivation. The first way occurs with the normal GPCR activation (with G-proteins) and then the G-protein subunits activate a specific protein named matrix metalloprotease (MMP), this protein in its turn cleaves the epidermal growth factor (EGF), the EGF binds to the EGFR leading to its activation. The second way does not have ligands, like the EGF, involved in the EGFR activation but mechanisms still need to be further understood.[35] The EGFR activation promotes the cell proliferation and growth so new therapeutic strategies are being studied to block the transactivation process by inhibiting both the receptor and the EGFR, for example.[35] Finally, there is another non-classic signalling mechanism called biphasic activation where there are two distinct phases (in the temporal line) of signalling activity, mediated by GPCR, and differentiated by their intensity and duration.[31]

1.4 Opioid receptors

Opioid receptors (OR) are a group of GPCR divided into 4 different types: μ (MOR), δ (DOR), κ (KOR) and nociceptin (NOP) opioid receptors. NOP is distinguished from the other receptors, in this subfamily, in the kind of the ligands it binds.[6] These receptors are present in two regions: central nervous system and gastrointestinal tract. They belong to the rhodopsin-like receptors group (GPCR class A), when coupled with specific agonists the opioid receptors can induce several cellular responses such as analgesia.[36,37] The analgesia is a common effect in all OR but striking differences were found between different OR in the kind of singular effects they produce when activated: for example the MOR causes respiratory depression, KOR causes dysphoria and DOR reduces anxiety.[37] The development of specific OR agonists is important in pain relief studies.[29,36,37] The OR also have a regulatory role on the calcium levels in the cells. The receptor activation, as said before, leads to the dissociation of the G-protein, the Gβγ subunit binds to the voltage-gated calcium channels (VGCC) and cuts the calcium flux through the channel. The OR can also associate directly with the VGCC, promoting the internalization of the complex and leading to a reduction of VGCC in the cellular membrane.[38]

Nonetheless, the prolonged use of opioids (constant stimulation) can increase the organism tolerance to these compounds, reducing their important clinic effects. The opioid use disorder is the basis of the opioid crisis, a huge socio-economic issue worldwide that is directly related to drug abuse.[36,39] The detailed study of opioid receptors is important to further increase the drug development speed in this area and to lead to new clinical responses to this crisis.[36,39] Beyond the generated tolerance by opioid abuse, there is an increasing difficulty by the individual to stop consuming these drugs due to harsh symptoms provoked by the absence of these substances in the organism. Some treatments are currently available to reduce the symptoms provoked by this absence. The methadone, for example, is used as an alternative to heroin. This drug binds to MOR and stays more time in the body than heroin. As consequence, there is a decrease of the absence symptoms resulting from the addiction.[37]

Other drugs like buprenorphine have a lesser stimulation degree (in MOR) than methadone, thus this drug promotes less aggressive effects resulting from MOR activation. Naloxone actuates differently, being a OR antagonist, this drug will block the OR activation. It is commonly used to prevent the dangerous consequences of opioid overdose.[37,40] Other allosteric modulators were also developed to avoid the opioid crisis problematic such as the diterpene alkaloid ignavine (MOR modulator). This modulator is safer to use and can increase the OR agonist efficacy without potential side effects.[2]

2. Computational methodologies

2.1 Homology modelling

The fact that GPCR are MP makes them particularly difficult to isolate and obtain a reliable three-dimensional (3D) structure. This is an obstacle and hampers further studies on these proteins and their physiological function. Instead of using experimental techniques like the X-ray crystallography or the nuclear magnetic resonance (NMR), with known limitations for this kind of proteins, the use of computational methods emerged as an alternative way to build a protein model.[41] Currently, there are two main technical approaches to model a protein: *ab initio* methods and homology modelling.[42] If we have a protein with a sequence which has no significant similarity with other proteins, with known structures, then is necessary to apply *ab initio* methods. This approach predicts the protein folding taking into account that the best structure is the one that have the lowest energy (native structure). This requires the use of force-fields that will predict the atoms behaviour. This is very computationally exhaustive and normally is only used in small proteins.[41]

The homology modelling method consists in the prediction of a 3D structure of a protein, with an unknown structure, based on another protein with a known structure (template), that has a certain level of similarity between the amino acid sequences (>25-35%). Above this level is acceptable that both structures share a similar folding. This process has some advantages comparing with the *ab initio* methods: can be used on large proteins and has more accuracy.[41–43] The homology modelling process is divided into several steps (usually 4), with the first one consisting of the choice of the best template. We need to take into account the sequence similarity between the template candidates and the protein in study.[43] Normally, this procedure is performed by programs specialized in sequence alignments such as the Basic Local Alignment Search Tool (BLAST)[44], which does pairwise sequence alignments, or the ClustalOmega[45] which does multiple sequence alignments (MSA). The similarity parameter is the most important one but there are other factors to consider when we choose the best template such as the structure resolution or the presence of other molecules (e.g. ligands)[43]. The second step consists in the alignment between the target and the template sequence; the third step is where the building of the model happens. There are several approaches to build a model, like the construction with spatial restraints. In this case, the program will build a model respecting the stereochemistry resulting from the chosen template.[43,46,47] The other way to have a protein model is through a rigid-body assembly approach, in this case the objective is to start this process by using conserved features present in the template. Such features could be elements of secondary structure, in the model these regions will be assembled together, and the gaps will be filled between them.[41,43,48]

Finally, we need to validate the constructed model, this step is important to estimate the quality of the model. The main way to evaluate it is through the calculation of the protein free energy and compare it with a group of proteins structures in native conformation (using, for example, Protein Structure Analysis (ProSA)

webserver)[49,50], however some other methods could be used.[42] It is also important to mention that as the percentage identity increases, between the sequences, the higher the quality will be of the final model.[41]

2.2 Molecular dynamics simulation

To study a system containing many particles there are 2 different approaches used in the computational simulation area: Monte-Carlo (MC) and molecular dynamic (MD) simulations.[51,52] MD simulations are a type of computer simulations that use classical mechanics (molecular mechanics) to study the dynamic processes behind a certain biological structure. A biologic structure is not a rigid entity and undergoes conformational changes[53] so the dynamic process represents nothing more than the change of the positions and velocities of the atoms, from a system, throughout the simulation time. The follow-up of the dynamic process is important because allows measuring several parameters of the biomolecule[53], usually, these parameters are a function of the position and the linear momentum (depends from the mass and velocity of the atoms) from the particles[51], hence the importance to understand the change of positions and velocities (particles trajectories) during the MD simulation.[52] This gives an instantaneous value of the parameter at a particular time but, in the end, the objective is to average all the obtained values, at each time of the simulation, to obtain a final value of the parameter (called time average).[52] This connection between the parameter (of thermodynamic basis) and the particles trajectories is well described in the next equation[52]:

$$< A >= \frac{1}{m}\sum_{i=1}^{m} A(p^n, r^n)$$

Equation 1: General equation to calculate thermodynamic parameters in MD simulations

The $< A >$ represents the average value of the thermodynamic parameter, m is the number of timesteps of the simulation, *n* is the number of particles of the system, A is the thermodynamic parameter value in a specific timestep, *p* is the momentums and *r* the positions from the *n* particles of the system.[52] Equation 1 represents a general description of this connection, there are variations to this equation depending upon the parameter in study.[52] To obtain a time average from the parameter in study the MD simulation protocol (or algorithm) usually follows these steps: system construction, energy minimization, equilibration and simulation.[51,54] In the construction of the system step, normally a molecule will be inserted in a simulation space, usually called a box. The simulation box can have several shapes, being the cubic shape the most used one, to avoid the boundary effect (due to the system size) the simulation box will be multiplied[54], this is important because, normally, the number of particles in the simulation is too small so a good part of the particles is interacting directly with the boundaries of the simulation box, this strategy simply puts similar simulation boxes in each direction of the original one so the closest particles to the boundary will not be affected by this condition.[52] In

the case of a biomolecule, it becomes necessary to mimic the real conditions where it is so nowadays there are ways to fill the simulation box with a solvent (water)[54] using, for example, the Transferable intermolecular Potential (TIP) models.[55] In the case of opioid receptors (MP), there is an additional constraint in the first step of the MD simulation which is the introduction of the lipid bilayer into the simulation, this poses the challenge of the phospholipid choice. Usually, to reduce the simulation complexity, the phospholipids available are simplified by using only two models of tail chains (palmitoyl and oleyl), the most frequently present chains in phospholipids. However, other chain models can be chosen.[56] After the construction of the system and the introduction of the thermodynamic parameters is fundamental to input the initial positions and velocities of each particle of the system to initiate the simulation[51,54], the original positions can be extracted, in the case of a protein, from the structure file obtained by experimental methods (e.g. X-ray crystallography) or from a model constructed by homology modeling[54], the initial velocities can be obtained using the Maxwell-Boltzmann distribution, represented in the following Equation 2:

$$P(v_{ix}) = \sqrt{(\frac{m_i}{2\pi kT})} exp\ (\frac{-m_i v_{ix}^2}{2kT})$$

Equation 2: Maxwell-Boltzmann distribution equation

Where k is the Boltzmann constant, m_i is the mass of the particle i, v_{ix} is the velocity of the particle i at the direction x and T is the temperature. $P(v_{ix})$ is the probability of the particle i has the velocity v_{ix}. The Maxwell-Boltzmann distribution is a Gaussian distribution (with the sinusoidal shape).[57]

This allows the start of the last two steps: the system equilibration and the simulation. The equilibration is an important feature, the initial positions and velocities of the system describe the original conformation of the biomolecule, this step has as objective that the system reaches its equilibrium conformation before the simulation starts[52], to do this is necessary to accompany the evolution of several thermodynamic and structural parameters, during the equilibration, until they do not vary with time[52], this step is widely used to refine structures (relaxation) coming from homology modelling techniques.[53]

The last step and the most important one is the simulation, the simulation is divided in timesteps, in these sections of time it will be calculated the forces applied on each particle, this is important to predict the trajectory of the particles in all timesteps, to do that is important to use Newton's equation of motion[52] described in the Equation 3:

$$\frac{d^2 x_i}{dt^2} = \frac{F_{xi}}{m_i}, F_{xi} = -\nabla_i U(r_i)$$

Equation 3: Newton's equation of motion

The m_i is the mass, $U(r_i)$ is the potential energy (sum of all interactions) and F_{xi} is the force acting on the particle i along the axis coordinate x_i.[52,54]

In Equation 3 the forces applied on the particle are a function of the potential energy of the particle[54].

To calculate the potential energy of a particle there are series of equations (force fields) that characterize 2 different interactions in a biomolecule: bonded and non-bonded interactions.[53,54,58] The bonded interactions can be further divided into 3 terms: bond stretching, valence angle bending and torsion angles. These interactions are mathematically described by the following equations:

$$v(l) = \frac{k}{2}(l - l_0)^2$$

Equation 4: Bond stretching term

$$v(\theta) = \frac{k}{2}(\theta - \theta_0)^2$$

Equation 5: Valence angle bending term

$$v(\omega) = \sum \frac{V_n}{2}[1 + (n\omega - \gamma)\,]$$

Equation 6: Torsion (dihedral) angles term

In the Equation 4 the k is the stretching constant of the bond, $v(l)$ is the potential energy, l_0 is the bond length at the equilibrium state whereas l is the bond length determined in the timestep. For the Equation 5 the k is the force constant of the bond, θ_0 is the valence angle at the equilibrium state, θ is the valence angle determined in the timestep and $v(\theta)$ is the potential energy.[52]

The Equation 6 represents the torsion angles between 4 consecutive atoms, in the case of proteins these torsion angles could include the widely known *phi* (φ) and *psi* (Ψ) angles, among others.[59] The ω is the torsion angle, γ , $v(\omega)$ is the potential energy, n is the periodicity, γ is the phase factor and V_n is considered the energy barrier to rotation.[52]

Furthermore, the non-bonded interactions represent electrostatic and van der Waals interactions between the atoms, these interactions can be mathematically described by the following equations:

$$U_{Coloumb} = \frac{q_a q_b}{4\pi\varepsilon_0 r_{ab}}$$

Equation 7: Coulomb law equation (electrostatic interactions)

$$U_{Lennard-Jones} = 4\varepsilon[\left(\frac{\sigma}{r}\right)^{12} - (\frac{\sigma}{r})^{6}]$$

Equation 8: Lennard-Jones potential equation (van der Waals interactions)

In Equation 7 the q_a and q_b are the charges from atoms a and b, r_{ab} is the distance between these atoms and ε_0 is the permittivity of free space.[54,60] In the Lennard-Jones potential (Equation 8) the ε is the well depth whereas the σ is the collision diameter. The r is the distance.[52,54,60] These interactions represent the forces that can actuate in an atom during a MD simulation, the sum of all these interactions give us the potential energy of the system at a particular time.

After the calculation of the non-bonded and bonded interactions and the determination of the potential energy it is possible to calculate the force in Newton's equation of motion (Equation 3), this will be useful in the next procedure.[51] To predict the trajectory of the particles is fundamental to use a mathematical process (integration) and apply it directly on Equation 3, this will give two solutions, one will determine the new position of the particle and the other one will give the new velocity.[54]

This procedure is repeated on each timestep of the simulation until the end, with this information is possible to predict the thermodynamic parameter (time average), throughout the simulation, taking into account what is stated above in Equation 1. It is important to refer that a MD simulation normally runs in a constant number of particles, volume and total energy (thermodynamic parameters)[61], these parameters are computed in the system construction step.

As said before the MD simulations use force fields with pre-determined parameters (see Equations 4,5,6,7 and 8) allowing the potential energy calculation between particles, they can be distinguished by the way these parameters were determined.[55] Also, these force fields are normally referred to as having an empirical basis because the parameters of their equations were obtained using experimental studies.[55] The force fields can be all-atom, united-atom or coarse-grained. While all-atom force fields provide the best description of the atomic particles considering the existence even of the hydrogen atoms, thus requiring more calculations, united-atom force fields are simpler because they do not consider the hydrogen atoms, uniting both carbon particles with the respective hydrogens. Coarse-grained force fields are the simplest as they simplify the structure omitting, for example, side chains. The advantages of this kind of force fields are the possibility of using larger timescales for the simulation.[62]

Several packages (containing force fields) can be used such as Chemistry at Harvard Molecular Mechanics (CHARMM)[63,64], Assisted Model Building with Energy Refinement (AMBER)[65], Groningen Molecular Simulation (GROMOS) or Optimized Potential for Liquid Simulation (OPLS). AMBER is a package for MD simulations (initially it was based on united atom force-fields) being the latest force fields developed for this package: Parm99[66] and Parm99SB.[67] The Parm99 uses a more accurate representation of the atomic charges, they were derived using a model named Restrained Electrostatic Potential (RESP)[68], the atomic charges are used to determine the q_i parameter (described in Equation 7). The Parm99SB is an attempt to improve the

torsional parameters (Equation 6), the optimization was conducted using several glycine and alanine conformations to approximate the molecular mechanics model to the respective quantum mechanics model.[67] The last iteration of OPLS package includes the OPLS-AA/L[69] force field, this force field revaluates torsional parameters and non-bonded parameters (to all amino acids in dipeptide form) from previous force-fields according to with quantum mechanics (QM) calculations[55,69]. GROMOS still uses a united-atom approach to this day, the GROMOS96[70] force-field is a recent version of this package being optimized for alkanes, a group of molecules widely present in biomembranes.[70]

So, the main difference between these force fields resides in the definition of the torsional, Coulomb (charges) and Lennard-Jones parameters.[59] The majority of these force fields are optimized for proteins, however, when is necessary to simulate a membrane protein there are currently reparametrized force-fields to simulate the lipid bilayer components like the LIPID14[71] or the CHARMM36[72] force fields.[56]

The empirical force fields described above have several advantages, including the allowing of simple and fast calculations of the potential energy but they lack the correct prediction of some phenomena that occur in real life, like the bond breaking, so models that use QM take into account the presence of electrons in the atoms, they are described, for example, with the density functional theory (DFT).[61]

2.3 Molecular docking

The molecular docking has the aim of providing the possible binding conformations between two molecules, giving a unique score for each one. This computational area is very important in drug discovery studies.[73,74] These techniques can be applied protein-ligand, protein-nucleic acid and protein-protein complexes, each with a unique set of features to study.[75] Usually, a docking protocol can be defined by the following steps: search for every possible ligand binding conformation (sampling) and evaluate them (scoring).[76,77]

For protein-ligand docking, the first step uses docking algorithms that take into account the flexibility factor from the ligand and, normally, maintain the proteins as a rigid monomer.[76,77] These algorithms are divided in three groups: stochastic; deterministic and systematic. The systematic algorithms allow the ligands to be full flexible, however this generates a greater number of possible candidates due to the fact that all torsion angles (for each bond) are rotated in search for the best configuration. The deterministic ones imply that the best conformation is directly dependent of the input structures. The stochastic algorithms are similar to the systematic ones, but with a particular difference, instead of testing every possible rotation this method changes randomly the rotation values in order to expedite the search.[74,76]

Protein flexibility can also be accounted, despite the challenges imposed by taking into account and numerous methods incorporate this parameter in docking strategies like, for example soft, selective, ensemble and on-the-fly docking.[76,77] These docking methods can be divided into two small groups: one group considers protein flexibility during the docking run, the other has a set of protein structures (with different conformations) created before the docking.[78] Soft docking belongs to the second group, this method uses a relaxed Lennard-Jones potential to avoid stereochemical conflict between the ligand and the protein allowing more

conformation configurations in the protein binding pocket. This method is the simplest approach to protein flexibility.[77,78] Ensemble docking also belongs to this group, however, uses a different strategy: the protein flexibility is not restrained in the binding pocket, as what happens with soft docking, but the generated conformations are built before docking (like soft docking), the objective is to construct a group of different conformations of the same protein (the can be built using MD simulations), each conformation will be docked against the ligand, although during the docking the protein conformation will not alter the respective configuration.[77]

Selective docking already belongs to the first group, the flexibility will be only considered in some binding regions (side-chains) of the protein and will be tested during the docking.[76] On-the-fly docking is the most realistic approach to protein flexibility, this method considers the proteins, as an all, a flexible entity. There are several ways to use this docking method like, for example: let the ligand dock with the protein (in a rigid configuration) and then the protein will alter his conformation to test several different configurations.[77]

In Protein-Protein docking there are several approaches, depending on the docking algorithms applied: rigid, semi-flexible and flexible docking.[73,74,79] The rigid docking, which considers both structures as rigid molecules[79] (without movement in the backbone or side-chains), has the advantage of requiring less computational calculations. Semi-flexible docking allows the flexibility of one of the structures (usually a ligand), maintaining the other structure in a rigid format. Flexible docking allows both structures as flexible molecules, being the best approach to binding real conditions but requires more computational calculations.[73,74]

The second step consists in giving a score to every binding configuration and assess them in order to find the best one, this is accomplished through scoring functions.[76,80] Several kinds of scoring functions evaluate these energies like, for example: knowledge-based, force field-based or empirical-based scoring functions.[81] The first one predicts the quality through information coming from other complexes with known 3D structures, however is limited by the structures available as primary information sources. The second one uses force fields, to calculate the energy, in a similar way to the MD simulations. The last one uses specific data from other experiences, with the intuit to determine the binding affinity, to calculate the binding free energy. These functions, in general, are similar to those used to score the homology models.[76] The last decades were prolific in the development of docking programs containing these two steps, these programs are, for example: AutoDock[82], GOLD[83] or DOCK[84]. GOLD is a webserver that performs protein-ligand docking, allowing the proteins to have some flexibility, thus being more accurate at predicting the best binding conformation. DOCK, for example, uses energy minimization to each conformation, similar to the stochastic algorithms.[75,83–85]

2.4 Normal mode analysis

Normal mode analysis (NMA) is an alternative method to study molecular dynamics, the objective of this technique is to introduce oscillations in the molecule maintaining the same conformation (contrary to what happens in MD simulations), this characteristic allows fewer calculations performed by a computer.[86–88] A molecule or complex needs to be in the lowest energy possible (energy minimization) to perform this kind of

analysis.[86,87] Being at the lowest energy configuration indicates that the protein is close to the native conformation or, in other words, near the equilibrium state. Each normal mode has its own specific frequency (which is fixed for all atoms), so the protein oscillates with the atoms moving at the same frequency. Normally the modes used to study dynamics correspond to the lowest frequencies. This happens because these range of frequencies correspond to the biggest conformational changes in a protein.[89] This has obvious advantages comparing with MD. Large proteins, for example, could be studied without expending large amounts of time to perform the dynamic processes. Even more important is the structural changes that encompass large protein sections such as helices, for example. These elements, normally, move at slower pace than individual residues.[89]

One way to study protein dynamics through NMA is considering the structure as an elastic network model (ENM), this model is even more simple than coarse-grained ones. In this case the α-carbons are the only atoms represented in the structure and they are considered as a simple point. The points are linked by "springs", this simplifies the potential energy calculation by only considering the sum of the Hookean potentials (between each pair of atoms) as the potential energy of the protein.[90,91]

3. Methods

A thorough review on the literature was proceeded in order to find all functional couplings between OR and G-proteins. The obtained results can be visualized in Table 1.

Table 1: Specific OR/G-protein coupling based on scientific literature. Several experimental methods were used to assess the different partners that each OR is capable to couple, such as: immunoprecipitation; *pertussis* and cholera toxin treatment; antisense oligodeoxynucleotide administration or chimeric G-protein use, for example. The functional couplings are marked by their respective references.

Receptors	Functional couplings
DOR	Gi1[92]; Gi2[92]; Gi3[92]; Go[92]; Gob[92]; Gz[92]; G14[92]; G15[92,93]; Arrestin 2[94,95]; Arrestin 3[96]
KOR	Gi1[92]; Gi2[92]; Gi3[92]; Go[92]; Gob[92]; Gz[92]; Gs(sh)[97]; Gs(lo)[97]; Arrestin 2[95]; Arrestin 3[98]
MOR	Gi1[21,92,99,100]; Gi2[21,92,99,100]; Gi3[21,92,99,100]; Go[21,92,99,100]; Gob[21,92,100]; Gz[21,92,100]; Gs(sh)[101,102]; Gs(lo)[101,102]; Gq[100]; G11[100]; G15[103]; Arrestin 2[104]; Arrestin 3[96,105]
NOP	Gi1[92]; Gi2[92]; Gi3[92]; Go[92]; Gob[92]; Gz[92]; G12[106]; G14[92,106]; Arrestin 2[107]; Arrestin 3[108]

3.1 Homology Modelling (receptors)

Homology modelling was applied to all members of the OR family, three distinct structures were used as templates: active KOR stabilized by a nanobody (human) (Protein DataBank ID: 6B73)[109]; Muscarinic M2 receptor (M2R) bound to arrrestin 2 (human) (Protein Data Bank ID: 6U1N)[110]; Neurotensin receptor 1 (NTSR1) bound to arrestin 2 (human) (Protein DataBank ID: 6PWC)[111]. Active KOR was used to model the receptors for G-protein coupling, whereas the NTSR1 and M2R were used to model the receptors for arrestin coupling. All templates were refined, removing the expendable monomers (nanobody, arrestins, …), leaving only the receptor structures (chain R in NTSR1 and M2R, chain A in KOR).

Target sequences were extracted from UniProt[112] (UniProt sequence codes: DOR (human)-P41143; KOR (human)-P41145; MOR (human)-P35372; NOP (human)-P41146; M2R (human)-P08172; NTSR1 (human)-P30989) of the opioid receptor members, further refinement of the sequences was needed so it was cut the N terminus and the C terminus for all sequences. The refined sequences (templates and OR) were aligned using ClustalOmega[45]. This webserver provides very accurate MSA even with thousands of sequences as input. The

objective was to see if the percentage identity values were reasonable enough to continue with these templates (Table 2). After the alignment the templates were used for model building.

Table 2: Percentage identity between all OR, NTSR1 and M2R. The sequences were extracted from UniProt[112] and refined posteriorly. The data was obtained from ClustalOmega[45] MSA program. The N-terminus and C-terminus of all receptors were cut.

	NTSR1 (human)	M2R (human)	DOR (human)	KOR (human)	MOR (human)	NOP (human)
NTSR1 (human)	100.00					
M2R (human)	20.86	100.00				
DOR (human)	24.74	26.07	100.00			
KOR (human)	26.39	27.40	68.49	100.00		
MOR (human)	24.21	26.52	70.99	69.52	100.00	
NOP (human)	27.02	25.09	59.39	61.17	58.90	100.00

The receptor homology modelling used MODELLER 9.21[46,47]. The program requires an alignment between each target sequence and the respective template sequence, was used ClustalOmega[45] results. Then, it was defined the TM limits (Table 3) and the disulphide bridges in the OR. This is necessary because the MODELLER 9.21 is based on the satisfaction of spatial restraints where is introduced data about the OR structure, forcing MODELLER to perform the homology modelling without violating the imposed restraints[47]. 100 models were built for each condition.

Table 3: OR residue numbering for the first and last residues in each OR subdomain (TM, ECL and ICL) as defined for the construction of the models.

	TM1	ICL1	TM2	ECL1	TM3	ICL2	TM4	ECL2	TM5	ICL3	TM6	ECL3	TM7	H8
DOR	6-44	45-48	49-78	79-83	84-118	119-127	128-154	155-172	173-210	211-215	216-254	255-259	260-288	289-302
KOR	6-36	37-40	41-70	71-75	76-110	111-119	120-146	147-166	167-205	206-210	211-249	250-253	254-282	283-295
MOR	6-37	38-41	42-71	72-76	77-111	112-120	121-147	148-165	166-203	204-208	209-247	248-251	252-280	281-293
NOP	6-37	38-41	42-71	72-76	77-111	112-120	121-147	148-165	166-203	204-208	209-247	248-251	252-280	281-293

The 100 constructed models were evaluated using different scores (the scores values, for each condition, are located in Annex 1 to Annex 3). A pre-selection, of the 20 best models, was realized using Discrete Optimized Protein Energy (DOPE) score.[113] This is a statistical potential constructed thanks to information (residue distances) derived from hundreds of native structures, this potential is correlated to the free energy of the protein which is tied to the quality of the model given that the lowest free energy belongs to the template. Other parameter, the molecules probability density function (molpdf)[47] also known as MODELLER objective function was used. The ideal is that the best model has the lowest values for these two last parameters. The ProSA webserver derived Z-score[49] identifies models with possible errors in the structure. This score is calculated through a potential similar to the DOPE score principle and determines the difference between model energy and an average of structures energies experimentally resolved. These last three scores are more suitable to soluble proteins. The last parameters (LGscore[114] and MaxSub[115]), which are derived from Protein Quality (ProQ)[116] webserver, are based on the superimposition between the model and the template structures. The Root Mean Square Deviation (RMSD) is calculated between all equivalent residues and incorporated in specific equations. The target sequences were submitted to PSIPRED[117] web service to determine their secondary structure, this information was incorporated during the calculation of LGscore[114] and MaxSub.[115]

3.2 Homology Modelling (partners)

Homology modelling, of G-proteins and arrestins, was applied for the following partners: Gi1, Gi2, Gi3, Goa, Gob, Gz, Gs(lo), Gs(sh), Gq, G11, G12, G14, G15, arrestin 2 and arrestin 3. Partner sequences were extracted from UniProt (UniProt sequence codes: Gi1 (human)-P63096; Gi2 (human)-P04899; Gi3 (human)-P08754; Go (human)-P09471-1; Gob (human)-P09471-2; Gz (human)-P19086; Gs(lo) (human)-P63092-1; Gs(sh) (human)-P63092-2; Gq (human)-P50148; G11 (human)-P29992; G12 (human)-Q03113; G14 (human)-O95837; G15(human)-P30679; arrestin 2 (human)-P49407-1; arrestin3 (human)-P32121-1). The Gi protein

(chain A) from the Rhodopsin-Gi complex (human) (Protein DataBank ID: 6CMO)[118] was used as template to the following partner models: Gi1; Gi2; Gi3; Go; Gob; Gz; Gq; G11; G12; G14; G15. The Gs protein (chain A) from the β2AR-Gs complex (PDB ID:3SN6)[119] was used to model: Gs(sh); Gs(lo). Finally, the arrestins were modelled with the arrestin 2 (chain A) either from the NTSR1-Arrestin 2 (PDB ID: 6PWC)[111] or the M2R-Arrestin 2 (PDB ID: 6U1N)[110] templates. The partner models were built using SWISS-MODEL.[120] This webserver performs homology modelling based on rigid-body assembly and in an automatic way, only requiring the input of the template file and the target sequence, generating one correspondent model.[43]

3.3 Membrane construction and MD simulations (equilibration)

After modelling the receptors and partners the next step consisted in embedding the receptor models in a lipid bilayer, upon extraction of the correct protein orientation inside the lipid bilayer from Orientations of Proteins in Membranes (OPM)[121] webserver. The resulting .pdb file was submitted to CHARMM-GUI[122] webserver to create a box containing the system (protein + membrane), water and ions. Lipids were added to the membrane (1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) and cholesterol (CHL)) in the following ratio (POPC: CHL/9:1), the box was filled with NaCl ions (0.15 M).

MD simulations were performed for all models to relax the structures (equilibration), using GROMACS 2018.4[123] package and CHARMM36 force field[72] (all-atom force field). Each simulation was divided into 8 steps (each one with a duration of 50 nanoseconds) with protein and lipid restraint forces decreasing sequentially.

3.4 Refinement of complexes

After concluding the equilibrations, the next objective was to combine the relaxed structures (receptor models) with corresponding partners (G-protein and arrestin models) to form the complexes. To do that was used superimposition templates. The OR-Gi/o complexes were formed using the Rhodopsin-Gi (PDB ID: 6CMO)[118]. The OR-Gs with the β2AR-Gs (PDB ID: 3SN6)[119], the OR-Gq/11 with either the Rhodopsin-Gi or the Muscarinic M1 receptor (M1R)-G11 (PDB ID: 6OIJ).[124] The OR-Arrestins with either the NTSR1-Arrestin 2 (PDB ID: 6PWC)[111] or the M2R-Arrestin 2 (PDB ID: 6U1N).[110] This procedure was realized using PyMOL.[125] Then, in order to refine the complexes was used the refinement tool of the High Ambiguity Driven protein-protein DOCKing (HADDOCK).[126]

3.5 Structural and Dynamic analysis

Several structural parameters, from the complexes, were analysed including: interhelical distances (TM3-TM6; TM3-TM7), electrostatic distances, interacting residues, residue interaction percentages, hydrogen bonds (HB), salt bridges (SB), solvent accessible surface area (SASA) and the number of surface or buried atoms. The TM3-TM6 distance was measured between the residues 3.50 and 6.30, whereas the TM3-TM7 distance was between residues 3.50 and 7.53 (using the Ballesteros-Weinstein numbering)[14], this information gives an

indication of the intracellular cavity flexibility when bound with different partners.[127] Electrostatic distances were determined by Protein Interaction Property Similarity Analysis (PIPSA)[128], this online program calculates the electrostatic potentials from proteins and compares them. The University of Houston Brownian Dynamics (UHBD)[129] was used to solve the Poisson-Boltzmann equation in order to calculate the protein potentials (under the following conditions: 300 Kelvin and an ionic strength of 50 millimolar) and then these results were converted into distances and presented in a heatmap. All interacting residues (specifically the α-carbons in the backbone) under 8 Angstroms (Å) of distance, between each other, residue interaction percentages (residue type and group) and SB, with a top limit of 4 Å of distance between residues, were determined through Python language scripts. The BioCOmplexes Contact MAPS (COCOMAPS)[130] is a webserver that analyses complex interfaces and identified HB, in OR complexes, thanks to the HBPLUS program[131]. It also determined the SASA through NACCESS[132] program, this software uses a sphere that mimics the water molecule (solvent) and introduces a van der Waals radius for each atom (in spheric form) from the receptor structure. The sphere rolls through the receptor and any contacts it makes with the atoms (without clashing with neighbouring atoms) allow to compute the SASA of the structure.[133] The InterProSurf[134] was used to determine the number of buried and surface atoms in a complex interface.

All the charts were built using *matplotlib*[135] package, for Python programming language, whereas the interaction plots (interacting residues) were built using the *circlize*[136] (in R). The *ggplot2*[137] (in R) was used regarding the inter-helices distance map construction. In the end was conducted a dynamical analysis (using NMA) to see the following characteristics of the complexes: flexibility change between the two respective receptor structures (monomer and in complex) and fold changes occurring during structure fluctuations, also between OR in monomer and OR in complex. Bio3d[138] was used to evaluate these two parameters. This package is specific for R programming language and is useful to compare protein structures by superimposing them and discarding the regions where the residues have the most conformational variation. This allows to identify the core regions. The flexibility changes were determined based on the Bhattacharyya coefficient (BC), this is a score that measures the similarity between two structures. As the BC value increases the higher the similarity, between the structures, will be.[139] Finally, the flexibility change values were used to build a two-dimensional map trough multidimensional scaling (MS) method.[140]

4. Results and discussion

In total 57 different OR-Partner complexes were analysed: DOR-Arr2_6PWC; DOR-Arr3_6PWC; DOR-Arr2_6U1N; DOR-Arr3_6U1N; DOR-Gi1; DOR-Gi2; DOR-Gi3; DOR-Go; DOR-Gob; DOR-Gz; DOR-G14_6CMO; DOR-G14_6OIJ; DOR-G15_6CMO; DOR-G15_6OIJ; KOR-Arr2_6PWC; KOR-Arr2_6U1N; KOR-Arr3_6PWC; KOR-Arr2_6U1N; KOR-Gi1; KOR-Gi2; KOR-Gi3; KOR-Go; KOR-Gob; KOR-Gz; KOR-Gslo; KOR-Gssh; MOR-Arr2_6PWC; MOR-Arr2_6U1N; MOR-Arr3_6PWC; MOR-Arr3_6U1N; MOR-Gi1; MOR-Gi2; MOR-Gi3; MOR-Go; MOR-Gob; MOR-Gz; MOR-Gslo; MOR-Gssh; MOR-Gq_6CMO; MOR-Gq_6OIJ; MOR-G11_6CMO; MOR-G11_6OIJ; MOR-G15_6CMO; MOR-G15_6OIJ; NOP-Arr2_6PWC; NOP-Arr2_6U1N; NOP-Arr3_6PWC; NOP-Arr3_6U1N; NOP-Gi1; NOP-Gi2; NOP-Gi3; NOP-Go; NOP-Gob; NOP-Gz; NOP-G12; NOP-G14_6CMO; NOP-G14_6OIJ. In some cases, a distinction was made between complexes with the same proteins but modelled or superimposed with different templates. These templates are referenced by their respective Protein Data Bank (PDB) codes.

4.1 Structural analysis

4.1.1 Interhelical distances

An analysis of interhelical distances was made with the purpose to measure the differences in TM3-TM6 and TM3-TM7 distances in order to evaluate the consequences, in these parameters, of different partner couplings. Figure 3 shows that when OR couples with Arrestins_6U1N partners they lead to higher TM3-TM6 and TM3-TM7 distances, when comparing with OR-Arrestins_6PWC. The only exceptions to this observation are the NOP-Arrestins complexes. Although NOP_Arrestins_6U1N have higher TM3-TM6 distance, they display a slightly lower TM3-TM7 distance in comparison with NOP-Arrestins_6PWC. Overall, it is important to note that both highest and lowest TM3-TM7 differences are observed for KOR complexes, when coupled with Arrestins_6U1N and Arrestins_6PWC, respectively. On the other hand, concerning TM3-TM6 distance, the highest difference is also verified for Arrestins_6U1N, when complexed with NOP, while complex KOR-Gob displays the lowest TM3-TM6 distance. Another difference was observed, in interhelical distances, by using different superimposition templates. The OR-Gq/11_6OIJ have higher TM3-TM6 and TM3-TM7 distances than OR-Gq/11_6CMO.

Regarding the G-protein subfamilies, they are generally clustered together, clearly separated from Arrestins. However, in MOR complexes there is more dispersion throughout the TM3-TM7 distance axis. Other exceptions are noticed like, for example, the KOR-Gs complexes, in these cases they are further away from the respective OR-G-protein clusters. All These clusters concerning OR-G-protein complexes seem to be present in a narrow range of TM3-TM6 distance (between 13,5 Å and 15 Å).

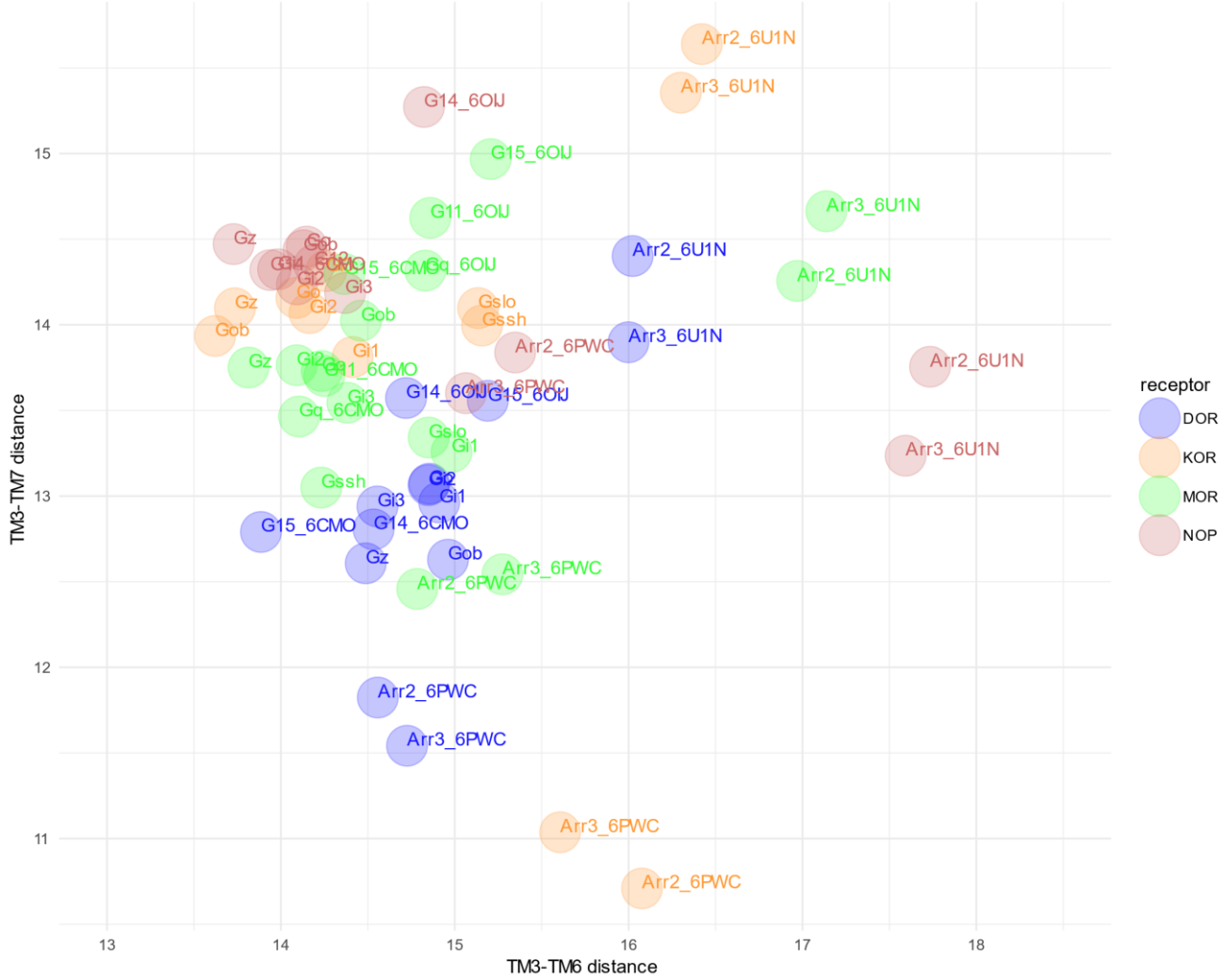


Figure 3:TM3-TM6 and TM3-TM7 interhelical distances in all OR-Partner complexes. These distances were measured between the residues 6.30, or 7.53, and 3.50 (using the Ballesteros-Weinstein numbering)[14]. Each OR has a specific colour, indicated in the map. The map was built using *ggplot2*[137] (in R programming language).

These results show unusual differences when comparing with other GPCR studies. Upon ligand coupling, the GPCR is subjected to conformational changes that allows the intracellular cavity opening. Resolved structures from GPCR in active state show that some of these changes are conserved. For example, the TM6 outward movement is considered as a characteristic motion of the GPCR activation process.[15,141] The extension of this movement is different depending upon the partner, being the TM6 displacement more pronounced when the GPCR couples with Gs, compared with other G-protein subfamilies, and even more in comparison with A rrestins.[119,142] However, the interhelical distances map showed the opposite. In general, the OR have a higher TM3-TM6 distance when coupled with Arrestins than when they are coupled with G-proteins. This assumption directly clashes with the literature. This problem can possibly be template-based (G-protein coupling). The template, in this case, is the KOR, in active state, stabilized by a nanobody. It is known that, in some cases, when the GPCR is stabilized by a peptide, or nanobody, the TM6 movement is smaller when comparing with GPCR-partner complexes.[141,143] This is confirmed by the resolved structure of the rhodopsin (opsin) (PDB ID:

4ZWJ) comparing with β2AR-Gs (PDB ID: 3SN6). In the opsin-peptide complex the TM6 displacement is smaller (6 Å) than the displacement promoted by the β2AR-Gs complex (14 Å).[119,143] Another example is the difference in TM6 movement between the same β2AR-Gs complex and β2AR-nanobody.[119,144]

OR-Arrestins_6PWC complexes have a more pronounced TM7 inward movement than OR-Arrestins_6U1N. Exceptionally, NOP-Arrestins seem to have a similar TM7 displacement between the two groups. The OR-G-protein complexes also have a smaller TM7 inward movement than OR-Arrestins_6PWC, following the same pattern of the OR-Arrestins_6U1N.

There are two possible strategies in order to overcome the discrepancies observed in OR-G-protein complexes results. The first possibility is to use a different template to modulate the receptor in G-protein complexes. Ideally, it should be used an active state of human OR structure. Since it is not available, it could be used one from different species. Another suitable hypothesis as a OR template could be a different GPCR from the class A. Nevertheless, it needs to have a reasonable sequence similarity with OR in order to produce good and reliable homology models. Other alternative is to submit these complexes to extensive all-atom MD simulations. In another study, using a OR bound to Gi, it was found that when the complex was submitted to a long MD simulation the TM6 outward movement was 3.2 Å. The total distance, between TM3 and TM6, went from 13.7 Å to 16.9 Å. It is important to note that this complex was modelled using a KOR template stabilized by a nanobody. The TM3-TM6 distance from the starting structure is very similar to the results obtained in this work, regarding the OR-G-proteins. The only difference is the displacement provoked by the MD simulation.[145]

4.1.2 Electrostatic distances

Electrostatic distances are represented by a heatmap with a colour scale. The lowest distances are represented by the red colour whereas the highest ones are represented by the purple colour. Furthermore, there is a similarity tree, above and at the left of the heatmap, with multiple branches that adds more visual information. The OR-Gi/o complexes are grouped in the same branch with the exception of the OR-Gz, which is separated and fairly distant from its group. Only KOR-Gi/o (Annex 4-Figure C) has all complexes in the same branch. In some cases, the two groups of OR-Arrestins are in different branches (DOR (Annex 4-Figure A) and MOR) whereas in others they are in the same one (KOR (Annex 4-Figure C) and NOP (Annex 4-Figure B)). The same can be applied to OR-Gq/11, where the different groups (superimposed with different templates) are in distinct branches, with no exception. All OR-Gs complexes have low, in general, electrostatic distances in comparison with their subfamily relatives. NOP-Gz seems to be the more related complex with NOP-G12. Figure 4 represents the electrostatic distance heatmap for MOR-Partner complexes.

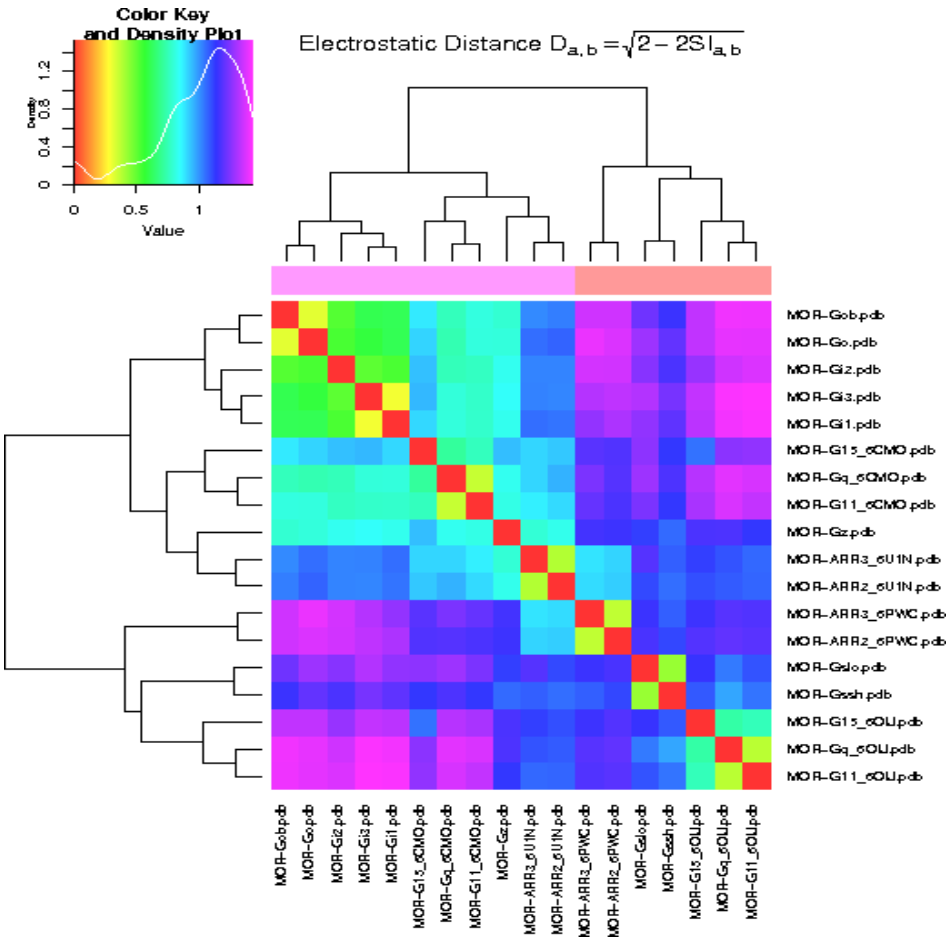


Figure 4: Electrostatic distances from all MOR-Partner complexes. The heatmap was built using PIPSA[128] online server and the distances are marked by different colours. In the equation $D_{a,b}$ represents the electrostatic distance between complex a and b. $SI_{a,b}$ is the similarity index between the two complexes, calculated from each structure electrostatic potential.[128]

Apparently, the use of different superimposition templates has a great impact on the electrostatic distances, even with the use of the same homology modelling templates. This significantly affects the OR-Gq/11 complexes, for example. OR-Arrestins, on the other hand, despite using different modelling templates are in the same branch in half of the heatmaps. Figure 4 clearly shows these discrepancies where the MOR-Gq/11_6OIJ are more related to MOR-Gs than MOR-Gq/11_6CMO. At the same time, the MOR-Arrestins complexes are clustered differently. Regarding the OR-Gi/o and OR-Gs, these complexes show to be clustered by subfamily, being OR-Gz one exception. This is in agreement with a similar study englobing the dopamine receptor family, in which Gz complexes seem to be less related to the remaining OR-Gi/o subfamily partners. This is probably explained by the biggest phylogenetic distance of Gz among this G-protein subfamily.[146]

4.1.3 OR-Partner interactions

As stated before, interaction plots were built in order to disclose specific interaction patterns between OR-Partner complexes. In terms of residue numbering (receptor) the TM residues follow the Ballesteros-Weinstein numbering[14], whereas the ICL residues follow the Table 3 numbering. In the case of the partners the residues follow the GPCRdb[4,5] numbering. The interaction pattern symbology chosen to this study has some rules: the residues that participate in the interactions are defined by the one-letter amino acid code, in which only the first and the last residue have the respective numbering; "x" symbol represents an amino acid position that does not interact but connects different interacting amino acids inside the same subdomain.

The two tables (Table 4 and 5) below show the main specific interaction patterns for OR-Partner complexes.

Table 4: Specific interaction patterns in the OR

Interaction patterns	Complexes
$R^{3.50}x_2A - V/I^{3.54}$	OR-Arrestins_6U1N; OR-Gq/11_6CMO; OR-Gq/11_6OIJ
$A{-}V/I^{3.54}$	OR-Gi/o
$A^{3.53}V$	OR-Gs
$V^{113/114/121}/I^{114}x_2LD^{117/118/125}$	OR-Arrestins_6PWC
$P^{112/113/120} - V/I - K/Rx_2D^{117/118/125}$	OR-Arrestins_6U1N
$P^{112/113/120} - V/Ix_2LD^{117/118/125}$	OR-Gq/11_6OIJ; OR-Gs
$S^{207/209/214}G$	OR-Gi/o; OR-Gq/11_6CMO
$LSG^{208/210}$	OR-Gs
$E^{6.25}x_7Ix_2M/L^{6.36}$	OR-G14_6OIJ; OR-G15_6OIJ
$E^{6.25}Kx_2NLRRI^{6.33}$	OR-Gs
$D^{8.47}E$	OR-Arrestins_6PWC
$L^{7.56} - D^{8.47}ENF$	OR-Gi/o
$L^{7.56} - D^{8.47}EN$	OR-Gq/11_6CMO; OR-Gs
$Y^{7.53}x_2L^{7.56} - D^{8.47}ENF$	OR-G14_6OIJ; OR-G15_6OIJ

Table 5: Specific interactions patterns for G-proteins and Arrestins partners

Interaction patterns	Complexes
$R^{33/37/40} - R/G$	OR-Gq/11_6OIJ
$D^{67/68}x_2VL^{71/72}$	OR-Arrestins_6PWC
$Y^{63/64} - G^{64/65}x_4DVLGL^{73/74}$	OR-Arrestins_6U1N
$N^{194/198}/S^{198}/K^{201} - I^{195/199}/V^{199}/T^{202}$	OR-Gq/11_6OIJ
$F^{244/245} - N/S$	OR-Arrestins_6PWC
$R^{285/286}G$	OR-Arrestins_6U1N
$E^{318/319}xY - T^{321/322}/C^{321}/S^{322}$	OR-Gi/o
$V^{319}/I^{323}/R^{338} - I/L - Y/F - S^{322/326/341}$	OR-Gq/11_6CMO
$V^{319}/I^{323}/R^{338}$	OR-Gq/11_6OIJ
$S^{335/349}Tx_2GDGx_2Y^{344/358} - C^{345/359}Y$	OR-Gs
$D^{341/342}x_2Ix_2N^{347/348}$	OR-Gi/o
$D^{342/346/361}x_2Lx_2 - N/Y - Lx_2 - F/Y/I - NL^{354/358/373}$	OR-Gq/11_6CMO
$K^{341/345}/R^{360}x_2 - I/V - Lx_2 - N/Y - L - R/K/D - E - F/Y/I - NL^{354/358/373}$	OR-Gq/11_6OIJ
$R^{366/380}x_2IQx_2HLRQYELL^{380/394}$	OR-Gs

4.1.4 OR-Arrestins interaction plots

The OR subdomains that interact with Arrestins modelled with NTSR1 (Arrestins_6PWC), are the ICL2, TM6 and H8 in all cases, while the TM2, TM3, ICL2, TM5 and TM6 are the prevalent interacting OR subdomains when modelled with M2R (Arrestins_6U1N). There are no interactions with the TM1 subdomain for any complexes containing Arrestins. Concerning specific subdomain interactions, the ICL1 interacts with Arrestins (finger loop) in only three complexes, all modelled by the NTSR1 template, through the K^{41} amino acid. TM2 also interacts with the finger loop of the Arrestins (through $T^{2.39}$ residue) in almost all complexes, with the

exception of DOR-Arrestins_6PWC (Figure 6 and Annex 5-Figure A) and KOR-Arrestins_6PWC (Annex 6-Figures A and B). The TM3 seems to be involved in interactions mainly in OR-Arrestins_6U1N complexes. For OR-Arrestins_6PWC complexes, TM3 only interacts with KOR-Arrestins_6PWC complexes (through $AV^{3.54}$ residues). In the OR-Arrestins_6U1N complexes the interaction pattern ($R^{3.50}x_2A-V/I^{3.54}$) differs slightly from KOR-Arrestins_6PWC. This subdomain interacts with the finger loop domain, this conception does not apply to NOP-Arr3_6U1N where it was observed that the TM3 interacts also with the C-loop, as can be seen in Figure 5.

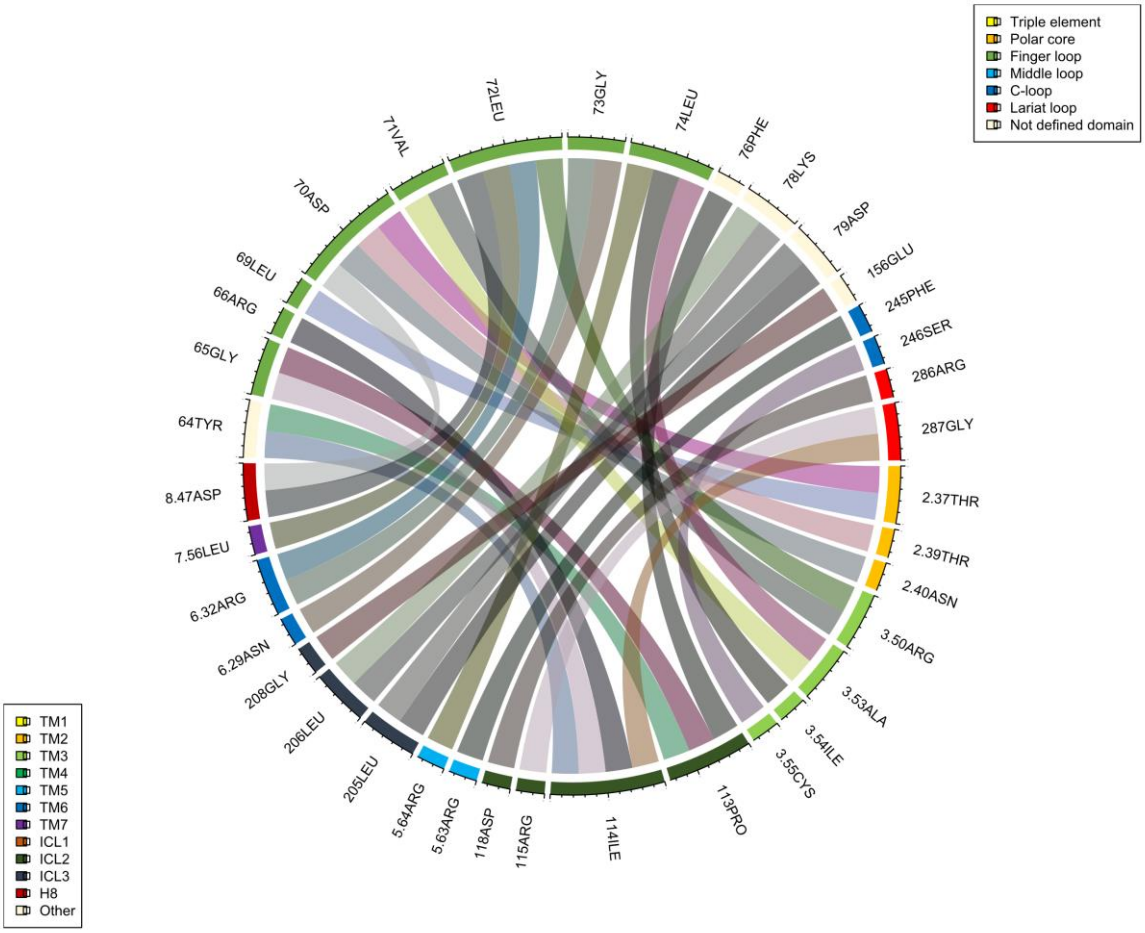


Figure 5: Interaction plot of the NOP-Arr3_6U1N complex. Several motifs (specific interaction patterns) are represented in this plot, such as: the TM3 motif ($R^{3.50}x_2A-V/I^{3.54}$) or the ICL2 motif ($P^{112/113/120}-V/I-K/Rx_2D^{117/118/125}$). Like all interaction plots, this one was built using circlize.[136]

The ICL2 interacts in all OR-Arrestin complexes with similar interaction patterns, namely $V^{113/114/121}/I^{114}x_2LD^{117/118/125}$ for OR-Arrestins_6PWC and $P^{112/113/120}-V/I-K/Rx_2D^{117/118/125}$ for OR-Arrestins_6U1N. Figure 6, which has the interaction plot of the DOR-Arr2_6PWC complex, shows the interaction of ICL2 with arrestins C-loop ($F^{244/245}-N/S$ motif). This is a common interaction for all OR-Arrestins_6PWC. For KOR complexes and MOR-Arr2_6PWC (Annex 7-Figure A) ICL2 also interacts with

the finger loop. This marks a difference to OR-Arrestins_6U1N complexes, in which the ICL2 interacts with the finger loop and the lariat loop ($R^{285/286}G$ motif).

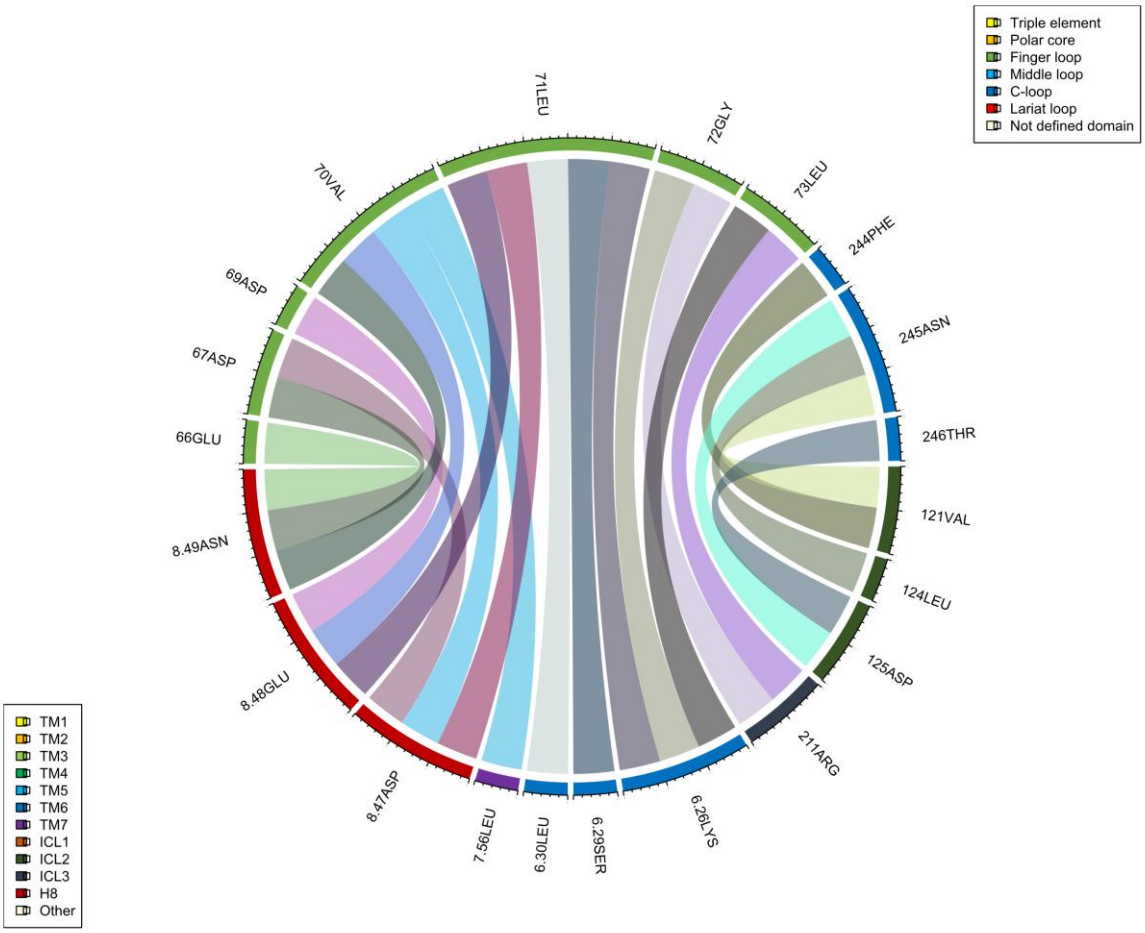


Figure 6: Interaction plot of the DOR-Arr2_6PWC complex. In this case ICL2 interacts only with C-loop through a specific interaction pattern ($V^{113/114/121}/I^{114}x_2LD^{117/118/125}$ motif).

Not a single interaction was observed between TM4 and Arrestins subdomains. TM5 interacts, in all OR-Arrestins_6U1N complexes ($R^{5.64}$), with the C-loop ($F^{244/245}$) and finger loop. On the other hand, there is no presence of interactions involving this subdomain in OR-Arrestins_6PWC, with the exception of both KOR-Arrestins_6PWC complexes. Remarkably, the ICL3 does not interact in NOP-Arrestins_6PWC (Annex 8-Figures A and B) and KOR-Arrestins_6U1N (Annex 6-Figures C and D), being present in all other Arrestins complexes.

In all OR-Arrestins complexes, TM6 is involved in interactions through a common residue ($S/N^{6.29}$). This residue interacts majorly with the finger loop. $L^{7.56}$ is also an amino acid widely present in OR-Partner interactions including OR-Arrestins, however, there are a few exceptions like the MOR-Arrestins_6U1N (Annex 7-Figures C and D) and NOP-Arr3_6PWC (Annex 8-Figure B). $L^{7.56}$ is interconnected with the

remaining interacting residues from the H8 helix ($D^{8.47}E$ for OR-Arrestins_6PWC and $D^{8.47}$ for OR-Arrestins_6U1N, except for KOR-Arrestins_6U1N and MOR-Arr3_6U1N).

From the Arrestin side, beyond the already mentioned interactions with C-loop and lariat loop, the majority of the interactions occur with the finger loop. There are distinct interaction patterns for OR-Arrestins_6PWC ($D^{67/68}x_2VL^{71/72}$) and OR-Arrestins_6U1N ($G^{64/65}x_4DVLGL^{73/74}$). In all cases, the finger loop interacts with several OR subdomains at the same time.

Some interesting features are shown in OR-Arrestins interactions. One of the most evident differences between OR-Arrestins_6U1N and OR-Arrestins_6PWC is the number of interactions between the receptors and the partners. Clearly, the latter complex group has a smaller number of interactions comparing with the former. The same was observed in the dopamine receptor family.[146] This could be related with the 90º rotation of Arrestin 2 in NTSR1-Arrestin 2 structure comparing with M2R-Arrestin 2. At the same time, the finger loop is the most interacting subdomain for both complex groups. This reflects the way the arrestins interact and couple with OR. Other studies showed that after receptor recognition by the arrestin (through the phosphorylated C-terminus), the finger loop would insert in the intracellular cavity interacting with several receptor residues, changing the inactive conformation to a more extended one (active state). However, remains to be seen if this behaviour is common in all GPCR-Partner complexes.[110,111,147,148]

It is reported, in the NTSR1-Arrestin 2 structure[111] (PDB ID: 6PWC), that the finger loop makes part of the complex interface. The finger loop, in this structure, makes contacts with TM7 and H8 through several residues, ranging between the E^{66} to L^{71}. In OR-Arrestins_6PWC complexes ($D^{8.47}E$ motif) the $L^{68/69}$ residue is only involved in interactions in the KOR-Arrestins_6PWC, being absent from the other ones. Furthermore, the $L^{68/69}$ residue, in KOR-Arrestins_6PWC, bounds to TM3. Although $D^{67/68}$ bounds to TM7 and H8 residues in almost all complexes, the KOR-Arrestins_6PWC shows a different behaviour. $D^{67/68}$, beyond the already shown interactions, also interacts with TM3 and ICL2 in these two complexes (ICL2 only interacts with this residue in the KOR-Arr2_6PWC). In the NTSR1-Arrestin 2 structure it is reported that the TM5, TM6, ICL1 and ICL2 surround the finger loop in the intracellular cavity, however, as shown before, in some complexes the finger loop interacts with other OR subdomains. This differences between the modelling template and the KOR-Arrestins_6PWC are noteworthy. It is likely that the NTSR1-Arrestin 2 structure represents only one possible conformation of the complex. Moreover, it is referred that the Arrestin 2 does not insert so deeply comparing with the visual arrestin. Therefore, the interactions between this Arrestin subdomain and TM7-H8 residues are far more dynamic. This hypothesis can possibly explain the different behaviour seen in KOR-Arrestins_6PWC. Other interactions with TM5 were observed in OR-Arrestins_6PWC, even though in the template interface there are no interactions of this kind. On the opposite, the TM6 interacts uniquely with the finger loop, which is in line with the template 6PWC. Some complexes have direct interactions between TM5 and the finger loop, although the NTSR1-Arrestin 2 does not have residues from these two subdomains interacting with each other. The ICL1 (K^{41}) interacts with the finger loop

in some complexes but is absent from others. These differences, as said before, could be linked with the significant dynamic of the finger loop in the complex interface.[111]

Other significant differences were seen in the complex models, when comparing with the NTSR1-Arrestin 2 modelling template. The Arrestins in OR-Arrestins_6PWC appear to interact with ICL1 in only three cases and all of them are interacting with the finger loop, therefore implying that, for the other cases, the ICL1 does not have a role on interactions with OR, unlike what happens with the template. The ICL2, on the other hand, is present in all complex models, establishing interactions with finger loop and C-loop, instead of the lariat loop as reported in the template. This draw the hypothesis that ICL2 is not deeply inserted into the reported cleft (with middle loop, bottom loop and lariat loop), present in arrestin, and interacts with finger loop present in the intracellular cavity. However, further studies involving long MD simulations will be necessary to better understand the structural and dynamic behaviour of this loop and its interaction with Arrestins. The ICL3 interaction pattern is in agreement with the NTSR1-Arrestin 2 structure, even though only one interacting ICL3 residue is involved in interactions, in all cases. This is a remarkable difference to other GPCR, with long ICL3, where there are multiple ICL3 residues interacting with the receptor, pointing out the different ways of the arrestin recognition.[111,142]

The OR-Arrestins_6U1N have a more embracing finger loop motif ($Y^{63/64} - G^{64/65} x_4 DVLGL^{73/74}$) comparing with the OR-Arrestins_6PWC ($D^{67/68} x_2 VL^{71/72}$). This is aligned to the differences found in both templates regarding the finger loop interactions. In the NTSR1-Arrestin 2, the finger loop interacts through a small handful of residues with TM7 and H8 (as stated before), while the M2R-Arrestin 2 finger loop has more interacting residues and, at the same time, interacts with more GPCR subdomains. Like the template, two particular interactions seem to be conserved in all OR-Arrestins_6U1N and includes E/DRY motif (TM3) residues: the interaction between $D^{69/70}$ and $T^{2.39}$ or $V^{70/71}$ and $V^{3.50}$. The change of $R^{3.50}$ (in NTSR1) by a $V^{3.50}$ (in OR) seems do not affect the interaction profile.[110] On the other hand, the interaction with $E^{8.48}$ (NPx_2YxF motif) is absent in OR-Arrestins_6U1N but it is present in the other Arrestin complex group.[149] The finger loop motif interacts, extensively, with TM6, TM5, TM3, TM2 and ICL2, giving a possible indication of the OR subdomains surrounding the intracellular cavity. The template structure also refers to the ICL2 special positioning inside a hydrophobic cleft between the two Arrestins domains. The results obtained, in the modelled complexes, show that ICL2 interacts with residues from both domains confirming the authors conclusions.[110] The V2 vasopressin receptor (V2R)-Arrestin 2 complex crystal implied the possibility of conformational change and engagement of the middle loop with the receptor residues, however these results show no interactions containing the middle-loop. However a different conformation may be adopted by V2R-Arrestin 2 changing the interaction profile comparing with the complexes in study.[26,148]

A comprehensive work published by *Mafi et al.*, using multiple biased and nonbiased agonists, showed that ICL2, ICL3 and TM6 are crucial to stabilize the OR-Arrestin complex through polar interactions (hydrogen bonds and salt bridges). Even more important, they affirm the existence of two salt bridges that appeared in two different complexes, one with a full agonist and the other with a partial agonist. These interactions have

ICL2, ICL3 or TM6 participation. The paper reported strong affinity between the ICL2 and the Arrestin 3 (complex with full agonist). This affinity is reflected by the formation of a group of polar interactions that stabilize and regulate arrestin binding, which may be a possible explanation for the fact that almost all ICL2 residue interact with Arrestin in OR. This is valid for both OR-Arrestins_6PWC and OR-Arrestins_6U1N. One particular interaction, between ICL2 ($D^{118/125/127}$) and the lariat loop ($R^{285/286}$), seems to be conserved in OR-Arrestins_6U1N which is not verified in OR-Arrestins_6PWC. This linkage is crucial to connect both components of the complex and allows the finger loop to establish other polar interactions with the receptor. The other interactions, with finger loop participation, combined with the results found in this study confirm a conserved group of interactions, further validating the paper results. The authors affirm that the $D^{67/68}$ residue participates in a hydrogen bond formation with $T^{2.37}$, however, in OR-Arrestins_6PWC it seems that $D^{67/68}$, as said before, interacts with multiple OR domains but not with the TM2. On the other side, OR-Arrestins_6U1N have a similar residue playing that role, the $D^{69/70}$, which could form a possible hydrogen bond with the TM2 residue. In some cases, other residues like the $T^{2.38}$ or the $T^{2.39}$ also interact with $D^{69/70}$ at the same time. It is possible that they also form hydrogen bonds with the finger loop. Altogether, the different results presented by the OR-Arrestins_6PWC, might be derived from the different conformation of the arrestin when comparing with OR-Arrestins_6U1N. The authors used, as template, the rhodopsin-visual arrestin structure (PDB ID: 5WOP)[150] which has a similar conformation to the M2R-Arrestin 2 structure.

The $D^{67/68}$ interacts, extensively, with H8 residues. Particularly, the interactions with $N^{8.49}$, $E^{8.48}$ or $D^{8.47}$ may form hydrogen bonds with $D^{67/68}$ in order to stabilize the complex. Nonetheless, some interactions reported by the authors are not verified in this study like, for example, the interaction between $R^{119/120/127}$ (in ICL2) and $R^{65/66}$ (in finger loop). This interaction is important because allows the reorientation of $D^{67/68}$ or $D^{69/70}$ in order to form the hydrogen bond with the TM2. These complexes may have a different interaction that proportionate the reorientation of the residue.[147]

The second salt bridge was not verified in OR-Arrestins complexes, although there is the presence of a possible hydrogen bond between $L^{206/208/213}/L^{207/209/214}$ (in ICL3) and $K^{77/78}$ (with the exception of KOR-Arrestins_6U1N and MOR-Arrestins_6U1N) in OR-Arrestins_6U1N. Furthermore, a second hydrogen bond (also consequence of the salt bridge formation) may possible be present in OR-Arrestins_6U1N but was not verified in the other complex group (OR-Arrestins_6PWC). The interaction between $K^{77/78}$ and $V^{5.68}$ (with the exception of NOP-Arrestins_6U1N (Figure 5 and Annex 8-Figure C)) adds more proof to the presence of a second salt bridge.

The authors also used a different complex, with a partial agonist, and found a different interaction (salt bridge) containing a ICL3 residue ($R^{204/206/211}$) instead of the TM6 residue. In fact, the arginine is present in some complexes, surprisingly even in some OR-Arrestins_6PWC, but interacts with residues from the N-domain instead of the C-domain, contrary to what happens in the complex with the partial agonist. Nevertheless, to

produce any conclusive evidence is necessary to submit the complexes to a long MD simulation to have a more definitive proof about the type of interactions established between each OR-Arrestin complex.[147]

4.1.5 OR-Gi/o interaction plots

The Gi/o group comprises 24 complex models in total, in which the prevalent interacting OR subdomains are the TM3, ICL3 and H8. No interactions involving the ICL1 and TM1 are reported for this subgroup. $T^{2.39}$ and $T^{2.37}$ amino acids are the TM2 residues that participate in interactions with Gi/o proteins (H5 subdomain). Nonetheless, these residues do not appear in OR-Gz, DOR-Go (Annex 9-Figure C), KOR-Go (Annex 10-Figure D) or NOP-Gi3 (Annex 12-Figure C) complexes. TM3 interacts through a common interaction pattern ($A{-}V/I^{3.54}$) for all Gi/o proteins. These residues interact with H5 subdomain like what happens with TM2. The ICL2, with a few exceptions (DOR-Gz (Annex 9-Figure E) and MOR-Gz (Annex 11-Figure F)), also shows interactions with other G-protein subdomains ($P^{112/113/120}$ and $V^{113/114/121}/I^{114}$). Besides the interactions with H5 subdomain there are also particular interactions between $V^{113/114/121}/I^{114}$ and hns1 subdomain (R/K^{32}) in some complexes.

TM5 ($V^{5.68}$) interacts with H5 subdomain except in 3 MOR-Gi/o complexes. The ICL3 ($S^{207/209/214}G$) interacts with multiple G-protein subdomains (h4s6, S6 and H5) with a specific interaction pattern. $I^{6.33}$ is the TM6 residue commonly present in this subgroup (except OR-Gz complexes). Lastly, the TM7 and H8 share an interaction pattern ($L^{7.56} - DENF^{8.50}$) with the exception of OR-Gz complexes. As stated before, the H5 subdomain interacts with multiple OR subdomains simultaneously, as can be seen in Figure 7.

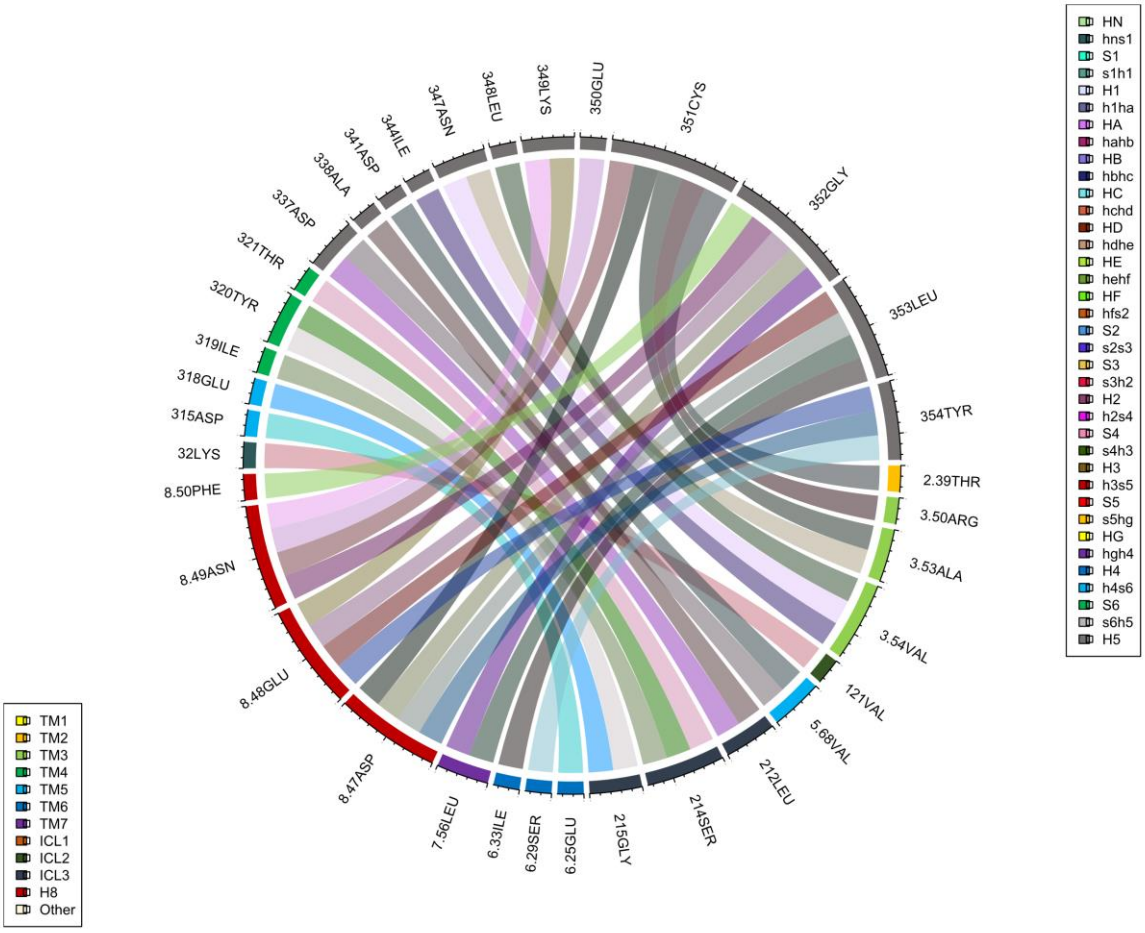


Figure 7: Interaction plot of DOR-Gi3 complex. The H5 subdomain interacts (with the $D^{341/342}x_2Ix_2N^{347/348}$) with multiple OR subdomains like the H8 and TM7 ($L^{7.56}-DENF^{8.50}$ motif), or the ICL2 ($S^{207/209/214}G$ motif).

This subdomain has an interaction pattern ($D^{341/342}x_2Ix_2N^{347/348}$) and represents, by a large margin, the majority of the interactions from the G-protein side. The h4s6 and S6 also participate in interactions with OR as already told before. Curiously these subdomains largely interact with ICL3, however there are also, in some cases, reported interactions with TM5 and TM6, these subdomains share an interaction pattern for the Gi/o subgroup ($E^{318/319}xY-T^{321/322}/C^{321}/S^{322}$).

In OR-Gi/o complexes, and ultimately in all OR-G-protein complexes, the H5 subdomain comprises the majority of G-protein residues that interact with the receptors. The H5 is a α-helix located in the G-protein C-terminus and has a similar role to the finger loop, penetrating in the intracellular cavity during the activation process. A study using a mimetic peptide of the finger loop (complexed with rhodopsin) and directly comparing with similar complex but with the C-terminus of the Gt protein (Gi/o subfamily). The results indicated a common interaction motif between the finger loop (arrestins) and the H5 helix (G-proteins), this motif was identified with the following nomenclature: $E/Dx-I/Lx_3GL$. Thus, having a similar recognition pattern. However, this study showed no correlation between Or-Arrestins finger loop motifs and OR-Gi/o H5 motifs.

In fact, no motif showed similarity to the aforementioned one despite some common residues, particularly in OR-Arrestins. Other contradicting results were found, for example, the low presence of interacting residues from TM5 and TM6 (in OR-Gi/o). One of the big differences between the GPCR-Arrestins and GPCR-G-proteins is the role of TM5 and TM6 in the interaction with either the finger loop or the H5 helix. Supposedly, the number of TM6 and TM5 residues interacting with the H5 should be higher comparing with the finger loop but, in a significant number of cases, the number of TM6 residues interacting with the finger loop is higher than the residues interacting with H5 ($I^{6.33}$ motif).[149] Nonetheless, two important leucines were found in some OR-Gi/o complexes (L^{348} and L^{353}). These leucines, when mutated, make the OR-Gi/o coupling unviable. They appear in several different active structures, being common in their binding interfaces.[118]

An active KOR-Gi structure[145] (submitted to long MD simulation) was proved useful to make a comparison with the obtained OR-Gi/o complexes. For example, the authors report a strong salt bridge between the R^{32} and the $D^{117/118/125}$ (ICL2). This salt bridge is not present in OR-Gi/o complexes, instead there are interactions between this particular arginine with proline and valine residues from ICL2.This salt bridge has an important stabilizing role, similar to the salt bridges enunciated before in the OR-Arrestins. Another salt bridge involving the ICL1 was reported, however ICL1 residues do not interact with H5 (through D^{312}) in the OR-Gi/o, even though it seems this ionic interaction is present in other GPCR with different partners (Gs). Another missing salt bridge in the OR-Gi/o is the interaction of $K^{6.26}$ and the $E^{318/319}$. The lysine is completely absent in the complexes whereas the glutamate interacts with ICL3 instead of the TM6, except in a few cases. This salt bridge has a major regulatory role in the interactions between TM6 or ICL3 with the H5.[145]

4.1.6 OR-Gq/11 interaction plots

The Gq/11 subgroup is divided between those superimposed by rhodopsin-Gi structure and M1R-G11 structure. TM3, ICL3, TM6, TM7 and H8 are subdomains present in all OR-Gq/11_6CMO whereas TM2, TM3, ICL2, TM4, ICL3, TM6, TM7 and H8 are present in all OR-Gq/11_6OIJ. A specific TM2 residue ($T^{2.39}$) interacts in almost OR-Gq/11_6CMO (except NOP-G14_6CMO (Annex 15-Figure A)) and in all OR-Gq/11_6OIJ with the H5 subdomain. Both groups share a common interaction pattern in TM3 ($R^{3.50}x_2A-V/I^{3.54}$), this subdomain overwhelmingly interacts with H5. The OR-Gq/11_6OIJ complexes have a specific interaction pattern ($P^{113/120}-V/Ix_2LD^{118/125}$) in ICL2, more residues in this subdomain interact with Gq/11_6OIJ proteins than in the OR-Gq/11_6CMO complexes, where not all complexes have interactions containing this subdomain (like in the case of MOR-G15_6CMO (Annex 14-Figure E).

TM4 ($T^{4.38}$) only participates in interactions with G-proteins in OR-Gq/11_6OIJ complexes, presenting a remarkable difference between these two subgroups. The threonine is a common TM4 residues between all these complexes despite the presence of other residues in some cases. These residues interact with HN and hns1 residues. TM5 ($V^{5.68}$) does not interact in OR-Gq/11_6OIJ, unlike what happens with OR-Gq/11_6CMO

with the exception of MOR-G11_6CMO (Annex 14-Figure C) and MOR-G15_6CMO. In the ICL3 both subgroups have distinct interaction patterns for OR-Gq/11_6CMO ($S^{207/214}G$) and OR-Gq/11_6OIJ ($G^{208/215}$) however they seem to interact with the same subdomains (h4s6 and S6) with the addition of the H5 for the OR-Gq/11_6CMO complexes. The residue $I^{6.33}$ is shared by all complexes of this group, however more residues participate in the interaction with G-proteins, like what happens with the ICL3, in this case, both groups also link with common G-protein subdomains (h4s6 and H5).

Lastly in the TM7 and H8 similar interactions can be found comparing with the Gi/o group, in this scenario the interaction pattern is $L^{7.56} - DEN^{8.49}$, the residues interact with H5.

From the G-protein side, what was seen for the H5 is similar between Gi/o and Gq/11 groups in the sense that H5, in both cases, makes the majority of interactions (in G-protein) with OR and interacts with multiple subdomains. The interaction patterns are $D^{342/346/361}x_2Lx_2 - N/Y - Lx_2 - F/Y/I - NL^{354/358/373}$, for OR-Gq/11_6CMO, and $K^{341/345}/R^{360}x_2 - I/V - Lx_2 - N/Y - L - R/K/D - E - F/Y/I - NL^{354/358/373}$ for OR-Gq/11_6OIJ.

Other G-proteins subdomains have an important role in the interaction profile like the interaction pattern $V^{319}/I^{323}/R^{338} - I/L - Y/F - S^{322/326/341}$ in h4s6 and S6 subdomains (in OR-Gq/11_6CMO), in the OR-Gq/11_6OIJ only the residues $V^{319}/I^{323}/R^{338}$ are common in this subgroup. Is important to note that these subdomains overwhelmingly interact with ICL3 and TM6. Common residues in OR-Gq/11_6OIJ were also found, they are present in hns1 ($R^{33/37/40} - R/G$) and in s2s3 and S3 subdomains ($N^{194/198}/S^{198}/K^{201} - I^{195/199}/V^{199}/T^{202}$).

A small number of OR-Gq/11 complexes were analysed in this study but, nonetheless, some interesting results were found. As said before, in this complex group two different superimposition templates were used. Until now there are only two GPCR structures in active state and coupled with Gq/11 partners deposited in public databases. One of them was used to superimpose (M1R-G11 structure), so in order to make a fair comparison between the two complex groups it was used the other available structure: the 5-hydroxytryptamine receptor 2A (5-HTR_{2A}) coupled with Gq. Two conserved residues in the Gs subfamily were identified in the structure and may be linked to the subfamily specificity: $Q^{233/237/240}$ and $N^{240/244/247}$ residues. However, in the obtained interaction plots these residues are missing.[151]

One notable difference between the two complex groups is the amount of interacting ICL2 residues. In the OR-Gq/11_6OIJ ($P^{112/113/120} - V/Ix_2LD^{117/118/125}$ motif) there are more ICL2 residues, interacting with the receptors, than in the OR-Gq/11_6CMO, giving a possible evidence of a small rotation of the H5 (Gq/11_6CMO) compared with H5 in the other group. The $V/I^{113/114/121}$(ICL2) is an important conserved residue and has direct role in the Gq/11 association with GPCR. This residue participates in the OR-Gq/11_6OIJ interface but is not present in OR-Gq/11_6CMO, this may be direct consequence of the interface similarity between the M1R-G11 and 5-HTR_{2A}-Gq. On the other hand, this proves the importance of the presence of this residue in the complex interface containing Gs proteins. In muscarinic receptors, an interaction

(with $R^{33/37/40}$ and ICL2 arginine residue) is conserved and it seems to be fundamental during the Gq/11 coupling process. A similar interaction was also seen in the serotonin receptor. In the OR-Gq/11 complexes the $R^{33/37/40}$ also interacts with ICL2 residues in similar positions ($D^{118/125}$ and $L^{117/124}$) and, interestingly, with a TM4 ($T^{4.38}$ motif) residue. Also, in some cases, interacts with R^{120} (in MOR-Gq/11_6OIJ complexes). This may indicate a similar hydrogen bond formation comparing with the other already determined GPCR-Gq/11 structures. These interactions are absent in OR-Gq/11_6CMO, which is comprehensible due to the fact of the superimposition template has a different conformation (GPCR-Gi/o coupling) comparing with the M1R-G11 or 5-HTR_{2A}-Gq. Apparently, the TM5 does not participates in interactions with Gq/11_6OIJ subdomains. But it is important to refer the possibility of direct interaction between TM5 and TM6 residues in order to stabilize the TM6, which couples with Gq/11.[124,151]

4.1.7 OR-Gs and OR-G12/13 interaction plots

For the Gs group was found that the common interacting OR subdomains are: TM3, ICL2, ICL3, TM6, TM7 and H8. $A^{3.53}V$ residues are common in this group and they interact with H5. ICL2 also has a specific interaction pattern for OR-Gs ($P^{112/113}Vx_2LD^{117/118}$), beyond the interactions with H5 the ICL2 also interacts with hns1 ($R^{38}A$) and S3 ($V^{203/217}$) as seen in Figure 8.

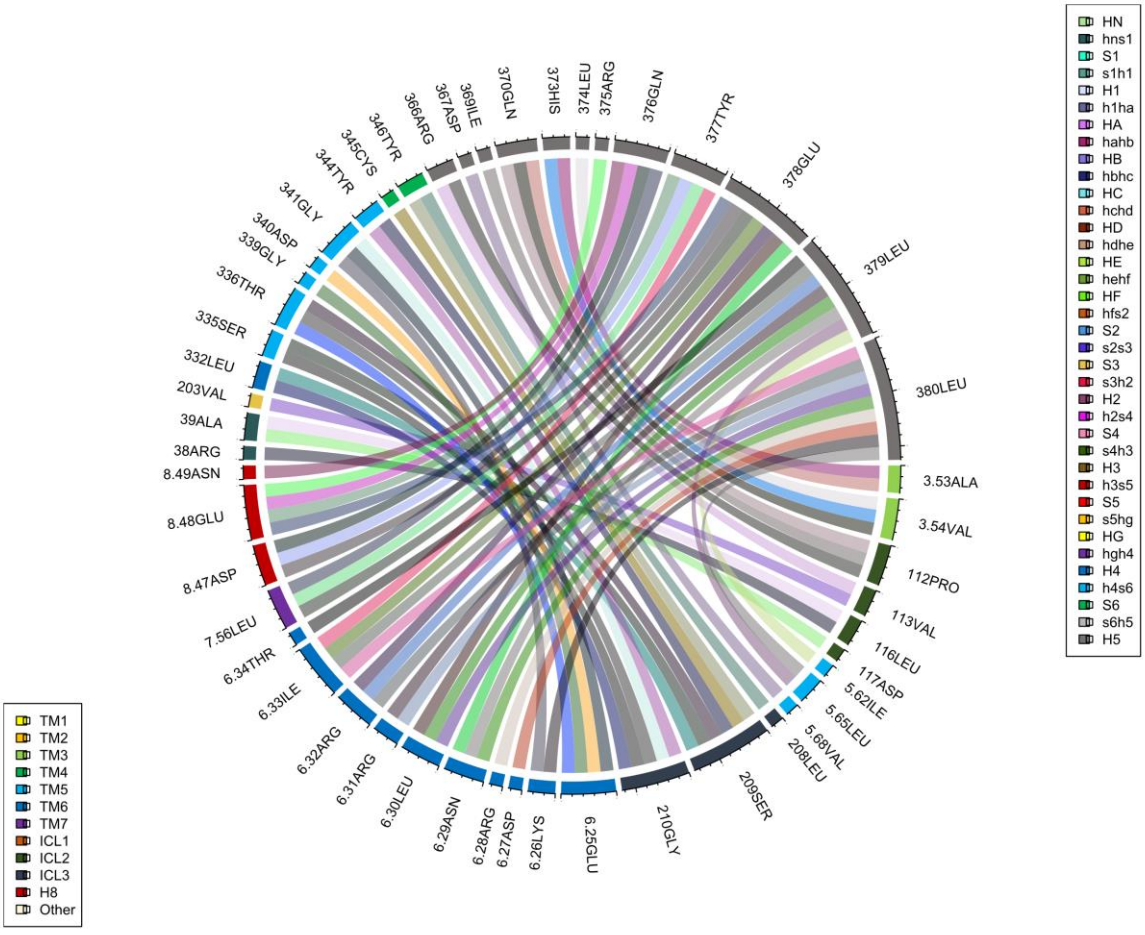


Figure 8: Interaction plot from the KOR-Gssh. Were found some interactions patterns specific for OR-Gs ($E^{6.25}Kx_2NLRRI^{6.33}$ motif) in TM6 or in H5 ($R^{366/380}x_2IQx_2HLRQYELL^{380/394}$ motif).

There are also TM5 interactions with the exception of the MOR-Gslo (Annex 16-Figure C). $LSG^{208/210}$ is the specific ICL3 interaction pattern for OR-Gs complexes, this subdomain interacts with h4s6 ($S^{335/349}Tx_2GDGx_2Y^{344/358}$) and S6 ($C^{345/359}Y$) residues. The TM6 interaction pattern differs greatly from other groups ($E^{6.25}Kx_2NLRRI^{6.33}$), these TM6 residues, like the ICL3, also interact with S6. Finally, the TM7 and H8 completely interact with H5 through the following interaction pattern: $L^{7.56} - DEN^{8.49}$. In the G-protein side is important to mention that the H5 ($R^{366/380}x_2IQx_2HLRQYELL^{380/394}$) also interacts with multiple OR subdomains, repeating a similar behaviour from other G-protein groups. There is also a common residue from H4 ($L^{332/346}$) interacting with ICL3.

NOP-G12 (Annex 17) complex is the only example of the OR-G12/13. The $T^{2.39}$ residue also interacts in this complex with H5. The $R^{3.50}x_2AI^{3.54}$ motif, widely common in OR-Partner, interacts exclusively with H5. The $RLLSG^{208}$ motif (ICL3) residues contact with H5 and S6 (LFH^{348}) subdomains. The $L^{7.56} - DENF^{8.50}$ motif, similar in OR-Gi/o, also has interactions with H5. The H5, like what happens in the majority of OR-G-proteins, has a wide coverage in the interface of the complex interacting with multiple OR subdomains ($HA^{365}x_2D^{368}x_2L^{371}x_2NLK^{376}xIMLQ^{381}$ motif).

Only 4 OR-Gs complexes were analysed within the scope of this study, nonetheless some interesting features were observed. Some active state GPCR-Gs structures are already available like the β2AR-Gs crystal structure. One important interaction, that allows the Gs coupling specificity, involves the F^{139} (in β2AR), the equivalent residue ($V^{113/114}$) is also present in OR-Gs complexes. The valine interacts with two coinciding Gs residues (also present in β2AR-Gs structure interactions), in this case the $R^{366/380}$ and the $V^{217/203}$ residues.[119]

One difference regarding these complexes and the β2AR-Gs structure is the TM7-H8($L^{7.56} - D^{8.47}EN$) motif presence. While in the adrenoreceptor there are no interacting residues in these two subdomains, in OR-Gs complexes they interact extensively with H5. This behaviour is similar to the A_{2A}AR-Gs structure.[152] Previous studies demonstrated the existence of a selectivity barcode for GPCR-Gs with common interacting residues in G-proteins (important during GPCR activation) but also specific residues, for each G-protein, crucial to selectively bind to a specific partner. Some of these residues were also found in the OR-Gs interface. The $R^{366/380}x_2IQx_2HLRQYELL^{380/394}$ motif, for example, contains several positions identified as crucial for Gs selectivity. In β2AR-Gs, these residues interact with TM5 and ICL3, with few contacts containing the TM6 and ICL2. In this work it seems the ICL3 is the only exception to this assumption because it has no meaningful interactions with H5.[153]

4.1.8 Residue interaction percentages

Several metrics were analysed to give a full description of the OR-Partner complex interface, one of them (interaction percentages) allowed to understand the most important residues and residue classes in complex

interfaces. The OR-G-protein complexes (receptor side) showed higher percentage of arginine and leucine residues except in DOR-Gz (Annex 21) where the arginine presence is similar to glutamate and valine values, DOR-Gi1 (Annex 21) and DOR-Gi2 (Annex 21) also shown similar values between arginine and valine.
KOR-Gz (Annex 21) and KOR-Gi3 (Annex 21) repeated the same pattern from DOR-Gi1 and DOR-Gi2 in respect to the comparison of arginine and valine percentages. MOR-Gq/11_6CMO (Annex 21) have a higher leucine residue number comparing with MOR-Gq/11_6OIJ (Annex 21). MOR-Gslo (Annex 21) has a higher leucine residue number comparing with MOR-Gssh (Annex 21). A vast majority of the OR-G-protein complexes have low percentages and even in some cases no presence of cysteine, glutamine, histidine, phenylalanine and tryptophan residues. OR-Gi/o, OR-Gs and OR-Gq/11_6CMO complexes have no presence of tyrosine residues, the opposite was seen regarding the OR-Gq/11 complexes, the only exception to this fact is MOR-G11_6CMO (presence of tyrosine).
The residue percentages from the partner side show a different situation, in this case for many complexes the leucine residue is not the most prevalent residue in the interface, however with few exceptions. In general, the Gi/o group most frequent residue is the aspartate, a huge difference of the aspartate values between OR-Gi/o and OR-Gs, OR-Gq_6CMO and OR-Gq_6OIJ (Annex 20) was denoted, whereas the opposite was seen with arginine values. Nevertheless, is important to mention that there is no presence of tryptophan residues in all complexes.
The Or-Arrestin complexes (partner side) also show some interesting results, the leucine is the most frequent residue ranging between ~15% to ~25% of the total residues in the interface however the leucine values are substantially higher in OR-Arrestins_6U1N comparing with OR-Arrestins_6PWC. Arginine values are higher for OR-Arrestins_6PWC with the exception of the NOP-Arrestins complexes where the opposite occurs. Aspartate and glycine are more frequent residues in OR-Arrestins_6U1N interfaces, on the other hand glutamine and phenylalanine residues are more frequent residues in OR-Arrestins_6PWC. Interacting glutamine, histidine, methionine and tryptophan residues are absent from the interface.
In the receptor side the same pattern (comparing with the partner side) in the arginine percentages can be seen. In general, the leucine percentages are substantially lower in relation to the partner side, like what happens with partner the leucine percentages are higher in OR-Arrestins_6U1N (Annex 21). Aspartate percentages are lower than in the partner side. There is a big difference concerning the NOP-Arrestins isoleucine percentages (considerably higher) comparing with the remaining complexes. Interacting glutamine and tryptophan residues are absent from the receptor.
The residue group interaction percentages (partner side) also give important information about the GPCR-Partner interfaces. The OR-Gi/o complexes have higher percentage of nonpolar aliphatic residues with the exception of OR-Gz where the acid negative residues are more numerous in DOR-Gz (Annex 18), in KOR-Gz the acid negative, nonpolar aliphatic and the polar uncharged residues have similar percentages, in MOR-Gz (Annex 18) the acid negative and polar uncharged residues have higher percentage values than nonpolar aliphatic residues, and finally the NOP-Gz (Annex 18) follows the same pattern of the KOR-Gz. In the OR-

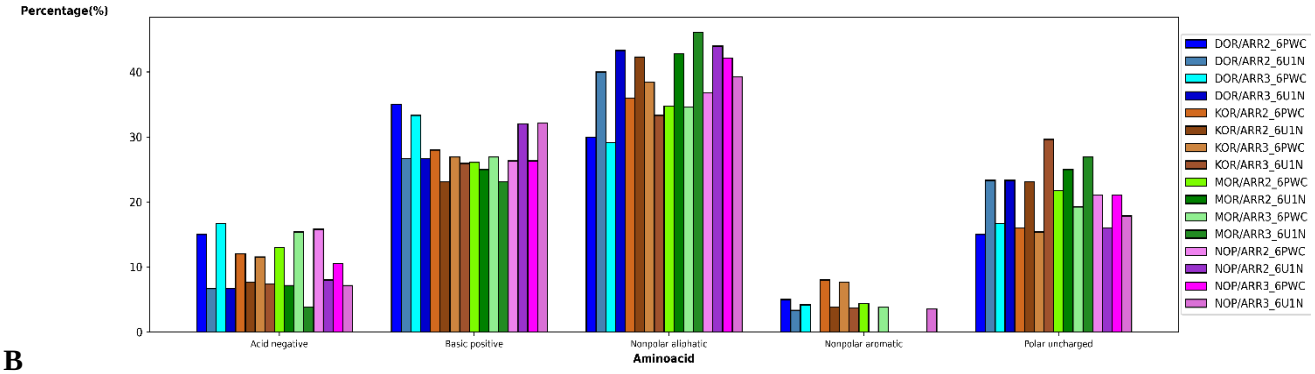
Gs (Annex 18) complexes the basic positive residues are the most numerous residue group with a slight lead to nonpolar aliphatic residues. OR-Gq/11 complexes have identical percentage values between basic positive and nonpolar aliphatic residues however no patterns were extracted from this data to differentiate OR-Gq/11_6CMO (Annex 18) and OR-Gq/11_6OIJ (Annex 18). In the receptor side, unlike what happens in the partner side, for all complexes, with no exceptions, the biggest percentage values belong to nonpolar aliphatic residues. Remarkably, interacting nonpolar aromatic residues are almost non-existent in the receptors, even in the KOR-G-protein complexes (Annex 19) this group is completely absent.

The OR-Arrestins (partner side) the most prevalent aminoacid group is nonpolar aliphatic, nonetheless the OR-Arrestin_6U1N have higher percentages values than OR-Arrestin_6PWC. The same conclusion can be applied to the acid negative group. The basic positive and nonpolar aromatic groups show different patterns, in these cases the OR-Arrestins_6PWC have higher percentage values than OR-Arrestins_6U1N.

In the receptor the most prevalent aminoacid group is the nonpolar aliphatic with the exceptions of the DOR-Arr2_6PWC and DOR-Arr3_6PWC. Nonpolar aromatic percentages are considerably lower than the partner percentages. The OR-Arrestins_6PWC complexes have higher percentages, for acid negative residues, than the OR-Arrestins_6U1N. Same pattern happens with the basic positive residues with the exception of NOP-Arrestins complexes. The opposite happens with polar uncharged residues where the OR-Arrestins_6U1N have higher percentages comparing with OR-Arrestins_6PWC, with the exception of the NOP-Arrestins complexes.

The role of electrostatic interactions in the formation of the GPCR-Partner complex is crucial because promotes affinity during the partner coupling with the receptor. Although is important to denote the importance of hydrophobic contacts in the stabilizing process of the complex. The results obtained in these complexes show subtle differences in previous studies on active state structures. Regarding the OR-Arrestins they contradict some of the established knowledge about the nature of the interface interactions.[147,154] Supposedly, the arrestin is positively charged whereas the receptor is negatively charged (total net charge). In Figure 9A, however, the receptors have more positively charged residues, with a clear difference between the acidic and basic residue percentages. Although this could be influenced by the absence of the C-terminus in the receptors. The arrestin recognizes the receptor, with the negatively charged phosphorylated residues in C-terminus, through the basic residues in the arrestin N-domain.[24,147,154] It is reported that the intracellular cavity has a plentiful of positive charges that allows the formation of electrostatic interactions with the finger loop (rich in negatively charged residues). This characteristic is fundamental because of the incredible diversity of GPCR, allowing the arrestins to couple with so many different receptors despite their differences.[111,142,154]

A



B

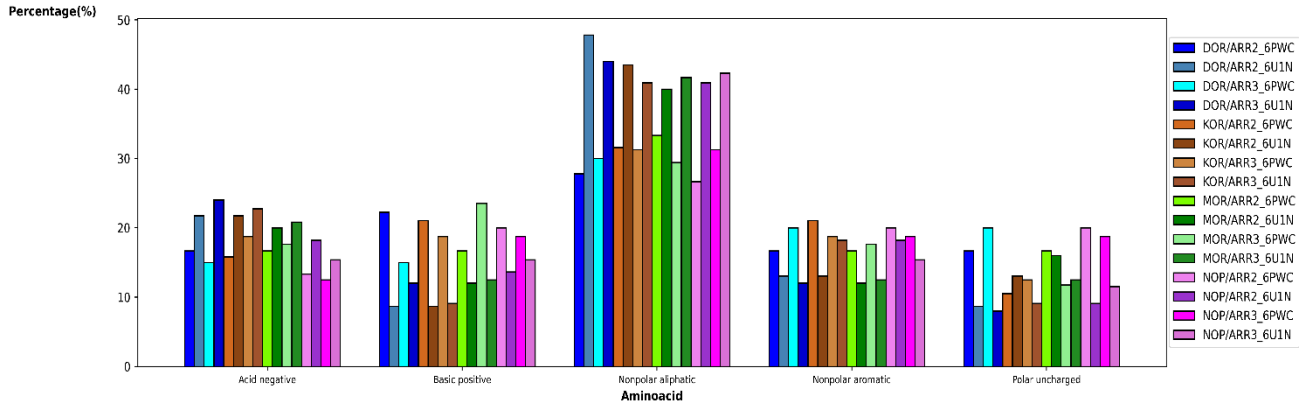


Figure 9: Electrostatic profile of the OR-Arrestins interface. **A)** Residue group interaction percentage values to all aminoacid groups (in OR) **B)** Residue group interaction percentage values to all amino acid groups (in arrestins) from OR-Arrestins complexes.

Taking into account that the finger loop comprises the majority of the interacting residues in the arrestin, is easy to understand that what is expected is more basic residues in the receptor and more acidic residues in the partner. However, the OR-Arrestins_6PWC do not follow this assumption regarding the electrostatic nature of the interacting residues from the arrestin side (Figure 9B). A more detailed analysis on the specific residues that constitute the OR-Arrestins_6PWC interface confirm the results in Figure 9 (see Figure 10).

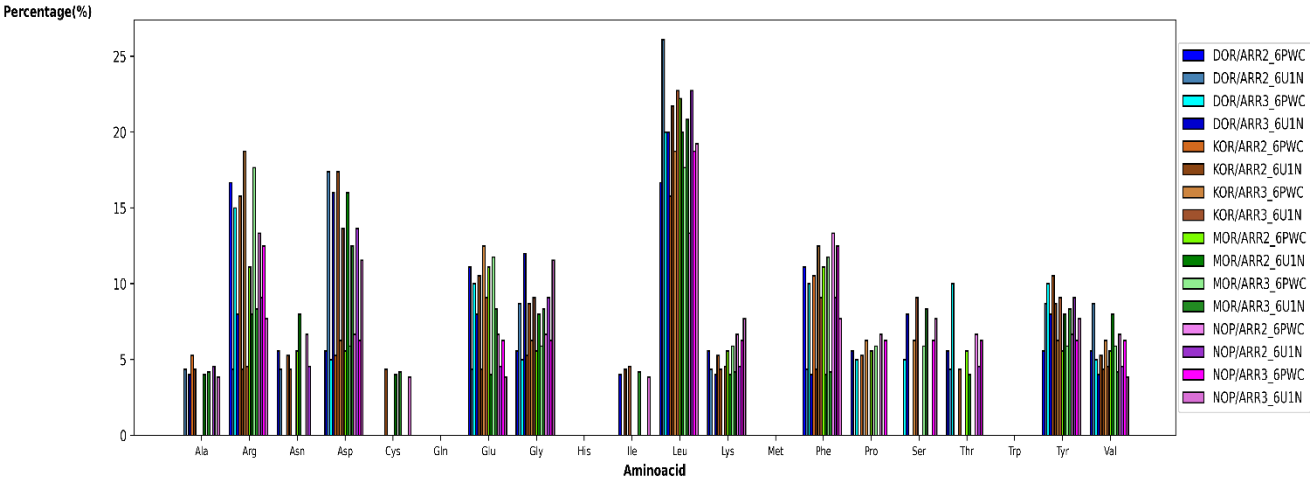


Figure 10: Interaction percentages for each residue type from OR-Arrestins complexes (in arrestins)

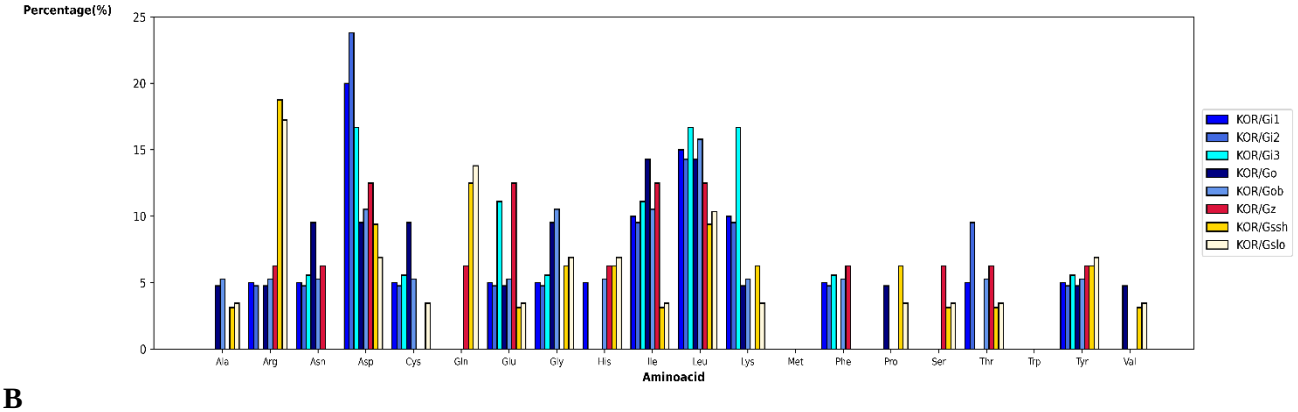
The arginine interaction percentages (in OR-Arrestins_6PWC) are incredibly inflated compared with the other percentages, from the same residue, in the other complex group modelled by the M2R-Arrestin 2 structure. It is important to note that arginine has positive charge whereas the aspartate has negative charge. The other acidic residues (histidine and lysine) are not sufficient to offset the sum of the glutamate and aspartate percentages.

The biggest number of interacting residues belong to the nonpolar aliphatic group (both in the receptor and partner) with few exceptions. In M2R-Arrestin 2 interface the finger loop interacts with the receptor with a mixture of hydrophobic and electrostatic interactions.[110] A comparison of the obtained finger loop motifs (see OR-Partner interactions section) confirms the presence of hydrophobic residues in both motifs. Nonetheless, the share of hydrophobic residues (finger loop) is substantially higher in OR-Arrestins_6U1N being a possible explanation for the fact that this complex group has higher interaction percentage in the nonpolar aliphatic group compared with the other one.

In relation to OR-G-protein complexes a curious situation happens. Indeed, the OR-Gq/11 and OR-Gs (in the G-protein side) seems to have a high number of basic residues comparing with the acidic residues, instead the OR-Gi/o (Annex 18) has the opposite. However, in the receptor, the positive charges outnumber, in general, the negative charges in the complex interfaces, creating a hindrance for the electrostatic interaction formation during the complex creation for the OR-Gq/11 and OR-Gs cases. In the D2 dopamine receptor-Gi1 complex, the surface potential of the receptor binding interface is overwhelmingly positive whereas in the G-protein is clearly negative, indicating the important presence of residues from opposite charges in both sides of the complex interface and their role during the complex formation.[155] Other complexes, like the KOR-Gi previously described, also report fundamental electrostatic interactions in coupling regulation. These interactions also follow the same charge pattern as the dopamine receptor.[145] This are in accordance with OR-Gi/o results where it seems the ionic residues are distributed at the same way of the already mentioned

complexes. Active state structures containing Gs also describe the same patterns seen before in GPCR-Gi/o.[152] Thus, it was supposed that OR-Gs would present similar results, outstandingly this was not observed. Instead, high interaction percentage levels, from arginine residue, were seen in the Gs binding surface (Figure 11A). The hydrophobic contacts, like what happens with the OR-Arrestins, are the most prevalent interaction in the binding interface (for OR-Gi/o). The nonpolar aliphatic group has the bigger interaction percentage, however in the OR-Gq/11 and OR-Gs both nonpolar aliphatic and basic positive residues have similar percentages in the G-protein side. In fact, sometimes, the basic residue group surpasses the nonpolar aliphatic one (see Figure 11B).

A



B

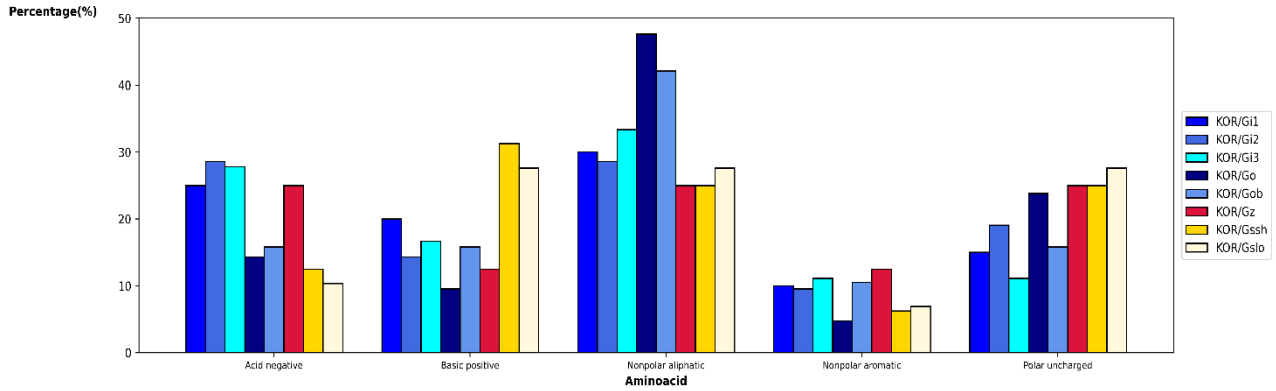


Figure 11: A) Interaction percentages for each residue type from KOR-G-protein complexes (in G-proteins) **B)** Residue group interaction percentage values to all aminoacid groups (in G-proteins) from KOR-G-protein complexes

The CXC chemokine receptor 2 bound to Gi showed the central role from two leucines (hydrophobic residues), that are present in H5, and the formation of extensive hydrophobic contacts with other hydrophobic residues (leucines, isoleucines, valines, …) in the TM that surround this helix.[156] It was observed the same leucine residues in the cannabinoid receptor-Gi complex.[157] Other GPCR-Gi complexes also report similar results.[158,159] This kind of interactions inserts inside a hydrophobic pocket containing the H5 and the other TM

already mentioned. This may well explain why the nonpolar aliphatic group is predominant in GPCR-Gi interface. The hydrophobicity, in the binding cavity, seems a feature that is present in other complexes containing different G-protein subfamilies.[151] However as mentioned before, and taking into account the importance of H5 subdomain in the OR-G-protein interface, the OR-Gq/11 and OR-Gs complexes seem to not have a clear favouring for hydrophobic residues, at the same time the electrostatic factor is not in place for these groups because of the preference for basic residues in the receptor and partner.

4.1.9 SASA

SASA values are very low for all complexes in ICL1 with the exception of OR-Arrestins_6PWC complexes. In ICL2 the OR-Gs (Annex 22), OR-Gq/11_6CMO (Annex 22) and OR-Gq/11_6OIJ (Annex 22) have higher SASA than OR-Gi/o (Annex 22), with the exception of MOR-Gq_6CMO which has lower SASA than MOR-Gi1. In the same loop the OR-Arrestins_6U1N (Annex 22) have higher SASA than OR-Arrestins_6PWC (Annex 22). In general, higher SASA, comparing with the other GPCR subdomains, is present in ICL3 for all complexes with the exception of OR-Arrestins_6PWC and OR-Arrestins_6U1N. OR-Gz complexes have a significant lower SASA than the remaining OR-Gi/o complexes. In H8 a similar pattern was also found regarding the OR-Gz and OR-Gi/o SASA differences. OR-Arrestins_6PWC have higher SASA, in H8, than OR-Arrestins_6U1N.

The SASA results are related with the receptor available area to interact with partners. If a OR subdomain has low SASA thus it will have residual interactions with either arrestin or G-protein. The ICL1 results are in line with the interacting residues results (interaction plots). The differences seen between OR-Arrestins_6PWC and OR-Arrestins_6U1N may be explained by the templates used. The NTSR1-Arrestin 2 mentions interactions containing ICL1 residues whereas the M2R-Arrestin 2 has no ICL1 participation in the complex interface.[110,111] Other study showed important interactions, with ICL1 residues, between KOR and Gi protein.[145] However, it seems this subdomain is irrelevant in OR-G-proteins interfaces based on the interaction plots and SASA results. Regarding ICL2 and ICL3, it was predictable that OR-Partner complexes had higher SASA values, than ICL1 or H8, because these structures are fundamental in GPCR-partner coupling. In OR-Arrestins_6U1N the ICL2 has an abnormally higher SASA values, comparing with ICL3, than other complexes. This is justified by the M2R-Arrestin 2 interaction profile where the ICL2 interacts, in conjunction with the receptor TM, with the arrestin interdomain (polar core).[110] The H8 difference between the two OR-Arrestins subgroups is remarkable and is in line with the interaction plots results. In the circular graphics (for OR-Arrestins_6PWC) the H8 has a more important role, in interacting with the arrestins, than the importance of this structure in OR-Arrestins_6U1N.

4.1.10 HB and SB

The HB and SB are totally absent from ICL1 in DOR-G-protein (Annex 23) and KOR-G-protein (Annex 23) complexes. In the MOR-G-protein (Annex 23) complexes only MOR-G11_6OIJ and MOR-G15_6OIJ have

HB and SB in ICL1. Only NOP-G14_6OIJ has HB and SB (ICL1) in NOP-G-protein (Annex 23) complexes. ICL2 HB and SB are present in more complexes than in ICL1 case. The exceptions are: KOR-Gi2, KOR-Gz, KOR-Gslo, MOR-Gz, MOR-Gq_6CMO, NOP-Gi1, NOP-Go and NOP-Gob. DOR-Gi3 have the highest number of HB and SB in ICL3 in DOR-Gi/o complexes whereas the DOR-G14_6CMO have the highest number in DOR-Gq/11 complexes. Considering the same intracellular loop, KOR-Gs complexes have more HB and SB than KOR-Gi/o with the exception of KOR-Gob which has the same number of HB and SB than KOR-Gslo. MOR-Gi2 and MOR-G15_6OIJ have the lowest number of HB and SB considering, respectively, MOR-Gi/o and MOR-Gq/11 complexes. NOP-Gi1 has the highest number of HB and SB (in NOP-G-protein complexes). In the H8, OR-Gz have no HB and SB unlike the other complex subgroups with the exception of MOR-Gq_6OIJ.

Unlike what happens with the majority of OR-G-protein complexes, many OR-Arrestins have HB and SB in ICL1 (except DOR-Arr2_6U1N, KOR-Arr3_6U1N, MOR-Arr3_6PWC, NOP-Arr2_6U1N and NOP-Arr3_6U1N). DOR-Arr3_6U1N, KOR-Arr2_6PWC, KOR-Arr3_6PWC, MOR-Arr2_6U1N, MOR-Arr3_6PWC and NOP-Arr3_6U1N have no HB and SB in ICL2. Only KOR-Arr2_6U1N has no HB and SB in ICL3. Finally, only OR-Arrestins_6PWC have HB and SB in H8.

The absence of HB or SB in ICL1, from OR-Gi/o and OR-Gs complexes, is in line with the results in the interaction plots. The ICL1 seems to not be involved in the coupling process, with a few exceptions from the OR-Gq/11 group. Other GPCR showed, previously, that ICL1 is not a crucial subdomain for the complex assembling.[124,157,160,161] However, the OR-Gq/11 complexes, that reveal these kind of contacts, present novel interactions that may demonstrate a more important role for ICL1 and can possible be specific for OR.

On the contrary, ICL2 forms HB or SB in almost all OR-G-proteins complexes. The already mentioned study, involving the KOR-Gi complex submitted to MD simulation, demonstrated that ICL2 forms a significant number of HB and SB with the H5 subdomain, however no mention was made about the formation of these kind of interactions with ICL1, being at the same level of the results obtained in this work.[145] In other structures, with different G-proteins, appear also HB with ICL2 residues participation, clearly demonstrating a more participative role than ICL1 in GPCR-G-protein complexes.[124,160,162,163]

ICL3, in general, seems to have a significantly superior number of HB and SB compared with the other subdomains in study. The ICL3, similar to what happens with ICL2, is considered as being a major component of the binding interface, so it is not surprising to see elevated numbers from this parameter.[124] The secretin receptor ICL3, for example, reports more stable electrostatic interactions and HB than the ICL2, so these are not the first GPCR to have this kind of pattern.[163] The number of HB and SB, in H8, is somewhat comparable to those in ICL2 albeit with some differences. Van der Waals interactions are formed in the A_{2A}AR-Gs structure where the glutamate residue in H5 makes interactions with arginines between TM7 and the H8.[152,159] On the other hand, an electrostatic interaction is seen, in the bile acid receptor, between glutamate in Gs and arginine in H8[164], so the formations of SB between H8 from OR and H5 from G-proteins are not discarded at

all. Nonetheless, this pattern is similar to the interactions between the finger loop and the intersection between the two aforementioned GPCR subdomains, thus is not surprising to see these results.

OR-Arrestins have some remarkable differences, compared to OR-G-proteins complexes, like a clear pattern of HB or SB formation using ICL1 residues. ICL2 and ICL3 also form a considerable amount of HB and SB, with few exceptions. The study, using MOR-Arrestin 3 submitted to long MD simulation, showed the large presence of SB in ICL2 and ICL3, even using different agonists. These SB actuate as complex stabilizers, pointing to a possible similar role for the interactions found in this work.[147]

β1 adrenoreceptor-Beta arrestin structure also pointed the presence of HB in ICL1 and ICL2, although the same was not seen in ICL3 for this case.[165]

The most striking difference between OR-Arrestins_6PWC and OR-Arrestins_6U1N is, without doubt, the H8 subdomain. In fact, only Or-Arrestins_6PWC has HB and SB in H8 as shown in Figure 12.

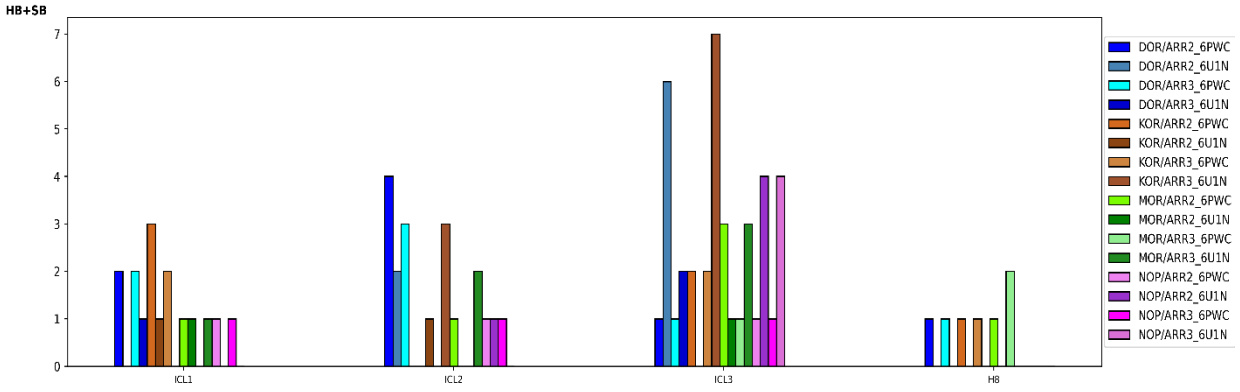


Figure 12: Sum of the HB and SB in OR-Arrestins in 4 specific OR subdomains

This difference may possibly be due to the use of different templates. The NTSR1-Arrestin 2 structure reports several interactions between negatively charged finger loop residues (glutamate and aspartate) and the residues situated in the junction containing the TM7 and H8.[111] The MOR-Arrestin 3 structure study also mentioned a SB between finger loop and this subdomain, being in line with the results found in the complexes modelled with this template.[147] The M2R-Arrestin 2 also mentions the formation of HB and SB with finger loop residues (aspartate), however it is not pointed to the H8 region, instead forms these interactions with TM2 and TM3.[110] Other studies confirm the possibility of a SB creation between finger loop and H8.[149]

4.1.11 Surface atoms/buried atoms

The number of buried and surface atoms revealed that OR-Arrestins_6U1N has a higher number of buried atoms than OR-Arrestins_6PWC, with the notable exception of MOR-Arrestins complexes where the opposite happens (Figure 13). For OR-G-protein, the number of buried atoms is considerably higher for OR-Gs (Annex 24) comparing with the remaining groups. Complexes superimposed with M1R-G11 present more buried atoms than those superimposed with rhodopsin-Gi structure, except in the MOR-G15_6CMO (Annex 24) and

MOR-G15_6OIJ (Annex 24). OR-Gz (Annex 24) has the lesser number of buried atom within the OR-Gi/o group (Annex 24).

A similar metric, widely used to study GPCR-partner complexes, is the buried surface area. This parameter is directly related to the number of buried atoms in a complex interface. OR-Gs have, by far, the largest number of these kind of atoms comparing with other complexes including OR-Arrestins.

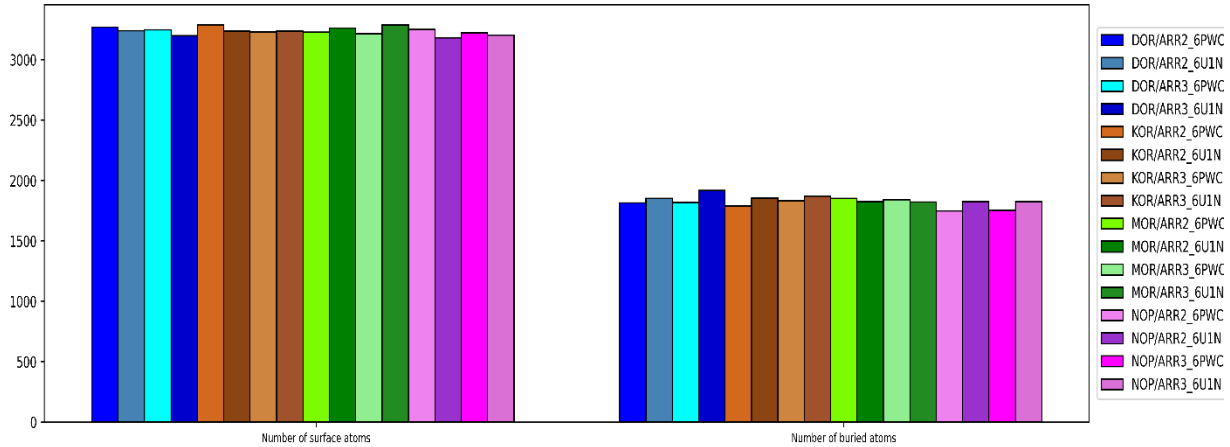


Figure 13: Number of surface and buried atoms in OR-Arrestins complexes

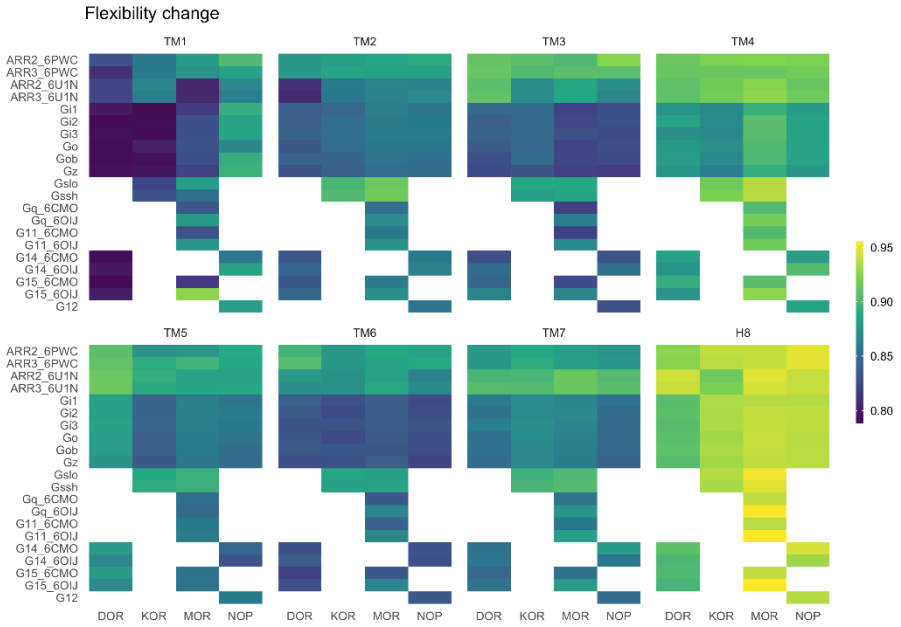
This reaffirms the results presented in other GPCR-partner structures. A big difference, regarding the buried surface area, was found between the rhodopsin-arrestin (1350 $Å^2$)[142] and the β2AR-Gs (2576 $Å^2$)[119]. The NTSR1-Gi1 (1199 $Å^2$)[160] structure also allowed to make a comparison with OR-Gi/o complexes. Despite this group has a smaller number of buried atoms comparing with OR-Gs, it seems that is not the case comparing with OR-Arrestins. These two complex groups have very similar values for this parameter.

4.2 Dynamic analysis

OR-Arrestins and OR-Gs have slightly higher BC values than OR-Gi/o and OR-Gq/11 complexes. Regarding OR-Gi/o, the DOR and KOR complexes in TM1 have particularly low BC values in comparison with MOR and NOP complexes. The same differences apply to OR-Gq/11 in TM1. In general, H8 structure has the highest BC values for almost all complexes (Figure 14A).

The average fluctuation fold change values show a clear distinction between OR-Arrestins and OR-Gs: in all OR subdomains these complexes have lower average fluctuation than OR-Gi/o and OR-Gq/11. TM1 and TM4, in particular, have high average fluctuation values, in OR-Gi/o and OR-Gq/11, than the remaining OR structures (Figure 14B).

A


Flexibility change
TM1
TM2
TM3
TM4
TM5
TM6
TM7
H8
ARR2_6PWC
ARR3_6PWC
ARR2_6U1N
ARR3_6U1N
Gi1
Gi2
Gi3
Go
Gob
Gz
Gslo
Gssh
Gq_6CMO
Gq_6OIJ
G11_6CMO
G11_6OIJ
G14_6CMO
G14_6OIJ
G15_6CMO
G15_6OIJ
G12
DOR
KOR
MOR
NOP
0.95
0.90
0.85
0.80

B

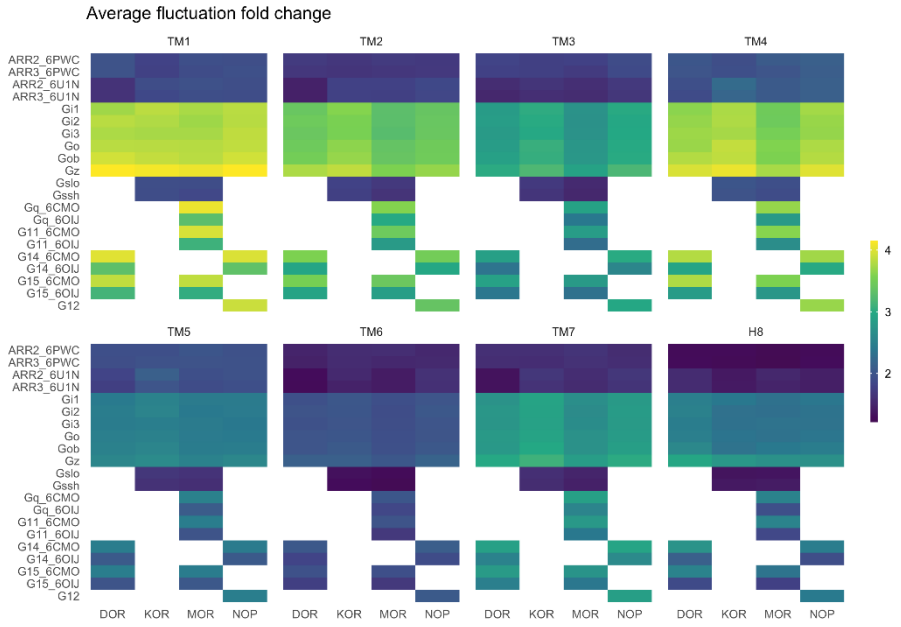

Average fluctuation fold change
TM1
TM2
TM3
TM4
TM5
TM6
TM7
H8
ARR2_6PWC
ARR3_6PWC
ARR2_6U1N
ARR3_6U1N
Gi1
Gi2
Gi3
Go
Gob
Gz
Gslo
Gssh
Gq_6CMO
Gq_6OIJ
G11_6CMO
G11_6OIJ
G14_6CMO
G14_6OIJ
G15_6CMO
G15_6OIJ
G12
DOR
KOR
MOR
NOP
4
3
2

Figure 14: A) Values for flexibility changes, between OR in monomer and OR in complex with partners, in all OR subdomains, measured through BC **B)** Values for average fluctuation fold changes between OR in monomer and OR in complex with partners, in all OR subdomains **C)** Map built, using MS method, with the flexibility changes values. The dots size is directly proportional to average fluctuation fold change values.

The MS map shows a very clear distinction between OR-Arrestins/OR-Gs and OR-Gi/o. The OR-Gq/11 seems to be near the OR-Gi/o subgroup with few exceptions. Only DOR-Arr2_6U1N and DOR-Arr3_6U1N are completely distant from their respective group (Figure 14C).

The dynamic analysis results clearly demonstrate some differences between different OR-Partner subgroups. In general, the coupling between OR and OR-Gi/o and OR-Gq/11, promotes higher flexibility in OR structures than in OR-Gs and OR-Arrestins. Similar results were found in the dopamine receptors.[146] These results also demonstrate that H8 is a rigid structure, even after the conformational changes, during the formation of the complex, the BC value is very high indicating strong similarity between the H8 structures in OR monomer and OR in complex.

At the same time the flexibility change values allowed to separate very well the several complex subgroups with the exception of the two structures of DOR-Arrestins_6U1N.

5. Conclusions

Future work may be centred on the use of a different homology modelling template for OR-G-proteins complexes, in order to overcome the results found in interhelical distance map. For example, using the active state MOR receptor (mouse) bound to Gi/o protein (PDB ID: 6DDF). Moreover, extensive MD simulation should be applied in order to provide further insights about the dynamic behaviour of the complexes, as well as the specific interactions that underlie OR-Partner selectivity.

This work provided a thorough description of the different OR-Partner structures with interesting results that are important for the development of new drugs that can interfere with complex binding interfaces and be a key weapon to solve the opioid crisis problematic. 57 models were built, and they covered a wide array of different partners. The interhelical distance map showed results that are not in line with literature. OR-Gi/o complexes had lower TM3-TM6 distance than OR-Arrestins, however resolved structures, available in public databases, demonstrated the opposite. The electrostatic distances clearly demonstrated differences between complexes. For example, the impact of using different superimposition templates on the similarity between structures modelled with the same template.

Interacting residues in the binding interface were found and more important specific interaction patterns could be extracted. In particular, interaction pattern differences were found between the OR-Gq/11subgroups, like for example: the $D^{342/346/361}x_2Lx_2 - N/Y - Lx_2 - F/Y/I - NL^{354/358/373}$ motif in OR-Gq/11_6CMO and $K^{341/345}/R^{360}x_2 - I/V - Lx_2 - N/Y - L - R/K/D - E - F/Y/I - NL^{354/358/373}$ motif in OR-Gq/11_6OIJ.

The electrostatic profile of the interactions, which form the complex interface, also had results for OR-Arrestins_6PWC that are the opposite of what was found in the literature. There is a bigger number of basic residues than acidic residues in the two sides of the binding interface. On the other hand, the OR-Arrestins_6U1N seem to follow the results from other studies regarding the electrostatic profile. This is incompatible for the formation of electrostatic interactions, although these interactions are regarded as crucial for the GPCR-Partner coupling. Nonetheless is important to mention the different arrestin conformations adopted in both templates. Other structural parameters (SASA, number of surface and buried atoms) were in line with the literature. This work also demonstrated that OR-Gi/o and OR-Gq/11 are highly dynamic complexes comparing with OR-Arrestins and OR-Gs, concerning the average fluctuation fold change and flexibility change values.

6. References

1. Lu, M. & Wu, B. Structural studies of G protein-coupled receptors. *IUBMB Life* **68**, 894–903 (2016).
2. Wold, E. A., Chen, J., Cunningham, K. A. & Zhou, J. Allosteric Modulation of Class A GPCRs: Targets, Agents, and Emerging Concepts. *J. Med. Chem.* **62**, 88–127 (2019).
3. Wang, J., Hua, T. & Liu, Z. J. Structural features of activated GPCR signaling complexes. *Curr. Opin. Struct. Biol.* **63**, 82–89 (2020).
4. Kooistra, A. J. *et al.* GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res.* 1–9 (2020). doi:10.1093/nar/gkaa1080
5. Munk, C. *et al.* GPCRdb: the G protein-coupled receptor database – an introduction. *Br. J. Pharmacol.* **173**, 2195–2207 (2016).
6. Katrich, V., Cherezov, V. & Stevens, R. C. Structure-Function of the G-protein-Coupled Receptor Superfamily. *Annu. Rev. Pharmacol. Toxicol.* 531–556 (2013). doi:10.1146/annurev- pharmtox-032112-135923
7. Rosenbaum, D., Rasmussen, S. & Kobilka, B. The Structure and Function of G-Protein Coupled Receptors. *Nature* 356–363 (2009). doi:10.1023/A:1015969301407
8. Erlandson, S. C., McMahon, C. & Kruse, A. C. Structural Basis for G Protein-Coupled Receptor Signaling. *Annual Review of Biophysics* **47**, 1–18 (2018).
9. Fredriksson, R., Lagerström, M. C., Lundin, L. G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
10. Venkatakrishnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
11. Unal, H. & Karnik, S. S. Domain coupling in GPCRs: The engine for induced conformational changes. *Trends Pharmacol. Sci.* **33**, 79–88 (2012).
12. Katritch, V., Cherezov, V. & Stevens, R. C. Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* **33**, 17–27 (2012).
13. Schiöth, H. B. & Lagerström, M. C. Structural diversity of g proteincoupled receptors and significance for drug discovery. *Nat. Rev. Drug Discov.* **7**, 339–357 (2008).
14. Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. in *Methods in Neurosciences* **25**, 366–428 (Academic Press, 1995).
15. Manglik, A. & Kruse, A. C. Structural Basis for G Protein-Coupled Receptor Activation. *Biochemistry* **56**, 5628–5634 (2017).
16. Preininger, A. M., Meiler, J. & Hamm, H. E. Conformational Flexibility and Structural Dynamics in

GPCR-Mediated G Protein Activation: A Perspective. *J. Mol. Biol.* **425**, 2288–2298 (2013).

17. Gurevich, V. V. & Gurevich, E. V. Molecular mechanisms of GPCR signaling: A structural perspective. *Int. J. Mol. Sci.* **18**, 1–17 (2017).
18. Syrovatkina, V., Alegre, K., Dey, R. & Huang, X.-Y. Regulation, Signaling and Physiological Functions of G-proteins. *J. Mol. Biol.* **428**, 3850–3868 (2016).
19. Latorraca, N. R., Venkatakrishnan, A. J. & Dror, R. O. GPCR dynamics: Structures in motion. *Chem. Rev.* **117**, 139–155 (2017).
20. Hilger, D., Masureel, M., Kobilka, B. K., Struct, N. & Biol, M. Structure and dynamics of GPCR signaling complexes. *Nat Struct Mol Biol* **25**, 4–12 (2018).
21. Chan, J. S. C., Chiu, T. T. & Wong, Y. H. Activation of Type II Adenylyl Cyclase by the Cloned μ-Opioid Receptor: Coupling to Multiple G Proteins. *J. Neurochem.* **65**, 2682–2689 (1995).
22. Suzuki, N., Hajicek, N. & Kozasa, T. Regulation and physiological functions of G12/13-mediated signaling pathways. *NeuroSignals* **17**, 55–70 (2009).
23. Rajagopal, S. & Shenoy, S. K. GPCR desensitization: Acute and prolonged phases. *Cellular Signalling* **41**, 9–16 (2018).
24. Ostermaier, M. K., Schertler, G. F. X. & Standfuss, J. Molecular mechanism of phosphorylation-dependent arrestin activation. *Curr. Opin. Struct. Biol.* **29**, 143–151 (2014).
25. Mayer, D. *et al.* Distinct G protein-coupled receptor phosphorylation motifs modulate arrestin affinity and activation and global conformation. *Nat. Commun.* **10**, 1–14 (2019).
26. Park, J. Y., Lee, S. Y., Kim, H. R., Seo, M. D. & Chung, K. Y. Structural mechanism of GPCR-arrestin interaction: Recent breakthroughs. *Arch. Pharm. Res.* **39**, 293–301 (2016).
27. Chaturvedi, M., Maharana, J. & Shukla, A. K. Terminating G-Protein Coupling: Structural Snapshots of GPCR-β-Arrestin Complexes. *Cell* **180**, 1041–1043 (2020).
28. Kim, K. & Chung, K. Y. Many faces of the GPCR-arrestin interaction. *Arch. Pharm. Res.* (2020). doi:10.1007/s12272-020-01263-w
29. Abdallah, K. & Gendron, L. The Delta Opioid Receptor in Pain Control. in *Delta Opioid Receptor Pharmacology and Therapeutic Applications* (ed. Jutkiewicz, E. M.) 147–177 (Springer International Publishing, 2017). doi:10.1007/164_2017_32
30. Smith, J. S. & Rajagopal, S. The β-Arrestins: Multifunctional regulators of G protein-coupled receptors. *J. Biol. Chem.* **291**, 8969–8977 (2016).
31. Wang, W., Qiao, Y. & Li, Z. New Insights into Modes of GPCR Activation. *Trends Pharmacol. Sci.* **39**, 367–386 (2018).
32. Wingler, L. M. & Lefkowitz, R. J. Conformational Basis of G Protein-Coupled Receptor Signaling Versatility. *Trends Cell Biol.* 1–12 (2020). doi:10.1016/j.tcb.2020.06.002
33. Shukla, A. K. Biasing GPCR signaling from inside. *Sci. Signal.* **7**, 2–4 (2014).
34. Pavlos, N. J. & Friedman, P. A. GPCR Signaling and Trafficking: The Long and Short of It. *Trends*

Endocrinol Metab **28**, 213–226 (2017).

35. Köse, M. GPCRs and EGFR – Cross-talk of membrane receptors in cancer. *Bioorganic Med. Chem. Lett.* **27**, 3611–3620 (2017).
36. Stein, C. Opioid Receptors. *Annu. Rev. Med.* **67**, 1–19 (2016).
37. Wang, S. Historical Review: Opiate Addiction and Opioid Receptors. *Cell Transplant.* **28**, 233–238 (2019).
38. Weiss, N. & Zamponi, G. W. Opioid Receptor Regulation of Neuronal Voltage-Gated Calcium Channels. *Cell. Mol. Neurobiol.* (2020). doi:10.1007/s10571-020-00894-3
39. Al-Hasani, Raem and Bruchas, M. R. Molecular mechanisms of opioid receptor-dependent signaling and behavior. *Anesthesiology* **115**, 1363–1381 (2011).
40. Eisenstein, T. K. The Role of Opioid Receptors in Immune System Function. *Front. Immunol.* **10**, 1–20 (2019).
41. Mullins, J. G. L. Structural modelling pipelines in next generation sequencing projects. in *Advances in Protein Chemistry and Structural Biology* **89**, 117–167 (Elsevier Inc., 2012).
42. Schmidt, T., Bergner, A. & Schwede, T. Modelling three-dimensional protein structures for applications in drug design. *Drug Discov. Today* **19**, 890–897 (2014).
43. Muhammed, M. T. & Aki-Yalcin, E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chem. Biol. Drug Des.* **93**, 12–20 (2019).
44. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
45. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, (2011).
46. Sali, A. & Bundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of molecular biology* **234**, 779–815 (1993).
47. Webb, B. & Sali, A. Comparative protein Structure Modeling using MODELLER. *Curr. Protoc. Bioinforma.* **54**, 5.6.1-5.6.37 (2016).
48. Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E. & Thornton, J. M. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **326**, 347–352 (1987).
49. Wiederstein, M. & Sippl, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, 407–410 (2007).
50. Sippl, M. J. Recognition of errors in three-dimensional structures of proteins. *Proteins Struct. Funct. Genet.* **17**, 355–362 (1993).
51. Frenkel, D. *Understanding Molecular Simulation: From algorithms to applications*. (Academic Press Inc., 1996).
52. Leach, A. R. *Molecular Modelling: Principles and applications*. (Pearson Education Limited, 2001).
53. Hospital, A., Goñi, J. R., Orozco, M. & Gelpí, J. L. Molecular dynamics simulations: Advances and

applications. *Adv. Appl. Bioinforma. Chem.* **8**, 37–47 (2015).

54. Baştuğ, T. & Kuyucak, S. Molecular dynamics simulations of membrane proteins. *Biophys. Rev.* **4**, 271–282 (2012).
55. Dauber-Osguthorpe, P. & Hagler, A. T. *Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there? Journal of Computer-Aided Molecular Design* **33**, (Springer International Publishing, 2019).
56. Goossens, K. & De Winter, H. Molecular Dynamics Simulations of Membrane Proteins: An Overview. *J. Chem. Inf. Model.* **58**, 2193–2202 (2018).
57. Bian, X., Li, Z., Deng, M. & Karniadakis, G. E. Fluctuating hydrodynamics in periodic domains and heterogeneous adjacent multidomains: Thermal equilibrium. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **92**, 1–20 (2015).
58. Ribeiro, J. M. L. & Filizola, M. Insights From Molecular Dynamics Simulations of a Number of G-Protein Coupled Receptor Targets for the Treatment of Pain and Opioid Use Disorders. *Front. Mol. Neurosci.* **12**, 1–13 (2019).
59. Guvench, O. & MacKerell, A. D. Comparison of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* **443**, 63–88 (2008).
60. Allen, M. P. Introduction to Molecular Dynamic Simulations. in *Computational Soft Matter: From Synthetic Polymers to Proteins, Lecture Notes* (eds. Attig, N., Binder, K., Grubmuller, H. & Kremer, K.) 1–28 (2004).
61. Tuckerman, M. E. & Martyna, G. J. Understanding Modern Molecular Dynamics: Techniques and Applications Mark. *J. Phys. Chem. B* **104**, 159–178 (2000).
62. Weng, J. & Wang, W. Molecular Dynamics Simulation of Membrane Proteins. in *Protein Conformational Dynamics, Advances in Experimental Medicine and Biology* **805**, 305–329 (2014).
63. Vanommeslaeghe, K. *et al.* CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM ALL-Atom Additive Biological Force Fields. *J. Comput. Chem.* **31**, 671–690 (2009).
64. Vanommeslaeghe, K. & MacKerell, A. D. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* **52**, 3144–3154 (2012).
65. Case, D. A. *Amber 18. Amber 2018 Reference Manual* (2018).
66. Cieplak, P., Kollman, P. A., Wang, J., Cieplak, P. & Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? - Wang - 2000 - Journal of Computational Chemistry - Wiley Online Library. *J. Comput. …* **21**, 1049–1074 (2000).
67. Hornak, V. *et al.* Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Genet.* **65**, 712–725 (2006).
68. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A well-behaved electrostatic potential

based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).

69. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **105**, 6474–6487 (2001).
70. Schuler, L. D., Daura, X. & Van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **22**, 1205–1218 (2001).
71. Dickson, C. J. *et al.* Lipid14: The amber lipid force field. *J. Chem. Theory Comput.* **10**, 865–879 (2014).
72. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 Dihedral Angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
73. Lexa, K. W. & Carlson, H. A. Protein flexibility in docking and surface mapping. *Q. Rev. Biophys.* **45**, 301–343 (2012).
74. Salmaso, V. & Moro, S. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Front. Pharmacol.* **9**, 1–16 (2018).
75. de Ruyck, J., Brysbaert, G., Blossey, R. & Lensink, M. F. Molecular docking as a popular tool in drug design, an in silico travel. *Adv. Appl. Bioinforma. Chem.* **9**, 1–11 (2016).
76. Guedes, I. A., de Magalhães, C. S. & Dardenne, L. E. Receptor-ligand molecular docking. *Biophys. Rev.* **6**, 75–87 (2014).
77. Antunes, D. A., Devaurs, D. & Kavraki, L. E. Understanding the challenges of protein flexibility in drug design. *Expert Opin. Drug Discov.* **10**, 1301–1313 (2015).
78. Buonfiglio, R., Recanatini, M. & Masetti, M. Protein Flexibility in Drug Discovery: From Theory to Computation. *ChemMedChem* **10**, 1141–1148 (2015).
79. Gromiha, M. M., Yugandhar, K. & Jemimah, S. Protein–protein interactions: scoring schemes and binding affinity. *Current Opinion in Structural Biology* (2017). doi:10.1016/j.sbi.2016.10.016
80. Jain, A. N. & Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput. Aided. Mol. Des.* **22**, 133–139 (2008).
81. Leelananda, S. P. & Lindert, S. Computational methods in drug discovery. *Beilstein J. Org. Chem.* **12**, 2694–2718 (2016).
82. Goodsell, D. S. & Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins Struct. Funct. Bioinforma.* **8**, 195–202 (1990).
83. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).
84. Allen, W. J. *et al.* DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* **36**, 1132–1156 (2015).

85. Gioia, D., Bertazzo, M., Recanatini, M., Masetti, M. & Cavalli, A. Dynamic docking: A paradigm shift in computational drug discovery. *Molecules* **22**, 1–21 (2017).
86. Bauer, J. A., Pavlovíc, J. & Bauerová-Hlinková, V. Normal mode analysis as a routine part of a structural investigation. *Molecules* **24**, (2019).
87. Wako, H. & Endo, S. Normal mode analysis as a method to derive protein dynamics information from the Protein Data Bank. *Biophys. Rev.* **9**, 877–893 (2017).
88. *Normal Mode Analysis Theory and Applications to Biological and Chemical Systems*. *Mathematical and Computational Biology Series* (Chapman & Hall/CRC, 2006).
89. Skjaerven, L., Hollup, S. M. & Reuter, N. Normal mode analysis for proteins. *J. Mol. Struct. THEOCHEM* **898**, 42–48 (2009).
90. Miyashita, O. & Tama, F. Hybrid methods for macromolecular modeling by molecular mechanics simulations with experimental data. in *Integrative Structural Biology with Hybrid Methods* (eds. Nakamura, H. & et al.) **1105**, 199–217 (Springer International Publishing, 2018).
91. Togashi, Y. & Flechsig, H. Coarse-grained protein dynamics studies using elastic network models. *Int. J. Mol. Sci.* **19**, (2018).
92. Inoue, A. *et al.* Illuminating G-Protein-Coupling Selectivity of GPCRs. *Cell* **177**, 1933–1947 (2019).
93. Zhu, T., Fang, L. Y. & Xie, X. Development of a universal high-throughput calcium assay for G-protein-coupled receptors with promiscuous G-protein Gα15/16. *Acta Pharmacol. Sin.* **29**, 507–516 (2008).
94. Vicente-Sanchez, A. *et al.* Tolerance to high-internalizing δ opioid receptor agonist is critically mediated by arrestin 2. *Br. J. Pharmacol.* **175**, 3050–3059 (2018).
95. Cheng, Z. J., Yu, Q. M., Wu, Y. L., Ma, L. & Pei, G. Selective interference of β-arrestin 1 with κ and δ but not μ opioid receptor/G protein coupling. *J. Biol. Chem.* **273**, 24328–24333 (1998).
96. Molinari, P. *et al.* Morphine-like opiates selectively antagonize receptor-arrestin interactions. *J. Biol. Chem.* **285**, 12522–12535 (2010).
97. Hampson, R. E., Mu, J. & Deadwyler, S. A. Cannabinoid and kappa opioid receptors reduce potassium K current via activation of G(s) proteins in cultured hippocampal neurons. *J. Neurophysiol.* **84**, 2356–2364 (2000).
98. Morgenweck, J., Frankowski, K. J., Prisinzano, T. E., Aubé, J. & Bohn, L. M. Investigation of the role of βarrestin2 in kappa opioid receptor modulation in a mouse model of pruritus. *Neuropharmacology* 600–609 (2015). doi:10.1016/j.neuropharm.2015.08.027.Investigation
99. Burford, N. T., Tolbert, L. M. & Sadee, W. Specific G protein activation and μ-opioid receptor internalization caused by morphine, DAMGO and endomorphin I. *Eur. J. Pharmacol.* **342**, 123–126 (1998).
100. Sánchez-Blázquez, P., Gómez-Serranillos, P. & Garzón, J. Agonists determine the pattern of G-protein activation in μ-opioid receptor-mediated supraspinal analgesia. *Brain Res. Bull.* **54**, 229–235

(2001).

101. Wang, D. *et al.* Contribution of adrenomedullin to the switch of G protein-coupled μ-opioid receptors from Gi to Gs in the spinal dorsal horn following chronic morphine exposure in rats. *Br. J. Pharmacol.* **173**, 1196–1207 (2016).
102. Effects, D., Boda, K. & Gintzler, A. R. Dual Effects of DAMGO [D-Ala2,N-Me-Phe4,Gly5-ol]-Enkephalin and CTAP (D-Phe-Cys-Tyr-D-Trp-Arg-Thr-Pen-Thr-NH2)on Adenylyl Cyclase Activity: Implications for μ-Opioid Receptor Gs Coupling. *J. Pharmacol. Exp. Ther.* **310**, 256–262 (2004).
103. Offermanns, S. & Simon, M. I. Gα15 and gα16 couple a wide variety of receptors to phospholipase C. *Journal of Biological Chemistry* **270**, 15175–15180 (1995).
104. Manabe, S. *et al.* Possible biased analgesic of hydromorphone through the G protein-over β-arrestin-mediated pathway: cAMP, CellKey™, and receptor internalization analyses. *J. Pharmacol. Sci.* **140**, 171–177 (2019).
105. Mori, T. *et al.* Usefulness for the combination of G protein- and β-arrestin-biased ligands of μ-opioid receptors: Prevention of antinociceptive tolerance. *Mol. Pain* **13**, 1–9 (2017).
106. Yung, L. Y. *et al.* Gα(L1) (Gα14) couples the opioid receptor-like1 receptor to stimulation of phospholipase C. *J. Pharmacol. Exp. Ther.* **288**, 232–238 (1999).
107. Mittal, N. *et al.* Select G-protein coupled receptors modulate agonist-inducedsignaling via a ROCK, LIMK and β-arrestin 1 pathway. *Cell* **23**, 1–7 (2013).
108. Malfacini, D. *et al.* Pharmacological profile of nociceptin/orphanin FQ receptors interacting with G-proteins and β-arrestins 2. *PLoS One* **10**, 1–22 (2015).
109. Che, T. *et al.* Structure of the Nanobody-Stabilized Active State of the Kappa Opioid Receptor. *Cell* **172**, 55-67.e15 (2018).
110. Staus, D. P. *et al.* Structure of the M2 muscarinic receptor–β-arrestin complex in a lipid nanodisc. *Nature* **579**, 297–302 (2020).
111. Yin, W. *et al.* A complex structure of arrestin-2 bound to a G protein-coupled receptor. *Cell Res.* **29**, 971–983 (2019).
112. Bateman, A. et al. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
113. Shen, M. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 (2006).
114. Cristobal, S., Zemla, A., Fischer, D., Rychlewski, L. & Elofsson, A. A study of quality measures for protein threading models. *BMC Bioinformatics* **2**, (2001).
115. Siew, N., Elofsson, A., Rychlewski, L. & Fischer, D. MaxSub: An automated measure for the assessment of protein structure prediction quality. *Bioinformatics* **16**, 776–785 (2000).
116. Wallner, B. & Elofsson, A. Identification of correct regions in protein models using structural, alignment, and consensus information. *Protein Sci.* **15**, 900–913 (2006).

117. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
118. Kang, Y. *et al.* Cryo-EM structure of human rhodopsin bound to an inhibitory G protein. *Nature* **558**, 553–558 (2018).
119. Rasmussen, S. G. F. *et al.* Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555 (2011).
120. Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
121. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, 370–376 (2012).
122. Sunhwan, J., Taehoon, K., G. Iyer, V. & Wonpil, I. CHARMM-GUI: A Web Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
123. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).
124. Maeda, S., Qu, Q., Robertson, M. J., Skiniotis, G. & Kobilka, B. K. Structures of the M1 and M2 muscarinic acetylcholine receptor/G-protein complexes. *Science (80-.).* **364**, 552–557 (2019).
125. The PyMOL Molecular Graphics System; Schrödinger, LLC.
126. Van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
127. Sandhu, M. *et al.* Conformational plasticity of the intracellular cavity of GPCR−G-protein complexes leads to G-protein promiscuity and selectivity. *Proc. Natl. Acad. Sci.* **116**, 11956–11965 (2019).
128. Richter, S., Wenzel, A., Stein, M., Gabdoulline, R. R. & Wade, R. C. webPIPSA: a web server for the comparison of protein interaction properties. *Nucleic Acids Res.* **36**, 276–280 (2008).
129. Davis, M. E., Madura, J. D., Luty, B. A. & McCammon, J. A. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Comput. Phys. Commun.* **62**, 187–197 (1991).
130. Vangone, A., Spinelli, R., Scarano, V., Cavallo, L. & Oliva, R. COCOMAPS: A web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics* **27**, 2915–2916 (2011).
131. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *Journal of Molecular Biology* **238**, 777–793 (1994).
132. Hubbard, S. J. & Thornton, J. M. 'NACCESS', computer program. (1993).
133. Lee, B. & Richards, F. M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **55**, (1971).
134. Negi, S. S., Schein, C. H., Oezguen, N., Power, T. D. & Braun, W. InterProSurf: A web server for

predicting interacting sites on protein surfaces. *Bioinformatics* **23**, 3397–3399 (2007).

135. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
136. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
137. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer International Publishing, 2016). doi:10.1007/978-3-319-24277-4
138. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 (2006).
139. Fuglebakk, E., Echave, J. & Reuter, N. Measuring and comparing structural fluctuation patterns in large protein datasets. *Bioinformatics* **28**, 2431–2440 (2012).
140. Gower, J. C. Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. *Biometrika* **53**, 325 (1966).
141. Carpenter, B. & Tate, C. G. Active state structures of G protein-coupled receptors highlight the similarities and differences in the G protein and arrestin coupling interfaces. *Current Opinion in Structural Biology* **45**, 124–132 (2017).
142. Kang, Y. *et al.* Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**, 561–567 (2015).
143. Scheerer, P. *et al.* Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502 (2008).
144. García-Nafría, J. & Tate, C. G. Cryo-EM structures of GPCRs coupled to Gs, Gi and Go. *Mol. Cell. Endocrinol.* **488**, 1–13 (2019).
145. Mafi, A., Kim, S. K. & Goddard, W. A. The atomistic level structure for the activated human κ-opioid receptor bound to the full Gi protein and the MP1104 agonist. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 5836–5843 (2020).
146. Preto, A. J. *et al.* Understanding the binding specificity of G-protein coupled receptors towards G-proteins and Arrestins: application to the dopamine receptor family. *J. Chem. Inf. Model.* **60**, 3969–3984 (2020).
147. Mafi, A., Kim, S. K. & Goddard, W. A. Mechanism of β-arrestin recruitment by the μ-opioid G protein-coupled receptor. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 16346–16355 (2020).
148. Shukla, A. K. *et al.* Structure of active β-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. *Nature* **497**, 137–141 (2013).
149. Szczepek, M. *et al.* Crystal structure of a common GPCR-binding interface for G protein and arrestin. *Nat. Commun.* **5**, 1–8 (2014).
150. Zhou, X. E. *et al.* Identification of Phosphorylation Codes for Arrestin Recruitment by G Protein-Coupled Receptors. *Cell* **170**, 457-469.e13 (2017).
151. Kim, K. *et al.* Structure of a Hallucinogen-Activated Gq-Coupled 5-HT2A Serotonin Receptor. *Cell*

182, 1574-1588.e19 (2020).

152. Carpenter, B., Nehmé, R., Warne, T., Leslie, A. G. W. & Tate, C. G. Structure of the adenosine A2A receptor bound to an engineered G protein. *Nature* **536**, 104–107 (2016).
153. Flock, T. *et al.* Selectivity determinants of GPCR-G-protein binding. *Nature* **545**, 317–322 (2017).
154. Kang, Y. *et al.* A structural snapshot of the rhodopsin-arrestin complex. *FEBS J.* **283**, 816–821 (2016).
155. Yin, J. *et al.* Structure of a D2 dopamine receptor–G-protein complex in a lipid membrane. *Nature* **584**, 125–129 (2020).
156. Liu, K. *et al.* Structural basis of CXC chemokine receptor 2 activation and signalling. *Nature* **585**, 135–140 (2020).
157. Hua, T. *et al.* Activation and Signaling Mechanism Revealed by Cannabinoid Receptor-Gi Complex Structures. *Cell* **180**, 655-665.e18 (2020).
158. Wasilko, D. J. *et al.* Structural basis for chemokine receptor CCR6 activation by the endogenous protein ligand CCL20. *Nat. Commun.* **11**, 1–9 (2020).
159. García-Nafría, J., Nehmé, R., Edwards, P. C. & Tate, C. G. Cryo-EM structure of the serotonin 5-HT1B receptor coupled to heterotrimeric Go. *Nature* **558**, 620–623 (2018).
160. Kato, H. E. *et al.* Conformational transitions of a neurotensin receptor 1–Gi1 complex. *Nature* **572**, 80–85 (2019).
161. Ma, X. *et al.* Analysis of β2AR-Gs and β2AR-Gi complex formation by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 23096–23105 (2020).
162. Liu, X. *et al.* Structural Insights into the Process of GPCR-G Protein Complex Formation. *Cell* **177**, 1243-1251.e12 (2019).
163. Dong, M. *et al.* Structure and dynamics of the active Gs-coupled human secretin receptor. *Nat. Commun.* **11**, 1–17 (2020).
164. Yang, F. *et al.* Structural basis of GPBAR activation and bile acid recognition. *Nature* **587**, 499–504 (2020).
165. Lee, Y. *et al.* Molecular basis of β-arrestin coupling to formoterol-bound β1-adrenoceptor. *Nature* **583**, 862–866 (2020).

7. Annexes

Annex 1: The 20 best models for each OR modelled with active KOR stabilized by a nanobody. The chosen model is marked in blue

OR models	Z-score	molpdf	DOPE score	LGscore	MaxSub
DOR.B99990078	-3.72	4759.37598	-40153.93750	2.81	0.24
DOR.B99990032	-3.71	4966.41260	-40133.26562	2.83	0.25
DOR.B99990036	-3.74	4809.60107	-40064.00391	2.90	0.24
DOR.B99990009	-3.81	4697.17090	-40044.42188	2.91	0.24
DOR.B99990095	-3.79	4811.34912	-40037.90625	2.78	0.24
DOR.B99990064	-3.73	4765.12402	-40021.49219	2.84	0.23
DOR.B99990035	-3.58	5072.75879	-40010.10938	03.02	0.23
DOR.B99990074	-3.75	4850.88721	-39996.56250	3.23	0.29
DOR.B99990096	-3.8	5086.31445	-39978.92969	2.83	0.23
DOR.B99990056	-3.92	4845.61475	-39977.05078	2.63	0.23
DOR.B99990050	-3.57	4777.05762	-39974.23828	2.45	0.21
DOR.B99990006	-3.83	4847.67529	-39953.78125	2.74	0.23
DOR.B99990016	-3.82	4824.20117	-39945.96484	2.65	0.22
DOR.B99990022	-3.76	4923.03223	-39939.16016	2.60	0.23
DOR.B99990040	-3.67	4791.07471	-39912.83203	2.93	0.26
DOR.B99990030	-3.76	4759.77979	-39906.18750	2.99	0.26
DOR.B99990077	-3.69	4698.48877	-39901.22266	2.58	0.23
DOR.B99990019	-3.74	4801.07568	-39893.04297	2.77	0.23
DOR.B99990087	-3.72	4885.90527	-39891.58984	2.74	0.21
DOR.B99990076	-3.74	4832.16699	-39889.39062	2.88	0.23

OR models	Z-score	molpdf	DOPE score	LGscore	MaxSub
KOR.B99990066	-1.95	4282.78467	-40779.39062	4.13	0.42
KOR.B99990036	-1.84	4127.31104	-40609.62500	4.13	0.41
KOR.B99990091	-1.77	4212.48779	-40603.80859	4.17	0.44
KOR.B99990037	-1.84	4171.57764	-40551.39062	4.12	0.42
KOR.B99990018	-1.87	4288.68311	-40548.70312	4.16	0.43
KOR.B99990032	-1.89	4101.14111	-40506.89062	4.13	0.43
KOR.B99990090	-1.86	4252.47656	-40506.00000	4.14	0.41
KOR.B99990093	-1.91	4135.55127	-40505.12891	4.20	0.46
KOR.B99990087	-1.96	4369.98145	-40502.24609	4.14	0.40
KOR.B99990071	-1.9	4234.75635	-40500.41406	4.14	0.40
KOR.B99990038	-1.84	4143.48193	-40495.55859	4.13	0.43
KOR.B99990002	-1.81	4281.85742	-40487.48047	4.20	0.44
KOR.B99990099	-1.73	4158.19971	-40473.94531	4.19	0.42
KOR.B99990027	-1.86	4163.50244	-40462.75781	4.17	0.42
KOR.B99990020	-1.9	4129.50293	-40462.71484	4.13	0.42
KOR.B99990086	-1.84	4207.87598	-40457.63281	4.14	0.40
KOR.B99990056	-1.91	4295.44141	-40455.93750	4.11	0.42

KOR.B99990062	-1.89	4172.39551	-40443.96875	4.14	0.41
KOR.B99990097	-1.82	4116.88037	-40442.83203	4.13	0.41
KOR.B99990054	-1.95	4436.91504	-40442.21094	4.13	0.41

OR models	Z-score	molpdf	DOPE score	LGscore	Maxsub
MOR.B99990039	-1.97	4709.92627	-40873.00781	4.14	0.36
MOR.B99990015	-2.14	4781.11963	-40872.26562	4.14	0.39
MOR.B99990083	-1.89	4536.46387	-40862.47656	4.14	0.38
MOR.B99990084	-1.92	4578.17969	-40842.67188	4.10	0.38
MOR.B99990081	-2.07	4579.89941	-40813.49609	4.09	0.38
MOR.B99990088	-1.92	4659.84131	-40809.19141	4.07	0.38
MOR.B99990090	-1.9	4580.39014	-40727.02344	4.10	0.39
MOR.B99990064	-1.92	4815.56592	-40703.24219	4.19	0.36
MOR.B99990003	-1.95	4554.38232	-40696.78516	4.16	0.39
MOR.B99990074	-1.86	4551.60010	-40693.82031	4.13	0.40
MOR.B99990035	-1.94	4588.84082	-40692.22656	4.13	0.41
MOR.B99990040	-1.91	4606.87695	-40690.74219	4.12	0.41
MOR.B99990082	-1.93	4645.30566	-40669.99219	4.13	0.38
MOR.B99990048	-1.9	4653.97314	-40654.95312	4.06	0.36
MOR.B99990053	-1.92	4540.21289	-40653.20703	4.17	0.38
MOR.B99990005	-2.05	4645.69824	-40641.12109	4.15	0.37
MOR.B99990095	-1.94	4674.05469	-40640.91406	4.15	0.39
MOR.B99990055	-1.98	4720.19971	-40638.52734	4.08	0.36
MOR.B99990036	-2.01	4707.69775	-40636.18750	4.14	0.42
MOR.B99990049	-1.96	4598.06885	-40618.63672	4.06	0.39

OR models	Z-score	molpdf	DOPE score	LGscore	Maxsub
NOP.B99990051	-2.41	4466.97949	-38309.62109	4.18	0.39
NOP.B99990056	-2.44	4411.88232	-38196.69922	4.21	0.36
NOP.B99990001	-2.49	4571.15869	-38169.99609	4.23	0.37
NOP.B99990028	-2.35	4374.12402	-38150.09766	4.24	0.36
NOP.B99990045	-2.51	4510.98486	-38052.59766	4.16	0.35
NOP.B99990032	-2.37	4627.27344	-38021.66797	4.23	0.36
NOP.B99990062	-2.45	4467.39355	-38014.03125	4.23	0.36
NOP.B99990075	-2.52	4499.74854	-38012.01562	4.22	0.38
NOP.B99990050	-2.42	4473.80908	-37995.82031	4.18	0.36
NOP.B99990014	-2.31	4464.51123	-37990.82422	4.26	0.34
NOP.B99990054	-2.48	4488.76807	-37985.37500	4.11	0.35
NOP.B99990063	-2.32	4600.10693	-37984.01562	4.31	0.35
NOP.B99990097	-2.47	5888.48682	-37983.61328	4.16	0.34
NOP.B99990066	-2.5	4637.95850	-37981.10938	4.18	0.34
NOP.B99990083	-2.37	4661.06348	-37976.60547	4.25	0.38
NOP.B99990067	-2.51	4485.09863	-37966.56250	4.24	0.37
NOP.B99990030	-2.4	4484.11475	-37963.80469	4.21	0.37
NOP.B99990018	-2.36	4484.43311	-37957.25391	4.24	0.36
NOP.B99990082	-2.46	4528.18555	-37954.80859	4.29	0.35
NOP.B99990009	-2.38	4553.88916	-37947.64844	4.23	0.36

Annex 2: The 20 best models for each OR modelled with NTSR1-Arrestin 2 template. The chosen model is marked in blue

OR models	Z-score	molpdf	DOPE score	LGscore	MaxSub
DOR.B99990056	-2.73	6387.93506	-39229.99609	2.44	0.18
DOR.B99990036	-2.58	6269.90381	-39077.57422	2.52	0.18
DOR.B99990097	-2.78	6222.17139	-38984.12109	2.03	0.12
DOR.B99990086	-2.71	6219.08105	-38982.98828	2.49	0.18
DOR.B99990075	-2.69	6407.69775	-38937.17969	2.37	0.18
DOR.B99990017	-2.74	6240.79590	-38928.13672	2.23	0.17
DOR.B99990059	-2.72	6209.81299	-38904.03125	2.45	0.18
DOR.B99990088	-2.79	6207.11426	-38804.67188	2.56	0.19
DOR.B99990016	-2.55	6323.07471	-38790.70703	2.67	0.20
DOR.B99990005	-2.61	6195.56738	-38756.12500	2.49	0.18
DOR.B99990003	-2.66	6280.92334	-38748.38281	2.16	0.16
DOR.B99990044	-2.43	6123.68701	-38740.56641	2.66	0.20
DOR.B99990064	-2.7	6216.49121	-38710.88281	2.25	0.17
DOR.B99990079	-2.57	6267.82471	-38689.57422	2.48	0.17
DOR.B99990089	-2.58	6312.15381	-38684.11328	2.42	0.18
DOR.B99990090	-2.68	6268.11768	-38659.31641	2.39	0.17
DOR.B99990031	-2.59	6405.68896	-38652.93359	2.15	0.15
DOR.B99990025	-2.81	6194.99072	-38636.74219	2.38	0.18
DOR.B99990030	-2.72	6276.31934	-38635.16016	2.57	0.20
DOR.B99990060	-2.64	6286.86279	-38628.41406	2.05	0.15

OR models	Z-score	molpdf	DOPE score	LGscore	MaxSub
KOR.B99990027	-0.73	6068.19043	-38511.57031	4.18	0.47
KOR.B99990015	-0.76	5981.23096	-38348.89453	4.17	0.47
KOR.B99990046	-0.78	6063.07031	-38236.01953	4.11	0.44
KOR.B99990081	-0.79	6008.88525	-38228.55469	4.14	0.49
KOR.B99990006	-0.71	6039.65332	-38212.75781	4.23	0.45
KOR.B99990052	-0.69	6053.05713	-38192.12891	4.23	0.47
KOR.B99990014	-0.79	6040.97266	-38179.07812	4.19	0.46
KOR.B99990056	-0.59	5988.99658	-38160.16797	4.15	0.44
KOR.B99990043	-0.86	6105.70215	-38126.32812	4.09	0.47
KOR.B99990083	-0.76	5971.74707	-38117.77344	4.18	0.45
KOR.B99990066	-0.8	5948.00684	-38111.19531	4.15	0.46
KOR.B99990076	-0.64	6118.03320	-38110.24219	4.13	0.47
KOR.B99990090	-0.76	5990.70020	-38082.19531	4.12	0.48
KOR.B99990018	-0.66	5949.58398	-38070.16016	4.21	0.49
KOR.B99990072	-0.75	5961.23633	-38049.04688	4.13	0.49
KOR.B99990084	-0.77	6069.76709	-38033.77344	4.20	0.47
KOR.B99990075	-0.76	6114.23682	-38006.74219	4.18	0.48
KOR.B99990094	-0.69	6037.58447	-37983.65234	4.24	0.47
KOR.B99990059	-0.75	6128.64941	-37979.60938	4.27	0.48
KOR.B99990079	-0.8	6052.94580	-37960.63672	4.16	0.49

OR models	Z-score	molpdf	DOPE score	LGscore	MaxSub
MOR.B99990029	-1.19	5044.35010	-39446.74609	4.10	0.42
MOR.B99990061	-1.15	5043.31787	-39225.51172	4.09	0.42
MOR.B99990077	-1.2	4983.03516	-39217.70312	4.15	0.42
MOR.B99990038	-1.03	4983.20117	-39151.26172	4.12	0.42
MOR.B99990028	-1.24	5125.59229	-39148.50781	4.12	0.41
MOR.B99990060	-1.26	5046.49463	-39146.52734	4.07	0.34
MOR.B99990001	-1.15	5166.90918	-39140.41797	4.10	0.38
MOR.B99990098	-1.27	5115.51074	-39135.70312	4.02	0.43
MOR.B99990032	-1.16	4944.79492	-39097.55859	4.08	0.41
MOR.B99990083	-1.12	5143.34570	-39094.71875	4.10	0.42
MOR.B99990097	-1.21	5183.96680	-39057.70703	4.15	0.41
MOR.B99990053	-1.15	5031.20654	-39054.67969	4.13	0.43
MOR.B99990011	-1.06	5018.08496	-39043.14844	4.10	0.43
MOR.B99990069	-1.17	5029.85156	-39039.55078	4.09	0.43
MOR.B99990066	-1.14	4981.61719	-39035.26172	4.07	0.42
MOR.B99990072	-1.06	4969.22168	-39033.48828	4.07	0.44
MOR.B99990057	-1.13	5052.41211	-39024.19922	4.03	0.39
MOR.B99990089	-1.21	4945.86230	-38993.55078	4.11	0.39
MOR.B99990012	-1.22	5240.97705	-38987.31250	4.02	0.40
MOR.B99990023	-1.38	5146.90820	-38987.06250	4.10	0.37

OR models	Z-score	molpdf	DOPE score	LGscore	MaxSub
NOP.B99990048	-1.43	4755.51465	-37441.76562	4.11	0.40
NOP.B99990097	-1.49	4730.17139	-37165.12500	4.13	0.35
NOP.B99990077	-1.56	4719.51611	-37163.64844	4.18	0.37
NOP.B99990099	-1.53	4832.85254	-37120.19141	4.23	0.37
NOP.B99990011	-1.3	4767.94824	-37117.39062	3.96	0.34
NOP.B99990053	-1.52	5032.90771	-37104.89062	4.12	0.37
NOP.B99990069	-1.55	4910.97705	-37056.36719	4.15	0.38
NOP.B99990078	-1.56	4791.92383	-37055.25391	4.08	0.37
NOP.B99990046	-1.36	4742.65576	-37055.23828	4.02	0.36
NOP.B99990040	-1.37	4774.04736	-37042.63281	4.14	0.38
NOP.B99990014	-1.68	4815.85107	-37039.37500	4.15	0.38
NOP.B99990007	-1.61	4725.86865	-37034.62500	3.99	0.37
NOP.B99990050	-1.46	4949.38574	-37027.99609	3.95	0.37
NOP.B99990087	-1.43	4896.09863	-37027.07422	4.05	0.37
NOP.B99990054	-1.39	4812.35986	-37022.21484	4.18	0.36
NOP.B99990065	-1.4	4991.42090	-36986.55078	4.09	0.36
NOP.B99990084	-1.59	5052.56348	-36981.19922	4.08	0.37
NOP.B99990073	-1.33	4778.52100	-36975.84766	4.08	0.39
NOP.B99990018	-1.35	5148.37256	-36972.77734	4.05	0.39
NOP.B99990066	-1.29	4706.35840	-36967.74609	4.05	0.38

Annex 3: The 20 best models for each OR using the M2R-Arrestin 2 template. The chosen model is marked in blue.

OR models	Z-score	molpdf	DOPE score	LGscore	MaxSub
DOR.B99990052	-1.73	5286.25977	-38736.09766	2.32	0.16
DOR.B99990002	-2.09	4040.58325	-38582.31250	2.29	0.17
DOR.B99990085	-2.18	3849.07812	-38573.01953	02.06	0.13
DOR.B99990001	-2.04	3889.15015	-38559.46094	1.93	0.12
DOR.B99990070	-2	4006.39038	-38526.32812	02.02	0.13
DOR.B99990035	-2.03	4029.42651	-38463.20312	2.18	0.14
DOR.B99990013	-2.11	3944.44312	-38440.08984	2.38	0.16
DOR.B99990003	-1.87	4128.66309	-38432.64062	2.22	0.15
DOR.B99990093	-2.15	4010.66333	-38425.58984	2.29	0.17
DOR.B99990088	-1.94	3892.07983	-38405.92188	2.16	0.14
DOR.B99990074	-2.21	3955.65625	-38401.83594	02.03	0.14
DOR.B99990079	-2.09	3966.24463	-38392.57031	1.95	0.14
DOR.B99990049	-2.11	3885.04688	-38390.53125	2.44	0.17
DOR.B99990061	-2.15	3884.28198	-38389.26953	1.93	0.12
DOR.B99990040	-2.12	4190.22363	-38363.76562	1.95	0.12
DOR.B99990081	-2.14	3922.83276	-38361.00391	2.24	0.16
DOR.B99990096	-2.05	3980.39624	-38320.27734	2.53	0.18
DOR.B99990065	-2.15	3853.93872	-38296.14062	2.21	0.15
DOR.B99990015	-2.18	3937.67090	-38295.30469	02.01	0.14
DOR.B99990021	-2.25	4005.06982	-38292.53906	2.29	0.15

OR models	Z-score	molpdf	DOPE score	LGscore	MaxSub
KOR.B99990003	-0.28	4073.19971	-39046.81641	4.18	0.38
KOR.B99990031	-0.25	4017.88550	-38977.55859	4.11	0.36
KOR.B99990098	-0.44	4092.28345	-38950.15234	04.09	0.37
KOR.B99990027	-0.36	4219.50342	-38829.35938	4.23	0.35
KOR.B99990030	-0.3	4000.33423	-38822.37500	4.16	0.39
KOR.B99990051	-0.19	3898.49463	-38816.96875	04.08	0.40
KOR.B99990020	-0.25	3937.27441	-38814.42188	4.13	0.38
KOR.B99990002	-0.26	3943.93628	-38812.31250	4.12	0.39
KOR.B99990086	-0.3	3969.81885	-38798.88672	04.05	0.33
KOR.B99990071	-0.24	3960.58301	-38731.92578	4.11	0.37
KOR.B99990091	-0.18	4123.97852	-38708.09375	4.11	0.34
KOR.B99990018	-0.38	4063.82520	-38698.63672	4.11	0.37
KOR.B99990004	0.09	4078.30566	-38689.75781	4.10	0.40
KOR.B99990044	-0.11	3940.87402	-38681.18750	4.18	0.38
KOR.B99990034	-0.28	3981.98584	-38679.42578	4.12	0.36

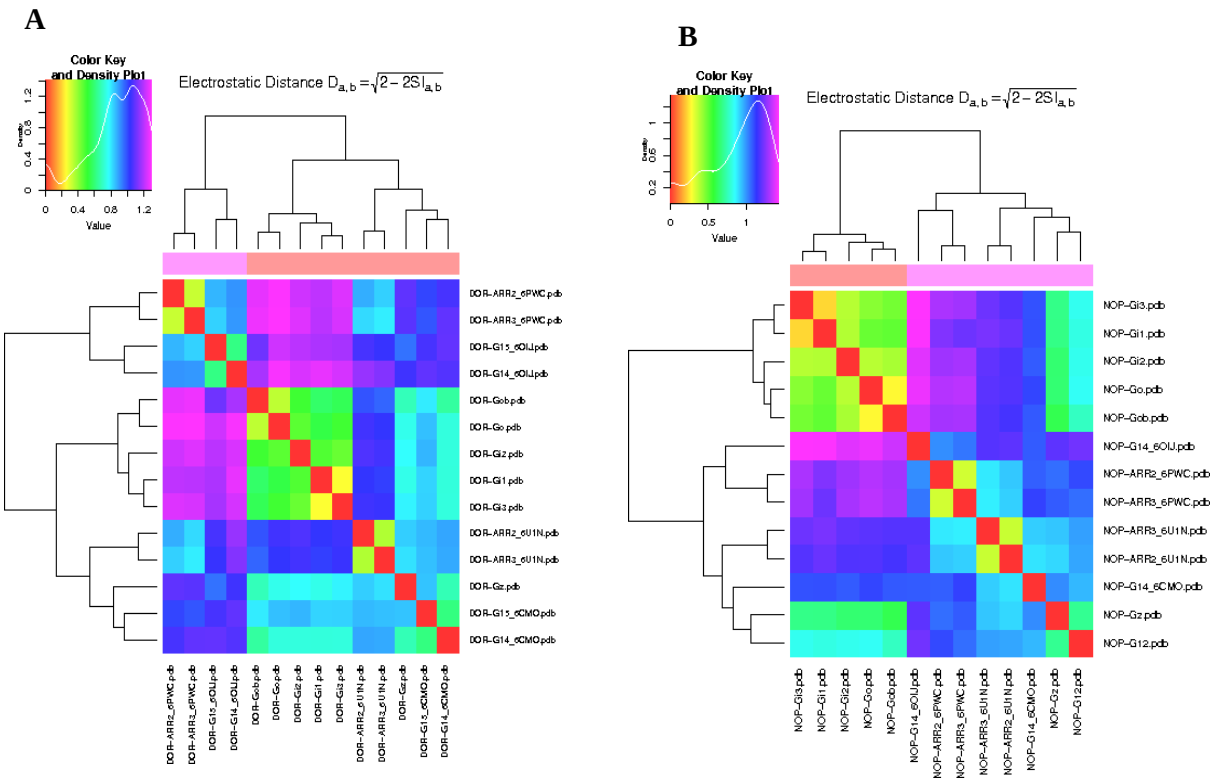
KOR.B99990025	-0.16	4145.17188	-38675.87109	04.04	0.33
KOR.B99990012	0.24	3964.44946	-38661.64844	4.13	0.37
KOR.B99990087	-0.32	4205.16260	-38655.33203	4.19	0.31
KOR.B99990089	-0.23	3917.84741	-38642.95703	04.06	0.38
KOR.B99990006	-0.24	4008.35864	-38594.31250	4.11	0.33

OR models	Z-score	molpdf	DOPE score	LGscore	MaxSub
MOR.B99990024	-0.57	4020.74609	-39299.64453	04.05	0.29
MOR.B99990012	-0.33	3930.86865	-39198.99609	04.08	0.33
MOR.B99990097	-0.38	4103.24561	-39162.07422	04.03	0.32
MOR.B99990001	-0.56	3936.13477	-39134.81250	4.14	0.35
MOR.B99990078	-0.43	4074.13281	-39071.50391	4.11	0.33
MOR.B99990025	-0.47	4022.64575	-39069.47656	3.98	0.31
MOR.B99990086	-0.7	3983.46045	-39028.63281	04.06	0.32
MOR.B99990016	-0.44	4065.59277	-39028.29297	04.04	0.31
MOR.B99990008	-0.4	3972.73804	-39025.39844	04.01	0.34
MOR.B99990083	-0.63	4189.60303	-39017.34375	04.03	0.33
MOR.B99990067	-0.61	4112.97559	-38994.41406	4.10	0.29
MOR.B99990085	-0.43	3952.07080	-38990.76172	04.03	0.33
MOR.B99990019	-0.46	3976.41821	-38980.45312	3.96	0.31
MOR.B99990006	-0.38	4162.34619	-38976.79297	04.03	0.35
MOR.B99990039	-0.42	4059.87158	-38973.42188	3.97	0.31
MOR.B99990089	-0.38	4106.34521	-38914.58203	04.07	0.33
MOR.B99990082	-0.36	4058.36060	-38901.15625	04.07	0.31
MOR.B99990045	-0.53	3997.81421	-38900.75391	4.12	0.31
MOR.B99990007	-0.34	4077.92261	-38898.12109	04.02	0.30
MOR.B99990014	-0.52	3997.15503	-38893.96094	04.06	0.28

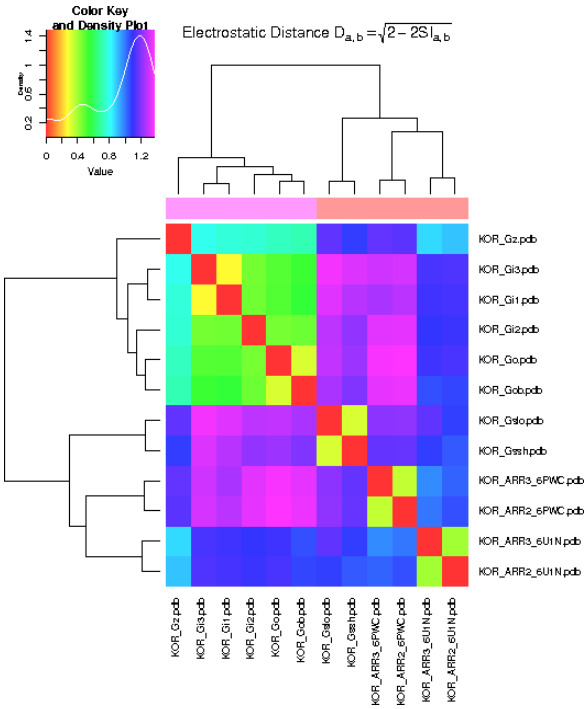
OR models	Z-score	molpdf	DOPE score	LGscore	MaxSub
NOP.B99990015	-1.07	3689.70557	-36778.55469	4.21	0.33
NOP.B99990052	-0.91	3693.27466	-36540.32031	4.21	0.33
NOP.B99990098	-1.22	3694.66406	-36341.45312	4.18	0.27
NOP.B99990096	-0.96	3682.33423	-36332.37109	4.13	0.32
NOP.B99990031	-1.06	3671.66724	-36302.58203	04.08	0.31
NOP.B99990009	-0.99	3703.85254	-36290.89453	4.18	0.32
NOP.B99990025	-1.13	3721.82397	-36285.18359	04.09	0.31
NOP.B99990083	-1.09	3631.37476	-36277.16797	4.15	0.31
NOP.B99990033	-1.12	3812.56567	-36266.87891	04.02	0.30
NOP.B99990063	-1.07	4512.14600	-36264.54688	4.20	0.30
NOP.B99990022	-0.92	3530.75854	-36259.64453	4.26	0.31

NOP.B99990036	-1.25	3766.60303	-36249.99219	4.16	0.31
NOP.B99990050	-1.31	3685.12305	-36241.48828	4.10	0.31
NOP.B99990056	-1.28	3634.51099	-36221.42969	04.09	0.27
NOP.B99990079	-1.16	3712.80029	-36209.31641	4.21	0.30
NOP.B99990037	-1.02	3761.07104	-36202.00391	4.18	0.31
NOP.B99990047	-1.07	3691.35449	-36201.05078	4.23	0.27
NOP.B99990001	-1.02	3865.72583	-36196.66797	4.29	0.32
NOP.B99990082	-1.12	3686.55273	-36189.30469	4.25	0.31
NOP.B99990094	-1.08	3553.03394	-36187.31250	4.11	0.29

Annex 4: Electrostatic distances, calculated by PIPSA webserver, and presented through heatmaps. A) DOR complexes B) NOP complexes C) KOR complexes

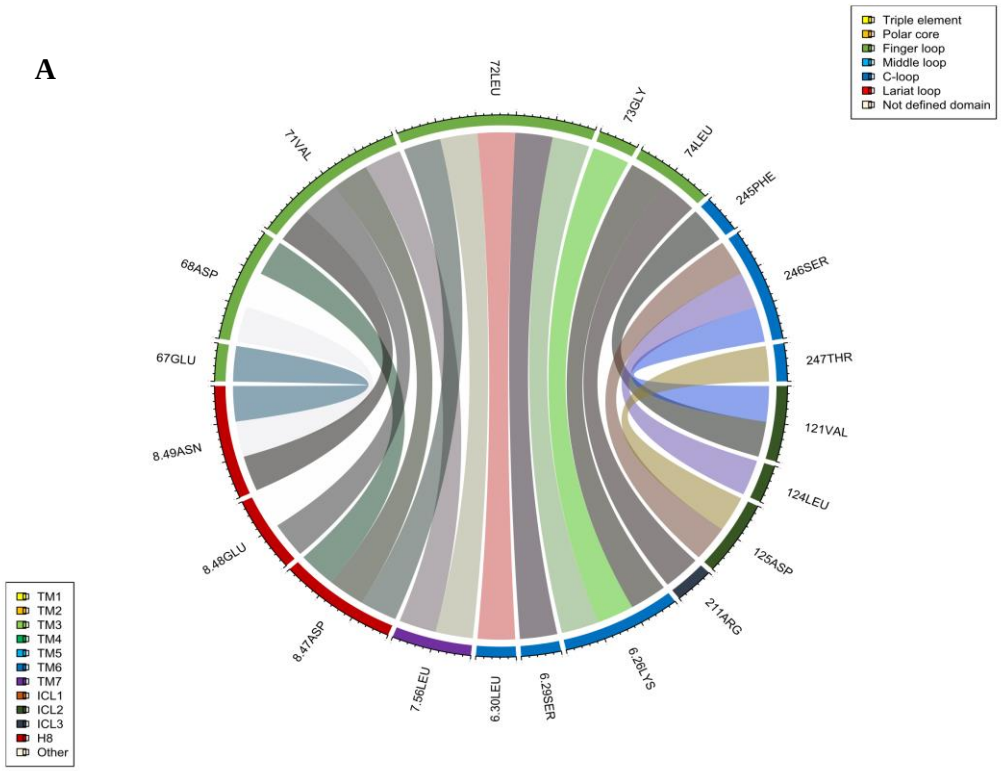


C

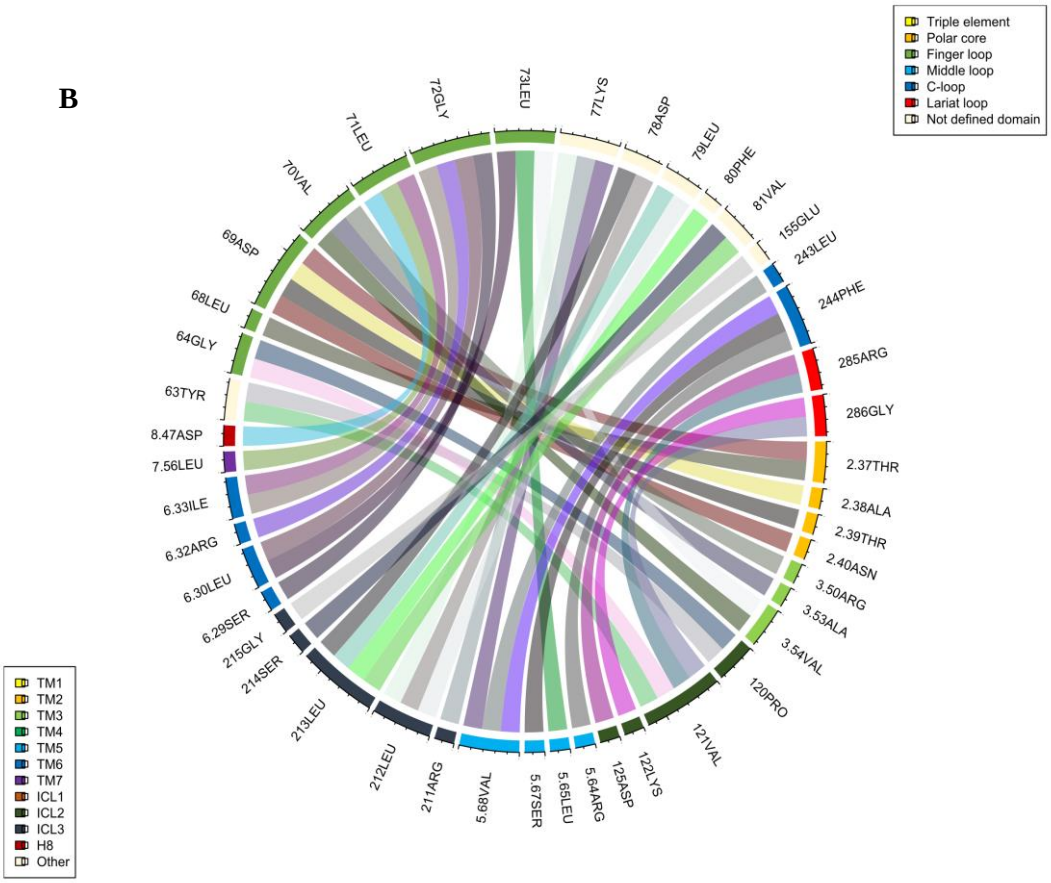


Annex 5: DOR-Arrestins interaction plots, built using *circlize* package for R language. A) DOR-Arr3_6PWC B) DOR-Arr2_6U1N C) DOR-Arr3_6U1N

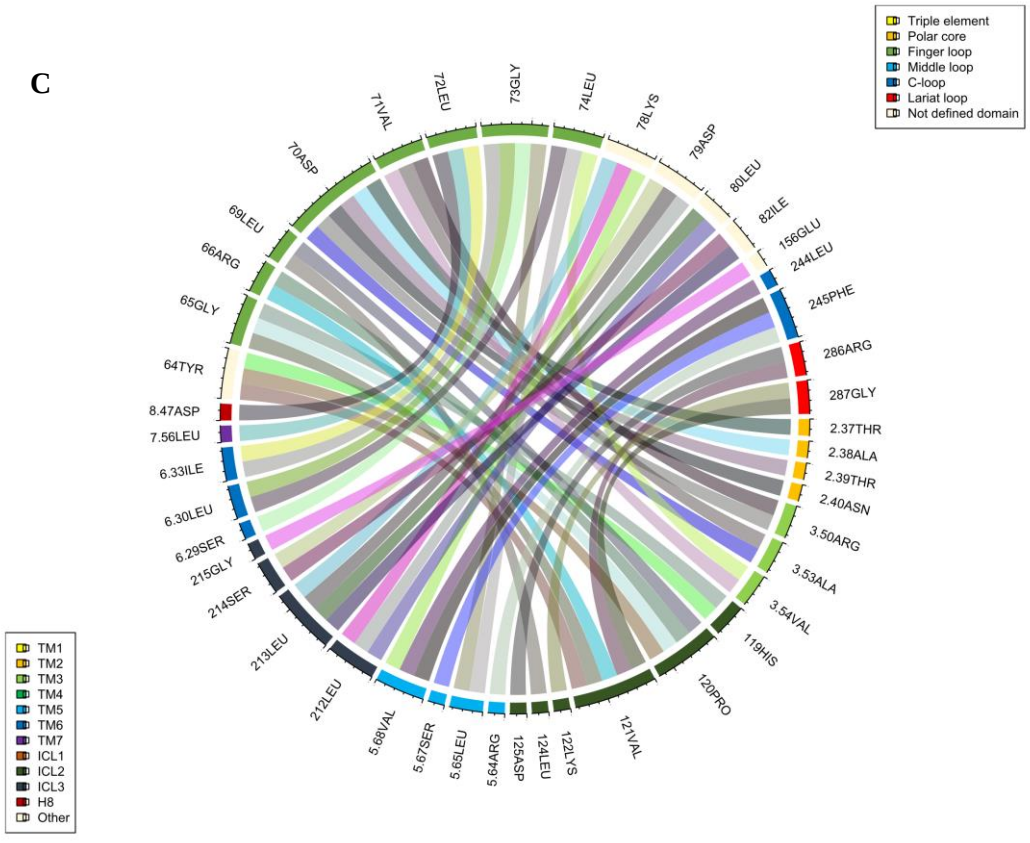
A



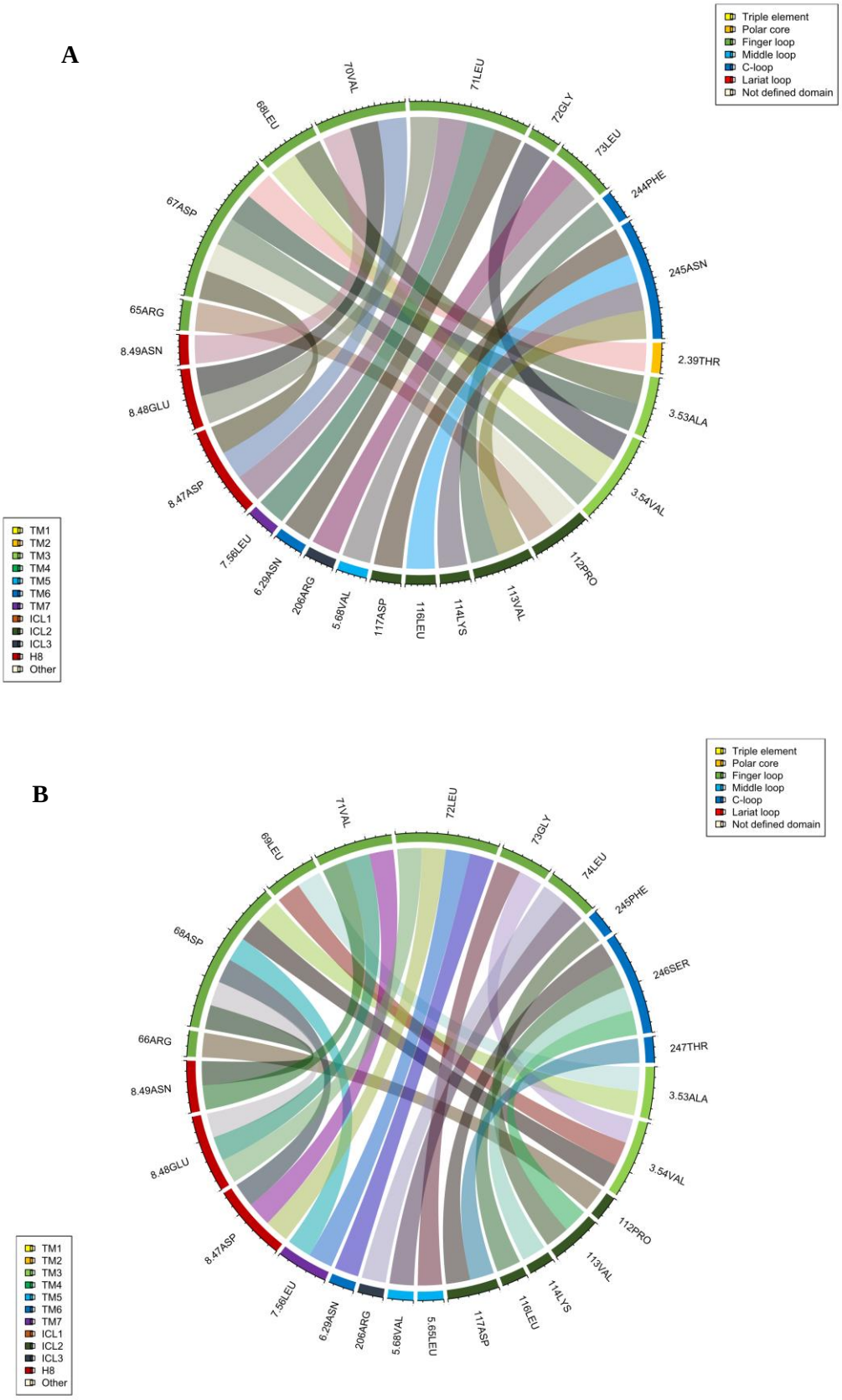
B


Triple element
Polar core
Finger loop
Middle loop
C-loop
Lariat loop
Not defined domain
73LEU
77LYS
78ASP
79LEU
80PHE
81VAL
155GLU
243LEU
244PHE
285ARG
286GLY
2.37THR
2.38ALA
2.39THR
2.40ASN
3.50ARG
3.53ALA
3.54VAL
120PRO
121VAL
122LYS
125ASP
5.64ARG
5.65LEU
5.67SER
5.68VAL
211ARG
212LEU
213LEU
214SER
215GLY
6.29SER
6.30LEU
6.32ARG
6.33ILE
7.56LEU
8.47ASP
63TYR
64GLY
68LEU
69ASP
70VAL
71LEU
72GLY
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

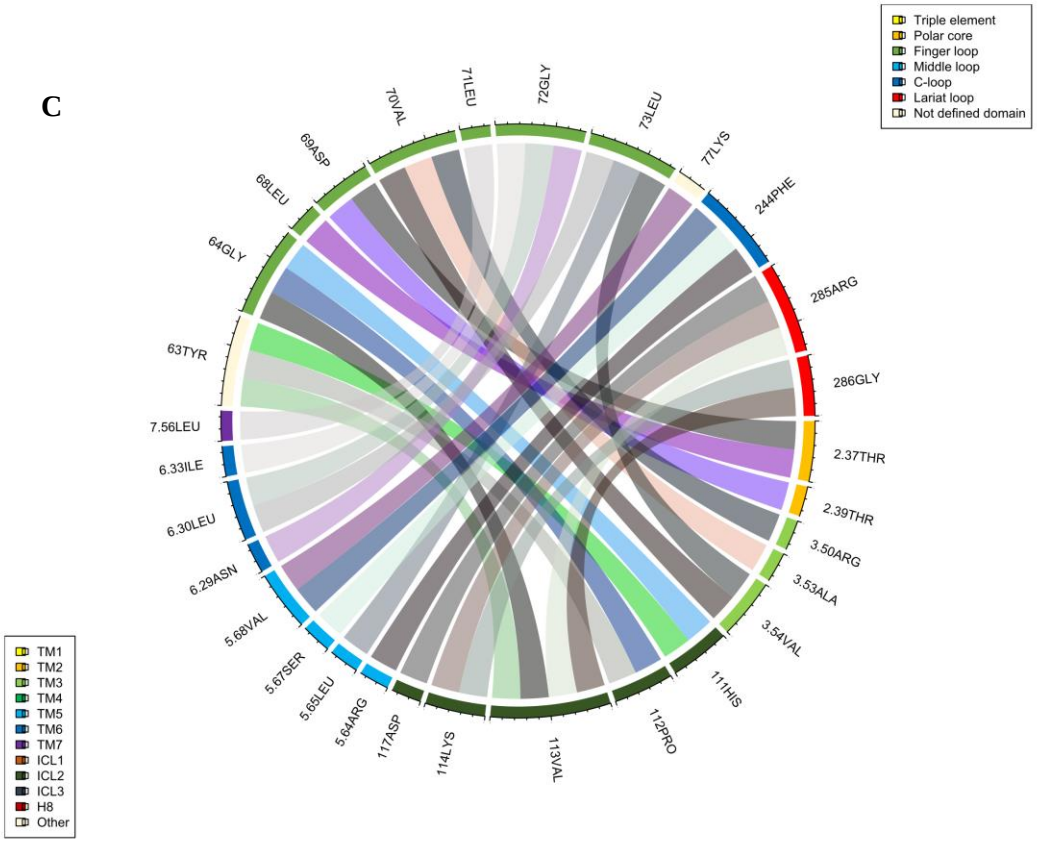
C


Triple element
Polar core
Finger loop
Middle loop
C-loop
Lariat loop
Not defined domain
73GLY
74LEU
78LYS
79ASP
80LEU
82ILE
156GLU
244LEU
245PHE
286ARG
287GLY
2.37THR
2.38ALA
2.39THR
2.40ASN
3.50ARG
3.53ALA
3.54VAL
119HIS
120PRO
121VAL
122LYS
124LEU
125ASP
5.64ARG
5.65LEU
5.67SER
5.68VAL
212LEU
213LEU
214SER
215GLY
6.29SER
6.30LEU
6.33ILE
7.56LEU
8.47ASP
64TYR
65GLY
66ARG
69LEU
70ASP
71VAL
72LEU
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

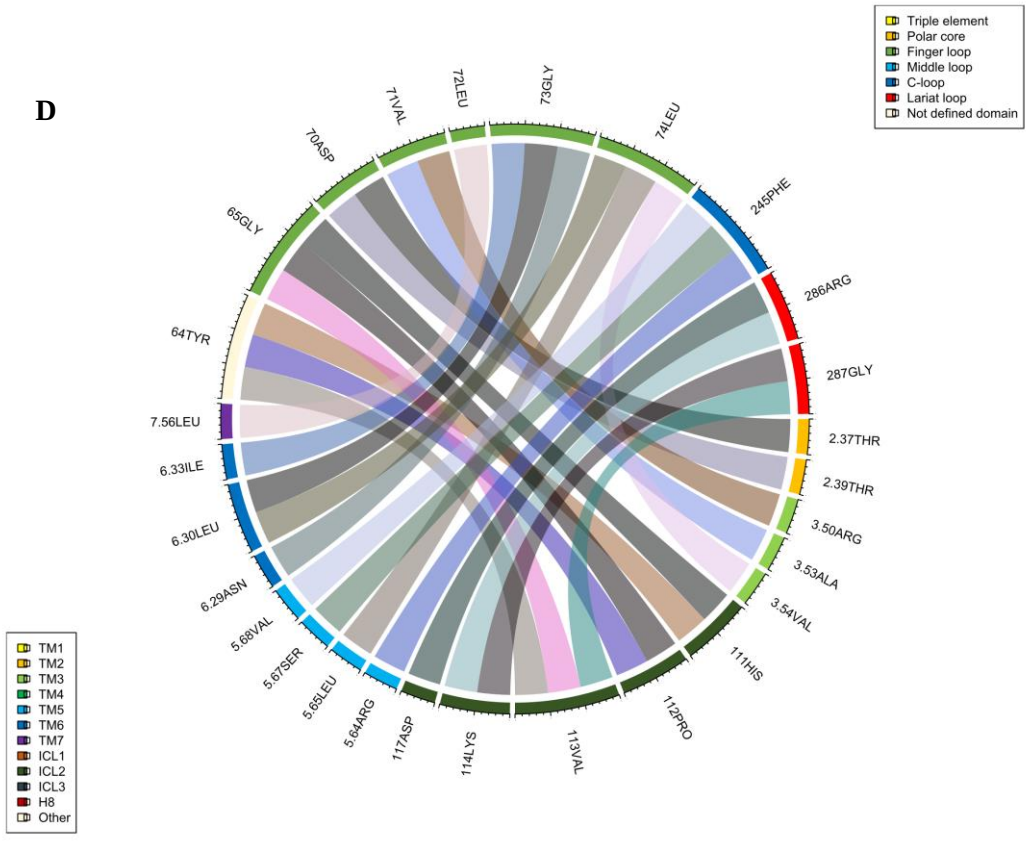
Annex 6: KOR-Arrestins interaction plots, built using *circlize* package for R language. A) KOR-Arr2_6PWC B) KOR-Arr3_6PWC C) KOR-Arr2_6U1N D) KOR-Arr3_6U1N



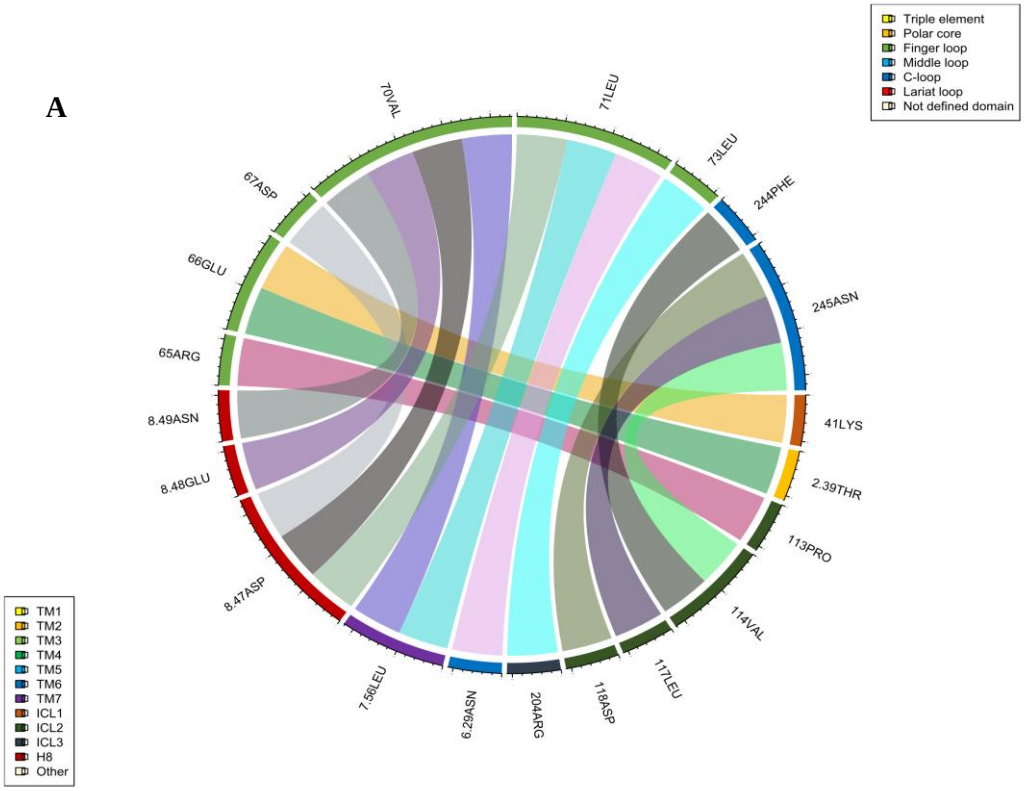
C

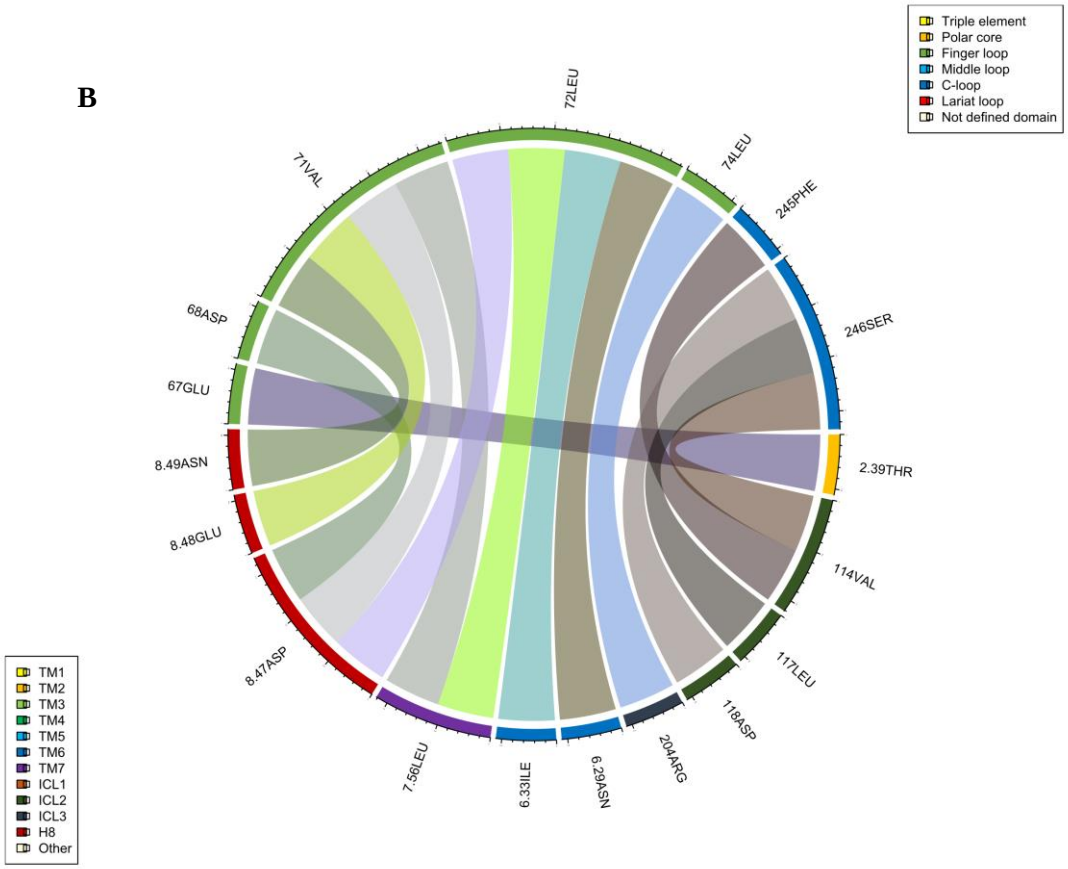

Triple element
Polar core
Finger loop
Middle loop
C-loop
Lariat loop
Not defined domain
71LEU
72GLY
73LEU
77LYS
244PHE
285ARG
286GLY
2.37THR
2.39THR
3.50ARG
3.53ALA
3.54VAL
111HIS
112PRO
113VAL
114LYS
117ASP
5.64ARG
5.65LEU
5.67SER
5.68VAL
6.29ASN
6.30LEU
6.33ILE
7.56LEU
63TYR
64GLY
68LEU
69ASP
70VAL
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

D

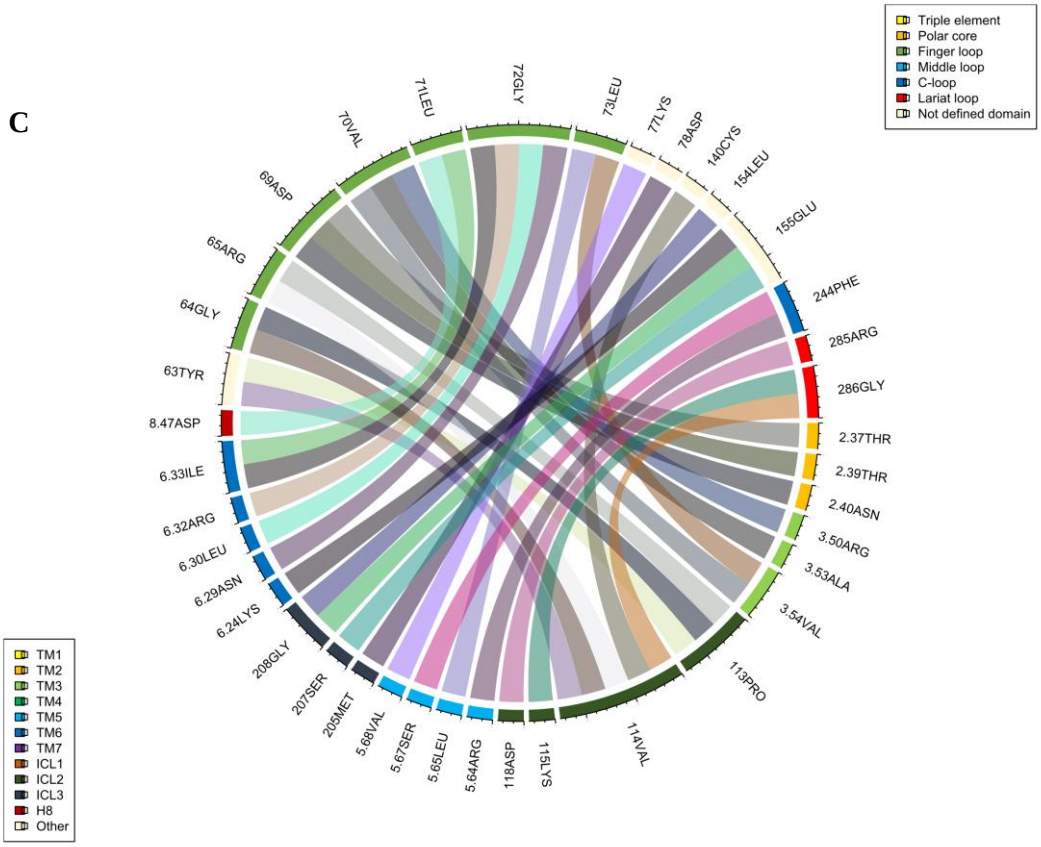

Triple element
Polar core
Finger loop
Middle loop
C-loop
Lariat loop
Not defined domain
72LEU
73GLY
74LEU
245PHE
286ARG
287GLY
2.37THR
2.39THR
3.50ARG
3.53ALA
3.54VAL
111HIS
112PRO
113VAL
114LYS
117ASP
5.64ARG
5.65LEU
5.67SER
5.68VAL
6.29ASN
6.30LEU
6.33ILE
7.56LEU
64TYR
65GLY
70ASP
71VAL
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

Annex 7: MOR-Arrestins interaction plots, built using *circlize* package for R language. A) MOR-Arr2_6PWC B) MOR-Arr3_6PWC C) MOR-Arr2_6U1N D) MOR-Arr3_6U1N

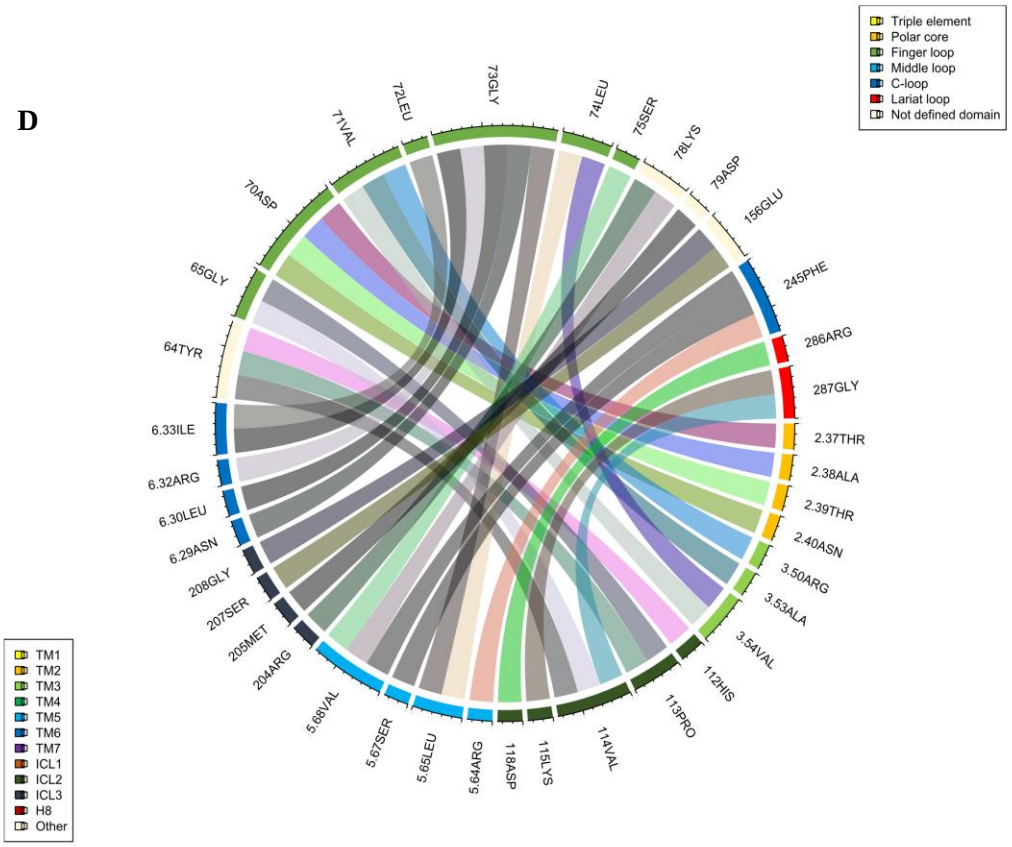




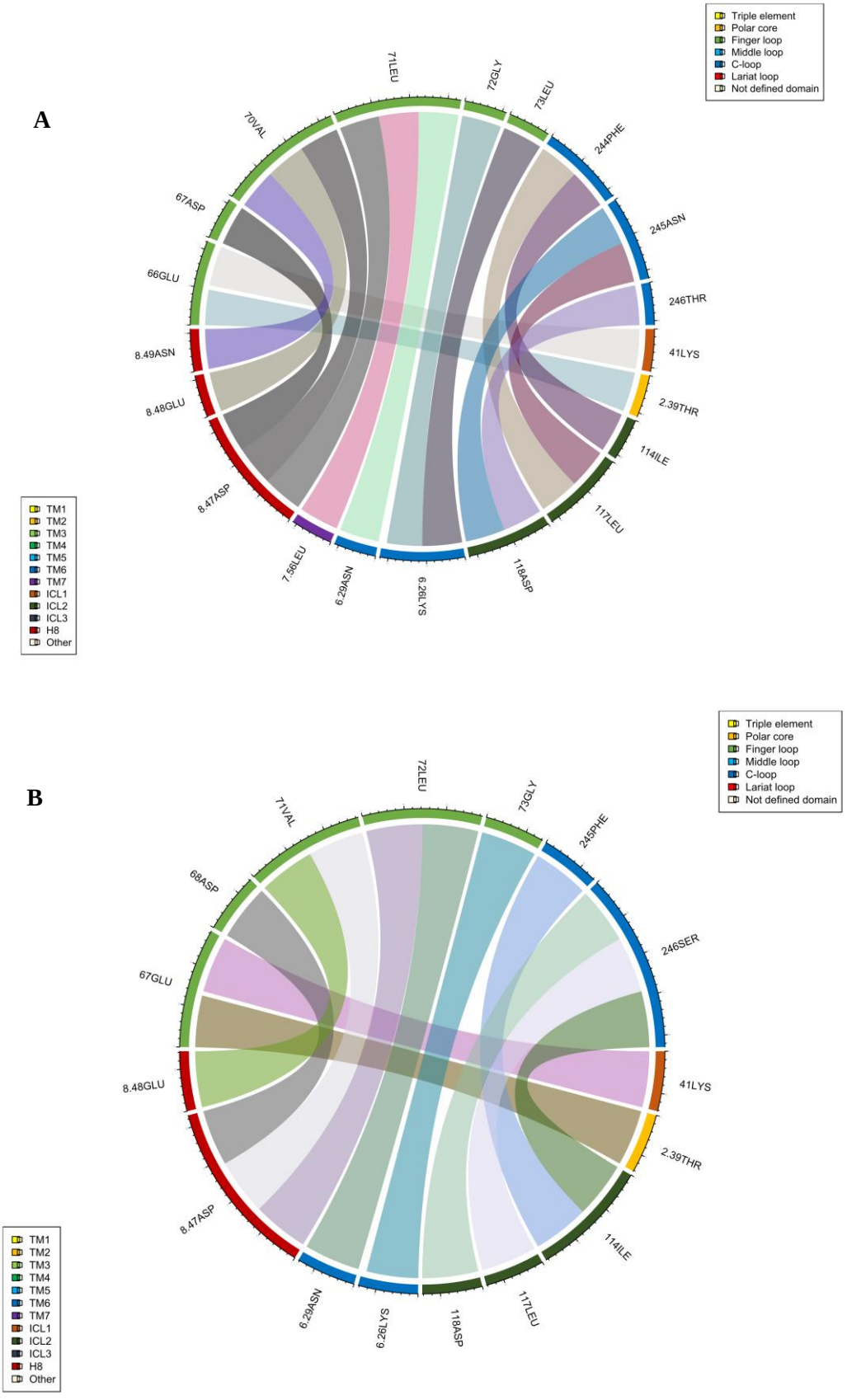
C


Triple element
Polar core
Finger loop
Middle loop
C-loop
Lariat loop
Not defined domain
72GLY
73LEU
77LYS
78ASP
140CYS
154LEU
159GLU
244PHE
285ARG
286GLY
2.37THR
2.39THR
2.40ASN
3.50ARG
3.53ALA
3.54VAL
113PRO
114VAL
115LYS
118ASP
5.64ARG
5.65LEU
5.67SER
5.68VAL
205MET
207SER
208GLY
6.24LYS
6.29ASN
6.30LEU
6.32ARG
6.33ILE
8.47ASP
63TYR
64GLY
65ARG
69ASP
70VAL
71LEU
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

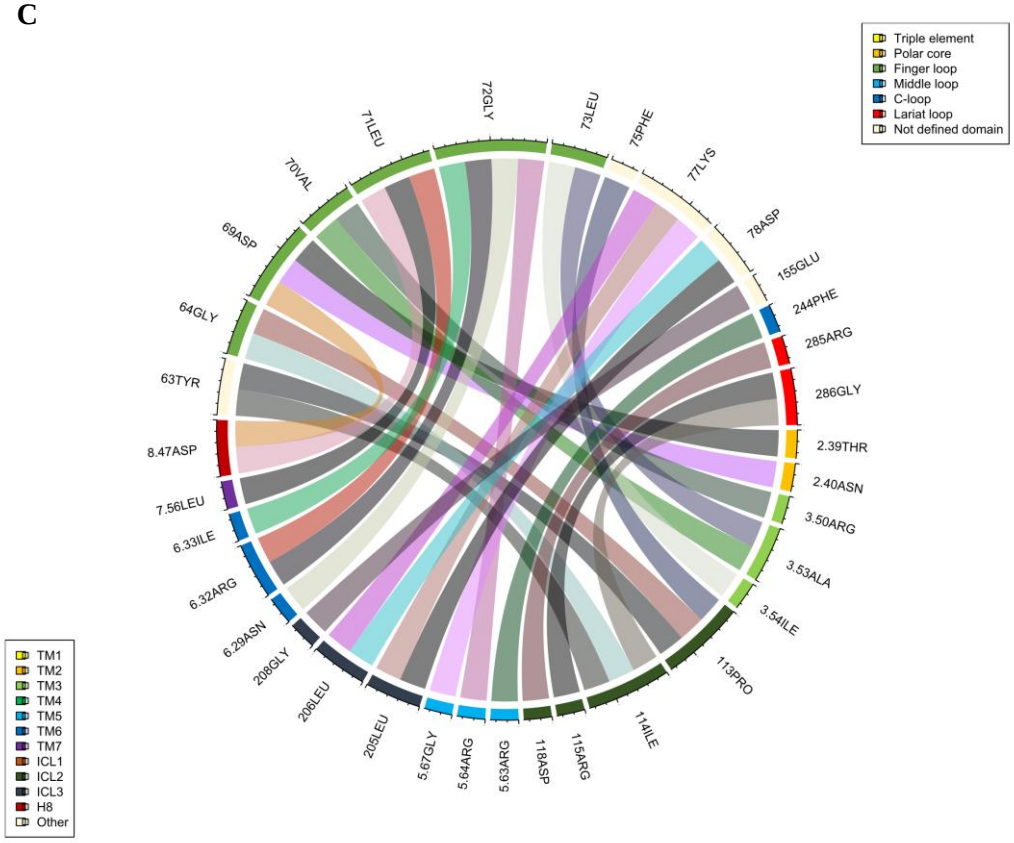
D


Triple element
Polar core
Finger loop
Middle loop
C-loop
Lariat loop
Not defined domain
72GLY
74LEU
75SER
78LYS
79ASP
156GLU
245PHE
286ARG
287GLY
2.37THR
2.38ALA
2.39THR
2.40ASN
3.50ARG
3.53ALA
3.54VAL
112HIS
113PRO
114VAL
115LYS
118ASP
5.64ARG
5.65LEU
5.67SER
5.68VAL
204ARG
205MET
207SER
208GLY
6.29ASN
6.30LEU
6.32ARG
6.33ILE
64TYR
65GLY
70ASP
71VAL
72LEU
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

Annex 8: NOP-Arrestins interaction plots, built using *circlize* package for R language. A) NOP-Arr2_6PWC B) NOP-Arr3_6PWC C) NOP-Arr2_6U1N

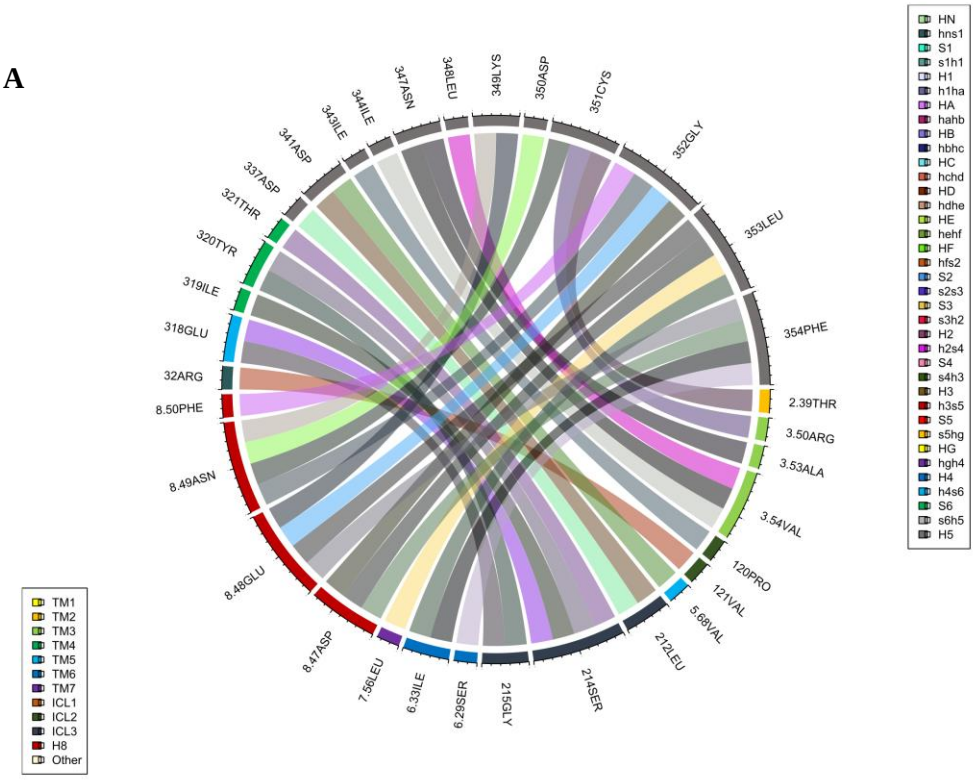


C

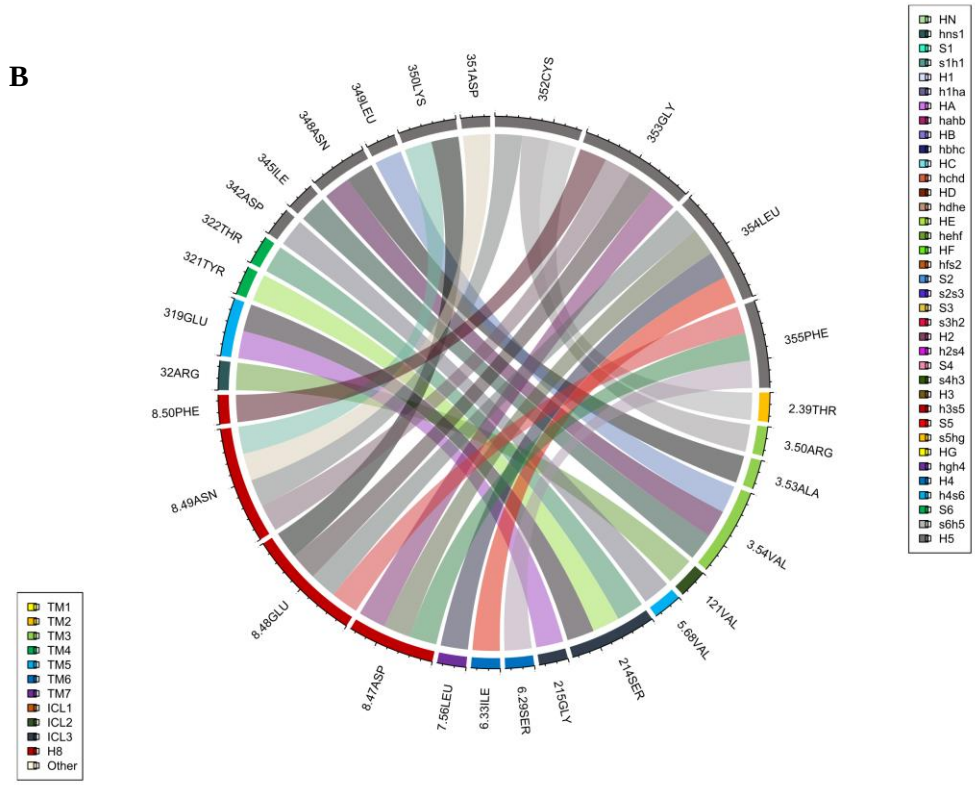

Triple element
Polar core
Finger loop
Middle loop
C-loop
Lariat loop
Not defined domain
72GLY
73LEU
75PHE
77LYS
78ASP
155GLU
244PHE
285ARG
286GLY
2.39THR
2.40ASN
3.50ARG
3.53ALA
3.54ILE
113PRO
114ILE
115ARG
118ASP
5.63ARG
5.64ARG
5.67GLY
205LEU
206LEU
208GLY
6.29ASN
6.32ARG
6.33ILE
7.56LEU
8.47ASP
63TYR
64GLY
69ASP
70VAL
71LEU
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

Annex 9: DOR-Gi/o interaction plots, built using *circlize* package for R language. A) DOR-Gi1 B) DOR-Gi2 C) DOR-Go D) DOR-Gob E) DOR-Gz

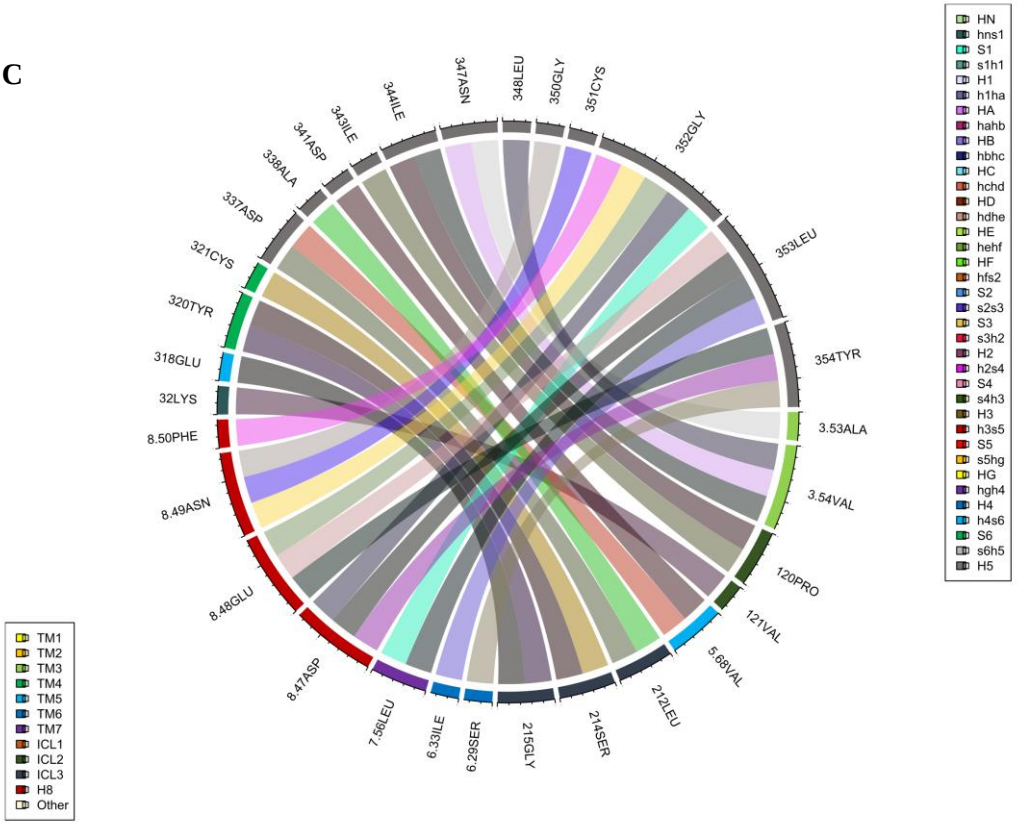
A



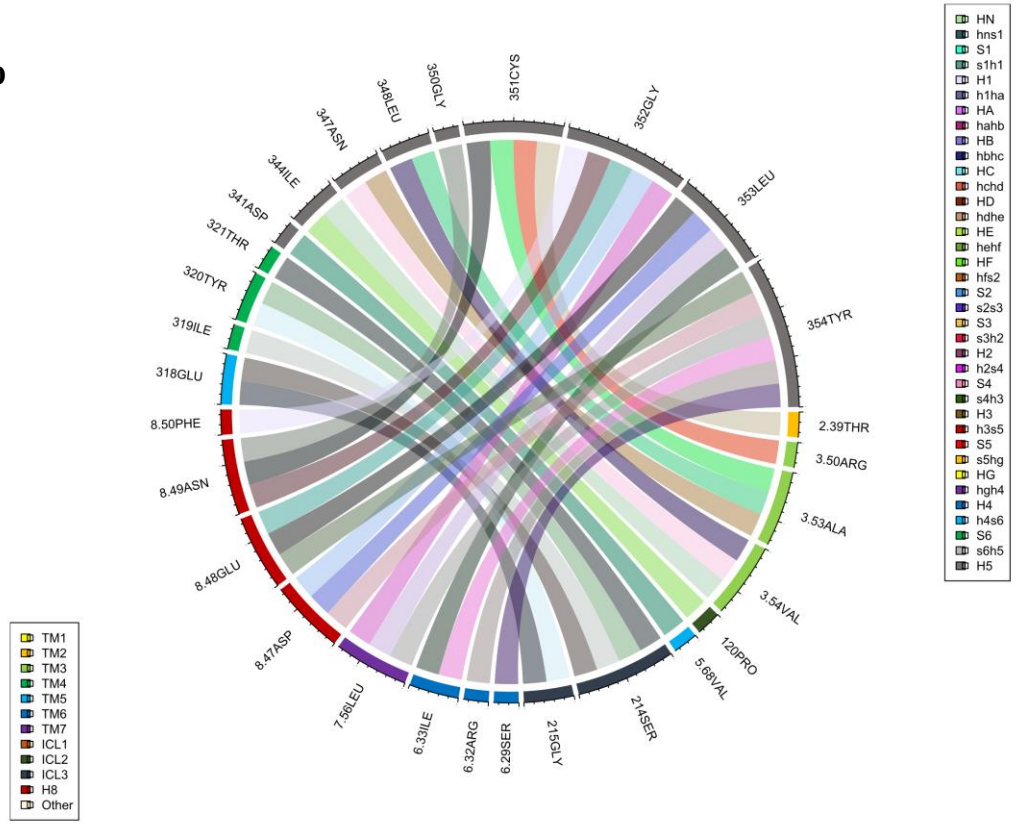
B



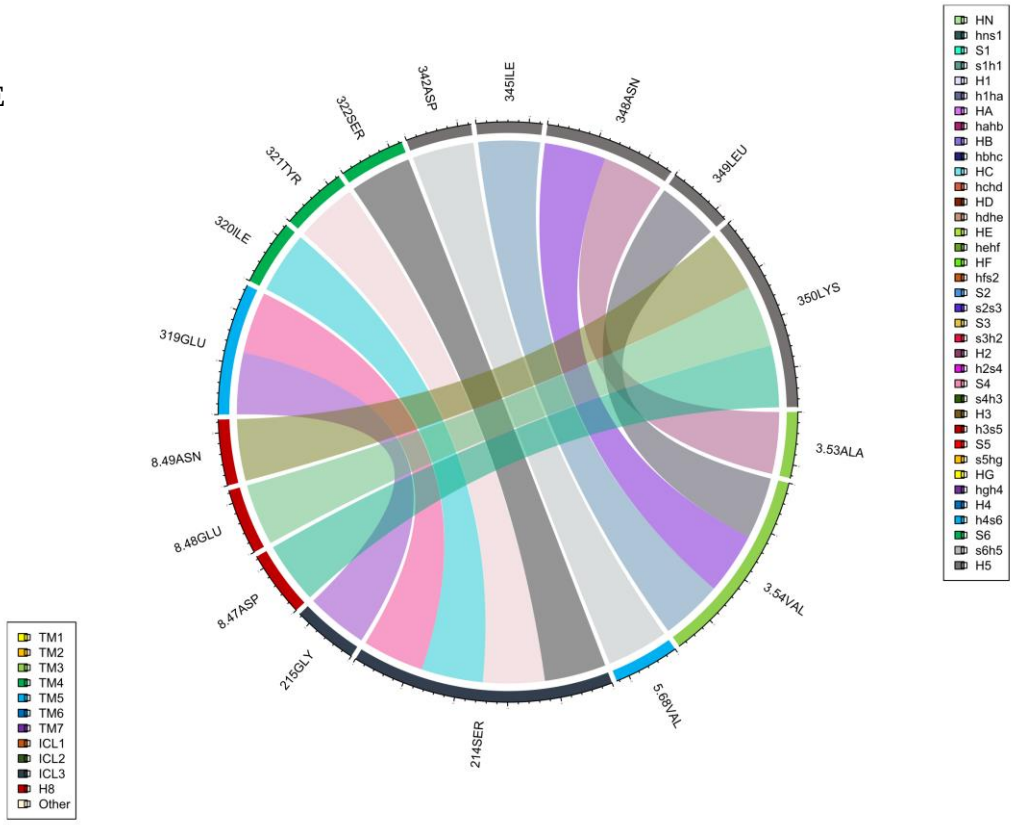
C


347ASN
348LEU
350GLY
351CYS
352GLY
353LEU
354TYR
3.53ALA
3.54VAL
120PRO
121VAL
5.68VAL
212LEU
214SER
215GLY
6.29SER
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
32LYS
318GLU
320TYR
321CYS
337ASP
338ALA
341ASP
343ILE
344ILE
HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

D

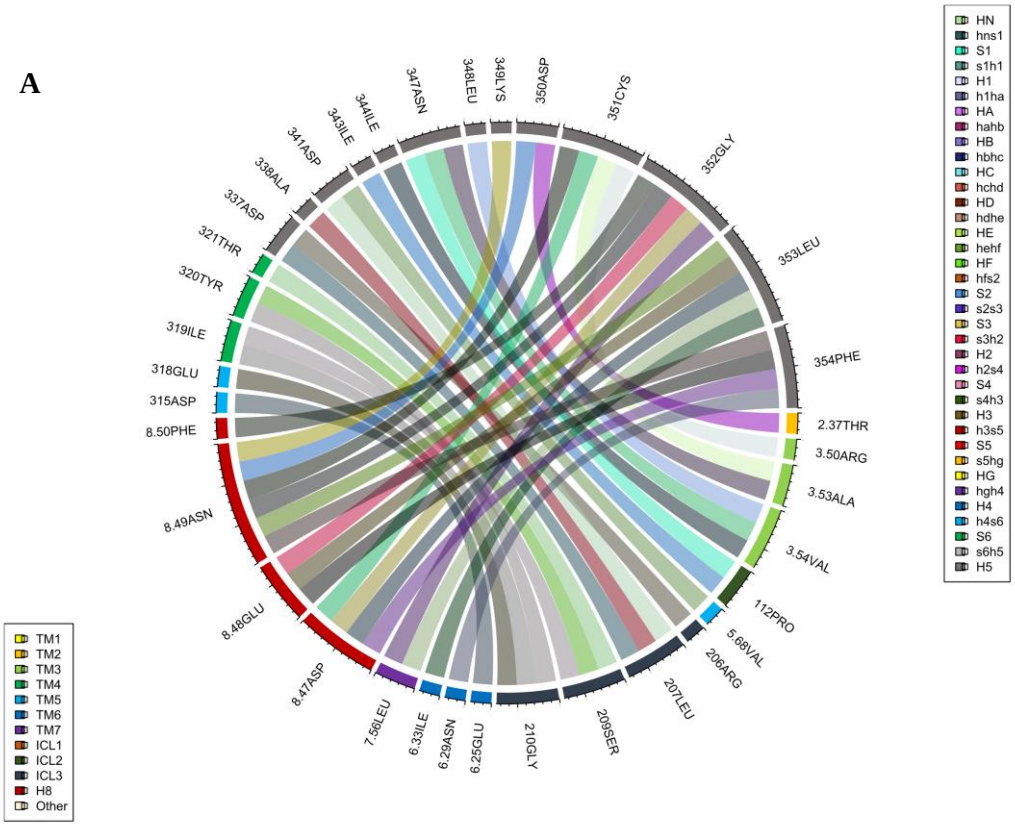

347ASN
348LEU
350GLY
351CYS
352GLY
353LEU
354TYR
2.39THR
3.50ARG
3.53ALA
3.54VAL
120PRO
5.68VAL
214SER
215GLY
6.29SER
6.32ARG
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
318GLU
319ILE
320TYR
321THR
341ASP
344ILE
HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

E

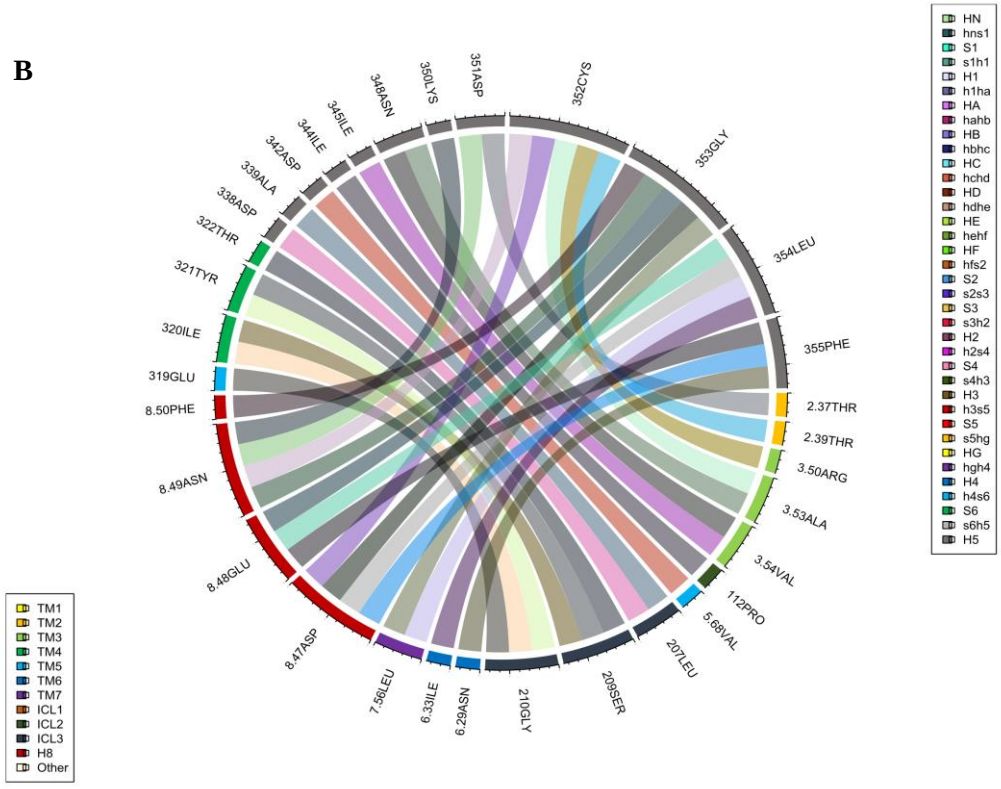

342ASP
345ILE
348ASN
349LEU
350LYS
3.53ALA
3.54VAL
5.68VAL
214SER
215GLY
8.47ASP
8.48GLU
8.49ASN
319GLU
320ILE
321TYR
322SER
HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

Annex 10: KOR-Gi/o interaction plots, built using *circlize* package for R language. A) KOR-Gi1 B) KOR-Gi2 C) KOR-Gi3 D) KOR-Go E) KOR-Gob F) KOR-Gz

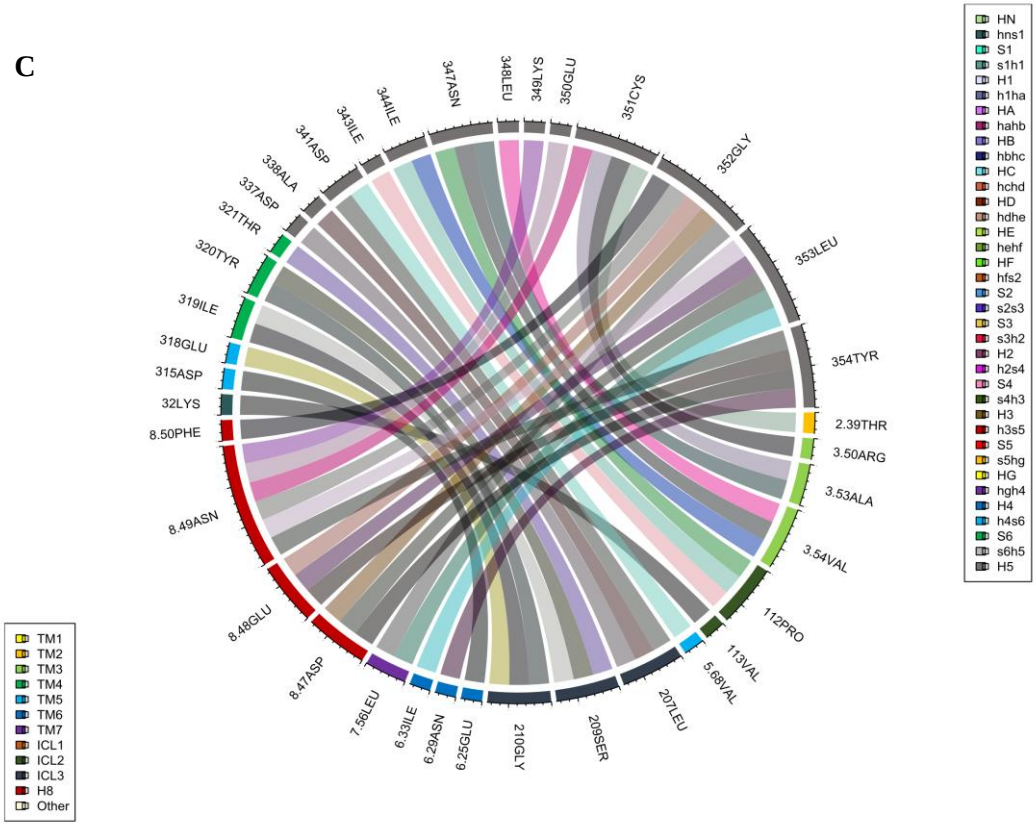
A



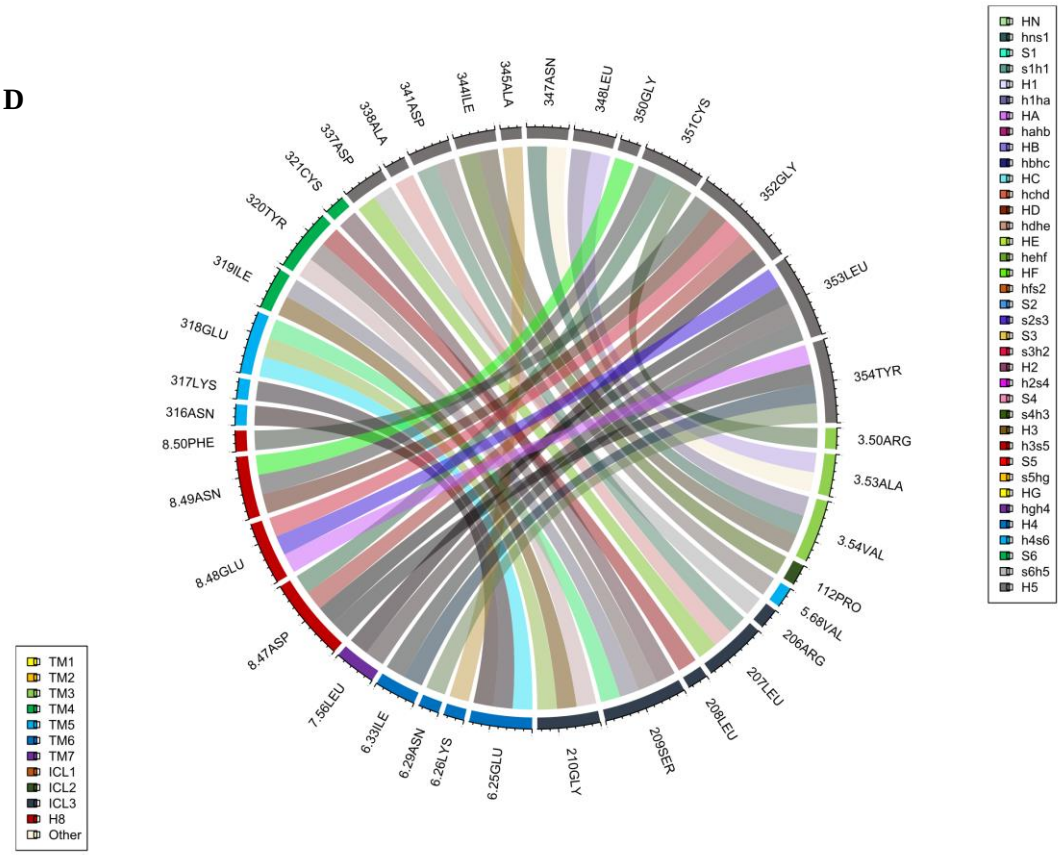
B



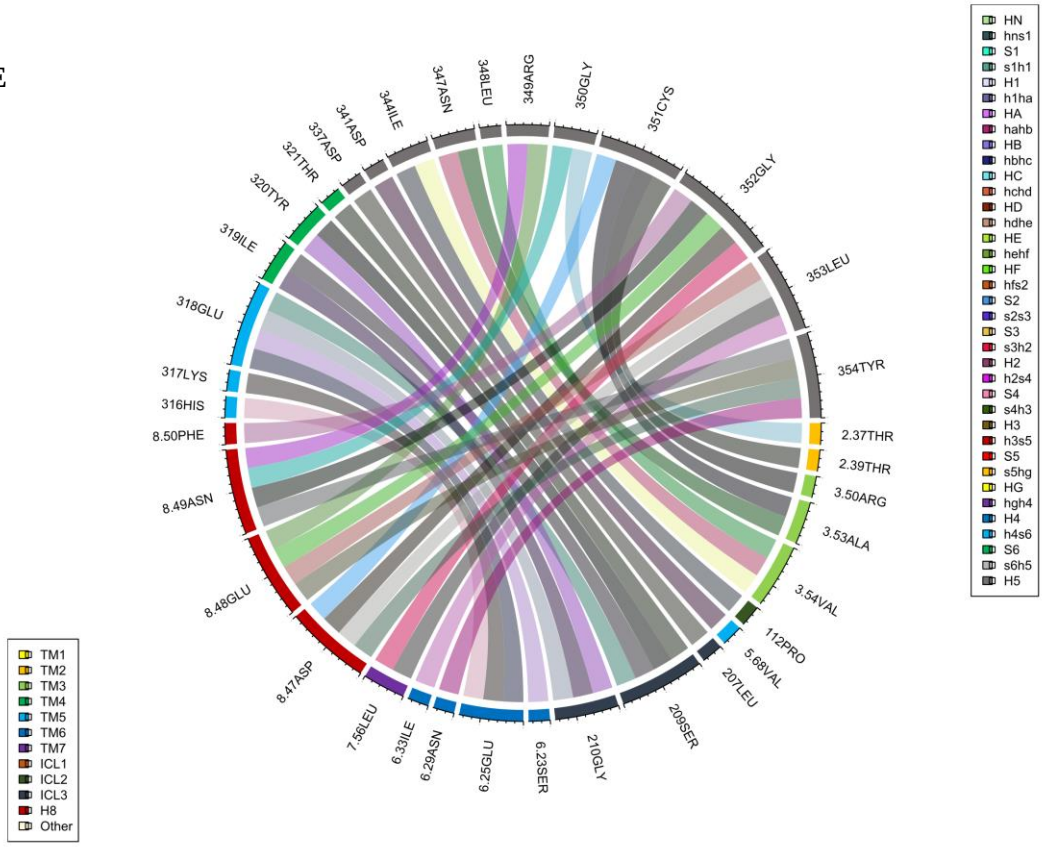
C


HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
347ASN
348LEU
349LYS
350GLU
351CYS
352GLY
353LEU
354TYR
2.39THR
3.50ARG
3.53ALA
3.54VAL
112PRO
113VAL
5.68VAL
207LEU
209SER
210GLY
6.25GLU
6.29ASN
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
32LYS
315ASP
318GLU
319ILE
320TYR
321THR
337ASP
338ALA
341ASP
343ILE
344ILE

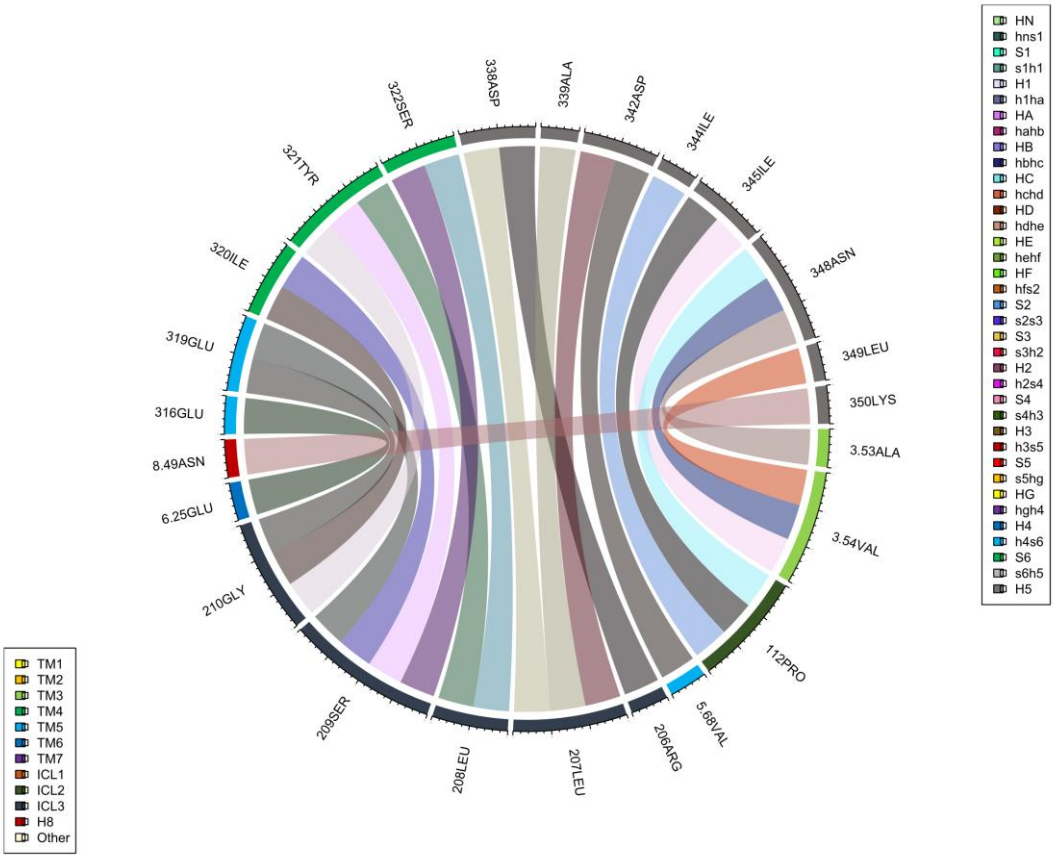
D


HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
345ALA
347ASN
348LEU
350GLY
351CYS
352GLY
353LEU
354TYR
3.50ARG
3.53ALA
3.54VAL
112PRO
5.68VAL
206ARG
207LEU
208LEU
209SER
210GLY
6.25GLU
6.26LYS
6.29ASN
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
316ASN
317LYS
318GLU
319ILE
320TYR
321CYS
337ASP
338ALA
341ASP
344ILE

E

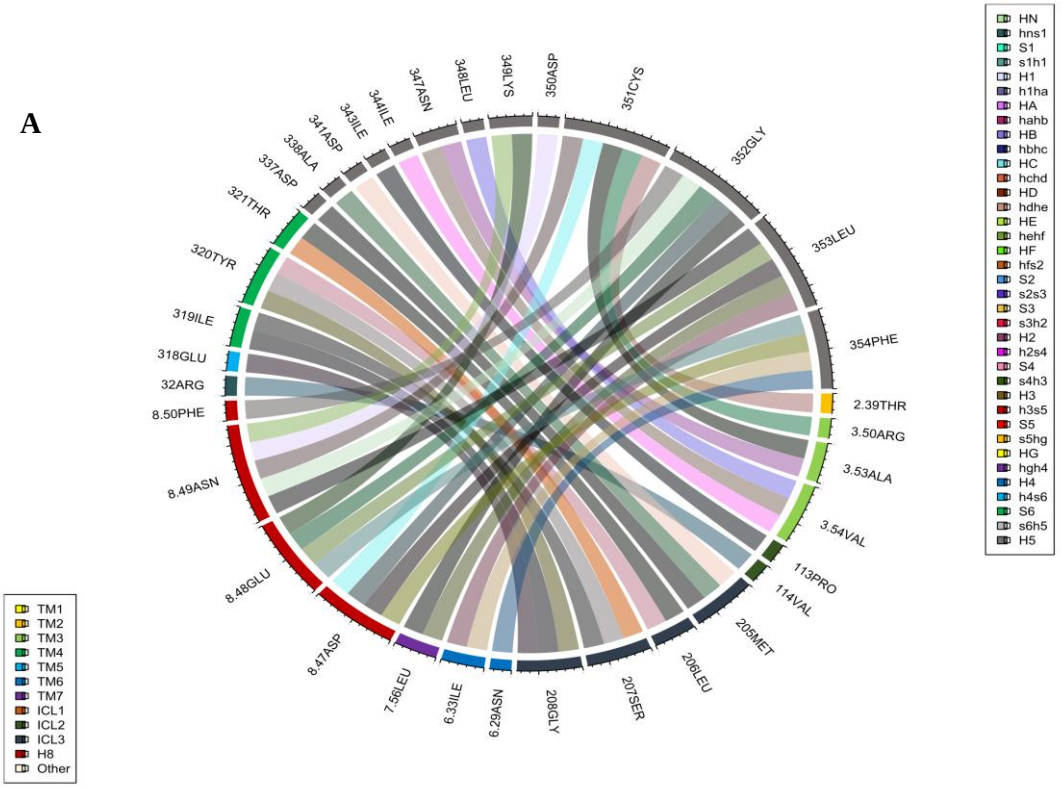

HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
350GLY
351CYS
352GLY
353LEU
354TYR
2.37THR
2.39THR
3.50ARG
3.53ALA
3.54VAL
112PRO
5.68VAL
207LEU
209SER
210GLY
6.23SER
6.25GLU
6.29ASN
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
316HIS
317LYS
318GLU
319ILE
320TYR
321THR
337ASP
341ASP
344ILE
347ASN
348LEU
349ARG

F

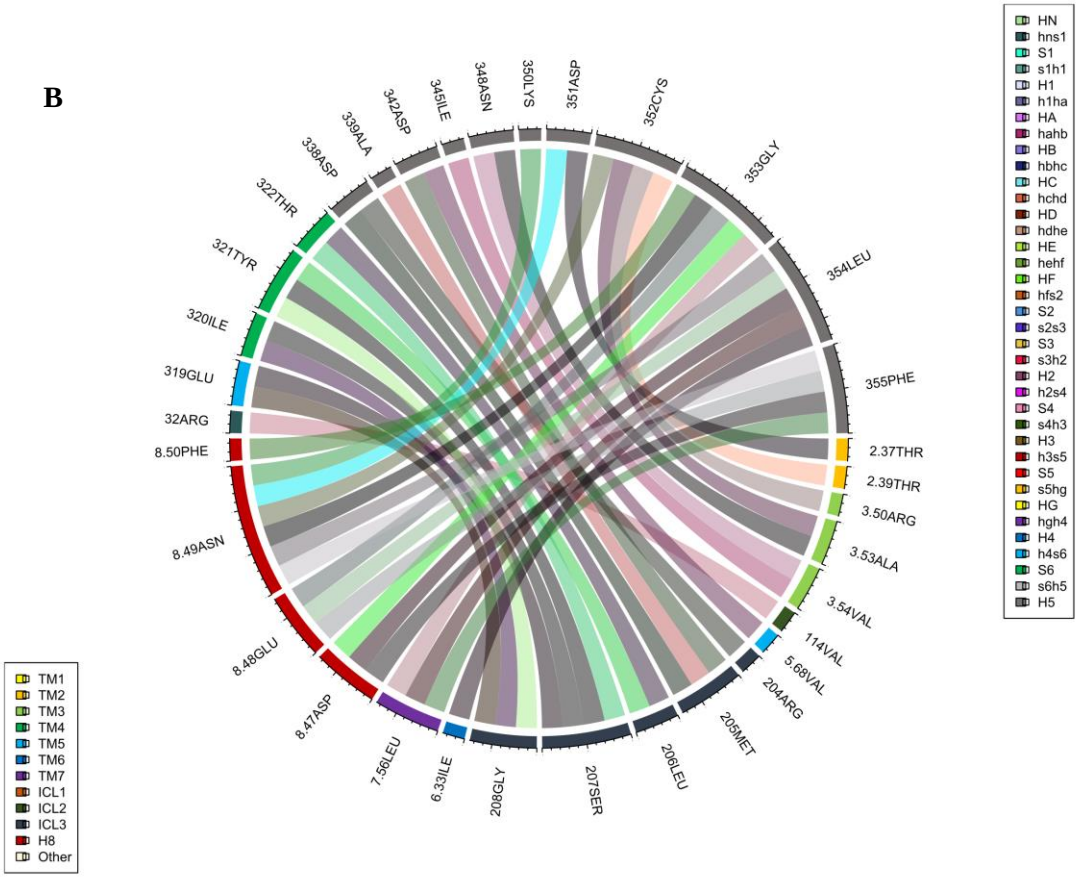

HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
338ASP
339ALA
342ASP
344ILE
345ILE
348ASN
349LEU
350LYS
3.53ALA
3.54VAL
112PRO
5.68VAL
206ARG
207LEU
208LEU
209SER
210GLY
6.25GLU
8.49ASN
316GLU
319GLU
320ILE
321TYR
322SER

Annex 11: MOR-Gi/o interaction plots, built using *circlize* package for R language. A) MOR-Gi1 B) MOR-Gi2 C) MOR-Gi3 D) MOR-Go E) MOR-Gob F) MOR-Gz

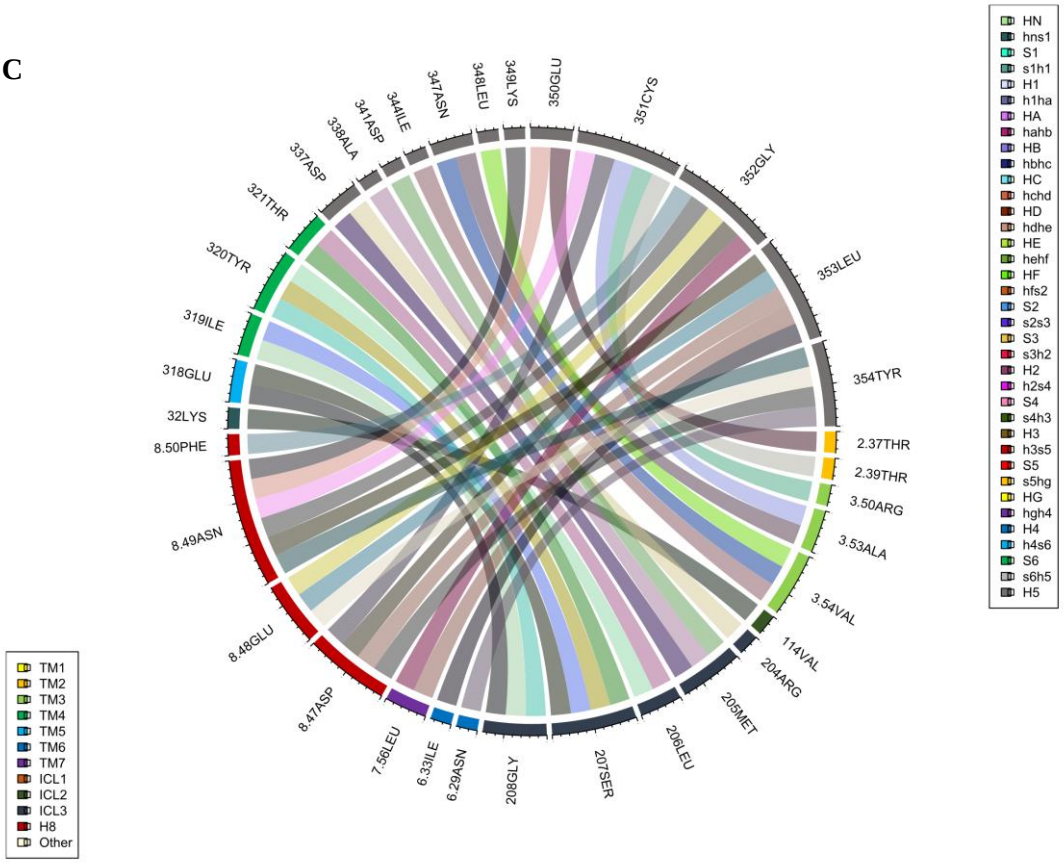
A



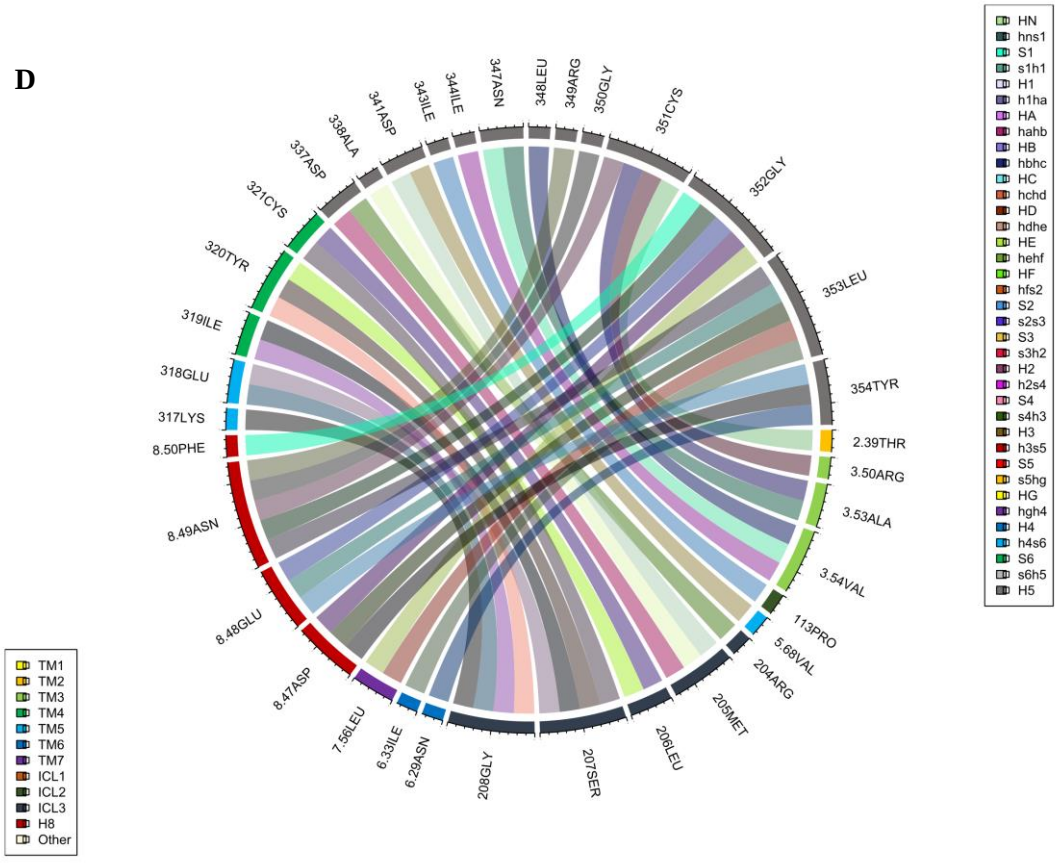
B



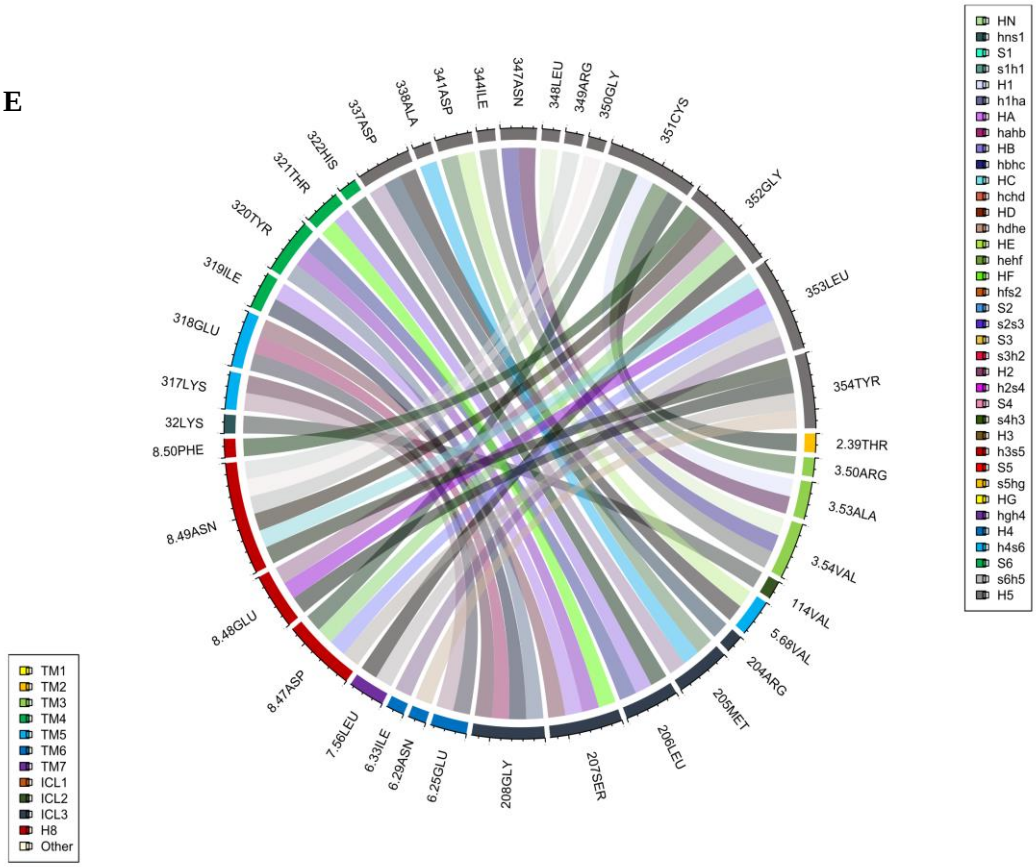
C


HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
350GLU
351CYS
352GLY
353LEU
354TYR
2.37THR
2.39THR
3.50ARG
3.53ALA
3.54VAL
114VAL
204ARG
205MET
206LEU
207SER
208GLY
6.29ASN
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
32LYS
318GLU
319ILE
320TYR
321THR
337ASP
338ALA
341ASP
344ILE
347ASN
348LEU
349LYS
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

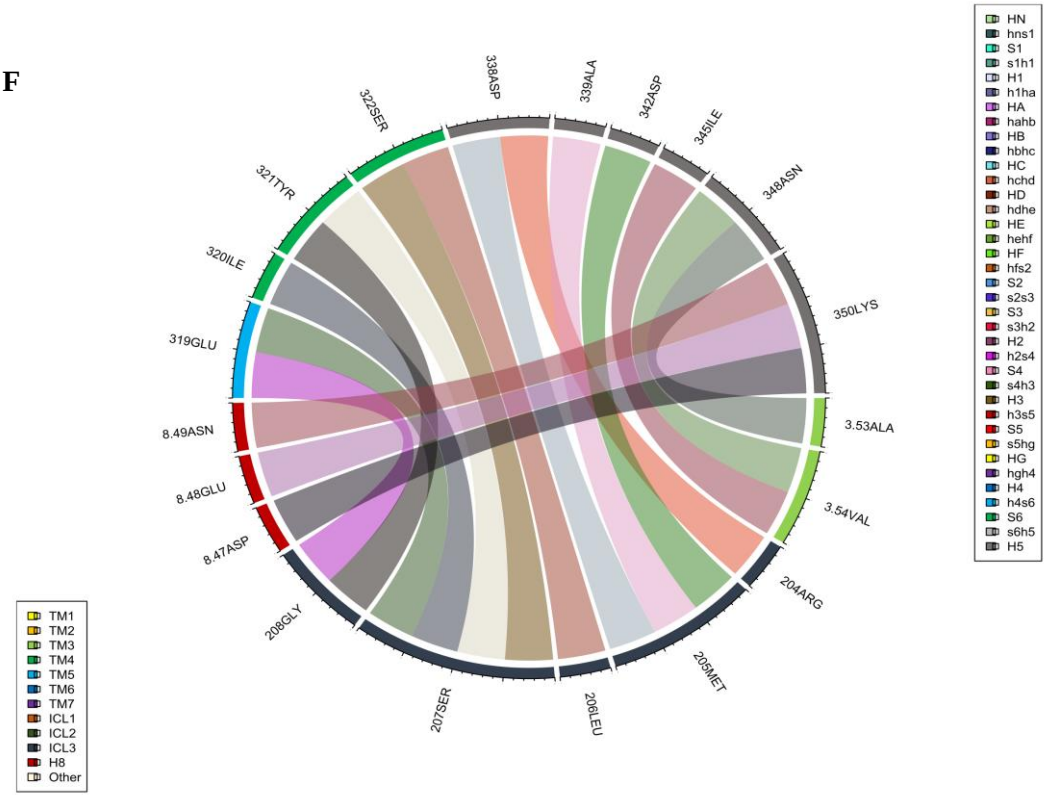
D


HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
349ARG
350GLY
351CYS
352GLY
353LEU
354TYR
2.39THR
3.50ARG
3.53ALA
3.54VAL
113PRO
5.68VAL
204ARG
205MET
206LEU
207SER
208GLY
6.29ASN
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
317LYS
318GLU
319ILE
320TYR
321CYS
337ASP
338ALA
341ASP
343ILE
344ILE
347ASN
348LEU
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other

E

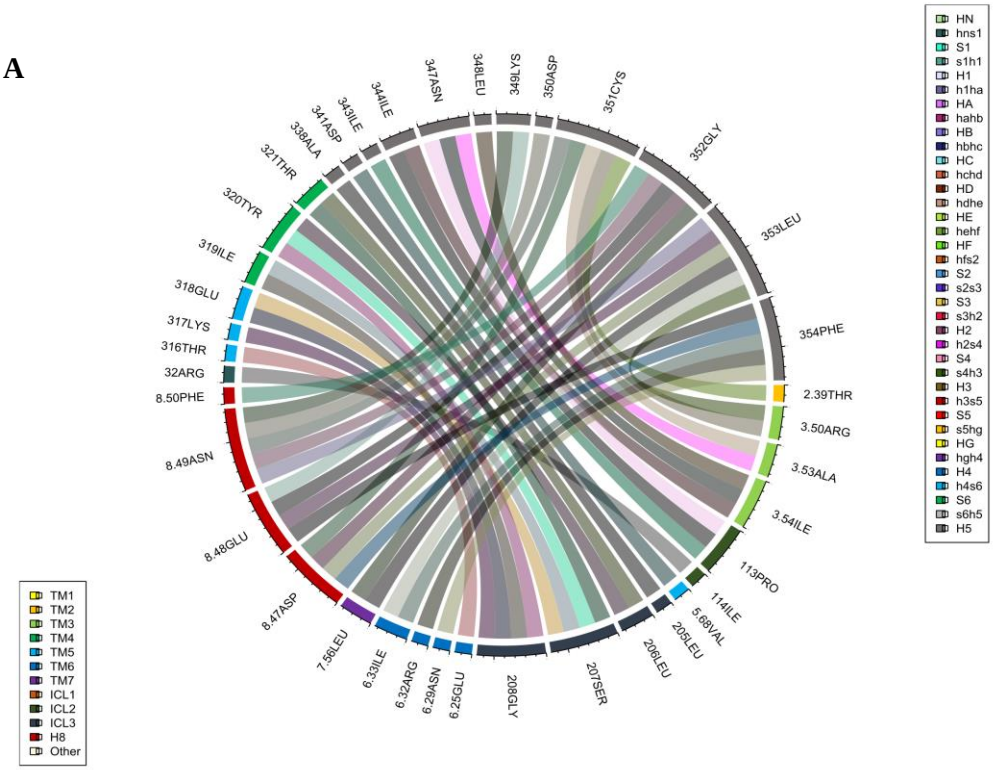

HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
322HIS
321THR
320TYR
319ILE
318GLU
317LYS
32LYS
8.50PHE
8.49ASN
8.48GLU
8.47ASP
7.56LEU
6.33ILE
6.29ASN
6.25GLU
208GLY
207SER
206LEU
205MET
204ARG
5.68VAL
114VAL
3.54VAL
3.53ALA
3.50ARG
2.39THR
354TYR
353LEU
352GLY
351CYS
350GLY
349ARG
348LEU
347ASN
344ILE
341ASP
338ALA
337ASP

F

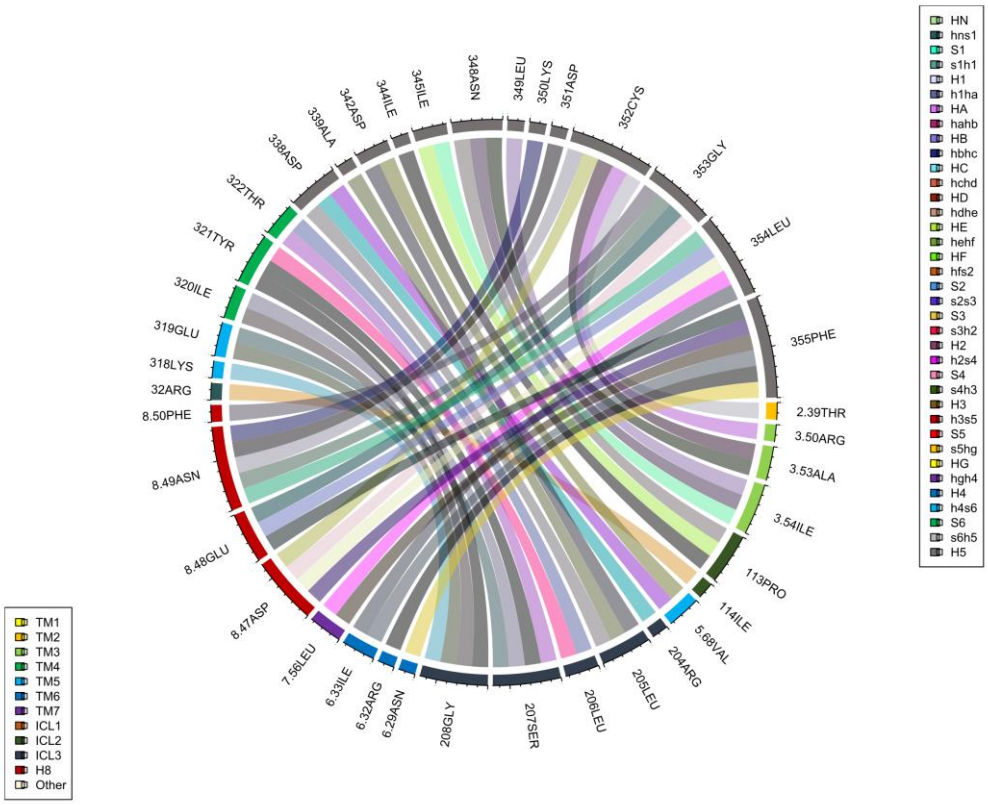

HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
322SER
321TYR
320ILE
319GLU
8.49ASN
8.48GLU
8.47ASP
208GLY
207SER
206LEU
205MET
204ARG
3.54VAL
3.53ALA
350LYS
348ASN
345ILE
342ASP
339ALA
338ASP

Annex 12: NOP-Gi/o interaction plots, built using *circlize* package for R language. A) NOP-Gi1 B) NOP-Gi2 C) NOP-Gi3 D) NOP-Go E) NOP-Gob F) NOP-Gz

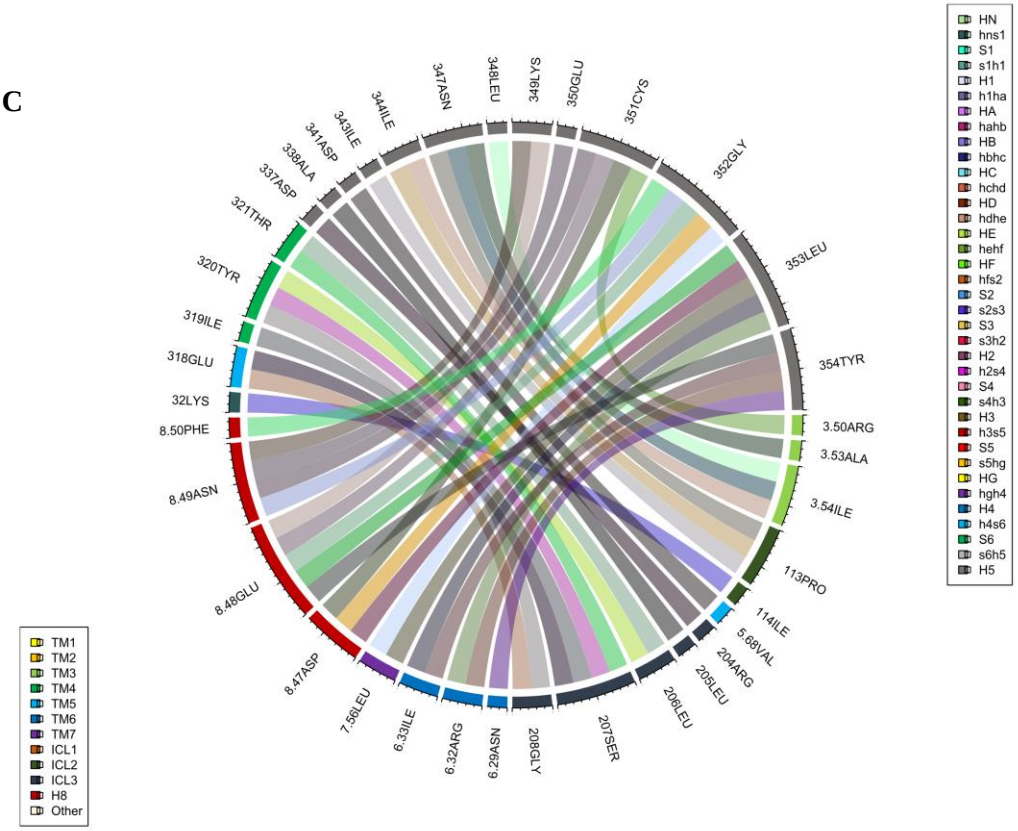
A



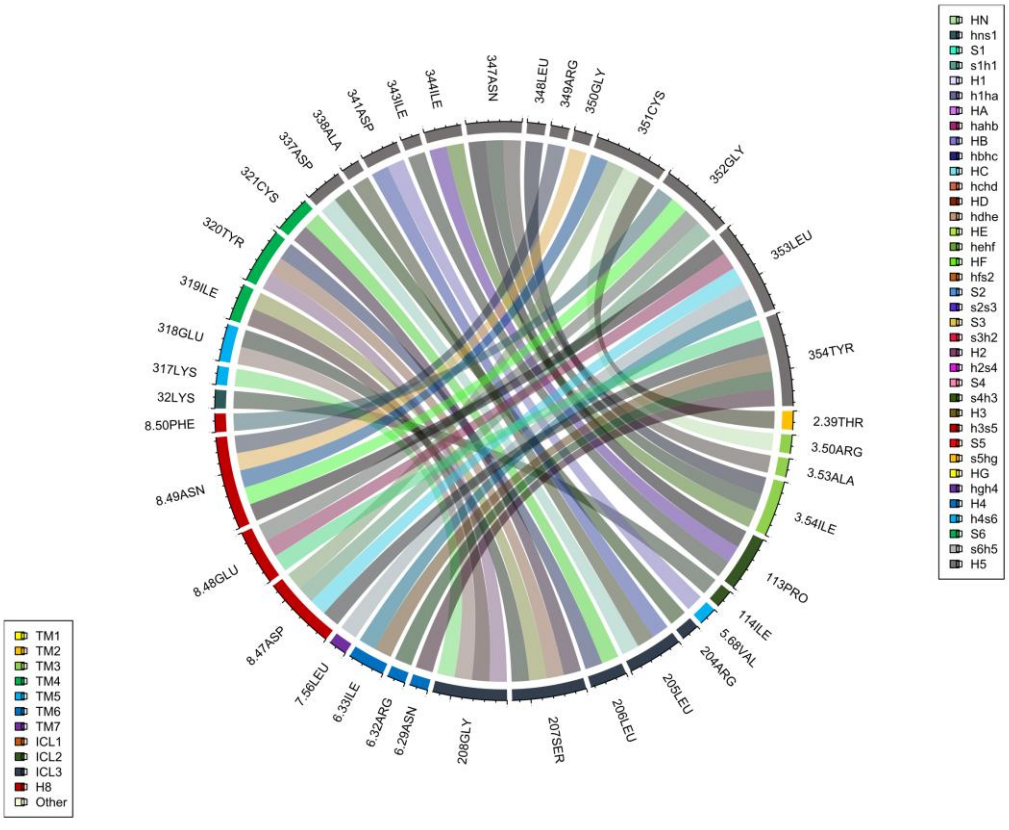
B



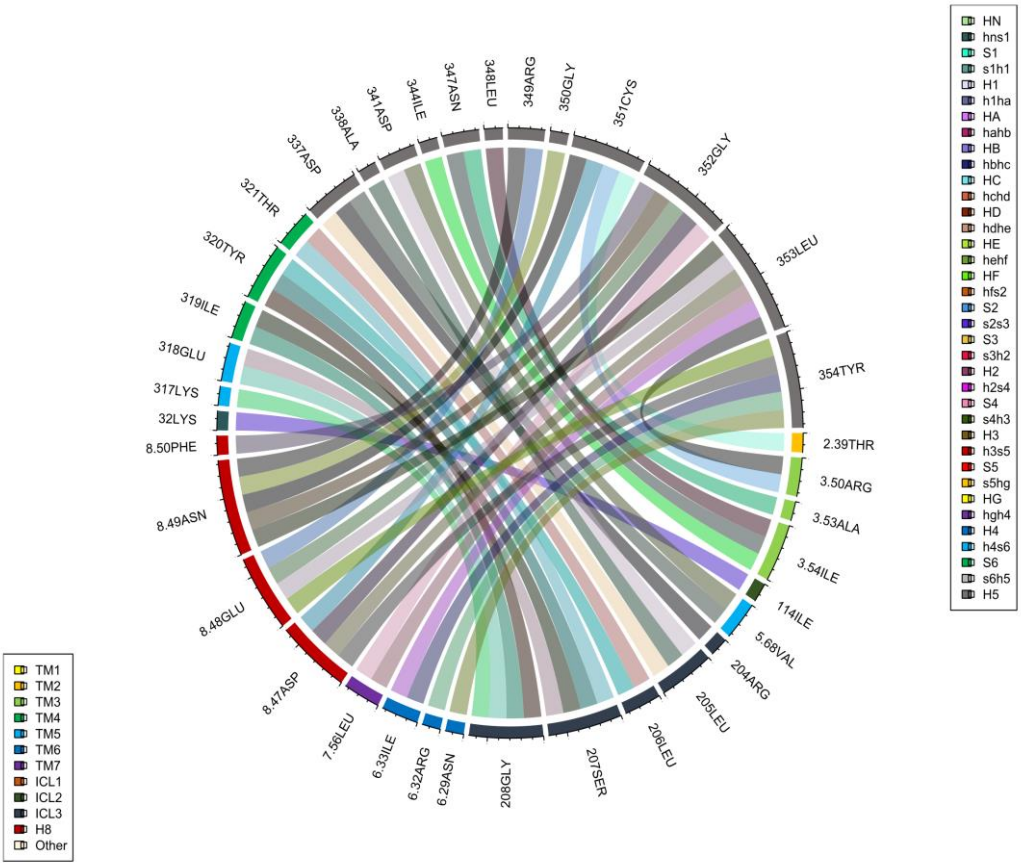
C


HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
347ASN
348LEU
349LYS
350GLU
351CYS
352GLY
353LEU
354TYR
3.50ARG
3.53ALA
3.54ILE
113PRO
114ILE
5.68VAL
204ARG
205LEU
206LEU
207SER
208GLY
6.29ASN
6.32ARG
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
32LYS
318GLU
319ILE
320TYR
321THR
337ASP
338ALA
341ASP
343ILE
344ILE

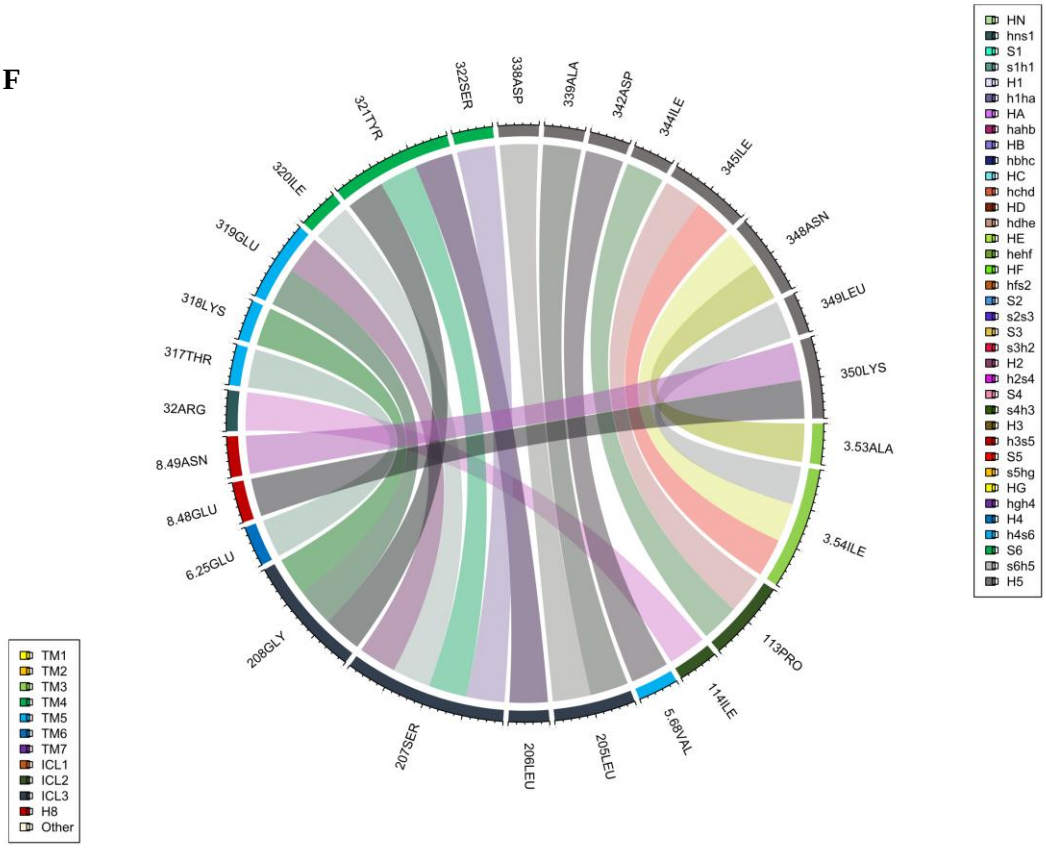
D


HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
347ASN
348LEU
349ARG
350GLY
351CYS
352GLY
353LEU
354TYR
2.39THR
3.50ARG
3.53ALA
3.54ILE
113PRO
114ILE
5.68VAL
204ARG
205LEU
206LEU
207SER
208GLY
6.29ASN
6.32ARG
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
32LYS
317LYS
318GLU
319ILE
320TYR
321CYS
337ASP
338ALA
341ASP
343ILE
344ILE

E

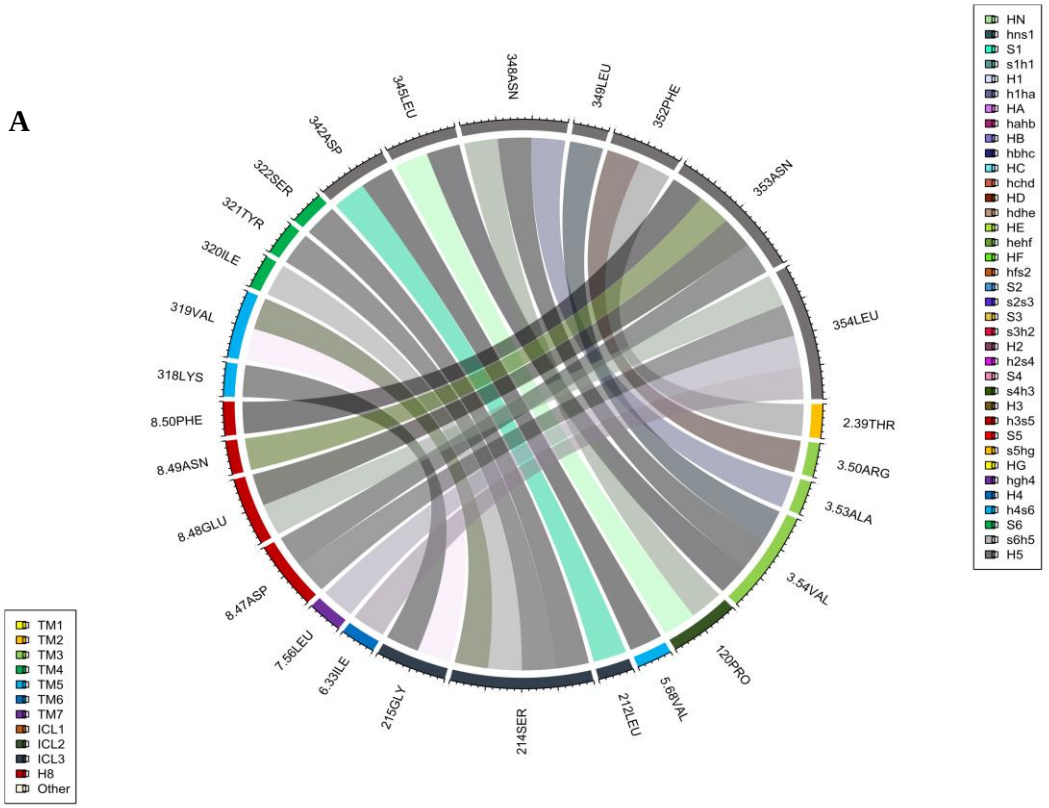

HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
347ASN
348LEU
349ARG
350GLY
351CYS
352GLY
353LEU
354TYR
2.39THR
3.50ARG
3.53ALA
3.54ILE
114ILE
5.68VAL
204ARG
205LEU
206LEU
207SER
208GLY
6.29ASN
6.32ARG
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
32LYS
317LYS
318GLU
319ILE
320TYR
321THR
337ASP
338ALA
341ASP
344ILE

F

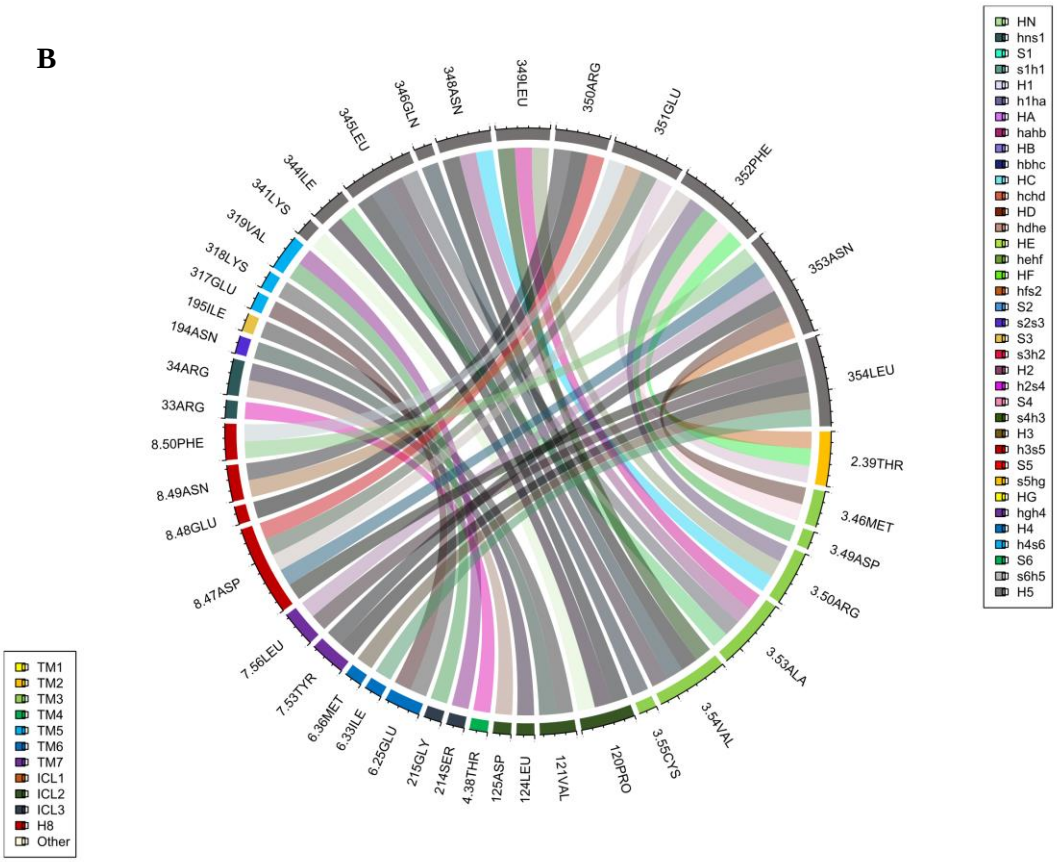

HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
322SER
338ASP
339ALA
342ASP
344ILE
345ILE
348ASN
349LEU
350LYS
3.53ALA
3.54ILE
113PRO
114ILE
5.68VAL
205LEU
206LEU
207SER
208GLY
6.25GLU
8.48GLU
8.49ASN
32ARG
317THR
318LYS
319GLU
320ILE
321TYR

Annex 13: DOR-Gq/11 interaction plots, built using *circlize* package for R language. A) DOR-G14_6CMO B) DOR-G14_6OIJ C) DOR-G15_6CMO D) DOR-G15_6OIJ

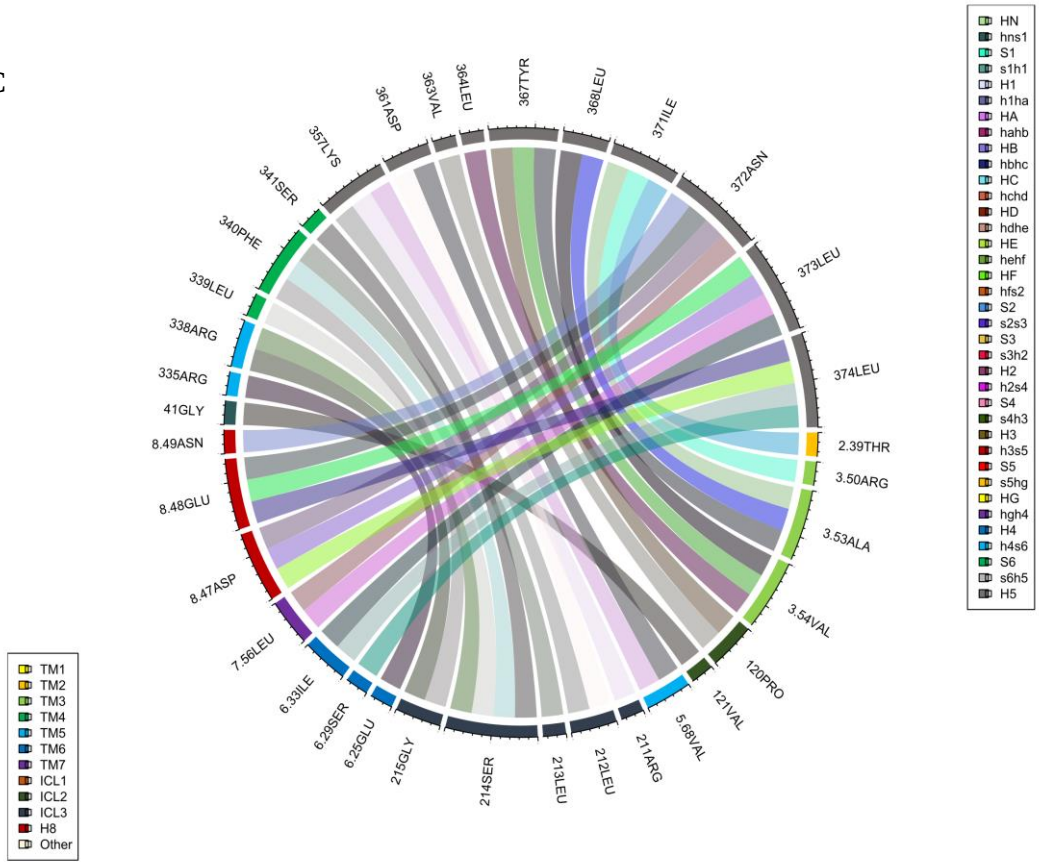
A



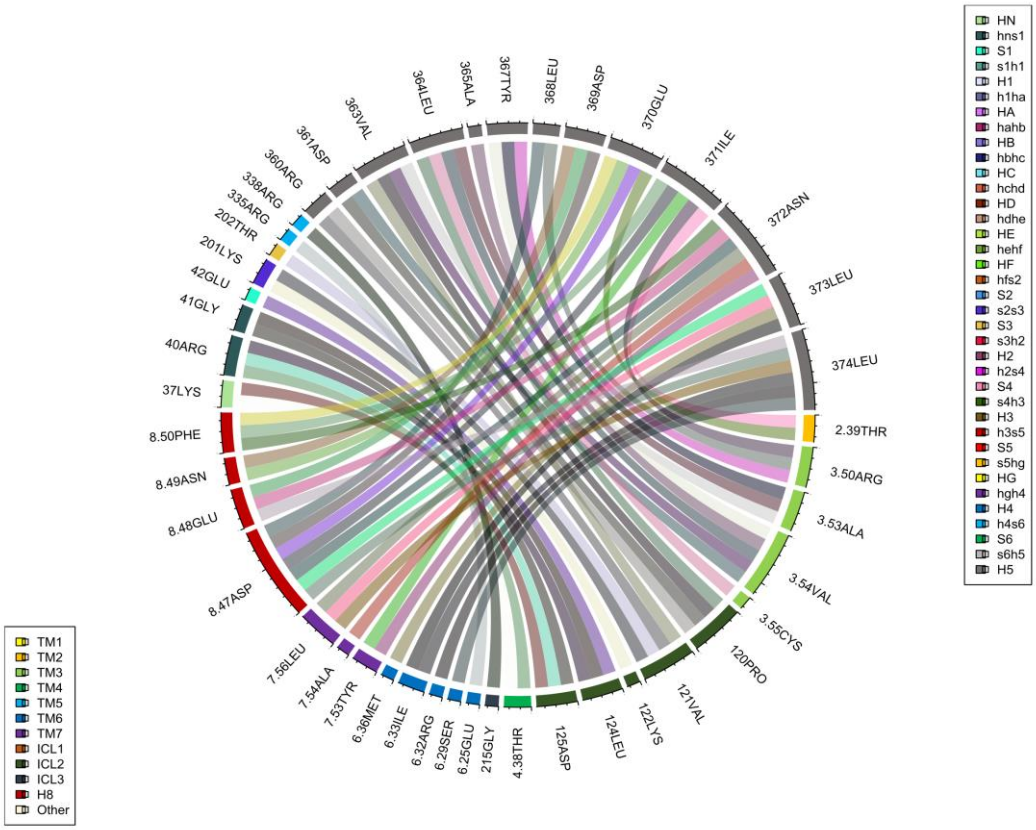
B



C

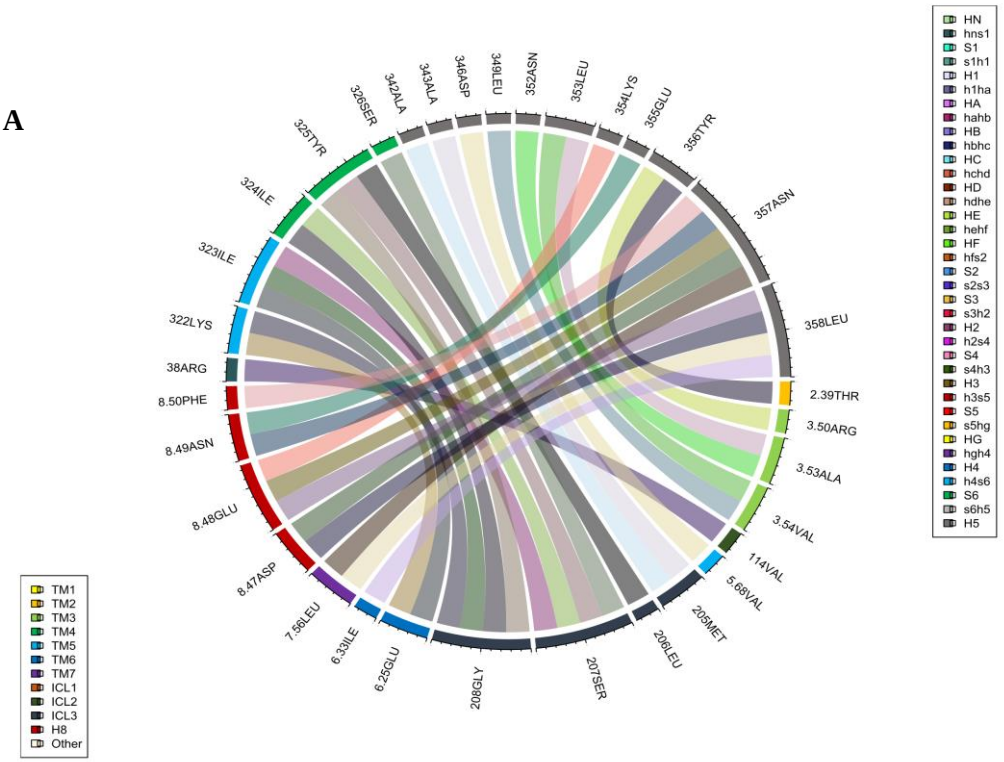

HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
361ASP
363VAL
364LEU
367TYR
368LEU
371ILE
372ASN
373LEU
374LEU
2.39THR
3.50ARG
3.53ALA
3.54VAL
120PRO
121VAL
5.68VAL
211ARG
212LEU
213LEU
214SER
215GLY
6.25GLU
6.29SER
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
41GLY
335ARG
338ARG
339LEU
340PHE
341SER
357LYS

D

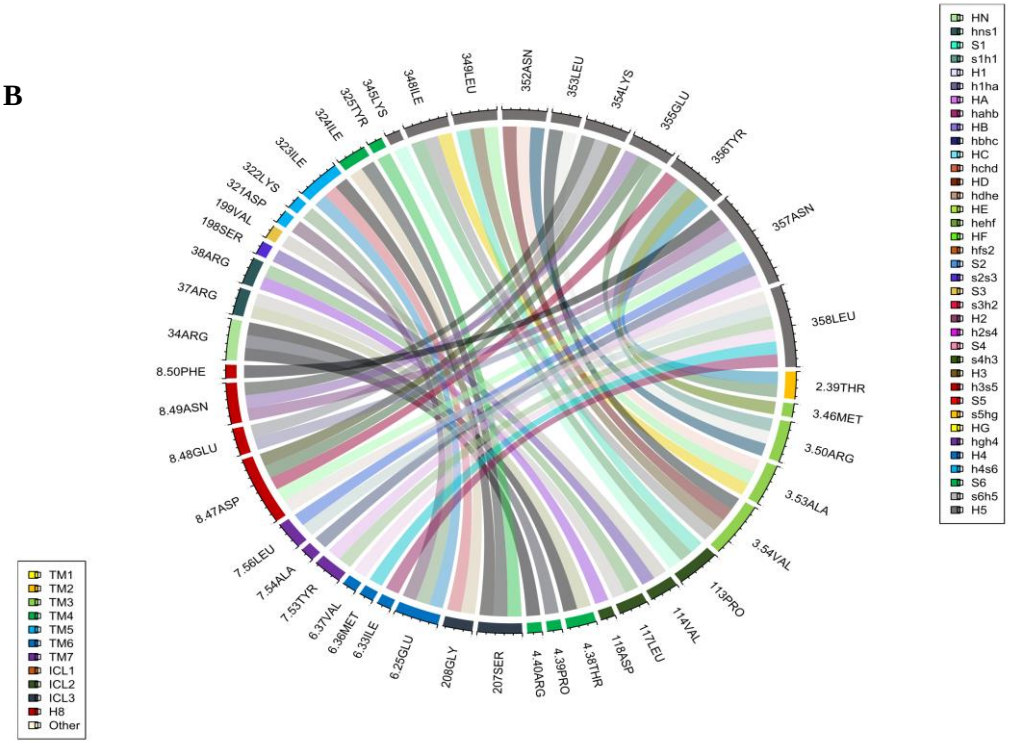

HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
363VAL
364LEU
365ALA
367TYR
368LEU
369ASP
370GLU
371ILE
372ASN
373LEU
374LEU
2.39THR
3.50ARG
3.53ALA
3.54VAL
3.55CYS
120PRO
121VAL
122LYS
124LEU
125ASP
4.38THR
215GLY
6.25GLU
6.29SER
6.32ARG
6.33ILE
6.36MET
7.53TYR
7.54ALA
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
37LYS
40ARG
41GLY
42GLU
201LYS
202THR
335ARG
338ARG
360ARG
361ASP

Annex 14: MOR-Gq/11 interaction plots, built using *circlize* package for R language. A) MOR-Gq_6CMO B) MOR-Gq_6OIJ C) MOR-G11_6CMO D) MOR-G11_6OIJ E) MOR-G15_6CMO F) MOR-G15_6OIJ

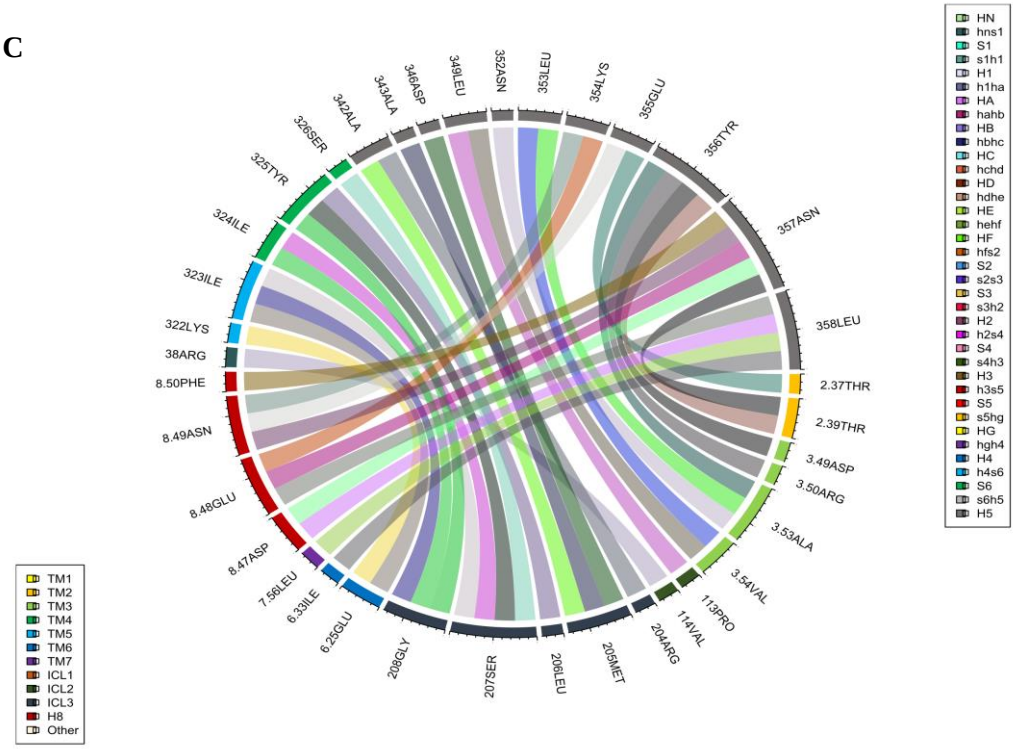
A



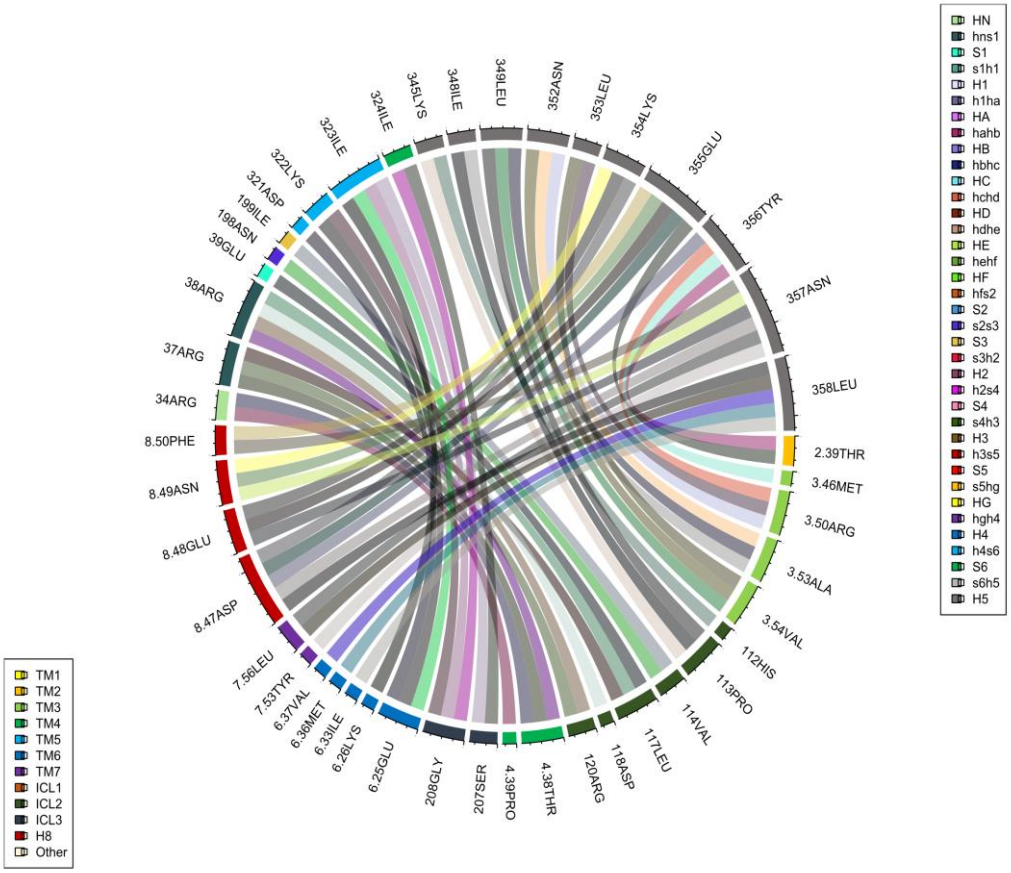
B



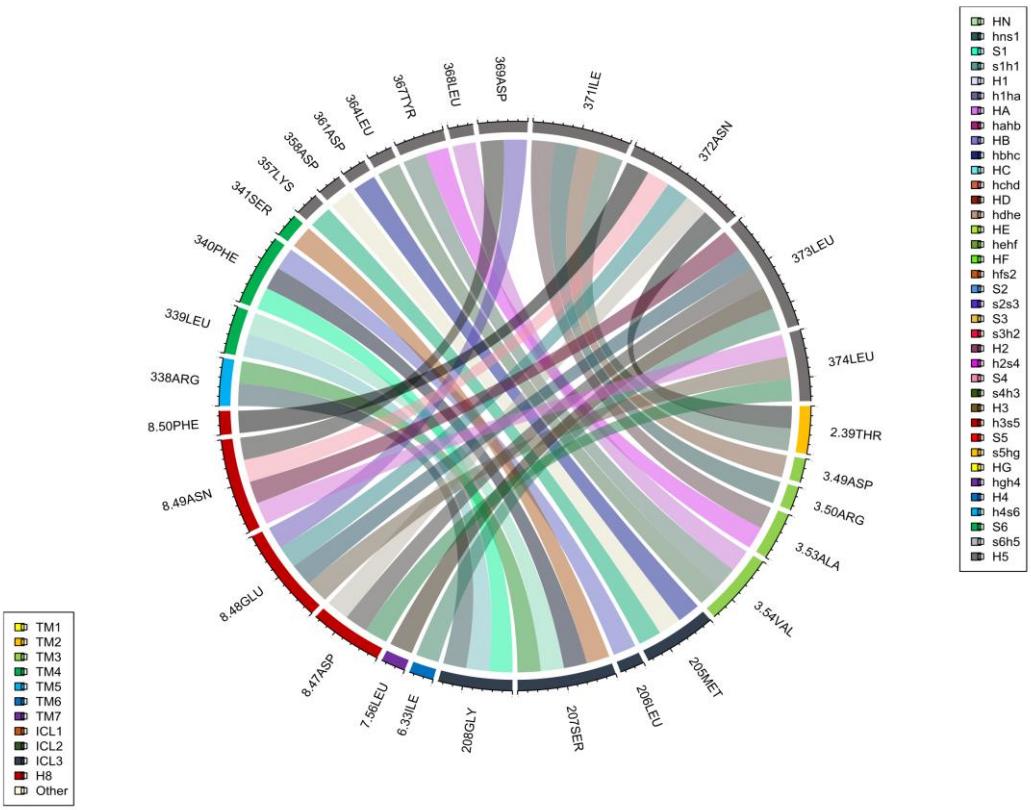
C


HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
352ASN
353LEU
354LYS
355GLU
356TYR
357ASN
358LEU
2.37THR
2.39THR
3.49ASP
3.50ARG
3.53ALA
3.54VAL
113PRO
114VAL
204ARG
205MET
206LEU
207SER
208GLY
6.25GLU
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
38ARG
322LYS
323ILE
324ILE
325TYR
326SER
342ALA
343ALA
346ASP
349LEU

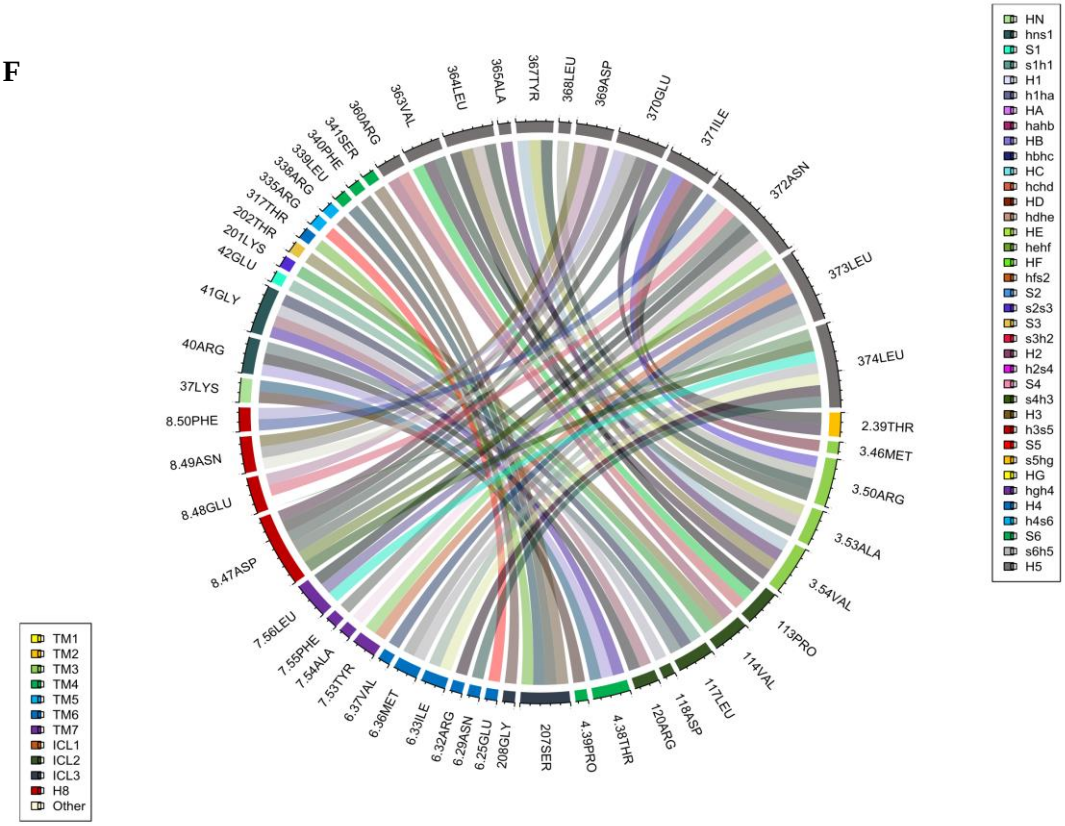
D


HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
352ASN
353LEU
354LYS
355GLU
356TYR
357ASN
358LEU
2.39THR
3.46MET
3.50ARG
3.53ALA
3.54VAL
112HIS
113PRO
114VAL
117LEU
118ASP
120ARG
4.38THR
4.39PRO
207SER
208GLY
6.25GLU
6.26LYS
6.33ILE
6.36MET
6.37VAL
7.53TYR
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
34ARG
37ARG
38ARG
39GLU
198ASN
199ILE
321ASP
322LYS
323ILE
324ILE
345LYS
348ILE
349LEU

E

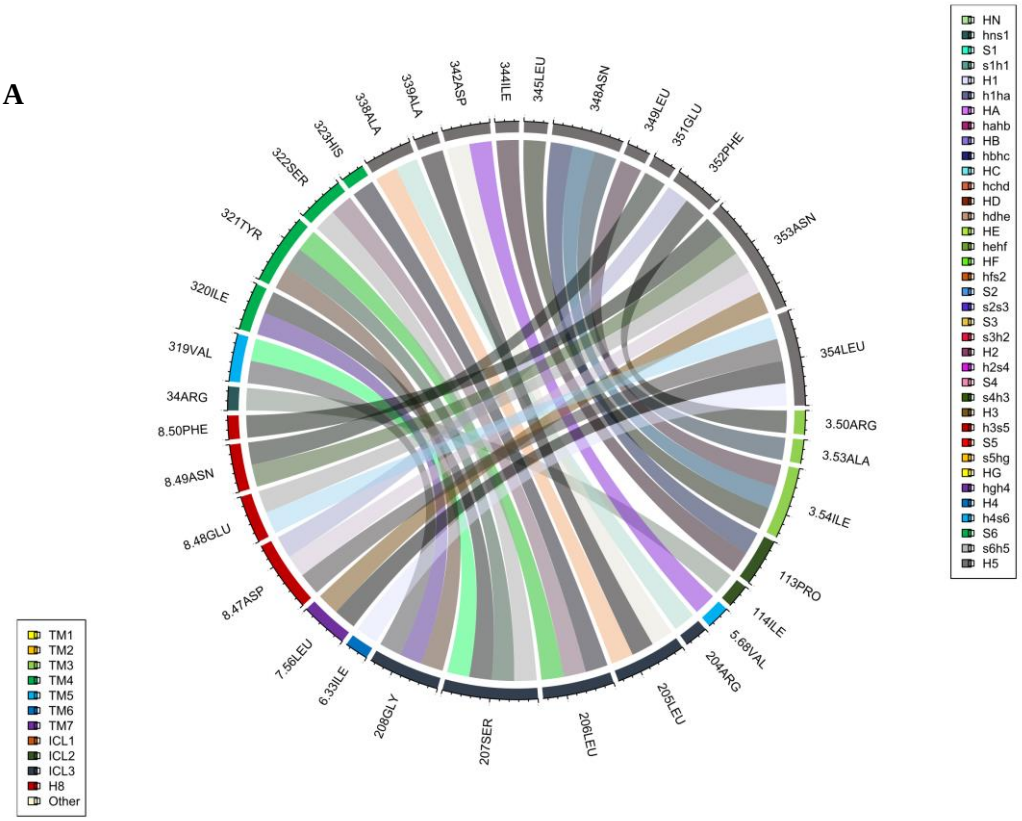

HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
369ASP
371ILE
372ASN
373LEU
374LEU
2.39THR
3.49ASP
3.50ARG
3.53ALA
3.54VAL
205MET
206LEU
207SER
208GLY
6.33ILE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
338ARG
339LEU
340PHE
341SER
357LYS
359ASP
361ASP
364LEU
367TYR
368LEU

F

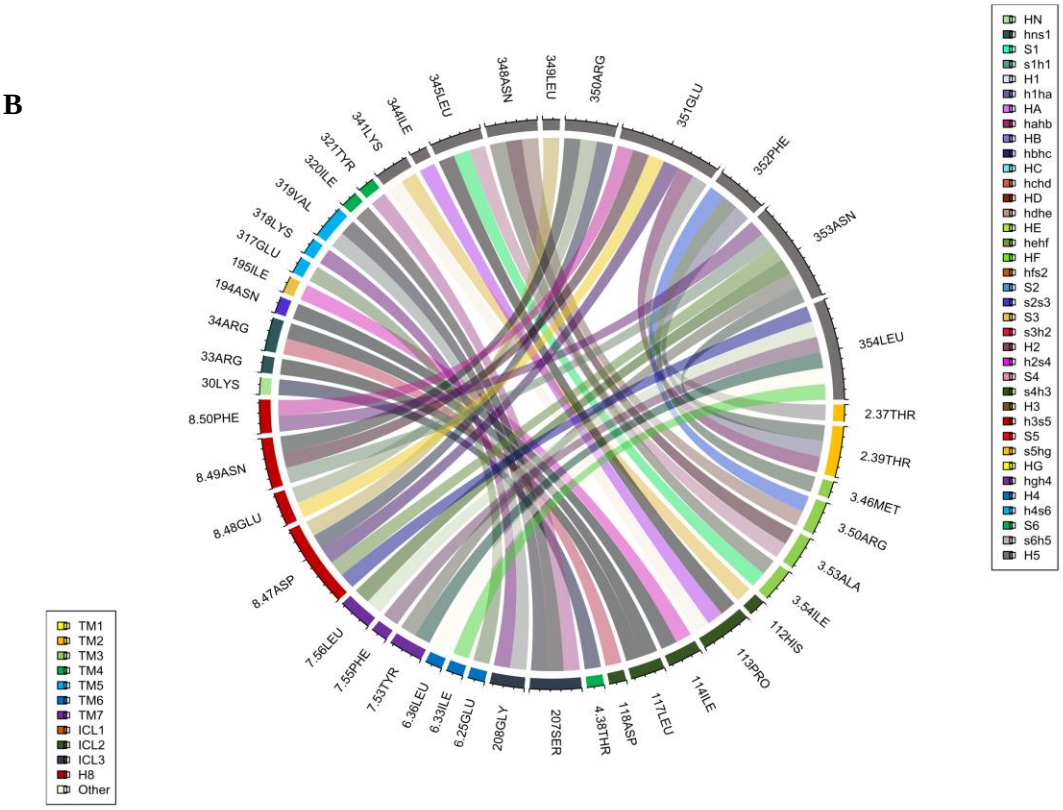

HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
365ALA
367TYR
368LEU
369ASP
370GLU
371ILE
372ASN
373LEU
374LEU
2.39THR
3.46MET
3.50ARG
3.53ALA
3.54VAL
113PRO
114VAL
117LEU
118ASP
120ARG
4.38THR
4.39PRO
207SER
208GLY
6.25GLU
6.29ASN
6.32ARG
6.33ILE
6.36MET
6.37VAL
7.53TYR
7.54ALA
7.55PHE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
8.50PHE
37LYS
40ARG
41GLY
42GLU
201LYS
202THR
317THR
335ARG
338ARG
339LEU
340PHE
341SER
360ARG
363VAL
364LEU

Annex 15: NOP-Gq/11 interaction plots, built using *circlize* package for R language. A) NOP-G14_6CMO B) NOP-G14_6OIJ

A

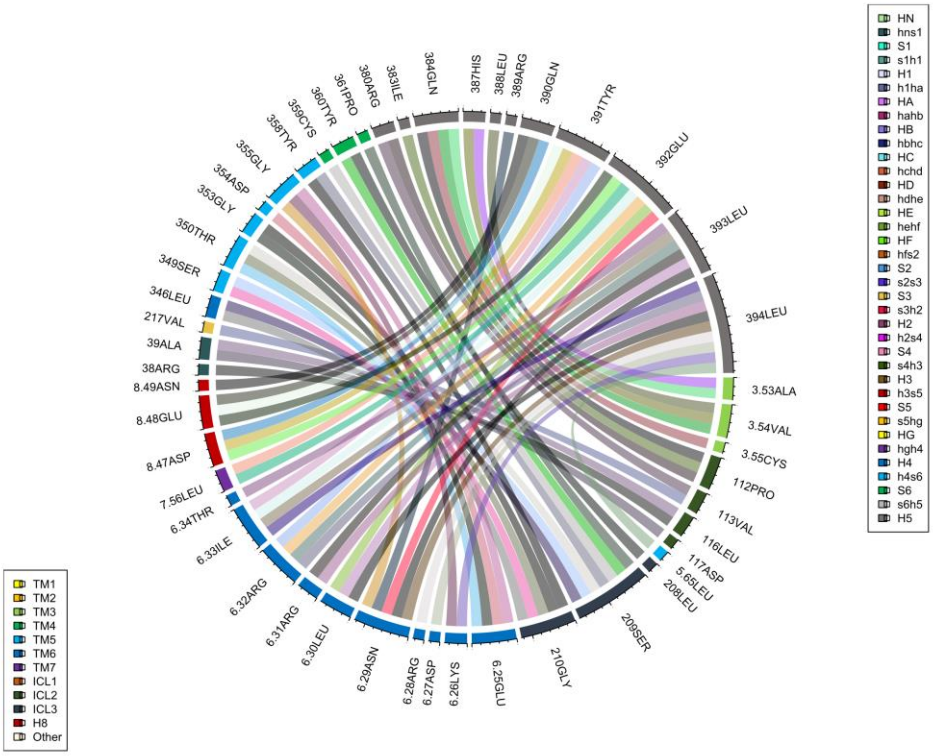


B

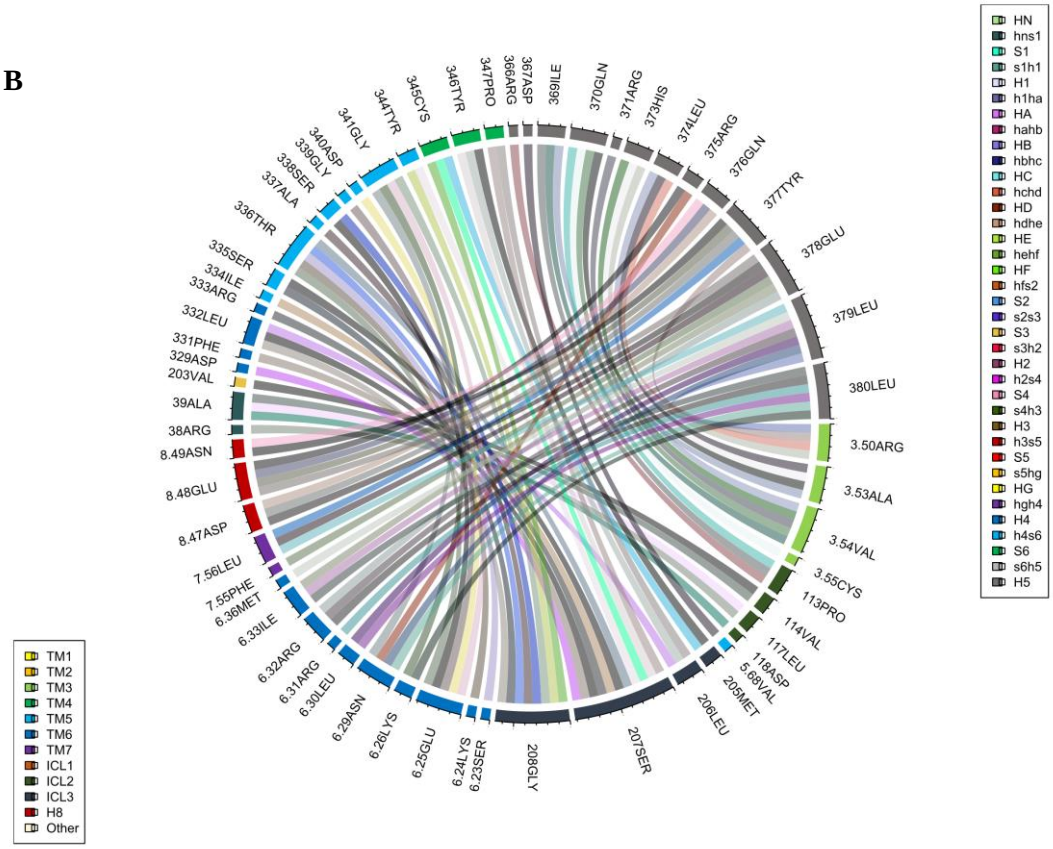


Annex 16: OR-Gs interaction plots, built using *circlize* package for R language. A) KOR-Gslo B) MOR-Gssh C) MOR-Gslo

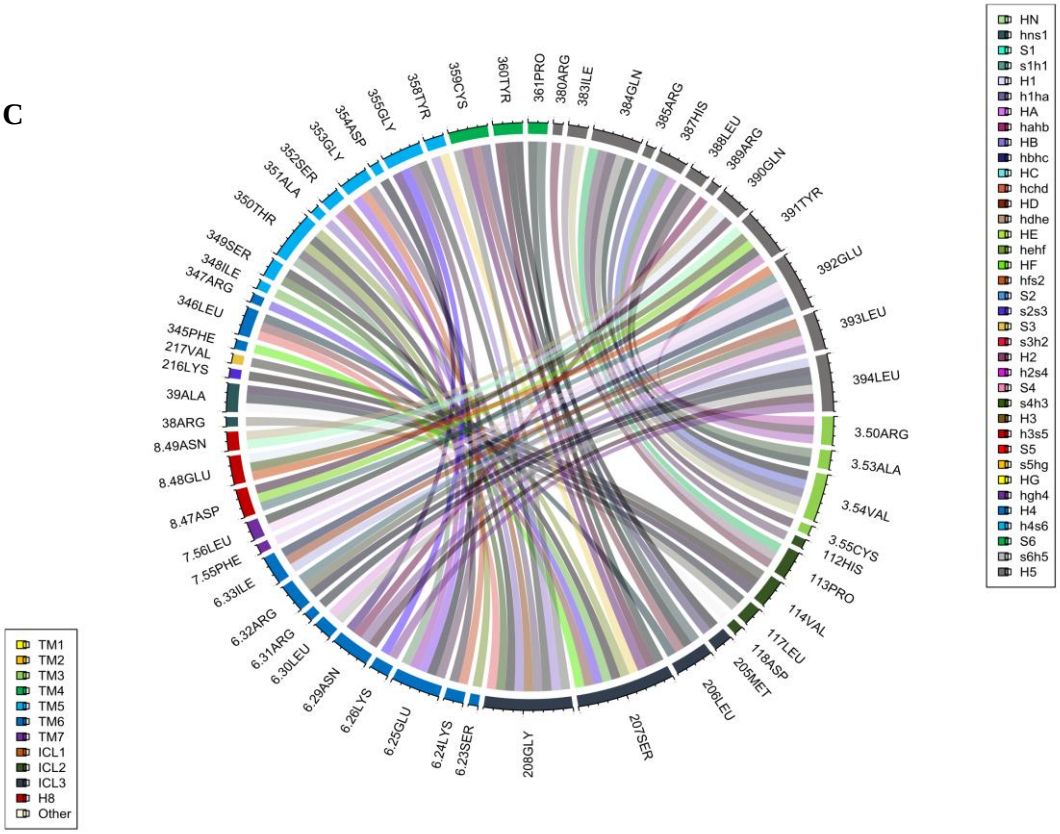
A



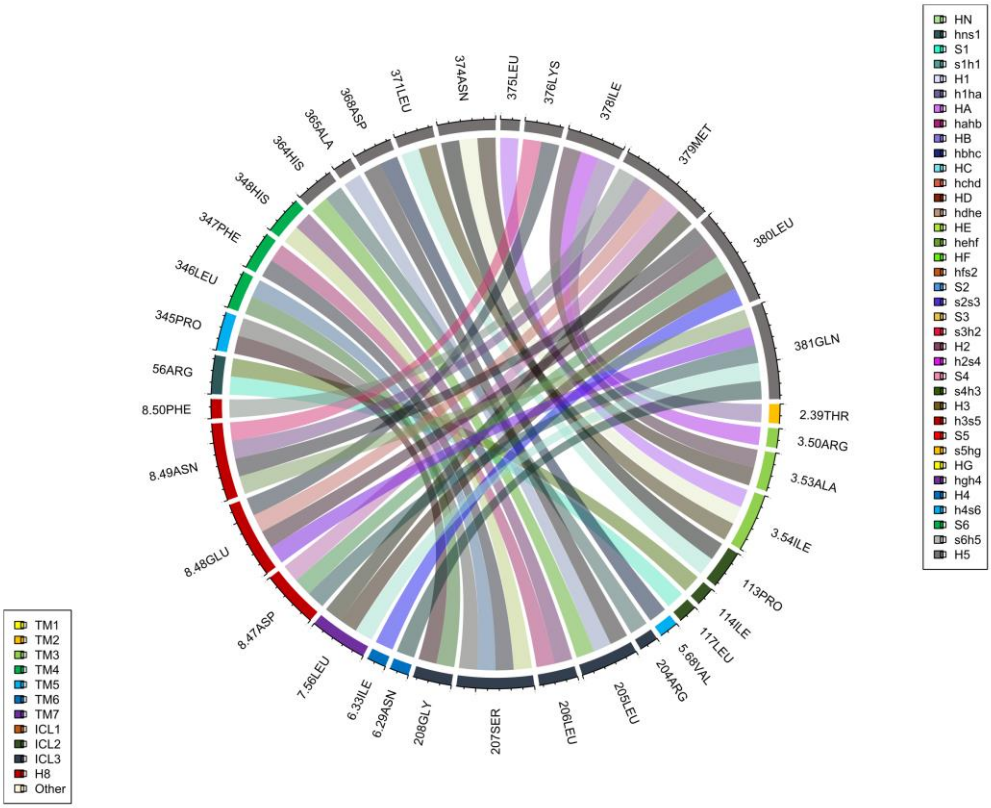
B



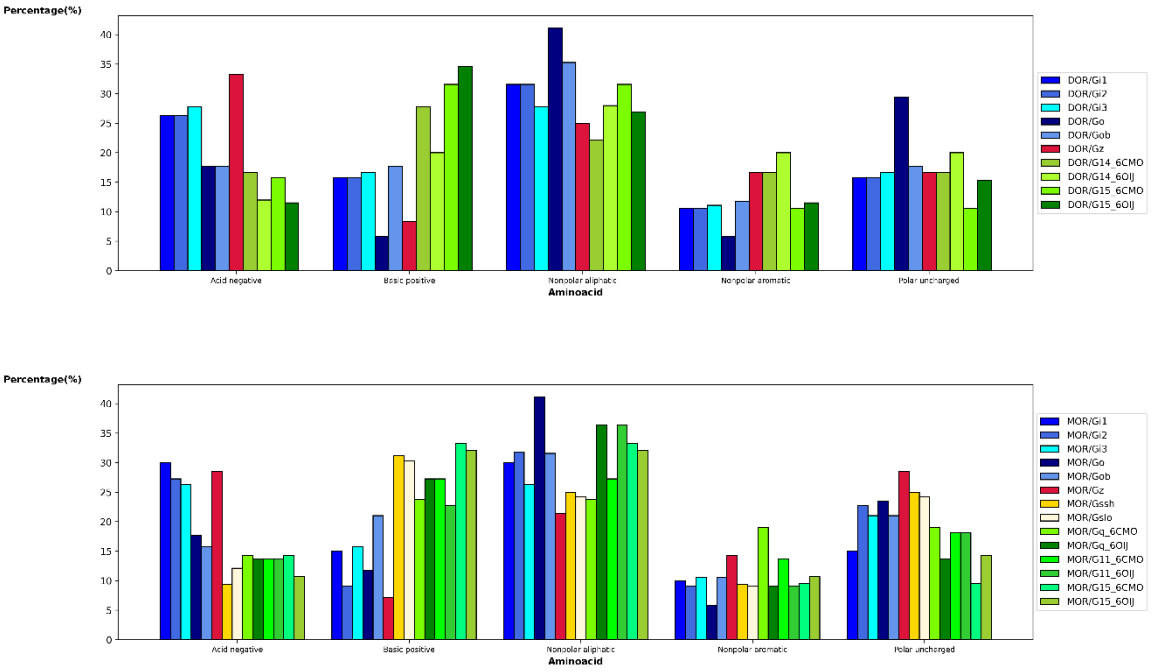
C

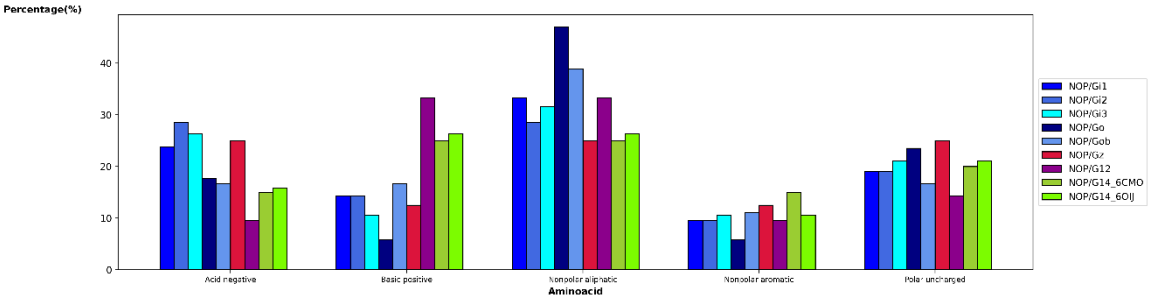

HN
hns1
S1
s1h1
H1
h1ha
HA
hahb
HB
hbhc
HC
hchd
HD
hdhe
HE
hehf
HF
hfs2
S2
s2s3
S3
s3h2
H2
h2s4
S4
s4h3
H3
h3s5
S5
s5hg
HG
hgh4
H4
h4s6
S6
s6h5
H5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
ICL1
ICL2
ICL3
H8
Other
360TYR
361PRO
380ARG
383ILE
384GLN
385ARG
387HIS
388LEU
389ARG
390GLN
391TYR
392GLU
393LEU
394LEU
3.50ARG
3.53ALA
3.54VAL
3.55CYS
112HIS
113PRO
114VAL
117LEU
118ASP
205MET
206LEU
207SER
208GLY
6.23SER
6.24LYS
6.25GLU
6.26LYS
6.29ASN
6.30LEU
6.31ARG
6.32ARG
6.33ILE
7.55PHE
7.56LEU
8.47ASP
8.48GLU
8.49ASN
38ARG
39ALA
216LYS
217VAL
345PHE
346LEU
347ARG
348ILE
349SER
350THR
351ALA
352SER
353GLY
354ASP
355GLY
358TYR
359CYS

Annex 17: NOP-G12 interaction plot, built using *circlize* package for R language.

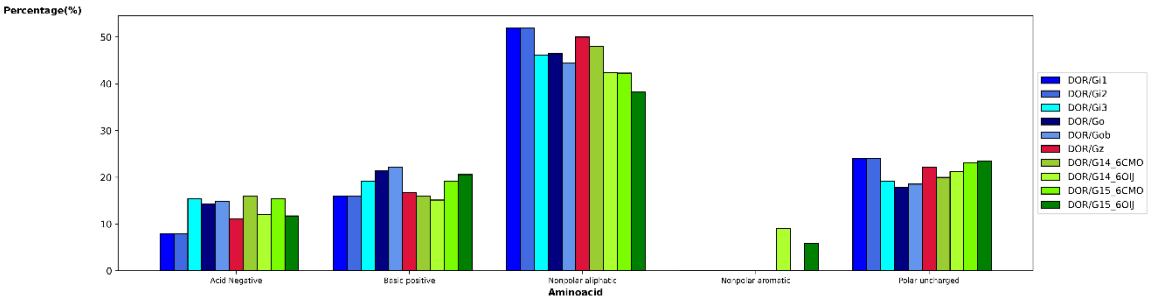


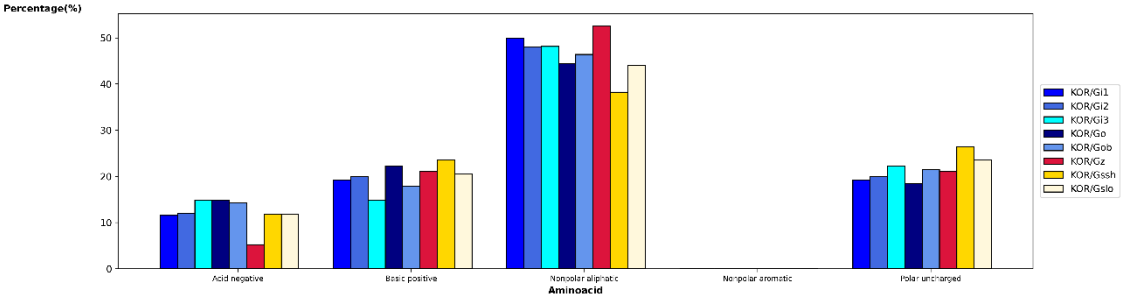
Annex 18: Residue group interaction percentage values to all aminoacid groups (in G-protein side)

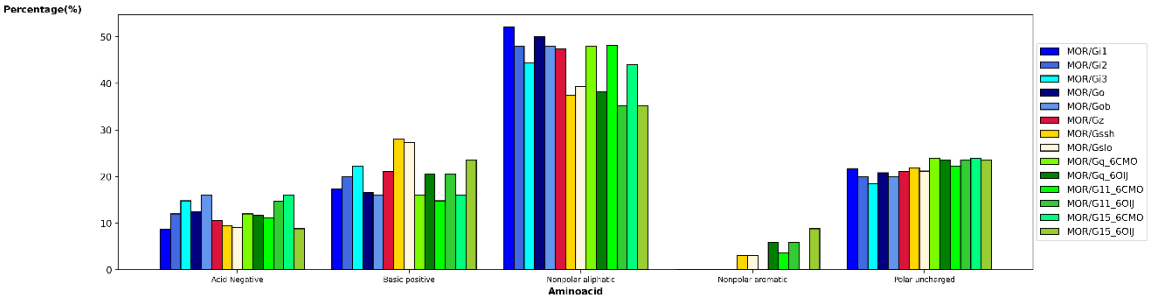


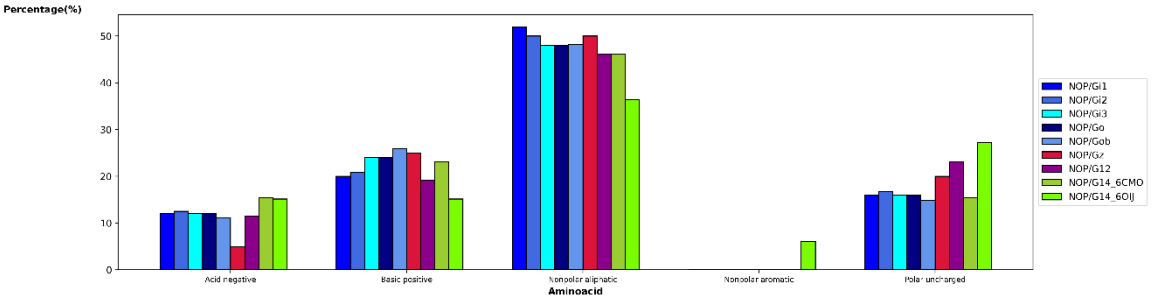


Annex 19: Residue group interaction percentage values to all aminoacid groups (in OR side)

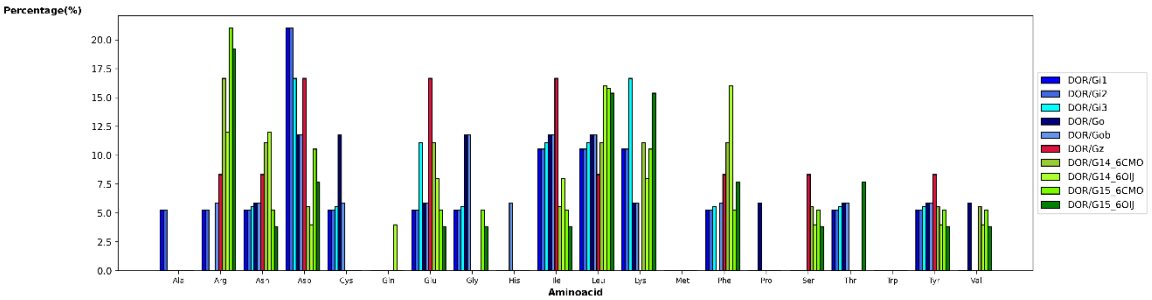


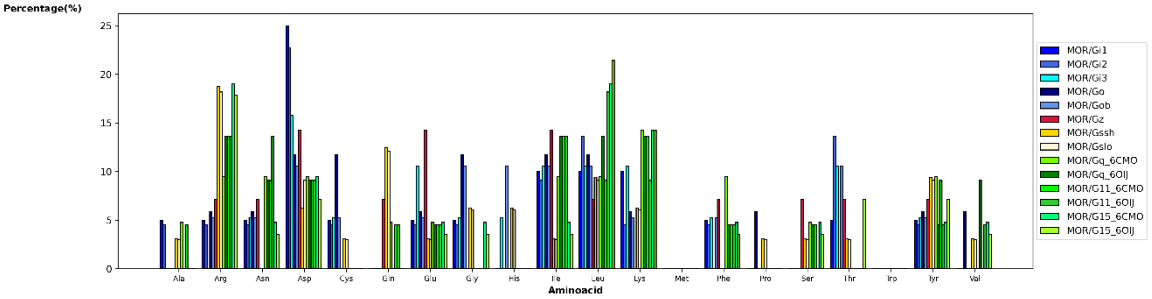


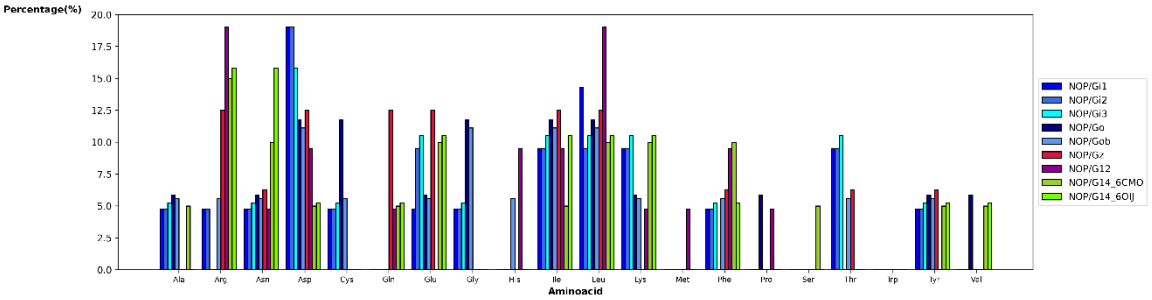




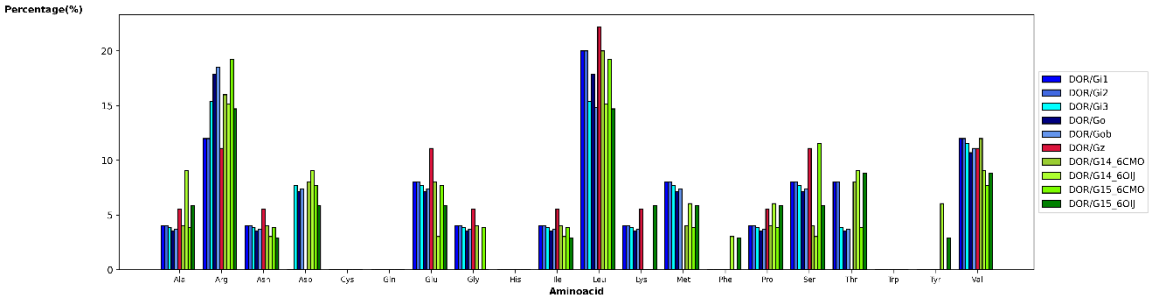
Annex 20: Residue type interaction percentage values to all aminoacid groups (in partner side)

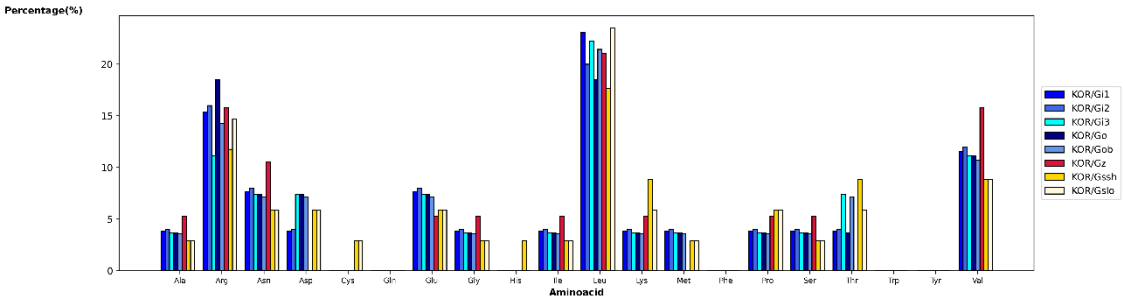


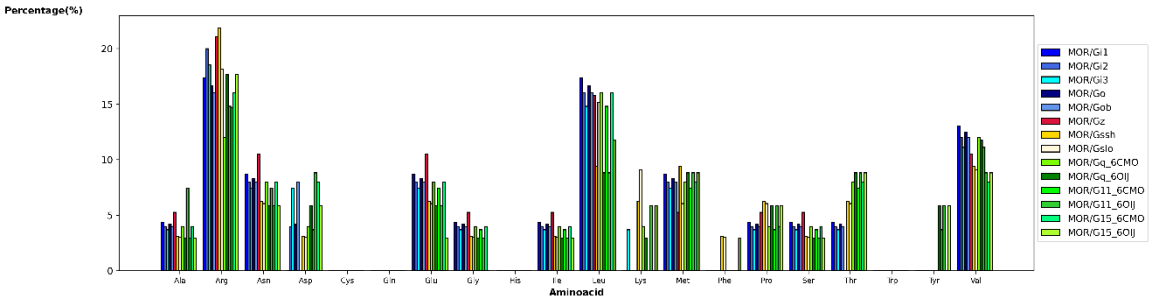


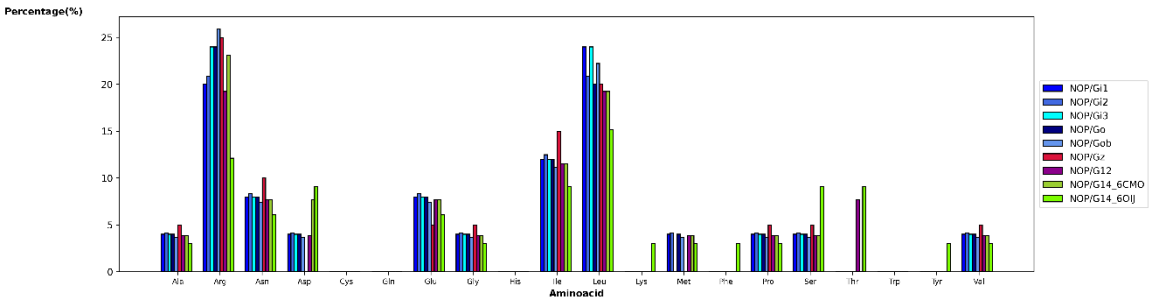


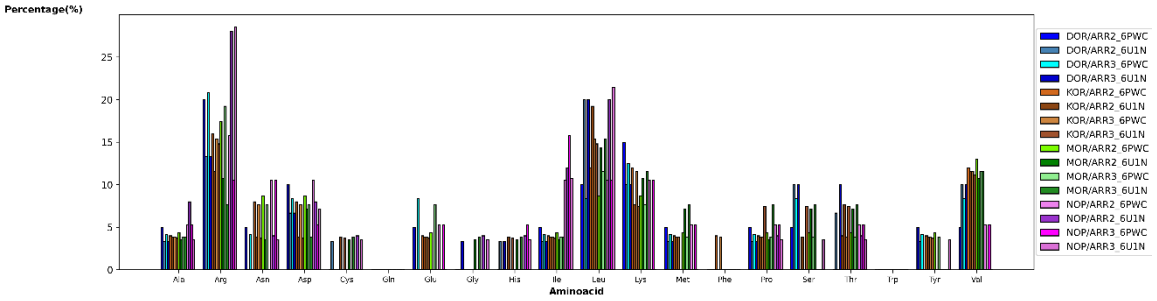
Annex 21: Residue type interaction percentage values to all aminoacid groups (in OR side)



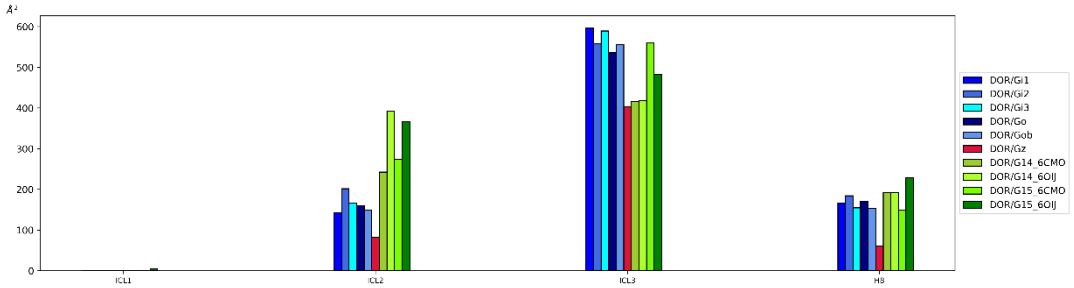


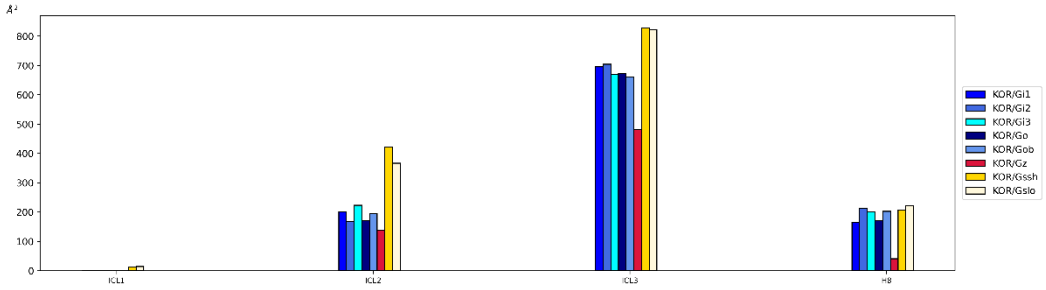


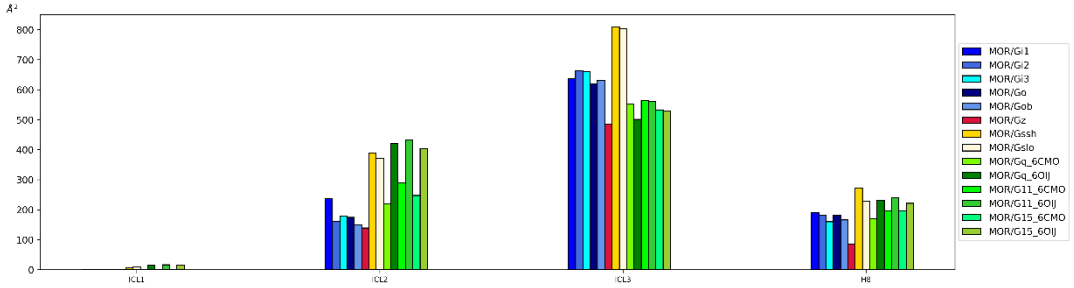


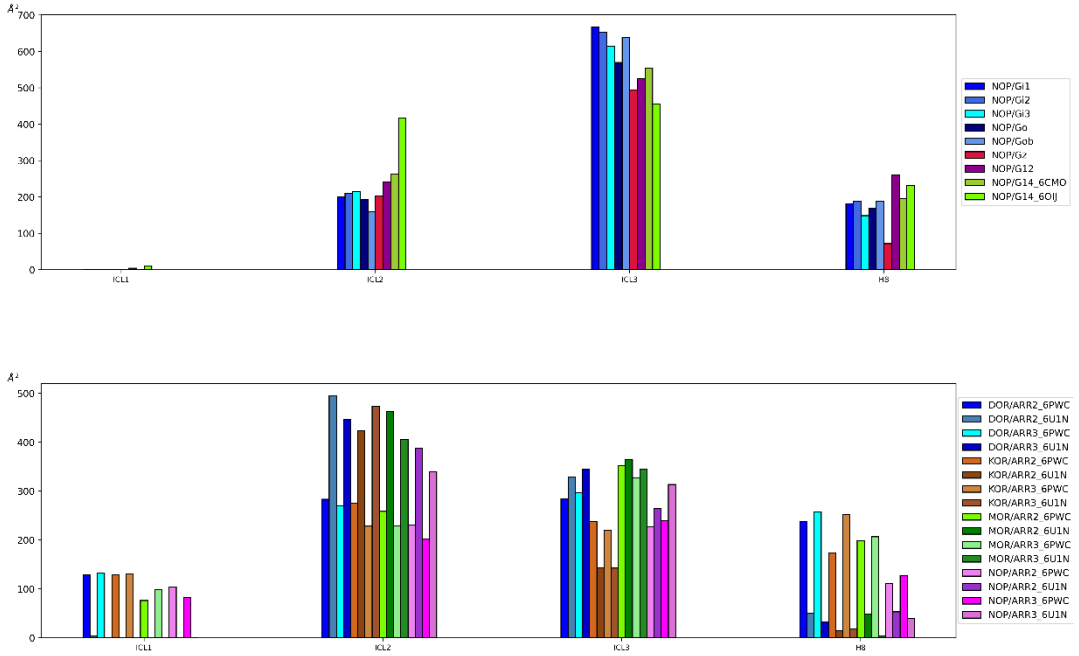


Annex 22: SASA values for all OR-Partner complexes, determined by COCOMAPS webserver

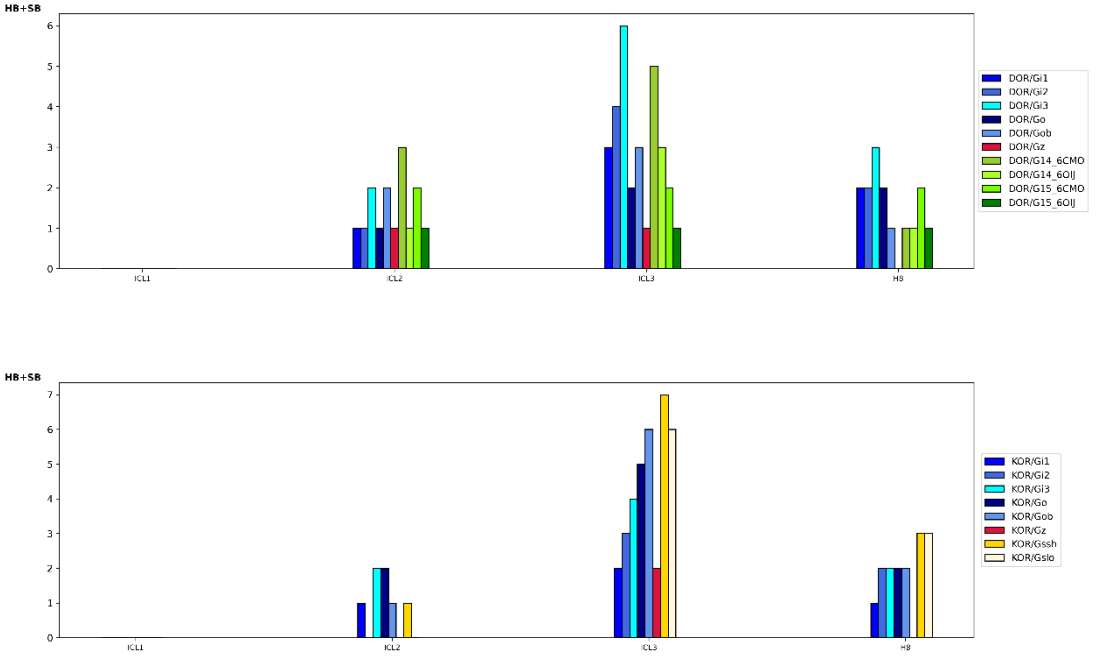


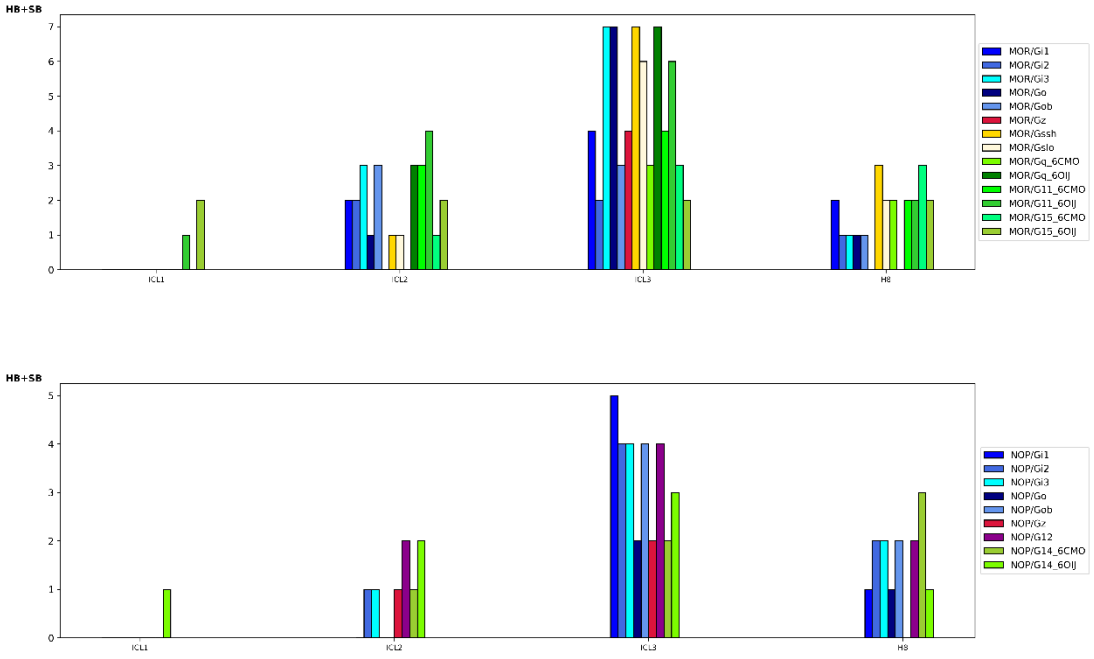




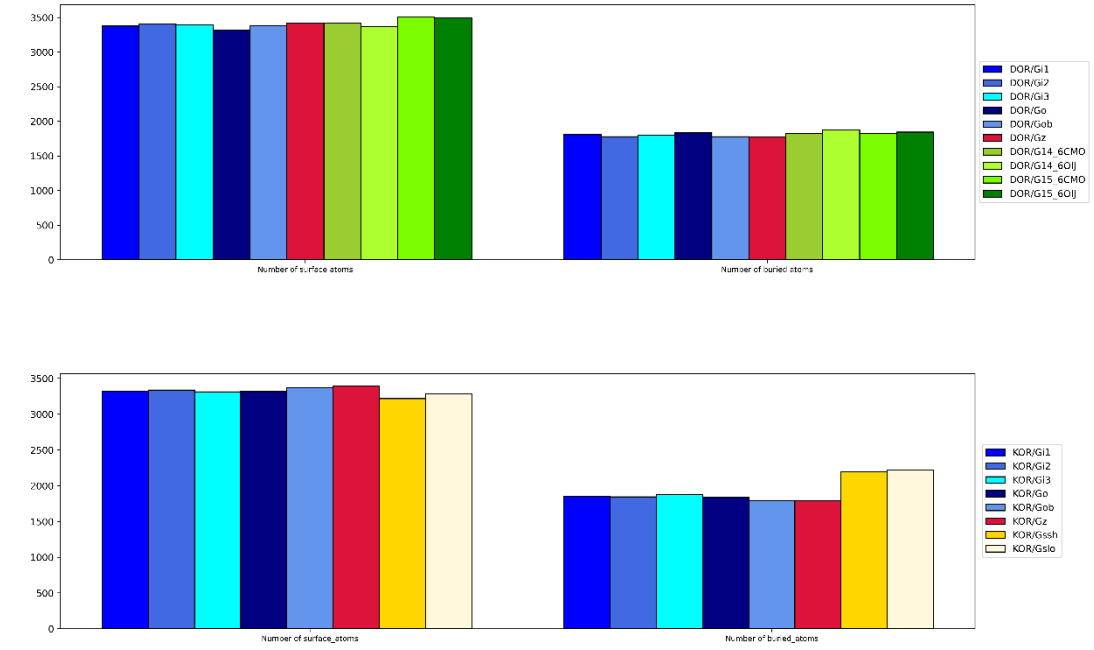


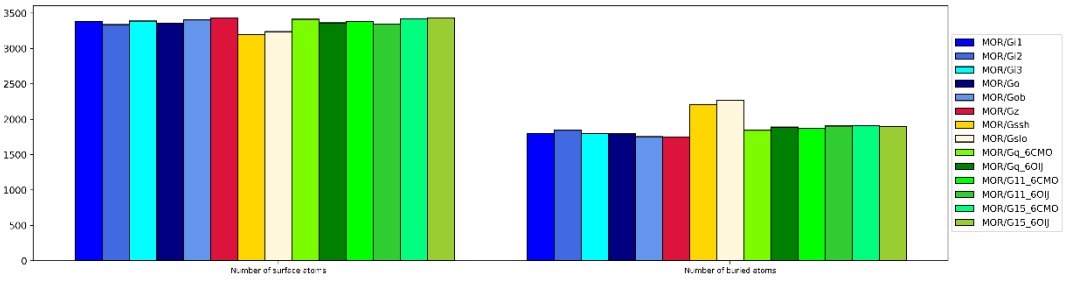
Annex 23: Number of HB and SB for OR-G-proteins complexes

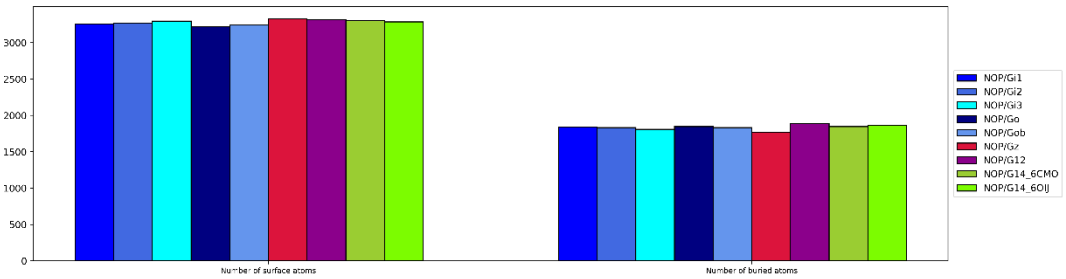




Annex 24: Number of surface and buried atoms for OR-G-proteins complexes




3500
3000
2500
2000
1500
1000
500
0
Number of surface atoms
Number of buried atoms
MOR/Gi1
MOR/Gi2
MOR/Gi3
MOR/Go
MOR/Gob
MOR/Gz
MOR/Gssh
MOR/Gslo
MOR/Gq_6CMO
MOR/Gq_6OIj
MOR/G11_6CMO
MOR/G11_6OIj
MOR/G15_6CMO
MOR/G15_6OIj


3000
2500
2000
1500
1000
500
0
Number of surface atoms
Number of buried atoms
NOP/Gi1
NOP/Gi2
NOP/Gi3
NOP/Go
NOP/Gob
NOP/Gz
NOP/G12
NOP/G14_6CMO
NOP/G14_6OIj