UNIVERSITAT DE
BARCELONA

In silico design of antibodies for biomedical applications

José Antonio Amengual Rigo

UNIVERSITAT DE
BARCELONA

In silico design of antibodies for biomedical applications

José Antonio Amengual Rigo

Barcelona, 2020

Memòria presentada per José Antonio Amengual Rigo per optar al grau de Doctor per la Universitat de Barcelona

UNIVERSITAT DE BARCELONA – FACULTAT DE BIOLOGIA PROGRAMA DE DOCTORAT EN BIOMEDICINA 2020

Aquest treball ha estat realitzat en el grup Electronic and Atomic Protein Modeling (EPAM), del Barcelona Supercomputing Center (BSC)

Prof. Victor Guallar Tasies

Director

Prof. Josep Lluís Gelpi Buchaca

Tutor

José Antonio Amengual Rigo

Doctorand

To my family and friends

Abstract

Proteins are large macromolecules constituted by amino acids that are responsible for most of the biological processes within a cell. Proteins showing high complementary affinity may bind forming protein-protein complexes. In this context, antibodies are proteins that recognize abnormal particles in the body (known as epitopes), and are elicited by means of random recombinatory events followed by strict screening selection processes. Along their production, antibodies can be modified by mutation events leading to potent antibody variants.

In this sense, there is an industrial and biomedical interest for the artificial optimization of antibodies. The rise of the computational era together with the deeper understanding of structural biology allowed the design and implementation of predictive algorithms for simulating the effects of mutations in protein-protein complexes. This process usually involves, among others, the prediction of changes in Gibbs free energy upon mutation and the use of other computational simulations for unveiling motions and binding patterns, such as Molecular Dynamics and Monte Carlo techniques.

During this thesis, we have developed and implemented predictive algorithms focused on the design of potent antibody variants. We developed UEP, an open-source code for predicting the effects of mutations in protein-protein complexes. UEP differs from the state-of-the-art and employs other sources of knowledge rather than experimental binding affinity determinations upon mutation. Moreover, we designed a PELE protocol to simulate the binding affinity of antibodies against hypermutated HIV-1 viral isolates. Finally, we describe three different computational workflows for antibody optimization. We particularly focused on the challenge of increasing the binding potency of the N6 antibody, one of the best antibodies against HIV-1. Each computational workflow has been evaluated experimentally by our collaborators from Irsicaixa, and such combined computational and experimental effort resulted in the design of an improved variant of the N6 antibody against HIV-1.

Table of contents

List of abbreviations

ΔG	Gibbs Free Energy
ΔΔG	Gibbs Free Energy change
AIDS	Acquired Immunodeficiency Syndrome
ANM	Anisotropic Network Models
bNAb	Broadly Neutralizing Antibody
CD4bs	CD4 binding site
CDR	Complementarity-determining region
EC50	Half Maximal Effective Concentration
FR	Framework region
gp120	Glycoprotein 120
GPU	Graphical Processing Unit
HIV-1	Human Immunodeficiency Virus 1
IC50	Half Maximal Inhibitory Concentration
MC	Monte Carlo
MD	Molecular Dynamics
PDB	Protein Data Bank
PELE	Protein Energy Landscape Exploration
PPI	Protein-Protein Interaction
SASA	Solvent Accessible Surface Area

Introduction

Proteins

Proteins are large and complex macromolecules formed by the linear addition of amino acids. Proteins play a variety of crucial roles in living organisms including protection against dangerous elements, serving as structural basis, transportation of molecules through the body, catalysis of chemical reactions, storage of ions, signalling processes, among others (Alberts *et al.*, 2002). Proteins are coded by genes, and they are synthetized in two stages: translation and transcription. A scheme of these two stages is illustrated in **Figure 1**.

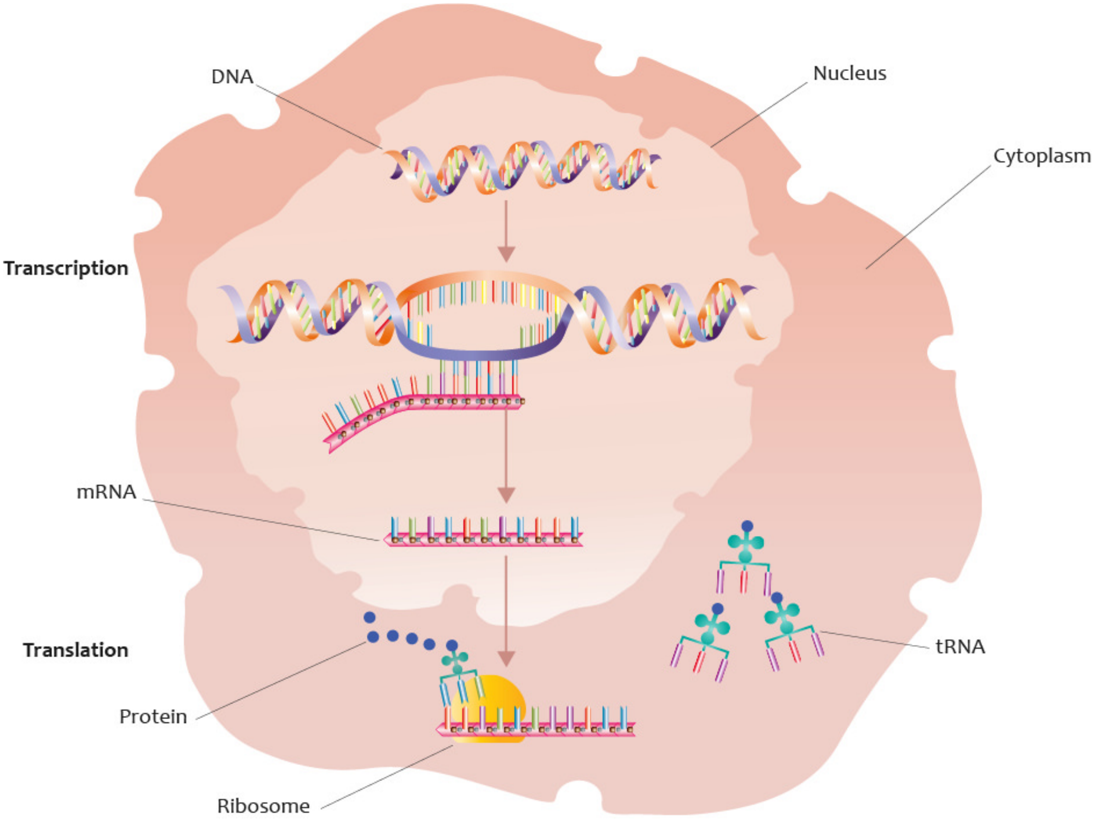


Figure 1. Flow of genetic information from DNA to mRNA, and from mRNA to proteins. The processes of transcription and translation in an animal cell (https://flickr.com/photos/119980645@N06/13083355814) from NHS National Genetics and Genomics Education Centre, licensed by CC BY 2.0 (https://creativecommons.org/licenses/by/2.0/).

Transcription takes place in the nucleus of the cell, and it is the process where genes coded in the deoxyribonucleic acid (DNA) are used as a template to generate messenger ribonucleic acids (mRNA). Then, mRNA molecules leave the nucleus and go to the cytoplasm, where translation occurs. During translation, ribosomes link amino acids in the order specified by mRNA molecules to generate a polypeptide chain. This flow of genetic information is part of the famous central dogma of molecular biology described by Francis Crick in 1958 (Fh, 1958), which originally states that once the information has been trespassed into a protein, it can not get out again and be transferred into other biopolymer (such as DNA or RNA).

Proteins may be organized in four structural levels. A scheme of the four levels of protein structure is illustrated in **Figure 2**.

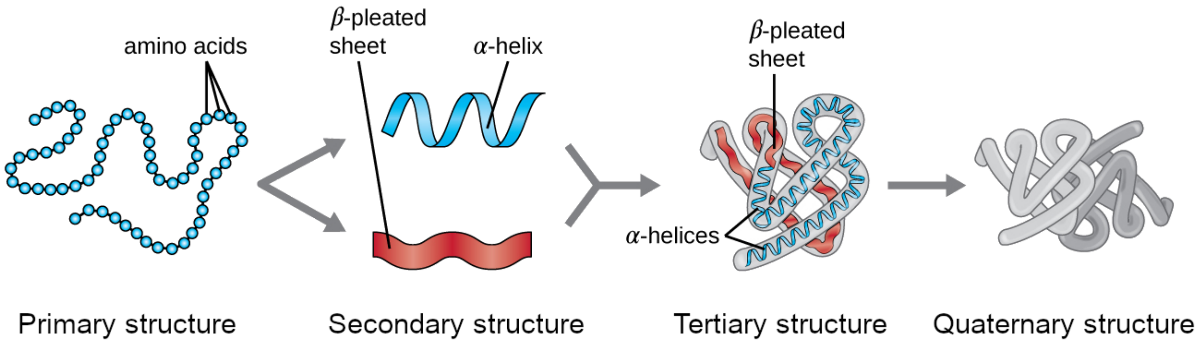


Figure 2. Four levels of protein structure conformation. OSC Microbio 07 04 (https://openstax.org/resources/85eae7ae8e3b242fa836f16b6abedfe5104e8ec6) from OpenStax Microbiology, licensed by CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/). Adapted version.

The primary structure is represented by the amino acid sequence of the protein and lacks a three-dimensional structure. The nascent polypeptide chain from the ribosomes is thermodynamically unstable, and quickly folds into semi-stable conformations driven by secondary structure conformations. Secondary structure units are classified into helices (α-helix), planar sheets (β-sheet), and stretched regions known as loops. Eventually, the protein adopts a completely folded conformation, also known as a tertiary structure, where multiple secondary structure units are organized into sophisticated protein domains. Monomeric proteins adopt the tertiary structure as a native conformation, however, oligomeric proteins can achieve another layer of structural complexity by adopting quaternary structures, where two or more polypeptide chains arrange to form a native conformation.

It has been assumed, under the hypothesis of “one protein coding gene = one protein”, that there are at least 20.000 unmodified canonical proteins in the human body (Ponomarenko *et al.*, 2016). However, the canonical sequence of proteins can be modified upon different processes including alternative splicing, single amino acid polymorphisms and posttranslational modifications, and certainly, the amount of different proteins in the body is expected to be much higher. Such protein tuning modifications are subjected to evolutionary pressure conditions and play a major role in natural selection and speciation processes. Remarkably, such important effects are achieved (in eukaryotic organisms) by the different combination of 21 proteinogenic amino acids: the 20 found in the standard genetic code, plus selenocysteine.

Amino acids differ one from each other by their physicochemical properties. A graphical scheme of the 21 proteinogenic amino acids is illustrated in **Figure 3**. Here, some features are included such as size, polarity and charge. Those features are mainly driven by the properties of the amino acids side chains. Hence, side chains may be classified into electrically charged, polar, hydrophobic and special cases. Charged side chains may be positive (Arginine -Arg, R-, Histidine -His, H- and Lysine -Lys, K-) or negative (Aspartic acid -Asp, D- and Glutamic acid -Glu, E-). Uncharged polar side chains may be small (Serine -Ser, S- and Threonine -Thr) or large (Asparagine -Asn, N- and Glutamine -Gln, Q-). Hydrophobic side chains may be small (Alanine -Ala, A-) or large (Valine -Val, V-, Isoleucine -Ile, I-, Leucine -Leu, L-, Methionine -Met, M-, Phenylalanine -Phe, F-, Tyrosine -Tyr, Y- and Tryptophan -Trp, W-). Special amino acids are the ones that behave differently from the other residues (Cysteine -Cys, C-, Selenocysteine -Sec, U-, Glycine -Gly- G-, and Proline -Pro, P-). Here, for instance, cysteine makes disulphide bonds, glycine induces protein flexibility and Proline induces tight turns in proteins. Histidine, however, shows different protonation states (as other residues, such as Glutamate, Aspartate and Lysines), and can be also classified in the uncharged polar group.

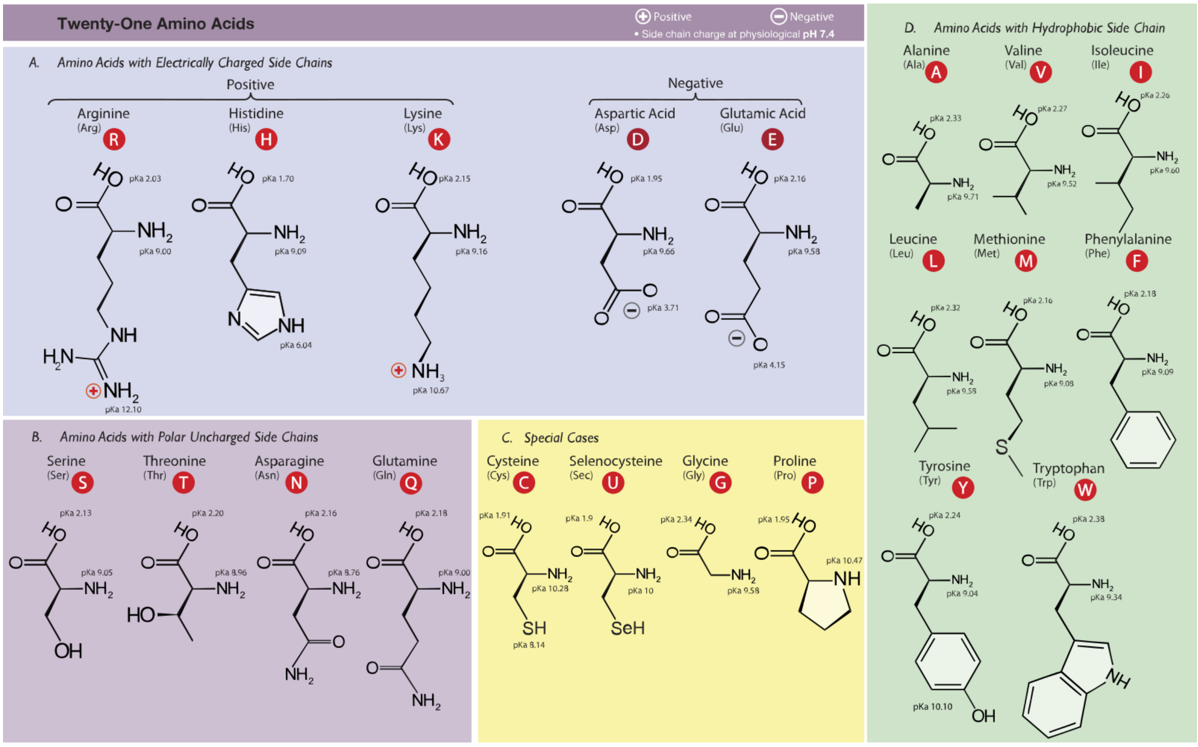


Figure 3. The 21 proteinogenic amino acids in eukaryotes are classified by the physicochemical properties of their side chains: positively charged, negatively charged, uncharged polar, hydrophobic and special cases. Amino Acids-wide. (https://commons.wikimedia.org/wiki/File:Amino_Acids-wide.svg) from Ben Hubert, licensed by CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0/deed.en).

As stated previously, proteins exert multiple biological functions. However, independently of the biological role, all proteins have the duty to serve as binders. In fact, the composition of the protein (this is, both amino acid sequence and three-dimensional conformation) determines the compounds that the protein can bind (known as ligands). Ligands are diverse, from small compounds such as ions or small drugs, to large molecules such as cofactors or proteins. The canonical protein-ligand complex is driven by transitory non covalent intermolecular interactions, which are mediated by hydrogen bonds, hydrophobic forces, electrostatic, π-π stacking and cation-π interactions. The main non covalent intermolecular interactions are depicted in **Figure 4**.

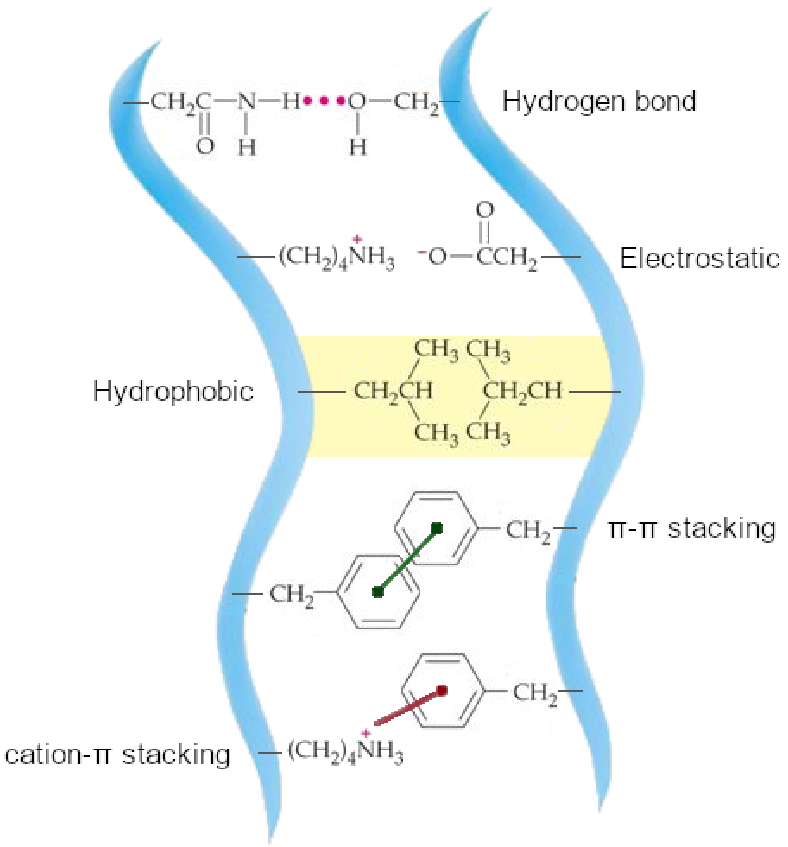


Figure 4. Non covalent intermolecular forces in proteins: hydrogen bond, hydrophobic forces, electrostatic, π-π stacking and cation-π interactions. Adapted from (https://slideplayer.com/slide/6613248/). Original credit to Pearson Prentice Hall, Inc (2010).

The interaction strength of a protein-ligand complex is described by the addition of the above mentioned intermolecular forces and it is defined as the binding affinity. However, independently of the binding strength, the binding process is always highly specific, in the sense that a protein is able to bind a range of tens of binders in a very selective way from the thousands that it encounters (with particular exceptions). This phenomenon is especially relevant for protein-protein interactions (PPI), where two or more proteins physically interact in a very specific manner by means of their binding sites. Protein-protein binding events can occur in at least three different ways, depending on the nature of the interacting partners: surface-string, helix-helix to form a coiled-coil and surface-surface. A scheme of those binding events is illustrated in **Figure 5**.

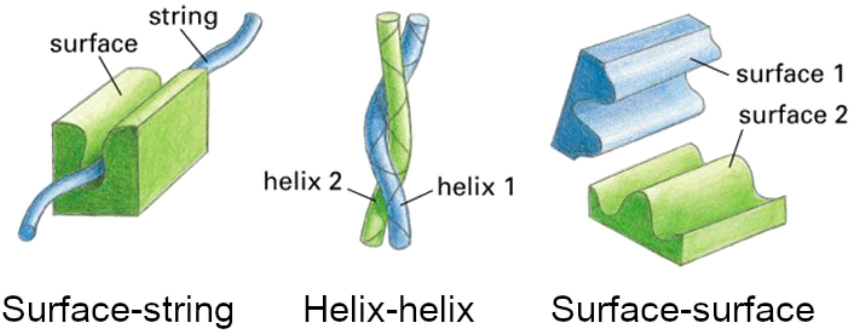


Figure 5. Three different protein-protein binding events: surface-string, helix-helix and surface-surface. Adapted from *Figure 3-41* of the Molecular Biology of the Cell book. 4th edition. Alberts B, Johnson A, Lewis J, *et al*.

The most common interaction found in protein-protein complexes consists of the two semi-rigid surfaces, where the shape of the interface perfectly matches for both proteins. This binding event results in a very tight interaction, favoured by a large number of weak contacts along the enormous binding surface. This *Lego*-like shape complementarity between the interacting partners usually leads to an extremely high specificity, which enables a protein to bind to just one (or few) partner/s from the enormous pool of encounters. Within a cellular environment, the combination of all PPIs is known as interactome, and its preservation (homeostasis) is crucial for the correct function of the host cell. In fact, interactome disorders are the basis of multiple aggregation-related diseases, such as Alzheimer's and Creutzfeldt–Jakob diseases. An unlimited source of protein variability of protein-protein networks may be found in the immune system and particularly in the context of antibodies. In the following section, we describe some details about the role of proteins in the immune system.

Role of proteins in the immune system

The immune system is a complex network of many biological structures and processes that serves as a protective mechanism against pathogens without compromising the host integrity. Two major subsets of the immune system can be differentiated in most of the species: innate immune system and adaptive immune system. An illustrative scheme of the main cellular components and the kinetics response of the innate and adaptive immune systems is depicted in **Figure 6**.

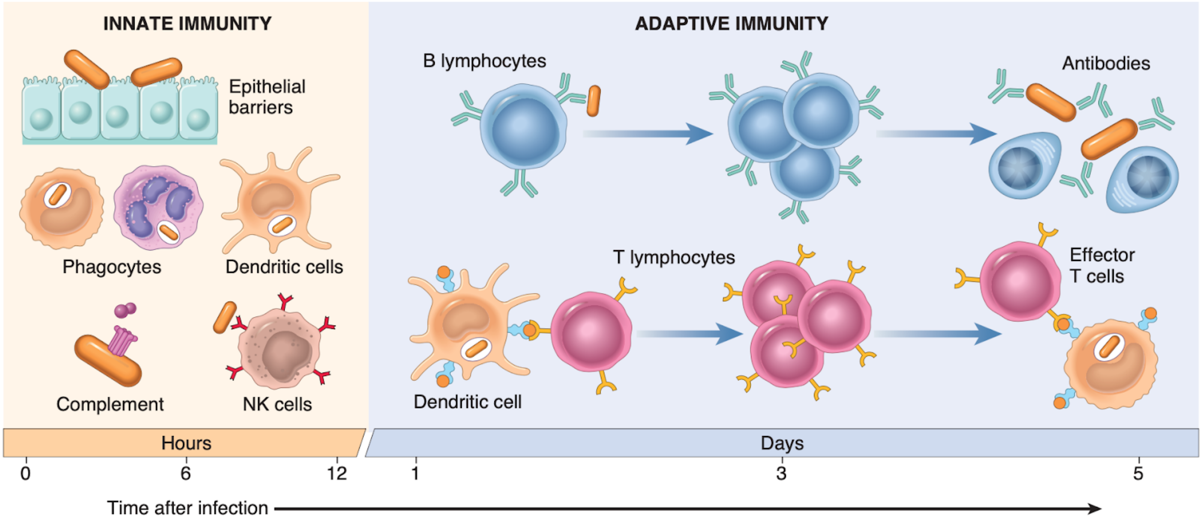


Figure 6. The principal components and kinetics of response of the innate and adaptive immune systems. Extracted from Figure 5.1 of the Robbins Basic pathology book. 10th edition. Kumar V, Abbas A, Aster J, et al.

The innate immune system is the first defensive line of our body against pathogens and it is activated by the chemical properties of the foreign body (antigen). The mechanism of action of the innate immunity is highly innespecific, in the sense that the effective response is similar among different antigens. Because of this reason, the innate immunity is able to exert a rapid response, involving the passive use of physical barriers (epithelial and mucous membranes) or the action of specialized immune cells including: phagocytic (mainly macrophages and neutrophils), dendritic, natural killer and other innate lymphoid cells. Moreover, the innate immunity also involves the release of proteins, including the complement system and cytokines to initiate the recruitment process of immune cells in the vicinity of the antigen source.

On the other hand, the adaptive immunity acts when the innate immunity is not able to completely neutralize the antigen source. Adaptive immunity generates a high-specific response that requires the processment and recognizement of the antigen source. Therefore, adaptive immunity is slower than the innate immunity. However, it includes immunological memory (which is the basis of vaccination), in the sense that the adaptive response towards previously known-antigens is faster and more efficient than during the first encounter. Compared to the innate immunity, the adaptive immunity shows a larger army of defensive mechanisms, including T lymphocytes (T cells), B lymphocytes (B cells) and their products, including antibodies. Moreover, there are two types of adaptive immunity: cellular (mediated by T cells) and humoral (mediated by B cells). An illustrative scheme of the main mechanism for both cellular and humoral immunities is depicted in **Figure 7.**

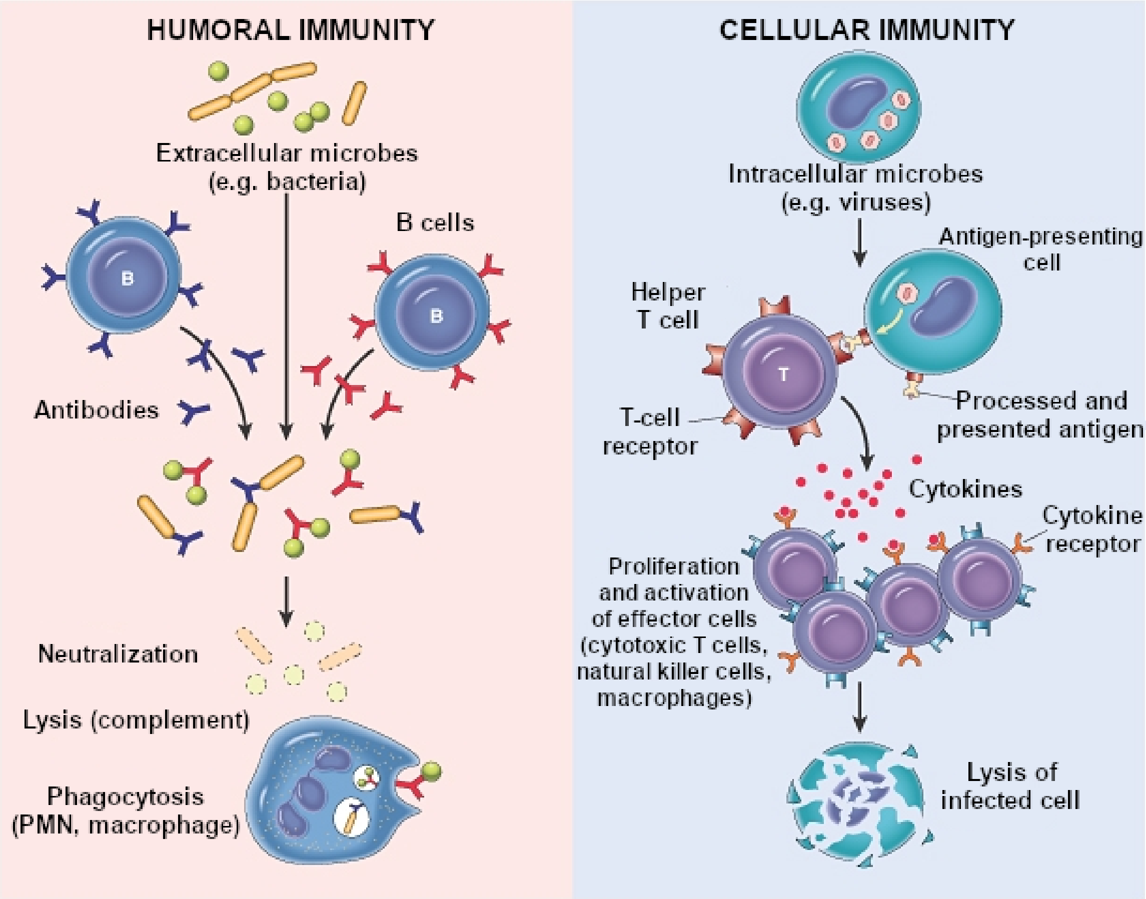


Figure 7. The principal components and kinetics of response of the innate and adaptive immune systems. Adapted from (https://www.pinterest.de/pin/686658274413051231). Original credit to Elsevier Science (USA), 2002.

There are multiple differences between both cell-mediated and humoral responses. However, the main difference relies on the mechanism of action. Hence, cellular responses target cells presenting antigens (including for instance, cancer cells, cells infected by viruses and/or transplanted tissues), while humoral responses secrete macromolecules to the extracellular fluid (mainly antibodies) to target pathogens. In this context, T cells (helper T cells, cytotoxic T-cells, natural killer cells, and macrophages) are the main drivers of cellular response, while humoral responses are mainly driven by B cells and antibodies (also assisted by CD4+ T helper cells).

Despite that all components of the humoral immunity play very important roles and most of them are irreplaceable, antibodies are highlighted to be one of the major contributors of the immune system. Hence, a simpler definition of humoral immunity can be understood as the generation of antibodies accompanied by accessory processes leading to the neutralization of pathogenic substances. There is a detail about the structure of antibodies, their production mechanism of antibodies and their importance in the following section.

Antibodies

As stated previously, antibodies are generated by B cells (or plasma cells). Antibodies, also known as immunoglobulines (Ig), are large Y-shaped proteins whose function is to bind antigens for their downstream removal. An scheme of the basic structure of an antibody is illustrated in **Figure 8**.

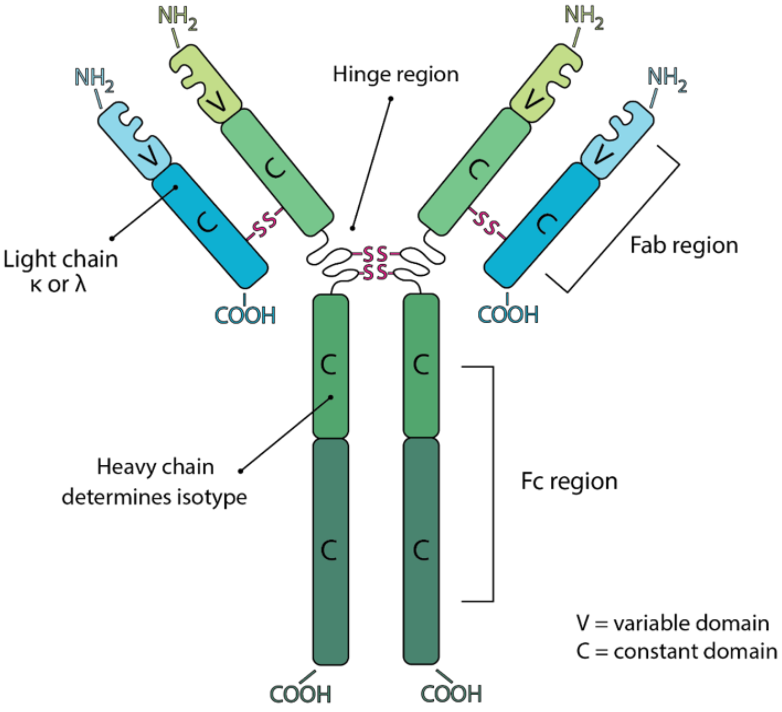


Figure 8. Basic structure of an antibody consisting of crystallizable fragment (Fc) and fragment-antigen binding (Fab) region, which is in charge of binding the antigen. Adapted from (https://bxcell.com/antibody-structure). Original credit to BioXCell.

Classically, antibodies are formed upon the arrangement of two polypeptide pairing chains known as heavy and light. Both heavy and light chains contain a crystallizable fragment (Fc) and a fragment-antigen binding (Fab) variable region. Heavy chains are composed by three to four constant Ig domains (depending on the antibody isotype), while light chains are composed by a constant and a variable Ig domain. Mammalians show five different antibody isotypes (IgA, IgD, IgE, IgG and IgM), which are determined by the type of the heavy chains (α, δ, ε, γ, and μ, respectively). A structural overview of the mammalian antibody isotypes is illustrated in **Figure 9**.

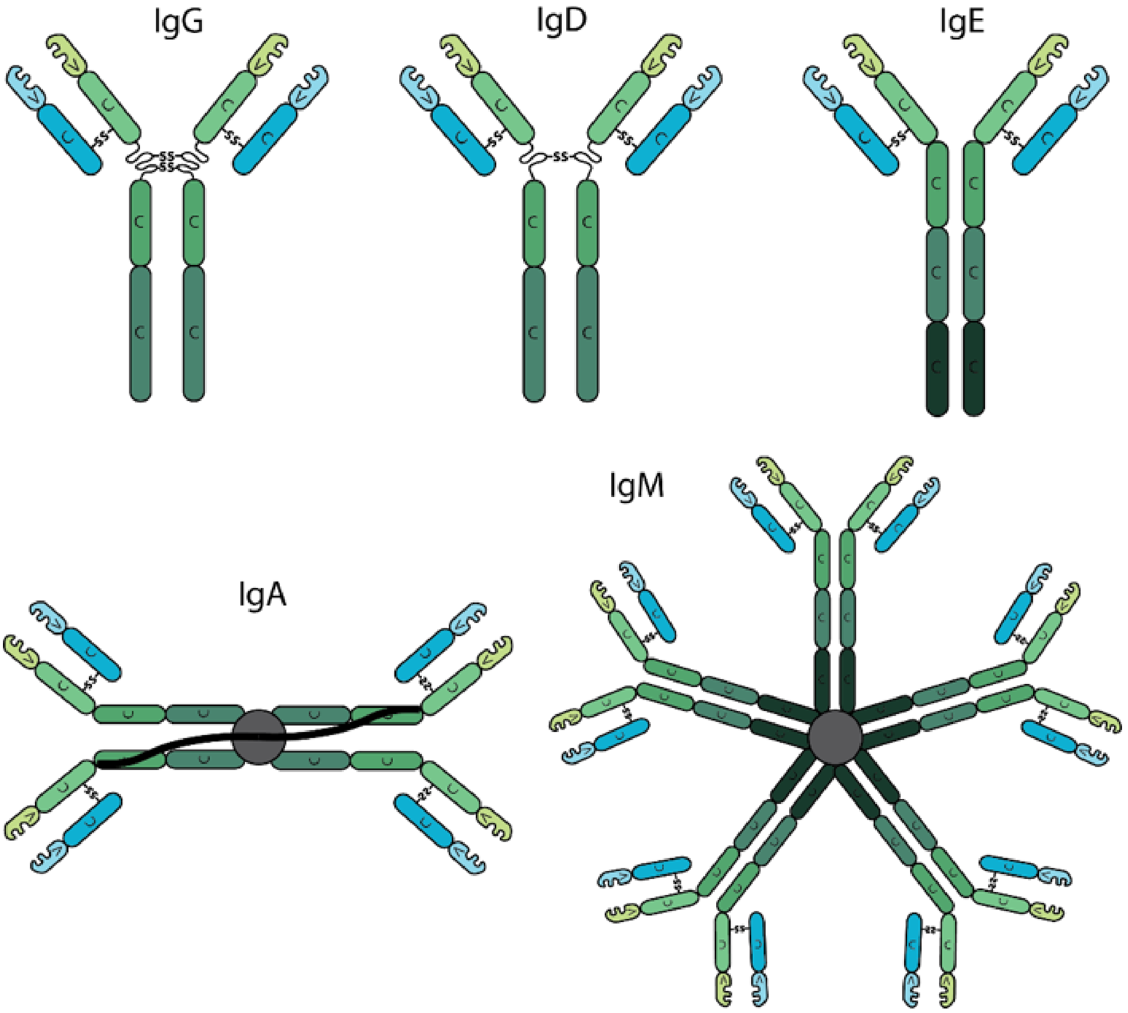


Figure 9. Basic structure of the mammalian antibody isotypes: IgG, IgD, IgE, IgA and IgM. The number of constant Ig domains and the overall three-dimensional structure differs among antibody isotypes. Adapted from (https://bxcell.com/antibody-structure). Original credit to BioXCell.

Antibody isotypes differ from each other on the heavy chain by:

- The number of constant domains.
- The amount and location of the interchain disulfide bonds.
- The number of attached oligosaccharide moieties.
- The length of the hinge region that adds flexibility to the antibody.

The constant domains of the Fc region are identical in all antibodies from the same isotype group. However, the variable domains of the Fab region differ from antibodies produced by different B cells clones. Regarding mammalian light chains, two types have been described (κ and λ). It is unknown if there exists functional differences between the κ and λ light chains, but both can occur in any of the above mentioned antibody isotypes.

It has been estimated that the human body may generate more than 10^{12} different antibodies, despite that the human genome contains roughly 40,000 genes (including coding and non coding). It is important to note that the mechanism for the generation of antibody diversity has evolved differently across species. In this way, birds, fish, sheeps, rabbits, camels, llamas, alpacas and humans show different antibody generation processes. For instance, camels, llamas and alpacas have evolved to generate camelid antibodies, which are smaller than regular antibodies. In this sense, those special antibodies can be made up by only two heavy chains (missing light chains), or just by the antigen binding domain of the heavy chain (V_HH, also known as nanobodies).

Despite the above mentioned differences, the immune system of mammals has developed unique genetic mechanisms enabling the generation of almost an unlimited set of different antibodies. This milestone is accomplished by sophisticated molecular mechanisms consisting of joining multiple gene segments together before their transcription. This process is known as V(D)J recombination. A representative scheme of the V(D)J recombination is depicted in **Figure 10**. Here, V, D, and J genes are placed in tandem within a variable gene locus section followed by a constant one. The initial stage of the V(D)J recombination consists of removing the unwanted D and J gene segments, followed by a DJ recombination of the D and J exons. The same process is repeated for removing the unwanted V and D exons during V-DJ recombination. Errors during the recombination process may further increase the variability of the final antibody transcript, which will also include the constant domain gene.

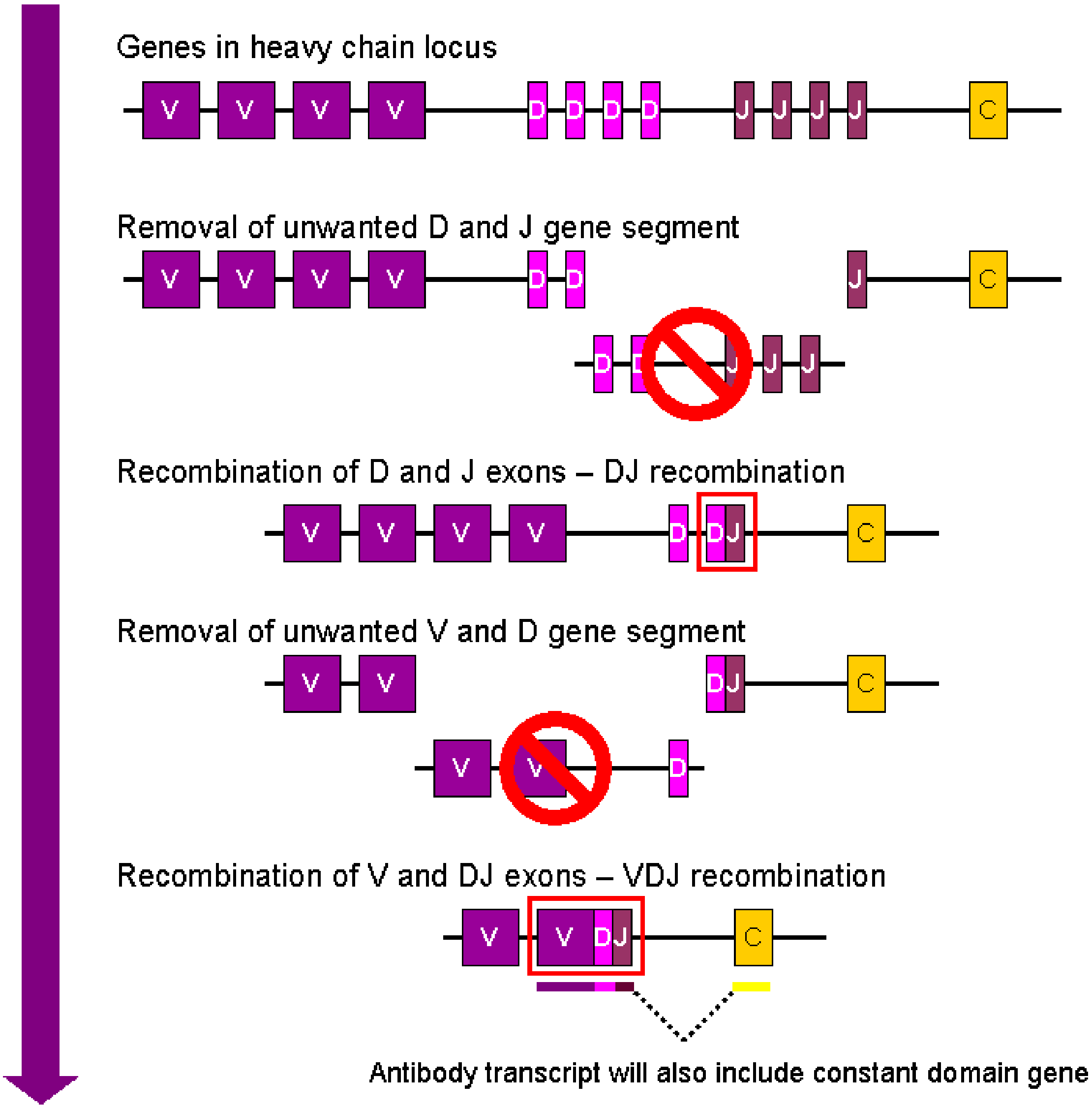


Figure 10. Representative scheme of the V(D)J recombination process during antibody diversity generation, involving removal of unwanted gene sections and recombination processes to generate a transcript coding for the full antibody sequence. Public domain image. Original credit to Gustavocarra.

Hence, billions of different antibodies are generated by recombining different gene segments, which are distant in the genome, into a new random pseudo-gene effector sequence. Moreover, the antibody diversity can be further increased by additional factors, which may include errors during the recombination process (decreasing or decreasing the amount of nucleotides at the site of gene-segment joining) or the acquisition of somatic mutations (which occurs at high frequency during antibody maturation processes).

Despite the enormous diversity of antibodies, the immune system minimizes the risk of generating antibodies binding to self-antigens. This achievement is reached through complex regulatory mechanisms, involving positive and negative selection of the B cells occurring at the bone marrow. The initial stage involves the positive selection of B cells expressing a functional BCR receptor complex, which eventually will bind to antigens and it will initiate the antibody response. During this process, B cells expressing a functional BCR (functional heavy and light chains, among other accessory proteins) are positively stimulated in order to ensure their proper development, while the others cease to exist. Later on, a negative selection process takes place, involving the active elimination of B cells that are able to recognize and bind to self-antigens. After this process, selected B cells migrate from the bone marrow to the spleen, where they keep differentiating into mature B cells. At this stage, B cells act as antigen-presenting cells in the sense that they endocitate antigens. Those antigens will be processed into large peptides, and presented into the B cell surface, through the major histocompatibility center II (MHC-II). At this moment, the CD4+ T cell repertoire (also called T helper cells) will screen the peptide-MHC-II complexes. Upon T cell binding the B cell is activated and it starts to further differentiate into more specialized cells, such as plasma B cells whose major role is to produce and secrete large amounts of antibodies. Those plasma B cells may undergo hypersomatic mutation and clonal selection processes, which will increase the specificity of the antibody towards the binding epitope.

Previously, we introduced a basic schematic structure of an antibody (**Figure 8**) and different isotypes (**Figure 9**). Three-dimensionally, antibodies form complex structures by combining multiple domain regions linked by loops. An overview of the three-dimensional structure of an IgG is illustrated in **Figure 11**.

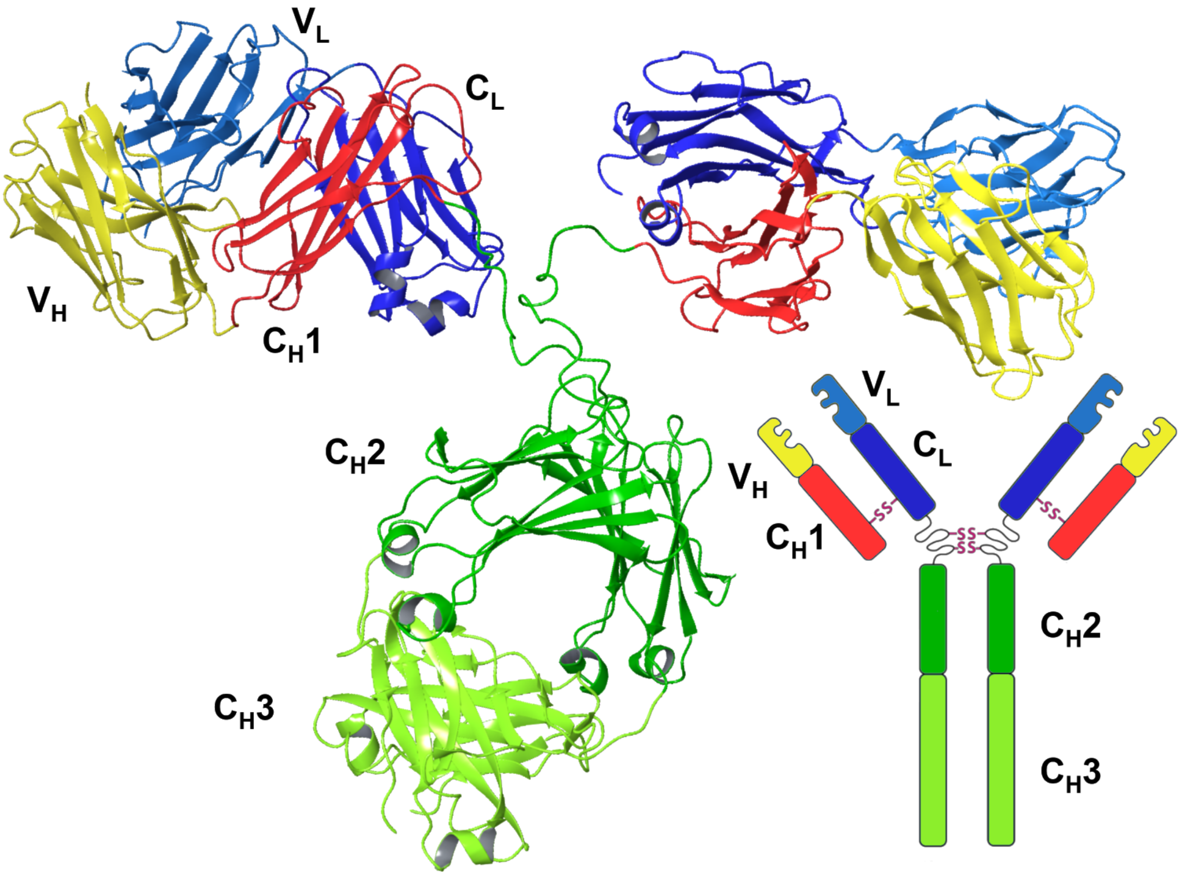


Figure 11. Three-dimensional structure of an IgG antibody consisting of three constant domains (C_H1, C_H2, C_H3) and a variable domain (V_H) for the heavy chain, and a constant and a variable domain for the light chain (V_L and C_L, respectively).

In IgG, heavy chains have four domains: three constant ones (C_H1, C_H2, C_H3) and a variable one (V_H). IgG light chains are formed by a constant (C_L) and a variable (V_L) domain. For both heavy and light chains, a short stretch known as a switch connects the variable and constant regions.

All antibody domains adopt a conserved structure termed as immunoglobulin fold, which consists of an antiparallel beta barrel motif. Such structure is formed by the interaction of two beta sheets packed tightly against each other and it is stabilized by different forces: i) hydrogen bonds between the beta strands of each sheet, ii) hydrophobic forces coming from opposite sheets in the interior of the fold, and ii) the formation of disulphide bonds between the sheets. Constant domains are typically formed by a 3-stranded sheet packed against a 4-stranded sheet. Variable domains show additional beta strands arranged in sheets of 4 and 5 strands.

Variable domains of the heavy and light chains are the ones that recognize and bind to the epitope. This process is mainly mediated by hypervariable loops, also known as complementarity-determining regions (CDRs). Each heavy and light variable domains show three different CDRs. A representative illustration of the antibody CDRs is shown in **Figure 12**. As can be observed, CDRs are the loops that connect the beta strands of the immunoglobulin fold. In total, the six CDRs are placed three-dimensionally close in space, cooperating for building an antigen-specific binding surface.

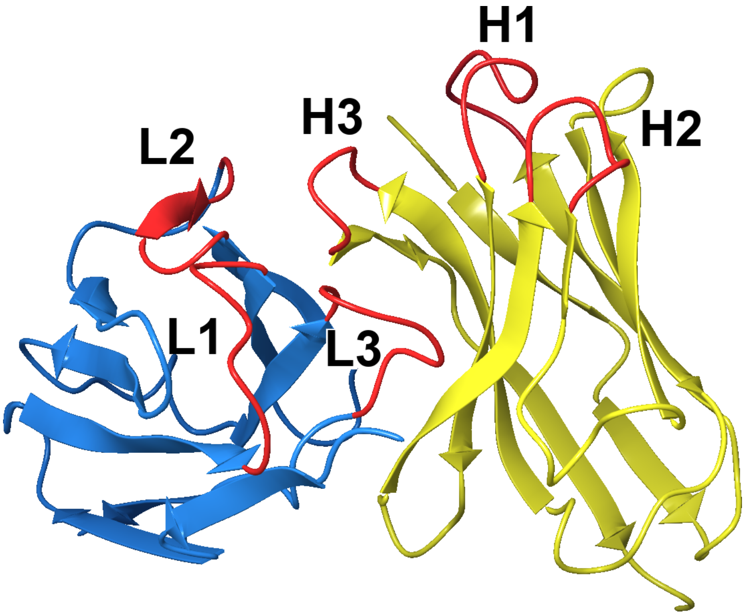


Figure 12. Heavy (yellow) and light (blue) antibody variable domains in a IgG antibody. Red loops represent the six hypervariable loops or CDRs for heavy (CDRH1, CDRH2 and CDRH3) and light (CDRL1, CDRL2 and CDRL3) variable domains.

In this way, all residues placed at the CDRs will play a major role for the recognition of the epitope. One of the main driving forces of binding antibodies comes from hydrophobic and van der Waals interactions. This is represented in **Figure 13**, where from the enormous antigen binding surface only a few polar contacts are formed (**Figure 13A**) compared to the entire interface (**Figure 13B**). In this example, the Protein Data Bank (PDB) entry 1JHL was used to represent this phenomena, containing a monoclonal antibody (heavy and light chains, in orange and green respectively) targeting an avian lysozyme (red).

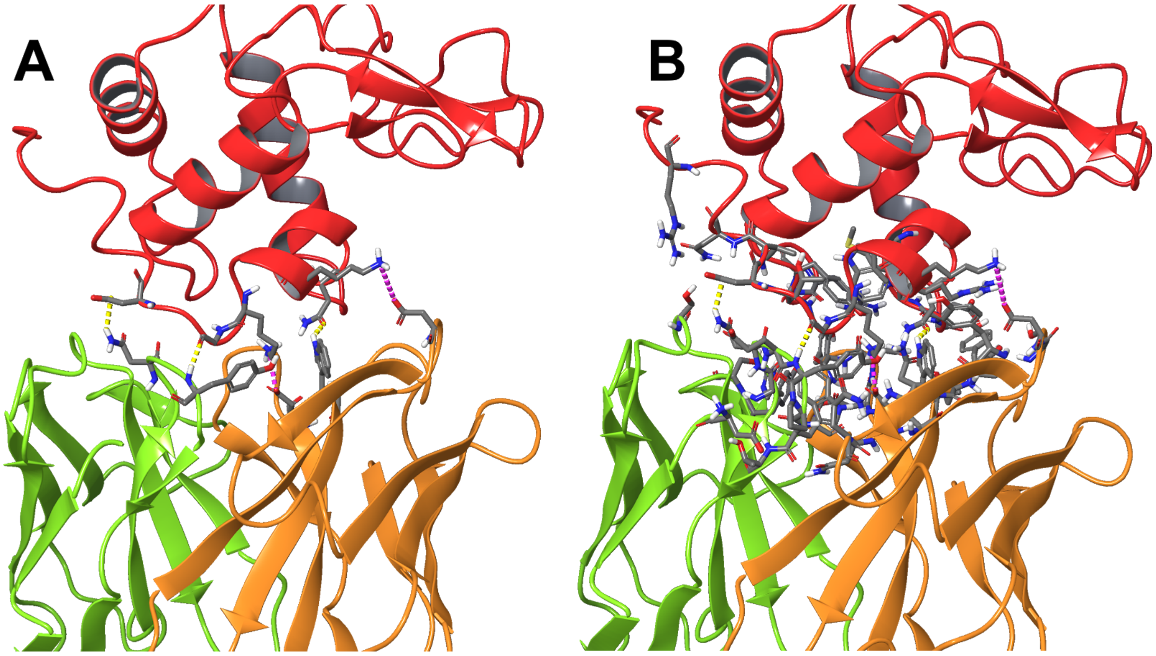


Figure 13. Most of the interacting binding surface of antibodies is determined by hydrophobic and van der Waal forces. Binding mode of a monoclonal antibody (heavy and light chains in orange and green, respectively) against an avian lysozyme (red), extracted from the 1JHL PDB entry. A) only a few polar intermolecular contacts are formed from B) the entire protein-protein binding interface. Hydrogen bonds and ionic bonds are represented by dashed yellow and pink lines.

In this thesis, we are particularly interested in the study of antibodies elicited against the human immunodeficiency virus 1 (HIV-1). The following section provides an overview of this virus and the antibodies that have been described to broadly neutralize its infectivity.

Human immunodeficiency virus

The human immunodeficiency virus (HIV) englobes two species, HIV-1 and HIV-2. HIV is a member of Lentivirus (from the retrovirus family) that infects humans and causes acquired immunodeficiency syndrome (AIDS). AIDS is a progressive disease that leads to the failure of the immune system. In 2019, the WHO estimated that nearly 38 million of people were dealing with AIDS, and about 700,000 died of AIDS-related complications. HIV-1 is more virulent and infective than HIV-1 (Gilbert *et al.*, 2003) and it is widely distributed around the globe compared to HIV-2, which is mainly restricted to West Africa (Reeves and Doms, 2002). Without treatment, the average survival time after HIV-1 infection is from 9 to 11 years. Advancements in the use of antiretroviral therapy largely increased the average survival time to nearly 40 years. However, there is no available cure for AIDS and most treatments aim to reduce the viral load (or viremia) in blood circulation, reducing the impact and the transmissibility of the virus. The HIV-1 genome takes advantage of a very sophisticated RNA splicing system to code nine genes from two copies of positive-sense single-stranded RNA of approximately 10kb. Those nine genes encode for an amount of 15 different proteins that have multiple roles. Depending on the function of the proteins, the genes encoding them can be classified in three different groups: viral structural elements (*gag*, *pol*, and *env*), essential regulatory elements (*tat* and *rev*) and accessory regulatory elements (*nef*, *vpr*, *vif* and *vpu*). A scheme of the RNA genome of HIV-1 is illustrated in **Figure 14**.

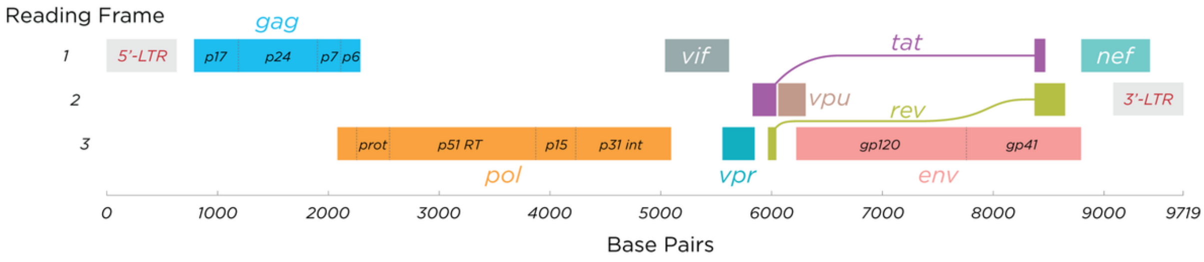


Figure 14. Structure of the RNA genome of HIV-1, encoding for nine different genes: gag, *pol*, *env*, *tat*, *rev*, *nef*, *vpr*, *vif* and *vpu*. The RNA genome has a length of approximately 10kb. *HIV-genome* from Thomas Splettstoesser (https://commons.wikimedia.org/wiki/File:HIV-genome.png), licensed under CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0/deed.en).

The genome of the virus is embedded within a viral capsid forming a virion. A representative scheme of the virion of HIV-1 is illustrated in **Figure 15**. HIV-1 virion is roughly spherical with a diameter of ~100nm (around 70 times smaller than a red blood cell). It contains all elements to ensure viral infection: the RNA genome embedded within a capsid and surrounded by protective proteins against nucleases (nucleocapsid), enzymes to ensure viral production (such as reverse transcriptase, integrase and protease) and surface proteins to initiate the viral entry into the host cell (envelope glycoprotein), among others.

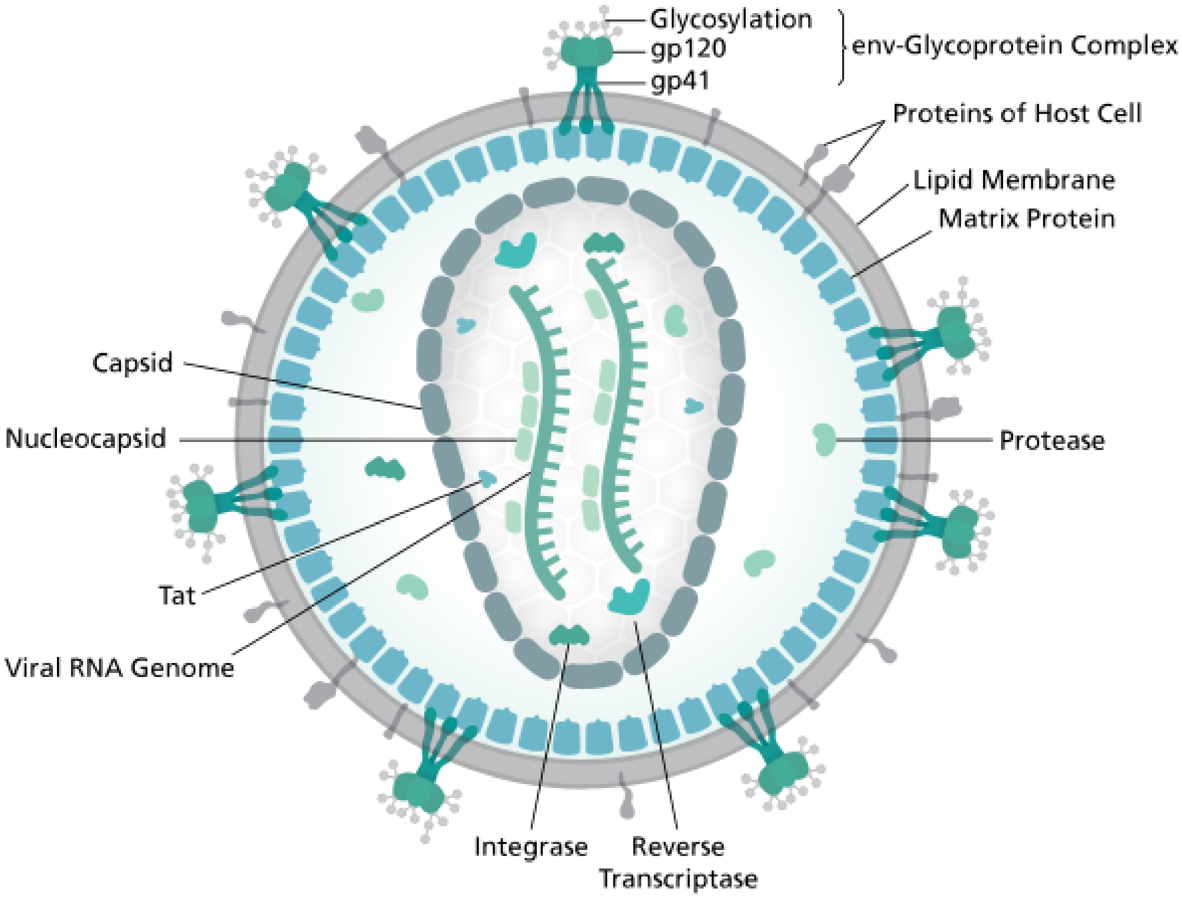


Figure 15. Structure of the HIV-1 virion containing all elements for viral transmission, including RNA genome, nucleocapsid, reverse transcriptase and the envelope glycoprotein, among others. *HI-virion-structure* from Thomas Splettstoesser (https://commons.wikimedia.org/wiki/File:HI-virion-structure_en.svg), licensed under CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0/deed.en).

The envelope glycoprotein is the major driver of the viral entry, and it is formed by two different proteins: gp120 and gp41. A representative scheme of the viral mechanism entry of HIV-1 is illustrated in **Figure 16**.

The target cells of the HIV are mainly CD4+ T cells and macrophages. The viral entry starts upon the binding between the gp120 of the virus and the CD4 receptor of the host cell. The region of the gp120 that binds to the CD4 receptor is known as CD4 binding site (CD4bs), and it is conserved among HIV-1. After this initial protein-protein interaction, the gp120 undergoes a conformational change and exposes its chemokine receptor binding domains, which will be able to initiate a secondary interaction with chemokine co-receptors. Numerous chemokine co-receptors have been described to bind to the chemokine receptor binding domains of the gp120, such as CXCR4 (mainly found in CD4+ T cells) and CCR5 (mainly found in macrophages) (Chan and Kim, 1998; Wyatt, 1998). At this stage, the N-terminal fusion peptide of the gp41 penetrates the cellular membrane of the host. After this process, the gp41 undergoes conformational changes that lead to the fusion of both viral and cellular membranes, allowing the entry of the viral capsid within the host cell.

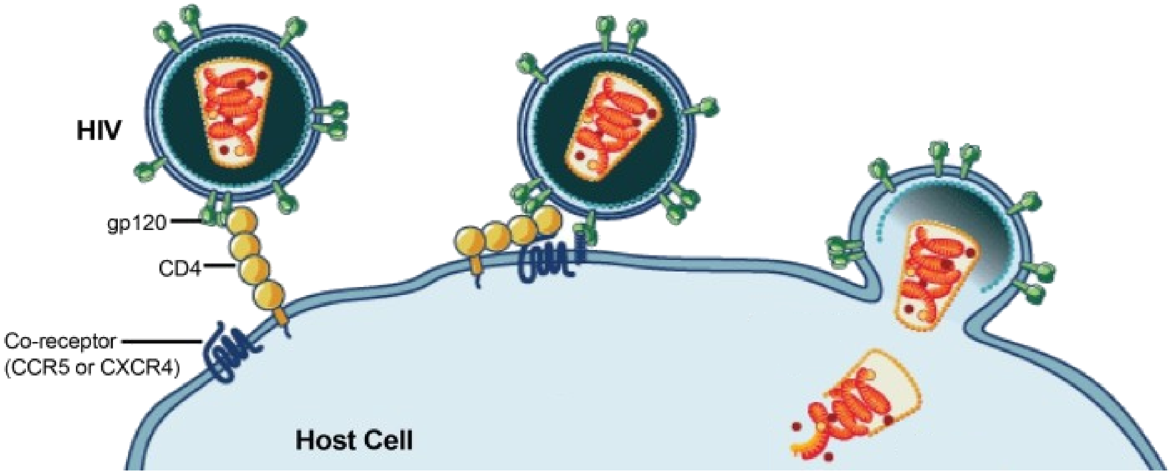


Figure 16. Mechanism of viral entry of HIV-1 into the CD4+ T cells and macrophages, mediated through CD4 and co-receptor proteins of the host cell (CXCR4 and CCR5, respectively). Adapted from *HIVs invasive strategy* (https://cen.acs.org/articles/92/i35/Aiming-HIVs-Weak-Spot.html) from Science Source/C&EN.

The envelope glycoprotein is the sole protein on the surface of the virus, being the major target of antibodies against HIV-1. However, the envelope glycoprotein is surrounded by a glycan shield that protects the virus from the immune system recognizement. In this context, the envelope glycoprotein of HIV-1 is one of the most glycosylated proteins described in the literature, where almost half of its mass is represented by N-glycan covalent modifications. A comparison of the glycosylation density between HIV-1 and other surface proteins, including the Spike proteins from SARS-CoV-1, SARS-CoV-2, MERS-CoV and the GP complex of LASV is illustrated in **Figure 17**.

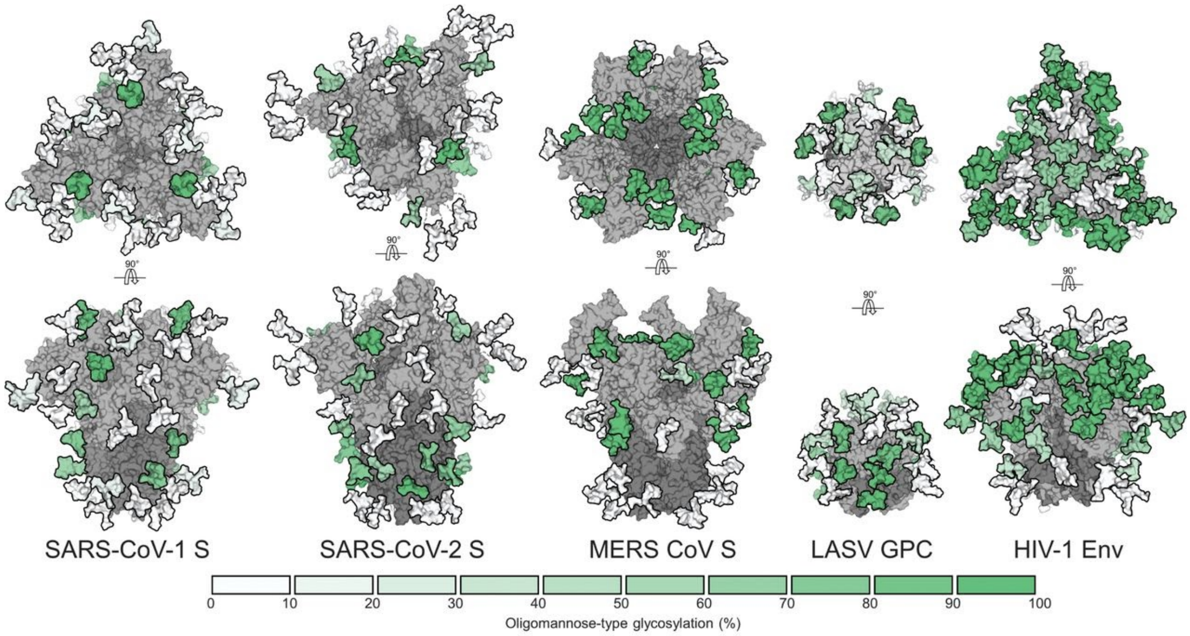


Figure 17. Different glycosylation levels of the HIV-1 envelope glycoprotein compared to other surface glycoproteins including spike protein from SARS-CoV-1, SARS-CoV-2 and MERS-CoV and the GP complex of LASV. Green color represents highly glycosylated epitopes. Figure 4 of "Site-specific glycan analysis of the SARS-CoV-2 spike", Science (2020) Vol. 369, Issue 6501, pp. 330-333, DOI: 10.1126/science.abb9983 (https://science.sciencemag.org/content/369/6501/330).

Overall, there are five phenomenons that impossibilities the immune system to completely deplete the HIV-1 from an infected patient:

- HIV targets immune cells, and therefore, decreases the potential of the immune system to combat the virus.
- The existence of latent infected CD4 cells that can not be effectively targeted (known as HIV-1 reservoir).
- The glycan shield, protecting the envelope glycoprotein from the immune system.
- The short replication cycle of the virus.
- The high mutation rate of the virus.

Nonetheless, the immune system is able to elicit antibodies against HIV-1, which can be strain-specific or broadly neutralizing antibodies (bNAbs). The main difference between both groups is that bNAbs are able to neutralize multiple viral isolates, while strain-specific antibodies are restricted to a few and similar viral isolates. The discovery of bNAbs in patient's samples has motivated their use as therapeutic tools against AIDS.

Over the past decades, large screening efforts have been made for isolating, describing and characterizing potent bNAbs. Some bNAbs have been reported to neutralize up to 90% of the viral isolates with great potency. The high efficacy of bNAbs for binding multiple viral isolates is achieved by targeting conserved regions of the virus, while also tolerating and accommodating mutations on the binding interface.

In this way, the efficacy of a bNAb is represented by a tradeoff between coverage and potency. On the one hand, coverage refers to the different strains that an antibody can bind. On the other hand, potency refers to the ability to strongly bind and neutralize the virus. A comparison of the coverage and potency of some of the most well-characterized bNAbs is illustrated in **Figure 18**.

So far, six different epitopes of the envelope glycoprotein have been described to elicit potent bNAbs: the CD4bs, the V1/V2 apex (or V2-apex), the V3 high-mannose loop (or V3-glycan), the membrane proximal external region of gp41 (MPER), the gp120-gp41 interface (Interface/FP) and the highly glycosylated 'silent' face of gp120 (silent face). However, the most well-characterized epitopes are the CD4bs, the V1/V2 apex, the V3 high-mannose loop and the MPER region. The relative position of the above mentioned four epitopes on the envelope glycoprotein and the occupancy of some bNAbs targeting them is illustrated in **Figure 19**.

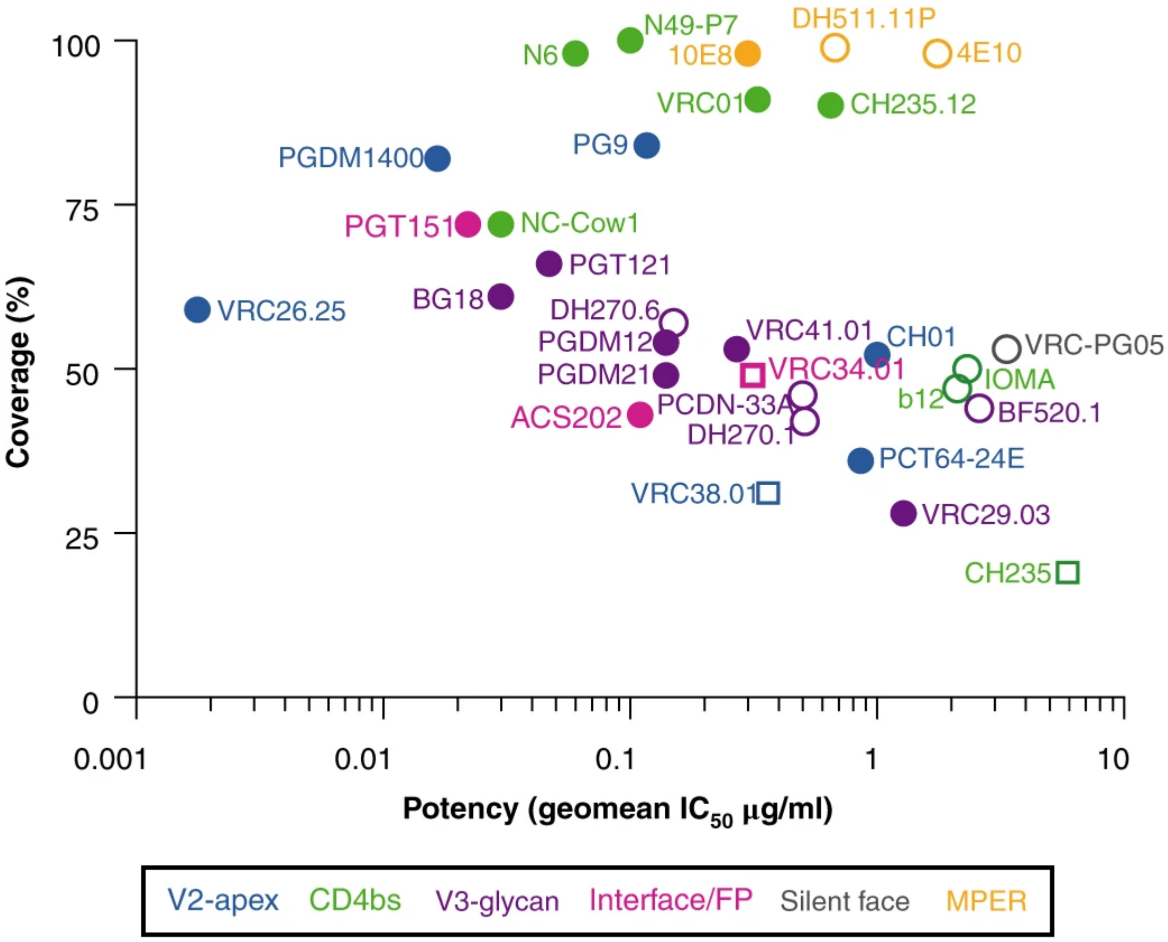


Figure 18. Strain coverage (%) and binding potency (µg/mL) of some of the most well-characterized bNAbs targeting different epitopes of the envelope glycoprotein. Adapted from Figure 1 of “Recent progress in broadly neutralizing antibodies to HIV”, Nature Immunology (2018) Vol. 19, p.p 1179–1188, DOI: 10.1038/s41590-018-0235-7 (https://www.nature.com/articles/s41590-018-0235-7).

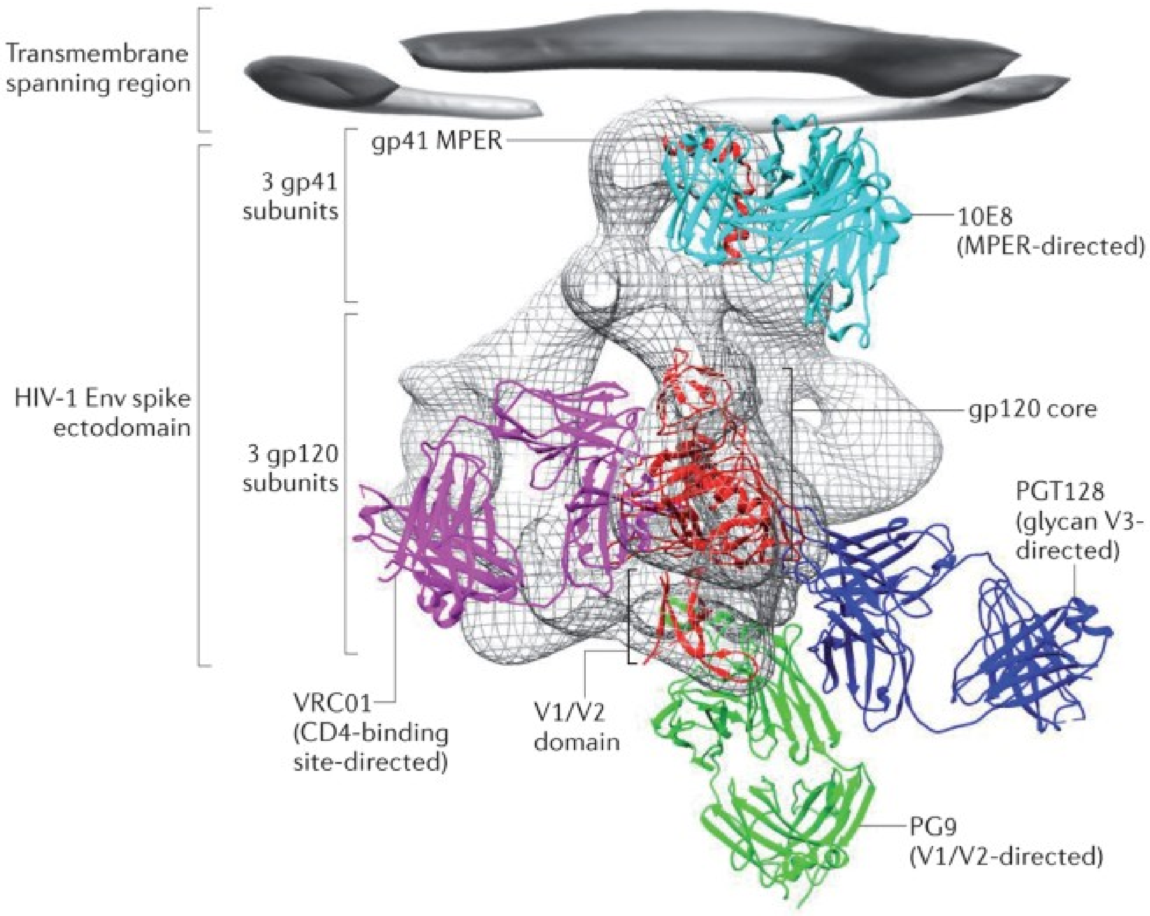


Figure 19. The four main epitopes of the HIV-1 envelope glycoprotein (the CD4bs, the V1/V2 apex, the V3 high-mannose loop and the MPER region) and the occupancy of some bNAbs targeting them (VRC01, PG9, PGT128 and 10E8, respectively). Adapted from Figure 1 of “Sites of HIV-1 vulnerability to neutralizing antibodies”, Nat Rev Immunol (2014) Vol. 13, p.p 693–701, DOI: 10.1038/nri3516 (https://www.nature.com/articles/nri3516).

From all bNAbs, the ones targeting the CD4bs and the MPER reach higher neutralization efficacy since they show a high strain coverage (usually up to 90%).

However, most of the research in HIV has been based on antibodies targeting the CD4bs (anti-CD4bs), because they can achieve higher potency than others while still preserving a great HIV-1 strain coverage. This phenomena is achieved by the blockage of the CD4bs, which initiates the viral entry by establishing the initial contact with the host cell. Hence, anti-CD4bs antibodies occupy the same three-dimensional epitope as the CD4 receptor. Despite the fact that the virus has evolved to escape the antibody recognizement (mainly by means of acquiring mutations and changing its glycosylation pattern), the biological role of the CD4bs must be preserved for the sake of the infectious process. Hence, the CD4bs cannot freely be modified to escape from bNAbs, since there exists a biological pressure towards preserving the binding with the CD4 receptor. Hence, mutations that hamper the binding with the CD4 receptor result in non-infective viruses which are (in principle) innocuous.

Despite that anti-CD4bs target the same epitope of the gp120, some differences arise from their mode of recognition and their heavy chain characteristics (Zhou *et al.*, 2019). In this way, those antibodies can be classified into two groups: VRC01-like class (VRC01, NIH45-46, 3BNC117, N6) and non-VRC01 classes (CH103, 8ANC131, VRC13 and VRC16) (Zhou *et al.*, 2015, 2013). From both groups, bNAbs from the VRC01-like class have drawn substantial attention to researchers since they mimic the binding mode of the CD4 receptor. In fact, VRC01-like class antibodies share some features with the CD3 receptor, including some known-key residues at certain positions that are crucial for a potent binding (Wu *et al.*, 2011; Zhou *et al.*, 2010, 2013). Some of those antibodies have been/are being evaluated in clinical trials, and relative therapeutic success has been observed. So far, none of them have been reported to cure AIDS since viral rebound occurs after days/weeks of stopping the treatment (such as during antiretroviral therapy). However, some computational efforts have been made aiming to better understand and characterize those antibodies. Such knowledge is intended to contribute for the generation of potent antibody variants against HIV-1. In this context, computational simulations may be key for the prediction of improved antibodies. The following sections will illustrate the current state-of-the-art of computational protein design.

Protein design

Recent advances in the availability of large computational resources have allowed the concept of computational protein design. This field takes advantage of experimental knowledge to develop computational algorithms that can be exploited to generate predictive models. Such predictive power can be used to engineer variants of proteins, including enzymes and protein-protein systems, for an improved application of interest. Enzyme engineering is probably one of the most representative applications of computational protein design. Many successful examples of enzyme design are described in the literature, since they have found their way into (mainly) industrial applications. During the last years, enzyme engineering has drawn the attention of industrial companies, especially for the *green chemistry* solutions that enzymes offer, in the sense that engineered enzymes can substitute chemical reactions that are environmentally harmful and costly. A recently described milestone of enzyme engineering consists of adding additional catalytic active centers into a protein (*Pluryzimes*^{TM}), aiming to increase the activity by having more than one active site, or to allow other different chemical reactions than the one found for the native enzyme (Santiago *et al.*, 2018; Alonso *et al.*, 2020). This demonstrates the usefulness of computational-aided protein design and the need of developing and improving algorithms to generate accurate predictions.

Regarding protein-protein design, a lesser amount of examples can be found in the literature compared to enzyme engineering, which is (probably) originated because of the two following main reasons. Firstly, protein-protein interactions occur by the formation of a large amount of weak and transitory interactions, which are challenging to model and estimate. Secondly, protein-protein systems are larger and thereby, more difficult to simulate than protein-small ligand studies. Hence, usually the bigger is the system the more expensive are the simulations. As a consequence, cheaper algorithms are necessary for protein-protein design projects, which in turn may decrease the performance of the prediction. Alchemical methods, such as the ones based on Free Energy Perturbations (FEP), have found a niche for ligand screening processes in drug discovery efforts. However, despite that they are described to be accurate, they demand large computational resources and more importantly their application in protein-protein systems is still limited (Clark *et al.*, 2017, 2019). The same phenomenon occurs for quantum simulations, which are still too expensive for performing large screening processes in protein-protein complexes (Bottaro and Lindorff-Larsen, 2018). However, with the rise of the supercomputational era during the following decades, techniques that are not widely-applicable at this moment may substitute the current state-of-the art for performing large screening processes aiming to design an improved protein-protein interaction. Currently, most of the work found in the literature is based on molecular mechanics (Siebenmorgen and Zacharias, 2020). The following section describes the concept of molecular mechanics and the basis of some approaches based on this concept, including the prediction of changes in Gibbs free energy upon mutation, Molecular Dynamics and Monte Carlo simulations.

Molecular mechanics

Molecular mechanics is a physics-based approach widely used to model atomistic processes. Molecular mechanics assumes that each atom is a solid sphere, which may be connected to other atoms by means of springs, simulating a covalent bond. Each atom has a radius and an electrostatic charge associated and it is placed in a three-dimensional space. The concept of force field is based on the above mentioned characteristics, which can be defined as a set of atoms in the three-dimensional space with associated potential energy values. Potential energy (*U*) can be mathematically modeled by means of a general equation consisting of multiple terms (**Equation 1**):

$$U = V_{BS} + V_{AB} + V_{PT} + V_{IT} + V_{LJ} + V_{EL} + V_{SO} \quad (1)$$

The first four terms (V_{BS}, V_{AB}, V_{PT} and V_{IT}) are known as bonded interactions, while the last three terms (V_{LJ}, V_{EL} and V_{SO}) are known as non-bonded interactions. Bonded interactions describe bond lengths, bond angles and dihedral angles of covalent bonds. On the other hand, non-bonded interactions describe non-covalent forces, such as van der Waals and electrostatic energies. Starting by the bonded interactions, V_{BS} represents the bond-stretching energy, which stands for the elastic interaction of covalent bonds between atoms. V_{AB} represents the angle-bending energy, standing for the interaction among three atoms connected by covalent bonds. V_{PT}, and V_{IT} represent the proper and improper torsional energies, standing for the interaction among four atoms connected by covalent bonds that form proper and improper dihedral angles, respectively. Regarding the non-bonded interactions, V_{LJ} represents the Lennard-Jones potential, which stands for the van der Waals energy (an attractive term) and the Pauli repulsion (a repulsive term). V_{EL} represents the potential energy according to Coulomb's Law. V_{SO} represents the solvation energy of the system, which may be considered by means of explicit or implicit waters. For explicit waters, the water molecules are three-dimensionally modeled and their energy contribution is taken into account. Some of the most widely used explicit water models are SPC (Berendsen *et al.*, 1981) and TIP3P (Jorgensen *et al.*, 1983). For implicit waters, the contribution of the waters is simplified by means of implicit solvent models, such as for instance OBC (Onufriev *et al.*, 2004) and VDGBNP (Zhu *et al.*, 2007).

The above mentioned descriptors aim to generally describe the essence of a force field. However, multiple force fields have been developed over the past years, such as OPLS (Jorgensen and Tirado-Rives, 1988), CHARMM (Vanommeslaeghe *et al.*, 2009), GROMOS (Reif *et al.*, 2012) and AMBER (Maier *et al.*, 2015). Despite that those force fields may follow different approaches, they usually take into account all described terms. However, they mainly differ on the parameterization process such as for instance the numerical terms on the partial charges or in the van der Waals parameters.

The following sections aim to provide some insights about the different molecular mechanics-based methodologies that have been used along this thesis. We will particularly focus on the prediction of changes in Gibbs free energy upon mutation, Molecular Dynamics simulations and Monte Carlo simulations.

Describing the ΔΔG upon mutation

Gibbs binding free energy changes upon mutation (ΔΔG) is determined by the intermolecular forces found between proteins. It is computed by the difference in Gibbs free energy (ΔG, also known as binding affinity) between the mutation and the native protein-protein complexes. There are many experimental methods for the determination of the ΔG including isothermal titration calorimetry, surface plasmon resonance, fluorescence, spectroscopy and stopped-flow assays (Geng *et al.*, 2016).

However, experimental ΔG determinations are expensive and time-consuming and therefore, there is an interest towards their computational prediction. In this way, the accurate estimation of the effects of mutations would contribute to decrease the amount of mutations to evaluate experimentally for a desired effect. Hence, computational techniques have the potential of guiding experimental mutagenesis studies, reducing time and costs.

There are many different ΔΔG predictors, which are mainly classified as physical energy descriptors, statistical potentials, shape complementarity, sequence conservation and more recently, machine learning-based techniques. A timeline of some representative ΔΔG binding predictors and databases of experimental ΔΔG's determinations is illustrated in **Figure 20**.

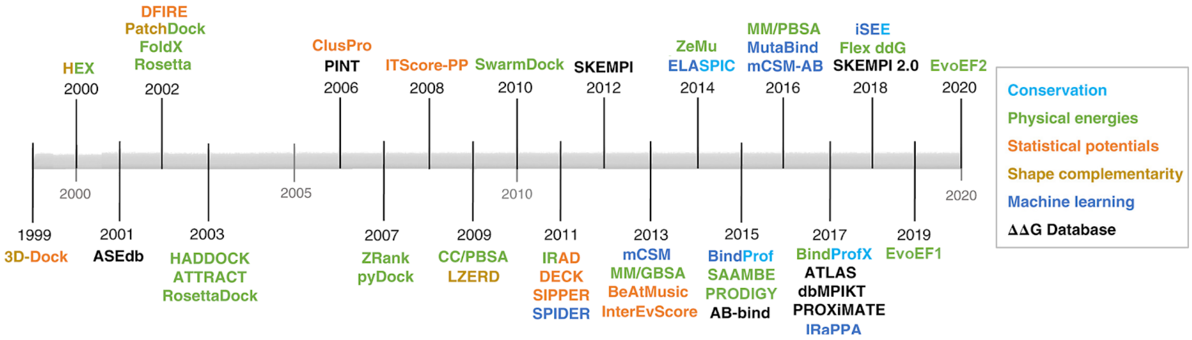


Figure 20. Timeline of some representative protein-protein binding ΔΔG predictors and databases. Adapted from Figure 2 of "Finding the ΔΔG spot: Are predictors of binding affinity changes upon mutations in protein–protein interactions ready for it?", *WIREs Comput Mol Sci.* (2019); 9:e1410, DOI: 10.1002/wcms.1410 (https://onlinelibrary.wiley.com/doi/full/10.1002/wcms.1410).

The following section will describe the basis of the ΔΔG predictors and the databases used in the "***Predicting ΔΔG: UEP***" result section.

ΔΔG Predictors

Seven different ΔΔG predictors have been benchmarked in this thesis including five physical energy descriptors (FoldX, EvoEF1, EvoEF2, pyDock, and PRODIGY), an statistical potential method (BeAtMuSiC) and a machine learning-based technique (mCSM). As described in the following lines, most of their force fields share some potential energy descriptors and partially overlap in some key aspects. However, they mainly differ on the amount of descriptors they use for the energy estimation. It is not intended here to describe the differences in the parameterization process of those techniques, but to provide a general overview of their scoring functions.

FoldX is one of the first published ΔG predictors for protein-protein complexes. The first version of the method was published in 2002 and it was available *via* a web-interface (Guerois *et al.*, 2002). FoldX has been updated through the years, including two updates of the method in 2005 (J. Schymkowitz *et al.*, 2005; J. W. H. Schymkowitz *et al.*, 2005). FoldX 5.0 is the latest release of the method (portable version), and it was published in 2019 (Delgado *et al.*, 2019). The core of the algorithm relies on the FOLDX force field (FOLDEF), which was designed for fast and accurate estimation of free energy changes upon mutations. FOLDEF is based on the linear combination of multiple energy terms weighted to experimental determinations. The FoldX energy (ΔG_{FOLDX}) accounts for six energy terms (**Equation 2**):

$$\Delta G_{FOLDX} = E_{VDW} + E_{ELECT} + E_{HB} + E_{SOLV} + E_{CLASH} + E_{ENTRO} \tag{2}$$

Here, E_{VDW}, E_{ELECT}, E_{HB}, E_{SOLV}, E_{CLASH} and E_{ENTRO} represent the total van der Waals, electrostatic, hydrogen bonds, solvation of polar and apolar groups, atomic clash penalties and entropic cost penalties contributions for a protein system, respectively. Differently to other ΔG predictors, FoldX offers a swiss-army of functions to manipulate three-dimensional structures, ranging for repairing them, reconstructing missing side chains, performing modelization of mutations, etc. Moreover, it is not limited to protein-protein structures, and it can generate base substitutions on three-dimensional structures of DNA and RNA.

EvoEF1 is a recent ΔG predictor published in 2019 (Pearce *et al.*, 2019). Its scoring function takes advantage of five different energy terms optimized on stability and binding affinity mutation data. The terms for the EvoEF1 energy (ΔG_{EvoEF1}) are (**Equation 3**):

$$\Delta G_{EvoEF1} = E_{VDW} + E_{ELECT} + E_{HB} + E_{DESOLV} - E_{REF} \tag{3}$$

Here, E_{VDW}, E_{ELECT}, E_{HB}, E_{DESOLV} and E_{REF} represent the total van der Waals, electrostatic, hydrogen bonding, desolvation energy of polar and apolar groups and reference energy terms for a protein system, respectively. The van der Waals energy is modified from the Lennard-Jones 12-6 potential, and the reference term is used to model the energy of the protein unfolded state. The authors described that EvoEF1 outperforms EvoEF2 for ΔΔG predictions. Similarly to FoldX, EvoEF1 also incorporates the needed functions for generating mutation files.

EvoEF2 was published in 2020 (Huang *et al.*, 2020), and it is recommended for *de novo* protein sequence design for a given protein scaffold. The EvoEF2 scoring function (ΔG_{EvoEF2}) uses nine different energy terms, preserving the five energy terms of EvoEF1 while taking into account four new terms to tackle difficult *de novo* protein designs (**Equation 4**):

$$\Delta G_{EvoEF2} = E_{EvoEF1} + E_{SS} + E_{AAPP} + E_{RAMA} + E_{ROT} \tag{4}$$

Here, the additional energy terms for EvoEF2 compared to EvoEF1 are E_{SS}, E_{AAPP}, E_{RAMA} and E_{ROT}, which describe the disulfide-bonding interactions, the energy for the calculation of amino acid propensities at a given backbone (φ/ψ) angles, the Ramachandran term given an amino acid for specific backbone (φ/ψ) angles and the energy term for modeling the side-chain rotamer, respectively. The authors described that EvoEF2 shows a decrease in performance for ΔΔG predictions, but an increase of performance for *de novo* design of proteins compared to EvoEF1. As EvoEF1, EvoEF2 incorporates the needed functions to generate mutation files.

PyDock is an algorithm specifically designed for protein-protein docking simulations and it was firstly published in 2007 (Cheng *et al.*, 2007). In 2013, a web server version of the method was released (Jiménez-García *et al.*, 2013). More recently in 2020, another web server version of the method was released, which provides details about the energetical contribution of each residue of the protein-protein interface (Romero-Durana *et al.*, 2020). Compared to other physical energy-based predictors, pyDock scoring function (ΔG_{pyDock}) only uses two terms (**Equation 5**):

$$\Delta G_{pyDock} = E_{ELEC} + E_{DESOLV} \tag{5}$$

Here, E_{ELECT}, and E_{DESOLV} represent the electrostatic and the desolvation energy contribution for a protein system, respectively. The authors described that the sum of electrostatics and desolvation energy accounts for more than 70% of the total energy contribution, showing enough predictive power for scoring rigid-body docking poses. In this sense, pyDock is not a widely used ΔΔG but its ranking as one of the best protein docking protocols in a recent blind prediction competition motivated its evaluation in this thesis (Lensink *et al.*, 2019). pyDock was not specifically designed to perform ΔΔG predictions and therefore, it is not intended to generate mutation files. Because of this reason, pyDock predictions must be performed on models generated by third-party applications (such as FoldX, EvoEF1 or EvoEF2).

PRODIGY was first published in 2015 (Vangone and Bonvin, 2015) and a web server version was released in 2016 (Xue *et al.*, 2016). According to the authors, the method predicts the ΔG based on a weighted network of interfacial contacts (ICs) corrected by a non-interface surface (NIS). The interfacial contacts are defined by the contacts within a 5.5 Å cut-off distance between the protein groups. A simplification of the PRODIGY scoring function ($\Delta G_{PRODIGY}$) is represented by (**Equation 6**):

$$\Delta G_{PRODIGY} = ICs_{char/char} + ICs_{char/apo} - ICs_{pol/pol} + ICs_{pol/apo} - NIS_{apo} - NIS_{char} \quad (6)$$

Here, $ICs_{char/char}$, $ICs_{char/apo}$, $ICs_{pol/pol}$, $ICs_{pol/apo}$, NIS_{apo} and NIS_{char} represent the ICs of charged/charged, the ICs of charged/apolar, the ICs of polar/polar, the ICs of polar/apolar, the NIS apolar and the NIS charged groups properties. The authors describe that their energy function, based on ICs, is less affected by sensitive conformational changes occurring upon binding. PRODIGY does not incorporate the needed tools to generate mutation files, and therefore, its predictions must be based on mutation files generated by third-party applications (such as FoldX, EvoEF1 or EvoEF2).

BeAtMuSiC was first released in 2013 (Dehouck *et al.*, 2013) as a web-server application. Differently to the other methods, BeAtMuSiC is specifically designed for the evaluation of ΔΔG and does not require the previous modeling of the mutation. The algorithm returns the predicted ΔΔG of a mutation compared to a native reference, and its scoring function ($\Delta\Delta G_{BeAtMuSiC}$) works as follows (**Equation 7**):

$$\Delta\Delta G_{BeAtMuSiC} = \alpha(\Delta\Delta W + \Delta\Delta V) \quad (7)$$

Here, α, ΔΔW and ΔΔV represent the accessibility of the solvent of the mutated residue, energy changes and volume changes induced by the mutation, respectively. The accessibility of the solvent is used as a sigmoid function to weight the other two parameters, and it is based on experimental binding affinity determinations upon mutation. Energy changes are described by 13 statistical potentials extracted from a dataset of known protein complexes. Those potentials include correlations between amino acids, pairwise inter-residue distances, torsion angles of the backbone, and solvent accessibilities. The term of changes in volume was introduced to account for possible packing effects originated by the mutation.

mCSM is one of the first machine learning-based techniques, and it was released in 2013 (Pires *et al.*, 2014) as a web server application. Several modifications of the original algorithm have been published along the past years, including algorithms specifically designed for antibody predictions (Pires and Ascher, 2016; Myung, Rodrigues, *et al.*, 2020; Myung, Pires, *et al.*, 2020), protein-ligand (Pires *et al.*, 2016), protein-nucleic acids (Pires and Ascher, 2017), and stability of membrane proteins upon mutation (Pires *et al.*, 2020). Differently to the other algorithms, mCSM scoring function is not formed by the linear combination of individual energy terms. It is based on graph-based signatures of the changes in contacts originated by the mutations weighted to the experimental binding determinations. The mCSM hierarchical procedure works as following: i) given a mutation with an experimental ΔΔG associated, the contacts of the wildtype and mutant residues with the partner protein are extracted, ii) for both wildtype and mutant, a pharmacophore based on the physicochemical properties of the contacts is generated, iii) the difference of the wildtype and mutant pharmacophores is computed, which will be the descriptors for the known experimental ΔΔG determinations, iv) the model is trained based on those descriptors and the associated experimental ΔΔG determinations, v) for generating predictions, the same hierarchical procedure is followed for the estimation of the ΔΔG under the trained model.

Databases related to ΔΔG predictors

In this section, we will introduce two databases that have been used in the context of this thesis: SKEMPI 2.0 and Interactome3D. They have been used in the "***Predicting ΔΔG: UEP***" result section.

SKEMPI 2.0 was released in 2018 (Jankauskaite *et al.*, 2019) and it is the updated version of the original database (Moal and Fernández-Recio, 2012). SKEMPI 2.0 is one of the largest databases manually curated of experimental data upon mutation for three-dimensionally solved protein-protein complexes. The database consists of 6187 unique mutations (from an amount of 7085) comprising 345 different PDB entries of protein-protein complexes. The authors raise several concerns about the quality of the data used for the construction of the database, which may be of great interest for computational design efforts. First of all, the authors cautiously state that the experimental determinations come from a broad range of studies that intrinsically add bias. This is because the data they have collected from the literature is biased towards the interests and the capabilities of the research community. For instance, almost three quarters of the database consists of single mutations, and more than a half of them are mutations to alanine. Moreover, mutations consisting of swapping charges (negative - positive) and mutations between aromatic residues are overrepresented. Furthermore, most of the mutation data belongs to the binding site, concretely at the core of the protein-protein interface. Taking everything together, there is no doubt that the data found in the literature shows some degree of bias, which will affect the development and benchmarking of algorithms using this source of data. An overview of the SKEMPI 2.0 database is illustrated in **Figure 21**.

Regarding the mutation data, SKEMPI 2.0 contains the information contained on the previous version of the database (SKEMPI) together with new data. Also, it contains data from other known databases, including dbMPIKT (Liu *et al.*, 2018), AB-BIND (Sirin *et al.*, 2016) and PROXiMATE (Jemimah *et al.*, 2017). Most of the ΔΔG values fall within a range of -3 to +7 kcal/mol. ΔG values are the most common binding measurement, and the most popular methods for determining binding affinity are surface plasmon resonance (SPR) and fluorescence (FL). Belonging to the protein-protein complex category, protease inhibitor and antibody-antigen systems are overrepresented compared to other systems.

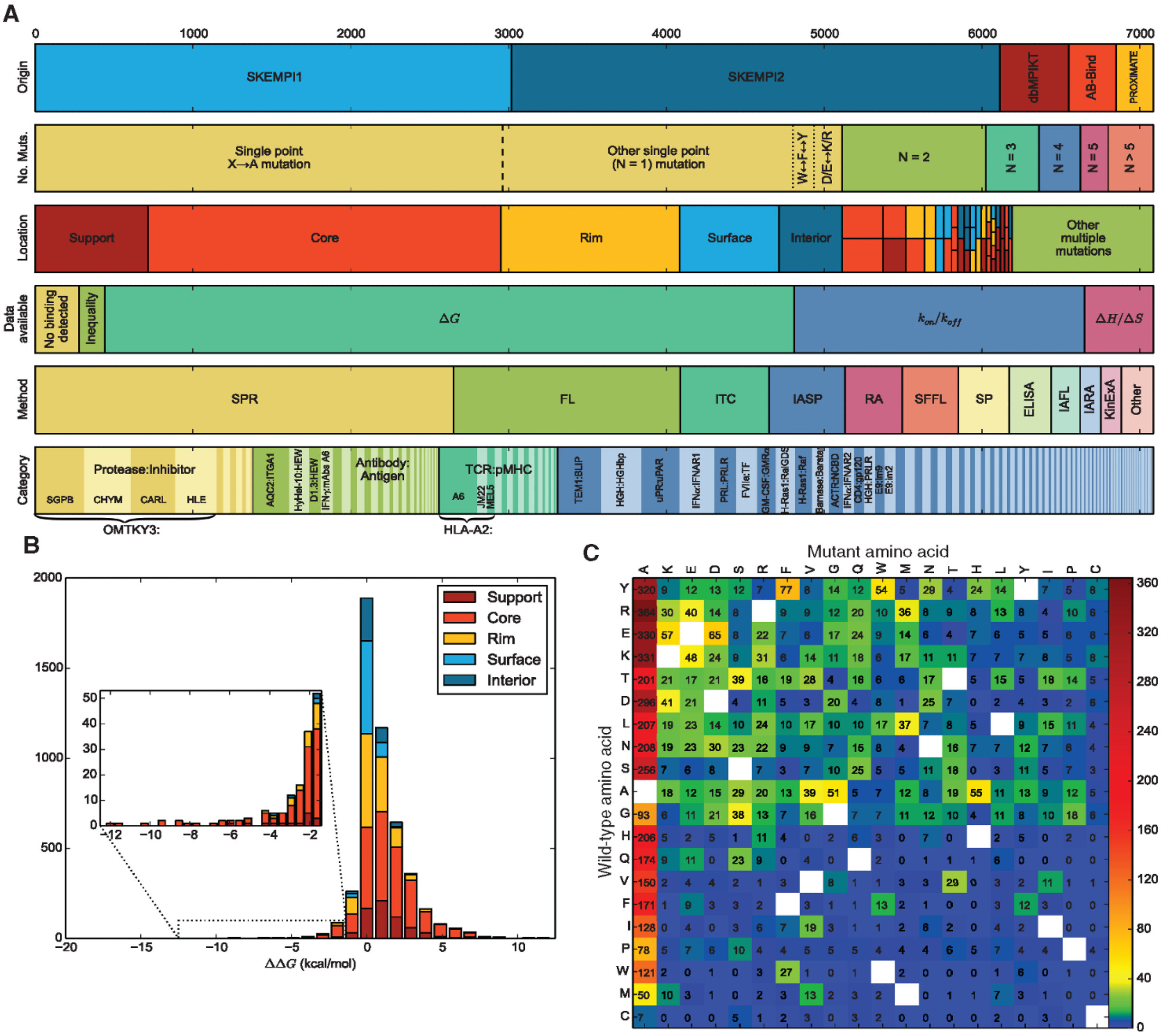


Figure 21. An overview of SKEMPI 2.0. (A) Mutations partitioned according to their origin, the number of altered residues, location within the complex, by the availability of additional kinetic and thermodynamic data, according to the experimental method used, and by category. (B) Distribution of ΔΔG. (C) Source and target amino acids for single point mutations. Figure 1 of "SKEMPI 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation", (https://academic.oup.com/bioinformatics/article/35/3/462/5055583), Bioinformatics (2018) Vol. 35, p.p 462–469, DOI: 10.1093/bioinformatics/bty635. Licensed by CC BY 4.0 (http://creativecommons.org/licenses/by/4.0/).

Interactome3D is a web service that contains structural annotation of protein-protein interaction networks. Despite that the database was first published in 2013 (Mosca *et al.*, 2013), it is being updated yearly. In this sense, the modeling pipeline of Interactome3D is fully automatized and handles two types of input data: a set of interactions defined by the user, or a list of organisms for modeling their full or partial interactomes. The authors describe the algorithm as a dynamic resource that collects the necessary structural data for single proteins and binary interactions, followed by an automatized protocol for selecting the best scaffolds for modeling, returning the three-dimensional coordinates of binary complexes. An overview of the automatized Interactome3D pipeline is illustrated in **Figure 22**. The first pipeline step consists of collecting the structures for each of the individual proteins in the interactome network. Here, the first stage consists of identifying the available experimental structures on the PDB, and to increase the structural coverage by using homology models using ModBase (Pieper *et al.*, 2014). Then, individual proteins are classified into three categories depending on the sequence coverage with three-dimensional data: complete experimental structures (covering >80% of the length of the protein with fully sequence identity), complete homology models (covering <80% of the length of the protein with fully sequence identity) and partial structures and models (the rest; here protein fragments are grouped to increase the protein length coverage). Next, the algorithm proceeds to identify interaction structures coming from experimental determinations or suitable templates for modeling them. For the modeling part, only are considered scaffolds having more than >30% sequence identity with the protein pairs to be modeled, by using protein-protein interactions found on the PDB or in 3did (Stein *et al.*, 2011). The models are built with Modeller (Sali and Blundell, 1993), and the three-dimensional coordinates are examined under different conditions. Finally, the algorithm ranks all structures and models assessing completeness and quality, obtaining a representative set of three-dimensional models for each interactome consisting of the top-ranked models for every protein-protein interaction.

In this thesis, we have used the representative Interactome3D database from 2019-01. This release contains 33,607 unique three-dimensional protein-protein complexes from a set of 18 species from different Kingdoms: Plant (1), Protist (1), Fungi (2), Animal (6), and Eubacteria (8). The three-dimensional complexes of

Interactome3D are classified in three groups, depending on their source: experimental structures (44.2%), homology models from generic PDB templates (35.0%) and domain-domain structural templates (20.8%).

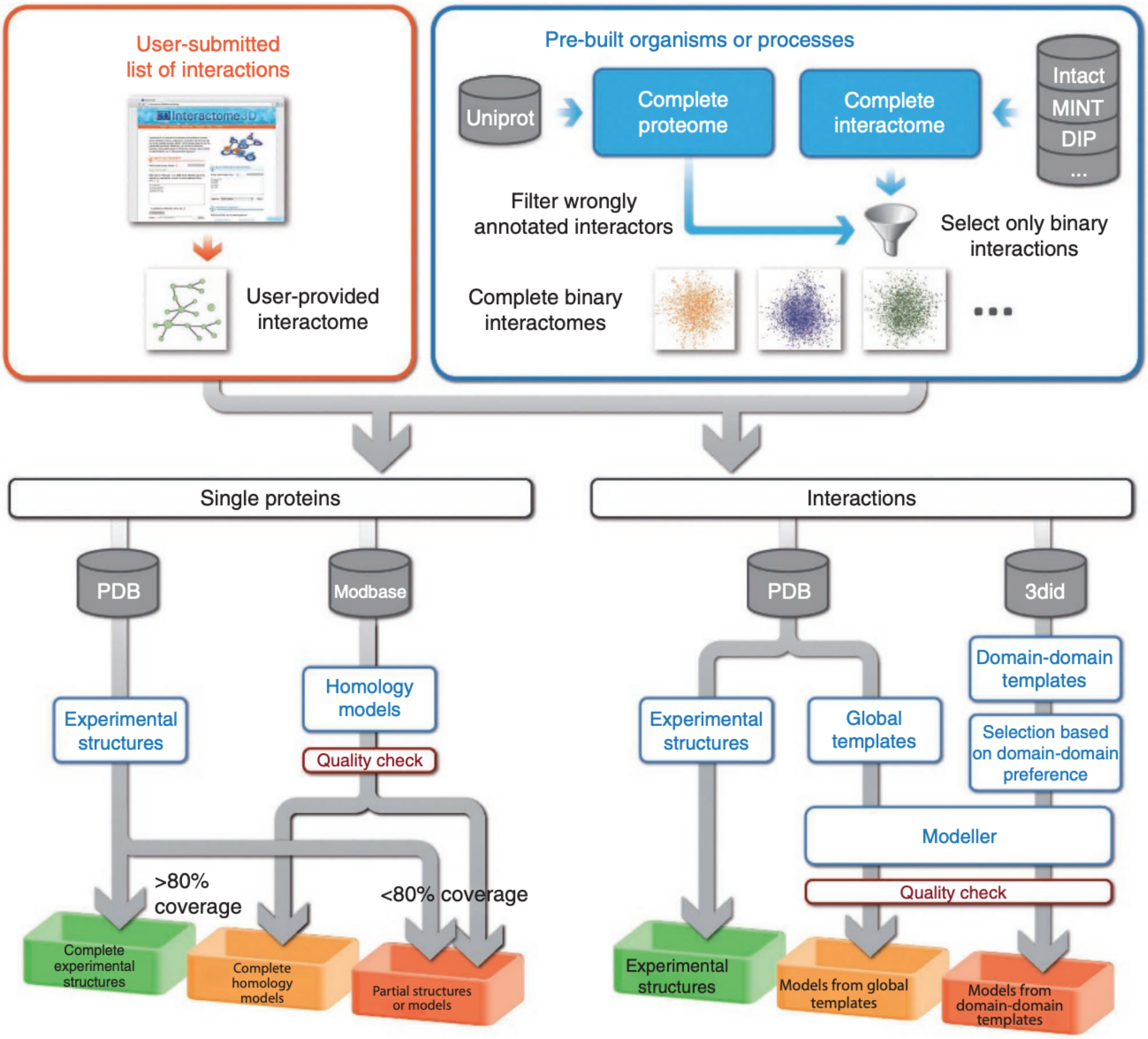


Figure 22. Overview of the Interactome3D pipeline. The fully automatic Interactome3D platform recollects experimentally resolved (PDB) and homology templates (ModBase) for single interacting proteins, and maps them into a binary interaction scaffold. Homology models for the protein-protein interaction are built using Modeller from PDB or 3did scaffolds. Figure 1 of “Interactome3D: adding structural details to protein networks”, Nature methods (2013), Vol.10, p.p 47–53, DOI: 10.1038/nmeth.2289 (https://www.nature.com/articles/nmeth.2289).

Molecular Dynamics

Molecular Dynamics (MD) is a computational method for predicting the time evolution of a system of interacting particles. In order to do so, MD integrates Newton's equations of motion under an energy potential described by a force field. Unless otherwise specified, MD employs a molecular mechanics force field. MD simulations can be used for investigating dynamic, thermodynamic and structural properties of the system (Braun *et al.*, 2019). A general workflow of an MD simulation consists of the following four steps:

- System preparation.
- Minimization/relaxation.
- Equilibration.
- Production.

The system preparation focuses on generating a valid three-dimensional structure fulfilling the requirements for an MD simulation, including modeling the system, solvation (if necessary) and applying a force field of choice. This is one of the most critical steps, and in many cases receives the least attention. The worst case scenario is, probably, when the system is incorrect (for instance, errors in the protonation state) but maintaining a correct chemical composition described by the force field, which propagates the error (without being noticed) in the further steps. This example does not only play an important role in MD, but also in all other molecular modeling computational techniques. After the system is properly built, minimization and relaxation steps are often used to stabilize the complex into an energy minima, aiming to guarantee the success of the further steps. Once the structure is at an energy minimum, the equilibration stage begins. The equilibration process consists of slowly adapting the dynamics of the system to the experimental conditions of the simulation, such as for instance temperature and pressure. There is no universal protocol for performing this equilibration process; its choice depends on the statistical production ensemble, being the most common ensemble variants: NPT (constant pressure and temperature), NVT (constant volume and temperature) and NVE (constant volume and energy). When the equilibration stage reaches the experimental designed conditions and the system remains stable, the production phase begins. The production phase consists of applying the equations of motions and saving the new coordinates periodically until some criteria is met, such as the total simulation time. A suggested workflow for MD simulations is illustrated in **Figure 23**.

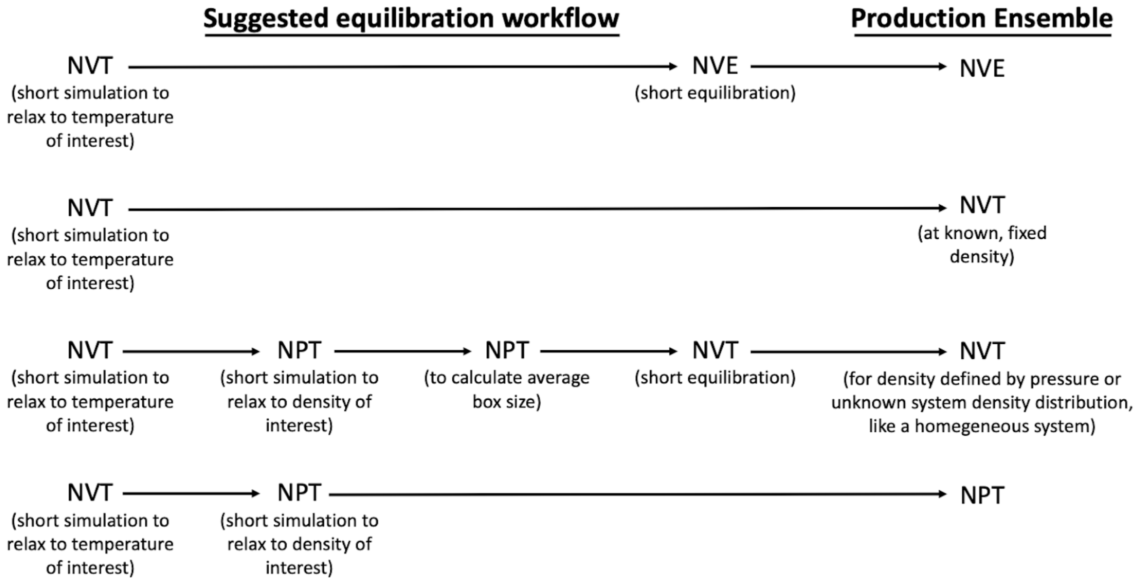


Figure 23. An suggested equilibration and production ensemble workflow for MD simulations. Usually, an initial phase at constant volume and temperature (NVT) is required to reach the desired temperature or energy of interest (even if the goal is to also bring the system to a target pressure). For NVE ensembles (first row), it is recommended to use an initial NVT phase followed by a short additional NVE equilibration. For NVT ensembles (second and third row), protocols may differ after the initial NVT phase depending if the density/volume is known or not. For NPT ensembles (fourth row), an initial NVT phase is followed by a NPT equilibration in order to equilibrate the system into a target pressure. After the equilibration step, the production ensemble is performed. Figure 5 of “Best Practices for Foundations in Molecular Simulations [Article v1.0]” (2018), Living J. Comp. Mol. Sci. 2019, 1(1), 5957, DOI: 10.33011/livecoms.1.1.5957.

MD simulations are relatively expensive compared to other cheaper computational approaches. This is because the displacement of all atoms of the system is computed for each timestep, which is an herculean effort that typically requires GPU parallelization. With the current technologies, the daily length of a regular MD simulation of a protein-protein system is in range between tens and hundreds of nanoseconds per day, depending on the amount of atoms of the system and the computational potency. Most of the processes happen in nature within the order of milliseconds to seconds, and therefore they would require extensive simulation processes. Therefore, one must know the current limitations of the MD simulations, and they must be used accordingly. Aiming to address those limitations, several enhancing sampling methods have been developed such as umbrella sampling MD or replica exchange. Most of those methods consist of forcing/directing the evolution of the dynamics of the system into a desired conformation/effect. Despite that those techniques are of great interest, they are out of the scope of this thesis.

Monte Carlo

Monte Carlo (MC) methods are based on probabilistic rules to generate a new configuration from the previous one (known as a MC step). The consecutive combination of MC steps generate a sequence of states that can be used to estimate structural and thermodynamic properties. However, MC simulations are time-independent, and dynamical properties based on the evolution over time can not be estimated. This is true for almost all MC algorithms with the exception of the Kinetic Monte Carlo techniques, where dynamical properties can be extracted from the simulations. The idea behind MC methods is that the ensemble of configurations generated reflect those that could be dynamically sampled by large MD simulations. There are different MC algorithms for performing biomolecular simulations, including PELE (Borrelli *et al.*, 2005), MCPRO (Jorgensen and Tirado-Rives, 2005) and ProtoMS (Woods *et al.*, 2018), among others. In the “***Predicting antibody binding efficacy: PELE-antibody***” result section we describe a PELE workflow for predicting antibody binding efficacy. Hence, we will focus on the description of PELE as a representative MC method.

PELE stands for Protein Energy Landscape Exploration, and the method was first published in 2005 (Borrelli *et al.*, 2005). PELE is based on the notion that an efficient conformational sampling can be obtained by the generation of a large ensemble of conformations by applying small/mid perturbations to the system in every PELE step. PELE, contrary to the other mentioned MC methods, takes into account a complex perturbation step including protein structure prediction algorithms. Hence, proteins simulated by PELE are not considered as rigid structures (such as in docking algorithms), but their structure is harmonically sampled over the entire simulation (giving the notion that the protein "breathes"). A basic scheme of a PELE step is illustrated in **Figure 24**.

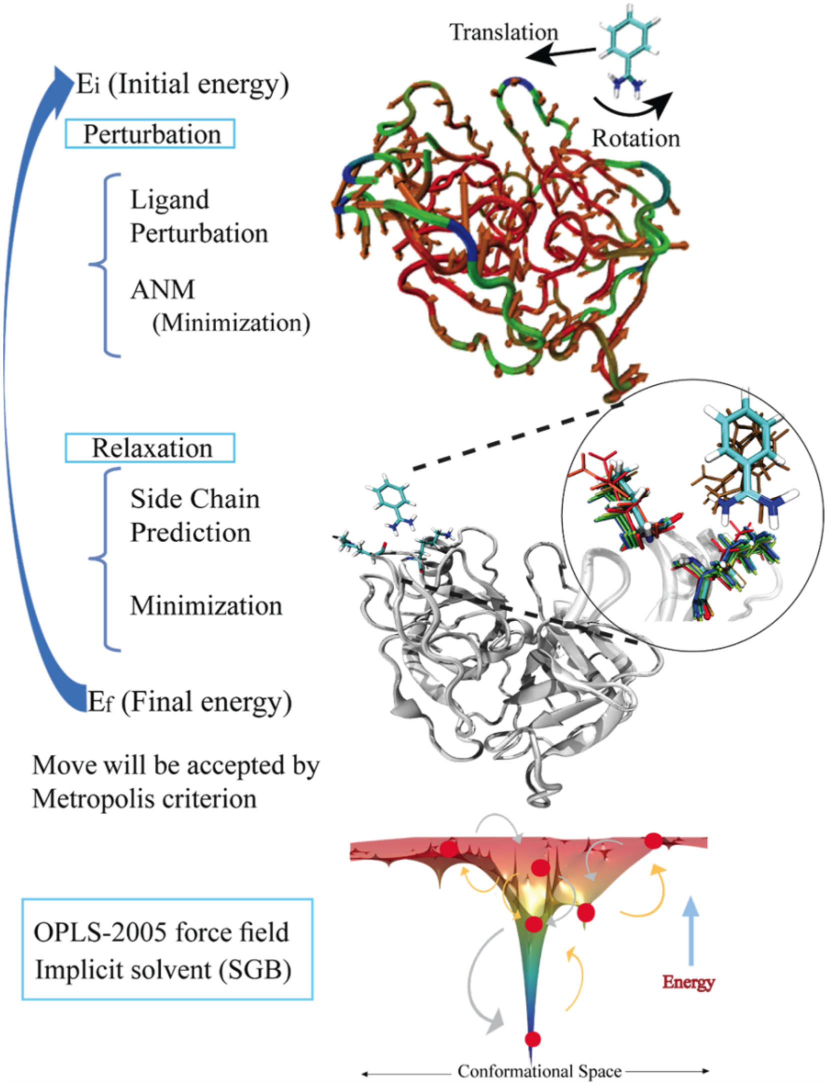


Figure 24. Schematic representation of a PELE step. (Image author: Ryoji Takahashi).

Each PELE step is composed of two phases: perturbation and relaxation. The algorithm is based on the concepts of MC minimization (Li and Scheraga, 1987), activation-relaxation fundamentals (Malek and Mousseau, 2000) and basin hoppin (Wales and Doye, 1997). The initial stage of PELE consists of measuring the initial total energy of the system (*Ei*). Then, PELE initiates the perturbation phase, where the ligand is randomly translated and rotated. Here, several movements are sampled and the one having the lowest energy is selected for the following stages. Next, a protein structure prediction protocol is applied by means of Anisotropic Network Models (ANM) (Atilgan *et al.*, 2001). This protocol is typically used for every 4-5 PELE steps. Then, an energy minimization protocol is applied where the alpha carbons of the protein backbone are typically constrained in order to favour protein backbone reorganization.

The relaxation process begins after the perturbation stage. Initially, a side chain prediction protocol is applied aiming to sample local interactions, such as hydrogen bonds or hydrophobic forces (Jacobson, Kaminski, *et al.*, 2002). Then, a minimization stage is used to relax the entire complex aiming to improve the acceptance of the movement. After the minimization, the total final energy (*Er*) is computed and is compared with the total initial energy. In short, the movement is always accepted if there is an energy gain after the movement. If not, the movement may be accepted or not depending on the Metropolis criterion (Metropolis and Ulam, 1949). If the movement is accepted, it will be used as the starting conformer for the next PELE step; otherwise, the previous conformer is used as starting point for the following PELE step.

PELE can use two different force fields, OPLS2005 (Kaminski *et al.*, 2001; Banks *et al.*, 2005) or AMBER99sbBSC0 (Pérez *et al.*, 2007), and two implicit solvents, OBC (Onufriev *et al.*, 2004) or VDGBNP (Zhu *et al.*, 2007). Several external packages developed in the group take advantage of the PELE algorithm for diverse finalities: enhancing the sampling efficiency to reduce the computational demand (Lecina *et al.*, 2017), estimating binding free energies through Markov State Models (Gilabert *et al.*, 2019, 2020), hit-to-lead drug design by means of dynamic ligand growing (Perez *et al.*, 2020) and more recently, MC perturbation of explicit waters within biological systems. PELE has been typically linked to drug design (Branco *et al.*, 2020; Díaz *et al.*, 2020; Saen-Oon *et al.*, 2019; Kotev *et al.*, 2018), ligand diffusion assays (Carro *et al.*, 2018) and enzyme design campaigns (Khersonsky *et al.*, 2018; Salas *et al.*, 2019; Santiago *et al.*, 2016; Pardo *et al.*, 2016).

Objectives

Along the introduction, the basis of the mechanisms of the immune system have been presented, paying special attention to antibodies and to the computational protein design field. Here, the main aim of this thesis is the development and application of computational algorithms and workflows for the design of antibodies for biomedical applications. To achieve this goal, we have conducted numerous computational studies which ultimately lead to the design of a potent antibody variant with experimental characterization. Hence, this thesis is presented as a “compendium” of research works, some of them already published in the literature, amining to define the following specific objectives:

- Elucidate the principles of protein-protein design and understand the basis of ΔΔG predictors. This led to the development of UEP, our estimator of the effects of mutations in protein-protein complexes.
- Unveil the principles of antibody binding affinity and to provide insights about the resistance mechanisms of hypermutated HIV-1 epitopes towards bNAbs.
- Apply computer-aided molecular simulations techniques to design an improved variant of the N6 bNAb, one of the most potent antibodies against HIV-1.

Results

In this section, we will describe the results of three computational projects for understanding the principles of protein-protein design. Firstly, we describe UEP, a method developed during this thesis for predicting the effects of mutations in protein-protein complexes. Moreover, the performance of UEP is compared with multiple ΔΔG state-of-the art predictors (see ***Describing the ΔΔG upon mutation: ΔΔG predictors*** section for a detail of the methods). Secondly, we will describe a protocol using PELE (see ***Molecular mechanics: Monte Carlo*** *section for a detail of the method*) for predicting antibody binding efficacy towards hypermutated HIV-1 epitopes. Thirdly, we will describe three *in silico* attempts using all above described methodologies (see ***Molecular mechanics*** section) for improving the binding efficacy of the N6 antibody (see ***Designing a potent broadly neutralizing antibody*** section for an overview of the N6 antibody).

Predicting ΔΔG: UEP

Almost all ΔΔGs for predicting the effects of mutations in protein-protein complexes fit their scoring function to experimental ΔΔG determinations to generate predictive power. However, as discussed in the section "***Databases related with ΔΔG predictors***", experimental data is limited and highly heterogeneous which may impact the reliability of the predictions.

Here, we propose UEP, a ΔΔG predictor trained on other sources of data different from experimental ΔΔG determinations. Our results indicate that UEP shows competitive predictive performance than the best evaluated state-of-the art algorithms to perform this task. The algorithm has been recently published (Amengual-Rigo, Fernández-Recio, *et al.*, 2020).

UEP algorithm

UEP algorithm takes advantage of the representative release of the Interactome3D database (2019-01 version) to generate a predictive model based on statistical potentials. Experimental binding determinations of mutations in protein-protein complexes (from SKEMPI 2.0) were used to determine the performance of UEP, together with the state-of-the art methods described in the “***ΔΔG predictors***” section: FoldX, EvoEF1, EvoEF2, pyDock, BeAtMuSiC, PRODIGY and mCSM.

All protein-protein complexes found in SKEMPI 2.0 are also found on the interactome Interactome3D database, and moreover, there is a high degree of redundancy between the complexes used in both databases. Aiming to minimize the risk of observing results biased towards this redundancy effect, we aimed to filter out the complexes from Interactome3D (that will be used to develop the predictive power of UEP) that are similar to the SKEMPI 2.0 ones. Hence, all possible pairwise sequence alignments between Interactome3D (33,607) and SKEMPI 2.0 (345) protein-protein complexes were performed using the BLOSUM62 matrix (Henikoff and Henikoff, 1992) and the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970). Sequence identity analysis revealed that most Interactome3D complexes (31,736 of 33,607) share less than 30% identity to any protein involved in a protein-protein complex of SKEMPI 2.0 (**Figure 25**). Thus, this threshold (30%) was used for selecting the complexes from Interactome3D to train UEP (31,736 protein-protein complexes sharing less than 30% sequence identity to any of the complexes found in SKEMPI 2.0). Overall, 13,773 experimental structures, 11,044 models from generic PDB templates and 6,919 from domain-domain structural templates were selected.

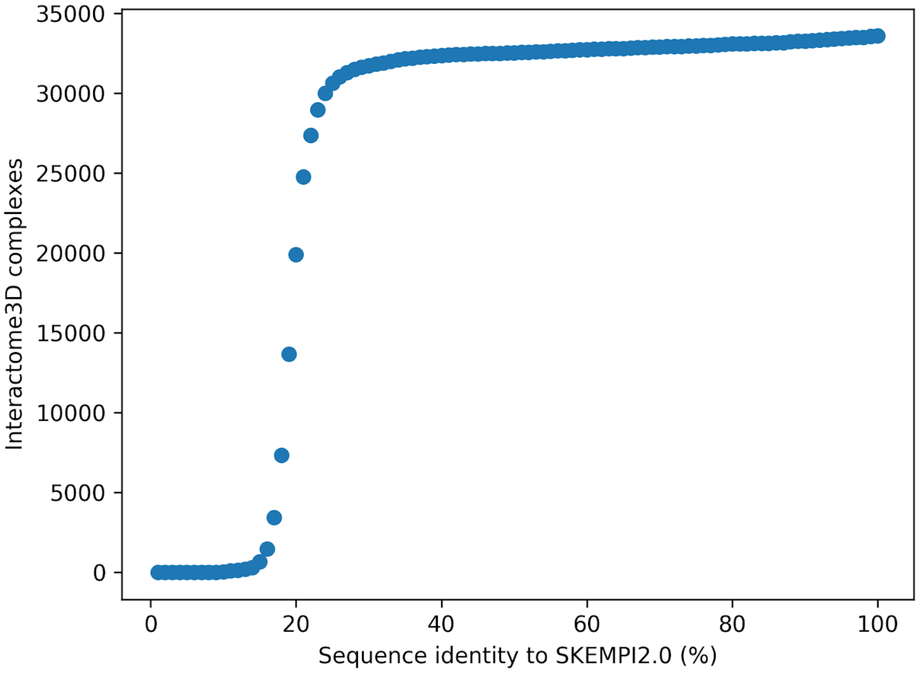


Figure 25. Sequence identity from pairwise alignments between the Interactome3D and SKEMPI 2.0 protein-protein complexes. Most of the Interactome3D complexes (31,736 of 33,607) share less than 30% to any of the proteins found in SKEMPI 2.0. Selected protein-protein complexes were used to train the predictive power of UEP.

Once that dissimilar protein-protein complexes from Interactome3D were selected, we proceeded to build the UEP algorithm. Our goal was to generate a fast algorithm to generate rapid scans on protein-protein interfaces, and therefore we designed the algorithm in such a way. A representative scheme of the UEP algorithm is illustrated in **Figure 26**. To start with, UEP algorithm takes advantage of interactome data of protein-protein complexes (the ones with low sequence identity <30% to SKEMPI 2.0). For every protein-protein complex, UEP scans the highly-packed residues of the protein-protein interface. The highly-packed term refers to positions having more than two intermolecular heavy-atom contacts within a distance of 5 Å (this is, being at least in contact with two different residues). We observed that mutations on the highly-packed region exert larger impacts on the experimental ΔΔG than the non-highly packed ones (**Table 1**), and therefore they are easier to be correctly predicted by any predictor. We also observed that mutations other than alanine exert larger impacts on the experimental ΔΔG than alanine mutations, and therefore, they are also easier to be correctly predicted by any method.

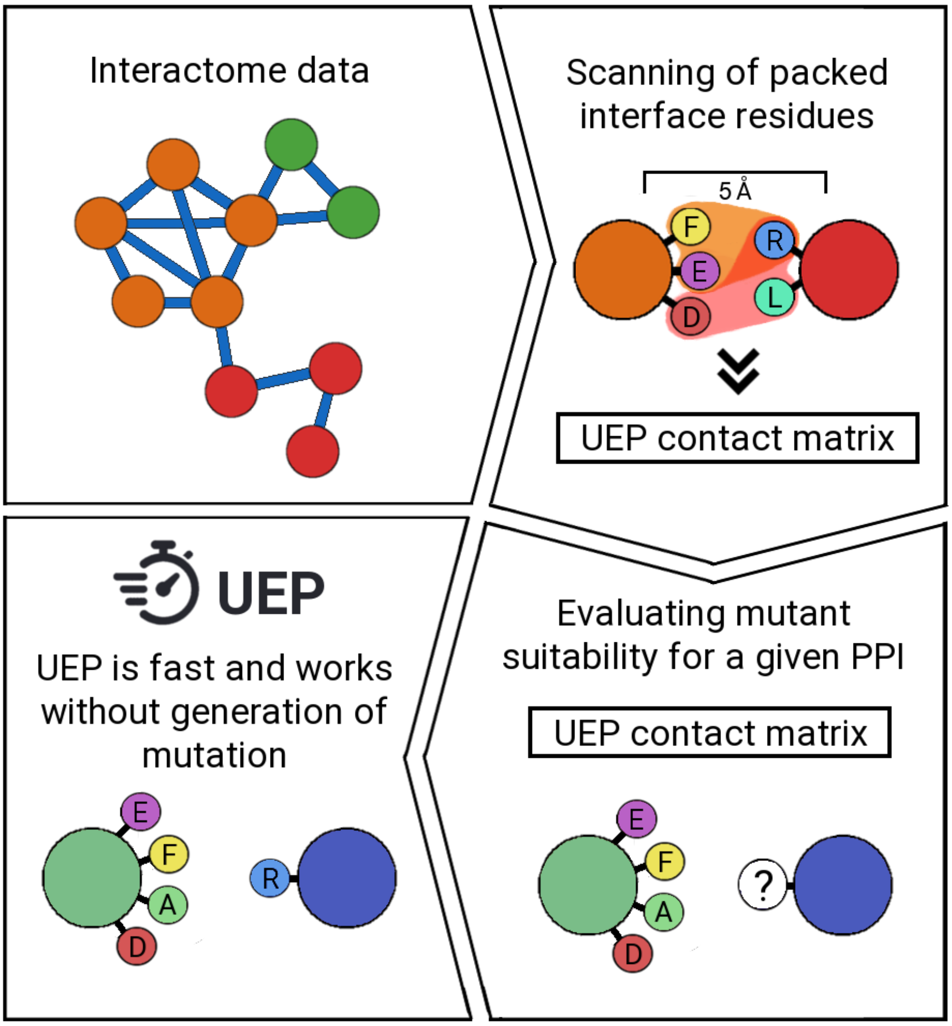


Figure 26. Scheme of the UEP algorithm. UEP is based on interactome data of protein-protein complexes. From those structures, UEP scans the packed interface residues to generate the UEP contact matrix. Such contact matrix can be used for the prediction of suitability of mutations for certain positions at a packed interface region. This process is fast and works without the generation of mutation files, allowing for computationally inexpensive large screening processes.

Mutation	**Alanine**		**Other than alanine**	
Highly-packed interface	YES	NO	YES	NO
ΔΔG increase (kcal/mol)	-0.79	-0.44	-2.24	-1.28
ΔΔG decrease (kcal/mol)	2.84	1.23	3.63	1.72

Table 1. Summary of experimental ΔΔG determinations in SKEMPI 2.0. Average ΔΔG determinations (kcal/mol) upon mutation depending on: (i) mutation nature: if the mutation is to alanine or to other than alanine, and ii) if the mutation data belongs to a highly-packed interface: more than two intermolecular heavy-atom contacts within a distance of 5 Å.

Once that the highly-packed residues have been located in the three-dimensional space, a statistical potential scheme is built: the amino acids forming intermolecular contacts with a highly-packed residue are taken into account following a three body scheme. In the three-body scheme, one residue of one protein is in contact with (at least) a pair of residues of the other protein. In the case that the number of contacts is higher than two, the combination of all possible pairs of residues (without repetition and order) is performed. An example is depicted here: (protein A: TRP 27) is in contact with (protein B: PRO 45, TYR 47 and SER 48). In this way, contacts that would be accounted into the contact matrix for TRP would be PRO-TYR, PRO-SER and TYR-SER. We observed that this contact scheme results in higher prediction accuracy than using the classical pairwise contact scheme.

Benefits of using the UEP three-body scheme compared to a pairwise architecture is shown in **Table 2-3**, for mutations to alanine and to other than alanine, respectively. A description of such statistical determinations (PPV, NPV, TPR, TNR and MCC) is provided in the "***UEP compared to the state-of-the art***" section. As can be observed, mutations other than alanine are easier to be predicted (in this case, by UEP), since the MCC values are higher (0.20 and 0.12, for the three-body and pairwise contact, respectively) than mutations to alanine (0.10 and -0.04, for the three-body and pairwise contact, respectively).

Three-body contact

	C+	C−		
P+	25	94	**PPV**	0.21
P−	97	769	**NPV**	0.89
TPR	0.20	**TNR**	0.89	**MCC**
				0.10

Pairwise contact

	C+	C−		
P+	11	112	**PPV**	0.30
P−	111	751	**NPV**	0.80
TPR	0.52	**TNR**	0.61	**MCC**
				-0.04

Table 2. Performance of the three-body contact potential compared to the classical pairwise contact scheme for 985 mutations to alanine evaluated in this work. P+/- indicate the prediction output, while the C+/- indicate experimental value (increasing and decreasing the ΔΔG, respectively. Statistical descriptors PPV, NPV, TPR, NPR and MCC are shown.

Three-body contact

	C+	C−		
P+	197	404	**PPV**	0.37
P−	101	549	**NPV**	0.83
TPR	0.66	**TNR**	0.58	**MCC**
				0.20

Pairwise contact

	C+	C−		
P+	155	367	**PPV**	0.30
P−	143	586	**NPV**	0.80
TPR	0.52	**TNR**	0.61	**MCC**
				0.12

Table 3. Performance of the three-body contact potential compared to the canonical pairwise contact scheme for 1251 mutations other than alanine evaluated in this work. P+/- indicate the prediction output, while the C+/- indicate experimental value (increasing and decreasing the ΔΔG, respectively. Statistical descriptors PPV, NPV, TPR, NPR and MCC are shown.

From those results we concluded that the predictive power towards alanine mutations is still limited (not only by UEP but also for all the other predictors, illustrated in the following sections), probably because of the short range of experimental ΔΔG determinations compared to mutations to other than alanine (**Table 1**). Here, the average experimental ΔΔG determinations of mutations to alanine located in the highly-packed interface are -0.79 and 2.84 kcal/mol (for mutations improving and decreasing the binding energy, respectively), while for mutations to alanine placed in the non-highly packed region are -0.44 and 1.23 kcal/mol (mutations improving and decreasing the binding energy, respectively). However, larger changes in experimental ΔΔG were observed for mutations to other than alanine for mutations placed in the highly-packed interface: -2.24 and 3.63 kcal/mol (mutations improving and decreasing the binding energy, respectively) and for mutations placed in the non-highly packed region: -1.28 and 1.72 (mutations improving and decreasing the binding energy, respectively).

The UEP contact matrix is generated after the scan of the selected complexes of Interactome3D. The matrix consists of the sum of all of the three-body contacts observed in all protein-protein complexes. After constructing the matrix, the method is able to make computationally inexpensive predictions of the mutations in highly-packed positions from any protein-protein complex. In order to do so, UEP only works on the native structure, and does not rely on the modelization of the mutation, which can be a time-consuming step. In this sense, the average time for modelling a single mutation using FoldX and EvoEF1/2 was ~30 and ~5 seconds, respectively, while the evaluation of the energy of the complex takes less than a second for every single predictor. Hence, the need to explicitly generate the models for the mutations highly increases the computational requirements in some orders of magnitude, which is counterproductive for large screening processes.

Instead of generating (or using) mutation files, UEP works on the native structure and approximates the contacts that the mutation would have by estimating changes in side-chain volume (Lin *et al.*, 2008). We observed that this simplification correctly approximates the real contacts of the mutation. After deciphering the contacts of the native and the mutant amino acids with the partner chain, two scores (for native and mutant residues) are obtained following the three-body scheme described previously. Both scores are normalized by the frequency of finding those residues (native and mutant, respectively) in the UEP contact matrix. Thus, a potential $\Delta\Delta G_{UEP}$ can be estimated from the ratio of such normalized scores (Moal and Fernandez-Recio, 2013).

$$\Delta\Delta G_{UEP} = -log\ \left(\frac{Normalized\ mutant\ score}{Normalized\ native\ score}\right) \quad (8)$$

In this sense, UEP does not perform additional normalizations based on physicochemical properties, such as hydrophobicity, polarity or charges (such as other predictors described along this thesis). We expected that our contact network would implicitly address such features by the observed contact frequencies in the interactome data. In the following section, we evaluated the ΔΔG predictions of UEP compared to the state-of-the art algorithms described previously.

UEP compared to the state-of-the art

UEP performance was evaluated for predicting SKEMPI 2.0 data together with five physical-energy descriptors (FoldX, EvoEF1, EvoEF2, pyDock, and PRODIGY), an statistical potential method (BeAtMuSiC) and a machine-learning based one (mCSM). All single mutations in SKEMPI 2.0 having no discrepancies on the experimental binding energies were used. This is, all determinations pointing out that the mutation increases or decreases the binding affinity in an unanimous way. This resulted in an amount of 2103 mutations to alanine and 1762 mutations to other than alanine. From those, 985 mutations to alanine and 1251 mutations to other than alanine are embedded within the highly-packed interface region, and can be scored by UEP (representing 264 out of 354 protein-protein complexes in the database). A comparison of the binding affinity data between highly-packed/non-highly packed regions and mutations to alanine/other than alanine has been previously exposed in **Table 1**. We splitted the SKEMPI 2.0 dataset in different ways to provide insights about the prediction mechanism by all methods regarding their ability to classify (improving/decreasing the binding affinity) and to correlate mutations (with the experimental binding values). For both classification and correlation studies, mutations to alanine and to other than alanine were evaluated independently.

Evaluation of the classification performance was based on the statistical analysis of the confusion matrix. Confusion matrix indicates the amount of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). Hence, the confusion matrix interprets the input conditions (C+ and C-) and the output predictions (P+ and P-) of a simulation case. A simple confusion matrix looks as following:

	C+	**C-**
P+	TP	FP
P-	FN	TN

Here, TP and TN represent the amount of mutations correctly predicted as such: this is, mutations predicted to increase (P+ and C+) or decrease (P- and C-) the binding affinity of the complex, respectively. FP represents the amount of mutations incorrectly predicted to increase the binding affinity, while in fact they are decreasing it (P+ and C-). FN represents the amount of mutations incorrectly predicted to decrease the binding affinity, while they are increasing it (P- and C+).

Several statistical descriptors for classification analysis can be computed from the confusion matrix: Positive Predictive Value (PPV), Negative Predictive Value (NPV), True Positive Rate (TPR), True Negative Rate (TNR) and Matthew's Correlation Coefficient (MCC).

- PPV (**Equation 9**) is also known as precision, and indicates the proportion of TP from the ones predicted to be as such (increasing the binding energy).

$$PPV = \frac{TP}{TP+FP} \tag{9}$$

- In a similar way to PPV, NPV (**Equation 10**) indicates the proportion of TN from the ones predicted to be as such (decreasing the binding energy).

$$NPV = \frac{TN}{TN+FN} \tag{10}$$

- TPR (**Equation 11**), also known as recall, sensitivity or power, indicates the fraction of TP from the amount of entries that are experimentally increasing the binding energy.

$$TPR = \frac{TP}{TP+FN} \tag{11}$$

- In a similar way to TPR, TNR (**Equation 12**) also known as specificity, represents the fraction of TN from the amount of entries that are experimentally decreasing the binding energy.

$$TNR = \frac{TN}{TN+FP} \tag{12}$$

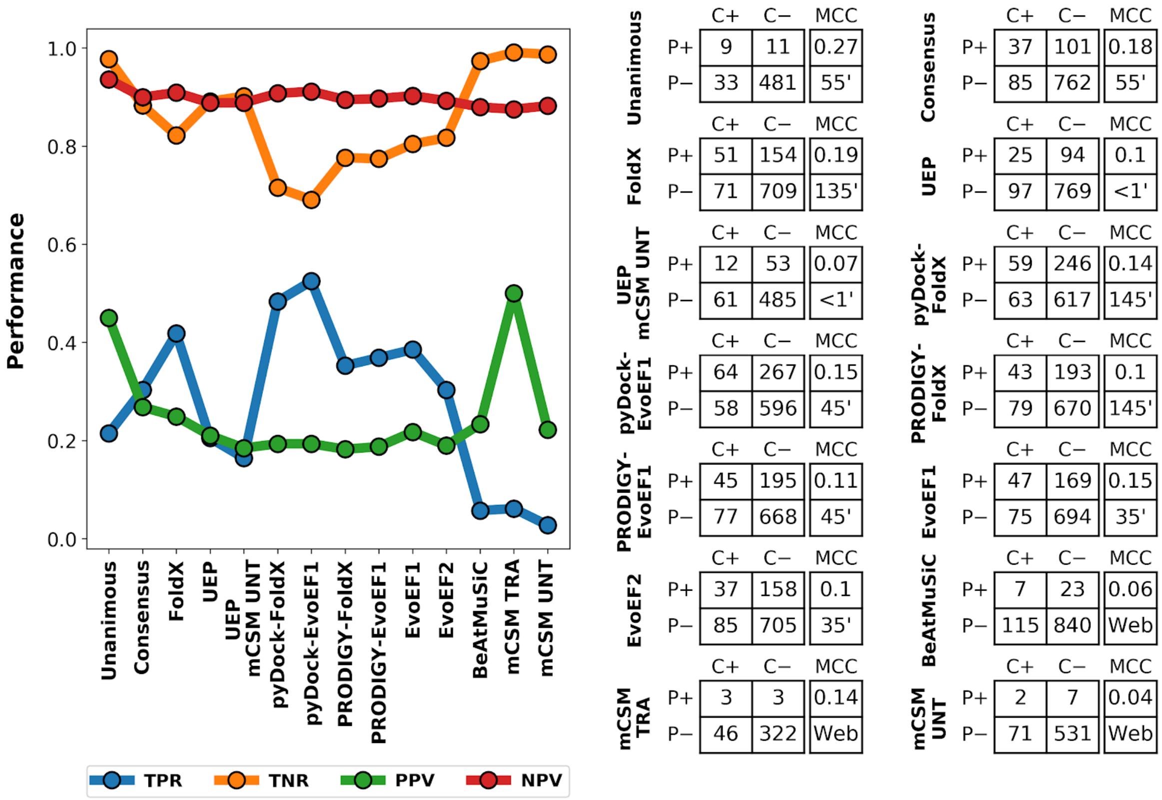
- MCC (**Equation 13**) is a statistical measurement widely used for classifying unbalanced data, which is the case for the SKEMPI 2.0 data, since only 122 out of 985 mutations to alanine, and 298 out of 1251 mutations other than alanine increase the binding affinity compared to the native counterpart.

$$MCC = \frac{(TPxTN)-(FPxFN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{13}$$

- On the other hand, evaluation of the correlation performance was based on the Pearson Correlation Coefficient (PCC) and the Root Mean Square Error (RMSE) between experimental and predicted ΔΔG.

As mentioned previously, mutations were grouped depending on their location: if they are placed on the highly-packed interface or not. Regarding the highly-packed interface group, **Figure 27** and **Figure 28** represent the confusion matrices from all methods for mutations to alanine and mutations to other than alanine, respectively.

Full data regarding the confusion matrices and the statistical descriptors is shown for highly-packed positions in **Table 4** and **Table 5** (for mutations to alanine and mutations to other than alanine, respectively) and for non-highly-packed positions in **Table 6** and **Table 7** (for mutations to alanine and mutations to other than alanine, respectively).



Unanimous

	C+	C−	MCC
P+	9	11	0.27
P−	33	481	55'

Consensus

	C+	C−	MCC
P+	37	101	0.18
P−	85	762	55'

FoldX

	C+	C−	MCC
P+	51	154	0.19
P−	71	709	135'

UEP

	C+	C−	MCC
P+	25	94	0.1
P−	97	769	<1'

UEP mCSM UNT

	C+	C−	MCC
P+	12	53	0.07
P−	61	485	<1'

pyDock-FoldX

	C+	C−	MCC
P+	59	246	0.14
P−	63	617	145'

pyDock-EvoEF1

	C+	C−	MCC
P+	64	267	0.15
P−	58	596	45'

PRODIGY-FoldX

	C+	C−	MCC
P+	43	193	0.1
P−	79	670	145'

PRODIGY-EvoEF1

	C+	C−	MCC
P+	45	195	0.11
P−	77	668	45'

EvoEF1

	C+	C−	MCC
P+	47	169	0.15
P−	75	694	35'

EvoEF2

	C+	C−	MCC
P+	37	158	0.1
P−	85	705	35'

BeAtMuSiC

	C+	C−	MCC
P+	7	23	0.06
P−	115	840	Web

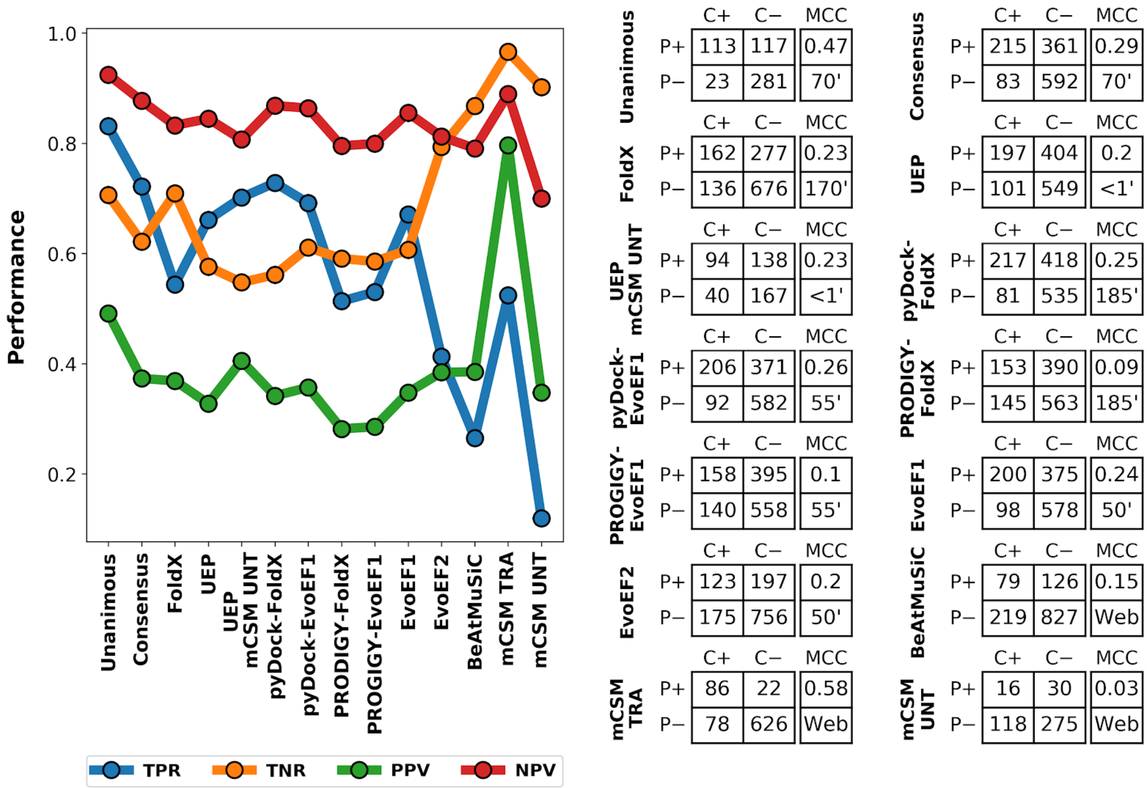
mCSM TRA

	C+	C−	MCC
P+	3	3	0.14
P−	46	322	Web

mCSM UNT

	C+	C−	MCC
P+	2	7	0.04
P−	71	531	Web

Figure 27. Performance of all predictors on the 985 alanine mutations embedded within the highly-packed region. Left panel shows the TPR, TNR, PPV, and NPV patterns. On the right panel, confusion matrices are shown (same data as **Table 4**): experimental conditions are represented in vertical (C+ or C−, if mutation increases or decreases experimental binding affinity, respectively) while predictions are represented in horizontal (P+ or P−, if mutation is predicted to increase or decrease the binding affinity). MCC scores and the approximated computer wall clock time are also provided. Time is not shown for web server based methods.



Unanimous	C+	C−	MCC
P+	113	117	0.47
P−	23	281	70'

Consensus	C+	C−	MCC
P+	215	361	0.29
P−	83	592	70'

FoldX	C+	C−	MCC
P+	162	277	0.23
P−	136	676	170'

UEP	C+	C−	MCC
P+	197	404	0.2
P−	101	549	<1'

UEP mCSM UNT	C+	C−	MCC
P+	94	138	0.23
P−	40	167	<1'

pyDock-FoldX	C+	C−	MCC
P+	217	418	0.25
P−	81	535	185'

pyDock-EvoEF1	C+	C−	MCC
P+	206	371	0.26
P−	92	582	55'

PRODIGY-FoldX	C+	C−	MCC
P+	153	390	0.09
P−	145	563	185'

PROGIGY-EvoEF1	C+	C−	MCC
P+	158	395	0.1
P−	140	558	55'

EvoEF1	C+	C−	MCC
P+	200	375	0.24
P−	98	578	50'

EvoEF2	C+	C−	MCC
P+	123	197	0.2
P−	175	756	50'

BeAtMuSiC	C+	C−	MCC
P+	79	126	0.15
P−	219	827	Web

mCSM TRA	C+	C−	MCC
P+	86	22	0.58
P−	78	626	Web

mCSM UNT	C+	C−	MCC
P+	16	30	0.03
P−	118	275	Web

Figure 28. Performance of all predictors on the 1251 mutations to other than alanine embedded within the highly-packed region. Left panel shows the TPR, TNR, PPV, and NPV patterns. On the right panel, confusion matrices are shown (same data as **Table 5**): experimental conditions are represented in vertical (C+ or C−, if mutation increases or decreases experimental binding affinity, respectively) while predictions are represented in horizontal (P+ or P−, if mutation is predicted to increase or decrease the binding affinity). MCC scores and the approximated computer wall clock time are also provided. Time is not shown for web server based methods.

Unanimous	C+	C−	PPV/NPV
P+	9	11	0.45
P−	33	481	0.94
TPR/NPR	0.21	0.98	MCC: 0.27

Consensus	C+	C−	PPV/NPV
P+	37	101	0.27
P−	85	762	0.90
TPR/NPR	0.30	0.88	MCC: 0.18

FoldX	C+	C−	PPV/NPV
P+	51	154	0.25
P−	71	709	0.91
TPR/NPR	0.42	0.82	MCC: 0.19

UEP	C+	C−	PPV/NPV
P+	25	94	0.21
P−	97	769	0.89
TPR/NPR	0.20	0.89	MCC: 0.10

UEP UNT	C+	C−	PPV/NPV
P+	12	53	0.18
P−	61	485	0.89
TPR/NPR	0.16	0.90	MCC: 0.07

pyDock-F	C+	C−	PPV/NPV
P+	59	246	0.19
P−	63	617	0.91
TPR/NPR	0.48	0.71	MCC: 0.14

pyDock-E	C+	C−	PPV/NPV
P+	64	267	0.19
P−	58	596	0.91
TPR/NPR	0.52	0.69	MCC: 0.15

PRODIGY-F	C+	C−	PPV/NPV
P+	43	193	0.18
P−	79	670	0.89
TPR/NPR	0.35	0.78	MCC: 0.10

PRODIGY-E	C+	C−	PPV/NPV
P+	45	195	0.19
P−	77	668	0.90
TPR/NPR	0.37	0.77	MCC: 0.11

EvoEF1	C+	C−	PPV/NPV
P+	47	169	0.22
P−	75	694	0.90
TPR/NPR	0.39	0.80	MCC: 0.15

EvoEF2	C+	C−	PPV/NPV
P+	37	158	0.19
P−	85	705	0.89
TPR/NPR	0.30	0.82	MCC: 0.10

BeAtMuSiC	C+	C−	PPV/NPV
P+	7	23	0.23
P−	115	840	0.88
TPR/NPR	0.06	0.97	MCC: 0.06

mCSM TRA	C+	C−	PPV/NPV
P+	3	3	0.50
P−	46	322	0.88
TPR/NPR	0.06	0.99	MCC: 0.14

mCSM UNT	C+	C−	PPV/NPV
P+	2	7	0.22
P−	71	531	0.88
TPR/NPR	0.03	0.99	MCC: 0.04

Table 4. Confusion matrices for 985 mutations to alanine embedded within the highly-packed interface region. P+/- indicate the prediction output, while the C+/- indicate experimental value (increasing and decreasing the ΔΔG, respectively. Statistical descriptors PPV, NPV, TPR, NPR and MCC are shown.

Unanimous	**C+**	**C−**	**PPV/NPV**
P+	113	117	0.49
P−	23	281	0.92
TPR/NPR	0.83	0.71	MCC: 0.47

Consensus	**C+**	**C−**	**PPV/NPV**
P+	215	361	0.37
P−	83	592	0.88
TPR/NPR	0.72	0.62	MCC: 0.29

FoldX	**C+**	**C−**	**PPV/NPV**
P+	162	277	0.37
P−	136	676	0.83
TPR/NPR	0.54	0.71	MCC: 0.23

UEP	**C+**	**C−**	**PPV/NPV**
P+	197	404	0.33
P−	101	549	0.84
TPR/NPR	0.66	0.58	MCC: 0.20

UEP UNT	**C+**	**C−**	**PPV/NPV**
P+	94	138	0.41
P−	40	167	0.81
TPR/NPR	0.70	0.55	MCC: 0.23

pyDock-F	**C+**	**C−**	**PPV/NPV**
P+	217	418	0.34
P−	81	535	0.87
TPR/NPR	0.73	0.56	MCC: 0.25

pyDock-E	**C+**	**C−**	**PPV/NPV**
P+	206	371	0.36
P−	92	582	0.86
TPR/NPR	0.69	0.61	MCC: 0.26

PRODIGY-F	**C+**	**C−**	**PPV/NPV**
P+	153	390	0.28
P−	145	563	0.80
TPR/NPR	0.51	0.59	MCC: 0.09

PRODIGY-E	**C+**	**C−**	**PPV/NPV**
P+	158	395	0.29
P−	140	558	0.80
TPR/NPR	0.53	0.59	MCC: 0.10

EvoEF1	**C+**	**C−**	**PPV/NPV**
P+	200	375	0.35
P−	98	578	0.86
TPR/NPR	0.67	0.61	MCC: 0.24

EvoEF2	**C+**	**C−**	**PPV/NPV**
P+	123	197	0.38
P−	175	756	0.81
TPR/NPR	0.41	0.79	MCC: 0.20

BeAtMuSiC	**C+**	**C−**	**PPV/NPV**
P+	79	126	0.39
P−	219	827	0.79
TPR/NPR	0.27	0.87	MCC: 0.15

mCSM TRA	**C+**	**C−**	**PPV/NPV**
P+	86	22	0.80
P−	78	626	0.89
TPR/NPR	0.52	0.97	MCC: 0.58

mCSM UNT	**C+**	**C−**	**PPV/NPV**
P+	16	30	0.35
P−	118	275	0.70
TPR/NPR	0.12	0.90	MCC: 0.03

Table 5. Confusion matrices for 1251 mutations to other than alanine embedded within the highly-packed interface region. P+/- indicate the prediction output, while the C+/- indicate experimental value (increasing and decreasing the ΔΔG, respectively. Statistical descriptors PPV, NPV, TPR, NPR and MCC are shown.

FoldX	C+	C−	PPV/NPV
P+	79	228	0.26
P−	193	618	0.76
TPR/NPR	0.29	0.73	MCC: 0.02

pyDock-F	C+	C−	PPV/NPV
P+	115	319	0.26
P−	157	527	0.77
TPR/NPR	0.42	0.62	MCC: 0.04

pyDock-E	C+	C−	PPV/NPV
P+	138	357	0.28
P−	134	489	0.78
TPR/NPR	0.51	0.58	MCC: 0.07

PRODIGY-F	C+	C−	PPV/NPV
P+	28	124	0.18
P−	244	722	0.75
TPR/NPR	0.10	0.85	MCC: -0.05

PRODIGY-E	C+	C−	PPV/NPV
P+	38	111	0.26
P−	234	735	0.76
TPR/NPR	0.14	0.87	MCC: 0.01

EvoEF1	C+	C−	PPV/NPV
P+	73	212	0.26
P−	199	634	0.76
TPR/NPR	0.27	0.75	MCC: 0.02

EvoEF2	C+	C−	PPV/NPV
P+	67	174	0.28
P−	205	672	0.77
TPR/NPR	0.25	0.79	MCC: 0.04

BeAtMuSiC	C+	C−	PPV/NPV
P+	28	71	0.28
P−	244	774	0.76
TPR/NPR	0.10	0.92	MCC: 0.03

mCSM TRA	C+	C−	PPV/NPV
P+	11	13	0.46
P−	107	294	0.73
TPR/NPR	0.09	0.96	MCC: 0.10

mCSM UNT	C+	C−	PPV/NPV
P+	20	53	0.27
P−	134	486	0.78
TPR/NPR	0.13	0.90	MCC: 0.04

Table 6. Confusion matrices for 1118 mutations to alanine located in the non-highly-packed interface region (the ones that UEP could not score because of a lack of intermolecular contacts). P+/- indicate the prediction output, while the C+/- indicate experimental value (increasing and decreasing the ΔΔG, respectively. Statistical descriptors PPV, NPV, TPR, NPR and MCC are shown.

FoldX	**C+**	**C−**	**PPV/NPV**
P+	75	132	0.36
P−	90	214	0.70
TPR/NPR	0.45	0.62	MCC: 0.07

pyDock-F	**C+**	**C−**	**PPV/NPV**
P+	108	171	0.39
P−	57	175	0.75
TPR/NPR	0.65	0.51	MCC: 0.15

pyDock-E	**C+**	**C−**	**PPV/NPV**
P+	105	156	0.40
P−	60	190	0.76
TPR/NPR	0.64	0.55	MCC: 0.17

PRODIGY-F	**C+**	**C−**	**PPV/NPV**
P+	73	151	0.33
P−	92	195	0.68
TPR/NPR	0.44	0.56	MCC: 0.01

PRODIGY-E	**C+**	**C−**	**PPV/NPV**
P+	74	144	0.34
P−	91	292	0.69
TPR/NPR	0.45	0.58	MCC: 0.03

EvoEF1	**C+**	**C−**	**PPV/NPV**
P+	66	161	0.29
P−	99	185	0.65
TPR/NPR	0.40	0.53	MCC: -0.06

EvoEF2	**C+**	**C−**	**PPV/NPV**
P+	69	111	0.38
P−	96	235	0.71
TPR/NPR	0.42	0.69	MCC: 0.1

BeAtMuSiC	**C+**	**C−**	**PPV/NPV**
P+	45	74	0.38
P−	120	272	0.69
TPR/NPR	0.27	0.79	MCC: 0.07

mCSM TRA	**C+**	**C−**	**PPV/NPV**
P+	25	6	0.81
P−	42	187	0.82
TPR/NPR	0.37	0.97	MCC: 0.46

mCSM UNT	**C+**	**C−**	**PPV/NPV**
P+	16	24	0.40
P−	82	129	0.61
TPR/NPR	0.16	0.84	MCC: 0.01

Table 7. Confusion matrices for 511 mutations to other than alanine located in the non-highly-packed interface region (the ones that UEP could not score because of a lack of intermolecular contacts). P+/- indicate the prediction output, while the C+/- indicate experimental value (increasing and decreasing the ΔΔG, respectively. Statistical descriptors PPV, NPV, TPR, NPR and MCC are shown.

Several conclusions can be extracted from the different simulation groups. Aiming to facilitate the comprehension of the results, MCC values from all groups were recopilated in **Table 8**. Here, it can be observed that the quality of the prediction differs depending on the location and the nature of the mutation. Overall, larger differences in MCC values have been observed between the location groups, where prediction of mutations placed in the highly-packed interface are more accurate than the non-highly packed group. The same pattern is observed for mutations other than alanine, which are easily to be correctly predicted than mutation to alanines.

Location	**Highly-packed interface**		**Non-highly-packed interface**	
Nature	**To alanine**	**To not alanine**	**To alanine**	**To not alanine**
Unanimous	0.27	0.47	n.d	n.d
Consensus	0.18	0.29	n.d	n.d
FoldX (FoldX)	0.19	0.23	0.02	0.07
UEP	0.10	0.20	n.d	n.d
UEP UNT	0.07	0.23	n.d	n.d
pyDock (FoldX)	0.14	0.25	0.04	0.15
pyDock (EvoEF1)	0.15	0.26	0.07	0.17
PRODIGY (FoldX)	0.10	0.09	-0.05	0.01
PRODITY (EvoEF1)	0.11	0.10	0.01	0.03
EvoEF1 (EvoEF1)	0.15	0.24	0.02	-0.06
EvoEF2 (EvoEF2)	0.10	0.20	0.04	0.10
BeAtMuSiC	0.06	0.15	0.03	0.07
mCSM TRA	0.14	0.58	0.10	0.46
mCSM UNT	0.04	0.03	0.04	0.01

Table 8. Summary of the MCC values from all predictors, depending on the location of the mutation (highly-packed interface or non-highly-packed interface) and nature of the mutation: mutations to alanine or to other than alanine (to not alanine). Method within parenthesis was used for the generation of the models. Mutations other than alanine are predicted accurately compared to mutations to alanine (large difference in MCC values). The same tendency was observed for comparing mutations depending on their position, where highly-packed mutations are predicted accurately for any method compared to non-highly packed positions.

The authors of the mCSM method reported which mutations from SKEMPI were used for the construction of their algorithm. We noticed that mCSM outperformed all other methods for predicting the entire dataset. Aiming to provide insights on this effect, we splitted the benchmark in two groups: those mutations reported to be used during the training of the mCSM algorithm (mCSM TRA) or not (mCSM UNT). After evaluating the performance of mCSM on both groups, we noticed a large drop in predictive performance for new mutations (MCCs of 0.04/0.04, and 0.03/0.01 for mutations to alanine and other than alanine, respectively) compared to the ones found in the training group (MCCs of 0.14/0.10 and 0.58/0.46 for mutations to alanine and other than alanine, respectively) (**Table 8**). As can be observed, the drop in performance is accentuated in the case of mutations to other than alanine.

It looks that the mCSM method shows a very high degree of overfitting towards already trained data, and that it encounters difficulties for classifying new mutations into improving and decreasing the binding energy of the complex. As can be observed in the confusion matrices (**Table 4-7**), mCSM tends to predict almost all new mutations as decreasing the binding energy of the complex, which explains the low MCC values. Because of this reason, it is very important to develop generalistic algorithms to avoid overfitting issues; otherwise their performance on real case scenarios will be lower than expected (as simulated here with a group of mutations not used for developing the mCSM algorithm). Aiming to discard possible effects related to difficulties for predicting this particular set of mutations, we compared the performance of UEP for classifying the same new group of mutations for mCSM. As shown in **Table 8**, UEP predictions on this set of mutations do not differ much in MCC terms compared to the entire UEP benchmark (MCC of 0.07 to 0.10, and 0.23 to 0.20 for mutations to alanine and to other than alanine, respectively). Hence, we discarded possible effects that make the prediction of mCSM UNT group more difficult than mCSM TRA, and we concluded that the main reason for this drop in performance is intrinsically related to the mCSM algorithm.

As may have been observed previously, some predictors called “Unanimous” and “Consensus” are described along this section. From the classification point of view, the decision of multiple algorithms may be consensuated. Therefore, the conclusion of multiple predictors would help to decrease particular errors of individual methods.

Hence, the decision of multiple predictors should provide a more robust indicator of the effect of the mutations, since it would involve the statistical sampling of multiple and different force fields. Here, we used the decision of three algorithms for deciding the effects of a mutation. In this sense and from the binary classification point of view of the effects of the mutations (this is improving or decreasing the binding affinity), a consensus decision is the choice of at least two out of three predictors, while a unanimous decision is when the three predictors agree. We evaluated the binary classification decision using the combination of the four best individual classifiers in our benchmark: UEP, pyDock, EvoEF1 and FoldX.

It is important to mention that following this approach results in a substantial increase in computational requirements. Hence, an increase in the classification performance must justify the increase in computational cost. Consensus and unanimous selection for mutations to alanine are represented in **Table 9**. On the other hand, consensus and unanimous selection of mutations other than alanine are represented in **Table 10**. In both tables, time for predicting the benchmarks is illustrated. We observed that the main bottleneck in computational time belongs to the generation of the mutation file, which can take 5 seconds (EvoEF1) to 30 seconds (FoldX). The groups of predictors that were used are: i) UEP, pyDock-EvoEF1 and EvoEF1, ii) UEP, pyDock-FoldX and FoldX, iii) UEP, EvoEF1 and FoldX, iv) FoldX, pyDock-FoldX and pyDock-EvoEF1, and v) EvoEF1, pyDock-EvoEF1 and pyDock-FoldX.

Predictors	Consensus	Unanimous	Time
UEP, pyDock-E, EvoEF1	0.19	0.10	55’
UEP, pyDock-F, FoldX	0.18	0.27	145’
UEP, EvoEF1, FoldX	0.16	0.21	175’
FoldX, pyDock-F, pyDock-E	0.18	0.27	185’
EvoEF1, pyDock-E, pyDock-F	0.18	0.18	185’

Table 9. MCC scores on the decision of three of the best classifiers of the effects of alanine mutations (985) placed on the highly-packed interface. Time for performing such simulations is shown.

Predictors	Consensus	Unanimous	Time
UEP, pyDock-E, EvoEF1	0.29	0.47	70’
UEP, pyDock-F, FoldX	0.29	0.47	185’
UEP, EvoEF1, FoldX	0.26	0.42	220’
FoldX, pyDock-F, pyDock-E	0.25	0.47	235’
EvoEF1, pyDock-E, pyDock-F	0.29	0.42	235’

Table 10. MCC scores on the decision of three of the best classifiers of the effects of mutations other than alanine (1251) placed on the highly-packed interface. Time for performing such simulations is shown.

For the design of an enhanced protein-protein interface, mutations other than alanine may be more suitable depending on the context. This is represented by the fact that we have observed that this group of mutations exert a larger impact on the experimental ΔΔG and therefore, they are easier to be correctly predicted by any method. Moreover, and without taking into account a few exceptions related to conformational dynamics, mutations other than alanine are more likely to improve the binding energy of the complex than mutations to alanine.

Because of this, the following sections will focus on the analysis of mutations other than alanine. As observed in **Table 10**, both consensus and unanimous decisions of any combination result in a classification improvement compared to any method alone (**Figure 28** and **Table 5**). From both selections, the unanimous one achieves higher classification performance than the consensus, but cautious must be taken into account. Not all mutations are predicted unanimously by the three methods, and therefore the amount of selected candidates is reduced (from 1251 to 534). In this context, consensus selection may be an alternative approach to tackle most of the real case scenarios despite achieving lower classification performance than the unanimous selection. From all combinations, the use of UEP, EvoEF1 and pyDock (the latter used for the energy evaluation of the models generated by EvoEF1) is the one achieving the best tradeoff in accuracy and computational time. As can be observed the time needed for evaluating the 1251 mutations is ~70 minutes (almost three times less than the second faster combination, 185’).

After observing that the use of the decisions taken by multiple algorithms increases the classification performance, we aimed to assess if the predictors show some predictive particularities towards any group of mutations. In this sense, we splitted the benchmark of 1251 mutations to other than alanine depending on the changes in volume (ΔV) and changes in hydrophobicity (ΔH). Side chain volumes and hydrophobicity indices were extracted from literature to perform this evaluation (Lin *et al.*, 2008; Eisenberg *et al.*, 1984).

A threshold of $0.1mm^3$ was used to determine whether a side chain mutation increases the volume ($\Delta V>0.1mm^3$, Gain group, consisting of 490 mutations), decreases ($\Delta V<-0.1mm^3$, Loss group, consisting of 423 mutations) or shows a similar volume compared to its original residue ($|\Delta V|\leq 0.1mm^3$, Neutral group, consisting of 338 mutations). Confusion matrices of each predictor regarding those volume subgroups are found at **Tables 11-19** for UEP, pyDock-FoldX, pyDock-EvoEF1, FoldX, PRODIGY-FoldX, PRODIGY-EvoEF1, BeAtMuSiC, EvoEF1 and EvoEF2, respectively.

UEP	**C+**	**C−**	PPV/NPV
P+	197	404	0.33
P−	101	549	0.84
TPR/NPR	0.66	0.58	**0.20**

Gain	**C+**	**C−**	PPV/NPV
P+	140	263	0.35
P−	19	68	0.78
TPR/NPR	0.88	0.21	**0.11**

Neutral	**C+**	**C−**	PPV/NPV
P+	42	88	0.32
P−	34	174	0.84
TPR/NPR	0.55	0.66	**0.19**

Loss	**C+**	**C−**	PPV/NPV
P+	15	53	0.22
P−	48	307	0.86
TPR/NPR	0.24	0.85	**0.09**

Table 11. UEP confusion matrices depending on gain, neutral or loss volume upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in volume, bot left: neutral mutations, bot right: mutations resulting in a loss in volume. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

pyDock-F	**C+**	**C−**	PPV/NPV
P+	217	418	0.34
P−	81	535	0.87
TPR/NPR	0.73	0.56	**0.25**

Gain	**C+**	**C−**	PPV/NPV
P+	124	203	0.38
P−	35	128	0.78
TPR/NPR	0.78	0.39	**0.17**

Neutral	**C+**	**C−**	PPV/NPV
P+	61	125	0.33
P−	15	137	0.90
TPR/NPR	0.80	0.52	**0.27**

Loss	**C+**	**C−**	PPV/NPV
P+	32	90	0.26
P−	31	270	0.90
TPR/NPR	0.51	0.75	**0.20**

Table 12. pyDock confusion matrices on the FoldX mutation models depending on gain, neutral or loss volume upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in volume, bot left: neutral mutations, bot right: mutations resulting in a loss in volume. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

pyDock-E	**C+**	**C−**	PPV/NPV
P+	206	371	0.36
P−	92	582	0.86
TPR/NPR	0.69	0.61	**0.26**

Gain	**C+**	**C−**	PPV/NPV
P+	117	172	0.40
P−	42	159	0.79
TPR/NPR	0.74	0.48	**0.21**

Neutral	**C+**	**C−**	PPV/NPV
P+	59	106	0.35
P−	20	156	0.89
TPR/NPR	0.74	0.60	**0.28**

Loss	**C+**	**C−**	PPV/NPV
P+	33	93	0.26
P−	30	267	0.90
TPR/NPR	0.52	0.74	**0.21**

Table 13. pyDock confusion matrices on the EvoEF1 mutation models depending on gain, neutral or loss volume upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in volume, bot left: neutral mutations, bot right: mutations resulting in a loss in volume. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

FoldX	**C+**	**C−**	PPV/NPV
P+	162	277	0.37
P−	136	676	0.83
TPR/NPR	0.54	0.71	**0.23**

Gain	**C+**	**C−**	PPV/NPV
P+	108	132	0.45
P−	52	199	0.79
TPR/NPR	0.67	0.60	**0.26**

Neutral	**C+**	**C−**	PPV/NPV
P+	33	83	0.28
P−	43	179	0.81
TPR/NPR	0.43	0.68	**0.10**

Loss	**C+**	**C−**	PPV/NPV
P+	22	62	0.26
P−	41	298	0.88
TPR/NPR	0.35	0.83	**0.16**

Table 14. FoldX confusion matrices depending on gain, neutral or loss volume upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in volume, bot left: neutral mutations, bot right: mutations resulting in a loss in volume. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

PRODIGY-F	**C+**	**C−**	PPV/NPV
P+	153	390	0.28
P−	145	563	0.80
TPR/NPR	0.51	0.59	**0.09**

Gain	**C+**	**C−**	PPV/NPV
P+	103	180	0.35
P−	56	142	0.72
TPR/NPR	0.65	0.43	**0.07**

Neutral	**C+**	**C−**	PPV/NPV
P+	34	104	0.25
P−	42	158	0.79
TPR/NPR	0.45	0.60	**0.04**

Loss	**C+**	**C−**	PPV/NPV
P+	16	97	0.14
P−	47	263	0.85
TPR/NPR	0.25	0-73	**-0.01**

Table 15. PRODIGY confusion matrices on the FoldX mutation models depending on gain, neutral or loss volume upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in volume, bot left: neutral mutations, bot right: mutations resulting in a loss in volume. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

PRODIGY-E	**C+**	**C−**	PPV/NPV
P+	158	395	0.29
P−	140	558	0.80
TPR/NPR	0.53	0.59	**0.10**

Gain	**C+**	**C−**	PPV/NPV
P+	104	195	0.35
P−	55	136	0.71
TPR/NPR	0.65	0.41	**0.06**

Neutral	**C+**	**C−**	PPV/NPV
P+	37	101	0.27
P−	39	161	0.81
TPR/NPR	0.49	0.61	**0.09**

Loss	**C+**	**C−**	PPV/NPV
P+	17	99	0.15
P−	46	261	0.85
TPR/NPR	0.27	0.72	**0.00**

Table 16. PRODIGY confusion matrices on the EvoEF1 mutation models depending on gain, neutral or loss volume upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in volume, bot left: neutral mutations, bot right: mutations resulting in a loss in volume. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

BeAtMuSiC	**C+**	**C−**	PPV/NPV
P+	79	126	0.39
P−	219	827	0.79
TPR/NPR	0.27	0.87	**0.15**

Gain	**C+**	**C−**	PPV/NPV
P+	46	70	0.40
P−	113	261	0.70
TPR/NPR	0.29	0.79	**0.09**

Neutral	**C+**	**C−**	PPV/NPV
P+	24	36	0.40
P−	52	226	0.81
TPR/NPR	0.32	0.86	**0.19**

Loss	**C+**	**C−**	PPV/NPV
P+	9	20	0.31
P−	54	340	0.86
TPR/NPR	0.14	0.94	**0.12**

Table 17. BeAtMuSiC confusion matrices depending on gain, neutral or loss volume upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in volume, bot left: neutral mutations, bot right: mutations resulting in a loss in volume. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

EvoEF1	**C+**	**C−**	PPV/NPV
P+	200	375	0.35
P−	98	578	0.86
TPR/NPR	0.67	0.61	**0.24**

Gain	**C+**	**C−**	PPV/NPV
P+	127	203	0.38
P−	32	128	0.80
TPR/NPR	0.80	0.39	**0.19**

Neutral	**C+**	**C−**	PPV/NPV
P+	44	103	0.30
P−	32	159	0.83
TPR/NPR	0.58	0.61	**0.19**

Loss	**C+**	**C−**	PPV/NPV
P+	29	69	0.30
P−	34	261	0.90
TPR/NPR	0.46	0.81	**0.23**

Table 18. EvoEF1 confusion matrices depending on gain, neutral or loss volume upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in volume, bot left: neutral mutations, bot right: mutations resulting in a loss in volume. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

EvoEF2	**C+**	**C−**	PPV/NPV
P+	123	197	0.38
P−	175	756	0.81
TPR/NPR	0.41	0.79	**0.20**

Gain	**C+**	**C−**	PPV/NPV
P+	69	63	0.52
P−	90	268	0.75
TPR/NPR	0.43	0.81	**0.26**

Neutral	**C+**	**C−**	PPV/NPV
P+	34	73	0.32
P−	42	189	0.82
TPR/NPR	0.45	0.72	**0.15**

Loss	**C+**	**C−**	PPV/NPV
P+	20	61	0.25
P−	43	299	0.87
TPR/NPR	0.32	0.83	**0.13**

Table 19. EvoEF2 confusion matrices depending on gain, neutral or loss volume upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in volume, bot left: neutral mutations, bot right: mutations resulting in a loss in volume. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

As observed, some predictors show better performance for classifying mutations depending on changes in amino acid volume. For instance, pyDock-EvoEF1, UEP and BeAtMuSiC tend to accurately classify mutations not involving drastic changes in mutation size. However, FoldX and EvoEF1 show a different behaviour. FoldX tends to better classify mutations involving a gain in mutation size, while EvoEF1 seems to better classify mutations involving a loss in mutation size. This information may be valuable for a protein-protein design campaign, where predictions of those methods could be taken into account depending on these observations.

Regarding the hydrophobicity classification, a threshold of 0.3 Eisenberg units (Eu) was used to determine whether a mutation increases (ΔH>0.3Eu, Gain group, consisting of 604 mutations), decreases (ΔH<-0.3Eu, Loss group, consisting of 466 mutations) or shows similar hydrophobicity (|ΔH|≤0.3Eu, Neutral group, consisting of 181 mutations). Confusion matrices of each predictor regarding those volume subgroups are found at **Tables 20-28** for UEP, pyDock-FoldX, pyDock-EvoEF1, FoldX, PRODIGY-FoldX, PRODIGY-EvoEF1, BeAtMuSiC, EvoEF1 and EvoEF2, respectively.

UEP	**C+**	**C−**	PPV/NPV
P+	197	404	0.33
P−	101	549	0.84
TPR/NPR	0.66	0.58	**0.20**

Gain	**C+**	**C−**	PPV/NPV
P+	96	172	0.36
P−	64	272	0.81
TPR/NPR	0.60	0.61	**0.19**

Neutral	**C+**	**C−**	PPV/NPV
P+	23	48	0.32
P−	15	95	0.86
TPR/NPR	0.61	0.66	**0.22**

Loss	**C+**	**C−**	PPV/NPV
P+	78	184	0.30
P−	22	182	0.89
TPR/NPR	0.78	0.50	**0.23**

Table 20. UEP confusion matrices depending on gain, neutral or loss hydrophobicity upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in hydrophobicity, bot left: neutral mutations, bot right: mutations resulting in a loss in hydrophobicity. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

pyDock-F	**C+**	**C−**	PPV/NPV
P+	217	418	0.34
P−	81	535	0.87
TPR/NPR	0.73	0.56	**0.25**

Gain	**C+**	**C−**	PPV/NPV
P+	133	261	0.34
P−	27	183	0.87
TPR/NPR	0.83	0.41	**0.23**

Neutral	**C+**	**C−**	PPV/NPV
P+	29	58	0.33
P−	9	85	0.90
TPR/NPR	0.76	0.59	**0.29**

Loss	**C+**	**C−**	PPV/NPV
P+	55	99	0.36
P−	45	267	0.86
TPR/NPR	0.55	0.73	**0.24**

Table 21. pyDock confusion matrices on the models generated by FoldX depending on gain, neutral or loss hydrophobicity upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in hydrophobicity, bot left: neutral mutations, bot right: mutations resulting in a loss in hydrophobicity. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

pyDock-E	**C+**	**C−**	PPV/NPV
P+	206	371	0.36
P−	92	582	0.86
TPR/NPR	0.69	0.61	**0.26**

Gain	**C+**	**C−**	PPV/NPV
P+	127	254	0.33
P−	33	190	0.85
TPR/NPR	0.79	0.43	**0.20**

Neutral	**C+**	**C−**	PPV/NPV
P+	32	50	0.39
P−	6	93	0.94
TPR/NPR	0.84	0.65	**0.40**

Loss	**C+**	**C−**	PPV/NPV
P+	47	67	0.41
P−	53	299	0.85
TPR/NPR	0.47	0.82	**0.27**

Table 22. pyDock confusion matrices on the models generated by EvoEF1 depending on gain, neutral or loss hydrophobicity upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in hydrophobicity, bot left: neutral mutations, bot right: mutations resulting in a loss in hydrophobicity. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

FoldX	**C+**	**C−**	PPV/NPV
P+	162	277	0.37
P−	136	676	0.83
TPR/NPR	0.54	0.71	**0.23**

Gain	**C+**	**C−**	PPV/NPV
P+	102	170	0.38
P−	58	274	0.83
TPR/NPR	0.64	0.62	**0.23**

Neutral	**C+**	**C−**	PPV/NPV
P+	21	31	0.40
P−	17	112	0.87
TPR/NPR	0.55	0.78	**0.30**

Loss	**C+**	**C−**	PPV/NPV
P+	39	76	0.34
P−	61	290	0.83
TPR/NPR	0.39	0.79	**0.17**

Table 23. FoldX confusion matrices depending on gain, neutral or loss hydrophobicity upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in hydrophobicity, bot left: neutral mutations, bot right: mutations resulting in a loss in hydrophobicity. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

PRODIGY-F	**C+**	**C−**	PPV/NPV
P+	153	390	0.28
P−	145	563	0.80
TPR/NPR	0.51	0.59	**0.09**

Gain	**C+**	**C−**	PPV/NPV
P+	76	158	0.32
P−	84	286	0.77
TPR/NPR	0.47	0.64	**0.11**

Neutral	**C+**	**C−**	PPV/NPV
P+	26	54	0.33
P−	12	89	0.88
TPR/NPR	0.68	0.62	**0.25**

Loss	**C+**	**C−**	PPV/NPV
P+	51	178	0.22
P−	49	188	0.79
TPR/NPR	0.51	0.51	**0.02**

Table 24. PRODIGY confusion matrices on the models generated by FoldX depending on gain, neutral or loss hydrophobicity upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in hydrophobicity, bot left: neutral mutations, bot right: mutations resulting in a loss in hydrophobicity. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

PRODIGY-E	**C+**	**C−**	PPV/NPV
P+	158	395	0.29
P−	140	558	0.80
TPR/NPR	0.53	0.59	**0.10**

Gain	**C+**	**C−**	PPV/NPV
P+	77	156	0.33
P−	83	288	0.78
TPR/NPR	0.48	0.65	**0.12**

Neutral	**C+**	**C−**	PPV/NPV
P+	24	0.58	0.29
P−	14	85	0.86
TPR/NPR	0.63	0.59	**0.18**

Loss	**C+**	**C−**	PPV/NPV
P+	57	181	0.24
P−	43	185	0.81
TPR/NPR	0.57	0.51	**0.06**

Table 25. PRODIGY confusion matrices on the models generated by EvoEF1 depending on gain, neutral or loss hydrophobicity upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in hydrophobicity, bot left: neutral mutations, bot right: mutations resulting in a loss in hydrophobicity. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

BeAtMuSiC	**C+**	**C−**	PPV/NPV
P+	79	126	0.39
P−	219	827	0.79
TPR/NPR	0.27	0.87	**0.15**

Gain	**C+**	**C−**	PPV/NPV
P+	61	101	0.38
P−	99	343	0.78
TPR/NPR	0.38	0.77	**0.15**

Neutral	**C+**	**C−**	PPV/NPV
P+	7	11	0.39
P−	31	132	0.81
TPR/NPR	0.18	0.92	**0.15**

Loss	**C+**	**C−**	PPV/NPV
P+	11	14	0.44
P−	89	352	0.80
TPR/NPR	0.11	0.96	**0.13**

Table 26. BeAtMuSiC confusion matrices depending on gain, neutral or loss hydrophobicity upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in hydrophobicity, bot left: neutral mutations, bot right: mutations resulting in a loss in hydrophobicity. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

EvoEF1	**C+**	**C−**	PPV/NPV
P+	200	375	0.35
P−	98	578	0.86
TPR/NPR	0.67	0.61	**0.24**

Gain	**C+**	**C−**	PPV/NPV
P+	110	209	0.34
P−	50	235	0.82
TPR/NPR	0.69	0.53	**0.19**

Neutral	**C+**	**C−**	PPV/NPV
P+	27	46	0.37
P−	11	97	0.90
TPR/NPR	0.71	0.68	**0.32**

Loss	**C+**	**C−**	PPV/NPV
P+	63	120	0.34
P−	37	246	0.87
TPR/NPR	0.63	0.67	**0.25**

Table 27. EvoEF1 confusion matrices depending on gain, neutral or loss hydrophobicity upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in hydrophobicity, bot left: neutral mutations, bot right: mutations resulting in a loss in hydrophobicity. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

EvoEF2	**C+**	**C−**	PPV/NPV
P+	123	197	0.38
P−	175	756	0.81
TPR/NPR	0.41	0.79	**0.20**

Gain	**C+**	**C−**	PPV/NPV
P+	71	122	0.37
P−	89	322	0.78
TPR/NPR	0.44	0.73	**0.16**

Neutral	**C+**	**C−**	PPV/NPV
P+	12	24	0.33
P−	26	119	0.82
TPR/NPR	0.32	0.83	**0.15**

Loss	**C+**	**C−**	PPV/NPV
P+	40	51	0.44
P−	60	315	0.84
TPR/NPR	0.40	0.86	**0.27**

Table 28. EvoEF2 confusion matrices depending on gain, neutral or loss hydrophobicity upon mutation. Top left: entire benchmark, top right: mutations inducing a gain in hydrophobicity, bot left: neutral mutations, bot right: mutations resulting in a loss in hydrophobicity. Statistical descriptors, such as PPV, NPV, TPR, NPR and MCC are shown (MCC in bold).

Similarly to the results observed for changes in amino acid size, some predictors show some particularities for predicting mutations depending on changes in hydrophobicity. In this case, most of the predictors show increased accuracy for predicting mutations with neutral changes in hydrophobicity. Some exceptions have been found, such as EvoEF2, which seems to accurately predict mutations resulting in hydrophobicity-loss, or BeAtMuSiC which shows a similar (but low) classification performance in all groups. MCC values of all predictors grouped by changes in amino acid size and volume are illustrated in **Figure 29**.

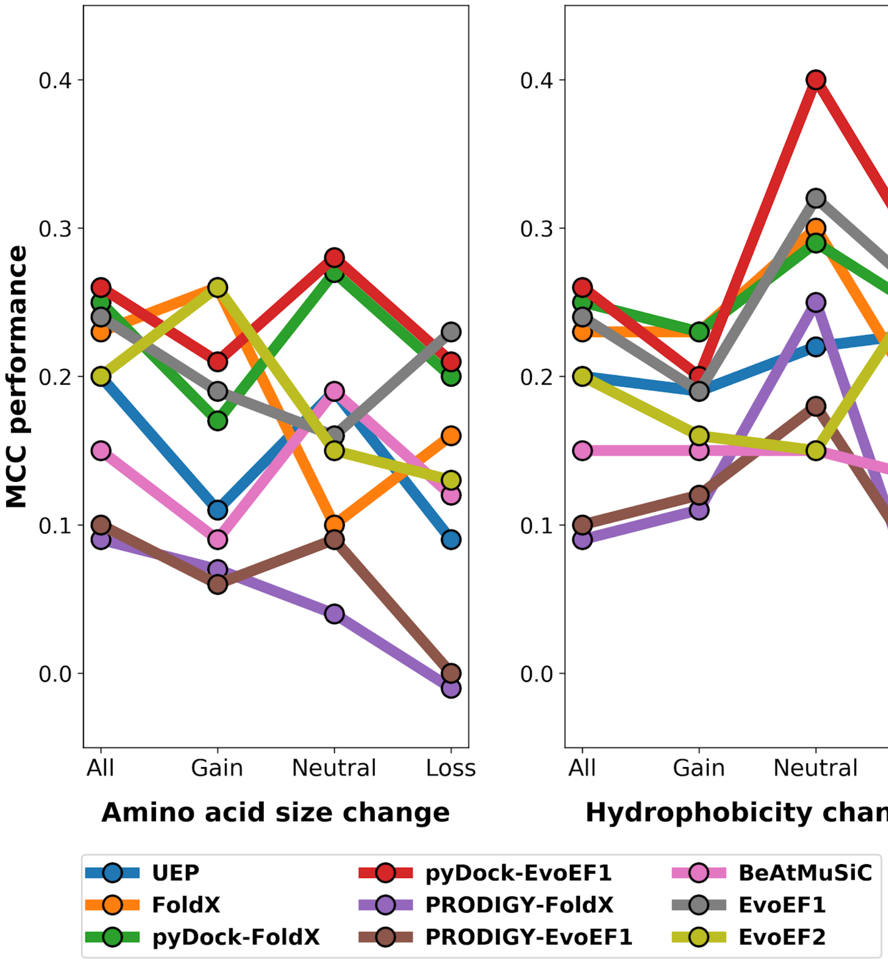


Figure 29. Performance of the binding affinity predictors upon mutations depending on changes in amino acid size change (left panel) and hydrophobicity (right panel). Subgroups represent MCC score for the entire group of mutations other than alanine embedded within the highly-packed interface (All), those who exert a gain in size/hydrophobicity (Gain), neutral changes (Neutral) and a loss (Loss).

Next, we aimed to calculate the similarity between predictors on the estimation of the effects of mutations. Thus, we investigated the similarity between pairs of predictors for classifying mutations into improving or decreasing the binding affinity. In this sense, we assessed the similarity percentage between predictors for all 1251 mutations other than alanine. A comparison of the similarity percentages of all pairs of predictors is illustrated in **Table 29**.

Predictors	Percentage	Predictors	Percentage
UEP/pyDock-F	59.1 (739)	pyDock-E/PRODIGY-F	56.7 (709)
UEP/pyDock-E	56.5 (707)	pyDock-E/PRODIGY-E	57.0 (713)
UEP/FoldX	64.0 (801)	pyDock-E/BeAtMuSiC	61.6 (771)
UEP/EvoEF1	65.3 (817)	pyDock-E/PRODIGY	56.7 (709)
UEP/EvoEF2	55.6 (696)	FoldX/EvoEF1	68.5 (857)
UEP/PRODIGY-F	57.6 (721)	FoldX/EvoEF2	71.3 (892)
UEP/PRODIGY-E	57.6 (721)	FoldX/PRODIGY-F	56.4 (705)
UEP/BeAtMuSiC	56.7 (709)	FoldX/PRODIGY-E	56.2 (703)
pyDock-F/pyDock-E	80.2 (1003)	FoldX/BeAtMuSiC	64.5 (807)
pyDock-F/FoldX	60.5 (757)	EvoEF1/EvoEF2	64.9 (812)
pyDock-F/EvoEF1	63.5 (795)	EvoEF1/PRODIGY-F	57.6 (721)
pyDock-F/EvoEF2	53.9 (674)	EvoEF1/PRODIGY-E	57.5 (719)
pyDock-F/PRODIGY-F	54.9 (687)	EvoEF1/BeAtMuSiC	59.2 (741)
pyDock-F/PRODIGY-E	55.1 (689)	EvoEF2/PRODIGY-F	56.0 (700)
pyDock-F-BeAtMuSiC	56.7 (709)	EvoEF2/PRODIGY-E	55.5 (694)
pyDock-E/FoldX	63.7 (797)	EvoEF2/BeAtMuSiC	69.4 (868)
pyDock-E/EvoEF1	63.5 (795)	PRODIGY-F/BeAtMuSiC	58.1 (727)
pyDock-E/EvoEF2	61.6 (770)	PRODIGY-E/BeAtMuSiC	57.0 (713)

Table 29. Percentage of similarity between all possible pairs of predictors. Percentage represents the number of mutations of the entire benchmark (1251 mutations other than alanine placed in the highly-packed region) that are predicted in the same way (this is, improving or decreasing the binding affinity of the complex). Numbers inside parentheses indicate the amount of mutations corresponding to such percentage.

These results indicate that the pairwise similarity of all pairs of predictors is in the range of 53.9-80.2%. The most dissimilar pair of predictors are pyDock (on the models generated by FoldX) and EvoEF2, and the most similar pair is constituted by pyDock predictions on the models generated by FoldX and EvoEF1. Overall, the agreement between predictors is quite low, and it seems that the modeling process plays a very important role in the final ΔΔG prediction. This is represented by the fact that pyDock predictions on the models generated by FoldX and EvoEF1 show a low similarity rate of 80.2%. This effect highlights the importance of the modeling process, and the need of standard protocols for generating mutations and the need of using the same force field from the generation of the mutation to the energy prediction.

Regarding the correlation ability with the experimental binding affinity values, **Figure 30 and 31** indicate the PCC and RMSE of all predictors for mutations to alanine and other than alanine, respectively. As can be observed, PCC values are low for all predictors: from 0.16 to 0.35 for mutations to alanine and from 0.06 to 0.43 for mutations other than alanine. UEP reached PCC values of 0.16 and 0.22 for mutations to alanine and other than alanine, respectively. Despite that those PCC values are low, we actually believe that observing correlations similar to the state-of-the art without using any experimental value of binding affinities is remarkable. This is even more relevant considering the fact that all other algorithms have been partially trained on the data we are using for the benchmarking process. In other words, it is very likely that those methods will work better for predicting pre-calibrated mutations on their scoring function than when predicting new mutations. This phenomenon is common during the design of any kind of predictive algorithm; however, large overfitting effects towards the training data must be avoided. This is the case of the mCSM algorithm, as we introduced previously. As can be observed, PCC values for mutations to alanine are 0.72 (trained) and 0.33 (new mutations), while PCC values for mutations to other than alanine are 0.88 (trained) and -0.34 (new mutations). We also benchmarked the same set of new mutations using UEP, and we did not observe a large drop in performance as mCSM, which corroborated that this group of mutations does not show peculiarities difficulting their prediction.

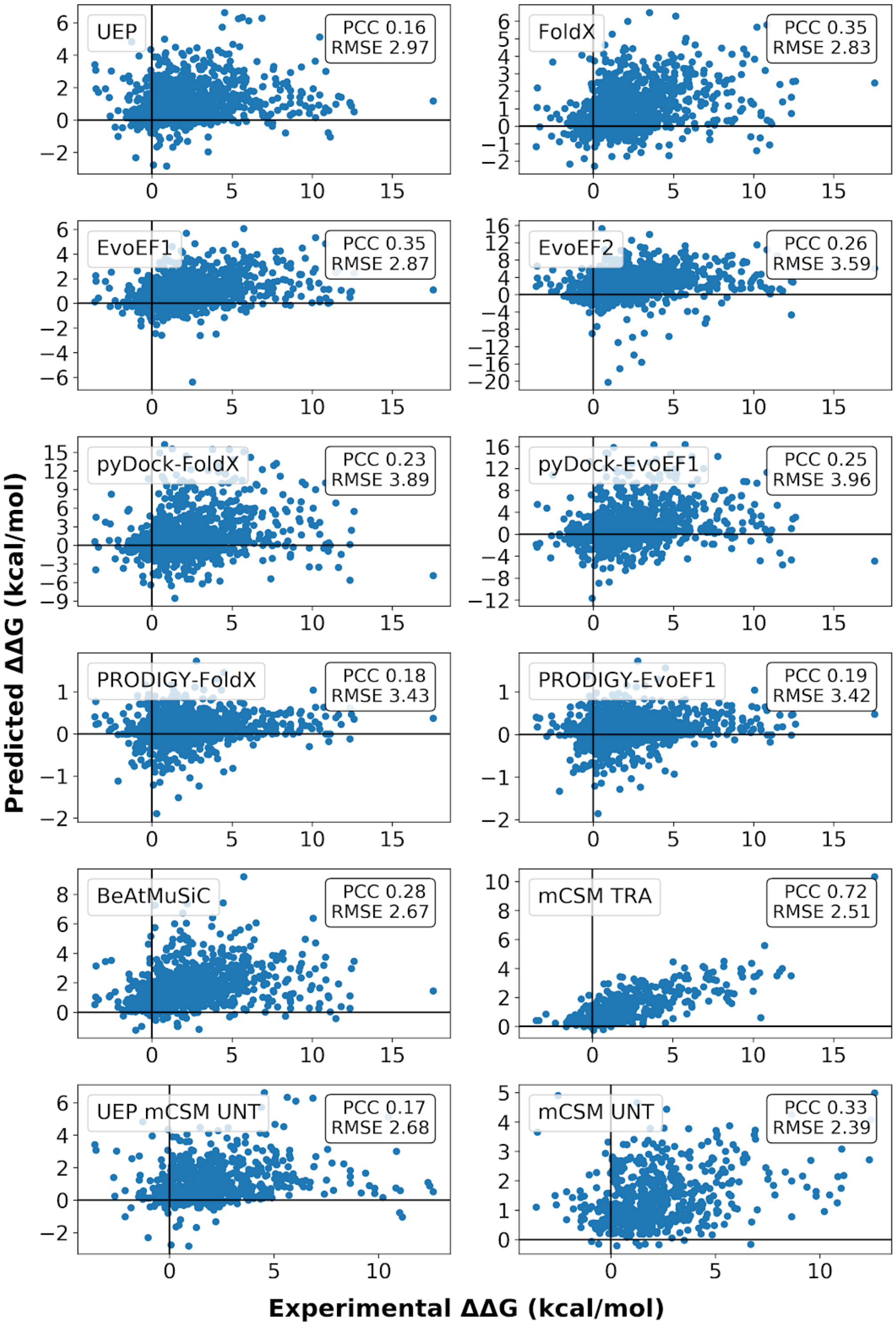


Figure 30. Performance of all tested protein-protein affinity predictors upon mutation on the 985 selected mutations to alanine of the SKEMPI 2.0. Predicted ΔΔG (kcal/mol) and Experimental ΔΔG (kcal/mol) are depicted in the vertical and horizontal axis of each subplot. PCC and RMSE statistics are shown for each predictor in the top right section of each graphic.

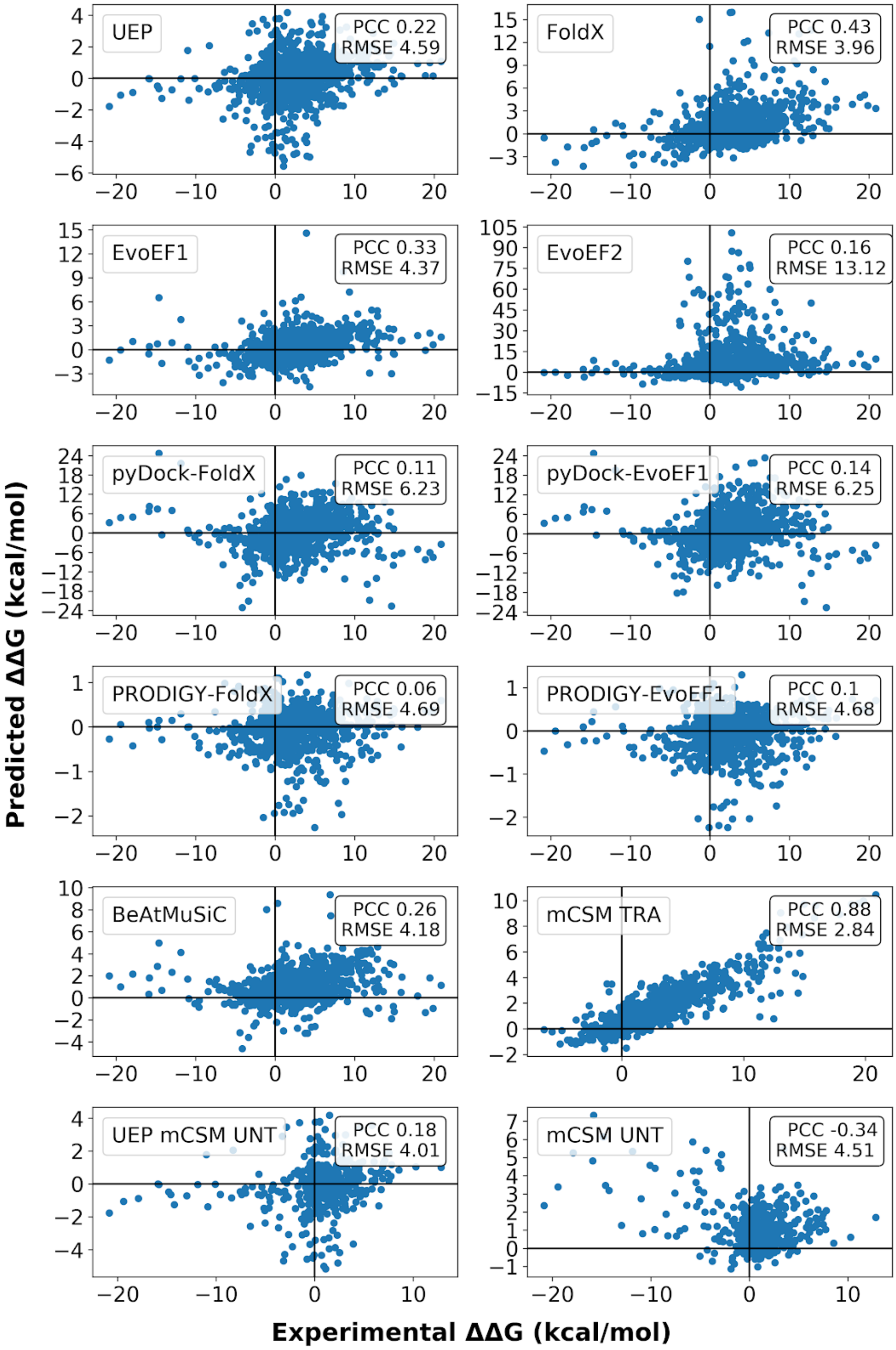


Figure 31. Performance of all tested protein-protein affinity predictors upon mutation on the 1251 selected mutations to other than alanine of the SKEMPI 2.0. Predicted ΔΔG (kcal/mol) and Experimental ΔΔG (kcal/mol) are depicted in the vertical and horizontal axis of each subplot. PCC and RMSE statistics are shown for each predictor in the top right section of each graphic.

Overall, several conclusions can be extracted from our experience in the field of ΔΔG predictions. First of all, we would like to highlight the challenge of making this kind of prediction taking into account the current limitations. As reported previously, the gain in binding affinity for most mutations is low (<1 and ~2 kcal/mol for alanine and other than alanine mutations, respectively). Here, understanding the limits of detection of computational techniques is a must. In this sense, it has been reported that the most accurate techniques, such as FEP protocols used in drug design efforts, work under an estimated error of 1kcal/mol (for “easier” benchmarks). In this sense, simpler techniques such as ΔΔG predictions are expected to work under larger errors than just 1kcal/mol. Therefore, it is very likely that the errors associated with the ΔΔG predictions evaluated in this thesis are higher than the observed experimental ΔΔG.

Another interesting conclusion belongs to the dependency of high-quality data for the development of useful predictors. Currently, available data is highly heterogeneous and is not sufficient for the design of improved algorithms. This is especially relevant for the study of mutations to other residues than alanine, which is underestimated on public databases such as SKEMPI 2.0. As reported along the introduction, all predictors follow different approaches to make ΔΔG predictions and in the end they show a similar performance for classification and correlation studies. Therefore, divergence in experimental determinations could play a major role for the observed low performance of all methods.

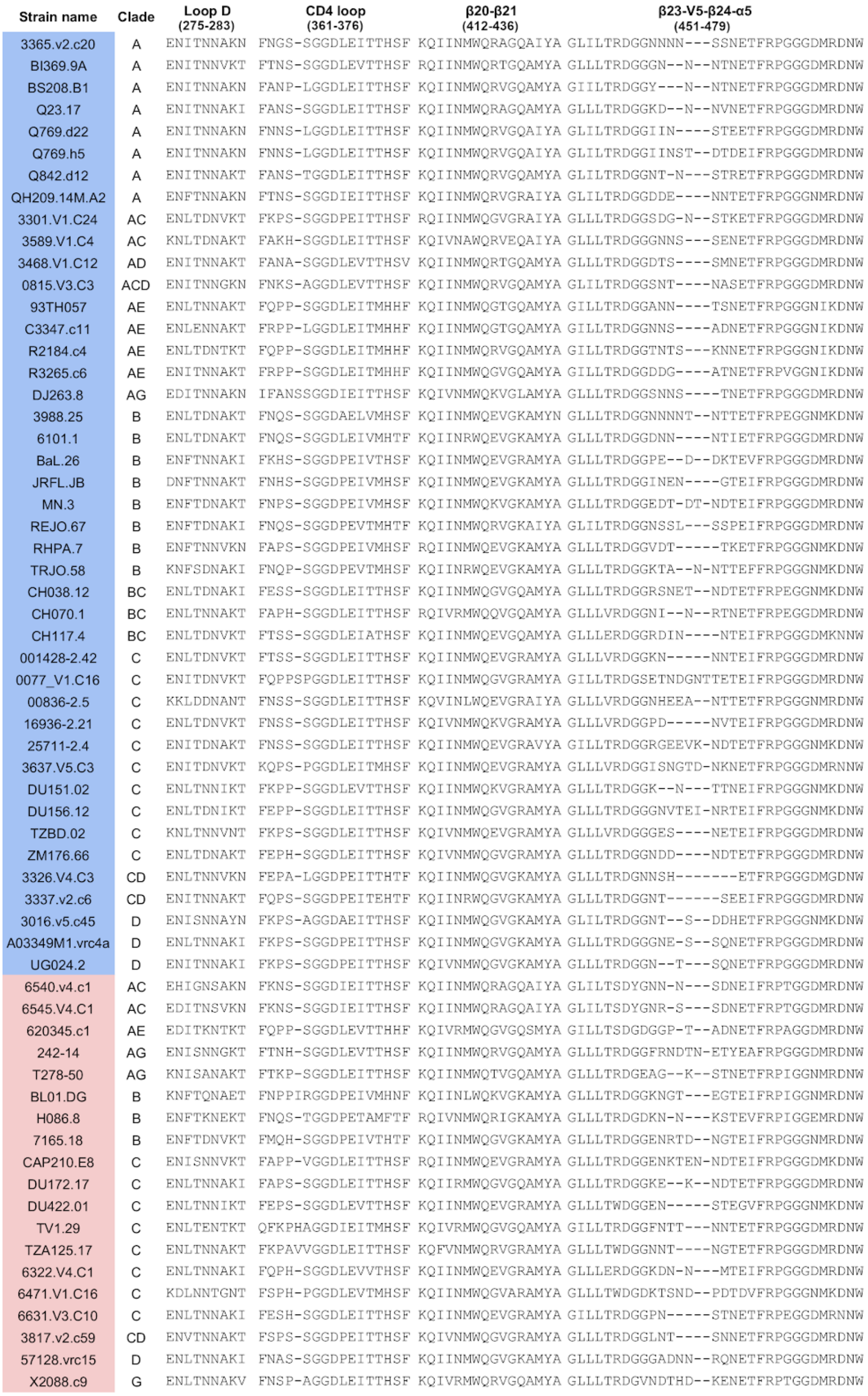
Predicting antibody binding efficacy: PELE-antibody

To date, PELE has been mostly restricted to small molecule recognition in the fields of enzyme design and drug discovery efforts. Additionally, there is no computational standard protocol to predict antibody binding efficacy towards an hypermutated epitope, which is a relevant topic for research against AIDS. In this section, we describe the potential use of PELE for predicting antibody binding affinity towards hypermutated HIV-1 strains. Previously, PELE was used for predicting drug efficacy on the HIV-1 protease using the patient's virus sequence (Hosseini *et al.*, 2016). In this thesis we significantly challenge such a modelling effort. Our system of study was based on anti-CD4bs bNAbs binding against multiple gp120 of HIV-1. Here, we benchmarked three different bNAbs: VRC01, NIH45-46 and 3BNC117 against multiple viral HIV-1 isolates by performing PELE simulations of the three-dimensional binding process. This work has been recently published (Amengual-Rigo, Carrillo, *et al.*, 2020).

Data collection and system preparation

Binding affinity of an antibody towards an epitope can be experimentally determined in multiple ways. The data that we have collected in this section was produced by an ELISA assay, where the binding is measured between captured antibodies and monomeric gp120 dissociated from pseudovirus proteins. Essentially, resistant gp120 strains are out of the detection limits of the method and have an IC50 determination >50 μg/mL.

Our case studies consisted of performing a large benchmark for VRC01, and a smaller benchmark for VRC01, NIH45-46 and 3BNC117 bNAbs. Here, a set of 45 sensible and 19 resistant gp120 strains were collected from the literature against VRC01 (**Figure 32**). From the best of our knowledge, all known resistant strains with available sequence in GenBank were collected. IC50 determinations of sensitive strains are illustrated in **Figure 33**. As can be observed, most sensitive gp120 strains have a binding value of < 0.2 μg/m, which makes challenging a correlation assay. Hence, we will focus on the binary classification point of view: if an antibody binds and neutralizes sensitive strains or not (resistant strains).



Strain name	Clade	Loop D (275-283)	CD4 loop (361-376)	β20-β21 (412-436)	β23-V5-β24-α5 (451-479)
3365.v2.c20	A	ENITNNAKN	FNGS-SGGDLEITTHSF	KQIINMWQRAGQAIYA	GLILTRDGGNNNN---SSNETFRPGGGDMRDNW
BI369.9A	A	ENITNNVKT	FTNS-SGGDLEVTTHSF	RQIINMWQRTGQAMYA	GLLLTRDGGGN--N--NTNETFRPGGGDMRDNW
BS208.B1	A	ENITNNAKN	FANP-LGGDLEITTHSF	KQIINMWQRVGQAMYA	GIILTRDGGY---N--NTNETFRPGGGDMRDNW
Q23.17	A	ENITNNAKI	FANS-SGGDLEITTHSF	KQIINMWQRAGQAMYA	GLLLTRDGGKD--N--NVNETFRPGGGDMRDNW
Q769.d22	A	ENITNNAKN	FNNS-LGGDLEITTHSF	KQIINMWQRVGQAIYA	GLILTRDGGIIN----STEETFRPGGGDMRDNW
Q769.h5	A	ENITNNAKN	FNNS-LGGDLEITTHSF	KQIINMWQRVGQAIYA	GLILTRDGGIINST--DTDEIFRPGGGDMRDNW
Q842.d12	A	ENITNNAKT	FANS-TGGDLEITTHSF	KQIINMWQRVGQAMYA	GLILTRDGGNT-N---STRETFRPGGGDMRDNW
QH209.14M.A2	A	ENFTNNAKN	FTNS-SGGDIEITTHSF	KQIINMWQRVGRAIYA	GLILTRDGGDDE----NNTETFRPGGGNMRDNW
3301.V1.C24	AC	ENLTDNVKT	FKPS-SGGDPEITTHSF	RQIINMWQGVGRAIYA	GLLLTRDGGSDG-N--STKETFRPGGGDMRDNW
3589.V1.C4	AC	KNLTDNAKT	FAKH-SGGDLEITTHSF	KQIVNAWQRVEQAIYA	GLLLTRDGGGNNS---SENETFRPGGGDMRDNW
3468.V1.C12	AD	ENITNNAKT	FANA-SGGDLEVTTHSV	KQIINMWQRTGQAMYA	GLLLTRDGGDTS----SMNETFRPGGGDMRDNW
0815.V3.C3	ACD	ENITNNGKN	FNKS-AGGDLEVTTHSF	KQIINMWQRVGQAMYA	GLILTRDGGSNT----NASETFRPGGGDMRDNW
93TH057	AE	ENLTNNAKT	FQPP-SGGDLEITMHHF	KQIINMWQGTGQAMYA	GILLTRDGGANN----TSNETFRPGGGNIKDNW
C3347.c11	AE	ENLENNAKT	FRPP-LGGDLEITMHHF	KQIINMWQGTGQAMYA	GILLTRDGGNNS----ADNETFRPGGGNIKDNW
R2184.c4	AE	ENLTDNTKT	FQPP-SGGDLEITMHHF	KQIINMWQRVGQAMYA	GILLTRDGGTNTS---KNNETFRPGGGNIKDNW
R3265.c6	AE	ENITNNAKT	FRPP-SGGDLEITMHHF	KQIINMWQGVGQAMYA	GILLTRDGGDDG----ATNETFRPVGGNIKDNW
DJ263.8	AG	EDITNNAKN	IFANSSGGDIEITTHSF	KQIVNMWQKVGLAMYA	GLLLTRDGGSNNS----TNETFRPGGGDMRDNW
3988.25	B	ENLTDNAKT	FNQS-SGGDAELVMHSF	KQIINMWQEVGKAMYN	GLLLTRDGGNNNNT--NTTETFRPEGGNMKDNW
6101.1	B	ENLTDNAKT	FNQS-SGGDLEIVMHTF	KQIINRWQEVGKAMYA	GLLLTRDGGDNN----NTIETFRPGGGDMRDNW
BaL.26	B	ENFTNNAKI	FKHS-SGGDPEIVTHSF	KQIINMWQEVGRAMYA	GLLLTRDGGPE--D--DKTEVFRPGGGDMRDNW
JRFL.JB	B	DNFTNNAKT	FNHS-SGGDPEIVMHSF	KQIINMWQEVGKAMYA	GLLLTRDGGINEN----GTEIFRPGGGDMRDNW
MN.3	B	ENFTDNAKT	FNPS-SGGDPEIVMHSF	KQIINMWQKVGKAMYA	GLLLTRDGGEDT-DT-NDTEIFRPGGGDMRDNW
REJO.67	B	ENFTDNAKI	FNQS-SGGDPEVTMHTF	KQIINMWQRVGKAIYA	GLILTRDGGNSSL---SSPEIFRPGGGDMRDNW
RHPA.7	B	ENFTNNVKN	FAPS-SGGDPEIVMHSF	RQIINMWQEVGKAMYA	GLLLTRDGGVDT-----TKETFRPGGGNMKDNW
TRJO.58	B	KNFSDNAKI	FNQP-SGGDPEVTMHSF	KQIINRWQEVGKAMYA	GLLLTRDGGKTA--N-NTTEFFRPGGGNMKDNW
CH038.12	BC	ENLTDNAKI	FESS-SGGDLEITTHSF	KQIINMWQGVGQAMYA	GLLLTRDGGRSNET--NDTETFRPEGGNMKDNW
CH070.1	BC	ENLTNNAKT	FAPH-SGGDLEITTHSF	RQIVRMWQQVGQAMYA	GLLLVRDGGNI--N--RTNETFRPEGGDMRDNW
CH117.4	BC	ENLTDNVKT	FTSS-SGGDLEIATHSF	KQIINMWQEVGQAMYA	GLLLERDGGRDIN----NTEIFRPGGGDMKNNW
001428-2.42	C	ENLTDNVKT	FTSS-SGGDLEITTHSF	KQIINMWQEVGRAMYA	GLLLVRDGGKN-----NNTEIFRPGGGDMRDNW
0077_V1.C16	C	ENITDNVKT	FQPPSPGGDLEITTHSF	KQIINMWQGVGRAMYA	GILLTRDGSETNDGNTTETEIFRPGGGDMRDNW
00836-2.5	C	KKLDDNANT	FNSS-SGGDLEITTHSF	KQVINLWQEVGRAIYA	GLLLVRDGGNHEEA--NTTETFRPGGGNMRDNW
16936-2.21	C	ENLTDNVKT	FNSS-SGGDLEITTHSF	KQIVNMWQKVGRAMYA	GLLLVRDGGPD-----NVTEIFRPGGGDMRDNW
25711-2.4	C	ENITDNAKT	FNSS-SGGDLEITTHSF	KQIINMWQEVGRAVYA	GILLTRDGGRGEEVK-NDTETFRPGGGNMKDNW
3637.V5.C3	C	ENITDNVKT	KQPS-PGGDLEITMHSF	KQIINMWQEVGRAMYA	GLLLVRDGGISNGTD-NKNETFRPGGGDMRNNW
DU151.02	C	ENLTNNIKT	FKPP-SGGDLEVTTHSF	KQIINMWQKVGRAMYA	GLLLTRDGGK--N---TTNEIFRPGGGNMKDNW
DU156.12	C	ENLTDNIKT	FEPP-SGGDLEITTHSF	KQIINMWQGVGRAMYA	GLLLTRDGGGNVTEI-NRTEIFRPGGGNMKDNW
TZBD.02	C	KNLTNNVNT	FKPS-SGGDLEITTHSF	KQIVNMWQEVGRAMYA	GLLLVRDGGGES----NETEIFRPGGGDMRDNW
ZM176.66	C	ENLTDNAKT	FEPH-SGGDLEITTHSF	KQIVNMWQGVGRAMYA	GLLLTRDGGNDD----NDTETFRPGGGDMRDNW
3326.V4.C3	CD	ENLTNNVKN	FEPA-LGGDPEITTHTF	KQIINMWQGVGKAMYA	GLLLTRDGNNSH-------ETFRPGGGDMGDNW
3337.v2.c6	CD	ENITNNAKT	FQPS-SGGDPEITEHTF	KQIINRWQGVGKAMYA	GLLLTRDGGNT------SEEIFRPGGGDMRDNW
3016.v5.c45	D	ENISNNAYN	FKPS-AGGDAEITTHSF	KQIINMWQGVGKAMYA	GLILTRDGGNT--S--DDHETFRPGGGNMKDNW
A03349M1.vrc4a	D	ENLTNNAKI	FKPS-SGGDPEITTHSF	KQIINMWQGVGKAMYA	GLLLTRDGGGNE-S--SQNETFRPGGGDMRDNW
UG024.2	D	ENITNNAKI	FKPS-SGGDPEITTHSF	KQIVNMWQGVGKAMYA	GLLLTRDGGN--T---SQNETFRPGGGDMRDNW
6540.v4.c1	AC	EHIGNSAKN	FKNS-SGGDIEITTHSF	KQIINMWQRAGQAIYA	GLILTSDYGNN-N---SDNEIFRPTGGDMRDNW
6545.V4.C1	AC	EDITNSVKN	FKNS-SGGDIEITTHSF	KQIINMWQRAGQAIYA	GLILTSDYGNR-S---SDNETFRPTGGDMRDNW
620345.c1	AE	EDITKNTKT	FQPP-SGGDLEVTTHHF	KQIVRMWQGVGQSMYA	GILLTSDGDGGP-T--ADNETFRPAGGDMRDNW
242-14	AG	ENISNNGKT	FTNH-SGGDLEVTTHSF	KQIINMWQRVGQAMYA	GLLLTRDGGFRNDTN-ETYEAFRPGGGDMRDNW
T278-50	AG	KNISANAKT	FTKP-SGGDLEITTHSF	KQIINMWQTVGQAMYA	GLLLTRDGEAG--K--STNETFRPIGGNMRDNW
BL01.DG	B	KNFTQNAET	FNPPIRGGDPEIVMHNF	KQIINLWQKVGKAMYA	GLLLTRDGGKNGT---EGTEIFRPIGGNMRDNW
H086.8	B	ENFTKNEKT	FNQS-TGGDPETAMFTF	RQIVNMWQRIGKAMYA	GLLLTRDGDKN-N---KSTEVFRPIGGEMRDNW
7165.18	B	ENFTDNVKT	FMQH-SGGDPEIVTHTF	KQIINMWQGVGKAMYA	GLLLTRDGGENRTD--NGTEIFRPGGGNMRDNW
CAP210.E8	C	ENISNNVKT	FAPP-VGGDLEITTHSF	RQIINMWQEVGRAMYA	GLLLTRDGGENKTEN-NDTEIFRPGGGDMKDNW
DU172.17	C	ENLTNNAKI	FAPS-SGGDLEITTHSF	KQIIRMWQGVGQAMYA	GLLLTRDGGKE--K--NDTETFRPGGGDMRDNW
DU422.01	C	ENLTNNIKT	FEPS-SGGDLEVTTHSF	KQIINMWQEVGRAMYA	GLLLTWDGGEN-----STEGVFRPGGGNMKDNW
TV1.29	C	ENLTENTKT	QFKPHAGGDIEITMHSF	KQIVRMWQGVGQAMYA	GILLTRDGGFNTT---NNTETFRPGGGDMRDNW
TZA125.17	C	ENLTNNAKT	FKPAVVGGDLEITTHSF	KQFVNMWQRVGRAMYA	GLLLTWDGGNNT----NGTETFRPGGGDMRDNW
6322.V4.C1	C	ENLTNNAKI	FQPH-SGGDLEVVTHSF	KQIINMWQEVGRAMYA	GLLLERDGGKDN-N---MTEIFRPGGGDMRDNW
6471.V1.C16	C	KDLNNTGNT	FSPH-PGGDLEVTMHSF	KQIINMWQGVARAMYA	GLLLTWDGDKTSND--PDTDVFRPGGGNMKDNW
6631.V3.C10	C	ENLTNNAKI	FESH-SGGDLEITTHSF	KQIINMWQEVGRAMYA	GILLTRDGGPN-----STNETFRPEGGDMRNNW
3817.v2.c59	CD	ENVTNNAKT	FSPS-SGGDPEITTHSF	KQIVNMWQGVGRAMYA	GLLLTRDGGLNT----SNNETFRPGGGDMRDNW
57128.vrc15	D	ENLTNNAKI	FNAS-SGGDPEITTHSF	KQIINMWQGVGKAMYA	GLLLTRDGGGADNN--RQNETFRPGGGDMRDNW
X2088.c9	G	ENLTNNAKV	FNSP-AGGDLEITTHSF	KQIVRMWQRVGQAMYA	GLLLTRDGVNDTHD--KENETFRPTGGDMRDNW

Figure 32. List of evaluated strains for the large VRC01 benchmark, consisting of a set of 45 sensitive (blue) and 19 resistant (red) gp120 strains. Clade identity, and the linear sequence in contact with the VRC01 antibody from the loop D, the CD4 loop, β20-β21 and β23-V5-β24-α5 are shown to illustrate the diversity of the HIV-1.

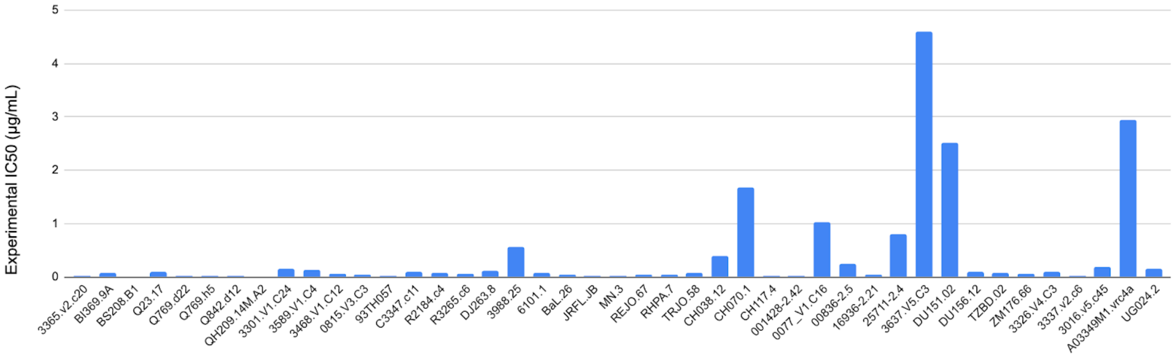


Figure 33. Experimental IC50 determinations (µg/mL) of the VRC01 sensitive strains evaluated in this work. As can be observed, most of all binding determinations have values <0.2 µg/mL.

Some HIV-1 clades (or subgroups) show a series of sequence and structural peculiarities that have been linked to certain resistance patterns. Because of this reason, the selection of sensitive strains (from 172 available sequences in GenBank) was (*per se*) not random: we aimed to select sensitive gp120 strains coming from similar clades than the resistant ones. By following this approach, we minimized the risk of observing biased results towards clade representation between sensitive and resistant groups. Hence, sensitive strains used in this work belong to the following clades: A, AC, AD, ACD, AE, AG, B, BC, C, CD and D, while resistant ones belong to: AC, AE, AG, B, C, CD, D and G.

One of the main limitations of this project is the lack of experimentally solved gp120 three-dimensional structures. In this sense, most of the gp120 strains have no available crystal structure, and this is even more accentuated in the case of resistant strains. This phenomenon may constitute a bottleneck for understanding and designing antibody variants and vaccines against AIDS. To overcome this limitation, we modelled all gp120 strains by homology modeling techniques, aiming to mimic the unknown-real conformation of the envelope glycoproteins. Prime from Schödinger (Jacobson *et al.*, 2004; Jacobson, Friesner, *et al.*, 2002) was used to generate the homology models. We used the gp120 protein co-crystallized with VRC01 found on the 3NGB PDB accession as template structure for the homology models. This gp120 is known as 93TH057, and it is widely used for unveiling the binding conformation of antibodies and therefore, is overrepresented on the PDB database.

After the generation of the homology models, we proceeded to glycosylate all viral structures. It has been reported that glycosylation levels shape the antibody binding effectivity. After analyzing all available envelope glycoproteins with three-dimensional structures on the PDB, we found out that the most common glycosylation consists of a covalent modification of asparagine amino acid with a molecule of N-acetyl glucosamine (NAG). This kind of glycosylation follows a linear motif of N-X-T/S, where N is the asparagine residue to be glycosylated, X is any amino acid, and T/S represents threonine and serine, respectively. Moreover, such asparagine residue must be exposed on the surface of the protein to be glycosylated. Hence, all exposed asparagines found in a N-X-T/S motif were glycosylated with a NAG molecule.

After generating the fully glycosylated gp120, we proceeded to generate the initial antibody-epitope conformation for initializing PELE simulations. Here, we aimed to simulate the three-dimensional binding process between the antibody (VRC01, NIH45-46 and 3BNC117) and the selected gp120s. Therefore, we placed the gp120 strains 10 Å away (in the vertical axis) from the docked conformation with the bNAbs, aiming to provide enough space to simulate the binding process (from out to in conformations).

PELE-antibody algorithm

Simulating the binding process of two proteins is not substantially different to simulating the binding process of a small molecule towards a target protein. Usually, the larger body remains fixed (in this case, the antibody) and the smaller one (the gp120) moves around the other. In fact, this is also the basis of most protein-protein docking algorithms. However, and especially for the analysis of the binding patterns in PELE simulations, it is important to start with an undocked conformation. This is because the acceptance of the PELE steps strongly relies on the contacts generated between the stimulating proteins. Hence, usually once that the interacting partners bind, PELE steps detaching the complex are not likely to be accepted.

In the "***State of the art: Monte Carlo***" section, the basics of the PELE algorithm is provided. The protocol used in this work is similar to the one described previously, as can be observed by comparing **Figure 24 and 34**.

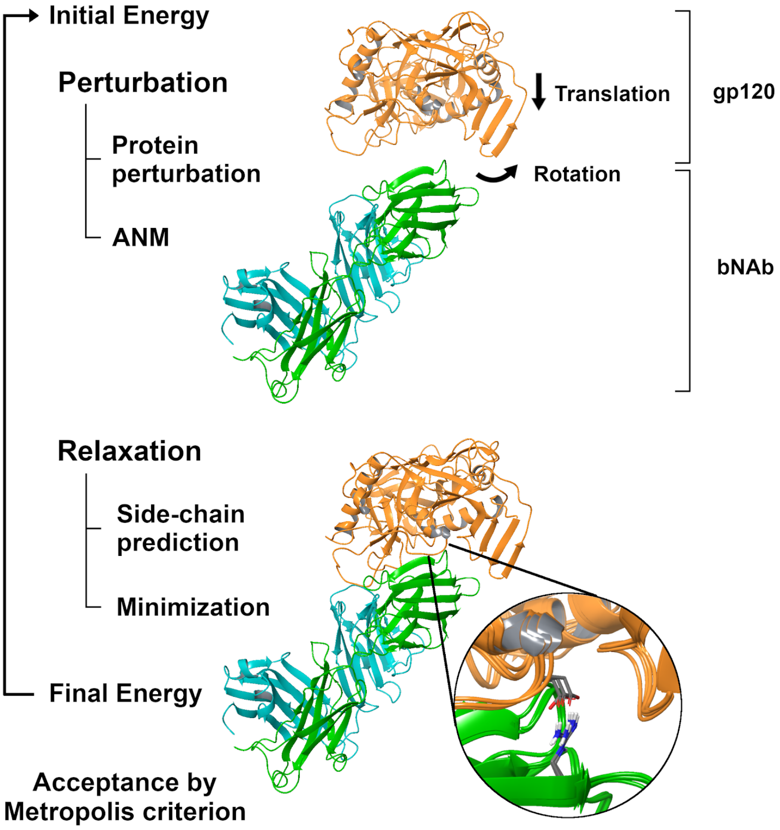


Figure 34. Graphical representation of a PELE step in protein-protein complexes. PELE consists of two phases: perturbation and relaxation. During the former, a protein is translated and rotated, and a backbone sampling protocol (ANM) is performed. During the latter, side chain conformations are sampled followed by a minimization step to generate energetically stable conformations. The movement may be accepted according to the Metropolis criterion.

However, some modifications were made to optimize the efficiency of the simulations, since that protein-protein simulations are computationally more costly than other smaller systems (such as protein-ligand). Hence, the acceptance of a PELE step takes longer for protein-protein systems than protein-ligand because of some particularities originated by the size of the system: i) clashes are common (a small rotation of one protein may originate several clashes with the partner protein), ii) side chain prediction takes longer because the interface involves larger amount of residues, and iii) the minimization process of the system involves more atoms and therefore, it takes longer. Aiming to optimize the simulation parameters, we: i) reduced the magnitude of translations and rotations (contributing to reduce the amount of clashes produced by large displacements), ii) restricted the side chain prediction step to the binding interface between the antibody and the epitope (decreasing the amount of rotamers to predict), iii) reduced the resolution of the side chain prediction step (expecting that the minimization process would relocate them into optimal energy conformations), iv) restricted the minimization step only to the binding interface.

After the fine-tuning of these parameters, we ran PELE simulations. We enabled ANM sampling every 4 PELE steps, and we used the OPLS2005 force field with the VDGBNP implicit solvent. PELE simulations were performed using 144 independent trajectories during 48h.

By performing many PELE steps over time the binding process of the epitope (gp120) and the antibody can be simulated. Our original hypothesis was that, taking everything together, the rate of reaching bound conformations would be different between sensitive and resistant strains. Concretely, we believed that reaching deeper bound conformations will be difficult or impossible for resistant ones (since they have been experimentally described to not bind to the antibody). Therefore, we believed that the analysis of these PELE simulations would provide hints that could be exploited for the prediction of antibody binding efficacy.

PELE-antibody results

Population analysis of the PELE simulations was conducted to evaluate the binding process of gp120-antibody systems. Such (statistical) analysis was mainly based on the solvent-accessibility surface area (SASA) of the gp120, which indicates the proportion of the protein that is exposed to the solvent. A representative scheme of such population analysis is illustrated in **Figure 35**.

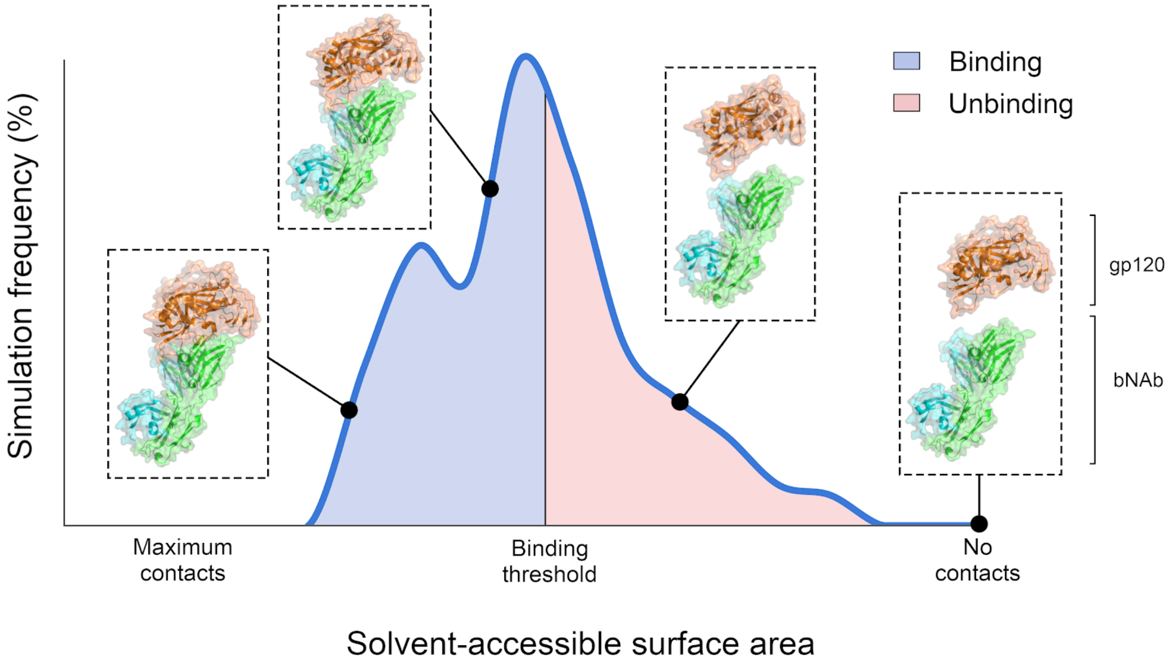


Figure 35. Graphical representation of a PELE population analysis. PELE simulations start from unbinding conformations, and over time, contacts between the epitope and the antibody may form. SASA of the gp120 indicates the amount of contacts formed with the antibody. Relative binding affinities can be determined by applying a SASA threshold along the simulation (**Equation 14)**.

Here, simulation starts from an unbinding state (SASA value of 1) where there are no contacts between the epitope (gp120) and the antibody (bNAb). As the PELE simulations progress, the amount of contacts between the epitope and the antibody may increase, which in turn will decrease the SASA value of such conformations (part of the gp120 would bind to the antibody instead of being exposed to the solvent). Population analysis of the PELE simulations for the large VRC01 benchmark is shown in **Figure 36**. As can be observed, the binding profiles of sensitive strains differ from the resistant ones: sensitive strains bind better to the antibody (lower SASA values) than resistant strains (higher SASA values).

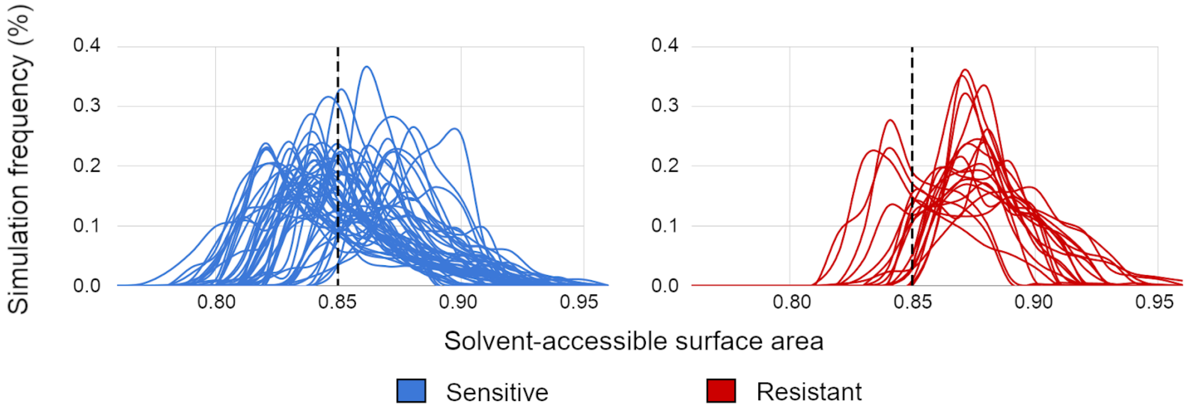


Figure 36. Frequency of SASA values of the PELE simulations in a set of 45 sensitive and 19 resistant gp120 HIV-1 strains against VRC01. Stripped line indicates a SASA threshold of 0.85, used for defining a binding and unbinding cut-off event for the prediction of a relative binding affinity (see **Equation 14**).

A contact SASA threshold of 0.85 was used to compute the predicted relative binding affinity of each strain against VRC01 (**Equation 14**). This threshold was selected by computing the mean of the SASA values from the maximum frequency peaks in all sensitive strains.

$$Predicted\ relative\ binding\ affinity = -\log\left(\frac{Amount\ of\ binding\ events}{Amount\ of\ unbinding\ events}\right) \quad (14)$$

By computing the predicted relative binding affinity, the complex information from the SASA distribution is converted into a numerical score, which facilitates the comprehension of the current output of the simulations. Predicted relative binding affinities for the VRC01 benchmark are reported in **Figure 37**. Favourable predicted binding affinities are those having a negative score, while desfavorable affinities have a positive score. As can be observed, most of the strains are correctly predicted by our contact-based approximation on the PELE simulations: 33 out of 45 sensitive strains and 16 out of 19 resistant ones.

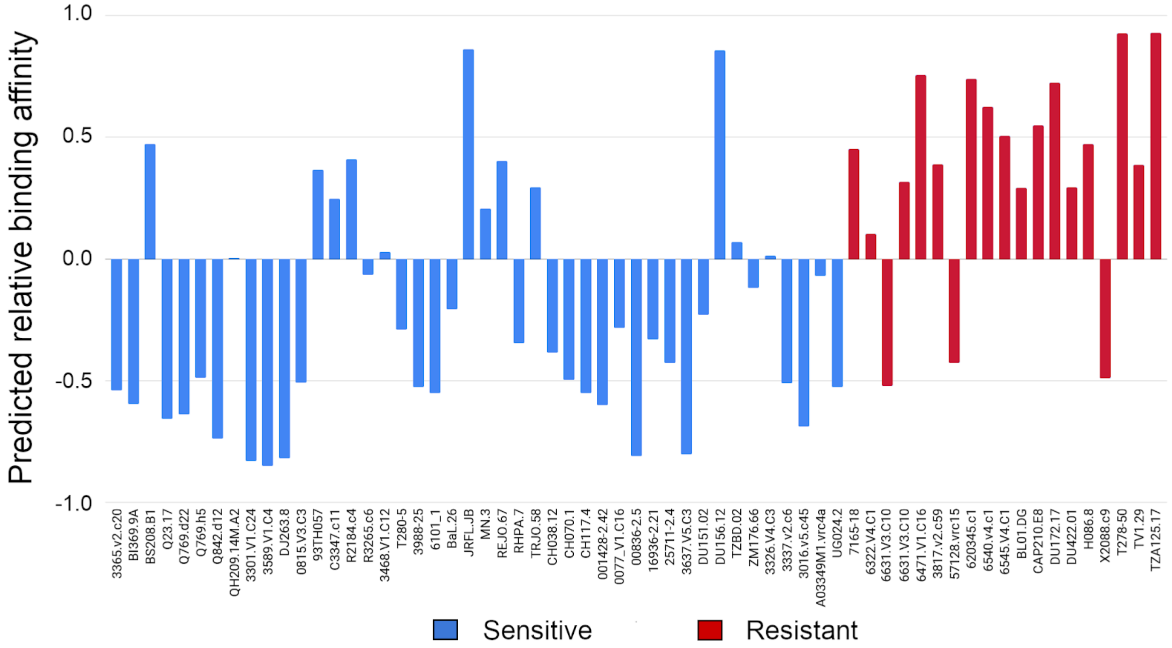


Figure 37. Predicted relative binding affinities of 45 sensitive and 19 resistant HIV-1 strains towards VRC01, extracted from **Figure 36** and **Equation 14**. Most strains were correctly predicted according to the experimental determinations: 33 out of 45 sensitive and 16 out of 19 resistant strains towards VRC01.

The distribution and the Receiver Operator Characteristic (ROC) curve of such predicted affinities is illustrated in **Figure 38**.

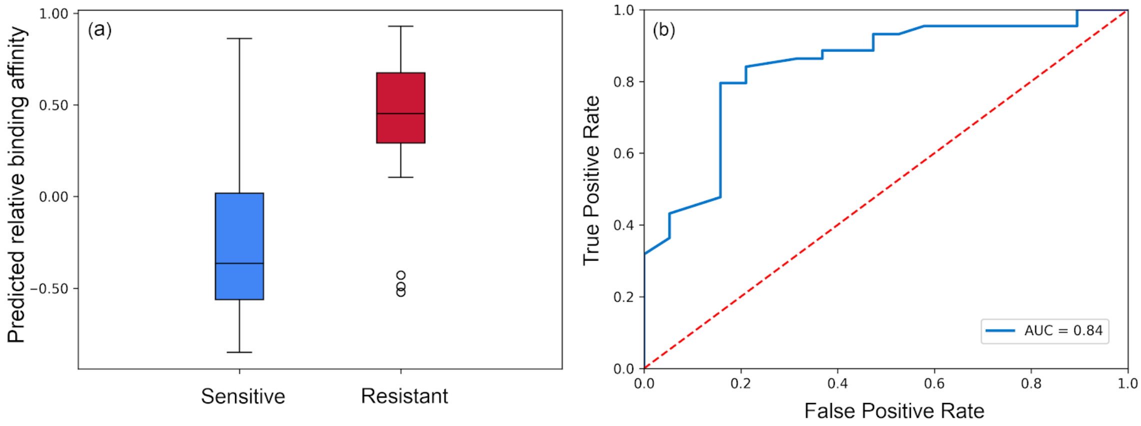


Figure 38. Statistics on the large VRC01 benchmark consisting of 45 sensitive and 19 resistant gp120. (a) Distribution of the predicted binding affinities; (b) Receiver Operator Characteristic (ROC) curve achieving an AUC of 0.84.

As can be observed, the predicted affinity distribution (**Figure 38a**) between sensitive and resistant groups differs substantially, as suggested by previous analyses. Moreover, analysis of the ROC curve (**Figure 38b**) indicates a strong predictive power for determining antibody binding efficacy for the VRC01 benchmark towards sensitive and resistant strains, achieving an area under the curve (AUC) of 0.84.

Next, we aimed to assess the applicability of this approach for the prediction on other anti-CD4bs bNAbs: NIH45-46 and 3BNC117. One of the main limitations of any classifier is the choice of the threshold, which can shape the output of the predictions. Here, we aimed to estimate the optimum threshold by computing the difference in contacts between the bNAbs, without the need of generating large benchmarks as we did for VRC01. We could estimate the optimum SASA threshold for a given uncharacterized antibody (anti-CD4bs) by the fact that the 93TH057 gp120 strain is overrepresented in crystal structures. In this case, the 93TH057 gp120 strain has been co-crystallized with VRC01 (PDB: 3NGB), NIH45-46 (PDB: 3U7Y) and 3BNC117 (PDB: 4JPV). Hence, the optimum SASA threshold for the interpretation of PELE simulations can be estimated for uncharacterized antibodies (**Equation 15**):

$$bNAb_{PELE} = VRC01_{PELE} - (VRC01_{BOUND} - bNAb_{BOUND}) \qquad (15)$$

Here, $bNAb_{PELE}$, $VRC01_{PELE}$, $VRC01_{BOUND}$ and $bNAb_{BOUND}$ represent the optimum SASA threshold for PELE simulations for a uncharacterized bNAb, the optimum SASA threshold for VRC01 (0.85 units of SASA), the SASA of the 93TH057 gp120 bound to the VRC01 crystal structure, and the SASA of the 93TH057 gp120 bound to the uncharacterized bNAb, respectively. Thus, we determined that the optimum SASA thresholds for NIH45-46 and 3BNC117 are 0.83 and 0.86, respectively.

For the PELE simulations, in this case we selected nine different gp120 strains from different clade identities having diverse binding affinities for the three antibodies. This is, to be sensitive for one antibody and resistant for the other two, or the other way around. Selected strains, clade identity and their binding affinity towards the three bNAbs (sensitive and resistant) are shown in **Figure 39**.

Strain name	Clade	VRC01	NIH45-46	3BNC117
3988.25	B	Sensitive	Sensitive	Resistant
7165.18	B	Resistant	Resistant	Sensitive
MN.3	B	Sensitive	Sensitive	Resistant
CH038.12	BC	Sensitive	Sensitive	Resistant
CAP210.E8	C	Resistant	Resistant	Sensitive
DU172.17	C	Resistant	Resistant	Sensitive
UG024.2	D	Sensitive	Resistant	Sensitive
3016.v5.c45	D	Sensitive	Resistant	Sensitive
57128.vrc15	D	Resistant	Resistant	Sensitive

Figure 39. List of the nine evaluated strains with the three bNAbs: VRC01, NIH45-46 and 3BNC117. Clade identity, and binding efficacy to the above mentioned bNAb is shown.

Predicted binding affinity scores of the nine selected gp120 towards the three bNAbs is illustrated in **Figure 40**. Our results indicate that, as observed for the large VRC01 benchmark, most sensitive strains are predicted as such (12 out of 14), as well for resistant ones (10 out of 13). These results demonstrate the wider applicability of the proposed technique using very limited knowledge without the need of performing extensive benchmarks.

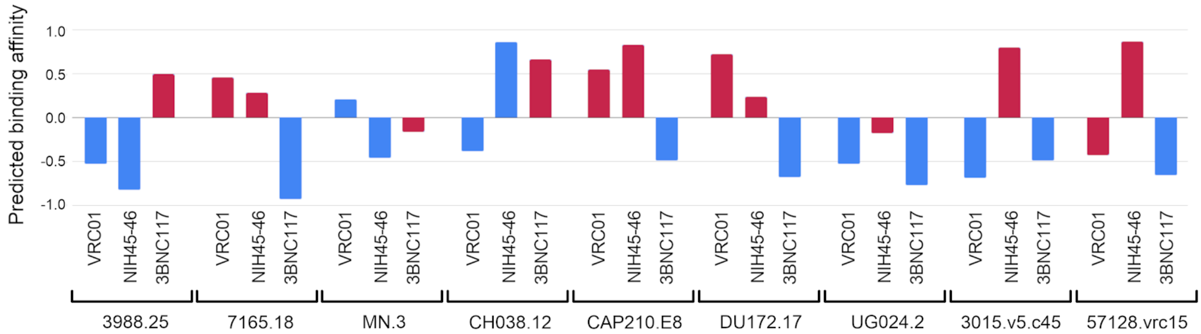


Figure 40. Predicted relative binding affinities using PELE simulations of 9 different gp120 HIV-1 strains towards VRC01, NIH45-46 and 3BNC117. Most of the sensitive strains (blue) are correctly predicted (12 out of 14), while also most of the resistant strains (red) are predicted as such (10 out of 13).

Finally, we aimed to investigate the structural basis that drives resistance in gp120 strains. After comparison of the three-dimensional models generated by PELE for sensitive and resistant strains, we observed that the antibody is unable to accommodate two important contact regions of the gp120, the loop D and the β23-V5-β24-α5 region. In fact, mutations in those regions have been associated with resistance patterns in previous studies (Huang *et al.*, 2016; Lynch *et al.*, 2015; Li *et al.*, 2011). Those mutations are typically placed in the position 279 (N/D-279-K/E/Q/R), 280 (N280D), 281 (A281T), 456 (R456W) and 458 (G458D). Multiple resistant strains evaluated in this work contain the above mentioned mutations, and all of them have been predicted to be resistant by our method. Those strains are TV1.29, DU422.01, TZA125.17, 6471.V1.C16, 620345.c1, BL01.DG and H086.8. The other resistant strains contain other mutations on the β23-V5-β24-α5 region that could play a major role for resistance mechanisms. Importantly, both loop D and β23-V5-β24-α5 are close in space (**Figure 41**), and mutations in any of both regions could affect the conformational dynamics of both loops, hindering the accommodation of the gp120.

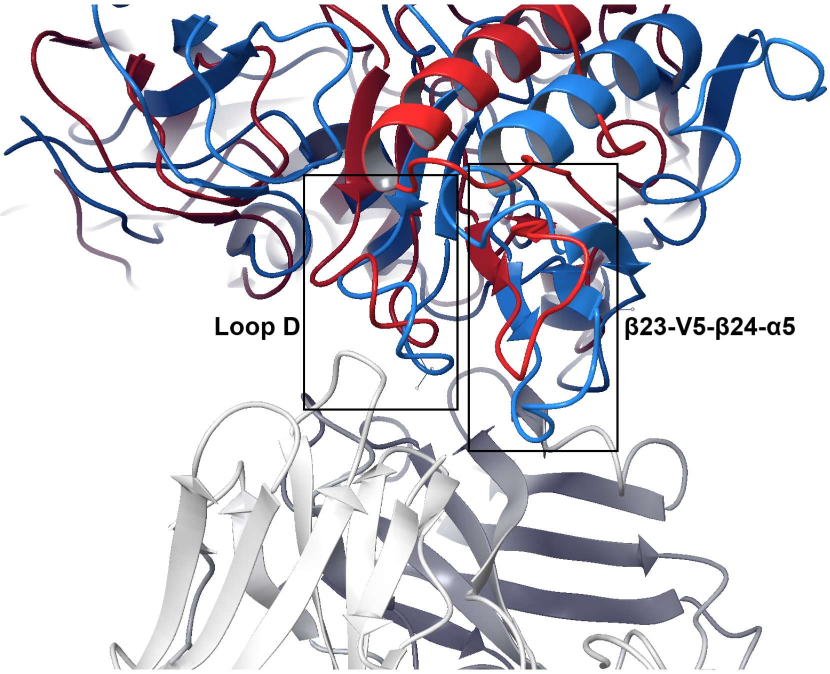


Figure 41. Representative binding mode of sensitive (Q482.d12, blue) and resistant (620345.c1, red) gp120 strains to VRC01. Analysis of the PELE simulations indicate that the antibody is unable to accommodate the loop D and the β23-V5-β24-α5 regions of the resistant gp120 strains.

Designing a potent broadly neutralizing antibody

As mentioned in previous sections, bNAbs may be elicited against HIV-1. The most potent bNAbs discovered so far bind to the CD4bs of the envelope glycoprotein gp120. Within this group, N6 is the most potent (**Figure 18**), showing one of the greatest strain coverage (~98%) combined with one of the greatest antibody potencies (0.066 µg/mL). The main advantage of the N6 antibody compared to other VRC01-class bNAbs is the ability to avoid steric clashes that originate resistance (Lynch *et al.*, 2015; Huang *et al.*, 2016). In this section, the distinctive features of the N6 antibody will be discussed for both heavy and light chains, which are key factors for understanding our rational design processes.

- The entire antibody adopts a rotation and shift displacement compared to other VRC01-class antibodies. This induces a different binding mode orientation that greatly contributes to the binding efficiency of the N6 (Huang *et al.*, 2016).
- Regarding the heavy chain (HC) of the N6 antibody, its CDRH2 loop contributes to the ~50% contact surface with the gp120, and it contains most of the essential-known residues of VRC01-class antibodies for a potent binding. Firstly, a large hydrophobic residue ($Tyr54_{HC}$) mimics the interaction of $Phe43_{CD4}$ with gp120. Secondly, important salt bridges between the antibody and the epitope ($Arg71_{HC}$ and $Asp368_{gp120}$) are conserved. Thirdly, a key conserved residue ($Trp100c_{HC}$) interacts with loop D $Asp/Asn279_{gp120}$. Lastly, the CDRH2 evolved to contain a $Gly_{60}GlyGly_{62}$ motif (not found in any other previously isolated CD4bs antibodies), which helps to accommodate variations on the V5 loop by avoiding side chain clashes.
- Regarding the light chain (LC) of the N6 antibody, the overall rotation of the antibody makes the CDRL3 and the N-terminal region move away from the highly variable V5 loop, which also helps to prevent clashes that drive resistance. Moreover, the CDRL1 contains another $Gly_{28}GlyGly_{30}$ motif that permits avoiding steric clashes with the loop D, especially with the glycosylation site on $Asn276_{gp120}$.

Overall, this set of features are responsible for the success of the N6 antibody. In this work, we aimed to design a potent variant of the N6 antibody. In this context, it is important to keep in mind that the N6 antibody has been isolated from blood samples of a patient (Z258) showing potent and broad serum against HIV-1. N6 antibody was the most potent antibody described in the serum of that patient, and to date N6 is the most potent antibody against AIDS. In this context, aiming to improve the potency of the N6 antibody is not a trivial task. In fact, a very little increase in antibody potency is expected to occur from the best possible designs, specially for single mutations; otherwise, such mutations would have been positively selected during the maturation process within the patient. Another layer of complexity resides in the fact that an improvement of the design must be observed against several viral isolates, which challenges the design process.

In this section, we will describe three different attempts of improving the binding potency of the N6 antibody. Firstly, we will describe our attempt to elongate the CDRH3 loop of the N6 with a sequence described to improve the binding potency of other VRC01-class antibodies. Secondly, we will describe a fully computational approach for identifying single mutations on the heavy and light chain of the N6 antibody that could improve the binding affinity towards multiple viral isolates. Lastly, we will describe a fully rational design process for improving the binding affinity of N6 by designing a double mutation on the light chain that helps to accommodate the gp120 strain. All of the above described attempts are accompanied by experimental validation performed by our collaborators from IrsiCaixa. The main topic of this section is to provide insights about the rational design process rather than providing an in-depth detail of the experimental validation process. However, a brief detail of the experimental protocol is described in the following lines:

Both N6 expressing plasmid for heavy and light chains were obtained from the NIH-AIDS referent research reagent (code 12967 and 12966, respectively). All N6 variants were generated by directed mutagenesis using overlapping oligonucleotides, which incorporated the mutations. The Phusion hotstart polymerase (Thermofisher) was used for polymerase chain reaction (PCR) amplification. The original N6 expressing plasmids were used as a template. N6 antibody and its variants were produced in vitro by co-transfecting both heavy and light plasmids using the Expi293 Cells Expression System (Thermofisher). Cell culture supernatants were harvested after 5 days from transfection and clarified by centrifugation (20 minutes at 3000g and filtration at 0,22µm). Concentration of antibodies in supernatants was determined by an in-house developed ELISA. Briefly, a goat anti-Human IgG (Jackson Immunology) was used at 1µg/mL as coating and N6 purified antibody was used as standard starting at 0,3µg/mL. All cell supernatants were evaluated for a different dilution using horseradish peroxidase (HRP)-conjugated goat anti-Human IgG (Jackson Immunology) as secondary antibody.

Antibodies were purified from clarified cell culture by protein A chromatography using HiTrap protein A HP columns and an ÄKTA Start equipment, and following standard methods.

The binding of N6 antibody and its variants was measured towards different HIV-1 envelope glycoproteins by ELISA using both HisTag-expressing monomeric recombinant gp120 (BaL isolate) and a trimeric recombinant gp140 protein (B41 SOSIP 664 his). Briefly, plates were coated anti-HisTag antibody (Thermofisher) and then gp120 or gp140 were used at 0,1µg/mL. All N6 variants were tested starting at 0,3µg/ml and seven 1/3 serial dilution to evaluate binding saturation. A HRP-conjugated goat anti-Human IgG was used as a secondary antibody.

Binding to full native HIV envelope glycoprotein from different isolates was evaluated by flow-cytometry. Expi293 cells were transfected with plasmid coding for the envelope glycoproteins. After 24 hours, cells were collected and incubated with N6 variants at 1µg/mL for 30 minutes at room temperature. A phycoerythrin (PE)-labelled goat anti-human IgG secondary antibody was used to detect cell surface bound antibodies. Untransfected Expi293 cells were used as negative control. Cells were analysed using a FACSCelesta cytometer from Becton Dickinson.

On the design of a chimera bNAb

VRC01-like class antibodies show a high degree of sequence conservation. Some examples of this phenomenon may be found on the heavy chains of VRC01-like class antibodies, such as for instance N6-VRC27 (91% sequence conservation), VRC07-NIH45-46 (98% sequence conservation) or VRC01-VRC07 (94% sequence conservation). Achieving high sequence similarity results in very similar three-dimensional antibody structures. A representation of the three-dimensional similarity between four different VRC01-like class bNAbs (VRC01, VRC07, NIH45-46 and N6) is illustrated in **Figure 42**. Color codes represent the 93TH057 gp120 (blue), the antibody heavy chain (orange), and the antibody light chain (yellow). As can be observed, all antibody-epitope complexes overlap in the three-dimensional space.

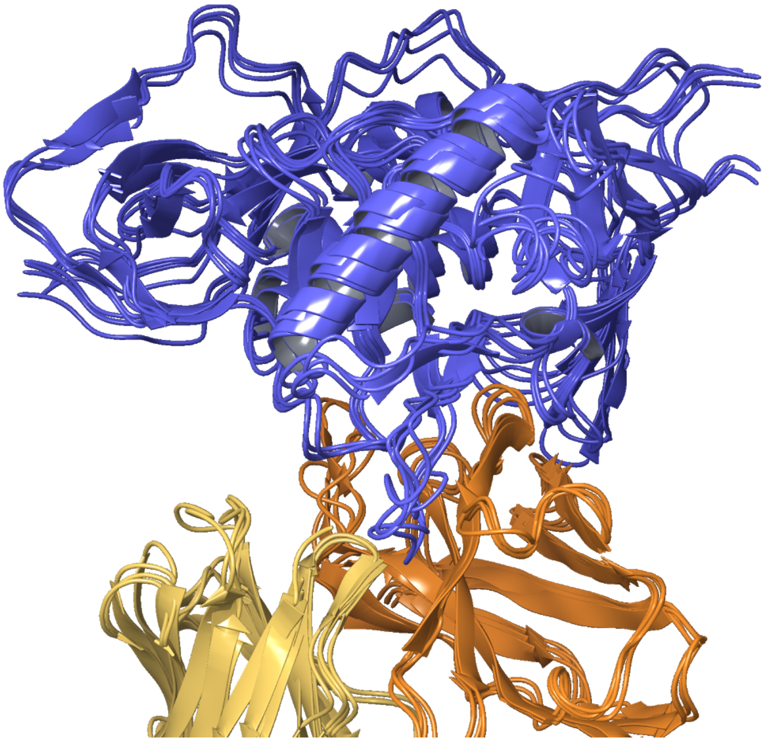


Figure 42. Binding mode of four VRC01-like class bNAbs: VRC01, VRC07, NIH45-46 and N6 (from 3NGB, 4OM0, 3UTY and 5TE6 PDB entries, respectively). The gp120, the antibody heavy chain and the antibody light chain are represented in blue, orange and yellow, respectively. As can be observed, the binding mode between the VRC01-like class antibodies is conserved and overlaps three-dimensionally.

However, despite showing very high sequence and structural similarity, some differences may be observed. Such differences are mainly located on the variable regions CDRH1, CDRH2 and CDRH3. CDR regions play an important role for gp120 binding and CDRH2 is the most important region, since it makes half of the contacts of the total antibody with the gp120. On the other hand, CDRH1 and CDRH3 play accessorial/supporting roles, and depending on the bNAb, such regions exert larger or lesser impact on the binding affinity. Interestingly, VRC01-like class antibodies accumulate high sequence diversity on the CDRH3. In this sense, there are two groups of VRC01-like class antibodies: those containing a 4-residue insertion (-ARDY-) and those which do not. This insertion has been described in NIH45-46 and VRC07, while it is not found on VRC01 either in N6. Here we show the linear sequence of the heavy chain of those antibodies.

```
          -------------FR1-------------- CDRH1
NIH45-46  EVRLSQSGGQMKKPGESMRLSCRASGYEFL NCPIN
VRC07     QVRLSQSGGQMKKPGDSMRISCRASGYEFI NCPIN
VRC01     QVQLVQSGGQMKKPGESMRISCRASGYEFI DCTLN
N6        RAHLVQSGTAMKKPGASVRVSCQTSGYTFT AHILF

          ------FR2----- ------CDRH2------
NIH45-46  WIRLAPGRRPEWMG WLKPRGGAVNYARKFQG
VRC07     WIRLAPGKRPEWMG WMKPRGGAVSYARQLQG
VRC01     WIRLAPGKRPEWMG WLKPRGGAVNYARPLQG
N6        WFRQAPGRGLEWVG WIKPQYGAVNFGGGFRD

          --------------FR3---------------
NIH45-46  RVTMTRDVYSDTAFLELRSLTSDDTAVYFCTR
VRC07     RVTMTRDMYSETAFLELRSLTSDDTAVYFCTR
VRC01     RVTMTRDVYSDTAFLELRSLTVDDTAVYFCTR
N6        RVTLTRDVYREIAYMDIRGLKPDDTAVYYCAR

          ------CDRH3----- ----FR4----
NIH45-46  GKYCTARDYYNWDFEH WGRGAPVTVSS
VRC07     GKYCTARDYYNWDFEH WGQGTPVTVSS
VRC01     GKNCD----YNWDFEH WGRGTPVIVSS
N6        DRSYGD---SSWALDA WGQGTTVVVSA
```

As can be observed, N6 sequence differs from the other antibodies. The main structural difference between N6 and the other antibodies is the lack of a conserved disulphide bond between the CDRH1 and the CDRH3 (highlighted in red).

Previous studies revealed that the -ARDY- insertion is responsible of an enhanced antibody potency in NIH45-46 and VRC07 compared to the VRC01 (Diskin *et al.*, 2011; Rudicell *et al.*, 2014). The authors demonstrated that this insertion elongates the CDRH3 loop, creating additional favourable contacts with the inner domain of the gp120, leading to an increase of the antibody binding potency. **Figure 43** indicates the three-dimensional structure of the CDRH3 of NIH45-46 (**Figure 43a**), VRC07 (**Figure 43b**), VRC01 (**Figure 43c**) and N6 (**Figure 43d**). Here, antibodies containing the 4-residue insertion -ARDY- (NIH45-46 and VRC07) achieved a well defined CDRH3 secondary structure consisting of an antiparallel beta-sheet organization, while the others (VRC01 and N6) show a disordered short loop.

On the other hand, the presence (or the lack) of the disulphide bond between the CDRH1 and CDRH3 regions is illustrated in **Figure 44** of NIH45-46 (**Figure 44a**), VRC07 (**Figure 44b**), VRC01 (**Figure 44c**) and N6 (**Figure 44d**) antibody.

Amino acid contacts between the CDRH3 and the gp120 strain are illustrated in **Figure 45** of NIH45-46 (**Figure 45a**), VRC07 (**Figure 45b**), VRC01 (**Figure 45c**) and N6 (**Figure 45d**) antibody. Antibodies containing the -ARDY- insertion achieve deeper contacts with the inner domain of the gp120, which is presumably one of the main drivers of increasing the antibody potency compared to VRC01.

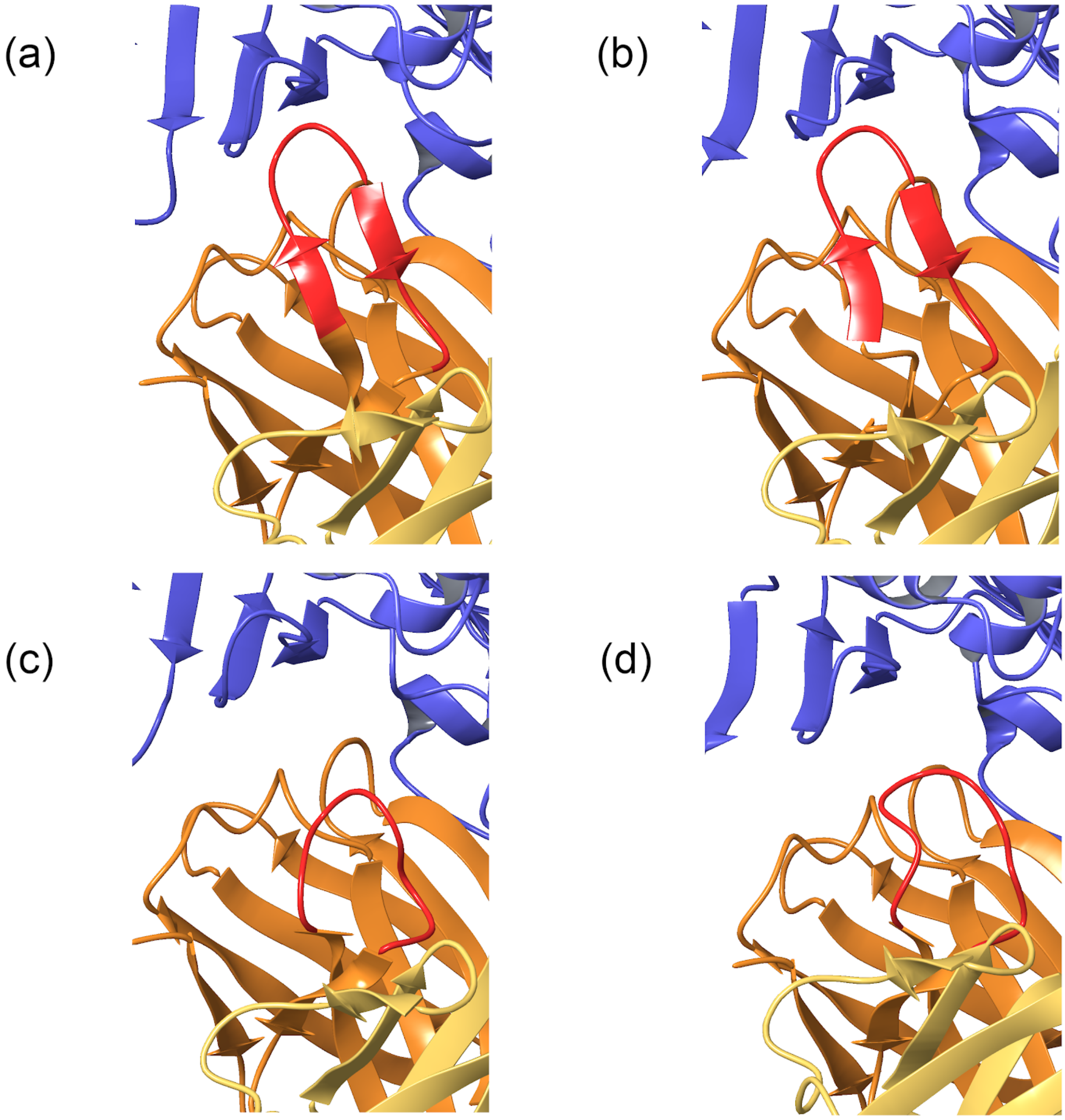


Figure 43. Structural conformation of the CDRH3 regions of (a) NIH45-46, (b) VRC07, (c) VRC01 and (d) N6 VRC01-like class antibodies. The gp120, the antibody heavy chain, the antibody light chain and the CDRH3 sequence are shown in blue, orange, yellow and red, respectively. An structurated antiparallel beta-sheet can be observed for the antibodies containing the -ARDY- insertion on the CDRH3, NIH45-46 and VRC07.

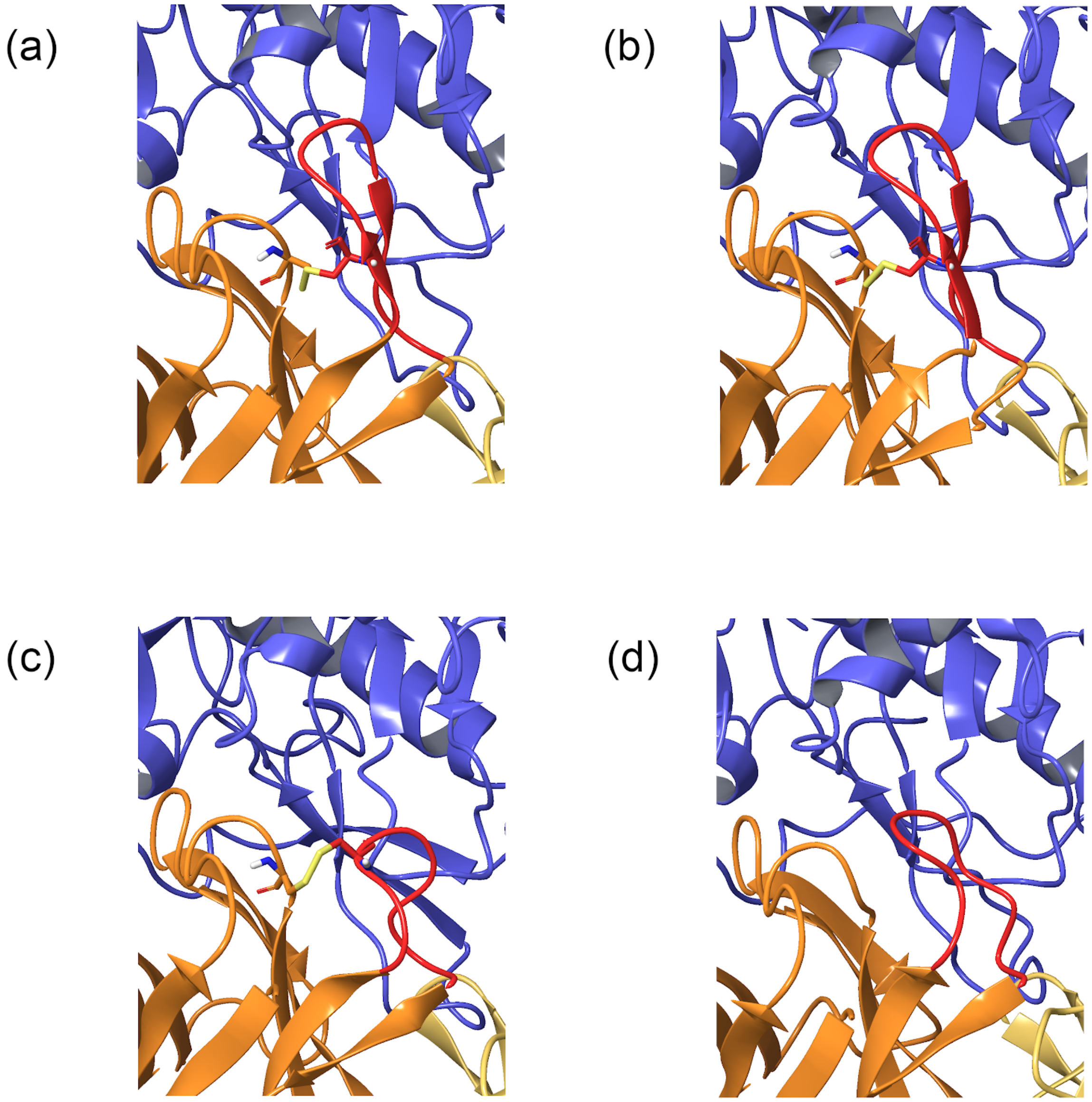


Figure 44. Presence of a disulphide bond on (a) NIH45-46, (b) VRC07, (c) VRC01, and not in (d) N6 VRC01-like class antibodies. The gp120, the antibody heavy chain, the antibody light chain and the CDRH3 sequence are shown in blue, orange, yellow and red, respectively. The disulphide bond is represented in yellow, and covalently connects the CDRH1 and the CDRH3.

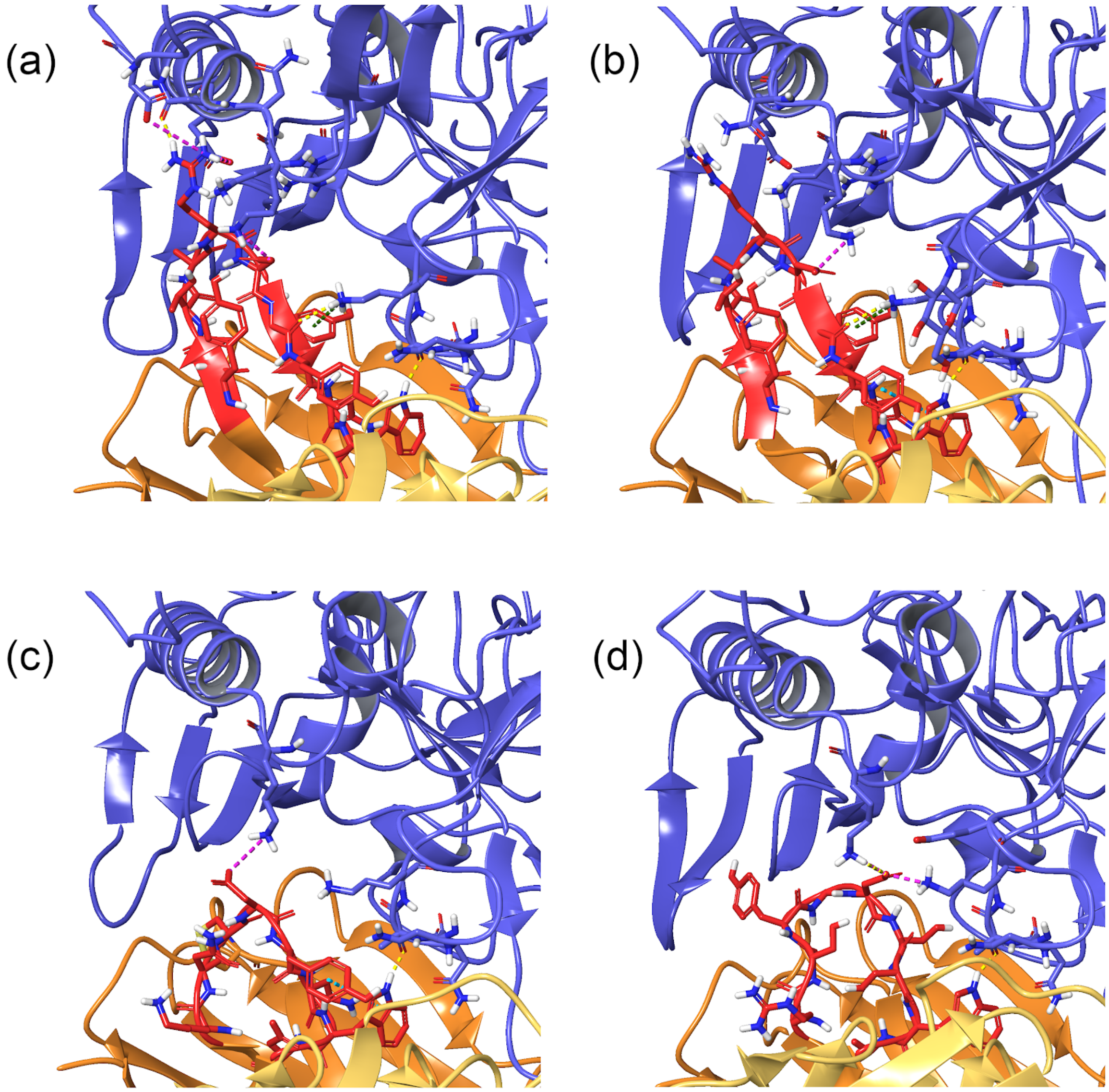
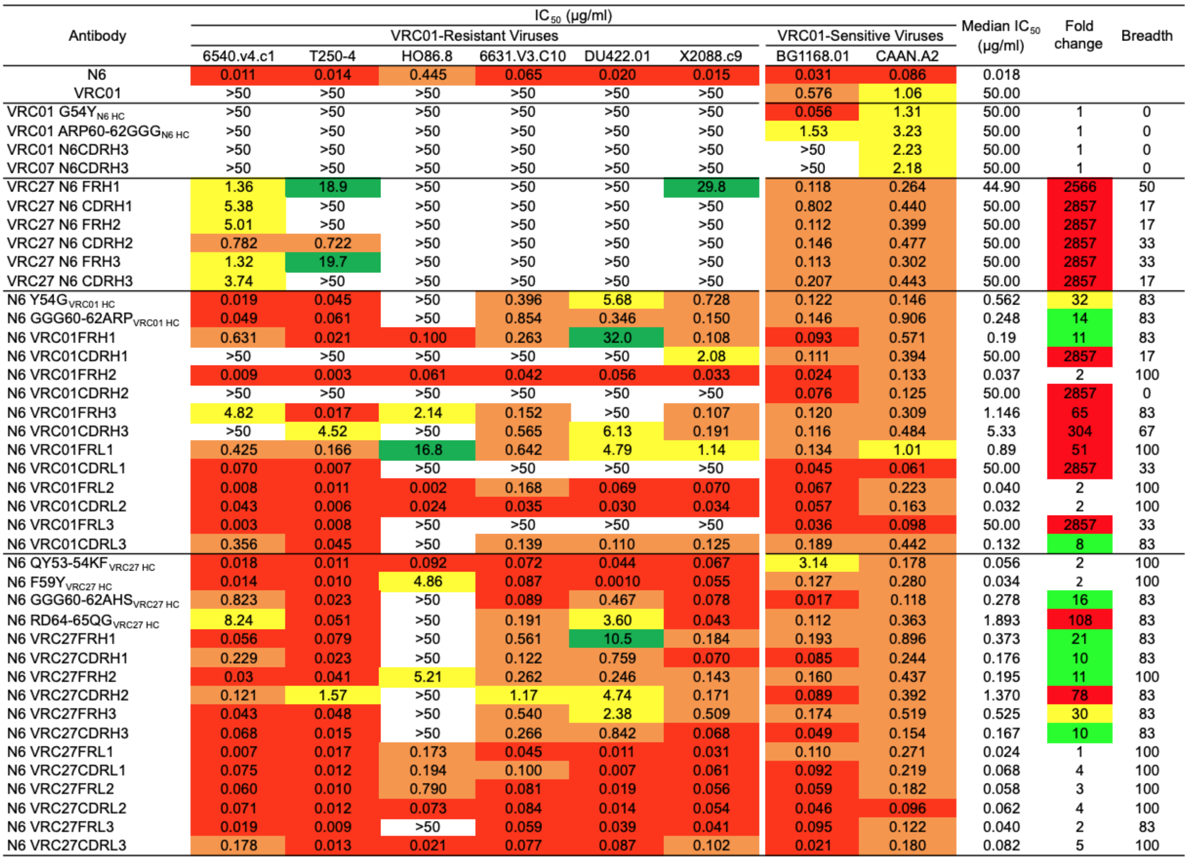


Figure 45. CDRH3 residues involved in binding with the gp120 for (a) NIH45-46, (b) VRC07, (c) VRC01, and (d) N6 VRC01-like class antibodies. The gp120, the antibody heavy chain, the antibody light chain and the CDRH3 sequence are shown in blue, orange, yellow and red, respectively. Dashed lines represent intermolecular bonds: hydrogen bond (yellow), salt bridge (pink) and π-π stacking (cyan) π-cation interactions (dark green).

Previous research has demonstrated that CDR and FR regions are exchangeable between these antibodies (Huang *et al.*, 2016). Hence, chimeras of those bNAbs can be produced and their impact on binding affinity can be evaluated. This is represented in **Figure 46**, where the authors of the above mentioned article exchanged all CDR and FR regions of the N6 antibody for those found on VRC01 and VRC27, and in the other way around.



Antibody	IC$_{50}$ (μg/ml)								Median IC$_{50}$ (μg/ml)	Fold change	Breadth
	VRC01-Resistant Viruses						VRC01-Sensitive Viruses				
	6540.v4.c1	T250-4	HO86.8	6631.V3.C10	DU422.01	X2088.c9	BG1168.01	CAAN.A2			
N6	0.011	0.014	0.445	0.065	0.020	0.015	0.031	0.086	0.018		
VRC01	>50	>50	>50	>50	>50	>50	0.576	1.06	50.00		
VRC01 G54Y$_{\text{N6 HC}}$	>50	>50	>50	>50	>50	>50	0.056	1.31	50.00	1	0
VRC01 ARP60-62GGG$_{\text{N6 HC}}$	>50	>50	>50	>50	>50	>50	1.53	3.23	50.00	1	0
VRC01 N6CDRH3	>50	>50	>50	>50	>50	>50	>50	2.23	50.00	1	0
VRC07 N6CDRH3	>50	>50	>50	>50	>50	>50	>50	2.18	50.00	1	0
VRC27 N6 FRH1	1.36	18.9	>50	>50	>50	29.8	0.118	0.264	44.90	2566	50
VRC27 N6 CDRH1	5.38	>50	>50	>50	>50	>50	0.802	0.440	50.00	2857	17
VRC27 N6 FRH2	5.01	>50	>50	>50	>50	>50	0.112	0.399	50.00	2857	17
VRC27 N6 CDRH2	0.782	0.722	>50	>50	>50	>50	0.146	0.477	50.00	2857	33
VRC27 N6 FRH3	1.32	19.7	>50	>50	>50	>50	0.113	0.302	50.00	2857	33
VRC27 N6 CDRH3	3.74	>50	>50	>50	>50	>50	0.207	0.443	50.00	2857	17
N6 Y54G$_{\text{VRC01 HC}}$	0.019	0.045	>50	0.396	5.68	0.728	0.122	0.146	0.562	32	83
N6 GGG60-62ARP$_{\text{VRC01 HC}}$	0.049	0.061	>50	0.854	0.346	0.150	0.146	0.906	0.248	14	83
N6 VRC01FRH1	0.631	0.021	0.100	0.263	32.0	0.108	0.093	0.571	0.19	11	83
N6 VRC01CDRH1	>50	>50	>50	>50	>50	2.08	0.111	0.394	50.00	2857	17
N6 VRC01FRH2	0.009	0.003	0.061	0.042	0.056	0.033	0.024	0.133	0.037	2	100
N6 VRC01CDRH2	>50	>50	>50	>50	>50	>50	0.076	0.125	50.00	2857	0
N6 VRC01FRH3	4.82	0.017	2.14	0.152	>50	0.107	0.120	0.309	1.146	65	83
N6 VRC01CDRH3	>50	4.52	>50	0.565	6.13	0.191	0.116	0.484	5.33	304	67
N6 VRC01FRL1	0.425	0.166	16.8	0.642	4.79	1.14	0.134	1.01	0.89	51	100
N6 VRC01CDRL1	0.070	0.007	>50	>50	>50	>50	0.045	0.061	50.00	2857	33
N6 VRC01FRL2	0.008	0.011	0.002	0.168	0.069	0.070	0.067	0.223	0.040	2	100
N6 VRC01CDRL2	0.043	0.006	0.024	0.035	0.030	0.034	0.057	0.163	0.032	2	100
N6 VRC01FRL3	0.003	0.008	>50	>50	>50	>50	0.036	0.098	50.00	2857	33
N6 VRC01CDRL3	0.356	0.045	>50	0.139	0.110	0.125	0.189	0.442	0.132	8	83
N6 QY53-54KF$_{\text{VRC27 HC}}$	0.018	0.011	0.092	0.072	0.044	0.067	3.14	0.178	0.056	2	100
N6 F59Y$_{\text{VRC27 HC}}$	0.014	0.010	4.86	0.087	0.0010	0.055	0.127	0.280	0.034	2	100
N6 GGG60-62AHS$_{\text{VRC27 HC}}$	0.823	0.023	>50	0.089	0.467	0.078	0.017	0.118	0.278	16	83
N6 RD64-65QG$_{\text{VRC27 HC}}$	8.24	0.051	>50	0.191	3.60	0.043	0.112	0.363	1.893	108	83
N6 VRC27FRH1	0.056	0.079	>50	0.561	10.5	0.184	0.193	0.896	0.373	21	83
N6 VRC27CDRH1	0.229	0.023	>50	0.122	0.759	0.070	0.085	0.244	0.176	10	83
N6 VRC27FRH2	0.03	0.041	5.21	0.262	0.246	0.143	0.160	0.437	0.195	11	100
N6 VRC27CDRH2	0.121	1.57	>50	1.17	4.74	0.171	0.089	0.392	1.370	78	83
N6 VRC27FRH3	0.043	0.048	>50	0.540	2.38	0.509	0.174	0.519	0.525	30	83
N6 VRC27CDRH3	0.068	0.015	>50	0.266	0.842	0.068	0.049	0.154	0.167	10	83
N6 VRC27FRL1	0.007	0.017	0.173	0.045	0.011	0.031	0.110	0.271	0.024	1	100
N6 VRC27CDRL1	0.075	0.012	0.194	0.100	0.007	0.061	0.092	0.219	0.068	4	100
N6 VRC27FRL2	0.060	0.010	0.790	0.081	0.019	0.056	0.059	0.182	0.058	3	100
N6 VRC27CDRL2	0.071	0.012	0.073	0.084	0.014	0.054	0.046	0.096	0.062	4	100
N6 VRC27FRL3	0.019	0.009	>50	0.059	0.039	0.041	0.095	0.122	0.040	2	83
N6 VRC27CDRL3	0.178	0.013	0.021	0.077	0.087	0.102	0.021	0.180	0.082	5	100

Figure 46. Contributions of N6, VRC01 and VRC27 paratope components to their breadth and potency. Adapted from Figure S5 of the article “Identification of a CD4-Binding-Site Antibody to HIV that Evolved Near-Pan Neutralization Breadth” , *Immunity*. (2016); 45(5): 1108–1121, DOI: 10.1016/j.immuni.2016.10.027 (https://www.cell.com/immunity/fulltext/S1074-7613(16)30438-1).

Taking into account that CDR exchanges between N6 and VRC01 produce antibody variants that can be expressed and produced, we envisioned a chimera consisting of the N6 scaffold where the CDRH1 and CDRH3 are exchanged by the NIH45-46 one. This construct would contain the -ARDY- insertion on the CDRH3, together with the two cysteines needed to form the disulfide bond between the CDRH1 and CDRH3. The sequence of the design is shown below (CHIMERA). Here, the sequence fragments coming from the N6 antibody are shown in black, while those coming from the NIH45-46 antibody are shown in blue.

```
         --------------FR1--------------- CDRH1
N6       RAHLVQSGTAMKKPGASVRVSCQTSGYTFT AHILF
NIH45-46 EVRLSQSGGQMKKPGESMRLSCRASGYEFL NCPIN
CHIMERA  RAHLVQSGTAMKKPGASVRVSCQTSGYTFL NCPIN

         ------FR2----- ------CDRH2------
N6       WFRQAPGRGLEWVG WIKPQYGAVNFGGGFRD
NIH45-46 WIRLAPGRRPEWMG WLKPRGGAVNYARKFQG
CHIMERA  WFRQAPGRGLEWVG WIKPQYGAVNFGGGFRD

         --------------FR3---------------
N6       RVTLTRDVYREIAYMDIRGLKPDDTAVYYCAR
NIH45-46 RVTMTRDVYSDTAFLELRSLTSDDTAVYFCTR
CHIMERA  RVTLTRDVYREIAYMDIRGLKPDDTAVYFCTR

         ------CDRH3----- ----FR4----
N6       DRSYGD---SSWALDA WGQGTTVVVSA
NIH45-46 GKYCTARDYYNWDFEH WGRGAPVTVSS
CHIMERA  GKYCTARDYYNWDFEH WGQGTTVVVSA
```

Next, we generated the three-dimensional structure of the chimeric variant by homology modeling techniques using Prime from Schrödinger. Then, we applied a series of energy minimizations by using the OPLS3e force field (Roos *et al.*, 2019) aiming to generate adequate conformers for the native and the chimera design structures. After generating energetically favourable conformations, we assessed their binding strength towards the 93TH057 gp120 strain. Aiming to do so, we docked the antibodies close to its native position on the CD4bs, and initiated PELE simulations. In this case, we allowed very small translations and rotations; our main goal was to analyze the interaction energy (kcal/mol) of the antibody-epitope while allowing side chain and backbone sampling. PELE simulations were carried out during 48 hours, and we used the OPLS2005 force field, the VDGBNP implicit solvent, and ANM motions were allowed for every four PELE steps. At this time, the

PELE-antibody protocol was not established and as stated above, we focused our analysis on the interaction energy of the antibody-epitope complex (**Figure 47**).

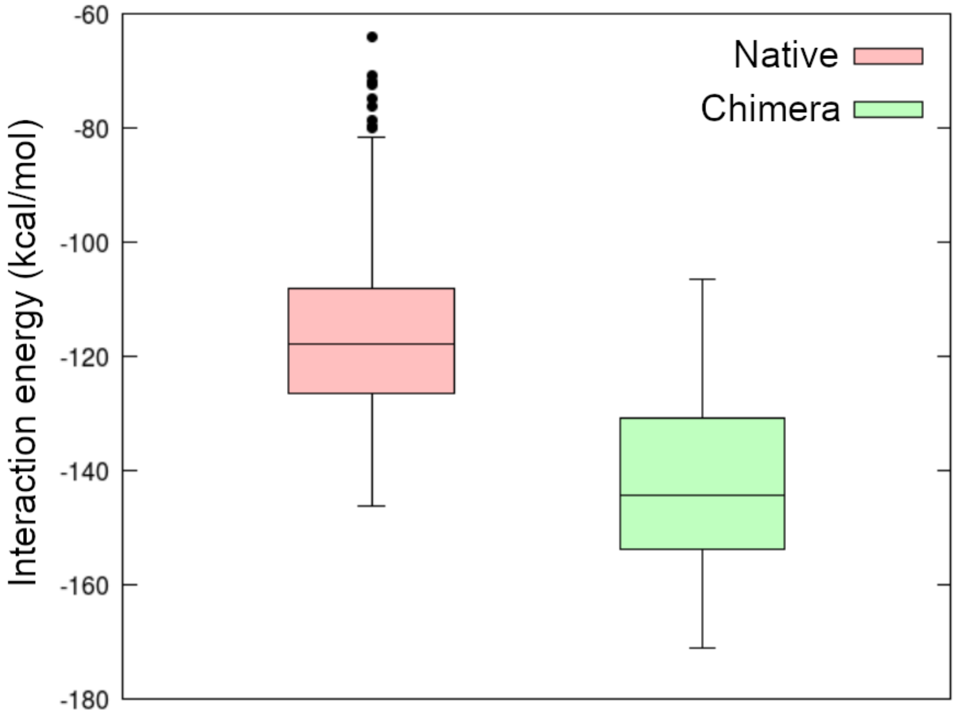


Figure 47. Interaction energy (kcal/mol) distribution of the antibody-epitope complex generated by PELE. As can be observed, the chimera antibody achieves stronger interaction energies with the 93TH057 gp120 strain, suggesting that this design could improve the binding potency of the native N6 antibody.

Here, the distribution of the interaction energies of antibody-epitope complex during PELE simulations is shown. Our chimera design achieved stronger interaction energies than the native N6 antibody, which may indicate that this design could improve the antibody potency of the N6 antibody. This increase in the interaction energy would be induced by the extra contacts originated by the -ARDY- insertion with the inner domain of the gp120, as observed in **Figure 45a**.

After these promising results, the antibody design sequence was sent to our experimental collaborators at IrsiCaixa, aiming to produce and experimentally evaluate the binding potency of the design. However, the antibody could not be produced on their experimental design setup, and therefore, we could not evaluate the effectiveness of this design.

Aiming to improve the N6 antibody: first round

At the very beginning of this thesis, our goal was to improve the N6 antibody by means of computational simulations. During this first round, we conducted multiple simulation assays aiming to increase the binding affinity of the antibody-epitope complex. Here, we envisioned a fully computational approach for identifying single mutations on the heavy and light chain of the N6 antibody that could improve the binding affinity towards multiple viral isolates. In this section, we will summarize the computational workflow that we followed, together with the experimental validations of the proposed designs.

As mentioned previously, we envisioned a method for predicting which single mutations on the N6 antibody could lead to a general binding potency improvement towards most (or all) viral isolates. Aiming to do so, first we needed to obtain a large amount of gp120 sequences to evaluate the binding of the antibody designs. Fortunately, literature contains a panel assay of 181 viral isolates representing clade diversity (Rudicell *et al.*, 2014; Huang *et al.*, 2016). Here, the main difficulty consisted of retrieving those sequences from GeneBank, since only the gp120 strain name (Virus ID) was provided. The main problem with this is that GeneBank codes do not correspond with the Virus ID for most of the viral isolates, and an extensive search followed by a manual data curation step was required.

After this selection process, an amount of 172 complete gp120 sequences were successfully located. Virus ID, clade identity, affinity to N6 (N6 aff; S, M and R represents sensitive, medium and resistant binding affinity towards N6) and GenBank code of all 172 complete gp120 sequences is illustrated in **Table 30**.

Virus ID	Clade	N6 aff	GenBank	Virus ID	Clade	N6 aff	GenBank
0260.v5.c36	A	S	HM215256.1	MN.3	B	S	HM215430.1
0330.v4.c3	A	S	HM215257.1	PVO.04	B	S	AY835444.1
0439.v5.c1	A	S	HM215258.1	QH0515.01	B	S	AY835440.1
3365.v2.c20	A	S	HM215298.1	QH0692.42	B	M	AY835439.1
3415.v1.c1	A	S	HM215299.1	REJO.67	B	S	AY835449.1
3718.v3.c11	A	S	HM215306.1	RHPA.7	B	S	AY835447.1
398-F1_F6_20	A	S	HM215312.2	SC422.8	B	S	AY835441.1
BI369.9A	A	S	DQ187019.1	SF162.LS	B	S	EU123924.1
BS208.B1	A	S	KX364401.1	SS1196.01	B	S	AY835442.1
KER2008.12	A	M	AY736809.1	THRO.18	B	M	AY835448.1
KER2018.11	A	S	AY736810.1	TRJO.58	B	S	AY835450.1
KNH1209.18	A	S	AY736813.1	TRO.11	B	S	AY835445.1
MS208.A1	A	S	DQ187010.1	WITO.33	B	S	AY835451.1
Q23.17	A	S	AF004885.1	YU2.DG	B	S	M93258.1
Q259.17	A	S	AF407152.1	CH038.12	BC	S	EF042692.1
Q461.e2	A	S	AF407156.1	CH070.1	BC	S	EF117255.1
Q769.d22	A	S	AF407158.1	CH117.4	BC	S	EF117262.1
Q769.h5	A	S	AF407159.1	CH181.12	BC	S	EF117259.1
Q842.d12	A	S	AF407160.1	CNE15	BC	S	HM215401.1
QH209.14M.A	A	S	FJ866118.1	CNE40	BC	S	HM215414.1
RW020.2	A	S	EU855131.1	CNE7	BC	S	HM215426.1
UG037.8	A	S	HQ702713.1	286,36	C	S	JQ362420.1
3301.V1.C24	AC	S	HM215294.1	288,38	C	M	JQ362421.1
3589.V1.C4	AC	S	HM215304.1	0013095-2.11	C	S	EF117267.1
6540.v4.c1	AC	S	HM215330.1	001428-2.42	C	S	EF117266.1
6545.V4.C1	AC	M	HM215332.1	0077_V1.C16	C	S	HM215254.1

0815.V3.C3	ACD	S	HM215260.1	00836-2.5	C	S	EF117265.1
6095.V1.C10	ACD	S	HM215323.1	0921.V2.C14	C	S	HM215262.1
3468.V1.C12	AD	S	HM215301.1	16055-2.3	C	S	EF117268.1
Q168.a2	AD	S	AF407148.1	16845-2.22	C	M	EF117269.1
620345.c1	AE	M	JQ362422.1	16936-2.21	C	S	EF117270.1
C1080.c3	AE	M	AY945712.1	25710-2.43	C	S	EF117271.1
C2101.c1	AE	S	AY945716.1	25711-2.4	C	M	EF117272.1
C3347.c11	AE	S	AY945721.1	25925-2.22	C	M	EF117273.1
C4118.09	AE	S	AY945722.1	26191-2.48	C	S	EF117274.1
CNE3	AE	S	HM215410.1	3168.V4.C10	C	S	HM215289.1
CNE5	AE	S	HM215415.1	3637.V5.C3	C	M	HM215305.1
CNE55	AE	S	HM215418.1	3873.V1.C24	C	M	HM215311.1
CNE56	AE	S	HM215419.1	6322.V4.C1	C	S	HM215326.1
CNE59	AE	S	HM215422.1	6471.V1.C16	C	R	HM215328.1
M02138	AE	M	AY713424.1	6631.V3.C10	C	S	HM215335.1
R1166.c1	AE	M	AY945728.1	6644.V2.C33	C	S	HM215336.1
R2184.c4	AE	S	AY945730.1	6785.V5.C14	C	M	HM215338.1
R3265.c6	AE	S	AY945732.1	6838.V1.C35	C	S	HM215341.1
TH966.8	AE	S	U08456.1	96ZM651.02	C	S	AF286224.1
TH976.17	AE	S	U08458.1	BR025.9	C	S	U15121.1
235-47	AG	S	EU513195.1	CAP210.E8	C	M	DQ435683.1
242-14	AG	M	EU513188.1	CAP244.D3	C	S	DQ435684.1
263-8	AG	S	EU513182.1	CAP45.G3	C	S	DQ435682.1
269-12	AG	S	EU513194.1	CNE30	C	M	HM215411.1
271-11	AG	S	EU513197.1	CNE31	C	M	HM215412.1
928-28	AG	S	EU513199.1	CNE53	C	S	HM215417.1
DJ263.8	AG	M	AF063223.1	CNE58	C	S	HM215421.1

T250-4	AG	S	EU513189.1	DU123.06	C	S	DQ411850.1
T251-18	AG	M	EU513196.1	DU151.02	C	S	DQ411851.1
T253-11	AG	S	EU513191.1	DU156.12	C	S	DQ411852.1
T255-34	AG	S	EU513184.1	DU172.17	C	S	DQ411853.1
T257-31	AG	M	EU513185.1	DU422.01	C	S	DQ411854.1
T266-60	AG	S	EU513193.1	MW965.26	C	S	U08455.1
T278-50	AG	R	EU513198.1	TV1.29	C	R	EU855132.1
T280-5	AG	S	EU513183.1	TZA125.17	C	M	JQ362423.1
T33-7	AG	S	EU513186.1	TZBD.02	C	S	JQ362424.1
3988,25	B	S	AY835436.1	ZA012.29	C	S	EU855133.1
5768,04	B	S	AY835435.1	ZM106.9	C	S	AY424163.2
6535,3	B	S	AY835438.1	ZM109.4	C	S	AY424138.2
7165,18	B	M	AY835437.1	ZM135.10a	C	M	AY424079.2
89.6.DG	B	S	U39362.2	ZM197.7	C	S	DQ388515.1
AC10.29	B	M	AY835446.1	ZM214.15	C	S	DQ388516.1
ADA.DG	B	S	AY426119.1	ZM215.8	C	S	DQ422948.1
Bal.01	B	S	DQ318210.1	ZM233.6	C	S	DQ388517.1
BaL.26	B	S	DQ318211.1	ZM249.1	C	S	DQ388514.1
BG1168.01	B	S	AY835443.1	ZM53.12	C	M	AY423984.2
BL01.DG	B	R	AY124970.1	ZM55.28a	C	S	AY423971.2
BR07.DG	B	M	AY124979.1	3326.V4.C3	CD	S	HM215296.1
BX08.16	B	S	GQ855765.1	3337.V2.C6	CD	S	HM215297.1
CAAN.A2	B	S	AY835452.1	3817.v2.c59	CD	M	HM215310.1
CNE10	B	S	HM215397.1	231965.c1	D	S	JQ361079.1
CNE12	B	S	HM215399.1	247-23	D	S	EU683891.1
CNE14	B	S	HM215400.1	3016.v5.c45	D	S	HM215283.1
CNE4	B	M	HM215413.1	57128.vrc15	D	M	AY736829.1

CNE57	B	S	HM215420.1	6405.v4.c34	D	S	HM215327.1
HO86.8	B	M	EF210732.1	A03349M1.4	D	M	HM215356.1
HT593.1	B	S	U08444.1	NKU3006.ec1	D	M	AY736835.1
HXB2.DG	B	S	K03455.1	UG021.16	D	S	U27399.1
JRCSF.JB	B	S	AY669726.1	UG024.2	D	S	U43386.1
JRFL.JB	B	S	U63632.1	X2088.c9	G	S	EU885764.1

Table 30. Manually curated search of the GeneBank identifiers for 172 gp120 viral isolated. Virus ID, clade identity, affinity to N6 (N6 aff) and GeneBank identifier is shown. Regarding N6 affinity, S, M and R represents sensitive, moderate and resistant affinities towards the antibody, respectively.

After successfully retrieving those sequences from GenBank, we aimed to obtain their three-dimensional structures from the PDB database. However, for the majority of the retrieved gp120 sequences there is no available three-dimensionally solved structure. Aiming to bypass this limitation, we modelled all gp120 lacking a three-dimensional structure on the PDB. The modeling process was performed by using homology modeling techniques using Prime from Schrödinger.

Homology models were constructed using the template structure from the PDB with high-sequence similarity to each gp120 sequence. A minimization round by using the OPLS3e force field was performed after the generation of the model. Next, all gp120 strains were aligned three-dimensionally and were carefully examined. After checking their integrity, we docked the N6 antibody into its correct binding orientation in all gp120 strains. After the docking process, a series of minimization rounds were performed until the total energy of the system converged into favourable conformations. Finally, the three-dimensional structures of all N6-gp120 complexes were carefully evaluated. No clashes were observed in any of the protein-protein complexes.

After the correct generation of the antibody-gp120 complexes, we located the residues involved on the protein-protein interface in all structures. Then, a position contact map of the interacting residues between the antibody and the gp120 structures was built from all protein-protein structures (**Table 31**). The position contact map was built using the common positions in the majority of the protein-protein complexes; concretely those found in at least 150 out of 172 protein-protein interfaces (considering a contact radius of 5 Å between the antibody and the gp120). A total of 71 positions were located on the protein-protein interface: 24 positions for N6 heavy chain, 11 positions for N6 light chain and 36 positions for gp120.

Heavy chain of N6	Light chain of N6	gp120
30, 33, 47, 50, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 64, 71, 73, 74, 99, 100, 100A, 100B, 100C.	1, 2, 25, 27, 28, 29, 32, 90, 91, 96, 97.	97, 122, 124, 275, 276, 278, 279, 280, 281, 282, 283, 365, 366, 367, 368, 370, 371, 425, 426, 427, 429, 430, 431, 432, 455, 456, 457, 458, 459, 460, 461, 462, 469, 473, 474, 475.

Table 31. Position contact map of the interacting residues between the N6 antibody and the 172 viral isolates extracted from GenBank. Those positions were selected by a cutoff radius of 5 Å and to be found in at least 150 out of 172 antibody-gp120 complexes.

After selecting the common interface positions of the antibody towards most of the gp120, we aimed to predict the effects of all possible single mutations on them. In order to do so, we performed the unanimous decision of FoldX, pyDock and UEP. As we described in the **UEP section**, the combination of these three predictors improves the classification performance compared to any method alone (EvoEF1 was not evaluated at that time). An amount of 700 mutations were evaluated for each N6-gp120 complex (20 different possible amino acids for each of the 35 different antibody positions). Considering all 172 N6-gp120 complexes, an amount of 120.400 mutations were evaluated for each predictor (700 mutations per complex for a total of 172 complexes), resulting in a total amount of 361.200 predictions made by FoldX, pyDock and UEP.

After generating all ΔΔG predictions, data was grouped and analyzed aiming to select mutations improving the binding energy for most N6-gp120 complexes. As mentioned previously, the selection was performed following the unanimous decision of the choice of the three predictors. Hence, all mutations unanimously predicted to increase the binding energy of the complex in a minimum of 150 out of 172 gp120 strains were selected. Our final selection consisted of 21 mutations from the 700 possible mutations on the antibody. Interestingly, selected mutations are located in only 6 different positions (from the 35 possible ones), indicating possible hot-spots regions. The three-dimensional location on the protein-protein interface of the 6 selected positions (heavy chain: 30, 52, 54, 57 and 74, and light chain: 96) is illustrated in **Figure 48**.

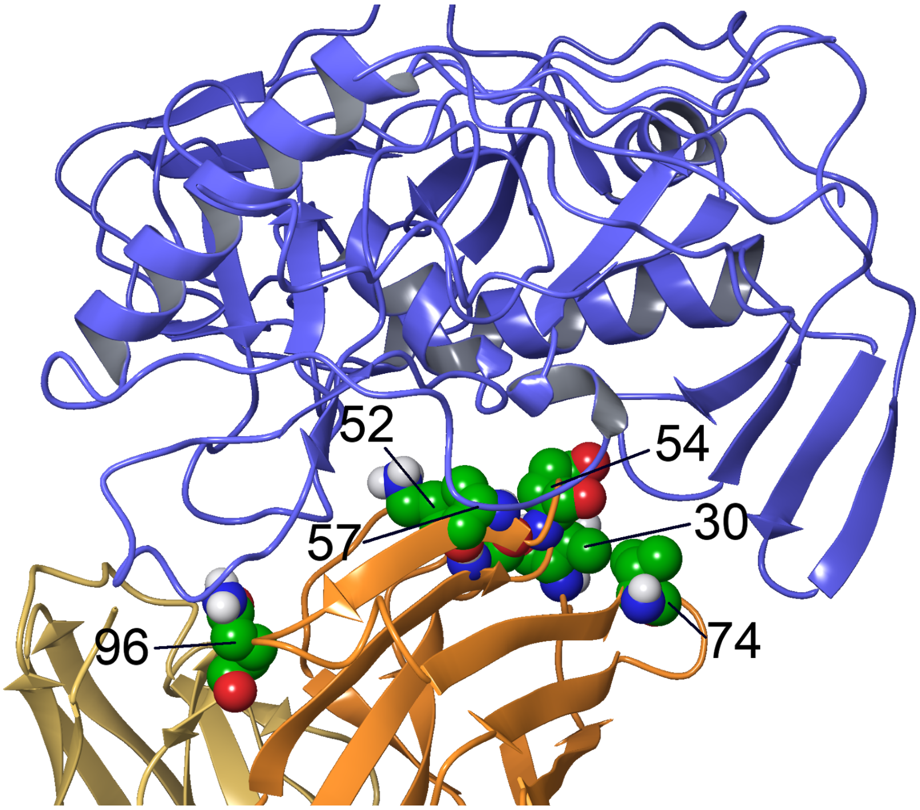
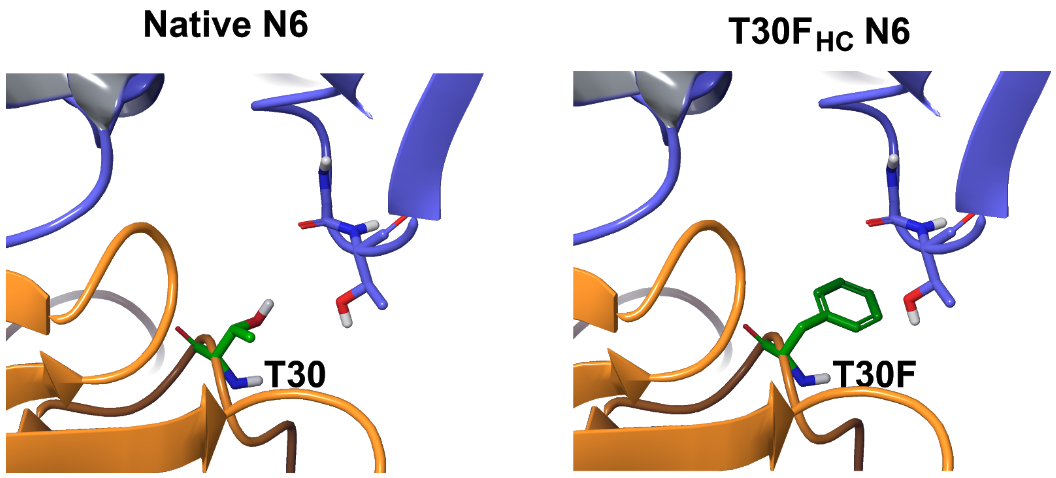


Figure 48. Selected positions for increasing the binding potency of the N6 antibody from the unanimous decision of FoldX, pyDock and UEP. An amount of 21 different mutations were selected on the positions of the heavy chain: 30, 52, 54, 57 and 74, and on the light chain: 96. Selection was performed under the binding energy improvement on at least 150 out of the 172 N6-gp120 strain complexes.

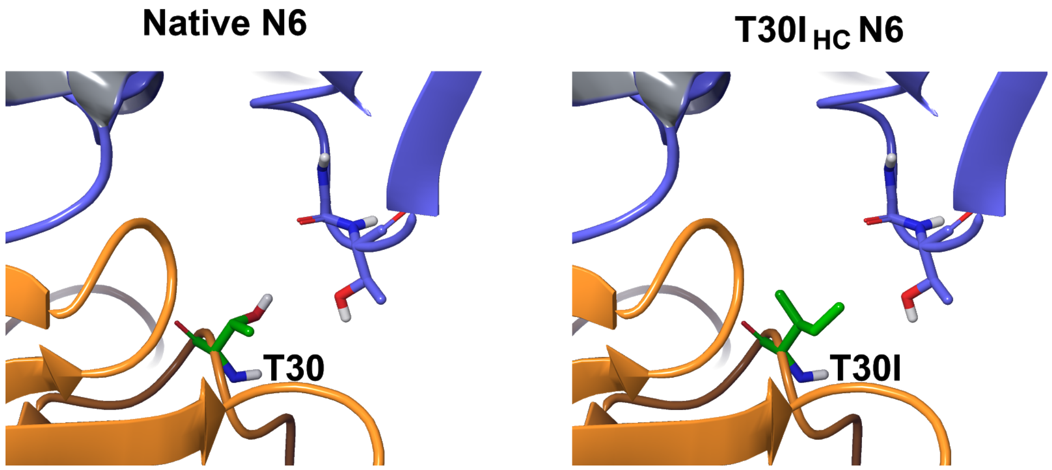
The 21 mutations selected after this process are T30F, T30I, T30L, T30M, T30V, T30W, T30Y, K52R, Q54K, Q54M, Q54R, Q54W, Q54Y, A57F, A57L, A57M, A57W, A57Y, V74M, V74R and Q96G (this one for light chain only). The three-dimensional modeling of those mutations y can be found in **Figures 49-69**, respectively.



Native N6

T30F$_{HC}$ N6

Figure 49. Three-dimensional model of the T30F$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.



Native N6

T30I$_{HC}$ N6

Figure 50. Three-dimensional model of the T30I$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

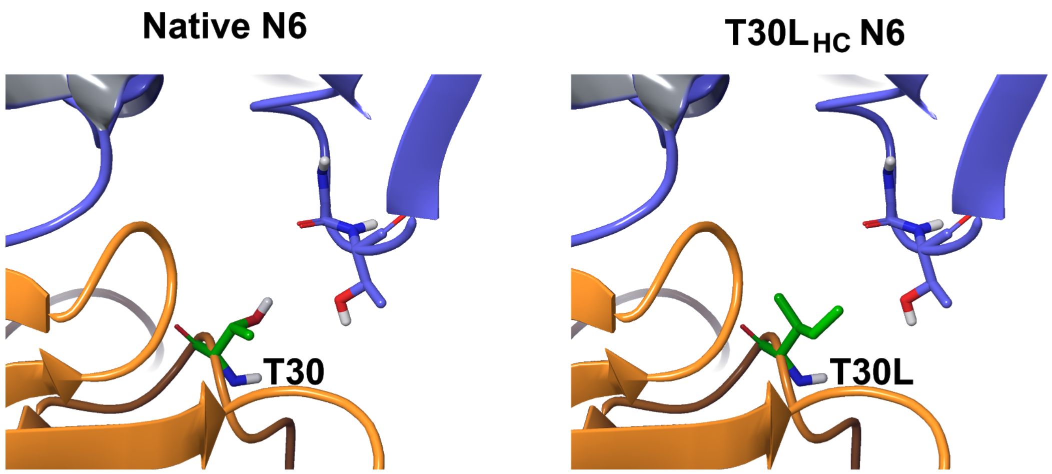


Figure 51. Three-dimensional model of the T30L$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

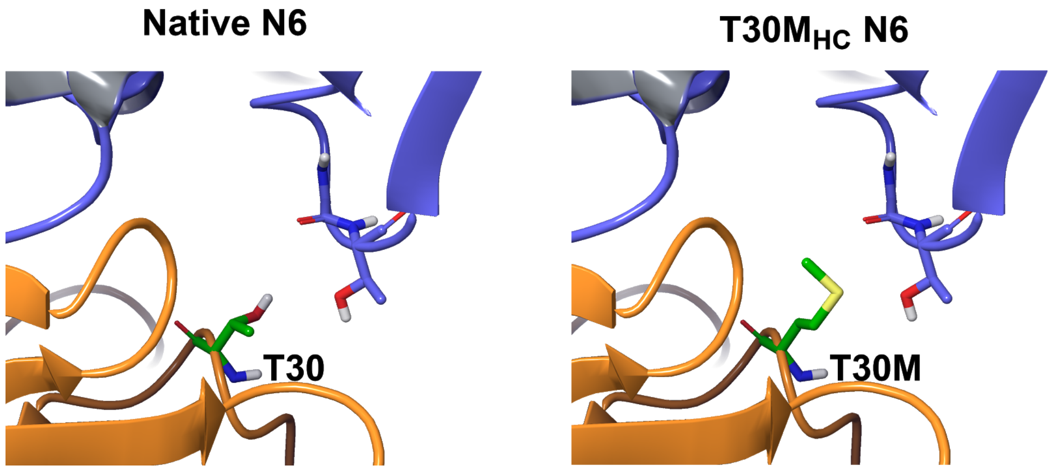


Figure 52. Three-dimensional model of the T30M$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

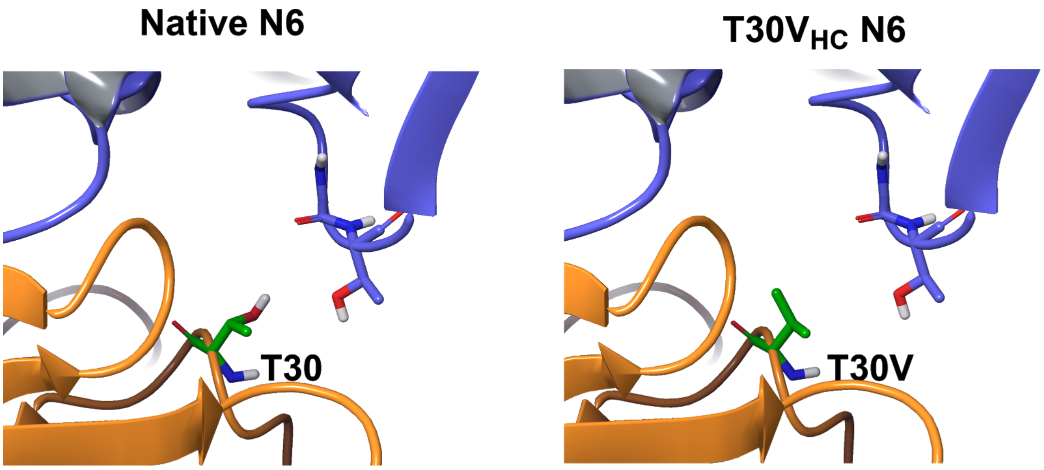


Figure 53. Three-dimensional model of the T30V$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

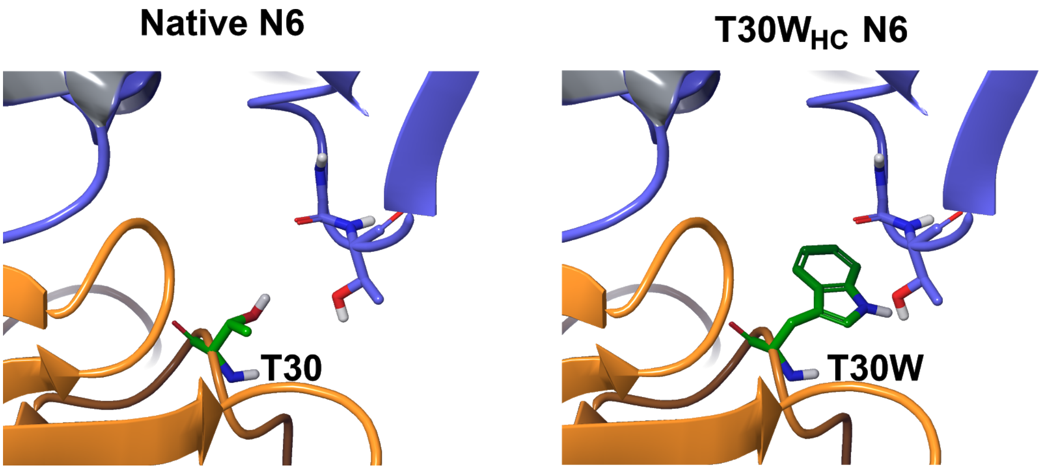


Figure 54. Three-dimensional model of the T30W$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

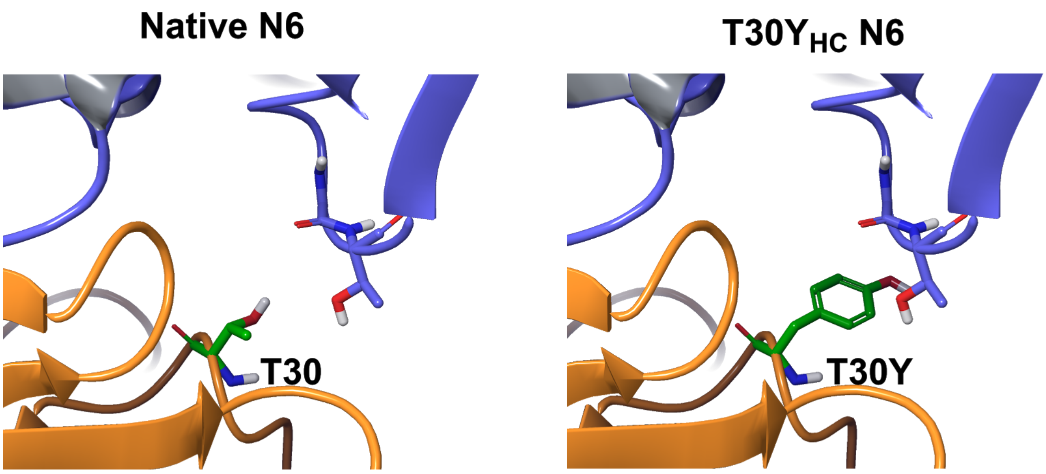


Figure 55. Three-dimensional model of the T30Y$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

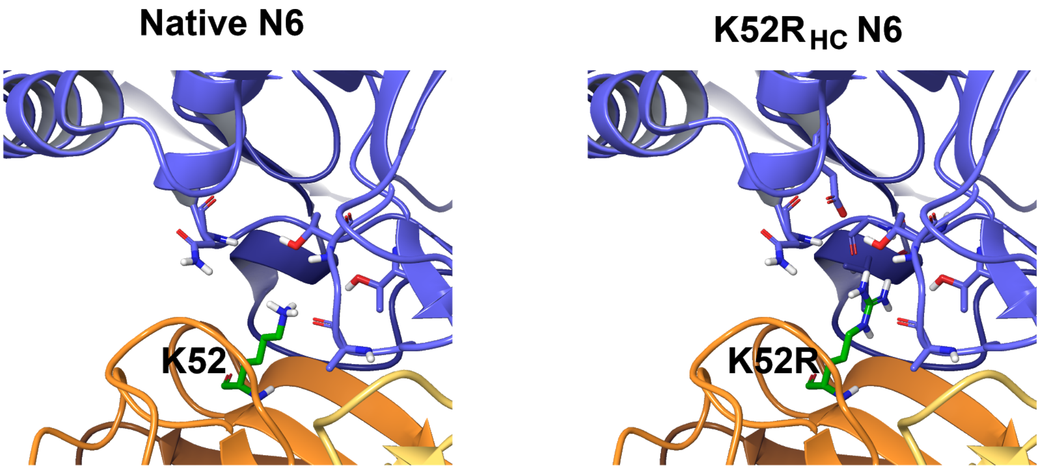
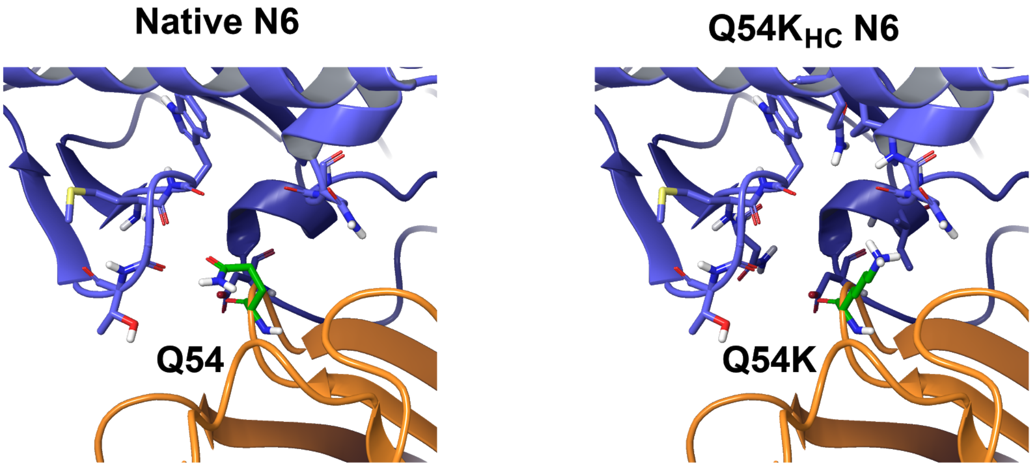
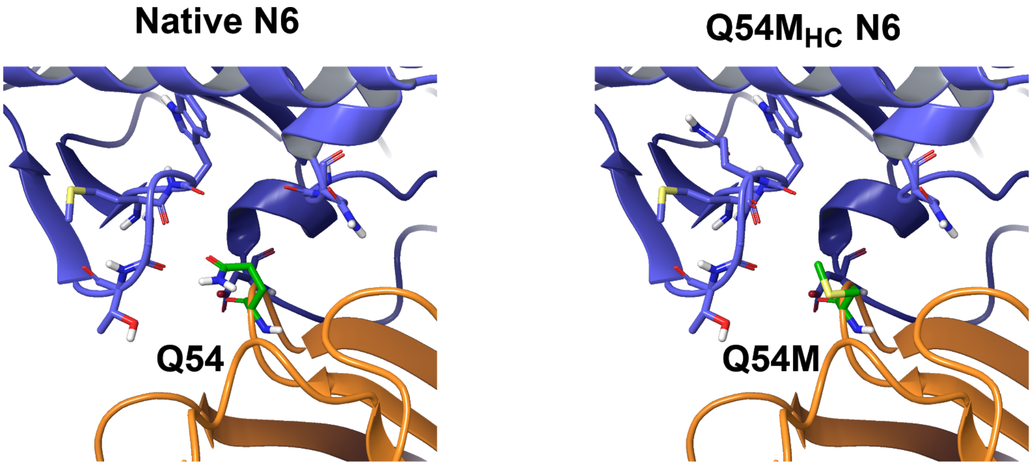


Figure 56. Three-dimensional model of the K52R$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown. Light chain is represented in yellow.


Native N6

Q54K$_{HC}$ N6

Figure 57. Three-dimensional model of the Q54K$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.


Native N6

Q54M$_{HC}$ N6

Figure 58. Three-dimensional model of the Q54M$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

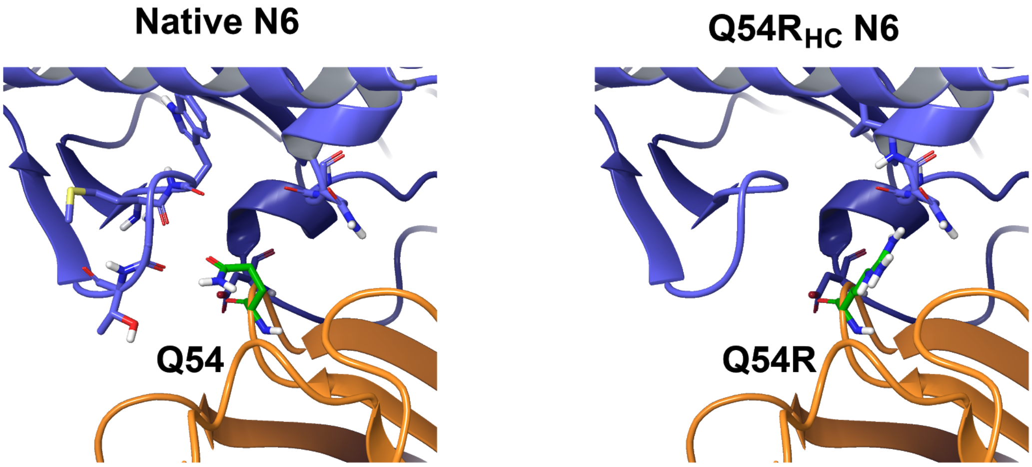


Figure 59. Three-dimensional model of the Q54R$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

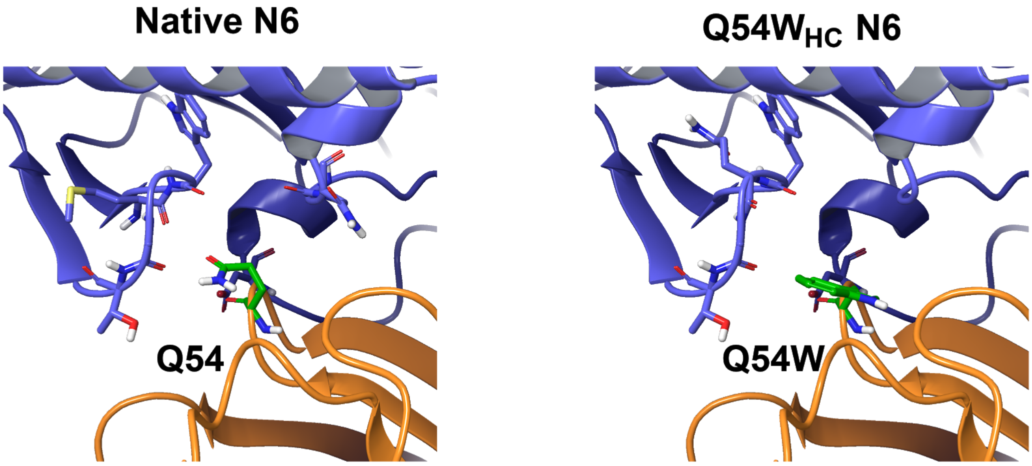


Figure 60. Three-dimensional model of the Q54W$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

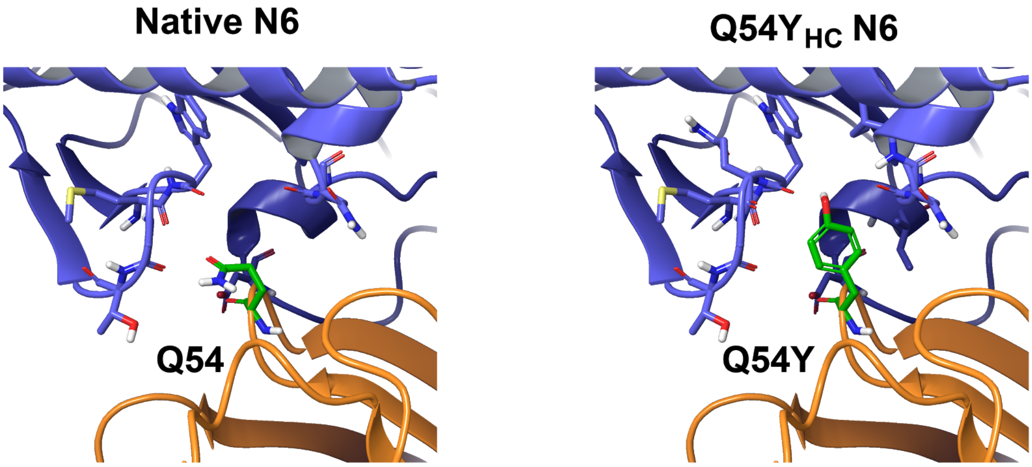


Figure 61. Three-dimensional model of the $Q54Y_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

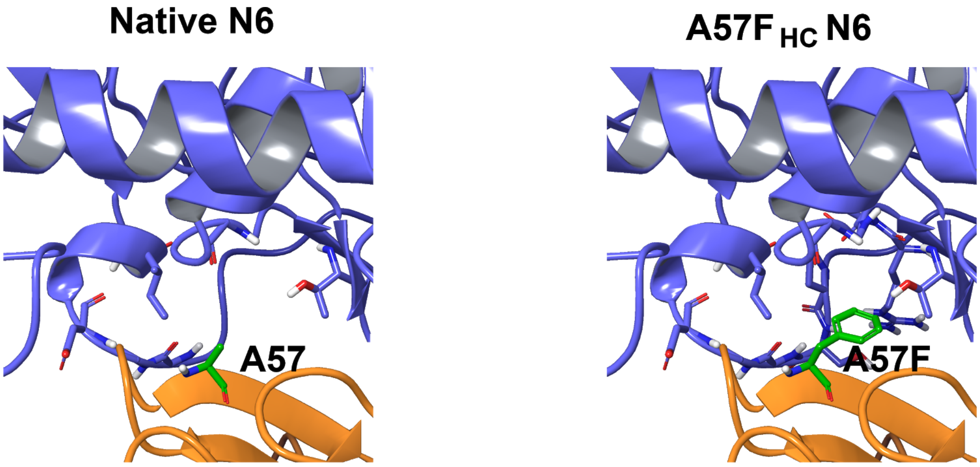
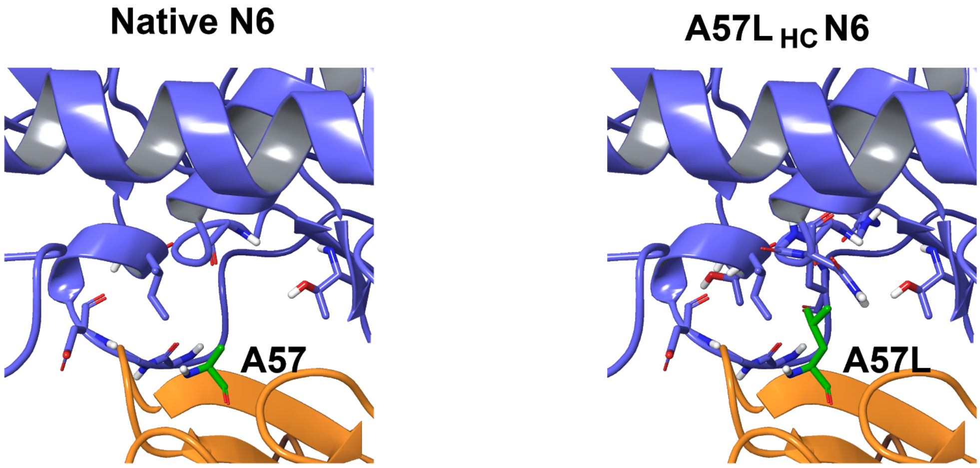


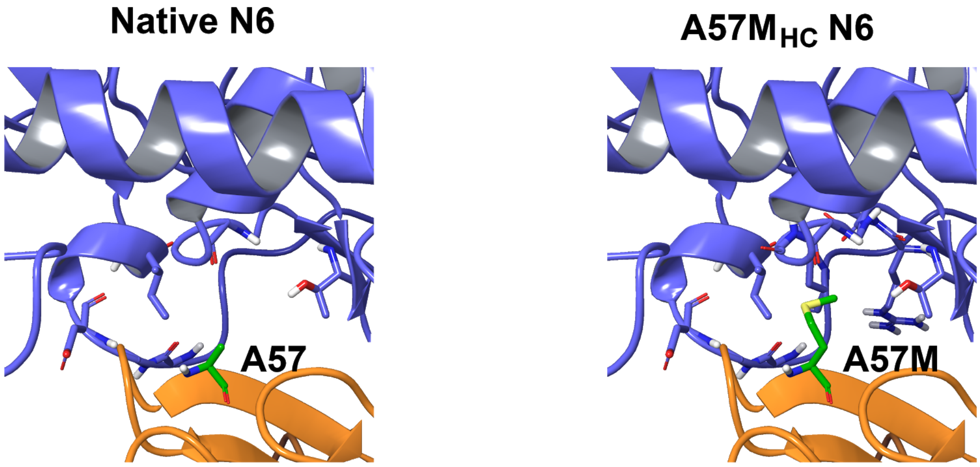
Figure 62. Three-dimensional model of the $A57F_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.



Native N6

A57L $_{HC}$ N6

Figure 63. Three-dimensional model of the A57L$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.



Native N6

A57M$_{HC}$ N6

Figure 64. Three-dimensional model of the A57M$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

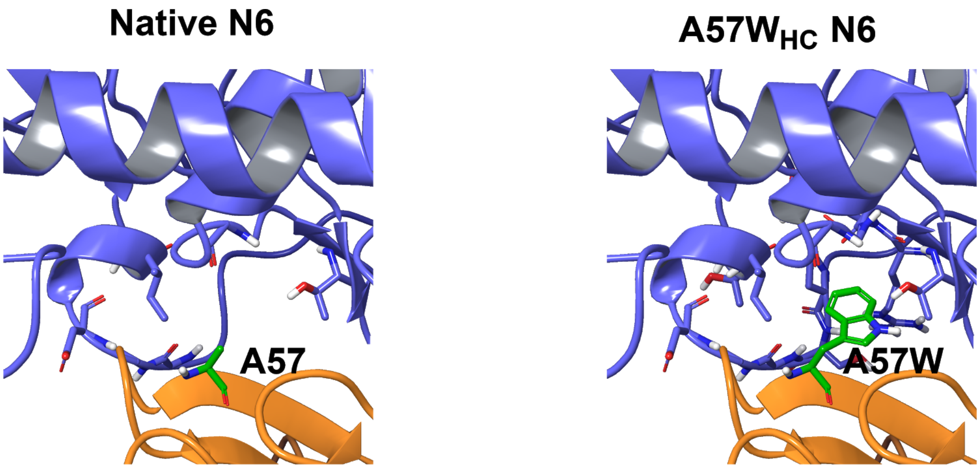


Figure 65. Three-dimensional model of the $A57W_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

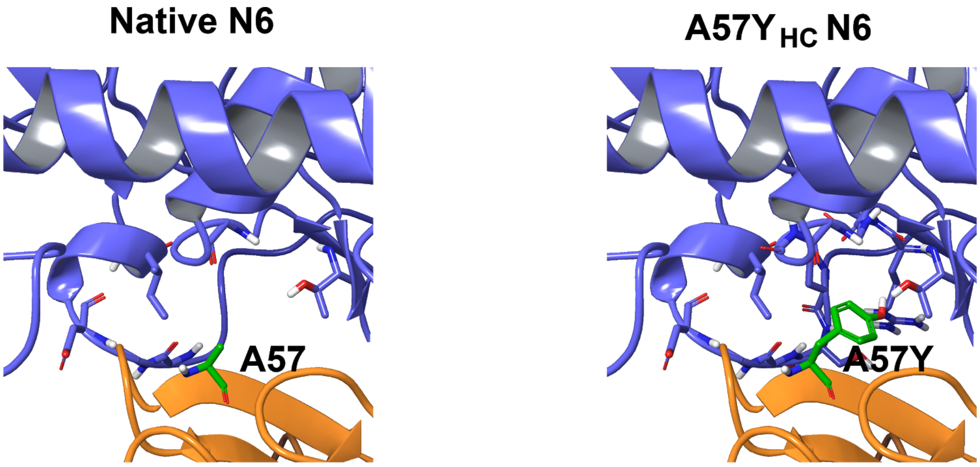


Figure 66. Three-dimensional model of the $A57Y_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

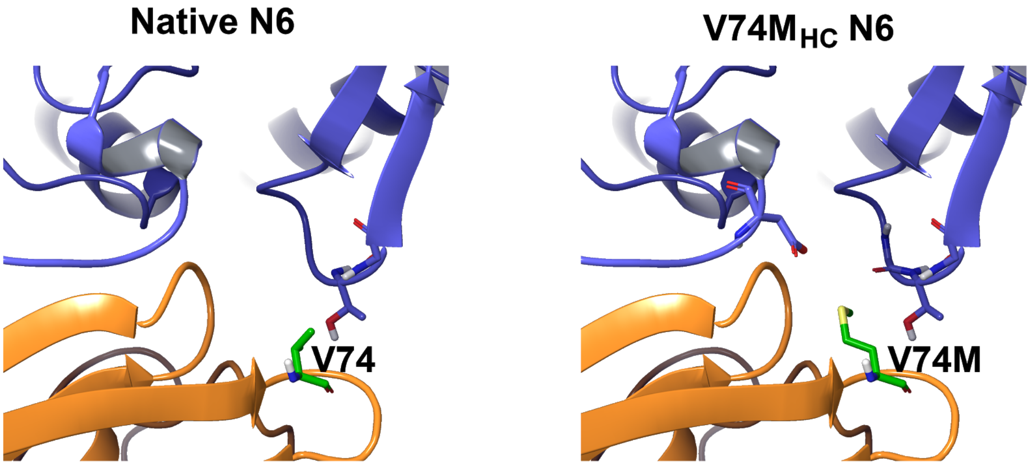


Figure 67. Three-dimensional model of the V74M$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

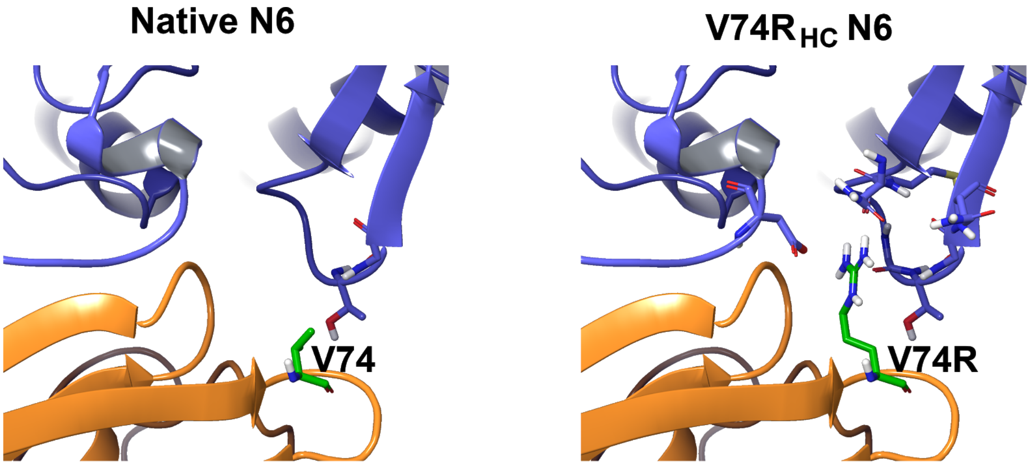


Figure 68. Three-dimensional model of the V74R$_{HC}$ mutation (mutation and heavy chain are colored in green and orange, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown.

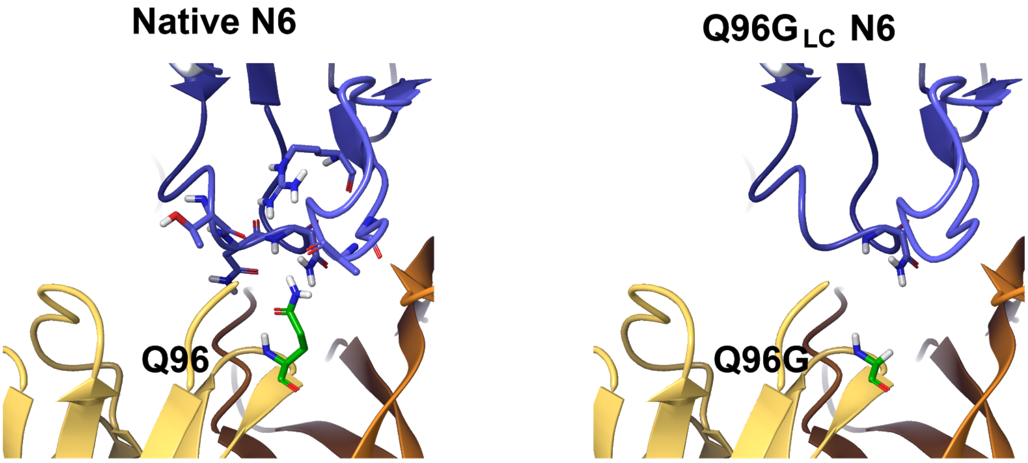


Figure 69. Three-dimensional model of the Q96G$_{LC}$ mutation (mutation and light chain are colored in green and yellow, respectively) of the N6 antibody. Intermolecular contacts with the gp120 (93TH057 as reference, colored in blue) within a radius of 5 Å are shown. Heavy chain is represented in orange.

In this section, we focused on the mutations located in the heavy chain of the antibody. Mutations on the light chain of the N6 antibody are the main topic of the following section “***Aiming to improve the N6 antibody: second round***”. In the following lines, heavy chain sequence conservation of some of the VRC01-like class antibodies is shown (N6, NIH45-46, VRC07, VRC01, N49-P7, CH235, CH235.12, b12, NC-Cow1 and IOMA). Here, the five selected positions of the N6 heavy chain are highlighted in red (T30, K52, Q54, A57 and V74).

Here, we will compare the proposed mutations by our computational pipeline with the amino acids variants found on the VRC01-like class antibodies for such positions. As can be observed on the alignments, all selected positions on the heavy chain (T30, K42, Q54, A57 and V74) show a high variability degree among the VRC01-like class antibodies. This is beneficial for the purpose of the design, since mutating highly-conserved residues could result in a decrease of the binding potency and even hamper the stability (and production) of the antibody.

	-------------FR1--------------	CDRH1
N6	RAHLVQSGTAMKKPGASVRVSCQTSGYTF**T**	AHILF
NIH45-46	EVRLSQSGGQMKKPGESMRLSCRASGYEF**L**	NCPIN
VRC07	QVRLSQSGGQMKKPGDSMRISCRASGYEF**I**	NCPIN
VRC01	QVQLVQSGGQMKKPGESMRISCRASGYEF**I**	DCTLN
N49-P7	-ADLVQSGAVVKKPGDSVRISCEAQGYRF**P**	DYIIH
CH235	QVQLVQSGAEVKKPGASVTVSCQASGYTF**T**	NYYVH
CH235.12	QVRLAQYGGGVKRLGATMTLSCVASGYTF**N**	DYYIH
b12	QVQLVQSGAEVKKPGASVKVSCQASGYRF**S**	NFVIH
NC-Cow1	QVQLRESGPSLMKPSQTLSLTCTVSGSSL**N**	DKSVG
IOMA	EVQLVESGAQVKKPGASVTVSCTASGYKF**T**	GYHMH

	------FR2-----	------CDRH2------
N6	WFRQAPGRGLEWVG	WI**K**P**Q**YG**A**VNFGGGFRD
NIH45-46	WIRLAPGRRPEWMG	WL**K**P**R**GG**A**VNYARKFQG
VRC07	WIRLAPGKRPEWMG	WM**K**P**R**GG**A**VSYARQLQG
VRC01	WIRLAPGKRPEWMG	WL**K**P**R**GG**A**VNYARPLQG
N49-P7	WIRRAPGQGPEWMG	WM**N**P**M**GG**Q**VNIPWKFQG
CH235	WVRQAPGQGLQLMG	WI**D**P**S**WG**R**TNYAQNFQG
CH235.12	WVRQAPGQGFELLG	YI**D**P**A**NG**R**PDYAGALRE
b12	WVRQAPGQRFEWMG	WI**N**P**Y**NG**N**KEFSAKFQD
NC-Cow1	WVRQAPGKALQWLG	SV**D**T**S**GN**T**-DYNPGLKS
IOMA	WVRQAPGRGLEWMG	WI**N**P**F**RG**A**VKYPQNFRG

	--------------FR3---------------
N6	RVTLTRD**V**YREIAYMDIRGLKPDDTAVYYCAR
NIH45-46	RVTMTRD**V**YSDTAFLELRSLTSDDTAVYFCTR
VRC07	RVTMTRD**M**YSETAFLELRSLTSDDTAVYFCTR
VRC01	RVTMTRD**V**YSDTAFLELRSLTVDDTAVYFCTR
N49-P7	RVSMTRD**T**SIETAFLDLRGLKSDDTAVYYCVR
CH235	RITMTRD**T**STSTVYMEMRSLRSEDTAVYYCAR
CH235.12	RLSFYRD**K**SMETLYMDLRSLRYDDTAMYYCVR
b12	RVTFTAD**T**SANTAYMELRSLRSADTAVYYCAR
NC-Cow1	RLSITKD**N**SKSRISLTVTGMTTEDSATYYCIT
IOMA	RVSMTRD**T**SMEIFYMELSRLTSDDTAVYYCAR

	-----------CDRH3---------	----FR4----
N6	DRSYGD------------SSWALDA	WGQGTTVVVSA
NIH45-46	GKYCTARDYY---------NWDFEH	WGRGAPVTVSS
VRC07	GKYCTARDYY---------NWDFEH	WGQGTPVTVSS
VRC01	GKNCD----Y---------NWDFEH	WGRGTPVIVSS
N49-P7	DRSNGSGKRFESS------NWFLDL	WGRGTAVTIQS
CH235	NVATEGS------------LLHYDY	WGQGTLVTVSA
CH235.12	NVGTAGS------------LLHYDH	WGSGSPVIVSS
b12	VGPYSWDDSPQD-------NYYMDV	WGKGTTVIVSS
NC-Cow1	AHQKTNKKECPEDYTYNPRCPQQYG	WSDCDCMGDRF
IOMA	EMFDSSADW--------SPWRGMVA	WGQGTLVTVSS

Of great interest is that 6 selected mutations from the 20 located on the heavy chain of the N6 antibody are present on some VRC01-like class antibodies, including T30I, T30L, Q54M, Q54R, Q54Y and V74M. Moreover, 8 of the 14 remaining single amino acid variants on the VRC01-like class antibodies share physicochemical properties with some selected mutations, despite not being exactly the same mutation. This is the case of the following selected variants: T30V, T30F, T30Y and T30W (I30 and L30 variants), K52R (K52 variant), Q54K (R54 variant), Q54W (F54 and Y54 variants) and V74R (K74 variant).

Overall, 14 out of 20 proposed variants on the heavy chain of the N6 antibody are conserved along the VRC01-like class antibodies. This indicates that such variants should be somehow beneficial or at least should not disrupt the binding affinity of the antibody with most of the gp120 strains; otherwise they would not be selected for potent bNAbs in nature. Moreover, this overlapping effect also indicates the robustness of the unanimous decision on the predictions of UEP, pyDock and FoldX, for its ability to screen an amount of 700 mutation possibilities on all 172 antibody-gp120 complexes, and select mutations already found in multiple potent bNAbs (that somehow should, in principle, work).

Aiming to further refine our selection process and to reduce the amount of candidates to evaluate experimentally, we ranked the designs using PELE simulations. Hence, we followed a similar pipeline as described in the previous section "***On the design of a chimera bNAb***". Here, the three-dimensional structure of the 93TH057 gp120 strain was docked together with the N6 designs in the same conformation as described on the crystal structure of the docked complex (PDB: 5TE6). Then, PELE simulations were initiated for every antibody design system, allowing very small rotations and translations to the gp120. In this way, we simulated how the N6 mutations accommodate the gp120 by mainly performing side chain and ANM backbone predictions. The analysis of the effects of the mutations was performed based on the SASA determinations, as described in the section "***Predicting antibody binding efficacy: PELE-antibody***". Changes in SASA exerted by all 21 mutations of the N6 are illustrated in **Figure 70**.

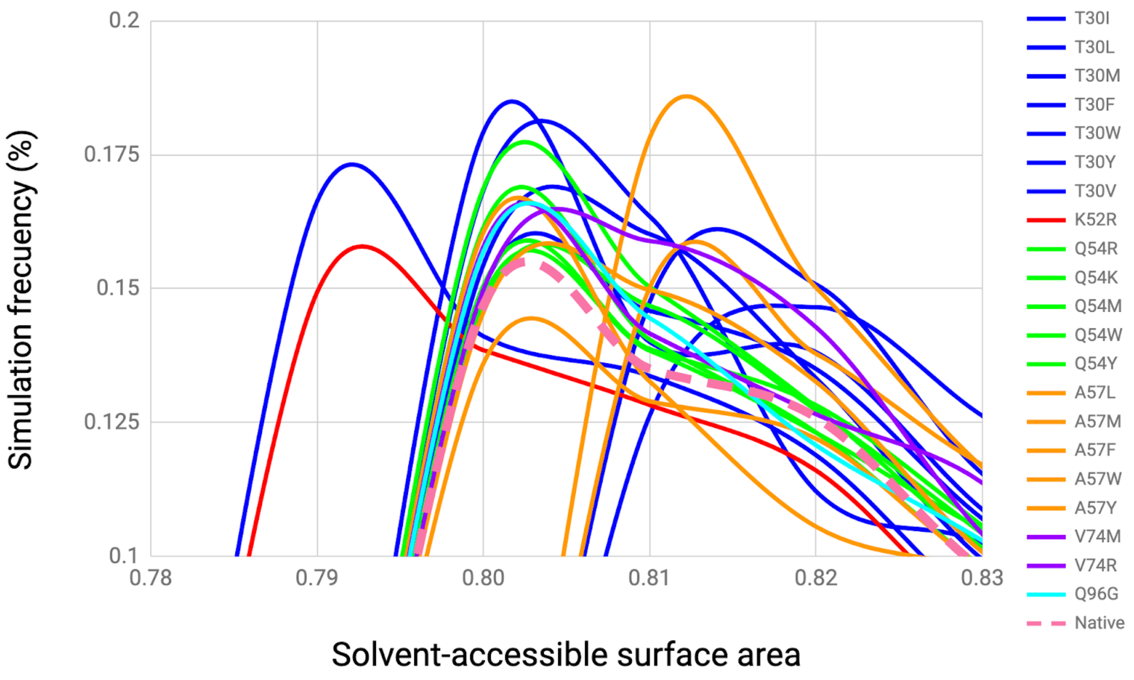


Figure 70. Simulation frequency of the solvent-accessible surface area determinations from PELE simulations in all N6 single mutant variants. Native N6 simulation is shown as a pink-dashed line. Mutations decreasing the amount of contacts along the PELE simulations compared to the native N6 antibody were discarded for experimental characterization.

Here, mutations affecting the same position are shown using the same color: T30 (blue), K52 (red), Q54 (green), A57 (orange), V74 (purple), Q96 (cyan) and native N6 antibody (pink dashed). As observed in the PELE simulations, only two of the N6 designs produce deeper contacts (SASA=0.79) compared to the native antibody (SASA=0.80): T30I and K52R. Most of the designs achieved the same unit of SASA than the native conformation (SASA=0.80) but showed higher simulation frequency for deeper units of SASA: T30L, T30M, Q54R, T30F, Q54W, A57Y, A57W and Q96G, among others. On the other hand, only five designs reached higher SASA values compared to the native antibody. Here, A57L particularly drew our attention since it shows a very well defined frequency peak at larger SASA than the native antibody (SASA=0.81).

By taking everything together, we proposed a set of 8 different mutations on the N6 heavy chain for their experimental characterization. All antibody designs were experimentally evaluated on different gp120 strains than the ones used during the prediction process: NL-43, BaL, AC-10, SVBP-12 and B41. The experimental characterization was performed under two different setups: a neutralization assay (EC50, ng/mL) using NL-43, BaL, AC-10 and SVBP-12, and an ELISA binding assay (EC50, ng/mL) using BaL monomer and the B41 trimer envelope glycoprotein. The experimental characterization of the antibody variants is illustrated in **Table 32**. Here, the color code indicates if the design achieves better binding determinations (no matter the amount of increase) than the native (green), a minimum loss (less than the double of the native antibody) (yellow) and large loss (higher than the double of the native antibody) (red). N.b indicates no binding data.

Several conclusions can be extracted from the results of the experimental characterization. The most important conclusion is that none of the designs lead to a significant increase in binding affinity compared to the native antibody. Despite that, Q54R design slightly improved the binding affinity towards half of the evaluated envelope glycoproteins (NL-42, BaL and the trimer B41) without compromising the binding affinity with the other viral isolates. Other designs, including T30I and K52R only resulted in improving the binding affinity towards a single viral isolate, without compromising the binding affinity of the other viral isolates.

On the other hand, T30M, Q54W and A57L decreased the binding affinity towards most viral isolates compared to the native antibody. Interestingly, A57Y and A57W increased the binding affinity of half of the neutralization assays for the gp120 monomers but completely disrupted the binding affinity towards the B41 trimer. In fact, all mutations located in position 57 hampered the binding affinity towards the B41 trimer, while mutations at position 54 improved it.

	Neutralization EC50 (ng/mL)				ELISA EC50 (ng/mL)	
	NL-43	**BaL**	**AC-10**	**SVBP-12**	**BaL**	**trimer B41**
N6	82,6	78,8	614,9	209,6	2,0	7,2
T30I	77,9	80,6	689,9	256,0	3,2	12,3
T30M	314,8	291,4	1558,0	785,1	2,7	8,3
K52R	102,1	100,3	871,1	166,8	3,4	13,0
Q54W	233,7	208,6	2763,9	662,9	4,9	5,5
Q54R	71,6	50,4	681,8	268,7	3,6	7,0
A57Y	144,8	60,2	N.b.	209,0	4,3	48,8
A57L	215,1	84,3	4418,9	437,4	3,0	195,3
A57W	108,2	64,1	N.b.	194,4	3,6	59,1

Table 32. Experimental determination (ng/mL) of the 8 evaluated N6 designs using neutralization and ELISA binding assays. Neutralization assays were performed on the gp120 monomers of NL-43, BaL, AC-10 and SVBP-12. ELISA binding studies were performed on a gp120 monomer (BaL) and using a gp120 trimer (B41). Color code represents changes in binding affinity compared to the native antibody. Here, green color indicates a gain in binding affinity, yellow represents a loss in binding affinity lower than the double of the native determination, and red indicates a loss in binding affinity higher than the double of the native determination. N.b indicates no binding data.

Overall, these results demonstrate the challenges of improving a potent bNAb, such as N6. For the improvement of this bNAb, the antibody variants must increase the binding affinity towards multiple viral isolates at the same time, and therefore, they can not be designed strain-specifically. In our case, none of the experimentally evaluated viral isolates was used during the prediction process, which also adds another layer of complexity for the success of the project.

Interestingly, Q54R mutation led to an slightly increase in antibody potency for half of the gp120 monomers (NL-43 and BaL), and towards the B41 gp120 trimer, without compromising the binding affinity of the other gp120 monomers (AC-10 and SVBP-12). However, the lack of experimental characterization of this design together with the fact that the improvement is modest impossibilities the proper evaluation of the impact of this mutation.

On the other hand, we have observed that neutralization and ELISA binding assays of the BaL viral isolate do not correlate. Regression analysis of both binding determinations achieved a correlation of $R^2=0.001$. According to the literature, it has been described that both experimental techniques show a certain degree of correlation (Holzmann *et al.*, 1996). In this sense, it is unclear which is the best determination technique and actually if the real experimental determinations are similar to the ones presented in this thesis.

Aiming to improve the N6 antibody: second round

After the herculean modeling effort described in the previous section, we were not able to increase the N6 binding potency experimentally. During the second round of design we followed a completely different approach than the one described previously. In this case, we decided to take a step back and follow a completely rational design protocol motivated by an extensive literature search together with previous observations during the first round of design.

During the first round, an amount of 21 mutations were selected from the unanimous decision of the predictions made by UEP, FoldX and pyDock. Only one mutation, Q96G, was selected on the light chain of the antibody (**Figure 69**) and also was predicted to achieve higher deeper contact frequencies on the PELE simulations (**Figure 70**). This mutation particularly drew our attention, since it was the only one decreasing the amino acid size compared to the natie residue (from a large glutamine residue to glycine, the shortest residue). At this point, we decided to continue exploring such phenomena, and we found out a crucial hydrogen bond between the CDRH3 and the loop D of the gp120 takes place in the vicinity of the Q96 position. This key hydrogen bond is formed between W100c$_{HC}$ and N/D279$_{GP120}$, and the W100c$_{HC}$ residue is conserved in VRC01-like class antibodies such as N6, VRC01, VRC03, NIH45-46 and N49-P7.

Regarding this intermolecular interaction, it has been already described that resistant strains towards the N6 antibody contain single point mutations at the position 279 that impede the formation of the hydrogen bond. In fact, mutagenesis studies revealed this phenomena (Huang *et al.*, 2016). It seems that the antibody binds to strains containing an asparagine or an aspartic acid in this position, but that it cannot bind to strains showing a glutamine (BL01.DG resistant strain), glutamic acid (TV1.29 resistant strain), arginine (6471.V1.C16 resistant strain) either an alanine (T278-50 resistant strain). From this observation, we followed the hypothesis that if the disruption of this intermolecular contact largely decreases the binding efficacy, its improvement may result in a potent antibody variant.

To date, N6 has been co-crystallized with only three different gp120 strains: 93TH057 (PDB: 5TE6, clade AE), X2088 (PDB: 5TE4, clade G) and DU172.17

(PDB: 5TE7, clade C). After inspection of the above mentioned crystal structures, we observed that the hydrogen bond between W100c$_{HC}$ and N279$_{GP120}$ is conserved on the three bound structures (**Figure 71**).

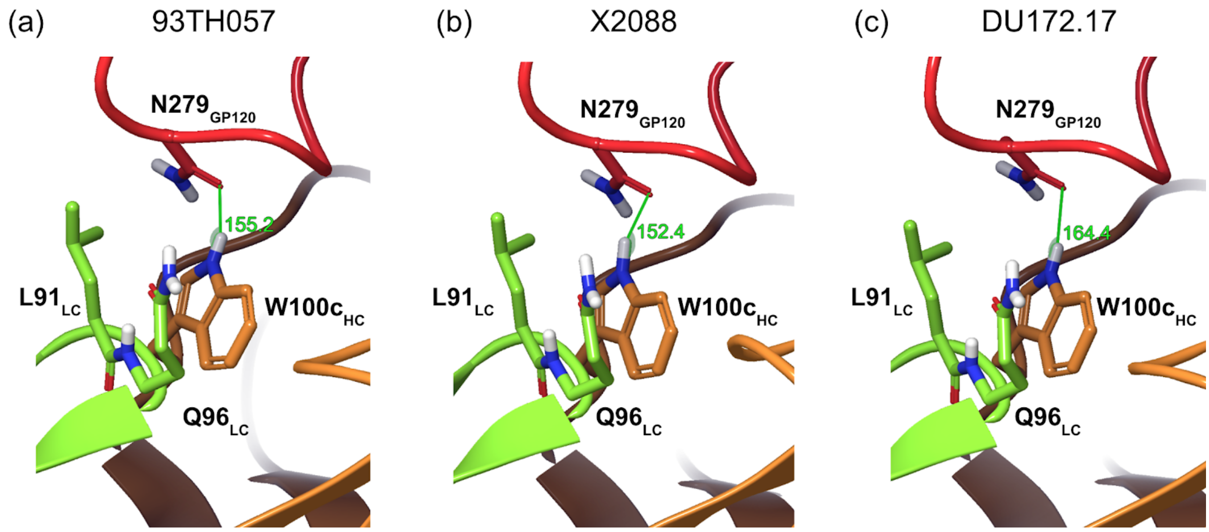


Figure 71. Conserved hydrogen bond between W100c$_{HC}$ and N279$_{GP120}$ in the three available crystal structures, (a) 93TH057 (5TE6), (b) X2088 (5TE4) and (c) DU172.17 (5TE7). Hydrogen bond angle is shown. Closest residues to his interaction from the CDRL3 (L91$_{LC}$ and Q96$_{LC}$) are highlighted.

An hydrogen bond is an interaction mainly dominated by electrostatic forces, where the distance and the orientation angle of the involved atoms play crucial roles. An illustration of the importance of the angle is represented in **Figure 72**. **Figure 72a** represents the donor (N-H, bound covalently by a σ bond) and the acceptor (O) groups of an hydrogen bond, and the angle (θ) involving the hydrogen bond atoms and the covalent σ bond. **Figure 72b** represents the distribution of hydrogen bond angles in high-resolution crystallographic structures of proteins and nucleic acids. Hydrogen bond distance and angle of the involved atoms are responsible for the strength of the interaction. Hence, strong hydrogen bonds show angles close between 170 and 180 degrees, and in general, the lower the hydrogen bond angle the weaker the interaction.

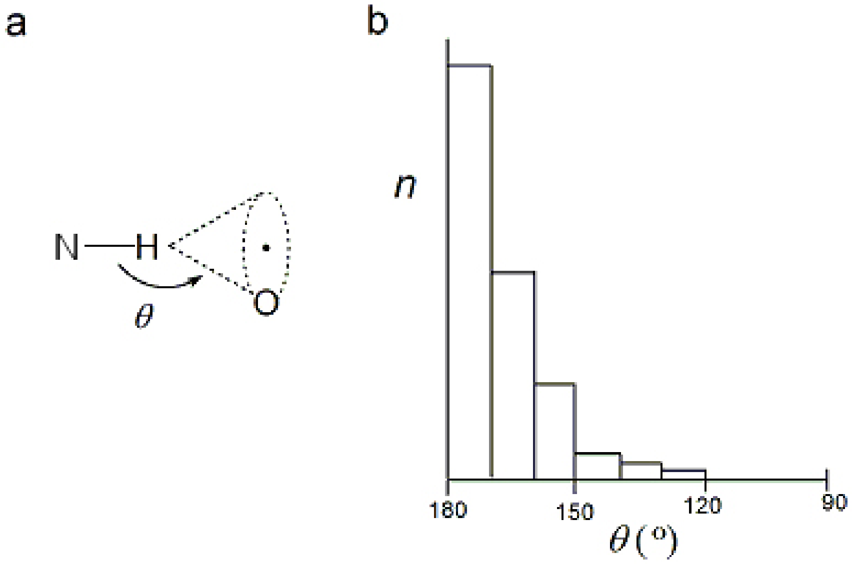


Figure 72. Geometry of a hydrogen bond. (a) represents the atoms involved in a hydrogen bond (donor and acceptor groups), and the θ angle they form. (b) represents the distribution of hydrogen bond angles in high resolution crystal structures of proteins and acid nucleics.

As detailed in **Figure 71**, hydrogen bond angles between W100c$_{HC}$ and N279$_{GP120}$ are 155.2º, 152.4º and 164.4º for 93TH057, X2088 and DU172.17, respectively. As mentioned previously, strong hydrogen bonds have angles between 170º and 180º, and the observed angles for N6 are not even close to this range. In this context, our goal was to improve this orientation angle aiming to enhance the strength of this key hydrogen bond. At this point, we particularly focused on the residues in the vicinity of W100c$_{HC}$, and concretely on two residues of the CDRL3: L91$_{LC}$ and Q96$_{LC}$. The side chain of those residues make contacts with the loop D, and Q96$_{LC}$ has been previously described to generate an important water-mediated hydrogen bond with the gp120 (Huang *et al.*, 2016).

However, and from our modelling experience, we hypothesized that both positions on the CDRL3 (L91$_{LC}$ and Q96$_{LC}$) should not be important for binding. And actually, after the visual inspection of the crystal available structures (**Figure 71**) we believed that the presence of their two large side chains could impede an efficient conformation of the loop D that ultimately would result in the observed "bad" hydrogen bond angles. Following this hypothesis, we aimed to generate antibody variants on those CDRL3 positions showing a reduction in amino acid size in those positions while keeping the physicochemical properties of the native amino acids. In this sense, we mutated L91$_{LC}$ to an alanine and Q96$_{LC}$ to a serine. We expected that by generating space in the CDRL3 region, the loop D could be better accommodated, allowing for a stronger hydrogen bond between W100c$_{HC}$ and N279$_{GP120}$ that ultimately would lead to an general increase in binding affinity for all viral isolates (**Figure 73**).

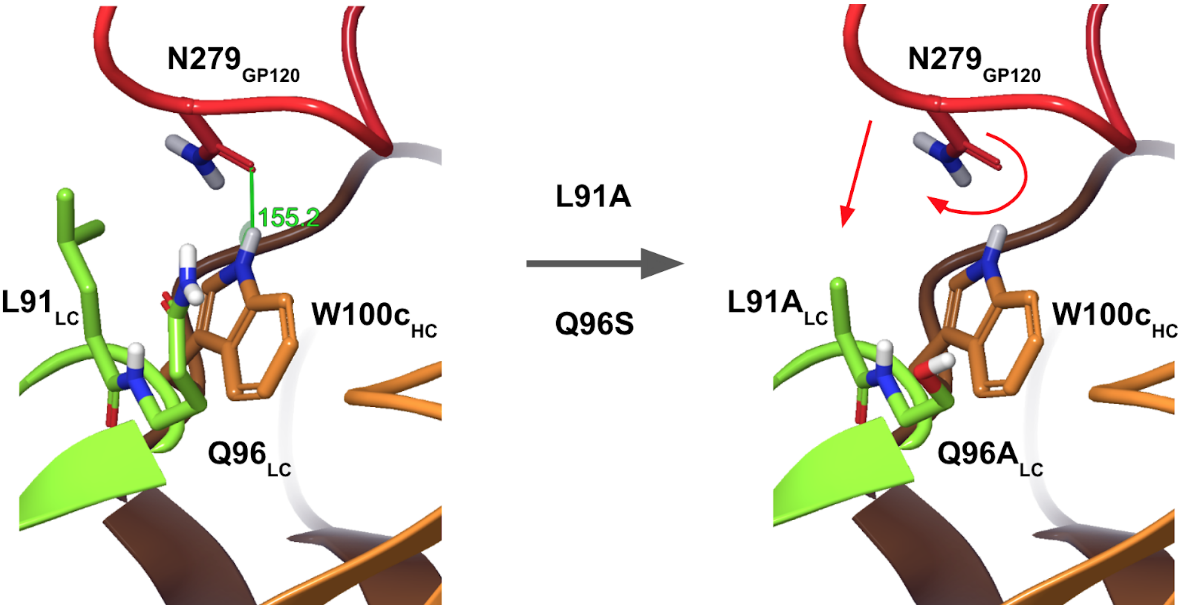


Figure 73. Double-mutation effect consisting of L91A, and Q96S on the CDRL3 loop of the N6 antibody. Removal of those large side chains would result in a better accommodation of the loop D, allowing for a stronger hydrogen bond between W100c$_{HC}$ and N279$_{GP120}$.

Aiming to estimate the effects of the double mutation on the dynamics of the loop D, we performed MD simulations of the full N6-gp120 bound conformation. Our template structure consisted of the PDB: 5TE6, where the co-crystal structure of the N6 together with the 93TH057 gp120 strain was experimentally solved. We used a MD platform generated in the group, which was recently published for estimating binding free energies (Gilabert *et al.*, 2020). This platform uses OpenMM (Eastman *et al.*, 2017) for automatizing the preparation, parametrization and equilibration of the system. Moreover, the platform uses the Amber14 (Maier *et al.*, 2015) force field for parameterizing the protein and GAFF (Wang *et al.*, 2004) using the Ambertools suite (Wang *et al.*, 2006) for the ligands. After the parameterization, the system was solvated using a cubic box with TIP3P (Jorgensen *et al.*, 1983) waters of 16 Å, and ions were also added for neutralizing it. Then, a minimization process consisting of 2000 steps using the L-BFGS optimization algorithm implemented by OpenMM was performed. Regarding the equilibration process, it consisted on a NVT simulation of 400 ps at 300 K with restrains to the heavy atoms of the protein (of 5 kcal mol^{-1} Å^{-2}), followed by 4 ns of NPT equilibration at 300 K and 1 atm with restraints to the protein alpha carbons (of 0.5 kcal mol^{-1} Å^{-2}). A time step of 2 fs and a nonbonded interaction cutoff of 9 Å were used. After the equilibration process, a production phase of 300 ns was simulated.

After the generation of the MD simulation, we analyzed the mobility and the structural fluctuation of the loop D along using MDLovoFit (Martínez, 2015). A comparison of the structural fluctuation of the loop D between the native and the L91;Q96 mutant is illustrated in **Figure 74**.

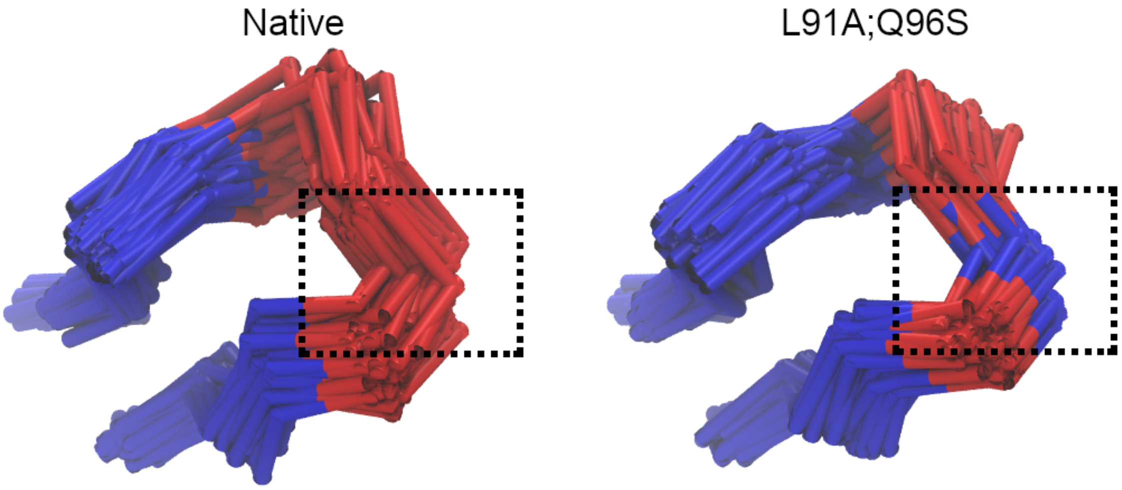


Figure 74. MDLovoFit structural fluctuations of the loop D along the MD simulations for the native and double mutant design of N6-gp120 complex. Blue represents rigid regions, and red represents mobile sections. Dashed-square points out the position N279$_{GP120}$ of the loop D, which is forming a key hydrogen bond with W100c$_{HC}$.

Here, blue color represents rigid regions while red color represents mobile regions of the loop D. The dashed-square indicates the position of N279$_{GP120}$. As can be observed, N279$_{GP120}$ remains more rigid on the simulations for the L91A;Q96S mutant than for the native antibody. Moreover, adjacent residues to N279$_{GP120}$ also show a lesser degree of mobility for the L91;Q96 mutant compared to the native complex; it can be observed that the dispersion of the backbone adjacent residues to N279$_{GP120}$ is higher for the native than for the L91A;Q96S mutant.

Next, we analyzed the distribution of the hydrogen bond angles between W100c$_{HC}$ and N279$_{GP120}$ along the MD simulation. The distribution of the hydrogen bond angles (in ranges of 10º, starting from 90º-100º to 170º-180º) is illustrated in **Figure 75**. To ease the representation, the maximum angle value for each range is used at the horizontal axis (100º and 180º for both previous examples, respectively). Moreover, only MD frames showing a hydrogen bond distance lower than 2.5 Å were taken into account for the distribution representation. As can be observed, the L91A;Q96S mutant shows a higher amount of MD frames in the optimal angle range (170º-180º) to generate a strong interaction.

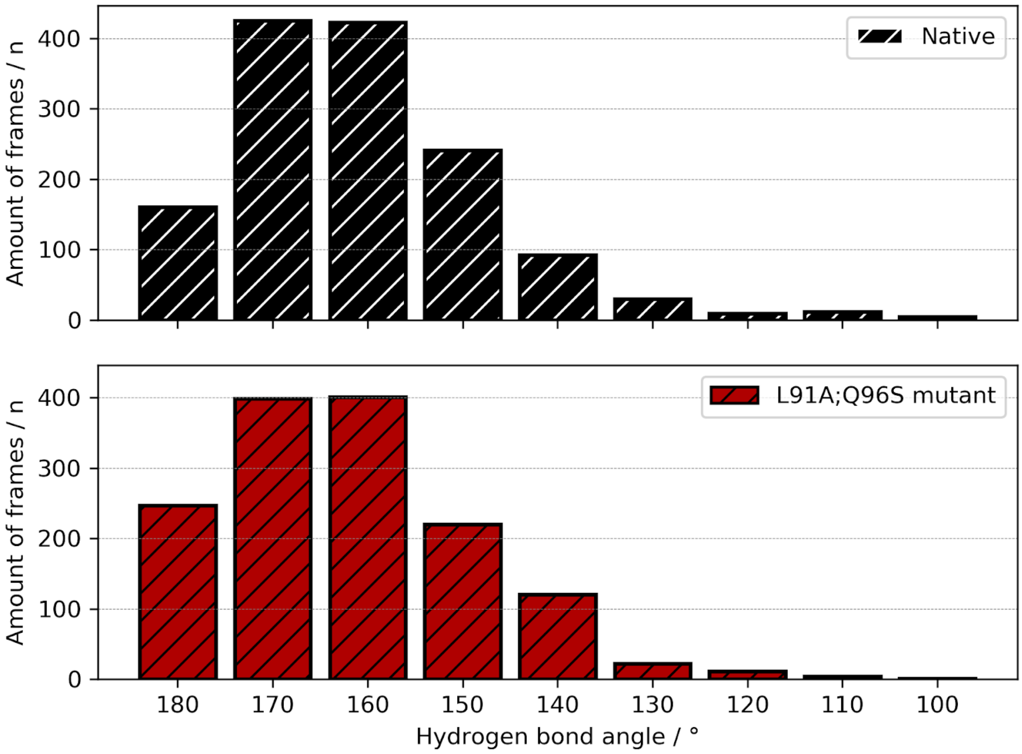


Figure 75. Distribution of the hydrogen bond angles between W100c$_{HC}$ and N279$_{GP120}$ along the MD simulations for the native and the double mutant. Only MD frames with an hydrogen bond distance < 2.5 Å were used. Data was grouped in range of 10º, starting from 90º-100º (right) to 170º-180º (left). Large amounts of frames were found in the optimal angle range (170º-180º) for the L91A;Q96S mutant compared to the native antibody-gp120 complex.

Then, the simulation data was normalized into frequencies of observing hydrogen bond angles within ranges of 5º (starting from 120º-125º to 175º-180º). Here, we show such frequency distribution for hydrogen bond distances less than 2.5 Å (**Figure 76**) and less than 1.8 Å (**Figure 77**). As can be observed, the L91A;Q96S mutation induces substantially better hydrogen bond angles between W100c$_{HC}$ and N279$_{GP120}$ along the MD simulations.

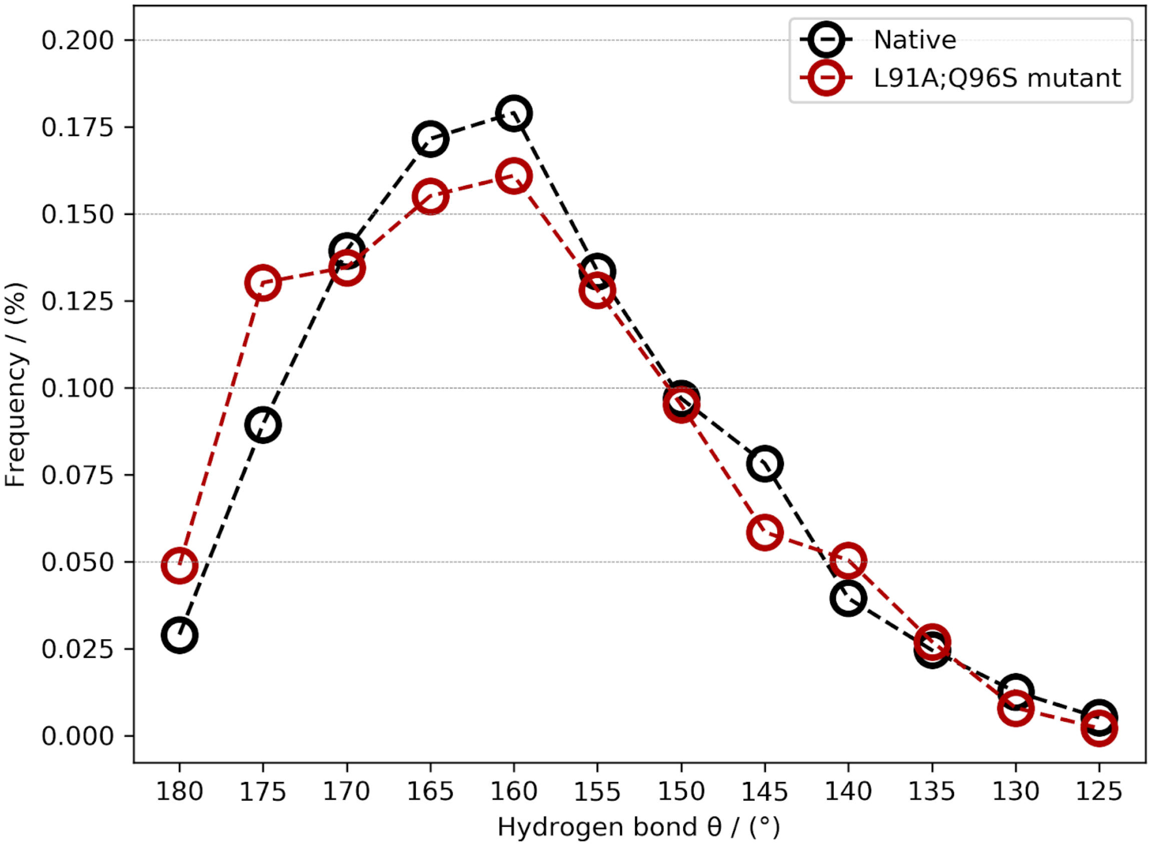


Figure 76. Frequency distribution of the hydrogen bond angle between W100c$_{HC}$ and N279$_{GP120}$ along the MD simulations for the native and the double mutant. Only MD frames with an hydrogen bond distance < 2.5 Å were used. Data was grouped in range of 5º, starting from 120º-125º (right) to 175º-180º (left). As can be observed, the L91A;Q96S mutant design achieves higher frequencies of stronger hydrogen bonds than the native complex.

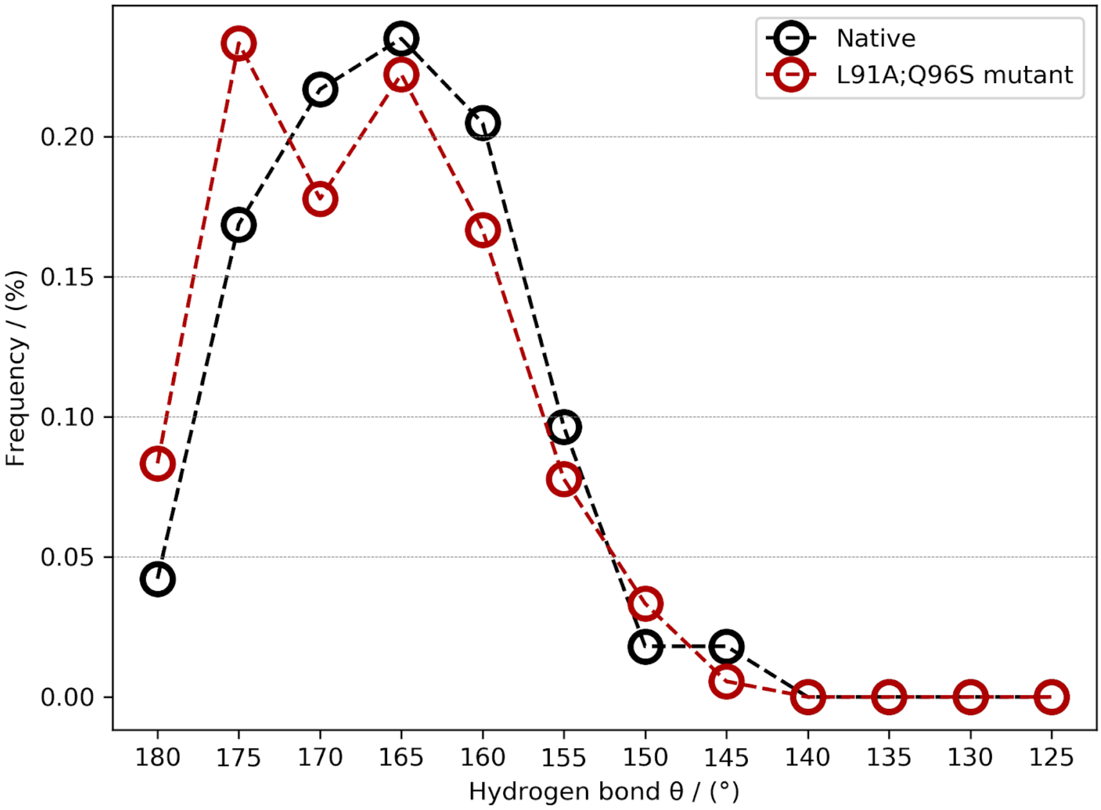
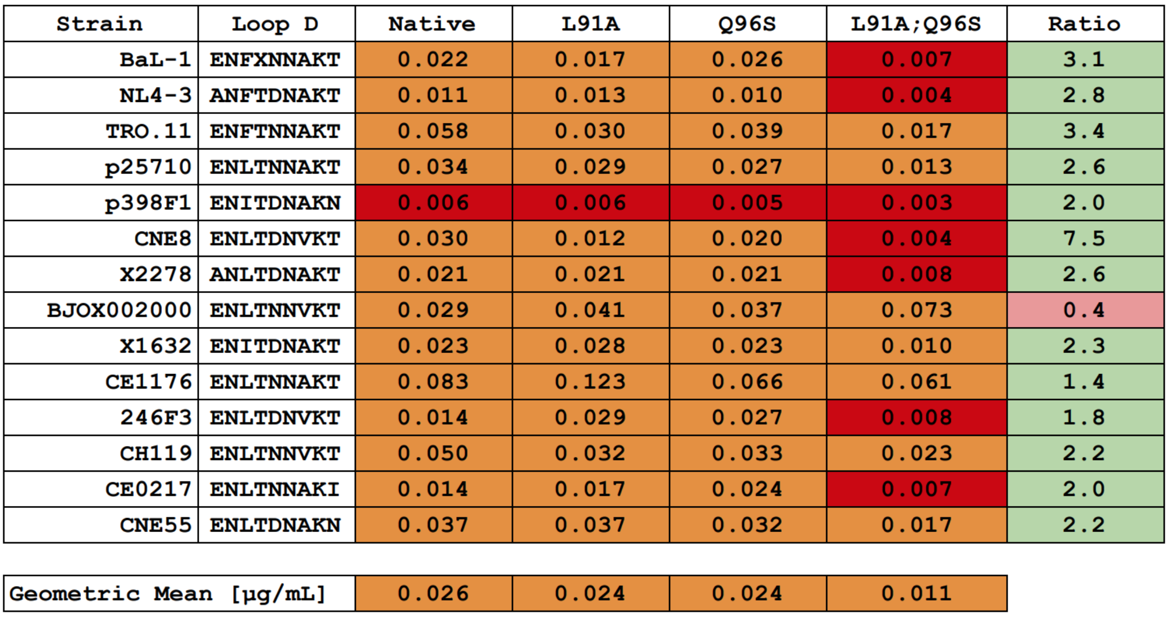


Figure 77. Frequency distribution of the hydrogen bond angle between W100c$_{HC}$ and N279$_{GP120}$ along the MD simulations for the native and the double mutant. Only MD frames with an hydrogen bond distance < 1.8 Å were used. Data was grouped in range of 5º, starting from 120º-125º (right) to 175º-180º (left). As can be observed, the L91A;Q96S mutant design achieves higher frequencies of stronger hydrogen bonds than the native complex.

We proposed to experimentally evaluate the single (L91A and Q96S) and the double (L91A;Q96S) N6 antibody variants to our collaborators at Irsicaixa. After antibody production and purification, the designs were evaluated in two antibody neutralization assays panels: a resistant one (consisting of 3 resistant gp120 strains) and a sensitive one (consisting of 14 sensitive gp120 strains). As expected, none of the antibody designs was able to neutralize any of the resistant strains (IC50 > 5 µg/mL). This can be explained because of the fact that those viral isolates are not able to form the hydrogen bond with the W100c$_{HC}$. However, the double mutant design showed a moderate increase in binding potency compared to the native and the single mutants (**Table 33**). Concretely, the L91A;Q96S mutant achieved a binding potency geometric mean of 0.011 µg/mL, and the native and the single mutations

L91A and Q96S achieved a binding potency geometric mean of 0.026 μg/mL, 0.024 μg/mL and 0.024 μg/mL, respectively. Comparison of the IC50 values between the native antibody and the double mutant design revealed that the design increases the binding potency towards thirteen out of fourteen gp120 sensitive strains (**Table 33** and **Figure 78**). **Figure 78** illustrates the changes in binding potency (in IC50 terms) between the native and the L91A;Q96S variant of the N6 antibody.



Strain	Loop D	Native	L91A	Q96S	L91A;Q96S	Ratio
BaL-1	ENFXNNAKT	0.022	0.017	0.026	0.007	3.1
NL4-3	ANFTDNAKT	0.011	0.013	0.010	0.004	2.8
TRO.11	ENFTNNAKT	0.058	0.030	0.039	0.017	3.4
p25710	ENLTNNAKT	0.034	0.029	0.027	0.013	2.6
p398F1	ENITDNAKN	0.006	0.006	0.005	0.003	2.0
CNE8	ENLTDNVKT	0.030	0.012	0.020	0.004	7.5
X2278	ANLTDNAKT	0.021	0.021	0.021	0.008	2.6
BJOX002000	ENLTNNVKT	0.029	0.041	0.037	0.073	0.4
X1632	ENITDNAKT	0.023	0.028	0.023	0.010	2.3
CE1176	ENLTNNAKT	0.083	0.123	0.066	0.061	1.4
246F3	ENLTDNVKT	0.014	0.029	0.027	0.008	1.8
CH119	ENLTNNVKT	0.050	0.032	0.033	0.023	2.2
CE0217	ENLTNNAKI	0.014	0.017	0.024	0.007	2.0
CNE55	ENLTDNAKN	0.037	0.037	0.032	0.017	2.2

Geometric Mean [μg/mL]	0.026	0.024	0.024	0.011

Table 33. Neutralization assays of 14 gp120 sensitive strains to N6 antibody designs, including the native form, single mutations L91A and Q96S, and double mutation L91A;Q96S. Loop D sequence for each HIV-1 strain, IC50 determinations and the experimental binding ratio between the native and the double mutant design are shown. Ratio higher than 1 indicates that the double mutant variant achieves higher binding potency than the native, and the other way indicates the opposite. As can be observed, the double mutation increases the binding potency of the N6 antibody towards 13 out of 14 gp120 sensitive strains.

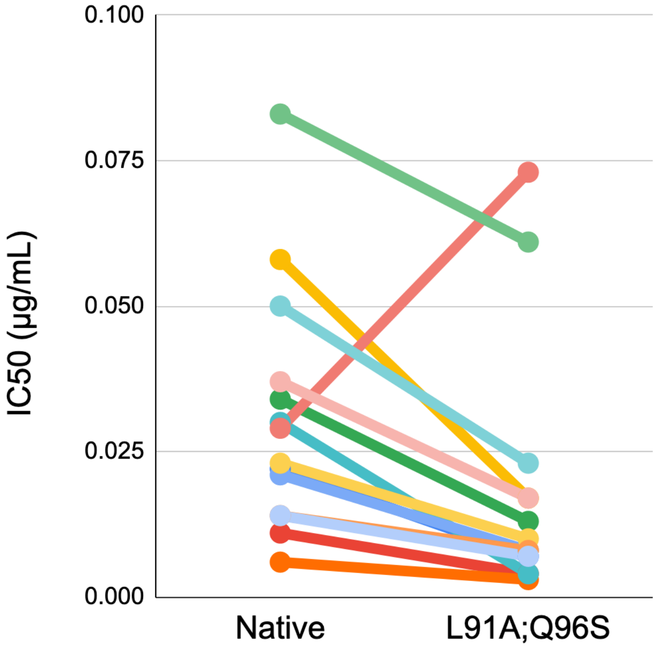


Figure 78. Changes in binding potency (IC50) between the native and the L91A;Q96S mutant of the N6 antibody for each gp120 sensitive strain. As can be observed, the double mutant design increases the binding potency towards 13 out of 14 gp120 strains.

Overall, our results demonstrate that the N6 antibody can be improved by means of computational predictions. However, the expected gain in antibody binding efficacy is low, since the binding potency of the N6 is extremely high. Moreover, it seems that several mutations are needed to observe differences in binding affinity. In this context, none of the two single mutations on the CDRH3 loop substantially increased the binding affinity compared to the native antibody (see geometric means). However, when the two mutations were performed at the same time a synergistic effect leading to an enhancement of the antibody potency was observed. In light of these events, we proposed the evaluation of a triple mutant consisting of the CDRL3 mutations L91A and Q96S together with the CDRH2 mutation Q54R. As observed during the first round of design, this mutation increased the binding affinity towards half of the envelope glycoproteins that were experimentally evaluated without disrupting the binding affinity with the others (**Table 32**). Hence, we would expect that the binding potency of this triple mutant could lead to an improved antibody variant.

Discussion

This thesis focuses on the design of protein-protein interfaces, with particular attention to the design of antibodies. The truth is that the principles that govern the binding between two proteins is not yet fully understood, probably because of the lack of experimental characterization of the effects of mutations. This is highlighted in the SKEMPI 2.0 database, which is one of the largest experimental databases of the effects of mutations. SKEMPI 2.0 contains ~5.200 different single mutations in ~300 protein-protein complexes, and most of them are mutations to alanine. One must to understand and accept that using this source of data for the design of a predictive algorithm presents serious flaws: the amount of available data is certainty low, the nature of the mutations and the diversity of the protein-protein complexes are highly heterogeneous and unbalanced and the experimental determinations show a high degree of estimation error, among others. Taking everything together, this situation originates a vicious cycle, which can lead to mistrust of structural biologists. However, our mission is to generate a predictive model accurate enough, taking into account the enormous difficulties we encounter along the way.

Algorithms positioned as the state-of-the-art aim to describe the ΔΔG upon mutation by using different approaches: physical energy descriptors, statistical methods or machine learning, among others. However, they do typically share a basic scheme: they describe features or changes originated by the mutation in the three-dimensional structure and they fit those features and the experimental ΔΔG into a mathematical equation. Hence, if we should not blindly trust the experimental ΔΔG as mentioned previously, why should we use it to generate a predictive model?

Following this concern, we aimed to design our own ΔΔG predictor by employing other sources of knowledge rather than experimental ΔΔG determinations. In this sense, we envisioned UEP, a contact-based method built from protein-protein data of Interactome3D. UEP algorithm is very simple, and it is based on the philosophy that evolution has designed protein-protein interfaces optimum enough for the maintenance of the cell homeostasis. Hence, despite that an individual protein-protein complex can be engineered leading to a stronger interaction, the analysis of all protein-protein interfaces should provide the rules behind intermolecular contacts for the design of a ΔΔG predictor.

One of the main interesting observations during this thesis (which is rather simpler) is the fact that mutations that are embedded in the highly-packed interface are the ones that exert larger differences in experimental ΔΔG, and moreover, they are easier to be correctly predicted by any predictor (**Table 8**). The same principle applies to the prediction of mutations to alanine and to other than alanine, where the latter exert larger effects in experimental ΔΔG and also are easier to be correctly predicted (**Table 8**). Hence, we believe that taking into account these effects during a protein design campaign may be a key factor for the success of the modeling effort.

Most of the state-of-the art predictors require the previous modelization of the mutation before the ΔΔG prediction, and most of them do not provide the tools for generating them. This is a problem, since the modelization step of the mutation (including for instance, the rotamer choice of the side chain) relies on the force field parameterization. Hence, estimating the energy of a protein-protein complex using a force field different from the one used for generating the mutation models may be counterproductive. In fact, we have observed that the modelization process exerts a large impact on the energy evaluation. This is clearly represented by the different energy predictions of pyDock made on the models generated by FoldX and EvoEF1, which only evaluate the electrostatics and solvation properties of the protein-protein complex. We observed that both pyDock predictions show a low similarity of 80.2% for mutations to other residues than alanine (this is, mutations predicted to improve or decrease the binding affinity compared to the native counterpart). Therefore, ideally, the same force field used for the energy prediction should be also used to select the optimal orientation of the mutation. In this context, UEP is not an usual ΔΔG predictor and it does not need the generation of mutation files since it works on the native protein-protein complex (such as BeAtMuSiC). We designed the algorithm to be fast and user friendly. Hence, UEP predicts the effects of all possible mutations in a protein-protein interface in less than a second, and returns a single CSV file with the prediction results. In this sense, other algorithms are computationally expensive and are unable to screen the entire protein-protein interface using a single command line. Usually, a run for each mutation is needed (with some exceptions), which leads to the generation of numerous files per mutation that are often not used.

UEP performance for the prediction of the effects of mutations is in line with the best state-of-the-art algorithms evaluated in our benchmark. Something that we do not explicitly state in our article is the fact that all other predictors have been pre-trained on experimental ΔΔG data to generate their scoring function. SKEMPI 2.0 contains almost all described experimental ΔΔG determinations, and it is very likely that all other predictors have been trained on the same (or partial) data than the one we evaluate in our benchmark. Aiming to represent this effect, we compared the performance of mCSM (whose authors explicitly stated which mutations were used to train the algorithm) for mutations used during the training process together with new mutations. Here, we observed a very important drop in predictive power for mCSM when comparing training and new mutation groups (**Figures 27-28, Table 8**). Interestingly, UEP performance on both groups of mutations did not result in a significant decrease, indicating that the set of new mutations does not contain particularities making their prediction challenging. Despite that the performance of all other methods is similar to the expected one in this field, a small degree of overfitting could also occur in some of them. In this way, we consider that achieving similar performance to the best algorithms in the field without using experimental ΔΔG determinations is a milestone, which indicates the usefulness of our approach.

Moving to the antibodies field, their discovery and characterization is an expensive and time-consuming process. Hence, this field could potentially take advantage of diverse computational algorithms to easen the antibody discovery and optimization process. Because of this reason, there is an interest for the prediction of antibody effectiveness towards an epitope (specially for hypermutated epitopes). This is of great importance in the HIV-1 research field, where bNAbs have been positioned as a potential therapeutic cure of the disease. HIV-1 is characterized by a short replication cycle and a very high mutation rate, which makes the virus recalcitrant to bNAb effectivity. Hence, being able to predict which is the optimal combination of bNAbs for a given patient will be of great importance for the design of efficient therapies against AIDS. The fact is that despite being of great interest there are not many protocols to predict if an antibody will bind to an hypermutated HIV-1 strain (of course, without taking into account protein-docking algorithms). Most available algorithms are based on sequence conservation, which limits their use for uncharacterized antibodies. In this sense, we envisioned a computational *ab initio* approach using PELE, a Monte Carlo algorithm, for the classification of resistant and sensitive hypermutated epitopes towards an antibody (**Figures 34 and 35**). Our protocol consists of evaluating the three-dimensional binding process between the epitope and the antibody, which is used to estimate a relative binding affinity (**Equation 15**). One of the main problems we encountered during this project belonged to the fact that resistant gp120 strains are poorly three-dimensionally described. Hence, we needed to perform homology models of all evaluated strains for the PELE simulations. From the predictive point of view this can be problematic, since the gp120 epitopes are highly hypermutated showing large variable loops (that are difficult to model). After bypassing this limitation by homology modeling techniques, our approach consisted in simulating the binding process from out-to-in positions; we placed the gp120 epitopes separated by 10 Å in the vertical axis from the docked position with the antibody, and we then simulated how the attachment process takes place.

Our results showed that the binding process of resistant and sensitive strains behave differently. Concretely, resistant strains do not accommodate into the antibody binding region as much as the sensitive ones. Hence, by analyzing the contacts that are being formed between the epitope and the antibody, the binding process can be mapped into a distribution pattern (**Figures 36**). In this sense, resistant strains can be easily determined, since they do not tend to generate a high amount of contacts with the antibody compared to the sensitive ones with an AUC of 0.84 on the VRC01 benchmark (**Figures 37 and 38**). One thing that must be considered is that those simulations can be only used for comparative studies: the estimation (resistant/sensitive) of an unknown epitope will be based on the similarity of its contacts distribution with known benchmarked epitopes. Hence, for the prediction of other systems, a comparative study or benchmarking should be performed prior to the evaluation of unknown effects. In fact, this is how many simulation processes work and should not represent an impediment for the use of this technology in other protein-protein scenarios. Aiming to extrapolate our method to other CD4bs bNAbs (NIH45-46 and 3BNC117), we conducted an approximation of the optimum contacts threshold for new antibodies that resulted into accurate antibody binding predictions (**Figure 40**). These results demonstrate that the antibody binding affinity towards an epitope can be simulated by means of MC techniques.

Regarding the antibody design field, one of the main goals of this thesis consists on improving the binding potency of an antibody. We have been collaborating with experimentalist researchers at IrsiCaixa in multiple different projects, and some of them have been included in this thesis. During this process, we have encountered many challenges for improving the efficacy of the N6 bNAb. The fact is that N6 antibody is the best bNAb in the market against HIV-1, and it is extremely efficient on its own. Probably, larger changes in experimental binding affinity could have been obtained if we worked on less potent antibodies. Nonetheless, we have discussed three projects aiming to improve the efficacy of the N6 antibody: a chimera bNAb case and two rounds of antibody design.

Regarding the chimera bNAb case, we envisioned an antibody variant of the N6 antibody by taking advantage of previous experimental determinations found in the literature. It is known that a four-residue insertion on the CDRH3 present in some VRC01-like class antibodies is responsible for an increase in binding potency. Moreover, it is also known that the CDR and FR regions of the VRC01-like class antibodies are exchangeable, in the sense that they can be produced and evaluated experimentally. This insertion on the CDRH3 loop is not found on the N6 design, and we envisioned that enlarging this loop could be beneficial for the binding potency, as previously described for NIH45-46 and VRC07 compared to VRC01. Aiming to do so, we designed an antibody variant of the N6 antibody by exchanging its CDRH1 and CDRH3 regions for the ones located in NIH45-46 antibody. It is important to exchange both regions at the same time, since both loops are covalently bound by a disulphide bond that stabilizes the insertion. After the generation of the three-dimensional variant of this antibody, we evaluated the binding potency by using PELE simulations. We observed that the insertion resulted in an increase in predicted binding affinity (kcal/mol) against the 93TH057 gp120 strain. Unfortunately, the antibody could not be produced in the laboratory and the experimental binding efficacy could not be assessed.

Next, we conducted the first round of design on N6 by means of *in silico* mutagenesis studies. Here, we manually recopilated all possible gp120 sequences with experimental characterization towards the N6 antibody (172 out of 181 described). We generated the three-dimensional models of the gp120 strains (if no crystal structure was available on the PDB) and we docked the N6 antibody on its original binding region on all 172 strains. After ensuring that there are no clashes in any of the models, we predicted all possible mutations on the binding interface of the antibody with the gp120. This prediction process was performed using FoldX, pyDock and UEP, resulting in 361.200 predictions. An amount of 21 designs that unanimously improve the binding affinity (predicted by the three methods) in at least 150 gp120 strains were selected. Then, all selected mutations were assessed using PELE simulations. PELE simulations were performed against the 93TH057 gp120 strain, where we evaluated the amount of tight contacts that could be formed between the antibody-epitope complexes. After PELE simulations, an amount of 8 designs were proposed to be experimentally evaluated, and we concluded that only one design (Q54R) could slightly improve the binding affinity, while the others did not result in a generalized improvement (**Table 32**).

It is important to mention that literature does not contain many cases of antibody improvement. It is generally accepted that candidates increasing the contacts in a particular region also increase the binding affinity towards the epitope. In fact, two published studies for antibody design in HIV-1 follow this approach (Diskin *et al.*, 2011; Kwon *et al.*, 2018). One of them introduces a missing hydrophobic residue to the heavy chain of the NIH45-46 antibody (G54W) that mimics the key F43 of the gp120. The other study is based on a surface-matrix screening of the interface regions of the 10E8 bNAb, aiming to introduce large residues (only phenylalanines and arginines) that could enhance the binding potency. In our case, none of the single mutations produced on the N6 antibody resulted in a large improvement of its binding potency. We should underline, however, that some increase in binding affinity was produced in certain virus clades, partly backing up the approach (that should probably be tested in a less optimized bNAb). Moreover, it is worth mentioning that several mutations proposed by our approach are also found in other VRC01-like class antibodies; indicating that this methodology can mimic the evolution of nature partially.

After receiving the feedback from our experimental collaborators, we decided to take a step back and to follow a different rational approach. Instead of enhancing the amount of contacts, we aimed to do the opposite while aiming to favour some key interactions already found in the N6 antibody. In this way, we focused on a key hydrogen bond interaction between W100c$_{HC}$ and N279$_{GP120}$. This hydrogen bond has been reported to be crucial in previous publications, and its disruption is believed to drive antibody resistance. This comes with the fact that some single mutations at the position 279 in the gp120 (to glutamine, arginine, glutamate and alanine) results in a complete loss of N6 binding. Moreover, it has been described that removal of the W100c$_{HC}$ largely decreases the binding efficacy of the antibody.

Taking everything together, we hypothesized that if the disruption of this hydrogen bond hampers the binding potency of the antibody, facilitating the formation of this hydrogen bond could lead into an enhanced variant. In this sense, we observed that the orientation angle of this hydrogen bond is not optimal for the three available N6-gp120 crystal structures (**Figure 71,72**), and we hypothesized that removing the large side chains of the CDRL3 in the vicinity of the W100c$_{HC}$ could help to optimize the geometry of this key hydrogen bond by accommodating better the loop D. Therefore, we decided to mutate L91 and Q96 into smaller amino acids with similar physicochemical properties: alanine and serine, respectively. Then, we performed MD simulations and observed that the L91A;Q96S mutation adds rigidity to the loop D (**Figure 74**) by stabilizing the hydrogen bond orientation between W100c$_{HC}$ and N279$_{GP120}$. (**Figures 75-77**). After these results, we proposed the characterization of the single (L91A and Q96S) and double mutants (L91A;Q96S) of the N6 light chain. Our collaborators expressed, produced and purified the antibody variants, and they evaluated their binding affinity against 3 resistant and 14 sensitive strains. As we expected, none of the designs could bypass resistance because the key hydrogen bond cannot be produced. However, a generalized increase in antibody potency was observed in 13 out of 14 sensitive strains (**Table 33** and **Figure 78**) for the double mutation, a major achievement.

Interestingly, no significant binding improvement was observed for any of the single mutation designs, as we observed during the first round of designs (especially for the Q54R case). It is likely that only when combining more than one mutation, binding increase can be observed upon a synergistic effect between multiple mutations. In this sense, a triple mutant consisting of the double mutation of the CDRL3 (L91A;Q96S) and the single mutation on CDRH2 (Q54R) could lead to an improvement in N6 potency. This hypothesis comes with the fact that Q54R mutation achieved an increase in antibody potency in half of the evaluated strains, without hampering the binding affinity with other strains (**Table 32**). We have proposed this triple mutation to our experimental collaborators, and its experimental characterization is on its way.

In fact, it is biologically logical that single mutations on potent antibodies are likely to not substantially improve the binding affinity. Potent antibodies have been generated under hypermutation and hyper screening processes within the host, and single mutations are more probable to occur than multiple mutations at the same time. Hence, for a given optimized scaffold it is very likely that most single mutations have been screened and discarded within the host, leading to probably the best template antibody. However, the probability that the host has generated multiple mutations on a single antibody variant is lower, and therefore it seems that is the only way to improve the binding affinity of potent antibodies. If this is true, the amount of combinations to evaluate computationally increases exponentially, and methods for reducing this combinatorial space will be required.

Overall, this thesis highlights the complexity of the prediction of the effects of mutations in protein-protein complexes, and especially on their application into the antibody design field. Probably in the future, protein design may be an automatized and trivial process, but at this time it is an herculean effort where structural biologists and experimentalists must work together to pave the way to the future generations of scientists.

Conclusions

The conclusions that can be extracted from this thesis are:

- Prediction of the effects of mutations in protein-protein complexes is still one of the main challenges in protein design due to (probably) the scarce experimental ΔΔG determinations. Methods using only this source data may tend to have biased predictions, limiting their use in de novo predictions. The development of novel methodologies using alternative sources of data may provide new insights on the prediction process.
- Three-dimensional protein-protein data may offer an alternative and complementary source of data for the estimation of the ΔΔG upon mutation, which can be a crucial factor for reducing overfitting towards biased ΔΔG determinations.
- From the binary classification point of view, prediction of antibody binding efficacy can be performed by analyzing the contacts of the three-dimensional binding process using PELE simulations. Hence, sensitive epitopes would bind tightly and the contact surface will be higher rather than the resistant ones.
- Biomedical antibodies are potent and improving their binding efficacy is a challenging process from the computational and experimental point of view. Here, we have followed several approaches aiming to improve the N6 antibody, one of the best bNAbs against HIV-1. Most of the single mutation designs we proposed were not successful, which highlights the complexity of this task. However, one of the designs consisting of a double mutation (L91A;Q96S) on the CDRL3 of the antibody achieved higher neutralization potency in 13 out of 14 evaluated sensitive strains than the native antibody (geometric mean of 0.011 µg/mL and 0.026 µg/mL, respectively). This design favours a known key hydrogen bond between $W100c_{HC}$ and $N279_{GP120}$.
- Single mutations are unlikely to be enough for improving the binding potency of potent antibodies. Our results suggest that the synergistic effect of multiple mutations may play a very important role during the *in silico* design of improved antibody variants.

References

Alberts,B. *et al.* (2002) Protein Function. In, *Molecular Biology of the Cell. 4th edition*. Garland Science.

Alonso,S. *et al.* (2020) Genetically engineered proteins with two active sites for enhanced biocatalysis and synergistic chemo- and biocatalysis. *Nature Catalysis*, **3**, 319–328.

Amengual-Rigo,P., Carrillo,J., *et al.* (2020) Predicting Antibody Neutralization Efficacy in Hypermutated Epitopes Using Monte Carlo Simulations. *Polymers*, **12**, 2392.

Amengual-Rigo,P., Fernández-Recio,J., *et al.* (2020) UEP: an open-source and fast classifier for predicting the impact of mutations in protein–protein complexes. *Bioinformatics*.

Atilgan,A.R. *et al.* (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.*, **80**, 505–515.

Banks,J.L. *et al.* (2005) Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.*, **26**, 1752–1780.

Berendsen,H.J.C. *et al.* (1981) Interaction Models for Water in Relation to Protein Hydration. *The Jerusalem Symposia on Quantum Chemistry and Biochemistry*, 331–342.

Borrelli,K.W. *et al.* (2005) PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput.*, **1**, 1304–1311.

Bottaro,S. and Lindorff-Larsen,K. (2018) Biophysical experiments and biomolecular simulations: A perfect match? *Science*, **361**, 355–360.

Branco,P.C. *et al.* (2020) Survivin modulation in the antimelanoma activity of prodiginines. *Eur. J. Pharmacol.*, **888**, 173465.

Braun,E. *et al.* (2019) Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living J Comput Mol Sci*, **1**.

Carro,J. *et al.* (2018) Multiple implications of an active site phenylalanine in the catalysis of aryl-alcohol oxidase. *Sci. Rep.*, **8**, 8121.

Chan,D.C. and Kim,P.S. (1998) HIV Entry and Its Inhibition. *Cell*, **93**, 681–684.

Cheng,T.M.-K. *et al.* (2007) pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins*, **68**, 503–515.

Clark,A.J. *et al.* (2017) Free Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the gp120 Glycoprotein of HIV-1. *J. Mol. Biol.*, **429**, 930–947.

Clark,A.J. *et al.* (2019) Relative Binding Affinity Prediction of Charge-Changing Sequence Mutations with FEP in Protein–Protein Interfaces. *Journal of Molecular Biology*, **431**, 1481–1493.

Dehouck,Y. *et al.* (2013) BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations. *Nucleic Acids Res.*, **41**, W333–9.

Delgado,J. *et al.* (2019) FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics*, **35**, 4168–4169.

Díaz,L. *et al.* (2020) Monte Carlo simulations using PELE to identify a protein–protein inhibitor binding site and pose. *RSC Advances*, **10**, 7058–7064.

Diskin,R. *et al.* (2011) Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science*, **334**, 1289–1293.

Eastman,P. *et al.* (2017) OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.*, **13**, e1005659.

Eisenberg,D. *et al.* (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.*, **179**, 125–142.

Fh,C. (1958) On protein synthesis. *Symp. Soc. Exp. Biol.*, **12**.

Geng,C. *et al.* (2016) Exploring the interplay between experimental methods and the performance of predictors of binding affinity change upon mutations in protein complexes. *Protein Eng. Des. Sel.*, **29**, 291–299.

Gilabert,J.F. *et al.* (2020) Combining Monte Carlo and Molecular Dynamics Simulations for

Enhanced Binding Free Energy Estimation through Markov State Models. *J. Chem. Inf. Model.*

Gilabert,J.F. *et al.* (2019) PELE-MSM: A Monte Carlo Based Protocol for the Estimation of Absolute Binding Free Energies. *J. Chem. Theory Comput.*, **15**, 6243–6253.

Gilbert,P.B. *et al.* (2003) Comparison of HIV-1 and HIV-2 infectivity from a prospective cohort study in Senegal. *Stat. Med.*, **22**, 573–593.

Guerois,R. *et al.* (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.*, **320**, 369–387.

Henikoff,S. and Henikoff,J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 10915–10919.

Holzmann,H. *et al.* (1996) Correlation between ELISA, hemagglutination inhibition, and neutralization tests after vaccination against tick-borne encephalitis. *J. Med. Virol.*, **48**, 102–107.

Hosseini,A. *et al.* (2016) Computational Prediction of HIV-1 Resistance to Protease Inhibitors. *J. Chem. Inf. Model.*, **56**, 915–923.

Huang,J. *et al.* (2016) Identification of a CD4-Binding-Site Antibody to HIV that Evolved Near-Pan Neutralization Breadth. *Immunity*, **45**, 1108–1121.

Huang,X. *et al.* (2020) EvoEF2: accurate and fast energy function for computational protein design. *Bioinformatics*, **36**, 1135–1142.

Jacobson,M.P. *et al.* (2004) A hierarchical approach to all-atom protein loop prediction. *Proteins*, **55**, 351–367.

Jacobson,M.P., Kaminski,G.A., *et al.* (2002) Force Field Validation Using Protein Side Chain Prediction. *The Journal of Physical Chemistry B*, **106**, 11673–11680.

Jacobson,M.P., Friesner,R.A., *et al.* (2002) On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.*, **320**, 597–608.

Jankauskaite,J. *et al.* (2019) SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, **35**, 462–469.

Jemimah,S. *et al.* (2017) PROXiMATE: a database of mutant protein-protein complex thermodynamics and kinetics. *Bioinformatics*, **33**, 2787–2788.

Jiménez-García,B. *et al.* (2013) pyDockWEB: a web server for rigid-body protein–protein docking using electrostatics and desolvation scoring. *Bioinformatics*, **29**, 1698–1699.

Jorgensen,W.L. *et al.* (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, **79**, 926–935.

Jorgensen,W.L. and Tirado-Rives,J. (2005) Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.*, **26**, 1689–1700.

Jorgensen,W.L. and Tirado-Rives,J. (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, **110**, 1657–1666.

Kaminski,G.A. *et al.* (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides†. *The Journal of Physical Chemistry B*, **105**, 6474–6487.

Khersonsky,O. *et al.* (2018) Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol. Cell*, **72**, 178–186.e5.

Kotev,M. *et al.* (2018) Pushing the Limits of Computational Structure-Based Drug Design with a Cryo-EM Structure: The Ca Channel α2δ-1 Subunit as a Test Case. *J. Chem. Inf. Model.*, **58**, 1707–1715.

Kwon,Y.D. *et al.* (2018) Surface-Matrix Screening Identifies Semi-specific Interactions that Improve Potency of a Near Pan-reactive HIV-1-Neutralizing Antibody. *Cell Reports*, **22**, 1798–1809.

Lecina,D. *et al.* (2017) Adaptive simulations, towards interactive protein-ligand modeling. *Sci. Rep.*, **7**, 8466.

Lensink,M.F. *et al.* (2019) Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment. *Proteins*, **87**, 1200–1221.

Lin,Z.-H. *et al.* (2008) New descriptors of amino acids and their application to peptide QSAR study. *Peptides*, **29**, 1798–1805.

Liu,Q. *et al.* (2018) dbMPIKT: a database of kinetic and thermodynamic mutant protein interactions. *BMC Bioinformatics*, **19**.

Li,Y. *et al.* (2011) Mechanism of Neutralization by the Broadly Neutralizing HIV-1 Monoclonal Antibody VRC01. *Journal of Virology*, **85**, 8954–8967.

Li,Z. and Scheraga,H.A. (1987) Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. U. S. A.*, **84**, 6611–6615.

Lynch,R.M. *et al.* (2015) HIV-1 Fitness Cost Associated with Escape from the VRC01 Class of CD4 Binding Site Neutralizing Antibodies. *Journal of Virology*, **89**, 4201–4213.

Maier,J.A. *et al.* (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.*, **11**, 3696–3713.

Malek,R. and Mousseau,N. (2000) Dynamics of lennard-jones clusters: A characterization of the activation-relaxation technique. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics*, **62**, 7723–7728.

Martínez,L. (2015) Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS One*, **10**, e0119264.

Metropolis,N. and Ulam,S. (1949) The Monte Carlo method. *J. Am. Stat. Assoc.*, **44**, 335–341.

Moal,I.H. and Fernandez-Recio,J. (2013) Intermolecular Contact Potentials for Protein–Protein Interactions Extracted from Binding Free Energy Changes upon Mutation. *Journal of Chemical Theory and Computation*, **9**, 3715–3727.

Moal,I.H. and Fernández-Recio,J. (2012) SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models. *Bioinformatics*, **28**, 2600–2607.

Mosca,R. *et al.* (2013) Interactome3D: adding structural details to protein networks. *Nat. Methods*, **10**, 47–53.

Myung,Y., Rodrigues,C.H.M., *et al.* (2020) mCSM-AB2: guiding rational antibody design using graph-based signatures. *Bioinformatics*, **36**, 1453–1459.

Myung,Y., Pires,D.E.V., *et al.* (2020) mmCSM-AB: guiding rational antibody engineering through multiple point mutations. *Nucleic Acids Res.*, **48**, W125–W131.

Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443–453.

Onufriev,A. *et al.* (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins*, **55**, 383–394.

Pardo,I. *et al.* (2016) Re-designing the substrate binding pocket of laccase for enhanced oxidation of sinapic acid. *Catalysis Science & Technology*, **6**, 3900–3910.

Pearce,R. *et al.* (2019) EvoDesign: Designing Protein–Protein Binding Interactions Using Evolutionary Interface Profiles in Conjunction with an Optimized Physical Energy Function. *Journal of Molecular Biology*, **431**, 2467–2476.

Pérez,A. *et al.* (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.

Perez,C. *et al.* (2020) FragPELE: Dynamic Ligand Growing within a Binding Site. A Novel Tool for Hit-To-Lead Drug Design. *J. Chem. Inf. Model.*, **60**, 1728–1736.

Pieper,U. *et al.* (2014) ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.*, **42**, D336–46.

Pires,D.E.V. *et al.* (2016) mCSM-lig: quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Sci. Rep.*, **6**, 29575.

Pires,D.E.V. *et al.* (2020) mCSM-membrane: predicting the effects of mutations on transmembrane proteins. *Nucleic Acids Res.*, **48**, W147–W153.

Pires,D.E.V. *et al.* (2014) mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, **30**, 335–342.

Pires,D.E.V. and Ascher,D.B. (2016) mCSM-AB: a web server for predicting antibody-antigen affinity changes upon mutation with graph-based signatures. *Nucleic Acids Res.*, **44**, W469–73.

Pires,D.E.V. and Ascher,D.B. (2017) mCSM-NA: predicting the effects of mutations on protein-nucleic acids interactions. *Nucleic Acids Res.*, **45**, W241–W246.

Ponomarenko,E.A. *et al.* (2016) The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.*, **2016**.

Reeves,J.D. and Doms,R.W. (2002) Human immunodeficiency virus type 2. *J. Gen. Virol.*, **83**, 1253–1265.

Reif,M.M. *et al.* (2012) New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field. *J. Chem. Theory Comput.*, **8**, 3705–3723.

Romero-Durana,M. *et al.* (2020) pyDockEneRes: per-residue decomposition of protein–protein docking energy. *Bioinformatics*, **36**, 2284–2285.

Roos,K. *et al.* (2019) OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.*, **15**, 1863–1874.

Rudicell,R.S. *et al.* (2014) Enhanced potency of a broadly neutralizing HIV-1 antibody in vitro improves protection against lentiviral infection in vivo. *J. Virol.*, **88**, 12669–12682.

Saen-Oon,S. *et al.* (2019) Atomistic simulations shed new light on the activation mechanisms of RORγ and classify it as Type III nuclear hormone receptor regarding ligand-binding paths. *Sci. Rep.*, **9**, 17249.

Salas,F. de *et al.* (2019) Engineering of a fungal laccase to develop a robust, versatile and highly-expressed biocatalyst for sustainable chemistry. *Green Chemistry*, **21**, 5374–5385.

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

Santiago,G. *et al.* (2016) Computer-Aided Laccase Engineering: Toward Biological Oxidation of Arylamines. *ACS Catalysis*, **6**, 5415–5423.

Santiago,G. *et al.* (2018) Rational Engineering of Multiple Active Sites in an Ester Hydrolase. *Biochemistry*, **57**, 2245–2255.

Schymkowitz,J. *et al.* (2005) The FoldX web server: an online force field. *Nucleic Acids Res.*, **33**, W382–8.

Schymkowitz,J.W.H. *et al.* (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 10147–10152.

Siebenmorgen,T. and Zacharias,M. (2020) Efficient Refinement and Free Energy Scoring of Predicted Protein–Protein Complexes Using Replica Exchange with Repulsive Scaling. *Journal of Chemical Information and Modeling*.

Sirin,S. *et al.* (2016) AB-Bind: Antibody binding mutational database for computational affinity predictions. *Protein Sci.*, **25**, 393–409.

Stein,A. *et al.* (2011) 3did: identification and classification of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.*, **39**, D718–23.

Vangone,A. and Bonvin,A.M. (2015) Contacts-based prediction of binding affinity in protein–protein complexes. *eLife*, **4**.

Vanommeslaeghe,K. *et al.* (2009) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, NA–NA.

Wales,D.J. and Doye,J.P.K. (1997) Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *The Journal of Physical Chemistry A*, **101**, 5111–5116.

Wang,J. *et al.* (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.*, **25**, 247–260.

Wang,J. *et al.* (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.

Woods,C.J. *et al.* (2018) ProtoMS 3.4. *ProtoMS*.

Wu,X. *et al.* (2011) Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science*, **333**, 1593–1602.

Wyatt,R. (1998) The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens. *Science*, **280**, 1884–1888.

Xue,L.C. *et al.* (2016) PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*, btw514.

Zhou,P. *et al.* (2019) Broadly resistant HIV-1 against CD4-binding site neutralizing antibodies. *PLoS Pathog.*, **15**, e1007819.

Zhou,T. *et al.* (2013) Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-class antibodies. *Immunity*, **39**, 245–258.

Zhou,T. *et al.* (2010) Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science*, **329**, 811–817.

Zhou,T. *et al.* (2015) Structural Repertoire of HIV-1-Neutralizing Antibodies Targeting the CD4 Supersite in 14 Donors. *Cell*, **161**, 1280–1292.

Zhu,K. *et al.* (2007) Improved Methods for Side Chain and Loop Predictions via the Protein Local Optimization Program: Variable Dielectric Model for Implicitly Improving the Treatment of Polarization Effects. *J. Chem. Theory Comput.*, **3**, 2108–2119.