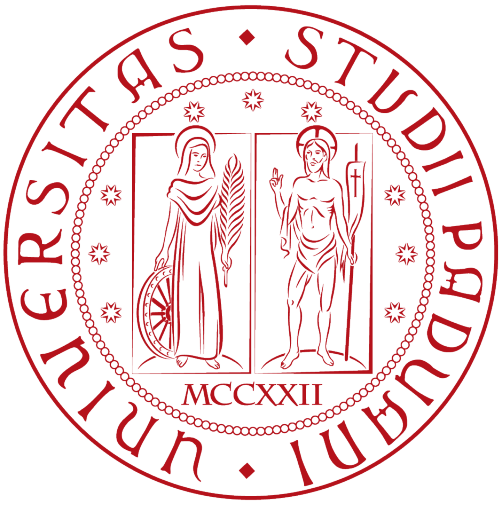

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento Di Scienze Del Farmaco
Corso Di Laurea Magistrale In Pharmaceutical Biotechnologies

TESI DI LAUREA

Production and Biophysical Characterization of Nanobodies that Recognize Proteins Mimicking SARS-CoV-2 Spike Epitopes.

RELATRICE:
Prof. Patrizia Polverino De Laureto

CORRELATORE ESTERNO:
Prof. Francisco Conejero Lara

UNIVERSIDAD DE GRANADA

Facultad de Ciencias

LAUREANDO:
Trolese Philipp

ANNO ACADEMICO 2022/2023

Contents

Abstract

Emerging SARS–CoV–2 variants of concern presenting an array of mutations in the Spike's S1 subunit are weakening the protective effects of vaccines and monoclonal antibodies, which are usually targeted to the receptor binding domain (RBD) domain. This calls for variant proof therapeutics that target more conserved and functionally important epitopes outside the S1 subdomain. Membrane fusion step during infection is driven by interaction between the highly conserved heptad repeat (HR) S2 regions HR1 and HR2, and thus pose as attractive targets for novel fusion inhibitors. HR1 and HR2 are known to become transiently exposed during the membrane fusion process and previous efforts to design peptidomimetic fusion inhibitors confirm that they can be targeted to achieve cell entry inhibition.

Herein, we used an HR1 based mimetic protein (L3C) to select camelid variable domains of heavy chain only (VHH) antibodies (also known as 'nanobodies' (NBs)) that bind to HR1 in vitro after alpaca immunization. Three of them belonging to a larger pool (NB113, NB248 and NB118) were investigated in this work. The first part centered around the expression, purification and structural characterization of the NBs. Results indicate that they can be produced with good yields and that they behave as expected folding in monomeric form and adhering to the canonical characteristics of NBs. In a second step we characterized the binding mode of the NBs to the HR1 mimetic protein by employing biophysical methods such as differential scanning and isothermal titration calorimetry. Results indicate a strong interaction with HR1 with affinities in the range of 100 nM. According to our results only NB118 is able to compete with the HR1–HR2 binding. NB113 binding to HR1 is only partially displaced by HR2, but no clear evidence of competition was obtained. Lastly NB248 appears to form a ternary complex with HR1 and HR2, showing no signs of binding displacement. Experiments aimed to map the binding epitopes of the NBs on HR1 revealed that all bind predominantly to the C–terminal region of HR1, implying that this region could be substantially more antigenic than the N–terminal part. This work explores for the first time the possibility of using S2 targeting NBs as SARS–CoV–2 fusion inhibitors and may help provide a structural framework for the quest to find therapeutic NBs with cross- neutralizing potential.

Keywords: *SARS–CoV–2, Nanobodies, Fusion Inhibitors, Protein Purification, Biophysics.*

Resumen

Variantes preocupantes emergentes del SARS-CoV-2 que presentan una serie de mutaciones en el subdominio S1 de la espícula están debilitando los efectos protectores de las vacunas y los anticuerpos monoclonales, que generalmente se dirigen al dominio de unión al receptor (RBD). Esto requiere nuevas terapias resistentes a variantes que apunten a epítopos más conservados e importantes desde el punto de vista funcional fuera del subdominio S1. La etapa de fusión de membrana durante la infección es impulsado por la interacción entre las regiones altamente conservadas de hepta-repetición (HR) de S2, HR1 y HR2, y por lo tanto, ambas representan dianas atractivas para nuevos inhibidores de fusión. Se sabe que HR1 y HR2 se exponen transitoriamente durante el proceso de fusión de las membranas y los esfuerzos previos para diseñar inhibidores de fusión peptidomiméticos confirman que se pueden dirigir para lograr la inhibición de la entrada del virus en la célula.

En este trabajo, utilizamos una proteína mimética de la región HR1 (L3C) para seleccionar anticuerpos camélidos de dominio variable sencillo de cadena pesada (VHH) (también conocidos como 'nanocuerpos' o, en Inglés 'nanobodies' (NBs)) que se unen a HR1 in vitro tras la inmunización de alpacas. Tres de ellos, pertenecientes a un grupo más grande (NB113, NB248 y NB118), fueron investigados en este trabajo. La primera parte se centró en la expresión, purificación y caracterización estructural de los NBs. Los resultados indican que pueden producirse con buenos rendimientos y que se comportan como se espera, plegándose en forma monomérica y cumpliendo con las características canónicas de los NBs. En una segunda etapa, caracterizamos el modo de unión de los NBs a la proteína mimética HR1 mediante métodos biofísicos como la calorimetría diferencial de barrido y la calorimetría isoterma de titulaciòn. Los resultados indican una fuerte interacción con HR1, con afinidades en el entorno de cien nM. Según nuestros resultados, solo NB118 es capaz de competir con la unión HR1-HR2. La unión de NB113 a HR1 solo es parcialmente desplazada por HR2, pero no se obtuvo evidencia clara de competición. Por último, NB248 parece formar un complejo ternario con HR1 y HR2, sin mostrar signos de desplazamiento de unión. Experimentos realizados con el objetivo de mapear los epítopos de unión de los NBs en HR1 revelaron que todos se unen predominantemente a la región C-terminal de HR1, lo que implica que dicha región podría ser sustancialmente más antigénica que la parte N-terminal. Este trabajo explora por primera vez la posibilidad de utilizar NBs dirigidos a S2 como inhibidores de fusión del SARS-CoV-2 y puede ayudar a proporcionar un marco estructural para la búsqueda de NBs terapéuticos con potencial de neutralización cruzada.

Palabras clave: *SARS–CoV–2, Nanocuerpos, Inhibidor de fusión, Purificación de proteínas, Biofísica.*

Sommario

Varianti emergenti preoccupanti del SARS-CoV-2, che presentano una serie di mutazioni nel sottodominio S1 della spinula, stanno indebolendo gli effetti protettivi di vaccini e anticorpi monoclonali, che di solito mirano al dominio che lega il recettore (RBD). Perciò si necessitano terapie in grado di resistere alle varianti e che mirano ad epitopi più conservati e funzionalmente importanti al di fuori di S1. Il passaggio di fusione delle membrane durante l'infezione è guidato dall'interazione tra le regioni conservate della ripetizione eptadica (HR) del dominio S2, HR1 e HR2, e rappresentano entrambe obiettivi interessanti per nuovi inibitori di fusione. Si sa che HR1 e HR2 vengono esposte transitoriamente durante il processo di fusione delle membrane e gli sforzi precedenti per progettare inibitori di fusione peptidomimetici confermano che possono essere bersagliate per ottenere l'inibizione dell'ingresso nella cellula.

In questo studio, abbiamo utilizzato una proteina mimetica basata su HR1 (L3C) per selezionare domini variabili (VHH) di anticorpi camelidi con solo la catena pesante (noti anche come "nanobodies" (NBs) in inglese), che si legano a HR1 in vitro dopo l'immunizzazione di alpache. Tre di essi appartenenti a un gruppo più ampio (NB113, NB248 e NB118) sono stati analizzati in questo lavoro. La prima parte si è concentrata sull'espressione, la purificazione e la caratterizzazione strutturale dei NBs. I risultati indicano che possono essere prodotti con buone rese e che si comportano come previsto, piegandosi nella forma monomerica e aderendo alle caratteristiche canoniche dei NBs. In un secondo passaggio, abbiamo caratterizzato la modalità di legame dei NBs alla proteina mimetica HR1 utilizzando metodi biofisici come la calorimetria differenziale a scansione e la caloimetria isotermica di titolazione. I risultati indicano una forte interazione con HR1, con affinità nei dintorni di 100 nM. Secondo i nostri risultati, solo NB118 è in grado di competere con il legame HR1-HR2. Il legame di NB113 a HR1 è solo parzialmente indebolito da HR2, ma non sono state ottenute evidenze chiare di competizione. Infine, NB248 sembra formare un complesso ternario con HR1 e HR2, senza mostrare segni di dispiazzamento del legame. Esperimenti svolti con l'obbiettivo di mappare gli epitopi di legame dei NBs su HR1 hanno rivelato che tutti si legano prevalentemente alla regione C-terminale di HR1, il che potrebbe implicare che detta regione sia più antigenica della parte N-terminale. Questo studio esplora per la prima volta la possibilità di utilizzare NBs che bersagliano S2 come inibitori di fusione del SARS-CoV-2 e potrebbe contribuire a fornire una struttura di base per la ricerca di NBs terapeutici con potenziale di neutralizzazione incrociata.

Parole chiave: *SARS–CoV–2, Nanocorpi, Inibitori di fusione, Purificazine di proteine, Biofisica.*

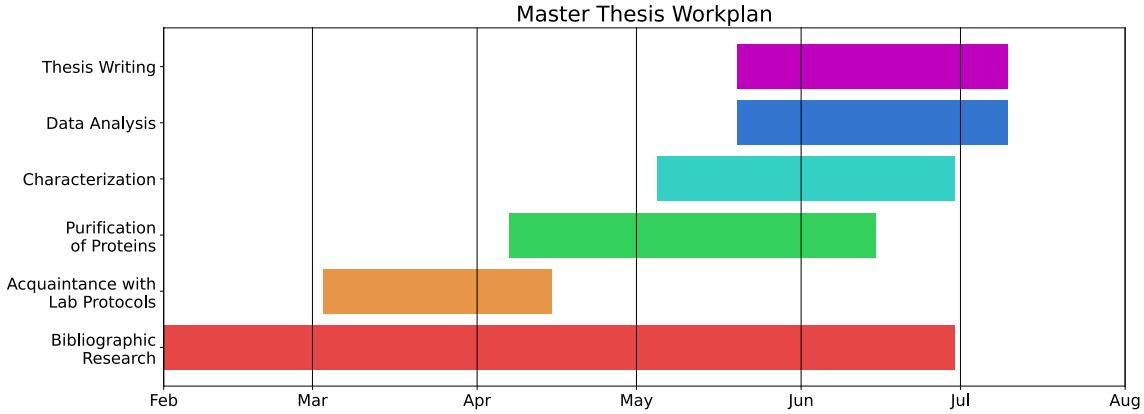

Figure 1: *Gantt diagram of the workplan followed for the completion of the master's thesis. The x–axis displays the time period in months from February 2023 to August 2023. The y–axis lists the main tasks carried out during the stay at Dr. Francisco Conejero Lara's lab. Notably, the activity of bibliographic research was spent the most time on, starting even before the actual work in the lab and lasting almost up to the deposit of the thesis. The first task once arrived at the lab, involved getting used to the equipment and instrumentation. Once acquaintance with the protocols was acquired the actual work relevant for this study started with the purification of the used proteins, which were continuously re–produced when more were needed. After successful production the proteins were characterized by the biophysical methods described in this work. The stay at the lab concluded with the analysis of the obtained data and writing of this manuscript.*

1 Introduction

1.1 COVID–19 pandemic

2023 will mark the fourth year of the Covid–19 pandemic. In this time Sars–CoV–2 caused more than 700 million infections and close to 7 million deaths worldwide [1]. Fortunately, our growing knowledge of the molecular basis of the disease and unprecedented advances in vaccine technology improved our abilities to fight the pandemic reaching the point where Sars–CoV–2 has become less of a disruption to our lives, making it possible to resume pre–pandemic activities. Nevertheless, frequently emerging variants of concern (VOCs), like Omicron sublineages with improved immune evasion and transmission, still pose a challenge to the control and management of the Covid–19 pandemic [2, 3]. Adding this to the fact that immunization provided by current vaccines has been shown to decline in effectiveness after several months and the lack of antiviral strategies to treat the infection [4], it is important not to rest on one's laurels and to continuously put more effort into the research of new treatments and variant proof immunization strategies by targeting more conserved and functionally essential sites.

1.2 Structural Biology of SARS–CoV–2 Cell Entry Mechanism

General Structure. Besides of many accessory proteins SARS-CoV-2 is made up of four main structural proteins, being the spike (S) glycoprotein, the small envelope (E) glycoprotein, the membrane (M) glycoprotein and the nucleocapsid (N) glycoprotein [5]. The S protein promotes cell entry by using the angiotensin converting enzyme 2 (ACE2) as its host receptor [6]. The spike protein is a trimer of heterodimers composed of the N-terminal S1 and the C-terminal S2 subunits [7]. The S1 subunit is responsible for interactions with the host receptor and consists of a N-terminal domain (NTD), a receptor binding domain (RBD) and two structurally conserved C-terminal subdomains (CTD1 and CTD2). The three RBDs of the S trimer, which make up the protein's apex, can exist in either one of two conformations: "up," which makes them receptor accessible, or "down," which makes them receptor inaccessible [8]. SARS–CoV–2 cell entry is characterized by dramatic conformational changes in the S–protein, making it possible to distinguish clearly between a pre–fusion and a post–fusion conformation. The driving force for these structural rearrangements are the interactions between the S1 and S2 subunits. The S2 region is covered and locked in the pre–fusion conformation by three S1 subunits [7]. In the pre–fusion S2 structure the central trimeric core is organized around a coiled-coil made by the central helix (CH) and the connector domain (CD). The HR1 is bent and the HR2 is part of the stem of the spike, although its structure is not known. The 'fusion-peptide proximal region' (FPPR), a structural component in S2 seems to assist stabilize the pre–fusion conformation

of the S trimer and clamp down the RBD [7]. The complete sequence of the Sars–CoV–2 spike protein with all its relevant domains is displayed in figure 2 (a).

Cell Entry Mechanism. Two protease cleavage events, the first at the S1–S2 junction and the second at the S2' site, internal to the S2 subunit, are necessary for SARS–CoV–2 to enter cells [9]. Furin pre–cleaves the polybasic sequence at the S1–S2 junction as the virus matures in an infected cell before ACE2 binding, but the S2' site is only cleaved after RBD binding to ACE2. The S2' location is made accessible to proteolysis by conformational changes brought on by binding to ACE2. Transmembrane serine protease 2 (TMPRSS2) expression levels in the target cell determine which of two entrance mechanisms the virus uses. The virus-ACE2 complex is internalized by the cell if it has insufficient amounts of TMPRSS2 on its surface [10]. This is done through a clathrin–mediated endocytosis step into endolysosomes, where cathepsins, which must be activated in an acidic environment, conduct the S2' cleavage. On the other hand, the S2' site is cleaved at the cell surface if there is enough TMPRSS2 present [6]. As a result, the activated spike protein is only visible on the cell surface during the cell surface entrance process. Both processes involve the S2' cleavage exposing the fusion peptide (FP), which is forced into the cell membrane by the drastic conformational changes caused by the dissociation of S1 and S2.

Transition to the Post–Fusion Conformation. In order overcome the natural repulsion between the virus and the cell membranes, the S2 subunits fold into an energetically stable state as a result of the large conformational changes they undergo. As a result of proteolytic activation and FP insertion into the cell membrane, the S protein acquires a metastable state, which is prone to change into a lower–energy state prior to membrane fusion [11]. The FPPR and CTD2 have been identified to play a crucial role in modulating S protein fusogenic structural rearrangements [9]. Another essential element for the structural rearrangement of the S protein necessary for membrane fusion is the '630 loop' in CTD2. It has been demonstrated that it is mostly disordered in the Wuhan–Hu–1 strain but ordered when the D614G mutation, which stabilizes the S trimer, is present [12]. The spike protein with the D614G mutation in the pre–fusion conformation is depicted in figure 2 (b). The FPPR and the 630 loop help maintaining the RBDs in the down conformation but move out of their positions when the adjacent RBD flips up [7]. When ACE2 captures the RBD–up conformation, the FPPR and the 630 are expelled from their positions. The FPPR shift may help expose the S2' site for cleavage, while departure of the 630 loop from the hydrophobic surface of CTD2 can destabilize this domain freeing the N–terminal segment of S2 from S1. Altogether, this shift ends up in the dissociation of S1 from S2 owing to the pre–cleavage at the S1–S2 junction by furin. Concurrently with this fusogenic shift, HR1 acquires a coiled coil structure extending the central helix and provides the thrust to propel the FP into

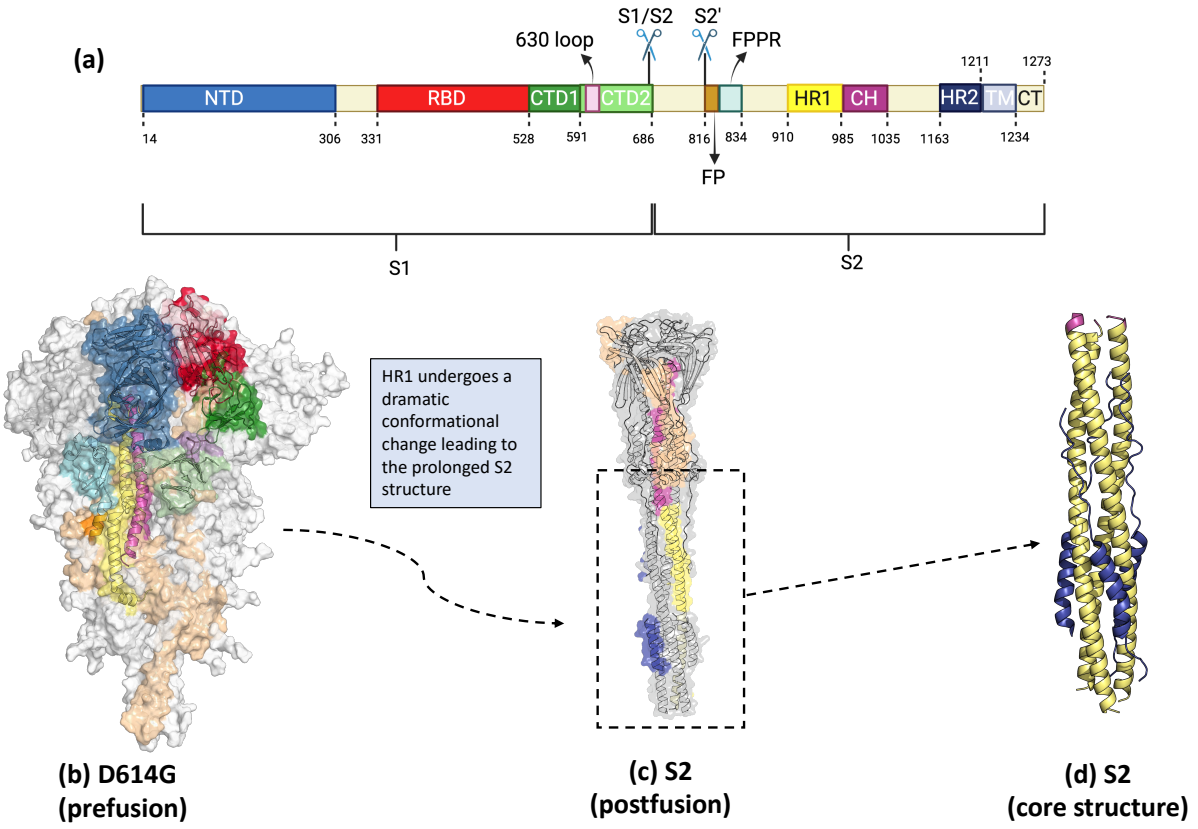

Figure 2: ***(a)*** *Domain organization of the Sars–CoV–2 S protein with all relevant sites highlighted (made with BioRender).* ***(b)*** *Cryo-electron microscopy structures of detergent-solubilized full-length S trimers from the D614G variant with all three RBDs in the down conformation (PDB ID: 7KRQ).* ***(c)*** *EM image of the S2 post–fusion structure (PDB ID: 6XRA) highlighting the extended conformation of HR1.* ***(d)*** *Core structure of the post–fusion conformation showing the 6HB (PDB ID: 6LXT) formed by HR1 and HR2. In* ***(b) - (d)*** *Subdomains of one monomer are colored according to the color scheme used in* ***(a)****, the other two monomers are displayed in grey. The protein images have been rendered in PyMol.*

the target–cell membrane [13]. This allows the fusogenic shift towards a stable postfusion structure where HR1 and HR2 bind forming the 6 helix bundle (6HB), displayed in figure 2 (c). As a consequence of HR2 folding back, the fusion peptide and transmembrane segments are located at the same end of the molecule [14]. This proximity leads to the membranes with which they interact to bend inwards, resulting in membrane fusion. Viral RNA is released into the cytoplasm of the host cell through a fusion pore formed when the viral and cellular membranes fuse, where it can uncoil and reproduce.

Post–Fusion Structure. In the post–fusion S2 structure the core helix and HR1 combine to produce an extended central three–stranded coiled coil visible in figure 2 (d). The HR2 peptide and its upstream linker region bind into the hydrophobic groove at the interface

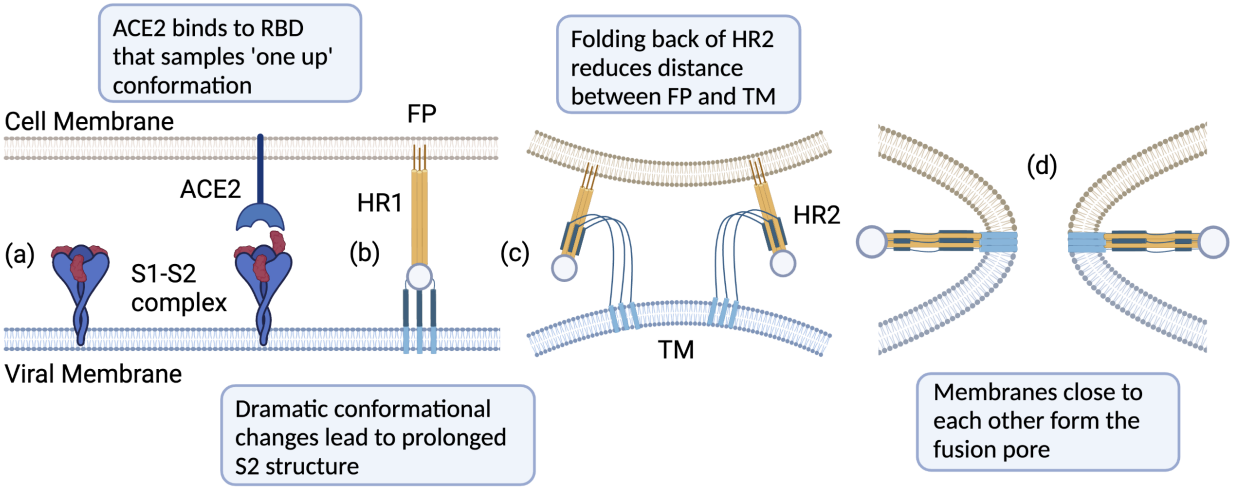

Figure 3: *Schematic drawing of the fusion mechanism of SARS–CoV–2 with the cellular membrane illustrating the main structural transitions (made with BioRender).* ***(a)*** *RBD fluctuates between 'three down' and 'one up' conformation. The latter is bound by ACE2 once achieved.* ***(b)*** *Binding to ACE2 results in the cleavage at the S2' site that initiates a cascade of refolding events in S1 altogether ending with the dissociation of S1 and a prolonged conformation.* ***(c)*** *Folding back of HR2 completes the post–fusion structure and reduces the distance between the two membranes.* ***(d)*** *The latter step facilitates membrane fusion and ultimately pore formation.*

between two HR1–CH helices of the neighboring protomers in an antiparallel manner [13], thus forming a highly stable 6HB. This structural rearrangement into the 6HB comes to a conclusion only during and shortly after pore formation. It is widely believed that the energy generated during the folding processes that result in the production of 6HB is essential for breaking through the high energy barrier needed to fuse two hydrophobic, negatively charged membranes.

1.3 Current Inhibition Strategies

Targeted inhibitors of SARS-CoV-2 are designed to prevent the virus from completing various steps in its life and infection cycle, such as entry, proteolytic processing, RNA synthesis, and assembly. The search for the "perfect" inhibitor is very complex because each stage may have several protein targets and within those multiple target-able domains. As a result, the emphasis of this section will be on outlining the inhibitory tactics that can be used during the cell entry stage. The homotrimeric spike protein, which is located on the surface of the virion, is essential for cell entrance. Targeted disruption of spike–ACE2 interaction takes advantage of inhibitors such as neutralizing antibodies, small-molecule inhibitors, and peptide inhibitors.

Anti–spike Antibodies. This group is the most numerous amongst the spike inhibitors,

with more than 300 anti–SARS–CoV–2 monoclonal antibodies (mAbs), including more than 20 candidates in clinical trials [15]. To date, a few anti-SARS-CoV-2 mAbs and mAb cocktails have been approved or granted emergency use authorization (EUA), e.g. bebtelovimab, sotrovimab, regdanvimab and bamlanivimab plus etesevimab. The general structure of an antibody is described in section 1.5. Identification of initial promising potent wild–type mAbs includes screening of antibodies from the memory B cells of convalescent patients or humanized mice exposed to SARS–CoV–2. Subsequently, they can also be engineered to improve their effector function. Although IgG1 mAbs have been the main emphasis up until this point, interest is developing towards polyclonal antibodies and other formats of antigen recognizing proteins. Most antibodies in this group target the RBD domain of the spike protein to block the interaction with ACE2.

The S1 subunit, and particularly the RBD domain, induce the dominant neutralizing antibody (nAbs) response in the host and serves as primary antigen for vaccine design [16, 17]. However, selective pressure from the host acts on the S1 domain leading to the generation of RBD variants with retained or even amplified ACE2 binding [18]. For instance, the omicron (B.1.1.529) variant displays a large number of mutations in the RBD domain and has been demonstrated to be able to escape from neutralizing antibody responses [19]. To counteract neutralization escape it might be crucial to target highly conserved, non–overlapping epitopes within and especially outside the RBD.

Anti–spike small molecules, peptides and engineered proteins. Small molecules such as clofazimine have shown promising results as broad coronavirus inhibitors [20], although so far no results from late-stage clinical trials have been reported. Clofazimine interferes with Spike–ACE2 interaction but the exact inhibitory mechanism remains unknown.

Evisbep, a thermostable designed ankyrin repeat protein (DARPins), is an antibody mimetic protein that targets the spike RBD to block the interaction with ACE2. Although showing no clinical benefits in a phase III trial [21] in hospitalized patients, results were more promising in a phase II/III trial on outpatients with symptomatic COVID–19 [22].

The heptad repeat regions (HR) provides for an attractive target in therapeutic drug discovery, due to its significance in membrane fusion. While the RBD is highly mutable and not ideal for broad–spectrum inhibition, the HR and the mode of interaction between HR1 and HR2 is highly conserved among human coronaviruses (HCoVs). Jiang et al. studied several potent anti–spike peptide inhibitors derived from the HR1 and HR2 α-helix regions of SARS-CoV-2, severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV) and HIV-1 spikes [23–25], implying antiviral efficiency of peptides derived from conserved regions of viral proteins. For instance, their group reported of a 5-helix construct based on the S2 6–HB structure without one HR2

region able to inhibit various SARS–CoV–2 variants [26]. Moreover, trimers of S2 HR1 polypeptides stabilized by conjugation to a foldon sequence (HR1MFd) have been shown to exhibit broad coronavirus inihibitory activity [27]. Nevertheless, despite their remarkable potency and low toxicity, their use as antiviral agents might be hampered by a poor pharmacokinetic and pharmacodynamic profile.

Unlike the S1 domain the S2 is covered and locked in the pre–fusion conformation by three S1 subunits. As a consequence, very few neutralizing antibodies are produced against S2 during infection [28]. Even so, S2–targeting nAbs would be more interesting because it is more conserved than S1. Hence, the S2 domain could serve as an attractive template for future vaccine formulation studies to obtain broadly neutralizing antibodies.

1.4 In–house Designed Mimetic Proteins

Francisco Conejero Lara's group (i.e. the group in which this study was carried out) previously designed artificial chimeric proteins based on the 6HB in the post–fusion structure lacking the HR2 helices [29]. This resulted in highly stable antiparallel trimers of HR1 helices, named L3A, L3B and L3C. Moreover, it was shown that these mimetic proteins can bind to the HR2 region with high affinity, both as a synthetic peptide (i.e. V39E) and in a recombinant trimeric spike protein, showing inhibitory activity towards SARS–CoV–2 cell infection. Figure 4 shows the crystal structure of L3B bound to V39E. V39E includes the sequence from 1164 to 1202 of the spike protein and is thus used to represent the HR2 region. Among the investigated mimetic proteins L3C was revealed to be the most stable showing the best inhibitory activity and antigenic profile. Hence, L3C will be used in this work as an HR1 mimic. Given the structural similarity and the similar binding affinity for V39E between L3B and L3C the same binding mode as shown in figure 4 can be assumed for L3C. The protein–protein interaction between L3B and V39E was briefly analyzed with Chimera [30]. By coloring the surface of L3B according to the hydrophobicity of the residues on the Kyte–Doolittle hydrophobicity scale [31], it becomes clear that the main driving forces of the interaction between these protein are of this nature. The most structured part of V39E binds on the N–terminal region of L3B in a major hydrophobic groove. Moreover, analysis with the web–based tool ANCHOR [32] revealed that the residues that contribute the most to the stabilization of the binding lie in this region as well. A very small alpha–helix of V39E binds to minor hydrophobic region on the C–terminal region. These findings resonate well with a previous paper [33], where smaller mimetic proteins each of which in composed of about a half of L3C were designed (i.e. N2C and C2C; see figure 4) to better understand the interaction between HR1 and HR2. It resulted that the N-terminal half of HR1 harbors the main HR2 binding determinants, but it has a lower conformational stability, whereas the C-terminal half is more stable but does not bind HR2 in isolation. Nevertheless, it co-

operates significantly with the N subdomain, contributing to the overall HR1-–HR2 binding energy.

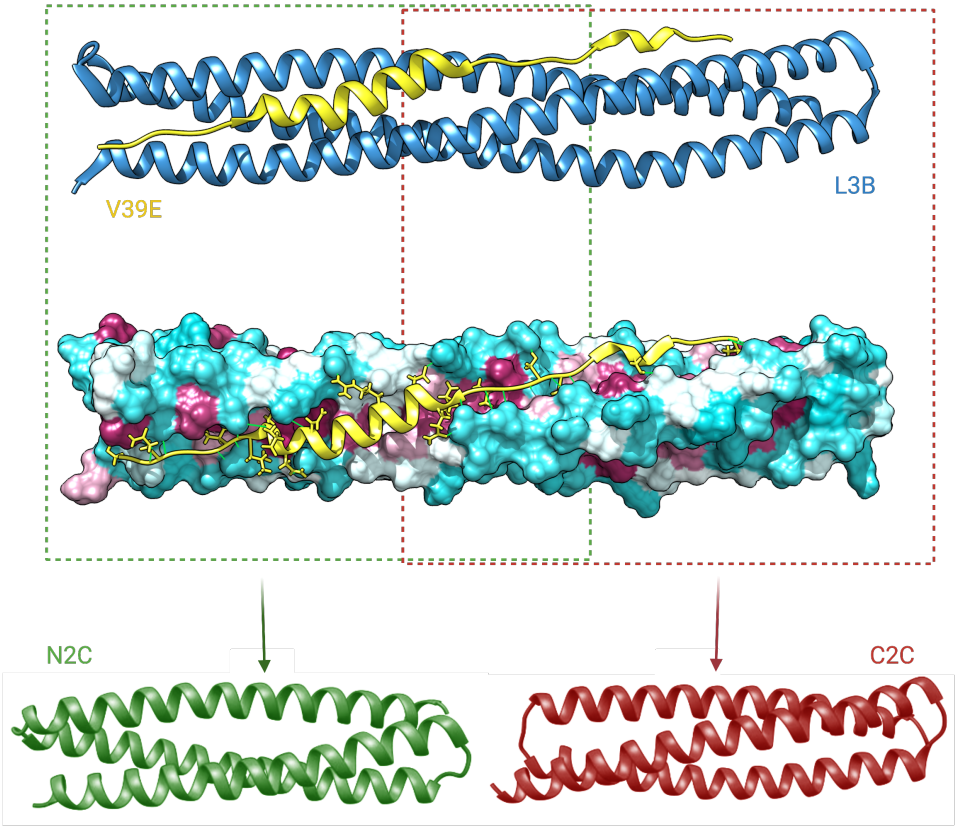

Figure 4: ***Above:*** *Crystal structure of L3B (blue) bound to V39E (yellow) (PDB ID: 7ZR2).* ***Center:*** *Molecular surface of L3B colored according to the hydrophobicity of residues on the Kyte–Doolittle scale, with cyan being hydrophilic and magenta hydrophobic. Interfacial residues of V39E building H–bonds with L3B are depicted as sticks and the H–bonds are shown in green. The green and the red dashed boxes indicate the corresponding regions on L3C named N2C and C2C respectively, which are represented* ***below*** *as 3D–structures. The images have been rendered in Chimera.*

1.5 Nanobodies

The exploration of cross-reactive antibodies targeting SARS–CoV–2 S2 protein represents a potential path for the development of pan–sarbecovirus therapeutics. Despite, having proved as a valuable asset during the COVID–19 pandemic, the choice of mAbs for treatments of lung infections is not flawless. They call for exceedingly high intravenous doses of a few grams [34] and these relatively big macromolecules only partially cross the plasma-pulmonary barrier when treating lung infections [35]. In addition, mAbs are inefficient in a pandemic situation where the pressure of new emerging variants of concern necessitates rapid development of medications easy to optimize, due to their inability of being manufactured

swiftly and cheaply. All of these elements might restrict the use of mAbs in therapeutic settings.

In the last decades, nanobodies (NBs) have been actively exploited for diagnostics and therapy owing to their remarkable characteristics that favor their use over conventional antibodies [36]. NBs are variable domains of heavy-chain-only antibodies (VHHs) derived by the immunization of camelid animals. Although having a molecular mass of 12-15 kDa, which is one-tenth of a conventional antibody, they successfully retain the capacity to bind to various antigens. Additionally, NBs with fluorescence, affinity, or epitope tags can be easily made in recombinant form in bacterial, yeast, plant, and mammalian cell lines without compromising their stability or affinity. Recent advances in NB engineering and diverse applications in biotechnology as potent therapeutic and diagnostic agents have been significantly influenced by the durability and adaptability of NBs in biophysical and therapeutic contexts.

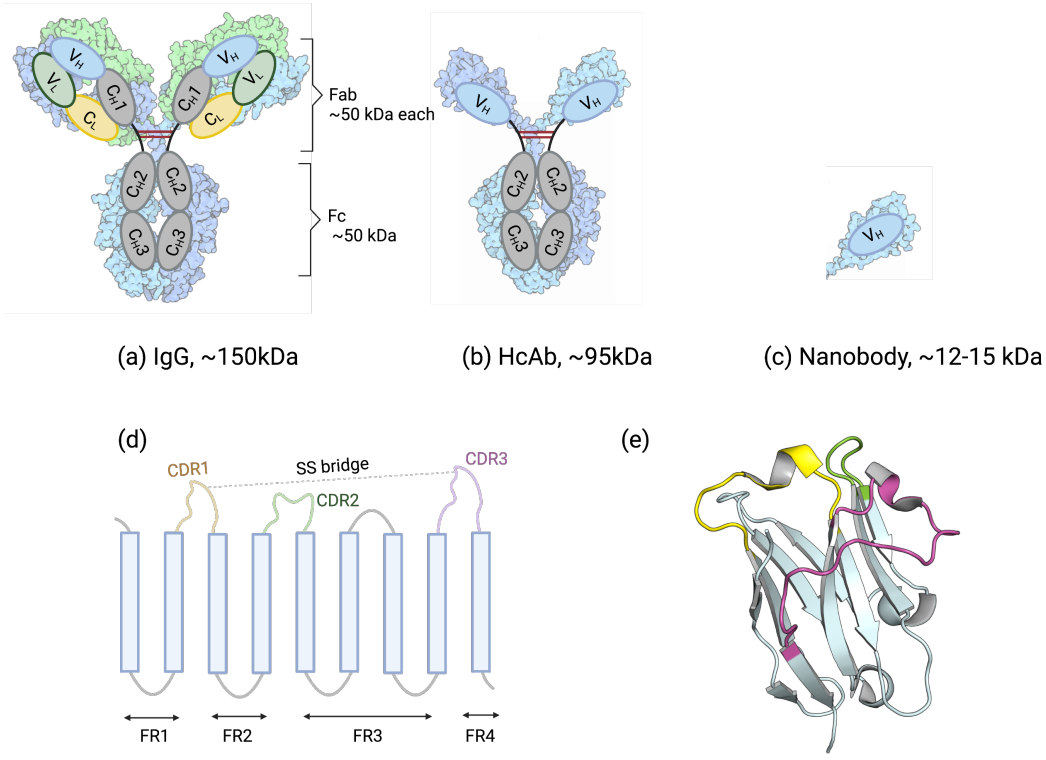

Figure 5: ***(a)-(c)*** *Comparison between the structures and molecular weights of an IgG, a camelid heavy chain antibody and a VHH, also known as nanobody.* ***(d)*** *Schematic structure of a nanobody displaying the framework (FR) regions, the complementarity determining loops (CDR) and the disulfide bridge between CDR1 and CDR3;* ***(e)*** *3D structure of a Nb with the CDR loops highlighted with the same color scheme as* ***(d)****. (Made with BioRender)*

From Antibodies to Nanobodies. Before talking about the general structure of a NB the next few paragraphs will be dedicated to describing the structure of a common immunoglobulin, in order to be able to better appreciate the differences between the two, which are highlighted as well in figure 5. As a model for the description of antibodies the isoform IgG will be used, since it is the one used for all therapeutic antibodies approved for marketing to date [37].

Antibodies are Y–shaped glycoproteins that consist of four polypeptide sub–units – two identical heavy chains (HC) and two identical light chains (LC). The HC of an IgG is comprised of four domains (VH, CH1, CH2, CH3), with a flexible hinge region connecting CH1 and CH2. The LC is comprised by two sub–domains: the VL and the CL. Within this basic H2L2 structure antibodies have two major sub–domains: the crystallyzable Fragment (Fc) and the antigen-binding Fragment (FAb). The Fc forms the tail of the Y. The hinge region connects the Fc and the FAbs granting both rotational flexibility as well as independent orientation of the FAbs with respect to each other and the Fc.

The two FAb regions form the arms of the Y and are comprised of four domains: the heavy chain variable domain (VH) linked to constant domain 1 (CH1), and the light chain variable domain (VL) linked to a constant domain (CL). Each variable region is composed by three hypervariable loops and four framework regions. The hypervariable loops are disordered in structure and characterized by a high ratio of amino acid variation. They directly contact the antigen's surface and are therefore also called complementarity determining regions (CDRs). On the other hand the framework regions have a more conserved amino acid sequence and separate the CDRs. They form a beta–sheet structure, which orients the six CDR loops closely together to generate the antigen binding site. The smallest portion to retain the antigen binding ability on its own is called variable fragment (Fv). It consists of the VL and VH sub–domains connected either by a flexible peptide linker or by a novel disulfide bond.

Some other interesting antibody formats are present in nature as part of the immune system of different animals. For instance, camelids (e.g., lamas and alpacas) possess a second type of γ–antibodies besides IgG that make up as much as 50 % of the camelid IgG-like repertoire in which each of paired chains are comprised by a V_{HH} domain linked through a hinge to two constant domains – C_H2 and C_H3 [38, 39]. In these heavy chain only antibodies (hcAb) the C_H1, C_L and V_L domains of common IgG's are absent. Pairing of two of these chains forms a normal Fc and two antigen-binding VHH domains with a molecular weight of about 95 kDa. Single domain antibodies or nanobodies made up by only the VHHs alone have a molecular weight of 12-15 kDa and fully retain their antigen binding activity without the Fc making them the smallest occurring intact antigen binding fragments [40].

Nanobody Structure. X-Ray structural studies in 1996 elucidated the camelid VHH fragment of having a relatively conserved Ig–like fold, with four framework regions, three

CDRs and a conserved intra–domain disulfide bridge [41]. The schematic and 3D strcture of a NB can be observed in figure 5 (d) and (e) respectively. Structurally they are composed of 2 β–sheets: one being made up by 4 antiparallel β–strands and the other one by 5 antiparallel β–strands. Nevertheless, a few characteristics seem to be specific to the camelid VHH structure. First, four conserved residues in the framework 2 region – Phe37, Glu44, Arg45 and Gly47 – known as the "VHH tetrad" contribute in making the surface naturally more hydrophilic [42]. The VHH tetrad lies in a position that in an Fv would make contact with the VL domain. Additionally, these four residues are believed to give camelid derived NBs the ability to reversibly unfold when heated – a two state reversible process that is unique among antibody structures. This results in their exceptional solubility and stability [43].

The CDR3 loop is longer than the CDR–H3 loop of typical mAbs, measuring between 10 and 24 residues with an average of 17. As compared to typical human IgGs, the increased length may help the VHH CDR3 to protrude into deep depressions on a protein's surface, such as the active site of an enzyme or buried clefts of receptors, opening up a wider range of potential targets. Additionally, the lengthy CDR3 loop helps the VHH domain antibody maintain its solubility and stability by partially burying residues that would typically be in touch with VL in a Fv [44].

To stabilize the long CDR3 loop occasionally an additional disulfide bridge may form between Cys residues of the CDR3 loop and either the CDR1 loop or in the framework region between CDR1 and CDR2. This inter–loop disulfide bridge is unique to camelid derived VHHs [42]. Finally, it appears that the conformations of the CDR1 and CDR2 hypervariable loops of the VHH domains frequently deviate from their conventional canonical forms [45].

1.6 Nanobodies as SARS–CoV–2 Inhbitors

Desirable Properties over Conventional mAbs. The unique structure of NBs offer multiple advantages over conventional mAbs. Their small size, convex shape and extended CDR3 loop confer them the ability to bind to antigen sections believed to be obstructed [46], exhibiting equilibrium dissociation constants within the nano– and picomolar range [40]. Additionally, a more precise and effective drug distribution is made possible by their reduced size, which also enhances imaging capabilities [47]. NBs are excellent candidates for multimerization because of their smaller size and monomeric structure, which allows for multivalency, multiparatopicity, and multispecificity, which can each boost avidity and permit binding of several antigens [48].

NBs also have a variety of advantageous biochemical traits in addition to the benefits that their tiny size has already bestowed upon them, such as i) high stability when exposed to higher temperatures [43], ii) low aggregation rate due to the hydrophilic FR2 region and

iii) low susceptibility to proteolytic cleavage and pH changes [49]. Combined these traits may enable NBs to be administered by alternative routes such as oral or intraperitoneal [46]. It's interesting to note that numerous studies have demonstrated that NBs are either non–immunogenic or have a low immunogenicity [40]. This is due to thier high sequence identity with the VH of human IgGs. To further mitigate the risk of adverse side reactions it is possible to 'humanize' the nanobody on the condition that the VHH tetrad is preserved [50]. Lastly, the simplicity to modify, produce and purify nanobodies [48] represents an advantage from an economic point of view as well. Both traditional prokaryotic expression systems like Escherichia coli and eukaryotic systems like Saccharomyces cerevisae and Pichia pastoris are capable of producing them [51–53]. These combined benefits make it possible to produce NBs for diagnostics and therapy at a comparable economic cost while offering a potentially wider range of antigen recognition than their traditional mAb competitors.

Anti SARS–Cov–2 Nanobodies. A wide array of nanobodies interacting with the S1 subunit of the S protein are already available. Recent investigation of the structural biology of NBs in regard to their binding with the S protein led to the realization that NB combinations can make use of different binding epitopes presented on the diverse conformational landscape of the S protein to acquire neutralizing and mutation resistant characteristics. Full–length mAbs would often have to deal with steric hindrance when interacting with more epitopes close in space. On the other hand, nanobodies are enabled by their small size and thermal stability to exploit different binding modes on previously unappreciated epitopes [46]. Thus, some of the most effective antiviral medicines to date include highly selective neutralizing NBs and their multivalent forms [54, 55]. Furthermore, in this significant research [56] the preclinical effectiveness and efficiency of utilizing an ultrapotent aerosolizable NB for inhalation treatment of SARS–CoV–2 infections in a susceptible COVID–19 model were shown.

Neutralizing NB can inhibit SARS–Cov–2 in a variety of ways. A study [57] presents a trivalent NB (mNb6–tri) able to lock all three RBDs of the S protein in the 'down' conformation. Another work [58] studies two NBs (Nb20 and Nb21) able to competitively inhibit spike–ACE2 interaction by binding to two distinct non–overlapping epitopes. According to structural studies of cryo–EM NB–bound structures [59], spike–targeting nanobodies can be clustered in three groups based on the targeted epitope: Class I targets ACE2–binding sites and disrupts host receptor binding; Class II binds highly conserved epitopes and retains activity against VOCs; Class III recognizes unique epitopes that are likely inaccessible to antibodies. While Class I and II NBs classically inhibit cell entry by sterically competing with ACE2 binding, members of Class III display very diverse action mechanism. For instance, NB17 [59] can lock all RBDs in the 'up' conformation, inducing premature S1 cleavage resulting in loss of function of a spike by cellular protease activity.

2 Hypothesis and Objectives

While the employed NBs described in section 1.6 against SARS–CoV–2 are very promising, it might be crucial to target epitopes outside of S1 subunit as well. The S2 subunit is more conserved and plays a very important role in the cell entry of the virus. Findings described in section 1.3 indicate that the different S2 domains are temporarily exposed during infection and thus target–able. L3C is clearly able to bind to HR2 and to interfere with the HR1–HR2 interaction in the virus that leads to the cascade of refolding events necessary for membrane fusion. Moreover, L3C has been shown to capture antibody responses in sera from COVID-19 convalescent patients, which indicates that it mimics S2 epitopes that are immunologically relevant during the course of the disease [29]. By this rationale, an NB that binds to L3C with high affinity, should display HR1 binding characteristics on the virus and be able to inhibit cell entry by blocking the cascade of refolding events. In this work, NBs obtained by immunizing alpacas with L3C are biophysically tested in regards to their ability to bind HR1 and displace HR1–HR2 interaction.

This study will focus on three NBs extracted from a pool of eight obtained by immunizing alpacas with L3C and successive selection through bio–panning. The objectives of this work are to assure that the selected NBs can be produced and purified by the standard methods with good yield and subsequently characterize them to assess if they are stable, behave as expected and can be stored under selected conditions. The second part of this work revolves around the biophysical investigation of the binding of the NBs to the HR1 region of SARS-CoV–2. This is done by using protein and peptide mimetics, among them L3C, V39E, N2C and C2C, which have been briefly described in section 1.3. The main questions to be answered are i) do the NBs bind L3C, and thus HR1, and do they do so with enough strength to displace the binding with V39E and ii) to which region of HR1 do the NBs bind to?. These questions will be tried to be answered by employing the methods of differential scanning calorimetry and isothermal titration calorimetry.

3 Materials and Methods

3.1 Investigated Nanobodies

This section is dedicated to describe the analyzed NBs and to give insights about their characteristics by briefly analyzing their amino acid sequence. The NBs are N–terminally tagged with a signal peptide that translocates them to the periplasmic region and C–terminally with a $(HIS)_6$ tag to aid with the first purification step. In this work three NBs from the pool were analyzed: NB113, NB118 and NB248. Since no crystal structure for the NBs is available, for visualization purposes computational models displayed in figure 6 have been built with the SWISS–MODEL software [60]. Identification of the CDR loops was possible thanks to comparison with other published works that focused on NB structure and sequence data analysis [61].

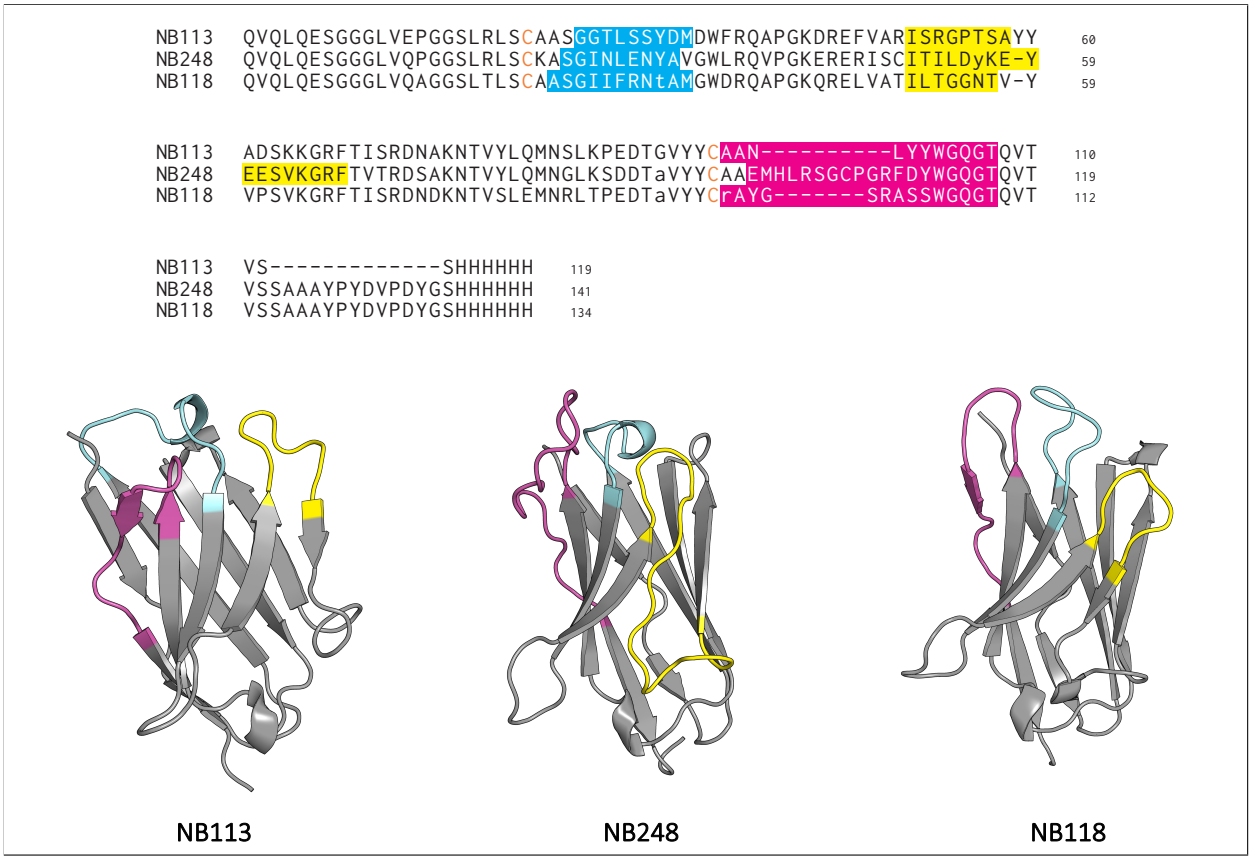

Figure 6: *Sequence comparison and Computational Model of the nanobodies with the CDR1 in cyan, the CDR2 in yellow and the CDR3 in magenta. (Images rendered in PyMol)*

Table 1 shows a comparison of the main parameters of interest of the two NBs obtained by submitting the amino acid sequence to the ProtParam tool of Expasy [62].
Sequence analysis with FELLS [63] and PsiPred [64] reveals the presence of an Immunoglobulin V-set domain, as expected. Moreover, both sequences are predicted to contain about

Table 1: *Molecular characteristics of the nanobodies obtained by analyzing amino acid sequence with the ProtParam tool of Expasy.*

	NB113	**NB248**	**NB118**
Nr. of amino acids	119	141	134
Molecular weight (g mol^{-1})	13158.52	15698.47	14441.97
Theoretical pI	8.61	6.71	8.01
Ext. coefficient	22920	25900	21430

40 % β–sheet structure and almost no α–helices. As expected, both NBs are predicted by the Expasy PROSITE tool to form only one disulfide bridge (NB113: CYS22–CYS95; NB248: CYS22–CYS96; NB118: CYS22–CYS95). Interestingly, neither of the NBs displays the above described conserved VHH tetrad. Nevertheless, polar and/or charged amino acids are still present at those positions, which should guarantee a good solubility with a negative tendency to aggregation.

3.2 Cell Trasformation

The lyophilized plasmids containing the gene for the proteins and an amipicillin resistance (pET303 for L3C; pHEN6 for nanobodies) were resuspended in 20 µL of filtered and sterilized H_2O. *E. coli* BL21 (DE3) cells were thawed on ice. 1 µL of the DNA solution was added to the cells, which was then left on ice for 5 min. The tubes were successively heated for 30 s at 42° C in a water bath and placed again on ice for 5 min. 800 µL of Super Optimal Catabolite (SOC) were added and the tubes agitated for 45 minutes at 37° C. In the meantime, petri dishes with the agar to plate the bacteria were prepared. Once, the 45 min had passed 200 µL of the sample were plated with the spread plate method to obtain the "diluted" colonies. The remaining sample was centrifuged for 5 min at 10000 rpm, 200 µL of the supernatant collected and the remaining one discarded. The pellet was resuspended in the collected supernatant and 200 µL plated as described before to obtain the "concentrated" colonies. The petri dishes were finally incubated at 37° C overnight. One of the grown colonies was taken with a micropipette tip and put in 5 ml LB with 0.5 mM AMP to obtain a pre-inoculum that was incubated overnight at 38°C. 500 ml of the pre-inoculum were added to 500 ml of 66 % glycerol to store and freeze the transformed cells.

3.3 Protein Purification

With a micropipette tip a piece of ice was scraped of the frozen cells in glycerol and put in 10 ml LB (10 g L^{-1} Tryptone, 5 g L^{-1} Yeast Extract, 10 g L^{-1} NaCl) media supplemented with 0.1 M Glucose, 0.1 mM $MgCl_2$ and 0.1 mg mL^{-1} Ampicillin to let the cells grow overnight. 5 ml of the culture obtained by the pre-inoculum were added in 500 ml of TB media (12

g L^{-1} Tryptone, 2.3 g L^{-1} KH_2PO_4, 12.5 g L^{-1} K_2HPO_4, 24 g L^{-1} Yeast Extract, 2.5 mL L^{-1} Glycerol) with 5 mM glucose, 0.5 mM AMP and 1 mM $MgCl_2$ to perform the scale-up. The flasks were agitated at 37° C and protein expression induced when the OD reached approximately 0.7 for L3C and 1.3 for the nanobodies at 600 nm.

The negative and the positive control samples were taken before and after inducing with IPTG respectively. The cells were then cultured overnight at 28°C. The cells were then harvested by centrifugation at 6000 rpm for 10 min and the supernatant discarded. The obtained pellet was resuspended in 20 ml lysis buffer (Tris 50 mM, NaCl 500 nM, EDTA 1 mM, β–mercaoptoethanol 1 mM, pH 7.5), a protease inhibitor added and lysed using 6 cycles of ultrasonication of 1 min at 50% duty cycles. The supernatant containing the proteins was divided into Eppendorf tubes and centrifuged at 14000 rpm for 30 min.

The proteins were purified by a first step of immobilized metal affinity chromatography (AC) with a Ni–NTA column using an FPLC system. The proteins were loaded onto the column with buffer A_{AC} (Tris 50 mM, NaCl 500 mM,pH 7.5), the loaded column washed with 5% Buffer B_{AC} (Tris 50 mM, NaCl 500 mM, 500 mM Imidazole, pH 7.5) and eluted with a 5 ml gradient with an end concentration of 20% Buffer B_{AC}. A polyacrylamide gel electrophoresis in presence of SDS (SDS-PAGE) was run to check which fractions contained the proteins (20 µL sample + 20 µL Laemmli buffer). The appropriate fractions were pooled and put into a membrane to perform dialysis in 3 L of Tris Buffer (50 mM, pH 7) for 4-5 h followed by a longer round in 3 L of approximately 14 h.

The proteins were then recovered from the dialysis bag, collected in a falcon tube and centrifuged for about 20 min at 4000 rpm. Subsequently, the proteins were purified by a second step of cation exchange chromatography (CEX). The column was loaded with Buffer A_{IC} (50 mM Tris, 25 mM NaCl) which was previously filtered and degasified. And the proteins eluted with 30% buffer B_{IC} (50 mM Tris, 1 M NaCl) with a gradient of 50 ml. The fractions were collected and those ones identified to contain enough protein by SDS–PAGE analysis pooled into a dyalisis bag to perform again 2 rounds (4–5 h in 3 L and 12–14 h in 2 L) of dialysis in a sodium phosphate (NaP) buffer (50mM Na_2HPO_4/chNaH2PO4, pH 7.4).

Finally, the proteins were transferred into a falcon tube and centrifuged for about 20 min at 4000 rpm and the supernatant collected to measure the concentration by UV absorbance measurement. If in an acceptable range the protein solution can be aliquoted in cryo–tubes, frozen in N_2 and stored at -80°C for further analysis. If necessary, the protein solution was concentrated with centrifuge ultrafilters or diluted to achieve a final concentration in the range of 100 µM. Then, it was frozen in liquid N_2 and stored at -80º C for further analysis.

3.4 Dynamic Light Scattering – DLS

DLS is a technique to measure the size of particles in the sub µm region. DLS measures Brownian motion and relates this to the size of the particles. Brownian motion is the random movement of particles suspended in a medium, due to molecular collisions. The larger the particle, the slower the Brownian motion will be. The velocity of the Brownian motion is defined by a property known as the translational diffusion coefficient D, from which the size of a particle can be calculated using the Stokes–Einstein equation:

$$r_H = \frac{kT}{6\pi\eta D} \tag{1}$$

, where r_H is the hydrodyanamic radius, k is the Boltzmann constant, η the viscosity and T the absolute temperature. The radius obtained by DLS refers to the radius of a sphere that diffuses within a liquid medium equivalently to the actual particle and is hence referred to as the hydrodynamic radius. Effectively, the instrument measures the rate at which the intensity of the scattered light fluctuates. The rate at which these intensity fluctuations occur depend on the size of the particles with small particles giving rise to fluctuations more rapidly. The fluctuations in light intensity measured over time are quantified via a second order correlation function, from which the translational diffusion coefficient can be obtained by applying a fitting algorithm.
A Malvern Zetasizer µV instrument was used to measure the apparent hydrodynamic radii of the proteins at 25° C. Frozen samples were thawed to be put into dialysis in NaP buffer and once finished the concentration was adjusted to 15 µM and measured in a cuvette of 1.5 mm path length. instrument's software was used for data collection and processing. DLS sets were recorded at 25°C with a total number of 30 acquisitions per run, with each acquisition lasting 10 s. The most appropriate measurements were then chosen to build an average and obtain the size volume distribution.

3.5 Circular Dichroism – CD

CD is a spectroscopic technique used to gain information about protein conformation. Especially, far–UV CD spectra (170–250 nm) yield information on the amount and the type of secondary structure a protein contains. CD relies on the differential absorption of left and right circularly polarized components of plane-polarized radiation by chromophores which either possess intrinsic chirality or are placed in chiral environments. Linear polarized light can be obtained by the sum of two circularly polarized components E_R and E_L. The resulting vector will oscillate along a vertical line, when E_R and E_L rotate in opposite directions. If either one of E_R or E_L is absorbed to different extents the resulting radiation would be elliptically polarized, i.e. the resulting vector would trace out an ellipse. If linearly polarized

light passes through a chiral compound the resultant radiation is elliptically polarized. The difference in absorption of the two components is converted into ellipticity θ following this equation:

$$\theta = tan^{-1}\left(\frac{b}{a}\right) \tag{2}$$

, where b and a are the minor and major axes of the resultant ellipse respectively. A CD spectrum measures the variation of θ as a function of the wavelength, which can be related to the secondary structure because the electronic transitions the radiation causes are influenced by the geometries of the polypeptide backbones.

CD measurements were performed with a Jasco J-715 spectropolarimeter (Tokyo, Japan) equipped with a temperature–controlled cell holder. For the measurement frozen samples of the protein were put in dialysis with a NaP buffer the day prior to measurement. After recollecting the samples from the dialysis membrane, the concentration was adjusted to 20 µM. Far-UV CD spectra (200 – 260 nm) were recorded in a quartz cuvette with a path length of 1 mm with following parameters: a scan rate of 100 nm/ min, 1 nm step resolution, 1 s response and 1 nm bandwidth. The resulting spectra were the average of 5 scans. Spectra were corrected by subtracting the baseline obtained by measuring the buffer only and finally the CD signal was normalized to ellipticity per residue number ($[\theta] = \text{deg cm}^2\,\text{dmol}^{-1}\,\text{res}^{-1}$). Percentage of secondary structure was calculated with the K2D3 software [65].

3.6 Lyophilization Test

To assess if the produced proteins are stable enough to withstand lyophilization a lyophylization test was performed. For this purpose, the samples were put into dialysis with pure water at pH 2.5. Subsequently, two identical samples with the adequate concentration for the characterization methods (being DLS and CD) were prepared. One of the samples was then lyophilized. The lyophilized product was the re–suspended with NaP buffer and the two samples analyzed identically with the previously described methods and the results compared.

3.7 Differential Scanning Calorimetry – DSC

DSC is a thermoanalytical technique which measures the difference in the heat capacity (C_p) of a sample and a reference as a function of temperature. The heat capacity is defined as the amount of heat needed to change the temperature of a given sample by 1°C. Peaks in heat capacity changes allow the detection of transitions, such as heat–induced denaturation of a protein. In this case the peak would be of endothermic nature. Thermodynamic parameters such as the enthalpy of unfolding can be deduced from this measurement by

following equation:

$$\Delta H_U = \int_{T_0}^{T_1} \Delta C_p dT \tag{3}$$

, where T_0 and T_1 indicate the temperatures where the endothermic peak begins and ends respectively. The thermal transition midpoint is reached at the maximum height of the peak and is defined as the melting temperature T_m, where 50% of the protein is unfolded. The higher this value, the more thermally stable the protein.

DSC measurements to investigate the thermal stability of the proteins were performed with a VP–DSC microcalorimeter (Microcal Inc., Northampton, MA). Scans were run from 10 °C to 120 °C with a scan rate of 90 °C min^{-1}. Protein concentration was typically 20 µM. The thermograms measured with the samples were corrected by subtracting the baselines obtained by measuring the buffer in both cells. Reheating runs were carried out to determine the thermal reversibility of the denaturation process, with the same parameters used for the main scan. For every NB a scan with only the NB plus multiple scans in complex with other proteins (L3C, V39E, C2C, N2C) were performed. For the analysis and the fitting of the thermograms of the NBs alone a two state unfolding fitting model [66] is used, which is described in detail in Appendix A.

3.8 Isothermal Titration Calorimetry – ITC

ITC is a physical technique that directly measures the heat exchange involved in a bimolecular reaction and allows to obtain a variety of thermodynamic parameters of the reaction. The instrument consists of two cells kept at the same temperature and pressure: one is the main cell for the investigated macromolecule, the other one is the reference cell with the solvent. Once the ligand is titrated into the main cell, macromolecular binding results in either heat discharge or consumption causing a temperature change in the main cell. The instrument will adjust the power supply to keep the temperature of the main cell equal to the reference cell. By integrating the power over time the enthalpy of the reaction can be calculated. The heat discharged or consumed all along the calorimetric reaction is proportional to the fraction of bound ligand and increased ligand concentration leads to saturation of substrate and finally less heat is discharged or consumed. Amount of exchanged heat Q upon ligand inclusion is defined by following equation:

$$Q = V_0 \Delta H_b [M]_t \left(\frac{K_b[L]}{1 + K_b[L]} \right) \tag{4}$$

, where V_0 is the sample cell volume, ΔH_b is the enthalpy of binding, $[M]_t$ is the total concentration of the macromolecule, K_b is the binding constant and $[L]$ is the concentration of free ligand.

Samples of the NBs to be investigated and of L3C protein were thawed and put into membranes to perform two rounds of dialysis in 50 mM phosphate buffer (pH 7.4). Then the proteins were centrifuged and either concentrated or diluted to reach the desired concentrations stated below.
ITC measurements were carried out in a Microcal VP-ITC calorimeter (Malvern Panalytical, Malvern, UK). Protein concentration in the measuring cell (2 ml) was about 10 µM, whereas ligand concentration in the syringe (500 µL) was about 200 µM. The titration was carried out reversely, where L3C was added to the NB in the cell. 30 injections of 5 µL were made at an interval of 480 s. After integration of the peaks and normalization, the obtained binding isotherms were fitted using a binding model of n independent sites, making it possible to determine of the binding constant K_b, the binding enthalpy ΔH_b, and the binding stoichiometry n. The standard Gibbs energy and entropy of binding were calculated according to following equation:

$$\Delta G = -RT \cdot lnK_b = \Delta H_b - T \cdot \Delta S_b. \tag{5}$$

4 Results

4.1 Purification

NB113 expression was induced in 1 L of media at an OD of 0.8 at 600 nm. The affinity chromatogram in figure 7 (a) shows two peaks, a thinner one and a wider one. Moreover, a small peak appeared when eluting with 100% B. After the AC purification elutions #5–#9 and #12 were collected, dyalized and then re-purified with a CEX chromatography. The chromatograms in figure 7 (b) shows some noise at the end of the peak, which is due to some air bubbles. SDS–PAGE confirmed that the obtained fractions after CEX were quite free of impurities. Elutions #1–#3 were collected and stored to obtain 5.6 mg of NB113.

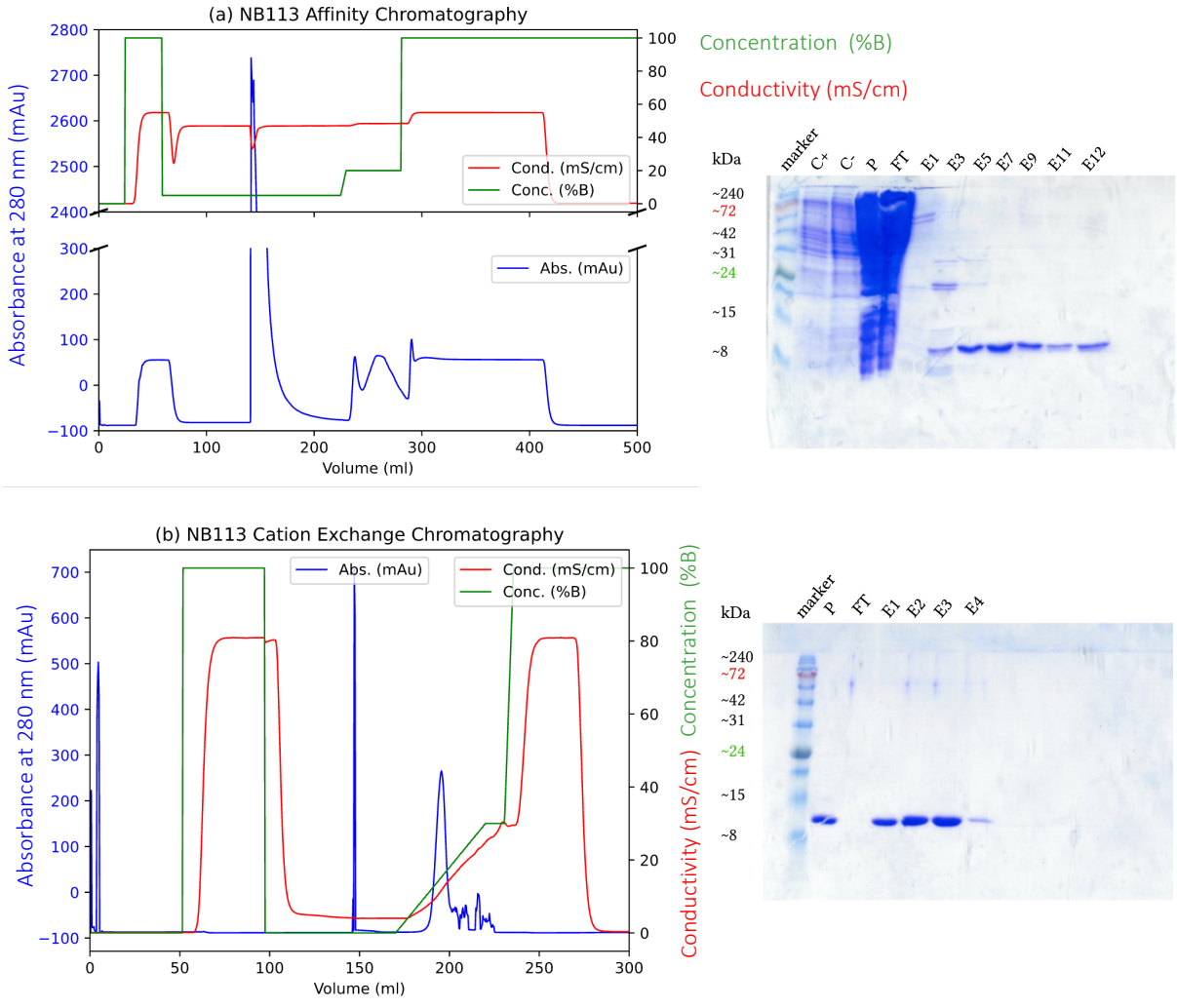

Figure 7: *Affinity* ***(a)*** *and cation exchange* ***(b)*** *chromatograms of NB113 accompanied by the relevant SDS–PAGE analysis.*

NB248 was purified two times, with only the second more successful purification being described here. NB248 expression was induced in 1 L of media at an OD of 1.3. The AC step in figure 8 went quite smoothly, although a very low amount of protein was recovered. Since SDS–PAGE revealed that the eluted fractions were relatively free of impurities, to avoid losing protein the CEX step was not performed. Indeed, the eluted fractions were dialyzed directly in the NaP buffer to freeze and store them. Although, fraction #2 showed a quite good amount of protein at a molecular weight of around 12 kDa, it presented high levels of impurities and was hence not included. Fractions #3 – #6 were collected to obtain 2.7 mg of NB248.

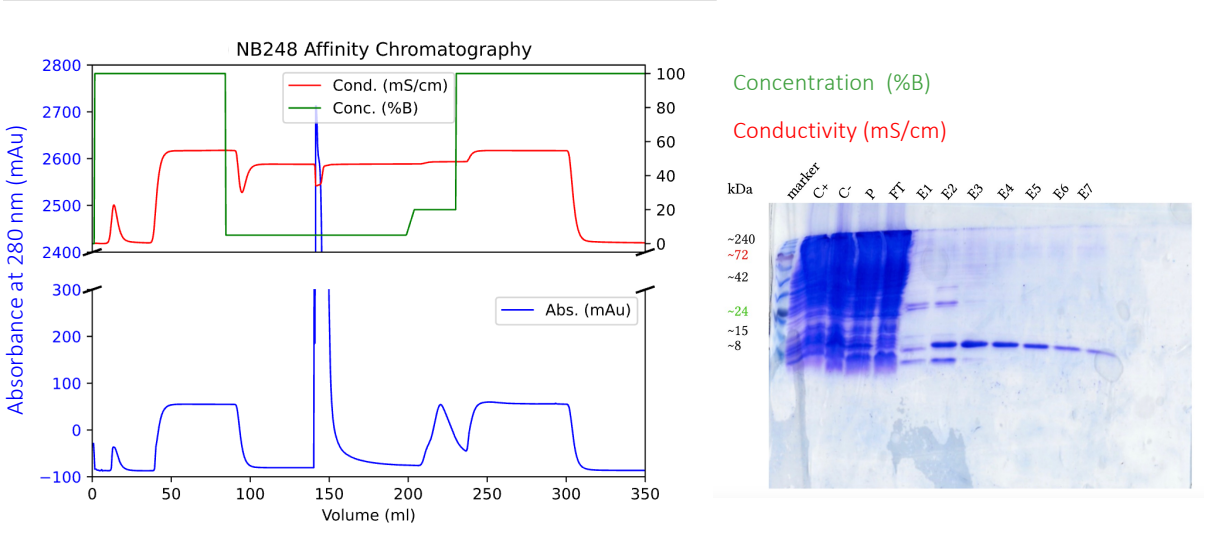

Figure 8: *Affinity chromatogram of NB248 accompanied by the relevant SDS–PAGE analysis.*

NB118 was purified two times as well, with the second purification having been more successful and thus being described here. NB118 expression was induced in 1 L of media at an OD of 1.6. The elution in the AC step (figure 9 (a)) occurred with two major peaks, with both containing the protein of interest as confirmed by SDS–PAGE. The second peak was eluted by increasing the concentration of B to 70%. Ultimately, a good amount of NB118 was recovered and elutions #4–#8 were collected, dialyzed and re–purified with CEX (figure 9 (b)). All three eluted fractions contained a good amount of NB and were free of impurities. Hence, they were collected to obtain 12.1 mg of NB118.

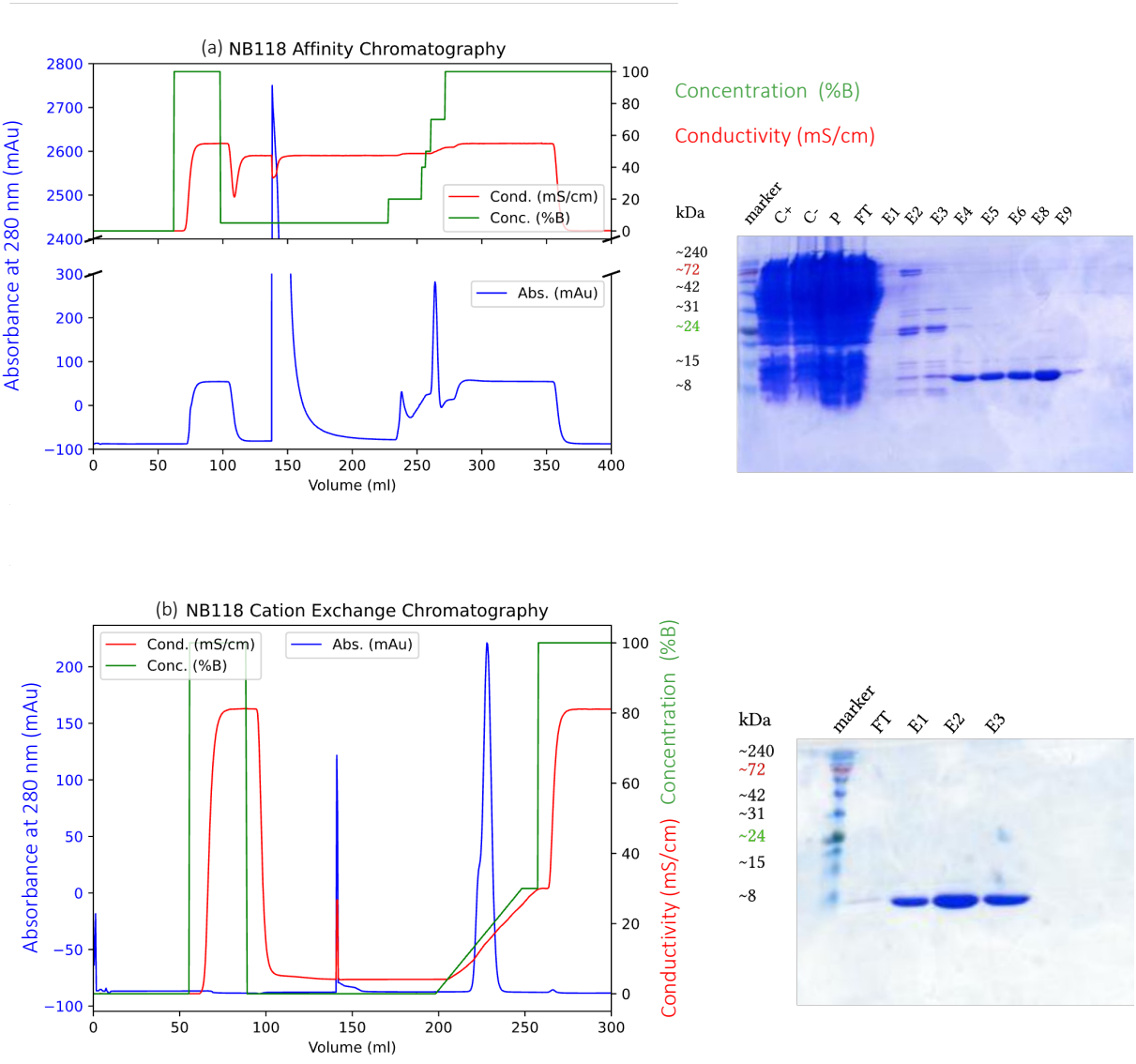

Figure 9: *Affinity* ***(a)*** *and cation exchange* ***(b)*** *chromatograms of NB118 accompanied by the relevant SDS–PAGE analysis.*

4.2 Initial Characterization

After successful production and purification of the nanobodies, first pre–characterization experiments were performed to assess if they behave as expected. These included DLS and CD measurements plus two DSC runs: one of the nanobody alone and a reheating run with the same parameters. The obtained results are depicted graphically in figure 10 and 11 and interesting values about the structure and stability of the proteins are listed in table 2.

DLS measurements returned hydrodynamic radii of about 2 nm for all NBs, which is in line with what is typically expected for this type of proteins in monomeric form. CD

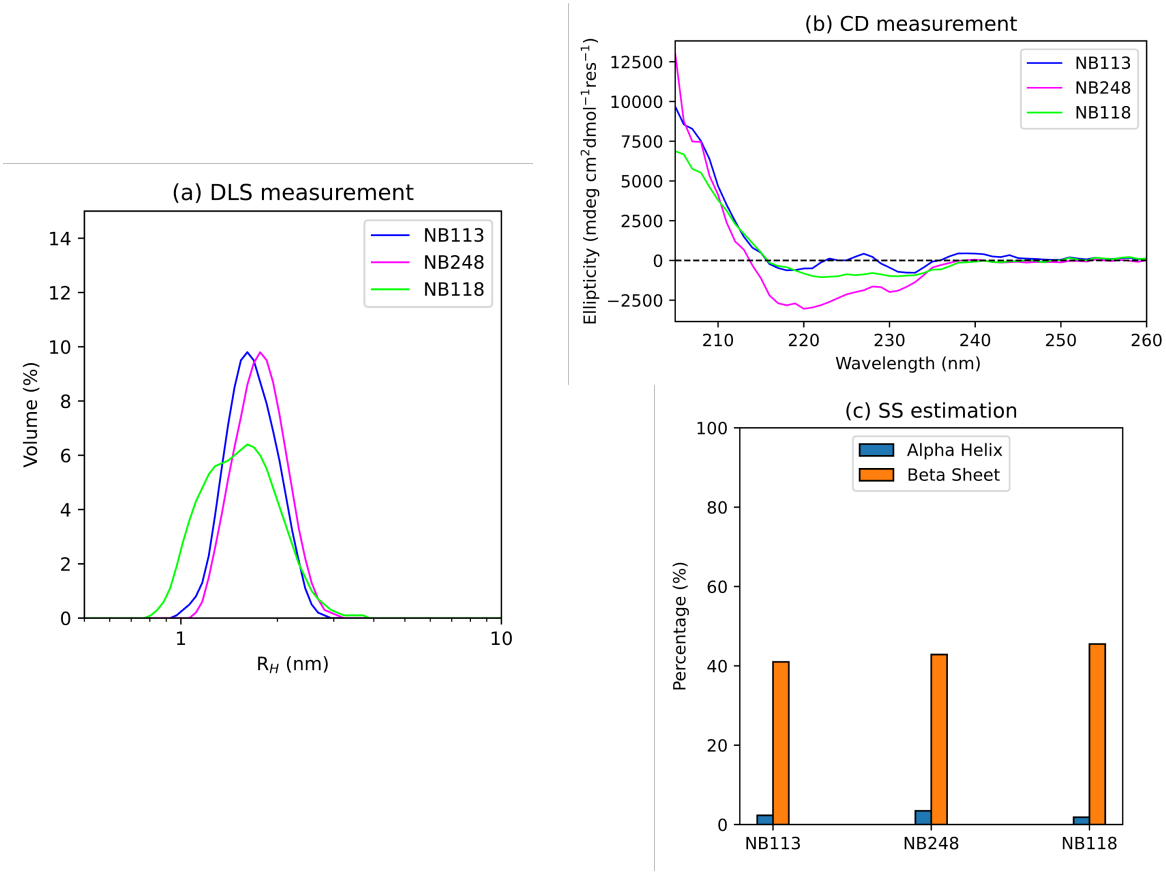

Figure 10: *Initial characterization experiments of the NBs which included* ***(a)*** *DLS measurements here depicted as a distribution of the volume contribution along the hydrodynamic radius;* ***(b)*** *CD measurements from 205–260 nm to infer information about the secondary structure calculated with the K2D3 software in* ***(c)*** *by using the CD spectra as input.*

measurements were only carried out from 260 to 205 nm because at the concentrations used the detector would saturate too much below 205 and result in a distorted spectrum. The CD signal for NB113 and NB118 are drastically lower than that of NB248 even though the estimated secondary structure amount is very similar. This can be due to the fact that measurement for NB113 and NB118 were performed with a lower concentration, which could have led to some imprecision. Analysis of the CD spectrum revealed the presence of about 40 % β–sheet structure for all NBs, which corresponds to the previous prediction done with the amino acid sequence alone. The presence of this backbone secondary structure can be observed in the spectra thanks to the peak at around 215-220 nm. Interestingly, the CD spectra show a peak around 230 nm, which is probably due to the presence of the aromatic residues Phe, Tyr and Trp which are known to contribute to the CD spectra in the presence of disulfide bridges [49]. It must be noted that in helical proteins this non–peptidic contribution is masked by the strongly negative signal of the α–helices. On the other hand, the positive CD band near 205 nm and the negative band at around 215–220 nm are most

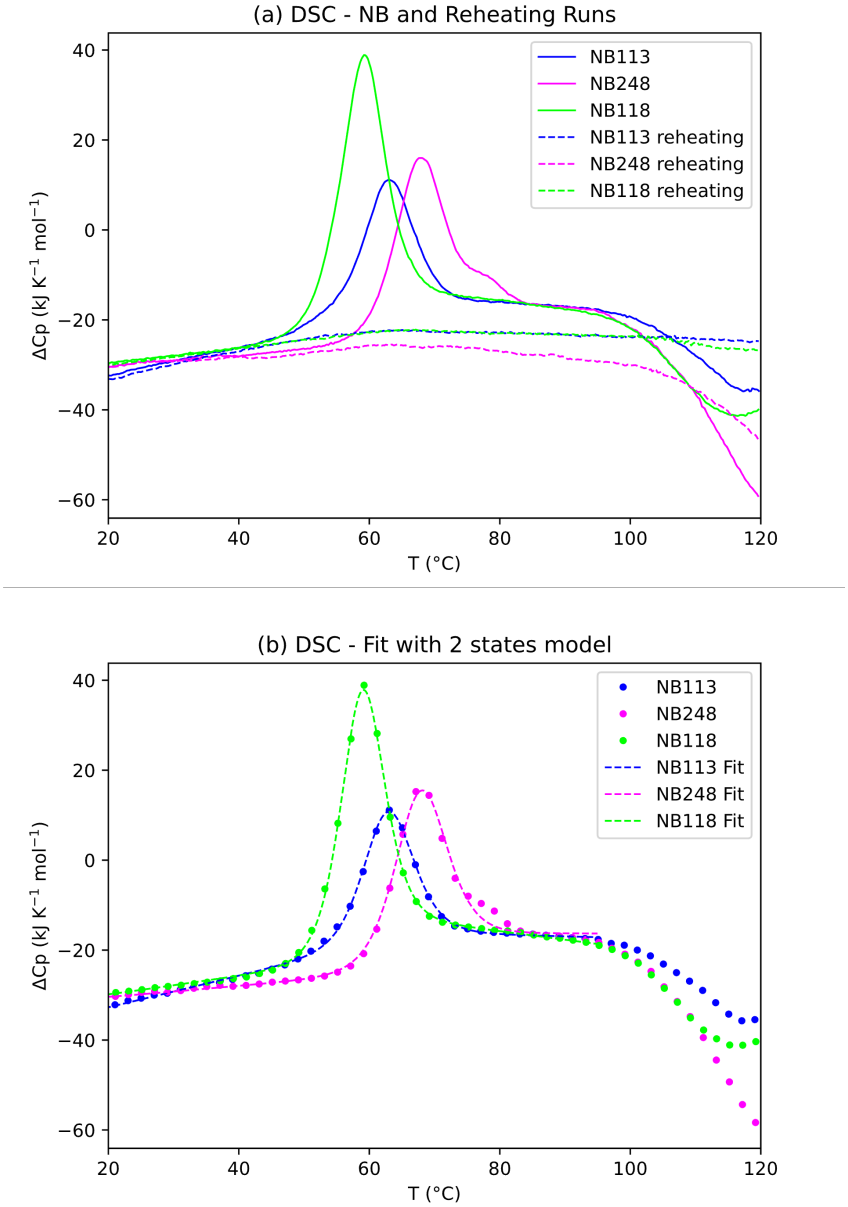

Figure 11: *Initial characterization experiment including* ***(a)*** *DSC scan of the NBs from 20° C to 120° C (solid line) and a reheating run with the same parameters as the first one (dashed line). In* ***(b)*** *the raw experimental data is represented as dots and the fits obtained with the 2 state unfolding model as dashed lines. The curves were only fitted until 90° C, to avoid influence of the SS bond rupture on the results.*

likely due to the peptide backbone, i.e., to the antiparallel β-sheet structure. Ultimately, it can be said that according to CD and DLS measurements the NBs are folded in monomeric form.

The DSC scans of the NBs revealed that they undergo a two–state thermal unfolding , with T_m values in the range of 58 to 68° C, which is in line with results in the scientific literature

Table 2: *Structural and stabilty properties of NB113, NB248 and NB118. The values for T_m and ΔH_m were obtained through the fit of the experimental thermograms with the 2 state unfolding model until 90° C.*

	r_H (nm)	α–Helix (%)	β–sheet (%)	T_m (°C)	ΔH_m (kJ mol^{-1})
NB113	1.9	2	41	62.6	249
NB248	2.0	3	43	67.5	376
NB118	1.9	2	46	58.6	410

for NBs. T_m values indicate that NB118 is the least thermally stable of the three, while ΔH_m values suggest that NB113 may be the one that builds the lowest amount of intramolecular interactions. Interestingly, the reheating runs show that the thermal denaturation in this temperature range is not reversible, even though NBs are well–known to possess this characteristic. This is due to the fact that the disulfide bridges cointained by the NBs break around 80 – 90 °C, indicated by the drop in ΔC_p after 100 °C. To observe reversibility the first measurement would have needed to be carried out only upon 80 °C, or an oxidizing agent added to reform the broken bridge.

4.3 Freeze and Lyophilization Test

To determine if the Nbs could be stored under frozen and lyophilized conditions without altering their structure, DLS and CD spectra were recorded for all proteins after (i) freeze and thawing them and (ii) after lyophilization and resuspension in the same volume of NaP–buffer. The CD spectra were normalized with the concentrations the NB samples had before lyophilizing them.

NB113 exhibits a reduction in hydrodynamic radius visible in graph 12 (a) after both lyophilization and freezing. Although no bigger aggregates of concern are visible in the Volume–r_H distribution, they cannot be excluded, because their volume contribution might be so small (<0.001 %) that they are not displayed in the size–r_H distribution. Another possibility is that the bigger contribution of aggregates shifts the autocorrelation function used to fit the light–scattering data, making the NBs look smaller. Indeed, when measuring the concentration of the lyophilized and re–suspended NB113 quite high levels of light scattering were observed even after centrifugation, indicating some aggregation of the protein. Hence, it cannot be excluded that the peak at around 1 nm belongs to some other unidentified peptide or impurity. On the other hand, light sacttering was not visible with the frozen and thawed sample. Contrarily, DLS measurements of NB248 (figure 12 (b)) show a good correspondence between the unaltered, the frozen and the lyophilized protein. DLS measurements of NB118 could only be made for the unaltered and the frozen sample, since with the lyophilized product no steady count rate could be reached, which may be due to the presence of large aggregates.

CD spectra for all NBs (figure 12 (c), (d) and (e)) reveal that although the secondary structure is maintained, there is a big reduction in signal mostly in the spectra of the lyophilized protein. This may be due to the loss of protein during the process and hence a reduction in concentration.

According to these results, it can be said, that NB248 is a much better candidate for being stored in lyophilized form than NB113 and NB118, for which the frozen storage form may be preferable.

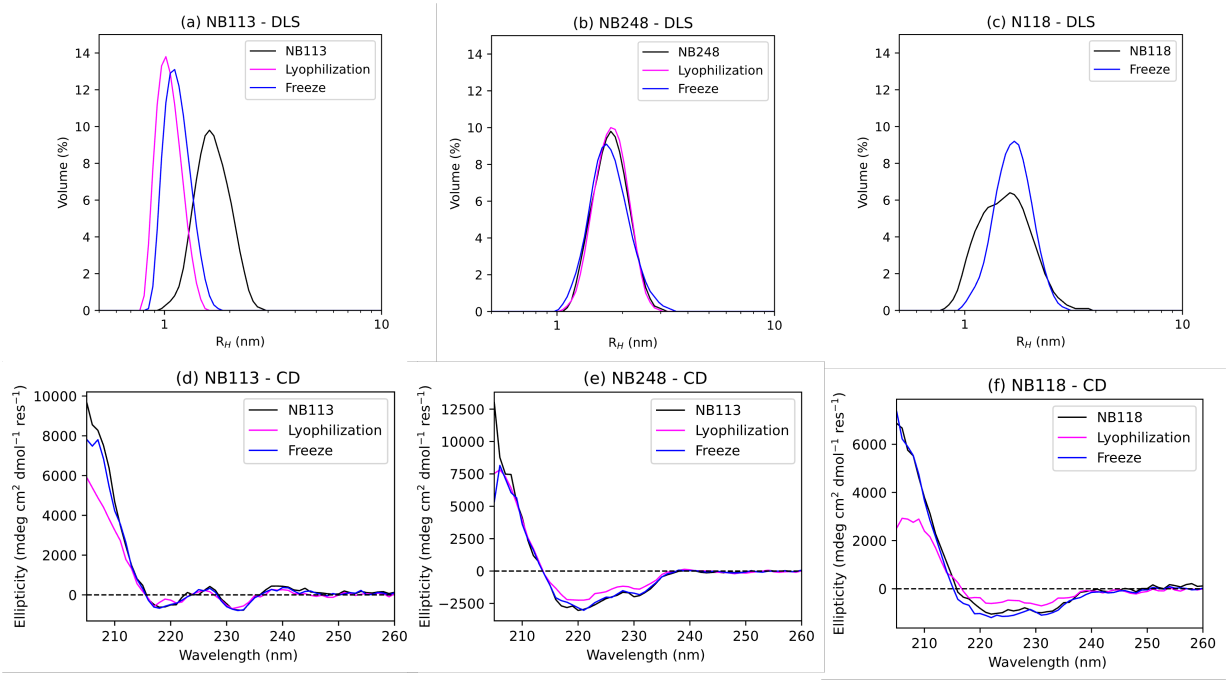

Figure 12: *Freeze and lyophilization tests of all NBs, which include comparison of DLS and CD measurements done for the unaltered protein (black), the frozen and thawed protein (blue) and the lyophilized and re–suspended protein (magenta). DLS measurements are represented as a Volume–r_H distributions. CD measurements were carried out from 205 to 206 nm.*

4.4 Investigating NB and L3C Interaction

In order to study the extent of the interaction between the NBs and L3C both DSC and ITC were employed. DSC was used to measure by how much L3C stabilizes the NBs, whereas ITC to measure the affinity of the NBs for L3C. Given that L3C unfolding occurs at a T_m of 104° C, binding of NBs to L3C should produce a stabilization of the NBs visible by a increase of T_m and the enthalpy of unfolding Δ H_m. Areas under the unfolding peaks (i.e. Δ H_m) were calculated through integration with Origin. Relevant figures are displayed in

Appendix B. DSC and ITC results should allow the ranking of NBs based on L3C binding strength.

DSC measurements with the NBs were carried out with an excess of L3C in a 1:2 (NB:L3C) molar ratio. The raw experimental DSC thermograms are displayed in figure 13 and the inferred thermodynamic parameters listed in table 3. In presence of L3C, all NBs experience a strong increase of melting temperature compared to the NBs alone. NB113 and NB118 display a comparable shift in T_m, but the change in ΔH_m is much larger for NB113, pointing to a stronger interaction with L3C. Notably, NB113 had the smallest enthalpy of unfolding on its own. NB248 experiences the smallest T_m shift, but the change in ΔH_m ranks second in front the value for NB118. Interestingly, according to these results no clear correlation between change in melting temperature and change in enthalpy of unfolding exists. $\Delta\Delta H_m$ values contain both (i) the contribution from the interactions built between the NBs and L3C and (ii) the contribution from the internal NB structure stabilization L3C binding produces. Hence, it is not possible to deduce which NB may form the most interactions with L3C.

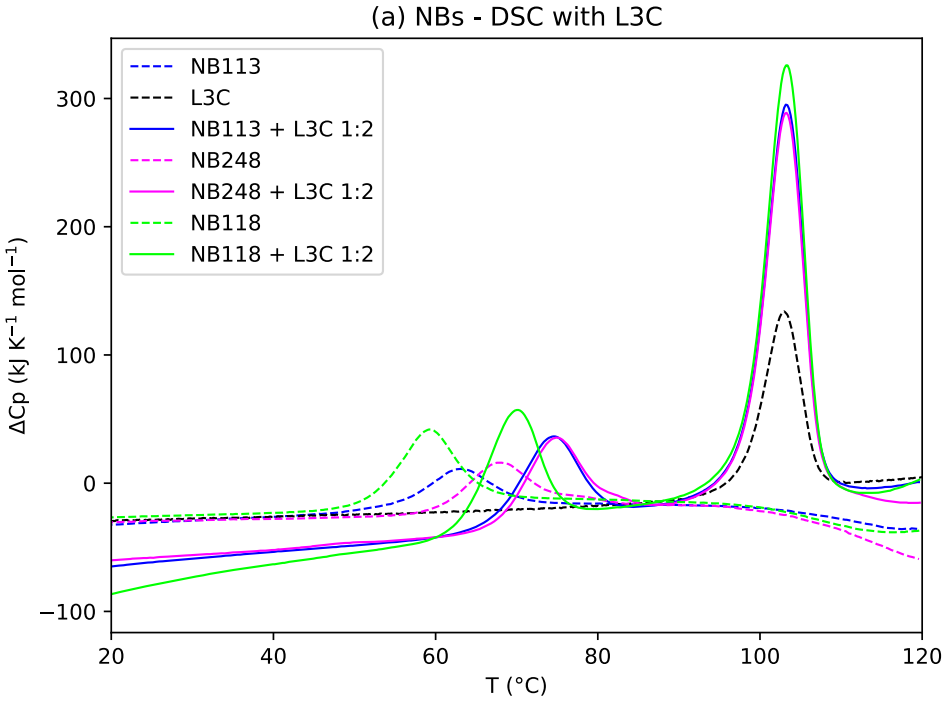

Figure 13: *Comparison between the thermograms of NB113 (blue), NB248 (magenta) and NB118 (green) in presence of an excess of L3C at a molar ratio of 1:2. The dashed lines represent the thermograms of the NBs alone.*

Figure 13 (a) shows the ITC plots obtained by titrating the NBs in the sample cell with L3C from the syringe. Notably, the first injections produce larger changes in heat with respect to later injections, due to the protein binding sites becoming saturated as the experiment

Table 3: *Changes of thermodynamic parameters of unfolding of the NBs after addition of L3C to the Nbs. ΔH_m values were calculated by integrating the peaks of the thermograms with Origin.*

	Tm (° C)	T$_m$(+L3C) (° C)	ΔTm (° C)	ΔH$_m$ (kJ mol^{-1})	ΔH$_m$ (+L3C) (kJ mol^{-1})	$\Delta\Delta$H$_m$ (kJ mol^{-1})
NB113	62.6	74.2	11.6	246	522	276
NB248	67.5	74.8	7.3	398	577	179
NB118	58.6	69.3	10.7	402	470	68

proceeds. The thermodynamic parameters of the binding between the Nbs and L3C are listed in table 4 and visually represented in figure 14 (b). As expected, the measurements follow a sigmoidal binding isotherm with 1:1 stochiometry. DSC experiments of NB118 could not be carried out due to time issues, but comparable results to NB113 and NB248 are expected.

Measured binding enthalpies were negative indicating an exothermic reaction. A negative change of entropy is common to all binding events due to the relative translational–rotational immonilization. Here, it might be mostly ascribable to the CDR loops of the NBs acquiring a less disordered structure upon binding, since L3C already possesses a high intrinsic stability and thus will have a lower contribution to the entropic penalty. Nevertheless, at the temperature at which the experiment was carried out, i.e. 25° C, the Gibbs energy assumed negative values indicating a spontaneous reaction. Moreover, the Gibbs free energy keeps a negative value over the whole temperature range in which the NBs are not denatured. Given the much more negative enthalpy values, it can be said, that the binding process between the NBs and L3C is enthalpy–driven. NB113 exhibits a lower dissociation constant than NB248, implicating a stronger interaction with L3C, which is also confirmed by a bigger thermal stabilization observed in the DSC experiments.

Table 4: *Thermodynamic parameters of binding between the Nbs and L3C measured by ITC.*

	n	K$_b$ (10^7 M^{-1})	K$_d$ (nM)	ΔH$_b$ (kJ mol^{-1})	TΔS$_b$ (kJ mol^{-1})	ΔG$_b$ (kJ/mol)
NB113	0.96 ± 0.01	1.06 ± 0.19	94.34 ± 16.82	-48.20 ± 0.73	-8.12	-40.08 ± 0.73
NB248	1.08 ± 0.01	0.88 ± 0.20	114.16 ± 25.80	-52.93 ± 1.04	-13.35	-39.58 ± 1.04
NB118	/ ± /	/ ± /	/ ± /	/ ± /	/	/ ± /

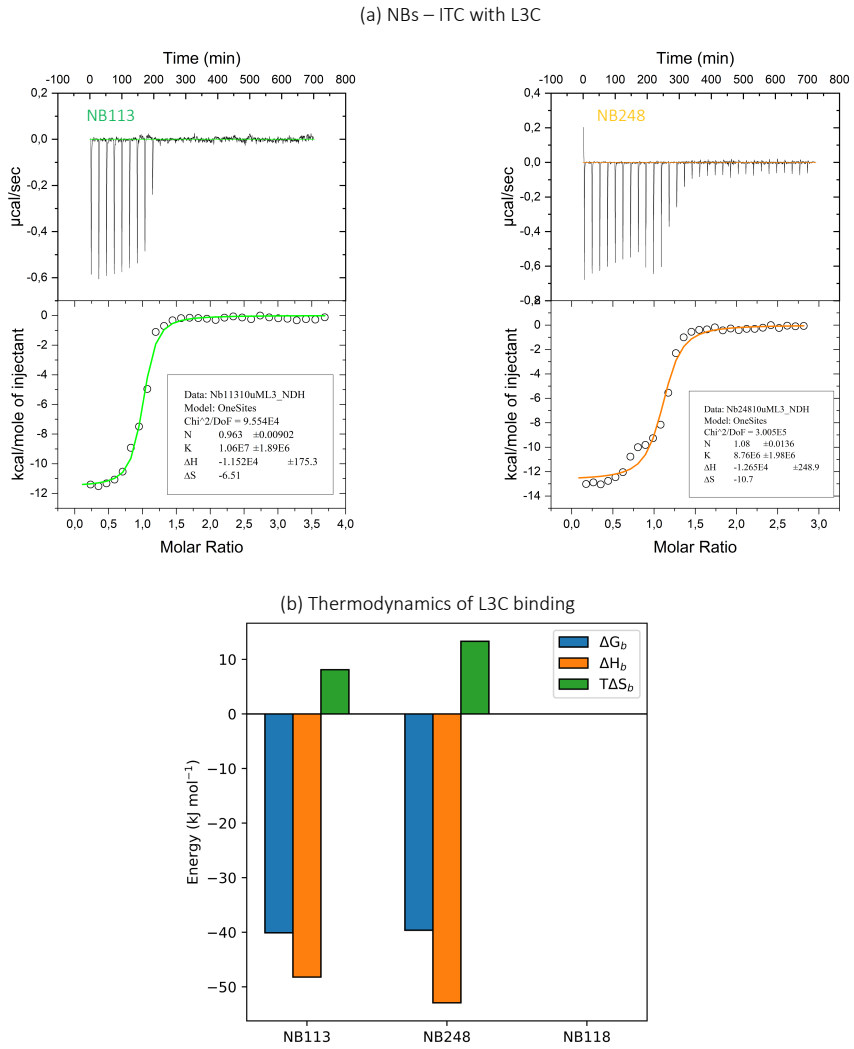

Figure 14: ***(a)*** *ITC experiments where the NBs have been titrated with L3C. The upper panels show the experimental thermogram corrected with the baseline. The lower panels show the heats normalized per mole of injected ligand. The green (NB113) and orange (NB248) lines correspond to the fit of the data with a model of n independent binding sites, done with Origin for ITC.* ***(b)*** *Thermodynamic parameters of the NBs binding to L3C.*

4.5 Assaying Binding Displacement with HR2

To assay whether the nanobodies are able to interfere with the interaction between HR1 and HR2, we carried out DSC scans with the NBs mixed with L3C and V39E. The samples were prepared with a 1:2:2 molar ratio (NB:L3C:V39E) in two ways: (i) the NB and L3C were mixed and after 15–20 min V39E added and (ii) L3C and V39E were mixed and the NB added after 15–20 min. Before starting the measurement the samples were pre–heated at 37° C for 15 min to pre–equilibrate them. The experimental graphs resulting from these measurements are displayed in figure 15. To infer more significant conclusions the changes in melting temperatures and enthalpy of unfolding of the NBs in the mixtures are listed in

table 5 and represented visually in image 16. The enthalpies of unfolding were calculated by integrating the peak of unfolding with Origin. Relevant images for this procedure are displayed in the Appendix B.

First, the thermograms of the mixtures (i) and (ii) are almost identical, indicating that the order of addition does not influence the result. Clearly, the presence of V39E only slightly changes the thermograms of NB113 and NB248 in comparison the the ones with L3C. On the other hand, the changes are more substantial for NB118. NB113 experiences a small reduction in melting temperature and a small increase in enthalpy of unfolding. NB248 is the only one to exhibit an increase in T_m, whereas slightly ΔH_m decreases. NB118 shows clear reductions in both T_m and ΔH_m.

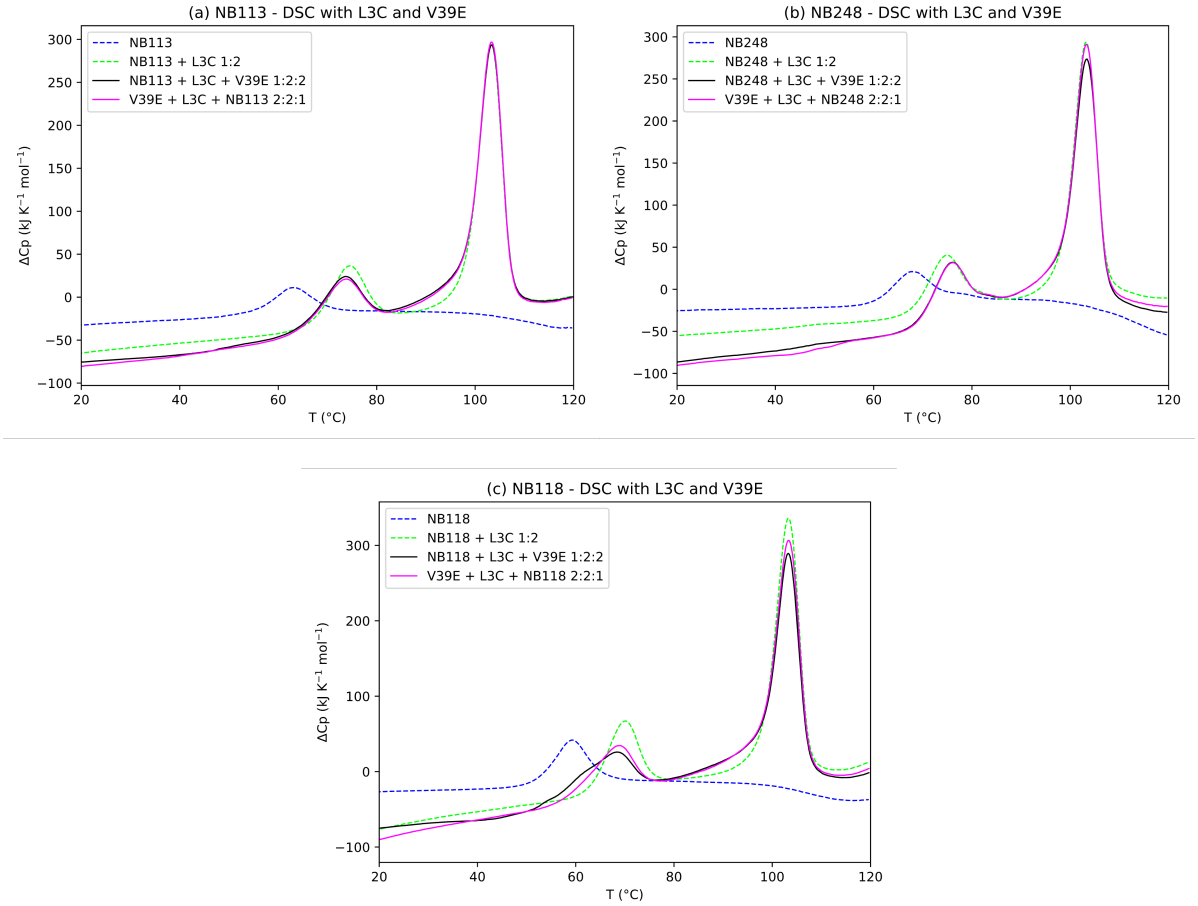

Figure 15: *Displacement experiments between the NBs and the V39E HR2 peptide for the binding to L3C carried out by DSC for* ***(a)*** *NB113,* ***(b)*** *NB248 and* ***(c)*** *NB118. The thermograms were recorded at a molar ratio of 1:2:2 (NB:L3C:V39E). The thermograms of the ternary mixtures are represented as solid lines, while the graphs of the NBs alone and of the NBs + L3C as dashed lines.*

Reduction in T_m in the presence of an excess of V39E compared to the T_m observed for the NB–L3C complex is a sign of overlapping between the epitope the NB binds to and

Table 5: *Changes in melting temperature and enthalpy of unfolding upon addition of L3C and V39E. The changes in enthalpy of unfolding were calculated by integrating the are under the unfolding peak in Origin.*

	T_m(+L3C) (° C)	T_m(+L3C+V39E) (° C)	ΔT_m (° C)	ΔH_m(+L3C) (kJ mol^{-1})	ΔH_m(+L3C+V39E) (kJ mol^{-1})	$\Delta\Delta H_m$ (kJ mol^{-1})
NB113	74.2	73.4	-0.8	522	547	25
NB248	74.8	75.9	1.1	577	540	-37
NB118	69.3	67.1	-2.2	470	403	-67

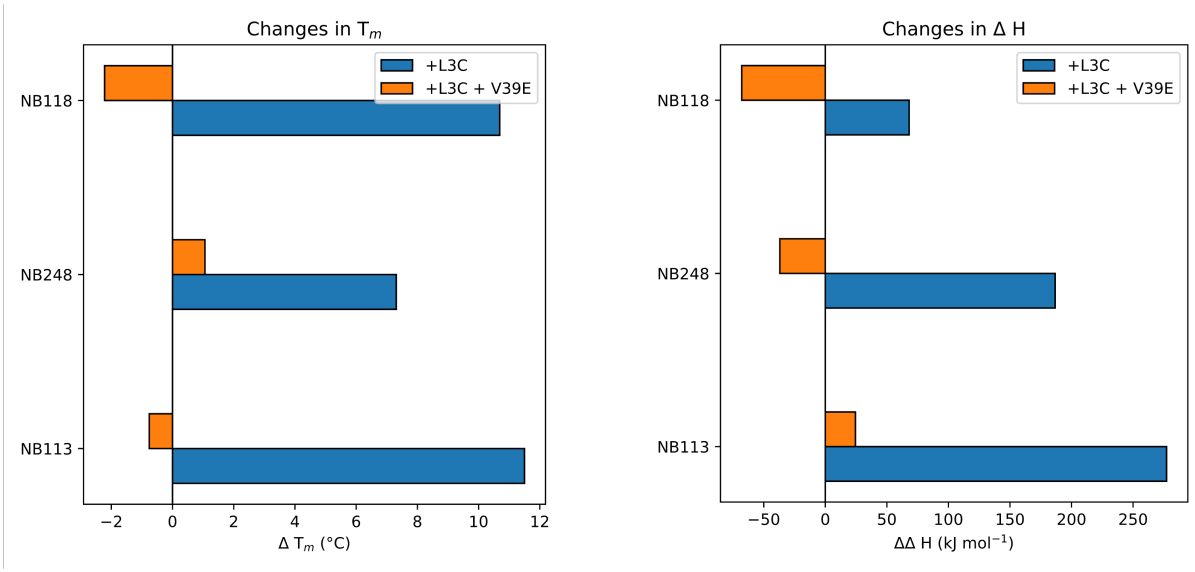

Figure 16: *Visual representation of the changes in T_m (**left**) and ΔH_m (**right**). The changes upon addition of L3C to the NB alone are depicted in blue, whereas the differences between the addition of L3C alone and the addition of L3C + V39E are shown in orange.*

the binding interface of V39E on L3C. Hence, the binding the NBs to L3C is weakened by V39E. To confirm if the NB epitopes overlap with the HR1–HR2 epitopes the changes in unfolding enthalpy must be taken in consideration. In case that considerable overlapping in present, binding of V39E should at least partially displace the Nb from L3C and therefore, the thermograms should show a reduction in the area under the peak of unfolding. This effect is driven by the compensation of the endothermic dissociation of the NB from L3C by the concomitant exothermic binding of V39E to L3C. NB118 shows both a considerable decrease in T_m and a reduction in ΔH_m after L3C and V39E addition similar in magnitude but of opposite sign compared to the changes produced by L3C alone (see figure 16). Both of these results indicate that Nb118 competes with the V39E peptide for binding at the HR2 binding groove.

In the case of NB113 we observe a slight reduction in T_m, meaning that the addition of HR2 weakens L3C binding, but the enthalpy of unfolding of the NB113–L3C complex in absence and in presence of V39E is very similar and much higher than the one of NB113 alone. This could be explained by either one of two ways. First, since NB113–L3C binding is partially displaced by V39E, it could be that the NB113 epitope only partially overlaps with the HR2 binding groove on L3C. The second possibility is that the binding of V39E allosterically alters the epitope of NB113 on L3C leading to a slightly weakened binding.
As NB113, NB248 shows a change in ΔH_m comparable to that obtained with L3C alone. Interestingly, it is the only one that exhibits an increase in T_m with V39E compared to the NB248–L3C complex, indicating that some additional interactions are established with L3C–V39E complex, probably forming a ternary complex. This means that the L3C–V39E complex presents a more stable NB248 epitope, probably due to allosteric stabilization. Hence, competition between NB248 and V39E for L3C binding can be excluded.

4.6 Determining the Binding Region

To determine in which regions of L3C the Nbs epitopes are located, we performed DSC scans of the NBs in presence of either one of N2C and C2C, which approximately represent the N– and C– terminal halves of L3C (see figure 4). The measurements were carried out with an excess of N2C or C2C at a molar ratio of 1:2 (NB:N2C or NB:C2C). As before the mixtures were heated for 15 min a 37° C to pre–equilibrate them. The raw thermograms obtained from the measurements are displayed in figure 17 (a)–(c). Figure 17 (d) visually represents the changes in melting temperature in presence of N2C or C2C compared to the NB alone. The same changes are listed in table 6.

Table 6: *Changes in melting temperature upon addition of N2C or C2C compared to the melting temperature of the NBs alone.*

	T_m(NB) (° C)	ΔT_m(+N2C) (° C)	T_m(+C2C) (° C)	ΔT_m (+N2C) (° C)	ΔT_m (+C2C) (° C)
NB113	62.6	62.9	71.2	0.3	8.6
NB248	67.5	68.8	72.5	1.3	5.0
Nb118	58.6	59.2	67.8	0.6	9.2

NB118 exhibits the highest increase in T_m in mixture with C2C. It also shows a very slight stabilization in presence of N2C which may suggest that the NB118 epitope might lie near the central region of L3C where C2C and N2C overlap. NB113 is strongly stabilized by the addition of C2C as well. The shift in T_m NB113 exhibits with N2C is not significant indicating that the epitope of this NB lies exclusively on the C–terminal half of L3C. NB248 shows the smallest stabilization in mixture with C2C although it is still significant. Nevertheless, it does not bind to the HR2–binding groove as it doesn't compete with V39E. NB248 also

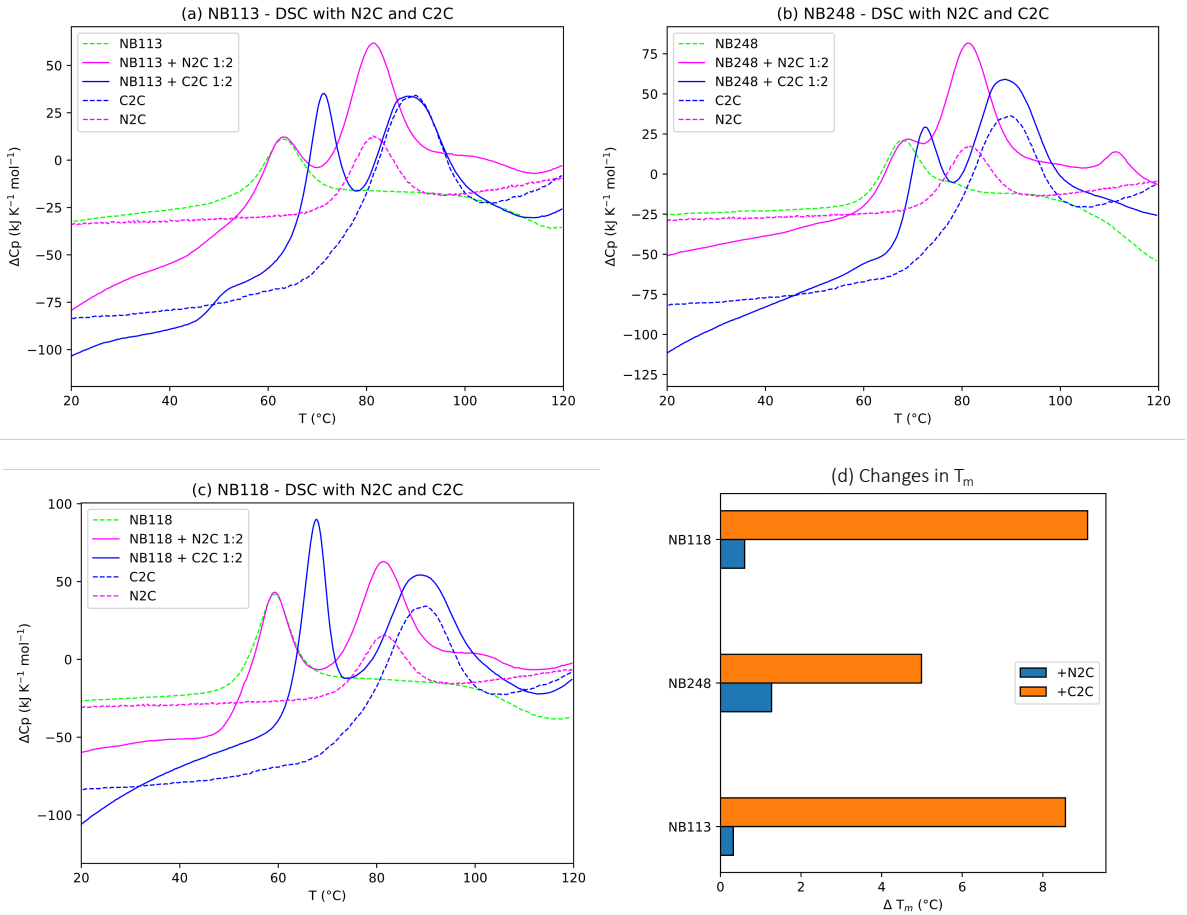

Figure 17: ***(a)–(c)****DSC scans of the NBs in presence of N2C (solid, magenta) or C2C (solid, blue). The thermograms of the NBs alone (green, dashed), N2C alone (magenta, dashed) and C2C alone (blue, dashed) are also shown.* ***(b)*** *Visual representation of the changes in T_m upon addition of N2C (blue) or C2C (orange) compared to the T_m of the NBs alone.*

displays a shift in T_m with N2C, but this could also be due to the partial overlapping of the unfolding peaks of NB248 and N2C alone.

Ultimately, these results indicate that the analyzed NBs primarily recognize the C–terminal region of L3C, probably implying that these region is more antigenic than the N–terminal region and that it produces more neutralizing heavy chain antibodies after alpaca immunization.

5 Discussion

Current SARS–CoV–2 vaccines provoke a neutralizing antibody response directed towards the RBD–subdomain of the spike protein, thus disrupting the interaction between the virus particle and the receptor on the host cell. Reports about variants of concern, described an Omicron spike evolving beyond immune evasion, thanks to a large number of mutations in the RBD [2]. Structural studies investigating SARS–CoV–2 cell entry mechanism have pointed out the crucial role S2 has in the membrane fusion step. Additionally, the heptad repeat regions HR1 and HR2 within S2 are highly conserved. During cell entry, HR1 and HR2 undergo a cascade of refolding events culminating in membrane fusion, similarly to the proposed model for membrane fusion catalysed by HIV envelope protein, in which gp120 dissociation triggers refolding of gp41 to complete the fusion process. These findings spawned the idea of using HR1–HR2 conformational intermediates as drug targets by preventing subsequent folding into the 6 helix bundle. Importantly, the prehairpin intermediate of HIV is a clinically validated target for the antiretroviral drug enfuvirtide [67], a 36-aa synthetic peptide that binds to HR1 and inhibits HIV entry. On the same note, successful attempts to design SARS–CoV–2 entry inhibitors as HR1–peptide mimetics and HR2 derived peptides have been reported by many research groups [26, 29]. A recent article [68] reports of a recombinant subunit vaccine targeting HR1 able to elicit neutralizing antibodies responses, but the idea of utilizing NBs to target the HR regions of the S2 subunit of the spike protein remains unexplored up to date.

Previous works describe the possibility of employing gp41 conformational intermediates as targets of HIV neutralizing antibodies, pointing out the difficulties in targeting the site with large molecules such as IgGs. Miller et al. [69] reported of a human IgG1 molecule, D5–IgG1, binding to an epitope overlapping the same HR1 hydrophobic pocket bound by peptidomimetic fusion inhibitors. D5–IgG1 was obtained by using peptide mimetic immunogens and subsequent selection of human mAbs that specifically target HR1 in vitro. Sabin et al. [70] presented the nAb HK20, which binds the same hydrophobic groove on HR1 as D5, given with a different contact angle raising both potency and breadth of neutralization. Moreover, they reported of the scFv fragment of HK20 having at least a 15–fold higher neutralization power, confirming the limited accessibility of the site for complete IgG1 molecules.

Given the similarities between HIV and SARS–CoV–2 in the cell entry step, by the same rationale of the papers just described, NBs targeted to S2 should experience little difficulties in accessing the HR1 hydrophobic groove given their small size. Here, NBs targeting HR1 were obtained by immunizing alpacas with an HR1–protein mimetic, L3C, and selected based on their binding to the immunogen. In this work, we characterized three NBs belonging to a larger pool obtained through this immunization process.

First of all, it was possible to produce and purify all NBs with good yield, obtaining proteins that behaved as expected, i.e. displaying (i) an immunoglobulin–like fold with majority of beta–sheet structure, (ii) a hydrodynamic radius of about 2 nm and (iii) a good thermostability. Additionally, the Nbs were tested on the ability to be stored in frozen and lyophilized condition, finding out that only NB248 is well suited for the latter one and that NB113 and NB118 should be stored frozen.

Having produced and purified the Nbs successfully our study based on the biophysical investigation of the ability of the NBs to interfere with HR1–HR2 binding commenced. Results pointed to a strong interaction between the NBs and HR1, with NB113 and NB118 being the tightest binders. According to our results only NB118 shows a clear capacity to compete and displace the HR2–derived peptide from its binding to L3C, meaning that at least to some extent, its epitope overlaps with the HR1 groove. HR2 was shown to slightly weaken NB113 binding to L3C, which can be explained either by (i) partially overlapping binding interfaces or (ii) by allosteric destabilization of the binding epitope on HR1 after HR2 binding. Finally, HR1–HR2 competition by NB248 could be excluded, since it seems to form a ternary complex with L3C and V39E.

Unfortunately all of the analyzed NBs bind to the C–terminal half of HR1, without showing significant interactions with the N–terminal part, which harbors the most energetically important part of the HR1–HR2 binding. Our results imply that the C–terminal region of L3C may be the most antigenic one. Interestingly, also D5–IgG1 was found to bind to the C–terminal region of the HR1 segment of gp41 in HIV, maintaining the ability to neutralize diverse HIV isolates. For that assessment a 5 helix mimic of gp41 was used, so it is not possible to say how well that data is translatable to L3C, which only contains the 3 HR1 helices. Finally, it remains to be seen how this data correlates with the results of SARS–CoV–2 inhibition assays, to draw definitive conclusions.

6 Conclusions and Future Perspectives

This work was inspired by the possibility of using nanobodies as SARS–CoV–2 fusion inhibitors targeting the S2 subdomain of the spike protein. Herein, an S2 based protein mimic, L3C, was used to select nanobodies generated in alpacas that bind to HR1 in vitro. We proceeded to biophysically investigate the capability of three NBs from the pool to interact the heptad repeat regions within S2. We found a strong interaction with HR1 and the data suggest competition with HR1–HR2 binding for one of the studied NBs (NB118). Moreover, there may be partial competition for NB113, although this competition needs additional experimental validation. Additional studies such as gel filtration, SPR, ELISA and X-ray crystallography could give us a better idea of the binding mode and competition. Furthermore, all NBs targeted the same subdomain, the C–terminal half, on the used immunogen, implying that that region preferentially elicits an antibody response.
For completion purposes it would be interesting to study NBs that bind to the N–terminal half of HR1, since it is known that this region plays a crucial role in HR1–HR2 binding. They could be obtained by repeating the NB selection process against the N2C miniprotein or by generating a library by immunizing alpacas with N2C directly. It is difficult to draw conclusions without having performed a proper SARS–CoV–2 inhibition assay, since it remains unclear how NB binding could affect neutralization of the virus. Theoretically, given the ability of NB118 to displace the interaction between HR1 and HR2 using model the model protein L3C, some extent of inhibitory activity should be observed in an assay with the whole virus as well, if the region bound by the NB becomes accessible at some point. Previous studies with peptidomimetic fusion inhibitors already confirmed the transient exposition of the HR1 hydrophobic groove during the cell–entry step. Hence, it remains to be seen (i) if the NB epitope is accessible in the context of the full Spike in the virion, (ii) if the NB epitope is neutralizing and (iii) if the strength of the HR1–HR2 interaction in the Spike S2 subunit is comparable to what observed in the model L3C protein.
It must be noted, that also the way SARS–CoV–2 is taken up by cells should be taken into account when trying to design neutralizing antibodies. Only if the target cells expresses enough TMPRSS2 does the hydrophobic groove become exposed on the cell surface. Should the virus be internalized via the endocytosis step, it must be guaranteed that the NBs enter the endolysosomes as well. Some ideas to do this would be either to conjugate the NBs to a molecule that anchors them to the cell membrane or to conjugate a NB targeting HR1 to a NB targeting the RBD or any other surface that is exposed during binding to ACE2. Both strategies should guarantee inclusion into the endolysosomes where the NBs should be able to reach their final target.

7 Acknowledgements

Lo primero de todo querría expresar mi profundo agradecimiento a Francisco Conjero Lara por la cálida bienvienida en su laboratorio y grupo de trabajo y además para ofrecerme la posibilidad de realizar mi Trabajo de Fin de Master aquí en la Universidad de Granada. Sus aportaciones, criticas y discusiones en las reuniones han sido fundamental para ultimar este trabajo. Además me gustaría extendir los agradicimientos a mi supervisor de thesis, Dani, que me hizo sentir inmediatamente cómodo y me guió con paciencia y experiencia, primero en inglés y despues en español también, ayudandome a acostumbrarme al idoma. Encima tengo que agradecer también a Mario que ha siempre sabido intervenir con maestría cuando aparecieron dificultades y problemas. En general me gustaría extender lo agradicimientos a todos los miembros del Departamento de Química Física que me hicieron sentir parte del equipo de inmediato.
Un grandissimo ringraziamento va anche alla Professoressa Patrizia Polverino de Laureto che ha reso possibile lo svolgimento di questo progetto Erasmus+ e mi ha fatto conoscere Francisco Conejero Lara e il suo ambito di studio. In fine ma non per importanza, ringrazio la mia compagna di laboratorio e di 'desayunos' con 'café y tostada', Sara, che rendeva meno pesanti certe giornate in laboratorio.

References

[1] Geneva: World Health Organization 2020. *WHO COVID-19 Dashboard*. Accessed: 07-04-2023. URL: https://covid19.who.int/.

[2] H. Tegally et al. "Emergence of SARS-CoV2-Omicron lineages BA.4 and BA.5 in South Africa." *Nat Med* 28 (2022), pages 1785–1790. URL: https://doi.org/10.1038/s41591-022-01911-2.

[3] K. Koelle, M. A. Martin, R. Antia, B. Lopman, and N. E. Dean. "The changing Epidemology of SARS-CoV-2". *Science* 375 (2022), pages 1116–1121. DOI: 10.1126/science.abm4915.

[4] J. S. Tregoning, K. E. Flight, S. L. Higham, Z. Wang, and B. F. Pierce. "Progress of the COVID-19 vaccine effort: viruses, vaccines and variants versus efficacy, effectiveness and escape." *Nat Rev Immunol* 21 (2021), pages 626–636. URL: https://doi.org/10.1038/s41577-021-00592-1.

[5] I. Astuti and Ysrafil. "Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response". *Diabetes and Metabolic Syndrome: Clinical Research and Reviews* 14.4 (2020), pages 407–412. DOI: https://doi.org/10.1016/j.dsx.2020.04.020.

[6] M. Hoffmann et al. "SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor." *Cell* 181 (2020), pages 271–280. DOI: 10.1016/j.cell.2020.02.052.

[7] Y. Cai, J. Zhang, T. Xiao, H. Peng, S. M. Sterling, R. M. Walsh, S. Rawson, S. Rits-Volloch, and B. Chen. "Distinct conformational states of SARS-CoV-2 spike protein". *Science* 369.6511 (2020), pages 1586–1592. DOI: 10.1126/science.abd4251.

[8] D. Wrapp, N. Wang, and K.S. Corbett KS et al. "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation." *Science* 367 (2020), pages 1260–1263. DOI: 10.1126/science.abb2507.

[9] C.B. Jackson, M. Farzan, B. Chen, and H. Choe. "Mechanisms of SARS-CoV-2 entry into cells." *Nat Rev Mol Cell Biol* 23.1 (2022), pages 3–10. DOI: doi:10.1038/s41580-021-00418-x.

[10] A. Bayati, R. Kumar, V. Francis, and P.S. McPherson. "SARS-CoV-2 infects cells after viral entry via clathrin-mediated endocytosis." *J Biol Chem* (2021), page 100306. DOI: 10.1016/j.jbc.2021.100306.

[11] S. C. Chiliveri, J. M. Louis, R. Ghirlando, and A. Bax. "Transient lipid-bound states of spike protein heptad repeats provide insights into SARS-CoV-2 membrane fusion." *Science advances* 7.41 (2021), eabk2226. URL: https://doi.org/10.1126/sciadv.abk2226.

[12] J. Zhang, Y. Cai, and T. Xiao et al. "Structural impact on SARS-CoV-2 spike protein by D614G substitution." *Science* 372 (2021), pages 525–530. DOI: 10.1126/science.abf2303.

[13] X. Fan, D. Cao, L. Kong, and X. Zhang. "Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein". *Nature Communications* 11.3618 (2020). DOI: 10.1038/s41467-020-17371-6.

[14] A.C. Walls, M.A. Tortorici, and J. Snijder et al. "Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion." *Proc Natl Acad Sci USA* 114.42 (2017), pages 11157–11162. DOI: 10.1073/pnas.1708727114.

[15] G. Li, R. Hilgenfeld, and R. Whitley et al. "Therapeutic strategies for COVID-19: progress and lessons learned." *Nat Rev Drug Discov* 22 (2023), pages 449–475. URL: https://doi.org/10.1038/s41573-023-00672-y.

[16] J. Yang, W. Wang, and Z. Chen et al. "A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity." *Nature* 586 (2020), pages 572–577. URL: https://doi.org/10.1038/s41586-020-2599-8.

[17] X. Ma, F. Zou, and F. Yu et al. "Nanoparticle Vaccines Based on the Receptor Binding Domain (RBD) and Heptad Repeat (HR) of SARS-CoV-2 Elicit Robust Protective Immune Responses." *Immunity* 53 (2020), pages 1315–1330. DOI: doi:10.1016/j.immuni.2020.11.015.

[18] W.T. Harvey, A.M. Carabelli, and B. Jackson et al. "SARS-CoV-2 variants, spike mutations and immune escape." *Nat Rev Micorbiol* 19 (2021), pages 409–424. URL: https://doi.org/10.1038/s41579-021-00573-0.

[19] W. Dejnirattisai et al. "SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses". *Cell* 185.3 (2022), pages 467–484. DOI: https://doi.org/10.1016/j.cell.2021.12.046.

[20] S. Yuan, X. Yin, and X. Meng et al. "Clofazimine broadly inhibits coronaviruses including SARS-CoV-2." *Nature* 593 (2021), pages 418–423. URL: https://doi.org/10.1038/s41586-021-03431-4.

[21] ACTIV-3/TICO Study Group. "Efficacy and Safety of Ensovibep for Adults Hospitalized With COVID-19 : A Randomized Controlled Trial." *Ann Intern Med* 175 (9) (2022), pages 1266–1274. DOI: 10.7326/M22-1503.

[22] Novartis Pharmaceuticals. *A Dose Finding, Efficacy and Safety Study of Ensovibep (MP0420) in Ambulatory Adult Patients With Symptomatic COVID-19 (EMPATHY)*. URL: https://beta.clinicaltrials.gov/study/NCT04828161.

[23] S. Jiang, K. Lin, N. Strick, and A.R. Neurath. "HIV-1 inhibition by a peptide." *Nature* 365.113 (1993). DOI: 10.1038/365113a0.

[24] L. Lu, Q. Liu, and Y. Zhu et al. "Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor." *Nat Commun* 5.3067 (2014). DOI: 10.1038/ncomms4067.

[25] S. Xia, Y. Zhu, and M. Liu et al. "Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein." *Cell Mol Immunol* 17 (2020), pages 765–767. URL: https://doi.org/10.1038/s41423-020-0374-2.

[26] L. Xing et al. "A Five-Helix-Based SARS-CoV-2 Fusion Inhibitor Targeting Heptad Repeat 2 Domain against SARS-CoV-2 and Its Variants of Concern". *Viruses* 14.3 (2022). DOI: 10.3390/v14030597.

[27] W. Bi, G. Chen, and B. Dang. "Novel Engineered SARS-CoV-2 HR1 Trimer Exhibits Improved Potency and Broad-Spectrum Activity against SARS-CoV-2 and Its Variants". *Journal of Virology* 96.13 (2022). DOI: 10.1128/jvi.00681-22.

[28] C. Wang, R. van Haperen, and J. Gutiérrez-Álvarez et al. "A conserved immunogenic and vulnerable site on the coronavirus spike protein delineated by cross-reactive monoclonal antibodies." *Nat Commun* 12.1715 (2021). URL: https://doi.org/10.1038/s41467-021-21968-w.

[29] M. Cano-Muñoz, D. Polo-Megías, A. Cámara-Artigas, and F. Conejero-Lara et al. "Novel chimeric proteins mimicking SARS-CoV-2 spike epitopes with broad inhibitory activity." *Int J Biol Macromol* 222 (2022), pages 2467–2478. DOI: 10.1016/j.ijbiomac.2022.10.031.

[30] E.F Pettersen, T.D. Goddard, and C.C. Huang et al. "UCSF Chimera–a visualization system for exploratory research and analysis." *J Comput Chem* 25.13 (2004). DOI: 10.1002/jcc.20084.

[31] J. Kyte and R.F. Doolittle. "A simple method for displaying the hydropathic character of a protein." *J Mol Biol* 157.1 (1982), pages 105–132. DOI: 10.1016/0022-2836(82)90515-0.

[32] L.M Meireles, A.S. Dömling, and C.J. Camacho. "ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery." *Nucleic Acids Res* 38 (2010), W407–W411. DOI: 10.1093/nar/gkq502.

[33] D. Polo-Megías, M. Cano-Muñoz, A.G. Berruezo, G. Laumond, C. Moog, and F. Conejero-Lara. "Exploring Highly Conserved Regions of SARS-CoV-2 Spike S2 Subunit as Targets for Fusion Inhibition Using Chimeric Proteins." *Int J Mol Sci* 23.24 (2022). DOI: 10.3390/ijms232415511.

[34] P. Chen, A. Nirula, and B. Heller et al. "SARS-CoV-2 Neutralizing Antibody LY-CoV555 in Outpatients with Covid-19." *N Engl J Med* 384.3 (2021), pages 229–237. DOI: 10.1056/NEJMoa2029849.

[35] J.S. Patton and P.R. Byron. "Inhaling Medicines: Delivering Drugs to the Body Through the Lungs." *Nat Rev Drug Discov* 6.1 (2007), pages 67–74. DOI: 10.1038/nrd2153.

[36] F. Chen, Z. Liu, and F. Jiang. "Prospects of Neutralizing Nanobodies Against SARS-CoV-2." *Frontiers in immunology* 12 (2021). URL: https://doi.org/10.3389/fimmu.2021.690742.

[37] The antibody society. *Therapeutic monoclonal antibodies approved or in review in the EU or US.* Accessed: 02.06.2023. URL: www.antibodysociety.org/resources/approved-antibodies..

[38] C. Hamers-Casterman, T. Atarhouch, and S. Muyldermans et al. "Naturally occurring antibodies devoid of light chains." *Nature* 363 (1993), pages 446–448. DOI: 10.1038/363446a0.

[39] E. De Genst, D. Saerens, S. Muyldermans, and K.Conrath. "Antibody repertoire development in camelids." *Dev Comp Immunol* 30 (2006), pages 187–198. DOI: 10.1016/j.dci.2005.06.010.

[40] S. Muyldermans. "Nanobodies: natural single-domain antibodies." *Annu Rev Biochem.* 82 (2013), pages 775–797. DOI: 10.1146/annurev-biochem-063011-092449.

[41] S. Spinelli, L. Frenken, and D. Bourgeois et al. "The crystal structure of a llama heavy chain variable domain." *Nat Struct Mol Biol* 3 (1996), pages 752–757. URL: https://doi.org/10.1038/nsb0996-752.

[42] S Muyldermans, T. Atarhouch, J. Saldanha, JA. Barbosa, and R. Hamers. "Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains." *Protein Eng* 7 (1994), pages 1129–1135. DOI: 10.1093/protein/7.9.1129.

[43] J.M. Pérez, J.G. Renisio, and J.J. Prompers et al. "Thermal unfolding of a llama antibody fragment: a two-state reversible process." *Biochemistry* 40 (2001), pages 74–83. DOI: 10.1021/bi0009082.

[44] S. Ewert, C. Cambillau, K. Conrath, and Plückthun. "Biophysical properties of camelid V(HH) domains compared to those of human V(H)3 domains." *Biochemistry* 41.11 (2002), pages 3628–3636. DOI: 10.1021/bi011239a.

[45] K. Decanniere, S. Muyldermans, and L. Wyns. "Canonical antigen-binding loop structures in immunoglobulins: more structures, more canonical classes?" *Journal of Molecular Biology* 300.1 (2000), pages 83–91. DOI: https://doi.org/10.1006/jmbi.2000.3839.

[46] I. Jovčevska and S. Muyldermans. "The Therapeutic Potential of Nanobodies." *BioDrugs* 34.1 (2020), pages 11–26. DOI: 10.1007/s40259-019-00392-z.

[47] D.A. Richards. "Exploring alternative antibody scaffolds: Antibody fragments and antibody mimics for targeted drug delivery." *Drug Discov Today Technol* 30 (2018), pages 35–46. DOI: 10.1016/j.ddtec.2018.10.005.

[48] J.P. Salvador, L. Vilaplana, and M.P. Marco. "Nanobody: outstanding features for diagnostic and therapeutic applications." *Anal Bioanal Chem* 411.9 (2019), pages 1703–1713. DOI: 10.1007/s00216-019-01633-4.

[49] M. Dumoulin, K. Conrath, and A. Van Meirhaeghe et al. "Single-domain antibody fragments with high conformational stability." *Protein Sci* 11 (2002), pages 500–515. DOI: 10.1110/ps.34602.

[50] C. Vincke, R. Loris, D. Saerens, S. Martinez-Rodriguez, S. Muyldermans, and K. Conrath. "General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold." *J Biol Chem* 284.5 (2009), pages 3273–3284. DOI: 10.1074/jbc.M806889200.

[51] D.T. Ta, E.S. Redeker, B. Billen, G. Reekmans, J. Sikulu, J.P. Noben, W. Guedens, and P. Adriaensens. "An efficient protocol towards site-specifically clickable nanobodies in high yield: Cytoplasmic expression in Escherichia coli combined with intein-mediated protein ligation." *Protein Eng Des Sel PEDS* 28 (2015), pages 351–363. DOI: 10.1093/protein/gzv032.

[52] Y. Wang, X. Li, X. Chen, J. Nielsen, D. Petranovic, and V. Siewers. "Expression of antibody fragments in Saccharomyces cerevisiae strains evolved for enhanced protein

secretion." *Microb Cell Fact* 20.134 (2021). URL: https://doi.org/10.1186/s12934-021-01624-0.

[53] R. Baghban, S.L. Gargari, M. Rajabibazl, S. Nazarian, and H. Bakherad. "Camelid-derived heavy-chain nanobody against Clostridium botulinum neurotoxin E in Pichia pastoris." *Biotechnol Appl Biochem* 63 (2016), pages 200–205. DOI: 10.1002/bab.1226.

[54] Q. Lu, Z. Zhang, and H. Li et al. "Development of multivalent nanobodies blocking SARS-CoV-2 infection by targeting RBD of spike protein." *J Nanobiotechnol* 19.33 (2021). URL: https://doi.org/10.1186/s12951-021-00768-w.

[55] C.J. Bracken, S.A. Lim, and P. Solomon et al. "Biparatopic and multivalent VH domains block ACE2 binding and neutralize SARS-CoV-2." *Nat Chem Biol* 17.1 (2021), pages 113–121. DOI: 10.1038/s41589-020-00679-1.

[56] S. Nambulli et al. "Inhalable Nanobody (PiN-21) prevents and treats SARS-CoV-2 infections in Syrian hamsters at ultra-low doses". *Science Advances* 7.22 (2021). DOI: 10.1126/sciadv.abh0319.

[57] M. Schoof, B. Faust, R. A. Saunders, S. Sangwan, V. Rezelj, and N. Hoppe et al. "An ultrapotent synthetic nanobody neutralizes SARS-CoV-2 by stabilizing inactive Spike". *Science* 370.6523 (2020), pages 1473–1479. DOI: 10.1126/science.abe3255.

[58] Y. Xiang, S. Nambulli, Z. Xiao, H. Liu, Z. Sang, W. P. Duprex, D. Schneidman-Duhovny, C. Zhang, and Y. Shi. "Versatile and multivalent nanobodies efficiently neutralize SARS-CoV-2". *Science* 370 (2020), pages 1479–1484. DOI: 10.1126/science.abe4747.

[59] D. Sun, Z. Sang, and Y.J. Kim et al. "Potent neutralizing nanobodies resist convergent circulating variants of SARS-CoV-2 by targeting novel and conserved epitopes." *Nat Commun* 12 (2021). URL: https://doi.org/10.1038/s41467-021-24963-3.

[60] A. Waterhouse et al. "SWISS-MODEL: homology modelling of protein structures and complexes". *Nucleic Acids Research* 46 (2018), pages 296–303. DOI: 10.1093/nar/gky427.

[61] L.S. Mitchell and L.J. Colwell. "Comparative analysis of nanobody sequence and structure data." *Proteins* 86 (2018), pages 697–706. DOI: 10.1002/prot.25497.

[62] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. Wilkins, R. Appel, and A. Bairoch. "Protein Identification and Analysis Tools on the ExPASy Server". In: *The Proteomics Protocols Handbook*. Edited by J. M. Walker. Humana Press, 2005, pages 571–607. DOI: 10.1385/1-59259-890-0:571.

[63] D. Piovesan, I. Walsh, G. Minervini, and S.C.E. Tosatto. "FELLS: fast estimator of latent local structure." *Bioinformatics.* 33.12 (2017), pages 1889–1891. DOI: 10.1093/bioinformatics/btx085.

[64] L. J. McGuffin, K. Bryson, and D. T. Jones. "The PSIPRED protein structure prediction server". *Bioinformatics* 16.4 (2000), pages 404–405. DOI: 10.1093/bioinformatics/16.4.404.

[65] C. Louis-Jeune, M.A. Andrade-Navarro, and C. Perez-Iratxeta. "Prediction of protein secondary structure from circular dichroism using theoretically derived spectra." *Proteins: Structure, Function, and Bioinformatics* 80 (2011). DOI: 10.1002/prot.23188.

[66] P. L. Mateo, F. Conejero-Lara, I. Luque, J. Ruiz-Sanz, J. C. Martinez, A. I. Azuaga, and E. S. Cobos. *In: Enthalpy and Internal Energy; Biocalorimetry: Differential Scanning Calorimetry of Protein Solutions*. Edited by W. Emmerich and Trevor M. Letcher. The Royal Society of Chemistry, 2017, pages 315–335. DOI: 10.1039/9781788010214-00315.

[67] J.P. Moore and R.W. Doms. "The entry of entry inhibitors: a fusion of science and medicine." *Proc Natl Acad Sci USA* 100.19 (2003), pages 10598–10602. DOI: 10.1073/pnas.1932511100.

[68] W. Pang, Y. Lu, and Y.B. Zhao et al. "A variant-proof SARS-CoV-2 vaccine targeting HR1 domain in S2 subunit of spike protein." *Cell Res* 32 (2022), pages 1068–1085. URL: https://doi.org/10.1038/s41422-022-00746-3.

[69] M.D. Miller, R. Geleziunas, and E. Bianchi et al. "A human monoclonal antibody neutralizes diverse HIV-1 isolates by binding a critical gp41 epitope." *Proc Natl Acad Sci USA* 102.41 (2005), pages 14759–14764. DOI: 10.1073/pnas.0506927102.

[70] C. Sabin, D. Corti, and V. Buzon et al. "Crystal structure and size-dependent neutralization properties of HK20, a human monoclonal antibody binding to the highly conserved heptad repeat 1 of gp41." *PLoS Pathog* 6.11 (2010), pages 1–11. DOI: 10.1371/journal.ppat.1001195.

Appendix A – Mathematical Fitting Model: 2 States Model

The 2 state model was used to fit the DSC data of the NBs in figure 11 (d). The 2 states model assumes that the proteins in solution are found in either one of the native of the unfolded state. This simple equilibrium is described by following scheme:

$$N \rightleftharpoons U. \tag{6}$$

The equilibroum constant for this process is defined as:

$$K = \frac{[U]}{[N]} = \frac{x_U}{x_N}, \tag{7}$$

where x_U and x_N represent the fractions of the protein in the unfolded and native state respectively and can be also written as:

$$x_N = \frac{1}{1+K}; \quad x_U = \frac{K}{1+K}. \tag{8}$$

Hence, the average enthalpy is:

$$\langle H \rangle = H_N x_N + H_U x_U. \tag{9}$$

Taking the native state as reference the excess enthalpy over that reference is:

$$\langle \Delta H \rangle = H_N x_N + H_U x_U - H_N = \Delta H_U x_U, \tag{10}$$

with ΔH_U being the unfolding enthaply per mole of protein. To obtain the excess molar heat capacity of the protein to the reference state the temperature derivative is taken to obtain:

$$\Delta C_p = \Delta C_{p,x_U} + \Delta H_U \frac{dx_U}{dT}. \tag{11}$$

By substitution with x_u and the van't Hoff equation $\frac{dK}{dT} = K\frac{\Delta H}{RT^2}$ following term is obtained:

$$\Delta C_p = \Delta C_{p,U} \frac{K}{1+K} + \frac{\Delta H_U^2}{RT^2} \frac{K}{(1+K)^2}. \tag{12}$$

Finally, a the temperature dependence of the heat capacity is defined, assuming a linear

function for the native state and a quadratic one for the unfolded state:

$$C_{p,N}(T) = d + e \cdot T; \tag{13}$$

$$C_{p,U}(T) = a + b \cdot T + c \cdot T^2; \tag{14}$$

$$\Delta C_p = (a - d) + (b - e) \cdot T + c \cdot T^2. \tag{15}$$

A reference temperature is needed to integrate the temperature dependence of the thermodynamic parameters. For sake of convenience it is set to the melting temperature T_m. Hence, the temperature dependence of the enthalpy is:

$$\Delta H_U = \Delta H_U(T_m) + \int_{T_m}^{T} \Delta C_p dT. \tag{16}$$

Solving this equation, substituting the result for ΔH_U into van't Hoff's equation and integrating it over dT allows the calculation of K. The final molar heat capacity obtained by the experiment consists of two terms: $C_{p,int}$, which is the difference between molar heat capacity between the folded native state $C_{p,N}$ and the unfolded state $C_{p,U}$; and $C_{p,exc.}$, which corresponds to the heat needed to break the interactions that keep the protein in the folded state. This thermogram is described by following equation:

$$C_p = C_{p,int} + C_{p,exc.} = C_{p,N} + \Delta C_{p,U} \cdot \frac{K}{1+K} + \frac{(\Delta H_U \cdot \Delta H_{vH})}{RT^2} \cdot \frac{K}{(1+K)^2}; \tag{17}$$

In a true two state model the van't Hoff enthaply ΔH_{vH} and Δ H_U should coincide. For more complex unfolding reactions this can not be assumed. Therefore, the ratio between these two values

$$r = \frac{\Delta H_U}{\Delta H_{vH}} \tag{18}$$

is used as a criteria to assess the validity of the two state model. The thermogram in equation (13) also contains the heat capacity contribution from the volume of water displaced by the protein:

$$\Delta C_{p,ap} = C_{p,N} + \Delta C_{p,U} - \frac{v_{pr} \cdot M_{pr}}{v_{H_2O}} \cdot c(H_2O). \tag{19}$$

The contribution from the displacement of water can be added assuming a constant specific volume of 0.73 $cm^3\,g^{-1}$ to correct the thermogram and obtain the partial molar heat capacity.

Appendix B – DSC Peak Integration

Appendix B shows the graphs relevant for the calculation of the enthalpies of unfolding from the corresponding raw experimental thermograms. The thermograms of the NB–L3C complexes were corrected by subtracting a thermogram of L3C alone multiplied by the molar ration, hence by two. From thermograms of the NB–L3C–V39E mixtures a thermogram of a L3C–V39E complex was subtracted multiplied by two. Subsequently, the area under the peak of unfolding was calculated by integration in Origin. The underlying graphs represent the corrected thermograms, with the baselines used for the integration and the value for the area under the unfolding peak.

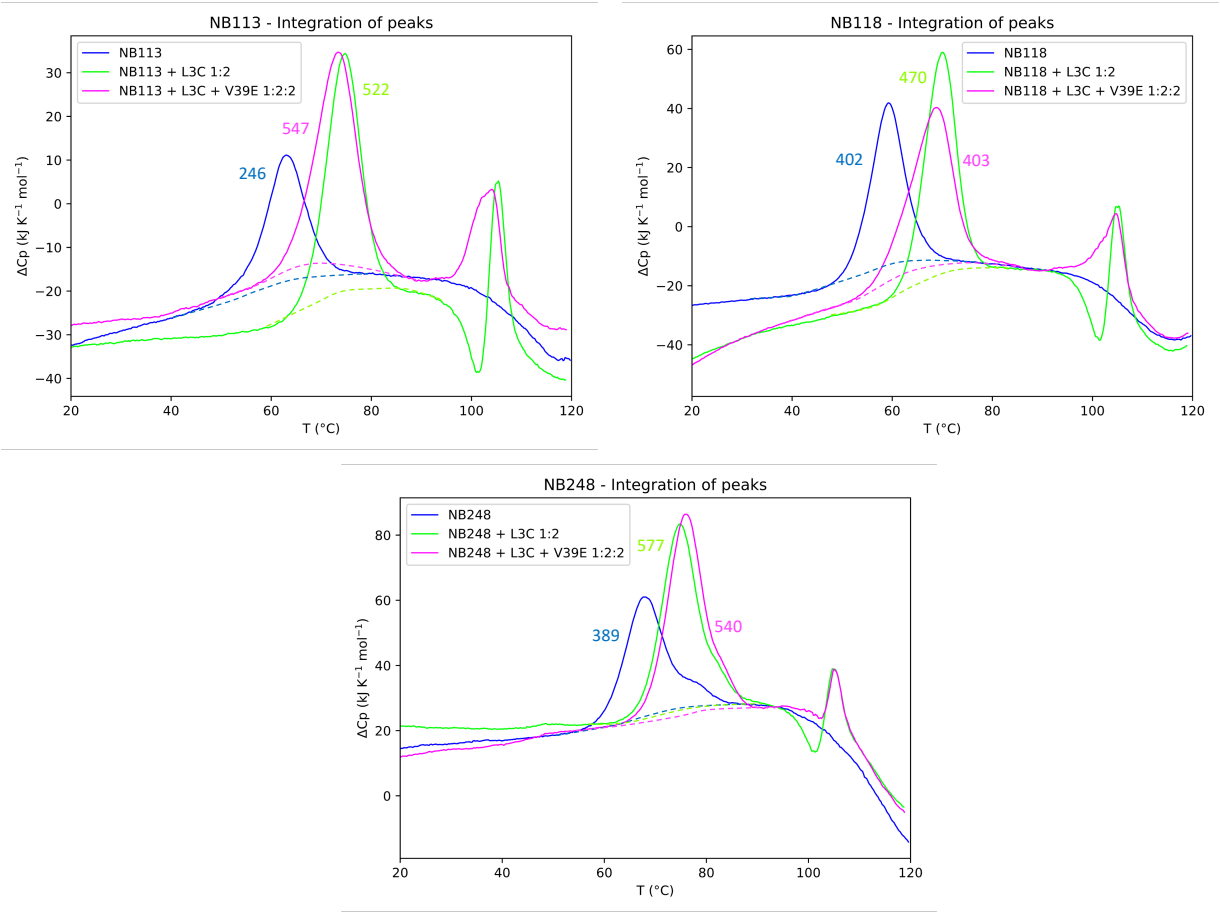

Figure 18: *Corrected graphs used for the calculation of the enthalpies of unfolding by integration of the peaks of unfolding. The dashed lines represent the baselines used for the integration. Integration was done with Origin.*