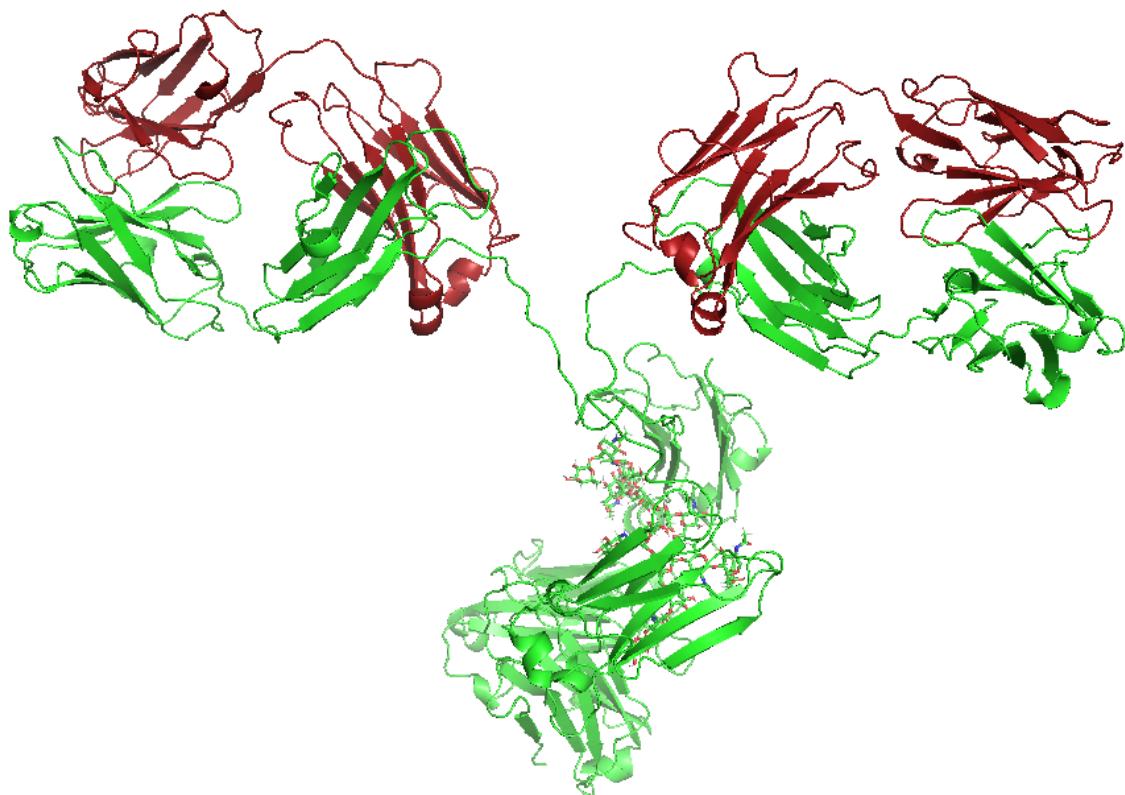


PROTEIN FUNCTION ANALYSIS

IMMUNOGLOBULIN G



ARES PERRA
MARIA RODRIGUEZ
MARIA COBO
MARIA ROY

INDEX

BACKGROUND	2
BASICS OF STRUCTURE	2
HYDROPHOBICITY	3
ISOTYPES	4
VDJ REARRANGEMENT	4
FUNCTION	5
IMMUNOGLOBULIN G ELICITS HUMORAL IMMUNITY	5
HOW DO THEY CARRY OUT THEIR FUNCTION? DO THEY REQUIRE THE INTERACTION WITH OTHER PROTEINS OR MOLECULES?	6
FOLD OF IgG	6
AVAILABLE STRUCTURES	8
FC REGION AS AN ESSENTIAL ROLE	8
CHROMOSOME TRANSLOCATION IN IGH	10
EVOLUTIONARY HISTORY OF IgG	12
PFAM DATABASE AND AVAILABLE HMM FOR IMMUNOGLOBULIN	12
EVOLUTIONARY HISTORY OF PROTEIN FAMILY	13
PROTEIN FAMILY ALIGNMENT (HMM)	14
CONSERVED REGIONS IN ALIGNMENTS	15
STRUCTURE	18
PDB STRUCTURES OF PROTEIN FAMILY AND PSIBLAST	18
SUPERIMPOSE OF STRUCTURES AND ANALYSIS	19
MONOCLONAL ANTIBODY ENGINEERING IN IgG	21
MODIFICATIONS AT DIFFERENT pH TO STUDY IgG HALF-LIFE, AFFINITY AND RECYCLING PHASE	24
IS THE MUTATION ON THE MONOCLONAL ANTIBODY COMPROMISING THE STABILITY OF THE RECEPTOR?	26
HOW COULD THE MUTATION BE AFFECTING THE PROTEIN-PROTEIN INTERACTION?	30
REFERENCES	34

BACKGROUND

- Basics of structure
- Isotypes
- VDJ rearrangement

BASICS OF STRUCTURE

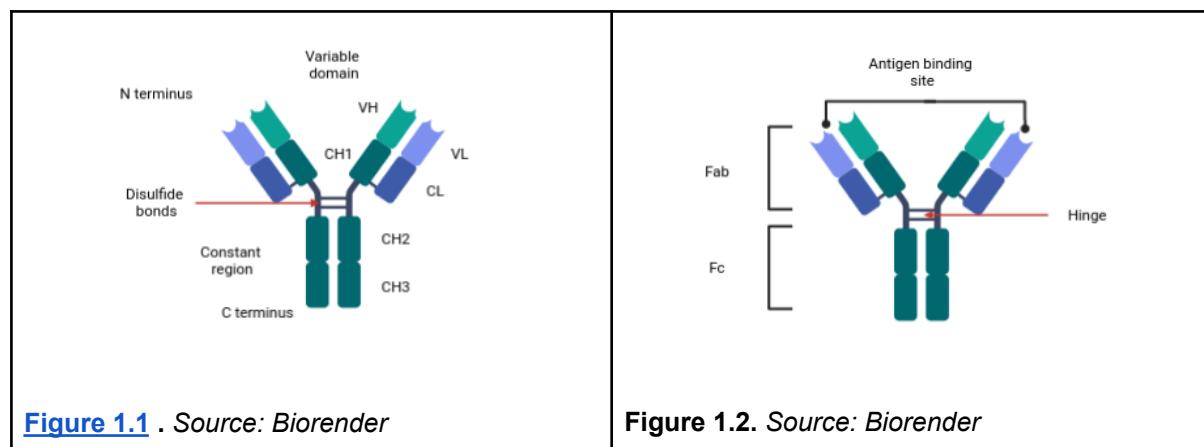
Antibodies are the **secreted form** of the **B-cell receptor**, they are identical except for a small portion of the C-terminus of the heavy-chain constant region. In the case of the B-cell receptor the C-terminus is a hydrophobic membrane-anchoring sequence, and in the case of antibody it is a hydrophilic sequence that allows secretion.

As shown in [Fig. 1.1](#) IgG antibodies are composed of **two different polypeptide chains**. One, the **heavy (H)** chain, the other, termed the **light (L)** chain. Each IgG molecule consists of **two heavy and two light** chains. The two heavy chains are linked by **disulfide bonds** and each heavy chain is linked to a light chain by a **disulfide bond**. In immunoglobulin molecules, those chains are identical, which results in two identical **antigen-binding sites**,

Note that the light chain is made up of **two domains**, whereas the heavy chain contains **four** (numbered from the amino-terminal end to the carboxy terminus: VH, CH1, CH2 and CH3). We also distinguish between two different types of fragments ([Fig. 1.2](#)): the ones contained in the **Fab** region, for Fragment antigen binding, composed of the complete light chain paired with VH and CH1 domains of the heavy chain. The other fragment, called the **Fc**, was originally observed to crystallize readily, that's why it is called Fragment crystallizable, which corresponds to the paired CH2 and CH3 domains.

There are two types of light chains, termed **lambda** (λ) and **kappa** (κ). By contrast, the heavy chain structure is what determines the function of the antibody.

The protein domains described above form globular domains. Thus, when fully folded and assembled, the antibody joins by a **flexible** stretch of polypeptide chain known as the **hinge** region. Which links Fc and Fab portions allowing independent movement of both Fab arms.



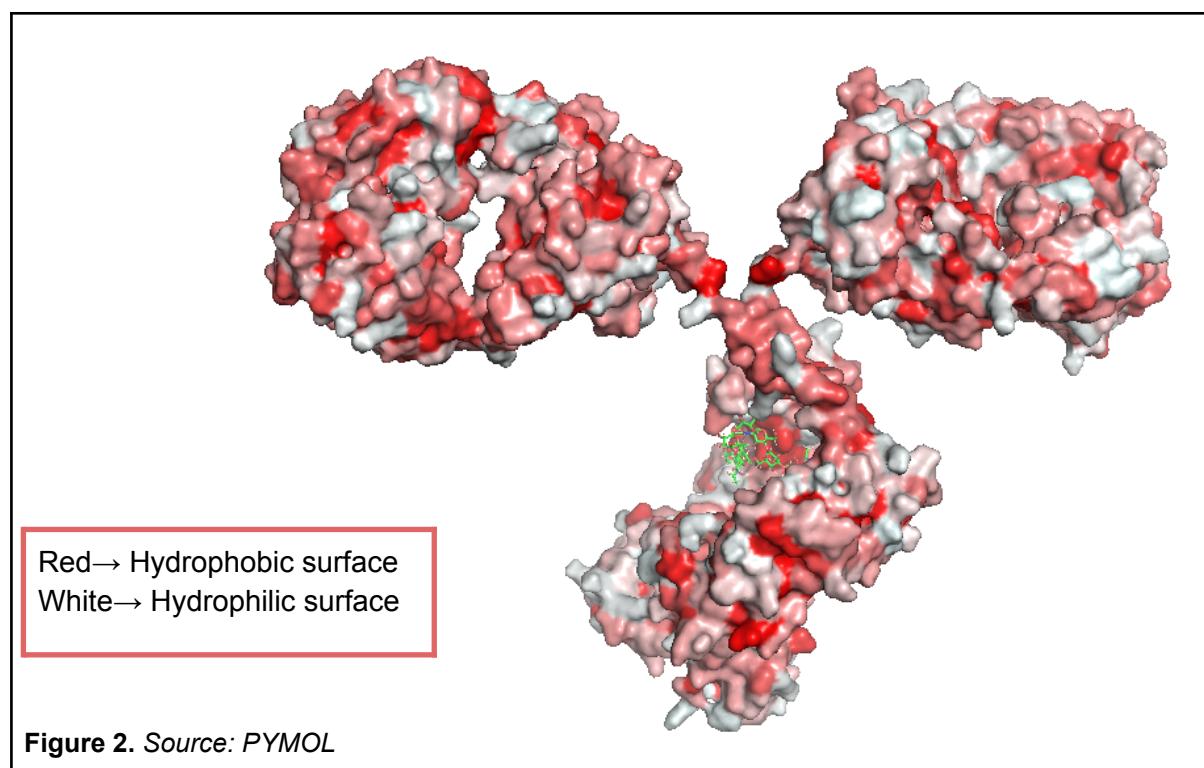
HYDROPHOBICITY

Immunoglobulin G, as we can extract from [Figure 2](#), is a globular protein composed of both hydrophobic and hydrophilic amino acid residues. IgG is considered to be an hydrophilic protein, thus, it has a high affinity for water and is able to dissolve in aqueous solutions.

The hydrophobicity will vary depending on the local environment. The protein is composed of peptide chains, two heavy chains and two light chains, that are folded into a compact three-dimensional structure. Meaning that the interior of IgG is typically more hydrophobic, while the surface is more hydrophilic due to the presence of charged and polar amino acid residues.

This correlates with the hydrophobicity of heavy and light chains of IgG. In general, the heavy chains of IgG tend to be more hydrophobic, since the residues on the interior can contribute to the stability of the protein by forming hydrophobic interactions with each other and by stabilizing the folded conformation of the protein.

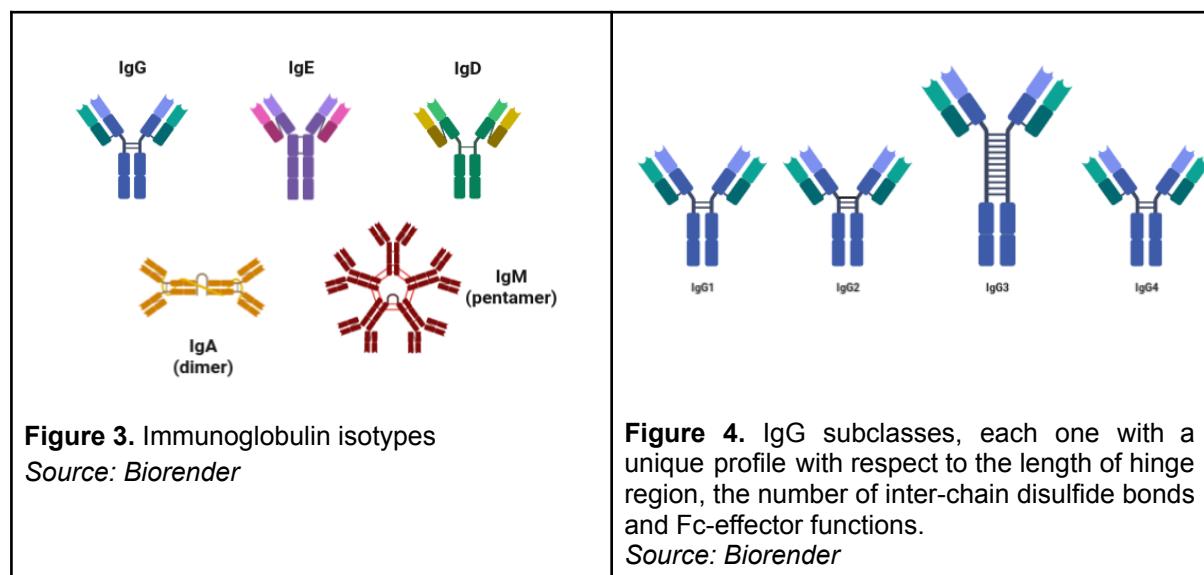
When referring to hydrophobicity, binding sites are also important. In overall, the hydrophobicity of the sites where antigen binds to IgG can play an important role in determining the specificity and strength of the interaction between the two molecules. By interacting with both hydrophobic and hydrophilic regions of the antigen, the binding site on IgG helps to ensure a tight and specific interaction that is able to recognize and neutralize the target antigen.



ISOTYPES

The functional differences between **heavy chain** isotypes lie mainly in the **Fc fragment**. Thus, we find five main functions or **isotypes**, some of them having subtypes. The five isotypes, shown in [Figure 3](#), are: immunoglobulin M (IgM), immunoglobulin D (IgD), immunoglobulin G (IgG), immunoglobulin A (IgA), and immunoglobulin E (IgE). Their heavy chains are also referred to by the corresponding lower-case Greek letter (μ , δ , γ , α , and ϵ , respectively).

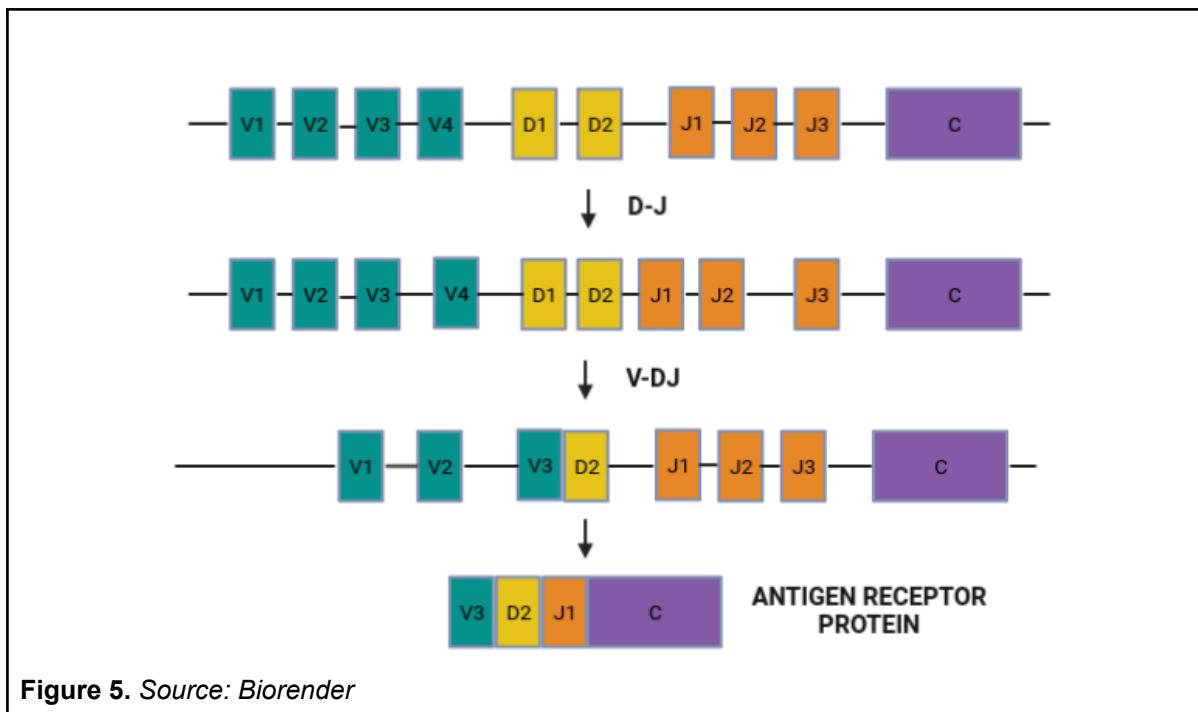
IgG is by far the most abundant immunoglobulin and has several subclasses (IgG1, 2, 3, and 4 in humans) represented in [Figure 4](#).



VDJ REARRANGEMENT

B cells can generate antibodies to a large variety of antigens, this is explained by **VDJ rearrangement**, where the exons encoding the antigen binding portions of the receptor (variable regions) are assembled by **chromosomal breakage** and **rejoining** in developing lymphocytes. The exons encoding to antigen binding domains are assembled from the V (variable), D (diversity) and J (joining) gene segments.

Assembly of a complete variable exon occurs in **two steps** (in the case of an Ig heavy chain gene), as shown below in [Figure 5](#). First, a D and a J segment are chosen from among several possibilities, and are brought together to form a D-J rearrangement. Then a V region is selected and joined with the D-J rearrangement to form a complete VDJ exon. Immunoglobulin light chain genes rearrange in a single step, involving V-J recombination, as D segments are absent from these loci.



This is a process that has a potential downside, that we will see in next points. Since it must break chromosomal DNA several times in order to generate a functional antigen receptor gene, this creates significant opportunities for error.

FUNCTION

IMMUNOGLOBULIN G ELICITS HUMORAL IMMUNITY

The primary function of immunoglobulins is to elicit humoral immunity by binding to the foreign antigen. The antibody-mediated humoral immune response kills the invading microbes and prevents infections from spreading to other regions of the body. This type of immune response is highly specific and complex.

We are going to focus on the IgG antibodies that are usually of higher affinity and are found in blood and in extracellular fluid, where they can neutralize toxins, viruses, and bacteria, opsonize them for phagocytosis, and activate the complement system by binding the receptor.

Immunoglobulins G belong to the Immunoglobulin superfamily.

Family:

- C1 set domains (antibody constant domain-like)
- V set domains (antibody variable domain-like)

HOW DO THEY CARRY OUT THEIR FUNCTION? DO THEY REQUIRE THE INTERACTION WITH OTHER PROTEINS OR MOLECULES?

The antigen-specific IgG antibodies can switch to a population with reduced sialic acid that is thus capable of mediating antigen clearance and a protective inflammatory response through the engagement of subclass-specific FcγRs on effector cells.

The IgG Fc contains distinct Fc receptor (FcR) binding sites: the leukocyte receptors **Fc gamma RI** and **Fc gamma RIIa** bind to a region in the Fc distinct from that recognized by neonatal FcR and protein A. To carry out their function they need to bind this receptor to cell-surface antigens.

So they require interaction with the antigens in order to carry out their function.

FOLD OF IgG

To find out the fold of our protein first we need to execute BLAST with our target sequence. In our case, we will be doing this for the four subclasses of IgG: IgG1, IgG2, IgG3 and IgG4 (all of these sequences were extracted from the Uniprot Database). Secondly, we use the PSSM (position substitution specific matrix) and execute PSI-BLAST on our target sequence with five iterations. Thirdly, we use Uniprot to create an accurate PSSM and lastly, we use this to search for templates in the PDB.

After we end this process, we now search the IDs that we got in the SCOP database. This will give us the fold of the protein. As a result we got that all four subclasses had the same protein IDs, therefore, their protein fold is the same: Immunoglobulin-like-beta-sandwich.

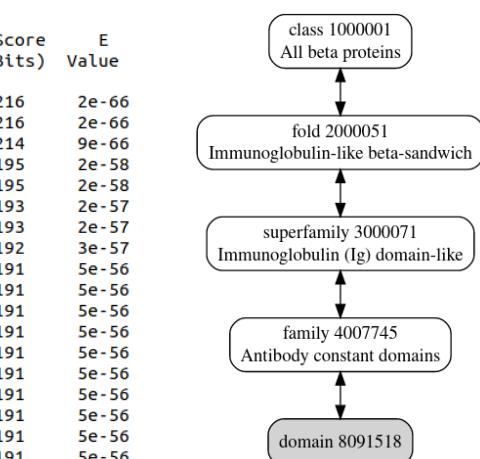
All Immunoglobulin-like domains are characterized by two antiparallel β-sheets and are known for assessing the misfolding tendency and the folding stability of a protein.

IgG1: 4byh → Immunoglobulin heavy constant gamma 1

```

32 Query= sp|P01857|IGHG1_HUMAN Immunoglobulin heavy constant gamma 1 OS=Homo
33 sapiens OX=9606 GN=IGHG1 PE=1 SV=1
34
35 Length=330
36
37 Sequences producing significant alignments:
38
39 1hzh_K mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN
40 1hzh_H mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN
41 2ig2_H mol:protein length:455 IGG1-LAMBDA KOL FAB (HEAVY CHAIN)
42 1igy_D mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.1.3
43 1igy_B mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.1.3
44 1igt_D mol:protein length:444 IGG2A INTACT ANTIBODY - MAB231
45 1igt_B mol:protein length:444 IGG2A INTACT ANTIBODY - MAB231
46 1mco_H mol:protein length:428 IGG1 MCG INTACT ANTIBODY (HEAVY CHAIN)
47 2rcj_T mol:protein length:523 Light chain
48 2rcj_S mol:protein length:523 Light chain
49 2rcj_P mol:protein length:523 Light chain
50 2rcj_O mol:protein length:523 Light chain
51 2rcj_L mol:protein length:523 Light chain
52 2rcj_K mol:protein length:523 Light chain
53 2rcj_H mol:protein length:523 Light chain
54 2rcj_G mol:protein length:523 Light chain
55 2rcj_D mol:protein length:523 Light chain
56 2rcj_C mol:protein length:523 Light chain

```

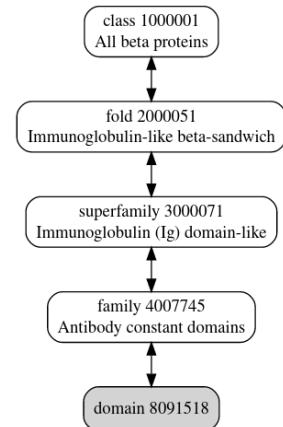


SCOP database: 1hzh → Fold: Immunoglobulin-like-beta-sandwich

IgG2: 4hag → Immunoglobulin heavy constant gamma 2

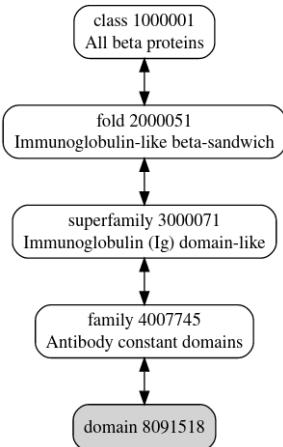
```

32 Query= sp|P01859|IGHG2_HUMAN Immunoglobulin heavy constant gamma 2 OS=Homo
33 sapiens OX=9606 GN=IGHG2 PE=1 SV=2
34
35 Length=325
36
37 Sequences producing significant alignments:
38
39 2ig2_H mol:protein length:455 IGG1-LAMBDA KOL FAB (HEAVY CHAIN) 220 9e-68
40 1hz_h mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN 219 2e-67
41 1hz_h mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN 219 2e-67
42 1igy_d mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.1.3 204 3e-62
43 1igy_b mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.1.3 204 3e-62
44 1mco_h mol:protein length:428 IGG1 MCG INTACT ANTIBODY (HEAVY CHAIN) 202 2e-61
45 2rcj_t mol:protein length:523 Light chain 202 3e-60
46 2rcj_s mol:protein length:523 Light chain 202 3e-60
47 2rcj_p mol:protein length:523 Light chain 202 3e-60
48 2rcj_o mol:protein length:523 Light chain 202 3e-60
49 2rcj_l mol:protein length:523 Light chain 202 3e-60
50 2rcj_k mol:protein length:523 Light chain 202 3e-60
51 2rcj_h mol:protein length:523 Light chain 202 3e-60
52 2rcj_g mol:protein length:523 Light chain 202 3e-60
53 2rcj_d mol:protein length:523 Light chain 202 3e-60
54 2rcj_c mol:protein length:523 Light chain 202 3e-60
55 1igt_d mol:protein length:444 IGG2A INTACT ANTIBODY - MAB231 200 3e-60
56 1igt_b mol:protein length:444 IGG2A INTACT ANTIBODY - MAB231 200 3e-60
  
```

**SCOP database: 1hzh → Fold: Immunoglobulin-like-beta-sandwich****IgG3: 6d58 → Immunoglobulin heavy constant gamma 3**

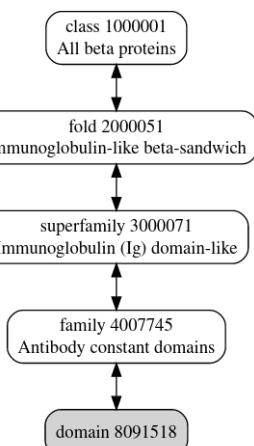
```

32 Query= sp|P01860|IGHG3_HUMAN Immunoglobulin heavy constant gamma 3 OS=Homo
33 sapiens OX=9606 GN=IGHG3 PE=1 SV=2
34
35 Length=377
36
37 Sequences producing significant alignments:
38
39 1hz_h mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN 204 6e-61
40 1hz_h mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN 204 6e-61
41 2ig2_h mol:protein length:455 IGG1-LAMBDA KOL FAB (HEAVY CHAIN) 202 2e-60
42 2rcj_t mol:protein length:523 Light chain 194 8e-57
43 2rcj_s mol:protein length:523 Light chain 194 8e-57
44 2rcj_p mol:protein length:523 Light chain 194 8e-57
45 2rcj_o mol:protein length:523 Light chain 194 8e-57
46 2rcj_l mol:protein length:523 Light chain 194 8e-57
47 2rcj_k mol:protein length:523 Light chain 194 8e-57
48 2rcj_h mol:protein length:523 Light chain 194 8e-57
49 2rcj_g mol:protein length:523 Light chain 194 8e-57
50 2rcj_d mol:protein length:523 Light chain 194 8e-57
51 2rcj_c mol:protein length:523 Light chain 194 8e-57
52 1igt_d mol:protein length:444 IGG2A INTACT ANTIBODY - MAB231 184 9e-54
53 1igt_b mol:protein length:444 IGG2A INTACT ANTIBODY - MAB231 184 9e-54
54 1igy_d mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.1.3 180 3e-52
55 1igy_b mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.1.3 180 3e-52
56 2gj7_B mol:protein length:227 Ig gamma-1 chain C region 162 2e-47
  
```

**IgG4: 4c55 → Immunoglobulin heavy constant gamma 4**

```

32 Query= sp|P01861|IGHG4_HUMAN Immunoglobulin heavy constant gamma 4 OS=Homo
33 sapiens OX=9606 GN=IGHG4 PE=1 SV=1
34
35 Length=327
36
37 Sequences producing significant alignments:
38
39 1hz_h mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN 215 8e-66
40 1hz_h mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN 215 8e-66
41 2ig2_h mol:protein length:455 IGG1-LAMBDA KOL FAB (HEAVY CHAIN) 212 6e-65
42 1igy_d mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.1.3 195 2e-58
43 1igy_b mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.1.3 195 2e-58
44 1mco_h mol:protein length:428 IGG1 MCG INTACT ANTIBODY (HEAVY CHAIN) 194 2e-58
45 1igt_d mol:protein length:444 IGG2A INTACT ANTIBODY - MAB231 194 4e-58
46 1igt_b mol:protein length:444 IGG2A INTACT ANTIBODY - MAB231 194 4e-58
47 2rcj_t mol:protein length:523 Light chain 193 5e-57
48 2rcj_s mol:protein length:523 Light chain 193 5e-57
49 2rcj_p mol:protein length:523 Light chain 193 5e-57
50 2rcj_o mol:protein length:523 Light chain 193 5e-57
51 2rcj_l mol:protein length:523 Light chain 193 5e-57
52 2rcj_k mol:protein length:523 Light chain 193 5e-57
53 2rcj_h mol:protein length:523 Light chain 193 5e-57
54 2rcj_g mol:protein length:523 Light chain 193 5e-57
55 2rcj_d mol:protein length:523 Light chain 193 5e-57
56 2rcj_c mol:protein length:523 Light chain 193 5e-57
  
```



SCOP database: 1hz → Fold: Immunoglobulin-like-beta-sandwich

AVAILABLE STRUCTURES

In the PDB we can search for proteins with an available structure. However, to know which are the available proteins of our protein family, we first need to know which IDs to search for. Thus, we use **PFAM**. PFAM has three programs and we execute them in the following order. First, we use hmmscan which from the output we will obtain the domain of the target sequence, in our case it is “C1-set”. Secondly, we execute hmmfetch to extract the HMM we want from the PFAM database. And lastly, the hmmsearch that will give us the protein IDs of the PDB that are in the **C1-set domain**.

```

1 # hmmsearch :: search profile(s) against a sequence database
2 # HMMER 3.2.1 (June 2018); http://hmmer.org/
3 # Copyright (C) 2018 Howard Hughes Medical Institute.
4 # Freely distributed under the BSD open source license.
5 # -----
6 # query HMM file:          domain_4byh.hmm
7 # target sequence database: /shared/databases/blastdat/pdb_seq
8 # -----
9
10 Query:      C1-set [M=83]
11 Accession:  PF07654.10
12 Description: Immunoglobulin C1-set domain
13 Scores for complete sequences (score includes all domains):
14   --- full sequence ---   --- best 1 domain ---  -#dom-
15   E-value  score  bias    E-value  score  bias    exp  N  Sequence Description
16   -----  -----  -----  -----  -----  -----  -----  -----
17   4e-82  273.3  18.7   6.6e-24  86.7  0.4   4.9  4  2rcj_C  mol:protein length:523 Light chain
18   4e-82  273.3  18.7   6.6e-24  86.7  0.4   4.9  4  2rcj_D  mol:protein length:523 Light chain
19   4e-82  273.3  18.7   6.6e-24  86.7  0.4   4.9  4  2rcj_G  mol:protein length:523 Light chain
20   4e-82  273.3  18.7   6.6e-24  86.7  0.4   4.9  4  2rcj_H  mol:protein length:523 Light chain
21   4e-82  273.3  18.7   6.6e-24  86.7  0.4   4.9  4  2rcj_K  mol:protein length:523 Light chain
22   4e-82  273.3  18.7   6.6e-24  86.7  0.4   4.9  4  2rcj_L  mol:protein length:523 Light chain
23   4e-82  273.3  18.7   6.6e-24  86.7  0.4   4.9  4  2rcj_O  mol:protein length:523 Light chain
24   4e-82  273.3  18.7   6.6e-24  86.7  0.4   4.9  4  2rcj_P  mol:protein length:523 Light chain
25   4e-82  273.3  18.7   6.6e-24  86.7  0.4   4.9  4  2rcj_S  mol:protein length:523 Light chain
26   4e-82  273.3  18.7   6.6e-24  86.7  0.4   4.9  4  2rcj_T  mol:protein length:523 Light chain
27   2.4e-57 193.9  10.1   9.9e-24  86.2  0.3   3.8  3  1hz_h  mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN
28   2.4e-57 193.9  10.1   9.9e-24  86.2  0.3   3.8  3  1hz_h  mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN
29   6.8e-57 192.4  10.9   5.3e-24  87.0  0.4   3.6  3  2ig2_h  mol:protein length:455 IGG1-LAMBDA KOL FAB (HEAVY
30   5.4e-56 189.6  15.2   4.8e-24  87.2  0.4   4.0  3  1mco_h  mol:protein length:428 IGG1 MCG INTACT ANTIBODY (H
31   3.2e-55 187.1  7.5    1.6e-23  85.5  0.5   4.0  3  1igy_b  mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.
32   3.2e-55 187.1  7.5    1.6e-23  85.5  0.5   4.0  3  1igy_d  mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.
33   5.3e-53 180.0  13.9   1.1e-22  82.9  1.6   3.9  3  1igt_b  mol:protein length:444 IGG2A INTACT ANTIBODY - MAB
34   5.3e-53 180.0  13.9   1.1e-22  82.9  1.6   3.9  3  1igt_d  mol:protein length:444 IGG2A INTACT ANTIBODY - MAB

```

This is the file that we obtain with hmmsearch and the IDs of the available structure of our protein family are in the column of “Sequence”. For example: 2rcj, 1hzh, 2ig2, 1mco, etc.

FC REGION AS AN ESSENTIAL ROLE

The function of IgG is to activate the immune system; to do so, the structure of the antibody has the **antigen-binding fragment** (Fab), the **Fc region**, also called fragment crystallizable region, joined with the hinge region. Looking in depth, we can say that the Fc region is essential because it ensures that IgG generates an appropriate immune response for a given antigen.

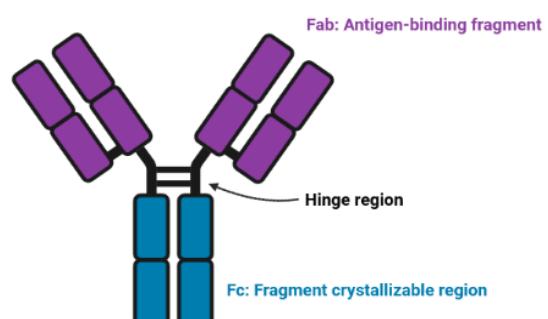


Figure 6. Parts of an antibody. Source: Biorender

The structure of the Fc region is composed of two identical protein fragments, derived from the second and third constant domains of the antibody's two heavy chains. It is located on the tail of the antibody. We can say that it is responsible for interacting with the cell surface receptors, which are called **Fc receptors**, and some proteins of the complement system, to promote removal of the antigen. If we take a look at the amino acids in this region, we can see that they are mostly **polar** meaning it is hydrophilic so it is in contact with H₂O molecules. The Fc region of IgG bears a highly conserved **N-glycosylation site**. Glycosylation of the Fc fragment is essential for Fc receptor-mediated activity. It consists of modifying appropriate **asparagine** residues of proteins with **oligosaccharide** structures.

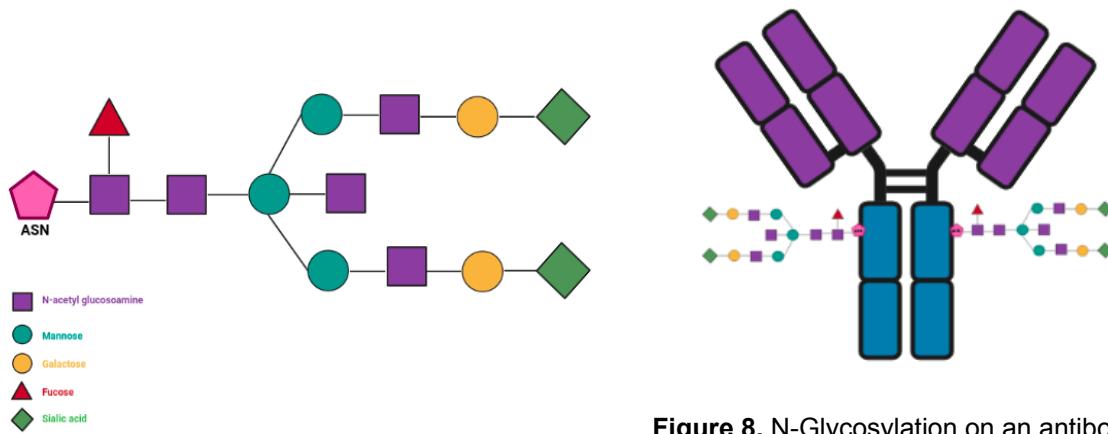


Figure 8. N-Glycosylation on an antibody.

Figure 7. N-Glycosylation. Source: Biorender

The Fc receptors are the key immune regulatory receptors connecting the antibody mediated humoral immune response to cellular effector functions. They are a group of glycoproteins belonging to the immunoglobulin superfamily. The ones specific for IgG found on leukocytes are **Fc γ R**, the high affinity receptor; Fc γ RIIa, Fc γ RIIb and Fc γ RIIc; and Fc γ RIIIa and Fc γ RIIIb.

If we look at the other types of immunoglobulins, we observe they have the same basic structure, containing this Fc region so we can say that it is essential as well. Even though, for each type of immunoglobulins the Fc receptors are from different classes such as IgA has Fc α R, IgM has Fc μ R.

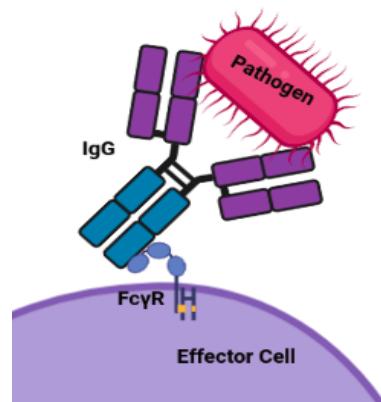


Figure 9. Fc region connected with Fc γ R. Source: Biorender

CHROMOSOME TRANSLOCATION IN IGH

Among the different subtypes of IgG, in the **IGHG1** gene - with **UniProt** entry: **P01857** - a chromosomal aberration is affecting this gene causing **Multiple myeloma** (MM).

MM is a malignant tumor of plasma cells, which still remains incurable with conventional chemotherapy.

It has been related with several translocations on the long arm of chromosome (q) 14 or with deletions on 13q.

Chromosome 14q32 translocations involve the immunoglobulin heavy chain (**IgH**) in multiple myeloma. These specific translocations are the hallmark of B cell malignancies, since they represent the mechanism of activation of several proto-oncogenes.

In IgH translocations's anatomy there are at least three enhancers that regulate transcription in B cells: the intronic enhancer ($E\mu$) located between the JH and switch μ ($S\mu$) sequences and two 3' IgH enhancers located downstream of the α constant regions ($E\alpha 1$ and $E\alpha 2$). In multiple myeloma the breakpoint usually occurs into switch regions (often the JH region), which are non-coding, repetitive sequences located 5' of the constant region genes and involved in class switch recombination.

These translocations involve three DNA breaks, with two breaks on the IgH locus and one on the reciprocal chromosome. In myeloma, there is a failure to join the switch regions (which would allow a new heavy chain gene to juxtapose to the VDJ region), DNA is still looped out, but the switch regions are joined to other chromosome ends that are [available](#).

The hematological malignancy affecting terminally differentiated B-cells and characterized by bone marrow infiltration by plasma cells secreting a monoclonal compound (IgG, IgA, IgD, or light chains), is characterized by diffuse involvement of the skeletal system, hyperglobulinemia, Bence-Jones proteinuria and anemia. Its complications involve bone pain, hypercalcemia, renal failure and spinal cord compression. The antibodies that are produced lead to impaired humoral immunity and patients have a high prevalence of infection.

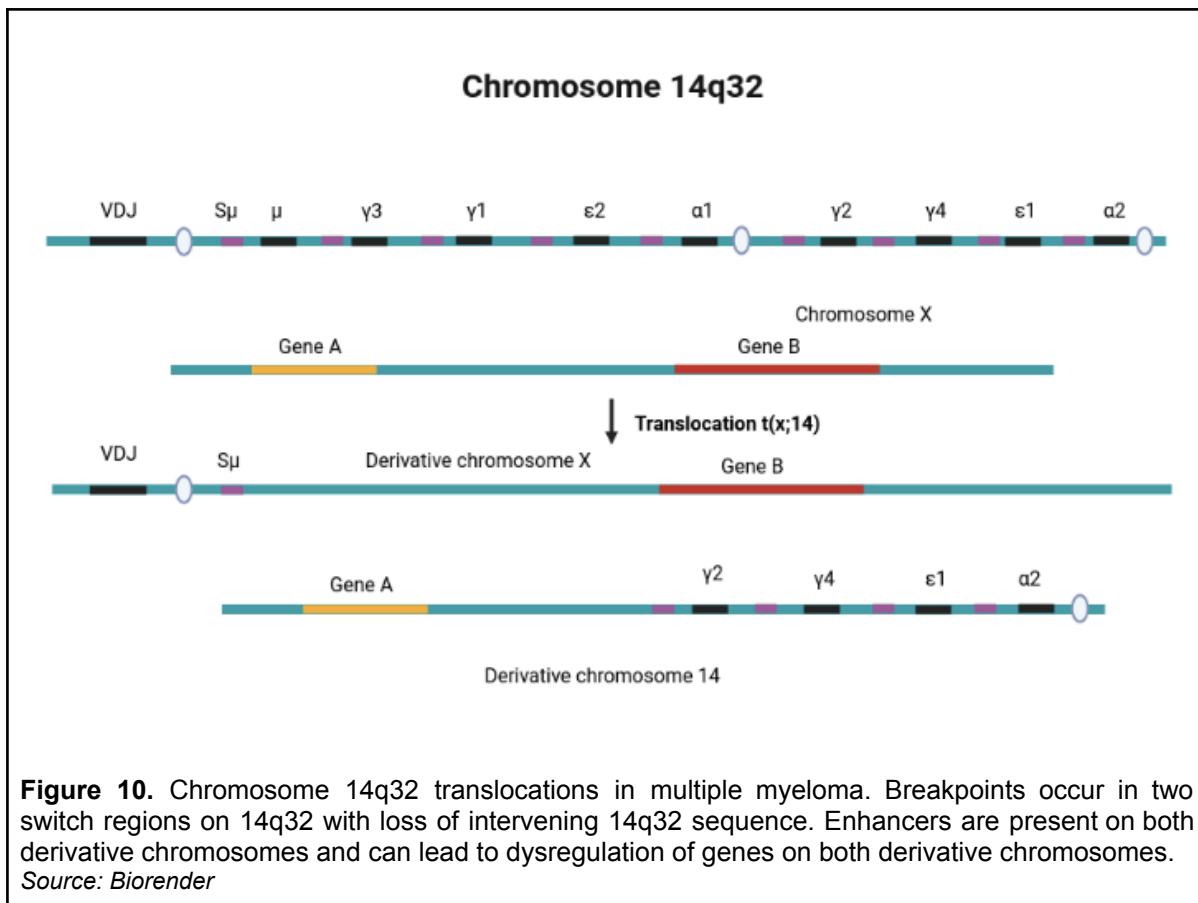
P01857 · IGHG1_HUMAN

Disease & Variants¹

Involvement in disease¹

Multiple myeloma (MM)

2 Publications



EVOLUTIONARY HISTORY OF IgG

PFAM DATABASE AND AVAILABLE HMM FOR IMMUNOGLOBULIN

The next step is to find the available **HMM** of our protein in the PFAM database. After executing `hmmscan` we get the following output.

```

1 # hmmscan :: search sequence(s) against a profile database
2 # HHMER 3.2.1 (June 2018); http://hhmer.org/
3 # Copyright (C) 2018 Howard Hughes Medical Institute.
4 # Freely distributed under the BSD open source license.
5 # -----
6 # query sequence file:          4byh.fasta
7 # target HMM database:         /shared/databases/pfam-3/Pfam-A.hmm
8 # -----
9
10 Query:   sp|P01857|IGHG1_HUMAN [L=330]
11 Description: Immunoglobulin heavy constant gamma 1 OS=Homo sapiens OX=9606 GN=IGHG1 PE=1 SV=1
12 Scores for complete sequence (score includes all domains):
13   --- full sequence ---   --- best 1 domain ---   -dom-
14   E-value score bias   E-value score bias   exp N Model   Description
15   ----- ----- -----   ----- ----- -----   -----
16   7.5e-59 195.5 9.2   5.8e-25 86.9 0.3   3.1 3 C1-set   Immunoglobulin C1-set domain
17   4.5e-19 68.0 2.1   1.4e-07 31.1 0.1   3.2 3 C2-set_2   CD80-like C2-set immunoglobulin domain
18   6.1e-14 52.1 14.9   0.00032 21.0 0.7   3.2 3 Ig_3   Immunoglobulin domain
19   5.6e-05 22.8 0.6   0.3   16.8 0.1   3.2 3 I-set   Immunoglobulin I-set domain
20   0.0008 19.3 10.8   0.019 14.9 0.6   4.2 3 ig   Immunoglobulin domain
21   ----- inclusion threshold -----
22   0.097 12.1 0.6   0.94 8.9 0.3   2.3 2 DUF2914   Protein of unknown function (DUF2914)
23
24
25 Domain annotation for each model (and alignments):
26 >> C1-set Immunoglobulin C1-set domain
27   # score bias c-Evalue i-Evalue hmmfrom hmmto alifrom alitо envfrom envto acc
28   ----- ----- ----- ----- ----- ----- ----- ----- -----
29   1 ! 54.2 0.1 4.1e-18 9.4e-15 3 80 .. 15 92 .. 12 95 .. 0.89
30   2 ! 59.1 0.7 1.2e-19 2.8e-16 4 83 .] 130 216 .. 127 216 .. 0.91
31   3 ! 86.9 0.3 2.6e-28 5.8e-25 2 81 .. 236 318 .. 235 320 .. 0.88

```

The `hmmscan` of our target sequence gives us HMMs for individual domains of the Immunoglobulins but not the combination of domains that we want for the whole Immunoglobulin. Therefore, we use the **jackhmmer** program to find a HMM specific to what we are interested in. If we write “`jackhmmer -h`” in the terminal, there will appear lots of built-in commands for `jackhmmer`, and one of them, in the section of “Options directing output”, is “`--chkhmm`”. What this built-in command does is that at the start of each iteration, it checkpoints the query HMM saving it to a file named. Also, we need to mention that `jackhmmer` must do five iterations.

This is the command line that we ended up with: `jackhmmer -N 5 --chkhmm uniprot 4byh.fasta/shared/databases/blastdat/uniprot_sprot.fasta`

```

1 HHMER3/f [3.2.1 | June 2018]
2 NAME sp|P01857|IGHG1_HUMAN-t4
3 DESC Immunoglobulin heavy constant gamma 1 OS=Homo sapiens OX=9606 GN=IGHG1 PE=1 SV=1
4 LENG 330
5 ALPH amino
6 RF yes
7 MM no
8 CONS yes
9 CS no
10 MAP yes
11 DATE Thu Feb 23 13:56:35 2023
12 NSEQ 3839
13 EFFN 39.016937
14 CKSUM 1244169655
15 STATS LOCAL MSV -11.3729 0.70054
16 STATS LOCAL VITERBI -12.8330 0.70054
17 STATS LOCAL FORWARD -5.4816 0.70054
18 HMM
    P      Q      R      S      T      E      F      G      H      I      K      L      M      N
    m->m  m->i  m->d  i->m  i->i  d->m  d->d
19
20 COMPO 2.68993 3.82574 3.02721 2.61352 3.57906 2.86011 3.69270 3.12652 2.70435 2.63821 4.09703 3.03785
2.92453 3.00404 2.93774 2.46949 2.61122 2.59837 4.22693 3.56291
21
2.68618 4.42225 2.77519 2.73123 3.46354 2.40513 3.72949 3.29354 2.67741 2.69355 4.24690 2.90347
2.73739 3.18146 2.89801 2.37887 2.77519 2.98518 4.58477 3.61503
22
0.02165 4.23953 4.96187 0.61958 0.77255 0.00000 *
23
1 0.59202 4.42020 3.88009 3.70800 4.46222 3.23157 4.67989 3.64645 3.73511 3.50940 4.51258 3.75045
4.00690 4.04388 3.94786 2.77359 3.67733 3.23388 5.85016 4.69873 1 A x -
24
2.68618 4.42225 2.77519 2.73123 3.46354 2.40513 3.72949 3.29354 2.67741 2.69355 4.24690 2.90347
2.73739 3.18146 2.89801 2.37887 2.77519 2.98518 4.58477 3.61503
25
0.02165 4.23953 4.96187 0.61958 0.77255 0.48576 0.95510
26
2 2.66528 4.89218 3.02698 2.42601 4.15269 3.47435 3.67836 3.57243 2.29733 2.79193 3.96991 3.00208
2.78502 2.76229 2.79854 1.93914 2.74469 3.23912 5.38919 4.05272 2 s x -
27
2.68618 4.42225 2.77519 2.73123 3.46354 2.40513 3.72949 3.29354 2.67741 2.69355 4.24690 2.90347
2.73739 3.18146 2.89801 2.37887 2.77519 2.98518 4.58477 3.61503
28
0.02022 4.30694 5.02933 0.61958 0.77255 0.48576 0.95510
29
3 2.56415 4.79695 3.17285 2.58580 4.02495 3.55052 3.77369 3.32101 2.42911 2.90918 3.89128 3.12259
2.91531 2.83514 3.02727 2.71701 1.66680 3.08408 5.33077 4.02271 3 t x -
30
2.68618 4.42225 2.77519 2.73123 3.46354 2.40513 3.72949 3.29354 2.67741 2.69355 4.24690 2.90347
2.73739 3.18146 2.89801 2.37887 2.77519 2.98518 4.58477 3.61503

```

Then, this will return five different files, one for each iteration. We must use the fifth iteration for our following step, the hmmsearch:

```
hmmsearch uniprot-5.hmm /shared/databases/blastdat/uniprot_sprot.fasta > 4byh_HMM.out
```

```
1 # hmmsearch :: search profile(s) against a sequence database
2 # HMMER 3.2.1 (June 2018); http://hmmer.org/
3 # Copyright (C) 2018 Howard Hughes Medical Institute.
4 # Freely distributed under the BSD open source license.
5 # -----
6 # query HMM file:          uniprot-5.hmm
7 # target sequence database: /shared/databases/blastdat/uniprot_sprot.fasta
8 # -----
9
10 Query:    sp|P01857|IGHG1_HUMAN-i4  [M=330]
11 Description: Immunoglobulin heavy constant gamma 1 OS=Homo sapiens OX=9606 GN=IGHG1 PE=1 SV=1
12 Scores for complete sequences (score includes all domains):
13   --- full sequence ---   --- best 1 domain --- -#dom-
14   E-value score bias      E-value score bias      exp N Sequence           Description
15   -----
16       0 8556.5 568.5     1.9e-49 173.5 4.9    144.0 134  sp|A2ASS6|TITIN_MOUSE Titin OS=Mus musculus GN=Ttn PE=1 SV=1
17       0 8513.8 571.8     4e-45 159.3 0.8    146.3 140  sp|Q8WZ42|TITIN_HUMAN Titin OS=Homo sapiens GN=TTN PE=1 SV=4
18       0 3342.0 121.8     1e-42 151.3 0.0    34.6 36   sp|Q96RW7|HMCN1_HUMAN Hemicentin-1 OS=Homo sapiens GN=HMCN1
19       0 3243.9 170.7     8.9e-44 154.8 5.7    41.6 44   sp|001761|UNC89_CAEEL Muscle M-line assembly protein unc-89
20       0 3237.9 206.0     9.1e-40 141.7 0.6    34.9 38   sp|A2AJ76|HMCN2_MOUSE Hemicentin-2 OS=Mus musculus GN=Hmcn2
21       0 3206.2 216.5     4e-39 139.5 0.2    36.1 39   sp|Q8NA2D|HMCN2_HUMAN Hemicentin-2 OS=Homo sapiens GN=HMCN2
22       0 2932.8 239.1     6.6e-34 122.4 0.2    47.8 51   sp|A2AAJ9|OBSCN_MOUSE Obscurin OS=Mus musculus GN=Obscn PE=2
23       0 2814.5 196.8     1.2e-42 151.1 1.0    55.5 46   sp|Q91T4U|TITIN_DROME Titin OS=Drosophila melanogaster GN=sl
24       0 2354.0 181.6     2.1e-34 124.1 0.1    42.3 43   sp|Q5VST9|OBSCN_HUMAN Obscurin OS=Homo sapiens GN=OBSCN PE=1
25       0 1480.5 142.5     1.9e-37 134.1 5.0    18.6 20   sp|P98160|PGBM_HUMAN Basement membrane-specific heparan sul
26       0 1418.9 69.9      7e-29 105.9 0.1    29.3 27   sp|Q23551|UNC22_CAEEL Twitchin OS=Caenorhabditis elegans GN=
27 1.8e-276 919.7 86.6     2.8e-30 110.5 2.9    12.6 12   sp|Q05793|PGBM_MOUSE Basement membrane-specific heparan sul
28 1.1e-225 752.8 26.5     9.4e-28 102.2 1.0    12.1 12   sp|Q3V1M1|IGS10_MOUSE Immunoglobulin superfamily member 10 0
29 8.6e-222 740.0 24.7     1.3e-28 105.0 0.0    11.9 12   sp|Q9NR99|MXRA5_HUMAN Matrix-remodeling-associated protein 5
30 2.9e-219 731.7 28.1     3.6e-31 113.4 0.0    16.2 15   sp|075147|OBSL1_HUMAN Obscurin-like protein 1 OS=Homo sapien
31 6.5e-218 727.2 117.6     1.6e-27 101.5 3.8    13.3 14   sp|Q06561|UNC52_CAEEL Basement membrane proteoglycan OS=Caen
32 3.9e-211 705.0 30.0     4.5e-27 100.0 1.1    12.3 12   sp|Q6WRH9|IGS10_RAT Immunoglobulin superfamily member 10 0
33 2.2e-199 666.3 19.1     3e-29 107.1 0.3    12.2 11   sp|Q6WRIO|IGS10_HUMAN Immunoglobulin superfamily member 10 0
34 4.1e-196 655.5 49.8     1.1e-41 147.9 0.2    9.2 9    sp|P11799|MYLK_CHICK Myosin light chain kinase, smooth musc
35 5.9e-186 622.2 15.8     8e-36 128.7 0.0    8.4 9    sp|Q4VA61|DSCL1_MOUSE Down syndrome cell adhesion molecule-l
36 5.9e-185 618.9 15.0     3e-35 126.8 0.0    8.4 9    sp|Q8TD84|DSCL1_HUMAN Down syndrome cell adhesion molecule-l
37 3.2e-184 616.5 39.5     5.9e-30 109.4 1.8    10.2 10   sp|Q15746|MYLK_HUMAN Myosin light chain kinase, smooth musc
38 3.6e-184 616.3 37.2     9.7e-30 108.7 1.9    10.1 9    sp|Q6PDN3|MYLK_MOUSE Myosin light chain kinase, smooth musc
39 2.7e-181 606.9 31.7     7e-24 89.5 0.7    13.4 12   sp|Q9BZZ2|SN_HUMAN Sialoadhesin OS=Homo sapiens GN=SIGLEC
40 4.8e-179 595.9 63.8     1.4e-24 91.8 5.0    13.6 14   sp|A7LCJ3|SN_PIG Sialoadhesin OS=Sus scrofa GN=SIGLECE1
41 8.7e-173 578.9 45.6     1.1e-19 75.7 0.7    13.2 13   sp|Q62230|SN_MOUSE Sialoadhesin OS=Mus musculus GN=Siglec
42 1.2e-158 532.5 17.5     8.1e-32 115.5 0.0    9.6 8    sp|Q9ERC8|DSCAM_MOUSE Down syndrome cell adhesion molecule h
```

EVOLUTIONARY HISTORY OF PROTEIN FAMILY

In order to study the evolutionary structure of the protein family, we have chosen 4 proteins with variated E-value and score from our hmmsearch file. We did a “Ctrl F” “gamma” in our output file to search for the four subclasses of IgG. Then, we searched for them in the Uniprot. The protein sequences we caught are the heavy chains from humans. This chain is the one that determines the function of the immunoglobulin and consequently characterizes it.

Why search for them in Uniprot? Uniprot is a non-redundant and non-biased database so it can better represent the evolutionary relation between sequences in the same family.

```
102   4.3e-98 333.4 8.7    2.4e-36 130.4 0.1    6.7 5  sp|Q7ZW34|CNTN5_DANRE Contactin-5 OS=Danio rerio GN=cntn5 PE
103   5e-97 329.9 15.5    5.6e-97 329.7 15.5   1.0 1  sp|P01857|IGHG1_HUMAN Ig gamma-1 chain C region OS=Homo sapi
104   1.1e-96 328.7 0.3    4.6e-26 96.6 0.0    7.9 7  sp|Q62682|CNTN3_RAT Contactin-3 OS=Rattus norvegicus GN=Cn
105   4.4e-96 326.8 0.1    1e-33 121.8 0.1    5.5 4  sp|Q02246|CNTN2_HUMAN Contactin-2 OS=Homo sapiens GN=cntn2 P
106   6.4e-96 326.2 6.0    1.4e-31 114.7 0.0    6.0 5  sp|P97527|CNTN5_RAT Contactin-5 OS=Rattus norvegicus GN=Cn
107   9.9e-96 325.6 6.0    1e-17 69.2 0.0    6.8 7  sp|Q14324|MYP2C_HUMAN Myosin-binding protein C, fast-type OS
108   1.6e-95 325.0 4.9    8.1e-31 112.3 0.0    6.2 5  sp|P68500|CNTN5_MOUSE Contactin-5 OS=Mus musculus GN=Cntn5 P
109   3e-95 324.1 20.0    1.8e-94 321.5 20.0   1.8 1  sp|P01860|IGHG3_HUMAN Ig gamma-3 chain C region OS=Homo sapi
110   3.1e-95 324.0 0.1    4.8e-18 70.3 0.0    6.7 6  sp|Q00872|MYP2C1_HUMAN Myosin-binding protein C, slow-type OS
111   1.3e-94 321.9 9.7    5e-38 135.9 0.1    6.9 5  sp|P97798|NEO1_MOUSE Neogenin OS=Mus musculus GN=Neo1 PE=1
112   1.4e-94 321.9 3.9    2.1e-32 117.5 0.1    4.5 4  sp|Q810U4|NRCAM_MOUSE Neuronal cell adhesion molecule OS=Mus
113   3.5e-94 320.6 12.1    4.9e-34 122.8 0.6    4.1 4  sp|P17948|VGFR1_HUMAN Vascular endothelial growth factor rec
114   3.7e-94 320.5 0.8    2.6e-23 87.6 0.0    6.4 5  sp|P14781|CNTN1_CHICK Contactin-1 OS=Galus gallus GN=cntn1
115   3.9e-94 320.4 3.9    1.2e-17 68.9 0.0    6.7 7  sp|Q5XKE0|MYP2C2_MOUSE Myosin-binding protein C, fast-type OS
116   4.1e-94 320.3 15.4   4.6e-94 320.1 15.4   1.0 1  sp|P01861|IGHG4_HUMAN Ig gamma-4 chain C region OS=Homo sapi
117   3.1e-93 317.4 6.8    7e-31 112.5 0.0    4.5 4  sp|Q92823|NRCAM_HUMAN Neuronal cell adhesion molecule OS=Hom
118   4.6e-93 316.9 0.0    1.2e-24 92.0 0.0    5.5 4  sp|Q28106|CNTN1_BOVIN Contactin-1 OS=Bos taurus GN=cntn1 PE=
119   4.7e-93 316.8 4.6    4.1e-32 116.5 0.2    4.5 4  sp|P97686|NRCAM_RAT Neuronal cell adhesion molecule OS=rat
120   5.1e-93 316.7 18.5   5.7e-93 316.5 18.5   1.0 1  sp|P01859|IGHG2_HUMAN Ig gamma-2 chain C region OS=Homo sapi
121   6.1e-93 316.5 8.9    5.1e-35 126.1 0.2    4.8 4  sp|P35331|NRCAM_CHICK Neuronal cell adhesion molecule OS=Gal
```

PROTEIN FAMILY ALIGNMENT (HMM)

To execute these alignments first we had to do a cat of all the sequences we wanted to align
“*cat P01857.fa P01859.fa P01860.fa P01861.fa > gamma.fa*”.

Then, we used the program to hmmpalign with our HMM of the whole Immunoglobulin and the target sequence “*hmmpalign uniprot-5.hmm gamma.fa > gamma_hmm.sto*”.

Moreover, as we know that the programs of HMMer work with alignments in Stockholm format, we then had to use Perl and transform the output file into clustalw format and back.
“*perl /shared/PERL/aconvertMod2.pl –in h –out c <gamma_hmm.sto> gamma_hmm.clu*”.

The HMM that we will be using for both alignments is the one we obtained from the “sp|P01857|IGHG1_HUMAN-i4”. Moreover, we have a multiple sequence alignment with the four sequences that we got from our HMM. The sequences aligned are: Immunoglobulin heavy constant gamma 1 (P01857), Immunoglobulin heavy constant gamma 2 (P01859), Immunoglobulin heavy constant gamma 3 (P01860) and Immunoglobulin heavy constant gamma 4 (P01861).

1 CLUSTAL W(1.60) multiple sequence alignment	
2	
3 sp P01857 IGHG1_HUMAN	ASTKGPSVFPLAPSSKSTSGTAALGCLVKDYFPEPVTVWSNSGALTSGVHTFPALQSS
4 sp P01859 IGHG2_HUMAN	ASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVWSNSGALTSGVHTFPALQSS
5 sp P01860 IGHG3_HUMAN	ASTKGPSVFPLAPCSRSTSGTAALGCLVKDYFPEPVTVWSNSGALTSGVHTFPALQSS
6 sp P01861 IGHG4_HUMAN	ASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVWSNSGALTSGVHTFPALQSS
7	
8 sp P01857 IGHG1_HUMAN	GLYSLSSVVTPSSSLGTQTYICNVNHKPSNTKVDKKVEPK-----
9 sp P01859 IGHG2_HUMAN	GLYSLSSVVTPSSNFGTQTYTCNVDHKPSNTKVDKTVERK-----
10 sp P01860 IGHG3_HUMAN	GLYSLSSVVTPSSSLGTQTYTCNVNHKPSNTKVDKRVELKtlgdthtcprcepks
11 sp P01861 IGHG4_HUMAN	GLYSLSSVVTPSSSLGTKYTCNVDHKPSNTKVDKRVESK-----
12	
13 sp P01857 IGHG1_HUMAN	-----SCDKHTCPCPAPELLGGPSVFLFPKPKD
14 sp P01859 IGHG2_HUMAN	-----CCVECPCPAPPV-AGPSVFLFPKPKD
15 sp P01860 IGHG3_HUMAN	dtpppcprcepkscdpppcprcepkscdTPPPCPRCPAPELLGGPSVFLFPKPKD
16 sp P01861 IGHG4_HUMAN	-----YGPPCPSCPAPEFLGGPSVFLFPKPKD
17	
18 sp P01857 IGHG1_HUMAN	LMISRTPEVCVVVDVSHEDPEVKFNWYVGVEVHNAKTKPREEQYNSTYRVVSVLTVLH
19 sp P01859 IGHG2_HUMAN	LMISRTPEVCVVVDVSHEDPEVQFNWYVGVEVHNAKTKPREEQFNSTFRVVSVLTVVH
20 sp P01860 IGHG3_HUMAN	LMISRTPEVCVVVDVSHEDPEVQFKWYVGVEVHNAKTKPREEQYNSTFRVVSVLTVLH
21 sp P01861 IGHG4_HUMAN	LMISRTPEVCVVVDVSQEDPEVQFNWYVGVEVHNAKTKPREEQFNSTYRVVSVLTVLH
22	
23 sp P01857 IGHG1_HUMAN	QDWLNGKEYKCKVSNKALPAPIEKTIKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVK
24 sp P01859 IGHG2_HUMAN	QDWLNGKEYKCKVSNKGLPAPIEKTIKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVK
25 sp P01860 IGHG3_HUMAN	QDWLNGKEYKCKVSNKALPAPIEKTIKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVK
26 sp P01861 IGHG4_HUMAN	QDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVK
27	
28 sp P01857 IGHG1_HUMAN	GFYPSDIAVEWESNGQPENNYKTPPVLDGSFFLYSKLTVDKSRWQQGNVFCSVMHE
29 sp P01859 IGHG2_HUMAN	GFYPSDISEWESENQQPENNYKTPPVMLDSDGSFFLYSKLTVDKSRWQQGNVFCSVMHE
30 sp P01860 IGHG3_HUMAN	GFYPSDIAVEWESSGQPENNYNTTPMLDSDGSFFLYSKLTVDKSRWQQGNIFSCSVMHE
31 sp P01861 IGHG4_HUMAN	GFYPSDIAVEWESNGQPENNYKTPPVLDGSFFLYSRLTVDKSRWQEGNVFCSVMHE
32	
33 sp P01857 IGHG1_HUMAN	ALHNHYTQKSLSLSPGK
34 sp P01859 IGHG2_HUMAN	ALHNHYTQKSLSLSPGK
35 sp P01860 IGHG3_HUMAN	ALHNRTQKSLSLSPGK
36 sp P01861 IGHG4_HUMAN	ALHNHYTQKSLSLSLGK

In this alignment, when we change the format of our file, we can see that we have an alignment with upper and lower case letters. We are interested in upper case letters. The upper case letters are the ones aligned with the HMM whereas the lower case letters have not aligned with the HMM. We will only be interested in the upper case letters region as they

are the ones with the HMM information. Thus, the lower case letters do not give us any information.

CONSERVED REGIONS IN ALIGNMENTS

For this part, we have only focused on the **heavy constant chain**. The reason why we only choose the constant region, is that the heavy chains contain a variable region of 110 amino acids long, that differs depending on the B cell that produced it. So, it is clear that if we compare them we will not find a similarity between them as they will vary for each isotype. In order to compare sequences, we decided to look at **subclasses** as they have a unique profile with respect to the length of the **hinge region**, the number of inter-chain disulfide bonds, and Fc-effector functions.

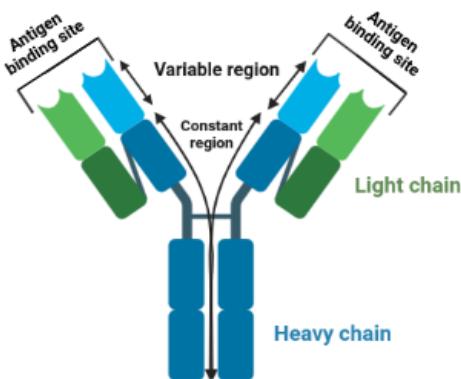


Figure 11. Parts of an antibody. Source: Biorender

CLUSTAL W(1.60) multiple sequence alignment



Hinge region (blue): It is a joint of heavy chains between the Fab and Fc regions. Its unique structure and position provide segmental flexibility, which is essential for normal functioning of antibodies. What differs between each subclass is the length of the region, which in fact, determines the flexibility of the molecule. We can conclude that the flexibility decreases in this order IgG3 > IgG1 > IgG4 > IgG2. Consequently, if we observe the hinge region in the alignment we can see variability that is responsible for each subclass flexibility but all the sequences start with a glutamic acid (E) and finish with a proline (P). In depth, proline has a unique chemical structure that affects the conformation of a protein and is known to be involved in stabilizing the structure of loops and turns in proteins. So in our immunoglobulin, as we know the hinge region is a loop we can say that this amino acid will be responsible for stabilizing the structure in order to continue with the Fc region that is made up of helices and beta-sheets.

Disulfide bonds: Are a covalent bond between two sulfur atoms ($-S-S-$) formed by the coupling of two thiol ($-SH$) groups. The typical amino acid where we find ($-SH$) group in its side chain, is Cysteine (C) which can easily be dimerized to cysteine in aqueous solution by forming a disulfide bond. This type of bonds are important in protein structure because they help proteins to fold and remain in their tertiary and quaternary structure. We found two types: interchain and intrachain disulfide bonds. Intrachain (yellow) are formed between two Cysteines within the same protein chain and are usually conserved. If you observe the alignment, the yellow color, that represents the intrachain disulfide bonds that put together the first cysteine with the last, you can affirm that for all subclasses are conserved. For the

contrary, interchains (dark and light green) are formed between Cysteines of individual chains of the same protein. In the case of immunoglobulin, you can find ones that get together with the light chain (light green) and the ones that get together with the heavy chain or heavy chain dimers (dark green). Neither of these are conserved at all as each subclass has a different quantity of disulfide bonds. We can observe that the IgG3 is the one with more disulfide bonds, exactly 11. IgG1 has 2 the same as IgG4. We can conclude that IgG3 has the highest flexibility to stabilize the tertiary structures of the protein.

Fab region constant part (purple line): Is the region that binds to antigens. For all antibodies it is composed of one constant and one variable domain of each of the heavy and the light chain. In the case of the sequences selected we only have the constant heavy chain. On the alignment, we can observe that the vast majority of amino acids that we observe are conserved for all the subclasses. Even though we have very little quantity of amino acids that have conservation between groups of strongly similar properties and others with weakly similar properties.

Fc region (dark blue line): As Fc-region is responsible for mediating the interaction with many of the effector functions of the immune system following antigen binding. It contains a binding epitope for the neonatal Fc receptor (FcRn). We have more variability between amino acids but we can even observe many of them conserved. In concrete, the first part has more variability, the one that goes after the hinge region between the amino acids PCP...KAK. The reason we find this variability is because here is where the different Fc receptors bind. So depending on the receptor that binds, the region will vary.

Asparagine (red): glycosylation in those amino acids is needed for the interaction with the Fc receptors and the ability to activate the complement pathway. If for the contrary, we had a deglycosylation by S.pyogenes EndoS or EndoS2 endoglycosidases (specific enzymes), it would block the interaction between IgG and Fc receptors, impairing the ability to activate the complement pathway. For this reason, we can observe that in all proteins this amino acid is conserved as it is needed for a right functioning of the molecule.

STRUCTURE

PDB STRUCTURES OF PROTEIN FAMILY AND PSIBLAST

In order to get a set of non biased structures in the PDB database that belong to our protein family, we use PSI-BLAST. PSI-BLAST is an iterative program that performs a sequence search with a Position Specific Substitution Matrix (PSSM).

As of previous knowledge, we know that PDB is a redundant database. Therefore, we use Uniprot, a non-redundant protein sequence database. The PSSMs built with this database will represent the evolutionary information of our target protein with much more accuracy. To find this PSSM we use the following command line:

```
psiblast -query target.fa -num_iterations 5 -out_pssm target_sprot5.pssm -out
target_sprot_5.out -db /shared/databases/blastdat/uniprot_sprot.fasta
```

Then, we use the PSSM obtained in the last command to make a sequence search in the PDB database:

```
psiblast -db /shared/databases/blastdat/pdb_seq -in_pssm target_sprot5.pssm -out
target_pdb_sprot5.out
```

The output file contains the PDB IDs of the proteins with known structure.

		Score (Bits)	E Value
36			
37	Sequences producing significant alignments:		
38			
39	1hzh_K mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN	216	2e-66
40	1hzh_H mol:protein length:457 IMMUNOGLOBULIN HEAVY CHAIN	216	2e-66
41	2ig2_H mol:protein length:455 IGG1-LAMBDA KOL FAB (HEAVY CHAIN)	214	9e-66
42	1igy_D mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.1.3	195	2e-58
43	1igy_B mol:protein length:434 IGG1 INTACT ANTIBODY MAB61.1.3	195	2e-58
44	1igt_D mol:protein length:444 IGG2A INTACT ANTIBODY - MAB231	193	2e-57
45	1igt_B mol:protein length:444 IGG2A INTACT ANTIBODY - MAB231	193	2e-57
46	1mco_H mol:protein length:428 IGG1 MCG INTACT ANTIBODY (HEAVY CHAIN)	192	3e-57
47	2rcj_T mol:protein length:523 Light chain	191	5e-56
48	2rcj_S mol:protein length:523 Light chain	191	5e-56
49	2rcj_P mol:protein length:523 Light chain	191	5e-56
50	2rcj_O mol:protein length:523 Light chain	191	5e-56
51	2rcj_L mol:protein length:523 Light chain	191	5e-56
52	2rcj_K mol:protein length:523 Light chain	191	5e-56
53	2rcj_H mol:protein length:523 Light chain	191	5e-56
54	2rcj_G mol:protein length:523 Light chain	191	5e-56
55	2rcj_D mol:protein length:523 Light chain	191	5e-56
56	2rcj_C mol:protein length:523 Light chain	191	5e-56
57	3b43_A mol:protein length:570 Titin	166	2e-46
58	2gj7_B mol:protein length:227 Ig gamma-1 chain C region	156	6e-46
59	2gj7_A mol:protein length:227 Ig gamma-1 chain C region	156	6e-46
60	1t89_B mol:protein length:224 recombinant IgG1 heavy chain	156	6e-46
61	1t89_A mol:protein length:224 recombinant IgG1 heavy chain	156	6e-46
62	1t83_B mol:protein length:224 IGG1	156	6e-46
63	1t83_A mol:protein length:224 IGG1	156	6e-46
64	2j6e_B mol:protein length:232 IG GAMMA-1 CHAIN C REGION	156	7e-46
65	2j6e_A mol:protein length:232 IG GAMMA-1 CHAIN C REGION	156	7e-46
66	2dts_B mol:protein length:223 Ig gamma-1 chain C region	156	9e-46
67	2dts_A mol:protein length:223 Ig gamma-1 chain C region	156	9e-46
68	2dtq_B mol:protein length:223 Ig gamma-1 chain C region	156	9e-46
69	2dtq_A mol:protein length:223 Ig gamma-1 chain C region	156	9e-46
70	2ql1_A mol:protein length:225 IGHM protein	155	1e-45
71	3c2s_A mol:protein length:225 IGHM protein	155	1e-45

Moreover, from our output file, we selected four proteins that we thought were significantly different from each other but always kept in mind that they had to be in the heavy chain.

The four PDB IDs that have been chosen: 1zhz (Immunoglobulin heavy chain), 1oqx (Immunoglobulin gamma-1 heavy chain), 2gj7 (Ig gamma-1 chain C region), 1e4k (Fc fragment of human IgG1).

SUPERIMPOSE OF STRUCTURES AND ANALYSIS

First of all, we superimpose our structures, selected before, by putting them in Pymol and running the next command which gives us the RMSD value. We selected only the fc region with the amino acids that appear on the PDB because it is the only way to have all the structures from the same part of the structure as many PDB only contains parts of the antibody. The command we use:

```
super 1zhz and resi 236-300, 1oqx and resi 236-300, object=aln →RMSD: 0.810
```

We repeat this command for the rest of the structures. When we do a superimposition with 2gj7 and 1oqx, we obtain a value of RMSD equal to 0.460. Finally, when we do the last one between 1e4k and 1oqx the RMSD value is 3.017.

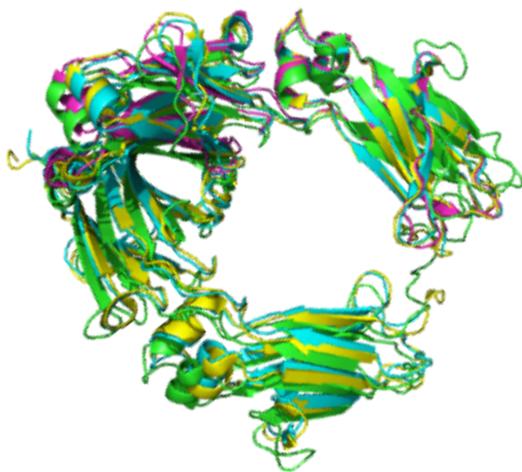


Figure 12. Superimposition of 1e4k (colored in green), 2gj7 (colored in purple), 1zhz (colored in yellow) and 1oqx (colored in blue). *Source: PyMOL*

After doing this, we can say that the smaller the RMSD value we obtain, the more similar the two structures are. So structurally, 2gj7 and 1oqx are the ones more similar. Then, 1zhz is also similar but looking at 1eqx we observe that the RMSD value is higher than the rest as it is not very similar.

Our structures are very similar but we can observe that one part of the molecule has more variability between structures. The region with more variability is the one that even 2gj7 (purple) doesn't appear. This part corresponds to the start of the amino acid sequence, where N-glycosylation occurs and where the Fc-receptors bind, for this reason we see this

small variability. We know that some of these structures bind a Fc-receptors III like in 1e4k or proteins such as glycoprotein E in 2gj7. For this reason they are not similar at all.

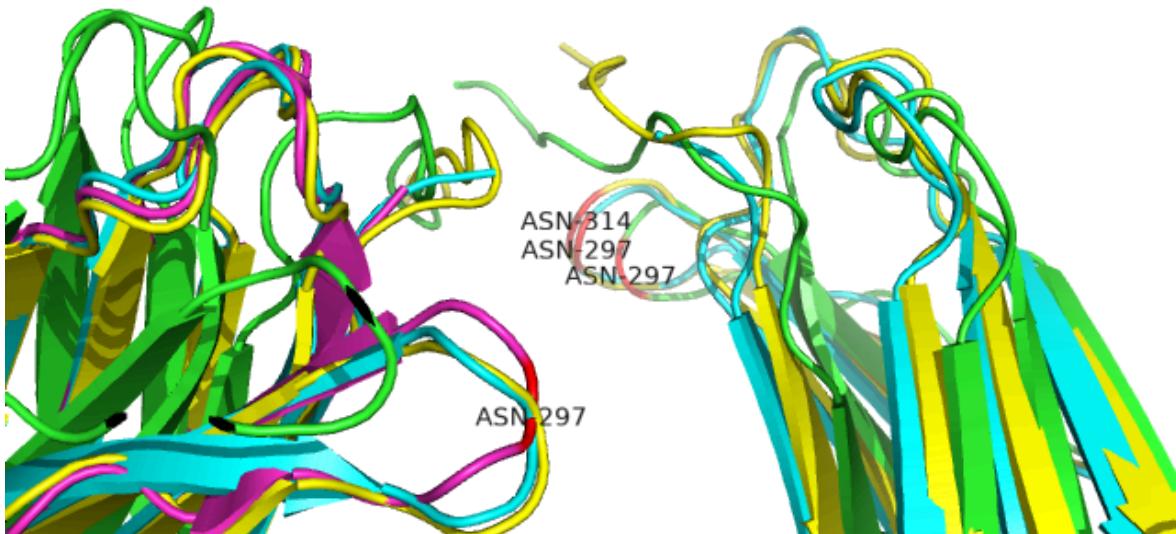


Figure 13. Marked in red the Asparagine amino acids. Source: PyMOL

We have selected the Asparagine residues that correspond to the ones that do glycosylation in the structure to observe where they appear and if they are conserved. These amino acids are needed for the interaction with Fc receptors, so in these structures are conserved because if they wouldn't appear, the molecule Fc region wouldn't function right. Even though, we can observe that one Asparagine residue that is on a loop is not in the same place as the others but they are on the same chain.

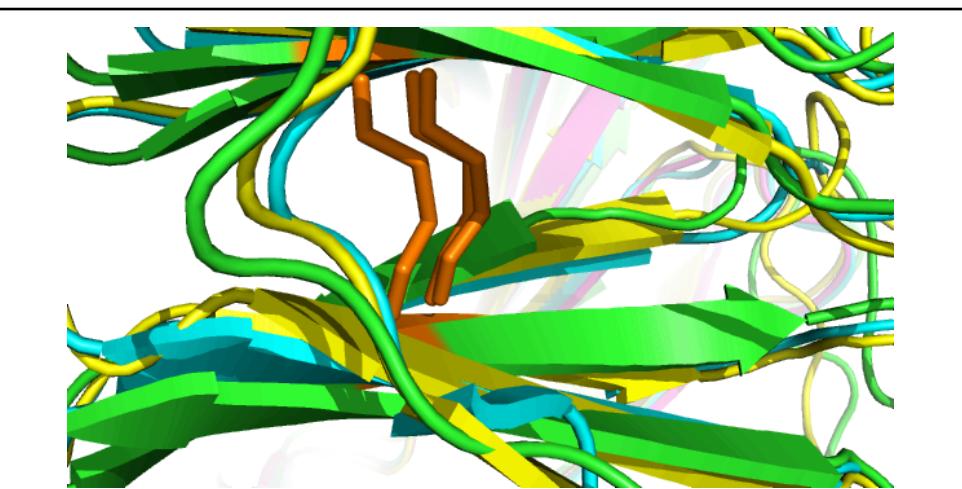


Figure 14. Marked in orange the disulfide bonds in structures 1hzh, 1ek4, 1oqx. Source: PyMOL

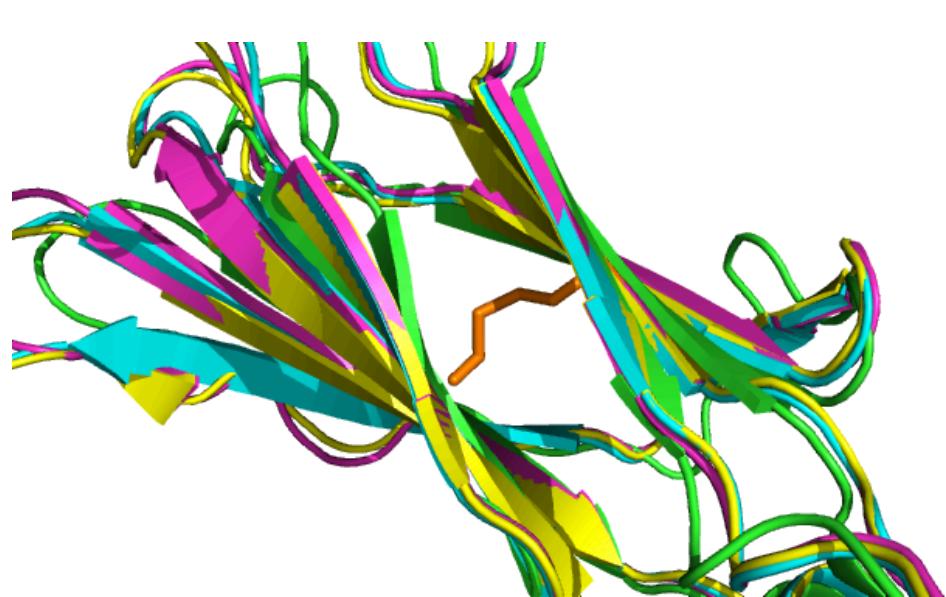


Figure 15. Marked in orange the disulfide bond in structure 2gj7. Source: PyMOL

As we are observing the Fc region we can only look at the intrachain disulfide bonds in the structures. The other ones don't appear in this region. We can observe that these disulfide bonds are conserved because they fix and stabilize the tertiary structure of a protein. In the case of IgG, the tertiary structure we find is the beta-sheets so when they make the disulfide bond they stabilize this structure by reducing the entropy of the denatured state. If the Cysteines were not conserved, the bonds wouldn't be able to be created, so the structure of the antibody wouldn't be able to stabilize the folding of a single polypeptide chain.

The function of the IgG is to act as a critical part of the immune response by specifically recognizing and binding to particular antigens. More exactly, the Fc region ensures that IgG generates an appropriate immune response for a given antigen. It does this by binding to specific proteins and cell receptors. It tries to maximize the receptor-mediated functions. To do so, these regions must be conserved because if we found a mutation on this region the effector functions wouldn't act. As a consequence, the interaction with immune cells wouldn't occur.

MONOCLONAL ANTIBODY ENGINEERING IN IgG

FcRn is a unique Fc receptor encoded by the **FCGR1** gene (chr 19), different from the Fc gamma receptor family; its binding site on IgG Fc is located at distance from that of Fc gamma receptor, not depending on glycosylation. FcRn functions as a recycling receptor that is responsible for maintaining IgG and albumin in the circulation, and bidirectionally transporting these two ligands across polarized cellular barriers. More recently, it has been appreciated that FcRn acts as an immune receptor by interacting with and facilitating antigen presentation of peptides derived from IgG immune complexes (IC).

Now we will focus on the interaction between therapeutic antibody and neonatal Fc receptor (FcRn), giving an understanding of the **FcRn-IgG interacting** that allows development of effective therapeutics and avoidance of potential adverse effects. In order to do that we used

a complex formed by the **Neonatal Fc receptor** (FcRn), serum **albumin**, **Fc receptor** and **Beta-2-microglobulin** (tumor marker) ([Figure 16.A](#)).

It's important to note that IgG binds FcRn at **acidic** pH (pH ~ 6) and then released at near **neutral** pH 7.4, either on the apical side of endothelial cells (recycling) or on the opposite side (transcellular transport).

Before working with its structure, we review some key aspects of these proteins. The FcRn heavy chain consists of three extracellular domains (α_1 , α_2 , and α_3), a transmembrane domain and a cytoplasmic tail of 44 amino acids. Several crystal structures of FcRn solved by [Burmeister](#) have shown that α_1 and α_2 domains form a platform of eight antiparallel β -strands with two α -helices, while β_2 is non-covalently associated with the α heavy chain. FcRn binds with high affinity to IgG and albumin through **non-overlapping sites at mildly acidic pH of 5.0-6.5** and exhibits no binding at neutral pH.

The interaction we are observing more in deeply is how FcRn interacts with the Fc portion of IgG, occurring at the CH₂ and CH₃ domain interface, and involves the **IgG Fc residues Ile253** and two central histidines: **His310** and **H435**.

On one hand, the **nitrogen of Isoleucine** (253) is interacting with **both oxygens of Glutamic Acid** (133) (Figure B), most likely this is a **hydrogen bond** known as **bifurcated**. This happens when a single hydrogen bond donor (in this case the nitrogen) forms hydrogen bonds with two acceptor groups (both oxygens of glutamic acid). They are relatively strong and contribute significantly to the stability of the protein structure, playing important roles in mediating interactions. Looking closely we observe that these interactions are occurring at 2.7 and 3.4 Å. In general, the optimal distance for hydrogen bonding is around 2 to 3 Å, still hydrogen bonds can be formed and be relatively strong.

So, if the nitrogen of isoleucine is interacting with both oxygens of glutamic acid at distance of 2.7 and 3.4 Å, it is possible that a bifurcated hydrogen bond could form and contribute to the stability of the protein structure.

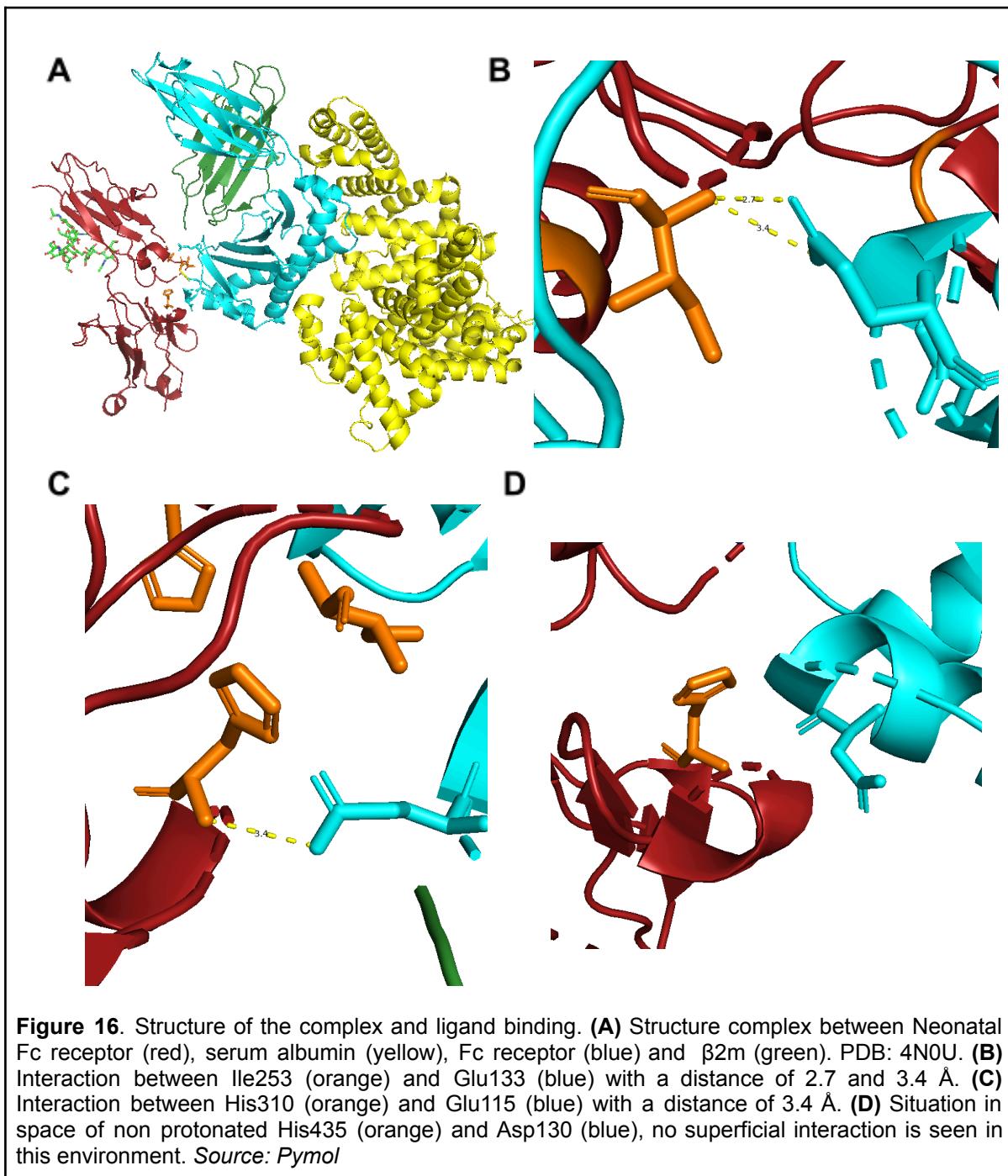
On the other hand, when talking about **histidines**, pH is a key for binding. Histidine residues become **protonated at pH ~6 which allows for interaction with the FcRn residues Glu115 and Asp130**. If we increased the pH, histidine protonation would be lost, this is why the interaction is pH dependent.

First interaction when protonated is happening between **Histidine** (310) and **Glutamic Acid** (115) (Figure C). If a histidine is protonated it means that the side chain has donated a proton, therefore the positively charged group of the protonated histidine can interact with the negatively charged carboxylate group of the glutamic acid through the formation of a **salt bridge** facilitating protein-protein interaction.

Then, interaction between **Histidine** (435) and **Aspartic Acid** (130), two acidic amino acids (can ionize and become protonated at low pH) is happening when histidine is protonated. As before, the **positively charged** group of histidine will interact with the **negatively charged** carboxylate group of the aspartic acid through the formation of a **salt bridge**, which is an electrostatic interaction occurring between oppositely charged groups. The formation of these types of bonds contribute to the stability of proteins by stabilizing the protein structure and facilitating protein-protein interactions.

The specific type of interaction when histidine is protonated depends on the protonation states of the two amino acids and the local environment of the protein.

Concluding that the mutation of the IgG residues **Ile253, His310 and His435 (IHH)** leads to complete abrogation of FcRn binding at pH 6.



MODIFICATIONS AT DIFFERENT pH TO STUDY IgG HALF-LIFE, AFFINITY AND RECYCLING PHASE

In the following we will see why the IgG Fc-FcRn is interesting for the development of certain drugs, taking into account that its clinical responses require a **long plasmatic half-life** (the increase in mAb¹ half-life allows maintenance of high plasma levels of therapeutic antibodies over longer periods of time) and a good drug exposure, both of which depend on the previous interaction.

Several modifications at different pH have been examined in order to study IgG half-life, affinity or the recycling phase. The following mutations and its results are extracted from "[Monoclonal Antibody Engineering and Design to Modulate FcRn Activities: A Comprehensive Review.](#)" Ramdani Y, Lamamy J, Watier H, Gouilleux-Gruart V.

Firstly, they performed mutations close to the FcRn binding site, increasing the affinity at acidic pH. Involving mainly the **CH2** and **CH3** domains of IgG, where **histidines**, main residues responsible for ligand binding, were mutated to **alanine**: I253A, H310A and H435A. Conversely, many combined or isolated mutations affecting T250, M252, S254, T256, T307, E380, M428, H433 or N434 have been described to enhance mAb affinity to FcRn only at acidic pH~6

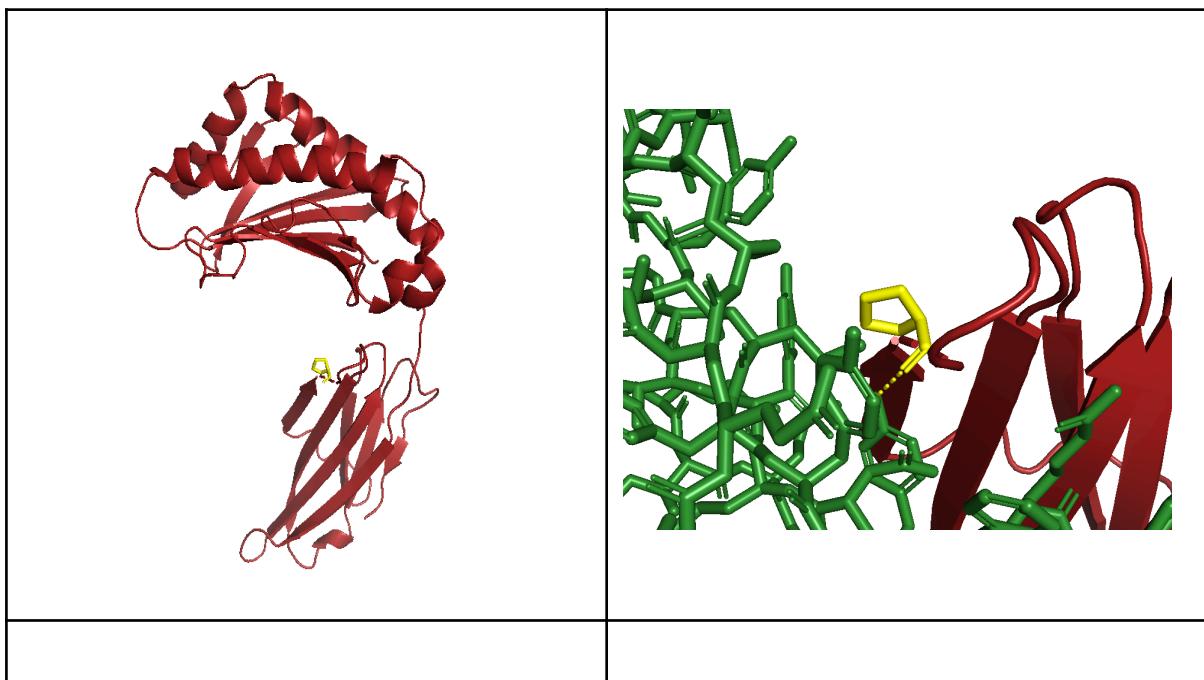
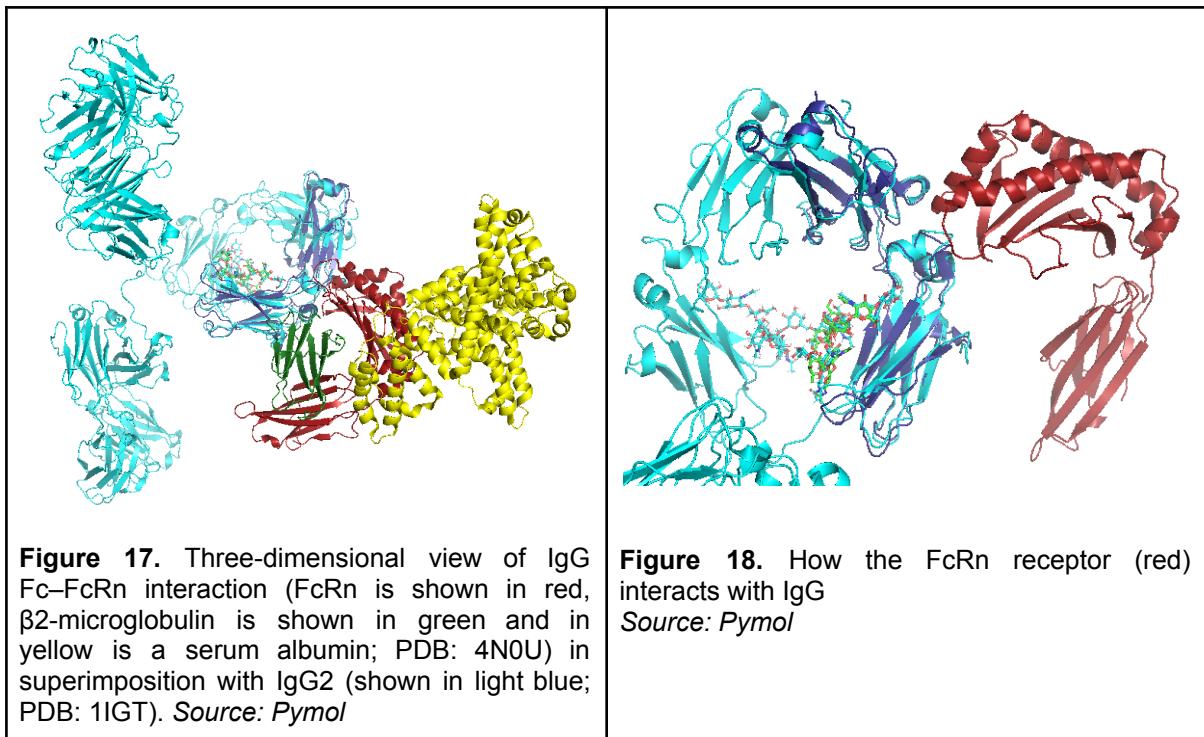
Some mutations or its combination increased **mAb half-life**, such as M252Y/S254T/T256E (YTE), T250Q/M428L (QL) and T307A/E380A/N434A (AAA), as reviewed by [Wang et al.](#) and [Kuo et al.](#)

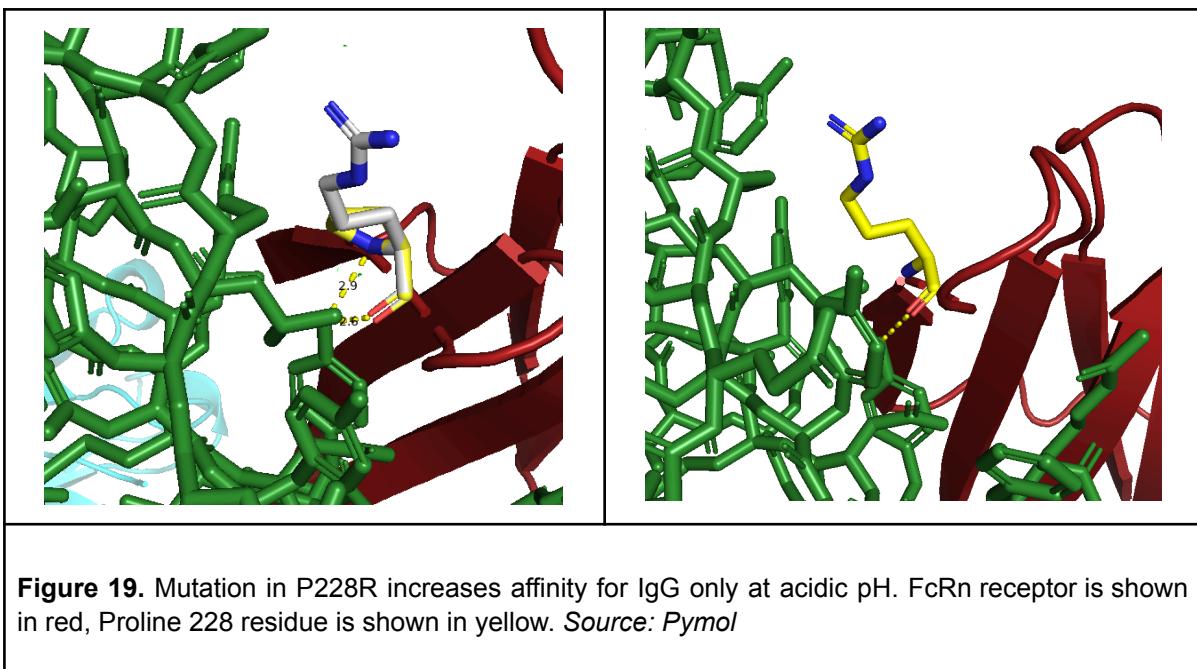
Secondly, they observed some modifications outside FcRn binding site that **increases affinity for IgG only at acidic pH**, i.e. **P228R** or L and **P230S** (located in the hinge region and interestingly, located at distance from the FcRn binding site).

However, they were seeing that FcRn recycles Immune Complexes as efficiently as free mAbs, which could be an issue resulting in plasma accumulation of the soluble antigen complexed with IgG. To solve this they generated antibodies capable of releasing their target endosomes and then being recycled by FcRn freed from their antigen. They generated a **M428L/N434A** mutant. Furthermore, for a neutral pH they saw that using mutations M252W/N434W, M252Y/N434Y, M252Y/N268E/N434Y (YEY) or M252Y/V308P/N434Y (YPY) they also performed better.

Finally, they found a strategy that is being used in advanced stages of clinical trials for the treatment of *myasthenia gravis* (a rare long-term condition that causes muscle weakness) for example. **Rozanolixizumab** (**S241P**, stabilize hinge region on IgG4) and **batoclimab** (**L234A, L235A** on IgG1) showed encouraging results in some phase trials of the mentioned disease and a reduction of endogenous IgG levels.

¹ Monoclonal antibody: type of targeted drug therapy, recognize specific proteins or cancer cells

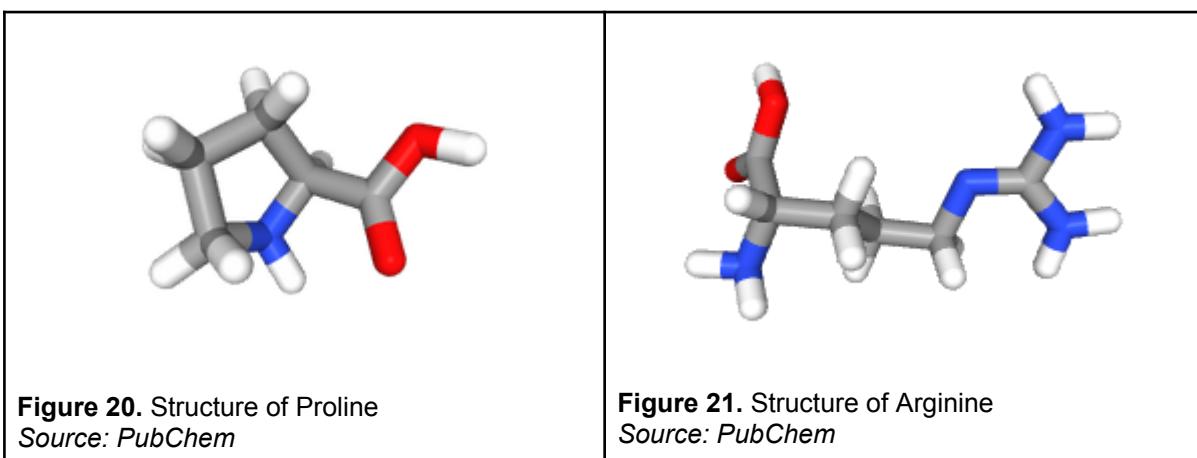




IS THE MUTATION ON THE MONOCLONAL ANTIBODY COMPROMISING THE STABILITY OF THE RECEPTOR?

Proline ([Figure 20](#)) is a special case of amino acid, since it is the only **cyclic** amino acid and cannot donate a hydrogen bond to stabilize an α helix or a β sheet (it would slightly bend due to the lack of hydrogen bond). In our case, proline is found at the beginning of a loop, which is more often found. Whereas, **arginine** ([Figure 21](#)) is a **positively charged** amino acid and can form salt bridges and stabilize protein structure.

When referring to evolution, this mutation can have an evolutionary advantage: providing a new **function** or improving the **stability** of the protein.



But, the impact of this change depends on the context of the mutation and the protein being studied. Therefore we performed a **PROSA analysis** to infer if any energy change was compromising the stability of the protein:

First we corrected both PDB files (one for the FcRn receptor and the other for the mutated version) since we only are interested in the chain A, which contains the position where the mutation is introduced. Then, after aligning both sequences and creating a ***pir file***², we ran **MODELLER** to obtain a reliable model for our target protein.

Now, we can **evaluate** the **quality** and reliability of our mutation model by using **statistical potentials**. Using **PROSA**, a program that enables us to view the energy profiles of the generated models, we compared results for both the FcRn receptor and the FcRn with the mutation introduced.

As expected we only observe a small change in the **Z-score** (measure of how far away a given energy value is from the mean energy of a set of conformations). The **mutated one** ([Figure 22](#)) is **-0.1 lower than the template** ([Figure 23](#)), which tells us that is a bit more stable (still not a significant result), we are not able to see a bigger difference since this is one of the multiple mutations we could introduce. Furthermore we have to take into account that this specific change **increases the affinity of the receptor with the IgG in an acidic environment** (pH ~ 6.5), which we cannot recreate with these tools (because we are not analyzing protein-protein interaction).

In addition, the energy profiles for this analysis do not give us more information, only one change in a sequence, which is not very drastically different, cannot be seen. But, by comparing these profiles we can identify that our model is **well modeled**, because **overall the energies are low**.

To conclude our results we also analyzed our protein structures with **QMEAN** ([Figure 24](#) and [Figure 25](#)). This program will plot a quality score based on their own score: **QMEANDisCo**. The global score is the average per-residue score and the error estimate is based on global scores estimated for a large set of models. QMEAN shows an extremely similar **overall quality score for the receptor and the mutated one**, perceiving small differences in those windows close to the change of amino acid, however, we cannot see any dramatic change as we anticipated.

² Protein information resource: annotated, non-redundant and cross-referenced database of protein sequences at the National Biomedical Research Foundation (NBRF).

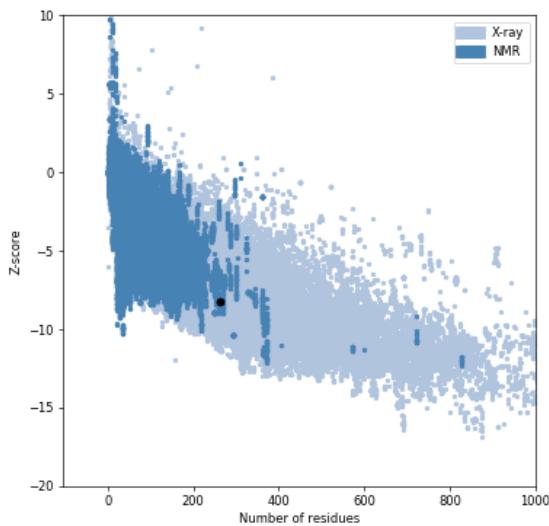
Overall model qualityZ-Score: **-8.27**

Figure 22. FcRn receptor results.
Source: PROSA

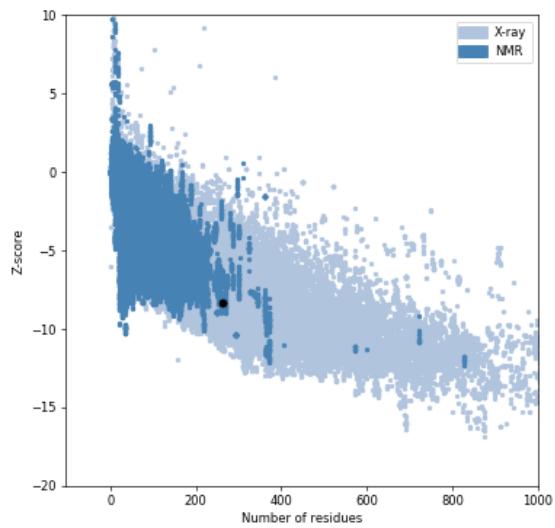
Overall model qualityZ-Score: **-8.37**

Figure 23. FcRn mutated receptor results.
Source: PROSA

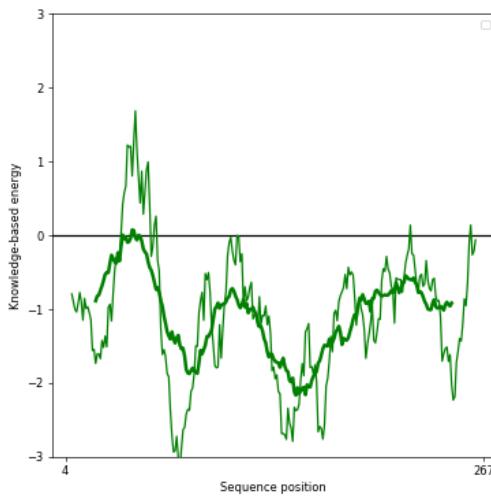


Figure 22.1. Energy profile for FcRn receptor
Source: PROSA

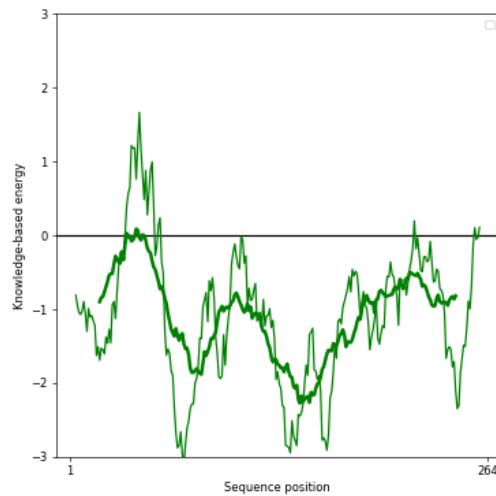
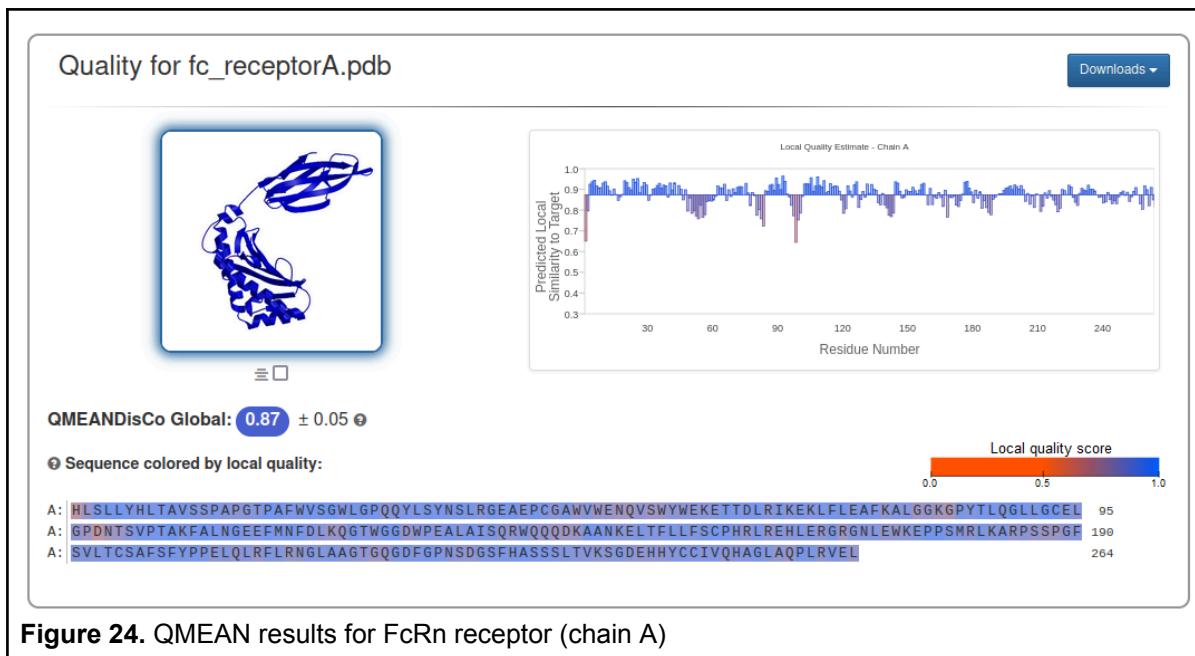
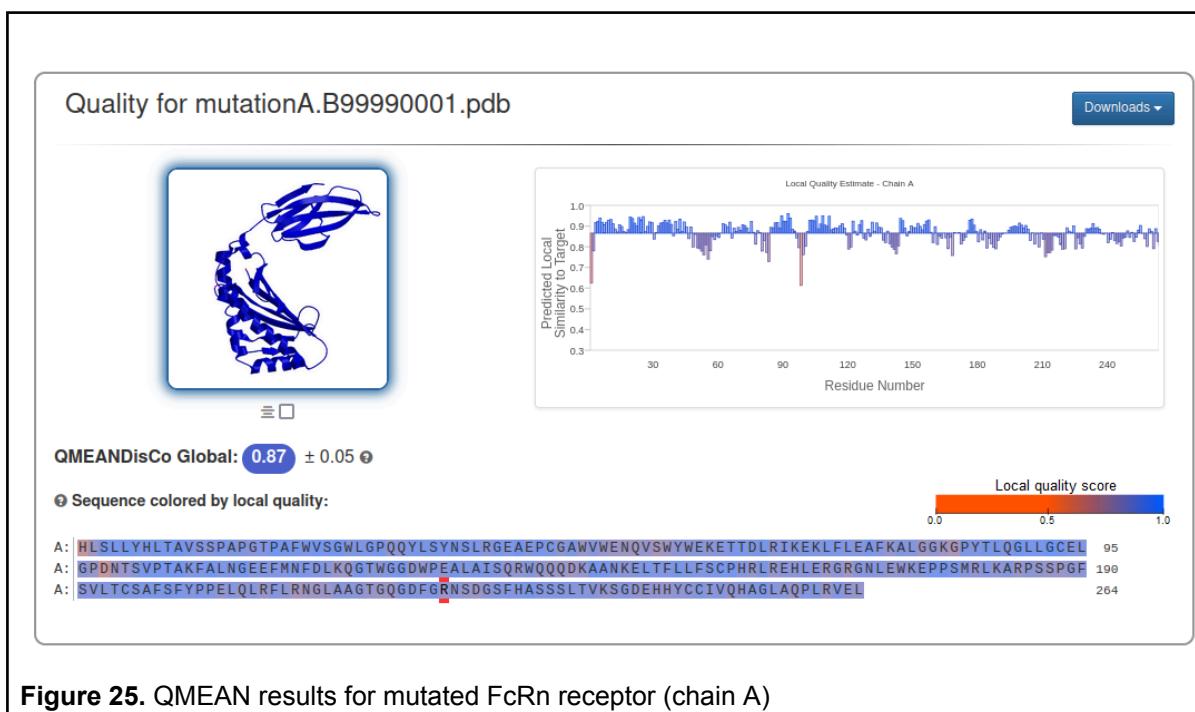


Figure 23.1. Energy profile for FcRn mutated receptor
Source: PROSA

**Figure 24.** QMEAN results for FcRn receptor (chain A)**Figure 25.** QMEAN results for mutated FcRn receptor (chain A)

HOW COULD THE MUTATION BE AFFECTING THE PROTEIN-PROTEIN INTERACTION?

In the previous point we saw that the change in amino acid is **not compromising the stability of the FcRn receptor**. Nevertheless, the important object to study would be the interaction between the **receptor and IgG**, which is in fact, increasing its binding affinity. In consequence, it can increase effector functions such as antibody dependent cell cytotoxicity or complement-mediated cell cytotoxicity.

In [Figure 19](#) we can see the interaction between Proline 228 and Tyrosine 10 and how, when introduced the mutation P228R, the distance is closer but the contact still remains between Arginine and Tyrosine. But, what do we think is in reality happening?

Our hypothesis is that this mutation is affecting the secondary structure of the receptor. Proline 228 is located in a coil between two β -strand situated in an antiparallel β -sheet, taking into account its properties, we know that it cannot donate a hydrogen bond, which makes it impossible to be part of a β -strand. The opposite of what arginine is providing. Therefore, what we suggest is that the **new amino acid disrupts the loop to be part of a strand** (β -strand is getting longer with its addition), which could somehow be changing the affinity with the immunoglobulin.

Our approach at that point was to use statistical potentials to predict secondary structures. To assess this prediction we used **PSIPRED** and **DSSP**. First we started with two pdb models: FcRn receptor and the mutated version, both only containing chain A where the amino acid is located. Then we executed PSIPRED to predict secondary structures on both proteins. But not all predictions are equally reliable. We should identify the confidence for our amino acid.

In Figure [26.1](#) and [26.2](#) there is the assignment of the predicted structure for each residue in the sequence. The change of residue makes the loop become a strand, in consequence of proline's loss, following that we should find a higher confidence for a loop prediction in the wild type than in our model. In fact, this is what we find, wild type shows a confidence of 8 for the coil structure, whereas, our model shows a score of 6.

Figure 26.1. Predicted structure for the wild type receptor. We show Proline in the highlighted red square with a confidence score of 8 predicting coil.

```

# PSIPRED HFORMAT (PSIPRED V4.0)

Conf: 941101444017999999713453203746701332018875356407884043446402
Pred: CCCCEHHHHCCCCCCCCCCEEEEEECCCHHHCCHHCCCCCCCCCCEEEECCEEECCC
AA: HLSLLYHLTAVSSPAPGTPAFWVSGWLGPQQQLSYNSLRGEAEPGCAWVWENQVSWYWEK
      10          20          30          40          50          60

Conf: 47884288999999998379887206784007638998988733566386010060267
Pred: CCCCCHHHHHHHHHHHHHHCCCCCCCCCHHHCCCECCCCCCCCCEEEECCCCCCCCCC
AA: ETTDLRIKEKLFLAEFKALGGKGPYTLQGLLGCELPDNTSVPTAKFALNGEEMNFDLK
      70          80          90          100         110         120

Conf: 8877888899999999881982144858954978999998088877568997530
Pred: CCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCC
AA: QGTWGWDWPEALAIISQRWQQDKAANKELTFLFSCPCHRRLREHLERGRNLIEWKEPPSMR
      130         140         150         160         170         180

Conf: 4616899980179874013149589999974531557888756898972642103765
Pred: EEECCCCCEEEEEECCHECCCCHHHHHHHHCCCCCCCCCCCCCCCCCCCCCEECC
AA: LKARPSSPGFSVLTCASFYPPELQLRFLRNGLAAGTGQQDFGRNSDGSFHASSLTVK
      190         200         210         220         230         240

Conf: 788412799987065346501039
Pred: CCCCEEEEEEECCCCCCCCCECC
AA: SGDEHHYCCIVQHAGLAQPLRVEL
      250         260

```

Figure 26.2. Predicted structure for our model receptor. Arginine is shown in the highlighted red square where the coil prediction score is lower than the original.

Then we contrasted the other output file, one containing the scores associated with the predictions done, **probabilities** of each amino acid to be a **coil, helix or strand** are shown respectively in Figure [27.1](#) and [27.2](#). If we look closely at the one residue that was changed (shown in Figure 22 in position 225, from P to R), for both amino acids, the highest confidence is for the prediction of coil. But the difference remains that when the **arginine** is

present in that particular position, the probability of being a **coil decreases** while the probability of being a sheet increases. Therefore, our initial suggestion is correct.

224 G C	0.920	0.017	0.065
225 P C	0.880	0.022	0.064
226 N C	0.945	0.022	0.026
227 S C	0.932	0.048	0.021

Figure 27.1. Probabilities of proline to be a coil, helix or sheet.

224 G C	0.721	0.113	0.176
225 R C	0.785	0.040	0.137
226 N C	0.895	0.032	0.059
227 S C	0.936	0.032	0.030

Figure 27.2. Probabilities of arginine to be a coil, helix or sheet.

To further contrast our hypothesis we aligned the results between PSIPRED and the DSSP results, if we compared the secondary structure prediction with the real secondary structure, we would obtain information about how the arginine is affecting the structure and if it is in reality making a change from a coil to a strand.

```
CLUSTAL W(1.60) multiple sequence alignment
mutationA.ss2Seq    QGTWGGDWPEALAISQRWQQQDKAANKELTFLLFSCPRLREHLERGRGNLEWKEPPSMR
mutationA.ss2Seq    HLSLLYHLTAVSSPAPGTPAFWVSGWLGPQQYLSYNSLRGEAEPGAWVWENQVSWYWEK
mutationA.ss2SS     CCCCEHHHHCCCCCCCCCCCCEEEEEECCCHHHHCCCCCCCCCCCCCEECCCEECC
mutationA.dsspSeq   HLSLLYHLTAVSSPAPGTPAFWVSGWLGPQQYLSYNSLRGEAEPGAWVWENQVSWYWEK
mutationA.dsspSS   --EEEEEEESS--TTS--SEEEEEEETTEEEEEEETTT---EE--GGGGGS---TTHHHH
mutationA.ss2Seq   LKARPSSPGFSVLTCASFYPPPELQLRFLRNGLAAGTGQGDFGRNSDGSFHASSSLTVK
mutationA.ss2Seq   ETTDLRIKEKLFLAEFKALGGKGPYTLQGLLGCELGPNTSVPTAKFALNGEEFMNFDLK
mutationA.ss2SS    CCCCCHHHHHHHHHHHHCCCCCCCCCHHHCCCEECCCCCCCCCCCCCEECCCCCCCCCCC
mutationA.dsspSeq   ETTDLRIKEKLFLAEFKALGGKGPYTLQGLLGCELGPNTSVPTAKFALNGEEFMNFDLK
mutationA.dsspSS   HHHHHHHHHHHHHHHHHHHH--SSS--EEEEEEEEEETTTEEEEEEETTTEEEEEEETT
mutationA.ss2Seq   SGDEHHYCCIVQHAGLAQPLRVEL
mutationA.ss2Seq   QGTWGGDWPEALAISQRWQQQDKAANKELTFLLFSCPRLREHLERGRGNLEWKEPPSMR
mutationA.ss2SS    CCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCC
mutationA.dsspSeq   QGTWGGDWPEALAISQRWQQQDKAANKELTFLLFSCPRLREHLERGRGNLEWKEPPSMR
mutationA.dsspSS   TTEEE--SHHHHHHHHHSSSSHSSHHHHHHIIIIHHHHHHHHHTHHHT--B--EEE

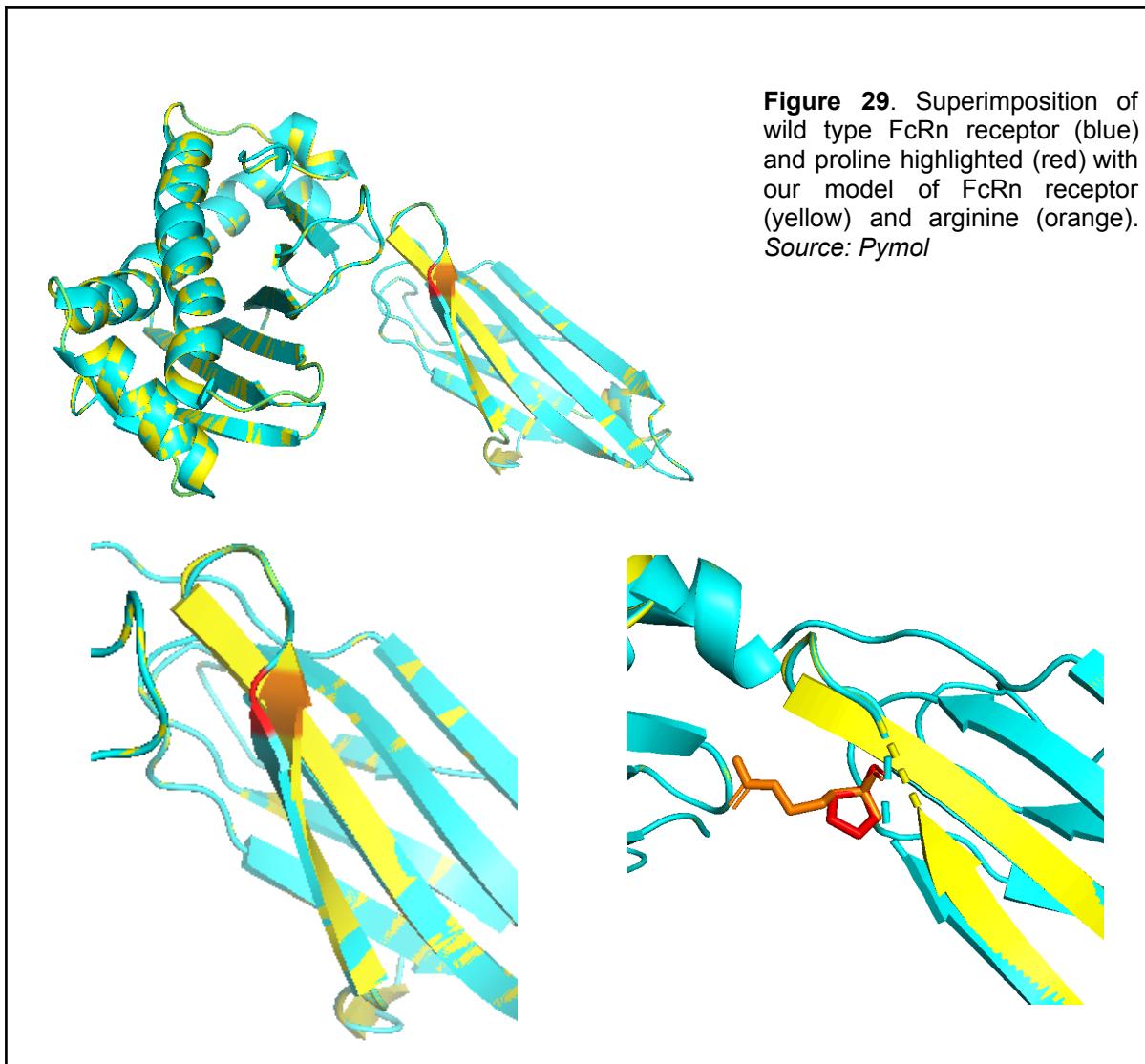
mutationA.ss2Seq   LKARPSSPGFSVLTCASFYPPPELQLRFLRNGLAAGTGQGDFGRNSDGSFHASSSLTVK
mutationA.ss2SS    EEECCCCCCEEEEEECCCEECCHHHHHHCCCCCCCCCCCCCCCCCEECCCEECC
mutationA.dsspSeq   LKARPSSPGFSVLTCASFYPPPELQLRFLRNGLAAGTGQGDFGRNSDGSFHASSSLTVK
mutationA.dsspSS   EEEEE--SSSEEEEEEBS--EEEEEESSB----EEEEEE--TTS--EEEEEEEEE

mutationA.ss2Seq   SGDEHHYCCIVQHAGLAQPLRVEL
mutationA.ss2SS    CCCCCCCCCCCCECCCC
mutationA.dsspSeq   SGDEHHYCCIVQHAGLAQPLRVEL
mutationA.dsspSS   TT--GGEEEEEE--TTSSS--EEE--
```

Figure 28. Clustalw alignment of predicted (PSIPRED) and real (DSSP) secondary structure. First line shows already the change in amino acid (mutated manually in PDB) showing a prediction of C (coil), after modeling the secondary structure we get a prediction of E (strand)

Finally ([Figure 29](#)) we inspected the obtained model using Pymol. We superimposed chain A of the model obtained in the previous steps with the wild type chain A receptor getting an **RMSD of 0.000**. However, **we remark how arginine is forming part of the strand making it longer, a fact that could be compromising the affinity of the protein.**

As said before, these are not concluding results, but they suggest the mechanism by which this mutation could be affecting the binding of the receptor.



REFERENCES

- Pratt G. *Molecular aspects of multiple myeloma*. Mol Pathol. 2002 Oct;55(5):273-83. doi: 10.1136/mp.55.5.273. PMID: 12354927; PMCID: PMC1187254.
- Harrison, C.J., Mazzullo, H., Ross, F.M., Cheung, K.L., Gerrard, G., Harewood, L., Mehta, A., Lachmann, H.J., Hawkins, P.N. and Orchard, K.H. (2002), *Translocations of 14q32 and deletions of 13q14 are common chromosomal abnormalities in systemic amyloidosis*. British Journal of Haematology, 117: 427-435.
- Janet Stavnezer, Carol E. Schrader; IgH Chain Class Switch Recombination: Mechanism and Regulation. *J Immunol* 1 December 2014; 193 (11): 5370–5378.
- Ramdani Y, Lamamy J, Watier H, Gouilleux-Gruart V. Monoclonal Antibody Engineering and Design to Modulate FcRn Activities: A Comprehensive Review. *Int J Mol Sci*. 2022 Aug 24;23(17):9604. doi: 10.3390/ijms23179604. PMID: 36077002; PMCID: PMC9455995.
- Wang Y, Tian Z, Thirumalai D, Zhang X. Neonatal Fc receptor (FcRn): a novel target for therapeutic antibodies and antibody engineering. *J Drug Target*. 2014 May;22(4):269-78. doi: 10.3109/1061186X.2013.875030. Epub 2014 Jan 9. PMID: 24404896.
- Janet Stavnezer, Carol E. Schrader; IgH Chain Class Switch Recombination: Mechanism and Regulation. *J Immunol* 1 December 2014; 193 (11): 5370–5378. <https://doi.org/10.4049/jimmunol.1401849>
- Wines BD, Powell MS, Parren PW, Barnes N, Hogarth PM. The IgG Fc contains distinct Fc receptor (FcR) binding sites: the leukocyte receptors Fc gamma RI and Fc gamma RIIa bind to a region in the Fc distinct from that recognized by neonatal FcR and protein A. *J Immunol*. 2000 May 15;164(10):5313-8. doi: 10.4049/jimmunol.164.10.5313. PMID: 10799893.
- Yanaka S, Yogo R, Inoue R, Sugiyama M, Itoh SG, Okumura H, Miyanoiri Y, Yagi H, Satoh T, Yamaguchi T, Kato K. Dynamic Views of the Fc Region of Immunoglobulin G Provided by Experimental and Computational Observations. *Antibodies (Basel)*. 2019 Jul 1;8(3):39. doi: 10.3390/antib8030039. PMID: 31544845; PMCID: PMC6784063.
- Burmeister W, Gustin L, Simister N, Blum M, Bjorkman P. Crystal structure at 2.2 Å resolution of the MHC-related neonatal Fc receptor. *Nature*. (1994) 372:336–43. doi: 10.1038/372336a0
- Justiz Vaillant AA, Jamal Z, Patel P, et al. Immunoglobulin. [Updated 2022 Nov 24]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK513460/>
- InterPro. Immunoglobulin C1-set: <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR003597/>
- Creative Biolabs. Immunoglobulin-Like β-Sandwich Protein Design: <https://www.creative-biolabs.com/immunoglobulin-like-sandwich-protein-design.html>