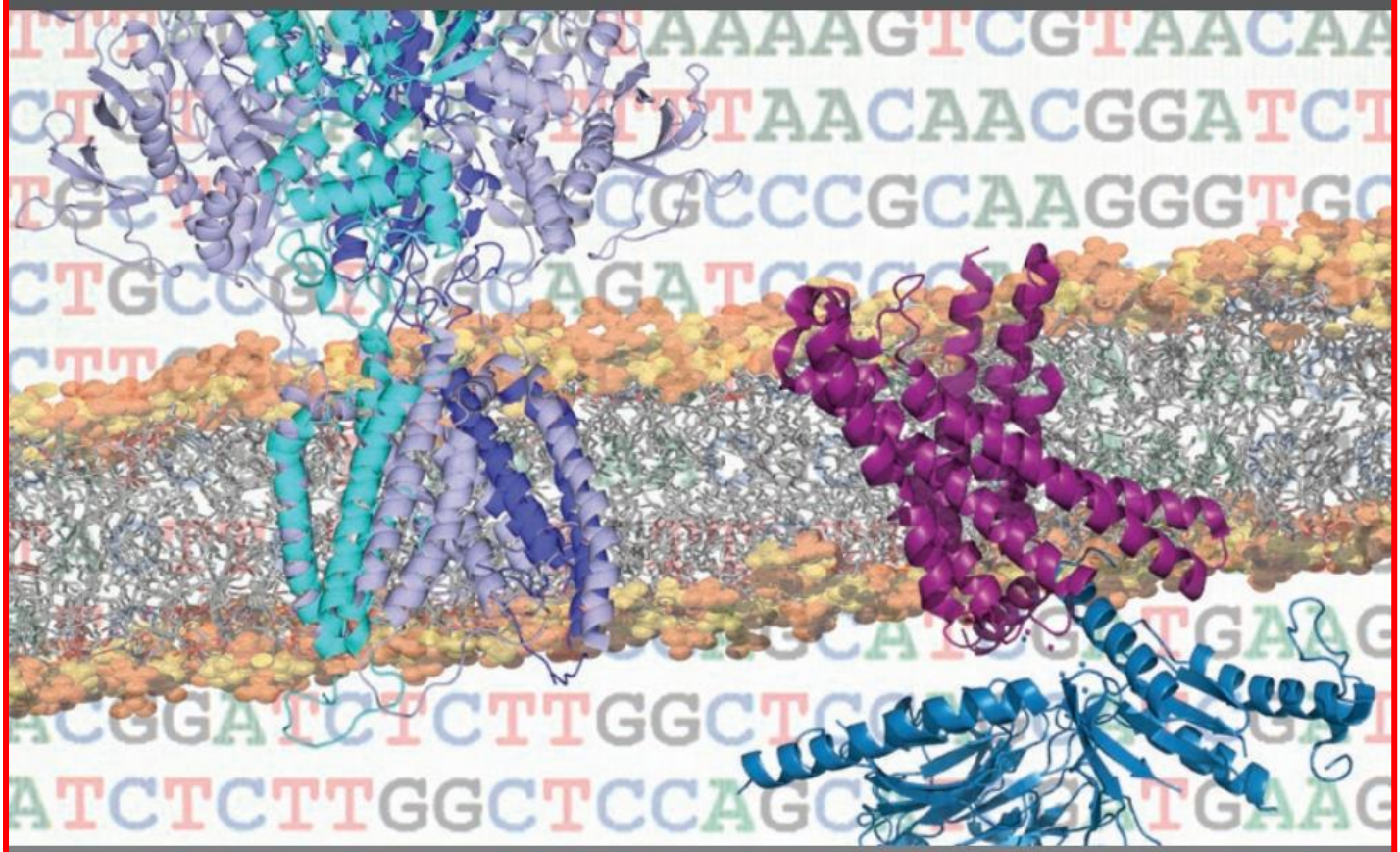
UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL DE CATALUNYA

DOCTORAL THESIS

Bioinformatics tools for membrane proteins: from sequences to structure and function

Adrián García Recio



Supervisors:
Dr. Arnau Cordomí Montoya and Dr. Mireia Olivella García

2021

UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL DE CATALUNYA
Escola de Doctorat

DOCTORAL THESIS

Bioinformatics tools for membrane proteins: from sequences to structure and function

Adrián García Recio

Supervisors: Dr. Arnau Cordomí Montoya and
Dr. Mireia Olivella García

PhD program in Bioinformatics

2021

Abstract

Membrane proteins (MPs) are a large group of proteins that play an essential role in the cell. This group includes receptors, ionic channels, transporters, and enzymes, accounting for 25% of proteins of the human genome. About 50% of the membrane proteins are pharmacological targets for various diseases. Moreover, there are pathogenic mutations that affect their folding, stability, and function reported for 90% of membrane proteins. The research on membrane proteins has grown along the last years together with the development of computational tools to deal with sequence, structural and functional data have become essential to understand all the information available on these proteins.

The main objective of this thesis is to develop Bioinformatics tools to model and analyze membrane proteins. In parts of the thesis, we have studied all membrane proteins, whereas in others we have focused on either N-methyl-D-aspartate receptors (NMDARs) or the G-protein-coupled receptors (GPCRs) families. Mutations with a certain negative impact in the function of these proteins may cause a malfunctionality that may eventually derive into a disease. In regard to these mutations, various predictors of the grade of pathogenicity are available but none of them is specific for membrane proteins. Because membrane proteins have different properties than globular proteins, we aimed to develop a tool, (TMSNP, available at http://lmc.uab.cat/tmsnp), that brings a searchable database of pathogenic and non-pathogenic mutations in transmembrane segments of the membrane proteins and provide a mutation prediction tool able to predict pathogenicity of non-reported transmembrane missense mutations.

In the field of NMDARs, because of the amount of sequence proteins that contain mutated variants, the stratification of uncertain annotated mutations from all available data have become crucial to develop personalized therapy. For this reason, we developed a public updated, manually curated, and nonredundant database, (GRINdb available at http://lmc.uab.cat/grindb), that comprises genetic, clinical, functional, and structural information about the largest repertoire of GRIN varied, with more than 4400 unique GRIN variant entries.

In the field of GPCRs, some computational scientists, molecular biologists and crystallographers are interested in finding the best possible template to model GPCRs. These models can be later used, for instance, to run molecular dynamics simulations. In regard to GPCRs, this thesis has developed two computational tools: a web application for rapid, intuitive and systematic visualization of the complete repertoire of crystallographic interfaces for GPCR homodimers available, (DIMERBOW, available at http://lmc.uab.cat/dimerbow), and a web application to introduce internal water molecules in GPCR structures using resolved water molecules from homologous structures (HomolWat, available at http://lmc.uab.cat/HW).

The general tools developed in this thesis as well as those specific for NMDARs and GPCRs, may help to better understand and characterize membrane proteins.

Acknowledgments

There were a lot of people involved in this thesis and I want to thank all of them. First of all, I want to thank two amazing persons that have been the pillars of this thesis and a big part of my scientific career: Mireia Olivella and Arnau Cordomí. Without their supervision, opportunities, and support, I could not have gotten to where I am right now. They will always have my respect and affection because they have been like parents in the academic world, teaching me, being sincere at all times and always trying to look out for my good.

Also, I would like to thank all co-authors of the papers presented as part of this thesis, and specially to Eduardo Mayol, Ramón Guixà, and Jose Carlos Gómez. I have learned a lot from you, and I am really grateful to you for it. I would also like to thank the entire LMC group at the UAB for having treated me like one more, especially Claudia, Nil, Marc, and Ramón for those good memories in the laboratory and in all the workshops and congresses we have attended. Furthermore, to the community of GRINs, the group of Xavier Altafaj, and the patients with their families that showed me to never surrender.

Besides, those who do not participate directly in the thesis, but have been an unconditional support at my side during this time and also important contributors to this work. To my childhood friends, Albert, Alonso, Javier and Sergi, I know that even though I was immersed in my doctorate all these years, they have always supported me, and are a great pillar in my life. My friends from the university, especially, Aina, Alberto, Ariadna, Dani, Mireia and Jordi, thank you for taking an interest in my work and for all the good times that we have had and will pass. Also, to all my other friends, but especially to Cristina and Cristian for sending me encouragement and trying to make me smile.

Last but not least, my whole family. Especially to my parents, José and Isabel, for teaching me and training me as the person I am today and for facilitating the way to develop professionally. I do not have enough words to be grateful to them. Finally, I could not also especially leave a person who has made the world easier for me, has endured my character in bad times and has had to put up with me. She is one of the reasons why I have come to the end of this thesis, my life partner Sara.

Part of this thesis also belongs to all of you. Thank you!

“If you have knowledge,
let others light their candles in it”.
~ Margaret Fuller

Abbreviations

ATD	Amino-terminal domain
ATP	Adenosine triphosphate
CG	Coarse-grained
CTD	C-terminal domain
GoF	Gain-of-function
GPCR	G-protein-coupled receptor
GRD	GRIN-related disorders
LBD	Ligand-binding domain
LoF	Loss-of-function
MD	Molecular dynamics
MP	Membrane protein
MSA	Multiple sequence alignment
NMDA	N-methyl-D-aspartate
NTD	N-terminal domain
OPM	Orientations of Proteins in Membranes
PDB	Protein data bank
RDBMS	Relational database management system
SNP	Single nucleotide polymorphism

TM	Transmembrane
TMD	Transmembrane domain

Table of Contents

Acknowledgments 5

Abbreviations 7

Chapter 1 – Introduction 11

1.1 Dogma of molecular biology 12

1.2 Proteins 13

1.2.1 Primary structure 14

1.2.2 Secondary structure 14

1.2.3 Supersecondary structure 16

1.2.4 Tertiary structure and domains 16

1.2.5 Quaternary structure 17

1.2.6 Proteins functions 19

1.2.7 Pathogenicity of mutations in proteins 20

1.2.8 Classification of proteins 23

1.3 The cell membrane 24

1.3.1 Structure and composition of the cell membrane 25

1.3.2 Membrane proteins 27

1.3.3 Specific features of membrane proteins 29

1.3.5 Disease and therapy in membrane proteins 30

1.4 N-methyl-D-aspartate receptors 30

1.4.1 Structure of the NMDAR 31

1.4.2 NMDAR signalling 33

1.4.3 NMDAR pathogenesis 35

1.4.4 Strategic therapy to modulate NMDAR activity 36

1.5 G Protein-Coupled Receptors 37

1.5.1 GPCR signaling and activation 38

1.5.2 GPCRs classification 40

1.5.3 GPCR oligomerization 43

1.5.4 GPCRs diseases and therapeutic strategy 45

Chapter 2 - Objectives 47

Chapter 3 - Methods 49

3.1 Bioinformatics and biomedical databases 50

3.2 Sequence alignment methods 52

3.3 Structural Bioinformatics 53

3.4 Programming languages, databases, and web applications 54

Chapter 4 - Results 57

4.1 List of articles derived from this thesis 58

4.2 Overview 59

4.3 Articles 61

Article I. TMSNP: a web server to predict pathogenesis of missense mutations in the transmembrane region of membrane protein 61

Article II. GRIN database: A unified and manually curated repertoire of GRIN variants 74

Article III. Identification of homologous GluN subunits variants accelerates GRIN variants stratification 102

Article IV. Disease-associated GRIN protein truncating variants trigger NMDA receptor loss-of-function 123

Article V. HomolWat: a web server tool to incorporate ‘homologous’ water molecules into GPCR structures 152

Article VI. DIMERBOW: exploring possible GPCR dimer interfaces 165

Chapter 5 - Global discussion 171

Chapter 6 - Conclusions 179

Bibliography 181

Chapter 1 – Introduction

Chapter 1 - Introduction

1.1 Dogma of molecular biology

The central dogma of molecular biology suggests that there is a unidirectionality in the expression of the information contained in the genes of a cell, that is, that DNA is transcribed as messenger RNA and that it is translated as protein, an element that finally carries out a cellular action. Also postulates that only DNA can duplicate itself and therefore reproduce and transmit genetic information to offspring. The power of this idea was originally developed by Francis Crick in 1957 (Crick, 1970). However, science is a dynamic field, and some studies amplify now the version of the original dogma (see Figure 1.1-1). The general transfers explain the standard flux of biological information as the DNA can be replicated to DNA, DNA information can be transcribed into mRNA, and proteins can be synthesized translating the information into mRNA. The special transfers detail that RNA is being replicated from RNA, DNA is being synthesized using a RNA template, and proteins are being synthesized directly from a DNA template without the use of mRNA. The rest of the possibilities, which are unknown transfers, are not thought to naturally occur.

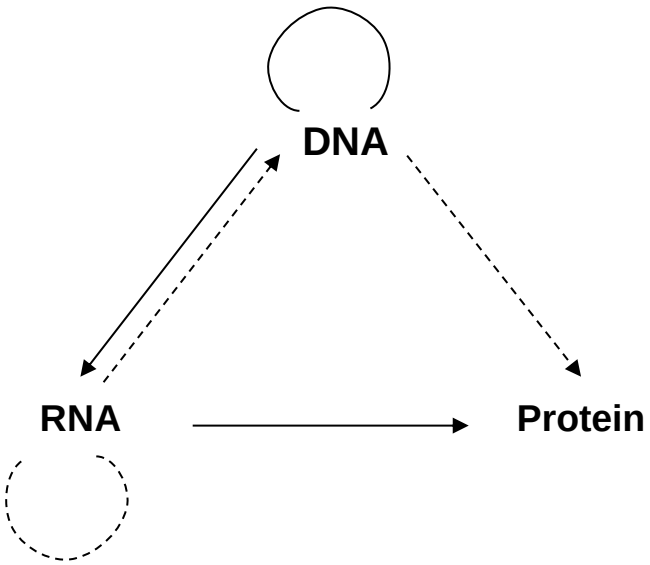


Fig 1.1-1. Central dogma of molecular biology. Continuous lines describe the original proposal of the "dogma" and the general processing of the genetic information. The non-continuous lines refer to the special cases discovered with next studies from Crick. Adapted from (Cobb, 2017).

1.2 Proteins

Proteins are the executors of the activity of the cells. These are macromolecules that account for most of the cell's dry mass (Alberts et al., 2018) and represent one of the main building blocks of cells. In the human genome there are more than 21.000 protein-coding genes (Pertea et al., 2018; Salzberg, 2018). In a period that has been predominant by Structural Biology for many years, a change of focus towards sequence analysis has pushed the advent of the genome projects and the resultant divergent sequence/structure deficit. The central challenge of Computational Structural Biology is therefore to rationalize the mass of sequence information into biochemical and biophysical knowledge and to decipher the structural, functional and evolutionary clues encoded in the language of biological sequences (Shenoy & Jayaram, 2010). The significance of the unique protein sequence, known as the protein's primary structure, dictates the 3D conformation of the folded protein. As Anfinsen postulates in his dogma, the environmental conditions at which folding occurs, the native structure is a unique, stable, and kinetically accessible minimum of the free energy (Anfinsen, 1973). This conformation will totally determine the function of the protein (Rajagopal, K. A., Indira, & Tan, T., 2021).

Proteins are involved in different functions in the cell such as catalyzing chemical reactions, synthesizing and repairing DNA, transporting materials across the cell, receiving and sending chemical signals, responding to stimulus, providing structural support, among others (*Biological Macromolecules*, 2020). To carry out this range of functions, proteins have specific sequences and structural properties. Specially, the amino acids sequences of proteins have been evolutive selected for its capacity to rapid adaptation to new selection pressures (*Biological Macromolecules*, 2020; Reynolds et al., 2013). Proteins fold into stable three-dimensional shapes, or conformations, that are determined by their amino acid sequence. The complete structure of a protein can be described at four different levels of complexity: primary, secondary, tertiary, and quaternary structure (see the Figure 1.2-2) (Sun et al., 2004).

1.2.1 Primary structure

The **primary structure** of a protein is defined as the sequence of amino acids linked together forming a polypeptide chain (Alberts et al., 2018). The sequence is determined by the translation of the coding regions of DNA that preserve the code to synthesize proteins. In order to create a specific primary protein structure sequence, twenty different amino acids can be used multiple times (see Figure 1.2-1). Each amino acid contains a common backbone of a carboxylic acid group and an amino group linked by an α-carbon atom but differs in the sidechain that branches from the α-carbon atom and provides specific physicochemical properties. Some side chains can be either acidic or basic, while others can be polar uncharged or non-polar. These characteristics allow understanding into whether the protein generally functions better in acidic or basic context, solubility in water or lipids, the temperature range for optimal protein function, and which segment of the protein are found on the protein interior being in contact with the external aqueous environment. Some amino acids included within the polypeptide chain can even create ionic bonds and disulfide bridges. The position of certain amino acids in the primary structure determines the secondary, tertiary, and quaternary structures (Sanvictores & Farci, 2020).

1.2.2 Secondary structure

The **secondary structure** relates to the 3-dimensional version of local folded configuration that forms within a polypeptide due to interactions between atoms of the backbone. The concept of the secondary structure was introduced by Linus Pauling and Robert Corey in 1951 using their studies with X-ray diffraction in amino acids and small peptides to get their 3D structure. Both authors suggested the two secondary structures of proteins (Eisenberg, 2003; Pauling et al., 1951; Pauling & Corey, 1951):

- **α-helices** are the most usual secondary structure, their backbone atoms are disposed of in a right-handed helical structure (see Figure 1.2-2A) where the carbonyl oxygen atom of each residue receives a hydrogen bond from the amide nitrogen four residues furthermore in the sequence, completing a twist every 3.6 amino acids.

- **β-sheets** are tracts of 3 to 10 consecutive amino acids, named strands, with the backbone in an extended conformation interacting laterally with hydrogen bonds each other in a parallel or antiparallel manner (see Figure 1.2-2B). The carbonyl oxygens in one strand hydrogen bond with the amino hydrogens of the adjacent strand. Antiparallel β-sheets are more stable due to a better matching of the hydrogen bond patterns.

Both types of structures are held in shape by hydrogen bonds formed between the carbonyl group of one residue and the amino group of another, despite they differ in the hydrogen bond pattern. To connect helices and β-sheets in proteins there are loops and turns. Most of these structures suppose irregular secondary structures, but there are exceptions as type I and II reverse turns, also named β-turns or β-bends, which are tight turns connecting adjacent, antiparallel β-strands. (Sun et al., 2004).

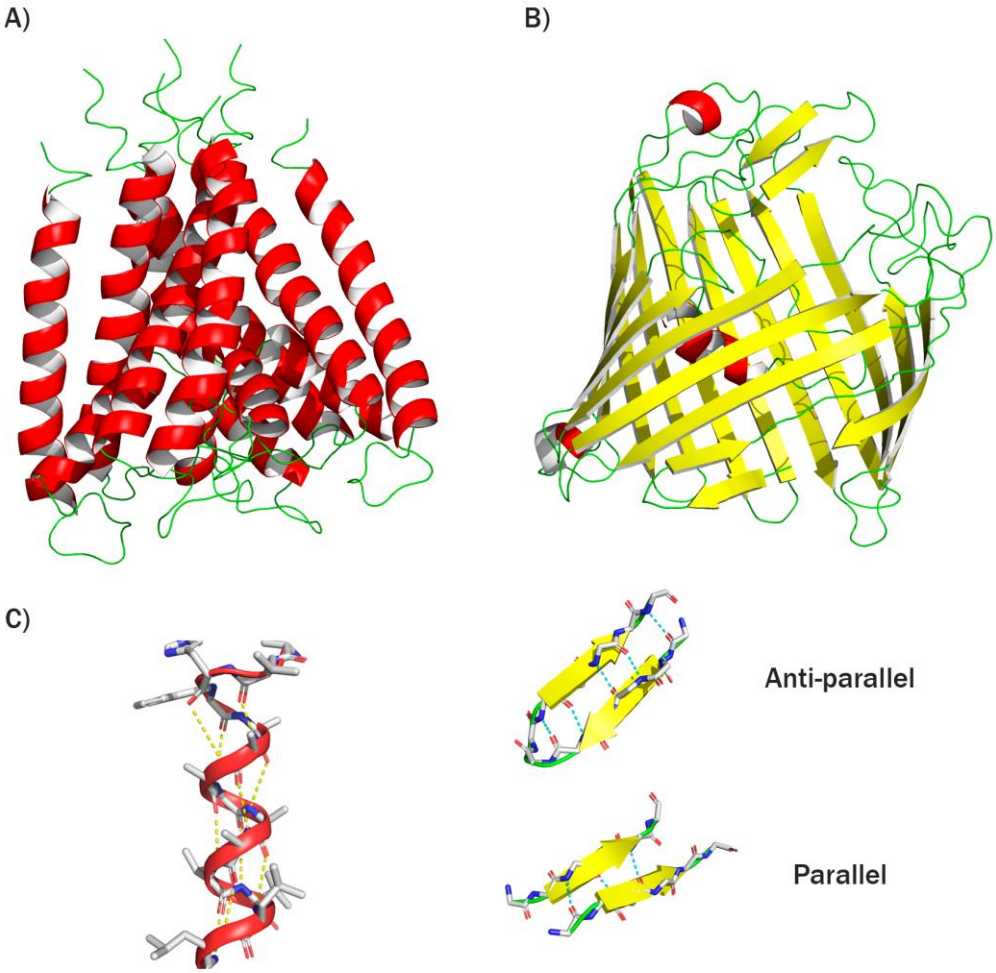


Figure 1.2-2 Typical structures of spanning TM proteins: A) α-helical ion channel (PDB:1BL8)**. B)** β-stranded barrel (PDB:2OMF). **C)** Hydrogen bonds between two atoms from different amino acids, one is the oxygen of the carbonyl, and the other is the hydrogen of the amino group of the main chain. Red represents α-helices, yellow arrows β-strands, and green lines random coils and β-turns, respectively. Image generated with PyMol (Schrödinger, 2021).

1.2.3 Supersecondary structure

The interactions by the residues of contiguous in the secondary structure patterns as helices and sheets form the **supersecondary structure** (*Biotechnology and Its Applications*, 2022). The main motifs are (see Figure 1.2-3): two α-helices joined by a β-turn, two β-sheets joined by a β-turn and two β-sheets and an α-helix joint by β-turns. The combination of these structures can form a range of diversity of 3D structures.

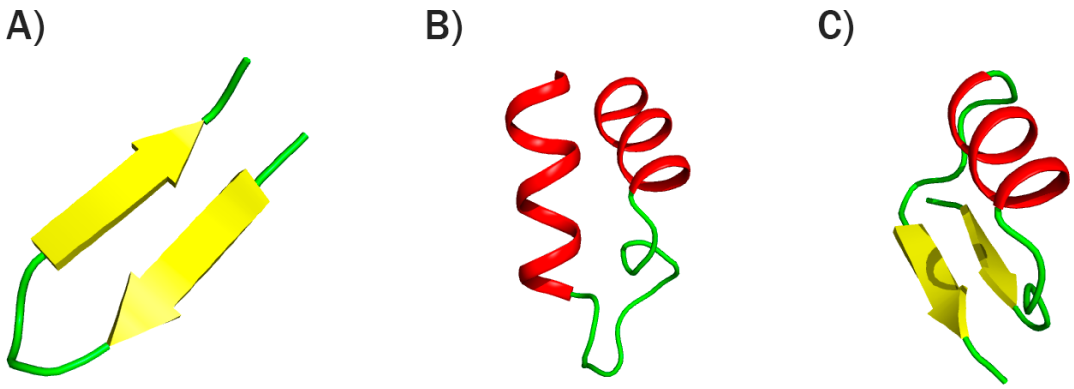


Figure 1.2-3 Type of supersecondary structures: A) two β-sheets joined by a β-turn**. B)** two α-helices joined by a β-turn. **C)** two β-sheets and an α-helix joint by β-turns. Image generated with PyMol (Schrödinger, 2021).

1.2.4 Tertiary structure and domains

The **tertiary structure** refers to the 3D arrangement of all the atoms that compose a protein molecule. It relates the precise space-based coordination of secondary structure elements and the location of all functional groups of a single polypeptide chain (Sun et al., 2004). This shape emerges from the mixture of the near interactions of the amino acids in the primary sequence and between amino acids from the distant parts of the primary sequence through disulfide, charge-charge, hydrophobic, or other non-covalent interactions. So, the weak noncovalent interactions with a much lesser

energetic contribution than covalent bonds are the main contributors to protein folding (Petsko & Ringe, 2004). The two most important are the van der Waals interaction and the hydrogen bond (Bitencourt-Ferreira et al., 2019; Tantardini, 2019).

Proteins can acquire new domains via some mechanisms. A **domain** is a distinct functional and structural unit in a protein and folds independently within its structural environment (Iliyas & Ansari, 2013; Pasek et al., 2006). Gene fusion, in which two adjacent genes become joined, is a major mechanism for multidomain protein formation in bacteria (Pasek et al., 2006). Nevertheless, the mechanisms for domain gain in eukaryotes are more varied, primarily because of their complex exon-intron gene structures. Some studies reveal that 40% of prokaryotic proteins consist of multiple domains while eukaryotes have around 65% of these multi-domain proteins (Ekman et al., 2005). Basing on the secondary structural elements to classified them into five main classes:

- **Alpha domains** composed of alpha-helices.
- **Beta domains** contain only beta-sheet.
- **Alpha/beta domains** contain beta strands with connecting helical segments.
- **Alpha + beta domains** that contain separated beta sheet and helical regions.
- **Cross-linked domains**, with an undefined secondary structure but stabilized by several disulfide bridges or metal ions.

1.2.5 Quaternary structure

The interaction of more than one polypeptide chain acting as a complex of protein designs the **quaternary structure** of a protein. This assembly of several polypeptide subunits permits many proteins to be functional. The weak noncovalent bonds that enable a polypeptide chain to fold into a specific configuration also allow proteins to join to each other to produce larger structures in the cell. Any region on a protein surface that interacts with another molecule through sets of noncovalent bonds is known as a binding site. A protein can contain binding sites for a range of molecules, large and small. If a binding site recognizes the surface of a second protein, the tight binding of two folded polypeptide chains on this site will produce a larger protein, whose quaternary structure has a precisely defined configuration. Each polypeptide

chain in such a protein is called a subunit, and each subunit may contain more than one domain (Alberts et al., 2018, p. 137).

If the protein is made of two subunits, the protein is called a dimer. In the case that these two subunits are identical, we named it as a homodimer, but if the subunits are different, we refer to them as heterodimer (see Figure 1.2-4). Also, we can find high-order oligomers, like trimers as the interaction of three subunits, tetramers as four subunits come close, etc.

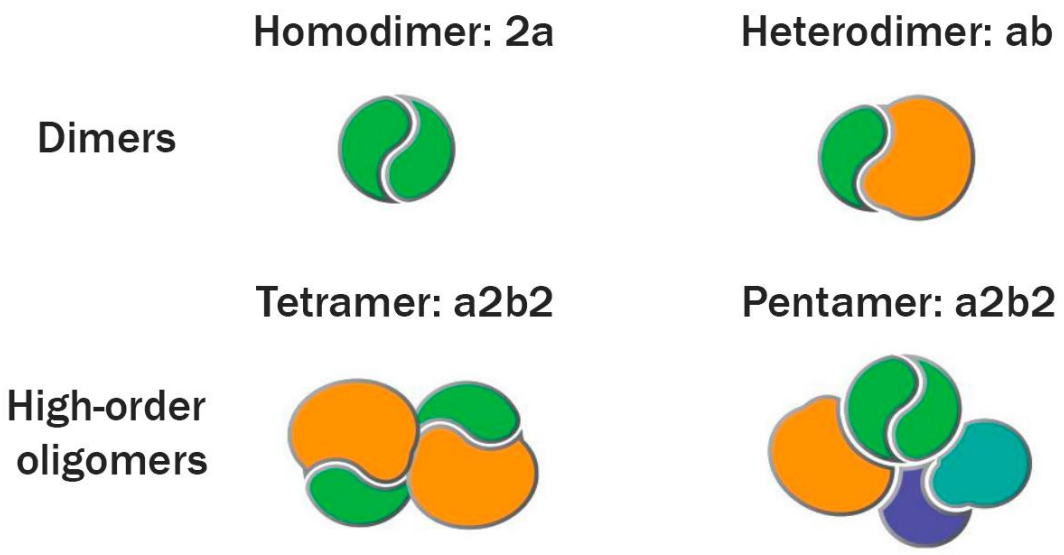


Figure 1.2-4. Schematic representation of different oligomers. Adapted from (Petsko & Ringe, 2004).

The subunits in a quaternary structure must be precisely arranged for the whole protein to function properly. Any change in the structure of the subunits or how they are connected causes notable changes in biological activity, with the possibility of causing a disease by the malfunction of this biological activity (Fedele et al., 2018; Guidolin et al., 2018; Szigetvari et al., 2019). For this reason, the study of these structures is relevant for the development of new therapies or treatments (Ryslik et al., 2016; Tarabini et al., 2019).

1.2.6 Proteins functions

Each type of protein consists of a precise sequence of amino acids that allows it to fold up into a particular 3D shape, or conformation. Proteins have precisely engineered moving parts whose mechanical actions are coupled to chemical events (Alberts et al., 2002). The bond of the proteins with other molecules enables functions as catalysts, mechanical support, and immune protection, generate movement, transmit nerve impulses, and control growth and differentiation, among others (see Table 1.2-1) (Alberts et al., 2018).

Type	**Function**
Enzymes	Catalyze covalent bond breakage or formation
Gene regulatory proteins	Bind to DNA to switch genes on or off
Motor proteins	Generate movement in cells and tissues
Receptor proteins	Detect signals and transmit them to the cell's response machinery
Signal proteins	Carry extracellular signals from cell to cell
Storage proteins	Storage of amino acids or ions
Structural proteins	Provides mechanical support to cells and tissues
Transport proteins	Carry small molecules or ions

Table 1.2-1. Type of proteins and main functions. Adapted from (Alberts et al., 2018).

1.2.7 Pathogenicity of mutations in proteins

The sequence of DNA that determines the structure of the proteins are crucial for the correct development of the activity of the cell. Although the cells have mechanisms to protect the DNA sequence from possible errors or damage, sometimes, it can occur that these mechanisms fail (Chatterjee & Walker, 2017). In those cases, the sequences suffer mutations changing the conformation of the structure, and that the function becomes pathogenic and unhealthy for the organism (J. D. Watson et al., 2008). The cause of a mutation may be by the change of one nucleotide, where in some cases, it is significant enough to cause disease, or a series of changes of several amino acids, such as truncation of the C-terminus by introducing an earlier stop codon and shortening the C-terminus.

The classification of each type of mutation is related to the pathological impact or effect of the mutated DNA sequence and depends if the mutation occurs in a coding region or not. For example, a nucleotide changes in an area of the sequence that has no associated function does not have the same impact as a change in a nucleotide belonging to a domain important for the functionality of the protein. Considering the effects of the DNA mutations, into the DNA sequence and protein sequences, there are two classifications to identify a mutation (Clancy, 2008):

A) Mutations based on the effect on the DNA sequence:

- Small-scale mutations (point mutation):
    - **Deletions**, the removal of one or more nucleotides from the DNA.
    - **Insertions,** the addition of one or more extra nucleotides into the DNA.
    - **Neutral mutation** changes in DNA sequence that is not beneficial nor prejudicial to an organism to survive.
    - **Substitutions** change of a single nucleotide to another.
- Large-scale mutations (chromosomal mutation):
    - Single-chromosome mutations:
        - **Deletions,** loss of the genes of the regions removed.
        - **Duplications**, repetitions of a chromosomal segment.

        - **Inversion**, the reverse of the orientation of a chromosome segment.
    - Two-chromosome mutations:
        - **Insertion**, the addition of a chromosomal segment of one chromosome to another.
        - **Translocation**, interchange of regions from nonhomologous chromosomes.

B) Mutations based on the effect on the protein sequence:

- **Point substitution** is the change of a single nucleotide. In the case that affect the DNA coding region, there are:
    - **Synonymous substitution,** replace of a codon with another that codes for the same amino acid, not modifying the protein sequence**.**
    - **Nonsynonymous substitution,** replace of a codon with another that codes for the same amino acid, modifying the protein sequence**.**
        - **Missense mutation** changes of a nucleotide to cause substitution of a different amino acid.
        - **Nonsense mutation** changes of a nucleotide that results in a premature stop codon, and often nonfunctional protein product.
- Changes of more than one nucleotide or big segments:
    - **Truncating mutation** changes in the nucleotide sequence of a gene resulting in the creation of an early stop codon.
    - **Frameshift** is a mutation resulting by insertion or deletion of a number of nucleotides that breaks up the reading frame of the codons, resulting in a totally distinct translation of the initial protein.

Mutations in proteins affect their structure and functions, which may lead to various diseases (Kulandaisamy et al., 2019). From the sequencing of the human genome in the 2000s, with the Human Genome Project, the great technological advances related to sequencing by the individual and community scientific projects have allowed the sequencing of a large number of populations with significantly higher throughput in lower cost and time (Buermans & den Dunnen, 2014; Kodama et al., 2012; Trinh et al., 2018). Currently, according to MutHTP there are 183395 disease variants in 4797

proteins and 17827 neutral variants in 3132 proteins. (Kulandaisamy et al., 2018, 2019). All the data related with the diseases are fast accumulating into public databases such as Uniprot, SwissVar (Mottaz et al., 2010), 1000 Genomes (1000 Genomes Project Consortium et al., 2015), ClinVar (Melissa J. Landrum et al., 2014), among others. Due to the large size of the data and their dispersion to different databases, making the annotation of the effects of the mutations by experimental approaches a hard task. This results in a lack of understanding of the effects of the mutation in the membrane protein (Kulandaisamy et al., 2019). The understanding of the role of mutations in disease pathology is crucial for the advance of precision/personalized medicine and it is necessary to identify the genetic differences between individuals (Vihinen, 2015). For this reason, bioinformatic tools become a big help to manage all this amount of data and try to correctly annotate the effects, as other of the objectives of this thesis (see chapter 2).

The acceleration of DNA sequencing from patients and population studies has resulted in extensive catalogs of human genetic variation, but the interpretation of rare mutations remains a challenge (Cummings et al., 2020). Most patients with rare disorders do not receive a molecular diagnosis and the variants and causative genes for more than half such disorders remain to be discovered. Despite that, the number of sequencing does not stop growing exponentially and bioinformatics tools are needed to organize and analyze all the new variants that must be analyzed (D. Huang et al., 2020; Richardson et al., 2016). For these reasons, a lot of studies are related to the sequencing and annotating of these variants in rare diseases (Turro et al., 2020; Venter et al., 2001). For example, many rare missense variants identified may or may not have an impact on gene function and can be problematic to interpret in a clinical setting. Thus, there is a need for a systematic approach to applying, reporting, and evaluating identified variants (Trinh et al., 2018). The molecular phenotyping of variants is a time-consuming process that can be shortened by the use of already existing data. Variant prioritization tools are widely used to predict the effect of mutations. These are mostly based on evolutionary conservation and expected impact on structure and function using evolutionary conservation parameters and physiochemistry properties of amino acids from sequence data, such as SIFT (Sim et

al., 2012), Provean (Choi & Chan, 2015) and MutationTaster (Schwarz et al., 2014), while some tools such as Polyphen-2 (Schwarz et al., 2014) also incorporate features related to structural data (Garcia-Recio et al., 2021). Using these structural computational algorithms, it is possible to predict if a mutation is pathogenic or not, causing a significant increase in the pathogenicity and functional parameters of non-annotated homologous variants contributing to clinical decision making.

1.2.8 Classification of proteins

Apart from the differentiation of proteins by function (see section 1.2.6) they can also be classified by structure (Sadowski & Jones, 2009). One of the most used classification schemes divides proteins based on their structure into:

- **Fibrous proteins** are sheet-like filamentous structures that perform mainly structural function. They are typically elongated and insoluble. An example of these types of proteins is collagen.
- **Globular proteins** are spherical compact structures soluble in water. These types of proteins are one of the most abundant ones. One example is the hemoglobin.
- **Membrane proteins** are present in the membranes and interact with them (see section 1.3.3). For these reasons, they are the target of medical drugs (Gong et al., 2019).

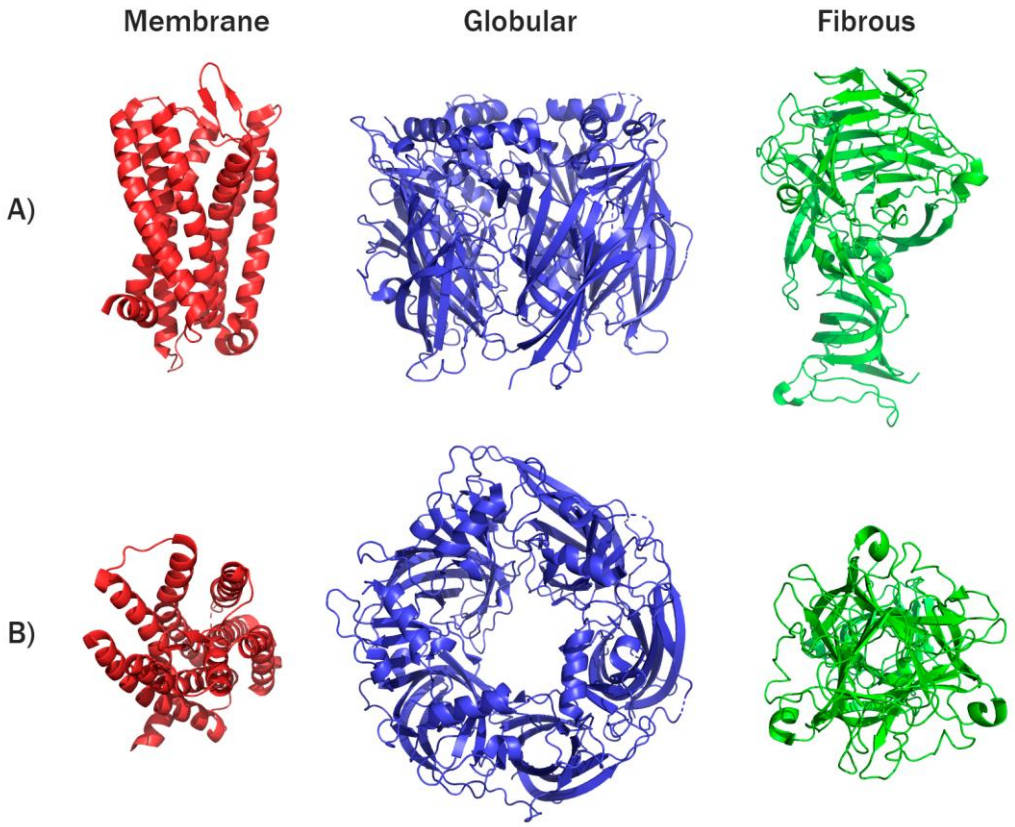


Figure 1.2-5. Type of proteins by structure. A) Lateral point of view. **B)** Top point of view. PDBID 4DKL (red) is a membrane protein. PDBID 7N0W (red) is a globular protein. PDBID 4UXE (green) is a fibrous protein. Image generated with PyMol (Schrödinger, 2021).

In this thesis we centered our view into the membrane proteins. Compared to globular proteins and other types of proteins, membrane proteins have always been a challenge due to their difficulty in establishing experimental favorable conditions to preserve their conformation correctly when they are isolated from their native environment.

1.3 The cell membrane

In order to maintain this “life”, cells are composed of different components that manage thousands of biochemical reactions necessary for the survival of these units. The structure of these components is based on an organization and interaction of several proteins related to a specific function and determine one of these elements that is the key to life, the cell membrane. The membrane cell acts as a physical barrier to separate

the interior of a cell from the outside by controlling the substances that enter and leave by their selective permeability (see section 1.3.2).

1.3.1 Structure and composition of the cell membrane

The biological cell membrane consists of an organization of assembling amphiphilic molecules as phospholipids, sterols, and other lipids and proteins to perform the transport and enzymatic functions (see Figure 1.3-1) (Schrum et al., 2010). The model used as biological membrane is based on the fluid-mosaic membrane model, proposed in 1972 by Singer and Nicolson (Nicolson, 2014; Singer & Nicolson, 1972). Mainly, the model describes the membrane as a back-to-back of phospholipids working as a semi-permeable bilayer barrier (see Figure 1.3-1A-B). For the actual eukaryotic cells, this bilayer of glycerophospholipid molecules is around ~30 Å as minimum size that surrounds each cell (Figure 1.3-1C). Thickness is controlled in part by a class of lipids, known as cholesterols, which are crucial for the formation of thicker and functionally important domains. (Gerle, 2019; H. Watson, 2015). Also, these lipids are contributing to the fluidity of the membrane and modulating the function of numerous types of different proteins inserted into the membrane with a variety of functions, called membrane proteins (see section 1.3) (Grouleff et al., 2015; H. Watson, 2015). Both elements together with the phospholipid bilayer develop one of the most important functions in the cell related with the transport of substances between the sections of the cell (Stillwell, 2016).

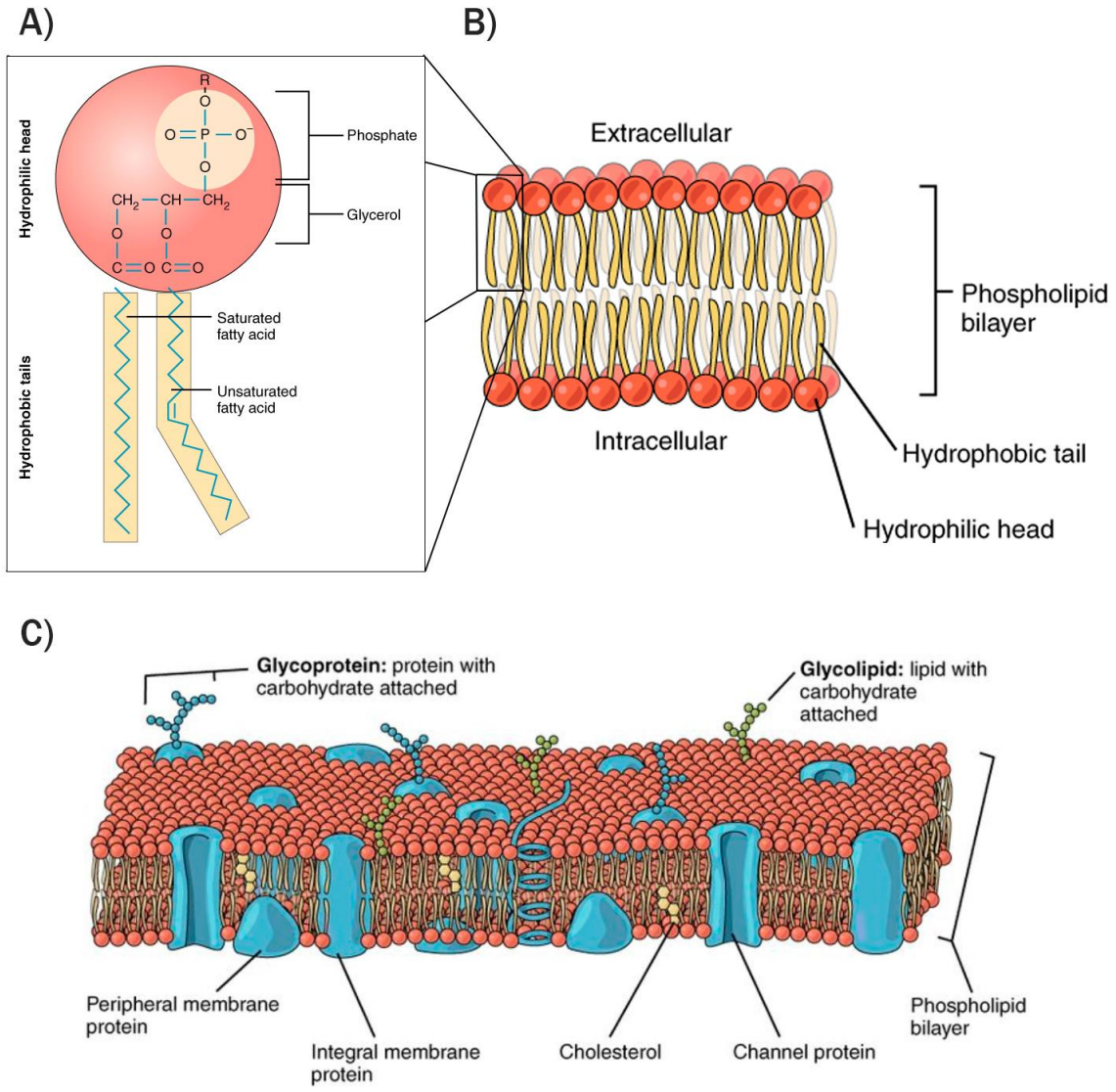


Figure 1.3-1 Phospholipid bilayer and phospholipid structure A) A phospholipid molecule consists of a polar phosphate “head,” which is hydrophilic and a non-polar lipid “tail,” which is hydrophobic. **B)** The phospholipid bilayer consists of two adjacent sheets of phospholipids, arranged tail to tail. The hydrophobic tails associate with one another, forming the interior of the membrane. The polar heads contact the fluid inside and outside of the cell. **C)** The cell membrane of the cell is a phospholipid bilayer containing many different molecular components, including proteins and cholesterol, some with carbohydrate groups attached. Adapted from (Gordon Betts et al., 2013)

1.3.2 Membrane proteins

Membrane proteins (MPs) are a large group of proteins that mediate the interaction between the cell environment and its interior. This huge group includes receptors, ionic channels, transporters, and enzymes, accounting for 25% of proteins of the human genome (see Figure 1.3-3) (Fagerberg et al., 2010). Membrane proteins are related with a diverse variety of vital functions and they are the key to the proper functionality of the cells (Almén et al., 2009). This variety of functions is specifically defined by the structure of these proteins and their location into the cell membrane (Cournia et al., 2015). A representation of the different types is shown in Figure 1.3-3. These types are:

- **Integral membrane proteins,** or also called transmembrane proteins, are totally attached to the membrane.
    - **Monotopic proteins** attach only by one side of the membrane and do not span the whole way across.
    - **Bitopic proteins** spanned across the membrane only once.
    - **Polytopic proteins** that cross the membrane more than once. In comparison to bitopic ones they have different transmembrane topology.
        - **Helix bundle proteins**.
        - **Beta barrel proteins**.
- **Peripheral membrane proteins** are partially attached to the lipid bilayer or to integral proteins by interactions.

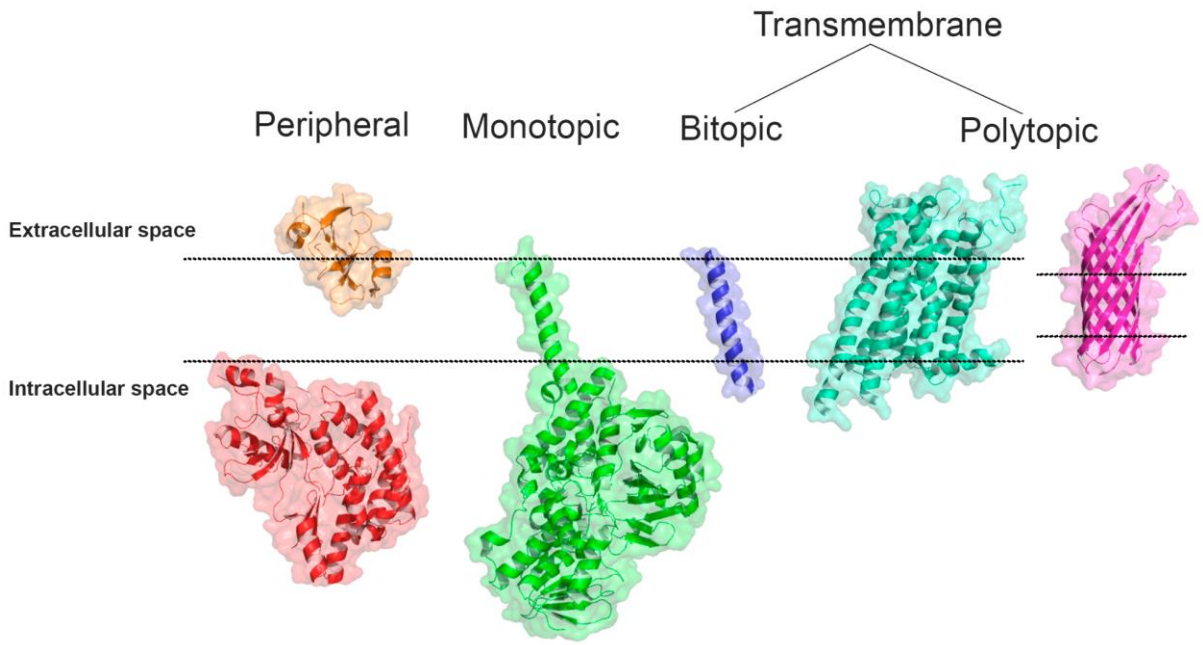


Figure 1.3-3. Representation of types of membrane proteins. PDBID 6CS9 (orange) and PDBID 4PLA (red) are peripheral proteins. PDBID 2Z5X (green) corresponds to a monotopic protein. PDBID 6MCT (blue) is a bitopic protein. PDBID 6BQG (cyan) is a transmembrane α-helical protein and PDBID 3QRA (magenta) represents a transmembrane β-barrel. Coordinates from OPM, image generated with PyMol (Schrödinger, 2021).

Besides, membrane proteins help the cell to interact with its environment carrying out diverse functions. These particular functions represented in the Figure 1.3-4 define the classification on:

- **Transporters** that carry and facilitate the passage of molecules and ions across the membrane.
- **Receptors** where relays signals between the internal and external environments activating cell processes.
- **Enzymes** are related in many catalytic activities like transferase, hydrolase or oxidoreductase.
- **Anchors** allow the interaction of the cell in different ways.

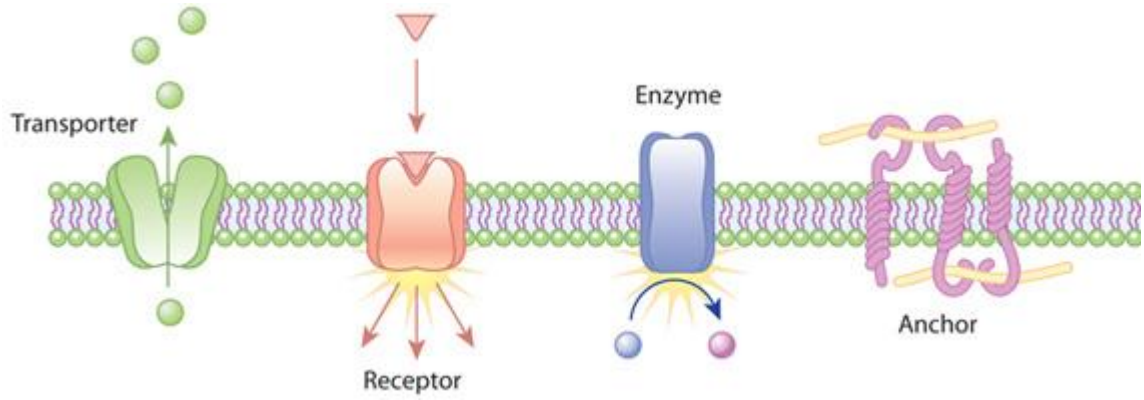


Figure 1.3-4: Examples of the action of transmembrane proteins. Adapted from (Nature Education, 2014)

1.3.3 Specific features of membrane proteins

Despite the membrane protein representation of proteins in sequence in the human genome is about 25% (Fagerberg et al., 2010), only ~2-3% (1411/48535) of crystal structures in Protein Data Bank are membrane proteins, due to the difficulty of their overexpression, purification and crystallization. The environment and function of membrane proteins are different from those of globular proteins and these lead to distinct properties. The transmembrane region of membrane proteins is mostly hydrophobic and consequently less prone to form hydrogen bonds and with little screening of electrostatic interactions. Compared with globular proteins there is a lack of polar–polar, polar–charged and charged–charged interactions and most of the interactions involve nonpolar residues. (Mayol et al., 2019). The lipid bilayer imposes a physical constraint that limits the number of folds that polypeptide chains can adopt when embedded in a membrane. Declining folds preservation in the transmembrane region of membrane proteins requires a lower degree of sequence conservation than in globular proteins. (Olivella et al., 2013). Consequently, due to the fact that of the limited high-resolution structural data on membrane proteins, computation methods to predict their 3D structure from the amino acid sequence are a useful tool. Homology models of proteins with undetermined experimental structure can also be built from homologous proteins of known structure and similar sequence, as templates (Olivella et al., 2013).

1.3.5 Disease and therapy in membrane proteins

Because of this grade of pathogenesis causing several diseases, about 50% of membrane proteins are the pharmacological target of various diseases, such as cancer, neurodegenerative diseases as Alzheimer's, diabetes and cystic fibrosis (Almeida et al., 2017; Meng et al., 2017; Overington et al., 2006; Suzuki et al., 2006). Most of these drugs are specifically to ion channels, as N-methyl-D-aspartate receptors (NMDARs), or receptors, as G-protein coupled receptors (GPCRs) cause of all the diseases related with these receptors (García-Recio et al., 2021; Santos-Gómez et al., 2021; Santos et al., 2017). There are two types of drugs:

- **Orthosteric**, which binds at the active site.
- **Allosteric**, which binds elsewhere on the protein surface, and allosterically changes the conformation of the protein binding site.

The different mechanisms through which the two drug types affect protein activity and their potential pitfalls call for different considerations in drug design (Nussinov & Tsai, 2012). In order to discover new drugs that target in a selective way we need more accurate models of these types of receptors using the actual structures availables. Because of that, bioinformatics tools are crucial to design drugs and one of the objectives of this thesis are related to it (see chapter 2) (García-Recio et al., 2020).

There are two groups of proteins that are really important because of their functionality and their relation to various diseases, such as cancer, neurodegenerative and cerebrovascular diseases, that are the most frequent causes of death and disability in the world. Furthermore, these groups of proteins are considered relevant and needed to study. These two groups of proteins are the N-methyl-D-aspartate receptors (NMDAR) and G-protein coupled receptors (GPCR), and they are studied specifically in this thesis (see sections 1.4 and 1.5).

1.4 N-methyl-D-aspartate receptors

The N-methyl-D-aspartate receptors (NMDARs) are a type of membrane proteins, specifically ion channels like neurotransmitters, that play an important role in the

neuronal signalling in the brain development, wiring high cognitive and motor functions, among others (Paoletti et al., 2013). NMDARs are heterotetramers composed of two pairs of subunits: one pair of glycine/D-serine-binding GluN1 subunits (encoded by the GRIN1 gene) and the other pair by two glutamate-binding subunits like GluN2A/GluN2B (encoded by GRIN2A and GRIN2DB genes, respectively) (Flores-Soto et al., 2012). Moreover GluN1, GluN2A and GluN2B subunits are the most predominant expressed subunits , although other subunits GluN2C/GluN2D/GluN3A/GluN3B are found at particular developmental stages and in precise neural subtypes (Chatterton et al., 2002; Madry et al., 2007; Pérez-Otaño et al., 2016; Smothers & Woodward, 2009). For a stable neuronal function and brain connectivity, NMDAR heteromeric channels require a precise spatio-temporal expression pattern. Functionally, the NMDAR acts as a detector responding to sustained glutamate release and its activation is mediated by a calcium influx flowing into signalling cascades, ending into neuronal responses (García-Recio et al., 2021).

1.4.1 Structure of the NMDAR

From PDB there are 96 structures ("Advanced search by Structure Title contains words NMDA"), such as domains, agonists, antagonists, among others, related with the NMDAR. To obtain the full structure of NMDAR some studies must model the structure using a template as reference, specially the CTD domain because of the difficulty to be crystallized. Due to these structurally unsolved regions the researchers only can investigate the sequence to study the subunits without knowing about their specific function and structure (see section 4.3, Article III) (Warnet et al., 2021). Only 82 structures are available as reference to modelling the NMDAR structure.

NMDARs are heterotetramers composed of GluN1 and GluN2 subunits, which bind glycine and glutamate, respectively, to activate their ion channels The structures crystallized reveal that activation and competitive inhibition, fundamental to the brain physiology, by both GluN1 and GluN2 antagonists occur by controlling the tension of the linker between the ligand-binding domain and the transmembrane ion channel of the GluN2 subunit (Chou et al., 2020). There are currently more than 90 structures of the NMDA receptor that cover the ATD; LBD and TM domains or only the ATD and the

LBD (Chou et al., 2020; Karakas & Furukawa, 2014; Zhang et al., 2018). The domains of NMDAR are (see Figure 1.4-1) (Lee et al., 2014; Tajima et al., 2016):

- **Extracellular domain** contains a modulatory domain named the **N-terminal domain** (NTD), or also **Amino-terminal domain** (ATD), and the Ligand-binding domain (LBD). GluN1 subunits bind the co-agonist glycine and GluN2/3 subunits bind the neurotransmitter glutamate.
- **Agonist-binding domain**, where three transmembrane segments and a loop reminiscent of the selectivity filter of potassium channels.
- **Ion channel forming transmembrane domain** (TMD) contributes residues to the channel pore and distributes the receptor conductance, high-calcium permeability, and voltage-dependent magnesium block.
- **Intracellular C-terminal domain** (CTD) as cytoplasmic domain contains residues that can be modified by protein kinases and protein phosphatases. The residues interact with structural, adaptor and scaffolding proteins.

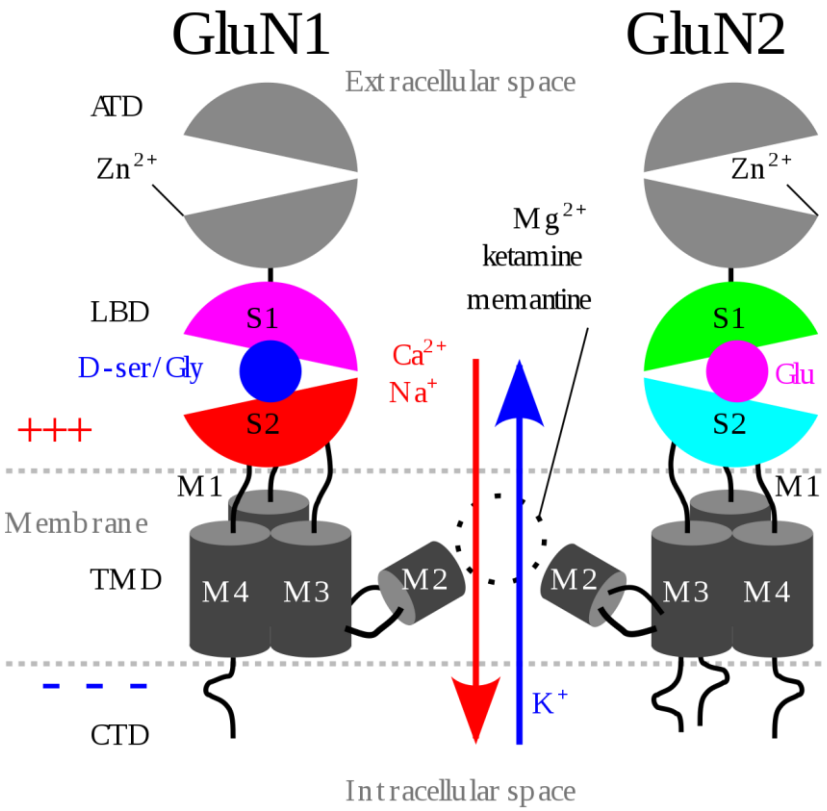


Figure 1.4-1 Cell membrane bound N1/N2/N1/N2 NMDA receptor. Typical NMDA-receptors have 4 subunits - only 2 are shown. Normal flow of Ca2+, Na+ and K+ ions is shown alongside with the regular membrane potential (+ and - signs) and some of the ligand binding sites. Adapted from (Keministi, 2019)

The extracellular amino-terminal and ligand-binding domains (ATD and LBD, respectively) are linked to the transmembrane domain (TMD) constituting the pore of the channel, and a cytoplasmic carboxy-terminal domain (CTD) a long cytoplasmic tail with no determined structure (see Figure 1.4-2). Compared with the extracellular domains, the structure and function of the C-terminal domain is much less known because of the difficulties of the crystallization, but they play multiple roles, such as a modifier of channel function, a regulator of cellular location and abundance, and signalling scaffold control the downstream signalling output (Warnet et al., 2021).

1.4.2 NMDAR signalling

The simultaneous binding of NMDAR coagonists, glutamate and glycine/D-serine, is mechanically transduced, allowing the channel pore opening and Na+ and Ca2+ influx (Mayer et al., 1984; Nowak et al., 1984). Moreover to the subunit composition the receptor activity is regulated by small ions and molecules, such as protons, zinc, spermine (Mony et al., 2011), membrane lipids (Korinek et al., 2015), and proteins, involved in posttranslational modifications (Lussier et al., 2015) and protein–protein interactions (Frank & Grant, 2017).

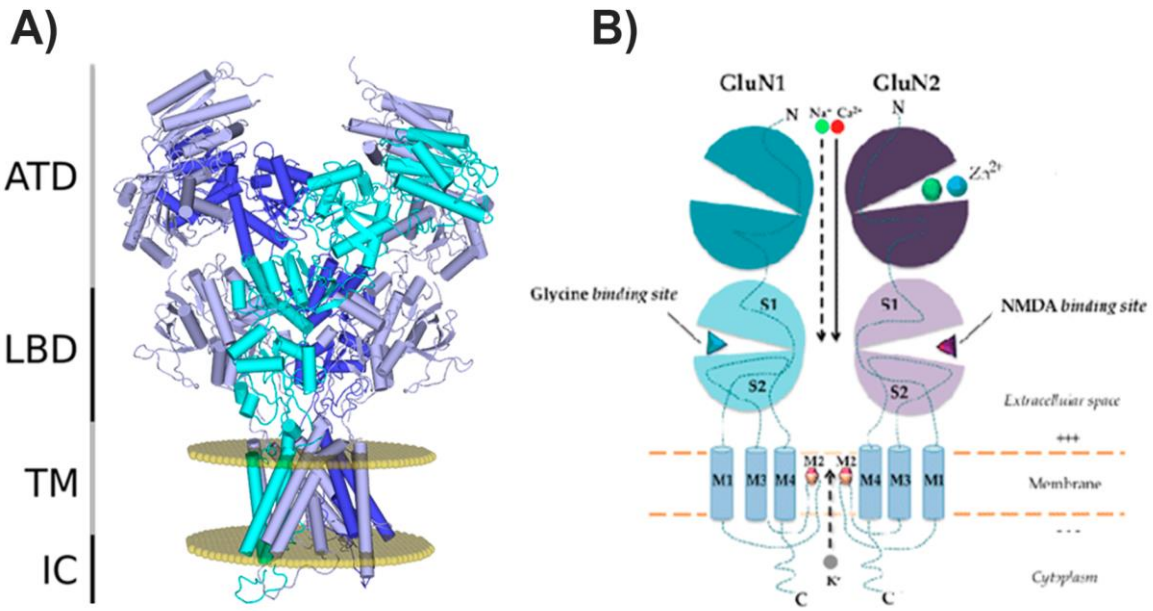


Figure 1.4-2. A) NMDAR 3d-structure with domains indicated. GluN1 subunits colored by blue, GluN2A subunits colored by dark blue and GluN2B subunits colored by cyan. The limits of the membrane are represented as yellow dots. B) Schema of GluN1/GluN2 NMDAR. Adapted from (Valdivielso et al., 2020).

NMDAR-mediated signalling is crucial for fundamental neuronal processes. Therefore, the dysregulation of the NMDAR function can cause synaptic dysfunction and disturb neuronal activity, which ultimately can lead to neurological conditions. Some years ago, the genetic reports describing the association between de novo mutations of GRIN genes associated with neurological disorders were released (de Ligt et al., 2012; O'Roak et al., 2011; Tarabeux et al., 2011). Since these reports, third-generation deep sequencing has accelerated the identification of novel GRIN variants in individuals affected by neurodevelopmental and psychiatric disorders and allowed them to initiate pioneer studies towards delineation of a clinical spectrum genetically defined as GRIN-related disorders (GRDs) (Burnashev & Szepetowski, 2015; Lemke et al., 2014, 2016; Lesca et al., 2013; Platzer et al., 2017). GRDs are a group of rare neurological disorders with a genetic etiology, namely de novo GRIN mutation (with almost exclusively autosomal dominant inheritance pattern), a primary cause of NMDAR-mediated dysfunction of excitatory neurotransmission. Clinically, GRDs' clinical spectrum includes intellectual disability, epilepsy, movement disorders, development

delay, autism spectrum disorder, and schizophrenia, among others (Monaghan & Jane, 2011). In addition, several functional studies have been conducted to characterize the direct impact of de novo GRIN variants gene products on NMDAR biogenesis and activity (Rao & Craig, 1997; Weston, 2017; XiangWei et al., 2018).

1.4.3 NMDAR pathogenesis

The initial genetic reports describe the association between de novo mutations of GRIN genes associated and neurological disorders (de Ligt et al., 2012; O'Roak et al., 2011; Tarabeux et al., 2011). Since then, next generation deep sequencing has accelerated the identification of novel GRIN variants in individuals affected by neurodevelopmental and psychiatric disorders defined as GRIN-related disorders (GRD) (Burnashev & Szepetowski, 2015; Lemke et al., 2014, 2016; Platzer et al., 2017). GRDs are a group of rare neurological disorders with a genetic etiology, namely de novo GRIN mutation, a primary cause of NMDAR-mediated dysfunction of excitatory neurotransmission. Clinically, GRDs' clinical spectrum includes intellectual disability, epilepsy, movement disorders, development delay, autism spectrum disorder, and schizophrenia, among others. Based on inherent, clinical and functional data, the grade of pathogenicity in GRIN variants can be classified into (García-Recio et al., 2021):

- **Disease-associated**, variants with reported functional alterations.
- **Neutral variants** that do not affect the NMDAR activity.
- **Uncertain pathogenicity**, variants with conflicting annotations.

Specifically, the GRIN variant's impact is experimentally defined by the evaluation of the level of expression of the receptor, follow by the analysis of the trafficking of the receptor into the membrane surface and, in the cases that the receptor is expressed in the membrane, do the meditation of the electrophysiological properties of the mutant NMDARs. The experimental functional information of these variants gives the possibility to do the stratification of the data into four categories important for the therapy selection (García-Recio et al., 2021):

- **Loss-of-function (LoF)**, variant that reduce the surface expression of the receptor and/or reduced NMDAR-mediated currents resulting from diminished

ion channel conductance, open probability, agonist(s) affinity(ies), and desensitization and/or deactivation rates.

- **Gain-of-function (GoF)**, variant that increase the surface expression of the receptor and/or NMDAR-mediated currents were increased as a result of changes in the gating/affinity properties
- **Complex**, variant of uncertain functional classification exhibiting both LoF- and GoF-associated changes.
- **Not affected,** variant that is not showing functional alterations in heterologous cell systems.

GRD translational research shows a gap between the growing genetic data and their evaluation toward functional stratification and therapeutic approaches evaluation. This stratification is critical to define precise NMDARs functional rescue aimed at personalized medicine to treat the diseases and reduce the lack of information. The majority of these variants do not have functional annotations and that have been classified as uncertain pathogenicity. Moreover, the data are spread among databases and the literature, together with data duplication and/or conflicting annotations, making their retrieval and interpretation more difficult (García-Recio et al., 2021).

1.4.4 Strategic therapy to modulate NMDAR activity

Due to the lack of clinical efficacy the development of NMDAR-targeted therapeutics has been disappointing. Despite this, NMDAR therapeutics continue to exhibit significant potential. Of the multiple drug binding sites on the various NMDAR subunits, many potential types of NMDAR antagonists exist, and some of these reveal distinct patterns of selectivity (Monaghan & Jane, 2011). Therapeutic approaches for GRDs are focused on the use of direct NMDAR modulators counteracting either variant effects, GoF or LoF, reducing or potentiating NMDARs activity, respectively. Regarding GRIN-GoF variants, the FDA-approved drug memantine (low-affinity NMDAR antagonist) has been evaluated for GRD associated with GRIN-GoF variants (Ogden et al., 2017; Pierson et al., 2014; Platzer et al., 2017). Similarly, radiprodil (GluN2B selective negative allosteric modulator) has been recently evaluated and constitutes a promising drug for the treatment of NMDARs containing GluN2B GoF mutations (Auvin

et al., 2020; Mullier et al., 2017) . The potential rescue of GRIN LoF variants has also been evaluated in preclinical and clinical studies, using both NMDARs orthosteric and allosteric modulators. In this line, a pilot clinical study showed that L-serine (precursor of D-serine, endogenous NMDAR coagonist) dietary supplement results in a clinical improvement in a patient harboring a GRIN2B LoF variant (Soto et al., 2019).

1.5 G Protein-Coupled Receptors

About 4% of genes (799/20595) in the human genome code for environment-sensing cell surface receptors, most of which are GPCRs (Gloriam et al., 2007; Kooistra et al., 2021; UniProt Consortium, 2019) G protein-coupled receptors (GPCRs) comprise the largest family of membrane proteins in mammals. They mediate several important physiological functions, as responding to a variety of ligands and are involved in the transmission of extracellular signals inside the cell, and are involved in the pathophysiology of serious diseases, such as retinitis pigmentosa, hypo- and hyperthyroidism, nephrogenic diabetes insipidus, several fertility disorders, and even carcinomas (Maurice et al., 2011; Schöneberg et al., 2004). About 34% of currently marketed drugs target human GPCRs, having a great potential in biomedical research and drug development (Hauser et al., 2017). Despite the high diversity in sequence, all GPCRs share one common structural feature, seven TM α-helices joined by a set of three extracellular and three intracellular loops (Katritch et al., 2013). GPCRs are classically described as monomeric transmembrane (TM) receptors that form a ternary complex along with the ligand and associated G protein (see Figure 1.5-1).

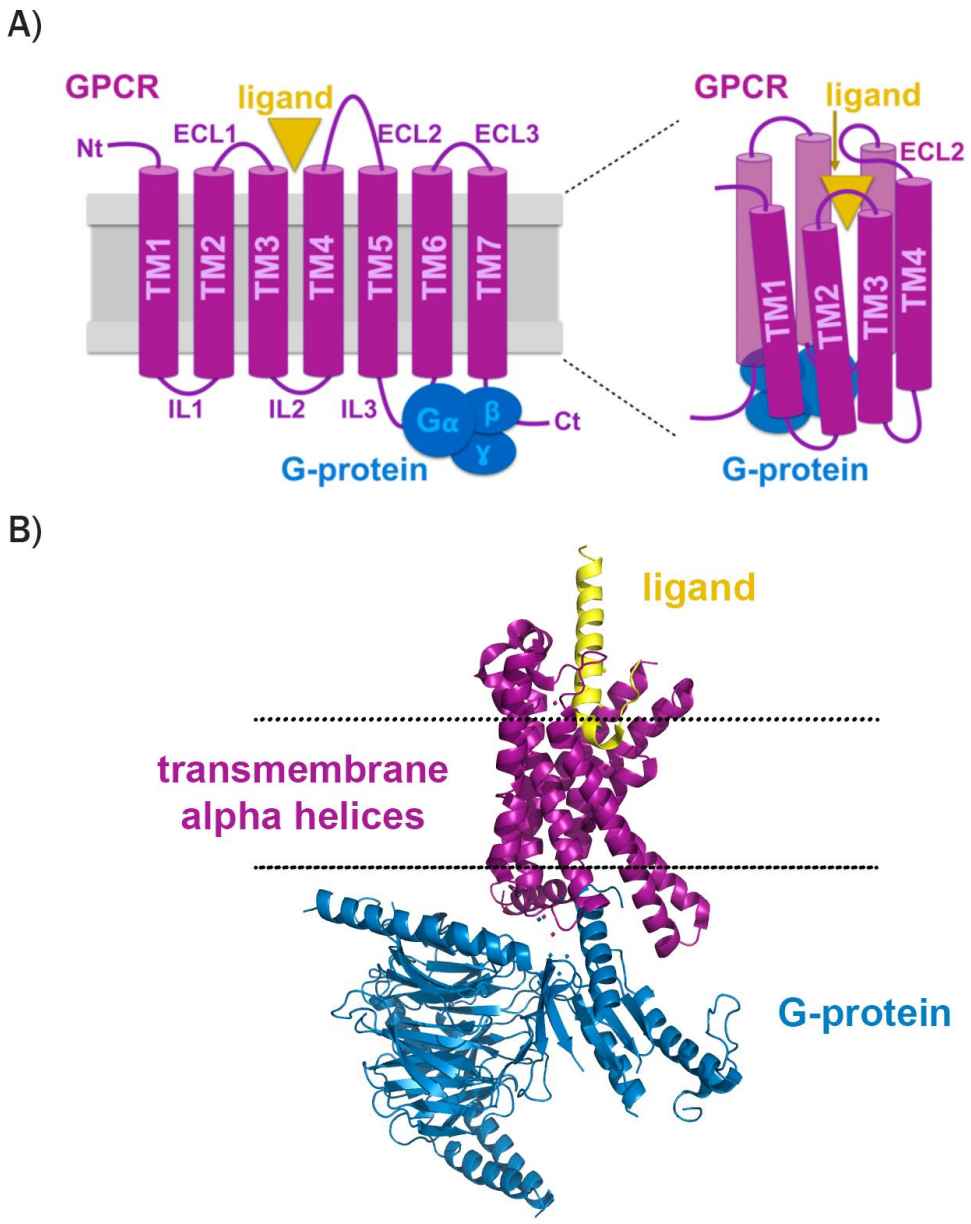


Figure 1.5-1. A) Schema of a GPCR. **B)** GPCR 3d-structure (PDBID 6P9X). Transmembrane alpha helices are colored in purple, ligand is colored in yellow, and the G-protein complex colored as blue. Adapted from (Schneider et al., 2018).

1.5.1 GPCR signaling and activation

GPCRs are the detonator of the intracellular signal transduction after a conformational reorganization by recruiting cytosolic proteins. There are two main paths of signal transduction carried out by these receptors (see Figure 1.5-2) (Dwivedi et al., 2018; Hilger et al., 2018):

- **G-protein dependent signalling path** that related with the coupling to a heterotrimeric G-protein (Gα, Gβ and Gγ), followed by the activation of second messengers, such as cyclic AMP (cAMP), calcium or inositol phosphate, which initiate a range of downstream signaling pathways.
- **G-protein independent signalling path** that involves the phosphorylation of the C-terminus by specific GPCR kinases and a recruitment of multifunctional proteins, arrestins.

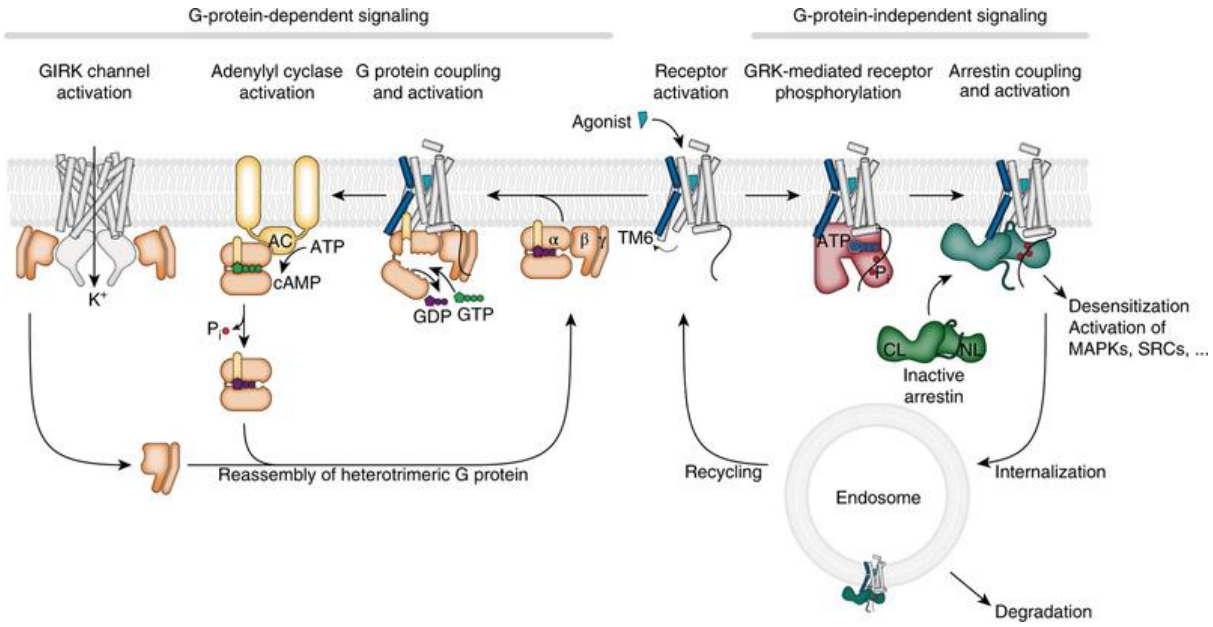


Figure 1.5-2. GPCR signal transduction mechanism. Adapted from (Hilger et al., 2018)

For specific receptors, ligands can activate one of these paths, but cannot the other path. This fact is known as bias signaling (Seyedabadi et al., 2019). The big volume of different crystal structures available (see Table 1.5-2), in different conformations, with different ligands, have assisted extensively to the comprehension of the activation process of this type of receptors. However, the entire conception of these processes needs further investigation of the conformational dynamics of receptor-transducer complexes. The crystal structures only show a static representation of the receptor and not reveal the complex activation of the mechanism of the receptor (Escuer, 2019). Another element that helps in the signal transduction are the different water networks that are formed in the active or inactive states (Venkatakrishnan et al., 2019).

1.5.2 GPCRs classification

Due to the amount of GPCRs structures with diverse roles in the cell regulation and signal transduction, the categorization of these proteins has been very helpful to their identification. The first GPCRs classification system used were based on the amino acid sequence and functional similarities classifying the proteins into six different classes from A to F (Attwood & Findlay, 1994; Kolakowski, 1994). As they have been investigated new classifications have been defined, like the Fredrikson classification that showed that human GPCRs can be divided into five main families (Fredriksson et al., 2003). Table 1.5-1 shows both classifications.

Kolakovski	Fredriksson	Description
A	Rhodopsin	Rhodopsin-like family
B	Secretin	Secretin receptor family
	Adhesion	
C	Glutamate	Metabotropic glutamate family
D	-	Fungal mating pheromone receptors
F	Frizzled/Taste2	Frizzled/smoothened receptors

Table 1.5-1. Table containing the classification of Kolakowski and the equivalence into Fredriksson classification.

Among these 5 main classes presented by Fredriksson (see Figure 1.5-1), class A is by far the largest and includes receptors like rhodopsin, olfactory, opioid, adrenergic, chemokine, angiotensin, histamine, dopamine or adenosine receptors. Despite their high diversity in sequence, all GPCRs share one common structural feature: seven TM α-helices joined by a set of three extracellular and three intracellular loops (Katritch 2013). Main structural differences among classes are located in the extracellular part.

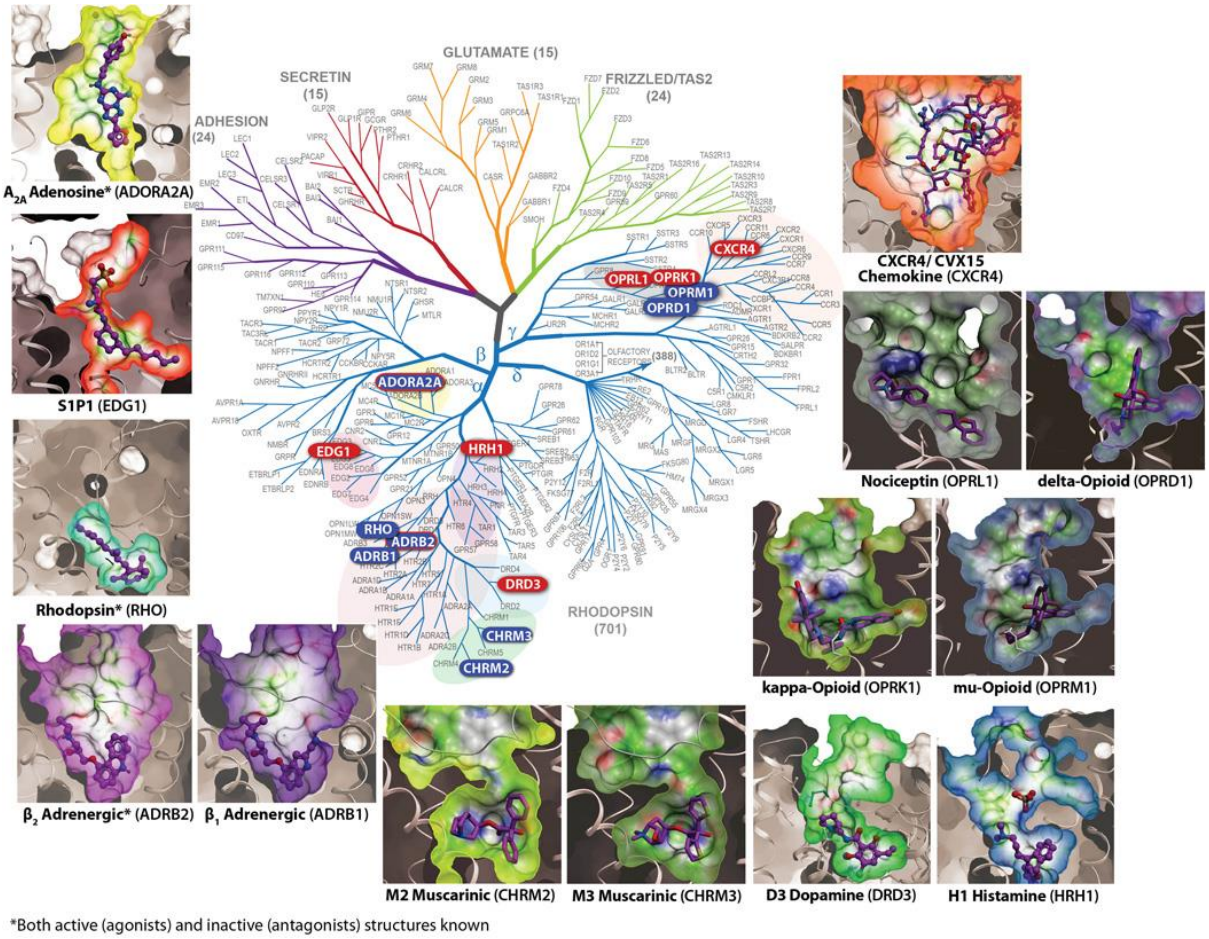


Figure 1.5-3. Phylogenetic tree representation of the human GPCR superfamily constructed using sequence similarity within the seven-transmembrane region (Stevens et al., 2013).

Recently, some studies propose new ways to classify the GPCRs based on the phylogeny using distance and character-based methods to relate better the evolution of GPCRs and provide a panoramic view of the GPCRs network (Hu et al., 2017; Kooistra et al., 2021).

Due to the environment of the cell membrane, where GPCRs are located, the methods to obtain the structure of these receptors have become a challenge for the researchers, as it commented in section 1.2.8. Long history, the advance of new technologies and methods provides a exponential increase to the actual number of unique receptors with a total of 113 receptors of GPCRs without consideration of the complex with ligand and/or G protein and their active state (see Table 1.5-2 and Figure 1.5-4) (GPCRdb,

2021). Even so, actually there are structures that are not isolated, or they do not have the good resolution to resolve the internal water molecules importance to the stability of the receptors (see section 1.5.2). Both elements are crucial to produce good models, because of that bioinformatic tools have become essential to reduce this blank of information and in this thesis, we present some tools that reduce this lack (see section 4.1).

Class (Family)	A (Rhodopsin)	B1 (Secretin)	B2 (Adhesion)	C (Glutamate)	D1 (Ste2-like fungal pheromone)	F (Frizzled)	T2 (Taste 2)	Total
Receptors**	85	14	1	8	1	4	0	113
Receptor - ligand complexes	309	41	2	41	1	13	0	407
Receptor - G protein complexes	40	14	1	3	1	2	0	61
Active-state receptors*	48	14	1	4	1	2	0	70

Table 1.5-2. Table containing the number of unique GPCR complexes. From (GPCRdb, 2021).

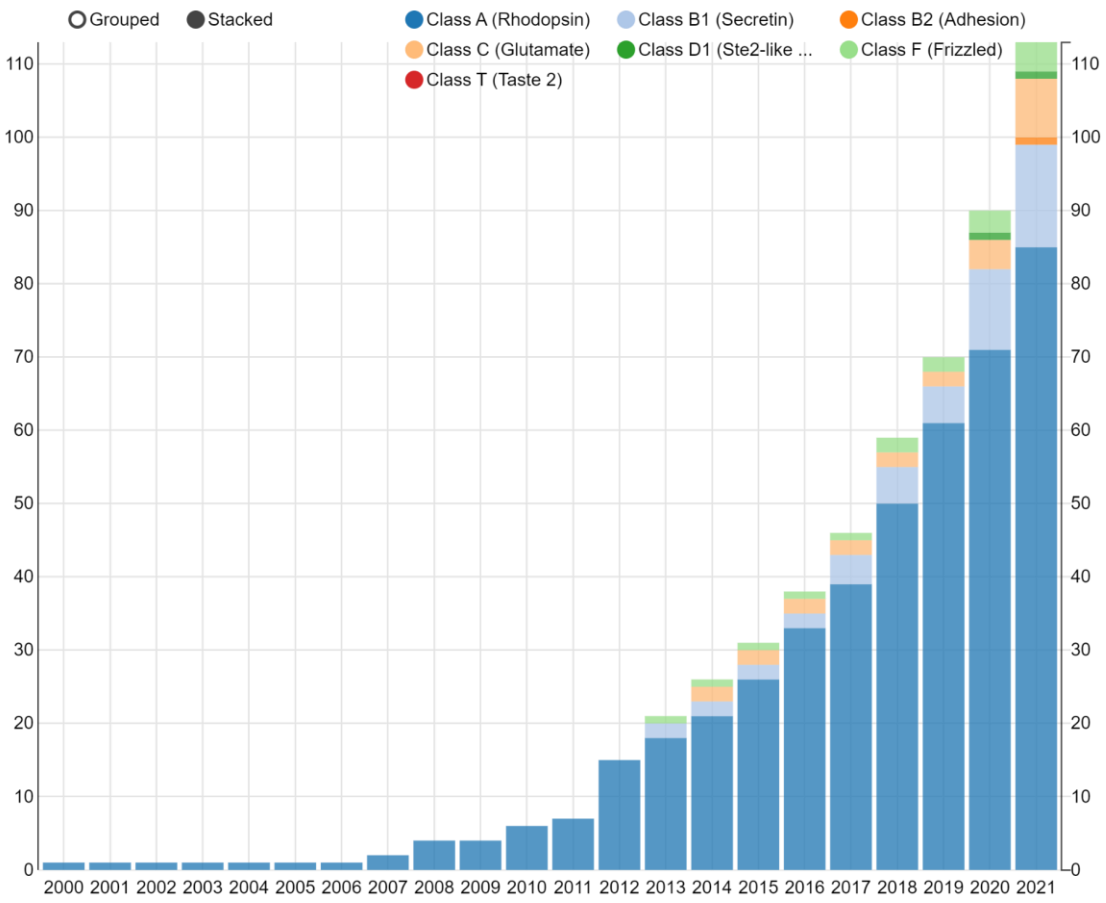


Figure 1.5-4. Number of unique receptors with structures of GPCRs by years. From (GPCRdb, 2021).

1.5.3 GPCR oligomerization

GPCRs interact with additional proteins to selectively modulate different intracellular signal transduction pathways. One case is the formation of GPCR homo-, hetero-dimers or higher-order oligomers that give opportunities of allosteric modulation between the protomers controlling their function in the active state (Han et al., 2009). They constitute novel signalling units with unique functional and regulatory properties opening new opportunities for drug discovery (Ferré, 2015; Ferré et al., 2014; Franco et al., 2016; Fuxe et al., 2010). These complexes have been thoroughly characterized by biophysical and biochemical methods (Guo et al., 2017; Sleno & Hébert, 2018; Xue et al., 2019) but the structural basis behind the allosteric communications between receptor protomers remains poorly understood. The pharmacological potential of such complexes is hampered by the limited information available on the type of complex

formed. A key aspect to understand their functionality is characterizing interaction interfaces between protomers (see Figure 1.5-5).

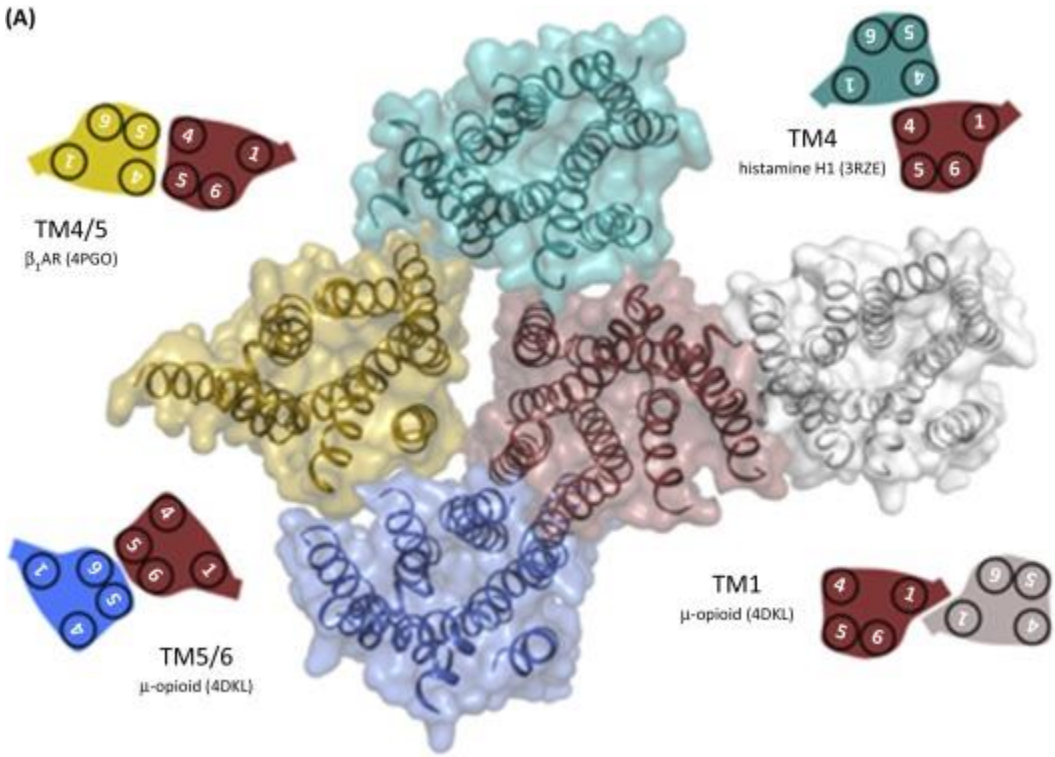


Figure 1.5-5. Extracellular view of representative receptor pairs superposed on a central protomer (in red). Interacting (symmetric) interfaces that may form a dimer are: TM1 (white), TM4 (gray), TM4/5 (yellow), or TM5/6 (blue) from reference Cordomí 2015 TIBS.

High-resolution crystal and cryo-EM structures available for GPCRs often reveal intimate association between monomers through the TM domains. These associations are compatible with the spatial restrictions imposed by the membrane and with previous experiments investigating the topology of such complexes. With only a few exceptions, these receptor pairs have twofold rotational symmetry; that is, they form head-to-head interfaces that occur mainly through TMs 1, 4, 4/5, and 5/6. Regardless of whether dimers and oligomers adopt these structural arrangements in vivo, there is no doubt that they constitute high-resolution structural information that can facilitate the task of modeling GPCR dimers and/or oligomers (Jastrzebska & Park, 2019).

In the last decade, there have been an increasing number of publications describing these complexes. The open question is which is the mechanism to integrate a new intracellular signal from the association of GPCRs protomers. Several studies have shown that GPCR heterodimerization affects pharmacological and cellular signaling responses compared to different receptor subtypes that are expressed alone. We aim to characterize the molecular architecture and the dynamic and functional properties of GPCR heteromers to understand the interaction between different components at the molecular level (see section 4.1).

1.5.4 GPCRs diseases and therapeutic strategy

As it described at the start of this section GPCRs play an important role in the mediate of several physiological functions, interact with a variety of ligands and they are involved in the transmission of extracellular signals inside the cell. For this reason, the mutations that cause damage to these receptors derive into serious diseases, such as neurodegenerative and cerebrovascular like Alzheimer's, Parkinson's and stroke. These neurological disorders occur as a result of neurodegenerative processes and represent one of the most frequent causes of death and disability worldwide with a significant clinical and socio-economic impact. (Guerram et al., 2016). The mutations in GPCR genes can severely alter their normal function and for example affect the exchange of GDP to GTP in the α-subunit of heterotrimeric G proteins. These mutations in GPCRs can lead to the inactivation of their activity (LoF) or ligand-independent activation of the activity (GoF) as the most frequent alterations (Schöneberg & Liebscher, 2021). Because of this pathogenesis the researchers highlight them as potential therapeutic targets for diseases (Hauser et al., 2017; Sriram & Insel, 2018). Currently, there are over 2 000 drug and in-trial agents (Kooistra et al., 2021).

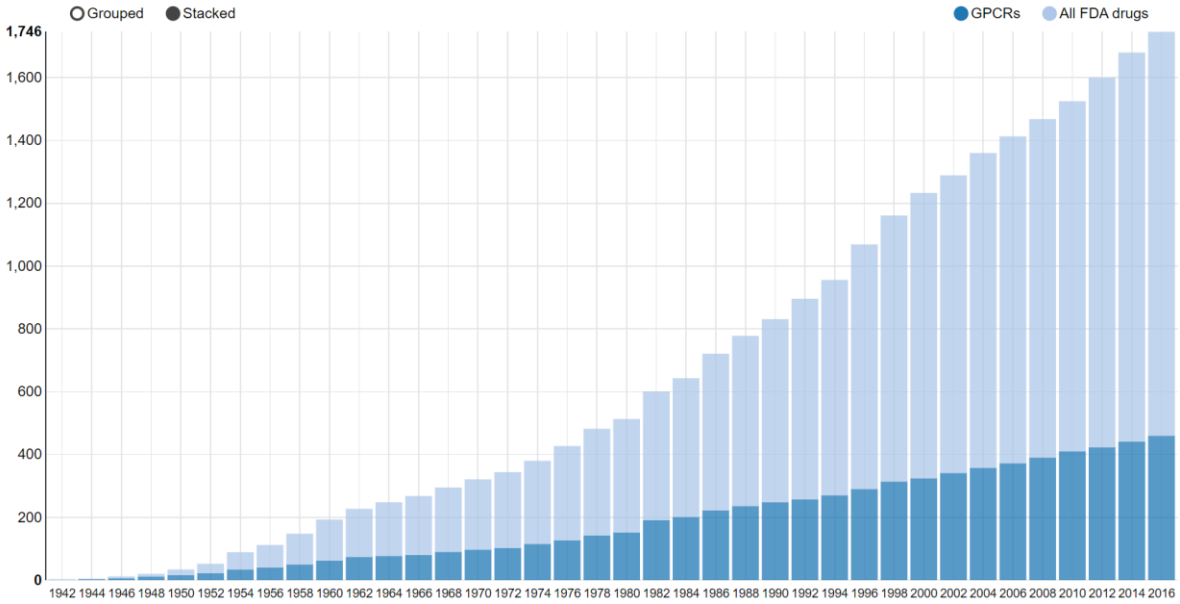


Figure 1.5-6. Number of drugs with annotated approval year related to GPCRs by years. From (GPCRdb, 2021).

With the advent of new transgenic and sequencing technologies, the number of monogenic diseases related to GPCR mutants has significantly increased, and our understanding of the functional impact of certain kinds of mutations has substantially improved (Schöneberg & Liebscher, 2021). To achieve this improvement researchers, need good model references and updated data of the GPCR. For this reason, the databases and computational tools related with GPCR are essential to the understanding of these receptors and help in the therapeutic strategies.

Chapter 2 - Objectives

Chapter 2 - Objectives

The main goal of this thesis is to develop bioinformatics tools to facilitate characterizing the structure and function of membrane proteins from their sequences and structures. With this ultimate purpose in mind, the following objectives were defined by parts.

Objectives for all membrane proteins:

- Develop a tool based on parameters related to sequence conservation to predict if a mutation into the transmembrane region of a membrane protein is pathogenic or not using machine learning.

Specific objectives for NMDARs:

- GRINdb: Create a database of all identified GRIN variants including its genetic, functional, clinical and structural information in order to stratify the variants.
- Exhaustive analysis of all GRIN truncating variants in order to predict the pathogenesis of any GRIN truncating variant based on subunit and sequence position.
- Construct an algorithm able to predict pathogenesis based on annotated homologous mutations.

Specific objectives for GPCRs:

- Develop a tool aimed to place internal water molecules in GPCR models based on the positions of other water molecules determined on homologous structures.
- Develop a tool to browse and compare crystallographic interfaces in GPCR structures as a tool to develop structural models of heteromers.

Chapter 3 - Methods

Chapter 3 - Methods

The general methods employed in this thesis. Full details are available in the corresponding study (see section 4.3).

3.1 Bioinformatics and biomedical databases

Like all fields of science, there has always been a reference, considered as the father or the mother of the field. Bioinformatics grew linked to the progress on computer-related technologies to find solutions to biological hypotheses. Margaret Dayhoff pioneered the use of computational tools in the comparison of protein sequences and the reconstruction of their evolutive histories using sequence alignments. This fact gave birth to Bioinformatics (McNeill, 2019). Margaret Dayhoff did a lot of crucial work and the first one considered the first database is the publication of books named “Atlas of Protein Sequence and Structure” containing the 65 protein sequences available at the moment. In 1984, the National Biomedical Research Foundation launched a free online database named Protein Information Resource containing over 283,000 protein sequences (Wu et al., 2003). Today the Protein Information Resource allows scientists all over the world to take an unknown protein, compare it to the thousands of known proteins in the database, and determine the ways in which it is alike and different.

Here remains the importance of the databases, but also their update and amplification over the years, the improvement of new technologies to organize and store data are crucial to maintain them. One of the big references related to databases and computational tools is the **National Center for Biotechnology Information** (NCBI) that develops information systems for molecular biology (Wheeler, 2004). This center is in collaboration with the DNA Database of Japan (DDBJ) and the European Molecular Biology Laboratory Nucleotide Sequence Database (EMBL-Bank) as well as from the scientific community. This collaboration provides data retrieval systems and computational resources as GenBank (Benson et al., 2008) and many other kinds of biological data, such as PubMed (*PubMed*, 2021), BLAST (Johnson et al., 2008), RefSeq (Johnson et al., 2008; Pruitt et al., 2007), dbSNP (Smigielski et al., 2000),

Online Mendelian Inheritance (Hamosh et al., 2000), Ensembl (Howe et al., 2021), UniProt (The UniProt Consortium et al., 2020)), among others.(Pruitt et al., 2007)

The massive amounts of data create challenges not only for data storage but also for organization, accessibility, data mining and analysis (Zheng et al., 2017). These facts are needed to think about it for the creation of big databases. The following freely available and public databases have been particularly relevant for this thesis.

ClinVar: an archive of human genetic variants and interpretations of their relationships to diseases and other conditions, maintained at the National Institutes of Health (NIH). ClinVar accessions submissions reporting human variation, interpretations of the relationship of that variation to human health and the evidence supporting each interpretation. The database (https://www.ncbi.nlm.nih.gov/clinvar/) is tightly coupled with dbSNP and dbVar, which maintain information about the location of variation on human assemblies. (Melissa J. Landrum et al., 2014; M. J. Landrum et al., 2020).

Gnomad: is currently the largest and most widely used available collection of population variation from harmonized sequencing data. The data is available through the online gnomAD browser (https://gnomad.broadinstitute.org/) (Gudmundsson et al., 2021). Includes genetic variation for 15,708 whole genomes in addition to 125,748 exomes yielding more than 240 million small genetic variants as well as structural variation (Koch, 2020).

PFAM: database of protein families and domains that is widely used to analyze novel genomes, metagenomes and to guide experimental work on particular proteins and systems, available at http://pfam.xfam.org/ (Mistry et al., 2020).

Protein Data Bank (PDB): a single worldwide archive of structural data of biological macromolecules (http://www.rcsb.org/pdb/) (Berman et al., 2000).

Universal Protein Resource (UniProt): databases to support biological and biomedical research by providing a complete compendium of all known protein sequence data linked to a summary of the experimentally verified, or computationally predicted, functional information about that protein (The UniProt Consortium et al., 2020).

Compared to the databases exposed above, the next databases are more specific:

Leiden Open-source Variation Database (LOVD): is genome-centered and can be used to store summary variant data, as well as full case-level data with information on individuals, phenotypes, screenings, and variants. While built on a standard core, the software is highly flexible and allows personalization to cope with the largely different demands of gene/disease database curators. LOVD follows current standards and includes tools to check variant descriptions, generate HTML files of reference sequences, predict the consequences of exon deletions/duplications on the reading frame, and link to genomic views in the different genome's browsers. It includes APIs to collect and submit data (https://www.lovd.nl/) (Fokkema et al., 2021).

GPCRdb: is a G protein-coupled receptor database system aimed at the collection and dissemination of GPCR related data. It holds sequences, mutant data and ligand binding constants as primary (experimental) data. Computationally derived data such as multiple sequence alignments, three dimensional models, phylogenetic trees and two-dimensional visualization tools are available (https://gpcrdb.org/). (Horn et al., 1998).

Orientations of Proteins in Membranes (OPM): database that provides a collection of transmembrane, monotopic and peripheral proteins from the PDB whose spatial arrangements in the lipid bilayer have been calculated theoretically and compared with experimental data (Lomize et al., 2006).

3.2 Sequence alignment methods

Basically, sequence alignment is a way of arranging sequences, such as DNA, RNA or protein, to identify the connection between two or more sequences and regions of similitude that may be a consequence of functional, structural, or evolutionary relationships between the sequences (M. Huang et al., 2019). Computational approaches to sequence alignment generally fall into two categories, global alignments and local alignments. Pairwise alignment, used to find the best match pairing alignment of only two sequences. Here, there are three primary methods of producing pairwise alignments: dot-matrix methods, dynamic programming (global

alignment Needleman-Wunsch algorithm (Needleman & Wunsch, 1970) and local alignment Smith-Waterman algorithm (Smith & Waterman, 1981)) and word methods.

Basic Local Alignment Search Tool (BLAST): is an algorithm and program for comparing primary biological sequence information, such as the amino-acid sequences of proteins or the nucleotides of DNA and/or RNA sequences. A BLAST search enables a researcher to compare a subject protein or nucleotide sequence (called a query) with a library or database of sequences and identify database sequences that resemble the query sequence above a certain threshold. For example, following the discovery of a previously unknown gene in mouse, a scientist will typically perform a BLAST search of the human genome to see if humans carry a similar gene; BLAST will identify sequences in the human genome that resemble the mouse gene based on similarity of sequence (Boratyn et al., 2013).

Multiple sequence alignment (MSA), which is an extension of pairwise alignment incorporating more than two sequences and has assumed a key role in comparative structure and function analysis of biological sequences. It often guides fundamental biological understanding into the sequence-structure-function relationships of nucleotide or protein sequence families (Nuin et al., 2006). This method led to the creation of phylogenetic trees that show the evolutionary relationships between sequences and describe which parts of the sequence are highly conserved and important for the function of the protein and many useful tools have been developed for constructing alignments. (Pirovano & Heringa, 2008).

PFAM alignment: a database of protein families that includes their annotations and multiple sequence alignments generated using hidden Markov models (Finn et al. 2008).

Predicted hydrophobic and transmembrane (PHAT) 75/73 matrix: a matrix specific for membrane proteins. (Ng et al. 2000).

3.3 Structural Bioinformatics

In the case of membrane proteins, due to the reduced number of determined structures compared to globular proteins because of difficulties associated with membrane

proteins, computational tools have also become essential for understanding the structure and function of these proteins (Gromiha & Ou, 2014). For this reason, some of the following computational tools help in the visualization and analysis of the structure of the proteins.

PyMOL: is an open-source molecular visualization system created by Warren Lyford DeLano. This tool can produce high-quality 3D images of small molecules and biological macromolecules, such as proteins. PyMOL is one of the few open-source model visualization tools available for use in structural biology. The Py part of the software's name refers to the program having been written in the programming language Python (Schrödinger, 2021).

Visual Molecular Dynamics (VMD): is a molecular modelling and visualization computer program. VMD is developed mainly as a tool to view and analyze the results of molecular dynamics simulations. It also includes tools for working with volumetric data, sequence data, and arbitrary graphics objects. Users can run their own Tcl and Python scripts within VMD as it includes embedded Tcl and Python interpreters ("VMD: Visual Molecular Dynamics," 1996).

GROMACS: is a molecular dynamics package. Molecular dynamics is defined as the calculations used with computational tools to create a simulated system of proteins, lipids, water and ions as a close example of a biological system. It was originally developed in the Biophysical Chemistry department of University of Groningen and is now maintained by contributors in universities and research centers worldwide. GROMACS is one of the fastest and most popular software packages available (Abraham et al., 2015).

3.4 Programming languages, databases, and web applications

It is known that the bioinformatic tools help the researchers to take another point of view of the data that they are studying, analyze big amounts of data at the same time and get the results fast. For this reason, we used this big potential of the computational tools to study the membrane proteins. At present, many bioinformatics tools are made available in the form of a web application facilitating the use of the tools by other

scientists. In our study, we used mainly Python 3 but also, we used some support languages for the web and server management like: HTML, CSS, JavaScript and Bash. Besides the base web server software, we also need a web framework to create and manage the applications in the server. These frameworks are a collection of packages or modules which allow the developers to create web applications without having to handle in detail with protocols, sockets or processes management, like an assistant. The detailed methodology is explained in each of the studies but some of the general tools used are explained next.

Python: is an interpreted high-level general-purpose programming language designed to help programmers write clear, logical code for all types of projects (Kuhlman, 2011). Python was created in the 1980s and today is one of the most popular programming languages (*PYPL PopularitY of Programming Language*, 2021, *The State of Developer Ecosystem 2021*, 2021). In this thesis, we used Python not only for its clear and logical structure, but for the big number of modules and packages, created by the big community associated with this language, that support this language to become a potent tool to analyze data.

Flask: is a web framework for Python based on Werkzeug and Jinja 2, including a built-in development server, unit testing support, and Web Server Gateway Interface (WSGI) to manage the communication between the web server software and the web application, mainly known as requests.

MySQL: is an open-source relational database management system (RDBMS). A relational database organizes data into one or more data tables in which data types may be related to each other; these relations help structure the data. SQL is a language programmer use to create, modify and extract data from the relational database, as well as control user access to the database. In addition to relational databases and SQL, an RDBMS like MySQL works with an operating system to implement a relational database in a computer's storage system, manages users, allows for network access, and facilitates testing database integrity and creation of backups.

Apache: is a free and open-source web server software started in 1995 and it has versions for each of the most used operating systems, like Microsoft Windows or Unix-

like systems, as MacOS and Linux. As any other type of software is on constant updating and we specifically have installed the version Apache2 in a based Linux-Ubuntu-Server operating system to create the base of our web server

Chapter 4 - Results

Chapter 4 - Results

4.1 List of articles derived from this thesis

This chapter contains the 6 articles performed during the thesis isolated into their correspondent topic.

I. **TMSNP: a web server to predict pathogenesis of missense mutations in the transmembrane region of membrane proteins.**
Garcia-Recio A, Gómez-Tamayo JC, Reina I, Campillo M, Cordomí A, Olivella M. NAR Genom Bioinform. 2021 Feb 23;3(1):lqab008. doi: 10.1093/nargab/lqab008. PMID: 33655207; PMCID: PMC7902201.

Contribution: Creation of the database of pathogenic and non-pathogenic mutations in transmembrane proteins, the filtering, homology reduction and dataset balancing, the feature extraction and the development of the web application to run the tool online.

II. **GRIN database: A unified and manually curated repertoire of GRIN variants.**
García-Recio A, Santos-Gómez A, Soto D, Julia-Palacios N, García-Cazorla À, Altafaj X, Olivella M. Hum Mutat. 2021 Jan;42(1):8-18. doi: 10.1002/humu.24141. Epub 2020 Nov 30. PMID: 33252190.

Contribution: Creation of the database by the data curation and functional classification available, develop a web server service to display the database into a web and the display of the mutations into the NMDA receptor model generated.

III. **Identification of homologous GluN subunits variants accelerates GRIN variants stratification**
Ana Santos-Gómez, Adrián García-Recio, Federico Miguez-Cabello, David Soto, Xavier Altafaj, and Mireia Olivella Hum Mol Genet. 2021 Oct. Under review

Contribution: Perform the comparative analysis of pathogenicity and functional

changes for equivalent subunits positions using the database presented in the article IV.

IV. **Disease-associated GRIN protein truncating variants trigger NMDA receptor loss-of-function.**
Santos-Gómez A, Miguez-Cabello F, García-Recio A, Locubiche-Serra S, García-Díaz R, Soto-Insuga V, Guerrero-López R, Juliá-Palacios N, Ciruela F, García-Cazorla À, Soto D, Olivella M, Altafaj X. Hum Mol Genet. 2021 Feb 25;29(24):3859-3871. doi: 10.1093/hmg/ddaa220. PMID: 33043365.

Contribution: Compiling and analyzing the database of GRIN protein truncating variants provided in the article IV about the genetic and clinical data for each subunit.

V. **HomolWat: a web server tool to incorporate 'homologous' water molecules into GPCR structures.**
Mayol E, García-Recio A, Tiemann JKS, Hildebrand PW, Guixà-González R, Olivella M, Cordomí A. Nucleic Acids Res. 2020 Jul 2;48(W1):W54-W59. doi: 10.1093/nar/gkaa440. PMID: 32484557; PMCID: PMC7319549.

Contribution: Participate in the development of the reference database and in the pipeline of the tool and did the implementation of the web service

VI. **DIMERBOW: exploring possible GPCR dimer interfaces.**
García-Recio A, Navarro G, Franco R, Olivella M, Guixà-González R, Cordomí A. Bioinformatics. 2020 May 1;36(10):3271-3272. doi: 10.1093/bioinformatics/btaa117. PMID: 32096817.

Contribution: Generation of the database of putative dimers, coarse-grained MD simulations, perform the analysis of the structures automatizing the pipeline using scripts and develop the web application to show the data.

4.2 Overview

Membrane proteins have been a challenge to researchers, but the field of Bioinformatics have paved the way to simplify their analysis and improve our

understanding, and this thesis is proof of this. To improve the prediction power of the actual methods used by the actual mutation predictors, we created a database of reported pathogenic and non-pathogenic missense mutations in the transmembrane region and developed a mutation prediction tool (TMSNP) able to predict pathogenicity for previously non-reported transmembrane missense mutations (Article I). Both the database and the predictor tools are freely available at http://lmc.uab.cat/tmsnp. Unlike this first study focused on all membrane proteins, the next studies (**Articles II - VI**) are related with two specific membrane protein families: N-methyl-D-aspartate receptors (NMDARs) and G protein-coupled receptors (GPCRs). Still, the work performed could be enlarged to other proteins as well.

In the field of NMDARs, mutations can affect the folding, stability and function of these receptors arriving to cause the malfunction of the ion channel and change of the flux of ions with the possibility of causing a neurological disease. For this reason, the annotations about clinical, genetic and functional information are important to develop therapeutic strategies, but this data is highly fragmented. Because of that we developed GRINdb, available at http://lmc.uab.cat/grindb, a public updated, manually curated, and nonredundant database comprising genetic, clinical, functional, and structural information of the largest repertoire of GRIN variants (**Article II**). GRINdb enabled the possibility of addressing two further studies accelerating the stratification of the GRIN variants to reduce the number of uncertain annotated variants enabling an improvement in the personalized treatments (**Article III and IV**).

Current methods to determine structures do not always reach enough resolution to determine the positions of internal water molecules. This is also true for GPCRs for which various tens of structures have been determined in the last 15 years. We develop HomolWat (available at http://lmc.uab.cat/HW) a open web application to introduce internal water molecules in GPCR structures using resolved water molecules from homologous structures (**Article V**). Also related to improving structural models for GPCRs, oligomers have become a target for the scientist, trying to find the best possible structural template to model GPCR dimeric structures or to identify the frequency of a specific dimeric interface. We developed DIMERBOW (available at http://lmc.uab.cat/dimerbow) a open web application for rapid, intuitive and systematic

visualization of the complete repertoire of crystallographic interfaces for GPCR dimers available in the Protein Data Bank (**Article VI**).

4.3 Articles

Article I. TMSNP: a web server to predict pathogenesis of missense mutations in the transmembrane region of membrane protein

ABSTRACT

The massive amount of data generated from genome sequencing brings tons of newly identified mutations, whose pathogenic/non-pathogenic effects need to be evaluated. This has given rise to several mutation predictor tools that, in general, do not consider the specificities of the various protein groups. We aimed to develop a predictor tool dedicated to membrane proteins, under the premise that their specific structural features and environment would give different responses to mutations compared to globular proteins. For this purpose, we created TMSNP, a database that currently contains information from 2624 pathogenic and 196 705 non-pathogenic reported mutations located in the transmembrane region of membrane proteins. By computing various conservation parameters on these mutations in combination with annotations, we trained a machine-learning model able to classify mutations as pathogenic or not. TMSNP (freely available at http://lmc.uab.es/tmsnp/) improves considerably the prediction power of commonly used mutation predictors trained with globular proteins.

INTRODUCTION

Whole genome and exome sequencing have revealed that Mendelian rare disease-causing missense mutations are more frequent than previously thought and collectively affect millions of patients worldwide (1). Thus, there is an urgent need to understand the relation between genotype and phenotype in order to identify disease-causing genetic variants within candidate variants. For this purpose, variant prioritization tools are widely used to predict the effect of mutations. These are mostly based on evolutionary conservation and expected impact on structure and function

using evolutionary conservation parameters and physico chemistry properties of amino acids from sequence data [SIFT (2), Provean (3), MutationTaster (4)], while some tools such as Polyphen-2 (5) also incorporate features related to structural data [see (6) for a review].

Membrane proteins represent 25% of all human proteins (7) and perform essential roles in cellular functions (8). Consequently, they are the target for 50% of drugs in the market (9). Moreover, 90% of membrane proteins present disease associated missense mutations that may affect protein folding, stability and/or function (10). Some of them have been related to various diseases, including cardiopathies, neurological diseases, cystic fibrosis and cancer (11,12). In fact, mutations in membrane proteins are more likely to cause diseases than in globular proteins (13). Membrane proteins differ from globular proteins in terms of amino acid composition, distribution, inter-residue interactions and structure (14,15). The main differences are in the transmembrane (TM) region of the proteins because of the different surface environments, that is lipid exposed versus water exposed. Current variant prioritization tools, which are mainly based on data from globular proteins, present low reliability for predicting the pathogenicity of mutations in membrane proteins (13). Thus, there is a need for computational tools specific for membrane proteins to understand its relation between sequence and structure (16). This is especially important given the scarce number of membrane protein structures compared to globular proteins due to experimental limitations (17). Mutation prediction tools and databases specific for membrane proteins are starting to emerge, such as Mut-HTP (18), Pred-MutHTP (10) or BorodaTM (19), which are all based on evolutionary conservation parameters, the former including structure descriptors and the latter focusing on regions with known structure.

With the aim of contributing to emerging mutation predictor servers for the TM region of membrane proteins, here we present TMSNP (accessible at http://lmc.uab.es/tmsnp/), a database of TM missense mutations (pathogenic and non-pathogenic) and a predictor server trained using evolutionary conservation parameters.

MATERIALS AND METHODS

Database of pathogenic and non-pathogenic mutations in transmembrane proteins

Our selected set of membrane proteins consisted of all human membrane proteins tagged as reviewed in the UniProt database (20,21). For each protein, we retrieved all disease-causing/pathogenic mutations associated with Mendelian disorders as reported in ClinVar (22) and SwissVar (23). We only kept mutations occurring in the TM helices because these are the regions that mostly differ from globular proteins. The ranges of TM segments were taken from the UniProt database (20,21). We also retrieved non-pathogenic missense mutations and their population allele frequency from GnomAD (24) and ClinVar (22). The database (accessible at http://lmc.uab.es/tmsnp/tmsnpdb) resulted in 196 705 non-pathogenic, 2624 pathogenic and 437 likely pathogenic mutations in the TM region of membrane proteins.

TMSNP predictor

Filtering, homology reduction and dataset balancing. To ensure that mutations used in the machine-learning models were linked to protein function and/or structure alteration, we discarded all mutations in proteins for which no single pathogenic disease-causing mutations have been reported, that is those likely involved in complex diseases or recessive inheritance and tagged as ‘non-pathogenic proteins’ (25). Thus, the obtained pathogenic and non-pathogenic missense mutations were used to classify human TM proteins as ‘pathogenic proteins’ (358 proteins), when at least one disease-causing pathogenic mutation has ever been reported for this protein and as ‘non-pathogenic proteins’ (2420 proteins), elsewhere. We next performed homology reduction by discarding mutations for proteins belonging to the same Pfam family (26) that resulted in the same amino acid change in the same aligned position. The dataset after homology reduction contained 2704 pathogenic and likely pathogenic mutations and 19 292 non-pathogenic mutations. Data were subsequently subsampled to obtain a balanced dataset (50% pathogenic and 50% non-pathogenic mutations), by selecting

the non-pathogenic mutations with the highest population allele frequency according to GnomAD (24). The final dataset used for training in the machine-learning model presented 5408 missense mutations, which implies a reduction of non-pathogenic mutations by 1/6.

Feature extraction*.* Multiple sequence alignments for the different families of the proteins in our dataset were taken from the Pfam database (26). For each missense mutation in the balanced dataset, we computed four variables related to evolutionary conservation and the likelihood that an amino acid change is tolerated in a position: (i) frequency of the wild-type amino acid in the Pfam alignment, (ii) frequency of the mutated amino acid, (iii) substitution matrix score (as a measure of similar physicochemical properties between wild-type and mutated amino acid) and (iv) entropy of the position (as a measure of sequence variability or information content) (27). For the substitution matrix score, we used the PHAT 75/73 matrix, which is specific for membrane proteins (28). The entropy of the position *i* is maximal (= 1) if all 20 amino acids at the position *i* present equal frequencies and is minimal (= 0) if only 1 amino acid has been observed at this position. Four additional variables: type of the reference and the mutated amino acids, Pfam and UniProt accession codes were included through one-hot-encoding of the qualitative variables. In the final dataset, each missense mutation had information encoded in eight variables contributing to a total of 569 features (20 features for reference and 20 for the mutated amino acids, 358 features for the UniProt accession codes and 167 for the Pfam accession codes).

Machine-learning models*.* We built three datasets from the 5408 missense mutations (see http://lmc.uab.es/tmsnp/datasets): (i) 8V dataset, containing all variables (569 features); (ii) 6V dataset, lacking UniProt and Pfam accession code variables, which are informative of the tendency of a protein or a protein family to pathogenesis, but still including one-hot encoded reference and mutated amino acids (60 features) and (iii) 4V dataset containing only conservation variables (wild-type and mutated frequencies, substitution matrix score and entropy). For each dataset, five different training (80%) and test (20%) sets were created by random sampling under certain restrictions for internal validation. For external validation, in the 8V dataset and 4V datasets, mutations with the same UniProt code were equally split between training set and test

set while for the 6V dataset Pfam accession codes were used to split mutations either in the training ser or in the validation set. Machine-learning models were built using Flame (https://github.com/phi-grib/flame; a Python modeling framework which wraps scikit learn (http://scikit-learn.sourceforge.net)) or Keras (https://keras.io/). Various predictive models using different algorithm settings, applicability domain and dataset were built and internally validated using *K*-fold (*K* = 5) cross validation. Specifically, we used Random Forest (RF), Gradient Boosting (XGBoost), Supporting Vector Machines (SVM) and a sequential neural network. The conformal prediction was used as an applicability domain technique (29) by testing our models at three different confidences: 95, 90 and 80%.

Web server

TMSNP web application tool was constructed using a Python backend (v.3.7) with the Flask framework (v.1.0.2). Both the application and the associated datasets used for training and testing the predictor were built automatically using Python/Bash scripts that collected the required data and stored it in a MySQL database (v.8.0.18), facilitating regular updates. All scripts can be found in a GitHub repository (https://github.com/adriangarciarecio/TMSNP).

RESULTS AND DISCUSSION

We initially constructed a database of missense mutations in human membrane proteins that exclusively focused on TM helices (accessible at http://lmc.uab.es/tmsnp/tmsnpdb; see ‘Materials and Methods’). The database currently contains 2624 pathogenic, 437 likely pathogenic and 196 705 non-pathogenic mutations. We used a subset of this database (see ‘Materials and Methods’) to develop machine-learning models able to classify mutations as pathogenic or not. We assessed three different algorithms (RF, SVM and XGBoost) and three different datasets (8V, 6V and 4V). All combinations showed good performance in both internal (5-fold cross-validation; Table 4.3-1) and external validation (independent 20% test set; Table 4.3-2) with none of them clearly outperforming the other two. RF showed the best performance on the 4V dataset, while XGBoost and SVM were the best on 6V and 8V,

respectively. Although we also tested a sequential neural network, we could not find advantages of using this method despite performance being close to the other algorithms used in this study. Models that use only four features reach a maximum accuracy of ~70%, without the increase in the confidence of the model bringing additional improvement. The two additional features (type of reference and mutated residues) included in the 6V dataset increase the average performance up to ~80% accuracy (at the maximum confidence) except for SVM which keeps at ~70%. The 8V dataset clearly shows the best performance both in internal cross validation and external validation. SVM provides the best models, which reach 94% accuracy with 46% coverage (95% confidence), or 85% with 86% coverage (80% confidence). XGBoost and RF models follow closely although with slightly worse performance (<5%).

Dataset	Algorithm	Confidence	Sensitivity	Specificity	MCC	Coverage	Accuracy
8V (569 features)	RF	0.95	0.92	0.88	0.80	0.42	0.90
	RF	0.90	0.89	0.83	0.72	0.62	0.86
	RF	0.80	0.82	0.77	0.58	0.88	0.79
	XGBOOST	0.95	0.96	0.93	0.89	0.39	0.94
	XGBOOST	0.90	0.92	0.88	0.80	0.58	0.90
	XGBOOST	0.80	0.85	0.81	0.66	0.86	0.83
	SVM	0.95	0.96	0.93	0.89	0.46	0.95
	SVM	0.90	0.92	0.90	0.82	0.63	0.91
	SVM	0.80	0.88	0.84	0.71	0.86	0.85
6V (44 features)	RF	0.95	0.91	0.71	0.63	0.13	0.81
	RF	0.90	0.87	0.72	0.60	0.29	0.79
	RF	0.80	0.79	0.67	0.46	0.58	0.72
	XGBOOST	0.95	0.88	0.73	0.60	0.08	0.79
	XGBOOST	0.90	0.81	0.74	0.55	0.25	0.77
	XGBOOST	0.80	0.76	0.69	0.45	0.57	0.72
	SVM	0.95	0.90	0.77	0.68	0.08	0.84
	SVM	0.90	0.81	0.74	0.55	0.25	0.77
	SVM	0.80	0.76	0.69	0.45	0.57	0.72
4V (4 features)	RF	0.95	0.86	0.59	0.46	0.12	0.72
	RF	0.90	0.77	0.64	0.42	0.29	0.71
	RF	0.80	0.71	0.63	0.34	0.57	0.67
	XGBOOST	0.95	0.87	0.72	0.59	0.09	0.79
	XGBOOST	0.90	0.82	0.72	0.54	0.21	0.76
	XGBOOST	0.80	0.73	0.67	0.40	0.48	0.70
	SVM	0.95	0.93	0.30	0.29	0.15	0.71
	SVM	0.90	0.90	0.41	0.37	0.34	0.69
	SVM	0.80	0.73	0.65	0.38	0.67	0.69

Table 4.3-1. Model statistics in cross-validation. The table shows quality metrics (5-fold) for the machine-learning models created using 8V, 6V and 4V datasets with different conformal significance. MCC stands for Matthews correlation coefficient, which is a measure that combines sensitivity and specificity. Coverage stands for the percentage of samples inside the applicability domain.

In order to check for possible overfitting of the models using the largest (8V) dataset, we performed feature selection for the three different algorithms. Supplementary Table S1 compares the performance of the models with the original and the reduced features using *K*-best feature selection performed with K = 60 or 30 (number of variables

reduced to 10% and ~5%, respectively). The mean accuracy loss for RF, XGBoost and SVM was, respectively, 2%, -2% and 3% for *K* = 60, and 4%, 0% and 6% for *K* = 30. These small differences suggest lack of overfitting and also point out that SVM is less robust than RF or XGBoost algorithms. In order to assess the presence of bias due to dataset balancing (see 'Materials and Methods'), we generated an additional dataset containing the first 3000 non-pathogenic mutations following those used in training. This bias might lead to unrealistic predictions, possibly translated to an excess of false positives. Supplementary Table S2 shows prediction results using 8V models at 95%, 90% and 80% confidence. Minimum true negatives/false positives ratio is ~4, being most of the predictions either negatives or out of the applicability domain and always sticking to the confidence restraints.

Dataset	Algorithm	Confidence	Sensitivity	Specificity	MCC	Coverage	Accuracy
8V (569 features)	RF	0.95	0.90	0.86	0.76	0.38	0.88
	RF	0.90	0.86	0.82	0.68	0.58	0.84
	RF	0.80	0.81	0.75	0.56	0.86	0.78
	XGBOOST	0.95	0.90	0.88	0.78	0.30	0.89
	XGBOOST	0.90	0.85	0.84	0.70	0.48	0.85
	XGBOOST	0.80	0.79	0.78	0.57	0.77	0.78
	SVM	0.95	0.91	0.89	0.80	0.39	0.90
	SVM	0.90	0.86	0.85	0.71	0.55	0.85
	SVM	0.80	0.77	0.79	0.56	0.81	0.78
6V (44 features)	RF	0.95	0.86	0.66	0.54	0.78	0.74
	RF	0.90	0.81	0.69	0.51	0.76	0.73
	RF	0.80	0.76	0.66	0.42	0.71	0.70
	XGBOOST	0.95	0.87	0.69	0.58	0.09	0.80
	XGBOOST	0.90	0.82	0.70	0.53	0.24	0.77
	XGBOOST	0.80	0.77	0.68	0.45	0.54	0.73
	SVM	0.95	0.90	0.77	0.68	0.08	0.84
	SVM	0.90	0.81	0.74	0.55	0.25	0.77
	SVM	0.80	0.76	0.69	0.45	0.57	0.72
4V (4 features)	RF	0.95	0.85	0.60	0.47	0.13	0.75
	RF	0.90	0.79	0.66	0.46	0.29	0.73
	RF	0.80	0.71	0.67	0.37	0.56	0.69
	XGBOOST	0.95	0.79	0.65	0.45	0.09	0.73
	XGBOOST	0.90	0.76	0.67	0.43	0.21	0.72
	XGBOOST	0.80	0.69	0.66	0.35	0.47	0.68
	SVM	0.95	0.92	0.38	0.37	0.14	0.74
	SVM	0.90	0.84	0.55	0.41	0.35	0.73
	SVM	0.80	0.72	0.68	0.40	0.67	0.70

Table 4.3-2. Model statistics at external validation. The table shows performance metrics in external validation (20% of the original dataset) for the machine-learning models created using 8V, 6V and 4V datasets with different conformal significance. MCC and coverage are described in Table 4.3-1.

These results demonstrate lack of sampling bias. Interestingly, the RF model at 95% confidence correctly predicts 658 non-pathogenic mutations with only 6 false positives out of 3000 non-pathogenic mutations.

Random Forest 8V was selected as the final model to be implemented in the web application. Although the performance of RF was not the best, it demonstrated to be

the more robust algorithm at different conditions. While XG Boost performed on average better than RF, it provides more importance to protein classification features rather than sequence conservation (amino acid frequencies, substitution matrix score and entropy), questioning its ability to generalize the predictions (Supplementary Table S3). On the other hand, SVM was discarded because it was less robust towards feature reduction and also because the implementation of the SVM algorithm using radial basis function kernel did not allow inspection of feature importance.

Feature importance analysis (Supplementary Table S4) shows that for the original RF model as well as for the 60 and 30 best feature-reduced models, conservation features contribute the most clearly driving the predictive power of the algorithms (30%, 60% and 75% of the total contribution, respectively). Pfam PF00520 (ion channel family) and UniProt P35498 (sodium channel protein) accession codes follow in contribution (∼2% each), probably indicating high sensitivity of this family of receptors to become pathogenic upon a mutation. Mutation to proline and to arginine (P_m and R_m features) appear as the next features in importance (∼1% each). This is compatible with the known distorting effects of these amino acids when present at TM helices (13). Accordingly, the loss of performance for the 6V or 4V dataset might be related to not using UniProt and Pfam codes as features, as they are related to different vulnerability of proteins and/or protein families to amino acid change in their transmembrane region.

TMSNP returns the unambiguous class prediction at the highest confidence possible. Predictions with a confidence below 0.75 are considered outside the domain of applicability. Table 4.3-3 shows the comparison between TMSNP models (8V dataset) generated at three levels of significance together with the results of SIFT, Polyphen-2 and Pred MutHTP. Pred-MutHTP is also specific for membrane proteins, whereas both SIFT and Polyphen-2 are not. As reflected by the equilibrated sensitivity/specificity and the higher accuracy at the different confidence levels, TMSNP not only provides a more balanced model but also performs very well when tested against a full non-pathogenic dataset of 3000 mutations. Importantly, TMPSNP brings higher specificity compared to the other methods. The robustness of the algorithm relies on the quality of the extracted conservation features, which are the most important according to the feature contribution analysis of the model. Noteworthy, TMSNP is using conformal prediction

as an applicability domain and uncertainty framework, providing predictions under confidence restraints. Our models demonstrated to be good at all confidence limits, which translates into the corresponding accuracy (i.e. a confidence of 0.8 sets to ~0.2 the maximum error rate). The higher specificity and accuracy of TMSNP compared to SIFT and PolyPhen-2, which also rely on similar conservation parameters, might be related to using a dataset for training specific for membrane proteins, as the used evolutionary conservation parameters differ between globular and membrane proteins. When compared to the previously reported membrane-specific predictor Pred-MutHTP, the higher specificity and accuracy of TMSNP might be related to the better curated non-pathogenic TM variants as the result of (i) only considering the highest allele frequencies in GnomAD and (ii) discarding non-pathogenic mutations in proteins for which no single causative disease mutation has been identified that might be affecting the structure and function of the protein without being related to pathogenesis (recessive inheritance and complex diseases). Compared to MutHTP database, TM SNP (i) contains mutations in the TM segments of membrane proteins but discards mutations in the extracellular and intracellular regions of membrane proteins, as these regions and domains of the proteins are exposed to an environment similar to globular proteins; (ii) is based on a bigger dataset of non-pathogenic variants as MutHTP does not include variants from GnomAD database (24); and (iii) does not include somatic mutations.

Predictor types	Method	Sensitivity	Specificity	MCC	Coverage	Accuracy
Specific for membrane proteins	**TMSNP** (0.95 confidence)	0.90	0.86	0.76	0.38	0.88
	TMSNP (0.90 confidence)	0.86	0.82	0.68	0.58	0.84
	TMSNP (0.8 confidence)	0.81	0.75	0.56	0.86	0.78
	Pre-MutHTP (0.95 confidence)	0.96	0.54	0.56	0.76	0.64
	Pre-MutHTP (0.90 confidence)	0.96	0.53	0.55	0.76	0.67
	Pre-MutHTP (0.80 confidence)	0.96	0.53	0.56	0.76	0.71
Non-specific for membrane proteins	**Polyhen-2**	0.93	0.35	0.35	1	0.64
	SIFT	0.88	0.52	0.42	1	0.70

Table 4.3-3. Sensitivity, specificity, Matthews correlation coefficient (MCC) and coverage of TMSNP model (8V dataset) and comparison to Pre-MutHTP, SIFT and Poyphen-2. Data are shown at various levels of significance in external validation. MCC and coverage are described in Table 4.2-1.

CONCLUSIONS

TMSNP is a regularly updated web server that presents two main functionalities: on the one hand, it brings a search able database of reported pathogenic and non-pathogenic mutations in TM segments of membrane proteins; on the other hand, it provides a mutation prediction tool able to predict pathogenicity for previously non-reported TM missense mutations. The predictive model developed specifically for membrane proteins allows to improve the prediction power compared to unspecific mutation predictor servers.

DATA AVAILABILITY

TMSNP is available at http://lmc.uab.es/tmsnp/.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

REFERENCES

1. Chong,J.X., Buckingham,K.J., Jhangiani,S.N., Boehm,C., Sobreira,N., Smith,J.D., Harrell,T.M., McMillin,M.J., Wiszniewski,W., Gambin,T. *et al.* (2015) The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am. J. Hum. Genet.*, **97**, 199–215.

2. Sim,N.-L., Kumar,P., Hu,J., Henikoff,S., Schneider,G. and Ng,P.C. (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.*, **40**, W452–W457.
3. Choi,Y. and Chan,A.P. (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, **31**, 2745–2747.
4. Schwarz,J.M., Cooper,D.N., Schuelke,M. and Seelow,D. (2014) MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods*, **11**, 361–362.
5. 5. Adzhubei,I., Jordan,D.M. and Sunyaev,S.R. (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.*, **7**, Chapter 7:Unit 7.20.
6. Niroula,A. and Vihinen,M. (2016) Variation interpretation predictors: principles, types, performance, and choice. *Hum. Mutat.*, **37**, 579–597.
7. Dobson,L., Lango,T., Reményi,I. and Tusnády,G.E. (2015)´ Expediting topology data gathering for the TOPDB database. *Nucleic Acids Res.*, **43**, D283–D289.
8. Gromiha,M.M. and Ou,Y.-Y. (2014) Bioinformatics approaches for functional annotation of membrane proteins. *Brief. Bioinform.*, **15**, 155–168.
9. Overington,J.P., Al-Lazikani,B. and Hopkins,A.L. (2006) How many drug targets are there? *Nat. Rev. Drug Discov.*, **5**, 993–996.
10. Kulandaisamy,A., Priya,S.B., Sakthivel,R., Frishman,D. and Gromiha,M.M. (2019) Statistical analysis of disease-causing and neutral mutations in human membrane proteins. *Proteins*, **87**, 452–466.
11. Kulandaisamy,A., Zaucha,J., Sakthivel,R., Frishman,D. and Michael Gromiha,M. (2020) Pred-MutHTP: prediction of disease-causing and neutral mutations in human transmembrane proteins. *Hum. Mutat.*, **41**, 581–590.
12. Hauser,A.S., Chavali,S., Masuho,I., Jahn,L.J., Martemyanov,K.A., Gloriam,D.E. and Babu,M.M. (2018) Pharmacogenomics of GPCR drug targets. *Cell*, **172**, 41–54.
13. Zaucha,J., Heinzinger,M., Kulandaisamy,A., Kataka,E., Salvador, ´ O.L., Popov,P., Rost,B., Gromiha,M.M., Zhorov,B.S. and ´Frishman,D. (2020)

Mutations in transmembrane proteins: diseases, evolutionary insights, prediction and comparison with globular proteins. Brief. Bioinform., bbaa132.

14. Olivella,M., Gonzalez,A., Pardo,L. and Deupi,X. (2013) Relation between sequence and structure in membrane proteins. *Bioinformatics*, **29**, 1589–1592.
15. Mayol,E., Campillo,M., Cordom´ı,A. and Olivella,M. (2019) Inter-residue interactions in alpha-helical transmembrane proteins. *Bioinformatics*, **35**, 2578–2584.
16. Almeida,J.G., Preto,A.J., Koukos,P.I., Bonvin,A.M.J.J. and Moreira,I.S. (2017) Membrane proteins structures: a review on computational modeling tools. *Biochim. Biophys. Acta Biomembr.*, **1859**, 2021–2039. ***6*** *NAR Genomics and Bioinformatics, 2021, Vol. 3, No. 1*
17. Burley,S.K., Berman,H.M., Christie,C., Duarte,J.M., Feng,Z., Westbrook,J., Young,J. and Zardecki,C. (2018) RCSB Protein Data Bank: sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci.*, **27**, 316–330.
18. Kulandaisamy,A., Binny Priya,S., Sakthivel,R., Tarnovskaya,S., Bizin,I., Honigschmid,P., Frishman,D. and Gromiha,M.M. (2018) ¨ MutHTP: mutations in human transmembrane proteins. *Bioinformatics*, **34**, 2325–2326.
19. Popov,P., Bizin,I., Gromiha,M., A,K. and Frishman,D. (2019) Prediction of disease-associated mutations in the transmembrane regions of proteins with known 3D structure. *PLoS One*, **14**, e0219452.
20. McGarvey,P.B., Nightingale,A., Luo,J., Huang,H., Martin,M.J., Wu,C. and UniProt Consortium (2019) UniProt genomic mapping for deciphering functional effects of missense variants. *Hum. Mutat.*, **40**, 694–705.
21. Consortium, T.U.The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
22. Landrum,M.J., Lee,J.M., Riley,G.R., Jang,W., Rubinstein,W.S., Church,D.M. and Maglott,D.R. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.*, **42**, D980–D985.
23. Mottaz,A., David,F.P.A., Veuthey,A.-L. and Yip,Y.L. (2010) Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar.

Bioinformatics, **26**, 851–852.

24. Karczewski,K.J., Francioli,L.C., Tiao,G., Cummings,B.B., Alfoldi,J., ¨ Wang,Q., Collins,R.L., Laricchia,K.M., Ganna,A., Birnbaum,D.P. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434–443.
25. Eilbeck,K., Quinlan,A. and Yandell,M. (2017) Settling the score: variant prioritization and Mendelian disease. *Nat. Rev. Genet.*, **18**, 599–612.
26. El-Gebali,S., Mistry,J., Bateman,A., Eddy,S.R., Luciani,A., Potter,S.C., Qureshi,M., Richardson,L.J., Salazar,G.A., Smart,A. *et al.* (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.*, **47**, D427–D432.
27. Pei,J. and Grishin,N.V. (2001) AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, **17**, 700–712.
28. Ng,P.C., Henikoff,J.G. and Henikoff,S. (2000) PHAT: a transmembrane-specific substitution matrix. *Bioinformatics*, **16**, 760–766.
29. Norinder,U., Carlsson,L., Boyer,S. and Eklund,M. (2014) Introducing conformal prediction in predictive modeling. A transparent and flexible alternative to applicability domain determination. *J. Chem. Inf. Model.*, **54**, 1596–160

Article II. GRIN database: A unified and manually curated repertoire of GRIN variants

ABSTRACT

Glutamatergic neurotransmission is crucial for brain development, wiring neuronal function, and synaptic plasticity mechanisms. Recent genetic studies showed the existence of autosomal dominant de novo GRIN gene variants associated with GRIN-related disorders (GRDs), a rare pediatric neurological disorder caused by N-methyl-D-aspartate receptor (NMDAR) dysfunction. Notwithstanding, GRIN variants identification is exponentially growing, and their clinical, genetic, and functional annotations remain highly fragmented, representing a bottleneck in GRD patient's stratification. To shorten the gap between GRIN variant identification and patient stratification, we present the GRIN database (GRINdb), a publicly available, non-redundant, updated, and curated database gathering all available genetic, functional, and clinical data from more than 4000 GRIN variants. The manually curated GRINdb outputs on a web server, allowing query and retrieval of reported GRIN variants, and thus representing a fast and reliable bioinformatics resource for molecular clinical advice. Furthermore, the comprehensive mapping of GRIN variants' genetic and clinical information along NMDAR structure revealed important differences in GRIN variants' pathogenicity and clinical phenotypes, shedding light on GRIN-specific fingerprints. Overall, the GRINdb and web server is a resource for molecular stratification of GRIN variants, delivering clinical and investigational insights into GRDs. GRINdb is accessible at http://lmc.uab.es/grindb.

INTRODUCTION

N-methyl-D-aspartate receptors (NMDARs) are ligand-gated ion channels mediating excitatory neurotransmission (Paoletti et al., 2013). Neuronal signaling via the NMDAR plays important roles in brain development, wiring, and high cognitive and motor functions, among others (Traynelis et al., 2010). NMDARs are heterotetramers mostly composed of the assembly of two glycine/D-serine-binding GluN1 subunits (encoded by GRIN1 gene) and two glutamate-binding subunits GluN2A/GluN2B (encoded by

GRIN2A and GRIN2B genes, respectively; Flores-Soto & Chaparro-Huerta, 2012). In addition to the predominantly expressed GluN1, GluN2A, and GluN2B subunits, NMDARs can also have the presence of GluN2C, GluN2D, GluN3A, and/or GluN3B subunits at particular developmental stages and in precise neuronal subtypes. For proper neuronal function and brain connectivity, NMDAR heteromeric channels require a precise spatio–temporal expression pattern, in an NMDAR heteromer-dependent manner. Functionally, the NMDAR acts as a coincident detector responding to sustained glutamate release. Following NMDAR activation, NMDAR-mediated calcium influx activates signaling cascades, leading to neuronal responses. Besides neurotransmitter-mediated regulation, the molecular composition of the NMDAR is critical for NMDAR- mediated neurotransmission.

The structure of NMDAR has been recently determined (Karakas & Furukawa, 2014; Lee et al., 2014; Tajima et al., 2016) and showed a topological organization of this receptor into four functionally related domains. The extracellular amino-terminal and ligand-binding domains (ATD and LBD, respectively) are mechanically linked to the transmembrane domain (TMD) that constitutes the pore of the channel, and a cytoplasmic carboxy-terminal domain (CTD) with no determined structure. Functionally, the NMDAR acts as a coincident detector of postsynaptic membrane depolarization (removing Mg^{2+} blockade) and glutamate release. In this context, the simultaneous binding of NMDAR coagonists (glutamate and glycine/ D-serine) is mechanically transduced, allowing the channel pore opening and Na^{+} and Ca^{2+} influx (Mayer et al., 1984; Nowak et al., 1984). In addition to the subunit composition, which determines NMDAR intrinsic biophysical properties, the receptor activity is regulated by small ions and molecules, such as protons, zinc, spermine (Paoletti et al., 2013), membrane lipids (Korinek et al., 2015), and proteins, involved in posttranslational modifications (Lussier et al., 2015) and protein–protein interactions (Frank & Grant, 2017).

As aforementioned, NMDAR-mediated signaling is crucial for fundamental neuronal processes. Concomitantly, the dysregulation of the NMDAR function can cause synaptic dysfunction and disturb neuronal activity, which ultimately can lead to

neurological conditions. Less than 10 years ago, the initial genetic reports describing the association between de novo mutations of GRIN genes associated, and neurological disorders were released (de Ligt et al., 2012; O'Roak et al., 2011; Tarabeux et al., 2011). Since then, third- generation deep sequencing has accelerated the identification of novel GRIN variants in individuals affected by neurodevelopmental and psychiatric disorders and allowed to initiate pioneer studies

toward the delineation of a clinical spectrum genetically defined as GRIN-related disorders (GRDs; Burnashev & Szepetowski, 2015; Lemke et al., 2014, 2016; Lesca et al., 2013; Platzer et al., 2017). GRDs are a group of rare neurological disorders (Lemke, 2020) with a genetic etiology, namely de novo GRIN mutation (with almost exclusively autosomal dominant inheritance pattern), a primary cause of NMDAR-mediated dysfunction of excitatory neurotransmission. Clinically, GRDs' clinical spectrum includes intellectual disability, epilepsy, movement disorders, development delay, autism spectrum disorder, and schizophrenia, among others.

In addition to genetic data, several functional studies have been conducted to characterize the direct impact of de novo GRIN variants gene products on NMDAR biogenesis and activity. Altogether, genetic and functional studies showed the need to stratify disease-associated GRIN variants and dichotomically dissect those functionally, causing either a gain-of-function (GoF) or a loss-of-function (LoF). This stratification, as well as gene-dependent functional and clinical outcomes definition, is critical to define precise NMDARs functional rescue, ultimately leading to GRDs personalized medicine, a clinical and in vestigational priority. Therapeutic approaches for GRDs have focused on the potential use of direct NMDAR modulators, potentially counteracting either GoF or LoF GRIN variants effects on NMDARs, for example, reducing or potentiating NMDARs activity, respectively. Regarding GRIN-GoF variants, the FDA-approved drug memantine (low-affinity NMDAR antagonist) has been evaluated for GRD associated with GRIN-GoF variants (Ogden et al., 2017; Pierson et al., 2014; Platzer et al., 2017). Similarly, radiprodil (GluN2B selective negative allosteric modulator) has been recently evaluated and constitutes a promising drug for the treatment of NMDARs containing GluN2B GoF mutations (Auvin et al.,

2020; Mullier et al., 2017). The potential rescue of GRIN LoF variants has also been evaluated in preclinical and clinical studies, using both NMDARs orthosteric and allosteric modulators. In this line, a pilot clinical study showed that L-serine (precursor of D- serine, endogenous NMDAR coagonist) dietary supplement results in a clinical improvement in a patient harboring a GRIN2B LoF variant (Soto et al., 2019).

As for many rare conditions with a genetic origin, GRD translational research shows a gap between growing genetic data (e.g., GRIN variants identification) and their evaluation toward functional stratification and therapeutic approaches evaluation. Further, GRIN variants genetic, clinical, and functional data are highly fragmented, redundant, and show certain inconsistencies, making both clinical and genetic advice arduous. In this scenario, we have developed the GRIN database (GRINdb), a clinical and investigational resource, with more than 4000 unique GRIN variants that constitute the largest public repository of GRIN variants and related data. More importantly, GRINdb provides a unique, nonredundant, curated, and updated GRIN variants database. Together with the genetic data, GRINdb delivers all available functional and clinical annotations of the GRIN variant of interest, which, in turn, might accelerate the patient's stratification and therapeutic decisions. Further, GRINdb allowed delineating important differences in GRIN pathogenesis and clinical phenotypes, shedding light on GRIN-specific fingerprints and GRDs' pathophysiology.

METHODS

GRINdb

For GRINdb creation, all available GRIN variants information (gene, nucleotide change, amino acid position, mutated amino acid, pathogenicity, clinical information, and functional annotations) were retrieved from the following sources: ClinVar (only pathogenic and likely pathogenic variants), LOVD (https://www.lovd.nl), Uniprot Database (UniProt Consortium, 2019), GnomAD (Karczewski & Francioli, 2017), “GRIN-Leipzig” (http://www.grin-database.de), CFERV (http:// functionalvariants.emory.edu/database/), and BCN-GRIN (http://lmc. uab.es/grindb/) and from previously published data. The genetic information was used to classify the

variants into the following categories: missense, truncating (nonsense, frameshift), and others (amino acid insertions and deletions, large DNA duplications, deletions and translocations, and chromosomal aberration). The amino acid position was used to locate the mutation into the corresponding topological domain (see Table S1 for domain positions), with the TMD positions assigned according to OPM Database (Lomize et al., 2012). Finally, redundant entries (identical GRIN variant) were unified under the same entry, while keeping clinical- and functional-associated information.

GRIN variants data curation and functional classification

Based on genetic, clinical, and functional data, GRIN variants pathogenicity was classified as “Disease-associated,” “Neutral,” or “Uncertain pathogenicity,” according to the following criteria. GRIN variants with reported functional alterations (LoF/GoF) were classified as disease-associated, while those not affecting NMDAR activity were classified as neutral variants. For GRIN variants devoid of functional annotations, their pathogenicity was defined according to the associated genetic and clinical data. GRIN variants associated with neurological disorders (ClinVar, LOVD, GRIN-Leipzig, BCN-GRIN, and CFERV GRIN variants with associated phenotypes) were classified as “Disease-associated,” while those variants exclusively reported in the GnomAD database were considered as neutral. Finally, GRIN variants with conflicting annotations (e.g., described as disease-associated and neutral variants in different databases) were classified as variants of uncertain pathogenicity.

In addition to pathogenic categories, the available functional information was manually curated and used for GRIN variants functional stratification into “loss-of-function” (LoF), “gain-of-function” (GoF), “Complex,” or “Not affected” categories. GRIN variants impact on receptor biogenesis (surface expression of mutant NMDARs) and bio physical changes (electrophysiological properties of mutant NMDARs) allowed to define LoF variants, based on the presence of a reduction of surface expression and/or reduced NMDAR-mediated currents resulting from diminished ion channel conductance, open probability, agonist(s) affinity(ies), and desensitization and/or deactivation rates. On the contrary, GRIN variants were annotated as GoF when the surface expression was increased and/or NMDAR-mediated currents were increased as a result of changes in

the gating/affinity properties mentioned previously. GRIN variants exhibiting both LoF- and GoF-associated changes were classified as “Complex” variants of uncertain functional classification, while those with not showing functional alterations in heterologous cell systems were defined as “Not affected.” Based on our recent findings (Santos-Gómez et al., 2020), GRIN2A and GRIN2B protein-truncating variants of the ATD, LBD, and TMD were classified as LoF variants. Thus, the classification of variants as GoF, LoF, or complex was entirely based on experimental information and not in a prediction algorithm.

Phenotypes associated with GRIN variants

The available clinical information allowed to group GRIN variants phenotypic traits into the following phenotype categories: intellectual disability, developmental delay, epilepsy, schizophrenia, autism spectrum disorders, attention-deficit and hyperactivity disorder, cortical visual impairment, hypotonia, speech disorder, movement disorders, and microcephaly. While all these phenotypes categories were automatically retrieved from available clinical information on reference databases and research articles, “epilepsy” phenotype category was grouping epilepsies and epileptic syndromes, such as epilepsy with continuous spike and wave during sleep, benign childhood epilepsy with centro-temporal spikes, Landau–Kleffner syndrome of childhood, focal epilepsy, and discrete epileptic seizures.

NMDA receptor molecular model

A molecular model for the triheteromeric NMDA receptor (GluN1)2- GluN2A-GluN2B was generated from 4PE5 X-ray crystal structure (Karakas & Furukawa, 2014). Modeller 9.20 (Webb & Sali, 2016) was used to model the lacking regions of the receptor and Scwrl4 (Krivov et al., 2009) to position the nondetermined side chains. The initial model was energy-minimized using GROMACS 5 (Hess et al., 2008). The CTDs were not modeled, as there is no available determined structure. Pathogenesis and functional annotations for all GRIN variants with available information were mapped on the NMDA receptor structural model using Pymol 2 (DeLano, 2002).

Web server

GRINdb is a freely available web application tool relying on Python 3.6.6 backend with Flask 1.0.2 framework (http://flask.pocoo.org). Python scripts monthly access data from ClinVar, LOVD, Uniprot, GnomAD, CFERV, GRIN-Leipzig, and BCN-GRINdbs. All data are stored in a MySQL database. The NMDAR molecular model containing the query GRIN variant is shown with an NGL Viewer (Rose & Hildebrand, 2015), allowing an interactive rotation in the x/y/z axes, together with zoom options (Rose & Hildebrand, 2015).

RESULTS AND DISCUSSION

GRINdb

The increasing number of identified GRIN genetic variants associated with neurological and psychiatric disorders deals with high fragmentation and redundancy of the corresponding data, which is split into several databases and research articles. We have constructed GRINdb, a unified, manually curated, nonredundant, and updated database containing all available genetic, clinical, functional, and structural information on GRIN variants. GRINdb currently contains 4415 unique GRIN variants (May 2020), classified into different genetic categories: 4008 missense variants, 295 protein-truncating variants (frameshift, nonsense, splice site), and 112 other variants (duplications, deletions, translocations; see Table 4.3-4). The database can be freely accessed at http:// lmc.uab.es/grindb/grindbdata.

	GRIN1	*GRIN2A*	*GRIN2B*	*GRIN2C*	*GRIN2D*	*GRIN3A*	*GRIN3B*	**Total**
Missense	313	803	592	602	364	596	738	4008
Truncations	22	57	54	43	17	26	74	293
Others	12	14	12	20	11	6	37	112
Total	347	874	658	665	392	628	849	4413

Table 4.3-4 Summary of variants in GRIN database. GRIN variants collected from ClinVar, SwissVar, GnomAD, GRIN-Leipzig, CFERV, and BCN-GRIN and bibliography including missense, truncating (nonsense, frameshift), and other variants (amino acid insertions and deletions, large DNA insertions, deletions and translocations, and chromosomal aberration)

GRINdb web server

To provide a user-friendly computational tool to retrieve all available information for the GRIN variant of interest, we have developed the GRINdb Web Server (http://lmc.uab.es/grindb/). In this GRIN variants explorer, the user is invited to introduce the gene name (GRIN1, GRIN2A, GRIN2B, GRIN2C, GRIN2D, GRIN3A, GRIN3B) and amino acid position as a query input. Optionally, the user can also select the mutated amino acid (in three- or one- letter code), for missense variants. Following GRIN variant query, GRIN web server explores GRINdb and retrieves—upon data availability—all associated functional and clinical data: reference database, mutation domain, and category (missense, truncation, other), as well as available functional annotations and clinical information with the corresponding bibliographic references (PubMed links). Classification into GoF/LoF/Complex, phenotype categories, and pathogenicity (disease-associated, neutral, and uncertain association to disease) is also shown (see Methods). This descriptive information is completed with a molecular model of the NMDAR showing the topological location of the mutated amino acid, displayed with an interactive NGL Viewer (Rose & Hildebrand, 2015; see Figure 4.3-1).

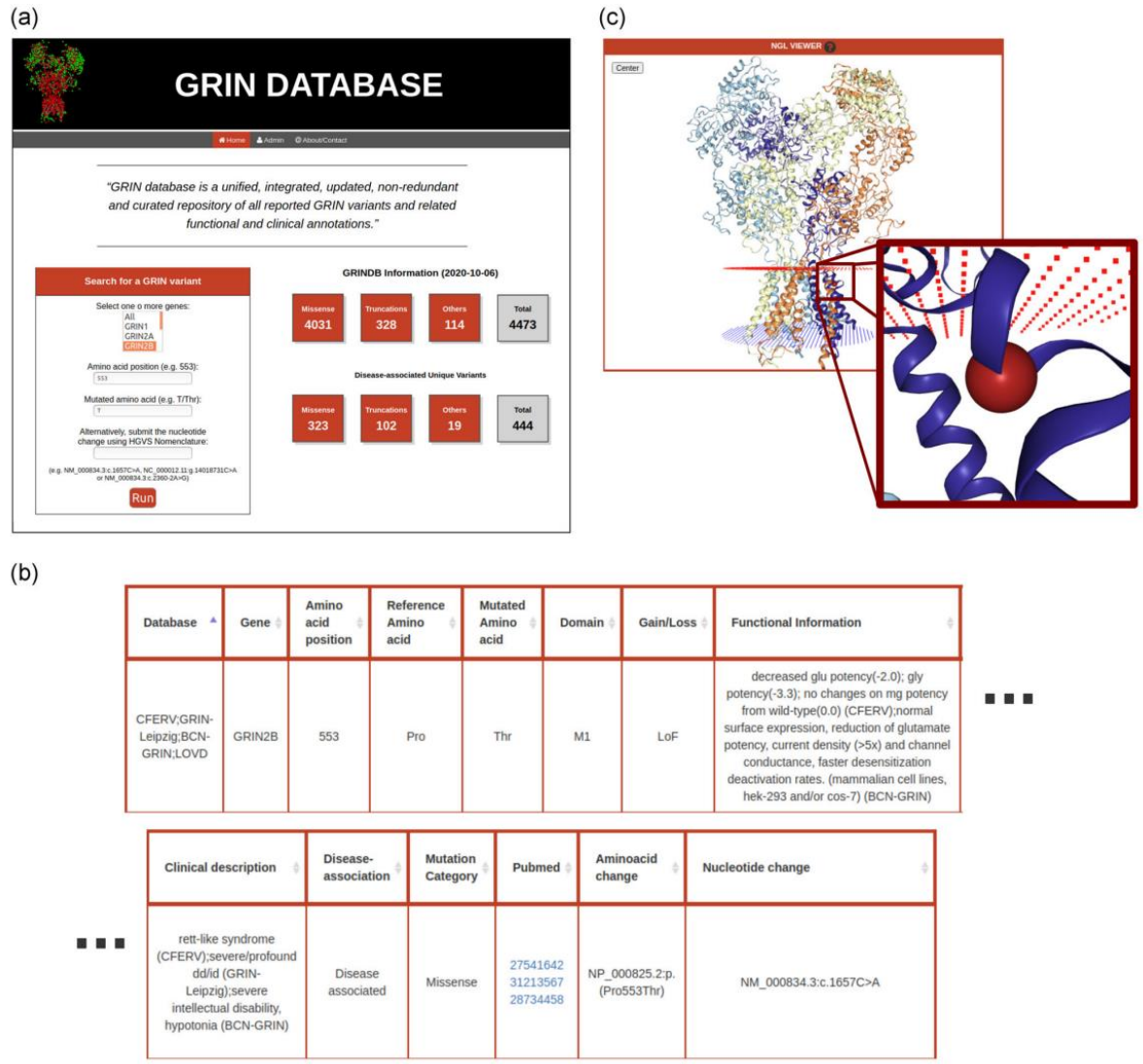


Figure 4.3-1 GRIN database web server example of use. (**a**) Input: the user can search for a specific GRIN variant (gene, amino acid position, and mutated amino acid); **(b**) output: the corresponding genetic, clinical, and functional information is displayed. (**c**) Additionally, the mutated amino acid is mapped on the N-methyl-D-aspartate receptor structural model with an NGL Viewer 2015. Variant GRIN2B Pro553Thr (NP_000825.2:p.Pro553Thr) has been used as an example. By integrating the fragmented information collected from databases and research articles, the variant can finally be annotated as loss-of-function (LoF) and the corresponding clinical phenotype is displayed

GRIN variants and disease association

The integration of functional and clinical information (disease- associated or neutral GRIN variants) into GRINdb was used to provide the first comprehensive analysis of GRIN variant-related pathogenesis. To this end, we analyzed the association of GRIN

variants with neurological conditions and classified them in three different categories. Disease-associated GRIN variants are GRIN variants unequivocally associated with GRD, with an autosomal dominant pattern and GRD clinical symptoms described. On the contrary, those GRIN variants described as genetic polymorphisms (neutral variants present in gnomAD and/or no functional changes associated) were considered as neutral. The third category of GRIN variants is constituted by those genetic variants with contradictory annotations facing the autosomal dominant inheritance pattern described in GRD and was thus classified as GRIN variants of uncertain pathogenicity (Table 4.3-5).

	GRIN1	*GRIN2A*	*GRIN2B*	*GRIN2C*	*GRIN2D*	*GRIN3A*	*GRIN3B*	Total
Associated with disease	74 (23.6%)	114 (14.2%)	111 (18.8%)	4 (0.7%)	9 (2.5%)	1 (0.2%)	3 (0.4%)	316
Neutral	222 (71.0%)	647 (80.5%)	445 (75.2%)	575 (95.5%)	347 (95.3%)	573 (96.1%)	704 (95.4%)	3513
Uncertain	17 (5.4%)	42 (5.2%)	36 (6.0%)	23 (3.8%)	8 (2.2%)	22 (3.7%)	31 (4.2%)	179
Total missense	313	803	592	602	364	596	738	4008

Table 4.3-5. Summary of GRIN missense variants classification into disease-association categories. Functional and clinical data of GRIN variants allowed their classification into variants with disease association, neutral, or with uncertain pathogenesis (see Methods for additional information on the criteria used for variants pathogenesis classification).

Further, functional data and clinical phenotype analysis of all collected GRIN missense variants (4008, scattered along the seven members of the GRIN gene family) revealed striking gene-dependent pathogenicity association. Interestingly, GRIN1, GRIN2A, and GRIN2B genes (encoding for GluN subunits, most widely expressed in excitatory neurons (Paoletti et al., 2013) present a significant percentage of variants associated with disease (ranging from 14.2% to 23.6%). Contrastingly, GRIN2C, GRIN2D, GRIN3A, and GRIN3B (encoding for GluN subunits expressed in specific excitatory neuron subpopulations; Paoletti et al., 2013) association with pathogenicity is scarce (from 0.2% to 2.5%), with 17 missense disease- associated variants that are randomly located across all domains. Interestingly, following database integration and curation, the originally reported 118 initial total disease-associated variants were reduced to 26, suggesting that genetic variation of GRIN2C, GRIN2D, GRIN3A, and GRIN3B genes could be more tolerated and less prone to cause neurological phenotypes. GRIN1,

GRIN2A, and GRIN2B genes pathogenicity were further analyzed toward the potential identification of GluN1, GluN2A, and GluN2B domains differential sensitivity to amino acid variation (Figure 4.3-2 and Table S2).

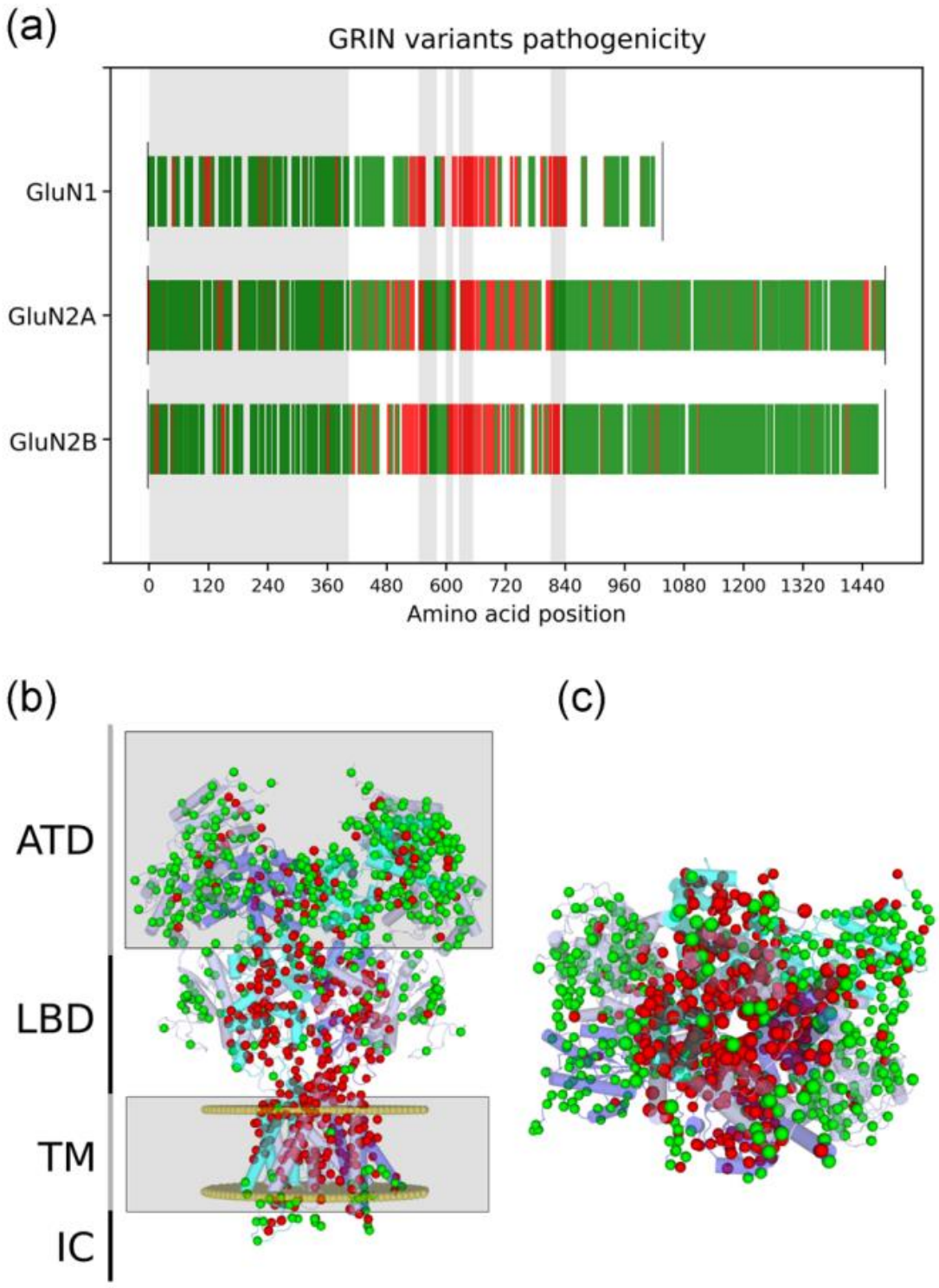


Figure 4.3-2 GRIN variants pathogenicity. (**a**) Scatter plot representing GRIN1, GRIN2A, and GRIN2B variants' disease association and distribution along the GluN1, GluN2A, and GluN2B protein sequence.

Spikes undefined color code: disease-associated GRIN variants in red; not associated with disease in green; no reported variant in white. Juxtaposed topological domains are represented in alternate colors (amino-terminal domain [ATD], M1, M2, M3, and M4 highlighted in gray) and the protein termination with a black spike. (**b**) GRIN1, GRIN2A, and GRIN2B variants pathogenicity mapped on the structural model of the trihetromeric (GluN1)2-GluN2A/GluN2B NMDA receptor. Lateral view of the structural model of the NMDAR, composed of two GluN1 (pale blue), one GluN2A (cyan), and one GluN2B (purple) subunits. Note the high density of disease-associated GRIN variants in the ligand-binding domain (LBD) and transmembrane domain (TMD). (**c**) Molecular model of the N-methyl-D-aspartate receptor (cytosolic side view). The pore channel and the GluN subunit interfaces are highly pathogenic. A position is considered as associated with disease (in red) if at least one disease-associated mutation is found. Neutral variants (in green) are those not reported as disease-associated, TM, transmembrane.

Among GRIN genes most prevalently associated with the disease, GRIN1 presents the lowest number of reported GRIN variants (313 GRIN1 vs. 803 and 592 variants, for GRIN1, GRIN2A, and GRIN2B, respectively) but the highest percentage of disease-associated variants (23.6%). On the contrary, GRIN2A presents the highest number of variants, but the lowest percentage of disease-associated variants (14.2%), suggesting a differential tolerance of GRIN genes in front of genetic variation, with a GRIN1 > GRIN2B > GRIN2A pathogenicity scaling.

Topologically, the majority (about 80%) of GRIN1, GRIN2A, and GRIN2B disease-associated variants are located within TMD and LBD (see Figure 4.3-2 and Table S2). This result is in accordance with previous reports compiling the data from around 100 GRIN variants (Amin et al., 2020; XiangWei et al., 2018). This differential domain sensitivity is exhibited by all GluN subunits, in line with the high phylogenetic conservation degree and structural constraints of the NMDAR channel pore. A more detailed analysis of GRIN disease- associated variants distribution along NMDAR subdomains revealed several hotspots. These subdomains highly vulnerable to genetic variance correspond to functionally crucial elements of the NMDARs, namely the agonist/coagonist binding pockets (LBD), the inner side of NMDAR channel pore, and the interface between GluN subunits in accordance with the rolling motion between subunits that is necessary to activate the receptor (Esmenjaud et al., 2019). Conversely, the ATD presents a low association with neurological conditions, with

occasional disease-associated variants located close to positive and negative allosteric binding sites. These sites are involved in NMDAR channel activity fine-tuning (Zhu et al., 2016), controlling NMDAR open probability and deactivation rate (Gielen et al., 2009; Yuan et al., 2009). Similarly, the CTD domains of GluN1, GluN2A, and GluN2B present a reduced number of disease-associated variants, suggesting that these intracellular domains are mainly neutral. Noteworthy, the presence of inherited GRIN variants within the CTD, both in asymptomatic carriers and in GRD individuals, was suggestive of a recessive inheritance pattern, potentially acting as a risk factor of complex disorders.

GRIN variants stratification and NMDAR dysfunction

Recently, several efforts have been conducted toward the evaluation of potential GRD therapies, aiming to rescue NMDAR dysfunctions in a personalized manner. While memantine (Ogden et al., 2017; Pierson et al., 2014; Platzer et al., 2017) and radiprodil (Auvin et al., 2020; Mullier et al., 2017) have been prescribed for patients with GRIN1 and GRIN2B GoF variants, respectively, the L-serine dietary supplement has been investigated in a single individual harboring a GRIN2B LoF variant (Soto et al., 2019). Nevertheless, before the therapeutic intervention, the functional annotation of the GRIN variant of interest is compulsory and constitutes the bottleneck in patient stratification and personalized therapies. Indeed, the molecular phenotyping of GRIN variants is a time-consuming process that can be shortened by the use of already existing data. Unfortunately, these data are spread among databases and the literature, together with data duplication and/or conflicting annotations, making their retrieval and interpretation difficult. To accelerate GRIN variants annotation and patient stratification, GRINdb offers a curated repertoire of available GRIN variants functional annotations.

	GRIN1	*GRIN2A*	*GRIN2B*
GoF variants	13 (39.4%)	23 (42.5%)	16 (25.4%)
LoF variants	20 (60.6%)	31 (57.4%)	47 (74.6%)
Variants with functional annotation	33 (44.6%)	54 (47.3%)	63 (56.8%)
Variants with no functional annotation	41 (55.4%)	60 (52.6%)	48 (43.2%)
Total	74	114	111

Table 4.3-6. Functional annotation of GRIN1, GRIN2A, and GRIN2B variants. Percentages are shown over reported annotations. Half of the GRIN variants contain a functional annotation, and LoF variants are more frequent than GoF variants. Abbreviations: GoF, gain-of-function; LoF, loss-of-function.

GRINdb analysis showed that disease-associated GRIN1, GRIN2A, and GRIN2B variants are similarly functionally annotated (44.6%, 47.3%, and 56.8%, respectively), with a higher number of LoF variants compared with GoF variants (Table 4.3-6). Importantly, GRIN variants mapping revealed no specific association between functional alterations (GoF or LoF) and topological domains, with the exception of a certain enrichment of GoF variants location at the Mg^{2+}-binding site at the transmembrane pore channel (see Figure 4.3-3 and Table S3). The analysis of GRIN variants' functional annotations revealed striking fingerprints of both LoF and GoF disease-associated variants. LoF variants mostly result from a decrease of NMDAR surface expression and/or a reduction of charge transfer across the mutant NMDAR. In contrast, GRIN-GoF variants are frequently associated with the presence of the following alterations, either individually or combined: reduction or abolishment of magnesium blockade, increase of agonist(s) potency(ies), and decrease of deactivation rate.

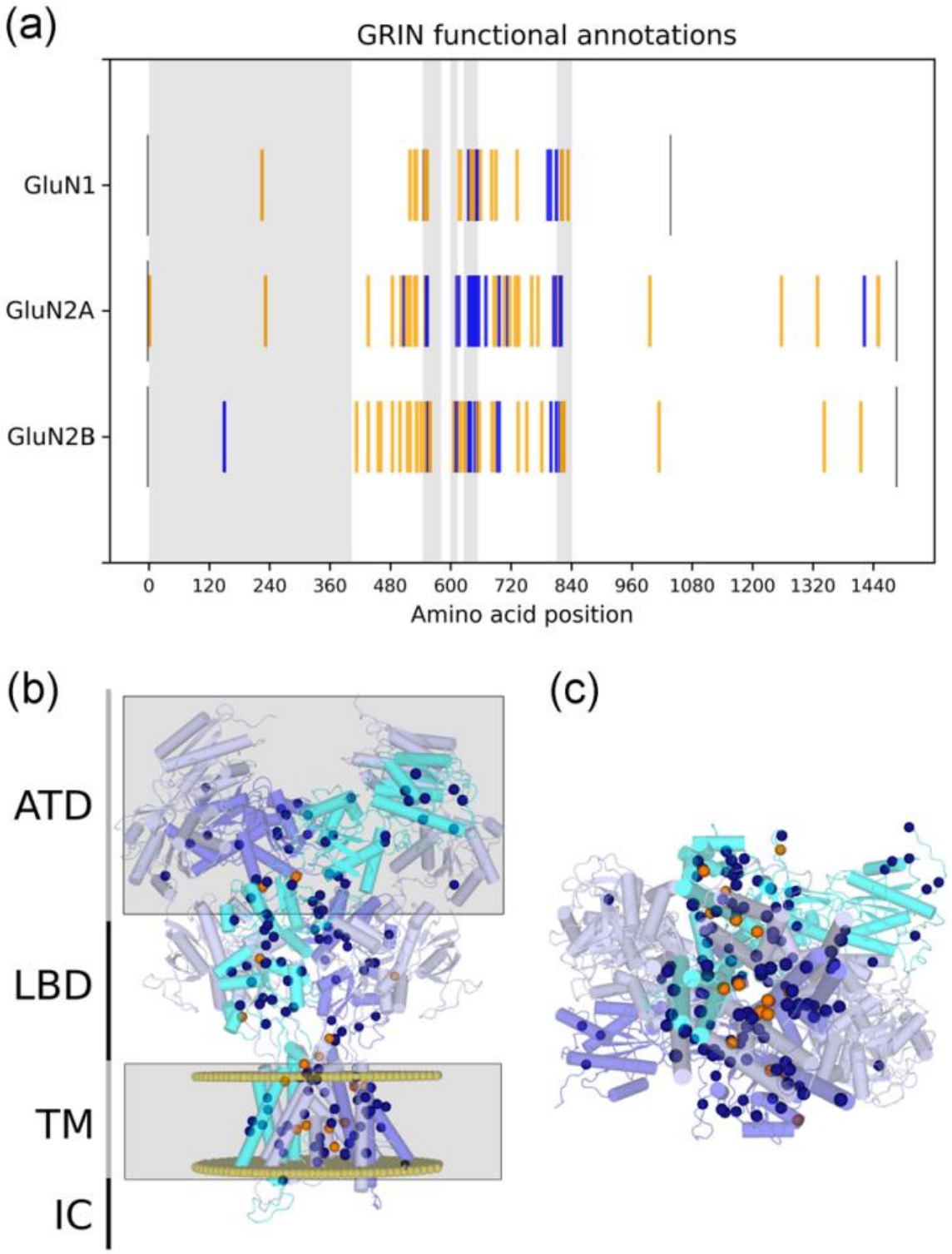


Figure 4.3-3 GRIN variants functional annotations. (**a**) GRIN1, GRIN2A, and GRIN2B variants functional annotations along the GluN1, GluN2A, and GluN2B sequence (proteins are limited by gray lines). Disease-associated GRIN missense variants causing a loss-of-function are represented by

orange lines in the corresponding amino acid position, while GoF variants are represented in blue lines. The domains ATD, M1, M2, M3, and M4 are highlighted in gray. (**b**) GRIN1, GRIN2A, and GRIN2B variants have functional annotations mapped on the structural model of the N-methyl-D-aspartate (NMDA) receptor. The structural molecular model is composed of two GluN1 (pale blue), one GluN2A (cyan), and one GluN2B (purple) subunits. LBD and TM domains concentrate most of the disease-associated variants. (**c**) NMDA receptor molecular model view from the cytosolic side. GoF mutations are mainly around the magnesium-binding site at the pore channel. ATD, amino-terminal domain; GoF, gain-of-function; IC, intracellular; LBD, ligand-binding domain.

GRIN variants clinical phenotypes analysis

Following GRINdb construction, GRIN variants-associated clinical phenotypes were compiled toward the delineation of the GRDs' clinical spectrum and the exploration of putative genotype- and domain-specific clinical symptoms. Further analysis of the most prevalent GRD-associated GRIN genes (GRIN1, GRIN2A, and GRIN2B) showed that disease-associated GRIN1 variants are generally clinically manifested by developmental delay (71%), intellectual disability (82%), and epilepsy (57%), although the scarce clinical data hampered the determination of putative clinical phenotypes correlation with GRIN variants GoF/LoF functional categories (Table 4.3-7 and Figure 4.3-4).

	GRIN1			*GRIN2A*			*GRIN2B*		
	GoF	LoF	Total	GoF	LoF	Total	GoF	LoF	Total
Reported phenotypes	7	16	56	18	26	99	12	40	91
Autism spectrum disorder	0	6% (1)	18% (10)	0	4% (1)	6% (6)	0	40% (16)	29% (26)
Cortical visual impairment	57% (4)	19% (3)	25% (14)	0	0	0	42% (5)	3% (1)	7% (6)
Developmental delay	86% (6)	81% (13)	71% (40)	83% (15)	46% (12)	38% (38)	100% (12)	75% (30)	68% (62)
Epilepsy (ECSWS, FE, BECTS, LKS)	57% (4)	56% (9)	57% (32)	83% (15)	100% (26)	89% (88)	83% (10)	40% (16)	47% (43)
Hypotonia	43% (3)	38% (6)	30% (17)	11% (2)	4% (1)	3% (3)	8% (1)	38% (15)	21% (19)
Intellectual disability	71% (5)	100% (16)	82% (46)	89% (16)	46% (12)	49% (49)	100% (12)	100% (40)	93% (85)
Microcephaly	0	6% (1)	4% (2)	0	0	1% (1)	50% (6)	8% (3)	14% (13)
Movement disorder	0	31% (5)	13% (7)	6% (1)	0	1% (1)	0	13% (5)	9% (8)
Schizophrenia	0	0	5% (3)	0	0	4% (4)	0	5% (2)	5% (5)
Speech disorder	0	19% (3)	7% (4)	50% (9)	73% (19)	57% (56)	0	0	0

Table 4.3-7. Clinical phenotypes of disease-associated GRIN1, GRIN2A, and GRIN2B variants. Percentages are shown over reported clinical phenotypes. Also, note that the clinical phenotypes are subunit- and GoF/LoF-dependent. Abbreviations: BECTS, benign childhood epilepsy with centro-temporal spikes; ECSWS, epilepsy with continuous spike and wave during sleep; FE, focal epilepsy;

GoF, gain-of-function; LKS, Landau–Kleffner syndrome of childhood; LoF, loss-of-function.

GRINdb analysis showed that GRIN2A variants are almost invariably associated with epilepsy (89%), with frequent speech disturbance (57%) and, to a lesser extent, intellectual disability (49%) and developmental delay (38%). Importantly, intellectual disability and development delay are more frequently associated with GRIN2A GoF variants (89% and 83%, respectively) than with GRIN2A LoF variants (46% in both cases), while speech disorder is more frequently associated with GRIN2A LoF variants, compared with GRIN2A GoF variants (73% vs. 50%). Interestingly, the domain-based analysis showed a strong presence of speech disorders in individuals harboring GRIN2A variants affecting the ATD. Genotype–phenotype analysis showed that ATD is the less vulnerable domain of GluN subunits, suggesting that speech disorder might not be recorded in databases for those individuals with severe developmental delay and/or intellectual disability. Regarding GRIN2B, disease-associated missense mutations are almost invariably associated with intellectual disability (93%) and development delay (68%). The occurrence of epilepsy is variable (47%) and more frequently associated with GoF than with LoF variants (83% vs. 40%). GRINdb analysis also showed that GRIN2B variants are frequently associated with the presence of autism spectrum disorders phenotypic traits in GRIN2B LoF but not GRIN2B GoF variants (40% vs. 0%). GRIN2B variants functional alterations were also associated with additional clinical traits, such as hypotonia (38% in LoF vs. 8% in GoF GRIN2B variants) and microcephaly (0% in LoF vs. 50% in GoF GRIN2B variants).

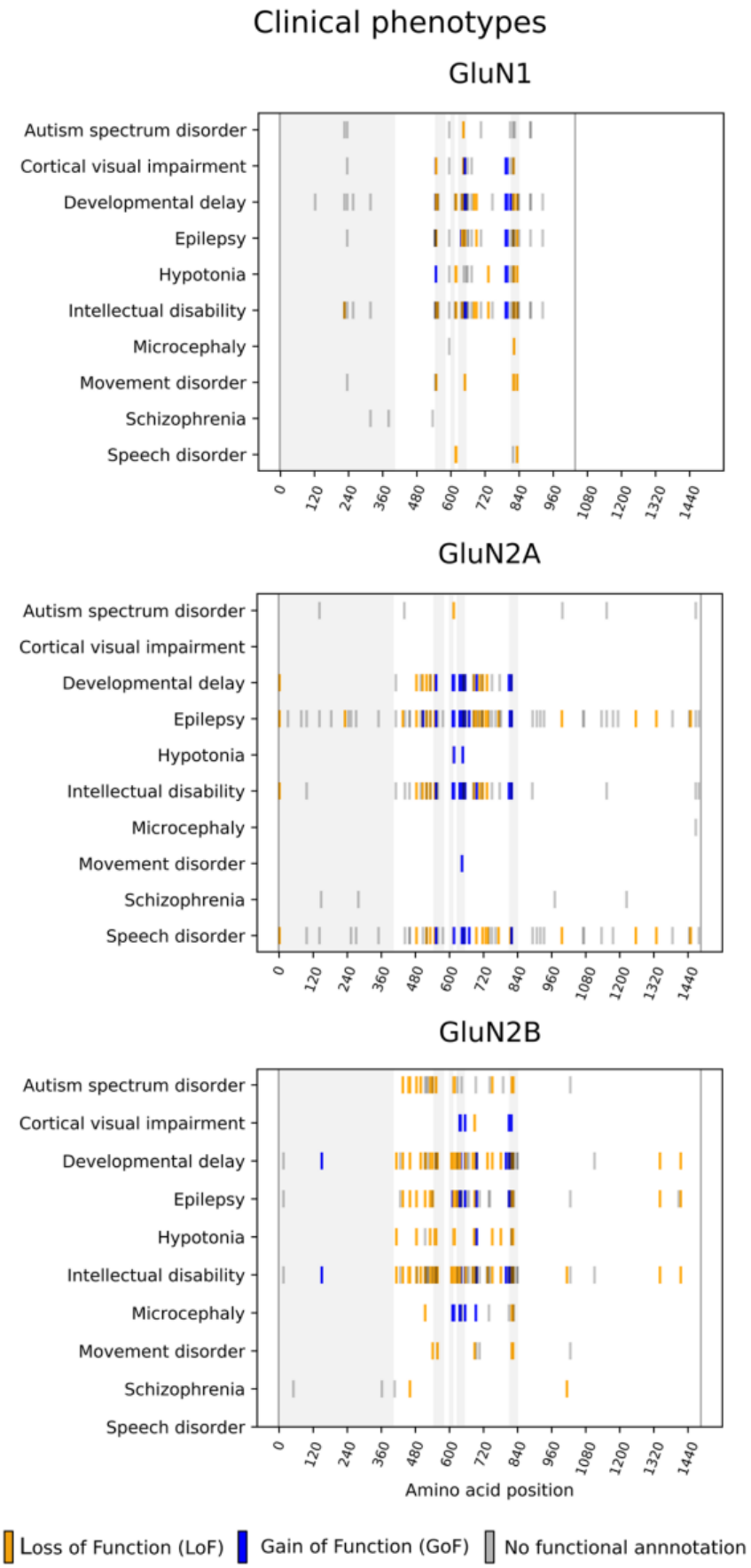

Clinical phenotypes
GluN1
Autism spectrum disorder
Cortical visual impairment
Developmental delay
Epilepsy
Hypotonia
Intellectual disability
Microcephaly
Movement disorder
Schizophrenia
Speech disorder
0
120
240
360
480
600
720
840
960
1080
1200
1320
1440
GluN2A
GluN2B
Amino acid position
Loss of Function (LoF)
Gain of Function (GoF)
No functional annnotation

Figure 4.3-4. GRIN variants clinical phenotypes. Scatter plot ofGRIN1, GRIN2A, and GRIN2B loss-of-function (LoF) and gain-of-function (GoF) genetic variants associated with clinical phenotypes. The spikes represent disease-associated GRIN variants (LoF, orange spike; GoF, blue spike) distribution along GluN1, GluN2A, and GluN2B proteins (GluN protein terminus, black spike). The domains ATD, M1, M2, M3, and M4 are highlighted in gray. Note that the clinical phenotypes are subunit- and GoF/LoF-dependent.

Noteworthy, along with this analysis, we noticed a nonsystematic analysis of NMDAR surface expression. As this aspect is crucial to determine the overall functional outcome of GRIN GoF, these in complete (pending) variant annotations were indicated as GoF(P) in GRINdb.

CONCLUSIONS

The GRINdb research resource constitutes an updated, manually curated, and nonredundant database comprising genetic, clinical, functional, and structural information of the largest repertoire of GRIN variants, with more than 4400 unique GRIN variants entries. From a clinical angle, the comprehensive data compiled in the GRINdb represents a valuable bioinformatic resource shortening the gap between the growing number of GRIN variants identification (resulting from the implementation of deep genomic sequencing data in the clinical practice) and patient stratification. Noteworthy, the query of individual GRIN variants throughout the GRIN web server provides important molecular diagnosis insights of GRIN variants' functional outcomes (disease-association and functional annotation, e.g., GoF/LoF), which is compulsory for GRD personalized medicine.

The analysis of more than 4000 missense GRIN variants sheds light on the pathogenesis map almost completely covering GluN subunits aminoacid positions, revealing their weighted association with neurological conditions in a gene-, subunit- and region- dependent manner. For GRIN1, GRIN2A, and GRIN2B genes, around 18% of its variants are disease-associated and mainly concentrated in the LBDs and TMDs. Contrastingly, reported GRIN2C, GRIN2D, GRIN3A, and GRIN3B variants have a very limited association with neurological conditions. Indeed, about 95% of the

reported missense variants affecting these genes are neutral. Further, the variants of these genes are homogeneously distributed across all domains, in contrast to GRIN1, GRIN2A and GRIN2B mutation hotspots present in the agonist-binding domain and the channel pore. In summary, these association studies suggest that although the variants affecting these genes could be involved in complex neurological diseases (schizophrenia, attention deficit hyperactivity disorder), the majority of the variants are neutral and not disease-causing. This reduced vulnerability is indicative of a weaker pathophysiological impact on glutamatergic neurotransmission-related processes, which might result from their intrinsic spatio–temporal expression patterns (restrained to a small set of glutamatergic synapses), a putative functional redundancy with other gene products and/or asymptomatic conditions. Despite the pathogenic map of GRIN variants being extensive, the functional annotations are still fragmented. Indeed, about half of disease-associated GRIN variants lack functional annotation and concomitantly their classification into LoF or GoF remains elusive. In the context of GRD, this annotation is essential toward patient stratification and personalized medicine (e.g., positive or negative allosteric modulators and subunit specificity) and thus further efforts are required to achieve this critical task.

The number of reported phenotypes brings the opportunity to characterize the clinical phenotype and suggests that clinical phenotype is subunit-, domain- and LoF/GoF-dependent. In addition, the clinical phenotypes are in accordance with GRIN protein-truncating variants-associated clinical phenotypes, which represent a paradigmatic genetic condition of NMDAR LoF (Santos-Gómez et al., 2020). However, GRINdb data analysis presents some intrinsic limitations resulting from the partial functional annotations of GRIN variants and the heterogeneous and not comprehensive clinical description of GRD individuals. Therefore, there is an urgent need to fully report the clinical phenotypes of each domain in each subunit and according to LoF and GoF and to homogenize and extensively assess the clinical phenotypes of GRD individuals. This would lead to delineate GRD genetic and functional determinants, as well as defining GRD patient stratification and prognosis.

The enrichment of the database with new GRIN variants, which is daily growing, will

allow to also improve the predictive model of GRIN pathogenesis and clinical phenotype and the effect of disease- associated variants on receptor structure and function, thus shortening even more the gap between variant identification and personalized medicine.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at http://lmc.uab.es/grindb/grindbdata and the database is accessible at http://lmc.uab.es/grindb.

WEB RESOURCES

The following web resources were used: ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), LOVD (https://www.lovd.nl), Uniprot Database (https://www.uniprot.org), GnomAD (https://gnomad.broadinstitute.org), GRIN-Leipzig (http://www.grin-database.de), and CFERV (http://functionalvariants.emory.edu/database/).

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

- Amin, J. B., Moody, G. R., & Wollmuth, L. P. (2020). From bedside-to- bench: What disease-associated variants are teaching us about the NMDA receptor. The Journal of Physiology. Advance online publication. https://doi.org/10.1113/JP278705
- Auvin, S., Dozières-Puyravel, B., Avbersek, A., Sciberras, D., Collier, J., Leclercq, K., Mares, P., Kaminski, R. M., & Muglia, P. (2020). Radiprodil, a NR2B negative allosteric modulator, from bench to bedside in infantile spasm syndrome. Annals of Clinical and Translational Neurology, 7(3), 343–352.
- Burnashev, N., & Szepetowski, P. (2015). NMDA receptor subunit mutations in neurodevelopmental disorders. Current Opinion in Pharmacology, 20, 73–82.
- DeLano, W. L. (2002). PyMOL: An open-source molecular graphics tool. CCP4 Newsletter on Protein Crystallography, 40(1), 82–92.
- de Ligt, J., Willemsen, M. H., van Bon, B. W. M., Kleefstra, T., Yntema, H. G., Kroes, T., Vulto-Van Silfhout, A. T., Koolen, D. A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B. B. A., Brunner, H. G., Veltman, J. A., & Vissers, L. E. L. M. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. The New England Journal of Medicine, 367(20), 1921–1929.
- Esmenjaud, J.-B., Stroebel, D., Chan, K., Grand, T., David, M., Wollmuth, L. P., Taly, A., & Paoletti, P. (2019). An inter-dimer allosteric switch controls NMDA receptor activity. The EMBO Journal, 38(2):e99894. https://doi.org/10.15252/embj.201899894
- Flores-Soto, M. E., Chaparro-Huerta, V., Escoto-Delgadillo, M., Vazquez- Valls, E., González-Castañeda, R. E., & Beas-Zarateaf, C. (2012). Structure and function of NMDA-type glutamate receptor subunits. Neurología (English), 27, 301–310. https://www.sciencedirect.com/science/article/pii/S2173580812000739
- Frank, R. A., & Grant, S. G. (2017). Supramolecular organization of NMDA receptors and the postsynaptic density. Current Opinion in Neurobiology, 45, 139–147.

- Gielen, M., Siegler Retchless, B., Mony, L., Johnson, J. W., & Paoletti, P. (2009). Mechanism of differential control of NMDA receptor activity by NR2 subunits. Nature, 459(7247), 703–707.
- Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation, 4(3), 435–447.
- Karakas, E., & Furukawa, H. (2014). Crystal structure of a heterotetrameric NMDA receptor ion channel. Science, 344(6187), 992–997.
- Karczewski, K., & Francioli, L. (2017). The Genome Aggregation Database (gnomAD). MacArthur Lab. https://ibg.colorado.edu/cdrom2019/nealeB/Gnomad/boulder_190307.pdf
- Korinek, M., Vyklicky, V., Borovska, J., Lichnerova, K., Kaniakova, M., Krausova, B., Krusek, J., Balik, A., Smejkalova, T., Horak, M., & Vyklicky, L. (2015). Cholesterol modulates open probability and desensitization of NMDA receptors. The Journal of Physiology, 593(10), 2279–2293.
- Krivov, G. G., Shapovalov, M. V., Dunbrack, R. L., Jr. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins, 77(4), 778–795.
- Lee, C.-H., Lü, W., Michel, J. C., Goehring, A., Du, J., Song, X., & Gouaux, E. (2014). NMDA receptor structures reveal subunit arrangement and pore architecture. Nature, 511(7508), 191–197.
- Lemke, J. R. (2020). Predicting incidences of neurodevelopmental disorders. Brain: A Journal of Neurology, 143(4), 1046–1048.
- Lemke, J. R., Geider, K., Helbig, K. L., Heyne, H. O., Schütz, H., Hentschel, J., Courage, C., Depienne, C., Nava, C., Heron, D., Møller, R. S., Hjalgrim, H., Lal, D., Neubauer, B. A., Nürnberg, P., Thiele, H., Kurlemann, G., Arnold, G. L., Bhambhani, V., … Syrbe, S. (2016). Delineating the GRIN1 phenotypic spectrum: A distinct genetic NMDA receptor encephalopathy. Neurology, 86(23), 2171–2178.
- Lemke, J. R., Hendrick, R., Geider, K., Laube, B., Schwake, M., Harvey, R. J., James, V. M., Pepler, A., Steiner, I., Hörtnagel, K., Neidhardt, J., Ruf, S., Wolff,

M., Bartholdi, D., Caraballo, R., Platzer, K., Suls, A., De Jonghe, P., Biskup, S., & Weckhuysen, S. (2014). GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. Annals of Neurology, 75(1), 147–154.

- Lesca, G., Rudolf, G., Bruneau, N., Lozovaya, N., Labalme, A., Boutry-Kryza, N., Salmi, M., Tsintsadze, T., Addis, L., Motte, J., Wright, S., Tsintsadze, V., Michel, A., Doummar, D., Lascelles, K., Strug, L., Waters, P., de Bellescize, J., Vrielynck, P., … Szepetowski, P. (2013). GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. Nature Genetics, 45(9), 1061–1066.
- Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I., & Lomize, A. L. (2012). OPM database and PPM web server: Resources for positioning of proteins in membranes. Nucleic Acids Research, 40(Database issue), D370–D376.
- Lussier, M. P., Sanz-Clemente, A., & Roche, K. W. (2015). Dynamic regulation of N-methyl-D-aspartate (NMDA) and α-amino-3- hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors by posttranslational modifications. The Journal of Biological Chemistry, 290(48), 28596–28603.
- Mayer, M. L., Westbrook, G. L., & Guthrie, P. B. (1984). Voltage-dependent block by Mg^{2+} of NMDA responses in spinal cord neurones. Nature, 309(5965), 261–263.
- Mullier, B., Wolff, C., Sands, Z. A., Ghisdal, P., Muglia, P., Kaminski, R. M., & André, V. M. (2017). GRIN2B gain of function mutations are sensitive to radiprodil, a negative allosteric modulator of GluN2B- containing NMDA receptors. Neuropharmacology, 123, 322–331.
- Nowak, L., Bregestovski, P., Ascher, P., Herbet, A., & Prochiantz, A. (1984). Magnesium gates glutamate-activated channels in mouse central neurones. Nature, 307(5950), 462–465.
- O'Roak, B. J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J. J., Girirajan, S., Karakoc, E., MacKenzie, A. P., Ng, S. B., Baker, C., Rieder, M. J., Nickerson, D. A., Bernier, R., Fisher, S. E., Shendure, J., & Eichler, E. E. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nature Genetics, 43(6), 585–589.

- Ogden, K. K., Chen, W., Swanger, S. A., McDaniel, M. J., Fan, L. Z., Hu, C., Tankovic, A., Kusumoto, H., Kosobucki, G. J., Schulien, A. J., Su, Z., Pecha, J., Bhattacharya, S., Petrovski, S., Cohen, A. E., Aizenman, E., Traynelis, S. F., & Yuan, H. (2017). Molecular mechanism of disease-associated mutations in the Pre-M1 helix of NMDA receptors and potential rescue pharmacology. PLOS Genetics, 13(1), e1006536.
- Paoletti, P., Bellone, C., & Zhou, Q. (2013). NMDA receptor subunit diversity: Impact on receptor properties, synaptic plasticity and disease. Nature Reviews Neuroscience, 14(6), 383–400.
- Pierson, T. M., Yuan, H., Marsh, E. D., Fuentes-Fajardo, K., Adams, D. R., Markello, T., Simeonov, D. R., Holloman, C., Tankovic, A., Karamchandani, M. M., Schreiber, J. M., Mullikin, J. C., Tifft, C. J., Toro, C., Boerkoel, C. F., Traynelis, S. F., & Gahl, W. A. (2014). GRIN2A mutation and early-onset epileptic encephalopathy: Personalized therapy with memantine. Annals of Clinical and Translational Neurology, 1(3), 190–198.
- Platzer, K., Yuan, H., Schütz, H., Winschel, A., Chen, W., Hu, C., Kusumoto, H., Heyne, H. O., Helbig, K. L., Tang, S., Willing, M. C., Tinkle, B. T., Adams, D. J., Depienne, C., Keren, B., Mignot, C., Frengen, E., Strømme, P., Biskup, S., … Lemke, J. R. (2017). GRIN2B encephalopathy: Novel findings on phenotype, variant clustering, functional consequences and treatment aspects. Journal of Medical Genetics, 54(7), 460–470.
- Rose, A. S., & Hildebrand, P. W. (2015). NGL Viewer: A web application for molecular visualization. Nucleic Acids Research, 43(W1), W576–W579.
- Santos-Gómez, A., Miguez-Cabello, F., García-Recio, A., Locubiche, S., García-Díaz, R., Soto, V., Guerrero-López, R., Juliá-Palacios, N., Ciruela, F., García-Cazorla, À., Soto, D., Olivella, M., & Altafaj, X. (2020). Disease-associated GRIN protein truncating variants trigger NMDA receptor loss-of-function. Human Molecular Genetics. Advance online publication. https://doi.org/10.1093/hmg/ddaa220
- Soto, D., Olivella, M., Grau, C., Armstrong, J., Alcon, C., Gasull, X., Santos-Gómez, A., Locubiche, S., Gómez de Salazar, M., García- Díaz, R., Gratacòs-

Batlle, E., Ramos-Vicente, D., Chu-Van, E., Colsch, B., Fernández-Dueñas, V., Ciruela, F., Bayés, À., Sindreu, C., López-Sala, A., … Altafaj, X. (2019). L-Serine dietary supplementation is associated with clinical improvement of loss-of-function GRIN2B-related pediatric encephalopathy. Science Signaling, 12(586), eaaw0936. https://doi.org/10.1126/scisignal. aaw0936

- Tajima, N., Karakas, E., Grant, T., Simorowski, N., Diaz-Avalos, R., Grigorieff, N., & Furukawa, H. (2016). GluN1b-GluN2B NMDA receptor structure in non-active-2 conformation. https://doi.org/10. 2210/pdb5fxi/pdb
- Tarabeux, J., Kebir, O., Gauthier, J., Hamdan, F. F., Xiong, L., Piton, A., Spiegelman, D., Henrion, É., Millet, B., Fathalli, F., Joober, R., Rapoport, J. L., DeLisi, L. E., Fombonne, É., Mottron, L., Forget- Dubois, N., Boivin, M., Michaud, J. L., Drapeau, P., … Krebs, M. O. (2011). Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia. Translational Psychiatry, 1, e55.
- Traynelis, S. F., Wollmuth, L. P., McBain, C. J., Menniti, F. S., Vance, K. M., Ogden, K. K., Hansen, K. B., Yuan, H., Myers, S. J., & Dingledine, R. (2010). Glutamate receptor ion channels: Structure, regulation, and function. Pharmacological Reviews, 62(3), 405–496.
- UniProt Consortium. (2019). UniProt: A worldwide hub of protein knowledge. Nucleic Acids Research, 47(D1), D506–D515.
- Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. Current Protocols in Protein Science, 86, 2.9.1–2. 9.37.
- XiangWei, W., Jiang, Y., & Yuan, H. (2018). De novo mutations and rare variants occurring in NMDA receptors. Current Opinion in Physiology, 2, 27–35.
- Yuan, H., Hansen, K. B., Vance, K. M., Ogden, K. K., & Traynelis, S. F. (2009). Control of NMDA receptor function by the NR2 subunit amino-terminal domain. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 29(39), 12045–12058.
- Zhu, S., Stein, R. A., Yoshioka, C., Lee, C.-H., Goehring, A., Mchaourab, H. S., & Gouaux, E. (2016). Mechanism of NMDA receptor inhibition and activation.

Cell, 165(3), 704–714.

Article III. Identification of homologous GluN subunits variants accelerates GRIN variants stratification

ABSTRACT

The clinical spectrum of GRIN-related neurodevelopmental disorders (GRD) results from gene- and variant-dependent primary alterations of the NMDA receptor disturbing glutamatergic neurotransmission. Despite GRIN gene variants functional annotations are dually critical for stratification and precision medicine design, currently genetically diagnosed disease-associated GRIN variants outnumber their respective functional annotations. Based on high-resolution crystal 3D models and topological domains conservation between GluN1, GluN2A and GluN2B subunits, we have generated a high RMSD score GluN1-GluN2A-GluN2B subunits structural superimposition model. We thus hypothesized that functional annotations of specific GRIN missense variants could predict the functional changes in equivalent structural positions in other GluN subunits. First, we validated this computational algorithm experimentally, using an in silico library of GluN2B-equivalent GluN2A artificial variants, designed from disease-associated GluN2B variants. Further, the structural algorithm was exhaustively applied to the full GRIN missense variants repertoire, consisting of 4525 variants. The analysis of this structure-based model revealed an absolute predictive power for GluN1, GluN2A and GluN2B subunits, both in terms of pathogenicity-association (neutral vs. disease-associated variants) and functional impact (loss-of-function, neutral, gain-of-function). Importantly, GRIN structure-based predictive algorithms duplicated the assignment of pathogenic GRIN variants, reduced by 30% the number of GRIN variants with uncertain pathogenesis and increased by 70% the number of annotated variants. Finally, GRIN structural predictive algorithm has been implemented into GRIN variants Database (http://lmc.uab.es/grindb), providing a computational tool that accelerates GRIN missense variants stratification concomitant with clinical therapeutic decision for this neurodevelopmental disorder.

INTRODUCTION

GRIN-related neurodevelopmental disorders (GRDs) constitute a group of rare genetic diseases with a broad clinical spectrum including intellectual disability, epilepsy, movement disorders, development delay, autism spectrum disorder and schizophrenia [1–4]. The primary cause of GRD is the presence of de novo GRIN mutations -with almost exclusively autosomal dominant inheritance pattern- that results in the presence of dysfunctional GluN subunits of the N-methyl D-Aspartate receptor (NMDAR). The NMDAR belongs to the glutamate ionotropic receptors family, and plays a pivotal role in neuronal development, synaptic plasticity and neuron survival [5]. NMDA receptors are composed by two obligatory GluN1 subunits, and a combination of two additional GluN subunits (GluN2A-D, GluN3A,B) encoded by GRIN1, GRIN2A-D and GRIN3A,B, respectively [5]. The NMDA receptor tetrameric architecture consists of an Amino Terminal Domain (ATD), a Ligand Binding Domain (LBD), a Transmembrane Domain (TMD) and a large Carboxy Terminal Domain (CTD, with unsolved structure). Concurrent binding of glycine to the GluN1 subunit and glutamate to the GluN2 subunit is required for NMDAR activation [6] and recent studies shed light on the conformational changes due to receptor activation and inhibition [7, 8]. Since the identification of GRD genetic aetiology [9–11], the number of GRIN variants has exponentially grown, being currently represented by 4520 reported unique variants (http://lmc.uab.es/grindb, July 2021)[12]. Despite some mutations affecting GRIN2C, GRIN2D, GRIN3A and GRIN3B genes have been described, the vast majority (95%) of GRD-associated mutations are associated with GRIN1, GRIN2A and GRIN2B genes. Genotype-phenotype association studies are partially accompanied by functional studies of GRIN variants gene products, allowing to assign both their pathogenicity likelihood and their functional stratification. Currently, a total of 445 GRIN variants are considered as disease-associated variants in patients with neurodevelopmental disorders, while 3800 variants are classified as neutral and 275 variants are considered to be of uncertain pathogenicity. Overall, the number of GRIN variants of certain pathogenicity annotation (i.e. genetic and experimental data supporting variant-associated NMDAR dysfunction) is still limited. The identification of novel GRIN variants in paediatric patients needs experimental validation, both to

assign likely pathogenicity (i.e. evaluation of mutation-related protein dysfunction) and to annotate the functional outcomes that are roughly classified into loss-of-function (LoF), gain-of-function (GoF) or complex effects. In turn, GRIN variant stratification is crucial to provide the molecular diagnosis supporting the selection of a given personalized therapeutic strategy option. Importantly, in the context of the temporal dimension of this neurodevelopmental disorder, accelerating GRIN variants annotation is crucial to early define precision therapeutic arms, either to rescue LoF (L-serine[13]) or GoF variants (memantine[14], radiprodil [15, 16], dextromethorphan).

The lack of high-throughput experimental methods for GRIN variants annotation and their intrinsic position-dependent phenotypic alterations result in a growing gap between GRIN variants genetic identification and their relative GluN subunits functional annotations. Together with numerous efforts to annotate GRIN variants, computational tools provide a powerful strategy to accelerate GRIN variants annotation. In this context, we recently released the GRIN variants Database (GRINdb) [12], comprehensively compiling available genetic, clinical and functional data related to GRIN variants. In addition to providing individual GRIN variants information with clinical outputs, GRIN variants Database analysis can accelerate GRIN variants annotation. In this regard, GRINdb revealed the subunit and domain variables defining GluN truncating variants pathogenicity [17].

In this study, based on GluN subunits structural conservation within the NMDAR tetramer, we have developed and experimentally validated -both in vitro and in silico- a structural computational algorithm significantly extending the pathogenicity and functional parameters of non-annotated (“orphan”) GRIN homologous variants. The algorithm has been implemented into the open access and most comprehensive GRIN Variants Database, providing additional tools for pathogenicity and functional GRIN variants stratification, overall contributing to clinical decision making.

METHODS

Structural Alignment of GluN subunits of the NMDA receptor

GluN1, GluN2A and GluN2B subunits Amino Terminal Domain (ATD), Ligand Binding Domain (LBD) and Transmembrane Domains (TM) (see Supp. Table 1 for detailed domain coordinates) were structurally superimposed using PyMOL 3 [18]. The residue ranges for each domain were determined using OPM [19]. 6IRA X-ray crystal structure [20] was used for GluN1 and GluN2A subunits and 6WHR X-ray crystal structure [8] was used for GluN2B structure. Alpha carbon root mean square deviation (RMSD) and sequence identity was computed for each subunit pair based on the structural sequence alignment. The sequence alignment obtained from the structural alignment was used to identify equivalent positions between GluN1, GluN2A and GluN2B subunits. For the structurally unsolved regions, sequence alignment was used to identify equivalent positions.

Topological conservation of GluN subunits of the NMDA receptor

A molecular model of the tri-heteromeric NMDA receptor (GluN1)2-GluN2A-GluN2B was generated from 4PE5 X-ray crystal structure [21]. Modeller 9.20[22] was used to model the lacking regions of the receptor and Scwrl4 [23] to position the non-determined amino acid sidechains. The initial model was energy-minimized using GROMACS 5 [24]. The carboxy terminal domains were not modeled, as there is no available determined structure. Each residue of the NMDA molecular model was colored according to its conservation (for equivalent positions, between GluN subunits) using Pymol 3 [18].

Computational design and experimental validation of homologous GluN artificial variants

In order to evaluate the structural superimposition model, a collection of nine domain representative GluN2A variants was designed using the structural algorithm, homologous to nine previously reported GluN2B pathogenic variants. Functional annotations were performed for the following pairs of GRIN2B patient-associated variants - GRIN2A in silico-designed variants: 2B(G459R)-2A(G458R); 2B(G484D)-

2A(G483D); 2B(T532A)-2A(T531A); 2B(G543R)-2A(G542R); 2B(G689S)-2A(G688S); 2B(R693G)-2A(R692G), 2B(G820A)-2A(G819A), 2B(G820E)-2A(G819E), 2B(M824V)-2A(M823V).

Plasmids

The expression plasmids for rat HA-GluN1, GFP-GluN2A and GFP-GluN2B were kindly provided by Dr. Nakanishi and Dr. Vicini [25], respectively. Nucleotide changes for the production of GRIN variants were achieved by oligonucleotide-directed mutagenesis, using the QuickChange II XL site-directed mutagenesis kit according to the manufacturer's instructions (Stratagene), and verified by Sanger sequencing.

Cell culture and transfection

HEK-293T and COS-7 cell lines were obtained from the American Type Culture Collection and maintained at 37°C in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal calf serum and antibiotics (100 units/ml penicillin and 100 mg/ml streptomycin) and D-2-amino-5-phosphonopentanoic acid (D-AP5, Abcam; 0.5-1 mM final concentrations, for HEK-293T and COS-7 cells, respectively) to prevent excitotoxicity. Transient expression of NMDARs in HEK-293T cells was achieved with polyethylenimine (PEI)-based transfection methods, and NMDAR-mediated currents were recorded 24 hours after transfection. COS-7 cells were transfected with LipofectamineTM 2000 (Invitrogen) following the manufacturer's instructions, and cells were fixed 24 hours post-transfection for further immunofluorescence analysis. Cells were transfected with equimolar amounts of GluN1 and GFP-GluN2A/GluN2B subunits (1:1) for immunofluorescence experiments and using a 1:2 (GluN1:GluN2) ratio for electrophysiological recordings.

Immunofluorescence analysis

Transiently transfected COS-7 cells were washed in PBS and fixed with 4% paraformaldehyde. Surface expression of NMDARs was achieved by immunolabeling the extracellular GFP tag (GFP cloned in-frame within GluN2 subunits ATD) and incubating with anti-GFP antibody (Clontech) for 1h at RT, under non-permeabilizing conditions. After washing, cells were incubated with anti-rabbit IgG-Alexa555

secondary antibodies (Life Technologies), for 1h at RT. The total amount of GFP-tagged GluN subunits was detected by the GFP endogenous fluorescent signal emitted by GFP-GluN2A/GluN2B constructs. Coverslips were mounted in ProLong antifade mounting medium (Life Technologies) and images were acquired in a Nikon Eclipse 80i microscope (63×/1.4 N.A. immersion oil objective).

Electrophysiological recordings of NMDAR-mediated whole-cell currents in HEK293T cells

Electrophysiological recordings were performed 24 h after transfection, perfusing the cells continuously with extracellular physiological bath solution (in mM): 140 NaCl, 5 KCl, 1 CaCl2, 10 glucose, and 10 HEPES, adjusted to pH 7.42 with NaOH. Glutamate (1 mM, Sigma-Aldrich) and glycine (50 microM; Tocris) were co-applied for 5 sec by piezoelectric translation (P-601.30; Physik Instrumente) of a theta-barrel application tool made from borosilicate glass (1.5 mm o.d.; Sutter Instruments) and the activated currents were recorded in the whole-cell configuration at a holding potential of –60 mV, acquired at 5 kHz and filtered at 2 kHz by means of Axopatch 200B amplifier, Digidata 1440A interface and pClamp10 software (Molecular Devices Corporation). Electrodes with open-tip resistances of 2–4 MΩ were made from borosilicate glass (1.5 mm o.d., 0.86 mm i.d., Harvard Apparatus), pulled with a P-97 horizontal puller (Sutter Instruments) and filled with intracellular pipette solution containing (in mM): 140 CsCl, 5 EGTA, 4 Na2ATP, 0.1 Na3GTP and 10 HEPES, adjusted to pH 7.25 with CsOH. Glutamate and glycine-evoked currents were expressed as current density (-pA/pF; maximum current divided by input capacitance, as measured from the amplifier settings) to avoid differences due to surface area in the recorded cells. The kinetics of deactivation and desensitization of the NMDAR responses were determined by fitting the glutamate/glycine-evoked responses at Vm –60 mV to a double-exponential function in order to determine the weighted time constant (τw,des):

$$\tau_{w,des} = \tau_f \left(\frac{A_f}{A_f + A_s} \right) + \tau_s \left(\frac{A_s}{A_f + A_s} \right)$$

where Af and τf correspond to the amplitude and time constant of the fast component of desensitization and As and τs are the amplitude and time constant of the slow component of desensitization.

Statistical analysis

Comparison between experimental groups was evaluated using Prism9 (GraphPad Software, Inc.), applying a One-Way Analysis of Variance (ANOVA) followed by a Bonferroni post-hoc test (cell surface expression experiments) or or Mann-Whitney U-test (for electrophysiology experiments). Data are presented as the mean ± SEM from at least three independent experiments.

Comparative analysis of pathogenicity and functional changes for equivalent GluN1, GluN2A and GluN2B subunits positions

In order to evaluate the pathogenesis and functional annotations of GluN1, GluN2A and GluN2B subunits missense variants in topologically equivalent positions -according to the structural superimposition model-, reported GRIN variants were retrieved from the GRIN variants Database [12]. GRIN variants pathogenicity likelihood (disease-associated, uncertain pathogenesis or neutral) and functional annotations (LoF/GoF/Complex) for equivalent positions were compared for (i) homologous mutations, i.e. same initial and final amino acid, (ii) different initial amino acids with same final amino acid, (iii) same initial amino acid with different final amino acids.

RESULTS

Structural alignment of GluN1, GluN2A and GluN2B subunits domains

Prior to GluN subunits structural alignment, the topology of monomeric GluN subunits was extracted from the recently reported heterotetrameric N-methyl D-aspartate receptor (NMDAR) crystal structures [8, 20]. Further, a pairwise structural alignment of GluN1, GluN2A and GluN2B subunits domains has been performed, delivering the annotation of GluN subunits equivalent positions across the extracellular and transmembrane domains (Supp. Figure 1). Quantitative analysis of GluN subunits structural alignment accuracy revealed a high structural conservation degree between

GluN subunits domains. Indeed, root mean square deviation (RMSD) values revealed the topological conservation (ranging from very high to high) between GluN subunits (see Table 4.3-8). Further, the resulting GluN subunit pairs structural alignment allowed to compute and to define the corresponding amino acid positions identity (Figure 4.3-5). In terms of subunit sequence identity, the analysis revealed the presence of a high sequence identity between GluN2A and GluN2B subunits and a moderate conservation between GluN1-GluN2A or GluN1-GluN2B subunits. In terms of functional domains, the analysis showed a high sequence identity and structure conservation within the LBD and TMD, and to a lesser extent in the amino acid sequence of the ATD. The domain-specific structural conservation is coincident with the distribution of GRIN variants disease-association. Indeed, GRIN variants pathogenicity is more elevated in amino acids located within the LBD and TMD, while GRIN missense variants within ATD are variably associated with neurological conditions. Despite some particular NMDAR subdomains exhibiting limited sequence identity, overall RMSD values supported the generation of a GluN subunits structural superimposition model, potentially allowing the identification of precise equivalent positions between GluN subunits. This is in compliance with a higher structure than sequence conservation in proteins [26], which is particularly more pronounced in membrane proteins [27].

		GluN2A-GluN1	GluN2B-GluN1	GluN2A-GluN2B
ATD	*Identical residues*	60 (out of 401)	48 (out of 402)	220 (out of 402)
	Sequence Identity	14,96%	11,94%	54,73%
	RMSD (ATD1)	2.034	3.106	1.096
	RMSD (ATD2)	1.062	1.324	0.989
LBD	*Number of identical residues*	104 (out of 317)	99 (out of 317)	265 (out of 317)
	Sequence Identity	32,81%	31,23%	83,60%
	RMSD	1.229	1.653	0.832
TMD	*Number of identical residues*	44 (out of 140)	46 (out of 140)	131 (out of 140)
	Sequence Identity	31,65%	33,09%	92,86%
	RMSD	1.899	1.774	1.312

Table 4.3-8. Amino acid sequence identity and root mean standard deviation (RMSD) between GluN subunits, based on pairwise structural alignment of GluN subunit topological domains. ATD, amino terminal domain; LBD, ligand binding domain; TMD, transmembrane domain.

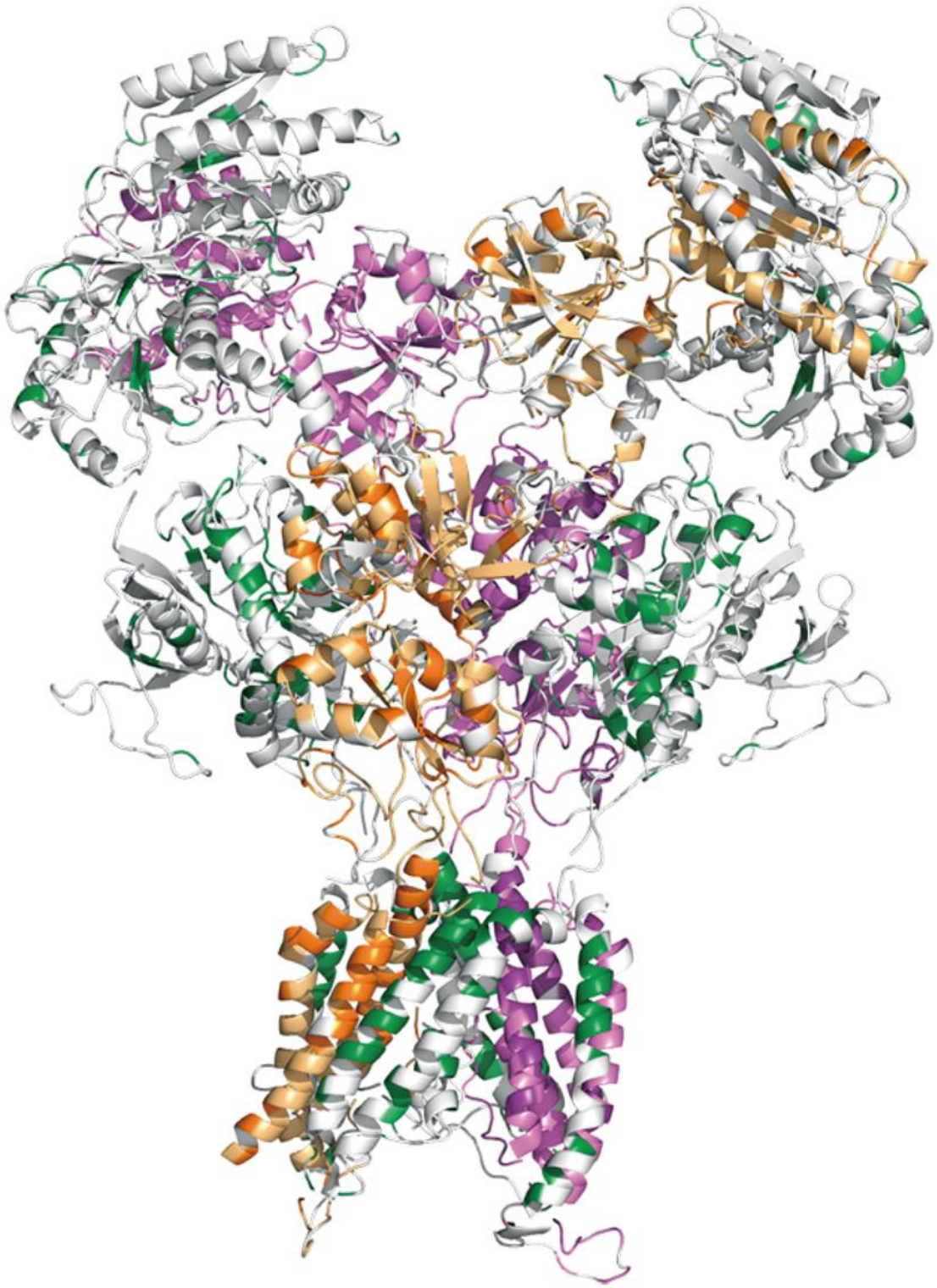
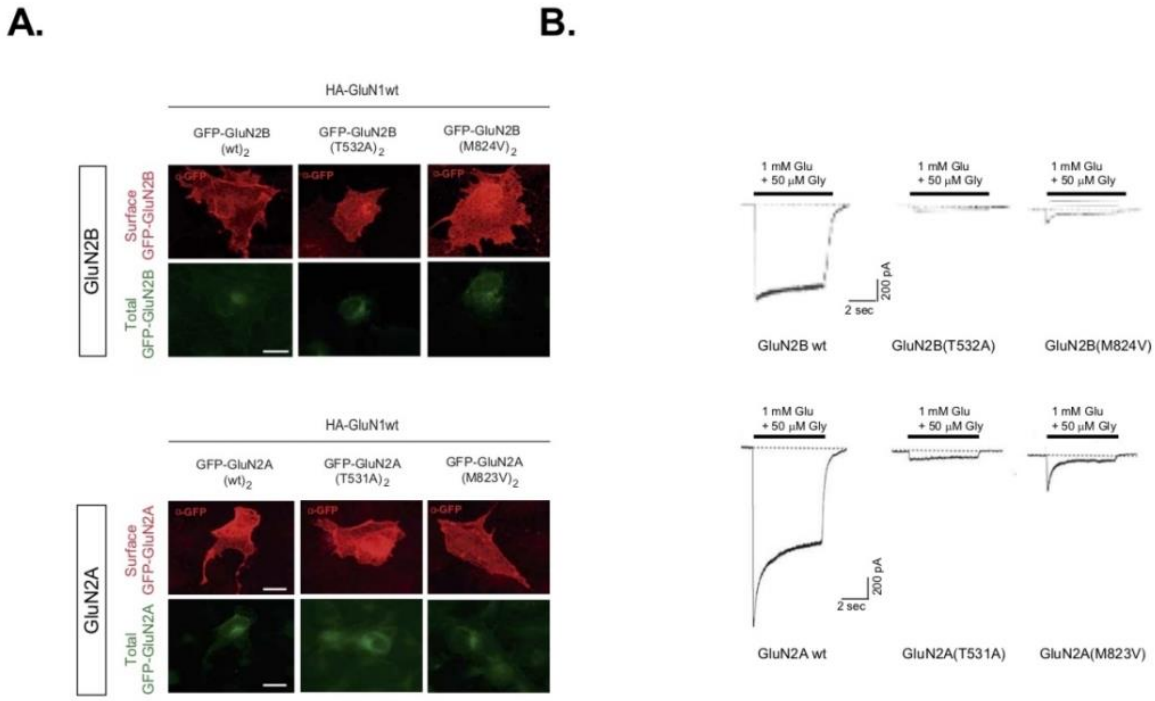


Figure 4.3-5. **Structural molecular model of the $(GluN1)_2$-GluN2A-GluN2B tri-heteromeric NMDA receptor showing amino acid sequence conservation across GluN subunits.** Based on domain pairwise structural alignment, amino acid positions of the GluN1, GluN2A and GluN2B subunits were

compared. Residue conservation in each position of the GluN1, GluN2A and GluN2B subunits are coloured following an intensity-gradient code, to depict amino acids conservation. Residue conservation color code: white residues indicate non-conserved residues; pale colors indicate partial residue conservation (between two GluN subunits) and dark color indicates residue conservation in the three subunits. GluN subunits color code: GluN2A, violet; GluN2B, orange; GluN1 subunits, green (gradient colour, illustrating amino acid conservation) and in white.

Structural model-based design of equivalent GluN2B-GluN2A subunits missense variants

Upon the generation of the GluN subunits structural model, the algorithm predictive power was experimentally evaluated *in vitro*. Based on GRD patients-associated *GRIN2B* variants referred to our group, we conducted the functional annotation of *in silico*-designed *GRIN2A* ("artificial") variants putatively equivalent and compared their functional outcomes to those from *GRIN2B* disease-associated variants. The artificial *GRIN2A* variants were selected to representatively cover *GRIN* mutation vulnerable domains (e.g. prevalently associated with *GRIN de novo* pathogenic mutations), e.g. localised at the glutamate binding domain (6 variants) and the transmembrane domain (3 variants). The comparative analysis of the functional impact of GluN2A-GluN2B homologous pairs was systematically performed in parallel, using mammalian heterologous expression systems. First, mutant NMDAR surface expression was assessed by immunofluorescence analysis in COS-7 cells transiently expressing di-heteromeric $(GluN1wt)_2$-$(GluN2Amut)_2$ or $(GluN1wt)_2$-$(GluN2Bmut)_2$ and showed surface trafficking patterns conservation between artificial GluN2A variants and equivalent GRD-associated GluN2B pairs (Supp. Fig. 2). Indeed, *GRIN2B* mutations leading to mild or unaffected NMDAR surface trafficking were compared with their relative pairs, showing an overall conserved impact (Supp. Fig. 2). This pattern conservation was specially noticed in a pair of variants showing a drastic effect, e.g. abolishing mutant NMDARs surface trafficking (*i.e.*, GluN2A(p.G458R)-GluN2B(p.G459R) pair).



C.

GRIN variant	Topological Domain	Surface expression (%)	Normalised current density (% vs. wild-type)	S-S (%)	Desensitization rate (msec)	Deactivation rate (msec)	Functional classification
GluN2B-WT		100 ± 1.5 (N=290)	100 ± 7.74 (N=90)	58.59 ± 1.89 (N=80)	2.16 ± 0.26 (N=34)	0.83 ± 0.32 (N=47)	
GluN2A-WT		100 ± 3.9 (N=42)	100 ± 27.02 (N=36)	47.70 ± 1.78 (N=35)	1.01 ± 0.10 (N=21)	0.26 ± 0.01 (N=23)	
GluN2B-G459R	LBD	1.0 ± 0.6 (N=28) ****	NA	NA	NA	NA	LoF
GluN2A-G458R		3.0 ± 0.8 (N=26) ****	NA	NA	NA	NA	LoF
GluN2B-G484D	LBD	85.9 ± 7.8 (N=24) ns	24.61 ± 6.26 (N=6) ***	53.73 ± 7.41 (N=7) ns	1.23 ± 0.23 (N=6) ns	0.23 ± 0.04 (N=8) ***	LoF
GluN2A-G483D		89.2 ± 6.7(N=20) ns	50.09 ± 11.64 (N=10) ns	53.84 ± 4.41 (N=9) ns	0.82 ± 0.07 (N=6) ns	0.33 ± 0.05 (N=8) ns	LoF
GluN2B-T532A	LBD	117.6 ± 7.0 (N=26) ns	24.69 ± 5.19 (N=11) ****	69.46 ± 4.46 (N=9) ns	0.86 ± 0.36 (N=2) ns	0.22 ± 0.18 (N=2) ns	LoF
GluN2A-T531A		105.7 ± 7.6 (N=20) ns	3.74 ± 1.10 (N=6) ****	51.13 ± 6.92 (N=7) ns	1.15 ± 0.38 (N=4) ns	0.31 ± 0.06 (N=6) ns	LoF
GluN2B-G543R	LBD	91.6 ± 8.8 (N=35) ns	12.33 ± 4.25 (N=9) ****	53.99 ± 9.15 (N=8) ns	ND	1.77 ± 0.44 (N=4) ***	Complex
GluN2A-G542R		120.6 ± 7.7 (N=23) ns	2.20 ± 0.32 (N=8) ****	22.71 ± 2.44 (N=8) ****	ND	0.41 ± 0.05 (N=7) **	Complex
GluN2B-G689S	LBD	103.9 ± 12.0 (N=27) ns	17.72 ± 4.42 (N=5) ***	96 ± 2.78 (N=5) ****	ND	0.08 ± 0.04 (N=5) ***	LoF
GluN2A-G688S		98.2 ± 6.5 (N=21) ns	12.48 ± 1.99 (N=7) **	68.6 ± 5.84 (N=6) ***	ND	0.32 ± 0.09 (N=4) ns	LoF
GluN2B-R693G	LBD	88.8 ± 4.3 (N=20) ns	9.80 ± 2.74 (N=8) ****	ND	ND	0.35 ± 0.09 (N=3) ns	LoF
GluN2A-R692G		132.9 ± 9.6 (N=24) *	65.34 ± 13.94 (N=12) ns	ND	ND	0.31 ± 0.02 (N=12) ns	LoF
GluN2B-G820A	M4	106.7 ± 7.4 (N=27) ns	2.08 ± 0.97 (N=5) ****	ND	ND	ND	LoF
GluN2A-G819A		130.1 ± 9.9 (N=28) *	ND	ND	ND	ND	LoF
GluN2B-G820E	M4	110.7 ± 6.7 (N=23) ns	ND	ND	ND	ND	LoF
GluN2A-G819E		96.6 ± 6.6 (N=20) ns	ND	ND	ND	ND	LoF
GluN2B-M824V	M4	119.6 ± 2.6 (N=66) ***	70.36 ± 28.21 (N=7) ns	23.17 ± 3.02 (N=6) ****	0.24 ± 0.03 (N=6) ****	0.48 ± 0.10 (N=7) ns	LoF
GluN2A-M823V		133.7 ± 9.2 (N=23) *	38.95 ± 21.31 (N=4) ns	13.40 ± 3.49 (N=4) ****	0.62 ± 0.21 (N=4) ns	0.23 ± 0.01 (N=4) ns	LoF

Figure 4.3-6. Comparative analysis of functional alterations in GluN2A-GluN2B predicted equivalent pairs. A, Representative images of immunofluorescence detection of surface trafficking from artificial GluN2A variants-containing NMDARs (upper panels), designed upon respective patients-associated GluN2B variants (lower panels). Transfected COS-7 cells heterologously expressing GFP-GluN2A/B subunits (green channel, intracellular expression) were immunostained and surface expression of mutant GluN2A/B labeled (red channel). **B**, Representative electrophysiological traces of NMDAR-mediated currents in HEK-293T cells transfected with GluN1 and GluN2A/2B pairs (upper

panels: $(GluN1wt)_2$-$(GluN2Amut)_2$ mediated currents; lower panels, $(GluN1wt)_2$-$(GluN2Bmut)_2$). **C**, Summary of functional annotations of GluN2A/2B pairs. Data representing mean +/- SEM. art, artificial GRIN2A variant; LBD, ligand binding domain; TMD, transmembrane domain; ns, non-significant statistical difference; * P < 0.05, ** P < 0.01, *** P < 0.001; ND, not detectable; NA, not assigned.

Characterisation of the biophysical behaviour of mutant NMDARs was assessed in HEK-293T cells transiently expressing mutant di-heteromeric receptors. Similarly to surface trafficking pattern effect between GluN2A-GluN2B equivalent pairs, whole-cell patch clamp experiments showed an overall conserved electrophysiological impact (Figure 4.4-6B, 4.4-6C, Supp. Figure 3). The integration of putative biophysical parameters disturbances (normalized current amplitude, channel gating kinetics) induced by GRIN2A-GRIN2B variants pairs showed an overall coincident functional output in GluN2A-GluN2B variants pairs (Figure 4.3-6). In summary, the experimental evaluation of the structural algorithm predictive power showed the reliability of the model, and further comprehensive analysis of functionally annotated *GRIN* variants was computationally performed.

Evaluation of GRIN predictive structural model likelihood in comprehensive *GRIN* variants repertoire

The predictive power of the GRIN structural model was investigated in the full repertoire of functionally annotated and stratified *GRIN* variants, retrieved from *GRIN* variants database (GRINdb). Disease-association and functional stratification potential correlation of annotated putative GluN1, GluN2A and GluN2B subunits equivalent positions were assessed.

Comparative analysis of homologous mutations

First, 87 missense variants corresponding to 33 homologous variants, *i.e*. same amino acid in equivalent positions between subunits, were identified. For each pair or trio of homologous variants, qualitative disease-association was compared. Mutations with uncertain pathogenesis were discarded, resulting in 28 neutral variants and 50 disease-associated filtered variants (see Supp. Table 2, Figure 4.3-7A). Importantly, in terms of disease-association, the comparison of these 78 variants showed a 100%

coincidence, indicating a complete disease-association pattern conservation across the annotated GluN1-GluN2A-GluN2B subunits missense variants.

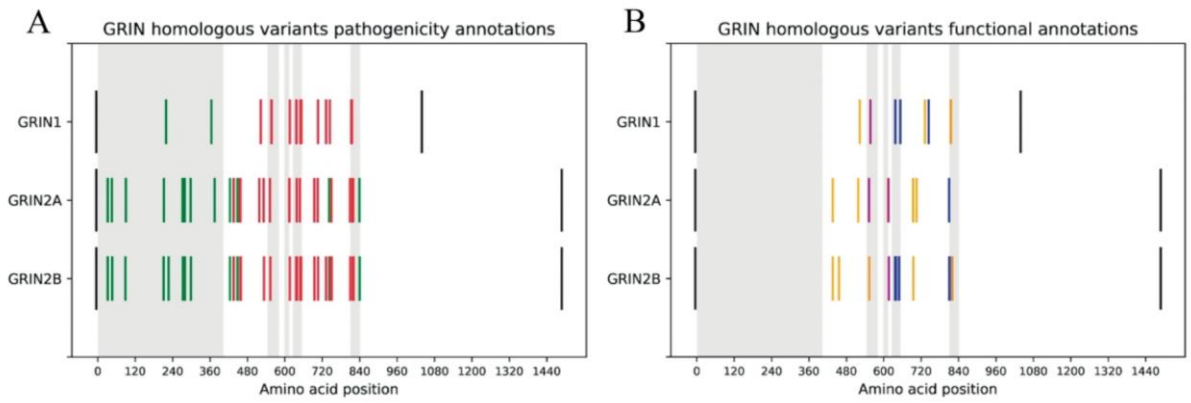


Figure 4.3-7. A. Comparative analysis of *GRIN* homologous variants disease-association. Homologous variants present the same pattern of pathogenesis across GluN1, GluN2A and GluN2B subunits. Disease-associated variants are colored in red and neutral variants are colored in green. Black bars represent initial and final amino acids of canonical GluN1, GluN2A and GluN2B subunits. Grey rectangles represent the amino-terminal domain, and transmembrane domains (TM1, TM2, TM3, TM4, respectively). **B.** Functional annotations of *GRIN* homologous mutations. LoF variants are colored in yellow, while GoF mutations are colored in blue.

The functional annotations for homologous mutations were extracted for comparison, resulting in 8 pairs of homologous variants with available LoF/GoF/Complex functional classification (see Supp. Table 2). The functional classification of these 8 pairs of homologous mutations was coincident (see Figure 4.3-7B), strongly supporting that pathogenicity and functional disturbances induced by homologous variants affecting ATD, LBD and TMD domains of GluN1, GluN2A and GluN2B are coincident.

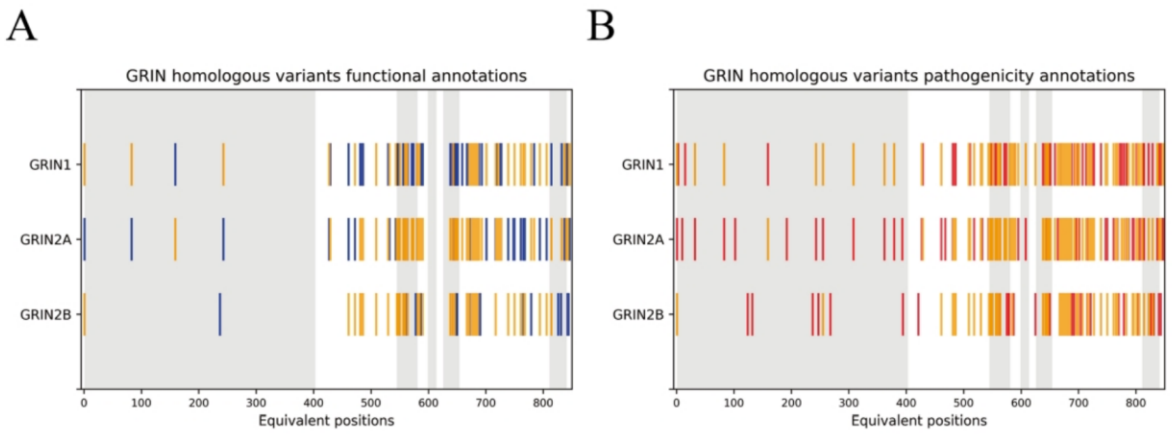
Comparative analysis of non-homologous mutations in equivalent positions

First, the pathogenesis and functional annotations of 27 variants of 13 equivalent positions with different initial amino acid and identical final amino acid were compared (see Supp. Table 3). With the exception of GluN1(K685R)-GluN2B(P687R), disease-association of non-homologous mutations in equivalent positions (12 pairs variants out of 13) was coincident in. Amongst them, in terms of functional classification, the only pair with available functional annotations showed coincident functional outcomes.

Second, the pathogenesis and functional annotations of 210 mutations from 72 equivalent positions with same initial amino acid and different final amino acid were compared (see Supp. Table 4). From these variants, 59 pairs (out of 72) were coincident in terms of disease-association and functional annotations were coincident in 23 out of 30 pairs. Interestingly, some variants affecting the same initial residue mutated into different final residues revealed a differential pathological outcome. Indeed, GluN2B(p.A639V) *vs*. GluN2B(p.A639S) comparison showed either a neutral effect or disease-associated. These data suggest that besides residue topological location, the final amino acid intrinsic physicochemical properties strongly determine the structural, functional and pathological impact. Overall, these findings indicate that non-homologous mutations of *GRIN1*, *GRIN2A* and *GRIN2B* genes can neither be used to extrapolate disease-association nor functional annotations.

GRIN structural algorithm-based extension of *GRIN* variants disease-association and functional annotations

Upon experimental and computational validation of the structural comparison algorithm between GluN1, GluN2A and GluN2B subunits, the model was used to extend the pathogenesis and functional annotations for homologous missense mutations affecting these GluN subunits. By applying the structural comparison algorithm, the initial number of disease-associated variants was duplicated (290 variants computationally annotated, representing a 100% increase) while the initial number of variants with uncertain patogenesis was reduced from 98 to 71 (representing a 28% decrease of non-stratified *GRIN1*, *GRIN2A*, *GRIN2B* missense variants) (see Figure 4.3-8A). Concomitantly, functional annotations of 87 *GRIN1*, 167 *GRIN2A* and 157 *GRIN2B* missense variants were increased (representing a 68% increase for *GRIN1*, *GRIN2A* and *GRIN2B* functional annotations; see Figure 4.3-8B). Figure 4.4-8C quantitatively summarizes the result of GRIN structural algorithm application to GRINdb, showing both an increase in the number of disease-associated variants and in the classification as LoF/GoF/Complex using GRIN structural algorithm.



C

	GluN1	*GluN2A*	*GluN2B*	*Total variants*
Disease-associated variants	156 (75)	188 (94)	216(108)	560 (270)
	108%	100%	100%	102,17%
Uncertain pathogenesis variants	13 (19)	28(39)	30(40)	71(98)
	-31,58%	-28,21%	-25,00%	-27,55%
Functionally annotated variants	87(38)	167(98)	157(109)	411(245)
	128,95%	70,41%	44,04%	67,70%

Figure 4.3-8. A. GluN variants pathogenicity annotations based on current annotations (in red) and on GRIN structural algorithm-inferred annotations from homologous mutations (in yellow). **B.** GluN variants functional annotations based on current available annotations (in blue) and GRIN structural algorithm-inferred annotations from homologous mutations (in yellow). **C.** Summary of GRIN structural algorithm-based computational annotation of non-classified GluN1, GluN2A and GluN2B subunits missense variants. GRIN predictive power of the algorithm was applied and showed an expansion of disease-associated (top row) *GRIN* variants, together with a reduction of *GRIN* variants number previously classified as “Uncertain pathogenesis” (central row) and an increase of functional annotations of *GRIN* variants affecting the ATD, LBD and TMD domains (bottom row).

GRIN structural algorithm model implementation into GRIN variants public database

Upon validation, the GRIN structural algorithm was implemented into the most comprehensive GRIN variants open access database (https://lmc.uab.es/grindb/home) (García-Recio et al.) This computational tool was adapted using a user-friendly interface, allowing the submission of individual GRIN variants queries, and retrieving the available disease-association and functional

annotation based on homologous variants identified by means of the GRIN structural algorithm.

DISCUSSION

In summary, by combining NMDAR structural models and intrinsic GluN subunits properties (e.g., topological conservation between GluN1, GluN2A and GluN2B subunits), we have developed a structure based GluN subunits sequence homology model. This model has been validated under standardized experimental approaches using *in silico*-designed GluN2A-GluN2B pairs, and further comprehensively evaluated in the bulk of previously annotated GluN1, GluN2A and/or GluN2B missense variants pairs and trios. The predictive power likelihood of the algorithm strongly enlarged the repertoire of disease-associated and annotated GRIN variants, providing a new functional diagnosis tool for the stratification of de novo GRIN1, GRIN2A and GRIN2B missense variants. Importantly, the implementation of this structural algorithm will accelerate GRIN variants stratification which, in the context of this variant-dependent pathophysiology, will contribute to decision-making for GRD personalised therapeutic approaches.

Disease-association and functional stratification of GRIN variants in GRD patients represents an important bottleneck to define the functional outcomes of GRIN variants and to envision and/or to evaluate personalised therapies. Indeed, following genetic diagnosis of a GRIN variant likely pathogenic in GRD individuals, disease-association assignment is required. Further, functional stratification (LoF/GoF/Complex) of GRIN disease-associated variants is crucial for clinical decision-making towards a precision therapy (e.g., radiprodil, memantine, and/or L-serine treatments)(Soto et al., 2019)). Currently, the prediction of *GRIN* variants pathogenicity is based on genetic association studies and/or the existence of previous functional studies on coincident *GRIN* variants. Despite genetic data currently providing valuable insights on vulnerable / tolerant genetic regions (García-Recio et al., 2021), the NMDAR tridimensional architecture is complex (multiple functional domains and interactions motifs) and precise genotype-phenotype correlation is hindered.

Recently, the implementation of next-generation sequencing in the clinical practice increased the amount of novel *GRIN* variants diagnosed, outnumbering the small throughput intrinsically associated with manual functional annotation studies (molecular biology, heterologous expression, microscopy and electrophysiological readouts). In order to circumvent this hurdle, functional annotation of novel ("orphan") *GRIN* variants can potentially benefit from the use of computational-based models with experimental validation. Indeed, as recently reported for *GRIN* protein truncating variants(Santos-Gómez et al., 2021), *in silico* approaches represent a fast and costless approach, ultimately accelerating GRD patient stratification and therapeutic decisions.

The structural homology algorithm hereby developed is based on the intrinsic higher conservation of 3D structure (over sequence conservation) of membrane proteins compared to globular proteins(Olivella et al., 2013). Based on this, we have built a high quality structural alignment of GluN1, GluN2A and GluN2B ATD, LBD and CTD domains. This alignment has identified equivalent positions in GluN subunits that are predicted to maintain the same role in the structure and function of NMDA receptors. The experimental validation of i*n silico*-designed *GRIN2A* variants putatively homologous to disease-associated *GRIN2B* variants has shown that pathogenesis and functional annotations can be extrapolated for homologous mutations. Importantly, by using this structural alignment and identifying conserved equivalent positions, we have extended the pathogenesis and functional annotations of *GRIN1*, *GRIN2A* and *GRIN2B* variants, duplicating the number of pathogenic variants, decreasing by 27.5% the number of variants with an uncertain pathogenesis. Additionally, the number of functional annotations have increased by 67.7%, therefore narrowing the gap between genetically diagnosed *GRIN* disease-associated variants and their functional stratification. Beyond the direct interest for current and future *GRIN de novo* variants annotation and therapeutic advice, this innovative multidisciplinary experimental/computational approach could be extended to other channelopathies of primary genetic aetiology, including but not limited to cardiopathies, skeletal muscle dystrophies, and metabolic disorders.

SUPPLEMENTARY INFORMATION

The supplementary information can be found here.

REFERENCES

1. Lesca G, Rudolf G, Bruneau N, Lozovaya N, Labalme A, Boutry-Kryza N, et al. GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. Nat Genet. 2013;45:1061–6.
2. Lemke JR, Hendrickx R, Geider K, Laube B, Schwake M, Harvey RJ, et al. GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. Ann Neurol. 2014;75:147–54.
3. Lemke JR, Geider K, Helbig KL, Heyne HO, Schütz H, Hentschel J, et al. Delineating the GRIN1 phenotypic spectrum: A distinct genetic NMDA receptor encephalopathy. Neurology. 2016;86:2171–8.
4. Burnashev N, Szepetowski P. NMDA receptor subunit mutations in neurodevelopmental disorders. Curr Opin Pharmacol. 2015;20:73–82.
5. Paoletti P, Bellone C, Zhou Q. NMDA receptor subunit diversity: impact on receptor properties, synaptic plasticity and disease. Nat Rev Neurosci. 2013;14:383–400.
6. Furukawa H, Singh SK, Mancusso R, Gouaux E. Subunit arrangement and function in NMDA receptors. Nature. 2005;438:185–92.
7. Wang H, Lv S, Stroebel D, Zhang J, Pan Y, Huang X, et al. Gating mechanism and a modulatory niche of human GluN1-GluN2A NMDA receptors. Neuron. 2021. doi:10.1016/j.neuron.2021.05.031.

8. Chou T-H, Tajima N, Romero-Hernandez A, Furukawa H. Structural Basis of Functional Transitions in Mammalian NMDA Receptors. Cell. 2020;182:357–71.e13. doi:10.1016/j.cell.2020.05.052.
9. Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I, et al. Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat Genet. 2010;42:1021–6.
10. Tarabeux J, Kebir O, Gauthier J, Hamdan FF, Xiong L, Piton A, et al. Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia. Transl Psychiatry. 2011;1:e55–e55.
11. de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med. 2012;367:1921–9.
12. García-Recio A, Santos-Gómez A, Soto D, Julia-Palacios N, García-Cazorla À, Altafaj X, et al. GRIN database: A unified and manually curated repertoire of GRIN variants. Hum Mutat. 2021;42:8–18.
13. Soto D, Olivella M, Grau C, Armstrong J, Alcon C, Gasull X, et al. L-Serine dietary supplementation is associated with clinical improvement of loss-of-function GRIN2B-related pediatric encephalopathy. Sci Signal. 2019;12. doi:10.1126/scisignal.aaw0936.
14. Pierson TM, Yuan H, Marsh ED, Fuentes-Fajardo K, Adams DR, Markello T, et al. GRIN2A mutation and early-onset epileptic encephalopathy: personalized therapy with memantine. Annals of clinical and translational neurology. 2014;1:190–8.
15. Mullier B, Wolff C, Sands ZA, Ghisdal P, Muglia P, Kaminski RM, et al. GRIN2B gain of function mutations are sensitive to radiprodil, a negative allosteric modulator of GluN2Bcontaining NMDA receptors. Neuropharmacology. 2017;123:322–31.
16. Auvin S, Dozières-Puyravel B, Avbersek A, Sciberras D, Collier J, Leclercq K, et al. Radiprodil, a NR2B negative allosteric modulator, from bench to bedside in infantile spasm syndrome. Ann Clin Transl Neurol. 2020;7:343–52.

17. Santos-Gómez A, Miguez-Cabello F, García-Recio A, Locubiche-Serra S, García-Díaz R, Soto-Insuga V, et al. Disease-associated GRIN protein truncating variants trigger NMDA receptor loss-of-function. Hum Mol Genet. 2021;29:3859–71.
18. DeLano WL. PyMOL. 2002. http://virology.wisc.edu/acp/Classes/DropFolders/Drop660_lectures/2013_660/L01_PyMOL_ 2013r.pdf.
19. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. 2012;40 Database issue:D370–6.
20. Zhang J-B, Chang S, Xu P, Miao M, Wu H, Zhang Y, et al. Structural Basis of the Proton Sensitivity of Human GluN1-GluN2A NMDA Receptors. Cell Reports. 2018;25:3582–90.e4. doi:10.1016/j.celrep.2018.11.071.
21. Karakas E, Furukawa H. Crystal structure of a heterotetrameric NMDA receptor ion channel. Science. 2014;344:992–7.
22. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Protein Sci. 2016;86:2.9.1–2.9.37.
23. Krivov GG, Shapovalov MV, Dunbrack RL Jr. Improved prediction of protein side-chain conformations with SCWRL4. Proteins. 2009;77:778–95.
24. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput. 2008;4:435–47.
25. Vicini S, Wang JF, Li JH, Zhu WJ, Wang YH, Luo JH, et al. Functional and Pharmacological Differences Between RecombinantN-Methyl-d-Aspartate Receptors. J Neurophysiol. 1998;79:555–66.
26. Rodionov MA, Blundell TL. Sequence and structure conservation in a protein core. Proteins: Structure, Function, and Genetics. 1998;33:358–66. doi:10.1002/(sici)1097- 0134(19981115)33:3<358::aid-prot5>3.0.co;2-0.
27. Olivella M, Gonzalez A, Pardo L, Deupi X. Relation between sequence and structure in membrane proteins. Bioinformatics. 2013;29:1589–92.

Article IV. Disease-associated GRIN protein truncating variants trigger NMDA receptor loss-of-function

ABSTRACT

De novo GRIN variants, encoding for the ionotropic glutamate NMDA receptor subunits, have been recently associated with GRIN-related disorders, a group of rare paediatric encephalopathies. Current investigational and clinical efforts are focused to functionally stratify GRIN variants, towards precision therapies of this primary disturbance of glutamatergic transmission that affects neuronal function and brain. In the present study, we aimed to comprehensively delineate the functional outcomes and clinical phenotypes of GRIN protein truncating variants (PTVs)—accounting for ~20% of disease-associated GRIN variants—hypothetically provoking NMDAR hypofunctionality. To tackle this question, we created a comprehensive GRIN PTVs variants database compiling a cohort of nine individuals harbouring GRIN PTVs, together with previously identified variants, to build-up an extensive GRIN PTVs repertoire composed of 293 unique variants. Genotype–phenotype correlation studies were conducted, followed by cell-based assays of selected paradigmatic GRIN PTVs and their functional annotation. Genetic and clinical phenotypes meta-analysis revealed that heterozygous GRIN1, GRIN2C, GRIN2D, GRIN3A and GRIN3B PTVs are non-pathogenic. In contrast, heterozygous GRIN2A and GRIN2B PTVs are associated with specific neurological clinical phenotypes in a subunit- and domain-dependent manner. Mechanistically, cell-based assays showed that paradigmatic pathogenic GRIN2A and GRIN2B PTVs result on a decrease of NMDAR surface expression and NMDAR-mediated currents, ultimately leading to NMDAR functional haploinsufficiency. Overall, these findings contribute to delineate GRIN PTVs genotype–phenotype association and GRIN variants stratification. Functional studies showed that GRIN2A and GRIN2B pathogenic PTVs trigger NMDAR hypofunctionality, and thus accelerate therapeutic decisions for this neurodevelopmental condition.

INTRODUCTION

Glutamate is the main excitatory amino acid neurotransmitter in the brain, with about 90% of excitatory synapses using glutamate for neuronal communication. Upon release in the synaptic cleft, glutamate acts on a group of transmembrane receptors functionally separated into metabotropic (mGluRs) and ionotropic glutamate receptors (iGluRs) families. The latter are pharmacologically divided into three main families, namely N-methyl-D-aspartate (NMDA), AMPA and kainate acid receptors. NMDA receptors (NMDARs) are critical for neuronal survival, circuits formation and synaptic plasticity processes, among others (1). Concomitantly, disturbance of the NMDAR function, by means of pharmacological, genetic or modulation of protein–protein interactions, can alter NMDAR-mediated currents which, in turn, can affect glutamatergic neurons, circuits and brain activity. Indeed, the alteration of NMDAR-dependent circuits have been associated with neurodevelopmental disorders (NDDs), epilepsy, neuropsychiatric diseases as well as neurodegenerative processes (2). The NMDAR is an heterotetramer resulting from the oligomerization of two GluN1 subunits and a combination of two additional subunits (GluN2A-D, GluN3A, GluN3B) whose distribution is finely regulated in a spatio-temporal manner (1). The subunit composition dictates the cellular and temporal expression pattern, subcellular localization and channel properties of the NMDAR, a coincidence detector of pre- and postsynaptic activities (1,3). With the emergence of next-generation sequencing, a growing number of genetic studies reported the association between genetic variants of GRIN genes (encoding for GluN subunits of the NMDAR) and neurological disorders. Initial studies describing de novo GRIN2A and GRIN2B mutations in discrete individuals with intellectual disability (ID) and/or epilepsy (4–6), together with exome analysis implementation in clinical diagnosis and functional studies, allowed to define GRDs (OMIM # 138 249, 138 253,138 252 and 602 717, for GRIN1-, GRIN2A-, GRIN2B- and GRIN2D-related disorders, respectively). GRDs are a group of rare pediatric encephalopathies resulting from the presence of pathogenic GRIN gene variants. To date, about 500 individuals harbouring likely pathogenic GRIN variants have been reported worldwide (www.grin-database.de), although GRD prevalence is probably underestimated as for other NDD of genetic origin. Clinically, this neurodevelopmental

condition is manifested by a spectrum of neurological and systemic alterations, including ID, hypotonia, communication

impairment, epilepsy, movement and sleep disorders and gastrointestinal disturbances (7–12). A comprehensive analysis of GRIN gene variants and associated clinical phenotypes, together with the resulting NMDAR structural modelling and functional annotations of GRIN variants, strongly supports the view of subunit- and domain-specificity of GRIN variants pathogenicity and clinical outcomes. Genetically, GRD mostly results from the presence of GRIN de novo variants, with an autosomal dominant inheritance pattern. GRIN variants are predominantly associated with single nucleotide variations causing missense mutations in particularly sensitive domains of GluN subunits. Besides missense mutations (present in 65% of GRD affected individuals), protein truncating variants (PTVs) have also been detected in 23% of individuals with GRD (13). PTVs are defined as genetic variants (nonsense single-nucleotide variants, frameshift insertions or deletions, large structural variants and splice-disrupting SNVs) disrupting transcription and leading to a shortened or absent protein from the mutated allele. Mechanistically, PTVs are hypothesized to provoke a protein loss of function (decrease of protein amounts) though it could potentially also cause gain-of-function effects. Nevertheless, from our knowledge, the functional consequences of GRIN PTVs remain elusive. Here, we present the first study integrating genetic, clinical and functional data related with GRIN pathogenic and nonpathogenic PTVs. Our data sheds light on to gene- and domain specific features of GRIN PTVs, enabling their pathogenicity stratification. Importantly, the delineation of the functional and clinical outcomes associated with GRD PTVs provides genetic and preclinical insights for the definition of personalized therapies for individuals harbouring GRIN PTVs.

RESULTS

Heterozygous GRIN PTVs pathogenicity is gene-dependent

Truncating variants of GRIN genes are predicted to introduce a premature stop codon or to disrupt the open reading frame and trigger a premature stop codon. Compared with missense variants, previously reported GRIN PTVs likely pathogenic are relatively

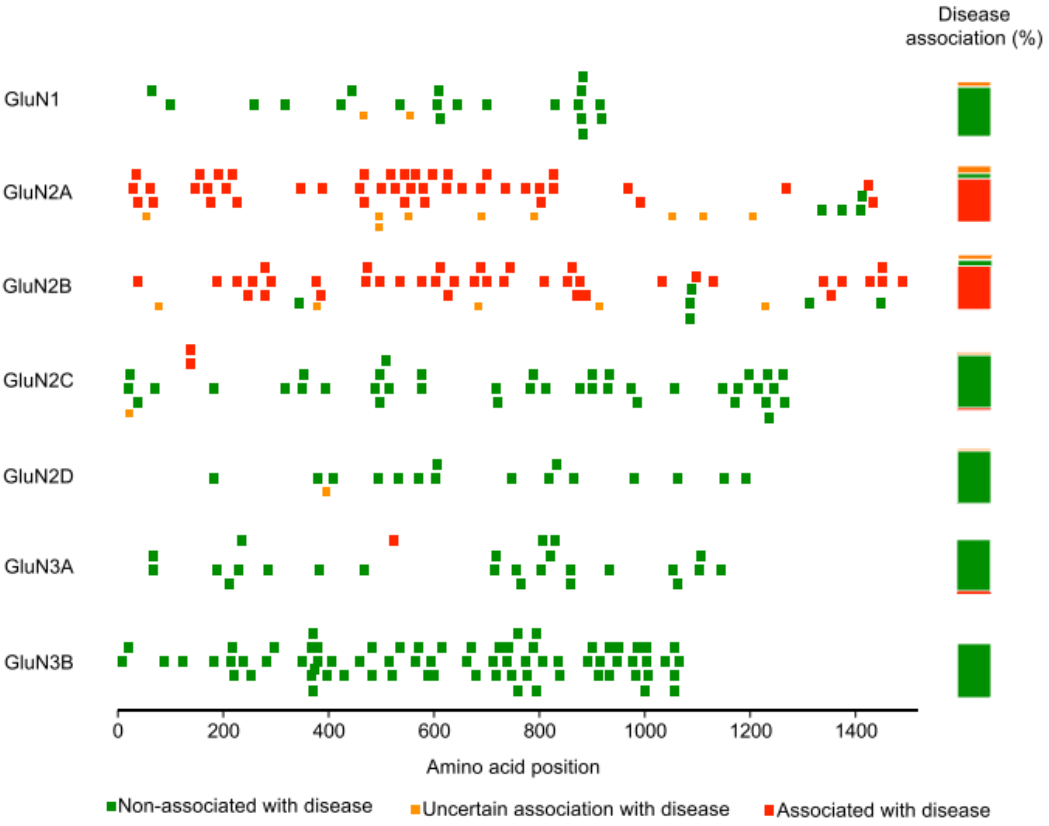
less abundant (59 out of 226, from (13)). In the present study, we compiled the genetic and clinical information of 10 clinical cases harbouring novel GRIN PTVs referred to our investigational network and proceeded to the functional annotation of representative GRIN PTVs. Remarkably, all pathogenic GRIN PTVs are clinically associated with mild-to-moderate ID and movement disorders (Fig. 4.3-9A), a relatively more homogeneous and milder phenotype compared with the heterogeneous and severe GRD phenotypes exhibited by individuals carrying pathogenic GRIN missense variants (7,8). Nevertheless, these initial observations were found in a reduced cohort of patients and were limited to PTVs affecting GRIN2A and GRIN2B gene products. In order to evaluate the GRIN-PTV genotype–phenotype association in a larger patients cohort and the potential impact on GRIN genes family, genetic databases were scanned and 293 unique GRIN PTVs (May 2020) and their related clinical symptoms were retrieved (Supplementary Material, Tables S1– S7). Genotype–phenotype analysis showed that GRIN PTVs are scattered along the seven members of the GRIN gene family. Disease-association analysis of GRIN PTVs showed that 63.1% are non-associated with neurological disorders (likely non pathogenic or with recessive inheritance pattern), whereas 30.7% are associated with neurological disorders (pathogenic or likely pathogenic) and 6.1% are of uncertain association with neurological symptoms (PTV conflicting description both in GRD individuals and in healthy population) (Fig. 4.3-9 and Supplementary Material, Tables S1–S7). Remarkably, genotype–phenotype analysis revealed a strong association of GRIN PTVs with neurological disorders in a gene-dependent manner (Fig. 4.3-9B and C). Gene-specific analysis showed that de novo heterozygous GRIN1 truncating variants (21) are invariably non-pathogenic. Noteworthy, the two only GRIN1 PTVs (p.R488Afs22Ter and biallelic GRIN1 (p.Q556Ter/Q556Ter) associated with neurological conditions were inherited, indicating a non-dominant inheritance pattern of these particular variants (Fig. 4.3-9, Supplementary Material, Table S1). Conversely, heterozygous GRIN2A and GRIN2B PTVs are mostly associated with neurological disorders (77.2 and 79.6% disease-association versus 7.0 and 11.1% non-disease-association, respectively: Fig. 4.3-9, Supplementary Material, Tables S2 and S3). The distribution of disease-associated variants along these genes revealed a high

vulnerability of truncations located at the amino terminal (ATD), agonist binding (ABD) and transmembrane (TMD) domains. In contrast, truncations affecting GluN2A and GluN2B carboxy terminal domains (CTDs) were heterogeneously associated with complex neurological disorders (e.g. schizophrenia or autism spectrum disorders (ASDs), for GRIN2A and GRIN2B, respectively). Despite the strong association between truncations affecting the non-CTD domains and neurological conditions, three GRIN2A (p.H595Wfs20Ter, p.H595Sfs60Ter and p.E58Ter) and a single GRIN2B (p.Q331Sfs5Ter) variants were reported in the gnomAD database. Functional annotation of these variants showed their deficiency to reach the cell surface, leading to a significant reduction of NMDAR-mediated currents (Supplementary Material, Fig. S1). Accordingly, these PTVs were conservatively annotated as variants of uncertain association with neurological conditions. Genetic analysis showed that de novo GRIN2C PTVs are not associated with clinical phenotypes (40 out of 41; Fig. 4.3-9C, Supplementary Material, Table S4), with the exception of GRIN2C (p.P132fs192Ter) detected in a subject with schizophrenia. Similarly, GRIN3A PTVs are neutral (25 de novo variants), with only one inherited GRIN3A variant (p.Q508Ter) associated with schizophrenia (Fig. 4.3-9B and C, Supplementary Material, Table S6). Truncations of GRIN2D and GRIN3B genes (16 and 74 variants, respectively) are invariably not associated with neurological disorders (Fig. 4.3-9B and C, Supplementary Material, Tables S5 and S7), with the exception of GRIN2D (c.1412G > A) transition potentially affecting GRIN2D splicing site and detected in an individual with schizophrenia (15). Overall, genetic association studies and variants distribution along NMDAR domains indicate a strong gene-dependent pathogenic effect of heterozygous GRIN truncations. In terms of functional annotation, these findings support the pathogenicity of heterozygous GRIN2A and GRIN2B truncations of the extracellular and transmembrane domains, while the pathogenicity of truncations affecting the CTD of GluN2A or GluN2B subunits is variable and not yet predictable. Moreover, heterozygous PTVs affecting GRIN1, GRIN2C-D and GRIN3A-B genes are almost non associated with disease, with the exception of discrete cases of schizophrenia.

A.

Clinical case	Gender	Gene	Variant	Genotype	Protein changes	Truncated domain(s)	Clinical symptoms
1	Male	GRIN2A	GRIN2A-E182fs22	c.544delG	p.E182Nfs22Ter	NTD-LBD-TMD-CTD	Mild intellectual disability. Focal epilepsy idiopathic pharmaco-resistant, evolving to ECSWS. Without hypotonia signs and absence of ASD traits. ADHD
2	Male	GRIN2A	GRIN2A-V452fs11	c.1354insT	p.V452Cfs11Ter	LBD-TMD-CTD	Mild intellectual disability, developmental delay, ECSWS, Psychomotor alterations, disturbance of learning, reading and writing skills
3	Female	GRIN2B	GRIN2B-R519ter	c.1555C>T	p.Arg519Ter	LBD-TMD-CTD	Mild intellectual disability. No epilepsy, no hypotonia, no ASD traits
4	Female	GRIN2B	GRIN2B-D786Mfs23	c.2355delA	p.D786Mfs23Ter	TMD-CTD	Mild intellectual disability, mild hypotonia, abnormal EEG during sleep without seizures, presence of ASD traits. Sleep and digestive problems
5	Female	GRIN2B	GRIN2B-E839ter	c.2515G>T	p.Glu839Ter	TMD-CTD	Mild intellectual disability, fine motor skills deficits, abnormal EEG, ECSWS, no ASD traits. Sleep problems, normal GI tract, early puberty, tall
6	Male	GRIN2B	GRIN2B-S34Qfs25	c.99insC	p.S34Qfs25Ter	NTD-LBD-TMD-CTD	Mild intellectual disability
7	Female	GRIN2B	GRIN2B-K670Vfs5	c.2010insA	p.K670Vfs5Ter	TMD-CTD	Mild intellectual disability, mild hypotonia
8	Female	GRIN2B	GRIN2B-het	p.12p13.1(13940689_14112006) del	Haploinsufficiency	GluN2B	Mild intellectual disability, expressive language deficits
9	Female	GRIN2B	GRIN2B-R847ter	c.2539-2540CG>TA	p.R847Ter	CTD	Mild intellectual disability. No hypotonia

B.



C.

	GRIN1	*GRIN2A*	*GRIN2B*	*GRIN2C*	*GRIN2D*	*GRIN3A*	*GRIN3B*	Total
Associated with disease	0 0,0%	44 77,2%	43 79,6%	2 4,7%	0 0,0%	1 3,8%	0 0,0%	90 30,7%
Uncertain association with disease	2 9,1%	9 15,8%	5 9,3%	1 2,3%	1 5,9%	0 0,0%	0 0,0%	18 6,1%
Non-associated with disease	20 90,9%	4 7,0%	6 11,1%	40 93,0%	16 94,1%	25 96,2%	74 100,0%	185 63,1%
Total	22	57	54	43	17	26	74	293

Figure 4.3-9. GRIN PTVs distribution and differential association with neurological disorders. (**A**) Heterozygous de novo GRIN2A and GRIN2B truncating variants in a European cohort of children with neurological disorders. GI, gastrointestinal tract; EEG, electroencephalogram. GluN subunit domains acronyms: ATD, amino terminal domain; LBD, ligand binding domain; TMD, transmembrane domain; CTD, carboxy terminal domain. (**B**) Scatter plot of PTVs of GRIN genes and their distribution along the amino acid sequence (x-axis). Right, Stacked bars representing the percentage of genetic variants (association, non-association, uncertain association with disease). Green squares: GRIN variants non-associated with disease; red squares: disease-associated GRIN variants; orange squares: uncertain genetic association and/or functional annotation. (**C**) Summary of GRIN truncating variants association with neurological conditions. The table compiles genotype–phenotype correlation of non-pathogenic, disease-associated and uncertain disease-association of GRIN truncating variants. GRIN PTVs (293) were collected from gnomAD, ClinVar, BCN-GRIN, GRIN-Leipzig and CFERV databases

Clinical phenotypes associated with pathogenic GRIN2A and GRIN2B PTVs

Individuals with GRDs display a spectrum of clinical symptoms not yet completely delineated. In order to establish the genotype–phenotype correlation of pathogenic GRIN PTVs, the clinical annotations for the different variants were manually curated and integrated. The analysis of the clinical symptoms exhibited by individuals harbouring PTVs of GRIN2A and GRIN2B revealed striking clinical hallmarks. Importantly, all pathogenic GRIN2A PTVs showed association with different types of epilepsy syndromes, ranging from mild focal epilepsy (benign epilepsy with centro-temporal spikes, BECTS) to epileptic encephalopathies (continuous spike-and-wave during slow wave sleep, ECSWS; Landau–Kleffner syndrome, LKS), with variable comorbidity with ID (57% of GRIN2A PTVs cases; Fig. 4.3-10). In contrast, while seizures are almost absent in subjects harbouring GRIN2B PTVs, genotype–phenotype analysis revealed a higher presence of ID and ASDs traits (85 and 40%, respectively; Fig. 4.3-10). Interestingly, GRIN2A and GRIN2B pathogenic truncations are associated with mild-to-moderate forms of ID, in comparison to disease-associated GRIN2A and GRIN2B missense variants that are associated with more severe ID (7,8).

A.

	GRIN2A PTVs	*GRIN2B* PTVs
Clinically Reported *	38	29
Developmental delay	20	18
Intellectual disability (ID)	25	27
Mild ID	10	5
Moderate ID	6	6
Severe ID	1	3
ID (% clinically reported cases)	65,8%	93,1%
Epilepsy	21	3
CSWS	5	1
BECT	3	0
LKS	1	0
Focal epilepsy	8	3
Epilepsy events	37	8
EPI (% clinically reported cases)	97,4%	27,6%
Hypotonia	1	2
Speech Disorder	15	0
Speech Disorder (% clinically reported cases)	39,5%	0
ASD traits	0	11
ASD (% clinically reported cases)	0,0%	37,9%
ADHD	1	1
SCZ	0	1

B.

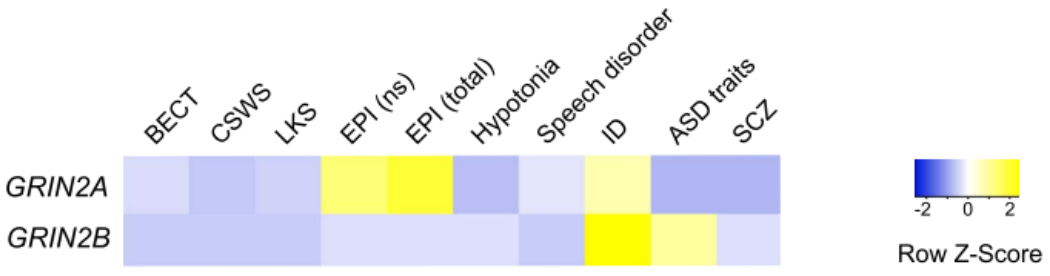


Figure 4.3-10. Clinical phenotypes associated with disease-associated GRIN2A and GRIN2B PTVs. (**A**) Clinical symptoms of individuals harbouring pathogenic GRIN2A or GRIN2B PTVs, represented as the percentage of individuals over the total number of reported clinical cases. (**B**) Phenotype-based heat maps displaying the relative abundance of clinical symptoms in clinically reported individuals harbouring GRIN2A and GRIN2B PTVs. EPI(ns), epilepsy without detailed epileptic symptoms description; EPI (total), individuals with epilepsy symptoms.

Heterozygous GRIN1 truncations do not alter NMDAR-mediated currents

Gene truncations provoke a shortening of the open reading frame, which in turn results in the lack of the corresponding distal domains. In the context of GRIN genes, truncations might affect different aspects of the NMDAR biogenesis and/or channel properties, although the pathophysiological mechanisms remain elusive. In order to determine the molecular and biophysical mechanisms underlying GRIN gene-dependent vulnerability to truncations, paradigmatic GRIN PTV were assessed along cell-based assays. First, we explored the gene dosage effect in GRIN1 PTVs, by transiently expressing different ratios of GRIN1 wild-type versus truncated constructs in cellular models. Electrophysiological experiments were conducted in HEK293T cells transiently expressing wild-type GRIN2A and GRIN1 (wild-type and truncating variants) at different stoichiometric conditions. Electrophysiological studies showed that GRIN1 truncations do not affect NMDAR-mediated current density in cells co-transfected with HA-Grin1 PTVs and wild-type GFP-Grin2A constructs (P-value > 0.05 for all comparisons; Fig. 4.3-11A and B). In contrast, biallelic GRIN1(p.Q556Ter) truncation resulted in the absence of measurable NMDAR-mediated currents in cells expressing GluN2Awt-GluN1(Q556Ter)2 (<0.1 pA/pF NMDARmediated currents, N = 10 patched cells; data not shown). Further analysis showed that heterozygous GRIN1 truncations do not alter channel properties (e.g. rise-time, desensitization and deactivation rates) of GluN2Awt-GluN1(wt/trunc) receptors (Fig. 4.3-11C–E). These data suggest that either GluN1 truncated subunits are not assembled into functional NMDARs—and do not affect NMDAR cell surface density—or, alternatively, they form functional NMDARs without interfering channel properties. In order to address these hypotheses, immunofluorescence analysis of NMDAR subunits was conducted. Transient co-transfection of wild-type GFP-GRIN2B and equimolar DNA amounts of HAGRIN1 (wt:PTV) showed no changes on GFP-GluN2B (NMDARs) surface

expression in COS-7 cells ($P > 0.05$; Fig. 4.3-11F). In contrast, transfection mimicking biallelic GRIN1(p.Q556Ter) truncation condition resulted in a complete loss of NMDAR surface expression ($P < 0.0001$; Fig. 4.3-11F), in agreement with previous data showing negligible NMDAR-mediated currents in HEK293T cells. The lack of NMDARs surface expression was also observed in cells transiently expressing other biallelic GRIN1 truncations ($P < 0.0001$; Supplementary Material, Fig. S2). Furthermore, a more detailed analysis of the different NMDAR subunits surface expression was conducted, based on the detection of differentially tagged-GluN subunits resulting from the expression of wild-type GFP-GRIN1 (GFP in the amino terminal domain, (14)), truncated HA-Grin1 and Flag-Grin2b constructs. Immunofluorescence analysis using anti-GFP antibody in non-permeabilized cells showed that GFP-GluN1 wild-type surface expression remains unaltered by the expression of truncated HA-GluN1 subunits ($P > 0.05$; Fig. 4.3-11G). In contrast, non-permeabilized cells incubated with anti-HA showed a lack of truncated HA-GluN1 subunit surface expression ($P < 0.0001$; Fig. 4.3-11H). Overall, these data suggest that truncated GluN1 subunits are unable to assemble and, consequently, do not reach the plasma membrane. Together with this observation, our findings showing that heterozygous GRIN1 truncations do not disturb overall NMDARs surface expression and NMDAR mediated currents indicate that GRIN1 monoallelic truncation is not limiting for NMDAR biogenesis in heterologous expression systems. These data are in agreement with the abovementioned genotype–phenotype correlations, showing no association between heterozygous GRIN1 truncations and neurodevelopmental conditions.

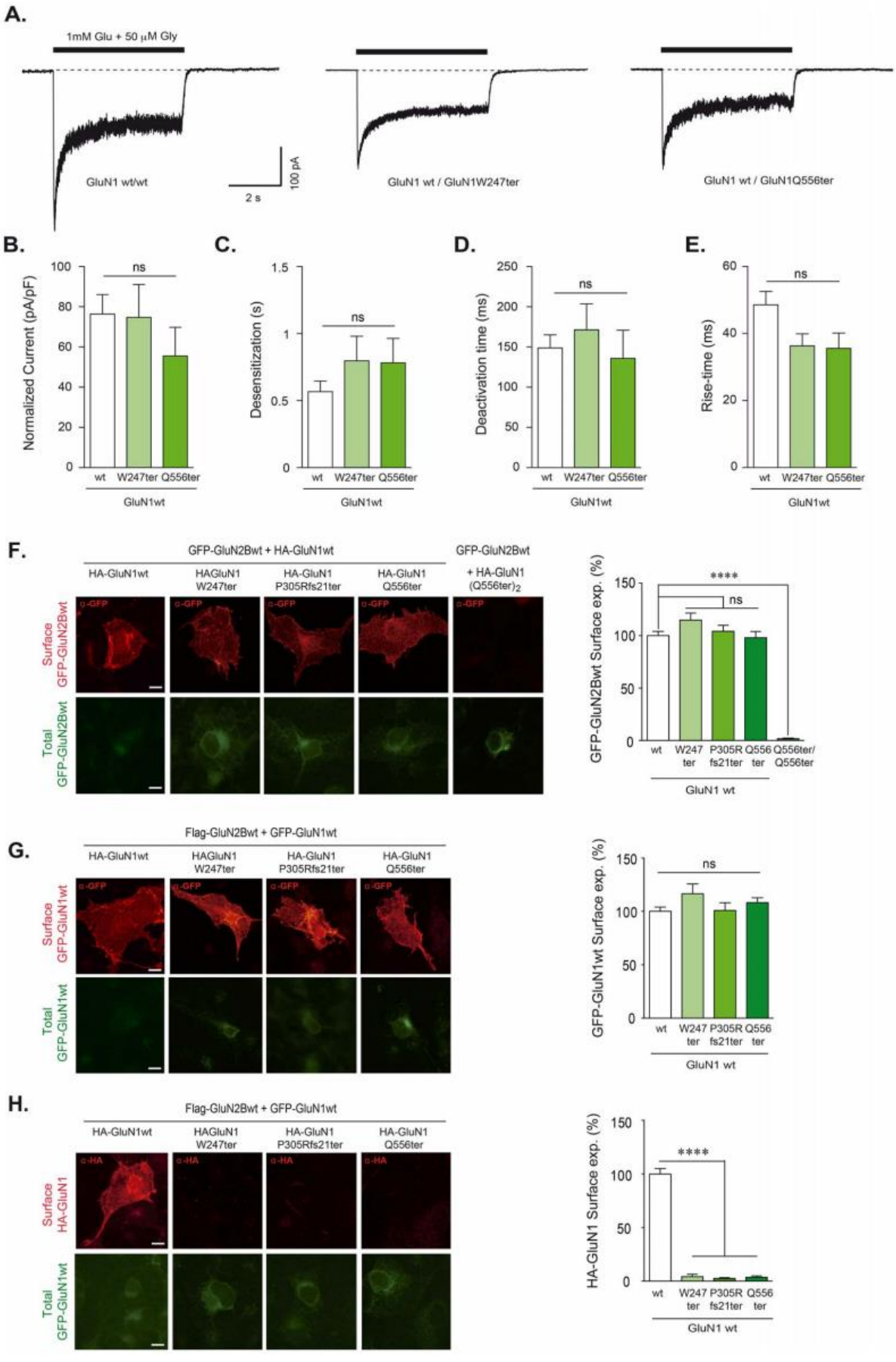

A.
1mM Glu + 50 μM Gly
GluN1 wt/wt
100 pA
2 s
GluN1 wt / GluN1W247ter
GluN1 wt / GluN1Q556ter
B.
Normalized Current (pA/pF)
ns
wt W247ter Q556ter
GluN1wt
C.
Desensitization (s)
ns
wt W247ter Q556ter
GluN1wt
D.
Deactivation time (ms)
ns
wt W247ter Q556ter
GluN1wt
E.
Rise-time (ms)
ns
wt W247ter Q556ter
GluN1wt
F.
GFP-GluN2Bwt + HA-GluN1wt
GFP-GluN2Bwt + HA-GluN1 (Q556ter)2
HA-GluN1wt
HAGluN1 W247ter
HA-GluN1 P305Rfs21ter
HA-GluN1 Q556ter
α-GFP
Surface GFP-GluN2Bwt
Total GFP-GluN2Bwt
GFP-GluN2Bwt Surface exp. (%)
****
ns
wt W247 ter P305R fs21ter Q556 ter Q556ter/ Q556ter
GluN1 wt
G.
Flag-GluN2Bwt + GFP-GluN1wt
HA-GluN1wt
HAGluN1 W247ter
HA-GluN1 P305Rfs21ter
HA-GluN1 Q556ter
α-GFP
Surface GFP-GluN1wt
Total GFP-GluN1wt
GFP-GluN1wt Surface exp. (%)
ns
wt W247 ter P305R fs21ter Q556 ter
GluN1 wt
H.
Flag-GluN2Bwt + GFP-GluN1wt
HA-GluN1wt
HAGluN1 W247ter
HA-GluN1 P305Rfs21ter
HA-GluN1 Q556ter
α-HA
Surface HA-GluN1
Total GFP-GluN1wt
HA-GluN1 Surface exp. (%)
****
wt W247 ter P305R fs21ter Q556 ter
GluN1 wt

Figure 4.3-11. Functional annotation of GRIN1 PTVs. (**A**) Representative whole-cell currents evoked by rapid application of 1 mM glutamate plus 50 μM glycine (0.5- second duration; −60 mV) in HEK-293 T cells expressing (GluN2Awt)2-(GluN1wt)2 (left trace), (GluN2A)2-GluN1wt/GluN1-PTVs (center and right traces). (**B**) Bar graph representing normalized peak currents of cells transfected with Grin2a together with equimolar amounts of Grin1 constructs (wt/mock/PTVs). (**C–E**) Bar graphs representing mean ± SEM of rise-time (panel **C**), desensitization (panel D) and deactivation (panel E) rates of NMDAR-mediated currents in HEK293 cells expressing (GluN2A)2-(GluN1wt)2, (GluN2A)2-(GluN1wt/GluN1-PTVs). ns, P-value > 0.05; one-way ANOVA with Dunnett's multiple comparison test or Kruskal–Wallis with Dunn's multiple comparison test. (**F–H**) Left, Immunofluorescence analysis of NMDAR surface expression in COS-7 cell line co-transfected with GFP-Grin2b and HA-Grin1 (wild-type and/or truncated) constructs (panel **F**) or Flag-Grin2b, GFP-Grin1 and HA-Grin1 (wild-type and/or truncated) constructs (panels **G, H**); scale bar: 10 μm; Right, Bar graphs representing the mean ± SEM of cell surface expression of NMDAR (N = 16–38 cells per condition, from 3 to 4 independent experiments; ns, P-value > 0.05; ∗∗∗∗, P-value < 0.0001, one-way ANOVA with Bonferroni's post hoc test).

Disease-associated GRIN2A PTVs functionally recapitulate GRIN2A haploinsufficiency

Genotype–phenotype studies indicate a strong relationship between truncating GRIN2A variants and NDDs. In order to shed light into the genetic dominant effect of GRIN2A PTVs, their potential pathophysiological impact was evaluated using cell-based assays. Electrophysiological experiments were conducted in HEK293T cells transiently expressing different Grin2a wt:PTV stoichiometric ratios. Normalized current density analysis showed that disease-associated GRIN2A PTVs drastically reduce NMDAR-mediated peak current (P-value < 0.001 for all comparisons; Fig. 4.3-12A and B) without disturbing NMDAR activation, desensitization or deactivation rates (P > 0.05 for all comparisons) in cells expressing GluN1(wt)2-(GluN2Awt/PTV), recapitulating GRIN2A haploinsufficiency condition (Fig. 4.3-12C–E). These data suggested the absence of GRIN2A PTVs dominant negative effect on the NMDARs wild-type pool, and further cell surface expression experiments were conducted to evaluate this hypothesis. Immunofluorescence analysis of COS-7 cells transiently co-transfected with HA-Grin1 and equimolar DNA ratios of GFP-tagged GRIN2A wild-type and Grin2a truncated constructs, showed a significant decrease of total HA-GluN1 surface expression (P < 0.0001; Fig. 4.3-12F). These data indicate that heterozygous

expression of GRIN2A-PTV, together with one GRIN2A wild-type allele, significantly decreases the surface expression of total NMDARs. Furthermore, in order to specifically determine GluN2A wild-type and GluN2A truncated subunits surface expression, COS-7 cells were co-transfected with GRIN1, GFP-GRIN2Awt and HA-GRIN2A wt/PTV (E182Nfs23ter or V452Cfs11ter) constructs. Immunofluorescence analysis showed that GRIN2A PTV allele resulted in a net reduction of surface expression of wild-type GluN2A subunit ($P < 0.0001$; Fig. 4.3-12G), and no surface expression was detected for the truncated GluN2A subunits ($P < 0.0001$; Fig. 4.3-12H). Interestingly, the lack of surface expression was confirmed in COS-7 co-transfected with GRIN1 and biallelic GRIN2A truncations affecting the NTD, LBD and TMD domains, while those affecting the CTD exhibited a normal surface trafficking (Supplementary Material, Fig. S2). Overall, these findings support the view that GRIN2A PTVs of the non-CTD functionally recapitulate GRIN2A haploinsufficiency condition, rather than exerting a dominant negative effect on wild-type GRIN2A allele.

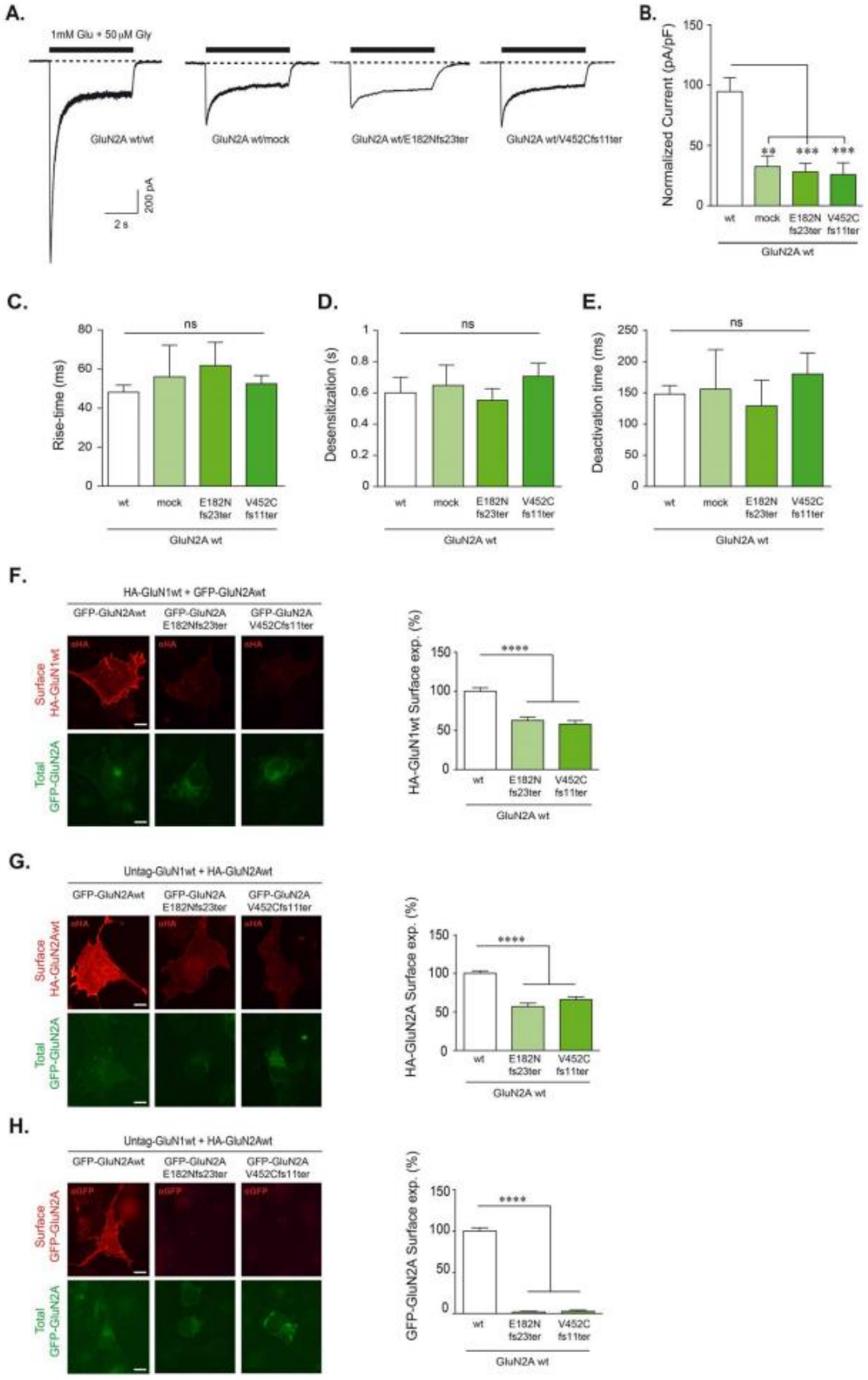

A.
1mM Glu + 50 µM Gly
GluN2A wt/wt
GluN2A wt/mock
GluN2A wt/E182Nfs23ter
GluN2A wt/V452Cfs11ter
200 pA
2 s
B.
Normalized Current (pA/pF)
150
100
50
0
**
***
***
wt
mock
E182N fs23ter
V452C fs11ter
GluN2A wt
C.
Rise-time (ms)
80
60
40
20
0
ns
wt
mock
E182N fs23ter
V452C fs11ter
GluN2A wt
D.
Desensitization (s)
1
0.8
0.6
0.4
0.2
0
ns
wt
mock
E182N fs23ter
V452C fs11ter
GluN2A wt
E.
Deactivation time (ms)
250
200
150
100
50
0
ns
wt
mock
E182N fs23ter
V452C fs11ter
GluN2A wt
F.
HA-GluN1wt + GFP-GluN2Awt
GFP-GluN2Awt
GFP-GluN2A E182Nfs23ter
GFP-GluN2A V452Cfs11ter
Surface HA-GluN1wt
Total GFP-GluN2A
αHA
HA-GluN1wt Surface exp. (%)
150
100
50
0
****
wt
E182N fs23ter
V452C fs11ter
GluN2A wt
G.
Untag-GluN1wt + HA-GluN2Awt
GFP-GluN2Awt
GFP-GluN2A E182Nfs23ter
GFP-GluN2A V452Cfs11ter
Surface HA-GluN2Awt
Total GFP-GluN2A
αHA
HA-GluN2A Surface exp. (%)
150
100
50
0
****
wt
E182N fs23ter
V452C fs11ter
GluN2A wt
H.
Untag-GluN1wt + HA-GluN2Awt
GFP-GluN2Awt
GFP-GluN2A E182Nfs23ter
GFP-GluN2A V452Cfs11ter
Surface GFP-GluN2A
Total GFP-GluN2A
αGFP
GFP-GluN2A Surface exp. (%)
150
100
50
0
****
wt
E182N fs23ter
V452C fs11ter
GluN2A wt

Figure 4.3-12. Functional annotation of representative de novo GRIN2A PTVs associated with neurological conditions. (**A**) Representative whole-cell currents evoked by rapid application of 1 mM glutamate plus 50 μM glycine (0.5-second duration; −60 mV) in HEK-293 T cells expressing (GluN1)2-(GluN2Awt)2 (left trace), (GluN1)2-GluN2Awt (hemizygous, 'mock' condition), (GluN1)2-GluN2Awt/PTVs. (**B**) Bar graph representing mean ± SEM normalized peak currents of cells transfected with Grin1 together with different stoichiometric amounts of Grin2a constructs (wt/mock/PTVs). (**C–E**) Left,Bar graphs representing mean ± SEM of rise-time (panel **C**), desensitization (panel **D**) and deactivation (panel **E**) rates of NMDAR-mediated currents in HEK293 cells expressing (GluN1)2-(GluN2Awt)2, (GluN1)2-GluN2Awt/− and (GluN1)2-GluN2Awt/PTVs. ns, P-value > 0.05; ∗∗, P-value < 0.01; ∗∗∗, P-value < 0.001; ∗∗∗∗, P-value < 0.0001; one-way ANOVA with Dunnett's multiple comparison test or Kruskal–Wallis with Dunn's multiple comparison test. (**F–H**) Immunofluorescence analysis of NMDAR surface expression in COS-7 cell line co-transfected with HA-Grin1 and GFP-Grin2a (wild-type and/or truncated) constructs (panel **F**) or Untag-Grin1, HA-Grin2a and GFP-Grin2a (wild-type and/or truncated) constructs (panels **G, H**); scale bar: 10 μm; Right, Bar graph representing the mean ± SEM of cell surface expression of NMDAR (N = 21–66 cells per condition, from 3 to 4 independent experiments; ns, P-value > 0.05; ∗∗∗∗, P-value < 0.0001; one-way ANOVA with Bonferroni's post hoc test).

Disease-associated GRIN2B PTVs functionally recapitulate GRIN2B haploinsufficiency

Despite the clustering of clinical symptoms with GRIN2A and GRIN2B PTVs (Fig. 4.3-10), we hypothesized that heterozygous GRIN2B PTVs pathophysiologic mechanisms would be similar to GRIN2A PTVs. Hence, cell-based functional annotation of representative GRIN2B truncations associated with GRD was conducted. Electrophysiological experiments were performed in HEK293T cells transiently expressing the combinations of GRIN1 and GRIN2B constructs (wild-type, R519Ter, D786Mfs23Ter, E839Ter) at different stoichiometric conditions (wt/wt versus wt/trunc). As for GRIN2A PTVs, disease-associated GRIN2B PTVs showed a significant reduction of NMDAR-mediated peak currents in whole cell patch clamp experiments ($P < 0.001$; Fig. 4.3-13A and B) without disturbing NMDAR gating parameters (e.g. rise time, activation and deactivation rates; Fig. 4.3-13C–E). Immunofluorescence studies were performed in COS-7 cells transiently co-transfected with HA-GRIN1wt and equimolar GFP-GRIN2Bwt and GFP-GRIN2Bwt/PTV ratios. Immunolabelling of surface HA-GluN1 subunit showed a significant decrease of total NMDARs surface

expression ($P < 0.0001$; Fig. 4.3-13E). In order to specifically evaluate potential alterations on GluN2B PTVs surface expression, HA-GRIN1, flag-GRIN2Bwt and GFPGRIN2Bwt/PTV (R519Ter, D786Mfs23Ter or E839Ter) were cotransfected in COS-7 cells. Immunofluorescence analysis showed an almost negligible detection of truncated GluN2B subunits surface expression ($P < 0.0001$; Fig. 4.3-13F), indicating GluN2B PTVs deficiency to assemble and/or to reach the plasma membrane. Interestingly, the lack of surface expression was confirmed in COS-7 cells co-transfected with Grin1 and biallelic Grin2b truncations affecting the NTD, LBD and TMD domains, while those affecting the CTD (non-related with neurological disorders) exhibited a normal surface trafficking (Supplementary Material, Fig. S2). Previous reports have described the modulatory effect of the GluN subunits domain CTD for a proper trafficking and retention of NMDARs in cell lines (16,17). In order to evaluate whether GRIN2B-CTD truncation might have an impact on surface density of the NMDAR in a neuronal context, disease-associated GRIN2B PTVs disrupting the TMD4 (E839Ter) and the proximal part of the CTD (R847Ter) were generated and expressed either in COS-7 cell line (together with GluN1 subunits) or in primary mouse neuronal cultures. Surface expression analysis of transfected GRIN2B PTVs revealed that, despite the CTD of GluN2B subunit, it is not required for surface expression of the NMDAR in COS-7 cells ($P < 0.0001$; Supplementary Material, Fig. S3A), truncation of the CTD results in a significant reduction of GluN2B(R847Ter) ($P < 0.001$ or < 0.0001 for neurons transfected with Grin2b versus Grin2b-E839Ter or Grin2b-R847Ter, respectively; Supplementary Material, Fig. S3B). Overall, these results support an haploinsufficiency-like effect of GRD-associated GRIN2B PTVs, rather than form a dominant negative effect on Grin2B wild-type alleles expressed in primary neurons.

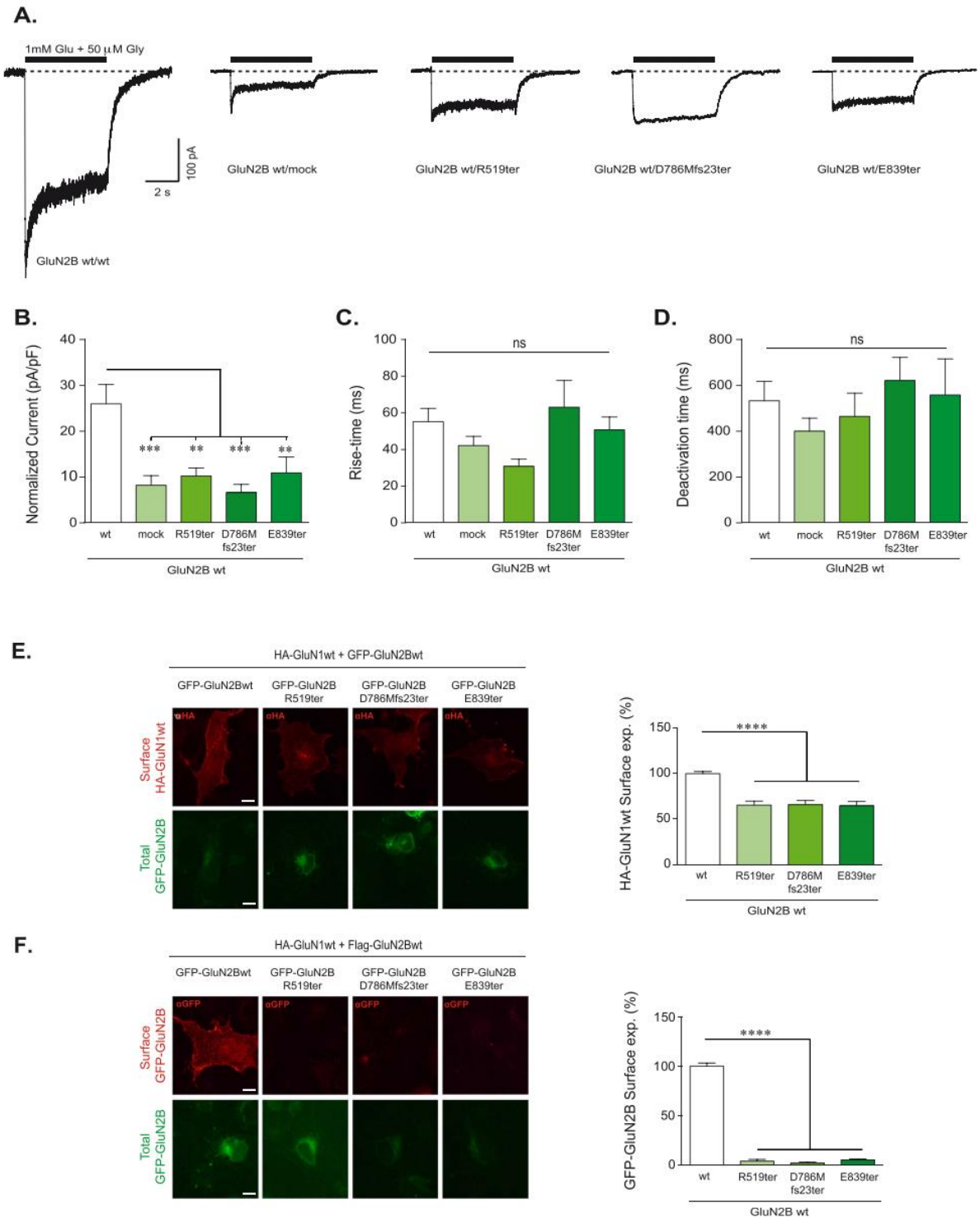


Figure 4.3-13. Functional annotation of representative de novo GRIN2B PTVs associated with neurological conditions. (**A**) Representative whole-cell currents evoked by rapid application of 1 mM glutamate plus 50 µM glycine (0.5-second duration; −60 mV) in HEK-293 T cells expressing (GluN1)2-(GluN2Bwt)2 (left trace), (GluN1)2-GluN2Bwt (hemizygous, 'mock' condition), (GluN1)2-

GluN2Bwt/PTVs. (**B**) Bar graph representing mean ± SEM normalized peak currents of cells transfected with Grin1 together with different equimolar amounts of Grin2b constructs (wt/mock/PTVs). (**C and D**) Bar graphs representing mean ± SEM of rise-time (panel **C**) and deactivation (panel **D**) rate of NMDAR-mediated currents in HEK293 cells expressing (GluN1)2-(GluN2Bwt)2, (GluN1)2-GluN2Bwt/− and (GluN1)2-GluN2Bwt/PTVs. ns, P-value > 0.05; ∗∗, Pvalue < 0.01; ∗∗∗, P-value < 0.001; one-way ANOVA with Dunnett's multiple comparison test or Kruskal–Wallis with Dunn's multiple comparison test. (**E and F**) Left, Immunofluorescence analysis of NMDAR surface expression in COS-7 cell line co-transfected with HA-Grin1 and GFP-Grin2b (wild-type and/or truncated) constructs (panel **E**) or HA-Grin1, Flag-Grin2b and GFP-Grin2b (wild-type and/or truncated) constructs (panel **F**); scale bar: 10 μm; Right, Bar graphs representing the mean ± SEM of cell surface expression of NMDAR (N = 20–86 cells per condition, from 3 to 4 independent experiments; P-value > 0.05; ∗∗∗∗, P-value < 0.0001, ns; one-way ANOVA with Bonferroni's post hoc test).

DISCUSSION

In the present study, we performed a comprehensive genotype–phenotype analysis of GRIN PTVs, delineating GRIN PTV pathogenicity scenario. Our findings indicate that GRIN PTVs' pathogenicity is both subunit- and domain-specific, with clinical outcomes dictated by the identity of the truncated GRIN gene. Brief ly, while heterozygous GRIN1, GRIN2C, GRIN2D, GRIN3A and GRIN3B truncations are not associated with autosomal dominant neurodevelopmental conditions, non-CTD PTVs of GRIN2A and GRIN2B genes are invariably related with NDDs, mostly associated with ID, epilepsy (GRIN2A PTVs) and/or autism traits (GRIN2B PTVs). Mechanistically, our data indicate that disease associated with GRIN2A and GRIN2B PTVs result in a reduction of NMDARs pool surface density, which invariably causes an overall loss of function. In general, PTVs lead to a nonsense-mediated decay, a cellular mechanism resulting in loss of function of the truncated protein. Nevertheless, our data showed that with the exception of two non-dominant GRIN1 variants (c.1342del/p.R448fs22Ter allele from maternal origin, uncertain pathogenicity; p.Q556Ter biallelic variant, reported by Lemke's group (9)), all reported heterozygous GRIN1 PTVs (20 variants) are not associated with neurological conditions. Theoretically, these findings would indicate that either GRIN1 gene expression is not critical or, alternatively, wild-type GRIN1 allele would be sufficient to overcome GRIN1 PTV. GluN1 subunit is invariably present in the NMDAR, which plays a critical role in excitatory glutamatergic neurotransmission

(1). Accordingly, GRIN1 gene product is widely considered a key element in excitatory neurotransmission and brain function. Therefore, the lack of association between GRIN1 functional haploinsufficiency and clinical outcomes might be attributed to wild-type GRIN1 allele ability to guarantee physiological NMDAR surface density. In line with this, our data showed that representative GRIN1 PTVs do neither affect NMDAR surface expression nor NMDAR-mediated currents. These findings are in agreement with elegant biochemistry experiments showing GluN1 subunit as a non-limiting subunit for NMDAR biogenesis, whereas GluN2 subunits play a limiting rate (18). Furthermore, Grin1+/− heterozygous mice do not display phenotypic abnormalities, indicating that Grin1 haploinsufficiency—mimicking Grin1 heterozygous PTVs—is non-pathogenic (19). On the contrary, GRIN1 missense mutations are strongly linked with NDDs (9,13). This might be due to a dominant negative effect of GRIN1 missense mutations that result in the incorporation of aberrant amino acids in the GluN1 subunits and abnormal functioning of the mutant GluN1 subunit-containing NMDARs. Genotype–phenotype correlation studies of GRIN2A and GRIN2B truncations showed similar pathogenicity patterns. Indeed, without considering non-consistent variants, an almost complete pathogenicity versus non-pathogenicity association was detected for GRIN2A (37 versus 0 cases, for GluN2A truncations upstream the non-CTD) and GRIN2B (24 versus 1 cases for GluN2B truncations upstream the non-CTD) genes. Interestingly, GRIN2A/B PTVs pathogenicity was variable in truncations affecting the CTD, similarly to GRIN2A and GRIN2B missense mutations (7,8) whose pathogenicity is concentrated at the ABD and TMD, while the CTD is less vulnerable to genetic variation. Mechanistically, our findings strongly suggest that GRIN2A- and GRIN2B-related pathophysiological alterations might result from similar dysfunctions. Indeed, surface expression studies showed that non-CTD GRIN2A and GRIN2B PTVs trigger a decrease of NMDARs surface density, probably due to a defect on subunits assembly and/or a reduction of trafficking to the cell surface. In agreement with this hypothesis, immunofluorescence data showed a strong perinuclear immune signal (density) of GluN2A or GluN2B PTVs, suggesting the inability of GluN2A and GluN2B truncating variants to exit the endoplasmic reticulum. GluN subunits retention in the ER results from the presence of a retention mechanism that has been proposed as a protein

quality control for the NMDAR (20). PTVs affecting the CTD of GluN2A or GluN2B subunits were found to be heterogeneously associated with NDDs. Functionally, the CTDs of GluN2A and GluN2B subunits are involved in receptor trafficking, docking to plasma membrane and coupling receptors to signalling cascades (3,21,22). Theoretically, GluN2A and GluN2B subunits CTD truncations would neither interfere with GluN subunits assembly nor with NMDAR trafficking to the plasma membrane. Nevertheless, the lack of the PDZ-binding domain (in CTD-truncated GluN2 subunits) would weaken mutant NMDARs-PSD scaffolding proteins interaction, as well as perturbing downstream signalling cascades. In this interactome scenario, we may hypothesize that gene variants affecting NMDAR-interacting proteins (structural) or proteins functionally regulating NMDAR-mediated signalling cascades (upstream and/or downstream) can have synergic/antagonistic effects, dictating the clinical outcome. Further whole exome analysis of gene variants in individuals harbouring GRIN2A and GRIN2B CTD truncations will contribute to identify potential protective and/or risk genes. While pathogenic GRIN missense variants lead to a continuum of functional outcomes, roughly stratified into gain of-function, loss-of-function and complex alterations, our study strongly support a pathophysiological scenario in which GRIN PTVs would be invariably associated with NMDAR hypofunctionality. Clinically, individuals harbouring GRIN2A and GRIN2B PTVs exhibit a global developmental delay (DD) and mild-to-moderate ID. In the context of GRD clinical spectrum, these mild cognitive and motor phenotypes might be explained by the reduction of GRIN2A and GRIN2B PTVs surface expression and NMDAR-mediated currents, without wild-type NMDARs interference. The reduction detected on current density was similar to the haploinsufficiency condition and roughly corresponded to a 50% reduction of current density. This percentage would be in agreement with the observed lack of surface delivery of GRIN2A truncated alleles products, according to an expected Mendelian inheritance pattern. This functional stratification (e.g. pathogenic GRIN2A and GRIN2B PTVs trigger NMDAR hypofunctionality) is critical for the selection of GRD personalized therapies. Despite no rescue experiments have been performed, we can hypothesize that GRIN2A and GRIN2B PTV could be attenuated by potentiating the NMDAR pool at the cell surface, by means of positive allosteric modulator(s) (PAMs)

administration and/or L-serine nutraceutical supplementation, respectively (23,24). In summary, the present study comprehensively delineates genetic and clinical data associated with GRIN PTVS. Data analysis and functional studies allowed to functional stratify GRIN PTVs, showing that disease-associated GRIN2A and GRIN2B PTVs result in an haploinsufficiency-like condition potentially treatable with NMDAR potentiators (e.g. nutraceutical L-serine supplementation, PAMs). Overall, these findings contribute to delineate GRDs and unequivocally accelerate GRIN variants functional annotation and decision-making of personalized therapeutic interventions.

Materials and Methods

Subjects

Genetic report and clinical information from individuals harbouring GRIN PTVs were provided by referring physicians. All legal guardians provided informed written consent for genetic testing in accordance with the respective national ethics guidelines and with approval of the local ethics committees in the participating study centers. Database of GRIN PTVs and clinical phenotypes GRIN PTVs were compiled from ClinVar (www.ncbi.nlm.ni h.gov/clinvar), gnomAD (gnomad.broadinstitute.org), GRIN variants databases (lmc.uab.es/grindb/truncations, www.grin-database.de, http://functionalvariants.emory.edu/database/i ndex.html) and Pubmed. Nucleotide change predicted protein sequence, clinical annotation and disease-association were automatically retrieved and integrated under unique entries. Pathogenic and likely pathogenic mutations were classified as disease-associated, while gnomAD variants were classified as non-pathogenic. Furthermore, a conservative annotation was applied and GRIN variants inconsistently annotated as pathogenic/non-pathogenic, as well as disease-associated GRIN variants with no clinical phenotypes described, were classified as GRIN variants of ‘Uncertain Pathogenesis’. Data are compiled in Supplementary Material, Tables S1–S7, and the updated repository accessible at lmc.uab.es/grindb/truncations.

Mice

Adult male and female CD-1 mice (IDIBELL Animal Facilities) were used. The IDIBELL Committee on Animal Use and Care and Autonomous Government (Generalitat de Catalunya) approved the protocol (number: 9108). Following the approved experimental protocol, all animals were supervised daily. Animals were housed and tested in compliance with the guidelines provided by the Guide for the Care and Use of Laboratory Animals and following the European Union guidelines (2010/63/EU).Mice were housed in groups of four in standard cages with ad libitum access to food and water and maintained under a 12 h dark/light cycle (starting at 7:30 a.m.), 22◦C temperature and 66% humidity (standard conditions).

Plasmids

The expression plasmids for rat GluN1, GFP-GluN2A and GFPGluN2B were kindly provided by Dr Vicini (14). Dr Nakanishi provided HA-GluN1 and HA-GluN2A plasmids (25,26), Dr SanzClemente the Flag-GluN2B vector and GFP-GluN1 plasmid was obtained from Dr. Barria (27). Nucleotide changes for the production of GRIN variants were achieved by oligonucleotide directed mutagenesis, using the QuikChange II XL site directed mutagenesis kit according to the manufacturer's instructions (Stratagene, La Jolla, CA, USA), and verified by Sanger sequencing. All GRIN constructs were generated using rat GRIN cDNA as template, providing a perfect match for all the truncating variants except three human GRIN2A variants (E182Nfs22Ter, H595Wfs20Ter and H595Sfs60Ter) that showed mild changes in the corresponding constructs generated from rat cDNA templates (Grin2a-E182Nfs23Ter, Grin2a-H595Rfs28Ter, Grin2a-H595Lfs60Ter and respectively).

Cell culture and transfection

HEK-293 T and COS-7 cell lines were obtained from the American Type Culture Collection and maintained at 37◦C in Dulbecco's modified Eagle's medium, supplemented with 10% fetal calf serum and antibiotics (100 units/ml penicillin and 100 mg/ml streptomycin). Furthermore, D-2-amino-5-phosphonopentanoic acid (D-AP5, Abcam, Cambridge, UK) was added to the medium (0.5-1 mM final concentrations, for

HEK-293 T and COS-7 cells, respectively) to avoid excitotoxicity. Transient transfection of HEK-293 T cells was achieved by the calcium phosphate method (Clontech, France) and cell extracts were obtained 48 hours after transfection. COS-7 cells were transfected with Lipofectamine™ 2000 (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions, and cells were fixed 24 hours posttransfection, for further immunofluorescence analysis. HEK293 T were transfected with equimolar amounts of plasmids encoding for GluN1 and GluN2A/GluN2B subunits, whereas COS7 cells were transfected with different combinations of tagged subunits, upon the considered allelic (heterozygous or biallelic) conditions. To analyze biallelic expression of GRIN PTV, cells were transfected with equimolar amounts of HA-GluN1 and GFPGluN2A/GluN2B subunits (1:1), while heterozygous condition was achieved by transfecting equimolar amounts of the wildtype and the truncated variants of the GRIN gene of interest (1:0.5:0.5). Expression experiments in dissociated mouse hippocampal neuron cultures were performed as previously described (24). Brief ly, mouse embryos (embryonic day E18) were obtained from pregnant CD1 females; their hippocampi were isolated and maintained in cold Hank's Balanced Salt Solution (HBSS, Gibco, Carlsbad, CA, USA) supplemented with 0.45% glucose (HBSS-Glucose). After carefully removing the meninges, the hippocampi were digested mildly with trypsin, washed in HBSS and resuspended in Neurobasal medium supplemented with 2 mM Glutamax (Gibco, Carlsbad, CA, USA) before filtering in 70-mm mesh filters (BD Falcon, Vaud, Switzerland). The cells were then plated onto glass coverslips (5 × 104 cells/cm2) coated with 0.1 mg/ml poly-L-lysine (Sigma-Aldrich, St. Louis, MO, USA), and 2 h after plating, the medium was substituted by complete growth medium (Neurobasal medium supplemented with 2% B27 and 2 mM Glutamax). Primary cultures were incubated at 37◦C in a humidified 5% CO2 atmosphere. Every 3–4 days, half of the conditioned medium was removed and replaced by fresh growth medium. Primary cultures were transfected with Lipofectamine™ 2000 (Invitrogen, Carlsbad, CA, USA) on day in vitro 11 (DIV11) for further surface expression analysis of GFP-GluN2B constructs.

Immunofluorescence analysis

Transiently transfected COS-7 cells were washed in PBS and fixed with 4% paraformaldehyde (PFA). Surface expression of NMDARs was achieved by immunolabelling the extracellular tags (GFP and HA, inserted in the ATD of the GluN subunits) and incubating the corresponding antibody (anti-GFP, Clontech, France; anti-HA, Covance Inc., Princeton, NJ, USA respectively) for 1 h at RT, under non-permeabilizing conditions. After washing, cells were incubated with secondary antibodies (anti-rabbit IgG or anti-mouse IgG, respectively) conjugated to Alexa fluorochromes (Life Technologies, Carlsbad, CA, USA), for 1 h at RT. The total amount of GFP-tagged GluN subunits was detected by the GFP endogenous fluorescence signal emitted by the GFPGluN2A/GluN2B construct. Coverslips were mounted in ProLong antifade mounting medium (Life Technologies, Carlsbad, CA, USA) and images were acquired in a Nikon Eclipse 80i microscope (63×/1.4 N.A. immersion oil objective). To analyze heterologous GluN2B subunit surface expression in primary neuronal cultures, at DIV14, neurons were washed and fixed with 4% PFA in PBS containing 4% sucrose. Surface expression of GFP-GluN2B was detected by incubating with antiGFP (Life Technologies, Carlsbad, CA, USA) during 1 hour at RT and visualized with an Alexa 488-conjugated goat anti-rabbit Ab (Molecular Probes, Carlsbad, CA, USA). The intracellular pool of GluN2B subunits was identified by permeabilizing cells with 0.1% Triton X-100 and labelling them with anti-GFP-Alexa 555- conjugated Ab (Molecular Probes, Carlsbad, CA, USA), as previously described (28). Images were acquired in a Leica TCSSL spectral confocal microscope (Leica Microsystems, Wetzlar, Germany) using a Plan-Apochromat 63×/1.4 N.A. immersion oil objective (Leica Microsystems, Wetzlar, Germany) and a pinhole aperture of 114.54 or 228 μm (for surface receptors). To excite fluorophores, confocal system is equipped with excitation laser beams at 488 and 546 nm. In each experiment, fluorescence intensity was measured in 5–15 COS-7 cells or 1–15 dendrites from at least 7 pyramidal neurons per condition. Fluorescence was quantified using Adobe Photoshop CS5 software (Adobe Systems Inc., San José, CA, USA) and the results are represented as the mean ± standard errors of the mean (SEM) of the surface immunofluorescence signal, analyzing at least three independent experiments.

Electrophysiological recordings of NMDA whole-cell currents in HEK293T cells

Electrophysiological recordings were obtained 38–48 hours after transfection, perfusing the cells continuously at RT with extracellular physiological bath solution (in mM): 140 NaCl, 5 KCl, 1 CaCl2, 10 glucose and 10 HEPES, adjusted to pH 7.42 with NaOH. Glutamate (1 mM, Sigma-Aldrich, St. Louis, MO, USA), in the presence of glycine (50 μM; Tocris, Bristol, UK), was applied for 5 seconds by piezoelectric translation (P-601.30; Physik Instrumente, Karlsruhe, Germany) of a theta-barrel application tool made from borosilicate glass (1.5 mm o.d.; Sutter Instruments, Novato, CA, USA), and the activated currents were recorded in the whole-cell configuration at a holding potential of −60 mV, acquired at 5 kHz and filtered at 2 kHz by means of Axopatch 200B amplifier, Digidata 1440A interface and pClamp10 software (Molecular Devices Corporation, San José, CA, USA). Electrodes with open-tip resistances of 2–4 MΩ were made from borosilicate glass (1.5 mm o.d., 0.86 mm i.d., Harvard Apparatus, Holliston, MA, USA), pulled with a PC-10 vertical puller (Narishige, Tokyo, Japan) and filled with intracellular pipette solution containing (in mM): 140 CsCl, 5 EGTA, 4 Na2ATP, 0.1 Na3GTP and 10 HEPES, adjusted to pH 7.25 with CsOH. Glutamate and glycine-evoked currents were expressed as current density (−pA/pF; maximum current divided by input capacitance as measured from the amplifier settings) to avoid differences due to surface area in the recorded cells. The kinetics of deactivation and desensitization of the NMDAR responses were determined by fitting the glutamate/glycine-evoked responses at Vm − 60 mV to a double exponential function in order to determine the weighted time constant (τ w,des),

$$\tau_{w,\mathrm{des}} = \tau_f \left(\frac{A_f}{A_f + A_s} \right) + \tau_s \left(\frac{A_s}{A_f + A_s} \right),$$

where Af and τf are the amplitude and time constant of the fast component of desensitization and As and τs are the amplitude and time constant of the slow component of desensitization.

Statistical analysis

Comparison between experimental groups was evaluated using Prism (GraphPad Software, Inc., San Diego, CA, USA), applying a one-way analysis of variance (ANOVA) followed by a Bonferroni post hoc test (cell surface expression experiments) or Dunn's multiple comparison (electrophysiology experiments). For single comparisons, either Student's t test (for parametric data) or Mann–Whitney U-test (for non-parametric data) was used. Data are presented as the means ± SEM from at least three independent experiments.

Supplementary Material

Supplementary Material is available at HMG online.

References

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

2. Lau, C.G. and Zukin, R.S. (2007) NMDA receptor trafficking in synaptic plasticity and neuropsychiatric disorders. Nat. Rev. Neurosci., 8, 413–426.

3. Traynelis, S.F., Wollmuth, L.P., McBain, C.J., Menniti, F.S., Vance, K.M., Ogden, K.K., Hansen, K.B., Yuan, H., Myers, S.J. and Dingledine, R. (2010) Glutamate receptor ion channels: structure, regulation, and function. Pharmacol. Rev., 62, 405–496.

4. de Ligt, J., Willemsen, M.H., van Bon, B.W.M., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C. et al. (2012) Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med., 367, 1921–1929.

5. Tarabeux, J., Kebir, O., Gauthier, J., Hamdan, F.F., Xiong, L., Piton, A., Spiegelman, D., Henrion, É., Millet, B., team, S2.D. et al. (2011) Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia. Transl. Psychiatry, 1, e55–e55.

6. Endele, S., Rosenberger, G., Geider, K., Popp, B., Tamer, C., Stefanova, I., Milh, M., Kortüm, F., Fritsch, A., Pientka, F.K. et al. (2010) Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat. Genet., 42, 1021–1026.

7. Platzer, K., Yuan, H., Schütz, H., Winschel, A., Chen, W., Hu, C., Kusumoto, H., Heyne, H.O., Helbig, K.L., Tang, S. et al. (2017) GRIN2B encephalopathy: novel findings on phenotype, variant clustering, functional consequences and treatment aspects. J. Med. Genet., 54, 460–470.

8. Strehlow, V., Heyne, H.O., Vlaskamp, D.R.M., Marwick, K.F.M., Rudolf, G., de Bellescize, J., Biskup, S., Brilstra, E.H., Brouwer, O.F., Callenbach, P.M.C. et al. (2019) GRIN2A-related disorders: genotype and functional consequence predict phenotype. Brain, 142, 80–92.

9. Lemke, J.R., Geider, K., Helbig, K.L., Heyne, H.O., Schütz, H., Hentschel, J., Courage, C., Depienne, C., Nava, C., Heron, D. et al. (2016) Delineating the GRIN1 phenotypic spectrum: a distinct genetic NMDA receptor encephalopathy. Neurology, 86, 2171–2178.

10. Lemke, J.R., Hendrickx, R., Geider, K., Laube, B., Schwake, M., Harvey, R.J., James, V.M., Pepler, A., Steiner, I., Hörtnagel, K. et al. (2014) GRIN2Bmutations in west syndrome and intellectual disability with focal epilepsy. Ann. Neurol., 75, 147–154.

11. Carvill, G.L., Regan, B.M., Yendle, S.C., O'Roak, B.J., Lozovaya, N., Bruneau, N., Burnashev, N., Khan, A., Cook, J., Geraghty, E. et al. (2013) GRIN2A mutations cause epilepsy-aphasia spectrum disorders. Nat. Genet., 45, 1073–1076.

12. Lesca, G., Rudolf, G., Bruneau, N., Lozovaya, N., Labalme, A., Boutry-Kryza, N., Salmi, M., Tsintsadze, T., Addis, L., Motte, J. et al. (2013) GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. Nat. Genet., 45, 1061–1066.

13. Xiang Wei, W., Jiang, Y. and Yuan, H. (2018) De novo mutations and rare variants occurring in NMDA receptors. Curr. Opin. Physiol., 2, 27–35.
14. Vicini, S., Wang, J.F., Li, J.H., Zhu, W.J., Wang, Y.H., Luo, J.H., Wolfe, B.B. and Grayson, D.R. (1998) Functional and pharmacological differences between recombinant N-methyl-Daspartate receptors. J. Neurophysiol., 79, 555–566.
15. Yu, Y., Lin, Y., Takasaki, Y., Wang, C., Kimura, H., Xing, J., Ishizuka, K., Toyama, M., Kushima, I., Mori, D. et al. (2018) Rare loss of function mutations in N-methyl-D-aspartate glutamate receptors and their contributions to schizophrenia susceptibility. Transl. Psychiatry, 8, 12.
16. Stroebel, D., Carvalho, S., Grand, T., Zhu, S. and Paoletti, P. (2014) Controlling NMDA receptor subunit composition using ectopic retention signals. J. Neurosci., 34, 16630–16636.
17. Tovar, K.R., McGinley, M.J. and Westbrook, G.L. (2013) Triheteromeric NMDA receptors at hippocampal synapses. J. Neurosci., 33, 9150–9160.
18. Horak, M., Chang, K. and Wenthold, R.J. (2008) Masking of the endoplasmic reticulum retention signals during assembly of the NMDA receptor. J. Neurosci., 28, 3500–3509.
19. Mohn, A.R., Gainetdinov, R.R., Caron, M.G. and Koller, B.H. (1999) Mice with reduced NMDA receptor expression display behaviors related to schizophrenia. Cell, 98, 427–436.
20. Horak, M. and Wenthold, R.J. (2009) Different roles of C-terminal cassettes in the trafficking of full-length NR1 subunits to the cell surface. J. Biol. Chem., 284, 9683–9691.
21. Sanz-Clemente, A., Nicoll, R.A. and Roche, K.W. (2013) Diversity in NMDA receptor composition: many regulators, many consequences. Neuroscientist, 19, 62–75.
22. Lussier, M.P., Sanz-Clemente, A. and Roche, K.W. (2015) Dynamic regulation of N-methyl-d-aspartate (NMDA) and α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors by posttranslational modifications. J. Biol. Chem., 290, 28596–28603.

23. Perszyk, R.E., Swanger, S.A., Shelley, C., Khatri, A., FernandezCuervo, G., Epplin, M.P., Zhang, J., Le, P., Bülow, P., GarnierAmblard, E. et al. (2020) Biased modulators of NMDA receptors control channel opening and ion selectivity. Nat. Chem. Biol., 16, 188–196.

24. Soto, D., Olivella, M., Grau, C., Armstrong, J., Alcon, C., Gasull, X., Santos-Gómez, A., Locubiche, S., de Salazar, M.G., GarcíaDíaz, R. et al. (2019) L-serine dietary supplementation is associated with clinical improvement of loss-of-function GRIN2B-related pediatric encephalopathy. Sci. Signal., 12, eaaw0936.

25. Tezuka, T., Umemori, H., Akiyama, T., Nakanishi, S. and Yamamoto, T. (1999) PSD-95 promotes Fyn-mediated tyrosine phosphorylation of the N-methyl-D-aspartate receptor subunit NR2A. Proc. Natl. Acad. Sci. U. S. A., 96, 435–440.

26. Taniguchi, S., Nakazawa, T., Tanimura, A., Kiyama, Y., Tezuka, T., Watabe, A.M., Katayama, N., Yokoyama, K., Inoue, T., Izumi-Nakaseko, H. et al. (2009) Involvement of NMDAR2A tyrosine phosphorylation in depression-related behaviour. EMBO J., 28, 3717–3729.

27. Barria, A. and Malinow, R. (2002) Subunit-specific NMDA receptor trafficking to synapses. Neuron, 35, 345–353.

28. Grau, C., Arató, K., Fernández-Fernández, J.M., Valderrama, A., Sindreu, C., Fillat, C., Ferrer, I., la Luna de, S. and Altafaj, X. (2014) DYRK1A-mediated phosphorylation of GluN2A at Ser (1048) regulates the surface expression and channel activity of GluN1/GluN2A receptors. Front. Cell. Neurosci., 8, 331.

Article V. HomolWat: a web server tool to incorporate ‘homologous’ water molecules into GPCR structures

ABSTRACT

Internal water molecules play an essential role in the structure and function of membrane proteins including G protein-coupled receptors (GPCRs). However, technical limitations severely influence the number and certainty of observed water molecules in 3D structures. This may compromise the accuracy of further structural studies such as docking calculations for molecular dynamics simulations. Here we present HomolWat, a web application for incorporating water molecules into GPCR structures by using template-based modelling of homologous water molecules obtained from high-resolution structures. While there are various tools available to predict the positions of internal waters using energy-based methods, the approach of borrowing water molecules from homologous GPCR structures makes HomolWat unique. The tool can incorporate water molecules into a protein structure in about a minute with around 85% of water recovery. The web server is freely available at http://lmc.uab.es/ homolwat.

INTRODUCTION

Water molecules confined inside cavities in a protein, named ordered or internal water molecules, play an essential role in the structure and function of proteins, ligand binding mechanisms, and catalytic reactions. In several proteins like G protein-coupled receptors (GPCRs), water molecules also mediate the core mechanism of activation (1–5). GPCRs are the largest family of membrane proteins with over 800 members in humans. Despite sharing a common seven - helical transmembrane architecture and similar conformational changes upon activation (1,6), they recognize a wide diversity of extracellular signals like hormones, neurotransmitters or entire proteins. As a result, they play a key role in signal transduction and have become targets of 35% of the currently approved drugs (7). Recent studies have demonstrated that the activation of these receptors involves a specific order of internal water molecules (8). Moreover, molecular dynamics simulations have shown that internal water molecules are highly

conserved among GPCRs and participate in their common activation mechanism (9). Overall, this emphasizes the important role that water molecules play in GPCR function.

Sequence homology and phylogenetic analyses have classified GPCRs into six families (or classes), namely A to F (10). Recent advances in protein engineering and structural biology have led to a rapid growth in the number of GPCR structures deposited in the Protein Data Bank (PDB). Thus, 10 years ago only 4 GPCR structures of class A had been deposited in the PDB whereas today this archive hosts 346 structures of GPCRs, 64 of which belong to unique receptor subtypes of four different GPCR classes (i.e. A, B1, C and F) (GPCRdb, http://gpcrdb.org/structure/statistics, 2020). However, technical limitations including resolution severely influence the number and certainty of solved water molecules in GPCR structures. In this context, molecular modeling can help improve and maximize internal hydration of these proteins. Current tools and methods to predict water placement in proteins (reviewed in (11)), range from knowledge-based or molecular mechanics methods to simulation approaches (e.g. molecular dynamics or Monte Carlo). One common pitfall of most of the former methods is the lack of experimental data supporting the modeled position. Another group of tools such as PyWater (12) and ProBIS H2O (12) has focused on identifying conserved or homologous water molecules in proteins using experimental data (12,13). The plethora of high-resolution GPCR structures recently deposited in the PDB (14) have simultaneously increased the number of solved waters, thus opening a door to improving the placement of internal water molecules by structural homology.

Here, we present HomolWat, a freely accessible web application (available at http://lmc.uab.cat/homolwat) aimed at incorporating internal water molecules into GPCR structures by using a molecular modeling method that borrows lacking water molecules from homologous structures. HomolWat relies on an up-to-date curated database of all internal water molecules from high-resolution structures of GPCRs deposited in the PDB. The tool uses this information to superpose water molecules from related structures in a hierarchical fashion. Water molecules that fit into receptor cavities not yet hydrated are incorporated into the model. Our method offers a novel,

fast and reliable way to place internal water molecules in GPCR structures.

MATERIALS AND METHODS

HomolWat reference database

We have constructed an up-to-date reference database with water molecules determined in all resolved GPCR structures in the PDB (15). To ensure that coordinates have been obtained using the latest experimental and computational methods we have downloaded the structures from PDB REDO whenever possible (16). Any non-protein molecules other than water, or GPCR orthosteric/allosteric ligands were removed from the structure. Additionally, auxiliary proteins used to assist in structure stabilization were also removed. Water molecules with low order (B-factor > 45 A2), (17) were discarded. Water molecules were classified as internal or external based on their circular variance, following a previously reported method (18). We discarded external water molecules (those with circular variance < 0.6, computed within a radius of 10 A around the water oxygen) for being incompatible with the membrane outside the crystal lattice. At the moment of writing HomolWat database contains 191 receptor chains from 150 high-resolution structures and 44 unique receptors, totaling 2448 internal water molecules (see http://lmc.uab.cat/homolwat/gpcr_table). The distribution of internal waters within GPCR classes and subfamilies is shown in Supplementary Figure S1.

Implementation of the web service

HomolWat relies on a Python (v.3.7) backend that uses the Flask web framework (v1.1.1). Data for the internal water molecules is stored in a MySQL database (v.8.0.18). The web server exploits the capabilities of the popular web based viewer NGL (v.2.0.0, (19)) for structure visualization. Preprocessing of new GPCR structures and placement of water molecules is fully automated within a routine using Python and Bash scripts. Sequence alignment is performed using Blast+ (v2.6.0+, (20)).

RESULTS

HomolWat pipeline

HomolWat protocol is schematically represented in Figure 4.3-14. The input of HomolWat is a file with the 3D structure in the PDB format (alternatively the user can select a PDB id and a specific chain) whereas the output is the same structure containing homologous internal water molecules obtained from the HomolWat reference database. This hydrated structure can be visualized interactively using NGL viewer (19) and downloaded for further use. The overall HomolWat protocol for water molecule placement contains steps as follows: First, HomolWat uses Blast+ to perform a multiple sequence alignment of the input structure sequence against all GPCR sequences from receptors hosted in Ho molWatDB, resulting in a list of homologous structures with water molecules sorted according to their Blast+ score. The user is then required to choose the GPCR functional state (i.e. inactive or active) in order to prioritize water molecules present in active or inactive structures. The user can decide whether to try to incorporate or not a sodium ion near the conserved residue D2.50 (numbering following Ballesteros and Weinstein scheme (21)) present in most active structures (8,22) from the structure with the highest sequence identity and best resolution. The ion is introduced when there are no water molecules within a radius of 2.1 A or protein atoms within 1.8 A (23). Moreover, users have the option to run the popular energy-based method Dowser+ (24) alongside the main HomolWat protocol and incorporate predicted water molecules that do not overlap with homologous waters. This option becomes more useful when few homologous water molecules exist. In addition, using a Blast+ score threshold, the user can limit the incorporation of water molecules from close homologs only. Next, Homolwat performs a global structural alignment of homologous structures to the query structure in descending order of Blast+ score and from highest to lowest resolution using the *align* and *super* functions of the visualization software PyMol (v.2.0.5, (25)). Subsequently, the position of each water molecule is refined through a local structural alignment of residues around 10 A using the *super* function and only those waters with a RMSD in the local structural alignment up to 2.0 A are kept. Water is assessed one by one in increasing

order of B-factor and incorporated into the model should they not clash with atoms from the query model or already incorporated water molecules (distance cutoff of 2.4 A (26)). The structure solvated with internal waters is shown interactively using an embedded NGL viewer (19) that also shows the source structure (PDB id, chain and Uniprot entry name) of each water molecule. The solvated structure, a list with the incorporated water molecules and a PyMOL session with the structures used for water positioning can be downloaded for further usage.

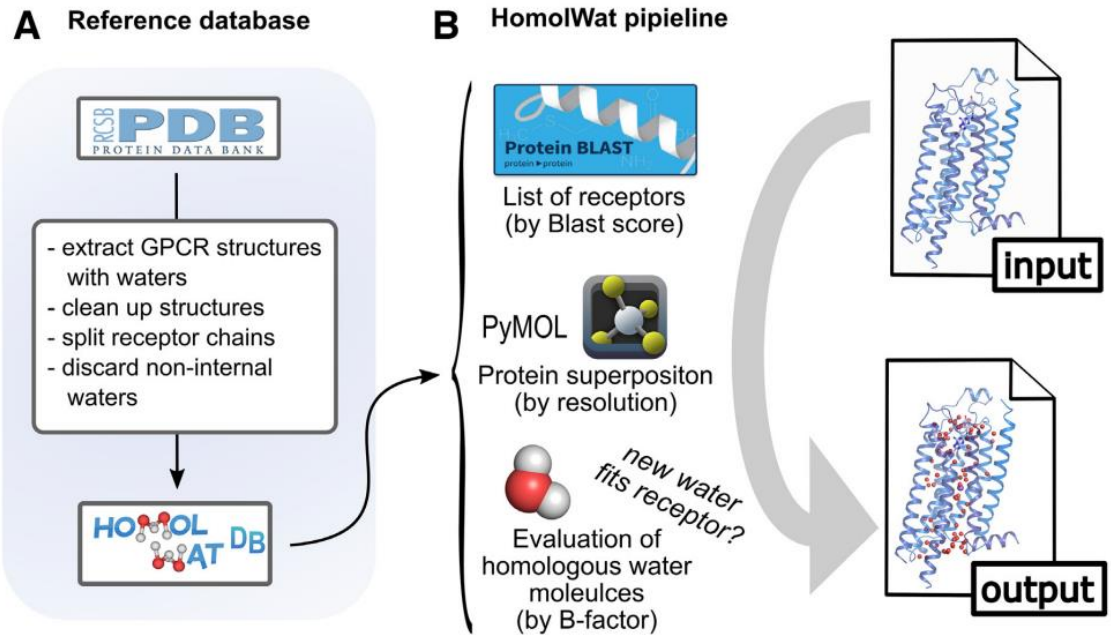
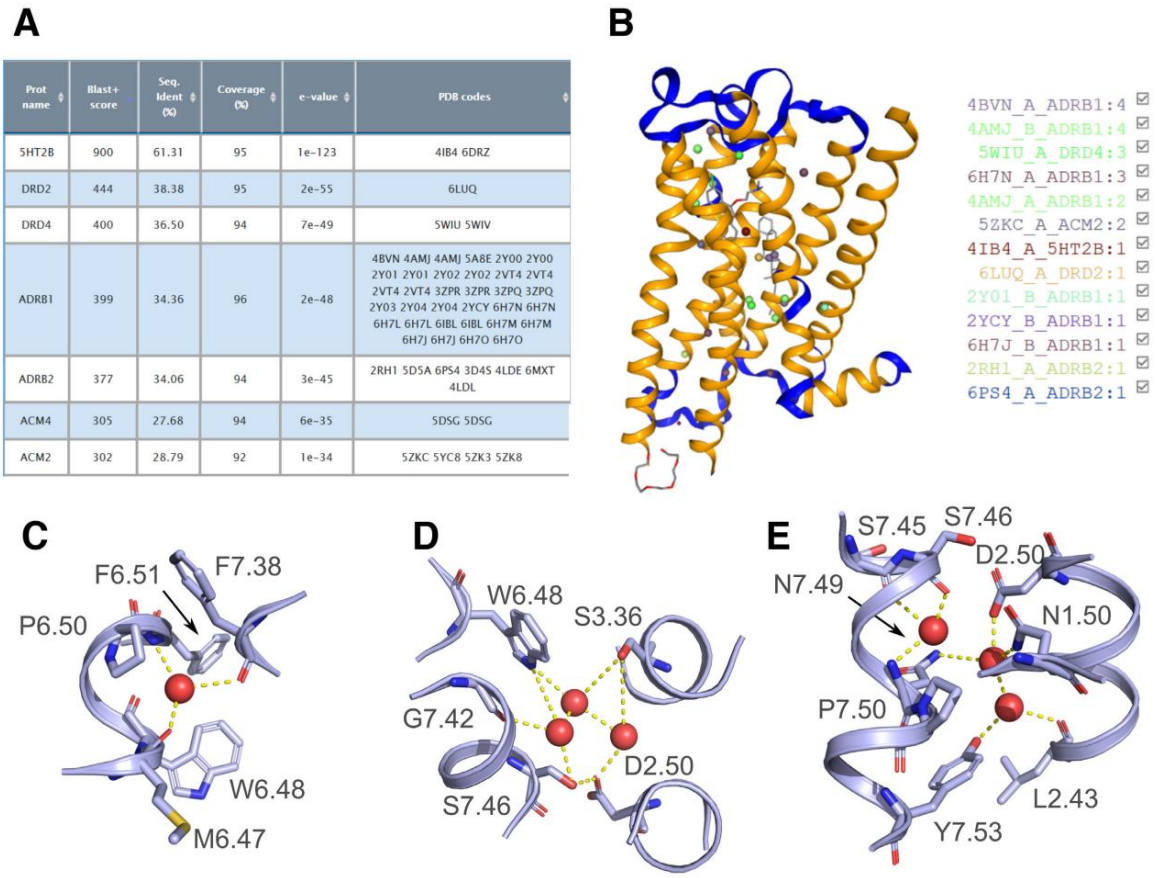


Figure 4.3-14. Schematic representation of the HomolWat reference database (**A**) and overall pipeline (**B**). In (B), the user provides an input 3D structure file of a GPCR—that could contain ligands or water molecules—in PDB format and obtains an output PDB file of the same structure with additional homologous internal water molecules inserted. As a first step, HomolWat extracts the protein sequence from the PDB file and runs Blast+ on all receptor sequences in HomolWatDB. This provides a list of receptors sorted by Blast score from high to low sequence homology. Subsequently, HomolWat uses this list to assess if a water molecule fits in the receptor, considering previously inserted water molecules. This assessment is performed for every water molecule across all receptors within chosen Blast+ score, and across every structure for each receptor.

Test case

To illustrate the use of HomolWat, we used our tool to place internal water molecules into the recently resolved crystal structure of the inactive serotonin 5-HT2A receptor at

2.9 A resolution (PDB id 6A94, (27)) where no internal water molecules could be determined. We selected chain A of this structure using the dropdown menu that loads preprocessed structures without fusion protein, nanobodies or other molecules used for crystallization purposes. We specified i) that the structure is in an inactive state, ii) that we would like to add the conserved sodium, iii) that we do not want to use Dowser+ to predict additional waters and that we will allow structures with a Blast+ score threshold of 300 (i.e. include all amine receptors). In about 15 seconds, HomolWat incorporated 25 water (Figure 4.3-15A–B) molecules from the homologous 5-HT2B receptor (61% sequence identity, 1 water), dopamine D2 receptor (38%, 1 water), dopamine D4 receptor (37%, 3 waters), 1-adrenergic receptor (34%, 16 waters), the $_2$-adrenergic receptor (34%, 2 waters) and the muscarinic M_2 receptor (29%, 2 waters). The hydrated model contains the conserved water at the large proline-associated kink in transmembrane helix 6 (3) along with other known functional water molecules (4,9) at regions that span from near the or thosteric site to more cytoplasmic locations near residues N1.50, D2.50, N7.49 and Y7.53 (3,9) (Figure 4.3-15C–E). Interestingly, HomolWat could not place the sodium ion as the mutated protein used for crystallization included a mutation (i.e. S3.39K) near residue D2.50, which introduces a large sidechain that hampers Na^+ binding.



Prot name	Blast+ score	Seq. Ident (%)	Coverage (%)	e-value	PDB codes
5HT2B	900	61.31	95	1e-123	4IB4 6DRZ
DRD2	444	38.38	95	2e-55	6LUQ
DRD4	400	36.50	94	7e-49	5WIU 5WIV
ADRB1	399	34.36	96	2e-48	4BVN 4AMJ 4AMJ 5A8E 2Y00 2Y00 2Y01 2Y01 2Y02 2Y02 2VT4 2VT4 2VT4 2VT4 3ZPR 3ZPR 3ZPQ 3ZPQ 2Y03 2Y04 2Y04 2YCY 6H7N 6H7N 6H7L 6H7L 6IBL 6IBL 6H7M 6H7M 6H7J 6H7J 6H7O 6H7O
ADRB2	377	34.06	94	3e-45	2RH1 5D5A 6PS4 3D4S 4LDE 6MXT 4LDL
ACM4	305	27.68	94	6e-35	5DSG 5DSG
ACM2	302	28.79	92	1e-34	5ZKC 5YC8 5ZK3 5ZK8

4BVN_A_ADRB1:4
4AMJ_B_ADRB1:4
5WIU_A_DRD4:3
6H7N_A_ADRB1:3
4AMJ_A_ADRB1:2
5ZKC_A_ACM2:2
4IB4_A_5HT2B:1
6LUQ_A_DRD2:1
2Y01_B_ADRB1:1
2YCY_B_ADRB1:1
6H7J_B_ADRB1:1
2RH1_A_ADRB2:1
6PS4_A_ADRB2:1

Figure 4.3-15. HomolWat test case: the crystal structure of the inactive serotonin 5-HT2A receptor (PDB id 6A94) without internal water molecules. (**A**) Screenshot of the results for the Blast+ alignment to receptors in HomolWatDB. (**B**) Screenshot of the output PDB structure with the introduced internal water molecules displayed in the NGL viewer. (**C-E**) Details of the clusters of water molecules associated with functionally important residues at different levels ranging from the orthosteric site to the G-protein binding crevice upon incorporation to its receptor core using HomolWat. The receptor is shown in light blue cartoons and water molecules are displayed as red spheres.

Validation

To validate the ability of HomolWat to introduce internal water molecules, we evaluated the percentage of recovery using a set of 19 Class A structures accounting for at least one unique structure from any GPCR with resolution <2.8 A and more than five internal water molecules, where we had previously removed all water molecules. To avoid redundancy, we excluded from the reference dataset the respective water molecules of those structures included in the test set. The results were compared to Dowser+ (24) predictions, which we chose, among various alternatives (see (11)), as

representative software able to find cavities in proteins and solvate them according to an energy criterion. Homol Wat places a median of 86 molecules per receptor, consistent with the number of experimentally determined water molecules within the protein core of rhodopsin (28). Figure 4.3-16 shows how many of the original internal water molecules for the tested 19 PDB structures were recovered by HomolWat using a cutoff radius <2 A. The percentages of recovery range between 41.2 and 100%, with a median of 84.6% (see Supplementary Table S1) and the average distance between original and recovered water is 0.68 A (see Supplementary Table S2). Larger recoveries are obtained for receptors with many homolog high-resolution structures, whereas poor recovery is obtained for receptors with a small number of determined structures with water molecules and low homology to other receptors. Our method outperforms previously described knowledge-based methods like Dowser+ (24).

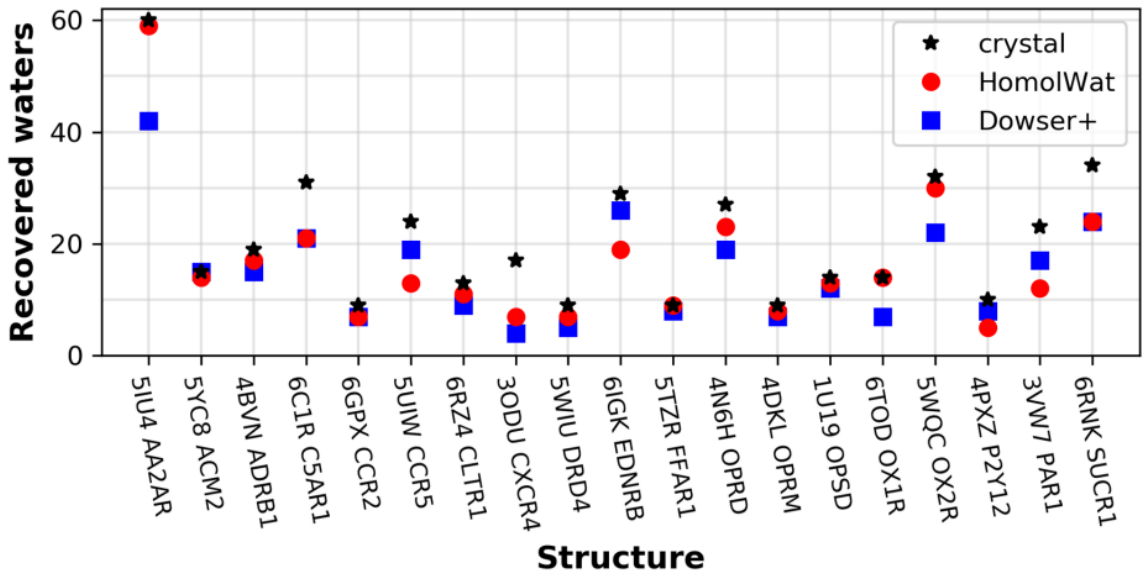


Figure 4.3-16. Validation of HomolWat and comparison to Dowser+. Number of water molecules recovered using Homolwat (red squares) and Dowser+ (blue dots) upon removal of the originally resolved water molecules (black star) in each PDB structure (mapped to their Uniprot accession code).

CONCLUSIONS AND FURTHER DIRECTIONS

HomolWat is a web application to introduce internal water molecules in GPCR structures using resolved water molecules from homologous structures. The tool uses a database of GPCR structures containing internal water molecules to place homologous water positions into GPCR models or experimental structures with few or no internal waters. Due to the foreseeable increase in the number of resolved high-resolution GPCR structures, HomolWat will be able to use more templates, hence increasing its current performance. Better GPCR models that explicitly introduce internal water molecules may for instance improve docking calculations or molecular dynamics simulations. HomolWat has been successfully applied as part of the pipeline used in the GPCRmd project (http://gpcrmd.org), a community-driven effort to create the first open, interactive and standardized database of GPCR molecular dynamics simulations (29). As shown in our test benchmark, HomolWat water placement pipeline showed a median recovery of 83%. The fact that HomolWat employs experimental knowledge to perform molecular modeling gives a clear competitive advantage to our tool when compared to methods based on energy calculations such as Dowser+ (24), Wa terMap (30) or Waterdock (31). Due to its automated fashion, this tool could expand to target water molecule placement in other protein families.

DATA AVAILABILITY

HomolWat is a web server freely accessible at http://lmc.uab.es/homolwat. The reference database can be obtained on request or downloaded from the website. The source code is available at https://github.com/EMayol/HomolWat.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Tehan,B.G., Bortolato,A., Blaney,F.E., Weir,M.P. and Mason,J.S. (2014) Unifying family A GPCR theories of activation. *Pharmacol. Ther. Dent.*, **143**, 51–60.
2. Yuan,S., Vogel,H. and Filipek,S. (2013) The role of water and sodium ions in the activation of the -opioid receptor. *Angew. Chem. Int. Ed.*, **52**, 10112–10115.
3. Pardo,L., Deupi,X., Dolker,N., Löpez-Rodríguez,M.L. and Campillo,M. (2007) The role of internal water molecules in the structure and function of the rhodopsin family of G protein-coupled receptors. *ChemBioChem*, **8**, 19–24.
4. Angel,T.E., Chance,M.R. and Palczewski,K. (2009) Conserved waters mediate structural and functional activation of family A (rhodopsin-like) G protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 8555–8560.
5. Yuan,S., Filipek,S., Palczewski,K. and Vogel,H. (2014) Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway. *Nat. Commun.*, **5**, 4733.
6. Venkatakrishnan,A.J., Deupi,X., Lebon,G., Heydenreich,F.M., Flock,T., Miljus,T., Balaji,S., Bouvier,M., Veprintsev,D.B., Tate,C.G. *et al.* (2016) Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature*, **536**, 484–487.
7. Hauser,A.S., Attwood,M.M., Rask-Andersen,M., Schioth,H.B. and ¨ Gloriam,D.E. (2017) Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discovery*, **16**, 829–842.
8. Liu,W., Chun,E., Thompson,A.A., Chubukov,P., Xu,F., Katritch,V., Han,G.W.,

Roth,C.B., Heitman,L.H., IJzerman,A.P. *et al.* (2012) Structural basis for allosteric regulation of GPCRs by sodium ions. *Science*, **337**, 232–236.

9. Venkatakrishnan,A.J., Ma,A.K., Fonseca,R., Latorraca,N.R., Kelly,B., Betz,R.M., Asawa,C., Kobilka,B.K. and Dror,R.O. (2019) Diverse GPCRs exhibit conserved water networks for stabilization and activation. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 3288–3293.
10. Kolakowski,L.F. Jr (1994) GCRDb: a G-protein-coupled receptor database. *Receptors Channels*, **2**, 1–7.
11. Nittinger,E., Flachsenberg,F., Bietz,S., Lange,G., Klein,R. and Rarey,M. (2018) Placement of water molecules in protein Structures: From Large-Scale evaluations to Single-Case examples. *J. Chem. Inf. Model.*, **58**, 1625–1637.
12. Jukic,M., Konc,J., Gobec,S. and Janeˇ ˇ zic,D. (2017) Identification of ˇ conserved water sites in protein structures for drug design. *J. Chem. Inf. Model.*, **57**, 3094–3103.
13. Patel,H., Grüning,B.A., G ¨unther,S. and Merfort,I. (2014) PyWATER: a PyMOL plug-in to find conserved water molecules in proteins by clustering. *Bioinformatics*, **30**, 2978–2980.
14. Qu,X., Wang,D. and Wu,B. (2020) Progress in GPCR structure determination. In: *GPCRs*. Elsevier, pp. 3–22.
15. Berman,H.M., Bhat,T.N., Bourne,P.E., Feng,Z., Gilliland,G., Weissig,H. and Westbrook,J. (2000) The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.*, **7**(Suppl), 957–959.
16. Joosten,R.P., Long,F., Murshudov,G.N. and Perrakis,A. (2014) The PDB REDO server for macromolecular structure model optimization. *IUCrJ*, **1**, 213–220.
17. Yuan,Z., Bailey,T.L. and Teasdale,R.D. (2005) Prediction of protein B-factor profiles. *Proteins*, **58**, 905–912.
18. Mezei,M. (2003) A new method for mapping macromolecular topography. *J. Mol. Graph. Model.*, **21**, 463–472.
19. Rose,A.S. and Hildebrand,P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.*, **43**, W576–W579.
20. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and

Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

21. Ballesteros,J.A. and Weinstein,H. (1995) [19]Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.*, **25**, 366–428.
22. Gutierrez-de-Terán,H., Massink,A., Rodríguez,D., Liu,W., Han,G.W., Joseph,J.S., Katritch,I., Heitman,L.H., Xia,L., IJzerman,A.P. *et al.* (2013) The role of a sodium ion binding site in the allosteric modulation of the A2A adenosine G protein-coupled receptor. *Structure*, **21**, doi:10.1016/j.str.2013.09.020.
23. Mancinelli,R., Botti,A., Bruni,F., Ricci,M.A. and Soper,A.K. (2007) Hydration of sodium, potassium, and chloride ions in solution and the concept of structure maker/breaker. *J. Phys. Chem. B*, **111**, 13570–13577.
24. Morozenko,A., Leontyev,I.V. and Stuchebrukhov,A.A. (2014) Dipole moment and binding energy of water in proteins from crystallographic analysis. *J. Chem. Theory Comput.*, **10**, 4618–4623.
25. Schrodinger,LLC. The PyMOL Molecular Graphics System, Version ¨ 2.0.5.
26. van Beusekom,B., Touw,W.G., Tatineni,M., Somani,S., Rajagopal,G., Luo,J., Gilliland,G.L., Perrakis,A. and Joosten,R.P. (2018) Homology-based hydrogen bond information improves crystallographic structures in the PDB. *Protein Sci.*, **27**, 798–808.
27. Kimura,K.T., Asada,H., Inoue,A., Kadji,F.M.N., Im,D., Mori,C., Arakawa,T., Hirata,K., Nomura,Y., Nomura,N. *et al.* (2019)Structures of the 5-HT2A receptor in complex with the antipsychotics risperidone and zotepine. *Nat. Struct. Mol. Biol.*, **26**, 121–128.
28. Fried,S.D.E., Eitel,A.R., Weerasinghe,N., Norris,C.E., Somers,J.D., Fitzwater,G.I., Pitman,M.C., Struts,A.V., Suchithranga M,D and Brown,M.F. (2019) G-protein-coupled receptor activation mediated by internal hydration. *Biophys. J.*, **116**, 207a.
29. Rodríguez-Espigares,I., Torrens-Fontanals,M., Tiemann,J.K.S., Aranda-

García,D., Ram´ırez-Anguita,J.M., Stepniewski,T.M., Worp,N., Varela-Rial,A., Morales-Pastor,A., Lacruz,B.M. *et al.* (2020) GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nat. Methods*, doi:10.1038/s41592-020-0884-y.

30. Abel,R., Young,T., Farid,R., Berne,B.J. and Friesner,R.A. (2008) Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.*, **130**, 2817–2831.
31. Sridhar,A., Ross,G.A. and Biggin,P.C. (2017) Waterdock 2.0: water placement prediction for Holo-structures with a pymol plugin. *PLoS One*, **12**, e0172743.
32. Bostock.M.. Ogievetsky.V. and Heer.J. (2011) D3: Data-Driven Documents. IEEE Trans. Vis. Comput. Graph.. 17. 2301–2309

Article VI. DIMERBOW: exploring possible GPCR dimer interfaces

ABSTRACT

Motivation: G protein-coupled receptors (GPCRs) can form homo-, heterodimers and larger order oligomers that exert different functions than monomers. The pharmacological potential of such complexes is hampered by the limited information available on the type of complex formed and its quaternary structure. Several GPCR structures in the Protein Data Bank display crystallographic interfaces potentially compatible with physiological interactions.

Results: Here we present DIMERBOW, a database and web application aimed to visually browse the complete repertoire of potential GPCR dimers present in solved structures. The tool is suited to help finding the best possible structural template to model GPCR homomers.

Availability: DIMERBOW is available at http://lmc.uab.es/dimerbow/.

INTRODUCTION

5-10% of genes in the human genome code for environment-sensing cell surface receptors, most of which are G protein-coupled receptors (GPCRs). This family of receptors respond to a huge variety of stimuli, including ions, amino acids, lipids, peptides, proteins and even photons. For long, GPCRs were described as monomeric transmembrane (TM) receptors, but it is now well accepted that they can form homo- and heteromer complexes in cells (Gurevich and Gure vich, 2018; BorrotoEscuela et al., 2014; Fuxe et al., 2010). They constitute novel signaling units with unique functional and regulatory properties that are opening new opportunities for drug discovery (Ferré,

2015; Fuxe et al., 2010; Franco et al., 2016). While GPCR heteromers have been thoroughly characterized by biophysical and biochemical methods (Sleno and Hébert, 2018; Xue et al., 2019; Guo et al., 2017), the structural basis behind the allosteric communications between receptor protomers remains poorly understood. Most GPCRs interact via their TM domain, except for class C GPCRs, where both extracellular and TM domains are involved (Koehl et al., 2019). Several crystal structures of GPCRs reveal crystallographic dimers compatible with the spatial constraints imposed by the membrane (Cordomí et al., 2015; Katritch et al., 2013; Simpson et al., 2010). Although some of these dimers might not represent biologically relevant interfaces outside the crystal lattice, they are certainly valuable templates to model GPCR homomers and heteromers. Stenkamp identified 71 parallel dimeric crystallographic interfaces (Stenkamp, 2018) in a systematic analysis of the Protein Data Bank (PDB) (Berman et al., 2000). Additionally, molecular dynamics (MD) simulations of crystallographic dimers have also been used to shed light on the structural basis of GPCR homodimers (Baltoumas et al., 2016; Johnston and Filizola, 2014). Here we present DIMERBOW, a web application (available at http://lmc.uab.cat/dimerbow) aimed at exploring the large and growing repertoire of crystallographic dimers of GPCRs involving the TM domain. DIMERBOW features a systematic analysis of protomer-protomer contacts along with the results from coarse-grained MD simulations of all feasible dimers. The database currently contains information from 97 interfaces and represents a unique tool to model GPCR dimers and oligomers. An automated pipeline for processing new interfaces, submitting simulations and performing data analysis guarantees an up-to-date resource.

METHODS

Database of putative dimers.

DIMERBOW relies on a database of putative GPCR dimers (Suppl. Note 1) from protomer-protomer pairs observed in structures deposited in the PDB (Berman et al., 2000).

Coarse-grained md simulations.

As a complementary tool to explore crystallographic interfaces, we used coarse-grained MD simulations with the MARTINI v2.1 force field (Marrink et al., 2007). We simulated a total of 97 dimeric interfaces totaling to 300 μs (Suppl. Note 2). All simulations were run using GROMACS v.2018.5 (Abraham et al., 2015).

Implementation and update.

DIMERBOW relies on a Python (v.3.7) backend that uses the Flask framework (v.1.0.2). Dimer data is stored in a MySQL database (v.8.0.18). Interactive plots and molecule visualizations employ Bokeh (v.1.3.5b) and NGL (v.2.0.0, (Rose and Hildebrand, 2015)), respectively. Preparation of experimental dimeric structures, submission of MD simulation and analysis pipelines are all automated using Python, Bash and R scripts.

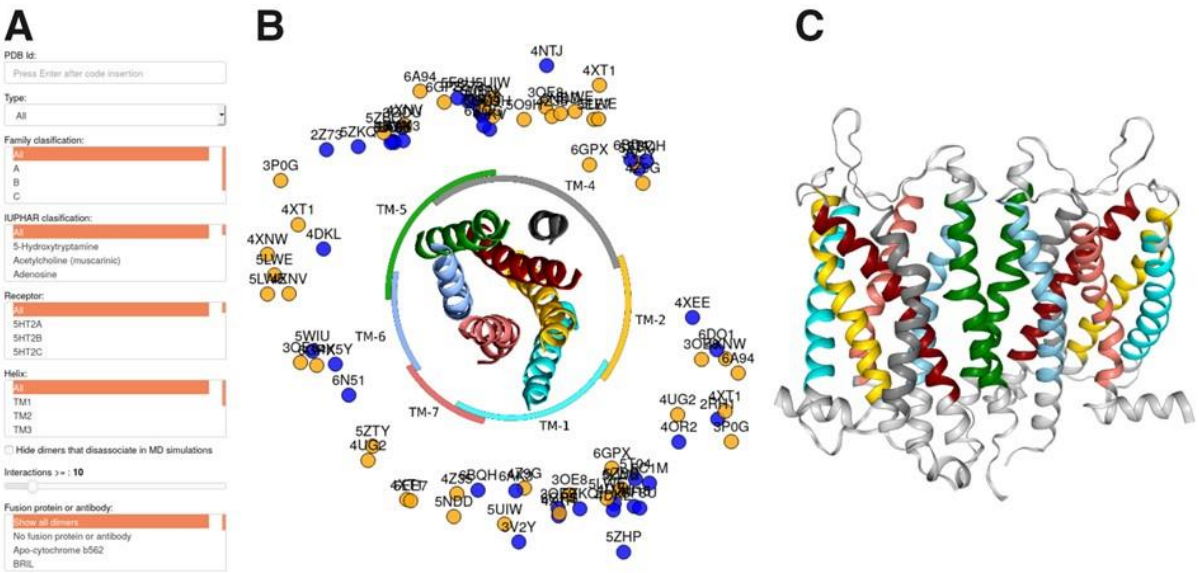


Fig. 4.3-17. The PDB dimer browser. **A**) filtering options, **B**) central panel displaying a schematic representation of all dimers, and **C**) interactive visualization of each dimer structure in NGL (Rose and Hildebrand, 2015).

RESULTS AND CONCLUSION

The main functionality of DIMERBOW is: i) browsing putative dimers from high-resolution structures of GPCRs deposited in the PDB (Fig. 4.3-17). As a

complementary tool, we also provide coarse-grained MD simulations of the putative dimers inserted in a model POPC membrane (Suppl. Fig. S1). In both browsers, a central panel features one (reference) protomer surrounded by the second protomers displayed as circles. Mouse actions on these circles can open an interactive 3D representation of each dimer in the right panel or display extra information including the number of residue-residue contacts, helices involved, and dimer stability in MD simulations. Finally, a toolkit in the left panel allows filtering by receptor type, helices involved, minimal number of interactions in the dimer interface, or interaction mode. Suppl. Table 1 summarizes the number of dimeric interfaces present in the current release. Regarding dimer's symmetry, Suppl. Table 1 shows a higher number of head-to-head (HH) dimers when compared to head-to-tail (HT) dimers. While most evidence supports the existence of HH dimers, previous (Pluhackova et al., 2016) and new (Köfalvi et al., 2020) reports also suggest a physiological role for HT dimers. Likewise, in line with previous reports (Guo et al., 2008; Liang et al., 2003; Katritch et al., 2013), a significant number of dimers involve TM1 (29%), TM4 (19%), and TM5 (26%) (see Suppl. Table 1). HH dimers tend to involve TM1 together with TM2 or TM7, and TM5 together with TM4 or TM6, whereas HT dimers mostly involve TM1 interacting with TM4.

DIMERBOW is a web application for rapid, intuitive, and systematic visualization of the complete repertoire of crystallographic interfaces for GPCR homodimers available in the PDB. The tool is suited to help computational scientists, molecular biologists and crystallographers interested in GPCR dimers and oligomers finding the best possible structural template to model GPCR homomers or heteromers or to ascertain how frequent a specific dimeric interface occurs. The tool has the potential to be expanded to other families of membrane proteins forming dimers or oligomers.

SUPPLEMENTARY DATA

Supplementary data are available at Bioinformatics online.

REFERENCES

- Abraham,M.J. et al. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 1-2, 19– 25.
- Baltoumas,F.A. et al. (2016) Molecular dynamics simulations and structure-based network analysis reveal structural and functional aspects of G-protein coupled receptor dimer interactions. J. Comput. Aided Mol. Des., 30, 489–512.
- Berman,H.M. et al. (2000) The Protein Data Bank. Nucleic Acids Res., 28, 235–242.
- Borroto-Escuela,D.O. et al. (2014) The G protein-coupled receptor heterodimer network (GPCR-HetNet) and its hub components. Int. J. Mol. Sci., 15, 8570–8590.
- Cordomí,A. et al. (2015) Structures for G-Protein-Coupled Receptor Tetramers in Complex with G Proteins. Trends in Biochemical Sciences, 40, 548–551.
- Ferré,S. (2015) The GPCR heterotetramer: challenging classical pharmacology. Trends Pharmacol. Sci., 1–8.
- Franco,R. et al. (2016) Basic Pharmacological and Structural Evidence for Class A G-Protein-Coupled Receptor Heteromerization. Front. Pharmacol., 7, 76.
- Fuxe,K. et al. (2010) The changing world of G protein-coupled receptors: from monomers to dimers and receptor mosaics with allosteric receptor-receptor interactions. J. Recept. Signal Transduct. Res., 30, 272–283.
- Guo,H. et al. (2017) Methods used to study the oligomeric structure of G-protein coupled receptors. Biosci. Rep., 37.
- Guo,W. et al. (2008) Dopamine D2 receptors form higher order oligomers at physiological expression levels. EMBO J., 27, 2293–2304.

- Gurevich,V.V. and Gurevich,E.V. (2018) GPCRs and Signal Transducers: Interaction Stoichiometry. Trends Pharmacol. Sci., 39, 672–684.
- Johnston,J.M. and Filizola,M. (2014) Differential stability of the crystallographic interfaces of mu- and kappa-opioid receptors. PLoS One, 9, e90694.
- Katritch,V. et al. (2013) Structure-function of the G protein-coupled receptor superfamily. Annu. Rev. Pharmacol. Toxicol., 53, 531–556.
- Koehl,A. et al. (2019) Structural insights into the activation of metabotropic glutamate receptors. Nature, 566, 79–84.
- Köfalvi,A. et al. (2020) Control of glutamate release by complexes of adenosine and cannabinoid receptors. BMC Biol., 18, 9.
- Liang,Y. et al. (2003) Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. J. Biol. Chem., 278, 21655–21662.
- Marrink,S.J. et al. (2007) The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. The Journal of Physical Chemistry B, 111, 7812–7824.
- Pluhackova,K. et al. (2016) Dynamic Cholesterol-Conditioned Dimerization of the G Protein Coupled Chemokine Receptor Type 4. PLOS Comput. Biol., 12, e1005169.
- Rose,A.S. and Hildebrand,P.W. (2015) NGL Viewer: a web application for molecular visualization. Nucleic Acids Res., 43, W576–9.
- Simpson,L.M. et al. (2010) Bioinformatics and molecular modelling approaches to GPCR oligomerization. Curr. Opin. Pharmacol., 10, 30–37.
- Sleno,R. and Hébert,T.E. (2018) The Dynamics of GPCR Oligomerization and Their Functional Consequences. Int. Rev. Cell Mol. Biol., 338, 141–171.
- Stenkamp,R.E. (2018) Identifying G protein-coupled receptor dimers from crystal packings. Acta Crystallogr D Struct Biol, 74, 655–670.

Chapter 5 - Global discussion

Chapter 5 - Global discussion

Proteins represent one of the main building blocks of cells. In the human genome, there are more than 21.000 protein-coding genes, and their sequences are the guide that dictates the structure of these vital proteins which, in turn, dictate their functional properties. Thus, there is a lot of interest in understanding the relationship between sequence and structure and function in proteins. We are living in the era of Big Data and biomedical data is growing exponentially. From the creation of the first databases, such as Protein Atlas, OMIM, and GenkBank, the research about proteins has been generating large amounts of data either on the sequence, structure, or functionality. When looking at available 3D structures in the PDB, most of the proteins with an available 3D structure are globular proteins, as the structure of membrane proteins is difficult to determine. The different environments of membrane proteins compared to globular proteins condition the structure and the sequence, requiring specific approaches to study membrane proteins. Consequently, membrane proteins need computational approaches to understand the relationship between sequence and structure. The main objective of this thesis is to contribute to the study of membrane proteins, by providing specific tools aimed to construct reliable models of membrane proteins and also to relate sequence variation to structure and function variation.

This thesis combines studies performed on all membrane proteins with studies centered on two major groups of membrane proteins: NMDA receptors and GPCRs. The tools developed for GPCRs and NMDA receptors represent the first step and could be extended in the future to all membrane proteins. The Bioinformatics tools used in this thesis rely on the analysis of Structural Data, Clinical Data, and Sequence Variation by the integration and the analysis of several databases.

The first part of the thesis focuses on the sequence of membrane proteins. The effect of mutations in various fields has been studied for a long time, and it is known that they have been a key point in the evolution of species, but sometimes these mutations affect the DNA sequences deriving a possible change in the protein sequence and that finally can cause negative effects. In the case of membrane proteins, 90% are affected by a mutation affecting their folding, stability, and function, leading to various diseases in

the worst cases. Even knowing this, the growth of technologies for DNA sequencing and SNP allele genotyping brought a large amount of information about mutations. Considering the number of membrane proteins in humans (5190), MutHTP provides information about mutations of the membrane proteins and that describes 183.395 are disease variants in 4797 proteins, related to diseases such as cancer, and neurodegenerative and cerebrovascular diseases, they are the most frequent causes of death and disability worldwide with an important clinical and socio-economic impact. Considering the objective of this thesis, there are two strong approaches to improve the design of therapeutic strategies: 1) sequence-based approaches to understand the implications of SNPs, 2) approaches to improving computational structural models of membrane proteins.

Going by parts, first to understand the implications of mutations is crucial to obtain new data from the current methodologies to comprehend the causing effect and to develop potential therapies. The result of these therapies is to minimize the malfunctionality of the protein or reverse the activity to a normal state. The next-generation sequencing methods have led to the obtaining of new mutations, but sometimes there is not enough to obtain all the information. In the absence of experimental information on specific mutations, predictive methods, such as machine learning, may be of help. Although, for a prediction to be relevant, it needs a strong pillar of data that supports the prediction. This connection is like a dog biting her tail. There are a lot of predictors related to the prediction of the pathogenesis of a mutation. Most predictors in the market are based on evolutionary conservation and expected impact on structure, function, and physicochemistry properties of amino acids from sequence data, such as SIFT, Provean, MutationTaster. Others, such as Polyphen-2, also incorporate features related to structural data, but none of them are specific for membrane proteins. To improve the prediction power of the actual methods, we developed TMSNP, which is both a searchable database of reported pathogenic and non-pathogenic missense mutations in the transmembrane segments of membrane proteins and a tool able to predict pathogenicity for previously non-reported transmembrane missense mutations collected in the database.

As I commented, the technologies to identify mutations are growing, getting better and more specific, also in variants related to rare diseases. This fact has resulted in the creation of a lot of different databases with the associated problem that is the possibility of fragmented information, that may lead to duplications, incomplete information, wrong annotations, or interpretations. One interesting article, titled, "Yet another database" (den Dunnen, 2018), discussed the decision to create a new database or collaborate with another database and store the collected information in that database. The article states that the over creation of databases is a problem. In the specific case of the effects of mutations, the dispersion of all available functional, clinical, and structural annotations of mutations hinders the evaluation of the grade of pathogenesis, a process called stratification of data, to help in the therapeutic design strategy. Related with the mutations on NMDA receptors, where the sequencing is also are in a exponentially growing state, the GRIN variant information have the annotations about clinical, genetic and functional information highly fragmented into different databases, such as Uniprot, Gnomad, ClinVar, LOVD, among others, where some data is contradictory between databases, representing a limitation in GRD patient's stratification. The stratification of these variants helps the researchers to identify if the receptors suffer an overstimulation (GoF) or a reduction (LoF) of their capabilities. Therefore, all of this available information checked is crucial to decide the type treatment and the selection of specific drugs to give to a patient, in order to revert the activity of the receptor to a normal state or the closest possible. For this reason, we developed GRINdb, a public manually curated non-redundant database comprising genetic, clinical, functional, and structural information of the largest repertoire of GRIN variants, with more than 4400 unique GRIN variants. From a clinical angle, the comprehensive data compiled in the GRINdb represents a valuable bioinformatic resource shortening the gap between the growing number of GRIN variants identification and patient stratification. This possibility of collecting all the data into a unique database opens new possibilities of discovery related with the acceleration of GRIN missense variants stratification linked with a therapeutic decision for neurodevelopmental disorders. The combination with the experimental validation of the variants represents a potential resource to stratify specific types of mutations, as

truncations, or the possibility to predict new variants using the homologues mutations that have the same grade of pathogenicity. In this prediction, the idea is to use mutations at specific positions in the sequence of one subunit and translate it to an equivalent position in the sequence of another subunit, such as homologous mutations having the same degree of pathogenicity. This allows us to map more variants from some subunits to others and amplify the information of these changes to help in the therapeutic strategy, allow us to map more variants from some subunits to others and amplify the information of these changes to help in the therapeutic strategy.

The second part of the thesis deals with improving computational structural models of membrane proteins. To achieve high quality reference structures, part of this thesis is focused on to offer this possibility, using as reference the GPCRs. As I describe in the introduction of this thesis, internal water molecules have a crucial role in the stability of the structure of a protein. One of the problems in the experimental isolation of the membrane proteins is that some important elements that participate in the functionality of the protein, such as water molecules, are lost. To find the best model for GPCRs, an interesting topic is the incorporation of these lost internal water molecules. For this reason, we developed Homolwat, to improve a prediction method to incorporate water molecules using homology structures of GPCRs. The web tool can incorporate water molecules into a protein structure in about a minute with around 85% of water recovery on our test set. The fact that HomolWat employs experimental knowledge to perform molecular modeling gives a clear competitive advantage to our tool compared to the methods based on energy calculations, making HomolWat unique. Because of that HomolWat has been successfully applied as part of the pipeline used in the GPCRmd project, a community-driven effort to create the first open, interactive and standardized database of GPCR molecular dynamics simulations (Rodríguez-Espigares et al., 2020). Due to the foreseeable increase in the number of resolved high-resolution GPCR structures, HomolWat will be able to use more templates, hence increasing its current performance. Better GPCR models that explicitly introduce internal water molecules may for instance improve docking calculations or molecular dynamics simulations.

It is accepted that GPCRs can interact between them forming oligomeric complexes of high-order that conduct to a new functionality compared with the function when they act as monomeric structures. These complexes have a real pharmacological potential, but the limited information available on the type of complex and its quaternary structures have been described as a challenge for the researchers that study this field. The conditions used to determine protein structures make in membrane proteins the quaternary structure very difficult to obtain. Hence, facilitating creating models is a small step that combined with new experiments can help characterize the membrane proteins (Navarro et al., 2016, 2018) and consequently help in the understanding of these proteins and in the design of new drugs for personalized medicine.

Today, the information of these complexes is growing thanks to Bioinformatics and several GPCR structures in the Protein Data Bank that contain potentially data compatible with physiological interactions to extract the data about these complexes. In consequence, some computational scientists, molecular biologists, and crystallographers are interested in GPCR dimers and oligomers trying to find the best possible structural template to model GPCR homomers or heteromers or to ascertain how frequent a specific dimeric interface occurs. To address this problem and provide the information about these complexes to researchers we develop DIMERBOW, an open web application for rapid, intuitive and systematic visualization of the complete repertoire of crystallographic interfaces for GPCR homodimers available in the Protein Data Bank.

This thesis opens new doors that can expand this research to other paths, hence our intention is to implement the technology created, using NMDAR and GPCR as a reference, to the rest of membrane proteins. It may be interesting to incorporate structural information TMSNP to predict pathogenicity from another point of view. A recent new computational method, named AlphaFold (Jumper et al., 2021), can predict protein structures with atomic resolution even in cases where there is no reference. This method uses machine learning to incorporate physical and biological knowledge about protein structure, taking advantage of multiple sequence alignments, into the design of the algorithm. They have demonstrated and validated their method

demonstrating that they are competitive up to an experimental level in most cases and are ahead of other methods. Therefore, it would be worth considering applying this method on some of our fronts in future projects to further improve the applied methodology. Perhaps the next step after structural waters and quaternary structure is the generation of models of large complexes involving many elements that are easy to analyze. Due to today's rapidly growing computational capacity, there are difficulties in the analysis of large complexes, such as GPCRs and G-protein. There are few complete models, and they are often separated into modules to be studied, making it difficult to see the overall picture. In the field of oligomerization, mutations may be a point of view to be considered. It may be possible that mutations in different membrane proteins, which on their own have no adverse effect. Being part of a complex, where interactions are crucial for the functioning of the complex, maybe of some relevance to the mutations as a whole, difficult the oligomerization.

For now, they are just hypotheses that will have to be postponed for future projects.

Chapter 6 - Conclusions

Chapter 6 - Conclusions

1) We use Bioinformatics tools to generate applications useful in the field of molecular modelling and the analysis of the data has made great advances in the classifying of mutations in membrane proteins and specifically in GPCRs and NMDARs.
2) We have constructed a public searchable database (TMSNP) of reported pathogenic and non-pathogenic mutations in TM segments of membrane proteins.
3) We have developed a mutation prediction tool able to predict pathogenicity for previously non-reported missense mutations in the TM region of membrane proteins.
4) We have created a non-redundant database (GRINdb), comprising genetic, clinical, functional, and structural information of the largest repertoire of GRIN variants, with more than 4400 unique GRIN variants entries.
5) We have performed the analysis of more than 4000 missense GRIN variants from GRINdb, revealing their weights associated with neurological conditions in a gene, subunit, and region-dependent manner.
6) We have developed a web application (HomolWat) to introduce internal water molecules in GPCR structures using resolved water molecules from homologous structures.
7) We have developed a web application (DIMERBOW) for systematic visualization of the complete repertoire of crystallographic interfaces for GPCR homodimers available in the PDB to be used to better understand the quaternary structure of GPCR complexes.

Bibliography

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74.

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. In *SoftwareX* (Vols. 1–2, pp. 19–25). https://doi.org/10.1016/j.softx.2015.06.001

Alberts, B., Hopkin, K., Johnson, A. D., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2018). *Essential Cell Biology: Fifth International Student Edition*. W.W. Norton & Company.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell*. Garland Science.

Almeida, J. G., Preto, A. J., Koukos, P. I., Bonvin, A. M. J. J., & Moreira, I. S. (2017). Membrane proteins structures: A review on computational modeling tools. *Biochimica et Biophysica Acta, Biomembranes*, *1859*(10), 2021–2039.

Almén, M. S., Nordström, K. J. V., Fredriksson, R., & Schiöth, H. B. (2009). Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biology*, *7*, 50.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, *181*(4096), 223–230.

Attwood, T. K., & Findlay, J. B. (1994). Fingerprinting G-protein-coupled receptors. *Protein Engineering*, *7*(2), 195–203.

Auvin, S., Dozières-Puyravel, B., Avbersek, A., Sciberras, D., Collier, J., Leclercq, K., Mares, P., Kaminski, R. M., & Muglia, P. (2020). Radiprodil, a NR2B negative allosteric modulator, from bench to bedside in infantile spasm syndrome. *Annals of Clinical and Translational Neurology*, *7*(3), 343–352.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Wheeler, D. L. (2008). GenBank. *Nucleic Acids Research*, *36*(Database issue), D25–D30.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242.

Biological Macromolecules. (2020, August 15). Libretexts. https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology/Book%3A_General_Biology_(Boundless)/3%3A_Biological_Macromolecules

Biotechnology and its Applications. (2022). https://doi.org/10.1016/c2018-0-00161-9

Bitencourt-Ferreira, G., Veit-Acosta, M., & de Azevedo, W. F., Jr. (2019). Van der Waals Potential in Protein Complexes. *Methods in Molecular Biology* , *2053*, 79–91.

Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., & Zaretskaya, I. (2013). BLAST: a more efficient report with usability improvements. *Nucleic Acids Research*, *41*(W1), W29–W33.

Buermans, H. P. J., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta*, *1842*(10), 1932–1941.

Burnashev, N., & Szepetowski, P. (2015). NMDA receptor subunit mutations in neurodevelopmental disorders. *Current Opinion in Pharmacology*, *20*, 73–82.

Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis*, *58*(5), 235–263.

Chatterton, J. E., Awobuluyi, M., Premkumar, L. S., Takahashi, H., Talantova, M., Shin, Y., Cui, J., Tu, S., Sevarino, K. A., Nakanishi, N., Tong, G., Lipton, S. A., & Zhang, D. (2002). Excitatory glycine receptors containing the NR3 family of NMDA receptor subunits. *Nature*, *415*(6873), 793–798.

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. In *Bioinformatics* (Vol. 31, Issue 16, pp. 2745–2747). https://doi.org/10.1093/bioinformatics/btv195

Chou, T.-H., Tajima, N., Romero-Hernandez, A., & Furukawa, H. (2020). Structural Basis of Functional Transitions in Mammalian NMDA Receptors. *Cell*, *182*(2), 357–371.e13.

Clancy, S. (2008). *Genetic Mutation*. Scitable. https://www.nature.com/scitable/topicpage/genetic-mutation-441/

Cobb, M. (2017). 60 years ago, Francis Crick changed the logic of biology. *PLoS Biology*, *15*(9), e2003243.

Cournia, Z., Allen, T. W., Andricioaei, I., Antonny, B., Baum, D., Brannigan, G., Buchete, N.-V., Deckman, J. T., Delemotte, L., Del Val, C., Friedman, R., Gkeka, P., Hege, H.-C., Hénin, J., Kasimova, M. A., Kolocouris, A., Klein, M. L., Khalid, S., Lemieux, M. J., … Bondar, A.-N. (2015). Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory. *The Journal of Membrane Biology*, *248*(4), 611–640.

Crick, F. (1970). Central Dogma of Molecular Biology. In *Nature* (Vol. 227, Issue 5258, pp. 561–563). https://doi.org/10.1038/227561a0

Cummings, B. B., Karczewski, K. J., Kosmicki, J. A., Seaby, E. G., Watts, N. A., Singer-Berk, M., Mudge, J. M., Karjalainen, J., Satterstrom, F. K., O'Donnell-Luria, A. H., Poterba, T., Seed, C., Solomonson, M., Alföldi, J., Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, Daly, M. J., & MacArthur, D. G. (2020). Transcript expression-aware annotation improves rare variant interpretation. *Nature*, *581*(7809), 452–458.

de Ligt, J., Willemsen, M. H., van Bon, B. W. M., Kleefstra, T., Yntema, H. G., Kroes, T., Vulto-van Silfhout, A. T., Koolen, D. A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B. B. A., Brunner, H. G., Veltman, J. A., & Vissers, L. E. L. M. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *The New England Journal of Medicine*, *367*(20), 1921–1929.

den Dunnen, J. T. (2018). Yet another database? *Human Mutation*, *39*(6), 755.

Dwivedi, H., Baidya, M., & Shukla, A. K. (2018). GPCR Signaling: The Interplay of Gαi and β-arrestin [Review of *GPCR Signaling: The Interplay of Gαi and β-arrestin*]. *Current Biology: CB*, *28*(7), R324–R327.

Eisenberg, D. (2003). The discovery of the -helix and -sheet, the principal structural features of proteins. In *Proceedings of the National Academy of Sciences* (Vol. 100, Issue 20, pp. 11207–11210). https://doi.org/10.1073/pnas.2034522100

Ekman, D., Björklund, A. K., Frey-Skött, J., & Elofsson, A. (2005). Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *Journal of Molecular Biology*, *348*(1), 231–243.

Escuer, E. M. (2019). *Development of Bioinformatic Tools for the Study of Membrane Proteins*.

Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M., & Berglund, L. (2010). Prediction of the human membrane proteome. *Proteomics*, *10*(6), 1141–1149.

Fedele, L., Newcombe, J., Topf, M., Gibb, A., Harvey, R. J., & Smart, T. G. (2018). Disease-associated missense mutations in GluN2B subunit alter NMDA receptor ligand binding and ion channel properties. *Nature Communications*, *9*(1), 957.

Ferré, S. (2015). The GPCR heterotetramer: challenging classical pharmacology. *Trends in Pharmacological Sciences*, 1–8.

Ferré, S., Casadó, V., Devi, L. a., Filizola, M., Jockers, R., Lohse, M. J., Milligan, G., Pin, J.-P., & Guitart, X. (2014). G protein-coupled receptor oligomerization revisited: functional and pharmacological perspectives. *Pharmacological Reviews*, *66*(2), 413–434.

Flores-Soto, M. E., Chaparro-Huerta, V., Escoto-Delgadillo, M., Vazquez-Valls, E., González-Castañeda, R. E., & Beas-Zarate, C. (2012). [Structure and function of NMDA-type glutamate receptor subunits]. *Neurologia* , *27*(5), 301–310.

Fokkema, I. F. A. C., Kroon, M., López Hernández, J. A., Asscheman, D., Lugtenburg, I., Hoogenboom, J., & den Dunnen, J. T. (2021). The LOVD3 platform: efficient genome-wide sharing of genetic variants. *European Journal of Human Genetics: EJHG*. https://doi.org/10.1038/s41431-021-00959-x

Franco, R., Martínez-Pinilla, E., Lanciego, J. L., & Navarro, G. (2016). Basic Pharmacological and Structural Evidence for Class A G-Protein-Coupled Receptor Heteromerization. *Frontiers in Pharmacology*, *7*, 76.

Frank, R. A., & Grant, S. G. (2017). Supramolecular organization of NMDA receptors and the postsynaptic density. *Current Opinion in Neurobiology*, *45*, 139–147.

Fredriksson, R., Lagerström, M. C., Lundin, L.-G., & Schiöth, H. B. (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Molecular Pharmacology*, *63*(6), 1256–1272.

Fuxe, K., Marcellino, D., Borroto-Escuela, D. O., Frankowska, M., Ferraro, L., Guidolin, D., Ciruela, F., & Agnati, L. F. (2010). The changing world of G protein-coupled receptors: from monomers to dimers and receptor mosaics with allosteric receptor-receptor interactions. *Journal of Receptor and Signal Transduction Research*, *30*(5), 272–283.

Garcia-Recio, A., Gómez-Tamayo, J. C., Reina, I., Campillo, M., Cordomí, A., & Olivella, M. (2021). TMSNP: a web server to predict pathogenesis of missense mutations in the transmembrane region of membrane proteins. *NAR Genomics and Bioinformatics*, *3*(1), lqab008.

García-Recio, A., Navarro, G., Franco, R., Olivella, M., Guixà-González, R., & Cordomí, A. (2020). DIMERBOW: exploring possible GPCR dimer interfaces. *Bioinformatics* , *36*(10), 3271–3272.

García-Recio, A., Santos-Gómez, A., Soto, D., Julia-Palacios, N., García-Cazorla, À., Altafaj, X., & Olivella, M. (2021). GRIN database: A unified and manually curated repertoire of GRIN variants. *Human Mutation*, *42*(1), 8–

18.

Gerle, C. (2019). Essay on Biomembrane Structure. *The Journal of Membrane Biology*, *252*(2-3), 115–130.

Gloriam, D. E., Fredriksson, R., & Schiöth, H. B. (2007). The G protein-coupled receptor subset of the rat genome. *BMC Genomics*, *8*, 338.

Gong, J., Chen, Y., Pu, F., Sun, P., He, F., Zhang, L., Li, Y., Ma, Z., & Wang, H. (2019). Understanding Membrane Protein Drug Targets in Computational Perspective. *Current Drug Targets*, *20*(5), 551–564.

Gordon Betts, J., OpenStax College, Desaix, P., Johnson, J. E., Johnson, E. W., Korol, O., Kruse, D., Poe, B., Wise, J., Womble, M. D., & Young, K. A. (2013). *Anatomy and Physiology*.

GPCRdb. (2021). *GPCRdb* [Png]. GPCRdb. https://gpcrdb.org/structure/statistics

Gromiha, M. M., & Ou, Y.-Y. (2014). Bioinformatics approaches for functional annotation of membrane proteins. In *Briefings in Bioinformatics* (Vol. 15, Issue 2, pp. 155–168). https://doi.org/10.1093/bib/bbt015

Grouleff, J., Irudayam, S. J., Skeby, K. K., & Schiøtt, B. (2015). The influence of cholesterol on membrane protein structure, function, and dynamics studied by molecular dynamics simulations. *Biochimica et Biophysica Acta*, *1848*(9), 1783–1795.

Gudmundsson, S., Singer-Berk, M., Watts, N. A., Phu, W., Goodrich, J. K., Solomonson, M., Consortium, Genome Aggregation Database, Rehm, H.

L., MacArthur, D. G., & ODonnell-Luria, A. (2021). *Variant interpretation using population databases: lessons from gnomAD*. http://arxiv.org/abs/2107.11458

Guerram, M., Zhang, L.-Y., & Jiang, Z.-Z. (2016). G-protein coupled receptors as therapeutic targets for neurodegenerative and cerebrovascular diseases. *Neurochemistry International*, *101*, 1–14.

Guidolin, D., Marcoli, M., Tortorella, C., Maura, G., & Agnati, L. F. (2018). G protein-coupled receptor-receptor interactions give integrative dynamics to intercellular communication. *Reviews in the Neurosciences*, *29*(7), 703–726.

Guo, H., An, S., Ward, R., Yang, Y., Liu, Y., Guo, X.-X., Hao, Q., & Xu, T.-R. (2017). Methods used to study the oligomeric structure of G-protein-coupled receptors. *Bioscience Reports*, *37*(2). https://doi.org/10.1042/BSR20160547

Hamosh, A., Scott, A. F., Amberger, J., Valle, D., & McKusick, V. A. (2000). Online Mendelian Inheritance in Man (OMIM). *Human Mutation*, *15*(1), 57–61.

Han, Y., Moreira, I. S., Urizar, E., Weinstein, H., & Javitch, J. A. (2009). Allosteric communication between protomers of dopamine class A GPCR dimers modulates activation. In *Nature Chemical Biology* (Vol. 5, Issue 9, pp. 688–695). https://doi.org/10.1038/nchembio.199

Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B., & Gloriam, D. E. (2017). Trends in GPCR drug discovery: new agents, targets and

indications. *Nature Reviews. Drug Discovery*, *16*(12), 829–842.

Higgins DG, S. P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, *73*(1), 237–244.

Hilger, D., Masureel, M., & Kobilka, B. K. (2018). Structure and dynamics of GPCR signaling complexes. *Nature Structural & Molecular Biology*, *25*(1), 4–12.

Horn, F., Weare, J., Beukers, M. W., Hörsch, S., Bairoch, A., Chen, W., Edvardsen, O., Campagne, F., & Vriend, G. (1998). GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Research*, *26*(1), 275–279.

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., El Houdaigui, B., Fatima, R., … Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*, *49*(D1), D884–D891.

Huang, D., Yi, X., Zhou, Y., Yao, H., Xu, H., Wang, J., Zhang, S., Nong, W., Wang, P., Shi, L., Xuan, C., Li, M., Wang, J., Li, W., Kwan, H. S., Sham, P. C., Wang, K., & Li, M. J. (2020). Ultrafast and scalable variant annotation and prioritization with big functional genomics data. *Genome Research*, *30*(12), 1789–1801.

Huang, M., Shah, N. D., & Yao, L. (2019). Evaluating global and local sequence alignment methods for comparing patient medical records. *BMC Medical*

Informatics and Decision Making, *19*(Suppl 6), 263.

Hu, G.-M., Mai, T.-L., & Chen, C.-M. (2017). Visualizing the GPCR Network: Classification and Evolution. *Scientific Reports*, *7*(1), 15495.

Iliyas, S., & Ansari, S. N. (2013). *A COMPARATIVE STUDY OF DIFFERENT PROPERTIES PROVIDED BY PROTEIN STRUCTURE VISUALIZATION TOOLS*. *3*(1), 2277–20731.

Jastrzebska, B., & Park, P. S.-H. (2019). *GPCRs: Structure, Function, and Drug Discovery*. Academic Press.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, *36*(Web Server issue), W5–W9.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589.

Karakas, E., & Furukawa, H. (2014). Crystal structure of a heterotetrameric NMDA receptor ion channel. *Science*, *344*(6187), 992–997.

Katritch, V., Cherezov, V., & Stevens, R. C. (2013). Structure-function of the G protein-coupled receptor superfamily. *Annual Review of Pharmacology and Toxicology*, *53*, 531–556.

Keministi. (2019). *Cell membrane bound N1/N2/N1/N2 NMDA receptor* [Svg]. https://commons.wikimedia.org/wiki/File:N1_N2_NMDA_receptor.svg

Koch, L. (2020). Exploring human genomic diversity with gnomAD. *Nature Reviews. Genetics*, *21*(8), 448–448.

Kodama, Y., Shumway, M., Leinonen, R., & International Nucleotide Sequence Database Collaboration. (2012). The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Research*, *40*(Database issue), D54–D56.

Kolakowski, L. F., Jr. (1994). GCRDb: a G-protein-coupled receptor database. *Receptors & Channels*, *2*(1), 1–7.

Kooistra, A. J., Mordalski, S., Pándy-Szekeres, G., Esguerra, M., Mamyrbekov, A., Munk, C., Keserű, G. M., & Gloriam, D. E. (2021). GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Research*, *49*(D1), D335–D343.

Korinek, M., Vyklicky, V., Borovska, J., Lichnerova, K., Kaniakova, M., Krausova, B., Krusek, J., Balik, A., Smejkalova, T., Horak, M., & Vyklicky, L. (2015). Cholesterol modulates open probability and desensitization of NMDA receptors. *The Journal of Physiology*, *593*(10), 2279–2293.

Kuhlman, D. (2011). *A Python Book: Beginning Python, Advanced Python, and Python Exercises*. Platypus Global Media.

Kulandaisamy, A., Binny Priya, S., Sakthivel, R., Tarnovskaya, S., Bizin, I., Hönigschmid, P., Frishman, D., & Gromiha, M. M. (2018). MutHTP:

mutations in human transmembrane proteins. *Bioinformatics* , *34*(13), 2325–2326.

Kulandaisamy, A., Priya, S. B., Sakthivel, R., Frishman, D., & Gromiha, M. M. (2019). Statistical analysis of disease-causing and neutral mutations in human membrane proteins. *Proteins*, *87*(6), 452–466.

Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., Hoffman, D., Jang, W., Kaur, K., Liu, C., Lyoshin, V., Maddipatla, Z., Maiti, R., Mitchell, J., O'Leary, N., Riley, G. R., Shi, W., Zhou, G., Schneider, V., … Kattman, B. L. (2020). ClinVar: improvements to accessing data. *Nucleic Acids Research*, *48*(D1). https://doi.org/10.1093/nar/gkz972

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, *42*(Database issue), D980–D985.

Lee, C.-H., Lü, W., Michel, J. C., Goehring, A., Du, J., Song, X., & Gouaux, E. (2014). NMDA receptor structures reveal subunit arrangement and pore architecture. *Nature*, *511*(7508), 191–197.

Lemke, J. R., Geider, K., Helbig, K. L., Heyne, H. O., Schütz, H., Hentschel, J., Courage, C., Depienne, C., Nava, C., Heron, D., Møller, R. S., Hjalgrim, H., Lal, D., Neubauer, B. A., Nürnberg, P., Thiele, H., Kurlemann, G., Arnold, G. L., Bhambhani, V., … Syrbe, S. (2016). Delineating the GRIN1 phenotypic spectrum: A distinct genetic NMDA receptor encephalopathy.

Neurology, *86*(23), 2171–2178.

Lemke, J. R., Hendrickx, R., Geider, K., Laube, B., Schwake, M., Harvey, R. J., James, V. M., Pepler, A., Steiner, I., Hörtnagel, K., Neidhardt, J., Ruf, S., Wolff, M., Bartholdi, D., Caraballo, R., Platzer, K., Suls, A., De Jonghe, P., Biskup, S., & Weckhuysen, S. (2014). GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. *Annals of Neurology*, *75*(1), 147–154.

Lesca, G., Rudolf, G., Bruneau, N., Lozovaya, N., Labalme, A., Boutry-Kryza, N., Salmi, M., Tsintsadze, T., Addis, L., Motte, J., Wright, S., Tsintsadze, V., Michel, A., Doummar, D., Lascelles, K., Strug, L., Waters, P., de Bellescize, J., Vrielynck, P., … Szepetowski, P. (2013). GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nature Genetics*, *45*(9), 1061–1066.

Lomize, M. A., Lomize, A. L., Pogozheva, I. D., & Mosberg, H. I. (2006). OPM: Orientations of Proteins in Membranes database. *Bioinformatics* , *22*(5), 623–625.

Lussier, M. P., Sanz-Clemente, A., & Roche, K. W. (2015). Dynamic Regulation of N-Methyl-d-aspartate (NMDA) and α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic Acid (AMPA) Receptors by Posttranslational Modifications. *The Journal of Biological Chemistry*, *290*(48), 28596–28603.

Madry, C., Mesic, I., Bartholomäus, I., Nicke, A., Betz, H., & Laube, B. (2007). Principal role of NR3 subunits in NR1/NR3 excitatory glycine receptor

function. *Biochemical and Biophysical Research Communications*, *354*(1), 102–108.

Maurice, P., Guillaume, J.-L., Benleulmi-Chaachoua, A., Daulat, A. M., Kamal, M., & Jockers, R. (2011). GPCR-interacting proteins, major players of GPCR function. *Advances in Pharmacology*, *62*, 349–380.

Mayer, M. L., Westbrook, G. L., & Guthrie, P. B. (1984). Voltage-dependent block by Mg2+ of NMDA responses in spinal cord neurones. *Nature*, *309*(5965), 261–263.

Mayol, E., Campillo, M., Cordomí, A., & Olivella, M. (2019). Inter-residue interactions in alpha-helical transmembrane proteins. *Bioinformatics* , *35*(15), 2578–2584.

McNeill, L. (2019, April 9). *How Margaret Dayhoff Brought Modern Computing to Biology*. Smithsonian Magazine. https://www.smithsonianmag.com/science-nature/how-margaret-dayhoff-helped-bring-computing-scientific-research-180971904/

Meng, X., Clews, J., Kargas, V., Wang, X., & Ford, R. C. (2017). The cystic fibrosis transmembrane conductance regulator (CFTR) and its stability. *Cellular and Molecular Life Sciences: CMLS*, *74*(1), 23–38.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419.

Monaghan, D. T., & Jane, D. E. (2011). Pharmacology of NMDA Receptors. In A. M. Van Dongen (Ed.), *Biology of the NMDA Receptor*. CRC Press/Taylor & Francis.

Mony, L., Zhu, S., Carvalho, S., & Paoletti, P. (2011). Molecular basis of positive allosteric modulation of GluN2B NMDA receptors by polyamines. *The EMBO Journal*, *30*(15), 3134–3146.

Mottaz, A., David, F. P. A., Veuthey, A.-L., & Yip, Y. L. (2010). Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. *Bioinformatics* , *26*(6), 851–852.

Mullier, B., Wolff, C., Sands, Z. A., Ghisdal, P., Muglia, P., Kaminski, R. M., & André, V. M. (2017). GRIN2B gain of function mutations are sensitive to radiprodil, a negative allosteric modulator of GluN2B-containing NMDA receptors. *Neuropharmacology*, *123*, 322–331.

Nature Education. (2014). *Contents of Essentials of Cell Biology*. Scitable. https://www.nature.com/scitable/ebooks/essentials-of-cell-biology-14749010/122996980/

Navarro, G., Cordomí, A., Casadó-Anguera, V., Moreno, E., Cai, N.-S., Cortés, A., Canela, E. I., Dessauer, C. W., Casadó, V., Pardo, L., Lluís, C., & Ferré, S. (2018). Evidence for functional pre-coupled complexes of receptor heteromers and adenylyl cyclase. *Nature Communications*, *9*(1), 1242.

Navarro, G., Cordomí, A., Zelman-Femiak, M., Brugarolas, M., Moreno, E., Aguinaga, D., Perez-Benito, L., Cortés, A., Casadó, V., Mallol, J., Canela,

E. I., Lluís, C., Pardo, L., García-Sáez, A. J., McCormick, P. J., & Franco, R. (2016). Quaternary structure of a G-protein-coupled receptor heterotetramer in complex with Gi and Gs. *BMC Biology*, *14*, 26.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, *48*(3), 443–453.

Nicolson, G. L. (2014). The Fluid-Mosaic Model of Membrane Structure: still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years. *Biochimica et Biophysica Acta*, *1838*(6), 1451–1466.

Nowak, L., Bregestovski, P., Ascher, P., Herbet, A., & Prochiantz, A. (1984). Magnesium gates glutamate-activated channels in mouse central neurones. *Nature*, *307*(5950), 462–465.

Nuin, P. A. S., Wang, Z., & Tillier, E. R. M. (2006). The accuracy of several multiple sequence alignment programs for proteins. *BMC Bioinformatics*, *7*, 471.

Nussinov, R., & Tsai, C.-J. (2012). The different ways through which specificity works in orthosteric and allosteric drugs. *Current Pharmaceutical Design*, *18*(9), 1311–1316.

Ogden, K. K., Chen, W., Swanger, S. A., McDaniel, M. J., Fan, L. Z., Hu, C., Tankovic, A., Kusumoto, H., Kosobucki, G. J., Schulien, A. J., Su, Z., Pecha, J., Bhattacharya, S., Petrovski, S., Cohen, A. E., Aizenman, E.,

Traynelis, S. F., & Yuan, H. (2017). Molecular Mechanism of Disease-Associated Mutations in the Pre-M1 Helix of NMDA Receptors and Potential Rescue Pharmacology. *PLoS Genetics*, *13*(1), e1006536.

Olivella, M., Gonzalez, A., Pardo, L., & Deupi, X. (2013). Relation between sequence and structure in membrane proteins. *Bioinformatics* , *29*(13), 1589–1592.

O'Roak, B. J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J. J., Girirajan, S., Karakoc, E., Mackenzie, A. P., Ng, S. B., Baker, C., Rieder, M. J., Nickerson, D. A., Bernier, R., Fisher, S. E., Shendure, J., & Eichler, E. E. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genetics*, *43*(6), 585–589.

Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there? *Nature Reviews. Drug Discovery*, *5*(12), 993–996.

Paoletti, P., Bellone, C., & Zhou, Q. (2013). NMDA receptor subunit diversity: impact on receptor properties, synaptic plasticity and disease. *Nature Reviews. Neuroscience*, *14*(6), 383–400.

Pasek, S., Risler, J.-L., & Brézellec, P. (2006). Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* , *22*(12), 1418–1423.

Pauling, L., & Corey, R. B. (1951). The pleated sheet, a new layer configuration of polypeptide chains. *Proceedings of the National Academy of Sciences of the United States of America*, *37*(5), 251–256.

Pauling, L., Corey, R. B., & Branson, H. R. (1951). The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, *37*(4), 205–211.

Pérez-Otaño, I., Larsen, R. S., & Wesseling, J. F. (2016). Emerging roles of GluN3-containing NMDA receptors in the CNS. *Nature Reviews. Neuroscience*, *17*(10), 623–635.

Pertea, M., Shumate, A., Pertea, G., Varabyou, A., Breitwieser, F. P., Chang, Y.-C., Madugundu, A. K., Pandey, A., & Salzberg, S. L. (2018). CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biology*, *19*(1), 208.

Petsko, G. A., & Ringe, D. (2004). *Protein Structure and Function*. New Science Press.

Pierson, T. M., Yuan, H., Marsh, E. D., Fuentes-Fajardo, K., Adams, D. R., Markello, T., Golas, G., Simeonov, D. R., Holloman, C., Tankovic, A., Karamchandani, M. M., Schreiber, J. M., Mullikin, J. C., PhD for the NISC Comparative Sequencing Program, Tifft, C. J., Toro, C., Boerkoel, C. F., Traynelis, S. F., & Gahl, W. A. (2014). mutation and early-onset epileptic encephalopathy: personalized therapy with memantine. *Annals of Clinical and Translational Neurology*, *1*(3), 190–198.

Pirovano, W., & Heringa, J. (2008). Multiple sequence alignment. *Methods in*

Molecular Biology , *452*, 143–161.

Platzer, K., Yuan, H., Schütz, H., Winschel, A., Chen, W., Hu, C., Kusumoto, H., Heyne, H. O., Helbig, K. L., Tang, S., Willing, M. C., Tinkle, B. T., Adams, D. J., Depienne, C., Keren, B., Mignot, C., Frengen, E., Strømme, P., Biskup, S., … Lemke, J. R. (2017). encephalopathy: novel findings on phenotype, variant clustering, functional consequences and treatment aspects. *Journal of Medical Genetics*, *54*(7), 460–470.

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, *35*(Database issue), D61–D65.

PubMed. (2021). https://pubmed.ncbi.nlm.nih.gov/about/

PYPL PopularitY of Programming Language. (2021). https://pypl.github.io/PYPL.html

Rajagopal, K. A., Indira, & Tan, T. (2021, March 6). *Structure & Function-Proteins I*. Libretexts. https://bio.libretexts.org/Bookshelves/Biochemistry/Book%3A_Biochemistry_Free_For_All_(Ahern_Rajagopal_and_Tan)/02%3A_Structure_and_Function/203%3A_Structure__Function-_Proteins_I

Rao, A., & Craig, A. M. (1997). Activity regulates the synaptic localization of the NMDA receptor in hippocampal neurons. *Neuron*, *19*(4), 801–812.

Reynolds, K. A., Russ, W. P., Socolich, M., & Ranganathan, R. (2013).

Evolution-based design of proteins. *Methods in Enzymology*, *523*, 213–235.

Richardson, T. G., Campbell, C., Timpson, N. J., & Gaunt, T. R. (2016). Incorporating Non-Coding Annotations into Rare Variant Analysis. *PloS One*, *11*(4), e0154181.

Rodríguez-Espigares, I., Torrens-Fontanals, M., Tiemann, J. K. S., Aranda-García, D., Ramírez-Anguita, J. M., Stepniewski, T. M., Worp, N., Varela-Rial, A., Morales-Pastor, A., Medel-Lacruz, B., Pándy-Szekeres, G., Mayol, E., Giorgino, T., Carlsson, J., Deupi, X., Filipek, S., Filizola, M., Gómez-Tamayo, J. C., Gonzalez, A., … Selent, J. (2020). GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nature Methods*, *17*(8), 777–787.

Ryslik, G. A., Cheng, Y., Modis, Y., & Zhao, H. (2016). Leveraging protein quaternary structure to identify oncogenic driver mutations. *BMC Bioinformatics*, *17*, 137.

Sadowski, M. I., & Jones, D. T. (2009). The sequence-structure relationship and protein function prediction. *Current Opinion in Structural Biology*, *19*(3), 357–362.

Salzberg, S. L. (2018). Open questions: How many genes do we have? *BMC Biology*, *16*(1), 94.

Santos-Gómez, A., Miguez-Cabello, F., García-Recio, A., Locubiche-Serra, S., García-Díaz, R., Soto-Insuga, V., Guerrero-López, R., Juliá-Palacios, N., Ciruela, F., García-Cazorla, À., Soto, D., Olivella, M., & Altafaj, X. (2021). Disease-associated GRIN protein truncating variants trigger NMDA receptor

loss-of-function. *Human Molecular Genetics*, *29*(24), 3859–3871.

Santos, R., Ursu, O., Gaulton, A., Bento, A. P., Donadi, R. S., Bologa, C. G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T. I., & Overington, J. P. (2017). A comprehensive map of molecular drug targets. *Nature Reviews. Drug Discovery*, *16*(1), 19–34.

Sanvictores, T., & Farci, F. (2020). Biochemistry, Primary Protein Structure. In *StatPearls*. StatPearls Publishing.

Schneider, J., Korshunova, K., Musiani, F., Alfonso-Prieto, M., Giorgetti, A., & Carloni, P. (2018). Predicting ligand binding poses for low-resolution membrane protein models: Perspectives from multiscale simulations. *Biochemical and Biophysical Research Communications*, *498*(2), 366–374.

Schöneberg, T., & Liebscher, I. (2021). Mutations in G Protein-Coupled Receptors: Mechanisms, Pathophysiology and Potential Therapeutic Approaches. *Pharmacological Reviews*, *73*(1), 89–119.

Schöneberg, T., Schulz, A., Biebermann, H., Hermsdorf, T., Römpler, H., & Sangkuhl, K. (2004). Mutant G-protein-coupled receptors as a cause of human diseases. *Pharmacology & Therapeutics*, *104*(3), 173–206.

Schrödinger, L. L. C. (2021). *The PyMOL Molecular Graphics System* (Version 2.5) [Linux Ubuntu]. https://pymol.org/2/#page-top

Schrum, J. P., Zhu, T. F., & Szostak, J. W. (2010). The origins of cellular life. *Cold Spring Harbor Perspectives in Biology*, *2*(9), a002212.

Schwarz, J. M., Cooper, D. N., Schuelke, M., & Seelow, D. (2014).

MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods*, *11*(4), 361–362.

Seyedabadi, M., Ghahremani, M. H., & Albert, P. R. (2019). Biased signaling of G protein coupled receptors (GPCRs): Molecular determinants of GPCR/transducer selectivity and therapeutic potential. *Pharmacology & Therapeutics*, *200*, 148–178.

Shenoy, S. R., & Jayaram, B. (2010). Proteins: sequence to structure and function--current status. *Current Protein & Peptide Science*, *11*(7), 498–514.

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, *40*(Web Server issue), W452–W457.

Singer, S. J., & Nicolson, G. L. (1972). THE FLUID MOSAIC MODEL OF THE STRUCTURE OF CELL MEMBRANES Reprinted with permission from Science, Copyright AAA, 18 February 1972, Volume 175, pp. 720–731. In *Membranes and Viruses in Immunopathology* (pp. 7–47). https://doi.org/10.1016/b978-0-12-207250-5.50008-7

Sleno, R., & Hébert, T. E. (2018). The Dynamics of GPCR Oligomerization and Their Functional Consequences. *International Review of Cell and Molecular Biology*, *338*, 141–171.

Smigielski, E. M., Sirotkin, K., Ward, M., & Sherry, S. T. (2000). dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Research*,

28(1), 352–355.

Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*, *147*(1), 195–197.

Smothers, C. T., & Woodward, J. J. (2009). Expression of glycine-activated diheteromeric NR1/NR3 receptors in human embryonic kidney 293 cells Is NR1 splice variant-dependent. *The Journal of Pharmacology and Experimental Therapeutics*, *331*(3), 975–984.

Soto, D., Olivella, M., Grau, C., Armstrong, J., Alcon, C., Gasull, X., Santos-Gómez, A., Locubiche, S., Gómez de Salazar, M., García-Díaz, R., Gratacòs-Batlle, E., Ramos-Vicente, D., Chu-Van, E., Colsch, B., Fernández-Dueñas, V., Ciruela, F., Bayés, À., Sindreu, C., López-Sala, A., … Altafaj, X. (2019). L-Serine dietary supplementation is associated with clinical improvement of loss-of-function GRIN2B-related pediatric encephalopathy. *Science Signaling*, *12*(586). https://doi.org/10.1126/scisignal.aaw0936

Sriram, K., & Insel, P. A. (2018). G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Molecular Pharmacology*, *93*(4), 251–258.

Stevens, R. C., Cherezov, V., Katritch, V., Abagyan, R., Kuhn, P., Rosen, H., & Wüthrich, K. (2013). The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. *Nature Reviews. Drug Discovery*, *12*(1), 25–34.

Stillwell, W. (2016). Membrane Transport. *An Introduction to Biological Membranes*, 423.

Sun, P. D., Foster, C. E., & Boyington, J. C. (2004). Overview of protein structural and functional folds. *Current Protocols in Protein Science / Editorial Board, John E. Coligan ... [et Al.]*, *Chapter 17*, Unit 17.1.

Suzuki, T., Araki, Y., Yamamoto, T., & Nakaya, T. (2006). Trafficking of Alzheimer's disease-related membrane proteins and its participation in disease pathogenesis. *Journal of Biochemistry*, *139*(6), 949–955.

Szigetvari, P. D., Muruganandam, G., Kallio, J. P., Hallin, E. I., Fossbakk, A., Loris, R., Kursula, I., Møller, L. B., Knappskog, P. M., Kursula, P., & Haavik, J. (2019). The quaternary structure of human tyrosine hydroxylase: effects of dystonia-associated missense variants on oligomeric state and enzyme activity. *Journal of Neurochemistry*, *148*(2), 291–306.

Tajima, N., Karakas, E., Grant, T., Simorowski, N., Diaz-Avalos, R., Grigorieff, N., & Furukawa, H. (2016). Activation of NMDA receptors and the mechanism of inhibition by ifenprodil. *Nature*, *534*(7605), 63–68.

Tantardini, C. (2019). When does a hydrogen bond become a van der Waals interaction? a topological answer. *Journal of Computational Chemistry*, *40*(8), 937–943.

Tarabeux, J., Kebir, O., Gauthier, J., Hamdan, F. F., Xiong, L., Piton, A., Spiegelman, D., Henrion, É., Millet, B., S2D team, Fathalli, F., Joober, R., Rapoport, J. L., DeLisi, L. E., Fombonne, É., Mottron, L., Forget-Dubois, N.,

Boivin, M., Michaud, J. L., … Krebs, M.-O. (2011). Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia. *Translational Psychiatry*, *1*, e55.

Tarabini, R. F., Timmers, L. F. S. M., Sequeiros-Borja, C. E., & Norberto de Souza, O. (2019). The importance of the quaternary structure to represent conformational ensembles of the major Mycobacterium tuberculosis drug target. *Scientific Reports*, *9*(1), 13683.

The State of Developer Ecosystem 2021. (2021). Jet Brains. https://www.jetbrains.com/lp/devecosystem-2021/

The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., … Teodoro, D. (2020). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489.

Trinh, J., Tadic, V., & Klein, C. (2018). How Do I Confirm that a New Mutation is Pathogenic? *Movement Disorders Clinical Practice*, *5*(2), 229.

Turro, E., Astle, W. J., Megy, K., Gräf, S., Greene, D., Shamardina, O., Allen, H. L., Sanchis-Juan, A., Frontini, M., Thys, C., Stephens, J., Mapeta, R., Burren, O. S., Downes, K., Haimel, M., Tuna, S., Deevi, S. V. V., Aitman, T. J., Bennett, D. L., … Ouwehand, W. H. (2020). Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*, *583*(7814), 96–102.

UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515.

Valdivielso, J. M., Eritja, À., Caus, M., & Bozic, M. (2020). Glutamate-Gated NMDA Receptors: Insights into the Function and Signaling in the Kidney. *Biomolecules*, *10*(7). https://doi.org/10.3390/biom10071051

Venkatakrishnan, A. J., Ma, A. K., Fonseca, R., Latorraca, N. R., Kelly, B., Betz, R. M., Asawa, C., Kobilka, B. K., & Dror, R. O. (2019). Diverse GPCRs exhibit conserved water networks for stabilization and activation. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(8), 3288–3293.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., … Zhu, X. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304–1351.

Vihinen, M. (2015). No more hidden solutions in bioinformatics. *Nature*, *521*(7552), 261.

VMD: Visual molecular dynamics. (1996). *Journal of Molecular Graphics*, *14*(1), 33–38.

Warnet, X. L., Bakke Krog, H., Sevillano-Quispe, O. G., Poulsen, H., & Kjaergaard, M. (2021). The C-terminal domains of the NMDA receptor: How intrinsically disordered tails affect signalling, plasticity and disease. *The*

European Journal of Neuroscience, *54*(8), 6713–6739.

Watson, H. (2015). Biological membranes. *Essays in Biochemistry*, *59*, 43–69.

Watson, J. D., Baker, T. A., & Bell, S. P. (2008). *Molecular Biology of the Gene*. Benjamin-Cummings Publishing Company.

Weston, M. C. (2017). and Bear the Diverse Functional Effects of Rare NMDA Receptor Variants [Review of *and Bear the Diverse Functional Effects of Rare NMDA Receptor Variants*]. *Epilepsy Currents / American Epilepsy Society*, *17*(6), 381–383.

Wheeler, D. L. (2004). Database resources of the National Center for Biotechnology Information. In *Nucleic Acids Research* (Vol. 33, Issue Database issue, pp. D39–D45). https://doi.org/10.1093/nar/gki062

Wu, C. H., Yeh, L.-S. L., Huang, H., Arminski, L., Castro-Alvear, J., Chen, Y., Hu, Z., Kourtesis, P., Ledley, R. S., Suzek, B. E., Vinayaka, C. R., Zhang, J., & Barker, W. C. (2003). The Protein Information Resource. *Nucleic Acids Research*, *31*(1), 345–347.

XiangWei, W., Jiang, Y., & Yuan, H. (2018). Mutations and Rare Variants Occurring in NMDA Receptors. *Current Opinion in Physiology*, *2*, 27–35.

Xue, L., Sun, Q., Zhao, H., Rovira, X., Gai, S., He, Q., Pin, J.-P., Liu, J., & Rondard, P. (2019). Rearrangement of the transmembrane domain interfaces associated with the activation of a GPCR hetero-oligomer. *Nature Communications*, *10*(1), 2765.

Zhang, J.-B., Chang, S., Xu, P., Miao, M., Wu, H., Zhang, Y., Zhang, T., Wang,

H., Zhang, J., Xie, C., Song, N., Luo, C., Zhang, X., & Zhu, S. (2018). Structural Basis of the Proton Sensitivity of Human GluN1-GluN2A NMDA Receptors. *Cell Reports*, *25*(13), 3582–3590.e4.

Zheng, H., Porebski, P. J., Grabowski, M., Cooper, D. R., & Minor, W. (2017). Databases, Repositories, and Other Data Resources in Structural Biology. *Methods in Molecular Biology* , *1607*, 643–665.


UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL DE CATALUNYA