UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Biologia

---

CORSO DI DOTTORATO DI RICERCA IN: Bioscienze e Biotecnologie
CURRICOLO: Biologia Cellulare
CICLO: 29°

# STRUCTURAL CHARACTERIZATION OF THE HUMAN IMMUNE RESPONSE TO THE MENINGOCOCCAL VACCINE ANTIGEN NHBA

Tesi redatta con il contributo finanziario di GSK Vaccines

**Coordinatore:** Ch.mo Prof. Paolo Bernardi
**Supervisore**: Ch.mo Prof. Cesare Montecucco
**Co-Supervisore**: Dr. Enrico Malito

**Dottorando:** Martina Maritan

## Table of Contents

# Abstract

*Neisseria meningitidis* is a human pathogen which can cause fatal sepsis and invasive meningococcal disease once it reaches the blood stream and the nervous system. Bexsero™ is the first broadly protective multicomponent vaccine against serogroup B *N. meningitidis*. Among its components, the Neisserial Heparin Binding Antigen (NHBA) represents the less structurally and functionally characterized antigen of Bexsero and therefore has been the target of the current study. NHBA is a surface-exposed lipoprotein composed by two domains (C- and N-terminal), and while the structure of the C-terminal domain was solved by NMR, the structure of N-terminal region is still unknown and predicted to be largely unstructured. Knowledge of the NHBA full-length structure can provide important insights towards the understanding of its biological function, as well as on its role as vaccine antigen.

In this work we combined the information derived from the profiling the B-cell repertoire in response to vaccination with Bexsero with structural biology, in order to shed light on NHBA structure and on the molecular bases of its recognition by the human immune system. By using fragments antigen binding (Fabs) derived from human monoclonal antibodies (mAbs), we sought to provide high-resolution epitope mapping of the NHBA immunogenic regions, which in turn could permit a deeper characterization of the molecular determinants of antibody binding and protective epitopes.

Here we present a structural characterization of the first high-resolution crystal structures of three free human anti-NHBA Fabs of NHBA. These structures reveal features compatible with the binding of NHBA regions as previously determined by other low-resolution methods. Moreover, these Fabs provided important tools for structural studies through co-crystallization experiments with various NHBA fragments and constructs. While using Fabs that bind to the NHBA N-terminal region complexed with full length or sub-full length antigen did not result in crystal growth, a construct including solely the C-terminal domain yielded crystals that allowed solving the structure of the complex with a C-terminal Fab binder. Structural analysis highlighted the conservation of the epitope and lead to the identification of the key residues involved into NHBA-Fab recognition. Additionally, the comparison between bound and unbound Fab revealed an interesting fitting mechanism occurring during antigen recognition that provides further details into the elucidation of antibody binding.

Combined together, these results enhance our structural and biophysical understanding of NHBA, and provide a platform for deeper analyses aimed at the elucidation of the molecular determinants of its immunogenicity.

# Sommario

*Neisseria meningitidis* è un patogeno umano obbligato e un potenziale agente eziologico di sepsi e meningite, qualora entri in contatto con il flusso sanguigno o il sistema nervoso. Bexsero è il primo vaccino multicomponente in grado di fornire ampia protezione contro il sierogruppo B *di N. meningitidis.* Tra i vari componenti di Bexsero, Neisserial Heparin Binding Antigen (NHBA) rappresenta l'antigene meno caratterizzato sia funzionalmente che strutturalmente e, per questa ragione, è stato selezionato come oggetto di studio nel presente lavoro. NHBA è una lipoproteina esposta sulla superficie del batterio e si compone di due domini (C- e N-terminale); mentre la struttura del dominio C-terminale è stata risolta tramite spettroscopia NMR, il dominio N-terminale è ancora strutturalmente non caratterizzato e predetto, per la maggior parte della sua sequenza, come non-strutturato. La conoscenza delle caratteristiche strutturali della proteina intera, quindi, potrebbe essere utile per chiarire il ruolo che NHBA svolge sia dal punto di vista biologico sia come antigene vaccinale.

In questo lavoro di tesi, informazioni derivanti dall'analisi del repertorio delle cellule B generate in risposta all'immunizzazione con Bexsero sono state combinante con tecniche di biologia strutturale (cristallografia a raggi X), in modo da delucidare la struttura completa di NHBA e le basi molecolari del suo riconoscimento da parte del sistema immunitario umano. Frammenti di anticorpi leganti l'antigene (Fabs), derivanti da anticorpi monoclonali umani (mAbs) sono stati utilizzati per la mappatura ad alta risoluzione degli epitopi immunogenici di NHBA, permettendo un'accurata caratterizzazione dal punto di vista molecolare delle regioni dell'antigene implicate nel legame con gli anticorpi.

Dall'analisi delle strutture cristallografiche ad alta risoluzione di tre Fabs umani diretti contro NHBA, è emerso che questi presentano a livello del paratopo caratteristiche che indicano la loro compatibilità con il legame a NHBA tramite regioni precedente mappate con tecniche a bassa risoluzione. Inoltre, questi Fabs si sono rivelati un importante strumento per studi strutturali attraverso esperimenti di co-cristallizzazione in complesso con NHBA. Purtroppo non è stato possibile ottenere cristalli idonei per esperimenti di diffrazione dalla co-cristallizzazione dei Fabs con l'intera NHBA o frammenti N-terminali di NHBA; al contrario, il dominio C-terminale dell'antigene in complesso con un Fab ha generato cristalli che hanno permesso di risolvere la prima struttura di un complesso tra C-terminale di NHBA e un Fab umano. L'analisi strutturale di questo complesso ha evidenziato l'estrema conservazione dell'epitopo e guidato l'identificazione dei residui chiave implicati nel riconoscimento tra NHBA e Fab. Inoltre, il confronto tra la forma complessata e non-complessata del Fab ha rivelato un

interessante meccanismo di legame che avviene nella fase di riconoscimento dell’antigene e che fornisce ulteriori dettagli sui meccanismi che regolano i legami tra anticorpi e antigeni.

I risultati ottenuti in questo studio hanno contribuito a migliorare la conoscenza strutturale e biofisica di NHBA e forniscono una base per una futura analisi strutturale più accurata, volta a delucidare i dettagli molecolari della risposta immunitaria a NHBA.

# Introduction

## Reverse Vaccinology & Structural Vaccinology

Vaccines are one of the most successful medical interventions against infectious diseases [1], and vaccination is thought to prevent an estimated 2.5 million deaths each year [2]. The aim of most vaccinations is to confer protection against microorganisms, and their action is based on the ability to elicit an immune response [3]. Vaccines have been historically developed by the “isolating, inactivating, and injecting” approach [4]. This empirical method used as vaccines mixtures of disease-causing agents after their attenuation (i.e. bacteria killed by heat or chemical treatment, and that retained immunogenicity) [5]. Today, novel technologies allow for a “rational” approach to vaccine development, with advances in areas such as genome sequencing and proteomics that almost 20 years ago provided the bases for a revolution in the field, the so-called Reverse Vaccinology (RV) [6-9].

The complete genome of a pathogen provides key information on genes that encode for potential antigens, which can thus be selected and tested as candidates for vaccine development by use of bioinformatics tools and without the need, in this first phase, to grow pathogens. For example, bioinformatics is used to select open reading frames (ORFs) that encode putative surface-exposed or secreted proteins, which can potentially be recognized by antibodies. This stage of identification can be performed *in silico* by searching for specific properties like the presence of signal peptide sequences or motifs, and of membrane spanning regions. Sequence homology analyses allow then to gain additional critical information on the candidate antigens such as homology both to other known virulence factors or protective antigens (from other pathogens), and to human proteins (to avoid autoimmune issues) [10]. In addition to genome sequencing, also mass spectrometry and proteomics play a significant role in modern vaccine development projects [11-14]. Once the initial candidates are identified and selected, these will then be produced as recombinant proteins for further characterization, and will then be tested for their immunogenicity in animal models. This will provide experimental evidence of their potential use as vaccine candidates.

The value of RV has been recently demonstrated with the discovery and development of Bexsero, the broadly protective vaccine against *Neisseria meningitidis* serogroup B [13] (**Fig.1.1**). In addition, RV has been also used to aid the identification of vaccine candidates for several other bacterial pathogens, and recently also for herpes simplex virus [15-19].

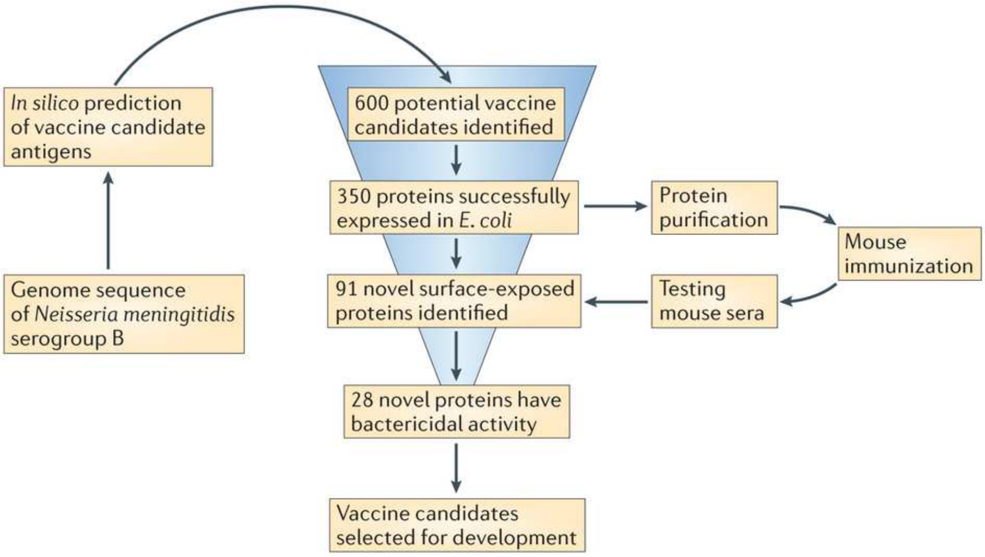
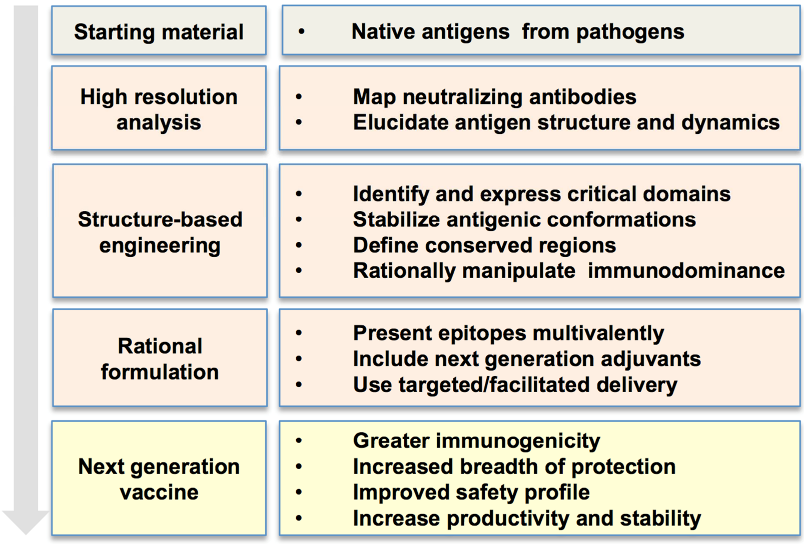


**Figure 1.1. Example of the RV approach applied to meningococcus serogroup B.** Bioinformatics is used for prediction of genes encoding vaccine candidate antigens that are i) exposed on the bacterial surface, ii) antigenically conserved, and iii) do not contain sequences that are similar to human proteins. Selected genes are expressed in *E. coli* and then the proteins are purified and used for mice immunization. The sera of the mice are tested for their bactericidal killing in the presence of complement. The 'best' antigens are then selected for vaccine development [3].

Technological advances in areas such as human immunology and structural biology are also providing key new tools for vaccine research and development. Modern capabilities to profile the antigen-specific B-cell repertoire and characterize the corresponding antibodies represent powerful tools for the study of the specific immune response upon immunization [20]. Using information and knowledge derived from the B-cell repertoire as a starting point for structural biology studies has also already given remarkable breakthroughs in vaccine research [21]. For example, several studies that use structure-based methods applied to vaccine research and development have been recently published on targets for different diseases like respiratory syncytial virus (RSV) [22], human immunodeficiency virus1 (HIV-1) [23], MenB [24] and group B streptococcus (GBS) [25]. Structural knowledge of protein antigens and of complexes between antigens and antibodies can provide the molecular or atomic bases to drive optimization and to enable design of more stable, homogeneous, immunogenic, and affordable vaccine antigens [26]. The combination of experimental structure-determination methods such as nuclear magnetic resonance (NMR), X-ray crystallography, and electron microscopy (EM), with computational methods like epitope predictions, structure modeling and design, can potentially provide the tools for a new revolution in vaccine research, what is now known as Structural Vaccinology (SV) [10].

One of the common applications of SV in vaccine research is the structural characterization of the immune response through epitope mapping, which provides the molecular details of antigen recognition by the host immune system. This translates into a three-dimensional map of the residues and atoms of a certain antigen that constitute its immunogenic regions [27-29], also known as epitopes, with obvious important implications for design strategies. The structure of vaccine antigens can also be used for structure-based optimization aimed at improving stability [30] and homogeneity of a candidate, stabilizing its fold [31], or reducing its degradation and aggregation tendencies. Structure is also a key starting point in order to engineer molecules in order to overcome sequence variability and the impact that this has on immune recognition. For example, SV can guide the design of chimeric antigens that display epitopes derived from multiple variants, resulting in the elicitation of a wider immune response, as previously reported for the meningococcal factor H binding protein (fHbp) [24]. In this study, starting from the structure of fHbp that was previously elucidated by NMR spectroscopy and X-ray crystallography, epitope mapping was performed using bactericidal antibodies against three sequence variants. Information on the variant specific epitopes recognized by these antibodies was then used for the design of an engineered fHbp molecule containing several cross-protective B cell epitopes on one single common scaffold. The choice of the scaffold in this case was the fHbp molecule itself, being its overall structure well conserved among the three variants, and being the variable regions only localized on its surface. Importantly, the newly generated chimeric antigen showed the ability to elicit broadly protective antibodies that could target different fHbp variants and sub-variants commonly circulating [24].



| | |
|---|---|
| **Starting material** | • **Native antigens from pathogens** |
| **High resolution analysis** | • **Map neutralizing antibodies**<br>• **Elucidate antigen structure and dynamics** |
| **Structure-based engineering** | • **Identify and express critical domains**<br>• **Stabilize antigenic conformations**<br>• **Define conserved regions**<br>• **Rationally manipulate immunodominance** |
| **Rational formulation** | • **Present epitopes multivalently**<br>• **Include next generation adjuvants**<br>• **Use targeted/facilitated delivery** |
| **Next generation vaccine** | • **Greater immunogenicity**<br>• **Increased breadth of protection**<br>• **Improved safety profile**<br>• **Increase productivity and stability** |

**Figure 1.2.** The structural vaccinology approach. Adapted from [26].

## *Neisseria meningitidis* and Meningococcal disease

*Neisseria meningitidis* is a gram negative bacterium of the bacterial family Neissiriaceae, which possess a spherical or kidney-like shape. It is a non-motile and non-sporulating encapsulated and piliated aerobic diplococcus, and as other gram negatives possesses an outer membrane that is composed of lipids, outer membrane proteins (OMPs), lipo-oligosaccharides (LOS), a peptidoglycan layer, and a lipopolysaccharide (LPS). The LPS plays a role in serum resistance and thus in the pathogenesis [32], while the polysaccharide capsule, which is present in almost all *N. meningitidi*s strains, is responsible for invasive disease, being crucial for the survival in the blood, the resistance to antibody/complement mediated killing, and inhibition of phagocytosis. According to the different structures and composition of the capsule, 12 different meningococcal serogroups have been defined [33, 34] of which five, A, B, C, W135 and Y, are responsible of more than 95% of the disease [35]. Further classification into serosubtype, serotype and immunotype is based on class 1 outer membrane proteins (PorA), class 2 or 3 (PorB) outer membrane proteins and lipopolysaccharide structure, respectively [36, 37].

*N. meningitidis* is a commensal that colonizes the mucosal epithelium of the nasopharynx of approximately the 10% of the population, with humans being its unique reservoir [38]. The rate of asymptomatic carriage increases with conditions such as household crowding, and at different ages with a peak in adolescents. Under particular circumstances, and in susceptible individuals, meningococcus can become pathogenic and can cause meningitis and sepsis [39]. Although the events that lead to bacterial invasion are not well understood, lack of a protective immune response and environmental factors that damage the nasopharyngeal mucosa, are thought to increase the incidence of invasive meningococcal disease. For unknown factors that depend both on the host and on the pathogen, meningococcus can invade the pharyngeal mucosal epithelium and disseminate into the blood-stream causing septicemia [40], or can cross the blood-brain barrier and enter the cerebrospinal fluid, causing meningitis [41] (**Fig. 1.3**).

From exposure to the onset of meningococal disease, the average incubation time is between two and ten days [37], with symptoms of stiff neck, high fever, sensitivity to light, confusion, headache, and vomiting. Although the disease is rare, it causes death in 10-15% of the cases, and typically within 24 to 48 hours of symptoms onset [37, 42]. Among survivors, 10% to 20% present permanent disability such as hearing loss, limb loss, cognitive dysfunction, visual impairment, educational difficulties, developmental delays, motor nerve deficits, seizure disorders, and behavioral problems [36].

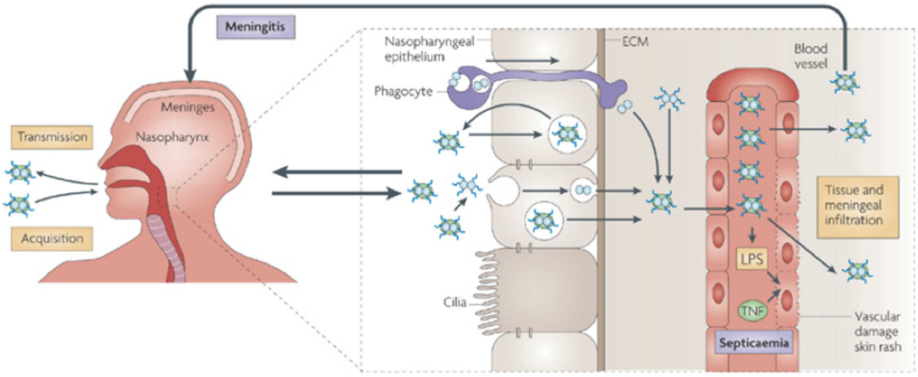


**Figure 1.3. Stages in the pathogenesis of *Neisseria meningitidis*.** The image summarizes the steps during meningococcal colonization and infection Adapted from [40]

## Meningococcal vaccines

The emergence of *N. meningitidis* strains with epidemic potential, as well as the change in serogroup circulation is still unpredictable [43]. Moreover, the rapid progression and the difficulties to diagnose meningococcal disease [36] make preventive vaccination the most effective option.

A number of glycoconjugate vaccines against meningococcus are currently available as monovalent (against serogroup A or C, i.e. MenAfriVac, Menjugate or NeisVac-C) or as tetravalent formulations (against serogroups A, C, W-135 and Y, i.e. Menactra, Menveo) [44]. However, for a long time the most critical target for vaccination remained meningococcus B, which is responsible for the 32% of all cases in the United States and for 45–80 % in Europe [45]. Differently than for other serogroup-targeting vaccines (see above), a polysaccharide-based vaccine approach could not be used for group B meningococcus because its capsular polysaccharide contains neuraminic acid, a mimic of the sialic acid found in human neural tissue. Thus, in addition to being a poor immunogen in humans, the MenB capsular polysaccharide could be recognized as a self-antigen with potentially devastating consequences of auto-immune reactions [46, 47]. Consequently, efforts for the identification of candidates to develop a vaccine against MenB have been focused on non-capsular antigens and on proteins.

Bexsero (4CMenB) is the first broadly protective recombinant multi-component vaccine against MenB that, as mentioned above, has been discovered and developed by Reverse vaccinology [48]. Originally, approximately 600 open reading frames were selected from the genome of *N. meningitidis* strain MC58 that were likely to encode potential surface exposed proteins [49]. The sequences of these initial candidates were analyzed for their conservation among a wide range of MenB strains, and were later expressed as recombinant proteins in *E. coli* in order to elucidate their biochemical properties and especially to test their ability to induce bactericidal killing in mice. This work revealed more than 90 novel surface-located proteins, of which 28 were able to induce bactericidal antibodies. The most promising candidates were then selected for further studies leading to the final identification of the three protein components that make Bexsero [6, 50], in combination with outer membrane vesicles (OMVs) from a New-Zealand epidemic strain (NZ98/254). The three protein antigens of Bexsero are: factor H-binding protein (fHbp), Neisserial Heparin-Binding Antigen (NHBA) and Neisserial adhesin A (NadA) [48, 49]. fHbp (sub-variant 1.1) and NHBA (peptide 2), are present as fusion proteins with two minor genome-derived *Neisseria* antigens (GNA), GNA2091 and GNA1030, respectively [13], which were included in the final vaccine formulation to further enhance immunogenicity of the main antigens and to facilitate large-scale manufacturing (**Fig. 1.4**).

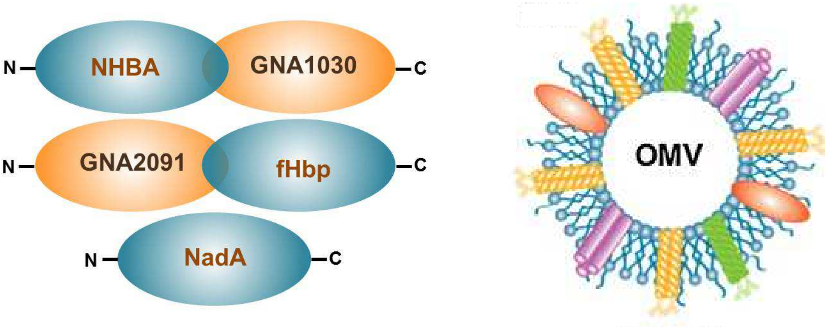


**Figure 1.4. Schematic representation of 4CMenB**. 4CMenB composition: the three main recombinant antigens fHbp, NadA, and NHBA and the fusion proteins GNA1030 and GNA2091 (left) combined with OMV from MenB strain NZ 98/254 (right).

Clinical trials have demonstrated safety and immunogenicity of Bexsero, which is able to induce a protective immune response in infants, children, adolescents, and adults, against the majority of MenB strains [51, 52]. Bexsero has been licensed in Europe in 2013 and in the U.S. in 2015.

A second recombinant protein-based vaccine against serogroup B, known as Trumenba, is now also available, and it has been discovered and developed by Pfizer. Trumenba has been licensed in the U.S. in 2014 for a target population of adolescents and young adults, and differently than Bexsero it is composed by equal amounts of two variants of lipidated fHbp (subfamily A05/var3.45 and subfamily B01/var1.55) [53]. During preclinical evaluation, Trumenba has been shown to have good breadth of coverage [54]. However, because it contains purified lipoproteins known as TLR-2 agonists, Trumenba is not suitable for use in infants [55].

## NHBA vaccine antigen

NHBA (Neisserial Heparin Binding Antigen, or GNA2132), is a surface-exposed lipoprotein able to induce antigen-specific bactericidal antibodies in animals and humans immunized with 4CMenB [50, 56, 57]. Moreover, NHBA is a protein specific to *Neisseria* species and its gene is ubiquitous in all Neisserial group B strains tested so far [58-61]. The NHBA-encoding gene is present also in *N. gonorrhoeae* and the commensal species *N. lactamica*, *N. polysaccharea*, and *N. flavescens* and, with the exception of *N. gonorrhoeae,* it is highly conserved [61]. Gene sequence analysis from genetically diverse group B strains revealed the existence of more than 600 peptides that possess some association with clonal complexes and sequence types [61, 62]. Examination of these gene sequences revealed also the presence of variable segments of NHBA at the level of the primary amino acid sequence, with the highest sequence diversity clustered in the N-terminal region of the protein (**Fig. 1.5**). In particular, sequence variability has been associated to an insertion of 60 amino acids (IB-insertion), present only in some strains and that allows a classification of NHBA peptides in so-called long and short variants. The long NHBA peptide 2 (p2) variant is the one included in Bexsero, being this the most frequently expressed peptide found through molecular epidemiological studies in Europe [63].

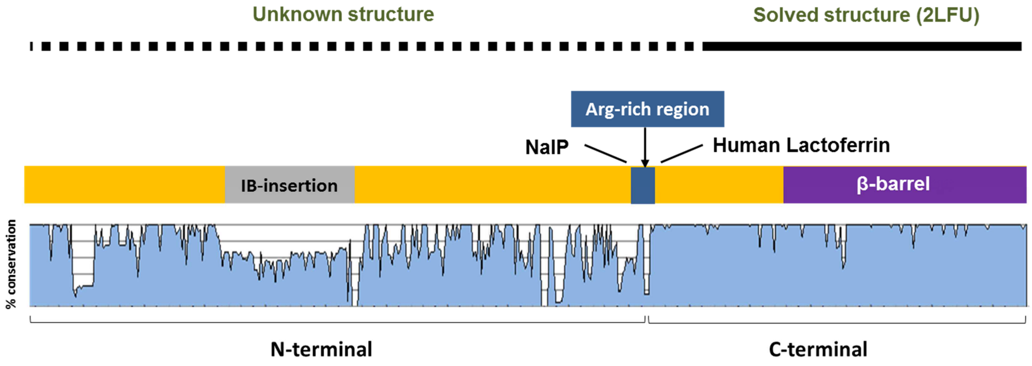


**Figure 1.5. NHBA variability and modular structural organization**. NHBA protein sequence reflects a modular structural organization, where it is possible to distinguish the N- (yellow) and C-terminal (purple) domains. The grey box represents the IB-insertion. Arg-rich motif (blue) is flanked by the NalP cleavage site and by the human lactoferrin cleavage site. The upper line illustrates the extent of the structural information currently available. The lower graph shows the percentage of amino acid conservation between the different peptides along the protein sequence.

The amino acidic protein sequence of NHBA allows to distinguish two independently folded domains (N- and C-terminus) separated by a central poly-Arginine motif. The structure of the C-terminal region (residues 246 to 427) of NHBA peptide 20 (p20) was first determined by NMR spectroscopy [64], revealing a single 8-stranded anti-parallel β-barrel and two β-strands in a hairpin conformation (**Fig. 1.6A**). The two β-strands are not well defined in the NMR structure, likely due to the experimental conditions under which the sample was analyzed. This fold resembles that of members of the lipocalin superfamily, which are typically involved in the transport of small hydrophobic compounds such as retinol in the plasma [65], and also shows clear structural similarity with transferrin binding protein B (TbpB) [66] and with the C-terminal domain of fHbp (**Fig. 1.6B**).

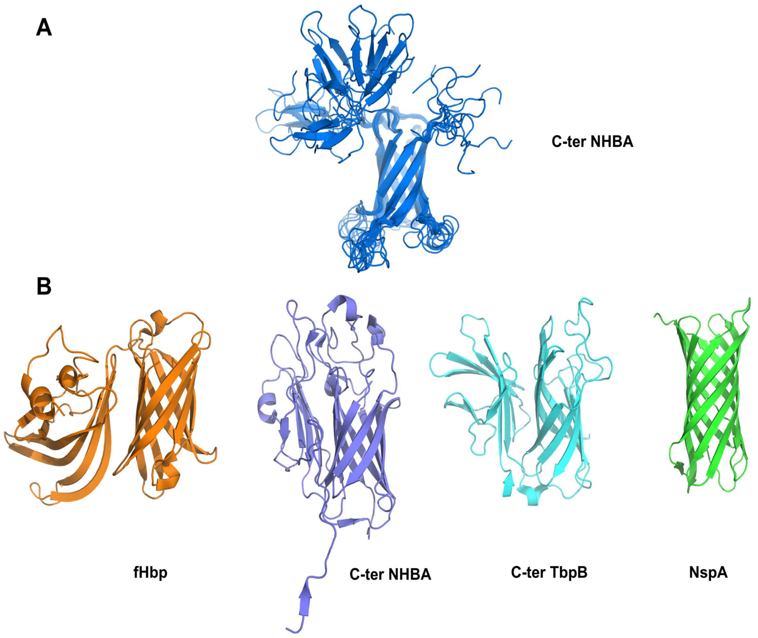


**Figure 1.6. NHBA C-terminal structure.** (**A**) Cartoon representation of NHBA C-terminal ensemble of structures determined by NMR spectroscopy (PDB ID 2LFU). (**B**) Cartoon representations of proteins with significant structural similarity to NHBA determined by X-ray crystallography. From left to right: fHbp (PDB 3KVD), the N- and C-terminal repeats of TbpB (PDB 3HOL) and NspA (PDB 1P4T).

The Arg-rich region appears to be extremely conserved among different NHBA peptides, underlying its importance in bacterial pathophysiology as also confirmed by experiments that showed how this region allows NHBA to bind heparin in vitro. Importantly, heparin binding is also associated to increased survival of bacteria in human serum [56]. A complex between NHBA and heparin on the meningococcal cell surface could recruit complement inhibitors, which in turn could prevent complement activation. Also, binding of heparin by NHBA *in vitro* could potentially translate *in vivo* to binding of glysoaminoglycans (such as heparan sulfate) that are present in mucosal secretions. In fact, several other bacterial adhesins have been reported to bind heparin and heparin sulfate [13]. Moreover, NHBA has been shown to be susceptible to cleavage mediated by two proteases, the meningococcal NalP and the human lactoferrin, that cleave NHBA upstream and downstream of the Arg-rich region respectively releasing two fragments into the extracellular milieu [56, 67].

Overall, NHBA appears to contribute to meningococcus interaction with the host cells and likely to be involved in biofilm formation. However, NHBA specific biological function remains unknown.

## Antibodies: tools for studying the immune system

The interaction between antibodies (Ab) and antigens (Ag) represent a key event for the clearance of microbes and foreign molecules by the immune system. Antibodies, which are generated by B-lymphocytes and activate the humoral immune response, mainly direct binding to antigens by coating the surface of the invader to prevent the pathogen from entering or damaging healthy body cells. Antibodies also stimulate other parts of the immune system, like the complement cascade, to destroy pathogens, and they can mark microbes by opsonization so that they can later be recognized and neutralized by effector immune cells [68, 69].

Antibodies are roughly Y-shaped glycoproteins that belong to the immunoglobulin (Ig) superfamily [70], and consist of a pair of β-sheets each formed of antiparallel β strands linked by a single disulfide bond. Antibodies are generally made of four separate polypeptide chains, two light (L; MW ~25 kDa) and two heavy (H; MW ~55 kDa). H and L chains of antibodies combine to form one Fc (Fragment crystallizable) domain and two Fab (Fragment antigen binding) modules. The Fc and the Fab are linked by a flexible "hinge" region, and can be separated by proteolytic cleavage. Furthermore, each Fab fragment is subdivided in two distinct structural domains named variable ($V_L$,$V_H$) and constant ($C_L$,$C_H1$), with the amino acid residues linking $V_L$ to $C_L$ and $V_H$ to $C_H1$ called switch residues [68, 71] (**Fig. 1.7**).

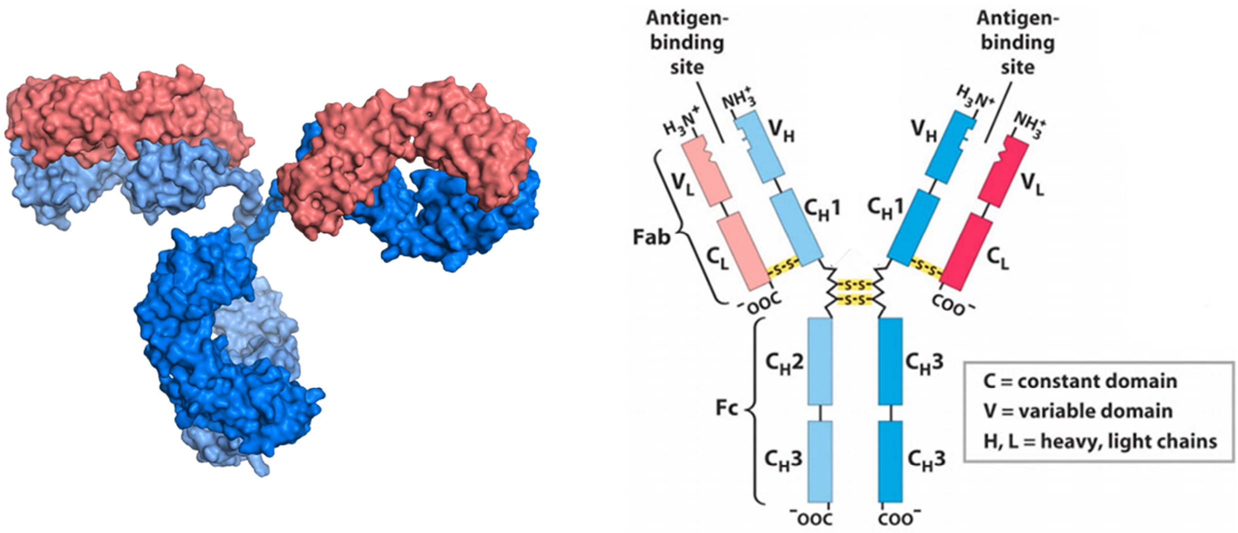


**Figure 1.7. Antibody structure**. (**A**) Crystal structure of the anti-canine lymphoma monoclonal antibody (PDB 1IGT) (**B**) Schematic representation of an IgG molecule divided in heavy (blue) and light (red) chains. Disulfides bonds are highlighted in yellow. Adapted from [68]

The Fab variable domains are responsible for antigen specificity, as they bind to the antigen through six hypervariable loops (L1, L2, L3, H1, H2, H3) that are commonly known as complementarity determining regions (CDRs) [72, 73]. Although also non-CDRs residues have been found to be involved in Ag recognition, these play a more modest role compared to the CDRs [74]. Loops forming the CDRs extend from the conserved structural framework of the Fabs and are highly variable in size and sequence. This allows antibodies to recognize an almost unlimited number of antigens [75]. Given their biological relevance, efforts in developing a rationale for CDRs identification and classification have been pursued [76-78]. These early studies were based on a limited number of structures, while the recent explosion of Fabs structures deposited in the Protein Data Bank (PDB) enabled a more exhaustive description of the CDRs [79].

The ensemble of antibody residues that in addition to the CDRs are required for Ag recognition, are often called “paratope”. Instead, molecular structures within any given antigen that contact the antibody are referred to as antibody or B-cell epitopes [80]. Knowing the structure of epitopes and of the epitope-paratope interfaces provides highly valuable information for the understanding not only of host-immune response, but also for the design of antigens with desired biochemical and immunological properties [81, 82]. Such knowledge can be mainly achieved by epitope mapping, which can be performed by several structural and non-structural methods [80, 83]. Experimental techniques of epitope mapping [84] include i) ELISA-based approaches, ii) Pep-Scan or Protein microarrays (elucidating linear peptides from the antigen protein sequence) [85-88]), iii) protein mutagenesis or phage display, iv) mass-spectrometry methods such as hydrogen-deuterium exchange coupled to mass-spectrometry (HDX-MS), and v) the more informative and high-resolution of all, Cryo-EM and X-ray crystallography, which can reveal the full picture of the often intricate three-dimensional arrangement of atoms and amino acids that make the epitope-paratope interface.

Importantly, in addition to their use to elucidate immunologically-relevant function and structures, antibody fragments (Fabs) can be used as chaperones to aid crystallization of their cognate antigens, specifically when the latter are particularly difficult to crystallize, as in the presence of protein containing flexible or unstructured regions or multiple domains [89]. The extreme stability of Fabs and their specificity generally stabilize antigen conformations or regions that otherwise might negatively impact the ability to form the crystal contacts necessary to obtain the ordered packing that distinguishes protein crystals.

# Aims

Despite its relevance for the generation of protective immunity as a component of the MenB vaccine Bexsero, the molecular details of NHBA biological function and of its recognition by the immune system remain yet not fully understood. A recent effort to profile the B-cell repertoire in response to vaccination with the Bexsero antigens enabled the isolation and characterization of human monoclonal antibodies (hu-mAbs) against these antigens (fHbp, NHBA, and NadA) [90, 91]. Specifically for this study, a set of four anti-NHBA hu-mAbs were isolated and later selected for further characterization including structural biology studies. Three of the initial four anti-NHBA hu-mAbs are the subject of the research work described in this thesis.

The specific aims of the current work were to exploit the availability of anti-NHBA hu-mAbs in order to: i) elucidate the structure of the N-terminal domain of NHBA, whose 3D structure remains elusive; and ii) elucidate the structural bases of the NHBA recognition by the immune system. Firstly, fragments antigen binding (Fabs) derived from the available hu-mAbs where used as tools to increase the probabilities of obtaining crystals of NHBA, which alone resulted recalcitrant to crystallization. Secondly, the Fabs were utilized in complex with NHBA for epitope mapping studies, to provide a high-resolution picture of those NHBA regions involved in the recognition by the human immune system.

Molecular biology, biochemical and biophysical methods, as well as X-ray crystallography, were the main techniques utilized in the work described in this thesis. The crystal structures of three unbound Fabs, called 12E1, 10C3, and 5H2, derived from the anti-NHBA hu-mAbs isolated in the study mentioned above, were solved at high-resolution. Moreover, the high-resolution crystal structure of a complex between Fab 5H2 and a C-terminal NHBA fragment (NHBA-C3) was also solved. To fully characterize the Fabs and the complex above, as well as NHBA alone, thermo-stability and binding studies were also performed.

These results allowed to perform an in-depth analysis of the biochemical and structural properties of the first human antibodies against a Bexsero antigen, and to our knowledge represent the first high-resolution structural characterization of a representative set of antibodies from the repertoire generated by immunization using an antigen that is already part of a commercially available vaccine. As such, this study provides insights that in addition to complement the full characterization of Bexsero, can also enable further and more advanced studies of protein design and engineering to expand our understanding of how vaccines work at the atomic level, with potential implications also for other future vaccine development programs.

# Materials & Methods

## Bacterial strains

The serogroup B *N. meningitidis* strains used in this study include the wild type MC58 (peptide 3), NZ98/254 (peptide 2), and 2996 (peptide 20). The *NHBA* gene fragments from *N. meningitidis* strains inserted into the pET-21b(+) vector (Novagen) were used for protein expression. Region 312-427 (hereafter indicated as NHBA-C3) of the GNA2132 sequence form strain 2996 was cloned in a pET21b(+) vector (Novagen), while region 114-332 (hereafter indicated as NHBA-114-332) from strain NZ98/254 was cloned into a pET24b(+) vector (Novagen). All NHBA expression constructs were cloned without the signal peptide and a C-terminal 6-His tag was inserted for purification purposes. The residues numbering used here refers to the full-length NHBA peptide 20 (p20), peptide 2(p2), and peptide 3 (p3), with corresponding UniProt accession numbers Q9JPP1, Q9JPH1 and Q9JQW0.

**Figure 2.1. Domain organization and constructs of NHBA.** Each expression construct had a C-terminal 6-His tag to enable affinity purification. The predicted domain organization and domain boundaries are shown schematically as boxes.

## NHBA-C3 mutants generation

Site-directed mutagenesis on NHBA-C3 (p20) DNA sequence was performed to produce the set of mutants listed below. Mutants were constructed using the polymerase incomplete primer extension cloning method (PIPE) method. The mutagenesis reactions were transformed into chemically competent E. coli Mach1™-T1$^R$ (Thermo Scientific). After sequencing, each plasmid was used to chemically transform *E. coli* BL21 (DE3) T1R cells (NEB) for protein production.

| **NHBA-C3 mutant** | **Plasmid characteristics** |
|---|---|
| NHBA-C3_Arg339Ala | pET21b(+) derivative for expression of recombinant NHBAp20 312-427, containing an Arg339Ala mutation, AmpR |
| NHBA-C3_Arg339Gly | pET21b(+) derivative for expression of recombinant NHBAp20 312-427, containing an Arg339Gly mutation, AmpR |
| NHBA-C3_Lys367Ala | pET21b(+) derivative for expression of recombinant NHBAp20 312-427, containing an Lys367Ala mutation, AmpR |
| NHBA-C3_Arg339Ala-Lys367Ala | pET21b(+) derivative for expression of recombinant NHBAp20 312-427, containing an Arg339Ala-Lys367Ala mutation, |
| NHBA-C3_Arg339Gly-Lys367Ala | pET21b(+) derivative for expression of recombinant NHBAp20 312-427, containing an Arg339Gly-Lys367Ala mutation, AmpR |
| NHBA-C3_Thr365Ala-Asp360Ala | pET21b(+) derivative for expression of recombinant NHBAp20 312-427, containing an Thr365Ala-Asp360Ala mutation, AmpR |
| NHBA-C3_Glu322Ala-Glu425Ala | pET21b(+) derivative for expression of recombinant NHBAp20 312-427, containing an Glu322Ala-Glu425Ala mutation, |

## Bioinformatics analyses

Various NHBA peptide sequences (p2, p3, p5, p1, p29, p18, p17, p10, p21) were submitted to several servers that perform secondary structure predictions, in each case using the standard parameters: Phyre2 [92, 93], SABLE [94], PSIPRED [95], and JPred4 [96]. Predictions of protein dynamics and disordered sequence regions were performed with DynaMine [97], IUPred [98] and MeDor [99].

## NHBA production and purification

Protein expression was performed in *E. coli* BL21D3 (T1r) cells (Invitrogen) [100], by using the EnPresso growth systems (BioSilta, Cambridge-shire, United Kingdom) supplemented with 100 mg/ml ampicillin. Bacteria were grown at 30°C for 12 h, and recombinant protein expression was induced by the addition of 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 25°C for 24 h. Bacteria expressing NHBA-114-332 were grown in DIFCO 3X medium (45g/L Yeast Extract, 4g/L $KH_2PO_4$, 16g/L $K_2HPO_4$, 15g/L glycerol) supplemented with 100mg/ml Kanamycin, 30 g/L glycerol and 0.25 g/L MgSO4, and the cells were then induced with 0.25 mM IPTG for 24 hours at 27 °C. Cells were harvested by centrifugation (6400g, 30 min, 4°C), re-suspended in 50 mM $NaH_2PO_4$, 300 mM NaCl, 20 mM imidazole, pH 7.5, and were lysed by sonication (Qsonica Q700). Cell lysates were clarified by centrifugation at 2800g for 30 min, and the supernatant was filtered using a 0.22 μm membrane (Corning filter system) prior to protein purification. NHBA proteins were purified at room temperature (RT, 18–26 °C) using an AKTA purifier 10 system (GE Healthcare) by Ni-affinity chromatography (5 mL HisTrap HP or HiTrap TALON crude, GE Healthcare) followed by size-exclusion chromatography on a Superdex 200 16/60 column equilibrated in 20 mM Tris–HCl, 150 mM NaCl, pH 8.0. A second purification step, consisting of an affinity purification based on the HiTrap Heparin HP (GE Healthcare), was included for the purification of NHBA-114-332 before the final size exclusion chromography (SEC) step common to all other proteins described here. The apparent molecular weight of the protein samples was calculated based on a calibration of the SEC column previously performed using a Standard Molecular Weight markers kit (Gel Filtration Standard, Biorad). The quality of the final NHBA samples was checked using 4–12% SDS–PAGE gradient gels in MES buffer.

## Fabs generation and purification

Gene fragments corresponding to the variable heavy and light chains of the α-NHBA human Fabs 12E1, 10C3, and 5H2 were codon optimized for mammalian expression and synthesized by Geneart (Life Technologies) with Eco31I site flanking gene extremities. After re-suspension in 50 μl nucleasefree water, the synthetic DNA strings were digested with the Eco31I restriction enzyme. Digested and purified DNA products were ligated into human pRS5a Igγ1, Igκ (for Fabs 5H2 and 12E1) or Igλ (for 10C3) expression vectors containing a human Ig gene signal peptide sequence and the Eco31I cloning site upstream of the human IgG1 and Igκ/λ constant regions. Fabs were cloned in *E. coli* DH5α cells, using a standard ligation protocol. In the pRS5a antibody expression vectors, the transcription is under the CMV promotor and clones can be selected based on resistance to ampicillin. The Fabs were expressed with cleavable C-terminal Strep-tag II on the heavy chain (WSHPQFEKGGSGGGSGGGSWSHPQFEK).

Fabs expression was achieved by transient transfection using Expi293 (Lifetechnologies) cells in suspension, according to the manufacturer's protocol. Equal amounts (15 μg each/30 ml of transfection volume) of vector DNA codifying FabH and FabL chains were used to transfect Expi293 cells. The cells were incubated at 37°C with a humidified atmosphere of 8% $CO_2$ in air on an orbital shaker rotating at 125 rpm. Cell culture supernatant was harvested after 36 and 72h post transfection, clarified with centrifugation for 30min at 4000 rpm and concentrated.

## Purification of fragments antigen binding

Human Fabs 5H2, 12E1, and 10C3 were all purified following three chromatographic steps, which include: i) initial affinity purification, ii) tag cleavage coupled with reverse affinity chromatography and iii) size exclusion chromatography. The first Strep-affinity chromatography step was performed using a Strep Trap HP column (GE Healthcare) in 100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 7.5, exploiting the highly selective double Strep tag at the C terminus of the Fab heavy chain. The elution of the Fabs was achieved by adding 2.5 mM D-desthiobiotin (IBA) in running buffer. Since the 28-amino-acid C-terminal peptide extension containing the double Strap-tag was thought to potentially interfere with crystallization, the sequence tags were removed by proteolytic cleavage at a preceding recognition sequence (ENLYFQG) using a recombinant tobacco etch virus (TEV) protease, prepared and purified in-house as described previously [101], at a ratio w/w Fab:TEV 50:1. The cleaved Fabs were separated by the free tags by performing a reverse affinity Strep-tag chromatography and collecting the flow-through. Fabs were further purified employing preparative size exclusion chromatography Superdex 200 26/600 (GE Healthcare) in buffer 20 mM Tris-HCl, 150 mM NaCl, pH 8. The homogeneity and the purity of the Fab proteins were confirmed by SDS-PAGE 4-12% Bis-Tris gel (Life Technologies) using MES 1X as buffer in both reducing and non-reducing conditions and performing a semi-analytical size exclusion chromatography on a Superdex 3.2/300 column. The final concentrations of the Fabs were calculated by measuring the absorbance at 280nm.

## Differential Scanning Calorimetry

The thermal stability of the NHBA proteins was assessed by differential scanning calorimetry (DSC) using a MicroCal VP-Capillary DSC instrument (GE Healthcare). NHBA samples were prepared at a protein concentration of 0.4 mg/mL in PBS buffer. The DSC temperature scan ranged from 20°C to 110°C, with a thermal ramping rate of 200°C per hour and a 4 second filter period. Data were analyzed by subtraction of the reference data for a sample containing buffer only, using the Origin 7 software. All experiments were performed in duplicate, and mean values of the melting temperature (Tm) were determined.

## Surface Plasmon Resonance

Surface plasmon resonance (SPR) was used to measure the binding affinity of full-length NHBA variants p2, p3 and p20, and of fragments NHBA-114-332 to the Fabs 12E1, 10C3 and 5H2. Moreover, SPR was used to assess the recognition of a set of NHBA-C3 mutants by Fab 5H2. All SPR experiments were performed using a Biacore T200 instrument equilibrated at 25 °C (GE Healthcare) and analysed through the Biacore T200 evaluation software (GE Healthcare).

Single-cycle kinetics (SCK) experiments were performed using a commercially available Human Fab Capture Kit (GE Healthcare) used to immobilize a mixture of monoclonal antibodies recognizing κ and λ subtypes of Fab fragment light chains by amine coupling on a carboxymethylated dextran sensor chip (CM5; GE Healthcare). A density level yielding ~5000-6000 response units (RU) was achieved. The immobilized anti-human Fab monoclonal antibodies were then used to capture ~800-1100 RU human Fab. Experimental SPR running buffer contained 10 mM Hepes, 150 mM NaCl, 3mM EDTA, 0.05% (v/v) P20 surfactant, pH 7.4 (HBS-EP). $K_D$ and kinetic parameters were calculated by performing a titration series of five consecutive injections of purified protein antigen diluted in HBS-EP at increasing concentration (flow rate of 30 μL/min) followed by a single final surface regeneration step with buffer containing 10 mM glycine pH 2.1 (180 s; 10 μL/min). All sensorgrams were corrected for nonspecific binding with readings from a protein-free reference flow cell. Data were analyzed using the standard SCK method [102] implemented by the Biacore T200 evaluation software (GE Healthcare). Each sensorgram was fitted with the 1:1 Langmuir binding model, including a term to account for potential mass transfer, to obtain the individual $k_{on}$ and $k_{off}$ kinetic constants. The individual values were then combined to derive the single averaged $K_D$ values reported.

To assess the binding of the NHBA-C3 mutants to Fab 5H2 captured onto the sensor chip described above, single equimolar injections of purified NHBA-C3 mutants at a concentration of 50 nM were performed with a flow rate of 10 µL/min. An anti-human Fab antibody–coated surface without captured Fab 5H2 was used as the reference channel. A blank injection of buffer only was subtracted from each curve, and reference sensorgrams were subtracted from experimental sensorgrams to yield curves representing specific binding. The data shown are representative of at least three independent experiments.

## Size-exclusion high-performance liquid chromatography (SE-HPLC) coupled with Multi-angle laser light scattering (MALLS)

SE-HPLC was performed at room temperature (18-26°C) on an analytical size exclusion TSK Super SW3000 column by loading 20 µl of each sample at a concentration of ~ 40µM. Samples were eluted isocratically in 0.1 M $NaH_2PO_4$, 0.4 M $(NH_4)_2SO_4$ buffer, at pH 7.5. Coupling SE-HPLC with Multi-angle laser light scattering (SE-HPLC/MALLS) NHBA p2 and NHBA-114-332 were analyzed for absolute molecular size in solution. Data analyses were carried out using the Astra V software (Wyatt) to determine the weight-average molecular mass (MW) in Daltons and the polydispersity index (MW/Mn) for each oligomer present in solution. Normalization of the MALLS detectors was performed in each analytical session by use of bovine serum albumin.

## Fab:NHBA complexes formation

Complexes between Fabs 10C3, Fab 12E1 or 5H2 with purified NHBA proteins p2, NHBA-114-332, and NHBA-C3, were freshly prepared by overnight co-incubation at 4°C followed by preparative size exclusion chromatography in 20 mM Tris-HCl, 150 mM NaCl, pH 8.0, on a Superdex 200 16/60 (GE Healthcare) column. Eluted fractions were run on an SDS-PAGE 4-12% Bis-Tris gel (Life Technologies) in both reducing and non-reducing conditions in order to assess protein content. Fractions containing both complex components were pooled and protein concentration was determined using the Bradford method (Protein Assay, Bio-Rad), using BSA as standard reference.

## Crystallization & X-ray data collection

Highly-purified samples of the different Fabs 10C3, 12E1, 5H2, NHBA p2, NHBA_114-332 and Fab:NHBA complexes were concentrated using centrifugal concentration devices with 10 kDa molecular weight cut-off membrane (Amicon Ultra-15 Centrifugal Filter Unit, Millipore). Concentrated samples were used for crystallization trials that were performed in 96-well low-profile Intelli-Plates (Art Robbins) using a nanodroplet sitting-drop vapour-diffusion format. Equal volumes (200 nL) of protein samples and crystallization buffers were mixed using a Crystal Gryphon liquid dispenser (Art Robbins

Instruments). Experiments were performed at 20°C and crystallization trays were imaged with the automatic imaging system RockImager-182 (Formulatrix). Crystallization trials were set up using different commercial sparse-matrix crystallization screens (Hampton Research and Molecular Dimension).

### ***Fabs crystallizations***

Crystals of Fab 12E1 (**Fig. 1.2A-B**) were obtained as multiple and stacked plates from a sample at a concentration of 19 mg/mL and using a condition containing 0.2 M potassium sodium tartrate, 0.1 M sodium citrate pH 5.6, 2 M ammonium sulfate. Single crystals were isolated from the clustered plates and were flash-frozen using 10% w/v of ethylene glycol as cryo-protectant prior to X-ray diffraction experiments. X-ray data collections were performed on beamline ID-23-1 of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France.

Crystals of Fab 10C3 (**Fig. 1.2C-D**) were obtained from a large number of different conditions (**Table 5**) and using a sample at a concentration of 17 mg/mL. Crystals were cryo-protected adding 20% w/v glycerol or 20% w/v ethylene glycol and then flash-frozen in liquid nitrogen. Diffraction data were collected on beamline ID-29 at ESRF.

Crystals of Fab 5H2 (**Fig. 1.2E-F**) were obtained from a sample at a concentration of 9 mg/mL in a condition containing 0.2 M ammonium sulfate, 0.1 M sodium-cacodylate pH 6.5, 30% w/v PEG 8K. In this case, the cryo-protectant used to freeze these crystals was made of crystallization mother liquor supplemented with 15% w/v glycerol. Data collection experiments were performed on beamline ID-29 of the ESRF.

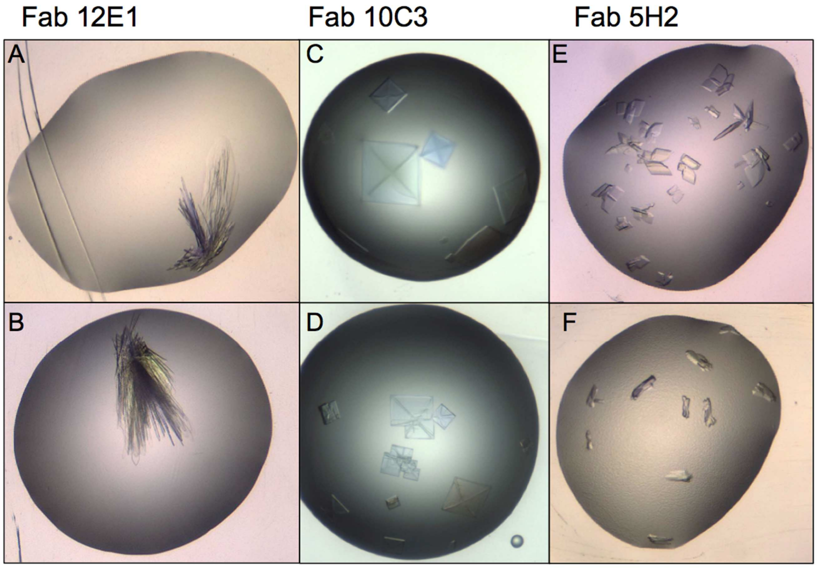


**Figure 2.2. Examples of anti-NHBA Fabs crystals.** Crystals of Fab 12E1 grown in (A) and (B): 0.2 M potassium sodium tartrate, 0.1 M sodium citrate pH 5.6, 2 M ammonium sulfate. Crystals of Fab 10C3 grown in (C) 0.2M lithium sulfate, 0.1M Bis Tris, 25% w/v PEG 3350 and (D) 0.2M sodium chloride, 0.1M Bis Tris, 25 % w/v, PEG 3350. Crystals of Fab 5H2 grown in (E) 0.2M ammonium sulfate, 0.1M sodium-cacodylate pH 6.5, 30% w/v PEG 8000 and (F) 0.2M Lithium sulfate, 0.1M Tris 8.5, 30 % w/v PEG 4000.

### *Fab:NHBA complexes: crystallization and soaking experiments*

Co-crystallization trials of Fab 10C3 with either full-length NHBA or NHBA fragments did not yield crystals suitable for diffraction experiments. Thus, for the complex 10C3:NHBA p2 *in-situ* proteolysis [103, 104] was attempted: the purified complex was concentrated at 30 mg/mL and α-chymotrypsin (JENA Bioscience) was added to the sample at a ratio protein:protease equivalent to 10000:1 w/w. The mixture was immediately used for setting up crystallization trials. Crystals appeared after ~60 days of incubation in three conditions (**Table 5**) and were analyzed for X-ray diffraction at the automated beamline ID30A-1 (MASSIF-1) at ESRF.

As an alternative strategy to obtain an Ab-Ag or Ab-epitope complex, unbound 10C3 crystals, obtained under different conditions (**Table 5**, group 3), were used for soaking experiments with a peptide made of a sequence corresponding to residues 247-271 of NHBA p2 protein (247-EKLSDADKISNYKKDGKNDGKNDKF-271). The peptide was synthetized by JPT Peptide Technologies with an acetylated N terminus and an amidated C terminus, and after resuspension in the mother liquor of the crystal target, this was soaked on top of unbound 10C3 crystals, using a 10-fold molar excess of peptide. Incubation times varied from 1, 5, 15, 30, 60, to 90 minutes. Soaked crystals that did not crack during the incubation were then cryo-protected using 15% w/v glycerol and flash-frozen in liquid nitrogen prior to X-ray diffraction experiments that were performed on beamline BM30A at ESRF.

Crystals of the complex 5H2:NHBA-C3 were obtained from a sample at a concentration of 5 mg/mL in 0.1 M TRIS pH 8.0 and 1.6 M lithium sulphate. Crystals were cryo-protected in mother liquor supplemented with 10% ethylene glycol and flash-frozen in liquid nitrogen prior to X-ray data collection experiments that were performed on beamline I02 at the Diamond Light Source, Oxford, UK

## Crystallographic structures determination

### *Fabs structures*

Where not specifically stated, all datasets were processed with XDS [105] and with the CCP4 suite of crystallographic programs [106], and manual building and refinement were performed using Coot [107], Phenix [108], and Buster [109].

Crystals of unbound Fab 12E2 diffracted up to 2.7 Å and belonged to space group $P2_12_12$ with a single molecule in the Asymmetric Unit (ASU), with a Matthews coefficient [110] of 2.66 $Å^3$/Da, for a solvent content of 53.8%. The structure of the unbound Fab 12E2 was solved using the automatic molecular replacement pipeline MoRDA[111] which automatically selected as search template the coordinates of the human anti-human Angiopoietin 2 Fab (PDB 4IMK). Full manual model building and refinement of the structure yielded final $R_{work}/R_{free}$ values of 18.0/26.3 % (**Table 2**). Excellent and continuous electron densities allowed modeling of the 12E1 Fab molecule including residues Gln1-Lys216 for the H chain, and Glu1-Arg216 of L chain. The final C-terminal six residues of the H chain and three residues of the L chain could not be modelled due to lack of electron densities.

Crystals of unbound Fab 10C3 diffracted up to 1.7 Å and belonged to spacegroup $P2_12_12_1$. The structure was solved by MR in *Phaser* [112], using the coordinates of the human anti-HIV-1 clade A/E gp120 Fab N5-i5 (PDB 4H8W) as the search model. The refined coordinates of this unbound 10C3 structure were later used in several attempts to solve the structures of 10C3:NHBA complexed and soaked crystals. These attempts were made either by Fourier difference methods [113], where unit cell dimensions of the different crystals were isomorphous, or by MR when otherwise. Since no bound NHBA or fragments thereof could be observed in the electron density maps of all datasets collected from either complexed or soaked crystals, the highest resolution (1.5 Å) dataset was selected for full refinement and to be used for i) the structure analyses presented in this work and ii) the data deposition in the PDB. These crystals belonged to space group $P2_12_12_1$, with a calculated Matthews coefficient of 2.33 Å$^3$ /Da that is compatible with the presence of one Fab molecule in the ASU (solvent content 47.3%). Final coordinates were refined to final $R_{\text{work}}$ and $R_{\text{free}}$ values of 18.0% and 21.1%, respectively (**Table 2**), and comprise residues 2–215 of the H chain and 2–211 of the L chain. The structure showed excellent electron densities throughout most the Fab chains, except for residues belonging to loop 130-134 and to the terminal linker of the H chain.

Unbound Fab 5H2 crystals diffracted to a resolution of 1.8Å and the diffraction data were automatically processed by EDNA framework fast processing system [114] which uses a sequence of XDS, XSCALE, Pointless, Aimless, Truncate. The crystals belonged to the monoclinic space group $P\ 2_1$, and a calculated Matthews coefficient [110] of 2.29 Å$^3$/Da, for a solvent content of 46.3%, suggested the presence of three Fab molecules in the ASU (chains H+L, M+N, O+P). The structure of unbound Fab 5H2 was solved by molecular replacement using the software package MoRDa [111] which automatically selected as templates input coordinates those of the anti-HIV-1 A32 Fab and of the anti-integrin LFA-1 FabAL-57 (PDBs 3TNM and 3HI6). The final model was refined to $R_{\text{work}}$ and $R_{\text{free}}$ values of 18.6% and 22.6%, respectively (**Table 2**). For the refinement both non-crystallographic symmetry (NCS) restraints, and Local Structure Similarity Restraints (LSSR) [109] were used. Excellent electron densities could be observed for residues 1-224 and 1-211 of the H and the L chains, for residues 2-224 and 1-210 of M and N chains and 1-224 of O chain, respectively. Residues 49-60 of the light chain P could not be modelled due to lack of density.

### *Fab:NHBA complex structure*

Crystals of the 5H2:NHBA-C3 complex diffracted to a resolution of 2.8 Å and belonged to the space group $P\ 6_122$. Two copies of the complex were present in the asymmetric unit with a Matthews coefficient of 2.91 Å$^3$/Da and a solvent content of 62.6%. Diffraction data of the 5H2:NHBA-C3 complex were processed by iMosflm [115] and scaled using AIMLESS in the CCP4 software suite [106]. Initial molecular replacement attempts using the program Phaser [116] and input coordinates from the antibody 5-51/O12 (PDB 4KMT [117]), with which 5H2 shared 82% sequence identity, and single models from the NMR structure 2LFU, failed. However, when the template coordinates for the Fab were substituted with those from the unbound 5H2 structure solved here, molecular replacement in Phaser gave a correct solution with the structure of the complex Fab 5H2:NHBA-C3, which was later refined to final $R_{\text{work}}$ and $R_{\text{free}}$ values of 19.2% and 23.8% (**Table 8**). NCS torsion-angle restraints [108] were used in the refinement of 5H2:NHBA-C3 complex, and excellent electron densities could be observed for almost the entire complex (residues Met290-Asp405 for NHBA-C3, residues Gln1-

Cys226 and Glu1-Glu212 for the Fab 5H2 H and L chain, respectively). The only regions that could not be modelled due to lack of electron density were those of residues 88-90 and of the C-terminal His-tag of both NHBA-C3 molecules present in the ASU.

## X-ray structure analysis

The quality of all the final refined models described in this study was assessed using MolProbity [118]. Fab elbow angles were calculated with phenix.fab_elbow_angle [108] and protein-protein interface areas were analysed and calculated using the Protein Interfaces, Surfaces and Assemblies service (PISA) available at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html, [119]). Pairwise structural comparisons of unbound-5H2 molecules (chains HL, MN, OP) with bound-5H2 (chains HL, IM), NHBA-C3 (chains A, B) and NHBA-C1 (PDB entry: 2LFU [64]) were performed using the Secondary Structure Matching (SSM) algorithm within Coot [107]. Electrostatic potentials were calculated using the APBS PyMOL plugin [120]. Figures were generated using PyMOL (http://www.pymol.org). Alignments were generated with ALINE [121].

# NHBA sequences

- **NHBA-NZ p2 (Uniprot accession number Q9JPH1)**

  MFKRSVIAMACIFALSACGGGGGGSPDVKSADTLSKPAAPVVSEKETEAKEDAPQAGSQG
  QGAPSAQGGQDMAAVSEENTGNGGAAATDKPKNEDEGAQNDMPQNAADTDSLTPNHTPAS
  NMPAGNMENQAPDAGESEQPANQPDMANTADGMQGDDPSAGGENAGNTAAQGTNQAENNQ
  TAGSQNPASSTNPSATNSGGDFGRTNVGNSVVIDGPSQNITLTHCKGDSCSGNNFLDEEV
  QLKSEFEKLSDADKISNYKKDGKNDGKNDKFVGLVADSVQMKGINQYIIFYKPKPTSFAR
  FRRSARSRRSLPAEMPLIPVNQADTLIVDGEAVSLTGHSGNIFAPEGNYRYLTYGAEKLP
  GGSYALRVQGEPSKGEMLAGTAVYNGEVLHFHTENGRPSPSRGRFAAKVDFGSKSVDGII
  DSGDGLHMGTQKFKAAIDGNGFKGTWTENGGGDVSGKFYGPAGEEVAGKYSYRPTDAEKG
  GFGVFAGKKEQD

- **NHBA-MC58p3 (Uniprot accession number Q9JQW0)**

  MFKRSVIAMACIFALSACGGGGGGSPDVKSADTLSKPAAPVVSEKETEAKEDAPQAGSQG
  QGAPSAQGSQDMAAVSEENTGNGGAVTADNPKNEDEVAQNDMPQNAAGTDSSTPNHTPDP
  NMLAGNMENQATDAGESSQPANQPDMANAADGMQGDDPSAGGQNAGNTAAQGANQAGNNQ
  AAGSSDPIPASNPAPANGGSNFGRVDLANGVLIDGPSQNITLTHCKGDSCSGNNFLDEEV
  QLKSEFEKLSDADKISNYKKDGKNDKFVGLVADSVQMKGINQYIIFYKPKPTSFARFRRS
  ARSRRSLPAEMPLIPVNQADTLIVDGEAVSLTGHSGNIFAPEGNYRYLTYGAEKLPGGSY
  ALRVQGEPAKGEMLAGAAVYNGEVLHFHTENGRPYPTRGRFAAKVDFGSKSVDGIIDSGD
  DLHMGTQKFKAAIDGNGFKGTWTENGSGDVSGKFYGPAGEEVAGKYSYRPTDAEKGGFGV
  FAGKKEQD

- **NHBA-2996 p20 (Uniprot accession number Q9JPP1)**

  MFERSVIAMACIFALSACGGGGGGSPDVKSADTLSKPAAPVVAEKETEVKEDAPQAGSQG
  QGAPSTQGSQDMAAVSAENTGNGGAATTDKPKNEDEGPQNDMPQNSAESANQTGNNQPAD
  SSDSAPASNPAPANGGSNFGRVDLANGVLIDGPSQNITLTHCKGDSCNGDNLLDEEAPSK
  SEFENLNESERIEKYKKDGKSDKFTNLVATAVQANGTNKYVIIYKDKSASSSSARFRRSA
  RSRRSLPAEMPLIPVNQADTLIVDGEAVSLTGHSGNIFAPEGNYRYLTYGAEKLPGGSYA
  LRVQGEPAKGEMLAGTAVYNGEVLHFHTENGRPYPTRGRFAAKVDFGSKSVDGIIDSGDD
  LHMGTQKFKAAIDGNGFKGTWTENGGGDVSGRFYGPAGEEVAGKYSYRPTDAEKGGFGVF
  AGKKEQD

- **NHBA-C3 p20**

  MASMLAGTAVYNGEVLHFHTENGRPYPTRGRFAAKVDFGSKSVDGIIDSGDDLHMGTQKF
  KAAIDGNGFKGTWTENGGGDVSGRFYGPAGEEVAGKYSYRPTDAEKGGFGVFAGKKEQDL
  EHHHHHH

- **NHBA-114-332 p2**

  MPNHTPASNMPAGNMENQAPDAGESEQPANQPDMANTADGMQGDDPSAGGENAGNTAAQGT
  NQAENNQTAGSQNPASSTNPSATNSGGDFGRTNVGNSVVIDGPSQNITLTHCKGDSCSGN
  NFLDEEVQLKSEFEKLSDADKISNYKKDGKNDGKNDKFVGLVADSVQMKGINQYIIFYKP
  KPTSFARFRRSARSRRSLPAEMPLIPVNQADTLIVDGEALEHHHHHH

# Results

## NHBA production and characterization

### *NHBA full length contains predicted partially unfolded regions*

Bioinformatics analyses were performed on a panel of long NHBA sequences (p1, p2, p3, p5, p29) to predict putative regions of disorder and to expand the analysis previously performed only on the short NHBA p20 sequence [64]. The Phyre2 [92] suite of tools for secondary structure prediction indicated a substantial lack of secondary structures elements in the region covering the first ~250 amino acids of NHBA (N-terminal domain), while the predictions for residues spanning from 341 to 492 (C-terminal domain) are in agreement with the presence of β secondary structures elements as observed in the experimentally-determined NMR structure (PDB 2LFU) [64](**Fig. 3.1**). Comparable outputs were obtained by using other secondary structure prediction tools such as SABLE [94], PSIPRED [95] and JPred4[96] (**Fig. 3.1**). Consistently, methods for the prediction of disorder (IUPred [98] and MeDor [99]) as well as of protein dynamics prediction (DynaMine [97]) indicated that the first 200 amino acids of NHBA are extremely flexible (**Fig. 3.2**). In agreement with these predictions, the NHBA N-terminal domain is enriched in charged and hydrophilic residues, which, as previously reported, is a common feature of intrinsically disordered proteins (IDPs) [122-124]. Moreover, these sequence-based predictions also suggest the presence of secondary structure elements (α-helices and β-strands) for the region made of residues ~240-340. It can be hypothesized that these 100 amino acids preceding the β-barrel can acquire a transient and context dependent secondary structure, similarly to other IDPs that fold transiently in the presence of specific interactors or under certain environmental conditions, in order to perform their biological function [122, 125-127].

Interestingly, it was previously shown how predictions of protein disorder for regions longer than 40 residues give less than 6% of false positives [128].

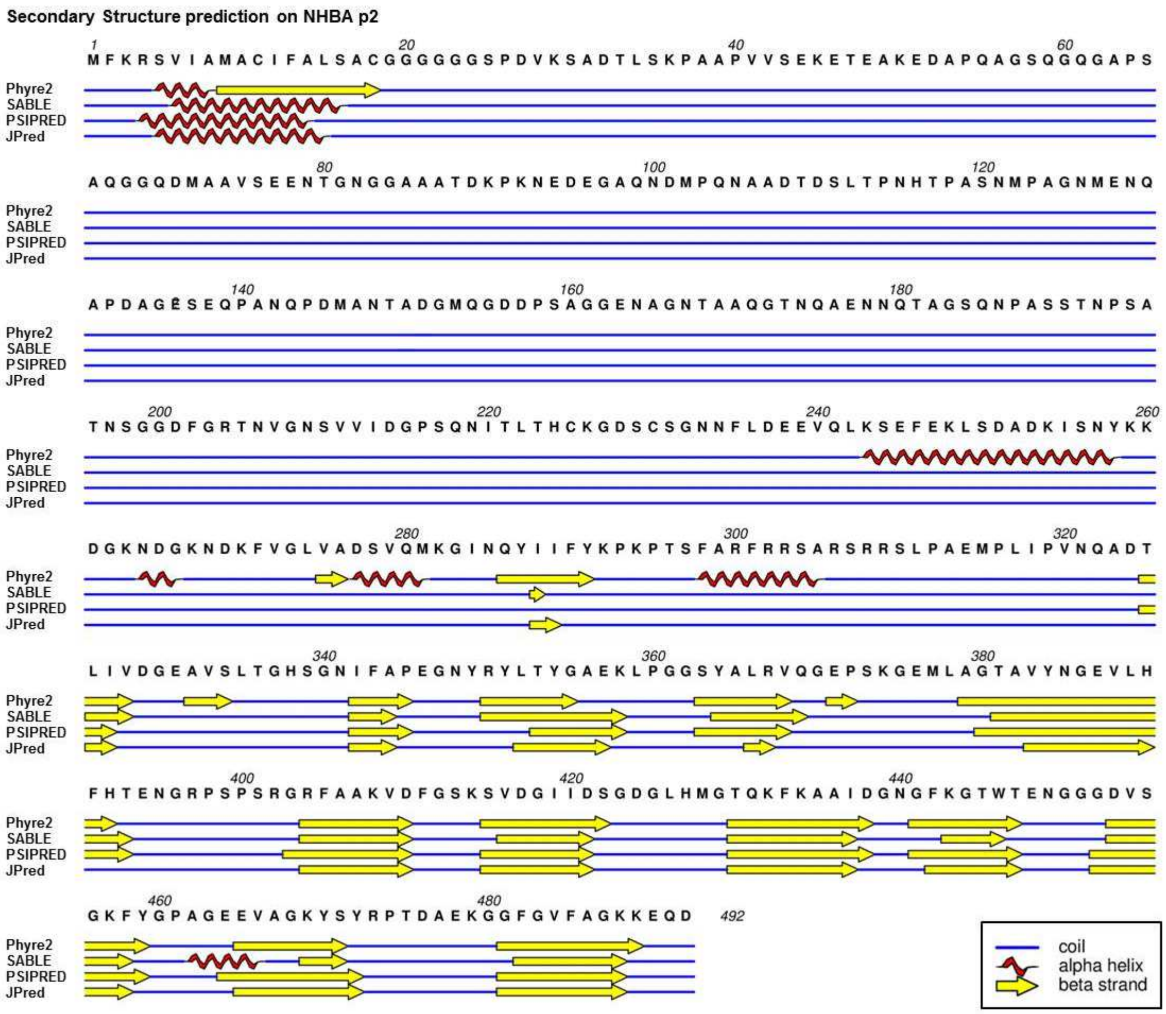


**Figure 3.1. NHBA p2 secondary structure predictions.** Outputs of secondary structure predictions on NHBA p2 sequence performed with Phyre2, SABLE, PSIPRED and JPred are summarized in the figure.

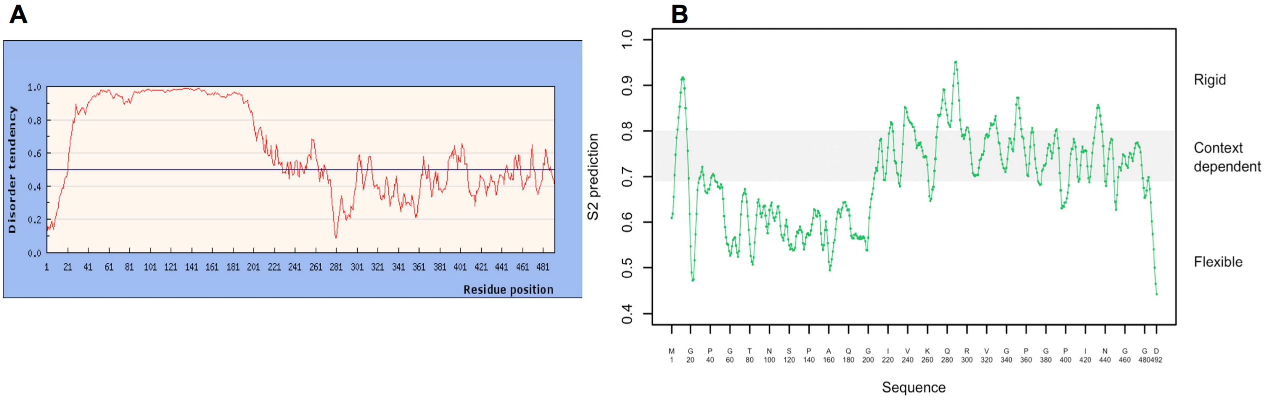


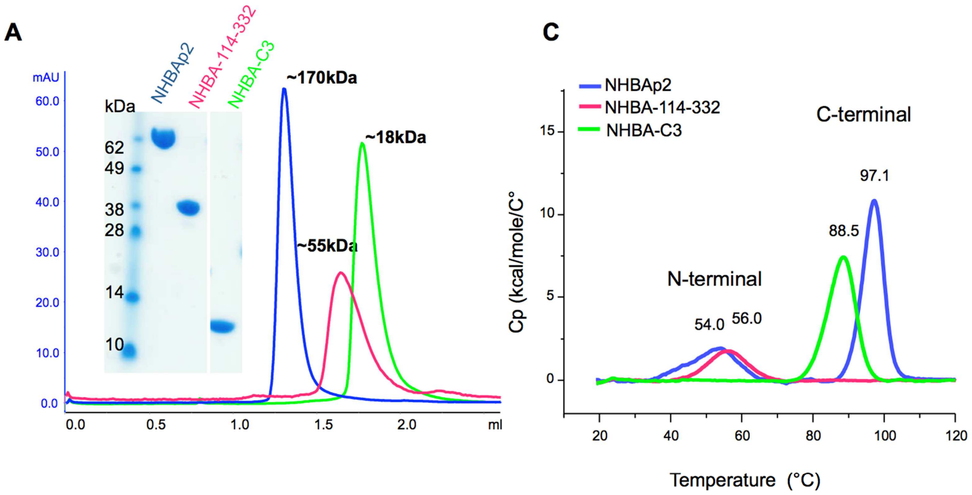
**Figure 3.2. NHBA p2 disorder predictions**. The outputs from the analyses performed with Disopred and with Dynamine are shown in **A** and **B**.

### ***Production and characterization of protein constructs for structural studies***

In order to gain molecular insights into the NHBA 3D organization, we sought to determine its X-ray structure using full length and sub-full length constructs, both alone and in complex with human Fabs. To perform crystallographic studies, high amount of pure proteins are needed, therefore recombinant NHBA constructs were produced in *E. coli* and purified as described in the Materials and Methods, yielding highly pure samples as shown by SDS-PAGE (**Fig. 3.3A**). While the band for the NHBA-C3 sample corresponded to the expected molecular weight (MW) of ~13kDa, both the NHBA p2 and NHBA-114-332 samples were observed to run at higher MW (~62kDa and ~38kDa, respectively) with respect to their calculated theoretical values (49.9kDa and 24kDa). In agreement, samples of NHBA p2 and NHBA-114-332 showed anomalous SEC profiles, eluting in correspondence of significantly higher MW than expected (~170kDa and ~55kDa, respectively) (**Fig. 3.3A**). Therefore, further analyses were performed on these samples by analytical size-exclusion high-performance liquid chromatography (SE-HPLC) coupled with multi-angle laser light scattering (MALLS), which showed how NHBA p2 presented a single species ≥ 97% pure with an absolute molecular mass of 51 kDa while NHBA_114-332 presented similar purity and an absolute molecular mass of 24 kDa (**Fig. 33B).** Therefore, both purified proteins were monomeric in solution, despite their anomalous properties both in SDS-PAGE and in SEC. Generally, hydrodynamic techniques such as size exclusion can aid in identifying IDPs, since they report on the radius of the protein which for IDPs is often larger than expected [129]. In conclusion, the behavior of constructs NHBA p2 and 111-342 during the purification seems in agreement with the previously predicted high flexibility and/or disorder of their N-terminal region [130].

To further characterize putative disorder in NHBA, and since protein stability often correlates with crystallization propensity [131], the thermal stability of NHBA was studied using differential scanning calorimetry (DSC). While NHBA full length (p2) presented two distinct peaks corresponding to melting temperature ($T_m$) transitions at 54.0 and 97.1 °C, the N- and C-terminal constructs (NHBA-114-332 and NHBA-C3) showed only one single $T_m$, at 56 or 88°C, respectively. These results suggested that the C-terminal domain of NHBA, previously shown to fold as a well-structured β-barrel [64] is significantly more thermo-stable than the N-terminal region, where instead the lower $T_m$ of the N-terminal domain agrees well with the predicted lack of fully structured regions.

Both purified samples of NHBA p2 and NHBA_114-332 were used to set up several crystallization trials as described in Materials & Methods, but these experiments did not yield crystals suitable for further structural studies.



**B**

| SEC-MALLS | NHBAp2 | NHBA_114-332 |
|---|---|---|
| Elution time (min) | 14.5 | 16.5 |
| Absolute Mw | $5.151\times10^{4}$ (±0.852%) | 2.440×104 (±2.040%) |
| Polydispersity Mw/Mn | 1.009 (±1.185%) | 1.005 (±2.886%) |
| Homogeneity Mz/Mn | 1.019 (±2.103%) | 1.010 (±5.006%) |

**Figure 3.3. Purification and characterization of NHBA and NHBA constructs.** (**A**) Overlay of size-exclusion chromatography profiles of purified NHBA samples. Above each peak the corresponding apparent molecular weight determined by calibration of the column using standard molecular weight markers (Bio-rad) is reported. The eluted material was run on a SDS-PAGE gel under non-reducing conditions (inset). (**B**) Table reporting the SEC-MALLS results for the determination of the elution time on an analytical column, the absolute molecular weight and polydispersity of NHBA p2 and NHBA_114-332 samples. In both cases, the MALLS data clearly indicated a single monodisperse species with an absolute molecular weight corresponding to their respective monomeric form. (**C**) DSC profiles of NHBA samples with the relative $T_m$ associated to each peak. The first transition is associated to the N-ter unfolding and the second to the C-ter unfolding. Samples are colored as follows: NHBA p2 in blue, NHBA_114-332 in magenta, and NHBA-C3 in green.

## Human Fabs production and characterization

Fragments antigen binding (Fabs) used in this study were purified using three chromatographic steps (see Materials and Methods), and the same purification protocol was applied to all three Fabs studied (10C3, 12E1, and 5H2) with no noticeable differences in the behavior and final quality levels of the purified samples. This protocol allowed achieving high amounts of purified Fab samples, with final average yields of 6 mg of Fabs from expression culture media of 100 ml in volume. Only one representative example of a Fab purification is described below.

SEC elution profiles of the Fabs consisted in a single symmetric peak with a retention time corresponding to an apparent molecular weight of ~35kDa (**Fig 3.4**). The purity of the Fabs was also checked by SDS–PAGE both in reducing and non-reducing conditions. As expected, while under reducing conditions distinct bands could be distinguished corresponding to the single Fab heavy and light chains, under non-reducing conditions a single band was detected at ~49kDa, which corresponds to a Fab heterodimer (**Fig 3.4**). The homogeneity of the purified Fabs were also confirmed by analytical SEC, which revealed a single symmetrical peak eluting at an apparent molecular weight of 35kDa, in agreement with the elution from the preparative SEC.

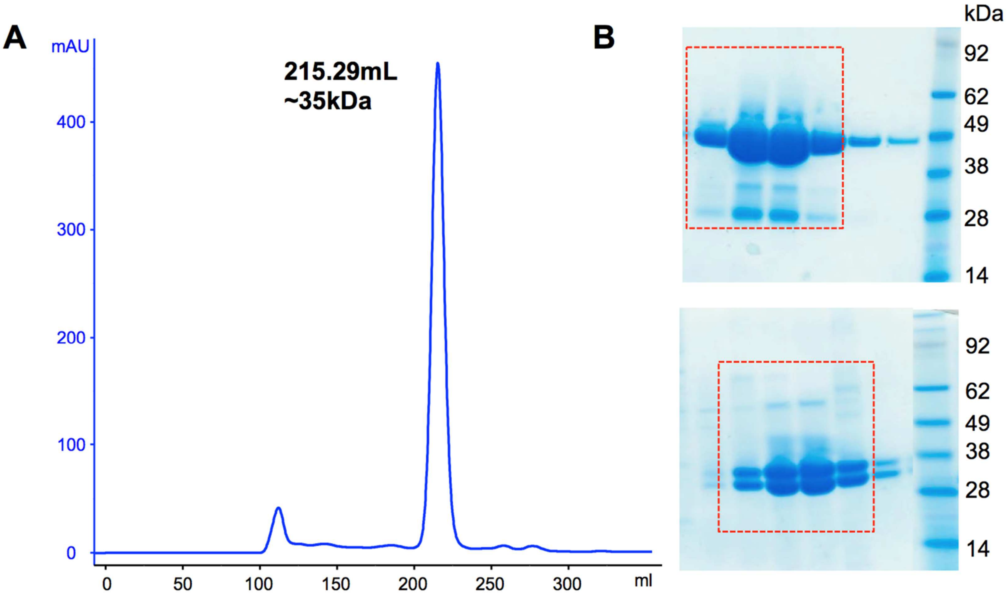


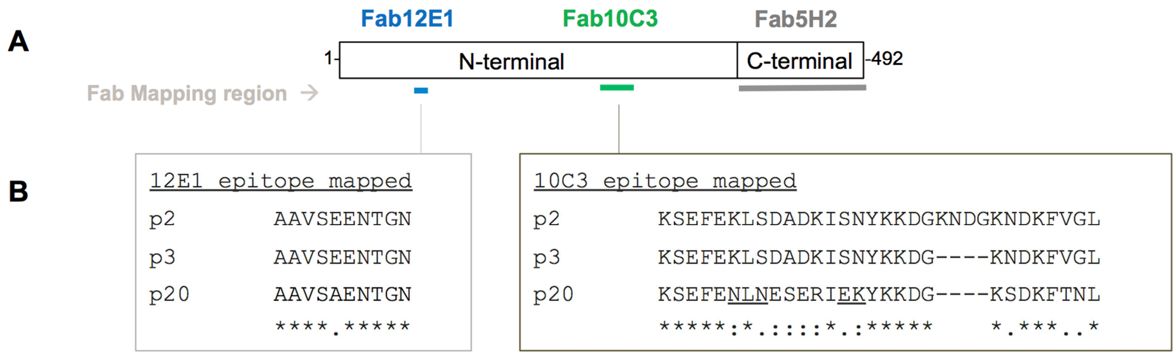
**Figure 3.4. Purification of Fab 10C3.** (**A**) Preparative size-exclusion chromatography profile of Fab 10C3. (**B**) Eluted fractions analyzed by SDS-PAGE under non-reducing (upper panel) and reducing (lower panel) conditions. A single band corresponding to the dimeric Fab is observed under non-reducing conditions; two separates bands corresponding to Fab HC and LC are observed under reducing conditions, where disulfides and inter-molecular bonds are disrupted by DTT and heat treatment. Similar results were obtained for 12E1 and 5H2.

## Fab:NHBA interactions studies

Mapping of the binding of Fabs 12E1, 10C3, and 5H2, to NHBA, has been performed by protein chip and it has been recently published [90]. This study showed how 12E1 recognizes a stretch of N-terminal residues of NHBA, while Fab 10C3 binds to an N-terminal region immediately before the boundary with the C-terminal domain, and 5H2 recognizes the full C-terminal domain. To further investigate the binding between the Fabs and NHBA, affinity constants and binding kinetics were calculated by surface plasmon resonance (SPR), performing single-cycle kinetics experiments (SCK). Three representative variants of NHBA were selected: two long (p2, the vaccine variant [13], and p3, the first *N. meningitidis* serogroup B sequenced strain [8]) and one short (p20, from which the NHBA-C3 construct was designed). These experiments showed that NHBA p2 and p3 are recognized by

Fabs 10C3, 12E1 and 5H2 with high affinities (pico to nano molar range) (**Table 1**). In particular, Fab 5H2 showed overall the highest binding affinity for all the tested NHBA variants.

The SPR results are also in agreement with the previously reported epitope mapping studies [90, 91] (**Fig. 3.5**). Specifically, SPR confirms that Fab 5H2 recognizes the conserved C-terminal β-barrel of NHBA, likely by binding to a very conserved epitope as inferred from its ability to recognize with high and comparable affinities all the variants tested (see for further details the results on section *NHBA β-barrel conservation and 5H2 cross-reactivity*, and **Fig. 3.24**). Similar considerations can be made for Fab 12E1, which maps onto the NHBA N-terminal domain and recognizes a conserved putative linear epitope. Finally, Fab 10C3 recognized both the long variants p2 and p3 but did not bind the short variant p20. A fully conserved region between NHBA p2 and p3 (**Fig 3.5**) that is not fully conserved in p20 could help explain the specificity of Fab 10C3 and define a “core epitope” to residues 243-KSEFEKLSDADKISNYKKDG-262.



```
12E1 epitope mapped
p2          AAVSEENTGN
p3          AAVSEENTGN
p20         AAVSAENTGN
            ****.*****
```

```
10C3 epitope mapped
p2          KSEFEKLSDADKISNYKKDGKNDGKNDKFVGL
p3          KSEFEKLSDADKISNYKKDG----KNDKFVGL
p20         KSEFENLNESERIEKYKKDG----KSDKFTNL
            *****:*.::::*.:*****    *.***..*
```

**Figure 3.5. Epitope mapping studies of NHBA p2.** (**A**) The top boxes show a schematic representation of NHBA p2, with colored lines referring to the regions of the 12E1 epitope (blue), 10C3 (green), and 5H2 (grey) as reported in [90] [91](**B**) The amino acid sequences of the 12E1 (left) and 10C3 (right) epitopes for variants p2, p3, and p20, are aligned. Identical (*), conserved (:) and semiconserved (.) residues are indicated below the alignment.

## Structural characterization of human Fabs and of a NHBA-Fab complex

### *Human Fab 12E1*

Crystals of unbound Fab 12E1 diffracted X-rays to a resolution of 2.7 Å. A full data set was acquired from a single 12E1 crystal fragment isolated from a cluster of plates and used for successful structure determination by molecular replacement (see *Material and Methods* and **Table 2**). The overall architecture and fold of the 12E1 Fab structure is consistent with the canonical β-sandwich immunoglobulin fold composed by two chains (H and L) and four domains (VL, CL, VH and CH, where

L indicates light chain, H heavy chain, V variable domain and C constant domain) (**Fig. 3.6**)**,** with four pairs of disulfide bridges that link residues Cys22-Cys96 in the VH and Cys142-Cys198 in the CH, and residues Cys23-Cys93 and Cys139-Cys199 in the VL and CL respectively The structure was analyzed with PISA (Protein Interfaces, Surfaces and Assemblies [119]) , which revealed an interaction surface between the H and L chains of 1735.5 Å$^2$, and the presence of 10 intermolecular hydrogen bonds (**Table 3**).

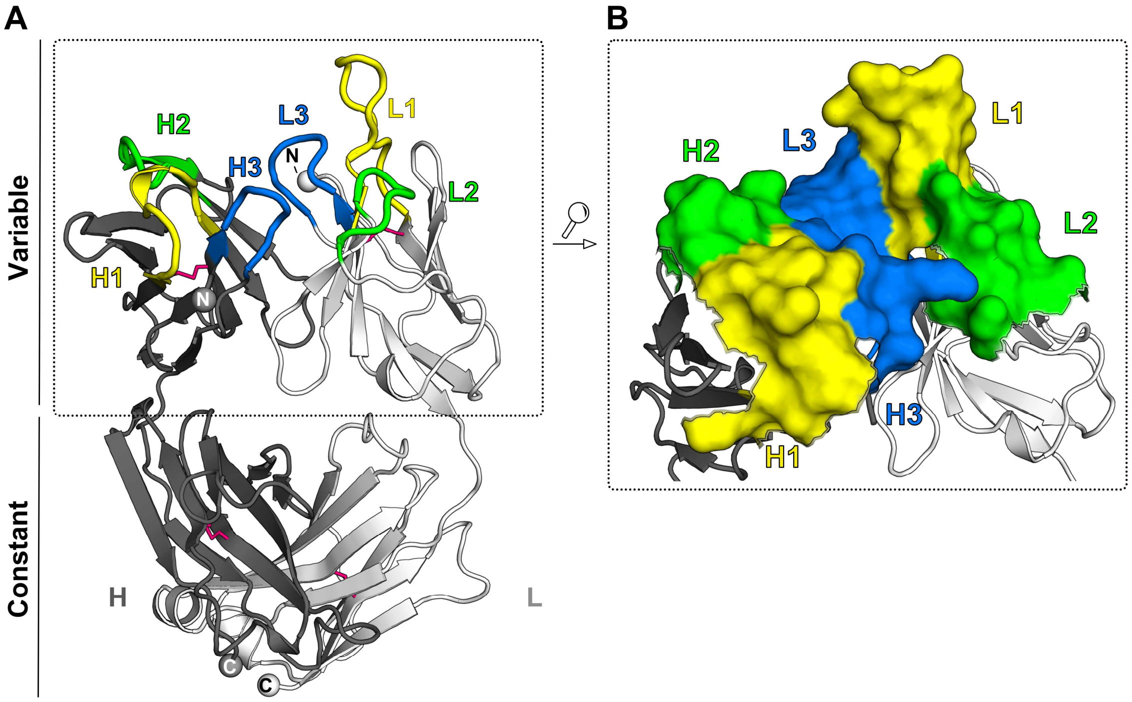


**Figure 3.6. Fab 12E1 crystal structure.** (**A**) Overall structure of Fab 12E1. The light (L) and heavy chains (H) are colored in light and dark grey respectively. CDR-L1and CDR-H1are colored in yellow, CDR-L2 and CRD-H2 are colored in green, CDR-L3 and CDR-H3 are shown in blue. Disulfide bonds are depicted with magenta sticks (**B**) Surface representation of a zoomed view of the 12E1 CDRs forming a concave surface at the interface between heavy and light chain.

To analyze and annotate the putative paratope of 12E1, the definition of CDRs as given by North et al. was used [79]. In this work, the classification of CDR "canonical" conformations [77, 78] have been recently revisited in light of the larger amount of antibody structural information available nowadays with respect to the canonical classifications that were originally introduced in the late 1990s. Therefore, this work now provides an enhanced and more detailed classification, resulting in revisited clustering of the structures of CDR loops. According to this new definition, regions of the Ramachandran plot that define CDRs clusters by conformation are annotated as follows: B for β-sheet region, P for polyproline II, A for α-helix, D for δ region (near α-helix but at more negative values of ϕ), L for left-handed helix, and G for γ region (ϕ>0° excluding the L and B regions) (see Fig 4 in reference [79]).

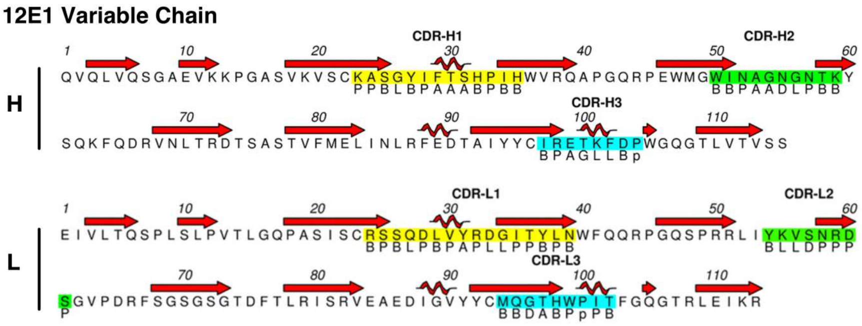


**Figure 3.7. 12E1 CDR regions.** CDRs for the variable heavy chain (**Top**) and for the variable light chain (**Bottom**) are shown with highlights over the sequence, while loop conformations are shown below the sequence, and secondary structures for each residue are shown above. Lower-case letters in loop conformation indicate cis residues.

In the H chain of 12E1, the CDR-H1 is made of residues Lys23–His35 and belongs to the cluster H1-13-1 with loop conformation PPBLBPAAABPBB, while the CDR-H2 is made of residues Trp50–Lys59, and belongs to cluster H2-10-1 with loop conformation BBPAADLPBB. Finally, the CDR-H3 is formed by Ile97-Pro104 (loop conformation BPAGLLBp) with the anchor region not clearly belonging to any of the classes defined by North et al (**Fig. 3.7A**). In the L chain, residues Arg24–Asn39 form a remarkably long CDR-L1 that falls within the L1-16-1 cluster and with loop conformation BPBLPBPAPLLPPBBPB, while the CDR-L2 is made of residues Tyr54–Ser61 and is within the L2-8-1 cluster and loop conformation BLLDPPPP. The CDR-L3 is made of residues Met94–Thr102 with conformation similar to the L3-9-cis7-1 cluster and loop conformation BBDABPpPB (**Fig. 3.7B**).

The overall architecture of the putative paratope shows a concave topology, with a deep groove at the interface between heavy and light chain delimited by the CDR-H3 on one side and by all of the CDR-Ls on the other side (**Fig. 3.6B**), with the CDR-L1 being the region that protrudes the most. The CDRs region has a calculated accessible surface area (ASA) of ~3,850 Å$^2$, as obtained with PISA [119]. The most abundant residues forming the exposed surface are Lys (13.35%), Arg (11.74%), Ser (9.83%), Tyr (9.70%), and Asn (7.56%) (**Table 4**). While the enrichment with aromatic residues and Ser of Fab paratopes is in agreement with other antibody paratope surfaces previously described [75, 132-135], the location of Ser on the surface of Fab 12E1 appears to be mostly peripheral, while Tyr and Trp are more equally distributed (**Fig. 3.8A**). A more noticeable feature of the 12E1 structure is the presence of a high number of positively-charged residues in the proximity of the putative paratope, mainly Arginine and Lysine that together form 28.8% of the entire paratope surface accessible area

(**Table 4**). This feature is not common among other Fabs, as long-chain hydrophilics are not frequently found in antibody paratopes [136], and it suggests a possible role in the recognition of NHBA (**Fig. 3.8B**). Also, the presence of these positively-charged patches in the paratope of 12E1 allows one to speculate on the apparently good complementarity with the overall acidic nature of the 10-residue epitope previously mapped (residues AAVSEENTGN [91]).

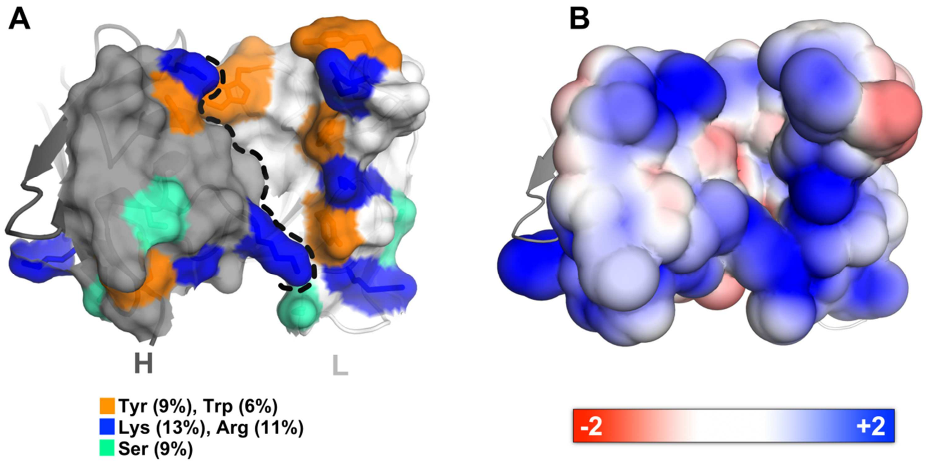


**Figure 3.8. 12E1 CDRs detailed view**. (**A**) Top view of the 12E1 CDR region, with the most represented residues colored in orange for Trp and Tyr, in blue for Lys and Arg, and in green for Ser. A dotted line separates HC from LC. (**B**) Same view as in A, with the electrostatic potential mapped onto the CDRs surface. The electrostatic potential distribution was calculated with APBS [120], where red and blue surfaces show negative and positive charges as contoured in the range from $-3\ k_BTe^{-1}$ (red) to $+3\ k_BTe^{-1}$ (blue), while white surfaces show neutral potentials.

Additionally, attempts of co-crystallizing 12E1 in complex with NHBA p2 were performed but did not yield crystals, while the low quality and number of crystals obtained for the unbound 12E1 did not allow to pursue further soaking experiments with epitope peptides.

### *Human Fab 10C3*

Crystals of unbound Fab 10C3 were obtained under a variety of conditions (**Table 5**, group 1), resulting in 7 initial complete X-ray diffraction datasets that were collected at resolutions between 1.7 and 2.2 Å. Among these, dataset #7 (**Table 5**), which diffracted at the highest resolution (1.7 Å), was selected for structure determination by MR (see *Materials and Methods*). The refined structure of unbound 10C3 was then used as a template to solve the structures of all the other 6 datasets either by Fourier difference methods [113], where unit cell dimensions were isomorphous, or by MR when not isomorphous. These refined coordinates were also later used as the model input template for attempts at solving the structure of the 10C3:NHBA complex by MR.

High resolution epitope mapping of the binding of Fab 10C3 to NHBA was attempted by performing a number of co-crystallization experiments of NHBA-10C3 complexes, as well as by soaking NHBA derived peptides onto crystals of unbound Fab 10C3. First, size exclusion chromatography was used to isolate complexes formed between Fab 10C3 and constructs NHBA_p2 or NHBA_114-332. The chromatographic profiles of the complexes showed in both cases one peak corresponding to the complexes 10C3:NHBA_p2 (apparent MW ~170kDa, calculated MW 96.2kDa) or 10C3:NHBA_114-332 (apparent MW ~85kDa, calculated MW 71.2kDa) and a second peak corresponding to the excess unbound Fab 10C3 at an apparent molecular weight of ~35kDa (calculated MW 46.3kDa) (**Fig. 3.9**). To explain the apparent molecular weights that were not in close agreement with those calculated, we performed SDS-PAGE on the eluted fractions in reducing conditions, which confirmed how in both cases the first peak contained both NHBA and 10C3, while the second peak contained Fab only. This was also in agreement with an overlay of elution profiles of NHBA_p2, 114-332, and 10C3 derived from separate runs (**Fig. 3.9**). Importantly, as already discussed above and as previously reported [130], purified NHBA samples eluted at apparent molecular weights larger than expected. Also, in a separate run unbound Fab 10C3 eluted after ~84.75.mL, but it could also be detected by SDS-PAGE in fractions belonging to the putative complexes that were eluting either at ~65 mL or ~73 mL. These results strongly suggest that both the complexes 10C3:NHBA_p2 and 10C3:NHBA_p2_111-332 were successfully formed and isolated.

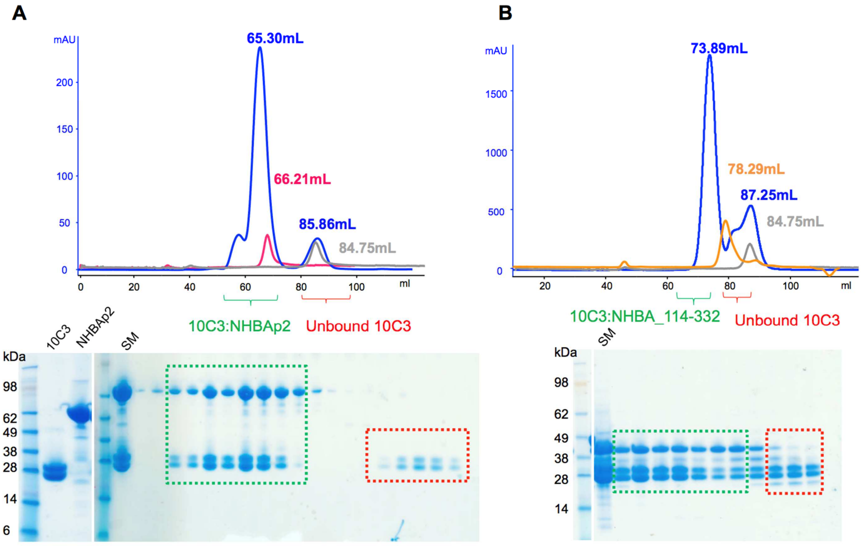


**Figure 3.9. Fab10C3:NHBA complex formation. Top:** SEC profiles of complexes 10C3 :NHBA p2 (**A**) and 10C3:NHBA_114-332 (**B**). **Bottom:** corresponding fractions run on SDS-PAGE gel. The first peak represents the complex (green) while the second peak is for the unbound Fab 10C3 (red box). Independent SEC runs: NHBAp2 (magenta), NHBA_113-332 (orange), 10C3 (grey). Retention volumes are indicated above each peak.

The purified complexes were concentrated and used to set up a number of crystallization experiments that were performed using different crystallization conditions, different sample concentrations, as well as using *in-situ* proteolysis in order to rid the samples of putative disordered regions that might hinder crystallization [104, 137]. Strikingly, only crystallization trials of the 10C3:NHBA_p2 complex and after *in-situ* proteolysis yielded crystals after ~60 days of incubation at 20°C, and under different crystallization conditions (**Table 5,** group 2). Six datasets of these putative 10C3:NHBA complexes were collected and their structures solved by MR using the refined coordinates of the unbound 10C3 (dataset #7) as search template. Disappointingly, despite the high binding affinity between Fab 10C3 and NHBA ($3.8x10^{-9}$ M, **Table 1**), none of the electron density maps calculated from these structures revealed the presence of a bound NHBA molecule in the Asymmetric Unit (ASU). Among other possible factors, this might be attributed also to the proteolytic process required to obtain these crystals, which might have affected the integrity of NHBA and thus induced its dissociation, or the complex might have dissociated due to the influence of the crystallization conditions.

Additional experiments were performed to obtain the structure of the NHBA epitope bound to Fab 10C3 by soaking a 17-residue long NHBA-derived peptide (made of residues 244-SEFEKLSDADKISNYKK-260) onto crystals of the unbound Fab. This peptide was designed starting from previous information derived from the epitope mapping by HDX-MS [90]. Although these soaked crystals diffracted X-rays at high resolution (**Table 5**, group 3), electron density maps calculated from several datasets again did not reveal a bound peptide (**Fig. 3.10**). The lack of bound peptide in the crystals of 10C3 might be due to several factors, such as: i) the physico-chemical properties of the synthesized peptide; ii) the peptide size/length that might have affected its ability to migrate through the crystal channels; iii) the lack of more precise information on the exact residues required for binding, which might have resulted in a putative sub-optimal design of the peptide; and iv) the likelihood of more conformational requirements (see antigen context) for 10C3 recognition.

Overall, these experiments resulted in additional 8 structures of unbound Fab 10C3 (**Table 5**, groups (2) and (3)), with dataset #15 being the one at the highest resolution (1.5 Å). Therefore, structure from dataset #15 was fully built and refined (**Table 2**) with the aim to use it for the analyses presented below (being at the highest resolution and thus presenting the highest level of structural details), as well as for the coordinates deposition in the PDB (in progress).

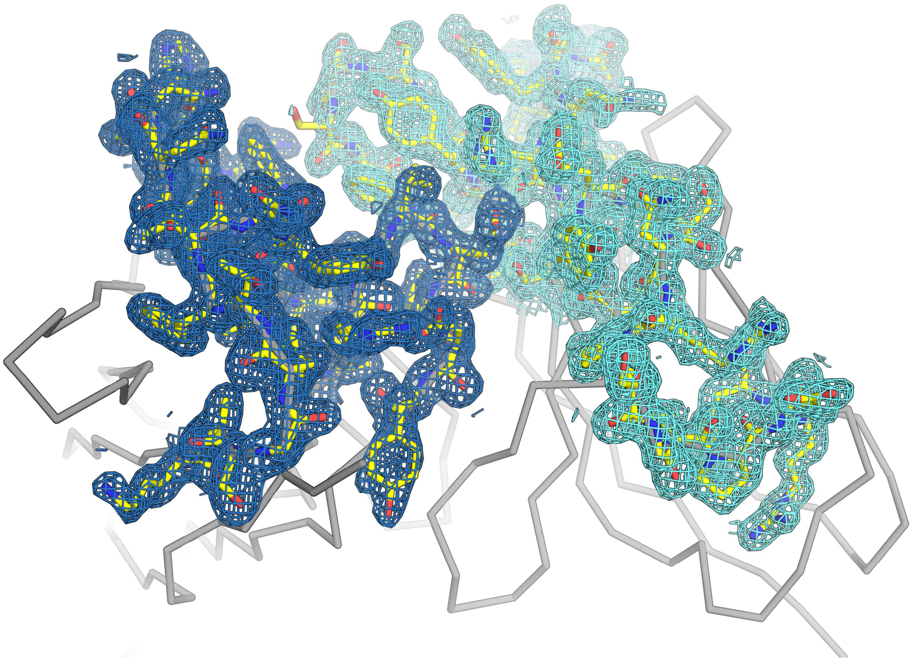


**Figure 3.10.** $2F_o$-$F_c$ electron density maps (HC: dark blue, LC: light blue) contoured at 1σ show the well-defined regions of the CDRs of 10C3 and lack of extra densities for a putative bound partner.

The overall structure of the Fab 10C3 presents the typical IgG fold (**Fig. 3.11**), stabilized by a total of four pairs of well-defined disulfide bridges located in the constant domains between residues Cys137 and Cys197 of the LC, and Cys142 and Cys198 of the HC. Disulfide bonds are also found in the variable domains between residues Cys22 and Cys90 of the LV, and Cys22 and Cys96 of the HV. The interface of the H and L chains of 10C3 has a total surface area of 1,491 Å$^2$, 10 hydrogen bonds, and 5 salt bridges (**Table 6**). The CDRs of Fab 10C3 were annotated, as for Fab 12E1, using the definition and nomenclature recently introduced by North et al. [79] (**Fig. 3.12**). The CDR-H1, made of residues Lys23-His35, falls within the H1-13-1 cluster and perfectly matches the loop conformation PPBLPAAABPBB; the CDR-H2 is made of residues Trp50-Asn59, falls into the H2-10-1 cluster and has a loop conformation BBPDDDLPBB; the CDR-H3 (Ala96-Tyr103) belongs to the H3-7-1 cluster with a loop conformation BPGADLAB. In the light chain, the CDR-L1, made of residues Thr23-Ser36, belongs to a L1-14-2 cluster, with loop conformation BBBAADAADBDPPB. The CDR-L2 comprises residues Tyr51-Ser58, lies within the L2-8-1 cluster and has a loop conformation BLLDPPPP, while the CDR-L3 is made of residues Ser91-Val100 and has a loop conformation BBPDGLLPPB, which is similar to those of the L3-10-1 cluster.

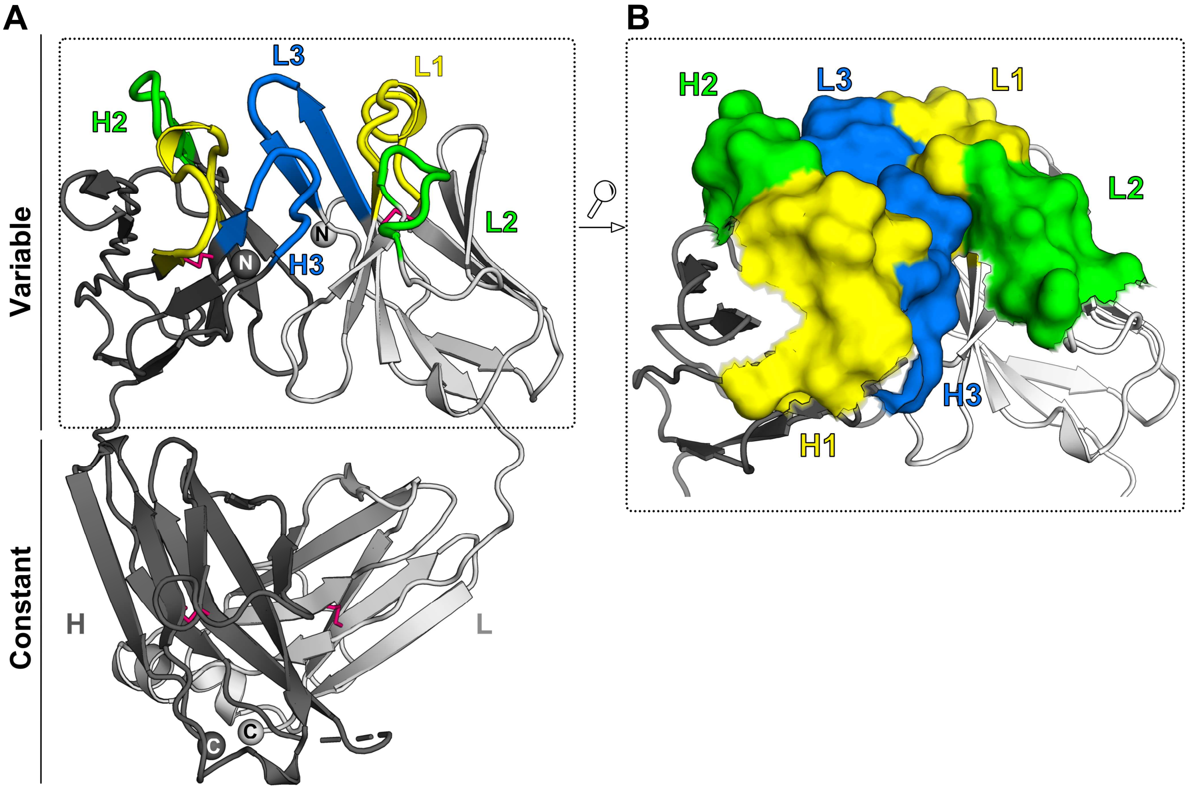


**Figure 3.11. Structure of unbound Fab 10C3.** (**A**) The overall structure of Fab 10C3 is shown, with the light and heavy chains colored in blue and gold, respectively. CDR-L1 and CDR-H1 are colored in yellow, CDR-L2 and CRD-H2 in green, CDR-L3 and CDR-H3 are in blue. Disulphide bonds are depicted with magenta sticks. (**B**) Surface representation of a zoomed view of the 10C3 CDRs region.

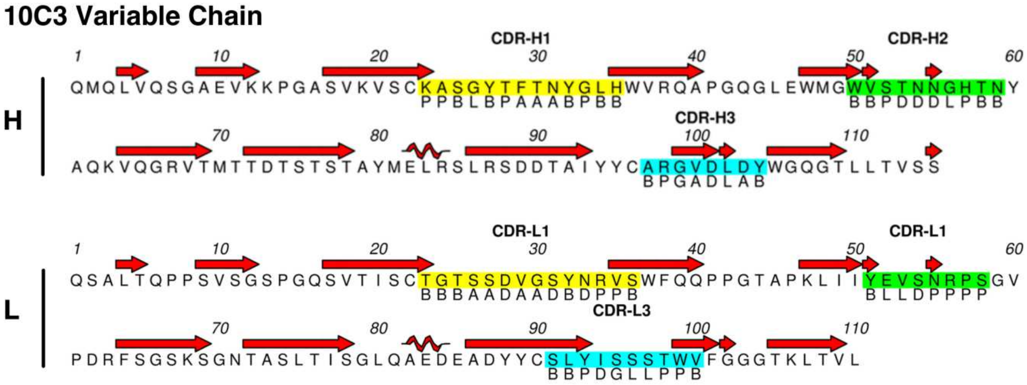


**Figure 3.12. 10C3 CDR regions.** CDRs for the variable heavy chain (**Top**) and for the variable light chain (**Bottom**) are shown with highlights over the sequence, while loop conformations are shown below the sequence, and secondary structures for each residue are shown above.

The total solvent exposed surface of the 10C3 CDRs has an area of ~3,600 Å$^2$, and it is dominated by polar uncharged amino acids that cover ~46% of the total accessible surface area (**Table 7**). Within the CDRs, a prevalence of Asn (15.17%) and of hydroxyl side-chain-containing residues like Ser (19.9%) and Thr (10.56%) can be observed, in agreement with previous studies on the typical composition of antibody paratopes (**Table 7**) [75, 133, 135]. These residues are clustered in the loop regions of CDR-H1, CHR-H2, CDR-L1 and CDR-L3 and contribute, along with several Tyr residues, to create a rim around a central positively-charged cavity at the interface between the H and L chains (**Fig.3.13A and B**). In addition, residues Asp101 and Asp103 from the CDR-H3 and Glu52 of CDR-L2 provide the paratope with a negatively charged area (**Fig. 3.13B**). Finally, several Tyr are solvent exposed, covering up to 11.71% of the ASA, and surrounded by short-chain hydrophilic amino acids (Asp, Asn, Ser, Thr).

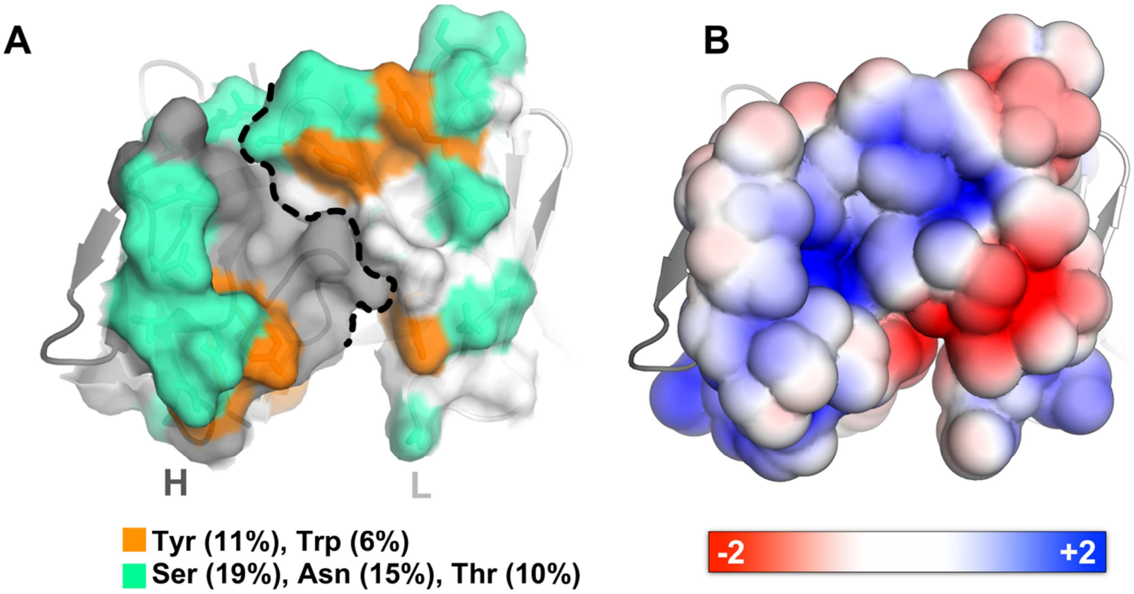


**Figure 3.13. Structure of Fab 10C3**. (**A**) Top view of the Fav 10C3 CDR region, with the most represented residues colored orange for Tyr and Trp and green for the polar Ser, Thr and Asn. A dotted line separates HC from LC. (**B**) Same view as in A, showing the surface of the CDRs colored according to the electrostatic potential distribution, which was calculated with APBS [120], where red and blue surfaces show negative and positive charges as contoured in the range from –3 $k_BTe^{-1}$ (red) to +3 $k_BTe^{-1}$ (blue), while white surfaces show neutral potentials.

In an attempt to speculate on the binding of 10C3 to NHBA, the paratope composition analyzed above can be related with the physicochemical properties of the 10C3 putative epitope previously identified by HDX-MS [90], which is made of a linear 32-residue sequence (243-KSEFEKLSDADKISNYKKDGKNDGKNDKFVGL-274). This region is localized in the N-terminal region of NHBA that is characterized by the absence of clearly predictable secondary structures, but with a putative slight tendency to acquire alpha-helical conformation (see *NHBA full length contains predicted partially unfolded regions* and **Fig.3.1A**). This segment is particularly rich in charged residues, especially Lys and Asp, which agrees with the presence of exposed positive and negative patches observed in the paratope surface (**Fig. 3.13B**). The latter suggests that electrostatic interactions could play a significant role in NHBA recognition by 10C3, a feature that has been already described as key specifically for antibody recognition of IDPs [136, 138].

Finally, the degree of similarities between the 15 unbound 10C3 structures was analyzed by performing superpositions of all the solved structures against the final high-resolution and fully refined coordinates of the reference dataset (#15). Using the secondary structure matching (SSM) algorithm within Coot [107], we aligned the Cα atoms of each solved structure (datasets #1 to #14) against the reference structure (dataset #15) and calculated the root mean square deviations (rmsd) value for each alignment (**Table 5**, column VII). The maximum rms deviations were observed for datasets #2 and #6, which diverged from the reference model with rmsd values of 1.01, and 0.97 Å, respectively, while the most superimposable structure (#7) had an rmsd value of 0.23 Å. Overall, these superpositions did not reveal significant differences in the 10C3 fold nor in the CDR arrangements (**Fig 3.14A and B**). Remarkably, also the distribution of side-chain rotamers for residues making the CDRs resulted conserved among all the structures compared. The only exceptions were residues His57 and Tyr104, localized in the CDR-H2 and the framework regions, respectively (**Fig. 3.14C**). In summary, a high degree of isomorphism was observed among all 15 structures of unbound Fab 10C3 described here (**Table 5**), although these were obtained from crystals grown in different crystallization conditions, made of different salt composition (i.e. sodium chloride, ammonium sulfate, calcium chloride, magnesium chloride, potassium thiocyanate) and concentrations, pH values (ranging from pH 4.2 to pH 6.5), and other precipitants and buffers (**Table 5**). These were virtually identical, with space group ($P2_12_12_1$), and with nearly identical cell parameters. This noticeable structure reproducibility has already been reported for other Fabs [139], and it highlights the extreme stability and rigidity of the Fab 10C3, an observation that agrees with the intrinsic and general structural stability of Fabs [78].

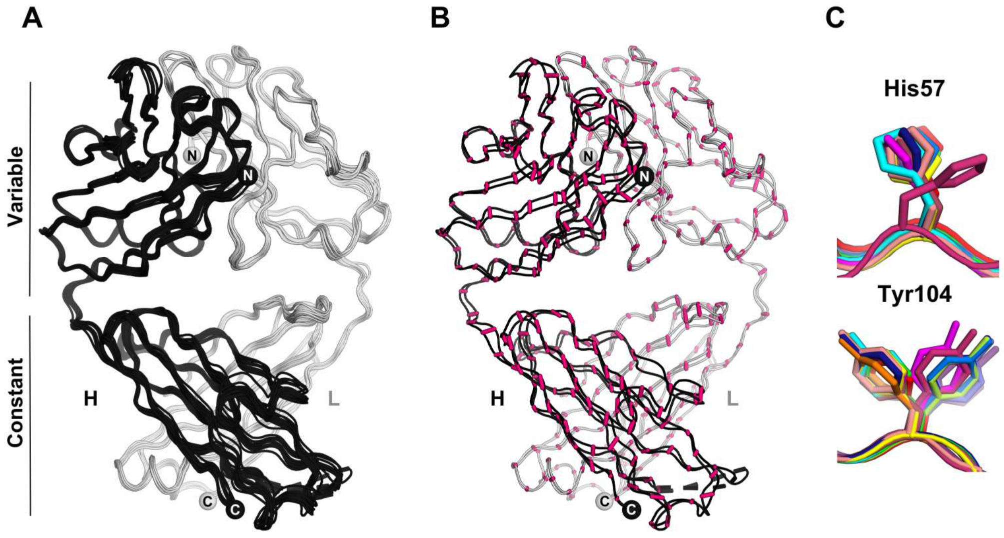


**Figure 3.14. Structural alignment of all 10C3 structures.** (**A**) All 10C3 structures solved in this work are shown as ribbon after superposition, and colored in black and white for the heavy (H) and light (L) chains. (**B**) Two highly divergent apo-10C3 structures are depicted as ribbons (structures from datasets #6 and #15), and colored as in (A). Regions of maximum divergence between Cα of the two structures are shown with magenta lines. (**C**) Zoom into the regions where differences in the position of side chain rotamers could be observed.

### *Human Fab 5H2*

The 1.8 Å crystal structure of Fab 5H2 displays the canonical sandwich immunoglobulin fold, similarly to those of the other Fabs studied here. Secondary structure superpositions among the three molecules that were observed in the ASU of the 5H2 crystals (chains for each heavy and light heterodimer named HL, MN, and OP), resulted in an rmsd value of 0.50 Å between copy HL and MN, and of 2.1 Å between HL and OP (**Fig. 3.15**).

Annotation of the Fab 5H2 CDRs was made based on the classification already used above for Fabs 12E1 and 10C3 [79]. The heavy chain CDR-H1 (residues Thr23-Thr37) belongs to a H1-15-1 cluster with loop conformation PBBLBPAAPPABBPB, the CDR-H2 (residues Tyr52-Lys60) belongs to cluster H2-9-1 and adopts a loop conformation BBPAALPBB, while the CDR-H3 (Ala98-Phe112) was assigned to the H3-7-1 class defined by North et al. [79] having BPPPBAADGBBPAB as loop conformation. In the L chain, the CDR-L1 (residues Arg24-Asn34) belongs to cluster L1-11-1 and adopts loop conformation BPABPDGDPBB, whereas the CDR-L2 (residues Tyr49-Ser56) and the CDR-L3 (residues Glu89-Thr96) belong respectively to clusters L2-8-1 and L3-8-1, with loop conformations BLLDPPPP and BDPABPPB (**Fig. 3.16**).

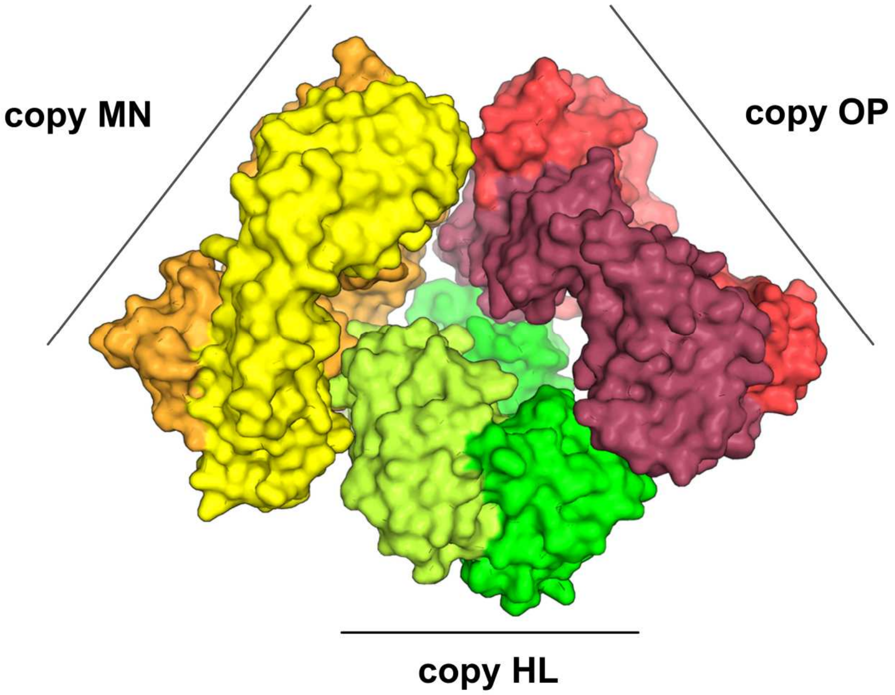


**Figure 3.15 ASU content of the unbound 5H2 structure.** The overall arrangement of the three copies of 5H2 is shown, and chains are labelled. Copy MN is colored in yellow (Light chain) and orange (Heavy chain), copy OP is colored in lime (Light chain) and green (Heavy chain), copy OP is colored in salmon (Light chain) and raspberry (Heavy chain).

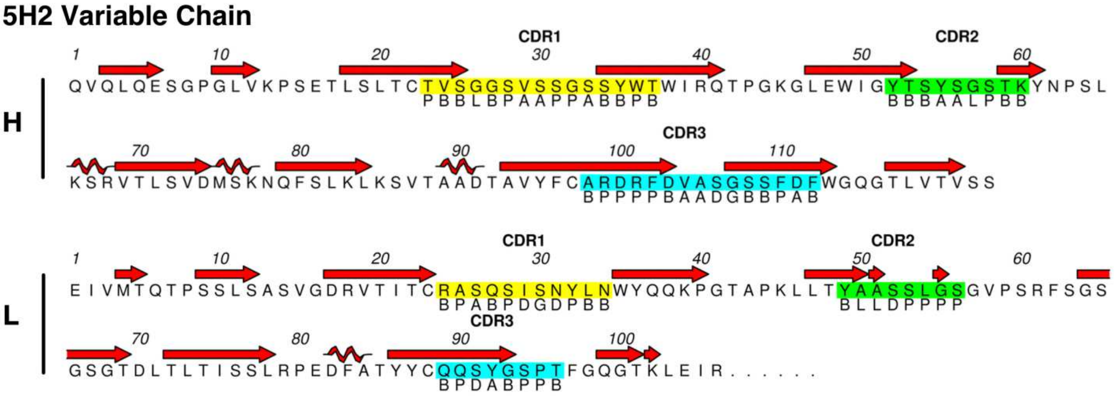


**Figure 3.16. Fab 5H2 CDRs region.** CDRs for the variable heavy chain (**Top**) and for the variable light chain (**Bottom**) are shown with highlights over the sequence, while loop conformations are shown below the sequence, and secondary structures for each residue are shown above.

## *Complex 5H2:NHBA-C3*

### *Structure determination*

The structure of the complex between Fab 5H2 and the C-terminal region of NHBA (construct NHBA-C3) was solved at 2.8 Å resolution (**Table 8**). Two copies of the complex were present in the ASU, disposed orthogonally to each other (**Fig 3.17A**). Structural superposition of these two copies of the complex resulted in an rmsd value of 0.43 Å for 544 aligned Cα atoms (from a total of 548 Cα atoms), therefore indicating very high similarity throughout all chains. For this reason, only one copy (chains HL) will be used for the description of the structure below.

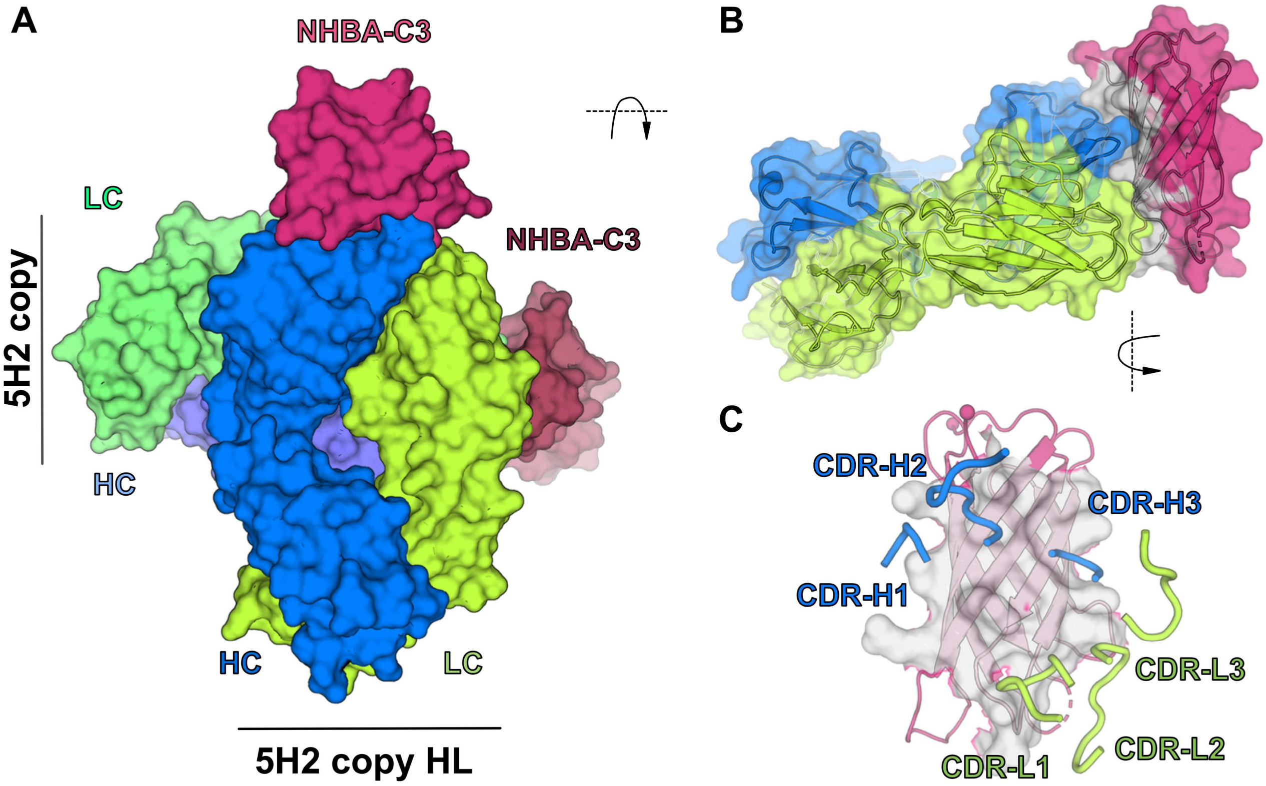


**Figure 3.17. Structure of the complex 5H2-NHBA-C3**. (**A**) The two copies of the complex are depicted with the 5H2 chains HL colored blue/green and chains IM colored violet/palegreen. NHBA molecules binding to HL and IM are colored magenta and raspberry, respectively. (**B**) Side view of the copy HL:NHBA-C3, with the epitope surface colored in grey. (**C**) Front view of NHBA-C3, with the 5H2-interacting interface colored in grey. Blue and green tubes show the 5H2 CDRs, while the rest of the Fab is omitted for clarity.

### *Structure analysis and details of the 5H2 epitope recognition*

The overall structure of the complex readily illustrates how Fab 5H2 recognizes a conformational epitope made of an entire face of the C-terminal β-barrel of NHBA through extensive interactions mediated by both the Fab H and L chains (**Fig 3.18**). NHBA-C3 is folded as an eight-stranded antiparallel β-barrel connected by seven loops of variable lengths, an architecture shared with other *N. meningitidis* proteins such as fHbp and TbpB [64], and also consistent with the previously determined NMR structure of an NHBA construct made of residues 245-427 [64] (PDB 2LFU). Structural superpositions of the NHBA-C3 structure from the complex determined here with each one of the ten NMR models deposited (PDB 2LFU) revealed an average rmsd of 1.85 Å, considering equivalent Cα atoms, with the highest rmsd corresponding to 2.62Å and the lowest to 1.58Å. The major differences could be observed in the loops, suggesting local inherent flexibility. However, the NMR models showed an average rmsd of 1.36Å when compared among themselves, in agreement with structural differences especially in the most N-terminal region of the 2 stranded β-sheet, which has coordinates not univocally assigned. These local variability in the NMR model could be attributed either to the inherent lower precision of the technique as compared to crystallographic structures, or to the sample environment (in solution).

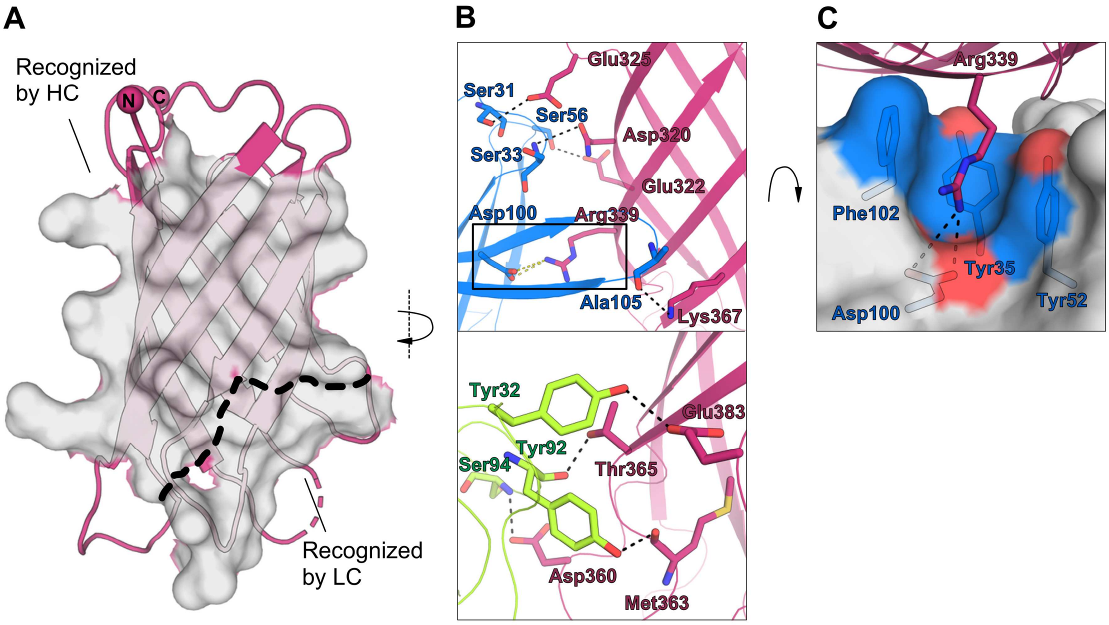


**Figure 3.18. Polar contacts between 5H2 and NHBA-C3. (A) The** 8-stranded β-barrel organization of NHBA-C3 and the surface of the 5H2 epitope are shown. A yellow dotted line splits the regions recognized by HC (top) and LC (bottom). (**B**) Top and bottom panel show the network of polar interactions between the HC or LC of Fab 5H2 and NHBA-C3. (**C**) Detail of the salt bridge between NHBA Arg339 and 5H2 Asp100.

The interface between NHBA-C3 and Fab 5H2 buries a total of ~1,700 Å$^2$ of surface area, split in ~750 Å$^2$ on NHBA-C3, and ~950 Å$^2$ on the Fab, as calculated by PISA [119]. These buried surfaces at the epitope-paratope interface correspond to ~11% and ~4% of the total surface area of NHBA-C3 and 5H2, respectively. Five strands forming the C-terminal face of the β-barrel (strands β1-β4 and β8) (**Fig. 3.18**) compose the epitope of NHBA, involving 27 antigen residues that contact 18 antibody residues. Loops forming the 5H2 CDRs fill the grooves of the conformational epitope surface, creating a remarkable epitope-paratope shape complementarity that involves the entire C-terminal face of the β-barrel (**Fig. 3.18**). The recognition is dominated by the 5H2 H chain, which contacts all three CDRs in a surface area of ~530 Å$^2$ (corresponding to 19 residues) of the antigen. Instead, the 5H2 L chain interacts with only ~210 Å$^2$ (corresponding to 8 residues) of the surface epitope. Also, the CDR-L2 is almost completely excluded from the recognition. The 5H2 H chain engages in polar contacts with Asn320, Glu322, Lys367, Glu425, and Arg339 of NHBA, through both main and side chains atoms (**Fig. 3.18B**, **Table 9**). Interactions between the conformational epitope and the H chain include a salt bridge (between NHBA Arg339 and 5H2 Asp100) and a hydrogen bond between the side-chains of NHBA Glu322 and 5H2 Ser56 (**Fig. 3.18B**). Additionally, epitope side chains of Glu425, Asn320 and Lys367 are positioned within H-bonding distance from backbone atoms of 5H2 Ser31, Ser33 and Ala105, respectively. There are also multiple interactions between NHBA-C3 and the 5H2 L chain, which include a network of well-positioned hydrogen bonds distributed around the NHBA loops made of residues 357-364 and 382-387. On loop 357-364, 5H2 Try92 forms H-bonds between the side chain OH group and the backbone O of Met363, and between the Tyr92 main chain O and the side chain O of Thr365. In addition, the OD2 of NHBA Asp360 is hydrogen-bonded to the backbone N of 5H2 Ser94. Finally, the side chain of NHBA Glu383 makes an H-bond with the hydroxyl group of 5H2 Tyr32, while the NHBA Asp360 makes interactions with the backbone N of 5H2 Ser94 (**Fig. 3.18B**, **Table 9**).

The distribution of the electrostatic potential on the surfaces of NHBA-C3 and 5H2 showed remarkable complementarity with several charged patches on the epitope interfacing opposite charged regions of 5H2 (**Fig. 3.19B**). Interestingly, a pocket can be observed at the interface between the H and L chains of 5H2, where Asp100 forms a negatively charged floor surrounded by a wall of hydrophobic residues made of Tyr35, Tyr52, and Phe102. This funnel-shaped pocket accommodates NHBA Arg339 which is positioned centrally on the NHBA epitope and forms a salt bridge between its side chains NH1 and the O of 5H2 Asp100. At the same time, the aromatic groups of Tyr35, Tyr52, and Phe102 surround and seem to trap the long aliphatic arginine side chain with favorable hydrophobic interactions. Similar complementarities could be observed for the NHBA negatively charged patches composed by Glu425-Asn320 (**Fig. 3.19B**), which interact with the predominantly positive loops CDR-H1 (Ser31-Gly32-Ser33) or NHBA Asp356-Asp60 facing a positive patch made of 5H2 Lys60 (chain H) and Gly93-Ser94 of the CDR-L3.

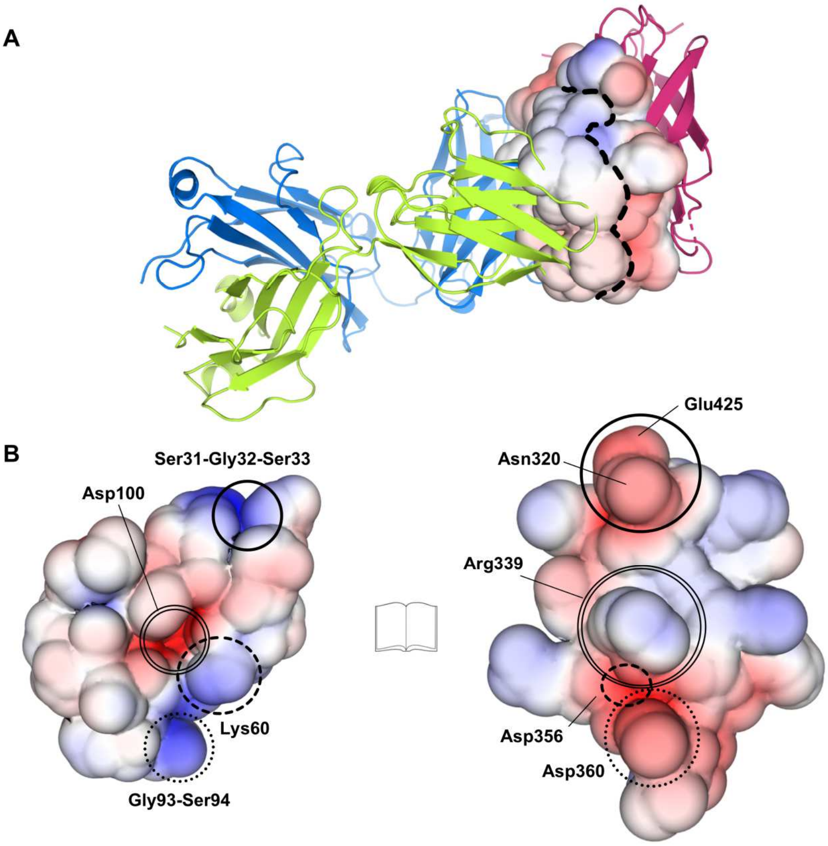


**Figure 3.19. Electrostatic potential of the complex interface**. (**A**) Side view of the Fab 5H2:NHBA-C3 complex, with the epitope-paratope surface colored according to the electrostatic potential distribution. A yellow dotted line divides the 5H2 and NHBA surfaces. (**B**) Open book view of the interfacing epitope-paratope surfaces. Circles with the same layout represent complementary regions.

### *Molecular bases of Fab 5H2 binding*

The high resolution of the X-ray complex structure enabled us to identify three main areas of NHBA-C3 that are contacted by 5H2: 1) the region interacting with the Fab light chain that engages polar contacts with NHBA residues Thr365 and Asp360, 2) the region interacting with Fab heavy chain comprising residues Glu322 and Glu425 3) the core region of the epitope made of residues Arg339 and K367. Hence, to investigate the precise relative importance of NHBA-C3 residues that interact with Fab 5H2, seven NHBA-C3 mutants were generated in the attempt of disrupt these interactions: Arg339Ala, Arg339Gly, Lys367Ala, Arg339Ala-Lys367Ala, Arg339Gly-Lys367Ala, Thr365Ala-Asp360Ala and Glu322Ala-Glu425Ala (**Fig. 3.20**). The constructs including the mutation of the conserved Arg339 to Ala or Gly were designed to assess the relevance of the long Arg side chain that protrudes with its polar head to reach inside a groove into the 5H2 paratope (see **Fig.3.18C**). Double mutants of Arg339 and Lys367 were designed in order to verify if there is a cumulative effect on Fab binding in presence of multiple mutations. Finally, the last two mutants (Thr365Ala-Asp360Ala and Glu322Ala-Glu425Ala) were generated for the evaluation of the role of 5H2 light and heavy chain in the binding to NHBA-C3.

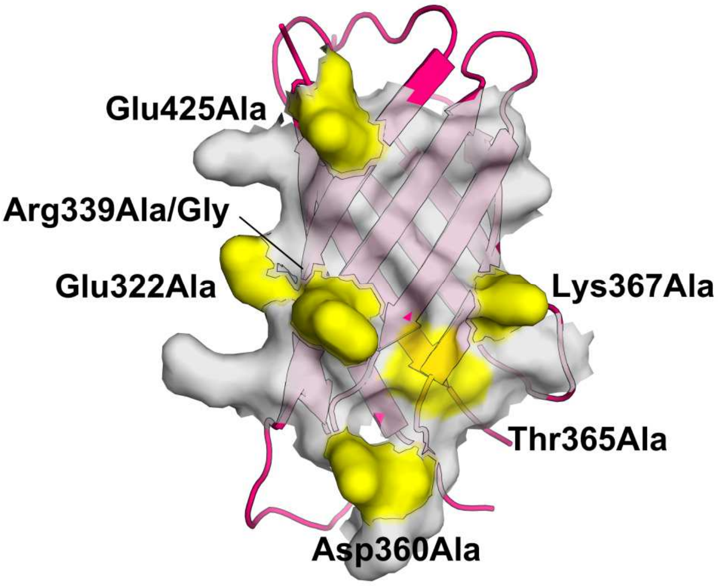
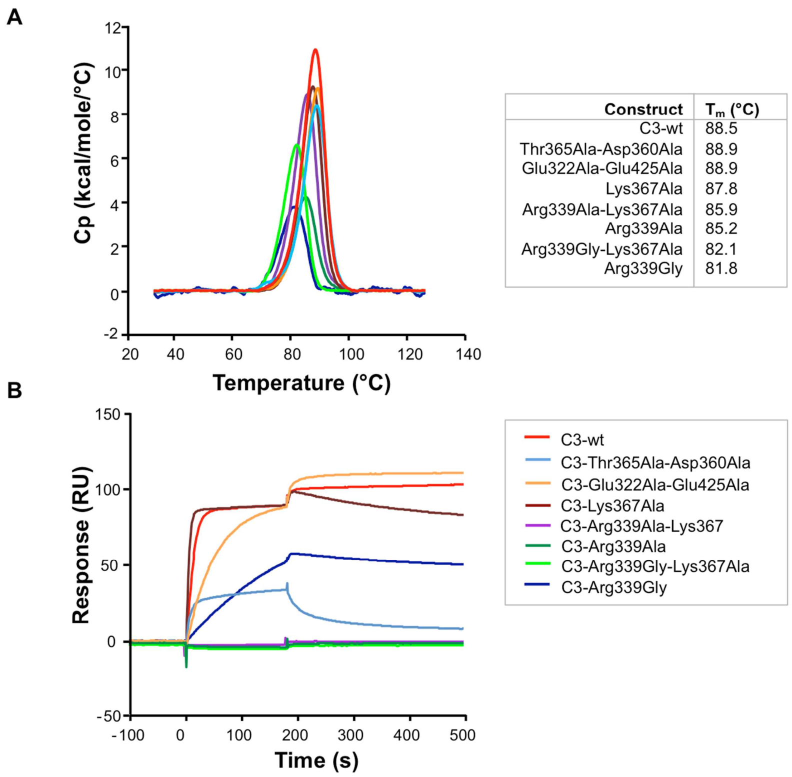


**Figure 3.20. Overview of NHBA-C3 mutated residues.** The 5H2 epitope is depicted as grey surface, while the mutated residues are highlighted in yellow and are labelled.

After successful expression and purification, the thermostability of the NHBA-C3 mutants was assessed by DSC, which revealed how mutations localized on the most peripheral regions of the epitope (i.e. Lys367Ala, Thr365Ala-Asp360Ala and Glu322Ala-Glu425Ala) did not significantly affect the melting temperature ($T_m$) with respect to the wild-type (**Fig. 3.21A**). Instead, mutations involving Arg339 (Arg339Ala, Arg339Gly, Arg339Ala-Lys367Ala, Arg339Gly-Lys367Ala) resulted in a significant thermal destabilizations, as they induced a decrease in the $T_m$ of up to ~7°C (**Fig. 3.21A**). Interestingly, the most dramatic effects were observed for constructs carrying the Arg339Gly mutation.

Finally, the effect of these mutations on the 5H2 binding was studied by SPR, by first capturing Fab 5H2 on a sensor chip, and then injecting, separately and at the same concentration, the seven NHBA-C3 mutants. These expertiments showed that mutations to Ala of residues Glu322, Glu425, or Lys367, involved in direct interactions with Fab heavy chain, are not critical for 5H2 binding, while mutation of residues Thr365 and Asp360, which make polar contacts with the 5H2 L chain, result in a 4-fold reduction in binding (**Fig. 3.21B**). However, the key determinants of the NHBA-5H2 interaction seem a combination of bonds mediated by both Arg339 and Lys367; while mutation of Arg339 to Gly results in a 2-fold reduction only, mutation to Ala does completely inhibit the binding. The same can be observed when mutating Arg339 to either Gly or Ala in combination with the Lys to Ala mutation in position 367. It is likely that the 5H2 Asp100 that forms a salt bridge with Arg339 (**Fig. 3.18C**) contributes to more repulsive interactions with Ala than with Gly, being the former bulkier and slightly more hydrophobic. The DSC data also suggest that the decreased thermo-stabilities of the Ala and Gly mutants of residue Arg339 (3.3 and ~6.7 degrees less stable than wild-type, respectively) do not result in major structural rearrangements or unfolding of NHBA. Such observation might explain the observed difference in binding of these mutants (i.e. the more destabilized mutant Arg339Gly shows residual binding by 5H2, while the less destabilized Arg339Ala is not recognized anymore by 5H2).



| Construct | $T_m$ (°C) |
|---|---|
| C3-wt | 88.5 |
| Thr365Ala-Asp360Ala | 88.9 |
| Glu322Ala-Glu425Ala | 88.9 |
| Lys367Ala | 87.8 |
| Arg339Ala-Lys367Ala | 85.9 |
| Arg339Ala | 85.2 |
| Arg339Gly-Lys367Ala | 82.1 |
| Arg339Gly | 81.8 |

**Figure 3.21. DSC and SPR studies of NHBA-C3 mutants. (A-Left)** DSC profiles of NHBA-C3 mutants, and (**Right**) observed $T_m$ values. **(B)** Sensorgrams profile of the SPR analysis performed using captured Fab 5H2 and NHBA-C3 mutants. The colors corresponding to each mutant are reported on the right.

### *Structural changes upon complex formation*

To investigate possible structural changes of Fab 5H2 upon binding to NHBA-C3, structural superpositions of the variable domains of the unbound and bound 5H2 structures were performed with SSM in Coot [107]. This revealed that Fab 5H2 undergoes local structural rearrangements upon binding of NHBA-C3 to better fit the interacting surfaces. Structural reorganization resulted particularly evident for residues Tyr35, Tyr52 and Phe102, which are located in the 5H2 H chain (**Fig. 3.22A and B**). Movement of the side chains of these residues seems to create a funnel-shaped pocket that accommodates NHBA Arg339 (**Fig. 3.22D**). While the side-chain of Phe102 is displaced by up to 10.5 Å, minor shifts can be observed for the side chains of Tyr35 and Tyr52. Additional conformational changes could be observed for the 5H2 CDR-H3 loop, likely influenced also by rearrangements of the nearby CDR-L2 (residues Tyr49-Ser60), which lacks electron density for one of the chains (O) thus suggesting intrinsic flexibility (**Fig. 3.22B and C**).

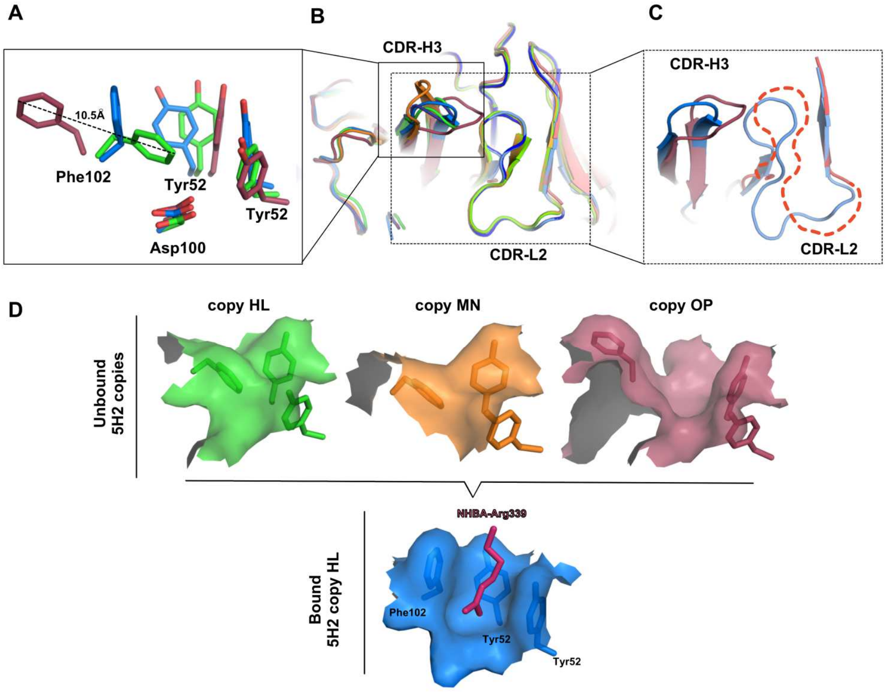


**Figure 3.22. Binding-induced conformational changes in the 5H2 structures**. (**A**) Rotamers of residues Phe102, Tyr35 and Tyr52 of the unbound Fab 5H2 chains HL (green) and chains OP (raspberry), compared to those of the NHBA-C3-bound Fab 5H2 chains HL (blue). (**B**) Top view of the superposition between unbound Fab 5H2 chains HL, MN (orange), OP and NHBA-C3-bound Fab 5H2 chains HL (blue) and IM (dark blue). Region with the major differences are boxed. (**C**) Zoom into CDR-H3 and CDR-L2 of the unbound Fab 5H2 chains OP and NHBA-C3-bound Fab5H2 chains HL, where a dotted line shows the missing atoms of the CDR-L2 from the OP chains. (**D**) Surfaces of the triad Phe102-Tyr52-Tyr35 of the unbound 5H2 structures are shown on top, compared with the NHBA-C3-bound 5H2 chains HL (bottom).

Measurements of the 5H2 elbow angles in its different states were also performed in order to study their structural deviations upon binding to NHBA. In the case of unbound-5H2, a remarkable oscillation was observed among the elbow angles calculated for each of the three unbound-5H2 molecules present in the ASU. These angles ranged from 122.58° and 122.20° for chains HL and MN, up to 146.10° for chains OP (**Fig. 3.23**). In contrast, the two ASU copies of the NHBA-C3-bound 5H2 displayed a rather similar elbow bend of 135.57° and 136.12°, respectively. As expected, this suggests that the hinge region between C and V chains of the bound 5H2 have restricted degrees of freedom if compared to their conformations when in the unbound state. It has been previously proposed that a relationship between the change in elbow angle and antigen binding exists [140-142]. Moreover, antigen recognition can lead to conformational changes in the variable domains that is subsequently transmitted to the constant domains [74]. Importantly, not all known structures of Fabs have different elbow angles for their free and unbound states [140, 143].

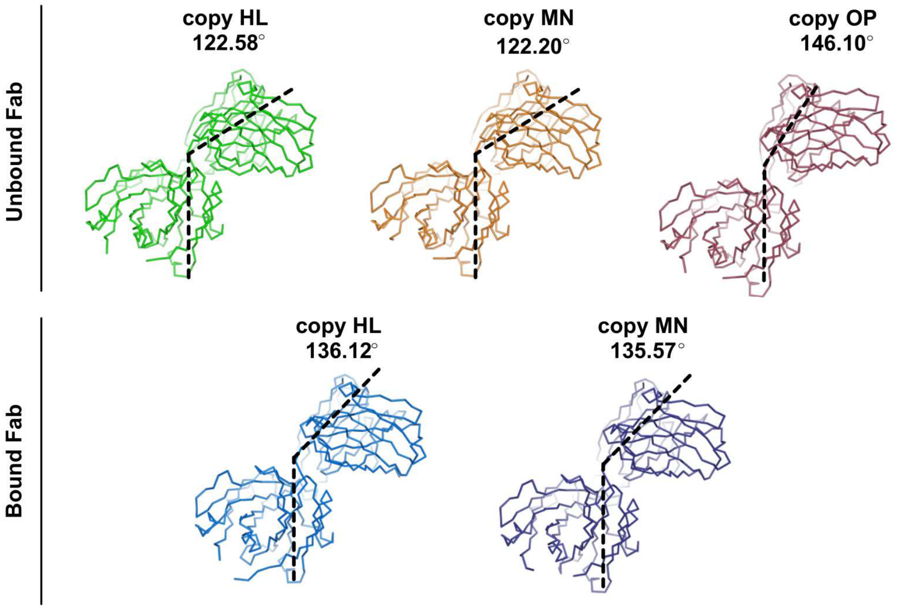


**Figure 3.23. The elbow angles of Fab 5H2.** All elbow angles of the five different 5H2 heterodimer structures solved here (3 copies from the unbound structure, and 2 copies from the NHBA-C3-bound structure) are shown as measured with phenix.elbow.

Taken together, our observations on the structural rearrangements of the 5H2 paratope upon binding of the NHBA-C3 seem to agree with the pre-existing equilibrium hypothesis in combination with the induced fit model as general mechanism of antibody-antigen recognition [144]. According to the pre-existing equilibrium hypothesis, antibodies can assume a set of multiple conformations in solution that are closely related to each other, with a shift towards the antigen binding preferred conformation as soon as they engage in complex formation. This could drive the so-called early binding mechanism that is then followed by induced-fit, considered as conformational changes of the binding region upon binding [145].

In the case of our 5H2:NHBA-C3 complex, we first observed a distribution of possible states from the structure of unbound 5H2, which likely reflects the intrinsic molecular motion of 5H2 even within the constraints of the crystal packing environment. Then, it is likely that the only 5H2 conformation observed in the complex corresponds to the one better fitting the antigen surface, and thus indicating a selection of the most favorable Fab conformation. This selection consisted specifically in conformational changes in the region of residues Try52-Tyr36-Phe102, to perfectly accommodate Arg339 from NHBA-C3 (**Fig.3.22**).

### *NHBA β-barrel conservation and 5H2 cross-reactivity*

In order to analyze the 5H2 cross-reactivity properties, a panel of short and long NHBA variants was chosen for sequence comparison. Sequences of the β-barrels from long (p2, p3, p5, p1, p29) and short (p18, p17, p10, p21) variants were aligned (**Fig. 3.24**) against the sequence of variant p20, from which the construct NHBA-C3 was obtained.

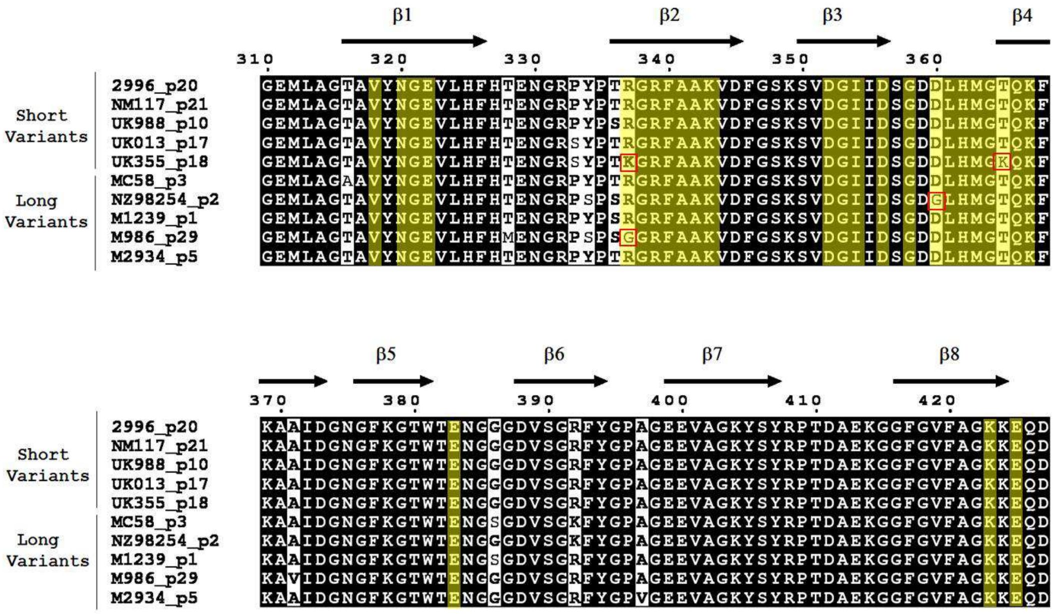


**Figure 3.24. Sequence alignment of NHBA long and short variants**. The 5H2 epitope is highlighted in yellow, while red boxes show the non-conserved residues of the 5H2 epitope.

The sequence conservation is almost absolute, with the most divergent variant being p17 that has a sequence identity of the 97.4% with p20. Only three variants (p18, p2 and p29) showed a few point mutations in the region of the mapped 5H2 epitope (**Fig. 3.24**). More in detail, Arg337, which is substituted with Lys in p18 or Gly in p29, is not involved in direct polar interactions with 5H2, thus it is unlikely to play a critical role once mutated. Instead, substitution of Thr365 into Lys in p29 could impact the interaction with 5H2, considering that the side-chain of Thr365 is engaged in hydrogen bonds with the 5H2 Tyr92 and that Lys has a higher steric hindrance compared to Thr. The other variant which is not fully conserved is p2, for which the experimental data already demonstrated that mutation of Asp360 into glycine does not affect the binding affinity (**Table 1**, see also *Fab-NHBA interaction studies*).

Together with the results of the binding experiments, these observations support the ability of Fab 5H2 to efficaciously recognize all NHBA variants studied here.

# Tables

## Table 1

Binding affinities of Fabs 12E1, 10C3 and 5H2 towards NHBA peptides 2, 3 and 20 calculated as described in the *Material and Methods* section.

| Fab | Protein | Kon (1/M*s) | Koff (1/s) | $K_D$ (M) |
|---|---|---|---|---|
| 12E1 | NHBA p3 | 8.4 $x10^{+4}$ | 3.1 $x10^{-4}$ | 5.1 $x10^{-10}$ |
| | NHBA p20 | 4.9 $x10^{+4}$ | 3.0 $x10^{-4}$ | 6.2 $x10^{-9}$ |
| | NHBA p2 | 2.1 $x10^{+5}$ | 5.9 $x10^{-5}$ | 2.7 $x10^{-10}$ |
| 10C3 | NHBA p3 | 4.34 $x10^{+5}$ | 5.41$x10^{-4}$ | 1.2 $x10^{-10}$ |
| | NHBA p20 | No binding | | |
| | NHBA p2 | 2.3 $x10^{+5}$ | 9.2 $x10^{-4}$ | 5.7 $x10^{-9}$ |
| | NHBA_114-332 | 4.9 $x10^{+5}$ | 0.149 | 3.5 $x10^{-9}$ |
| 5H2 | NHBA p3 | 4.2 $x10^{+5}$ | 1.7 $x10^{-4}$ | 4.0 $x10^{-10}$ |
| | NHBA p20 | 5.4 $x10^{+5}$ | 2.3 $x10^{-4}$ | 4.3 $x10^{-10}$ |
| | NHBA p2 | 2.9 $x10^{+5}$ | 1.4 $x10^{-4}$ | 4.9 $x10^{-10}$ |

## Table 2

Data collection and refinement statistics of Fabs 12 E1, 10C3 and 5H2

| | **Fab 12E1** | **Fab 10C3** | **Fab 5H2** |
|---:|---|---|---|
| **Wavelength (Å)** | 0.97932 | 0.979788 | 0.97721 |
| **Resolution range (Å)** | 23.4-2.7 (2.89 - 2.75) | 32.1 - 1.5 (1.55 - 1.50) | 47.0 - 1.8 (1.94 - 1.88) |
| **Space group** | $P\ 2_1\ 2_1\ 2$ | $P\ 2_1\ 2_1\ 2_1$ | $P\ 2_1$ |
| **Unit cell (*a*,*b*,*c*)** | 64.8, 82.1, 100.3 | 69.9, 79.8, 82.5 | 82.0, 105.7, 86.2 |
| **(α,β,γ)** | 90, 90, 90 | 90, 90, 90 | 90, 111.9, 90 |
| **Total reflections** | 65074 (6325) | 264157 (23995) | 325510 (32667) |
| **Unique reflections** | 14236 (1360) | 73405 (7068) | 109353 (10952) |
| **Multiplicity** | 4.6 (4.7) | 3.6 (3.4) | 3.0 (3.0) |
| **Completeness (%)** | 98.6 (97.5) | 99 (97) | 98.8 (99.2) |
| **Mean I/sigma(I)** | 9.9 (1.6) | 14.88 (1.06) | 6.6 (1.3) |
| **Wilson B-factor (Å$^2$)** | 53.12 | 18.09 | 29.80 |
| **R-merge** | 0.134 (0.897) | 0.05052 (1.139) | 0.082 (0.831) |
| **R-meas** | 0.151 (1.009) | 0.05934 (1.349) | 0.100 (1.008) |
| **R-pim** | 0.068 (0.451) | 0.03079 (0.7127) | 0.056 (0.563) |
| **CC1/2** | 0.993 (0.609) | 0.999 (0.446) | 0.994 (0.474) |
| **Reflections used in refinement** | 14204 | 73375 | 109345 |
| **Reflections used for R-free** | 717 | 3669 | 5540 |
| **R-work** | 0.1808 (0.2197) | 0.1800 (0.3057) | 0.1864 (0.2870) |
| **R-free** | 0.2638 (0.3366) | 0.2114 (0.3300) | 0.2265 (0.3423) |
| **Number of non-hydrogen atoms** | 3436 | 3857 | 10613 |
| **Macromolecules** | 3337 | 3156 | 9805 |
| **Solvent** | 73 | 701 | 726 |
| **Protein residues** | 432 | 420 | 1291 |
| **RMS(bonds, Å)** | 0.014 | 0.010 | 0.007 |
| **RMS(angles, °)** | 1.96 | 1.27 | 0.94 |
| **Ramachandran favored (%)** | 93.4 | 98 | 98 |
| **Ramachandran allowed (%)** | 5.6 | 1.9 | 2.1 |
| **Ramachandran outliers (%)** | 1 | 0 | 0 |
| **Rotamer outliers (%)** | 6.6 | 0.84 | 1.3 |
| **Clashscore** | 5.25 | 1.77 | 7.65 |
| **Average B-factor (Å$^2$)** | 56.44 | 23.88 | 34.65 |
| **Macromolecules** | 56.49 | 21.24 | 34.08 |
| **Ligands** | 70.63 | - | 50.33 |
| **Solvent** | 48.34 | 35.80 | 40.62 |

**Statistics for the highest-resolution shell are shown in parentheses**

**Table 3**
Description of the polar contacts between H and L chains of Fab 12E1.

| Bond type | Light chain | Dist. [Å] | Heavy chain |
|---|---|---|---|
| Hydrogen bond | L:GLN 43 [ NE2] | 3.18 | H:TYR 95 [ OH] |
| | L:SER 48 [ OG ] | 2.95 | H:GLY 106 [ O ] |
| | L:ILE 122 [ N ] | 3.33 | H:LYS 131 [ O ] |
| | L:SER 167 [ OG ] | 2.73 | H:PRO 169 [ O ] |
| | L:GLN 165 [NE2] | 3.36 | H:LEU 172 [ O ] |
| | L:SER 181 [OG] | 3.35 | H:SER 181 [ OG] |
| | L:TYR 92 [OH] | 3.06 | H:GLN 39 [ NE2] |
| | L:LEU 4 [O ] | 2.81 | H:ARG 44 [NH2] |
| | L:SER 213 [ O ] | 2.83 | H:LYS 131 [ NZ ] |
| | L:GLU 128 [OE2] | 3.19 | H:LYS 211 [ NZ ] |
| Salt bridge | L:ASP 60 [OD1] | 3.38 | H:LYS 101 [ NZ ] |
| | L:GLU 128 [OE2] | 3.19 | H:LYS 211 [ NZ ] |

## Table 4

Analysis of 12E1 CDRs amino acid composition. The contribution of single amino acids or groups of amino acids clustered on the basis of their biochemical properties to the total ASA (solvent accessible area calculated by PSA) are expressed both in percentage and Å$^2$. The total ASA is calculated as the sum of the ASA of each residues belonging to the CDRs.

**Fab 12 E1**

| type | % on ASA | Amino acid | ASA (Å$^2$) | % on ASA |
|---|---|---|---|---|
| aromatic | 18,93 | Tyr | 373,35 | 9,70 |
| | | Trp | 240,42 | 6,25 |
| | | Phe | 114,77 | 2,98 |
| positive | 28,84 | Lys | 513,73 | 13,35 |
| | | His | 144,41 | 3,75 |
| | | Arg | 451,67 | 11,74 |
| negative | 7,72 | Asp | 239,15 | 6,21 |
| | | Glu | 57,85 | 1,50 |
| polar uncharged | 28,97 | Ser | 378,39 | 9,83 |
| | | Asn | 290,85 | 7,56 |
| | | Gln | 103,75 | 2,70 |
| | | Thr | 225,57 | 5,86 |
| | | Pro | 116,60 | 3,03 |
| | | Cys | 0 | 0,00 |
| non polar | 15,54 | Ile | 262,38 | 6,82 |
| | | Gly | 176,57 | 4,59 |
| | | Leu | 88,66 | 2,30 |
| | | Val | 36,67 | 0,95 |
| | | Ala | 8,99 | 0,23 |
| | | Met | 24,99 | 0,65 |
| Total | 100,00 | | 3848,77 | 100,00 |

**Table 5**

List of all the solved 10C3 datasets. Column I: starting material used for the crystallization experiment; Column II: identification number of the dataset; Column III: crystallization conditions; Column VI: resolution of each dataset; Column V and VI: cell dimensions and SG; column VII: rmsd values obtained by SSM of each dataset to the reference one.

| starting crystallization material | # dataset | Mother liquor composition | Resolution | cell dimension (a b c) | SG | SSM r.m.s.d.(Å) |
|---|---|---|---|---|---|---|
| (1) Crystal of unbound Fab 10C3 | 1 | 0.2 M LiSO$_4$ , 0.1 M BIS-TRIS 5.5 pH , 25 %w/v PEG 3350 | 1.9Å | 69.30 76.58 82.14 | P 21 21 21 | 0.33 |
| | 2 | 0.1 M KSCN , 30 %w/v PEG MME 2000 | 1.8Å | 69.67 76.78 82.29 | P 21 21 21 | 1.01 |
| | 3 | 0.1 M KSCN , 30 %w/v PEG MME 2000 | 1.9Å | 69.72 77.69 82.49 | P 21 21 21 | 0.76 |
| | 4 | 0.2 M NH$_4$ Acet , 0.1 M BIS-TRIS 5.5 pH , 25 %w/v PEG 3350 | 1.9Å | 69.65 77.91 82.35 | P 21 21 21 | 0.67 |
| | 5 | 0.2 M (NH4)$_2$H Cit , 20 %w/v PEG 3350 | 1.8Å | 69.65 77.31 82.05 | P 21 21 21 | 0.88 |
| | 6 | 37.5 %w/v M1K3350, 0.1 M MB1 6.5 pH , 10 % MAA | 2.2 Å | 69.61 74.24 82.41 | P 21 21 21 | 0.97 |
| | 7 | 0.2 M NH$_4$ Acet , 0.1 M Na$_3$ Cit 5.6 pH , 30 %w/v PEG 4K | 1.7 Å | 69.74 78.69 82.89 | P 21 21 21 | 0.23 |
| (2) Crystals of complexed Fab 10C3 | 8 | 0.2 M Ammonium sulfate 0.1 M Sodium acetate pH 4.6, 25 % w/v PEG 4000 | 1.89 Å | 69.17 76.8 82.53 | P 21 21 21 | 0.56 |
| | 9 | 0.2 M Ammonium sulfate 30 % w/v PEG 4000 | 1.53 Å | 67.93 75.54 81.87 | P 21 21 21 | 0.44 |
| | 10 | 0.2 M Magnesium chloride hexahydrate 0.1 M MES pH6.0 20 % w/v PEG 6000 | 1.69 Å | 69.33 76.81 82.28 | P 21 21 21 | 0.76 |
| | 11 | 0.2 M Magnesium chloride hexahydrate 0.1 M MES pH6.0 20 % w/v PEG 6000 | 1.28 Å | 69.49 78.11 82.86 | P 21 21 21 | 0.36 |
| | 12 | 0.2 M Calcium chloride dihydrate 0.1 M MES pH6.0 20 % w/v PEG 6000 | 1.44 Å | 76.77 82.04 68.80 | P 21 21 21 | 0.58 |
| (3) Crystals of Soaked Fab 10C3 | 13 | 0.2 M NaCl , 0.1 M Na Phos Cit 4.2 pH , 20 %w/v PEG 8000 | 2.2Å | 70.19 78.15 83.86 | P 21 21 21 | 0.43 |
| | 14 | 0.2 M NaCl , 0.1 M Na Phos Cit 4.2 pH , 20 %w/v PEG 8000 | 1.5Å | 69.51 78.52 82.73 | P 21 21 21 | 0.24 |
| | **15** | **0.17 M (NH$_4$)$_2$SO$_4$ , 15 %v/v Glycerol , 25.5 %w/v PEG 4000** | **1.5 Å** | **69.91 79.83 82.58** | **P 21 21 21** | **reference** |

**Table 6**

Description of the polar contacts between H and L chains of Fab 10C3.

| Bond type | Light chain | Dist. [Å] | Heavy chian |
|---|---|---|---|
| Hydrogen bond | H:GLN 39 [ NE2] | 3.0 | L:GLN 40[ OE1 ] |
| | H:ASN 59 [ ND2] | 2.9 | L:SER 97[ O ] |
| | H:TYR 95 [ OH ] | 3.8 | L:THR 44 [ O ] |
| | H:LYS 145 [ NZ ] | 2.7 | L:THR 135 [ OG1] |
| | H:SER 174 [ OG ] | 3.0 | L:GLU 164 [ OE2] |
| | H:SER 174 [ N ] | 3.3 | L:GLU 164 [ OE2] |
| | H:SER 181 [ OG ] | 2.9 | L:TYR 181 [ OH ] |
| | H:GLN 39 [ OE1] | 3.1 | L:GLN 40 [ NE2] |
| | H:GLN 43 [ O ] | 3.5 | L:TYR 89 [ OH ] |
| | H:ASP 101 [ O ] | 3.1 | L:SER 36 [ OG ] |
| Salt bridge | H:TYR 95 [ OH ] | 3.7 | L:GLN 40 [ NE2] |
| | H:LYS 211 [ NZ ] | 2.9 | L:GLU 127[ OE1 ] |
| | H:LYS 211 [ NZ ] | 3.3 | L:GLU 127[ OE2 ] |
| | H:ASP 101 [ OD1] | 2.8 | L:ARG 34 [ NH1 ] |
| | H:ASP 101 [ OD1] | 3.6 | L:ARG 34 [ NH1 ] |

**Table 7**

Analysis of 10C3 CDRs amino acid composition. The contribution of single amino acids or groups of amino acids clustered on the basis of their biochemical properties to the total ASA (solvent accessible area calculated by PISA [119] ) are expressed both in percentage and Å$^2$. The total ASA is considered the sum of ASA of each residues belonging to the CDRs.

**Fab 10C3**

| type | % on ASA | Amino acid | ASA (Å$^2$) | % on ASA |
|---|---|---|---|---|
| aromatic | 18,16 | Tyr | 422,48 | 11,71 |
| | | Trp | 218,67 | 6,06 |
| | | Phe | 14,05 | 0,39 |
| positive | 13,47 | Lys | 101,79 | 2,82 |
| | | His | 134,04 | 3,72 |
| | | Arg | 249,96 | 6,93 |
| negative | 8,14 | Asp | 227,42 | 6,30 |
| | | Glu | 66,07 | 1,83 |
| polar uncharged | 46,53 | Ser | 718,75 | 19,92 |
| | | Asn | 547,36 | 15,17 |
| | | Gln | 0 | 0,00 |
| | | Thr | 381,09 | 10,56 |
| | | Pro | 31,46 | 0,87 |
| | | Cys | 0 | 0,00 |
| non polar | 13,71 | Ile | 51,88 | 1,44 |
| | | Gly | 202,27 | 5,61 |
| | | Leu | 51 | 1,41 |
| | | Val | 185,44 | 5,14 |
| | | Ala | 3,86 | 0,11 |
| | | Met | 0 | 0,00 |
| Total | 100,00 | | 3607,59 | 100,00 |

**Table 8**

Data collection and refinement statistics for 5H2-NHBA-C3 complex

| | **5H2:NHBA-C3** |
|---|---|
| **Wavelength (Å)** | 0.97950 |
| **Resolution range (Å)** | 91.05 - 2.86 (2.93 - 2.86) |
| **Space group** | *P* $6_1$ 2 2 |
| **Unit cell (*a,b,c*)**<br>**(*α,β,γ*)** | 119.45, 119.45, 364.22<br>90, 90, 120 |
| **Total reflections** | 710332 (50746) |
| **Unique reflections** | 36576 (3562) |
| **Multiplicity** | 19.4 (19.3) |
| **Completeness (%)** | 100 (100) |
| **Mean I/sigma (I)** | 18.4 (1.0) |
| **Wilson B-factor (Å$^2$)** | 88.02 |
| **R-merge** | 0.189 (3.765) |
| **R-meas** | 0.194 (3.866) |
| **R-pim** | 0.044 (0.875) |
| **CC1/2** | 0.999 (0.557) |
| **Reflections used in refinement** | 36559 (3559) |
| **Reflections used for R-free** | 1775 (182) |
| **R-work** | 0.1921 (0.3393) |
| **R-free** | 0.2384 (0.3750) |
| **Number of non-hydrogen atoms** | 8326 |
| **Macromolecules** | 8260 |
| **ligands** | 20 |
| **solvent** | 46 |
| **Protein residues** | 1100 |
| **RMS(bonds, Å)** | 0.014 |
| **RMS(angles, °)** | 1.93 |
| **Ramachandran favored (%)** | 96 |
| **Ramachandran allowed (%)** | 3.2 |
| **Ramachandran outliers (%)** | 0.55 |
| **Rotamer outliers (%)** | 11 |
| **Clashscore** | 5.45 |
| **Average B-factor (Å$^2$)** | 84.53 |
| **macromolecules** | 84.60 |
| **ligands** | 91.57 |
| **solvent** | 67.94 |

Statistics for the highest-resolution shell are shown in parentheses.

**Table 9**

Polar contacts between Fab 5H2 light chain (LC), heavy chain (HC) and NHBA-C3.

| POLAR CONTACTS 5H2-LC:NHBA-C3 | | | |
|---|---|---|---|
| | **LIGHT CHAIN** | **Dist. [Å]** | **NHBA-C3** |
| Hydrogen bonds | L:TYR 32[ OH ] | 3.47 | A:GLU 383[ OE2] |
| | L:TYR 92[ OH ] | 2.88 | A:MET 363[ O ] |
| | L:SER 94[ N ] | 2.90 | A:ASP 360[ OD2] |
| | L:ASN 31[ OD1] | 3.64 | A:LYS 367 NZ ] |
| | L:TYR 92[ O ] | 2.67 | A:THR 365[ OG1] |

| POLAR CONTACTS 5H2-HC:NHBA-C3 | | | |
|---|---|---|---|
| | **HEAVY CHAIN** | **Dist. [Å]** | **NHBA-C3** |
| Hydrogen bonds | H:SER 31[ OG ] | 3.44 | A:GLU 425[ OE2] |
| | H:SER 33[ N ] | 3.84 | A:ASN 320[ OD1] |
| | H:SER 56[ OG ] | 3.07 | A:GLU 322[ OE2] |
| | H:ALA 105[ O ] | 3.40 | A:LYS 367[ NZ ] |
| Salt bridges | H:ASP 100[ OD2] | 3.20 | A:ARG 339[ NH1] |
| | H:ASP 100[ OD1] | 3.97 | A:ARG 339[ NH1] |

# Discussion

Vaccination is widely considered as one of the most effective tools to prevent infectious diseases and preserve human health worldwide. At the root of vaccines' success is the mechanism by which vaccinations provide the vaccinee with the ability to respond to microbial infections with the prompt production of specific protective antibodies against a certain pathogen. Accordingly, understanding the characteristics of antibodies generated either upon infection or upon vaccination, as well as the exact role that vaccine antigens play for the immune recognition and ensuing response, represent powerful tools for vaccine research [21]. For many years, the rate limiting step in the analysis of the antibody repertoire has been the intrinsic complexities of the isolation of antibodies. Numerous technological advances in the last decade now enable the successful isolation and generation of monoclonal antibodies from B-cells, which naturally become key reagents for vaccines research, as they allow to discover key epitopes potentially useful in vaccine design [20]. For example, the use of monoclonal antibodies to reveal protective epitopes is at the base of the vaccine design strategies in programs such as those against diseases caused by human cytomegalovirus (CMV) [146, 147], influenza [148], and HIV-1 [149]. Together with the latest developments in human immunology, structural biology also plays a significant role in modern vaccine discovery and development programs [20] and the interplay between structural biology and human immunology can provide in-depth insights required to enable antigen design. The work described in this thesis is an example of the application or combination of the recent human B-cell technologies with structural biology, with a focus on the latter.

A recent large study for the isolation of human mAbs from subjects immunized with the MenB vaccine Bexsero, allowed the profiling of the B-cell repertoire induced by this vaccine, and enabled the identification of mAbs specific for all the three vaccine antigens [90, 91]. Importantly, this study revealed the presence of multiple protective epitopes in each antigen and the importance of their simultaneous recognition by different antibodies for functional bactericidal activity. In the specific case of NHBA, three antibodies (10C3, 12E1, 5H2) were shown to induce bactericidal killing in serum bactericidal assays (SBA) when combined together [91]. Interestingly, bactericidal killing was observed only when one mAb against the N-terminal (i.e. 10C3 or 12E1) and one targeting the C-terminal (i.e.5H2) were used simultaneously, while this effect could not be observed when the two anti-N-terminal mAbs were coupled without 5H2. This evidence guided their choice for further investigation performed in this work.

One of the aims of this study was the structural determination of those regions of NHBA (specifically the N-terminal domain) whose structures are still unknown. To achieve this, we sought to use complexes with Fabs, as these are generally known to stabilize the antigen they bind to, and consequently to increase chances of crystallization. This is the reason why Fabs have been defined as crystallization chaperones [89], with many examples in the PDB of difficult-to-crystallize proteins whose structures have been ultimately obtained thanks to their complexation with specific Fab fragments [150-152]. Disappointingly, various and numerous attempts to obtain crystals of full length or sub-full length NHBA fragments (that included the yet to be determined N-terminal region) complexed with human Fabs that were made available by the B-cell profiling study mentioned above, failed. This is somewhat in agreement with the predicted intrinsic disorder of the first 200 residues of NHBA that was already reported [64, 130]. Also, extensive bioinformatics analyses and additional experimental evidences collected in this study, such as a detected larger hydrodynamic radius in solution and unexpected elution profiles in size exclusion chromatography and SDS-PAGE, confirmed the hypothesis of predicted disorder. In principle, we reasoned that the choice of using Fabs that targeted the N-terminal disordered regions (see Fabs 10C3 and 12E1) as crystallization chaperones could stabilize those regions and aid the crystallization of the complexes. Also, Fab 10C3 seemed the most promising candidate as its putative binding site belonged to a region for which some structured elements (α-helices and β-strands) were predicted by our bioinformatics analyses. Moreover, thermo-stability studies on the construct NHBA_114-332 revealed an unfolding event at ~55°C, which being associated to this specific region of the N-terminal domain suggests the presence of some secondary structure elements. It is tempting to speculate that these regions might contain transient and metastable conformations that could be stabilized upon binding of Fab 10C3. However, our work confirmed the difficulties in crystallizing a protein complex containing one or more flexible regions, even if coupled with a putatively stabilizing binder such as the Fab used here. In fact, none of the complexes between Fabs 12E1 or 10C3 and NHBA fragments resulted in crystals suitable for consequent structural studies.

As a rescue strategy, *in-situ* proteolysis experiments were also performed on a 10C3:NHBA complex sample. Here, we hypothesized that the addition of protease traces in the crystallization mix could cause the degradation of unstable regions of the complexes, thus facilitating the crystallization of the remaining and stable portions. This approach is known as *in-situ* proteolysis, in which the proteolytic reaction is performed directly in the crystallization drops, and it has been previously demonstrated as an effective rescue strategy for crystallization of proteins recalcitrant to crystallization [103, 153, 154]. In these examples pertaining to other unstable or flexible proteins, the use of proteases did indeed help in identifying the stable domains by removing one or more regions that being unstructured were sensitive to proteolysis [155], and thus finally allowed their successful crystallization. Although several proteases, at different concentrations, and several complex:protease molar ratios were tested, our attempts resulted only in the crystallization of free Fab 10C3, suggesting dissociation of the complex despite the high affinity between the components. It is possible that the proteolytic cleavage affected extensive regions of NHBA, thus causing dissociation of the complex. This might provide a further but indirect evidence of NHBA inherent flexibility (see protease susceptibility), also in agreement with one of the reported hallmarks of IDPs [129, 156].

One important outcome of all the attempts made here to obtain the structures of NHBA-Fab complexes was the elucidation at high-resolution of the crystal structures of all the Fabs in their unbound or apo state. These now provide the first high resolution structural characterization of human Fabs against NHBA, from the antibody repertoire of Bexsero. Analyses of the structures of these Fabs allows the fine dissection of their paratope composition and features, and can now be exploited for applications in protein design to generate for example models of the interaction with their respective epitopes. For example, analyses of the CDRs of Fab 12E1 and 10C3 showed a prevalence of short side-chain-containing residues like Ser and Thr, alongside to Tyr, which is in agreement with previous general observations on the average composition of antibodies paratopes [132, 134, 136, 157]. Interestingly, the CDRs of both Fabs 12E1 and 10C3 also displayed a significant percentage of positive and negative residues as Lys, Arg, and Asp (**Tables 4** and **7**), positioned to form charged patches on the antibody surfaces which were surrounded by amino acids with hydrophilic functional groups. For Fab 12E1, an observed enrichment in positively-charged residues could be related to the fact that its putative epitope is mainly acidic (made of residues 73-AAVSEENTGN-82 [91]). Also, the 12E1 paratope displays a marked concave topology and it is also rich in hydrophobic residues (Tyr and Trp), both features previously reported to distinguish antibodies that bind disordered epitopes or intrinsically disordered proteins [138, 158]. Also for Fab 10C3 similar patches of charged residues, especially positively-charged, were detected, which again could correlate with the composition of its putative epitope (243-KSEFEKLSDADKISNYKKDGKNDGKNDKFVGL-274 [90]). In general, an overall enrichment in charged residues in the paratopes of antibodies that bind intrinsically-disordered proteins is thought to be due to more favorable putative charge-charge interactions that in turn might result in more interfacial complementarity as compared instead with the distribution of charged residues in other 3D complexes [158].

A second major result of this work is the first co-crystal structure of a NHBA:Fab complex, obtained by co-crystallizing Fab 5H2 with a C-terminal fragment of NHBA (construct C3). The high resolution epitope mapping provided by this co-crystal structure shows how an entire face of the NHBA β-barrel is recognized as a large conformational epitope by Fab 5H2. We called this side of the antigen "proximal" or C-terminal face of NHBA, with "distal" face the one putatively interacting or interfacing with the N-terminal portion of the protein, of which we do not have structural information yet. Importantly, the epitope is highly conserved across a panel of NHBA strains (**Fig. 3.24**), and analysis of the structure revealed the molecular bases of how only minor sequence variations in the epitope composition of a few NHBA variants do not have a significant impact in the recognition of Fab 5H2. Detailed analysis of the 5H2:NHBA-C3 interface shows the presence of an extensive network of intermolecular hydrogen bonds, among which a key salt bridge between Asp100 of 5H2 and Arg339 of NHBA, as confirmed by stability and binding studies. Additionally, structural comparisons of the unbound Fab 5H2 with the NHBA-C3-bound form shows unequivocal conformational rearrangements at the level of side-chains making the close local environment of Arg339 of NHBA. These changes seem to generate a funnel-shaped pocket specifically on the 5H2 surface of the 5H2:NHBA complex, which is consequently filled by the long side-chain of NHBA Arg339. In addition, this pocket has a peculiar mixed electrostatic (floor) and hydrophobic (walls) nature, and its formation also slightly changes the local topology of the paratope surface (**Fig. 3.18C**).

The crystal structures of unbound human Fabs described in this work represent valuable models for potential new computational studies of protein modeling and design. Algorithms such as Rosetta FlexPepDock or SnugDock [159, 160], which have been reported to perform highly accurate *de-novo* folding of peptides by calculating and selecting lowest-energy states of designed sequences, could be used to generate 3D models of the 10C3 and 12E1 epitope peptides starting from the low resolution epitope mapping information already available [90, 91]. In addition, “data-driven” docking of such peptide epitopes into the structurally determined Fabs paratopes could be explored. Moreover, computational protocols for *ab-initio* folding that integrate also the experimental information available can be used to generate models of the N-terminal NHBA, around the binding region of Fab 10C3. To increase the reliability of such *ab-initio* folded models, the docked Fab-epitopes obtained above could also be used to provide structural constraints for their refinement and partial validation.

Given the importance in the identification of the exact location of B-cell epitopes, many efforts have been performed in order to develop computational tools that can predict epitopes on a given antigen. However, given the complexity of Ab-Ag interactions, the performances of these methods have been disappointing [161, 162]. It is believed that one of the reasons of this weakness is the lack of large enough standard datasets, as data for building structure-based models are still scarce although the number of resolved antigen-antibody structures available in the PDB is steadily increasing. One proven enhancement towards the robustness of these prediction methods is the use of complementary experimental structural data, such as Fabs and Fab-complexes structures [163]. Therefore, the structures of unbound and bound-Fabs obtained in this study might contribute to enhancing the datasets that the scientific community can use to increase the performance of B-cell epitopes predictions.

Overall, the structural studies performed in this work provide for the first time high-resolution data on human-derived Fabs specific for the vaccine antigen NHBA, elucidating the details of the recognition by the human immune system of a conserved epitope, and enabling in-depth analyses of the fine, atomic features of their paratopes or interfacing residues. In addition to contributing to the elucidation of the properties of one of the Bexsero antigens, these results contribute to the growing knowledge of antibody-antigen recognition mechanisms, and at the same time they provide useful experimental data for further development of structure-driven computational tools.

# Acknowledgments

Dopo quest’immane fatica siamo finalmente giunti alla sezione dei ringraziamenti, difficile da scrivere tanto quanto la discussione praticamente. Perchè ripensando a questi tre anni si affollano nella mia mente troppe cose: persone, luoghi, fallimenti (a valanghe), successi, emozioni, risate, situazioni deliranti e momenti disperatissimi che non è facile sintetizzare in poche righe.

Inizierò quindi dall’ovvio ringraziando i miei genitori Ariella e Fabio e mia sorella Silvia, con me in ogni momento (con le dovute distaze per carità) e sempre di supporto nei momenti difficili, nonchè efficientissimi nei traslochi. Il ringraziamento si estende doverosamente a tutta la mia famiglia allargata con nonni, zii e cugini, che ad ogni ritorno a casa mi ricordano quanto la nobile stirpe Maritan & Co. sia meravigliosa, variegata e fiera portatrice del gene della pazzia nelle sue forme più disparate. Un grazie speciale alle mie cugine Marialaura, Francesca e dai mettiamoci anche la Gloria (scusa Frex), da sempre più amiche che cugine.

Un grazie dal cuore alle mie amiche secolari Chiara e Giorgia che non mi hanno mai fatto sentire lontana da casa per un solo momento, hanno ascoltato pazientemente i miei deliri, consigliato nei momenti peggiori e perdonato le mie sparizioni. Ringrazio Silvia, Claudia e Valentina che si sono sempre prodigate per riuscire a salutarmi anche cinque minuti negli incasinatissimi week end padovani e non hanno mancato di farmi sentire il loro affetto nonstante le distanze.

Ringrazio Leandro, che non solo porta il peso di esser mio moroso (santo subito!), ma che ha anche trovato il coraggio di stare insieme a me durante tutta l’epopea del dottorato, ascoltanto pazientemente fiumi di paranoie, racconti di esperimenti falliti e paturnie esistenziali di ogni sorta. Un punto fisso ora e spero per tanto tempo ancora (inteso fino a quando non sarò inserita nel testamento).

Grazie alla aitanti giocatrici della Fossa, squadra di pallavolo che non solo mi ha accolto con un calore e un affetto incredibili, ma mi ha fatto sentire da subito parte di una famiglia speciale. Non posso spiegare quanto allenarmi voi sia stato bello e soprattutto determinante per mantenere la mia salute mentale più o meno stabile, nonchè per battezzare il mio alter-ego pallavolistico, oggi conosciuto come “Carla”.

Voglio ringraziare tutte le persone che ho conosciuto in questi anni a GSK, per tutti gli insegnamenti ricevuti che, nel bene o nel male, mi hanno fatto diventare la persona che sono oggi. Un grazie spassionato a tutti i componenti dell’unità di biologia strutturale (past and present) Daniele, Ilaria, Maria, Eriberto, Francesca, Pietro, Danilo, Massi, Nathalie, Lassi e Jacinto. I ringraziamenti non basterebbero mai invece per le tre pazze che ho incontrato quando sono arrivata: Alessia, Ilaria e Irene. Semplicemente non sarebbe stato quello che è stato senza amiche come voi. Seguono ringraziamenti ai miei compagni di lamentela per eccellenza Gigi, Bruna, Lucia DI e Lucia EF, per aver dato un senso alle mie innumerevoli pause e ed esser stati il mio punto di riferimento per esternare i rarissimi momenti di disagio. Ringrazio tutta la serie di personaggi notevoli che hanno colorato in vari modi questi anni tumultuosi come Enea, Filippo, Giacomo, Luca, Manuele, Michela, Maria, Diego e tutti quelli che ho dimenticato, ma se mi avete offerto da bere almeno una volta vi ringrazio di cuore perchè avete saputo cogliere i miei bisogni essenziali.

*Dulcis in fundo*, il ringraziamento più grande che mi rimane da fare è alle due persone che mi hanno giudato più di tutti in questo percorso di crescita professionale e personale, Enrico e Matt. Ringrazio Matt per la gentilezza, l'umanita e anche la puntigliosità che mi ha dimostrato. Infine grazie Enrico per la pazienza infinita, per la presenza costante, per tutti gli insegnamenti, le correzioni, gli spunti di riflessione, i confronti, le idee e per tutta la passione che è riuscito a trasmettermi.

# References

1. Delany, I., R. Rappuoli, and E. De Gregorio, *Vaccines for the 21st century.* EMBO Molecular Medicine, 2014. **6**(6): p. 708-720.
2. WHO. *Global VaccineActionPlan2011–2020*. 2013.
3. De Gregorio, E. and R. Rappuoli, *From empiricism to rational design: a personal perspective of the evolution of vaccine development.* Nat Rev Immunol, 2014. **14**(7): p. 505-14.
4. Plotkin, S., *History of vaccination.* Proc Natl Acad Sci U S A, 2014. **111**(34): p. 12283-7.
5. Rappuoli, R., *Vaccines: Science, health, longevity, and wealth.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(34): p. 12282-12282.
6. Rappuoli, R., *Reverse vaccinology, a genome-based approach to vaccine development.* Vaccine, 2001. **19**(17-19): p. 2688-91.
7. Rappuoli, R., *Reverse vaccinology.* Curr Opin Microbiol, 2000. **3**(5): p. 445-50.
8. Tettelin, H., et al., *Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.* Science, 2000. **287**(5459): p. 1809-15.
9. Rinaudo, C.D., et al., *Vaccinology in the genome era.* The Journal of Clinical Investigation, 2009. **119**(9): p. 2515-2525.
10. Liljeroos Lassi, E.M., Ilaria Ferlenghi and Matthew James Bottomley, *Structural and computational biology in the design of immunogenic vaccine antigens.* 2015.
11. Vytvytska, O., et al., *Identification of vaccine candidate antigens of Staphylococcus aureus by serological proteome analysis.* Proteomics, 2002. **2**(5): p. 580-90.
12. Couto, N., et al., *Identification of vaccine candidate antigens of Staphylococcus pseudintermedius by whole proteome characterization and serological proteomic analyses.* J Proteomics, 2016. **133**: p. 113-24.
13. Serruto, D., et al., *The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens.* Vaccine, 2012. **30 Suppl 2**: p. B87-97.
14. Donnarumma, D., et al., *The role of structural proteomics in vaccine development: recent advances and future prospects.* Expert Review of Proteomics, 2016. **13**(1): p. 55-68.
15. Chiang, M.H., et al., *Identification of novel vaccine candidates against Acinetobacter baumannii using reverse vaccinology.* Hum Vaccin Immunother, 2015. **11**(4): p. 1065-73.
16. Montigiani, S., et al., *Genomic approach for analysis of surface proteins in Chlamydia pneumoniae.* Infect Immun, 2002. **70**(1): p. 368-79.
17. Naz, A., et al., *Identification of putative vaccine candidates against Helicobacter pylori exploiting exoproteome and secretome: a reverse vaccinology based approach.* Infect Genet Evol, 2015. **32**: p. 280-91.
18. Talukdar, S., et al., *Identification of potential vaccine candidates against Streptococcus pneumoniae by reverse vaccinology approach.* Appl Biochem Biotechnol, 2014. **172**(6): p. 3026-41.
19. Xiang, Z. and Y. He, *Genome-wide prediction of vaccine targets for human herpes simplex viruses using Vaxign reverse vaccinology.* BMC Bioinformatics, 2013. **14 Suppl 4**: p. S2.
20. Rappuoli, R., et al., *Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design.* The Journal of Experimental Medicine, 2016. **213**(4): p. 469-481.
21. Lanzavecchia, A., et al., *Antibody-guided vaccine design: identification of protective epitopes.* Curr Opin Immunol, 2016. **41**: p. 62-7.
22. McLellan, J.S., et al., *Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus.* Science, 2013. **342**(6158): p. 592-8.

23. Kwon, Y.D., et al., *Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env.* Nat Struct Mol Biol, 2015. **22**(7): p. 522-31.
24. Scarselli, M., et al., *Rational design of a meningococcal antigen inducing broad protective immunity.* Sci Transl Med, 2011. **3**(91): p. 91ra62.
25. Nuccitelli, A., et al., *Structure-based approach to rationally design a chimeric protein for an effective vaccine against Group B Streptococcus infections.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10278-83.
26. Dormitzer, P.R., J.B. Ulmer, and R. Rappuoli, *Structure-based antigen design: a strategy for next generation vaccines.* Trends Biotechnol, 2008. **26**(12): p. 659-67.
27. Malito, E., A. Carfi, and M.J. Bottomley, *Protein Crystallography in Vaccine Research and Development.* Int J Mol Sci, 2015. **16**(6): p. 13106-40.
28. Malito, E., et al., *Defining a protective epitope on factor H binding protein, a key meningococcal virulence factor and vaccine antigen.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(9): p. 3304-3309.
29. Malito, E., et al., *Structure of the meningococcal vaccine antigen NadA and epitope mapping of a bactericidal antibody.* Proc Natl Acad Sci U S A, 2014. **111**(48): p. 17128-33.
30. Godley, L., et al., *Introduction of intersubunit disulfide bonds in the membrane-distal region of the influenza hemagglutinin abolishes membrane fusion activity.* Cell, 1992. **68**(4): p. 635-45.
31. Lee, J.K., et al., *Reversible Inhibition of the Fusion Activity of Measles Virus F Protein by an Engineered Intersubunit Disulfide Bridge.* Journal of Virology, 2007. **81**(16): p. 8821-8826.
32. McNeil, G., M. Virji, and E.R. Moxon, *Interactions of Neisseria meningitidis with human monocytes.* Microb Pathog, 1994. **16**(2): p. 153-63.
33. Harrison, O.B., et al., *Description and Nomenclature of Neisseria meningitidis Capsule Locus.* Emerging Infectious Diseases, 2013. **19**(4): p. 566-573.
34. Branham, S.E., *SEROLOGICAL RELATIONSHIPS AMONG MENINGOCOCCI.* Bacteriological Reviews, 1953. **17**(3): p. 175-188.
35. Rouphael, N.G. and D.S. Stephens, *Neisseria meningitidis: biology, microbiology, and epidemiology.* Methods Mol Biol, 2012. **799**: p. 1-20.
36. Rosenstein, N.E., et al., *Meningococcal disease.* N Engl J Med, 2001. **344**(18): p. 1378-88.
37. Stephens, D.S., B. Greenwood, and P. Brandtzaeg, *Epidemic meningitis, meningococcaemia, and Neisseria meningitidis.* Lancet, 2007. **369**(9580): p. 2196-210.
38. Caugant, D.A. and M.C. Maiden, *Meningococcal carriage and disease--population biology and evolution.* Vaccine, 2009. **27 Suppl 2**: p. B64-70.
39. Brooks, R., et al., *Increased case-fatality rate associated with outbreaks of Neisseria meningitidis infection, compared with sporadic meningococcal disease, in the United States, 1994-2002.* Clin Infect Dis, 2006. **43**(1): p. 49-54.
40. Virji, M., *Pathogenic neisseriae: surface modulation, pathogenesis and infection control.* Nat Rev Microbiol, 2009. **7**(4): p. 274-86.
41. Pizza, M. and R. Rappuoli, *Neisseria meningitidis: pathogenesis and immunity.* Curr Opin Microbiol, 2015. **23**: p. 68-72.
42. van Deuren, M., P. Brandtzaeg, and J.W. van der Meer, *Update on meningococcal disease with emphasis on pathogenesis and clinical management.* Clin Microbiol Rev, 2000. **13**(1): p. 144-66, table of contents.
43. Khatami, A. and A.J. Pollard, *The epidemiology of meningococcal disease and the impact of vaccines.* Expert Rev Vaccines, 2010. **9**(3): p. 285-98.
44. Lewis, L.A. and S. Ram, *Meningococcal disease and the complement system.* Virulence, 2014. **5**(1): p. 98-126.

45. Scholten, R.J., et al., *Meningococcal disease in The Netherlands, 1958-1990: a steady increase in the incidence since 1982 partially caused by new serotypes and subtypes of Neisseria meningitidis.* Clin Infect Dis, 1993. **16**(2): p. 237-46.
46. Finne, J., M. Leinonen, and P.H. Makela, *Antigenic similarities between brain components and bacteria causing meningitis. Implications for vaccine development and pathogenesis.* Lancet, 1983. **2**(8346): p. 355-7.
47. Finne, J., et al., *An IgG monoclonal antibody to group B meningococci cross-reacts with developmentally regulated polysialic acid units of glycoproteins in neural and extraneural tissues.* J Immunol, 1987. **138**(12): p. 4402-7.
48. Giuliani, M.M., et al., *A universal vaccine for serogroup B meningococcus.* Proc Natl Acad Sci U S A, 2006. **103**(29): p. 10834-9.
49. Pizza, M., et al., *Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing.* Science, 2000. **287**(5459): p. 1816-20.
50. Welsch, J.A., et al., *Antibody to genome-derived neisserial antigen 2132, a Neisseria meningitidis candidate vaccine, confers protection against bacteremia in the absence of complement-mediated bactericidal activity.* J Infect Dis, 2003. **188**(11): p. 1730-40.
51. McQuaid, F., et al., *Persistence of bactericidal antibodies to 5 years of age after immunization with serogroup B meningococcal vaccines at 6, 8, 12 and 40 months of age.* Pediatr Infect Dis J, 2014. **33**(7): p. 760-6.
52. Prymula, R., et al., *A phase 2 randomized controlled trial of a multicomponent meningococcal serogroup B vaccine (I).* Hum Vaccin Immunother, 2014. **10**(7): p. 1993-2004.
53. Fletcher, L.D., et al., *Vaccine potential of the Neisseria meningitidis 2086 lipoprotein.* Infect Immun, 2004. **72**(4): p. 2088-100.
54. Jiang, H.Q., et al., *Broad vaccine coverage predicted for a bivalent recombinant factor H binding protein based vaccine to prevent serogroup B meningococcal disease.* Vaccine, 2010. **28**(37): p. 6086-93.
55. Richmond, P.C., et al., *Safety, immunogenicity, and tolerability of meningococcal serogroup B bivalent recombinant lipoprotein 2086 vaccine in healthy adolescents: a randomised, single-blind, placebo-controlled, phase 2 trial.* Lancet Infect Dis, 2012. **12**(8): p. 597-607.
56. Serruto, D., et al., *Neisseria meningitidis GNA2132, a heparin-binding protein that induces protective immunity in humans.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3770-5.
57. Giuliani, M.M., et al., *Measuring antigen-specific bactericidal responses to a multicomponent vaccine against serogroup B meningococcus.* Vaccine, 2010. **28**(31): p. 5023-30.
58. Bambini, S., et al., *Distribution and genetic variability of three vaccine components in a panel of strains representative of the diversity of serogroup B meningococcus.* Vaccine, 2009. **27**(21): p. 2794-803.
59. Jacobsson, S., et al., *Sequence constancies and variations in genes encoding three new meningococcal vaccine candidate antigens.* Vaccine, 2006. **24**(12): p. 2161-8.
60. Lucidarme, J., et al., *Characterization of fHbp, nhba (gna2132), nadA, porA, and sequence type in group B meningococcal case isolates collected in England and Wales during January 2008 and potential coverage of an investigational group B meningococcal vaccine.* Clin Vaccine Immunol, 2010. **17**(6): p. 919-29.
61. Muzzi, A., et al., *Conservation of meningococcal antigens in the genus Neisseria.* MBio, 2013. **4**(3): p. e00163-13.
62. Bambini, S., et al., *An analysis of the sequence variability of meningococcal fHbp, NadA and NHBA over a 50-year period in the Netherlands.* PLoS One, 2013. **8**(5): p. e65043.
63. Vogel, U., et al., *Predicted strain coverage of a meningococcal multicomponent vaccine (4CMenB) in Europe: a qualitative and quantitative assessment.* Lancet Infect Dis, 2013. **13**(5): p. 416-25.

64. Esposito, V., et al., *Structure of the C-terminal domain of Neisseria heparin binding antigen (NHBA), one of the main antigens of a novel vaccine against Neisseria meningitidis.* J Biol Chem, 2011. **286**(48): p. 41767-75.
65. Flower, D.R., A.C. North, and C.E. Sansom, *The lipocalin protein family: structural and sequence overview.* Biochim Biophys Acta, 2000. **1482**(1-2): p. 9-24.
66. Calmettes, C., et al., *Structural variations within the transferrin binding site on transferrin-binding protein B, TbpB.* J Biol Chem, 2011. **286**(14): p. 12683-92.
67. Casellato, A., et al., *The C2 fragment from Neisseria meningitidis antigen NHBA increases endothelial permeability by destabilizing adherens junctions.* Cell Microbiol, 2014. **16**(6): p. 925-37.
68. Lehninger, *Principles of Biochemistry*. 5 ed. 2008.
69. Janeway CA Jr, T.P., Walport M, et al., *Immunobiology: The Immune System in Health and Disease*. 5 ed. 2001: Garland Science.
70. Halaby, D.M., A. Poupon, and J. Mornon, *The immunoglobulin fold family: sequence analysis and 3D structure comparisons.* Protein Eng, 1999. **12**(7): p. 563-71.
71. Stanfield, R.L. and I.A. Wilson, *Antibody Structure.* Microbiol Spectr, 2014. **2**(2).
72. Davies, D.R. and G.H. Cohen, *Interactions of protein antigens with antibodies.* Proc Natl Acad Sci U S A, 1996. **93**(1): p. 7-12.
73. Davies, D.R., E.A. Padlan, and D.M. Segal, *Three-dimensional structure of immunoglobulins.* Annu Rev Biochem, 1975. **44**: p. 639-67.
74. Sela-Culang, I., V. Kunik, and Y. Ofran, *The Structural Basis of Antibody-Antigen Recognition.* Front Immunol, 2013. **4**.
75. Yu, C.M., et al., *Rationalization and design of the complementarity determining region sequences in an antibody-antigen recognition interface.* PLoS One, 2012. **7**(3): p. e33340.
76. Chothia, C. and A.M. Lesk, *Canonical structures for the hypervariable regions of immunoglobulins.* J Mol Biol, 1987. **196**(4): p. 901-17.
77. Martin, A.C. and J.M. Thornton, *Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies.* J Mol Biol, 1996. **263**(5): p. 800-15.
78. Al-Lazikani, B., A.M. Lesk, and C. Chothia, *Standard conformations for the canonical structures of immunoglobulins1.* Journal of Molecular Biology, 1997. **273**(4): p. 927-948.
79. North, B., A. Lehmann, and R.L. Dunbrack, Jr., *A new clustering of antibody CDR loop conformations.* J Mol Biol, 2011. **406**(2): p. 228-56.
80. Abbott, W.M., M.M. Damschroder, and D.C. Lowe, *Current approaches to fine mapping of antigen-antibody interactions.* Immunology, 2014. **142**(4): p. 526-35.
81. Wilson, P.C. and S.F. Andrews, *Tools to therapeutically harness the human antibody response.* Nat Rev Immunol, 2012. **12**(10): p. 709-19.
82. Back, J.W. and J.P. Langedijk, *Structure-based design for high-hanging vaccine fruits.* Adv Immunol, 2012. **114**: p. 33-50.
83. Gershoni, J.M., et al., *Epitope mapping: the first step in developing epitope-based vaccines.* BioDrugs, 2007. **21**(3): p. 145-56.
84. Ladner, R.C., *Mapping the Epitopes of Antibodies.* Biotechnology and Genetic Engineering Reviews, 2007. **24**(1): p. 1-30.
85. Hudson, E.P., M. Uhlen, and J. Rockberg, *Multiplex epitope mapping using bacterial surface display reveals both linear and conformational epitopes.* Sci Rep, 2012. **2**: p. 706.
86. Barlow, D.J., M.S. Edwards, and J.M. Thornton, *Continuous and discontinuous protein antigenic determinants.* Nature, 1986. **322**(6081): p. 747-748.
87. Van Regenmortel, M.H.V., *Mapping Epitope Structure and Activity: From One-Dimensional Prediction to Four-Dimensional Description of Antigenic Specificity.* Methods, 1996. **9**(3): p. 465-72.

88. Scietti, L., et al., *Exploring host-pathogen interactions through genome wide protein microarray analysis.* Sci Rep, 2016. **6**: p. 27996.
89. Griffin, L. and A. Lawson, *Antibody fragments as tools in crystallography.* Clin Exp Immunol, 2011. **165**(3): p. 285-91.
90. Masignani, V., *Characterization of the human antibody repertoire to type B meningococcal vaccine.* 19th International Pathogenic Neisseria Conference 2014. **Poster #35**
91. Giuliani, M., *Characterization of the human antibody repertoire to type B meningococcus vaccine.* 12th Congress of The European Meningococcal and Haemophilus Disease Society, 2015. **Poster #43**.
92. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nat. Protocols, 2015. **10**(6): p. 845-858.
93. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server.* Nat Protoc, 2009. **4**(3): p. 363-71.
94. Adamczak, R., A. Porollo, and J. Meller, *Combining prediction of secondary structure and solvent accessibility in proteins.* Proteins, 2005. **59**(3): p. 467-75.
95. Buchan, D.W.A., et al., *Scalable web services for the PSIPRED Protein Analysis Workbench.* Nucleic Acids Research, 2013. **41**(Web Server issue): p. W349-W357.
96. Drozdetskiy, A., et al., *JPred4: a protein secondary structure prediction server.* Nucleic Acids Res, 2015. **43**(W1): p. W389-94.
97. Cilia, E., et al., *The DynaMine webserver: predicting protein dynamics from sequence.* Nucleic Acids Research, 2014. **42**(Web Server issue): p. W264-W270.
98. Dosztanyi, Z., et al., *IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content.* Bioinformatics, 2005. **21**(16): p. 3433-4.
99. Lieutaud, P., B. Canard, and S. Longhi, *MeDor: a metaserver for predicting protein disorder.* BMC Genomics, 2008. **9 Suppl 2**: p. S25.
100. Studier, F.W. and B.A. Moffatt, *Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes.* J Mol Biol, 1986. **189**(1): p. 113-30.
101. van den Berg, S., et al., *Improved solubility of TEV protease by directed evolution.* J Biotechnol, 2006. **121**(3): p. 291-8.
102. Karlsson, R., et al., *Analyzing a kinetic titration series using affinity biosensors.* Anal Biochem, 2006. **349**(1): p. 136-47.
103. Tong, Y., et al., *Salvage or recovery of failed targets by in situ proteolysis.* Methods Mol Biol, 2014. **1140**: p. 179-88.
104. Dong, A., et al., *In situ proteolysis for protein crystallization and structure determination.* Nat Methods, 2007. **4**(12): p. 1019-21.
105. Kabsch, W., *XDS.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
106. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 235-42.
107. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
108. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
109. Bricogne, G., et al., *BUSTER version 2.11.4.* Cambridge, United Kingdom: Global Phasing Ltd, 2016.
110. Matthews, B.W., *Solvent content of protein crystals.* J Mol Biol, 1968. **33**(2): p. 491-7.
111. Vagin, A. and A. Lebedev, *MoRDa, an automatic molecular replacement pipeline.* Acta Crystallographica Section A, 2015. **71**(a1): p. s19.
112. McCoy, A.J., et al., *Phaser crystallographic software.* Journal of Applied Crystallography, 2007. **40**(Pt 4): p. 658-674.

113. Rould, M.A. and C.W. Carter, Jr., *Isomorphous difference methods.* Methods Enzymol, 2003. **374**: p. 145-63.
114. Incardona, M.-F., et al., *EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis.* Journal of Synchrotron Radiation, 2009. **16**(6): p. 872-879.
115. Battye, T.G., et al., *iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 271-81.
116. Zwart, P.H., et al., *Automated structure solution with the PHENIX suite.* Methods Mol Biol, 2008. **426**: p. 419-35.
117. Teplyakov, A., et al., *Antibody modeling assessment II. Structures and models.* Proteins, 2014. **82**(8): p. 1563-82.
118. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
119. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.
120. Lerner M. G., C., H. A. and A. Arbor:, *APBS Plugin for PyMOL*. 2006: University of Michigan.
121. Bond, C.S. and A.W. Schuttelkopf, *ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 5): p. 510-2.
122. Tompa, P., *The interplay between structure and function in intrinsically unstructured proteins.* FEBS Lett, 2005. **579**(15): p. 3346-54.
123. Uversky, V.N., *Biophysical Methods to Investigate Intrinsically Disordered Proteins: Avoiding an "Elephant and Blind Men" Situation.* Adv Exp Med Biol, 2015. **870**: p. 215-60.
124. Uversky, V.N., *Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins.* J Biol Chem, 2016. **291**(13): p. 6681-8.
125. Mitrea, D.M., et al., *Disorder-function relationships for the cell cycle regulatory proteins p21 and p27.* Biol Chem, 2012. **393**(4): p. 259-74.
126. Csizmok, V., et al., *Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling.* Chem Rev, 2016. **116**(11): p. 6424-62.
127. Meszaros, B., et al., *Molecular principles of the interactions of disordered proteins.* J Mol Biol, 2007. **372**(2): p. 549-61.
128. Dunker, A.K., et al., *Intrinsic disorder and protein function.* Biochemistry, 2002. **41**(21): p. 6573-82.
129. Ball, K.A., D.E. Wemmer, and T. Head-Gordon, *Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-β Peptides.* The journal of physical chemistry. B, 2014. **118**(24): p. 6405-6416.
130. Martino, A., et al., *Structural characterisation, stability and antibody recognition of chimeric NHBA-GNA1030: an investigational vaccine component against Neisseria meningitidis.* Vaccine, 2012. **30**(7): p. 1330-42.
131. Deller, M.C., L. Kong, and B. Rupp, *Protein stability: a crystallographer's perspective.* Acta Crystallographica Section F, 2016. **72**(2): p. 72-95.
132. Ramaraj, T., et al., *Antigen-antibody interface properties: composition, residue interactions, and features of 53 non-redundant structures.* Biochim Biophys Acta, 2012. **1824**(3): p. 520-32.
133. Mian, I.S., A.R. Bradwell, and A.J. Olson, *Structure, function and properties of antibody binding sites.* J Mol Biol, 1991. **217**(1): p. 133-51.
134. Kringelum, J.V., et al., *Structural analysis of B-cell epitopes in antibody:protein complexes.* Mol Immunol, 2013. **53**(1-2): p. 24-34.
135. Ofran, Y., A. Schlessinger, and B. Rost, *Automated identification of complementarity determining regions (CDRs) reveals peculiar characteristics of CDRs and B cell epitopes.* J Immunol, 2008. **181**(9): p. 6230-5.

136. Peng, H.P., et al., *Origins of specificity and affinity in antibody-protein interactions.* Proc Natl Acad Sci U S A, 2014. **111**(26): p. E2656-65.

137. Wernimont, A. and A. Edwards, *In situ proteolysis to generate crystals for structure determination: an update.* PLoS One, 2009. **4**(4): p. e5094.

138. Wong, E.T., D. Na, and J. Gsponer, *On the importance of polar interactions for complexes containing intrinsically disordered proteins.* PLoS Comput Biol, 2013. **9**(8): p. e1003192.

139. Cehlar, O., et al., *Crystallization and preliminary X-ray diffraction analysis of tau protein microtubule-binding motifs in complex with Tau5 and DC25 antibody Fab fragments.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2012. **68**(Pt 10): p. 1181-5.

140. Sela-Culang, I., S. Alon, and Y. Ofran, *A systematic comparison of free and bound antibodies reveals binding-related conformational changes.* J Immunol, 2012. **189**(10): p. 4890-9.

141. Sotriffer, C.A., et al., *Ligand-induced domain movement in an antibody Fab: molecular dynamics studies confirm the unique domain movement observed experimentally for Fab NC6.8 upon complexation and reveal its segmental flexibility.* J Mol Biol, 1998. **278**(2): p. 301-6.

142. Sotriffer, C.A., et al., *Elbow flexibility and ligand-induced domain rearrangements in antibody Fab NC6.8: large effects of a small hapten.* Biophys J, 2000. **79**(2): p. 614-28.

143. Stanfield, R.L., et al., *Antibody elbow angles are influenced by their light chain class.* J Mol Biol, 2006. **357**(5): p. 1566-74.

144. Keskin, O., *Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies.* BMC Struct Biol, 2007. **7**: p. 31.

145. Sundberg, E.J., *Structural basis of antibody-antigen interactions.* Methods Mol Biol, 2009. **524**: p. 23-36.

146. Pötzsch, S., et al., *B Cell Repertoire Analysis Identifies New Antigenic Domains on Glycoprotein B of Human Cytomegalovirus which Are Target of Neutralizing Antibodies.* PLoS Pathog, 2011. **7**(8): p. e1002172.

147. Kabanova, A., et al., *Antibody-driven design of a human cytomegalovirus gHgLpUL128L subunit vaccine that selectively elicits potent neutralizing antibodies.* Proc Natl Acad Sci U S A, 2014. **111**(50): p. 17965-70.

148. Krause, J.C., et al., *A broadly neutralizing human monoclonal antibody that recognizes a conserved, novel epitope on the globular head of the influenza H1N1 virus hemagglutinin.* J Virol, 2011. **85**(20): p. 10905-8.

149. Corti, D., et al., *Analysis of Memory B Cell Responses and Isolation of Novel Monoclonal Antibodies with Neutralizing Breadth from HIV-1-Infected Individuals.* PLoS ONE, 2010. **5**(1): p. e8805.

150. Uysal, S., et al., *Crystal structure of full-length KcsA in its closed conformation.* Proc Natl Acad Sci U S A, 2009. **106**(16): p. 6644-9.

151. Granata, V., et al., *Comparison of the crystallization and crystal packing of two Fab single-site mutant protein L complexes.* Acta Crystallographica Section D, 2005. **61**(6): p. 750-754.

152. McKinstry, W.J., et al., *Crystallization of the receptor-binding domain of parathyroid hormone-related protein in complex with a neutralizing monoclonal antibody Fab fragment.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2009. **65**(Pt 4): p. 336-8.

153. Civril, F. and K.P. Hopfner, *Crystallization of mouse RIG-I ATPase domain: in situ proteolysis.* Methods Mol Biol, 2014. **1169**: p. 27-35.

154. Bai, Y., T.C. Auperin, and L. Tong, *The use of in situ proteolysis in the crystallization of murine CstF-77.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2007. **63**(Pt 2): p. 135-8.

155. Koth, C.M., et al., *Use of limited proteolysis to identify protein domains suitable for structural analysis.* Methods Enzymol, 2003. **368**: p. 77-84.

156. Hansen, J.C., et al., *Intrinsic protein disorder, amino acid composition, and histone terminal domains.* J Biol Chem, 2006. **281**(4): p. 1853-6.

157. Koide, S. and S.S. Sidhu, *The importance of being tyrosine: lessons in molecular recognition from minimalist synthetic binding proteins.* ACS Chem Biol, 2009. **4**(5): p. 325-34.
158. MacRaild, C.A., et al., *Antibody Recognition of Disordered Antigens.* Structure, 2016. **24**(1): p. 148-57.
159. Raveh, B., et al., *Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors.* PLoS One, 2011. **6**(4): p. e18934.
160. Sircar, A. and J.J. Gray, *SnugDock: paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models.* PLoS Comput Biol, 2010. **6**(1): p. e1000644.
161. Ponomarenko, J.V. and P.E. Bourne, *Antibody-protein interactions: benchmark datasets and prediction tools evaluation.* BMC Struct Biol, 2007. **7**: p. 64.
162. Vajda, S., *Classification of protein complexes based on docking difficulty.* Proteins, 2005. **60**(2): p. 176-80.
163. Kringelum, J.V., et al., *Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking.* PLoS Comput Biol, 2012. **8**(12): p. e1002829.