UNIVERSIDADE D
COIMBRA

Raquel Pina Gouveia

The Structural and Functional Role of CACNG2 Mutations in Psychiatric Disorders

Dissertação no âmbito do Mestrado em Biologia Celular e Molecular, orientada pela Professora Doutora Irina Sousa Moreira e pela Professora Doutora Ana Luísa Carvalho e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

October 2020

Raquel Pina Gouveia

The Structural and Functional Role of CACNG2 Mutations in Psychiatric Disorders

Dissertação no âmbito do Mestrado em Biologia Celular e Molecular, orientada pela Professora Doutora Irina Sousa Moreira e pela Professora Doutora Ana Luísa Carvalho e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

October 2020

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D COIMBRA

Este trabalho foi realizado no CNC – Centro de Neurociência e Biologia Celular, Universidade de Coimbra e financiado por fundos nacionais através da Fundação para a Ciência e Tecnologia (FCT) e da Direção Geral do Ensino Superior (DGES) através da bolsa de investigação para o Curso de Verão em "Metodologias de Investigação Científica" - Módulo de I&D "Metodologias Avançadas para o Estudo do Cérebro", dentro do programa "Verão com Ciência".

This work was performed at CNC – Center for Neuroscience and Cell Biology, University of Coimbra and funded by national funds from the Portuguese Foundation for Science and Technology (FCT) and the Directorate General for Higher Education (DGES) through through fellowship for Curso de Verão in "Metodologias de Investigação Científica" - Módulo de I&D "Metodologias Avançadas para o Estudo do Cérebro", under the program "Verão com Ciência".

Agradecimentos

Em primeiro lugar quero agradecer à minha orientadora professora Dra. Irina Moreira por me ter proporcionado as condições necessárias à realização deste projecto. Quero também agradecer todas as oportunidades que me proporcionou desde o momento que me aceitou na rotação laboratorial.

Quero também agradecer à minha co-orientadora Ana Luísa Carvalho pelas ideias que proporcionou na realização deste projeto, mesmo não tendo conseguido focar-me em todas, apreciei tudo o que consegui aprender durante o processo.

De seguida quero agradecer a todos os meus colegas do laboratório por todo o encorajamento que me deram. Ao Carlos Barreto agradeço também a sua paciência enquanto me ensinava e pelo apoio todo que me deu durante todo o tempo que estive a trabalhar neste projeto. Ao António Gomes, Rita Melo e Salete Batista quero agradecer pela ajuda que me deram durante várias etapas deste projeto, mesmo quando já estavam bastante ocupados com outros projetos.

Quero deixar um agradecimento especial à minha mãe que mesmo não podendo estar presente para me ver concluir esta tese, foi uma grande inspiração para mim.

Por último quero agradecer ao resto minha família e à Rafaela Bettencourt pelo apoio constante que me deram durante todo o meu percurso.

A todos, muito obrigada!

Table of Contents

Abstract

Neuropsychiatric disorders incidence has increased worldwide and became an increasing economic and social burden. Intellectual disability and Schizophrenia, key examples of such disorders, were already associated with proteins related to homeostatic plasticity. Even though the number of proteins associated with these conditions is massive, the mechanism that leads to the disease is still poorly understood.

Interestingly, various studies showed that the same gene may be found mutated in patients diagnosed with different neuropsychiatric disorders. This may point to a common origin related to synaptic dysfunction. The CACNG2 gene, which codes for stargazin, was indeed described as a susceptible gene for psychiatric disorders. STG, an auxiliary subunit for AMPA receptors, plays an important role in modulating AMPAR function, transporting it to the synapse and assisting in the homeostatic synaptic scaling of AMPAR. There are already two key mutations, one reported in literature and another found on CNC lab "Synapse Biology" linked to intellectual disability and schizophrenia.

During my master thesis, *in silico* techniques such as modeling, dynamic simulation and structural and dynamical analysis were applied to these molecular systems to better understand the effects of these mutations on the structure of stargazin and in its interaction with AMPAR.

Keywords: Intellectual disability; Schizophrenia; Stargazin; AMPAR; Molecular dynamics.

Resumo

A incidência a nível global de doenças psiquiátricas foi aumentando, o que as torna num crescente fardo económico e social. O défice intelectual e a esquizofrenia, típicos exemplos desse tipo de doenças, têm sido associados a proteínas importantes para a modulação da força sináptica. Mesmo com o grande número de proteínas associadas a estas condições, o mecanismo que despoleta estas doenças é ainda pouco conhecido.

Curiosamente, vários estudos mostram que o mesmo gene pode ser encontrado mutado em pacientes diagnosticados com diferentes doenças psicológicas. Isto pode apontar para uma origem comum relacionada com disfunção neuronal. O gene CACNG2, que codifica a stargazin, foi de facto descrito como sendo o gene susceptível para doenças psiquiátricas. A stargazin, uma subunidade auxiliar dos receptores AMPA, tem o importante papel de modular a função do AMPAR, transportá-lo para a sinapse e assistir no escalamento homeostático do AMPAR. Duas mutações importantes foram encontradas nesta proteína, uma relatada na literatura e outra encontrada no laboratório "Synapse Biology" do CNC relacionadas com défice intelectual e com esquizofrenia.

Durante a minha dissertação de mestrado, técnicas *in sillico* como a modelação, simulações dinâmicas e análises estrutural e dinâmica foram aplicadas a estes sistemas moleculares para compreender os efeitos que estas mutações têm na estrutura da stargazin e na sua interacção com o AMPAR.

Palavras-Chave: Défice intelectual; Esquizofrenia; Stargazina; AMPAR; Dinâmica molecular.

List of Abbreviations

3D	three-dimensional
AAMD	all-atom or atomistic molecular dynamics approach
AMBER	assisted model building with energy refinement
AMP	adenosine monophosphate
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
AMPAR	AMPA receptor
ARC	activity-regulated cytoskeleton-associated protein
CACNG2	calcium voltage-gated channel auxiliary subunit gamma2
CaMKII	Ca2+/calmodulin-dependent protein kinase II
CGMD	coarse-grained molecular dynamics approach
CHARMM	chemistry at harvard molecular mechanics
CNS	central nervous system
CTB	C-terminal domain
DNA	deoxyribonucleic acid
DOPE	discrete optimized protein energy
ERBB2	receptor tyrosine-protein kinase erbB-2
GPCR	G protein-coupled receptor
GROMACS	groningen machine for chemical simulation
ID	intellectual disability
iGluR	ionotropic glutamate receptor
IP3	inositol trisphosphate
KAR	kainite receptors
LBD	ligand-binding domain
LTD	long-term depression

LTP	long-term potentiation
MAGI-2	membrane-associated guanylate kinase inverted 2
MD	molecular dynamics
mGluR	metabotropic glutamate receptor
MP	membrane protein
mRNA	messenger RNA
MSA	multiple sequence alignment
MUPP1	multi-PDZ domain protein 1
NAMD	nanoscale molecular dynamics
NMDA	N-methyl-D-aspartate
NMDAR	NMDA receptor
NTD	N-terminal domain
OMP-25	outer membrane protein 25
PDB	protein data bank
PBC	periodic boundary conditions
PKC	protein kinase C
PSD93	postsynaptic density protein 93
PSD95	postsynaptic density protein 95
PSI-BLAST	position specific iterated BLAST
RMSD	root mean square deviation
RMSF	root mean square fluctuation
RNA	ribonucleic acid
SAP-102	synapse-associated protein 102
SAP-97	synapse-associated protein 97
SCZ	schizophrenia
SNP	single nucleotide polymorphisms

STG	stargazin
TARP	transmembrane AMPAR regulatory proteins
TMD	transmembrane domain
VMD	visual-molecular-dynamics

List of Figures

List of Tables

Introduction

1 Psychiatric disorders

Psychiatric disorders are characterized by cognitive, emotional or behavior alterations that are linked to a dysfunction in the psychological, biological, or developmental processes underlying mental functioning[1]. The incidence of psychiatric disorders has been increasing worldwide, and as such, their economic and social burden has been also increasing significantly for most countries and is expected to continue rising[2].

As these disorders have neurological symptoms that are often similar between different disorders, and are heterogeneous between patients with the same diagnosis, the mechanisms that are responsible for them are still poorly understood[3]. Recent studies showed that psychiatric disorders have a genetic component, usually a combination of many genetic variants with a small effect and rarely other variants that have larger effects[4,5]. Multiple studies suggest that abnormalities in genes encoding synaptic proteins may participate in neuropsychiatric disorders by disrupting the excitatory/inhibitory balance in the central nervous system[4,6].

1.1 Intellectual Disability

Intellectual Disability (ID) is a neurodevelopmental disorder characterized by impairments in intellectual and adaptive functioning in three domains: conceptual, social and practical. This disorder can be divided into categories: syndromic, associated with clinical, morphologic, metabolic, radiological or biological anomalies; or non-syndromic, in which the only symptom is cognitive deficits[7]. The symptoms for this disorder start during the developmental period (early infancy) and are diagnosed based on the severity of deficits in adaptive functioning[8].

Globally, it is estimated that 1-3% of the population is affected by this disorder, more prevalent in males[7]. The fact that some forms of ID are caused by mutations in chromosome X (e.g. fragile X syndrome) could explain the higher prevalence in males. The etiology of ID is in most cases unknown. However, approximately 25% has a genetic cause, from chromosomal aneuploidy to single point mutations, and about 15% is caused by an environmental factor, like premature birth[9]. Some genes that encode for glutamatergic synaptic proteins were implicated in ID. Examples of those genes are GRIN2A and GRIN2B (NMDAR subunits) [10] and GRIA3 (AMPAR subunit)[11].

1.2 Schizophrenia

Schizophrenia (SCZ) is a chronic brain disorder characterized by symptoms that cause social or occupational dysfunction like hallucinations and disorganized speech

and behavior. To make a diagnosis, symptoms must have been present for six months of which at least one month the patient suffered from active symptoms. The symptoms usually appear in early adulthood, late teens or early 20s for men, and 20s or early 30s for women[12].

Worldwide, it is estimated that 0,4% of the population is affected by this heritable disorder[13,14]. However, not only gender (males have a higher chance of developing the disease) is a risk factor, but some environmental factors can also influence the onset and severity of the symptoms. From these factors, we can stress out like pre- and perinatal events (including birth complications and late parenthood) and stressful events during childhood[15,16].

Numerous mutated genes that encode proteins that are required for homeostatic plasticity, were associated with SCZ, as seen for ARC[17] and CaMKII[18]. Consequently, it was suggested that a possible explanation for the phenotype of the disease is an imbalance in the excitation/inhibition ratio in the prefrontal cortex since it is involved in executive and working memory functions[19].

2 Glutamatergic Synapses

The most abundant excitatory neurotransmitter in the central nervous system (CNS) is glutamate, first characterized by Frode Funnum in 1984[20,21]. This neurotransmitter enters vesicles in the presynaptic terminal and then, when those vesicles fuse with the membrane of the neuron at the active zone, it is released into the synaptic cleft. The glutamate will then act on postsynaptic glutamate receptors which can be classified as metabotropic glutamate receptors (mGluRs) or ionotropic glutamate receptors (iGluRs). Other than glutamate receptors, on the postsynaptic side, there are kinases, phosphatases and scaffold proteins, like PSD-95, that are responsible for organizing the receptors and their regulators[22,23].

Glutamatergic synapses are critical to regulate the viability and the function of neuronal circuits, and, ultimately, the behavior and responses to environmental cues[24].

2.1 Metabotropic Glutamate Receptors

GPCRs superfamily, the largest group of membrane proteins, can activate different signaling pathways (mainly mediated by G-proteins and arrestins effectors) to trigger a cellular response to a wide range of extracellular signals[25]. In this superfamily, Class C contains all eight subtypes of mGluRs (mGlu1-8), which can be divided into three subgroups according to their sequence similarity, G protein-effector coupling and pharmacological properties. Group I mGluRs include mGlu1 and mGlu5, which are coupled to Gq-protein. This group stimulates phospholipase C and synthesis of inositol trisphosphate (IP3) and diacylglycerol, leading to an increase of the intracellular levels of calcium and protein kinase C activity[26–28]. Group II mGluRs,

which include mGlu2 and mGlu3, and group III mGluRs, comprising mGlu4, mGlu6, mGlu7 and mGlu8, are preferentially coupled to Gi-proteins[29] and are negatively coupled to adenylyl cyclase, and as such, they inhibit the formation of cyclic adenosine monophosphate (AMP)[30].

Depending on the localization of the mGluRs, distinct mechanisms are used to modulate synaptic transmission. Group I receptors are localized in the postsynaptic compartment away from active zones, whereas Group III receptors are localized in the presynaptic compartment, near active zones. Group II receptors are localized perisynaptically[30]. These receptors, more specifically group I mGluRs, have been found to be essential for both Hebbian plasticity[31,32] and non-Hebbian plasticity[33].

2.2 Ionotropic Glutamate Receptors

The group of iGluRs includes kainite receptors (KARs), N-methyl-D-aspartate receptors (NMDARs) and α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors (AMPARs)[24].

2.2.1 NMDA Receptors

NMDAR are channels highly permeable to calcium that not only need glutamate to be activated but also a co-agonist (D-serine or glycine) and previous depolarization of the plasma membrane to remove the voltage-sensitive magnesium blockage[33].

These heterotetrameric receptors differ in subunit composition. According to sequence homology, these subunits can be divided into three subfamilies, GluN1 subunit, GluN2 subunits (GluN2A, GluN2B, GluN2C and GluN2D), and GluN3 subunits (GluN3A and GluN3B). Between subunits, the number of amino-acids varies from 900 to over 1,480, mostly due to the different sizes of the C-terminal. This area has the key role in regulating the trafficking of NMDAR[33].

NMDAR plays an important role in some forms of synaptic plasticity that was already associated with learning and memory. In fact, different subunits of NMDAR can promote long term potentiation (LTP) or long term depression (LTD) dependent on calcium influx by the receptor[34].

2.2.2 AMPA Receptor

AMPARs are the ionotropic glutamate receptors that mediate fast excitatory neurotransmission. The kinetics of these receptors are tuned by the composition and alternative RNA processing of the four subunits (GluA1–GluA4) that assemble in homomeric or heteromeric tetramers and by auxiliary factors[35,36]. These subunits can undergo post-transcriptional editing with relevant physiological impact. The best-studied example is the mutation glutamine-arginine in subunit GluA2, which leads to a

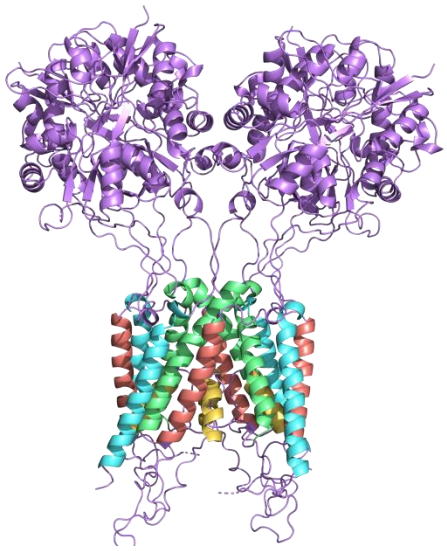

Figure 1 - Representation of the AMPAR with four GluA2 subunits

reduced permeability for calcium ions. This edition has a significant effect as it avoids glutamate-mediated excitotoxicity[37]. The subunits can also undergo alternative splicing resulting in two variants, flip or flop. The most notorious difference between these two variants is that flop desensitizes faster in response to glutamate compared to the flip variant [38]. AMPAR can also be phosphorylated in serine residues of the C-terminal, which will regulate trafficking, insertion and AMPAR-mediated currents. Apart from kinases, other proteins also interact with AMPAR and regulate their trafficking, insertion and gating properties [39].

AMPAR subunits consist of four domain layers: an extracellular N-terminal domain (NTD), a ligand-binding domain (LBD), a membrane-embedded transmembrane domain (TMD) and a C-terminal domain (CTD). The TMD is formed by four helices (M1-M4), of which M2 is a reentrant loop that forms the pore, and M1, M3 and M4 are transmembrane helices[40].

Depending on the brain region, cell type and development stage, the composition of AMPAR changes. While in the cortex, hippocampus, olfactory regions, basal ganglia, lateral septum and amygdala there is a predominant expression of GluA1, GluA2 and GluA3, in the cerebellum and reticular thalamic nuclei GluA4 is more abundant[39].

Pathological disruptions in synapse structure are believed to trigger the development of psychiatric disorders such as autism and schizophrenia. It is also suggested that AMPAR can contribute to the pathophysiology of schizophrenia[41].

2.3 Transmembrane AMPAR Regulatory Proteins

Transmembrane AMPAR regulatory proteins (TARPs) are a family of proteins that can regulate not only the trafficking and insertion at the synapse but also the channel properties of AMPAR. According to their sequence similarity and functional properties they are divided into different subfamilies, Type I (Ia and Ib) and Type II. Type Ia comprises γ2, also known as STG, and γ3, which rate of desensitization and deactivation of GluA1 is less extensive when compared to Type Ib TARPs. This later subfamily includes γ4 and γ8. Type II TARPs consist of γ5 and γ7, whose properties differ profoundly from Type I TARPs[42,43]. The C-terminal of Type I TARPs differs from Type II TARPs. More precisely, while Type I has a class I PDZ-domain and ten

phosphorylation sites, Type II has an atypical PDZ-domain and only two of the ten phosphorylation sites. This difference may affect which proteins they bind with and enable differential receptor regulation[44–46]. Furthermore, Type I TARPs are differently expressed in the brain. While STG is mostly expressed in the cerebellum and to a lesser extent in the cortex and hippocampus, γ-3 is enriched in the cortex, being also found in the hippocampus and olfactory bulb. On the other hand, TARP γ-8 is highly expressed in the hippocampus and to a lower extent in the cortex and olfactory bulb, whereas γ-4 is found in lower concentration in the cortex and olfactory bulb[47]. Type II TARP are enriched in the cerebellum and olfactory bulb, but they can still be found in the hippocampus. Whereas γ-5 can also be present in locus coeruleus and interpeduncular nucleus, γ-7 can also be found in the cortex, striatum, thalamus and brainstem[48,49].

2.3.1 Stargazin

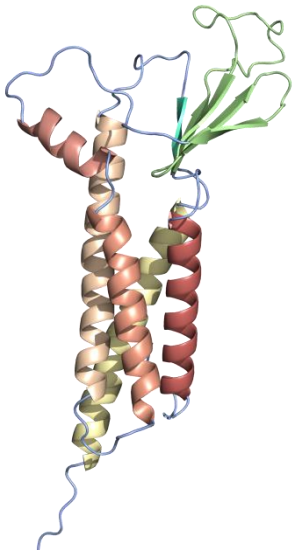

Figure 2 - Representation of the protein stargazin

The big interest on the stargazin (STG) arose from the identification of stargazer mice, which do not express that protein and displayed severe generalized spike, wave seizure disorder and an easily recognizable behavior, recurrently tipping their heads back to stare upward as if they were "gazing the stars"[50]. STG, a transmembrane protein, is encoded by the calcium voltage-gated channel auxiliary subunit gamma 2 (CACNG2) gene. This protein is key to regulate the trafficking, stabilization and homeostatic synaptic scaling of AMPAR and to modulate their gating properties (deactivation and desensitization) by slowing their rates of activation[44,51–53].

STG has 4 transmembrane domains, in which the third and fourth are crucial for the interaction between STG and AMPAR while the first and second face the lipid membrane[51]. STG also has two extracellular segments; the first between the first and second transmembrane domains with four β-sheets and a small helix, which is important for the interaction with the STG and the promotion of cell-cell adhesion and the second between the third and fourth transmembrane domains with only one β-sheet[51–55]. The C-terminal of this protein contains nine serine residues located in a short consecutive region of the cytoplasmic domain. The phosphorylation of these residues by protein kinase C (PKC) or Ca2+/calmodulin-dependent protein kinase II (CaMKII) is important to regulate synaptic recruitment and stabilization of AMPAR. Moreover,

when they are not phosphorylated, STG binds to negatively charged lipids in an electrostatic-dependent manner. This interaction is due to the existence of eight arginine residues located around the 9 serine residues in STG[56]. The C-terminal STG also contains a PDZ binding motif that allows this protein to bind with synaptic PDZ proteins able to target the AMPA receptors to synapses[57,58]. When this domain is phosphorylated by PKA, more precisely in the threonine 321, the interaction with PSD95 is disrupted[59]. Other binding partners of STG are:

Synaptic PDZ proteins: Proteins from the postsynaptic density that can bind to proteins with a PDZ domain were shown to interact with STG in vivo (PSD 95 and PSD93) or only in vitro (SAP-97, SAP-102, MUPP1 and OMP-25) This seems important because this proteins can mediate clustering of ion channels[60,61];

MAGI-2: Scaffold protein that can interact with STG and seems to be able assemble neurotransmitter receptors and cell adhesion proteins[62];

Npist: Protein is known to bind with ERBB2 receptor stabilizing it. It was observed that it binds to STG and might be implicated in the trafficking of AMPAR[63,64];

Erbin: Protein needed for the stabilization of STG and AMPAR surface expression[65];

Light chain of microtubule-associated protein 1A: Protein able to regulate neuronal differentiation and microtubule dynamics. It is believe that through its interaction with STG plays a role in the trafficking of AMPAR to the cell surface[66];

Serine racemase: Enzyme that converts l-serine into d-serin, co-agonist for the glycine site of NMDAR. It interacts with PSD-95 and STG, which aids its membrane localization and inhibits its activity[67];

ARC: Protein required for learning and memory. Its N-terminal lobe binds with STG and CamKII to be able to play its role in synapse[68].

STG is increasingly attracting attention in the last few years. Some computational studies were performed with this protein, including a study published in 2016 by Zhang et al. in which they investigated the interaction between Arc and STG. This interaction is established through a hydrophobic pocket from Arc and β-sheets from STG and is presumably responsible for mediating Arc-dependent homeostatic scaling of AMPAR[68]. In the same year, another group described that AMPAR and STG interaction is mediated by the last transmembrane domains (TMDs) and the first two beta-sheets of STG. While the last two TMDs interact with the first and last TMDs of GluA2, the beta-sheets interact with the ligand binding domain (LBD)[51].

In 2018, Twomey et al. used three different channel blockers to clarify the mechanisms of blockage of AMPAR. The blockers can diffuse into the open channel and bind to the central cavity. Since these blockers do not interfere with the gating machinery, the ion channel closes with the blocker inside and traps it within. For it to exit, AMPAR must open[69].

2.3.2 Role of CACNG2 gene in psychiatric disorders

The CACNG2 gene was associated as a potential susceptible gene for psychiatric disorders. In fact, this gene is located in the 22q chromosome region, shown to be a risk-conferring area for single nucleotide polymorphisms (SNPs), like those identified in patients with schizophrenia[70–72], bipolar disorder[73,74] and intellectual disability[75]. Furthermore, a study done in the anterior cingulate cortex of patients diagnosed with schizophrenia showed significant changes in the mRNA protein levels of other TARPs. More precisely, the levels of mRNA were only significantly decreased in γ-8, while the levels of proteins were significantly decreased in γ4, γ7 and γ-8 and increased in γ-3 and γ-5[76].

Recently, a missense mutation in the CACNG2 gene was found in an 8-year-old male with a non-syndromic intellectual disability. The mutation leads to the substitution of valine 143 for a leucine residue that affects the function of the protein. In fact, the V143L mutation in STG significantly decreases the STG's ability to interact with GluA1 and GluA2 AMPAR subunits[75].

Objectives

STG, by interacting with AMPAR, plays a major role in regulating its function, transporting it to the synapse and aiding in the homeostatic scaling of AMPAR. Mutations in STG were found to lead to some neuropsychiatric disorders, however the exact effects of these mutations on the structure and function of STG and the complex AMPAR:STG are still unknown. Thus, the main aim of this work was to detail the mechanistic effect of those mutations on STG and their consequences on STG's ability to interact with AMPAR.

Methodologies

1 Protein Sequence Alignment

Sequence alignment is a method for comparing two or more sequences of DNA, RNA or proteins by searching for aggregates of characters that are in the same order[77]. For that, we can use one of three types of sequence alignment: global, local or multiple sequence alignment. An example of a global alignment is the Needleman-Wunsch algorithm, a technique commonly used in bioinformatics. This algorithm uses a scoring system that gives a value for each match, mismatch, and gap and a two-dimensional matrix with a size $(|x| + 1) \times (|y| + 1)$, where $|x|$ and $|y|$ are the lengths of the sequences x and y. In the first cell of the first column and first row there is a 0 and subsequently the gap scores are added to these columns and rows. The score for each remaining cell is computed using the following equation:

$$M_{i,j} = \max [M_{i-1,j-1} + s_{i,j}, M_{i,j-1} + g, M_{i-1,j} + g] \quad \text{(Equation 1)}$$

Where $M_{i,j}$ is the element of the ith row and jth column in the matrix, $s_{i,j}$ is the substitution score and g is the gap penalty. The value on the last row and column in the matrix is the alignment score [78].

To represent the alignment of two amino-acids sequences it is usual to have each of them in a row. In this representation bundles of amino-acids that are identical or similar will be represented in the same column, while non-identical amino-acids will be represented by a gap in the other sequence or by being added to the other sequence[79]. The protein sequence alignment is normally used to align protein sequences of animals which are evolutionarily related[77].

1.1 Clustal Omega[80]

Clustal Omega[80] is a multiple sequence alignment (MSA) tool that can deliver alignments of virtually any number of sequences of amino-acids with high accuracy in a short amount of time. Sequences can be given in many formats like GCG and FASTA. The current sequencing limit is 4000 sequences and 4MB of data[81,82].

2 Homology Modeling

Even though the amount of three-dimensional (3D) structures deposited in the protein data bank (PDB)[83] has quickly increased in the last few years, the number of membrane protein structures is still far from desirable. This can be explained by the fact that the experimental technique needed to obtain crystals of membrane proteins is still a really challenging[84,85].

Several theoretical methods can be applied to overcome the absence of 3D structures of macromolecular proteins. Homology modeling, also nominated comparative modeling or template-based modeling, is the most reliable and common technique used in the field. This tool predicts the 3D structure of a protein when the sequence of amino-acids is available and the structure of a homologous protein is known[86,87].

The basis of the method is the observation that two proteins with a similar amino-acid sequence, will maintain the main fold and characteristic features of tertiary structure. The first step is the identification of the correct template by alignment of amino-acid sequences. This alignment is then used to construct the 3D structure of the protein, a model that should be optimized and its quality tested. As a general rule, to obtain a good template, a minimum of 30% of sequence identity is required between the target and the template, since below 25% of sequence identity it is difficult to assume a common ancestry. Usually, a higher sequence identity correlates with a more accurate model, however, functionally important similarities between proteins aren't always evident from just the comparison of the raw sequences. To be able to recognize those similarities one must compare the 3D structures, due to the fact that many proteins are overlooked as template structures because they share little sequence homology with the target sequence when they could potentially share structural similarity[87].

The accuracy of the predicted protein is critical to understand how good the performance of the prediction method is. Numerous authors have proposed different quality assessment measures that can be used to assure that the model has the correct structure[88,89]. Examples of these quality measurements are: i) Discrete optimized protein energy (DOPE), a statistical potential optimized for assessing the quality of models dependent on atomic distance[90]; ii) Z-score, a quality measure of the deviation of a protein structure energy in relation to an energy distribution derived from random conformations of the protein[91]; iii) LGscore, a structural score, is calculated using structural P-values of the protein[88]; and iv) MaxSub, another model quality measure, is calculated through the identification of the maximal subset of residues whose distance between the model and the experimental structure are shorter than 3.5 Å[89].

2.1 SWISS-MODEL[92–96]

SWISS-MODEL[92–96] is a fully automated protein homology modelling server that has been improving since its creation and now generates around 3000 models a day. This server uses homology modelling to create a 3D model of a protein from a target sequence by extrapolating experimental information from an evolutionary related protein structure that serves as a template [92].

2.2 MODELLER[97]

MODELLER[97] is a software widely used for homology modelling with or without restrictions. The input the user has to provide is the alignment of the sequence to be modeled and the atomic coordinates of the template structure. Additionally, restrains can be added which will operate on secondary structure and special bonds, like disulfide bridges[98].

2.3 PSIPRED[99]

The PSIPRED[99] is a simple and accurate protein structure prediction server that allows its users to get a prediction of a protein secondary structure based on the sequence of the protein in single letter amino-acid code format or in FASTA format. It incorporates three new methods (PSIPRED[100], GenTHREADER[101,102] and MEMSAT 2[103]) to predict structural information about a protein from its amino-acid sequence. This method incorporates two feed-forward neural networks that perform an analysis on the output obtained from PSI-BLAST (Position Specific Iterated BLAST). This method also uses a rigorous cross-validation procedure to evaluate its performance. MEMSAT 2[103] is an accurate secondary structure and topology of integral membrane proteins prediction method based on its ability to recognize topological models and to use a variety of sequence alignments generated by PSI-BLAST. The use of consensus information in the scoring of different topological models has increased its prediction accuracy. Finally, GenTHREADER[101,102] is an accurate fold recognition method that can be applied to proteomes or individual protein sequences with the aim of detecting superfamily relationships. It uses a conventional sequence alignment algorithm to create sequence-structure alignments, which are then analyzed. A variant of this method calculates several sequence profiles of the target sequence using PSI-BLAST, and uses these profiles to make an alignment to sequences of the fold library[104].

3 Molecular Dynamics Simulation

Molecular dynamics (MD) simulations have become an effective technique to understand macromolecular structure-function relationships[84,85,105]. It allows the study of the movement of atoms and molecules, which provides a dynamic view of the evolution of the system[84,85,105]. This technique has several uses including protein docking[106], folding and insertion onto the membrane[107], protein structure refinement[108], protein aggregation[109], large conformational protein changes[110], protein-ligand binding[111], protein crowding[112], channel gating[113] and protein-lipid interactions[109,114,115]. It can also be used in the drug development pipeline to characterize the interaction between proteins and other molecules[106,116–119] and the uptake of small molecules[120].

Major advances in both hardware and mathematical algorithms used in MD simulations lead to an increase in performance allowing longer run times on increasingly larger systems[121]. In a simulation procedure it is important to choose carefully the force fields (FFs), mathematical expressions that can describe the dependence of the potential energy of a system on the coordinates of its atoms[122], since the choice will affect how accurate the simulation is compared to experimental data[123]. The accuracy of FFs has been continuously improvement as can be observed in recent studies[124–126].

To perform a simulation, the 3D structure of a protein must be available in a public database (PDB)[83,123,127] or be modeled through homology modeling or *de novo* modeling[83,123,127]. When the structure of the protein of interest or similar to it is unavailable, MD can be used as a physics-based *ab initio* modeling tool, since it can predict protein folding[128]. Regarding membrane proteins, the lipid composition of biological membranes is an important factor to be considered. This composition varies depending on certain conditions such as cell type, cell cycle phase and environmental conditions. Since the interaction with specific lipids and their location in the inner or outer leaflet of the membrane can influence membrane protein functions, the built of a system as to consider the lipids used in the inner and outer leaflet and create a membrane with 0 leaflet tension[129,130]. After deciding the composition of the membrane, different approaches can be taken to insert the protein into the membrane as inflategro or -membed (flag applied to the command gmx mdrun from the GROMACS software). While inflategro, start by inflating the membrane before removing overlapping lipids, followed by deflation, -membed, starts by contracting protein's width and removing overlapping molecules, followed by a gradual protein decompression until it reaches its full size in a short MD simulation[131,132]. Alternatively, to automatically build the system, the CHARMM-GUI[133–140] tool can be used since it generates several input files that can be used to start the simulation and avoid an additional step of solvation.

To represent a MD system, a mixture of implicit and explicit methods can be applied[141–145]. An implicit solvent model is an approach that enhances the speed of MD simulations by treating the solvent (water) as a simplified continuum media while doing an atomistic representation of the rest of the system[146]. Implicit methods are useful to study large-scale phenomena of proteins[147], ligand docking[148], protein-protein interactions[149], creating free energy landscapes[150] and folding of proteins[151]. On the other hand, explicit representations of the system can be split into all-atom or atomistic molecular dynamics approach (AAMD), and coarse-grained molecular dynamics approach (CGMD)[84,152]. While AAMD is computationally more demanding, hindering longer time scales beyond millisecond in large systems, it can offer more detailed information about lipids and proteins' local structural properties. On the other hand, CGMD, with less accurate representation of the number of atoms in proteins and

lipids, enables system's MD simulations to reach longer time scales compared with the AAMD[84,152,153].

Depending on how the system is represented, different FFs are available. Depending on the constituents of the system, certain FFs may not have the proper parameterization needed to ensure reliable results. FFs can be categorized according to the environment's ability to influence atoms' partial charge as addictive or non-polarizable and non-addictive or polarizable. Despite the apparent advantage of non-additive FFs, their high cost and effort to develop hindered its progression and usage[125], making addictive FFs the most commonly used in MPs[154]. These FFs are integrated into a wide diversity of software used to run MD simulation with the most commonly used being Groningen machine for chemical simulation (GROMACS)[155,156], nanoscale molecular dynamics (NAMD)[157], chemistry at Harvard molecular mechanics (CHARMM)[135], and assisted model building with energy refinement (AMBER)[158].

To avoid the possibility of interaction problems near the system boundary, a simulation will usually use constrained spherical boundary models for solute and solvent or the cubic or rectangular periodic boundary conditions (PBC), which consists of the repetitions of the system in the 26 adjacent unit cells. This way, PBC approach leads to conservation of mass and number of particles since when a molecule leaves the system, another equivalent enters[159]. An important step in MD is the energy minimization according to the position of side-chains atoms. To compute the minimum energy there are a lot of methods but the most commonly used are steepest descent and conjugate gradient. Initial velocities are then assigned to each atom[159], which can be done using Maxwell-Boltzmann distribution equation:

$$P(v_{ix}) = \sqrt{\left(\frac{m_i}{2\pi kT}\right)} \exp\left(\frac{-m_i v_{ix}^2}{2kT}\right) \quad \text{(Equation 2)}[160]$$

As such, new velocities are assigned to slightly higher temperatures in small intervals until the desired temperature is achieved. Constraining forces are gradually removed as structural tensions dissipate by heating. The equilibration of kinetic and potential energies is done in the equilibration and when the solvent is equilibrated, all the constraints on the protein can be removed[159]. The production phase is carried out to generate a trajectory for the protein in compliance with specific equilibrium conditions during the desired amount of time. In this step, Newton's second law of motion:

$$F_i = m_i a_i \quad \text{(Equation 3)}[161]$$

$$F(x) = -\nabla U(x) = m\frac{d^2 x_i}{dt^2} \quad \text{(Equation 4)}[153]$$

are incorporated to be able to predict the trajectory of atoms. In an AAMD simulation atoms are represented as spheres that can interact with other atoms by means of a

potential energy function. This function takes into consideration both intramolecular and intermolecular interactions[162]. Due to the structural complexity of the systems used, there is a need to include a huge variety of atomic parameters, however FFs can be described using the follow equation[163]:

$$U(\vec{R}) = \sum_{bonds} K_b(b - b_0)^2 + \sum_{Urey-Bradley} K_{UB}(S - S_0)^2 + \sum_{angle} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_x(1 + \cos(n_x - \delta)) + \sum_{impropers} K_{imp}(\varphi - \varphi_0)^2 + \sum_{nonbond} \varepsilon \left[\left(\frac{R_{min}ij}{r_{ij}} \right)^{12} - \left(\frac{R_{minij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon_1 r_{ij}} \quad \text{(Equation 5)}[163]$$

A wide range of features can be obtained through the results of MD[159]. Examples of quality and stability evaluation features are the radius of gyration, root mean square deviations (RMSDs), and minimal distance evaluation[159]. The analysis of these features is essential to guarantee that the systems' equilibration was successfully attained and to provide the conformation change throughout the simulation[164]. QwikMD[165], a relatively new visual-molecular-dynamics (VMD)[166] plugin, can not only help set up the system to run MD simulations but also be used to make basic and advanced analyses of results.

3.1 CHARMM-GUI[133–140]

Among many web-based modeling tools, CHARMM-GUI[133–140] is an interface that prepares complex biomolecular systems for MD, by creating files that can be used in a number of programs, including GROMACS[155,156,167–172]. This interface typically simplifies both common and advanced simulation systems. For example, to simulate transmembrane proteins, the simulation will have to have water, ions, various lipid types and the protein. CHARMM-GUI[133–140] Membrane Builder helps generate all of that in just a few minutes or hours depending on the size of the system[84,137].

3.2 GROMACS[155,156,167–172]

There are a large number of software packages that implement MD with a high performance and that can be used in parallel in central processing units (CPUs) or graphics processing unit (GPUs) in a supercomputer[155]. GROMACS[155,156,167–172] is a free, versatile software to perform MD that, although mainly designed for biochemical molecules like proteins, is also used to study non-biological systems because of how fast it is in calculating non bonded interactions[167].

Materials and Methods

Case study: Stargazin

Model Construction

The model of STG was created by homology modelling using two structures: one with GluA2:STG complex (PDB-ID: 6DLZ[69]) and another with claudin-19 (PDB-ID: 3X29[173]). The sequence of STG was retrieved from UniProt[174] (Q9Y698). To determine sequence identity, Clustal Omega[175] was used to perform sequence alignment the sequence obtained from Uniprot[174] and sequences from the template structures: for 6DLZ[69] crystal, sequence identity was 99,47%; and for 3X29 was 35%.

First, 100 models were built using MODELLER[97] using the structure of the complex of GluA2 with STG as a template. A model was selected taking into account DOPE score, z-score[91,176], LGscore[177], MaxSub[177] and a final visual inspection. Afterwards, another STG model was built through SWISS-MODEL[92–96] using the crystal of claudin-19 as a template. Since there were a few missing residues in a key beta-sheet in the GluA2:STG crystal, and according to Twomey, E. C., *et al.* (2016)[51] that region is important for STG and AMPAR coupling, the final model was created with the beta-sheets region from the Claudin-19 derived model while the rest of the structure was constructed from the GluA2:STG derived model. Loops of the final model were refined using MODELLER[97]. Due to the lack of secondary structures and subsequent high structural instability, the C-terminal of this protein was removed from the final model. The mutant models of STG were built using mutagenesis tool of PyMOL[178], where residue 143 was substituted with a leucine to create the STG-ID model and the residue 148 was substituted with an asparagine to create the STG-SCZ model.

Molecular Dynamics Simulations

MD simulations of STG, including the WT and two mutated models, were performed using GROMACS 2018.4[179] and the CHARMM36 force field[180]. STG orientation in the membrane was obtained through the oriented crystal of GluA2:STG complex (PDB-ID: 6DLZ[69]). Systems were built using CHARMM-GUI[133–140] membrane builder with a bilayer membrane of POPC:Cholesterol (9:1 ratio) to replicate the physiological environment. These simulation boxes were also hydrated with TIP3 model of water and 0.15M of NaCl. This system had 81 POPC and 9 Cholesterol in each leaflet of the membrane, was hydrated with approximately 14800 molecules of water and was ionized with 39 molecules of sodium and 40 of chlorine. Using a steepest descent algorithm, the systems went through a minimization. To raise the temperature of the systems to 310K, a Berendsen-thermostat was used, and to increase the pressure

to 1 bar, a semi-isotropic pressure coupling algorithm was used. Simultaneously the constraint forces of lipids and protein were successively lessened.

For each system, three independent replicas of 1 μs each were produced. Simulations were performed in the isothermal–isobaric ensemble. Temperature coupling was done using Nose-Hoover thermostat with time constant of 1 ps whereas to maintain the pressure constant a semi isotropic Parrinello–Rahman barostat was used with a time constant of 5 ps and compressibility of 4.5 × 10-5 bar-1. Electrostatic interactions were performed with fast smooth Particle-Mesh Ewald, a cutoff of 1.2 nm. H-bonds were constrained using the linear constraint solver.

Case study: Complex AMPAR:STG

Model Construction

Models for each GluA2 subunit of AMPAR were built using MODELLER with the crystal of GluA2:STG complex (PDB-ID: 6DLZ[69]) as a template to fill the gaps present in the crystal. The sequence of GluA2 was retrieved from UniProt[174] (P42262) and with the sequence alignment was performed with Clustal Omega[175]. The amino terminal domain of the subunits was removed in the final model of AMPAR since it doesn't interact with STG. This model was selected using the same criteria used with the STG model. The complexes AMPAR:STG (WT, ID and SCZ variants) were obtained by superimposition of the STG models obtain prior and AMPAR models with 6DLZ[69] structure.

Molecular Dynamics Simulations

MD simulations of the complex AMPAR:STG, including the WT and two mutation models, were performed using GROMACS 2018.4[179] and the CHARMM36 force field[180]. The complex's orientation in the membrane was obtained through the oriented crystal of GluA2:STG complex (PDB-ID: 6DLZ[69]). Systems were built using CHARMM-GUI[133–140] membrane builder with a bilayer membrane of POPC:Cholesterol (9:1 ratio) to replicate the physiological environment. These simulation boxes were also hydrated with TIP3 model of water and 0.15M of NaCl. This system had 315 POPC and 35 Cholesterol in each leaflet of the membrane, was hydrated with approximately 130110 molecules of water and was ionized with 357 molecules of sodium and 371 of chlorine. Using a steepest descent algorithm, the systems went through a minimization. To raise the temperature of the systems to 310K, a Berendsen-thermostat was used, and to increase the pressure to 1 bar, a semi-isotropic pressure coupling algorithm was used. Simultaneously the constraint forces of lipids and protein were successively lessened.

For each system, three independent replicas of 0,5 μs each were produced. Simulations were performed in the isothermal–isobaric ensemble. Temperature coupling was done using Nose-Hoover thermostat with time constant of 1 ps whereas to maintain the pressure constant a semi isotropic Parrinello–Rahman barostat was used with a time constant of 5 ps and compressibility of 4.5×10^{-5} bar^{-1}. Electrostatic interactions were performed with fast smooth Particle-Mesh Ewald, a cutoff of 1.2 nm. H-bonds were constrained using the linear constraint solver.

Analysis

Solvent-Accessible Surface Area Analysis

Solvent-accessible surface area (SASA) analysis for each individual residue of four interfaces (1 STG + 2 AMPARs) followed a similar protocol reported by Magalhães *et al.*[181], splitting by ligand (STG) and receptor (2 AMPAR). For standardization purposes, the maximum value for each type of amino acid was extracted. These analyses were performed considering the whole complex, without STG and without AMPAR for every replica of the three systems. The mean was calculated for the three replicas and transformed in percentage using the maximum value of the amino-acids.

To find the most relevant residues for the interface between STG and AMPAR, the SASA results of the complex AMPAR:STG of each residue where subtracted to the SASA results of AMPAR and STG isolated. Residues with a difference below 20% and residues in which the standard error of the mean was superior to the value of the difference were excluded.

MMPBSA

The binding free energy of protein and ligands was calculated using molecular mechanics generalized Poisson-Boltzmann (MM-PBSA) method, in which the binding free energy of both proteins was calculated from snapshots of the trajectories of molecular dynamics (MD). The binding free energy was computed using this Equation:

$$\Delta G_{bind} = G_{complex} - (G_{STG} + G_{AMPAR}) \qquad \text{(Equation 6)}$$

The binding free energy was calculated using MMPBSA as implemented in AMBERtools[158]. Only the equilibrated regions of each replicate were considered with a snapshot per ns. The contribution of each residue of the complex to the biding free energy was calculated. All parameters were used as default. Averages and standard error of mean of each AMPAR's and STG's substructure was calculated

Interface Distances

In-house VMD[166] script was used to calculate the intermolecular distances between residues in the interfaces. Interfacial residues were defined as any residues whose atoms were within a 6 Å cut-off of the other chain to ensure that no meaningful interaction was lost during MD.

Root Mean Square Fluctuations

Root mean square fluctuations (RMSF) of Cα atoms in all STG's trajectories were calculated using Bio3D R package[182].

Cross Correlation Analyses

For each system, the cross correlation analysis was calculated for residue-level dynamic on the respective Cα trajectory using the Bio3D R package[182].

Results and Discussion

Part - I Effects of mutations on monomeric STG

To have an overall picture of how the ID and SCZ mutations could structurally affect the STG protein in its monomeric form (**Figure 3A**), a computational approach with different analyses was performed. The RMSF analyses, a measure of the displacement of a particular atom, or group of atoms, relative to the reference structure, averaged over the number of atoms allow to identify and understand the regions with the higher fluctuations. **Figure 3B-D** shows that the more affected region upon mutation is the first extracellular domain, more precisely the first four beta-sheets. Concerning the STG with SCZ mutation, in replica 3 (**Figure 3D**) the third transmembrane domain also presented higher fluctuations. At both mutations, the loop between the 4th helix (3rd TMD where the mutations occur) and B5, which is located at 2nd ECD, presents higher values when compared with STG WT. This data is relevant when analyzing the effect of the protein mutations.

The cross-correlation analysis (CCA) tracks the movements of two or more sets of time series data relative to one another. In both mutated systems, residues in the region of the beta-sheets move in the same direction, while the 3rd TMD and 4th TMD move in opposite directions (**Figure 3E-G**). Analyzing the SCZ mutation (**Figure 3G**), the last half of the 4th TMD goes in the opposite direction of the 1st TMD and the H2 (at the 1st ECD) moves in the same direction of the 1st TMD. This behavior is not observed in the WT system (**Figure 3E**), revealing the mutations impact in the conformational space of the STG in the monomeric form.

The RMSD was another parameter chosen to assess the effects of the mutation on the secondary structures of the protein. The RMSD was calculated for the helices (**Figure 4A, C and E**) and beta-sheets (**Figure 4B, D and F**). The overall obtained results showed that the helices stability seems to be greatly affected by the mutations excepting of H1 (1st TMD) in the ID system and H3 (2nd TMD) in the SCZ system. Several authors[1–4] described that both TMDs do not participate in the coupling between STG and AMPAR. Additionally, it was also observed an increase deviation for H2 (at 1st ECD), H4 (3rd TMD) and H5 (4th TMD) in the SCZ system when compared to WT. Taking into account that 3rd and 4th TMDs are known to play a key role in the interaction between both proteins, this fact opens the possibility of the mutation hindering the coupling. Regarding the beta-sheets, between WT and the system with the ID mutation, the results point to a slight tendency towards a reduced deviation. However, for the SCZ system, an increase in conformational flexibility is observed in all beta-sheets. Since the region of the beta-sheets, specifically the first four, are

considered relevant in the binding between STG and AMPAR, these results indicate a reduction in the binding of the mutated STG and the receptor.

To evaluate the stability of the mutated STGs we calculated the number of hydrogen bonds that were established during the MD simulation between the various STG substructures (**Figure 5**). The results suggest that both mutations can influence the number of intramolecular H-Bonds in the whole structure (**Figure 5A**), as it can be seen in **table 1**, even though it is more significant (with a p-value inferior to $2.2x10^{-16} < 0.05$) with the STG-SCZ. This effect, being more significant within the STG-SCZ, may be related to the results of the CCA, in which we also observed bigger effects on the movement of STG residues. Regarding the H-bonds between the helix residues (**Figure 5B**), in both mutations fewer H-bonds were established during MD simulation (both with a p-value inferior to $2.2x10^{-16} < 0.05$), which may be related to the increase in helix conformational flexibility found by the RMSD analysis in both mutations. Regarding the H-bonds between beta-sheets residues (**Figure 5C**) from both mutations the number of H-Bonds significantly increase, although the STG-ID has a sligtly greater increase, which may explain why this mutated STG has a small deviation in the region where the beta-sheets are located. Regarding H-Bonds between loop residues (**Figure 5D**) in both mutations the number of H-Bonds significantly increased, although the STG-ID had a bigger increase. Regarding the number of H-bonds between the various STG substructures and the rest of the protein (**Figure 5E-G**) the number of H-Bonds decreased significantly in both mutations, this effect being more pronounced in the SCZ mutation. The significant decrease in the number of hydrogen bonds between helix residues and the other sub-structures of STG-SCZ, may explain the increase in the deviation for three STG helices. Likewise, this effect between beta-sheets residues and the other STG-SCZ structures may be related to the deviation increase and loss of stability observed in the RMSD analysis. All these results indicate that the ID mutation has a higher effect on residues of beta-sheets and STG loops, while the SCZ mutation has a more generalized effect on the whole protein.

Table 1 – Average H-Bonds of STG at monomeric form and p-values obtained from Mann Whitney test.

	Average WT	Average ID	P-value (WT vs ID systems)	Average SCZ	P-value (WT vs SCZ systems)
Whole structure	166.85	165.33	8.5×10^{-6}	161.21	$<2.2 \times 10^{-16}$
Between helices	93,39	90.82	$<2.2 \times 10^{-16}$	91.30	$<2.2 \times 10^{-16}$
Between beta-sheets	23.07	24.82	$<2.2 \times 10^{-16}$	24.44	6.1×10^{-13}
Between loops	15.96	24.82	$<2.2 \times 10^{-16}$	18.15	$<2.2 \times 10^{-16}$
Between helices and the rest of STG	30.31	28.85	1.8×10^{-13}	23.08	$<2.2 \times 10^{-16}$
Between beta-sheets and the rest of STG	8.35	7.60	2.8×10^{-13}	6.36	$<2.2 \times 10^{-16}$
Between loops and the rest of STG	36.76	35.09	7.3×10^{-14}	29.05	$<2.2 \times 10^{-16}$

Taking into account the key role of 4th helix (3rd TMD) and 5th helix (4th TMD) and B1-4 for the STG:AMPAR binding[1], the structural conformational analysis point to a decrease protein-protein coupling when either of the two STG mutations is present. The results also point to a greater effect in the STG-SCZ mutated system.

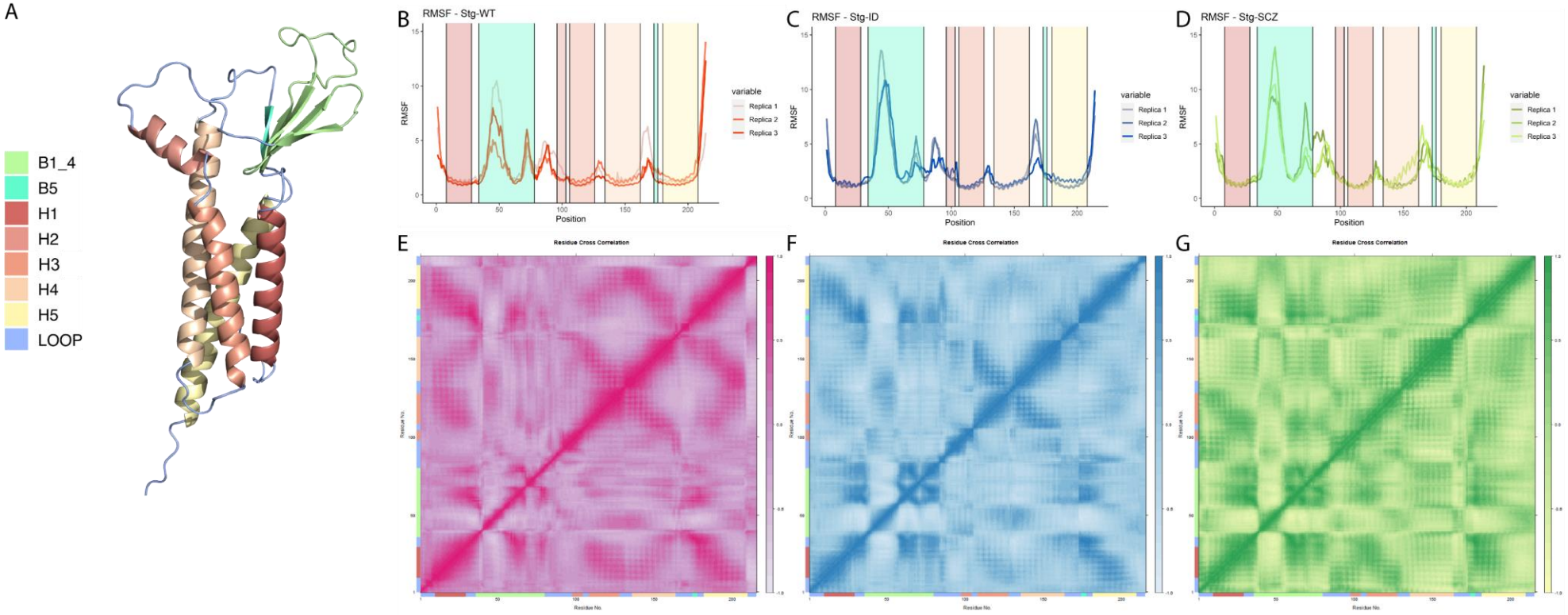

Figure 3 - Results of STG in the monomeric form. (**A**) Representation of STG as cartoon colored according to secondary structures (with color code for each of the sub-structures. (**B–D**) RMSF of the WT system and with both mutations with vertical bars representing each substructure of STG (**E–G**) Dynamical cross-correlation analysis of the wild type system (pink), ID mutation (blue) and SCZ mutation (green) with substructure annotation at the bottom and left side of each map.

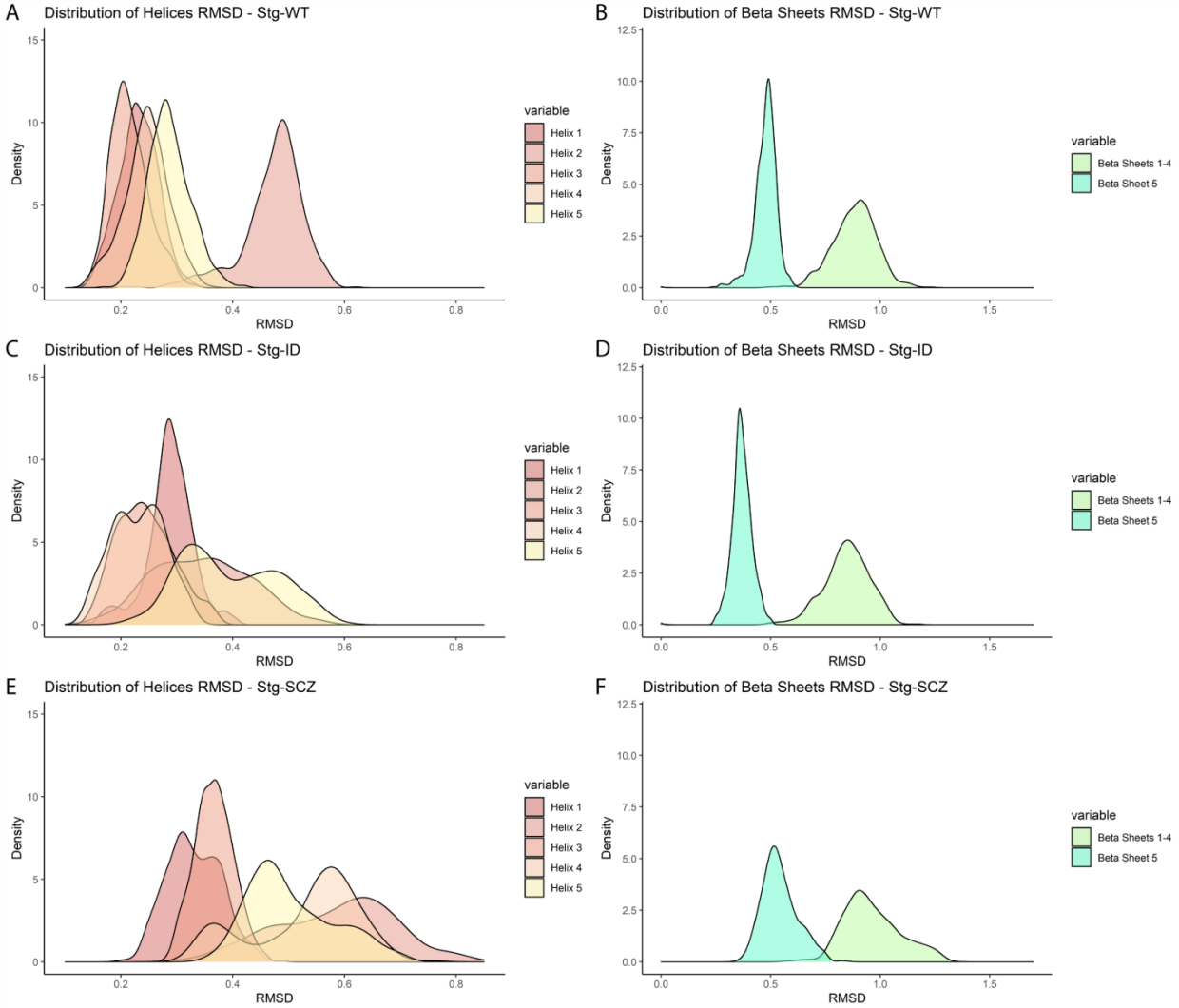

Figure 4 – RMSD density plots for Cα atoms of the secondary structures of the STG at monomeric form: (**A, C and E**) RMSD of the protein helices of all systems colored using the colors of the previous figure. (**B, D and F**) RMSD of the protein beta sheets of all systems colored using the colors of the previous figure.

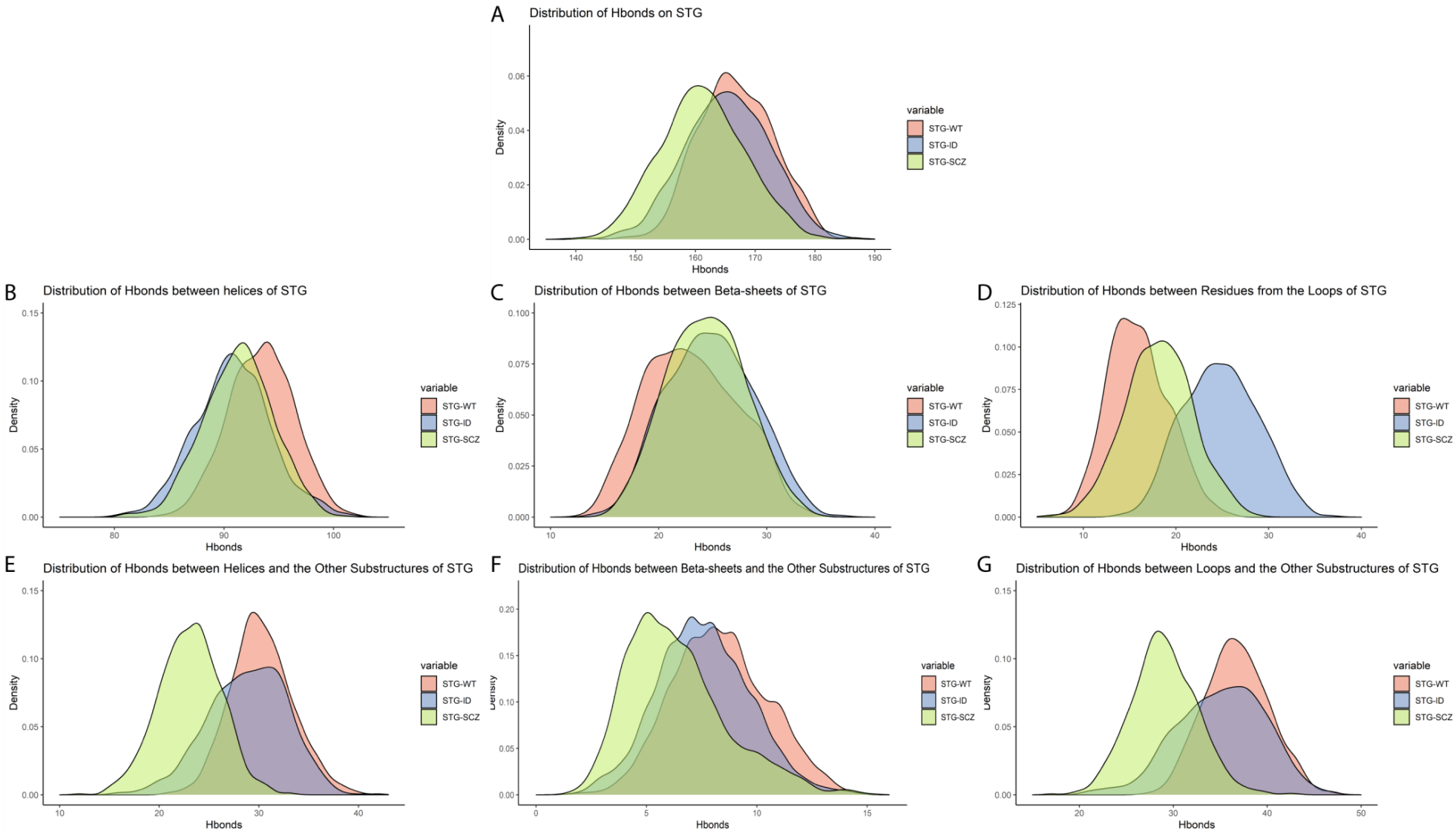

Figure 5 – Distribution of H-Bonds in the different substructures of STG in the monomeric form: The density plot where colored according to the system that the H-bonds where calculated (WT – pink; ID – blue; SCZ – green). The number of H-bonds formed was calculated for amino acids in the whole protein **(A),** between helix residues **(B),** between beta-sheet residues **(C),** between loop residues **(D),** between helix residues and the other substructures **(E),** between beta-sheet residues and the other substructures **(F),** and between loop residues and the other substructures **(G)**.

Part II - Effects of mutations on STG complexed with AMPAR

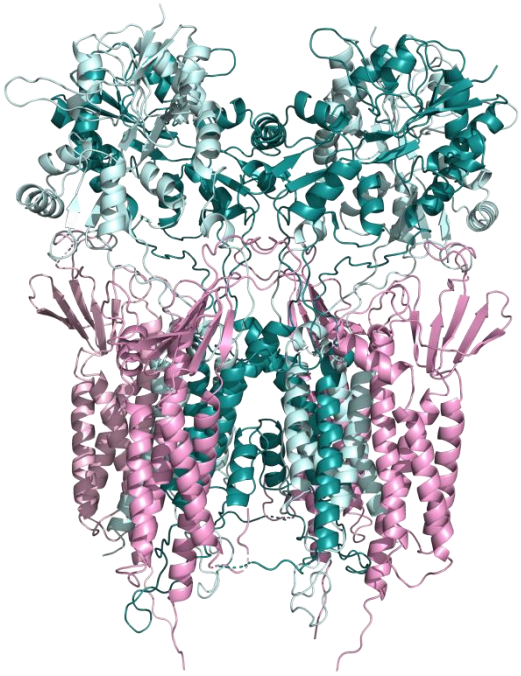

Figure 6 - Representation of the four GluA2 subunits (dark and light cyan) interacting with STG(pink).

Based on the results obtained for STG in monomeric form, the effect of mutations on the STG/AMPAR binding was also analyzed. To that purpose we created a system with four GluA2 subunits from AMPAR and four STGs. The X-ray structure of the system used for MD simulations allows us to perceive that the interaction between the two proteins occurs between 1 STG and two GluA2 (**Figure 6**). Thus, after dimer formation, we have 4 different possible interfaces comprising 1 STG and 2 GluA2 proteins.

The RMSF values (**Figure 7**) show a slightly increase of residue fluctuation of both proteins in the first B4. With the exception of 3 replicas of the ID system (**Figure 7 E-H**), the RMSF values were lower than in the complex with the STG-WT. Within the ID system, these higher fluctuations were observed in similar positions to what was already observed for the STG in monomeric form, precisely in the beginning/middle region of the beta-sheets. However, for the SCZ system (**Figure 7 I-L**) these higher fluctuations were more marked near the end of the region where these beta-sheets are located, which is located in a different position when compared to the STG in the monomeric form.

Similarly to what was done for the monomeric forms, we also performed CCA (**Figure 8**) and observed that the first four beta-sheets had a positive correlation in both mutated systems, however the 3rd TMD and 4th TMD showed a less accentuated movement. Additionally, in the STG-SCZ:AMPAR complex, the 1st TMD and 4th TMD showed a positive correlation.

The RMSD values were also calculated for the helices (**Figure 9**) and beta-sheets (**Figure 10**). The results showed that the H1 (1st TMD) and H3 (2nd TMD) of the mutated systems had a smaller deviation in relation to WT and that H4 (3rd TMD) and H5 (4th TMD) showed slightly greater deviations. When comparing these results with those obtained for the monomeric STG, it was observed that upon AMPAR binding, the STG-WT:AMPAR complex tends to gain greater conformational flexibility, while the

mutated forms seem to lose part of the conformational flexibility that they had previously, although flexibility is greater in the binding domains. Regarding the beta-sheets, the B5 of the STG-ID had smaller deviation when compared to STG-WT while the B1-4 tend to have an increased conformational flexibility. For the STG-SCZ:AMPAR complex all beta-sheets showed lower deviation when compared to the STG-WT:AMPAR complex.

The average distance between the interfacial residues of both macromolecular systems was also performed (**Figure 12**). Between the STG-WT/AMPAR complex and the complexes with the mutated STG, for amino acids with the smallest distance between both proteins, there is not much variation either in the STG (Leu147, Ser158, Glu191 and Val198) or in the GluA2 subunits of AMPAR (Gly556, Val559, Leu563 and Ser811). To be more specific the average for the STG-WT:AMPAR complex the distance between the pair Leu147 and Gly556 is 7.23±0.21 Å, between the pair Glu191 and Glu559 is 7.12±0.30 Å, between the pair Val198 and Leu563 is 6.79±0.31 Å and between the pair Ser158 and Ser811 is 5.47±0.74 Å. For the STG-ID:AMPAR complex the distance between the pair Leu147 and Gly556 is 7.18±0.48 Å, between the pair Glu191 and Glu559 is 7.32±0.16 Å, between the pair Val198 and Leu563 is 6.81±0.40 Å and between the pair Ser158 and Ser811 is 5.16±0.72 Å. For the STG-SCZ:AMPAR complex the distance between the pair Leu147 and Gly556 is 6.78±0.17 Å, between the pair Glu191 and Glu559 is 7.03±0.20 Å, between the pair Val198 and Leu563 is 6.90±0.36 Å and between the pair Ser158 and Ser811 is 4.89±0.40 Å. Of all these pairs of residues the one that consistently appears in all interfaces of both the WT and mutated systems with the lowest values is the pair Ser158 and Ser811. The average distance between both proteins in the binding region of the SCZ, especially regarding the first four beta-sheets, increases when compared to the WT system.

A detailed analysis of the interfacial residues of both proteins was also performed by measuring the coupling impact on the solvent exposition of the different interfacial residues**.** We observed that the average number of residues involved in the different interfaces were 52.25±2.75 from STG and 24.50±3.42 and 25.50±7.59 from both GluA2s for the STG-WT:AMPAR, 46.00±7.44 from STG and 21.50±6.24 and 24.00±8.91 from both GluA2s for the STG-ID:AMPAR, and 49.50±6.35 from STG and 24.25±4.03 and 23.00±4.24 from both GluA2s for the STG-SCZ:AMPAR. Regarding the interaction area, we observed that the area was 2247.08±210.68nm^2 for STG and 1144.04±246.36nm^2 and 947.18±218.29nm^2 for both GluA2s from the STG-WT:AMPAR, 1975.66±423.58nm^2 for STG and 965.68±234.74nm^2 and 910.97±269.40nm^2 for both GluA2s from the STG-ID:AMPAR and 2104.14±227.45nm^2 for STG and 1110.95±188.21nm^2 and 886.58±115.45nm^2 for both GluA2s from the STG-SCZ:AMPAR. The results showed that the STG-WT (**Figure 13**) had more residues with higher ΔSASA values and

therefore, more involved in the protein-protein coupling when compared with the STG-ID. To be more explicit, while the beta-sheets residues of the STG-WT that are present in interface tend to vary between Lys41 and Lys52, in the ID system, except for interface 4, the beta sheets are not present on the interface or vary between Lys41 and Val45 and Lys43 and Asn48. Contrary to expectations, the results for STG-SCZ:AMPAR complex do not show an obvious difference compared to the STG-WT:AMPAR complex. Additionally, the residues in the H4 (3rd TMD) and H5 (4th TMD) are similar between the three systems and Leu98 from STG is present in every interface. Regarding the results for AMPAR subunits (**Figure 14**), the residues of the M1 and M4 are similar in all systems, and number of residues varies mostly in the loops region between M1 and M2 and between M3 and M4. This behavior is more easily understood in two of the interfaces of the ID system, in which the number of residues decreases visibly. Ile594, from M2 is also present in all interfaces of all systems, suggesting that the ability of STG to modulate the gating properties of AMPAR is not affected by the mutations.

At complex systems, the number of intramolecular H-bonds (**Figure 11 A, C and E**) tends to decrease in comparison with MD simulations from the monomeric form of STG as described in **Table 1 and 2.** Comparison between the mutated systems and the WT, showed a significant increase of the number of H-Bonds in most interfaces, especially in the system with the mutation found in a SCZ patient.

Regarding the intermolecular H-Bonds between STG and AMPAR (**Figure 11 B, D and F**), the mutations led to a significant decrease in almost all interfaces (**Table 2**) in particular within the SCZ mutated system. We observed that the average H-bonds involved in the different interfaces were 17.10±4.29 for the STG-WT:AMPAR, 14.44±6.02 for the STG-ID:AMPAR, and 12.55± 4.83 for the STG-SCZ:AMPAR.

Table 2 – Means and p-values obtained from Mann Whitney test of the intramolecular and intermolecular H-Bonds for the different systems.

	Mean WT	Mean ID	P-value (WT vs ID systems)	Mean SCZ	P-value (WT vs SCZ systems)
Whole structure	156.03	155.84	0.90	160.26	$<2.2x10^{-16}$
	155.62	154.56	0.03	156.15	0.10
	148.68	153.92	$<2.2x10^{-16}$	156.07	$<2.2x10^{-16}$
	144.35	152.54	$<2.2x10^{-16}$	147.97	$2.8x10^{-13}$
Between Proteins	20.69	20.58	0.60	17.35	$<2.2x10^{-16}$
	17.74	15.35	$7.60\ x10^{-14}$	13.84	$<2.2x10^{-16}$
	15.40	14.47	$3.71\ x10^{-4}$	11.67	$<2.2x10^{-16}$
	14.60	7.34	$<2.2x10^{-16}$	7.35	$<2.2x10^{-16}$

Salt bridges were also calculated (**Figure 15**), and their average value for the different interfaces were 11.86±3.97 for the STG-WT:AMPAR, 10.85±3.88 for the STG-ID:AMPAR, and 10.73±4.69 for the STG-SCZ:AMPAR. Although the effect was not as clear as the one achieved for hydrogen bonds, we observed that both mutations diminished the number of salt bridges as well as their occupancy along the MD simulations. In terms of salt bridges between residues of the first four beta-sheets and AMPAR, STG-WT showed a higher number of salt bridges spread from Lys43 to Lys53. In STG-ID this number was reduced for all interfaces with the exception of interface 4, in which the salt bridges tend to be formed preferably with the residues from near the end of the region. This facts explains why the beginning/middle region showed higher fluctuations in the RMSF analysis. On the other hand, the STG-SCZ was the system with the lowest number of salt bridges formed in the beta-sheets region corroborating the results obtained with the cross correlation analysis in which the region of the beta-sheets was found more affected within the STG-SCZ system.

The stability of the STG:AMPAR complex in all systems was further evaluated using the well stablished MMPBSA method that allowed us to calculated the $\Delta G_{binding}$ and $\Delta\Delta G_{binding}$ energy. It was calculated not only the contribution of each residue but also the sum of each substructure of AMPAR and STG (**Figure 16**). The mutated complexes showed slightly higher $\Delta G_{binding}$ values, and therefore fewer stable interactions. All possible four interfaces showed a less stable AMPAR_H1, STG-B1-4,

and STG-H4 in the mutated systems. The impact on STG-H5 was more variable, especially for the SCZ system.

The free binding energy was also applied for the individual residues (**Figure 17 and 18**) and the obtained values of the mutated systems were subtracted from the values of the STG-WT:AMPAR complex system ($\Delta\Delta G_{binding}$). Thus, positives results indicate that the residues are less important for the interaction between STG and AMPAR, compared to the STG-WT. Regarding the STG-ID:AMPAR complex (**Figure 17**), one residue from the first four beta-sheets (Val45), loop between β4 and the 2nd TMD (Glu90), 3rd TMD (Ile150, Ile151 and Ile153), 4th TMD (Val198) and C-terminal (Thr211) showed higher values in most interfaces. All these regions are known to be essential for the biological function of STG, and all, except the C-terminal, have been described as responsible for AMPAR binding[1,4]. On the other hand, some residues from the loop between the 2nd TMD and 3rd TMD (His132) and from the 4th TMD (His205 and Lys206) showed lower free binding energy values. The Glu545 at M1, the loop between M1 and Arg566 at M2 and M4 (Tyr818 and Leu824) presented higher values in most interfaces. Excepting Arg566, all residues are located in key regions for the interaction between STG and AMPAR. Oppositely, only Leu832 from M4 had lower energy levels in almost interfaces. Regarding the STG-SCZ:AMPAR complex (**Figure 18**), the loop residues between β4 and the 2nd TMD (Glu90), 3rd TMD (Ile140), 4th TMD (Phe180, Tyr181 and His205) and C-terminal (Thr211 and Arg213) showed values higher in most interfaces, while other residues in the same regions, such as His 132, Ile 143, Phe187, Lys 206 and Arg 209 were shown to have lower energy. The residues from the N-terminal (Lys530 and Lys532), the M1 (Glu545 and Met548), the loop between M1 and M2 (Arg 566, Glu 571 and TRP 572), the loop between M3 and M4 (Lys 720) and M4 (Phe 817 and Leu 832) showed higher energy in most interfaces. On the other hand, the loop between M1 and M2 (Glu587), Ile594 at M2 and the loop between M3 and M4 (Pro533, Val535 and Ser809) had lower energy.

The MMPBSA results, the number of hydrogen bonds and salt bridges, suggest that, for both STG mutations, the interaction between STG and AMPAR becomes less stable with a fewer contact network. Taking in to account all the results obtained for both the STG in monomeric form and for the STG in complex with AMPAR, both mutations seem to be able to alter the behavior of STG, with the most affected region being the first extracellular domain, more precisely in the region where the first four beta-sheets are located. The results also point to the fact that simple single mutations affected the stability of the interaction between AMPAR and STG by propagating their signal towards the interfaces.

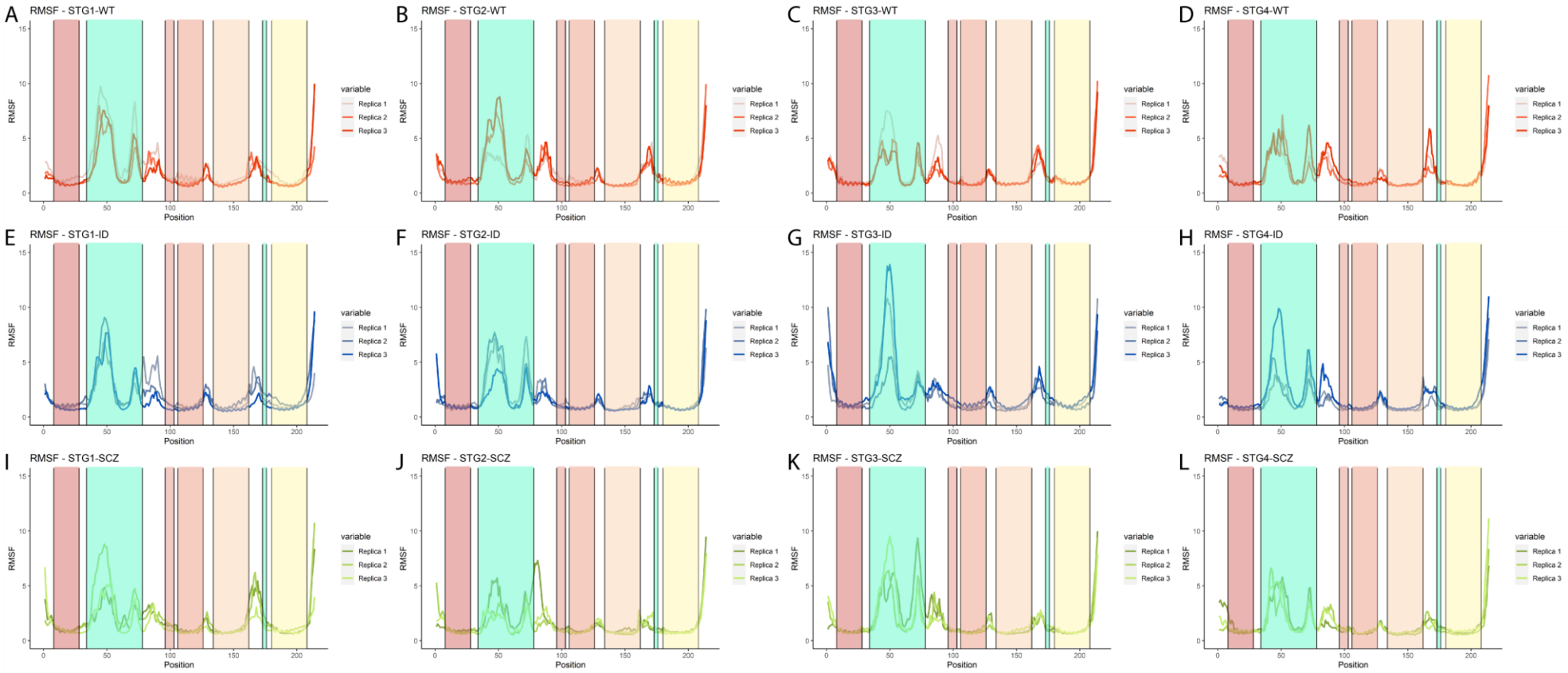

Figure 7 – RMSF results for each individual stargazin of the complex AMPAR:STG. (**A-D**) RMSF results of the wild type system, (**E-H**) ID system and (**I-L**) SCZ system with vertical bars representing each substructure of STG colored with the same colors as the same analysis done to the monomeric STG.

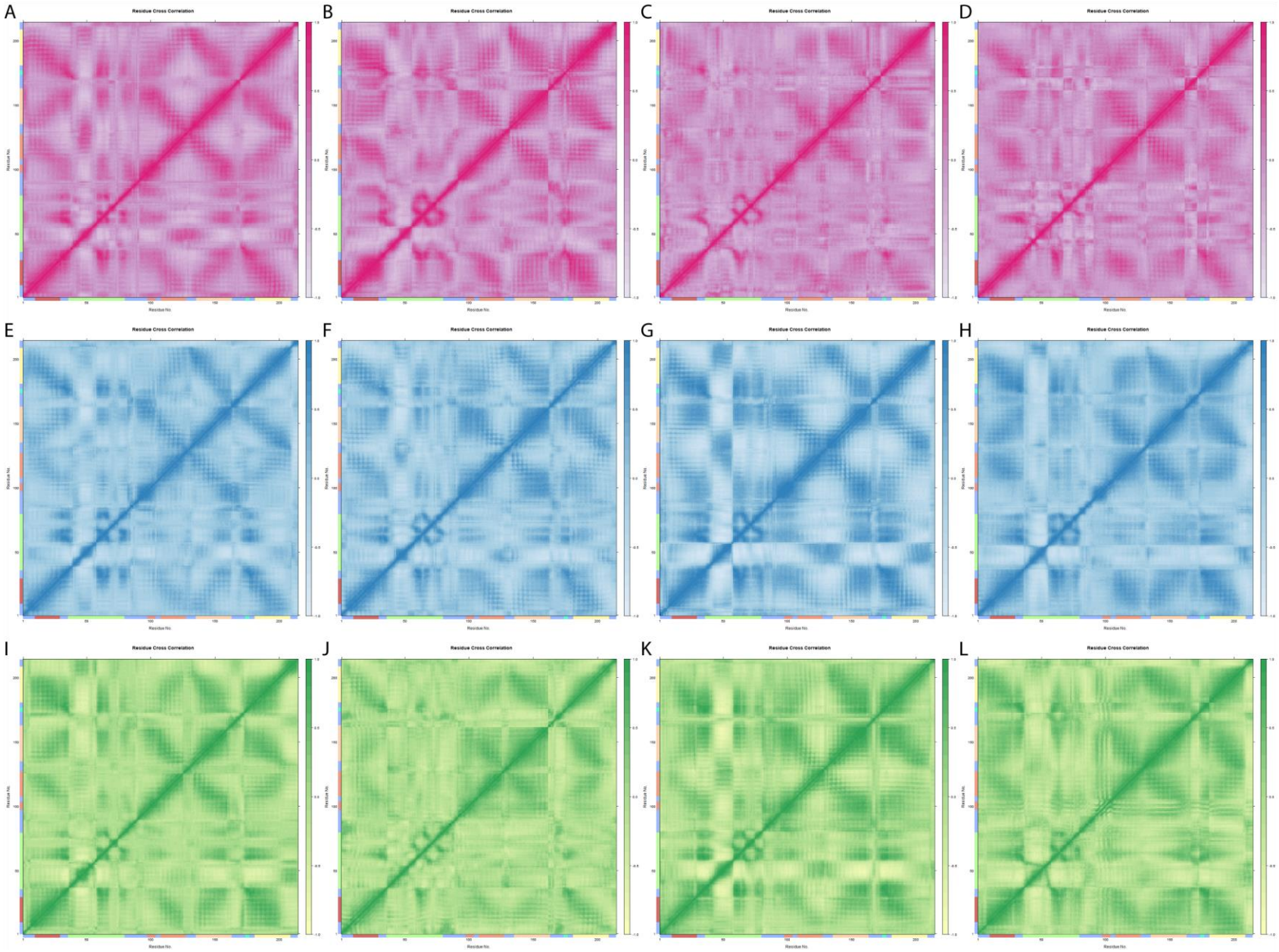

Figure 8 – Dynamical Cross Correlation Analysis for each individual STG from the complex AMPAR:STG. CCA results of the 4 STGs of wild type system (**A-D**), ID system (**E-H**) and SCZ system (**I-L**) with substructure annotation at the bottom and left side of each map.

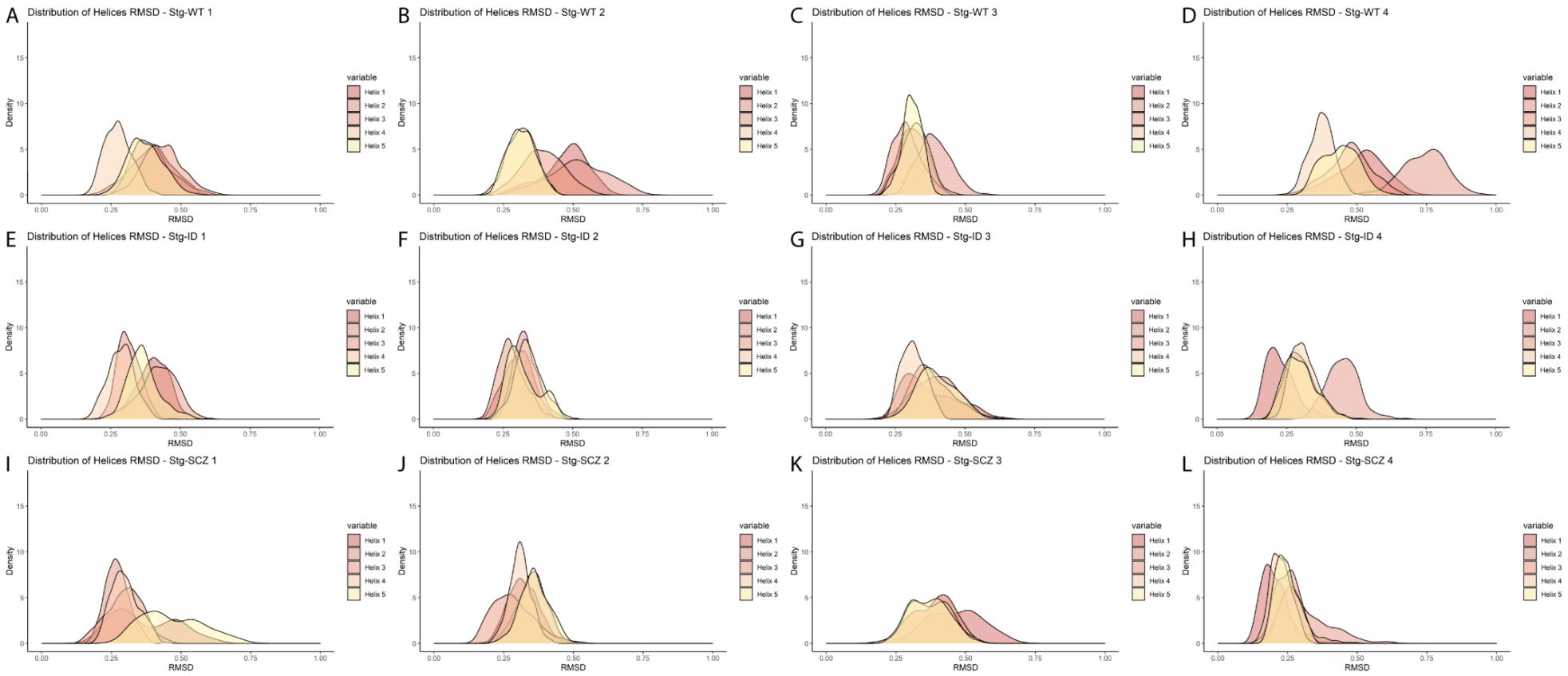

Figure 9 – RMSD density plots for Cα atoms of the helices of each individual stargazin from the AMPAR:STG complex. RMSD results of the wild type system (**A-D**), ID system (**E-H**) and SCZ system (**I-L**) colored using the colors of the previous figure.

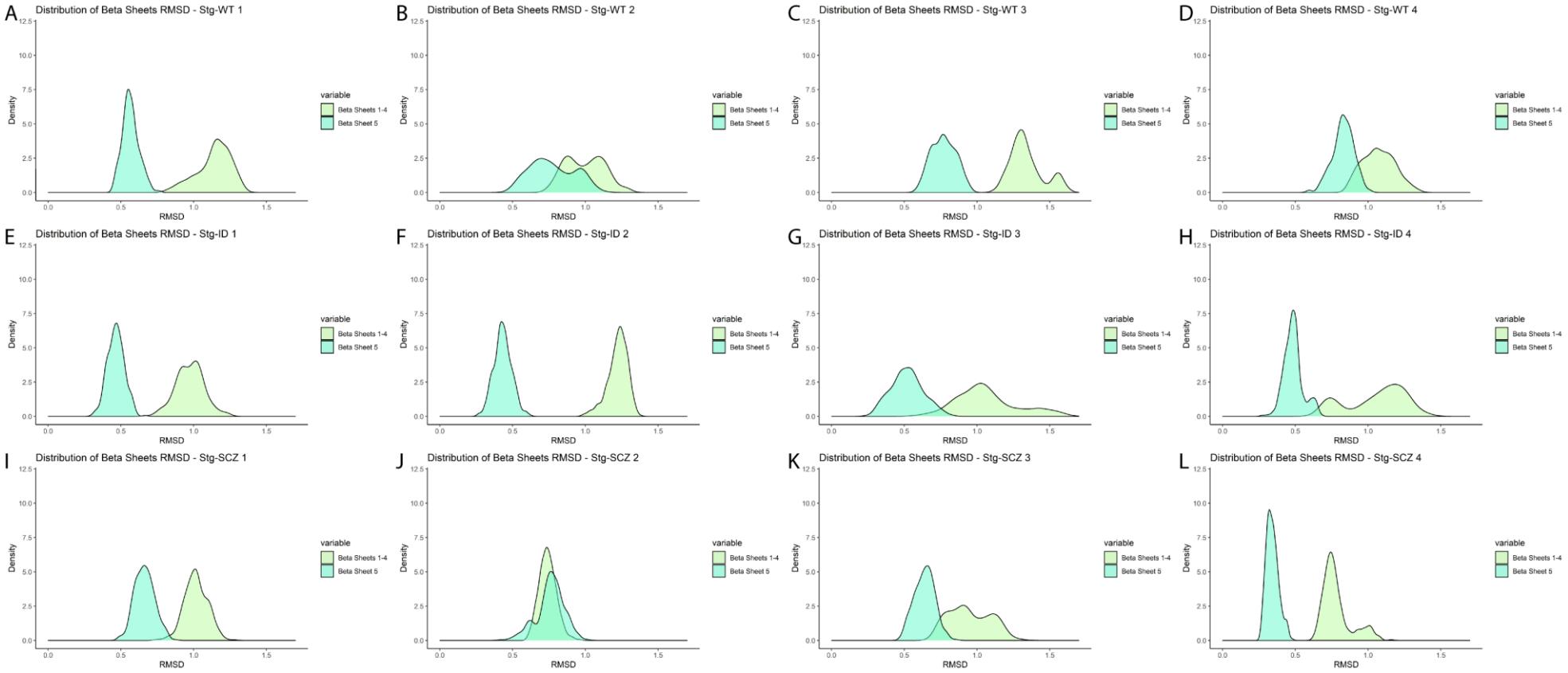

Figure 10 – RMSD density plots for Cα atoms of the beta-sheets of each individual stargazin from the AMPAR:STG complex. RMSD results of the wild type system (**A-D**), ID system (**E-H**) and SCZ system (**I-L**) colored using the colors of previous figures.

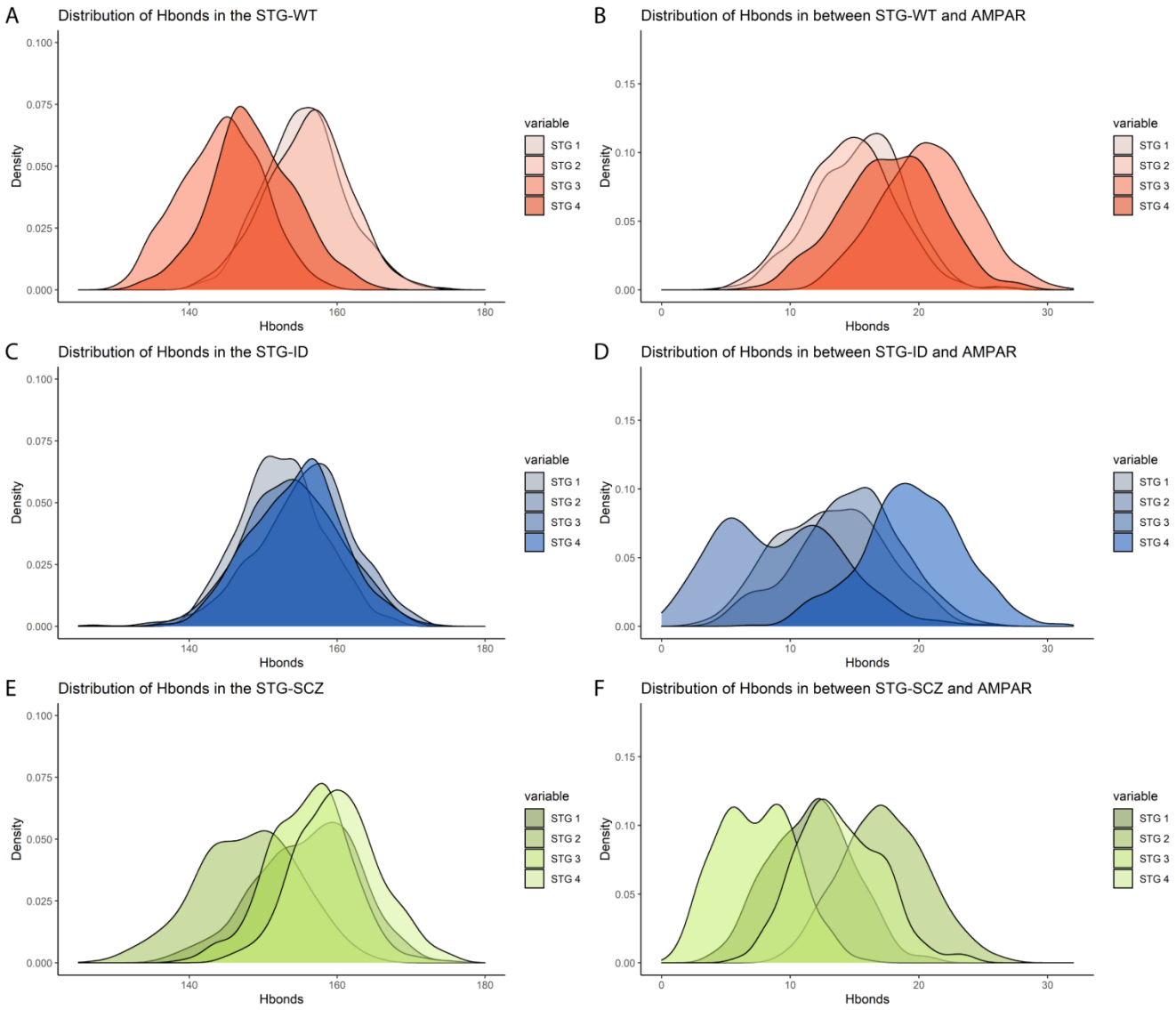

Figure 11 – Hydrogen bonds distribution for each STG of the AMPAR:STG complex. The intramolecular hydrogen bonds formed during the simulation where calculates for WT (**A**), ID (**C**) and SCZ system (**E**). Additionally the hydrogen bonds formed between STG and AMPAR during the simulation where also calculated for the WT system (**B**), ID (**D**) and SCZ system (**F**).

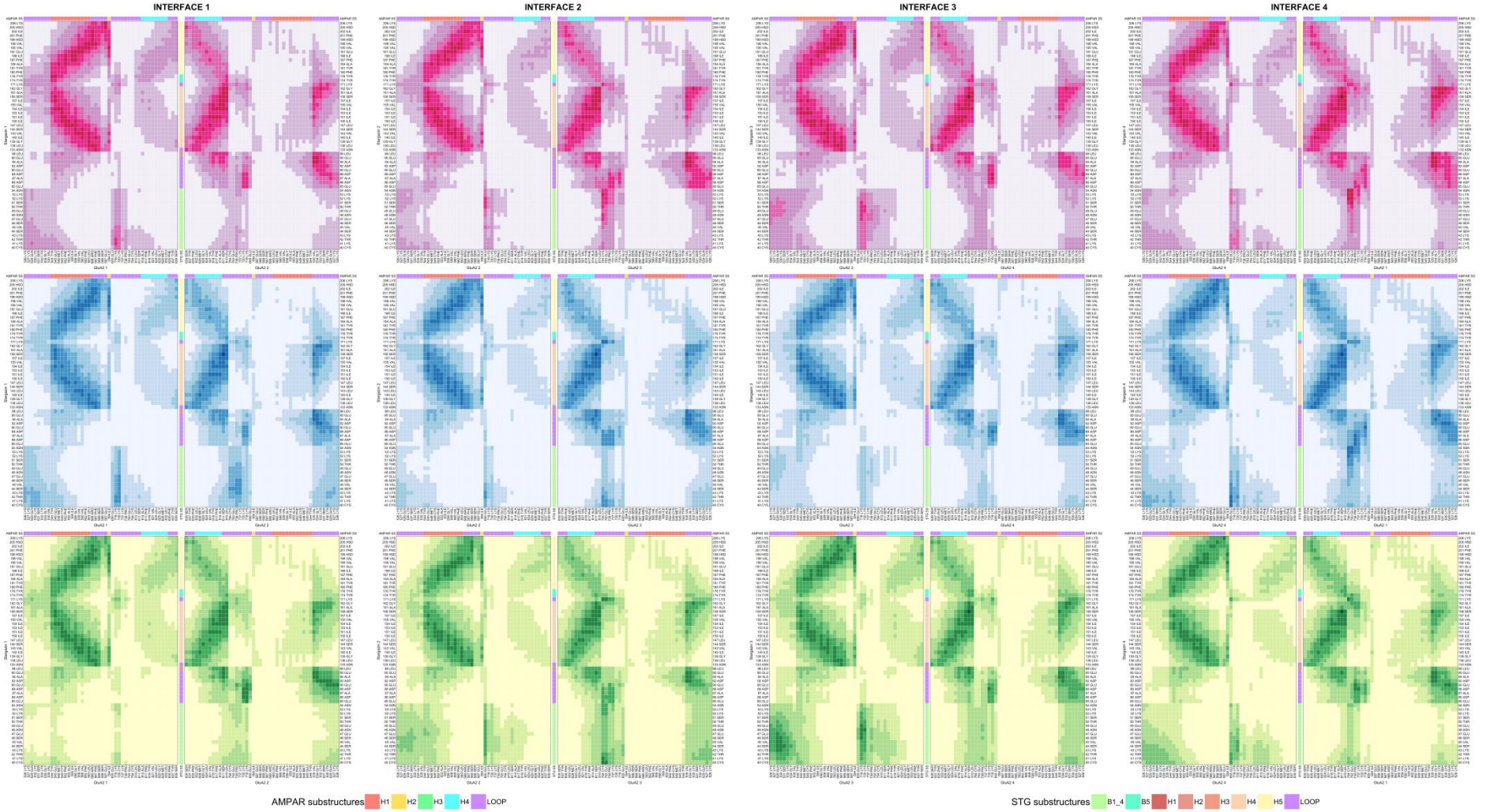

Figure 12 – Average distances between interface residues of STG and AMPAR. The average distances between residues present in the interface was calculated for the four possible interfaces present in the WT (in pink), ID (in blue) and SCZ system (in green).

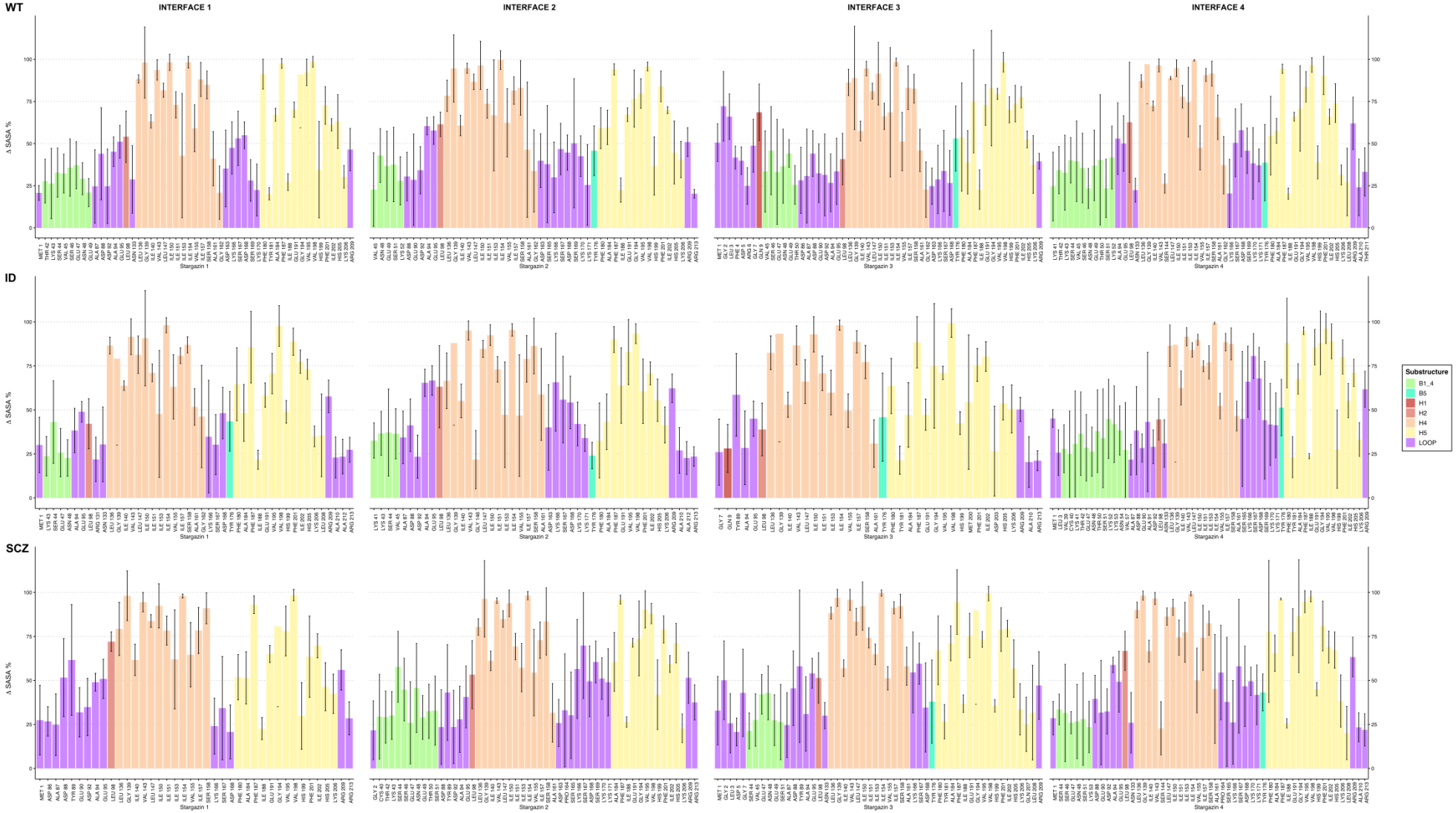

Figure 13 – Average Δ SASA values per residue of the four interfaces of STG. Δ SASA values for each system (WT, ID and SCZ) where calculated by subtracting the SASA values of the complex and the SASA value of the individual proteins (GluA2 and STG). In this figure only the Δ SASA of STG is represent, with the residues being colored according to what substructure they belong to.

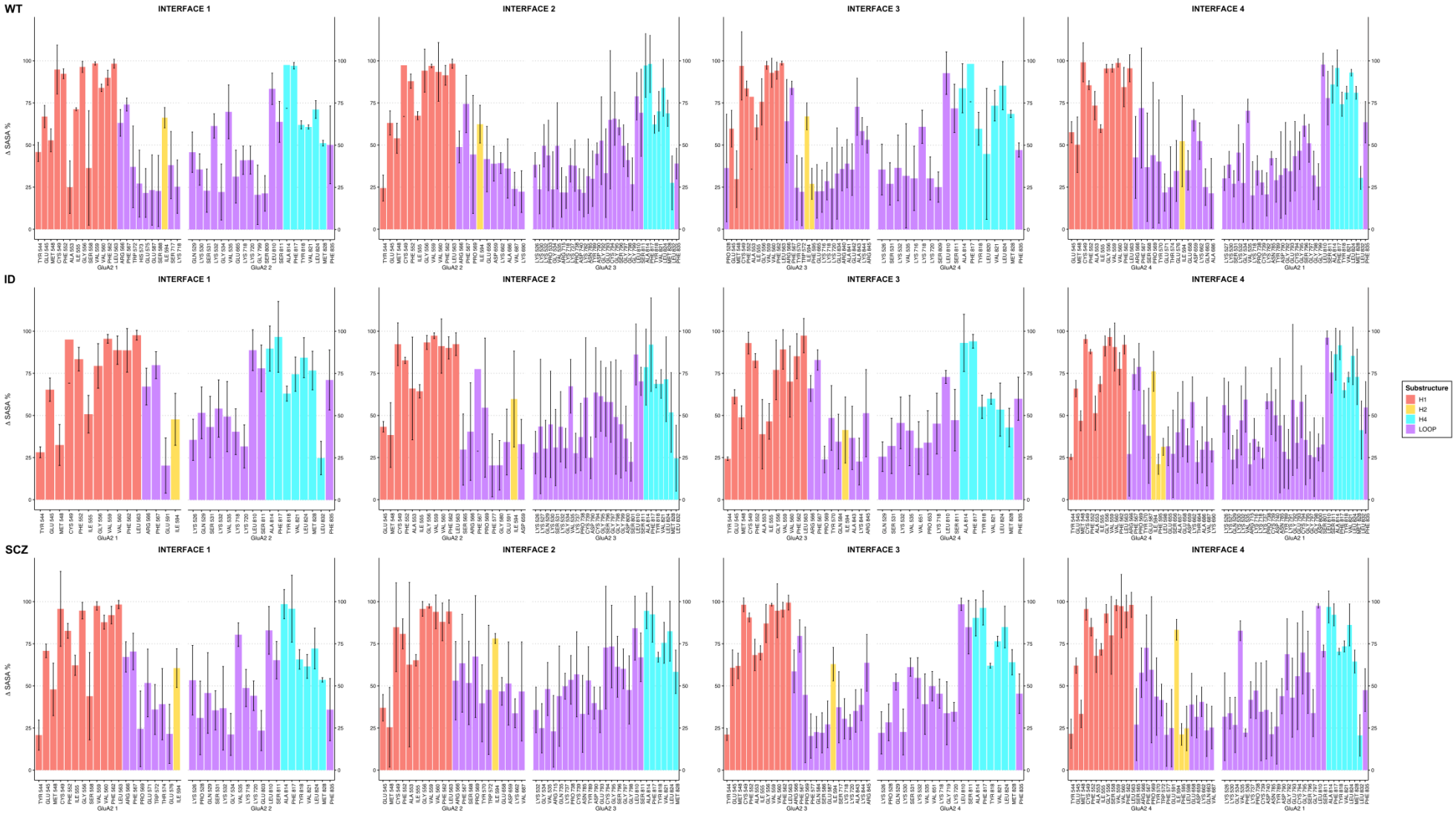

Figure 14 – Average Δ SASA values per residue of the four interfaces of AMPAR. Δ SASA values for each system (WT, ID and SCZ) where calculated by subtracting the SASA values of the complex and the SASA value of the individual proteins (GluA2 and STG). In this figure only the Δ SASA of AMPAR is represent, with the residues being colored according to what substructure they belong to.

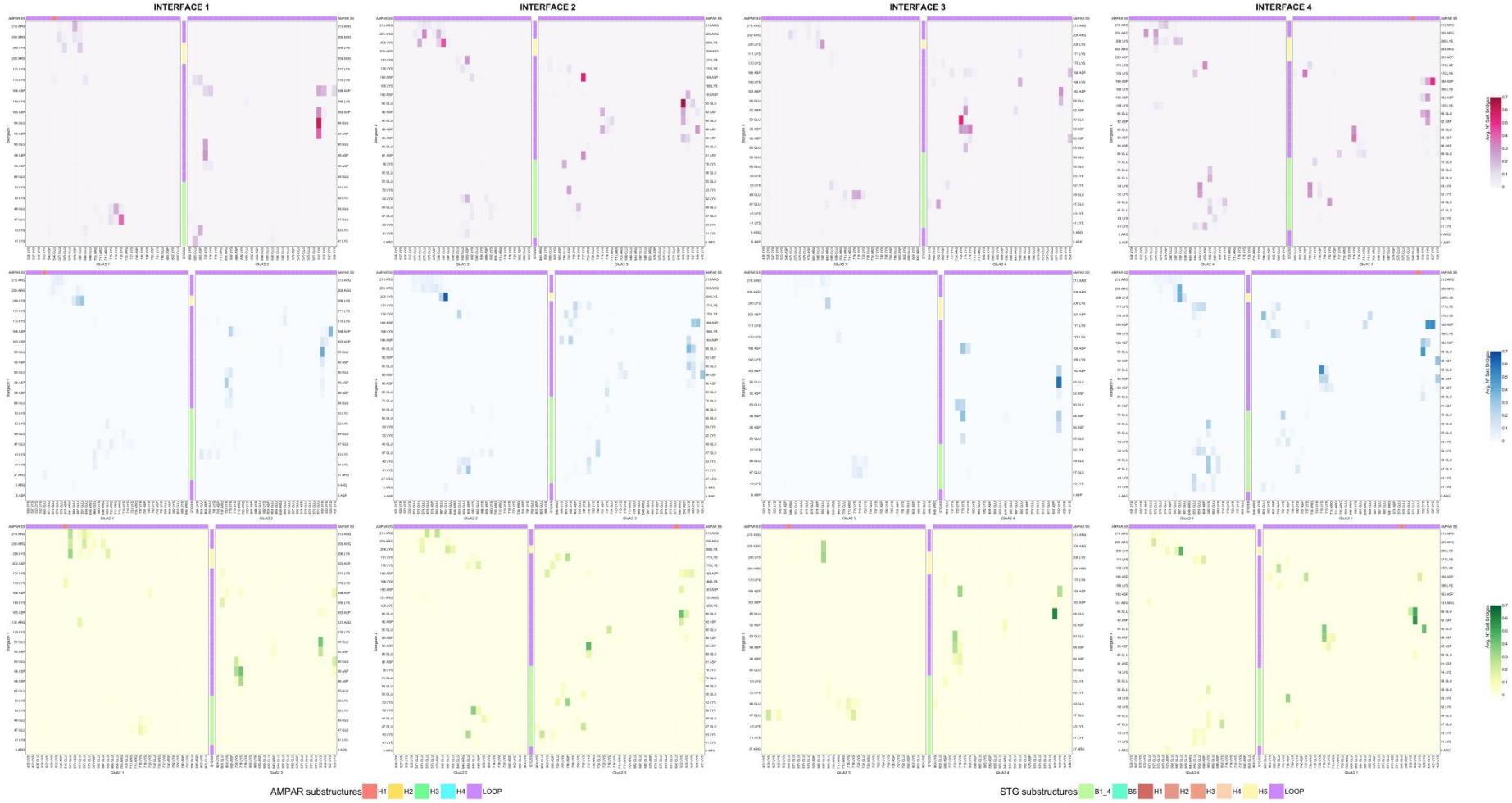

Figure 15 – Salt bridges between STG and AMPAR. The number of salt bridges and their occupancy along the simulation was calculated for of each interface in the WT (**pink**), ID (**blue**) and SCZ system (**green**). For an easier reading of this results substructure annotation was added at the top and between the two gluA2s that participate in each interface.

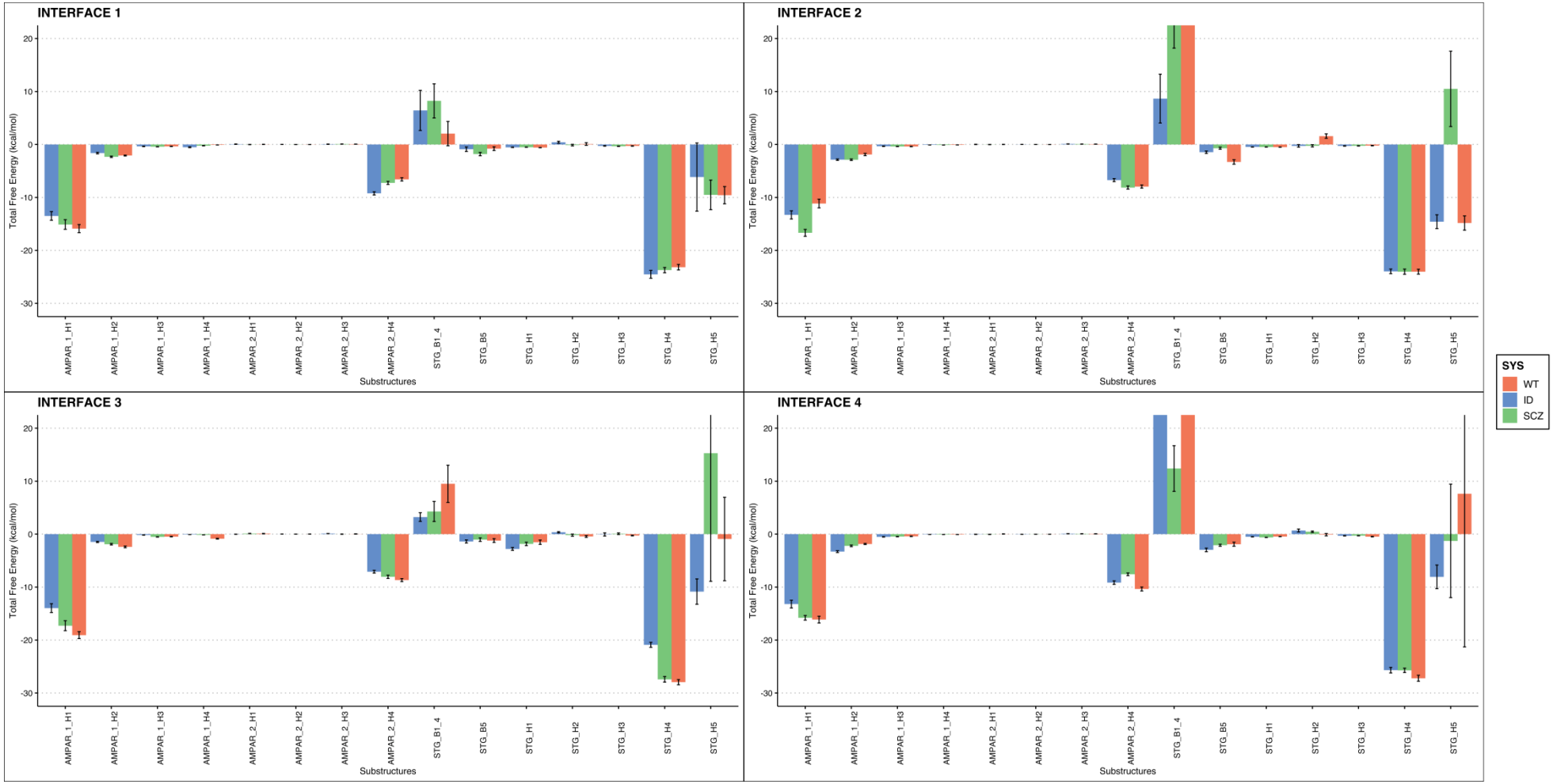

Figure 16 – ΔGbinding values for each secondary structure of STG and AMPAR. The average $\Delta G_{binding}$ values over the secondary structures of STG and AMPAR for the four interfaces of the complex was calculated using the MMPBSA method. The graphs were colored according to the system the values are representing.

Figure 17 – ΔΔGbinding values per residue of the four interfaces of STG. $\Delta\Delta G_{binding}$ values, obtained for the individual residues, were result of the subtraction of mutated systems and the WT system. On top the ◎◎Gbinding values of STG for the ID system can be observed, as well as on the bottom for the SCZ system. For an easier reading of the bars were colored according to which secondary structure the residues are part of.

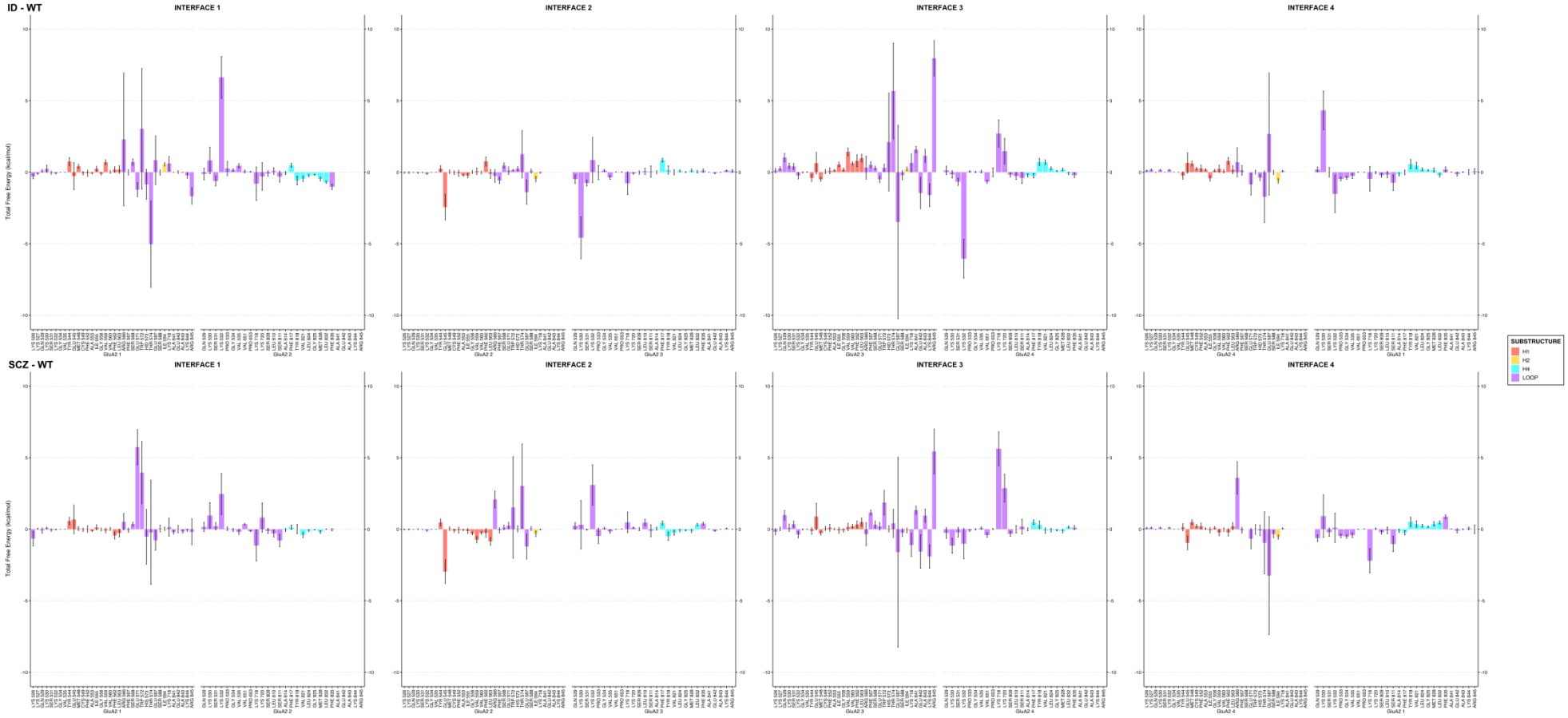

Figure 18 – ΔΔGbinding values per residue of the four interfaces of AMPAR. $\Delta\Delta G_{binding}$ values, obtained for the individual residues, were result of the subtraction of mutated systems and the WT system. On top the ΔΔGbinding values of STG for the ID system can be observed, as well as on the bottom for the SCZ system. For an easier reading of the bars were colored according to which secondary structure the residues are part of.

Conclusion

The main aim of this thesis was to study the effect of two mutations found in patients with psychiatric disorders on STG and its ability to interact with AMPAR. Through the multiple analysis performed, it was possible to conclude that both mutations have a large impact not only the first extracellular domain but also on the first, third and fourth TMDs. The results also point that the mutation found in a SCZ patient, S148N, could be more detrimental for the stability of the protein when compared to the other mutation found in an ID patient, V143L. Considering that the main interaction between STZ and AMPAR is conducted between third and fourth TMDs and M1 and M4 of GluA2, and the first extracellular domain also plays a role in the interaction[51], the overall results suggest that this mutation will difficult that ability of STG to bind with AMPAR.

In the second part of this work, the complex AMPAR:STG was studied to evaluate the effects of the point mutations on the overall complex structure and stability As already assessed for the monomeric STG, this mutation is still able to affect the same domains, first extracellular and the first, third and fourth TMD. However, the effect is not as striking as it was found for the unbound structure, which may indicate that AMPAR has some protective effects on STG after it binds to the receptor. Lastly, the effect of the mutations on the interaction between both proteins was evaluated, in which, initially, the region of the first extracellular domain was found to be the most affected by the mutations, especially with the V143L mutation. Considering that the extracellular domains are positioned closely to the ligand binding domain and the loop between that domain and TMD responsible to form the channel and it is able to regulate AMPAR function, it appears that the mutation V143L will affect more the ability of STG in regulating the activity of AMPAR. Nevertheless, not only by the decreased number of hydrogen bonds and salt bridges (as well as their occupancy), but also by calculating the free energy binding energies, it can be concluded that both mutations are able to destabilize the interaction between STG and AMPAR.

Altogether, this study contributes to a better understanding of how two point mutations in stargazing, found in patients with ID and increase SCZ, influence the binding of this protein to AMPAR, which can be ultimately responsible for the appearance of those disorders.

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (2013).

2. Trautmann, S., Rehm, J. & Wittchen, H. The economic costs of mental disorders. *EMBO Rep.* **17**, 1245–1249 (2016).

3. Kirov, G. *et al.* Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum. Mol. Genet.* **17**, 458–465 (2008).

4. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).

5. Itsara, A. *et al.* Population Analysis of Large Copy Number Variants and Hotspots of Human Genetic Disease. *Am. J. Hum. Genet.* **84**, 148–161 (2009).

6. Yizhar, O. *et al.* Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* **477**, 171–178 (2011).

7. Chelly, J., Khelfaoui, M., Francis, F., Che, B. & Bienvenu, T. Genetics and pathophysiology of mental retardation. *Eur. J. Hum. Genet.* **14**, 701–713 (2006).

8. What is Intellectual Disability? *American Phychiatric Association* (2013).

9. Srivastava, A. K. & Schwartz, C. E. Intellectual Disability and Autism Spectrum Disorders: Causal Genes and Molecular Mechanisms. *Neurosci. Biobehav. Rev.* **46**, 161–174 (2014).

10. Endele, S. *et al.* Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nat. Genet.* **42**, 1021–1026 (2010).

11. Wu, Y. *et al.* Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans. *PNAS* **104**, 18163–18168 (2007).

12. What is Schizophrenia? *American Phychiatric Association* (2013).

13. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a Complex Trait Evidence From a Meta-analysis of Twin Studies Patrick. *Arch Gen Psychiatry* **60**, 1187–1192 (2015).

14. Bhugra, D. The Global Prevalence of Schizophrenia. *PloS Med.* **2**, e151 (2005).

15. Os, J. Van, Kenis, G. & Rutten, B. P. F. The environment and schizophrenia. *Nature* **468**, 203–212 (2010).

16. Day, R. *et al.* Stressful life events preceding the acute onset of schizophrenia: a cross-national study from the world health organization. *Med. Psychiatry* **11**, 123–

205 (1987).

17. Managò, F. & Papaleo, F. Schizophrenia : What's Arc Got to Do with It ? *Front. Behav. Neurosci.* **11**, 1–11 (2017).

18. Robison, A. J. Emerging role of CaMKII in neuropsychiatric disease. *Trends Neurosci.* **1072**, 1–10 (2014).

19. Lisman, J. Excitation, inhibition, local oscillations, or large-scale loops: what causes the symptoms of schizophrenia? *Curr. Opin. Neurobiol.* **22**, 537–544 (2012).

20. Li, C. T., Yang, K. C. & Lin, W. C. Glutamatergic dysfunction and glutamatergic compounds for major psychiatric disorders: Evidence from clinical neuroimaging studies. *Front. Psychiatry* **9**, 1–11 (2019).

21. Fonnum, F. Glutamate: a neurotransmitter in the mammalian brain. *J. Neurochem.* **42**, 1–11 (1984).

22. Moretto, E., Murru, L., Martano, G., Sassone, J. & Passafaro, M. Glutamatergic synapses in neurodevelopmental disorders. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **84**, 328–342 (2017).

23. Okabe, S. Molecular anatomy of the postsynaptic density. *Mol. Cell. Neurosci.* **34**, 503–518 (2007).

24. Iacobucci, G. J. & Popescu, G. K. NMDA Receptors: Linking Physiological Output to Biophysical Operation. *Nat. Rev. Neurosci.* **18**, 236–249 (2017).

25. Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53**, 531–556 (2013).

26. Eichel, K. *et al. Catalytic activation of β-arrestin by GPCRs*. *Nature* vol. 557 (2018).

27. Uematsu, K. *et al.* Protein kinase A directly phosphorylates metabotropic glutamate receptor 5 to modulate its function. *J. Neurochem.* **132**, 677–686 (2015).

28. Niswender, C. M. & Conn, P. J. Metabotropic glutamate receptors: Physiology, pharmacology, and disease. *Annu. Rev. Pharmacol. Toxicol.* **50**, 295–322 (2010).

29. Nasrallah, C. *et al.* Direct coupling of detergent purified human mGlu5 receptor to the heterotrimeric G proteins Gq and Gs. *Sci. Rep.* **8**, 1–13 (2018).

30. Cartmell, J. & Schoepp, D. D. Regulation of neurotransmitter release by metabotropic glutamate receptors. *J. Neurochem.* **75**, 889–907 (2000).

31. Huber, K. M., Roder, J. C. & Bear, M. F. Chemical induction of mGluR5- and protein synthesis-dependent long-term depression in hippocampal area CA1. *J. Neurophysiol.* **86**, 321–325 (2001).

32. Shigemoto, R., Abe, T., Nomura, S., Nakanishi, S. & Hirano, T. Antibodies inactivating mGluR1 metabotropic glutamate receptor block long-term

depression in cultured Purkinje cells. *Neuron* **12**, 1245–1255 (1994).

33. Paoletti, P., Bellone, C. & Zhou, Q. NMDA receptor subunit diversity : impact on receptor properties , synaptic plasticity and disease. *Nat. Rev.* **14**, 383–400 (2013).

34. Massey, P. V *et al.* Differential Roles of NR2A and NR2B-Containing NMDA Receptors in Cortical Long-Term Potentiation and Long-Term Depression. *J. Neurosci.* **24**, 7821–7828 (2004).

35. Penn, A. C., Balik, A., Wozny, C., Cais, O. & Greger, I. H. Activity-Mediated AMPA Receptor Remodeling, Driven by Alternative Splicing in the Ligand-Binding Domain. *Neuron Rep.* **76**, 503–510 (2012).

36. Hollmann, M. & Heinemann, S. Cloned Glutamate Receptors. *Annu. Rev. Neurosci.* **17**, 31–108 (1994).

37. Seeburg, P. H., Higuchi, M. & Sprengel, R. RNA editing of brain glutamate receptor channels : mechanism and physiology. *Brain Res. Rev.* **26**, 217–229 (1998).

38. Sommer, B. *et al.* Flip and Flop : A Cell-Specific Functional Switch in Glutamate-Operated Channels of the CNS. *Science (80-.).* **249**, 1580–1585 (1989).

39. Santos, S. D., Carvalho, A. L., Caldeira, M. V. & Duarte, C. B. Regulation of AMPA receptors and synaptic plasticity. *Neuroscience* **158**, 105–125 (2009).

40. Greger, I. H., Watson, J. F. & Cull-candy, S. G. Structural and Functional Architecture of AMPA-Type Glutamate Receptors and Their Auxiliary Proteins. *Neuron Rev.* **94**, 713–730 (2017).

41. Hammond, J. C., Mccullumsmith, R. E., Funk, A. J., Haroutunian, V. & Meador-woodruff, J. H. Evidence for Abnormal Forward Trafficking of AMPA Receptors in Frontal Cortex of Elderly Patients with Schizophrenia. *Neuropsychopharmacology* **35**, 2110–2119 (2010).

42. Schnell, E. *et al.* Direct interactions between PSD-95 and stargazin control synaptic AMPA receptor number. *PNAS* **99**, 13902–13907 (2002).

43. Straub, C. & Tomita, S. The regulation of glutamate receptor trafficking and function by TARPs and other transmembrane auxiliary subunits. *Curr. Opin. Neurobiol.* **22**, 488–495 (2012).

44. Tomita, S. *et al.* Stargazin modulates AMPA receptor gating and trafficking by distinct domains. *Nature* **435**, 1052–1058 (2005).

45. Jackson, A. C. & Nicoll, R. A. The Expanding Social Network of Ionotropic Glutamate Receptors: TARPs and Other Transmembrane Auxiliary Subunits. *Neuron* **70**, 178–199 (2011).

46. Soto, D. *et al.* Selective regulation of long-form calcium-permeable AMPA

receptors by a novel TARP. *Nat. Neurosci.* **12**, 277–285 (2009).

47. Tomita, S. *et al.* Functional studies and distribution define a family of transmembrane AMPA receptor regulatory proteins. *J. Cell Biol.* **161**, 805–816 (2003).

48. Kato, A. S. *et al.* New Transmembrane AMPA Receptor Regulatory Protein Isoform , gamma -7 , Differentially Regulates AMPA Receptors. *J. Neurosci.* **27**, 4969–4977 (2007).

49. Kato, A. S., Siuda, E. R., Nisenbaum, E. S. & Bredt, D. S. AMPA Receptor Subunit-Specific Regulation by a Distinct Family of Type II TARPs. *Neuron* **59**, 986–996 (2008).

50. Letts, V. A. Stargazer—A Mouse to Seize! *Epilepsy Curr.* **5**, 161–165 (2005).

51. Twomey, E. C., Yelshanskaya, M. V., Grassucci, R. A., Frank, J. & Sobolevsky, A. I. Elucidation of AMPA receptor-stargazin complexes by cryo- electron microscopy. *Science (80-.).* **353**, 83–86 (2016).

52. Shaikh, S. A. *et al.* Stargazin Modulation of AMPA Receptors. *Cell Rep.* **17**, 328–335 (2016).

53. Ben-Yaacov, A. *et al.* Molecular Mechanism of AMPA Receptor Modulation by TARP/Stargazin. *Neuron* **93**, 1126-1137.e4 (2017).

54. Kevin Range, and D. M. Y. A. M. The α-Amino-3-hydroxyl-5-methyl-4-isoxazolepropionate Receptor Trafficking Regulator "Stargazin" Is Related to the Claudin Family of Proteins by Its Ability to Mediate Cell-Cell Adhesion. *J. Biol. Chem.* **280**, 19711–19720 (2005).

55. Vandenberghe, W., Nicoll, R. A. & Bredt, D. S. Stargazin is an AMPA receptor auxiliary subunit. *Proc. Natl. Acad. Sci.* **102**, 485–490 (2005).

56. Sumioka, A., Yan, D. & Tomita, S. TARP phosphorylation regulates synaptic AMPA receptors through lipid bilayers. *Neuron* **66**, 755–767 (2010).

57. Pedersen, S. W. *et al.* Site-Specific Phosphorylation of PSD-95 PDZ Domains Reveals Fine-Tuned Regulation of Protein-Protein Interactions. *ACS Chem. Biol.* **12**, 2313–2323 (2017).

58. Tomita, S., Nicoll, R. A. & Bredt, D. S. PDZ protein interactions regulating glutamate receptor function and plasticity. *J. Cell Biol.* **153**, 19–23 (2001).

59. Choi, J. *et al.* Phosphorylation of Stargazin by Protein Kinase A Regulates Its Interaction with PSD-95. *J. Biol. Chem.* **277**, 12359–12363 (2002).

60. Dakoji, S., Tomita, S., Karimzadegan, S., Nicoll, R. A. & Bredt, D. S. Interaction of transmembrane AMPA receptor regulatory proteins with multiple membrane associated guanylate kinases. *Neuropharmacology* **45**, 849–856 (2003).

61. Chen, L. *et al.* Stargazin regulates synaptic targeting of AMPA receptors by two distinct mechanisms. *Nature* **408**, 936–943 (2000).

62. Deng, F., Price, M. G., Davis, C. F., Mori, M. & Burgess, D. L. Stargazin and Other Transmembrane AMPA Receptor Regulating Proteins Interact with Synaptic Scaffolding Protein MAGI-2 in Brain. *J. Neurosci.* **26**, 7875–7884 (2006).

63. Mei, L. & Borg, J. ERBB2 oncogenicity : ERBIN helps to perform the job. *Mol. Cell. Oncol.* **2**, e995033 (2015).

64. Cuadra, A. E., Kuo, S. H., Kawasaki, Y., Bredt, D. S. & Chetkovich, D. M. AMPA Receptor Synaptic Targeting Regulated by Stargazin Interactions with the Golgi-Resident PDZ Protein nPIST. *J. Neurosci.* **24**, 7491–7502 (2004).

65. Tao, Y. *et al.* Erbin interacts with TARP γ-2 for surface expression of AMPA receptors in cortical interneurons. *Nat. Neurosci.* **16**, 290–299 (2013).

66. Ives, J. H., Fung, S., Tiwari, P., Payne, H. L. & Thompson, C. L. Microtubule-associated protein light chain 2 is a stargazin-AMPA receptor complex-interacting protein in vivo. *J. Biol. Chem.* **279**, 31002–31009 (2004).

67. Ma, T. M. *et al.* Serine racemase regulated by binding to stargazin and PSD-95: Potential N-methyl-D-aspartate-α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (NMDA-AMPA) glutamate neurotransmission cross-talk. *J. Biol. Chem.* **289**, 29631–29641 (2014).

68. Zhang, W. *et al.* Structural Basis of Arc Binding to Synaptic Proteins: Implications for Cognitive Disease. *Neuron* **86**, 490–500 (2015).

69. Twomey, E. C. *et al.* Mechanisms of channel block in calcium-permeable AMPA receptors. *Neuron* **99**, 956–968 (2018).

70. Benesh, J. L., Mueller, T. M. & Meador-Woodruff, J. H. AMPA receptor subunit localization in schizophrenia anterior cingulate cortex. *Schizophr. Res.* 1–9 (2020) doi:10.1016/j.schres.2020.01.025.

71. Yang, H. *et al.* The DAO Gene Is Associated with Schizophrenia and Interacts with Other Genes in the Taiwan Han Chinese Population. *PLoS One* **8**, e60099 (2013).

72. Liu, Y. L. *et al.* RASD2, MYH9, and CACNG2 Genes at Chromosome 22q12 Associated with the Subgroup of Schizophrenia with Non-Deficit in Sustained Attention and Executive Function. *Biol. Psychiatry* **64**, 789–796 (2008).

73. Miranda, A. *et al.* Study of 45 candidate genes suggests CACNG2 may be associated with lithium response in bipolar disorder. *J. Affect. Disord.* **248**, 175–179 (2019).

74. Nissen, S., Liang, S., Shehktman, T. & Kelsoe, J. R. Evidence for association of bipolar disorder to haplotypes in the 22q12.3 region near the genes stargazin,

ift27 and parvalbumin. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **159**, 941–950 (2012).

75. Hamdan, F. F. *et al.* Excess of De Novo Deleterious Mutations in Genes Associated with Glutamatergic Systems in Nonsyndromic Intellectual Disability. *Am. J. Hum. Genet.* **88**, 306–316 (2011).

76. Drummond, J. B., Tucholski, J., Haroutunian, V. & Meador-Woodruff, J. H. Transmembrane AMPA receptor regulatory protein (TARP) dysregulation in anterior cingulate cortex in schizophrenia Jana. *Schizophr. Res.* **147**, 32–38 (2013).

77. Mount, D. W. *Bioinformatics: Sequence and Genome Analysis*. (Cold Spring Harbor Laboratory Press, 2004).

78. Chua, S. & Foo, L. K. Tree Alignment Based on Needleman-Wunsch Algorithm for Sensor Selection in Smart Homes. *sensors* **17**, 1–15 (2017).

79. Al-Karadaghi, S. Introduction to Protein Sequence Alignment and Analysis. https://proteinstructures.com/Sequence/sequence-analysis.html.

80. Chojnacki, S., Cowley, A., Lee, J., Foix, A. & Lopez, R. Programmatic access to bioinformatics tools from EMBL-EBI update : 2017. *Nucleic Acids Res.* **45**, 550–553 (2017).

81. Duggan, K. Clustal Omega. https://www.ebi.ac.uk/seqdb/confluence/display/THD/Clustal+Omega. (2018).

82. Sievers, F. *et al.* Fast , scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 1–6 (2011).

83. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

84. Goossens, K. & Winter, H. D. Molecular Dynamics Simulations of Membrane Proteins : An Overview. *J. Chem. Inf. Model.* **58**, 2193–2202 (2018).

85. Hospital, A., Goñi, J. R., Orozco, M. & Gelpí, J. L. Molecular dynamics simulations: Advances and applications. *Adv. Appl. Bioinforma. Chem.* **8**, 37–47 (2015).

86. Al-Karadaghi, S. Introduction to Homology Modeling. https://proteinstructures.com/Modeling/Modeling/homology-modeling.html.

87. Pitman, M. R. & Menz, R. I. Methods for Protein Homology Modelling. in *Applied Mycology and Biotechnology* 37–59 (2006).

88. Cristobal, S., Zemla, A., Fischer, D., Rychlewski, L. & Elofsson, A. A study of quality measures for protein threading models. *BMC Bioinformatics* **2**, (2001).

89. 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).

90. Shen, M. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 (2006).

91. 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).

92. Waterhouse, A. *et al.* SWISS-MODEL : homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, 296–303 (2018).

93. Bienert, S. *et al.* The SWISS-MODEL Repository — new features and functionality. *Nucleic Acids Res.* **45**, 313–319 (2017).

94. Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L. & Schwede, T. Modeling protein quaternary structure of homo- and hetero- oligomers beyond binary interactions by homology. *Sci. Rep.* **7**, 1–15 (2017).

95. Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Struct. Bioinforma.* **27**, 343–350 (2011).

96. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss- PdbViewer : A historical perspective. *Electrophoresis* **30**, 162–173 (2009).

97. Sali, A. & Blundell, T. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol* **234**, 779–815 (1993).

98. Sali, A. MODELLER A Program for Protein Structure Modeling. https://salilab.org/modeller/manual/.

99. Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **41**, 349–357 (2013).

100. Jones, D. T. Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices. *J. Mol. Biol.* **292**, 195–202 (1999).

101. Lobley, A., Sadowski, M. I. & Jones, D. T. pGenTHREADER and pDomTHREADER : new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics* **25**, 1761–1767 (2009).

102. Mcguffin, L. J. & Jones, D. T. Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics* **19**, 874–881 (2002).

103. Nugent, T. & Jones, D. T. Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics* **10**, 1–11 (2009).

104. Mcguffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).

105. Allen, M. P. Introduction to Molecular Dynamics Simulation. **23**, 1–28 (2004).

106. Alford, R. F., Smolin, N., Young, H. S., Gray, J. J. & Robia, S. L. Protein docking and steered molecular dynamics suggest alternative phospholamban-binding sites on the SERCA calcium transporter. *J. Biol. Chem.* **295**, 11262–11274 (2020).

107. Im, W. & Brooks, C. L. Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6771–6776 (2005).

108. Dutagaci, B., Heo, L. & Feig, M. Structure refinement of membrane proteins via molecular dynamics simulations. *Proteins Struct. Funct. Bioinforma.* **86**, 738–750 (2018).

109. Kargar, F., Emadi, S. & Fazli, H. Dimerization of Aβ40 inside dipalmitoylphosphatidylcholine bilayer and its effect on bilayer integrity: Atomistic simulation at three temperatures. *Proteins Struct. Funct. Bioinforma.* **88**, 1540–1552 (2020).

110. Dhusia, K., Su, Z. & Wu, Y. Understanding the Impacts of Conformational Dynamics on the Regulation of Protein-Protein Association by a Multiscale Simulation Method. *J. Chem. Theory Comput.* **16**, 5323–5333 (2020).

111. Radwan, A. & Mahrous, G. M. Docking studies and molecular dynamics simulations of the binding characteristics of waldiomycin and its methyl ester analog to Staphylococcus aureus histidine kinase. *PLoS One* **15**, 1–16 (2020).

112. Goose, J. E. & Sansom, M. S. P. Reduced Lateral Mobility of Lipids and Proteins in Crowded Membranes. *PLoS Comput. Biol.* **9**, (2013).

113. Fagnen, C. *et al.* New Structural insights into Kir channel gating from molecular simulations, HDX-MS and functional studies. *Sci. Rep.* **10**, 1–14 (2020).

114. Jenkins, K. *et al.* Combining data integration and molecular dynamics for target identification in α-Synuclein-aggregating neurodegenerative diseases: Structural insights on Synaptojanin-1 (Synj1). *Comput. Struct. Biotechnol. J.* **18**, 1032–1042 (2020).

115. Corey, R. A., Stansfeld, P. J. & Sansom, M. S. P. The energetics of protein-lipid interactions as viewed by molecular simulations. *Biochem. Soc. Trans.* **48**, 25–37 (2020).

116. Niesen, M. J. M., Zimmer, M. H. & Miller, T. F. Dynamics of Co-translational Membrane Protein Integration and Translocation via the Sec Translocon. *J. Am. Chem. Soc.* **142**, 5449–5460 (2020).

117. Lee, Y., Lazim, R., Macalino, S. J. Y. & Choi, S. Importance of protein dynamics

in the structure-based drug discovery of class A G protein-coupled receptors (GPCRs). *Curr. Opin. Struct. Biol.* **55**, 147–153 (2019).

118. Harpole, T. J. & Delemotte, L. Conformational landscapes of membrane proteins delineated by enhanced sampling molecular dynamics simulations. *Biochim. Biophys. Acta - Biomembr.* **1860**, 909–926 (2018).

119. Venko, K., Roy Choudhury, A. & Novič, M. Computational Approaches for Revealing the Structure of Membrane Transporters: Case Study on Bilitranslocase. *Comput. Struct. Biotechnol. J.* **15**, 232–242 (2017).

120. Ferreira, R. J. & Kasson, P. M. Antibiotic Uptake across Gram-Negative Outer Membranes: Better Predictions towards Better Antibiotics. *ACS Infect. Dis.* **5**, 2096–2104 (2019).

121. Loschwitz, J., Olubiyi, O. O., Hub, J. S., Strodel, B. & Poojari, C. S. *Computer simulations of protein–membrane systems*. *Progress in Molecular Biology and Translational Science* vol. 170 (Elsevier Inc., 2020).

122. Zheng, L., Alhossary, A. A., Kwoh, C. K. & Mu, Y. Molecular dynamics and simulation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* vol. 2 550–566 (2019).

123. Shiref, H., Bergman, S., Clivio, S. & Sahai, M. A. The fine art of preparing membrane transport proteins for biomolecular simulations: Concepts and practical considerations. *Methods* 1–12 (2020) doi:10.1016/j.ymeth.2020.02.009.

124. Mustafa, G., Nandekar, P. P., Mukherjee, G., Bruce, N. J. & Wade, R. C. The Effect of Force-Field Parameters on Cytochrome P450-Membrane Interactions: Structure and Dynamics. *Sci. Rep.* **10**, 7284 (2020).

125. Marrink, S. J. *et al.* Computational Modeling of Realistic Cell Membranes. *Chem. Rev.* **119**, 6184–6226 (2019).

126. Sandoval-Perez, A., Pluhackova, K. & Böckmann, R. A. Critical Comparison of Biomembrane Force Fields: Protein-Lipid Interactions at the Membrane Interface. *J. Chem. Theory Comput.* **13**, 2310–2321 (2017).

127. Srivastava, A., Nagai, T., Srivastava, A., Miyashita, O. & Tama, F. Role of computational methods in going beyond x-ray crystallography to explore protein structure and dynamics. *Int. J. Mol. Sci.* **19**, 3401 (2018).

128. Dhingra, S., Sowdhamini, R., Cadet, F. & Offmann, B. A glance into the evolution of template-free protein structure prediction methodologies. *Biochimie* **175**, 85–92 (2020).

129. Doktorova, M. & Weinstein, H. Accurate In Silico Modeling of Asymmetric Bilayers Based on Biophysical Principles. *Biophys. J.* **115**, 1638–1643 (2018).

130. Sanders, C. R. & Mittendorf, K. F. Tolerance to Changes in Membrane Lipid

Composition as a Selected Trait of Membrane Proteins. *Biochemistry* **50**, 7858–7867 (2011).

131. Wassenaar, T. A., Ingólfsson, H. I., Böckmann, R. A., Tieleman, D. P. & Marrink, S. J. Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.* **11**, 2144–2155 (2015).

132. Wolf, M. G., Hoefling, M., Aponte-SantaMaría, C., Grubmüller, H. & Groenhof, G. g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *J. Comput. Chem.* **31**, 2169–2174 (2010).

133. Jo, S., Kim, T., Iyer, V. G. & Im, W. Software News and Updates CHARMM-GUI : A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).

134. Wu, E. L., Cheng, X., Jo, S., Rui, H. & Song, K. C. CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. *J. Comput. Chem.* 1–8 (2014) doi:10.1002/jcc.23702.

135. Brooks, B. R. *et al.* CHARMM : The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545–1614 (2009).

136. Additive, C. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).

137. Jo, S. *et al.* CHARMM-GUI 10 Years for Biomolecular Modeling and Simulation. *J Comput Chem* **38**, 1114–1124 (2017).

138. Jo, S., Lim, J. B., Klauda, J. B. & Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **97**, 50–58 (2009).

139. Lee, J. *et al.* Biomolecular Systems CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theory Comput.* (2018) doi:10.1021/acs.jctc.8b01066.

140. Jo, S., Kim, T. & Im, W. Automated Builder and Database of Protein / Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2**, e880 (2007).

141. Gupta, C. & Mertz, B. Protonation Enhances the Inherent Helix-Forming Propensity of pHLIP. *ACS Omega* **2**, 8536–8542 (2017).

142. Pallante, L. *et al.* In silico Investigations of the Mode of Action of Novel Colchicine Derivatives Targeting β-Tubulin Isotypes: A Search for a Selective and Specific β-III Tubulin Ligand. *Front. Chem.* **8**, 1–8 (2020).

143. Im, W. & Roux, B. Ion permeation and selectivity of OmpF porin: A theoretical study based on molecular dynamics, Brownian dynamics, and continuum

electrodiffusion theory. *J. Mol. Biol.* **322**, 851–869 (2002).

144. Gupta, C., Ren, Y. & Mertz, B. Cooperative Nonbonded Forces Control Membrane Binding of the pH-Low Insertion Peptide pHLIP. *Biophys. J.* **115**, 2403–2412 (2018).

145. Mertz, B., Feng, J., Corcoran, C. & Neeley, B. Explaining the mobility of retinal in activated rhodopsin and opsin. *Photochem. Photobiol. Sci.* **14**, 1952–1964 (2015).

146. Liu, Y., Haddadian, E., Sosnick, T. R., Freed, K. F. & Gong, H. A novel implicit solvent model for simulating the molecular dynamics of RNA. *Biophys. J.* **105**, 1248–1257 (2013).

147. Laradji, M., Sunil Kumar, P. B. & Spangler, E. J. Exploring large-scale phenomena in composite membranes through an efficient implicit-solvent model. *J. Phys. D. Appl. Phys.* **49**, 1–16 (2016).

148. Kellici, T. F., Ntountaniotis, D., Liapakis, G., Tzakos, A. G. & Mavromoustakos, T. The dynamic properties of angiotensin II type 1 receptor inverse agonists in solution and in the receptor site. *Arab. J. Chem.* **12**, 5062–5078 (2019).

149. Mizuhara, Y., Parkin, D., Umezawa, K., Ohnuki, J. & Takano, M. Over-Destabilization of Protein-Protein Interaction in Generalized Born Model and Utility of Energy Density Integration Cutoff. *J. Phys. Chem. B* **121**, 4669–4677 (2017).

150. Ulmschneider, M. B., Ulmschneider, J. P., Sansom, M. S. P. & Di Nolay, A. A generalized born implicit-membrane representation compared to experimental insertion free energies. *Biophys. J.* **92**, 2338–2349 (2007).

151. Ulmschneider, J. P. & Ulmschneider, M. B. Folding simulations of the transmembrane Helix of virus protein U in an implicit membrane model. *J. Chem. Theory Comput.* **3**, 2335–2346 (2007).

152. Latek, D. *et al.* Modeling of Membrane Proteins: From Bioinformatics to Molecular Quantum Mechanics. in *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes* 371–451 (2019). doi:10.1007/978-3-642-28554-7_12.

153. Han, X. *et al. Study of the variability of the native protein structure*. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* vol. 3 (Elsevier Ltd., 2019).

154. Inakollu, V. S., Geerke, D. P., Rowley, C. N. & Yu, H. Polarisable force fields: what do they add in biomolecular simulations? *Curr. Opin. Struct. Biol.* **61**, 182–190 (2020).

155. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2**,

19–25 (2015).

156. Páll, S., Abraham, M. J., Kutzner, C., Hess, B. & Lindahl, E. Tackling exascale software challenges in molecular dynamics simulations with GROMACS. *Solving Softw. challenges exascale* 3–27 (2015).

157. Phillips, J. C. *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 044130 (2020).

158. Case, D. A. *et al. Amber2020*. (2020).

159. Patodia, S., Bagaria, A. & Chopra, D. Molecular Dynamics Simulation of Proteins : A Brief Overview. *J. Phys. Chem. Biophys.* **4**, 1–4 (2014).

160. Moisan, M. & Pelletier, J. Appendix I Some Properties of the Maxwell-Boltzman (M-B) Velocity Distribution. in *Physics of Collisional Plasmas* 387–477 (2012). doi:10.1007/978-94-007-4558-2.

161. Schneider, R., Sharma, A. R. & Rai, A. Introduction to Molecular Dynamics. in *Computational Many-Particle Physics* vol. 739 3–40 (2008).

162. Casalini, T. Fundamentals and application of modeling in support of spinal cord injury repair strategies. in *Spinal Cord Injury (SCI) Repair Strategies* 279–306 (Elsevier Ltd, 2020). doi:10.1016/B978-0-08-102807-0.00014-4.

163. Hernández-Rodríguez, M., C. Rosales-Hernández, M., E. Mendieta-Wejebe, J., Martínez-Archundia, M. & Correa Basurto, J. Current Tools and Methods in Molecular Dynamics (MD) Simulations for Drug Design. *Curr. Med. Chem.* **23**, 3909–3924 (2016).

164. Doktorova, M., LeVine, M. V., Khelashvili, G. & Weinstein, H. A New Computational Method for Membrane Compressibility: Bilayer Mechanical Thickness. *Biophys. J.* **116**, 487–502 (2019).

165. Ribeiro, J. V. *et al.* QwikMD - Integrative Molecular Dynamics Toolkit for Novices and Experts. *Sci. Rep.* **6**, 1–14 (2016).

166. Yamada, Y. *et al.* VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

167. Abraham, M. J., Spoel, D. V., Lindahl, E., Hess, B. & . the G. development team. Gromacs User Manual Version 2019-beta1.

168. Bekker, H. *et al.* Gromacs: A parallel computer for molecular dynamics simulations. *Phys. Comput.* **92**, 252–256 (1993).

169. Berendsen, H. J. C., Spoel, D. van der & Drunen, R. van. GROMACS: A message-passing parallel molecular dynamics implementation. *Phys. Comput.* **91**, 43–56 (1995).

170. Lindahl, E., Hess, B. & Spoel, D. van der. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317 (2001).

171. Spoel, D. van der *et al.* GROMACS: Fast, Flexible and Free. *J. Comput. Chem.* **26**, 1701–1718 (2005).

172. Pronk, S. *et al.* GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).

173. Saitoh, Y. *et al.* Structural insight into tight junction disassembly by Clostridium perfringens enterotoxin. *Science (80-.).* **347**, 775–778 (2015).

174. The UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

175. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).

176. Sippl, M. J. Recognition of errors in three-dimensional structures of proteins. *Proteins Struct. Funct. Bioinforma.* **17**, 355–362 (1993).

177. Wallner, B. & Elofsson, A. Can correct protein models be identified? *Protein Sci.* **12**, 1073–1086 (2003).

178. L DeLano, W. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **40**, 82–92 (2002).

179. Abraham, M. J. *et al. GROMACS User Manual version 2018.3*. (2018).

180. Huang, J. & Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).

181. Magalhães, P. R. *et al.* Dynamical rearrangement of human epidermal growth factor receptor 2 upon antibody binding: Effects on the dimerization. *Biomolecules* **9**, (2019).

182. 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).