UNIVERSITÀ DEGLI STUDI DI TRIESTE
E
UNIVERSITÀ CA' FOSCARI DI VENEZIA

XXXVII CICLO DEL DOTTORATO DI RICERCA IN CHIMICA

Progetto finanziato da Fondazione AIRC per la Ricerca sul Cancro, Istituto Italiano di Tecnologia

Biochemical Approaches in Medical Research: Computationally Designed Nanobodies and Novel Inhibitors for HIV-1 Protease

Settore scientifico-disciplinare: **CHIM/03**

DOTTORANDO
FEDERICO NOLASCO

COORDINATORE
PROF. ENZO ALESSIO

SUPERVISORE DI TESI
PROF. RITA DE ZORZI

ANNO ACCADEMICO 2023/2024

ABSTRACT

La costante necessità della ricerca biomedica di aggiornarsi è fondamentale per garantire il progresso nella diagnosi, nel trattamento e nella prevenzione delle malattie. In questa tesi sono proposti due differenti approcci biochimici per la risoluzione di problemi di ambito biomedico.

Nella parte A l'algoritmo evolutivo BINDesignER/PARCE viene utilizzato per generare VHH ottimizzati per il riconoscimento di nuovi target terapeutici. Questa sezione si articola nell'analisi di un caso studio, in cui la specificità di un nanobody viene dirottata verso un target completamente differente rispetto a quello originale. In particolare, le regioni di riconoscimento dell'antigene vengono mutate fino a generare VHH ottimizzati per il riconoscimento del sito attivo del lisozima del bianco d'uovo di gallina. I due VHH con più alta affinità per il nuovo bersaglio, con costanti di legame 60-80 nM, sono in grado di legare il proprio epitopo solo in condizioni di pH che garantiscono la deprotonazione di un residuo istidinico sul sito di legame per l'antigene. Per confermare che il sito attivo dell'enzima è coinvolto nell'interazione con i nanobody, sono stati condotti con successo saggi di inibizione dell'enzima bersaglio utilizzando i VHH ottimizzati.

Successivamente, l'algoritmo PARCE è stato applicato alla generazione di VHH in grado di distinguere due fome estremamente simili di HER2, un noto marker tumorale per il cancro al seno, ma con una significativa differenza funzionale: P105, utilizzata come biomacatore del cancro al seno, e P100, che invece sembra coinvolto nell'inibizione la crescita tumorale. In questo lavoro, vengono mostrati i risultati delle fasi preliminari necessarie alla caratterizzazione dei VHH disegnati in silico, ovvero quelle relative alla produzione di proteine funzionali. Al fine di effettuare saggi di binding, è stata sintetizzata una libreria di peptidi in grado di mimare la sequenza dell'estremita C-terminale di P105.

La parte B di questo lavoro si concentra sulla caratterizzazione di una libreria di potenziali inibitori della proteasi di HIV-1, virus responsabile dell'AIDS. Nonostante quarant'anni di ricerca sul tema, l'alta frequenza di mutazione del virus e lo sviluppo di resistenze portano alla ricerca di composti sempre nuovi da impiegare come potenziali farmaci. Test condotti contro la forma "wild-type" dell'HIV-1 proteasi hanno permesso di individuare nella libreria ALES un lead compoud, che mostra un IC50 paragonabile a quello del Darunavir. Tramite un approccio strutturale, è stato possibile produrre cristalli di questo inibitore in complesso con diversi mutanti farmacoresistenti di HIV-1 proteasi. L'analisi delle strutture ha permesso di individuare l'inibitore nella tasca del sito attivo di tutte queste

proteine, a piena occupazione e sempre nel medesimo orientamento. Cristalli di uno dei mutanti sono stati ottenuti anche in complesso con un altro inibitore della stessa libreria.

ABSTRACT

The constant need for biomedical research to stay updated is essential for advancing the diagnosis, treatment, and prevention of diseases. This thesis proposes two different biochemical approaches to solving biomedical problems.

In Part A, the evolutionary algorithm BINDesignER/PARCE is used to generate VHHs optimized for recognizing new therapeutic targets. This section includes a case study in which the specificity of a nanobody is redirected to a completely different target compared to the original. Specifically, the antigen-recognition regions are mutated to create VHHs optimized to recognize the active site of Hen Egg-White Lysozyme. The two VHHs with the highest affinity for the new target, with binding constants of 60-80 nM, are able to bind their epitope only at pH conditions that ensure the deprotonation of a histidine residue at the antigen-binding site. To confirm that the enzyme active site is involved in the interaction with the nanobodies, successful inhibition assays of the target enzyme were conducted using the optimized VHHs.

Subsequently, the PARCE algorithm was applied to generate VHHs capable of distinguishing between two very similar forms of HER2, a well-known tumor marker for breast cancer, but with a significant functional difference: P105, used as a breast cancer biomarker, and P100, which appears to inhibit tumor growth. This work presents the preliminary results necessary for characterizing the in silico-designed VHHs, specifically focusing on the production of functional proteins. To perform binding assays, a peptide library was synthesized to mimic the C-terminal sequence of P105.

Part B of this work focuses on characterizing a library of potential inhibitors of the protease of HIV-1, the virus responsible for AIDS. Despite forty years of research, the high mutation rate of the virus and the development of drug resistance drive the need for new compounds as potential drugs. Inhibition tests against the "wild-type" form of HIV-1 protease identified a lead compound in the ALES library with an IC50 comparable to that of Darunavir. Using a structural approach, crystals of this inhibitor were produced in complex with various drug-resistant HIV-1 protease mutants. Structural analysis revealed the presence of the inhibitor in the active site pocket of all these proteins, with full occupancy and in the same orientation. Crystals of one mutant were also obtained in complex with another inhibitor from the same library.

Contents

PART A

1. INTRODUCTION

1.1 Nanobodies: a versatile and efficient alternative to conventional antibodies

Acquired immunity allows the immune system to recognize and effectively combat pathogens, decreasing the risk of infectious diseases. In mammals, this process is mediated by immunoglobulins, or antibodies, produced by lymphocytes. The structure of immunoglobulin γ (IgG) is highly conserved across mammals, consisting of two identical polypeptide heavy chains (H) and two identical light chains (L). While light chains contain two domains, heavy chains have four. The region responsible for antigen recognition comprises the N-terminal domains of both H and L chains, which vary in amino acid composition and are called VH and VL, respectively. The remaining domains have more conserved sequences and are labeled CH and CL. Proteolytic cleavage allows obtaining antibody fragments that retain the binding properties of IgGs. In particular, papain cleavage produces Fabs, fragments containing a whole light chain and the first two domains of the heavy chain (Figure 1).

A naturally occurring exception to the classical IgG structure is found in camelids (camels, llamas, vicuñas, and alpacas)[1]. Besides IgGs, these species produce also a different kind of antibodies, known as heavy-chain antibodies (HCAbs), that lack the light chain and the C1 domain of the heavy chain (Figure 1). The variable heavy chain domains of IgG2- and IgG3-type HCAbs are known as nanobodies or VHHs (an acronym for Variable domain of the Heavy chain of HCAbs). These 14-15 kDa proteins are the smallest antibody fragments maintaining the original antibody binding affinity and specificity[2]. The relationship between Fabs and IgGs is comparable to that between VHHs and HCAbs[3]: in camelids, VHHs are the antigen-binding domains of heavy chain antibodies (HCAbs), just as Fabs are the antigen-binding units of traditional IgGs. While both Fabs and VHHs maintain the binding specificity and affinity of their full-length counterparts, Fabs originate from the two-chain structure of typical IgGs, while VHHs are single-chain and single-domain structures.

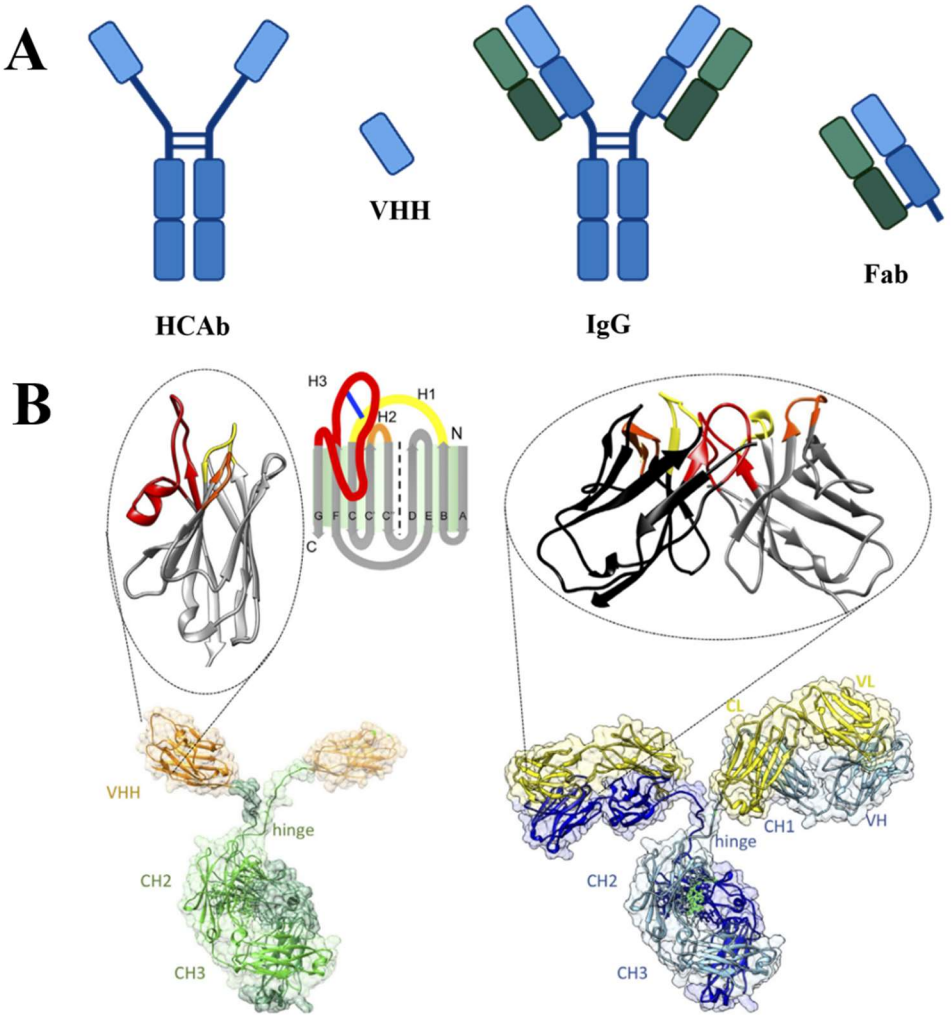


Figure 1: **Antibodies and antibody fragments.** (A) Schematic representation of an HCAb, its VHH fragment, an IgG, and its Fab fragment,. Heavy chains are colored in blue and light blue, light chains are colored in green and light green. Conserved domains are represented in darker colors, and variable domains in light blue and green. (B) Structures of a HCAb, a VHH, an IgG, and a Fab fragment. CDRs are represented in yellow, orange, and red. (Adapted from Muyldermans S., 2021[4]).

From a structural point of view, a VHH comprises nine β-strands that form two sheets, one with four strands and the other with five (Figure 2). The VHH framework consists of four conserved regions (FR1-4) and contains at least one disulfide bridge[5]. FRs are spaced by three loops of varying amino acid composition responsible for specific antigen binding, known as complementarity determining regions (CDR1-3)[6]. Compared to the variable domain of IgG heavy chains (VH), VHHs are smaller and their CDR3 has an extremely variable length. These features confer to VHHs the ability to bind targets that are difficult for conventional antibodies, such as proteins with almost planar surfaces or concave regions like target clefts and protein active sites. To compensate for the superior flexibility of CDR3,

a disulfide bond between a residue in this region and a residue in CDR1, CDR2, or FR2 contributes to stabilizing the framework [7,8].

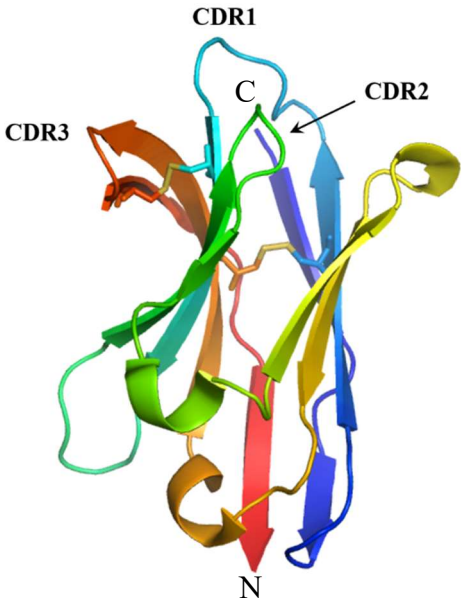


Figure 2: **VHH structure.** Crystal structure of a VHH (from PDB 1ZV5). Complementarity Determining Regions (CDRs) are indicated. Cysteine residues involved in the formation of disulfide bridges are represented as sticks.

VHHs exhibit a significant degree of resemblance to the VH domains of human antibodies. No immunogenic reactions resulting in clinical consequences have been seen thus far in *in vivo* applications of these antibodies. However, long-term clinical treatments usually employ nanobodies with a humanized framework to reduce the possibility of unanticipated reactions[2,9]. The amino acid sequence of a humanized framework is very similar to that of human antibodies[8]. In terms of time and cost of manufacturing, nanobodies are viable substitutes for conventional IgGs. Microbial systems such as bacteria or yeasts can effectively express these tiny antibodies, as recombinant proteins[5,10,11]. Libraries of recombinant nanobodies can be produced with either natural or engineered sequences, in their native form or associated with protein tags. For instance, the cysteine tag offers a convenient site for conjugating VHHs to solid supports like chips, resins, or functionalized surfaces, while other small tags such as the histidine tag or the streptavidin tag enable protein purification via affinity chromatography. Larger tags can enhance the performances of this small fragment by improving its solubility (e.g., Maltose Binding Protein tag, MBP-tag) or enabling easy detection during cell expression and functional analysis (e.g., tags containing GFP or other fluorescent proteins). Construct containing signal peptides can direct the VHHs toward otherwise inaccessible cellular compartments. Additionally, nanobodies can be

organized into homo- or heterooligomers with other VHHs to enhance avidity, bioavailability, and selectivity (Figure 3)[5,12].

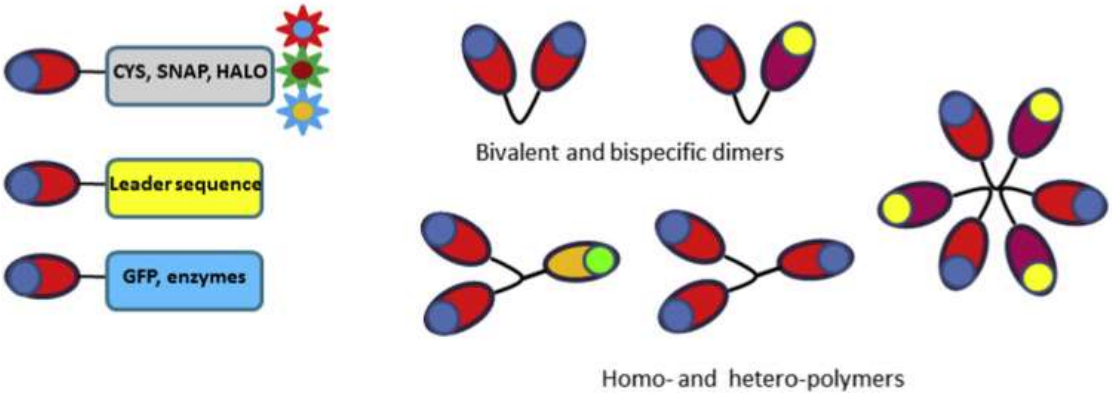


Figure 3: **VHH constructs for biotechnological applications.** On the left, a schematic representation of fusion proteins in which specific tags are inserted in the sequence of a single nanobody, to facilitate purification, detection or localization of the recombinant construct. On the right, homo- and heterooligomes of VHHs with potential for improving avidity or combining different selectivities. (Adapted from de Marco A., 2020[5].)

Two years after the VHH discovery, another kind of antibody lacking the light chains was found in sharks[13,14]. The immunoglobulin neoantigen receptors (IgNAR) have an even smaller antigen-binding unit, known as VNAR, with a molecular weight of ~12 kDa (Figure 4)[15]. Similarly to VHHs, VNAR possesses a particularly long CDR3 that enables its binding to regions hardly accessible to conventional antibodies; on the other hand, VNARs differ from VHHs in the absence of CDR2 and the presence of two highly variable loops (HV1-2). Although less popular than VHHs, VNARs are promising for several applications, particularly in the biomedical field[14].

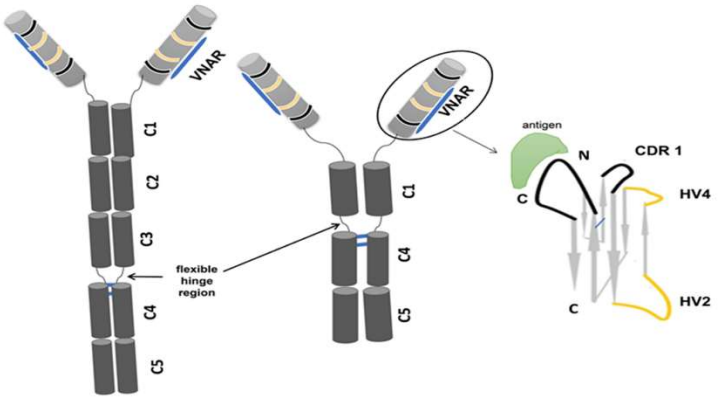


Figure 4: **Shark antibodies.** Schematic representation of immunoglobulin neoantigen receptors (IgNARs) and their antigen-binding domain VNAR. On the left and in the center of the image, two IgNARs produced by alternative splicing. On the right, compared to VHHs, VNARs lack the CDR2. (Adapted from Juma S. *et al.*, 2021[14].)

1.2 Applications of nanobodies across various fields

VHHs, or nanobodies, have been effectively used in several fields, such as biotechnology, medicine, and diagnostics. Thanks to their remarkable target specificity, they are also frequently employed as research instruments. This paragraph reports some examples of the major applications in each field.

VHHs are widely used in theranostics and molecular imaging[16,17], where they are employed as tracers in diagnosing conditions like cancer, inflammation, atherosclerosis, and dementia. For example, nanobodies can be coupled with chelators to carry radionuclides like ^{68}Ga or ^{89}Zr for immuno-positron emission tomography (immunoPET) imaging. They may also be conjugated to radiohalogens to create agents for immunoPET and immuno-single photon emission computed tomography (immunoSPECT). Lysine conjugation is frequently used for nanobody modifications, due to the presence of lysine residues on the surface of the nanobody; yet the lack of specificity of this reaction might result in heterogeneous products. Malemide-cysteine chemistry is a valid alternative for more regulated and site-specific conjugation. It exploits cysteine residues inserted into the nanobody in specific sites, often as a small terminal tag. Maleimide-based reagents react with cysteine thiols to form stable thioether bonds, allowing precise placement of the radioisotope chelator. Recombinant methods such as sortase-mediated ligation provide the nanobody a recognition tag, which allows for even more precise control. For instance, by recognizing a short peptide sequence (the LPXTG motif), the *Staphylococcus aureus*-derived enzyme Sortase A can catalyze the attachment of any chelator with a complementary tag[18].

Nanobodies are also a promising tool for the therapeutic field. While several monoclonal antibodies have been approved as therapeutic agents in cancer treatments[19] and transplant rejection[20], their size is often a significant limiting factor for tissue penetration and crossing the blood-brain barrier (BBB). To overcome these difficulties, research turned toward phage technology applications and the use of engineered antibody fragments. Notably, in February 2019, the FDA approved Sanofi's Caplacizumab, a bivalent nanobody against the Willebrand factor, a glycoprotein involved in the blood clotting cascade. It represents the first life-saving drug for patients with acquired thrombotic thrombocytopenic purpura (aTTP)[21]. In addition, toxin-binding VHHs have shown excellent results against poisoning pathogens[22].

Following the global spread of the SARS-CoV-2 pandemic, nanobody technology made it possible to create useful diagnostic tools[23], showing the same sensitivity and

precision as RT-qPCR, the gold standard detection method, but faster and easier to handle. Several nanobodies that target the SARS-CoV-2 spike protein have been employed for preliminary *in vitro* analyses[24,25], while others progressed to preclinical development for medicinal purposes (Figure 5). Their small size and modularity made them suitable for neutralizing viral particles and preventing virus entrance into host cells. In the Syrian hamster model of COVID-19, administration via the respiratory route allowed a nanobody homotrimer to directly reach the respiratory tract, the primary location of the SARS-CoV-2 infection. This localized delivery resulted in strong therapeutic effects, reducing viral load and mitigating symptoms[26].

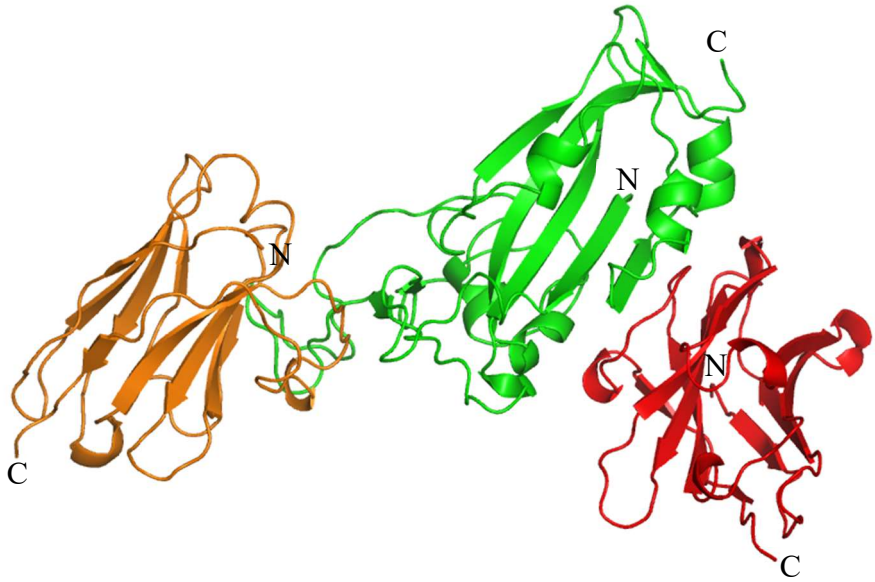


Figure 5: **VHHs against SARS-CoV-2.** Crystal structure of two VHHs bound to the receptor binding domain of the SARS-CoV-2 spike protein (PDB 7OAP[26]). The viral protein is in green, and VHHs are in red and orange.

In structural biology, nanobodies became indispensable tools, especially for stabilizing conformations of dynamic protein regions or locking the protein in a specific functional form (Figure 6). For instance, in G protein-coupled receptors (GPCRs), membrane proteins known to be especially difficult to crystallize, the presence of several flexible loops and their overall conformational flexibility, essential factors for their role in cellular signaling, make the study of their interactions with their signaling partners extremely difficult to study. Nanobody-stabilized GPCRs have advanced structural studies via cryo-electron microscopy (cryo-EM), X-ray protein crystallography (Figure 6), and NMR spectroscopy[27–30].

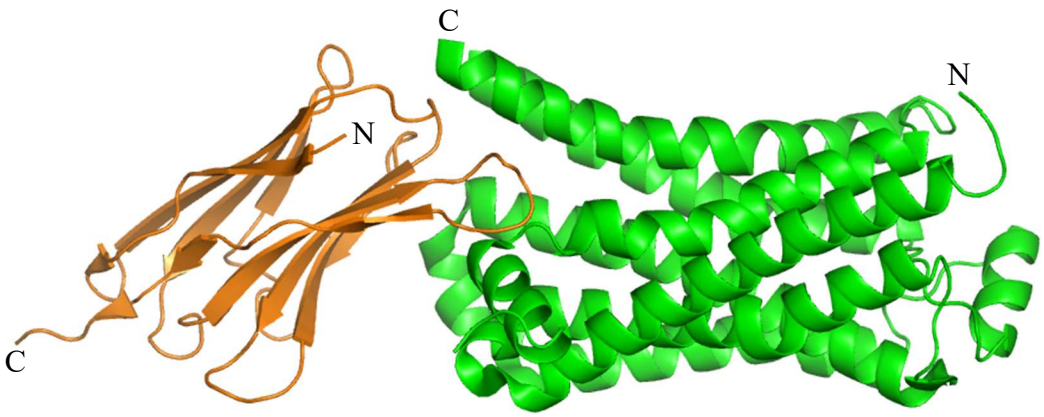


Figure 6: **VHH stabilizing a GPCR protein.** Crystal structure of the active state of the human β(2) adrenergic receptor (β(2)AR), stabilized by the binding of a VHH (PDB 3P0G[30]). The β(2)AR protein is colored green, the VHH orange.

As regards the biotechnological field, VHHs are extensively applied in affinity chromatography. Their high specificity and ease of functionalization enable directional immobilization on inert support materials, such as resins, enhancing the efficiency of chromatography techniques based on antigen-antibody interactions[31,32].

In agribiotechnology, nanobodies with specific action toward fungal glucosylceramide have been shown to prevent the growth of *Botrytis cinerea*, a fungus that infects tomato plants causing extensive economic damage[33]. Similarly, nanobodies specific against the broad bean mottle virus (BBMV) have been explored. Their transient expression in *Vicia faba* has the potential to confer immunity to the plant, providing targeted protection against BBMV infection[34].

1.3 Nanobody libraries

Three types of VHH libraries are used to generate specific nanobodies for different applications: immune, naïve, and synthetic libraries.

Immune libraries are developed by immunizing camelids with a target antigen. This immunization process enhances the immune response, enabling the selection of high-affinity, antigen-specific VHHs (Figure 7A). These libraries are the preferred choice for developing nanobodies for high-sensitivity assays or therapeutic applications since the animal adaptive response is usually effective. However, since this approach requires immunization and involves ethically charged animal welfare issues, it is limited to stable antigens that can be safely administered to animals, like recombinant, soluble, properly

folded proteins; toxic and highly infective agents are mostly avoided. Moreover, this method is time-consuming, since the immunization requires at least 2 months. In addition, it is necessary to immunize simultaneously more than one animal to obtain sufficiently large libraries (10^6-10^8) from which the best binders can be isolated. After completing the immunization schedule, blood samples are collected from the animals to isolate the activated B cells producing HCAbs. mRNA extracted from these cells is retro-transcripted in cDNA and amplified by polymerase chain reaction (PCR) techniques. Since camelids also have conventional antibodies, this step allows amplification of both VHHs from HCAb and VHs domains from the other immunoglobulins. Then, a second PCR step is required to isolate only VHH sequences.

Naïve libraries are produced from camelid immune cells without prior immunization (Figure 7B). This process is suitable for non-immunogenic or toxic targets that cannot be safely administered. To select the optimal nanobodies against a specific target, a large-size library (10^9-10^{11} sequences) is required. Consequently, a large quantity of blood must be collected from several animals (at least 10). Naïve libraries work as a starting point to isolate VHHs with moderate stability and binding affinity since the production of naïve libraries does not involve the natural affinity maturation characterizing immune libraries. After preliminary screens to isolate VHHs from naïve libraries, further optimization of the antibodies is usually required. Due to their large size and broad specificity, the same library can be used to isolate VHH against different targets.

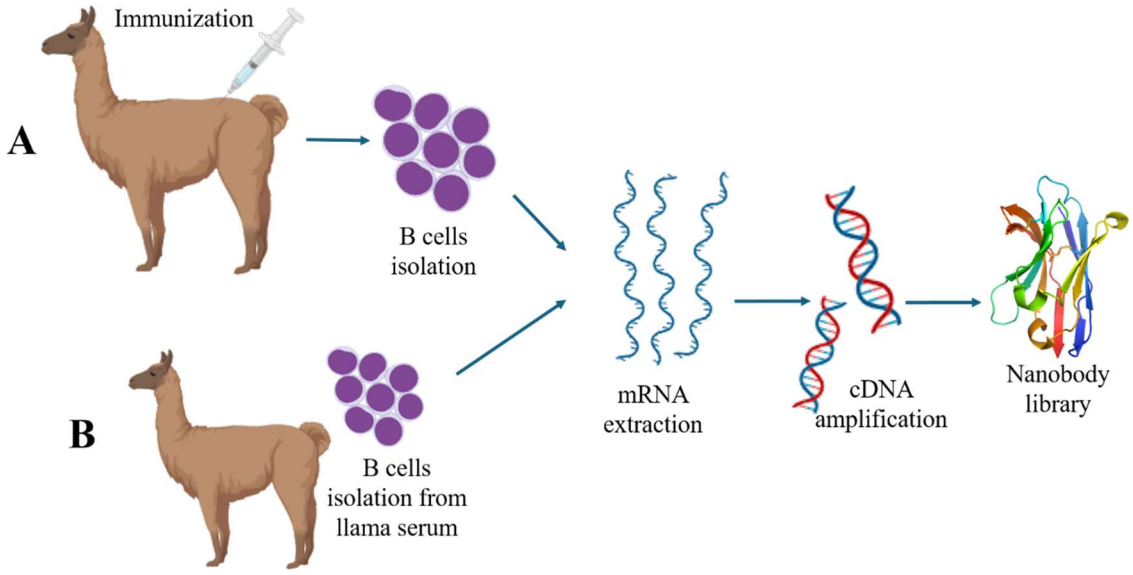


Figure 7: **Production of immune and naïve VHH libraries.** Schematic representation of strategies for producing (A) immune nanobody libraries and (B) naïve nanobody libraries.

Synthetic libraries of VHHs are generated by introducing randomized sequences into known VHH scaffolds, often chosen based on their structural stability, expression levels, or lack of immunogenic responses (Figure 8). Mutations are usually localized within the CDRs, particularly CDR3, the most variable region in length and sequence *in vivo*, and the most involved in antigen recognition. This approach removes the need for animal immunization, making it an ethically preferable option. Moreover, this strategy allows the selection of binders for a broad range of targets, including toxins, small molecules, and short peptides that usually do not generate immune responses. These libraries are customizable to address specific research needs, such as generating VHH with distinct paratope architectures to target particular antigen epitopes. The library size depends on the research necessity, going from 10^9 mutants to 10^{15} or more. In this case, the number of mutants is not a limiting step in the library generation since animal blood is not required, but the screening phase is computationally intensive. The success of these libraries relies heavily on their design, as inadequate synthetic libraries may produce binders with low affinity or incorrect folding. In particular, modifications to CDR3 length and mutations in key structural residues, crucial for stability and function, can affect nanobody production or compromise their performance.

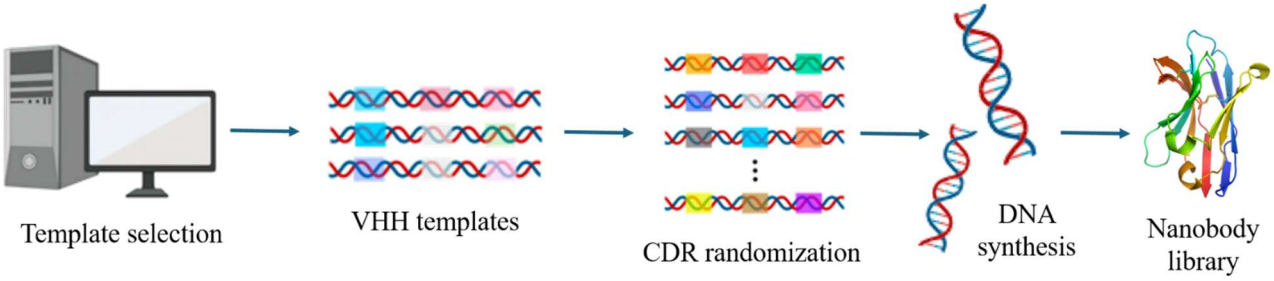


Figure 8: **Synthetic VHH libraries.** Schematic representation of the steps involved in the generation of a synthetic nanobody library.

1.4 Canonical nanobody selection techniques

After their generation, nanobody libraries have to be screened against the selected target. Two main steps are involved in this process: the first step consists in exploiting biological systems to display the libraries on a suitable surface; while the second step is the biopanning, i.e., the selection of antibody-displaying particles that bind to the selected antigen[4,35].

Phage display[36,37], the most widely used display system, was introduced by George Smith in 1985 (Figure 9). It involves displaying antibody fragments on bacteriophages like M13, allowing *in vitro* screening against immobilized antigens. In this system, antibody fragments are fused with one of the phage coat proteins, usually pIII or pVIII. Other phages, like T4[38], T7[39], and lambda phages[40] may be used, often to display large proteins. The gene encoding the antibody fragment is inserted into a phagemid vector as a fusion protein that includes the chosen coat protein. Recombinant phage particles produced using this method display the nanobody on their surface. During biopanning, phages that bind to the target are isolated and amplified infecting bacteria. The process is repeated multiple times (typically three or four) to select high-affinity nanobodies. Phage display can handle extremely large combinatorial libraries, which enhances the chance of identifying strong binders even against challenging targets. Although phage display allows for a vast diversity of nanobodies, certain sequences may be overrepresented due to biases in library construction or the amplification process, thus limiting the exploration of possible binders.

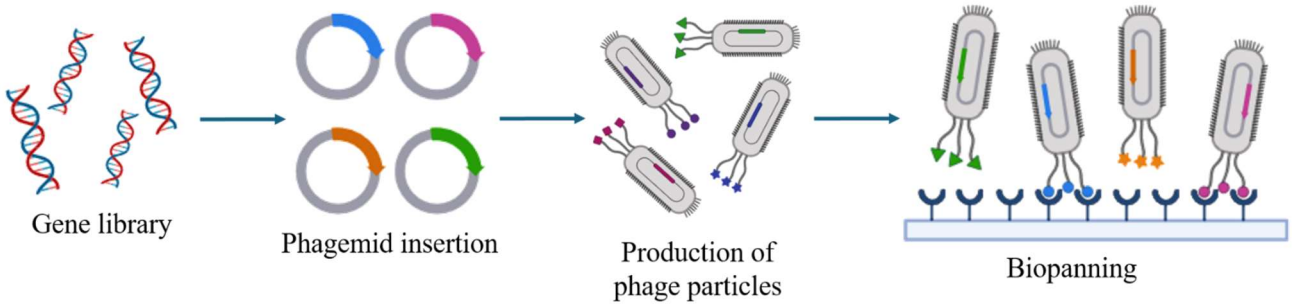


Figure 9: **Phage display and biopanning.** Schematic representation of the phage display technique, followed by biopanning to select nanobodies against a specific, immobilized target, within a library of VHHs.

Introduced in 1977, yeast display[41] is a versatile eukaryotic display system used to develop therapeutic antibodies and other binding proteins, especially suited to identify proteins requiring complex post-translational modifications and proper folding[42]. The system predominantly uses *Saccharomyces cerevisiae*, a yeast strain with cellular machinery compatible with mammalian proteins. The most common yeast display method involves the Aga1-Aga2 complex, the α-agglutinin heterodimer. Aga1 is the cell wall anchoring subunit of the complex, while Aga2 is linked to the yeast cell surface by forming disulfide bonds with Aga1[43]. In this yeast display method, the protein of interest is fused to Aga2 and thereby covalently bound to the surface of the yeast cell. Yeast display is compatible with fluorescence-based detection methods in the biopanning phase, enabling high-throughput selection. For example, flow cytometry may be used to sort cells and select those that display

high-affinity binding proteins. Antigen binding is detected by fluorescent labeling, allowing the selection of the strongest binders through repeated rounds of flow cytometry. Yeast display offers the unique advantage of quantifying *in situ* the binding affinity, considering that the fluorescence signal corresponds directly to the number of binding events, and therefore to the affinity of the antigen-binding domain for the target. The slow growth of yeast cells and the occurrence of events of nanobody over-glycosylation are the primary limitations of this technique and are related to the biology of the expression host. In addition, the presence of multiple protein scaffold copies on the yeast cell surface may result in undesired multivalent binding of protein targets with multiple antigenic regions. Another limit is the requirement of specialized equipment, such as fluorescence-activated cell sorting (FACS) instruments, to screen the library with this method.

The bacterial display method offers distinct advantages in the rapidity of expression, ease of handling, and compatibility with flow cytometry-based sorting. While the most used bacterial species is *Escherichia coli*[44], Gram-positive bacteria, like *Staphylococci, Streptococci*, and *Mycobacteria*, are employed for bulky antibody fragments[45]. In this method, the VHH is recombinantly fused to a bacterial protein exposed on the cell surface[46]. Carrier proteins, including outer membrane proteins, like OmpA[46], flagellar proteins[47], lipoproteins, and autotransporters[48], are commonly used in *E. coli* systems. Compared to the yeast display, this system is low-cost, ensures a rapid growth, and allows the quantification of binding affinity, as in the yeast display system. However, bacterial display relies on the fusion of the target protein with a carrier, which can affect the function and/or stability of some antibodies, making this system less effective than phage and yeast display. *E. coli* display, in particular, requires fusing the VHH with a carrier that allows its translocation across two cell membranes, with a significant chance of impacting the structure of the antibody displayed on the cell surface.

Alternative display techniques are available, albeit less popular, each with specific advantages in terms of high-throughput potential and binding properties. Among these, the ribosome display is worth mentioning, as it is employed in particular for VHHs[49]. This technique allows *in vitro* selection without the need for culturing living cells, enabling the screening of extremely large libraries.

1.5 Alterative computational approaches to high- affinity VHH generation

All techniques used for generating VHH libraries have a major problem: they are time-consuming and costly. Immune libraries require the use of animals, delays for the immune response, and a labor-intensive process of isolation of genes coding for the nanobodies. Naïve libraries do not require antigen delivery but they require the use of more animals, are much larger than immune libraries, and result in less potent binders. Finally, synthetic libraries do not require the use of animals but are very large and their success rate is strictly dependent on the selection of a suitable framework and careful design of regions to be mutated. In addition, following library generation, nanobody selection by display and biopanning techniques requires additional time and resources.

Another significant limitation of these approaches is the difficulty in targeting specific epitopes on the antigen with high selectivity: while the resulting VHHs often bind to the molecular target with high affinity, guiding the selection to a particular region of the antigen is generally not possible.

Alternative computational approaches employed for VHH generation can effectively optimize the time and cost of library production and selection of the strongest ligands.

1.5.1 CDR grafting

CDR grafting is a commonly employed technique for generating new ligands targeting a known antigen. In this approach, the CDRs from an existing larger antibody specific to the target (e.g., an IgG) are transferred onto the framework of a VHH[50]. Typically, this method requires further optimization, as the initial VHH generally exhibits only weak binding to the antigen, due to conformational differences in the antigen-binding surface induced by the different framework. To enhance binding affinity, mutants of the VHH are produced and screened to select the most effective ligands.

An innovative application of nanobody grafting has enabled the generation of VHHs targeting specific epitopes of intrinsically disordered proteins, challenging targets due to their low immunogenicity and complexity of expression.[51] In this approach, the first step is to identify peptides with high specificity and affinity for regions of the target protein by screening PDB structures involving the protein complexes. These peptides are then grafted onto a stable nanobody framework to produce effective VHHs.

1.5.2 *De novo* binder design

Instead of carefully designed modifications of preexisting compounds, the *de novo* approach in rational drug design involves developing new molecules from scratch. This method is widely used in the identification of new drugs and biomolecules, including nanobodies, where limited information is available on existing structures that could inspire rational design.

During the past few years, computational techniques such as deep learning and artificial intelligence have made great strides, profoundly influencing how research is done in an incredible range of fields and applications. For instance, by enabling the prediction of protein structures with extreme precision, the impact that AlphaFold[52,53] had on the structural biology field is huge. The following step would be to train this technology to *de novo* generate new binders. Such is the case with RFdiffusion[54], from BakerLab, a generative method exploiting denoising diffusion probabilistic models (DDPMs) in the generation of protein binders for specific motifs. RFdiffusion was trained on the available experimental structures of protein complexes, to produce new proteins with little resemblance to the initial training set. The method was applied to five targets: Influenza A H1 haemagglutinin, Interleukin-7 Receptor-α, Programmed Death-Ligand 1, Insulin Receptor, and Tropomyosin Receptor Kinase A. Binding assays and cryoEM structures were employed to validate the theoretical results. Recently, RFdiffusion was tuned to the design of single-domain antibodies (VHHs)[55]. While this approach demonstrates significant progress, there remain limitations, such as the modest binding affinities achieved and the still low success rate in producing binders.

1.5.3 Bindesigner/PARCE protocol

The BINDesignER protocol, known as PARCE (Protocol for Amino acid Refinement through Computational Evolution) in its Python version, is an iterative method for optimizing the affinity of a ligand for a specific epitope of an antigen, generating a very small library (10-20 sequences) of high-affinity mutants[56,57]. When applied to nanobody design, this method allows the simultaneous generation of a synthetic library of VHHs and the analysis of their interaction with the target, drastically lowering the time and cost required for canonical library construction, display, and biopanning.

The protocol introduces single random mutations into the nanobody CDRs, which are directly involved in target binding. After each mutation, molecular dynamics simulations are run on the VHH-target complex in a water environment to predict how the nanobody

would interact with the target (Figure 10). Different scoring functions[58–60] (SFs) based on docking evaluate the binding affinity after each mutation. These SFs consider factors such as hydrogen bonding, electrostatics, and van der Waals interactions. Since each SF provides slightly different predictions, the protocol uses a consensus approach: a mutation is retained by the algorithm only if at least half of the SFs indicate an improvement in binding affinity over the previous VHH. Otherwise, the mutation is discarded, and the algorithm starts again from the sequence of the previous successful iteration. This consensus method minimizes reliance on any single SF, enhancing prediction reliability. A threshold parameter can be adjusted to set the level of stringency in mutation introduction. The mutation-evaluation cycle is repeated iteratively. Over multiple cycles, by selecting beneficial mutations, the algorithm optimizes the starting nanobody structure, while progressively increasing binding affinity. After several cycles, the process reaches a point where improvements for additional mutations are minimal, suggesting that the VHH sequence has been optimized for high-affinity binding.

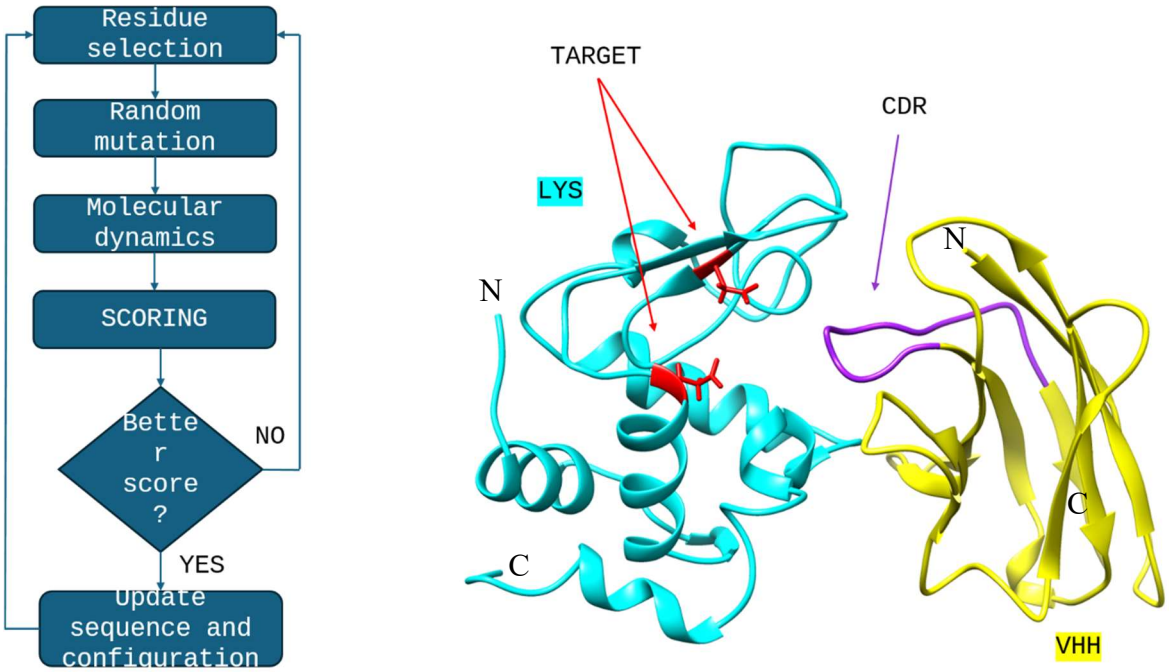


Figure 10: **The BINDesignER/PARCE algorithm.** On the left, a schematic representation of the algorithm used by the BINDesignER/PARCE software for the selection of optimized nanobodies. On the right, an example of the application of the algorithm to select a VHH (yellow, with the CDR3 in purple) that binds to a specific region of the target protein (in red), the Hen Egg White Lysozyme (in cyan).

The BINDesignER platform has been recently enhanced to incorporate a replica-exchange Monte Carlo (REMC) approach, enabling multiple simultaneous optimization

pathways[61]. By conducting parallel VHH optimizations with different mutation acceptance thresholds and periodically exchanging configurations across runs, this method accelerates the search for high-affinity binders. Additionally, this setup reduces the risk of the simulation becoming trapped in suboptimal local minima.

Recently, Locuaz[62], an iterative protocol that allows for parallel optimizations of VHHs based on docking, has been released. This method implements and optimizes the advantages of PARCE[62].

The BINDesignER/PARCE protocol has achieved excellent results in optimizing panned nanobodies for a specific target, increasing their affinity[56]. Its use goes beyond just the optimization of VHHs, as it has also been successfully employed in designing and selecting peptides to be used as ligands for target proteins[63]. All these applications have been experimentally validated, confirming the effectiveness of the method.

As a proof of concept, the protocol was used to optimize a VHH for a specific target, starting from the sequence of a nanobody against a completely different epitope. In other words, the algorithm was applied to change the specificity of a VHH, rather than improving its binding affinity, generating mutants that are highly specific for a target different from that of the progenitor[61]. In particular, starting from a VHH selective against HER2, a receptor involved in breast cancer, the algorithm was applied to target a region on the surface of the Hen Egg White Lysozyme close to the enzyme active site.

In the following chapters, the experimental validation of this proof-of-concept study will be discussed. In particular, three VHHs have been selected from the last optimization steps of the PARCE protocol.

1.6 EGFR family receptors

Epidermal growth factor (EGFR) receptors, or ErbB receptors, are transmembrane proteins with tyrosine kinase activity. EGFRs are involved in the regulation of many cellular processes, including proliferation, differentiation, migration, and cell survival.

The receptor family includes EGFR (or HER1), HER2, HER3, and HER4. These proteins consist of a large N-terminal extracellular region responsible for ligand recognition, a single transmembrane domain (TMD), and an intracellular tyrosine kinase domain (Figure 11)[64].

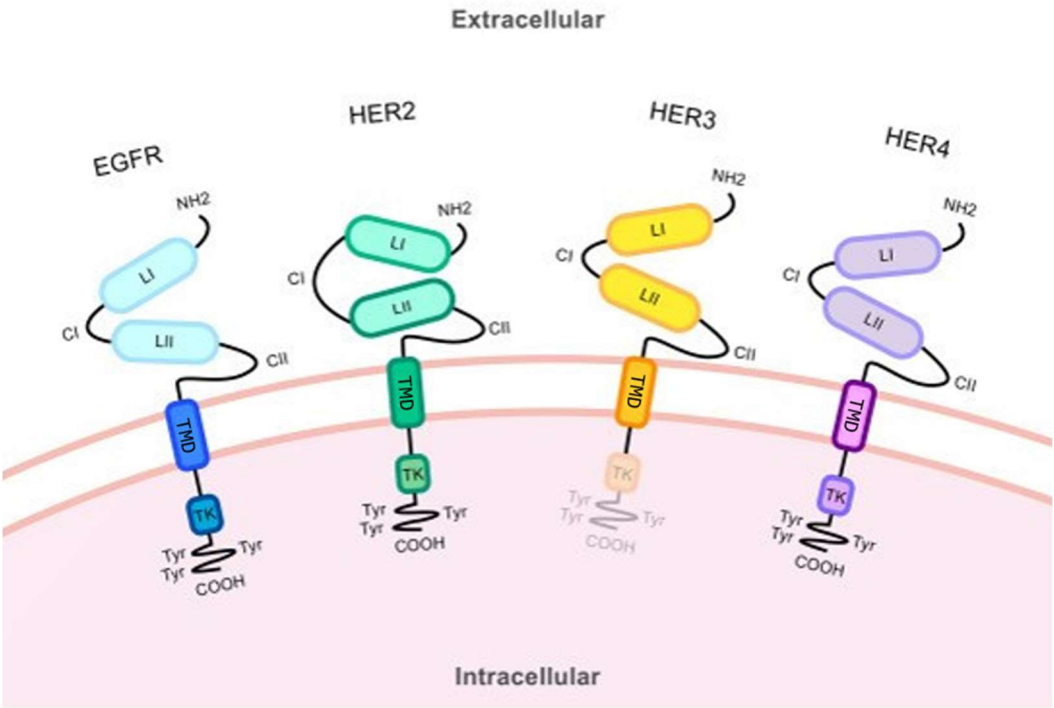


Figure 11: **Epidermal Growth Factor Receptors Family.** Schematic representation of the EGFR proteins. Receptors EGFR, HER2, and HER4 are characterized by: two N-terminal extracellular domains, LI and LII, involved in ligand binding; a transmembrane portion, TMD; a tyrosine kinase cytosolic domain, TK; and a tyrosine-rich C-terminal segment, location of the phosphorylation sites. Receptor HER3 is similar but it lacks the cytoplasmic portion (TK and tyrosine-rich segment). (Adapted from Hart V. *et al.*, 2020[65].)

The extracellular region includes the binding domains (LI and LII) and the cysteine-rich domains (CI and CII). The region between the extracellular portion of the protein and the transmembrane helix is known as juxtamembrane domain (JMD) and is notable for the presence of two proteolytic cleavage sites. In the case of receptors EGFR, HER2, and HER4, the intracellular region includes the tyrosine kinase domain (TK) and the trans-phosphorylation sites (tyrosine residues) in the C-terminal region, while HER3 lacks a functional tyrosine kinase domain. Ligand binding to the extracellular domain induces receptor dimerization, activating the cytoplasmic kinase activity. This activation results in the autophosphorylation of the C-terminal segment, which in turn initiates further downstream signaling events. To date, over 11 different EGF-like ErbB ligands have been identified[66]. While many of these ligands primarily facilitate the formation of homodimeric receptor complexes, the binding of most of them, if not all, to the extracellular portion have also a secondary effect of increasing the probability of formation of heterodimeric receptor complexes with HER2. Despite serving as a universal co-receptor for the ErbB family, HER2

remains the only ErbB receptor for which no high-affinity ligand has been identified, leaving it as the only "orphan receptor" in the family.

Structural analysis of the extracellular domains of ErbB family members reveals that, except for HER2, they maintain a closed, auto-inhibited conformation of the extracellular portion, until ligand binding induces a rearrangement to an open conformation that promotes dimerization[67,68]. In contrast, HER2 adopts an open conformation without a ligand, enabling it to interact with other activated receptors in the family (Figure 12)[69].

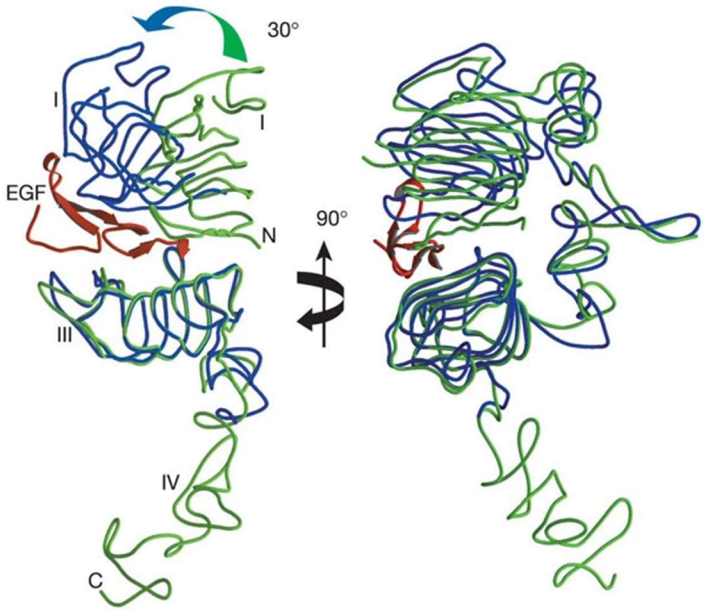


Figure 12: **Extracellular domain of EGFRs.** A comparison between the structure of the ECD of HER1 (blue) bound to its ligand EGF (red), and the structure of the ECD of HER2 (green) shows that the LI and LII domains adopt a similar reciprocal arrangement in the activated state of HER1 and the unbound state of HER2. In this comparison, domains LII of the two receptors were superimposed. In the left panel, the CI cysteine-rich segment has been omitted. The right panel was obtained by rotating the model on the left by 90° along the vertical axis. The structures are available from the PDB with accession codes 1NQL (HER1-EGF) and 1N8Y (HER2). (From Cho H.-S. *et al.*, 2003[69].)

1.6.1 The role of HER2 in breast cancer

Breast cancer, with over 2 million diagnosed cases and 600,000 deaths annually worldwide, is the most prevalent malignant tumor and the leading cause of cancer-related mortality among women[70]. The HER2 receptor plays a significant role in tumor progression by modulating various signaling pathways. HER2 overexpression in approximately 20% of invasive breast cancers makes it a crucial biomarker in diagnostic tests. The so-called HER2-

amplified breast cancers exhibit extremely high HER2 receptor levels (500,000–2,000,000 proteins per cell), far exceeding those in non-amplified tumors (25,000–185,000 proteins per cell)[71]. This overexpression is associated with aggressive tumor behavior and poor prognosis. HER2 levels are generally evaluated using two main diagnostic techniques: immunohistochemistry (IHC), where HER2 expression levels in cancer cells are compared with the normal expression levels of the receptor; and fluorescence in-situ hybridization (FISH), where HER2 amplification is determined by a HER2 gene copy number of six or more or a HER2/CEP17 ratio of 2.0 or greater (CEP17 is the centromere of chromosome 17)[72]. Both methods help in identifying HER2-positive cancers. However, in up to 25% of cases, HER2 status is discordant between the primary tumor and the metastases[73,74].

Through the activation of numerous cellular signaling pathways, HER2 plays a crucial role in the progression of tumors. In particular, HER2 is responsible for the phosphorylation of ERK 1/2 (also referred to as p44/p42). Once activated, this protein, a crucial part of the RAS/MAPK pathway, controls E2F transcription factors, which in turn govern cell cycle entry into the S phase, affecting cell proliferation, apoptosis, and differentiation. Ultimately, uncontrolled activation of ERK 1/2 results in altered gene expression and accelerated cancer progression[75].

Another important pathway activated by HER2 is the PI3K/Akt pathway. In this case, HER2 phosphorylates transcription factors that affect angiogenesis, epithelial-mesenchymal transition (EMT), protein synthesis, metabolism, apoptosis, and cell proliferation[16]. HER2 signaling is also involved in modulating pathways such as Notch and Wnt, and in activating kinases belonging to the Src family and to various transcription factor families. HER2 signaling, however, strongly depends on its dimerization: the formation of homodimers or complexes with other members of the EGFR family results in the activation of different pathways[76].

The estrogen receptor (ER) family is also implicated in breast cancer cases associated with rapid cell proliferation and poor prognosis. Notably, the expression levels of HER2 and of the estrogen receptor often display an inverse correlation in breast cancer tumors[77].

In addition, HER2 autophosphorylation may occur either via dimerization with other EGF receptors or after proteolytic cleavage of the extracellular domain. Truncated carboxy-terminal fragments (CTFs) retain kinase activity, driving tumor progression[65,71,78].

1.6.2 HER2 fragments from proteolytic cleavage

The dimerization of HER2 with other ErbB family receptors is not the only factor influencing the activation of its signal transduction cascade. An important role is played by proteolytic events that produce different fragments and lead to distinct outcomes.

The first HER2 fragment isolated from the medium of HER2-overexpressing cell lines and the serum of patients with metastatic breast cancer is a soluble protein of approximately 105 kDa, corresponding to the extracellular domain (ECD) of HER2[79]. This protein fragment, known as soluble HER2 (sHER2), $p105^{HER2}$, or simply P105, is produced through proteolytic shredding by zinc-containing membrane metalloproteases, such as ADAM (A Disintegrin and Metalloprotease) and MMP (Matrix Metalloprotease) (Figure 13)[80,81]. *In vivo*, the cleavage process can be stimulated by pharmacological and physiological triggers that increase metalloprotease activity. In breast cancer, ADAM 10, 15, and 17 are overexpressed and thus upregulated. However, *in vitro* experiments have shown that the ECD portion of HER2 is likely not a substrate for ADAM 17, with ADAM 10 appearing, instead, to be the principal enzyme involved[71]. The main HER2 cleavage position is located in the juxtamembrane region between residues 647 and 648, while a minor site involves residues 644 and 645 (Figure 13).

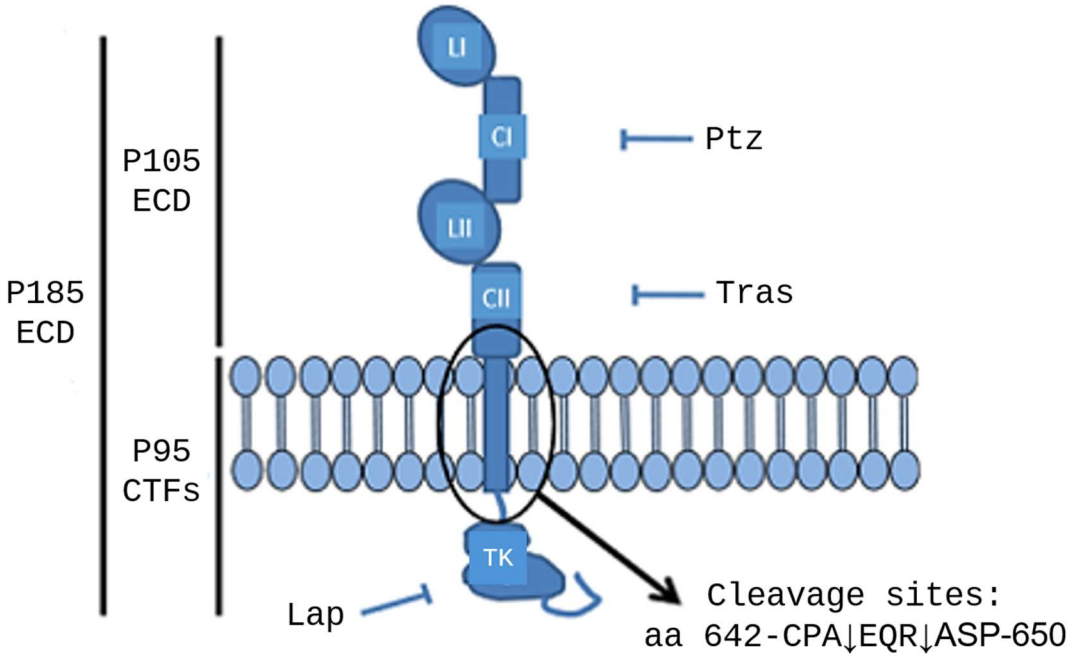


Figure 13: **Cleavage sites on the HER2 protein.** Schematic representation of the fragments of HER2 produced by proteolytic cleavage and the location of the cleavage sites, between residues 644 and 645, and residues 647 and 648, respectively. Ptz, Tras, and Lap refer to the positions of binding sites of the monoclonal antibodies developed against HER2, Pertuzumab and Trastuzumab, and the kinase inhibitor Lapatinib, respectively. (Adapted from Tsé C. *et al.*, 2012[71].)

Following proteolytic cleavage, the extracellular domain P105 is released into the extracellular matrix, while truncated protein fragments of about 95 kDa, known as carboxyl-tail fragments (CTFs), $p95^{HER2}$ or simply P95, remain anchored to the cell membrane, retaining kinase activity. Studies have shown that removing P105 from the whole receptor markedly increases tyrosine kinase activity, making truncated P95 forms 10–100 times more oncogenic than the full-length protein[82]. No longer possessing the extracellular domain that can modulate signal transmission, this form has a constitutively active kinase domain that imparts continuous growth and survival signals to cancer cells. Although both P105 and P95 are produced through proteolytic cleavage, a direct link between P95 levels in the membrane and P105 concentration in the serum remains unclear. Both P105 and P95 are used as tumoral markers to recognize breast cancer patients with poor outcomes[83–86].

1.6.3 HER2 splice variants

In addition to proteolytic fragments of the full-length HER2, shorter protein variants are produced through alternative translation initiation or splicing[65,87]. For instance, translation events that start at the methionine residues 611 or 687 of the HER2 transcript produce two fragments known as 611- and 687-CTF. These CTFs lack the extracellular domain required for dimerization and activation of the full-length HER2. While some HER2 CTFs remain inactive, 611-CTF is a notably oncogenic protein that forms highly active homodimers, promoting tumor progression and metastasis[88]. 611-CTF homodimers are particularly stable due to disulfide bond formation between exposed cysteine residues of the CII cysteine-rich segment.

Alternative splicing can generate HER2 isoforms, with altered signaling functionality. Δ16-HER2 exhibits an in-frame deletion of exon 16, resulting in a variant missing 16 amino acids in the juxtamembrane region (Figure 14, on the left)[89]. Commonly expressed in breast cancer, Δ16-HER2 constitutes about 9% of the wild-type HER2 transcript[90]. This deletion exposes unpaired cysteine residues, leading to cysteine bridge formation and constitutive dimerization[91]. The stable Δ16-HER2 homodimers enhance downstream signaling and are associated with increased Wnt and Notch pathways, contributing to tumor aggressiveness, breast cancer stem cell maintenance, and EMT.

Other splice variants lack the transmembrane domain, making them soluble and causing their release into the extracellular matrix. Herstatin, also known as "self-inhibitor"[92], is a secreted isoform of HER2 having a size of 68 kDa. Herstatin comprises the extracellular

domain LI and the CI cysteine-rich segment, followed by a new C-terminal segment of 79 residues (Figure 14, on the right). Herstatin binds to the CI cysteine-rich domain of the full-length HER2 through the C-terminal portion. The formation of the Herstatin-HER2 dimer hampers the homo- or heterodimerization of the full-length HER2. Since the Herstatin-HER2 complex does not activate the kinase activity of the receptor, Herstatin acts as an inhibitor of EGFR activation[93].

The P100 variant lacks the transmembrane and the intracellular domains compared to the full-length HER2 due to the retention of intron 15, which includes a premature termination codon and the polyA tail addition site (Figure 14, in the center). P100 has the same sequence of the P105 proteolytic fragments but for the absence of 11 or 14 C-terminal residues, depending on the position of the protease cleavage site on P105, i.e. at residues 644-645 or 647-648[94].

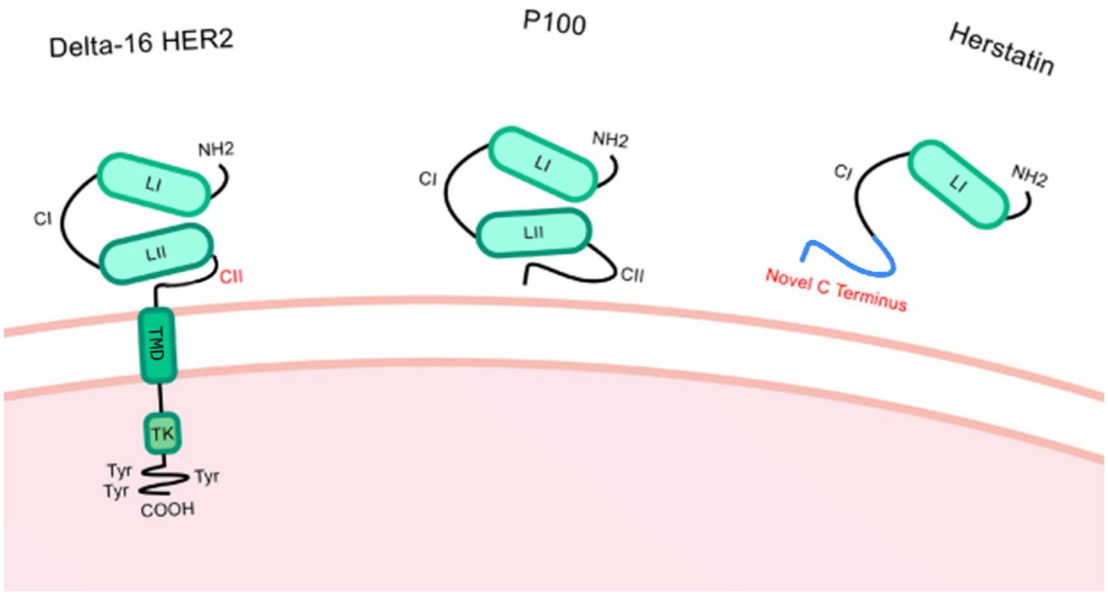


Figure 14: **HER2 splice variants.** On the left, the Δ16-HER2 isoform lacks a 16-residue segment in the CII portion. In the center, the P100 isoform lacks the transmembrane and the intracellular domains. On the right, the Herstatin isoform has only the LI domain and the CI cysteine-rich segment, followed by a new C-terminal segment (in blue). (Adapted from Hart V. *et al.*, 2020[65].)

Despite the little knowledge available regarding the P100 isoform, it is known that this secreted protein reduces the efficacy of drug treatments involving monoclonal antibodies targeting the full-length HER2, by acting as a competitive binder[94]. On the other hand, while high P105 levels are associated with poor prognosis, according to other studies P100 acts as an endogenous inhibitor of tumor cell proliferation, resulting in an opposite clinical outcome[95]. Both P105 and P100 are soluble forms of HER2 present in blood serum and their

selective detection is further hampered by their sequence similarity. Presently, no diagnostic or therapeutic tool can distinguish between P105 and P100. Considering the differences in the clinical outcomes between the two forms, a specific quantification of either would be a crucial step for tumor prognosis and the selection of a suitable treatment. Furthermore, a selective binder able to distinguish between the two soluble proteins would pave the way to the design of new potential therapies in HER2-overexpressing breast cancers.

1.6.4 HER2+ breast cancer therapies based on antibodies

Standard therapies for HER2+ breast cancer include monoclonal antibodies targeting specific domains of the HER2 extracellular portion. The two antibodies approved for treatment, Trastuzumab (Herceptin™) and Pertuzumab (Perjeta™), are widely used in clinical therapy. Trastuzumab targets the CII domain of HER2, blocking protein homodimerization and inhibiting the HER2-mediated mitogenic signaling, thereby reducing cell proliferation (Figure 13 and Figure 15). Pertuzumab binds to an apical region of HER2 extracellular domain, exerting the same effect of interfering with receptor dimerization (Figure 13 and Figure 15)[69,96,97]. When Trastuzumab and Pertuzumab are administered simultaneously, they bind the HER2 extracellular domain and form a ternary complex (Figure 15), with a strong synergy in inhibiting HER2 signaling, which is crucial for HER2+ breast cancer cell survival[98]. Clinical studies have found that combining Trastuzumab with Pertuzumab as an adjuvant yields better results than Trastuzumab alone, with improved survival rates of up to five years for women with metastatic HER2+ breast cancer[72].

Although these two monoclonal antibodies are commonly used against breast cancers, patients frequently develop acquired resistance to these medications, as evidenced by the high recurrence rate. In particular, the decrease in efficacy of Trastuzumab has been related to the or upregulation of the proteolytic cleavage of the receptor, thereby removing the antibody binding site[73,99]. Finding novel ligands that can be used alone or in combination becomes essential to offering therapeutic alternatives[100].

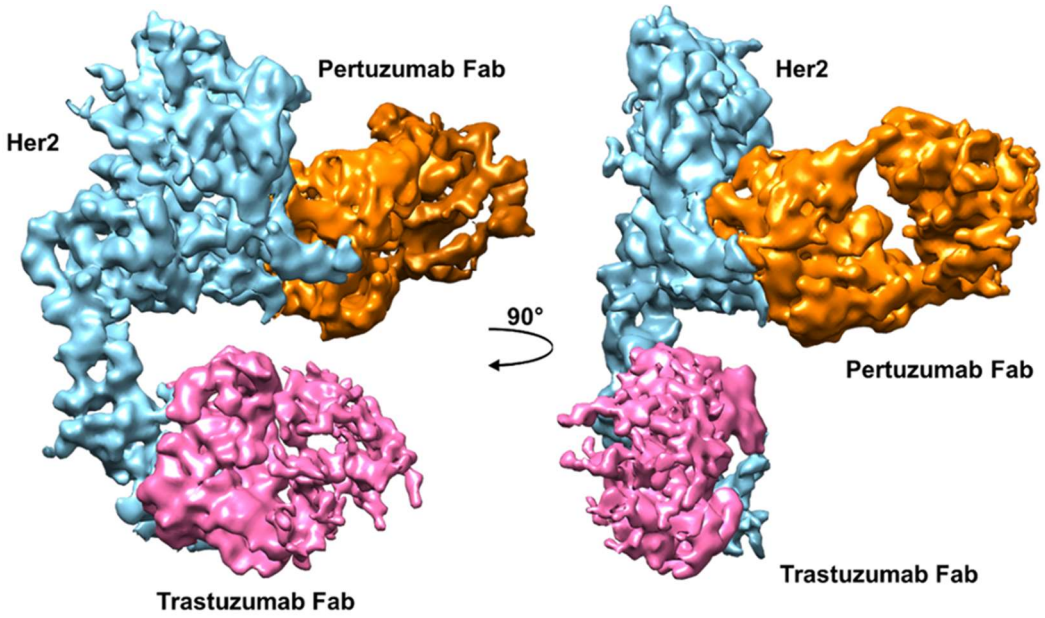


Figure 15: **HER2-Trastuzumab-Pertuzumab ternary complex.** Cryo-EM map of the ternary complex between the extracellular domain of HER2 and the Fab fragments of the monoclonal antibodies Trastuzumab and Pertuzumab. The right panel was obtained by rotating the map on the left by 90° along the vertical axis. The cryo-EM map is available from the Electron Microscopy Data Bank with access code EMD-7137. (From Hao Y. *et al.*, 2019.[98])

2. AIM

Due to their small size and long CDR3, nanobodies bind high with affinity clefts and concave surfaces on proteins, representing valuable tools for diagnostic and therapeutic applications. The possibility of producing recombinantly these small proteins in bacterial expression systems adds a further advantage compared to conventional antibodies. However, traditional methods for generating and selecting nanobody libraries are time-consuming and costly. The BINDesignER/PARCE protocol was developed to overcome these limitations by producing small libraries of nanobodies with high affinity for selected targets. Through an evolutionary algorithm, the protocol selects mutations on the CRDs of the nanobody that enhance epitope recognition. The algorithm has already been successfully applied to enhance the affinity of VHH nanobodies for specific targets.

Recently, in a proof-of-concept study, the BINDesignER/PARCE software has been applied to divert the specificity of a VHH towards a new target, unrelated to the initial epitope, i.e. the active site of the Hen Egg White Lysozyme. The first aim of this thesis is the experimental validation of the results provided by the algorithm in this study. VHHs with the CDR sequences suggested by the PARCE software are expressed in bacterial hosts and purified through chromatographic techniques. The stability of the VHHs in different solutions are tested with Thermal Shift Assays and this technique is also used for a preliminary evaluation of the nanobody-target binding. Further, the ability of these nanobodies to recognize their new target is explored using suitable techniques such as microscale thermophoresis (MST). The structural determinants that lead to the specificity of the VHHs for the new target are analyzed and the experimental results compared to the models of the VHH-target complexes provided by the software. Finally, inhibition assays are conducted to confirm that the binding of the nanobodies to the lysozyme active site has the potential to disrupt the enzymatic activity of the target.

The HER2 receptor is an important biomarker of breast cancers generally indicating a highly aggressive tumor with a poor prognosis. However, different soluble proteins related to HER2 are present in the serum of patients, each deriving from a specific physiologic event and with a markedly different influence on tumor progression. Current theranostic tools are not able to distinguish between two of the soluble forms of HER2, namely P100 and P105, that differ only for a short C-terminal peptide of 11–14 residues, but whose influence on the receptor activation might be divergent. While the soluble isoform P100 is thought to inhibit breast tumor growth, the presence in the serum of P105 is indicative of a proteolytic cleavage

event that leaves a highly active truncated form of the receptor into the cell membrane, leading to a higher probability of activation of mitotic pathways in the cancer cells.

To address the need for stable and specific ligands able to distinguish between the soluble forms of HER2, the PARCE software was applied to design VHHs targeting the HER2 P105 segment that differentiates this fragment from the P100 isoform. Considering the similarity between these two proteins, the computational algorithm has the significant advantage of selecting a specific target region on the protein surface, rather than targeting random epitopes like traditional display techniques. Starting from a VHH that binds FGFR1, PARCE generated a library of VHH sequences selective for P105 and without cross-reactivity with P100. The second objective of this thesis is the cloning, expression, and purification of the anti-P105 VHHs designed by the software. Considering that the target protein P105 is not readily available, a set of peptides mimicking the selected epitope are produced by chemical synthesis in order to test the binding affinity and selectivity of members of the designed VHH library.

.

3. MATERIALS AND METHODS

3.1 Nanobodies against lysozyme

This section outlines the computational and experimental protocols, materials, and analytical techniques used to characterize the optimized VHH against the Hen Egg White Lysozyme (HEWL). The study encompasses five main phases: (1) computational VHH optimization (2) VHH production, (3) evaluation of the VHH thermal stability, (4) experimental analysis of the interaction with the target (HEWL), (5) enzymatic inhibition of HEWL by the VHHs, and (6) crystallization trials for the protein complex. Each phase was carefully structured to maximize the accuracy and reproducibility of the results, employing standardized methods, where applicable.

3.1.1 *In silico* VHH optimization for lysozyme detection

The available crystal structure of an engineered dual-specific VHH antibody was used as a template (PDB code 4POY[101]) for the construction of an anti-Her2 VHH[100] (later referred to as VHH_0) by homology modeling. HADDOCK was used to dock VHH_0 into the lysozyme substrate binding pocket (PDB 1ZV5[102]). The lysozyme binding site was defined by residues 44, 52, 58-60, and 107-110, based on interactions with the co-crystallized camelid antibody D2-L29[102]. The VHH binding site included residues from its three CDRs: CDR1 (26-32), CDR2 (52-58), and CDR3 (98-106). The resulting VHH_0-lysozyme complex served as the starting point for MD simulations to evaluate its stability and compactness for further optimization. The VHH-lysozyme complex underwent minimization, solvation, and neutralization before 100 ps equilibration. A 250 ns MD simulation at 330 K used the AMBER99SB-ILDN force field[103], SPC water model, and periodic boundary conditions. Temperature and pressure were controlled with a thermostat[104] and barostat[105]. The analysis of the resulting MD production trajectory (stripped of solvent and counterions) was conducted using the VMD software package[106] and GROMACS[107].

Starting from VHH_0-lysozyme complex, the PARCE optimization protocol iteratively refined the VHH-HEWL interaction. In each step, a random CDR residue was mutated, and the conformational space of the mutant is sampled via a short MD simulation. Configurations are then scored using six scoring functions. A consensus score is calculated to evaluate the improvement of the mutant compared to its predecessor. If the score meets or

exceeds a predefined threshold, the mutation is accepted, and the new configuration becomes the starting point for the next iteration. If not, the mutation is discarded, and the procedure continues. After each mutation, the system was relaxed through three minimizations: side chain in vacuo, local residues with waters, and the entire system globally. The mutated system was equilibrated with a 100 ps NVT MD run, restraining the protein backbone, followed by an NPT MD simulation at 330 K to sample the VHH-lysozyme complex conformations. MD parameters applied are the same to those utilized during the previously described initial VHH_0-lysozyme complex preparation.

The VHHs selected for the best binding by the optimization algorithm were finally subjected to a longer MD. The last frame from the MD trajectories of each selected VHH-lyzozyme complex has been solvated and neutralized with ions, minimized by employing the steepest descent method.

The simulation design employed the program Scwrl[108] to perform a correct rotamer-based configuration of mutated residues, Gromacs v.2016.1[107] to carry out the minimization and the molecular dynamics (MD) simulations, and IRAD[109] , PIE*PISA[110], BLUUES[60], HADDOCK[111], BACH6[112], and FIREDOCK[59] to obtain scoring functions.

3.1.2 Production of anti-lysozyme VHHs

The nanobody gene sequences were adapted to *Escherichia coli* codon expression, purchased from Twist Bioscience, and cloned into pET-14b vectors. Nanobodies were fused to a C-terminal Cys-tag (-CGCC-) followed by a 6xHis (Figure 16), or to a GFP-6xHis tag.

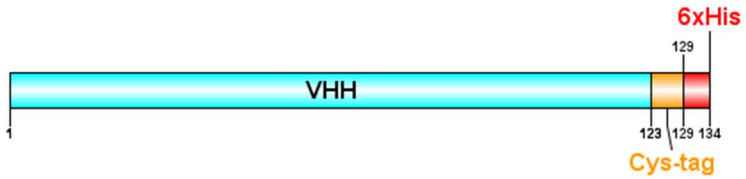


Figure. 16: **VHH fusion protein representation**

Proteins were expressed in the E. coli BL21(DE3) SOX strain, which carries an additional plasmid for the expression of a sulfhydryl oxidase and a prokaryotic disulfide bond isomerase (DsbC), required to facilitate the correct formation of disulfide bridges in bacteria[113–115]. Transformed cells were pre-inoculated in MDG medium in the presence of 34 µg/mL chloramphenicol and 50 µg/mL ampicillin. The cell culture was incubated

overnight at 37°C under agitation. Large-scale protein production was performed using ZYM 5052 autoinduction medium[116], inoculated with a 1:1000 dilution of the MDG preculture. Bacteria were grown for 4h at 37 °C, and then 0.5% arabinose was added to induce sulfhydryl oxidase and DsbC expression. The culture was initially incubated at 30 °C for 45 min and, finally, at 18 °C for 16 h.

Cells were harvested by centrifugation at 6000×g for 15 min at 4 °C, washed with TBS, resuspended in a lysis buffer containing 20 mM Tris pH 7.9, 300 mM NaCl, 1 mM PMSF, 10 μg/mL DNase I, and 10 mM $MgCl_2$, lysed with a cell homogenizer, and centrifuged for 30 min at 4430×g. The soluble fraction was incubated for 1h with TALON superflow resin (Cytiva) (300 μL/L of culture) and then loaded into a gravity chromatography column. The resin was washed with 20 CV of Wash 1 Buffer (20 mM Tris pH 7.9, 300 mM NaCl) and 20 CV of Wash 2 Buffer (20 mM Tris pH 7.9, 300 mM NaCl, 10 mM imidazole). The protein was eluted 4 times with 1CV of the IMAC Elution Buffer (20 mM Tris pH 7.9, 300 mM NaCl, and 150 mM imidazole). Protein polishing was performed at 4 °C by Size Exclusion Chromatography using a Superdex 75 10/300 (Cytiva) equilibrated with the SEC Buffer (20 mM Tris pH 7.9, 150 mM NaCl, 1 mM DTT).

Protein purity was estimated by SDS Gel Electrophoresis using a 15% acrylamide gel. Nanobody concentrations were assessed through absorption spectroscopy at 280 nm, assuming the theoretical extinction coefficients calculated from the amino acid sequences using the ProtParam tool (https://web.expasy.org/protparam/).

3.1.3 Microscale Thermophoresis

Microscale Thermophoresis (MST) assays of nanobodies and their targets were carried out by Prof. Daniela Marasco and Dr. Sara La Manna from the Department of Pharmacy, University of Naples Federico II.

MST is an analytical technique used to study molecular interactions by observing the movement of molecules in response to a temperature gradient. In MST, molecules are labelled with a fluorescent dye and placed in a capillary tube. A localized temperature gradient is generated using an infrared laser, causing the molecules to migrate, a phenomenon known as thermophoresis. The movement depends on size, charge, hydration shell, and binding state of each individual protein species.

To perform a titration MST experiment, a purified protein sample is mixed with its binding partner in different ratios and the fluorescence signal of the complex is recorded. Complex formation alters the physical properties of the fluorescently labelled species,

causing a change in their thermophoretic behavior. The technique allows to measure binding affinities and kinetics in real-time without the need for immobilization.

For this experiment, the Monolith NT 115 system (Nano Temper Technologies) (60 % LED and 40 % IR-laser power) was employed. The labeling of His-tagged VHHs was performed employing the His-Tag labeling Kit RED-tris-NTA, as previously reported[117]. A solution of the VHH was prepared in labeling buffer (Nano Temper Technologies) at 200 nM; a solution of the dye at a concentration of 100 nM was used. 100 µL of VHH and 100 µL of dye were mixed and then incubated in the dark for 30 min. To monitor the binding of the HEWL protein, a ligand solution in the concentration range of 12.5 µM – 0.381 nM, each obtained through serial dilutions, was added to the labelled VHH in 1:1 ratio. Standard capillaries were employed for the analyses, performed at 25 °C in 25 mM Hepes pH 7.5, 150 mM NaCl, 2 mM DTT. An equation implemented by the software MO-S002 MO Affinity Analysis, provided by the manufacturer, was used to fit data at different concentrations.

3.1.4 Thermal shift assays

Thermal shift assay (TSA) is used to study protein stability and ligand binding by monitoring the protein thermal unfolding profile. In this assay, a fluorescent dye that binds to hydrophobic regions is added to a protein solution. When unbound, the dye has minimal fluorescence, but as the protein unfolds with increasing temperature, hydrophobic regions of its core are exposed and the dye binds, increasing the fluorescence signal. In TSA experiments, the protein is subject to a temperature ramp while the fluorescence is detected continuously. When a ligand binds to a protein, it often induces a shift in unfolding temperature (T_u) since the interactions can affect the protein conformation and stability. By comparing T_u values with and without the ligand, TSA can detect ligand stabilization effects, making it useful for preliminary screening of potential drugs.[118]

In experiments performed with anti-HEWL nanobodies, the thermal denaturation of VHHs has been analyzed in buffers at different pH. 10 µM solutions of the VHHs were prepared either in a sodium phosphate buffer (100 mM) at pH 8.0, or in a MES buffer (100 mM) at pH 5.8. The solutions were supplemented by diluting to 2x a solution of SYPRO Orange Protein Gel Stain (5000x in DMSO, from Sigma-Aldrich). The thermal denaturation of the VHHs was analysed in the presence or absence of 10 µM HEWL. 50 µL of each solution were introduced in a white 96-well PCR plate. Each condition was tested in triplicate. The analyses were performed using a Biorad CFX Real-Time PCR thermocycler set on the FRET channel. After 2 min at 20 °C, the temperature was raised by 0.5 °C every

30 s, detecting the fluorescence emission at each increment. The temperature ramp was continued until 80 °C.

3.1.5 VHH-Lysozyme protein crystallization

To structurally determine the exact interactions the optimized VHHs and HEWL, a crystallographic approach was tested. VHH_129 and VHH_135 were crystallized using various commercial screens: Index-HT (Hampton), PEG/Ion-HT (Hampton), PACT Premier HT-96 (Molecular Dimensions). In this case, experiments were conducted by vapor diffusion in a sitting drop configuration. The drops were placed using the Mosquito LCP (SPT Labtech). Additionally, the 24 SturaFootprint conditions[119] were tested. Finally, the screening of similar conditions that successfully produced crystals of the nanobody D2-L29[120] in complex with lysozyme were applied. In these two last cases, hanging-drop experiments were carried out.

3.1.6 Lysozyme activity assays

To test the enzymatic activity, the standard turbidimetric protocol[121] was modified to screen the effect of pH on VHH binding and inhibition of lysozyme activity. Briefly, in the absence of lysozyme binders, *Micrococcus luteus* cells are lysed by the enzyme, causing a significant reduction of the optical density of the suspension; whereas, when binders/inhibitors of lysozyme are present, the efficiency of the enzyme is significantly reduced, and the optical density does not show a decrease.

Freeze-dried cultures of *Micrococcus luteus* (Sigma-Aldrich) were resuspended in an ice bath in a phosphate buffer solution (100 mM) at pH 8, or in a buffer solution (100 mM) at pH 5.8, to obtain a final optical density of ~0.750, measured at a wavelength of 450 nm against the respective buffer solution.

Lysozyme solutions were prepared at a final concentration of 0.04 mg/mL (~940 U/mL). In different experiments to test the inhibitory activity of the VHHs, lysozyme solutions were incubated in the presence of nanobodies, in a 1:5 lysozyme-to-nanobody ratio, for 30-45 minutes at 18°C, before adding them to the bacteria suspension.

The optical density measurements were performed at 25°C using a Shimadzu UV-1800 spectrophotometer. For each experiment, 1 mL of cellular suspension was warmed up to remove the condensation and inserted in a cuvette. Then, 37.5 µL of lysozyme solution, with or without the nanobody (or buffer for the negative control experiments), was added to

the cuvette and the solution was gently inverted. The optical density was monitored at the wavelength of 450 nm for ~5 min. Measurements were repeated at least twice.

The experiment was performed first with a solution containing only lysozyme (negative control), and then repeated with the VHH_0, used as a starting point for the optimization algorithm and with two of the VHH generated by the PARCE algorithm at cycles 129 and 135, referred to as VHH_129 and VHH_135, respectively. The nanobody D2-L29, described in the literature[122] was used as a positive control. A solution containing only the buffer was measured as well (blank).

In the negative control, the lysozyme activity led to a rapid decrease in turbidity due to the cell disruption. Inhibition of the lysozyme activity corresponds to a slower decrease in optical density. The slope of the curve obtained by reporting the decrease in optical density of the culture (ΔA450) against time was employed to calculate the lysozyme's biological activity in Enzyme Units (EU). One EU produces a decrease in the optical density of 0.001 absorbance units per minute at 25°C. The lysozyme activity per milliliter of culture was estimated as follows:

$$\frac{EU}{mL} = \frac{(\Delta A450/min\ Test - \Delta A450/min\ Blank)(df)}{(0.001)(0.0375)}$$

where *ΔA450/min Test* and *ΔA450/min Blank* are the slopes of the curves obtained fitting the data of a single test and of the blank, respectively, *df* is the dilution factor, 0.001 is the change in optical density as per Enzyme Unit definition, 0.0375 is the volume (in milliliters) of the enzyme solution.[123]

3.2 Nanobodies against HER2$_{P105}$

This section lists the experimental protocols developed specifically for the study of nanobodies designed to bind the HER2$_{P105}$ receptor fragment. The main phases of the VHH production and analysis are the following: (1) *in silico* VHH optimization and selection (2) preparation of the VHH gene-containing plasmids for protein expression, (3) small-scale expression and the optimization of the bacterial growing conditions, (4) large-scale production and purification of the VHH. Since the target production requires a mammalian expression system which was not available, a reductionist approach was considered. A peptide corresponding to the C-terminal region of P105 was produced. In order to test the specificity of these VHHs, several control peptides were synthesized including the last residues of P100, but with different modifications in the remaining residues of P105.

3.2.1 Anti P105 VHH computational optimization

The procedure follows the one used in the paragraph 3.1.1. Homology models of the HER2 and the anti-FGFR1 VHH C8[124] (referred as C8WT) proteins were created starting from the complex HER2 ECD-Trastuzumab Fab-Pertuzumab Fab complex (PDB: 6OGE[98]). In this case, VHH candidates came from three different optimization runs (group 100, 200, and 300), each starting from different HER2-VHH conformations of the same complex.VHH protein sequences were obtained from the PARCE algorithm. Atom distance matrices were calculated between the CDR (Complementarity Determining Region) residues of the VHH, and two subset of HER2 residues (P100: C608 T609 H610 S611; P105: C612 V613 D614 L615 D616 D617 K618 G619 C620 P621 A622 E623 Q624 R625 A626). Pairs of atoms within 4 Å of each other were considered as contacts. For a comprehensive evaluation that considered not only the values provided by the scoring functions, but also the contact analyses, four parameters were considered for the selection of the best binders. These included: the ratio of the number of contacts with P100 subset residues to the total number of P100 residues (Q_{100}), the ratio of the number of contacts with P105 residues to the total number of P105 residues (Q_{105}), the ratio of atomic contacts with P100 to the total contacts with the entire HER2 protein atoms (A_{100}), and the ratio of contacts with P105 residues to the total contacts with residues of the whole HER2 protein (R_{105}). The candidate binders were subjected to a longer MD as before (paragraph 3.1.1) to explore the recurrent poses.

3.2.2 Preparation of plasmids for protein expression containing the anti-HER2 VHH genes

VHH protein sequences were obtained from the PARCE algorithm. The Tobacco Etch Virus protease (TEV) consensus cleavage site (ENLYFQ|G) was added to the C-terminus of the VHH sequences. The gene sequences were optimized for *Escherichia coli* codon usage, adding NotI and NcoI restriction sites at the 5' and 3' end, respectively, and synthesized by Twist Bioscience. Genes purchased contained polyA sequences at both ends to protect the coding region from exonucleases. The *pET-14b* plasmid, containing *eGFP-6xHis* tag downstream of the cloning site, was selected as the expression vector.

The genes and vectors were digested with NotI-HF and NcoI-HF (New England BioLabs). The linear vector (5345 kbp) was separated from the cleaved insert by 1% agarose gel electrophoresis, excised from the gel, and then purified using the GenElute™ Gel Extraction Kit (Sigma-Aldrich). The gene fragment was used after digestion without further

purification. Gene cloning employed T4 DNA ligase (New England BioLabs). The VHH genes were successfully inserted using 12 ng of the gene, with insert-to-vector stoichiometric ratios of 1:3 and 1:4. The resulting *VHH-TEV cleavage site-eGFP-6xHis* (Figure 17) construct was analyzed by sequencing.

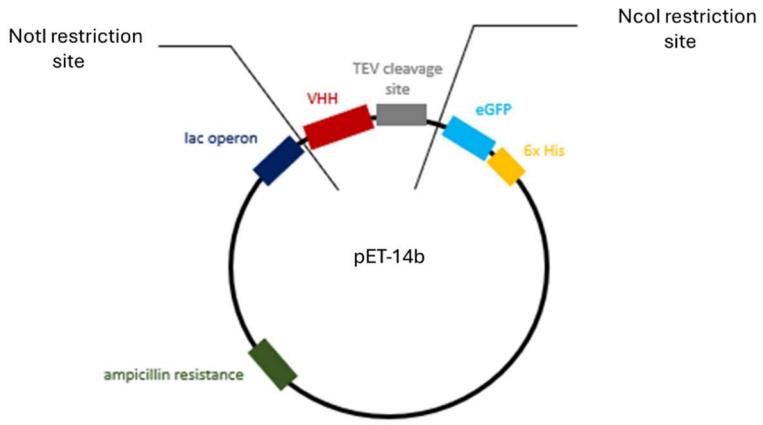


Figure 17: ***pET-14b****.Schematic representation of VHH construct used in this study, inserted into the vector* pET-14b

3.2.3 VHH solubility according to CamSol Structurally Corrected

The CamSol Structurally Corrected method[125] (http://www-vendruscolo.ch.cam.ac.uk/camsolmethod.html) utilizes the three-dimensional structure of a protein to calculate its solubility profile, taking into account the proximity of residues in the protein folding and their exposure to the solvent. This method was applied to the VHH sequences designed by the PARCE algorithm, allowing the visualization of regions of the VHH surface predicted to decrease the solubility.

For the analysis, the HER2-related chain was manually removed from the models of the VHH complexes generated by BINDEsignER/PARCE software. These modified PDB files were used as input. The analyses were conducted at pH 8 (the closest value to the experimental conditions allowed by the CamSol interface), with the Patch radius set to 10 Å. The patch parameter defines the radius of the sphere onto which the intrinsic solubility profile of each residue is projected. In the output PDB files B-factors of each atom were replaced with the solubility values calculated by CamSol for each amino acid residue: values below -0.8 indicate low solubility, while values above 0.8 indicate high solubility. Visualization of the color-coded structures was performed using the PyMOL molecular visualization software, using the B-factor coloring option.

3.2.4 Small-scale expression

29 µL of BL21(DE3) SOX cells were transformed with 1 µL of *pET14b* plasmid (10-100 ng/µL) containing the gene of interest, which consisted of the VHH sequence fused at the C-terminal end to a TEV protease cleavage site, eGFP, and a 6xHis tag (Figure 18). The plasmid contains the ampicillin (Amp) resistance gene, while the SOX cells contain an additional plasmid with chloramphenicol (Cam) resistance.

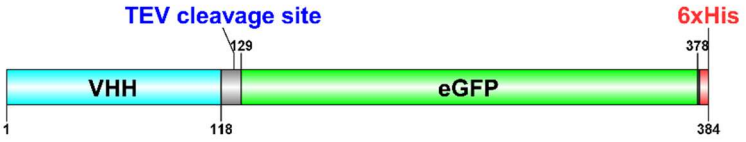


Figure 18: **Representation of the VHH::GFP fusion protein**.

Following transformation using the heat-shock protocol, 470 µL of LB medium was added to each sample and incubated at 37°C for 1 hour. After incubation, 100 µL of the transformed suspension was inoculated into 2 mL of LB medium, 100 µg/mL ampicillin, and 34 µg/mL chloramphenicol, and cultured overnight. A negative control, consisting of untransformed BL21(DE3) SOX cells, was included, without ampicillin in the medium. The effect of supplementing the culture with 1% glucose was tested. The next day, 1:500 inoculations were made into 10 mL of LB medium with antibiotics (Amp and Cam), and the cultures were incubated at 37 °C. Once the optical density at λ=600 nm (OD600) reached approximately 0.5, expression of the chaperones encoded by the SOX plasmid was induced by adding arabinose to a final concentration of 0.5%. The cultures were further incubated at 30 °C for 45 minutes. Afterward, expression of VHHs was induced by adding IPTG to a final concentration of 0.5 mM, followed by a reduction of the incubation temperature to 18 °C. The culture was incubated overnight (16-18 hours) under agitation.

After the induction time, fluorescence and OD600 of the culture were recorded using the plate reader Biotek Synergy H1 Hybrid Reader (Agilent), sampling the cultures into black and clear 96-well plates, respectively. Each measurement was performed in triplicate. To recover the cell pellet, cultures were centrifuged at 4450×g for 20 minutes. The supernatants were discarded and the cell pellets were resuspended in 1 mL of Lysis Buffer (20 mM Tris pH 7.6, 300 mM NaCl, 1 mM PMSF, 10 µg/mL DNaseI, 0.1 mM $MgSO_4$,

0.125% v/v protease inhibitor cocktail (Sigma-Aldrich), 0.2% v/v Triton X-100, 2 mg/mL lysozyme).

Three cycles of freezing (at -80 °C for 15 minutes) and thawing (at 30 °C for 25 minutes) were used to lyse cells. To extract the soluble fraction, lysates were centrifuged at 12000×g for 20 minutes at 4 °C. To analyze protein expression in inclusion bodies, the cellular debris separated by centrifugation was rinsed twice with the Inclusion Body Wash Buffer (2 M urea, 20 mM Tris pH 7.6, 250 mM NaCl, 1% v/v Triton X-100) and centrifuged at 12,000×g for 15 minutes at 4 °C. The resulting pellet was resuspended in Urea Buffer (8 M urea, 50 mM Tris pH 8, 150 mM NaCl, 2 mM DTT) and incubated overnight at 4 °C, under stirring, to solubilize protein aggregates.

The cellular total extract, the soluble fraction of the lysate, and the inclusion bodies pellet solubilized in urea were analyzed through SDS PAGE and Western Blot.

3.2.5 Western blot

A polyvinylidene fluoride (PVDF) membrane was pre-treated with methanol for five minutes. Proteins separated by SDS-PAGE were transferred to the membrane using the wet transfer method appliyg a tension of 100 V. After transfer, the membrane was incubated overnight in TBS containing 5% (w/v) milk powder, under agitation. The membrane was washed three times with TBS + 0.05% Tween 20 (TBST), shaking for 5 minutes each time. The membrane was, then, incubated for 3 hours in TBST + 1% (w/v) milk powder with a solution of anti-6xHis mouse monoclonal antibody conjugated to HRP (Sigma-Aldrich) at a 1:2000 (v/v) dilution. Afterward, the membrane was washed thrice with TBST and treated with the HRP substrate ECL Star (EuroClone) to produce a chemiluminescent signal, detected by the VWR Imager CHEMI Premium system.

3.2.6 VHH::GFP production scale-up and purification

The expression conditions optimized in the small-scale phase were applied to the scale-up phase of protein production. Cell lysis was carried out at 4 °C using the same Lysis Buffer and a Potter-Elvehjem homogenizer for cell lysis. The resulting lysate was clarified by centrifugation at 4 °C at 4430×g for 45 minutes. Because of the different behavior of wild-type VHH and its mutants, two different purification strategies were employed.

3.2.6.1 Soluble protein purification

The soluble fraction derived from 0.5 L of lysed bacterial culture was incubated with 1 mL Talon Superflow™ (Cytiva) resin at 4 °C for 1 h. The suspension was transferred to a gravity column to collect the flow-through. The resin was washed with 20 column volumes (CVs) of Wash Buffer (20 mM Tris pH 8, 300 mM NaCl), followed by 20 CVs of the same buffer supplemented with 5 mM imidazole. The elution was performed with the Elution Buffer (20 mM Tris pH 8, 300 mM NaCl,150 mM imidazole), and the solution containing the protein was dialyzed overnight against 20 mM Tris pH 8, 150 mM NaCl, 10 % v/v glycerol, and 2 mM DTT, to remove the imidazole. In the final purification step, a Superdex™ 200 Increase 10/300 GL column (Cytiva) was used for protein polishing via Size Exclusion Chromatography. Samples from all purification steps were analyzed with SDS-PAGE. Protein concentration was calculated with Bradford assay.

Removal of the eGFP-6xHis tag from the fusion protein was obtained by adding recombinant TEV protease, produced in-house as a fusion protein with a polyhistidine tag. The protease was added to the protein solution in a 1:10 protein substrate-to-protease molar ratio and the solution was dialyzed against a pH 8 buffer solution containing 20 mM Tris, 150 mM NaCl, and 1 mM β-mercaptoethanol for 18 h at 4 °C. Reverse IMAC separation was performed to isolate the cleaved VHH from other components (protease, uncleaved fusion protein, eGFP-6xHis tag): the dialyzed solution was treated with 1 mL Talon Superflow™ (Cytiva), collecting the flow-through. The flow-through was concentrated in an Amicon® Ultra Centrifugal Filter, 3 kDa MWCO (Millipore). A final SEC polishing step was performed in a Superdex™ 75 10/300 GL column (Cytiva) equilibrated with the following buffer: 25 mM Tris pH 7.6, 300 mM NaCl, and 0.1 mM TCEP.

3.2.6.2 C8WT Thermal Shift Assay

After polishing and tag removal, a VHH solution was diluted to 10 μM in each of the buffers listed in Table 1, at a concentration of 100 mM, or in the same buffer used for Size Exclusion Chromatography (Paragraph 3.2.6.1). Each solution was supplemented with 2x SYPRO Orange Protein Gel Stain (Sigma-Aldrich). The experiment was conducted as previously reported (Paragraph 3.1.4) in a white 96-well plate, with each condition analyzed in triplicate.

Table 1: *Buffer compositions used to assess the stability of VHH in different conditions.*

Buffer	pH
Bicin 100 mM	8.5
Sodium Phosphate 100 mM	7.9
Tris 100 mM	7.6
HEPES 100 mM	7.5
HEPES 100 mM	7
MES 100 mM	6.5
SEC buffer	7.6

3.2.6.3 Purification of anti-p105 VHH::GFP proteins expressed in inclusion bodies

Cellular debris from 0.25 L of culture, containing inclusion bodies (IB), was rinsed with Inclusion Body Wash Buffer and resuspended in 5 mL of Urea Buffer, as described in Paragraph 3.2.4. The suspension was incubated overnight at 4 °C, under stirring, to allow the solubilization of protein aggregates. After solubilization, samples were centrifuged at 13000×g for 30 minutes at 4 °C to remove insoluble material. The centrifugation protocol was repeated three times to ensure the efficient removal of insoluble components, thus enhancing the purity of the solubilized inclusion bodies.

Three different protein refolding protocols were applied on VHH 100_85. Results were analyzed and the optimal approach was chosen for the refolding of the other VHHs produced in inclusion bodies.

- **Protocol A: Dialysis**

  8 mL of a solution of 0.5 mg/mL unfolded protein, in urea 8M, was sequentially dialyzed against a larger volume (1 L) of the following buffers, with decreasing urea concentrations, to remove the chaotropic agent. At every step, the dialysis was incubated at 4 °C for at least 12 hours.

  - Step 1: 50 mM Tris pH 7.6, 150 mM NaCl, 6 M urea
  - Step 2: 50 mM Tris pH 7.6, 150 mM NaCl, 3 M urea
  - Step 3: 50 mM Tris pH 7.6, 150 mM NaCl, 1.5 M urea, 400 mM L-arginine, 3 mM L-glutathione (GSH), 1 mM oxidizes L-glutathione (GSSG)
  - Step 4: 50 mM Tris pH 7.6, 150 mM NaCl

    After the last dialysis step, the protein solution was centrifuged at 13000×g at 4 °C for 10 minutes, to remove the precipitate.

- **Protocol B: Dilution**

  1 mL of solution of the unfolded protein at 0.3 mg/mL, in Urea 8M, was rapidly diluted using a Refolding Buffer B, containing 20 mM Tris pH 7.6, 5 mM L-glutathione (GSH), 1 mM oxidized L-glutathione (GSSG), and 400 mM L-arginine. The buffer solution was added to the protein solution at a rate of 2.5 mL/min, up to a total volume of 10 mL. Then, the sample was incubated overnight at 4 °C, under stirring. The urea concentration in the final solution is 0.8 M.

- **Protocol C: Reverse dilution**

  5 mL of unfolded protein solution was added to 150 mL of Refolding Buffer C, containing 50 mM Tris pH 8, 10% v/v glycerol, 0.5% v/v Triton X-100, 3 mM L-glutathione (GSH), 0.3 mM oxidizes L-glutathione (GSSG), and 400 mM sucrose. The protein solution was added in aliquots of 0.5 mL every 30 minutes, while the buffer was stirred. At the end, the solution was incubated overnight at 4 °C, under stirring.

After refolding, samples were centrifuged at 13000×g for 10 minutes to remove the precipitate. The protein was concentrated using an Amicon® Ultra Centrifugal Filter, 10 kDa MWCO (Millipore). To evaluate the efficiency of the refolding process, samples were treated with Laemmli loading buffer, either with or without β-mercaptoethanol, and separated by SDS-PAGE. This analysis aimed at identifying aggregates formed through intermolecular disulfide bridge formation, by comparing samples from the same refolding protocol treated with the reducing and non-reducing loading buffer.

Since Protocol C produced the highest yield of refolded protein, it was applied for the refolding of all anti-P105 VHHs produced in inclusion bodies. After refolding, protein solutions used for MST analysis were dialyzed against TBS, supplemented with 10% glycerol and 2 mM DTT.

3.2.6.4 Dot Blot

For further confirmation of the presence of the VHH::GFP proteins, 2 µL of each sample was applied onto a methanol-preactivated PVDF membrane, forming drops of 3-4 mm. Once the drops had dried, the membrane was treated for 2 hours with TBS containing 5% (w/v) milk powder. The membrane was subsequently washed three times with TBS containing 0.05% Tween 20 (TBST), shaking the solution for 5 minutes at each washing step, and then treated with an anti-His tag monoclonal antibody conjugated with HRP, as previously described (Paragraph 3.2.4).

3.2.6.5 GFP-tag removal

The eGFP-tagged samples were treated with TEV-6xHis protease at a 1:10 protease-to-substrate molar ratio and dialyzed for 24 hours at 4 °C in a buffer containing 50 mM Tris

pH 7.6, 0.5 mM EDTA, and 1 mM DTT. Successful cleavage of the eGFP tag was confirmed by SDS-PAGE analysis. The cleaved samples were further purified through Size Exclusion Chromatography, using a Superdex 75 10/300 column (Cytiva).

3.2.7 Design of HER2 mimetic peptides

VHHs were developed to specifically recognize P105, but not P100. A peptide corresponding to the C-terminal region of P105 (aa 627-647) was synthesized (Table 2), Peptide P105). To confirm that epitope recognition is influenced by the sequence downstream of P100, several control peptides were also prepared (Table 2). The C-terminal region of P105 contains multiple disulfide bridges, and one of these (C4-C16) was preserved in the synthesized peptides. The cysteine at position 8, involved in another disulfide bridge *in vivo* with a cysteine located upstream of the synthesized sequence, was mutated to alanine. A negative control (Peptide P100) was synthesized with the C-terminal sequence of P100. Another control (Negative Control) has the same sequence followed by PEG O2Oc linkers, whose structural unit covers 3 amino acid residues (shown as XXX), to reach the same length of Peptide P105. To retain a conformation similar to Peptide P105, cysteine residues forming the disulfide bridge and the proline 17 residue have been maintained in the Negative Control. To evaluate the influence of electrostatic interactions in the VHH target recognition, a peptide was synthesized with the same sequence of Peptide 105 but for the inversion of the charged residues (Peptide P105 RC).

Table 2: *Peptide sequences synthesized to evaluate selectivity and specificity of the anti-P105 VHHs. Disulfide bridges are marked with a line.*

Name	Sequence	Monoisotopic MW [Da]
Peptide P105	Ac-PINCTHS**A**VDLDDKGCPAEQR-NH2	2307.04
Peptide P105 RC (reverse charge control)	Ac-PINCTHS**A**VRLRRDGCPARQD-NH2	2403.18
Peptide P105 NC (negative control)	Ac-PINCTHS**A**XXXXXXGCPXXXA-NH2	1644.76
Peptide P100	Ac-PIN**A**THS-NH2	779.38

3.2.7.1 HER2 mimetic peptide synthesis

Peptide systhesis was accomplished in collaboration with Dr. Andrea Caporale form Istituto di Cristallografia IC-CNR, Trieste (Italy).

Solid-phase syntheses were performed in using an Initiator+ Alstra microwave peptide automatic synthesizer (Biotage). C-terminal amidated peptides were synthesized on a Fmoc-Rink-Amide-PEG200-HypoGel resin (HypoGel® 200 RAM, 0.56 mmol/g loading, 100 mg, from Iris Biotech GmbH, Marktredwitz, Germany). Oxyme/DIC and TBTU/HOBt/Dipea were employed as standard coupling agents for all syntheses (Table 3.). N-terminal residues were amidated. Synthetic procedures followed the protocol described by Caporale *et al*.[126] and are summarized in Table 4, 5, 6, and 7.

Table 3: *protocol of coupling/decoupling and reagents used during the synthesis.*

	Reagents dissolved in NMP/DMF							
a	Oxime 0.2 M	4 eq	DIC 0.2 M	4 eq			50°C x 5 min	RT x 10 min
b	HOBt 0.2 M	4 eq	TBTU 0.2 M	4 eq.	DIPEA 2M	8 eq	50°C x 5 min	RT x 10 min
a*	Oxime 0.2 M	2 eq	DIC 0.2 M	2 eq			50°C x 7 min	RT x 15 min
b*	HOBt 2.0 M	2 eq	TBTU 0.2 M	2 eq.	DIPEA 2M	4 eq	50°C x 7 min	RT x 15 min
Ac	Ac_2O 5 M	10 eq			DIPEA 2M	10 eq		RT x 10 min
x	Pip 20%						50° x 2 min	RT x 1 min

Table 4: *Peptide P105 synthesis.*

N°	AA	Fmoc-AA	Deprot.	1st coupling	2nd coupling	3rd coupling	Acetilation
21	R	Fmoc-Arg-(Pbf)-OH	x	a	b		
20	Q	Fmoc-Gln(trt)-OH	x	a	b		
19	E	Fmoc-Glu(tBu)-OH	x	a	b		
18	A	Fmoc-Ala-OH	x	a	b		
17	P	Fmoc-Pro-OH	x	a	b		
16	C	Fmoc-Cys(Trt)-OH	x	a	b		
15	G	Fmoc-Gly-OH	x	a	b		
14	K	Fmoc-Lys(Boc)-OH	x	a	b		
13	D	Fmoc-Asp(tBu)-OH	x	a	b		
12	D	Fmoc-Asp(tBu)-OH	x	a	b		
11	L	Fmoc-Leu-OH	x	a	b		

10	D	Fmoc-Asp(tBu)-OH	x	a	b		
9	V	Fmoc-Val-OH	x	a	b		
8	A	Fmoc-Ala-OH	x	a	b		
7	S	Fmoc-Ser(tBu)-OH	x	a	b		
6	H	Fmoc-His(Boc)-OH	x	a	b		
5	T	Fmoc-Thr(tBu)-OH	x	a	b		
4	C	Fmoc-Cys(Trt)-OH	x	a	b		
3	N	Fmoc-Asn(Trt)-OH	x	a	b		
2	I	Fmoc-Ile-OH	x	a	b		
1	P	Fmoc-Pro-OH	x	a	b	x	Ac

Table 5: *Peptide P105 RC synthesis.*

N°	AA	Fmoc-AA	Deprot.	1st coupling	2nd coupling	3rd coupling	Acetilation
21	D	Fmoc-Asp(tBu)-OH	x	a	b		
20	Q	Fmoc-Gln(trt)-OH	x	a	b		
19	R	Fmoc-Arg-(Pbf)-OH	x	a	b		
18	A	Fmoc-Ala-OH	x	a	b		
17	P	Fmoc-Pro-OH	x	a	b		
16	C	Fmoc-Cys(Trt)-OH	x	a	b		
15	G	Fmoc-Gly-OH	x	a	b		
14	D	Fmoc-Lys(Boc)-OH	x	a	b		
13	R	Fmoc-Arg-(Pbf)-OH	x	a	b		
12	R	Fmoc-Arg-(Pbf)-OH	x	a	b		
11	L	Fmoc-Leu-OH	x	a	b	x	
10	R	Fmoc-Arg-(Pbf)-OH	x	a	b		
9	V	Fmoc-Val-OH	x	a	b		
8	A	Fmoc-Ala-OH	x	a	b		
7	S	Fmoc-Ser(tBu)-OH	x	a	b		
6	H	Fmoc-His(Boc)-OH	x	a	b		
5	T	Fmoc-Thr(tBu)-OH	x	a	b		
4	C	Fmoc-Cys(Trt)-OH	x	a	b		
3	N	Fmoc-Asn(Trt)-OH	x	a	b		
2	I	Fmoc-Ile-OH	x	a	b		
1	P	Fmoc-Pro-OH	x	a	b	x	Ac

***Table 6**: Peptide P105 NC synthesis*

N°	AA/ PEG	Fmoc-AA/ PEG	Deprot.	1st coupling	2nd coupling	3rd coupling	Acetilation
15	A	Fmoc-Ala-OH	x	a	b		
14	O2Oc	Fmoc-O2Oc-OH	x	a*	b*		Ac
13	P	Fmoc-Pro-OH	x	a	b		
12	C	Fmoc-Cys(Trt)-OH	x	a	b		
11	G	Fmoc-Gly-OH	x	a	b		
10	O2Oc	Fmoc-O2Oc-OH	x	a*	b*		Ac
9	O2Oc	Fmoc-O2Oc-OH	x	a*	b*		Ac
8	A	Fmoc-Ala-OH	x	a	b		
7	S	Fmoc-Ser(tBu)-OH	x	a	b		
6	H	Fmoc-His(Boc)-OH	x	a	b		
5	T	Fmoc-Thr(tBu)-OH	x	a	b		
4	C	Fmoc-Cys(Trt)-OH	x	a	b		
3	N	Fmoc-Asn(Trt)-OH	x	a	b		
2	I	Fmoc-Ile-OH	x	a	b		
1	P	Fmoc-Pro-OH	x	a	b	x	Ac

***Table 7**: Peptide P100 synthesis*

N°	AA	Fmoc-AA	Deprot.	1st coupling	2nd coupling	3rd coupling	Acetilation
7	S	Fmoc-Ser(tBu)-OH	x	A	b		
6	H	Fmoc-His(Boc)-OH	x	A	b		
5	T	Fmoc-Thr(tBu)-OH	x	A	b		
4	A	Fmoc-Ala-OH	x	A	b		
3	N	Fmoc-Asn(Trt)-OH	x	A	b		
2	I	Fmoc-Ile-OH	x	A	b		
1	P	Fmoc-Pro-OH	x	A	b	x	Ac

After completing peptide synthesis, the resins were washed in DMF and swelled in DCM (20 min) and finally washed twice with DMC, and dried. Cleavage of the peptides was performed at room temperature by adding 1 mL of a freshly prepared TFA/TIS/H_2O solution

(95:2.5:2.5 v/v/v), with mild stirring for 4 hours. Immediately after, the solution was concentrated to approximately 500 µL under nitrogen flow. The crude product was precipitated in 40 mL of cold methyl tert-butyl ether, followed by centrifugation at 6500 rpm for 10 minutes to recover the cerude peptide. Finally, the pellet was dissolved in 25% CH_3CN and lyophilized.

3.2.7.2 Peptide purification

Crude peptides were purified via semipreparative HPLC using a Phenomenex Gemini column (250 x 10 mm) on an UltiMate™.3000 HPLC System (Thermo Fisher). A linear gradient of 5% to 35% CH_3CN in H_2O, containing 0.05% TFA, was applied over 17 minutes at a flow rate of 0.5 mL/min, following the purification by measuring the absorption at 214 nm, 254 nm and 280 nm. The target peptide was identified by mass spectrometry utilizing an LTQ XL™ Linear Ion Trap Mass Spectrometer (Thermo Scientific) with an ESI source. P100 peptide required no further purification steps.

Peptides containing disulfide bridges were lyophilized after the first purification step and then cyclized using the following protocol[127]: 2-3 mg of peptide were dissolved in 20 mL of a 200 mM ammonium acetate solution at pH 9.1, with 2% v/v DMSO and stirred at room temperature for two days. The pH of the solution was monitored and adjusted by adding NaOH. The reaction progress was tracked using HPLC and mass spectrometry. Peptide P105 cyclization was followed by LC-MS using a Phenomenex Luna C18 analytical column (5µm 100Å 150x2.1 mm) connected to a 6120 Single Quadrupole LC/MS System (Agilent). Samples were analyzed at the beginning of the reaction, after one day, and after two days. The same protocol was applied for the cyclization of the other peptides (Peptide P105 RC and Peptide P105 NC). A final purification step by HPLC using the gradient settings reported above yielded the cyclized and pure peptides.

4. RESULTS AND DISCUSSION

4.1 Case study: nanobodies against lysozyme

This section reports an experimental validation of the PARCE algorithm in the design of nanobodies able to interact with a specific target. In particular, the sequence of a VHH, originally targeting the breast cancer-associated receptor HER2, was mutated using the algorithm to recognize a completely different target: the active site pocket of the protein Hen Egg White Lysozyme (HEWL). This proof-of-concept study highlights the versatility of the PARCE algorithm in reprogramming the specificity of antibody fragments, allowing them to target a wide range of antigens beyond their original specificity.

4.1.1 Anti-lysozyme nanobody production

Starting from an anti-HER2 nanobody (VHH_0, Figure 19, upper panel) and the crystallographic model of the HEWL (PDB code: 1ZV5), *in silico* optimization generated a small library of anti-HEWL virtual nanobodies, containing 14-15 mutations in the Complementarity-Determining Regions (CDRs), compared to the starting nanobody: in *in silico* calculations, VHH_114, VHH_129, and VHH_135 showed, respectively, a moderate, high, and intermediate score in binding the new target (Figure 19, lower panel). The small size of the nanobodies (~15 kDa) and the absence of post-translational modifications allow *E. coli* to be used as a host for protein heterologous expression.

An initial expression and purification screening on the VHH library has been performed using a construct containing a GFP-tag at the C-terminus of the protein. Later, the same VHH sequences were recloned and produced as chimeras with a Cysteine tag (Cys-tag) in place of the fluorescent tag, as is known from the literature that the tag influences the affinity of the nanobody toward its target [12]. Both constructs contain a Histidine tag (6xHis), following the GFP-tag or the Cys-tag, for ease of purification.

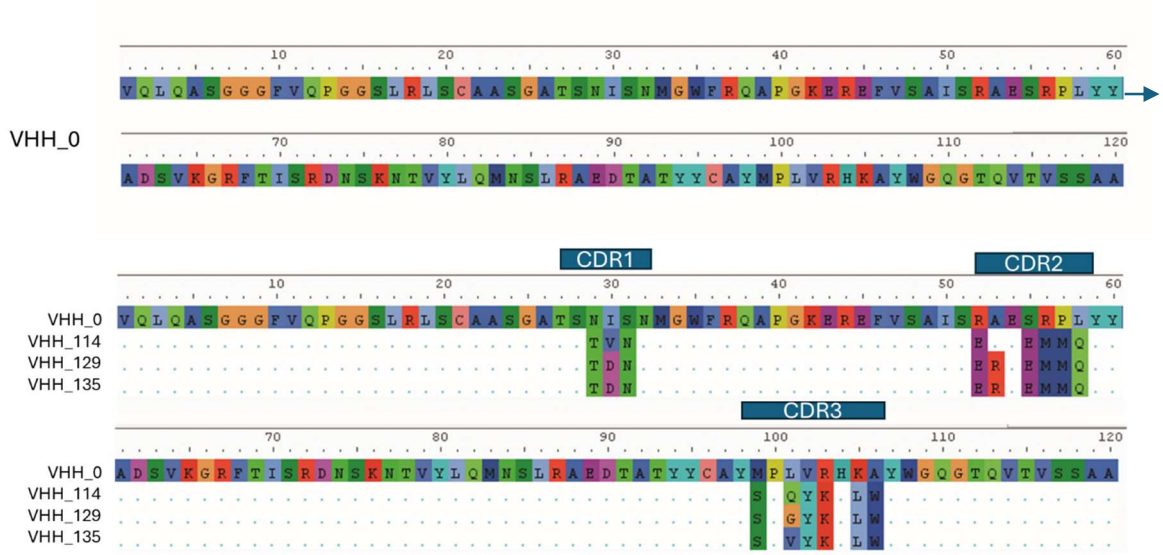


Figure 19: (Upper panel) Sequence of the nanobody used as input for the PARCE algorithm, VHH_0. *(Lower panel)* Alignment of the library of anti-lysozyme VHHs used in this study, in comparison with VHH_0. Mutations in the CDR segments are highlighted.

All VHHs have been expressed in good yield (more than 1 mg/L culture) and successfully purified: a well-defined main peak is visible in each of the chromatograms (Figure 20.A). SDS-PAGE analysis of mutants VHH_129 and VHH_135 revealed that, after the last SEC purification step, the samples did not show the presence of any contaminant (Figure 20.B)). Peaks appearing at lower retention volumes contain contaminants or protein aggregates and were therefore discarded. Comparing the chromatograms from VHH purifications, the predominant peak shows a retention volume that varies considerably depending on the VHH sequence: 12.6, 17.5, 14.3, and 13.0 mL for VHH_114, VHH_129, VHH_135, and VHH_0, respectively. The retention volume of VHH_129, in particular, is very high. Although they have roughly the same molecular mass and, presumably, the same size, the VHHs have a markedly different behavior during the chromatographic run. However, it is not uncommon for artificially developed nanobodies to display a non-specific affinity for resins used in chromatography[128], which can justify the increased retention volume of VHH_129. Anyway, it was possible to unquestionably identify the VHHs via SDS-PAGE analysis (Figure 20).

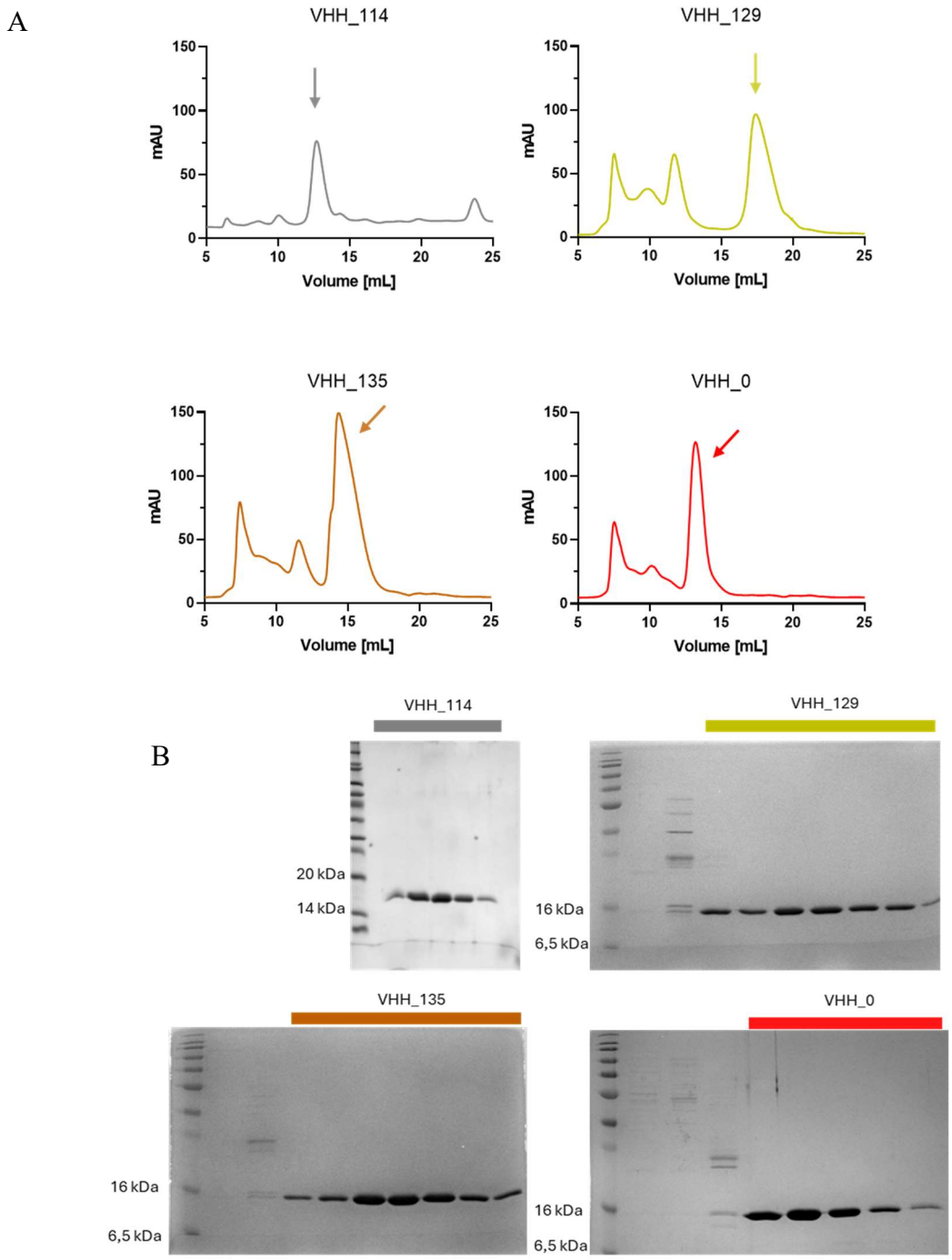


Figure 20 **Anti-lysozyme VHH purification.** (A) Typical results of Size Exclusion Chromatography purifications of the anti-HEWL library of VHHs and of the VHH_0 nanobody, reported for comparison. VHH_114, VHH_129, VHH_135, and VHH_0 are grey, yellow, orange, and red, respectively. (B) Images of SDS-PAGE analyses of fractions eluted from the SEC purification. The main peak fractions are marked with a colored rectangle.

4.1.2 Computational designed VHHs bind their target

MST experiments aimed at evaluating the ability of VHH_114, VHH_129, and VHH_135 to recognize HEWL. The nanobody constructs containing a His- and Cys-tag were treated with a fluorescent dye. In the MST experiments, the ligand HEWL was used as a titrant. The non-optimized ancestor VHH_0 was employed as a negative control, while VHH D2-L29, described in the literature to have a strong affinity for lysozyme[120], was used as a positive control (CTRL+). VHH_129 and VHH_135 provided dose-response curves of MST signals, as reported in Figure 21, but with different K_D values: for VHH_129, a value of 60 nM was registered, while VHH_135 showed a slightly weaker binding affinity, with a K_D of 80 nM. However, the two values are within the margin of error. No binding was observed for VHH_114. The fitting of experimental data for the positive control allowed an estimation of K_D of 14 nM, close to the value reported in the literature[120]. A micromolar affinity was registered for the negative control VHH_0 (K_D of 5 μM).

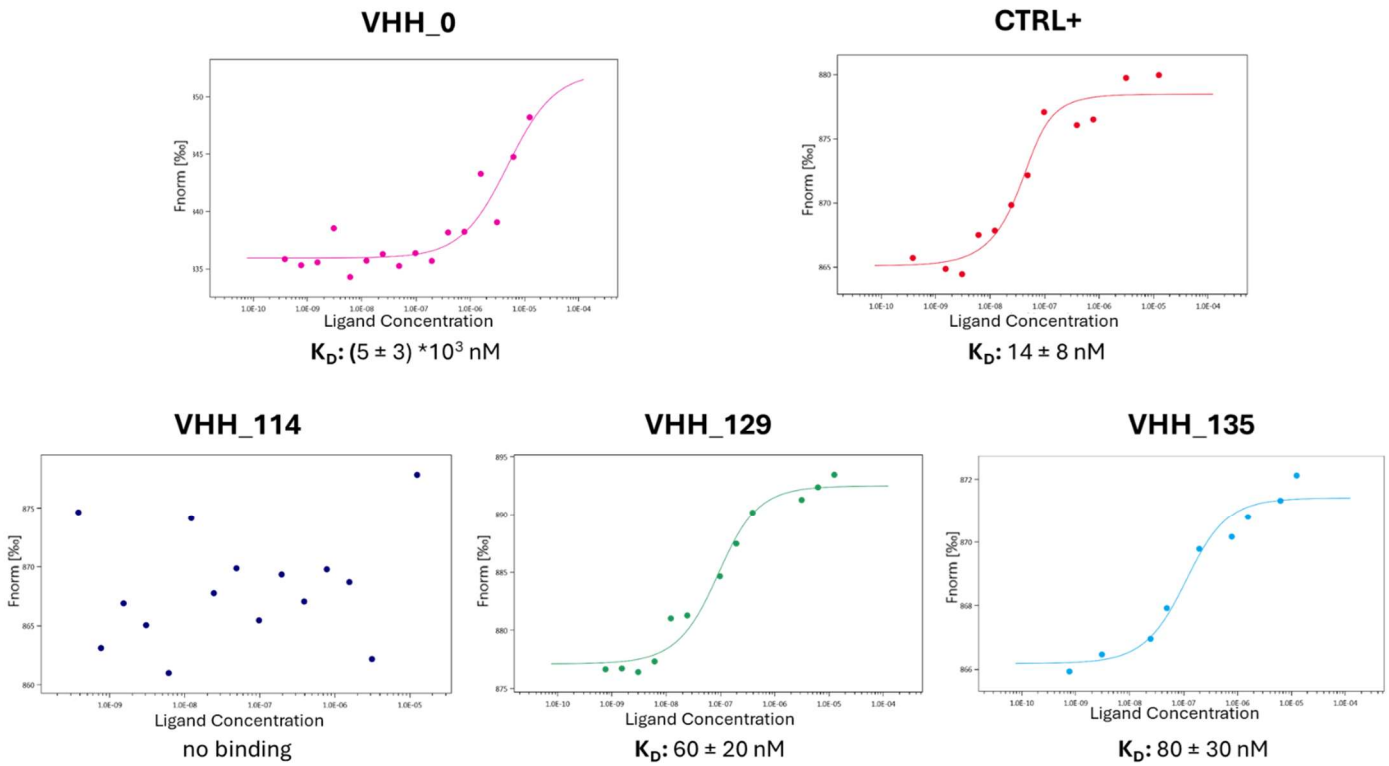


Figure 21: **MST experiments.** For each of the VHH analyzed, the graph reports the normalized fluorescence (Fnorm) against the logarithm of the concentration. By fitting the data, the dissociation constant (K_D) was determined for each sample and is reported below the graph.

Although VHH_0 binds to the new target with low affinity, the K_D values of two of the optimized VHHs are significantly lower compared to their progenitor, indicating that the PARCE optimization process achieved the intended outcome. Moreover, the binding constants align with the computational predictions: VHH_129 proves to be a slightly stronger binder, followed by VHH_135. VHH_114, identified as the weakest of the three in the computational analysis, did not bind the target in the MST experiments. The micromolar binding affinity of the negative control VHH_0 suggests a non-specific binding for this nanobody.

4.1.3 Lysozyme binding of VHH_129 and VHH_135 is pH-dependent

The nanobody optimization process of the PARCE algorithm introduces several mutations in the ancestor VHH_0, each evaluated against the previous iteration and retained only if an improvement in the binding affinity is observed. For each VHH mutant produced by the algorithm, the software outputs a model of the nanobody docked to the specific portion of the target protein previously selected, together with the values of the docking scores. A manual analysis of the models was performed to identify the main interactions responsible for the binding to the target.

In the models of VHH_129 and VHH_135, an interaction between the His 104 residue of CDR3 of the nanobody and the Trp 62 residue of HEWL was observed (Figure 22 A and B). Interestingly, this residue is conserved from the VHH_0 sequence (Figure 22 C). The docking model of the progenitor, however, shows that the histidine side chain is in a completely different position compared to the optimized VHHs, due to the significant differences in the protein sequence preceding and following residue 104 in the three VHHs, a region where mutations accumulated during *in silico* optimization.

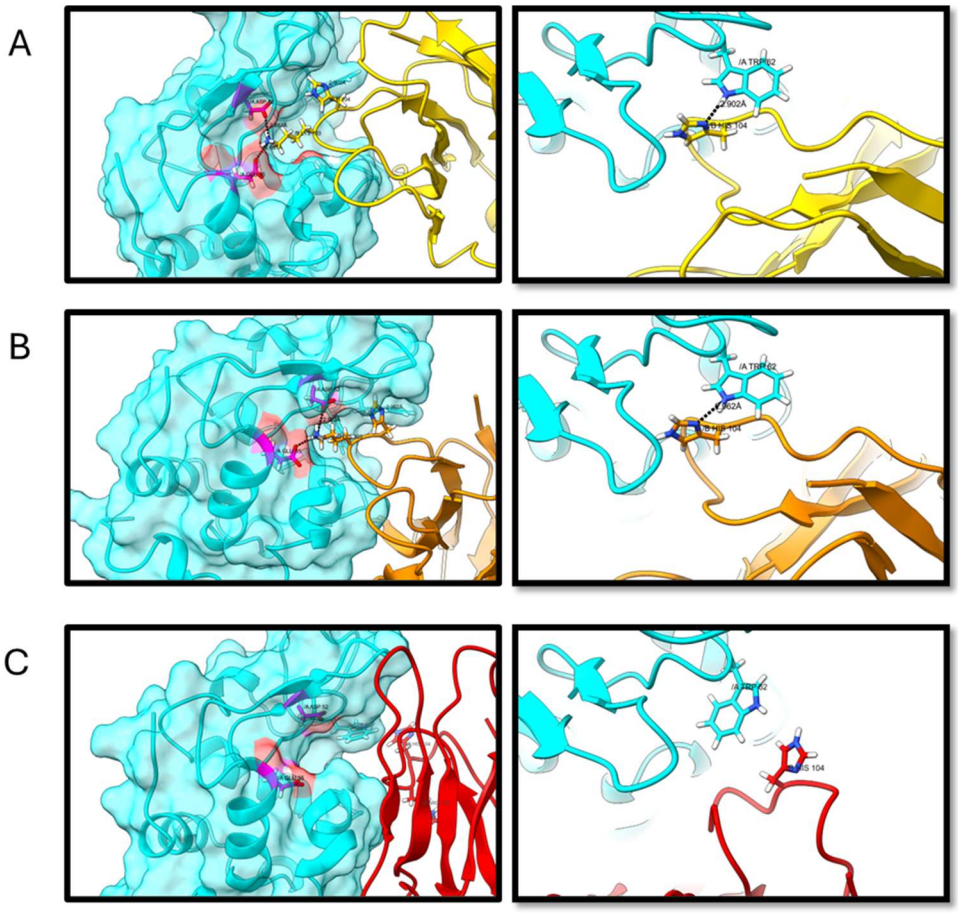


*Figure 22:***The role of His 104 in target binding**. Final MD simulations suggest that a primary role in the VHH_129 (A) and VHH_135 (B) binding to lysozyme is played by the interaction between the side chain of residue His 104 of the VHHs and a Trp residue of HEWL binding site. In the simulation of VHH_0 (C), both the His and the Trp show a different orientation, unsuitable for the formation of a hydrogen bonding interaction. Left panels: lysozyme active pocket. Right panels: focus on His104 from VHH and Trp62 from HEWL. Nanobodies VHH_129, VHH_135, and VHH_0 are represented in yellow, orange, and red, respectively, while the lysozyme is cyan. Dashed lines represent H-bond interactions, with the relative donor-to-acceptor distance. Residue His 104 of the nanobodies and Trp 62 of HEWL are represented as sticks.

While the final MD simulations were carried out using models of the nanobodies at physiological pH (7.4), in which the histidine 104 residues are in the deprotonated state, a significant change in the strength of this interaction is expected when the pH of the solution is decreased by changing the buffer. The H-bonding interaction is possible only when the histidine side chain is deprotonated, i.e., when it can act as a hydrogen bond acceptor. Considering the histidine theoretical pKa at 25 °C (6.04), we decided to investigate the binding of these nanobodies at a pH 8.0 and 5.8.

First, the thermal stability of the nanobodies was investigated in the selected buffer solutions, using the Thermal Shift Analysis method (TSA)[118]. TSA allows a fast assessment

of the thermal stability of the VHH mutants in different buffer conditions. The TSA employs a fluorescent dye to monitor protein unfolding in response to temperature changes. The dye binds to hydrophobic moieties thereby increasing the fluorescence signal. During protein unfolding, hydrophobic residues buried in the core of soluble proteins become exposed to the solvent, resulting in an increase in fluorescence. By plotting the fluorescence signal against temperature, a thermal unfolding curve is obtained. The inflection point on this curve, also known as the unfolding temperature (T_u), represents the temperature at which 50% of the protein is unfolded. The first derivative of the unfolding curve is often reported, in which the minimum corresponds to the unfolding temperature. Shifts in T_u suggest alterations in protein stability arising from mutations, environmental influences, or protein binding.

The TSA experiment performed on the VHH mutants and on the progenitor VHH_0 show that thermal stabilities of these proteins are generally lower at pH 8.0 compared to pH 5.8 (Figure Fig. 7, solid lines). However, while the unfolding temperatures of VHH_0 and the mutant VHH_135 at pH 8 are comparable (49.0 ± 0.50 °C and 49.0 ± 0.58 °C, respectively), VHH_129 shows a lower unfolding temperature (44.0 ± 0.39 °C), indicative of a lower protein stability. At pH 5.8, VHH_129 and VHH_135 are more stable, with unfolding temperatures of 52 ± 0.5 °C and 56 ± 0.29 °C, respectively.

In another set of TSA experiments, lysozyme binding was investigated using the same assay. The unfolding temperature of a protein might be affected by protein-protein or protein-ligand binding. The TSA technique is often employed in drug discovery to identify molecular interactions[129]. TSA experiments were conducted on solutions of the nanobodies and their target HEWL at pH 8. In the presence of lysozyme, the thermal unfolding of the VHHs at pH 8 occurred at (44.0 ± 0.39 °C, 49.0 ° ± 0.58 °C and 49.0 °C ± 0.50 °C for VHH_129, VHH_135, and VHH_0, respectively (Figure 23, dashed lines), a significant decrease from the unfolding temperatures of the unbound VHHs. Considering the possibility that the registered increase in fluorescence at lower temperatures was indicative of the unfolding of lysozyme, the same experiment was conducted at pH 8 on a lysozyme sample at the same concentration (10 μM). The curve, reported in Figure 23 as a solid light blue line, shows a higher unfolding temperature (69 °C) and a lower fluorescence value, probably due to the lower number of hydrophobic residues in the core of HEWL. At pH 5.8, on the other hand, unfolding temperatures registered in the presence of HEWL remain similar to those obtained for VHH solutions: 52 °C ± 0.5 °C for VHH_129/HEWL and 56 ± 0.29 °C for VHH_135/HEWL (Figure 23, right side).

The unfolding temperature shifts observed for HEWL-VHH solutions at pH 8 suggest a destabilizing effect due to the interaction of the nanobodies with the epitope. However, at pH 5.8, the presence of lysozyme does not significantly affect the unfolding temperature and the stability of the nanobodies. From these results, we infer that the nanobody-lysozyme binding occurs only at higher pH, with a destabilizing effect on the nanobody folding. The unchanged thermal behavior at pH 5.8 is consistent with the crucial role of the deprotonated histidine residue in position 104 in binding to the target, as predicted by docking analyses and molecular dynamics simulations.

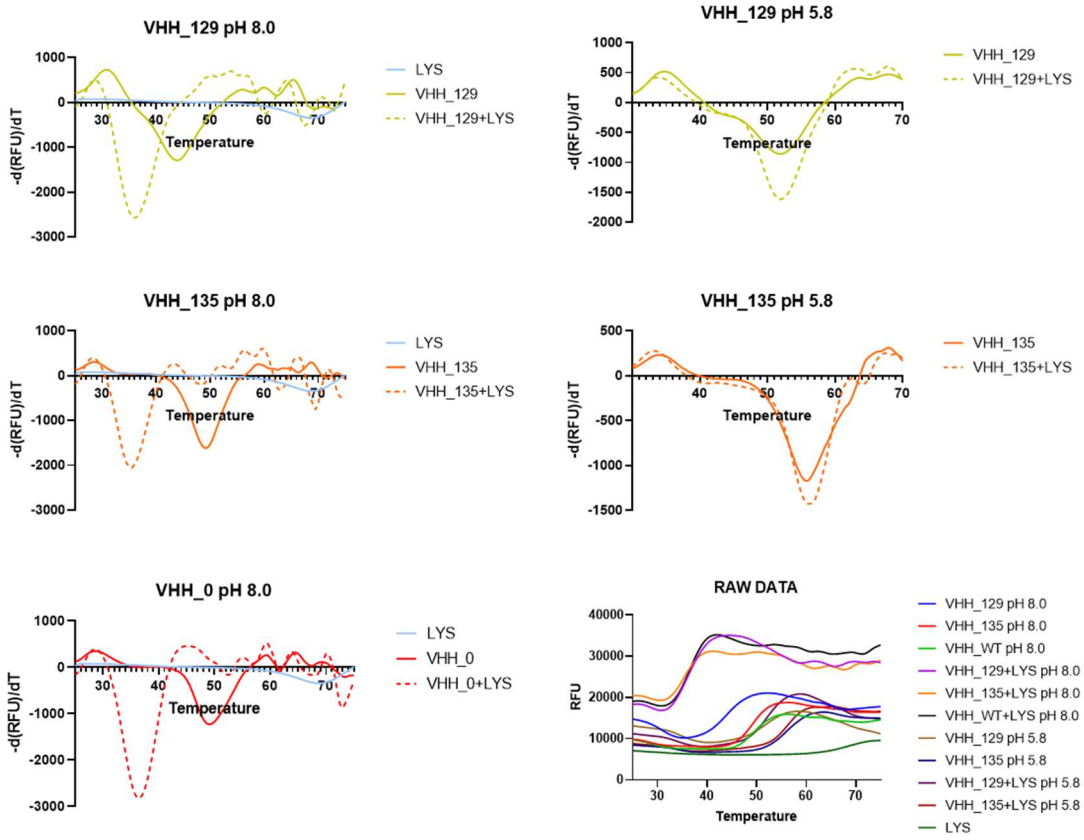


Figure 23: **Thermal shift assays**: VHH unfolding temperatures were observed in two different conditions, in the presence or absence of lysozyme. On the left side, results of the TSA experiments at pH 8.0 for VHH_129 (yellow), VHH_135 (orange), and VHH_0 (red) are shown, in comparison with the curve of a HEWL sample (light blue). On the right side, results of the optimized VHHs are reported at pH 5.8. Solid lines represent samples containing only one of the binding partners, and dashed lines represent solutions where VHHs and HEWL are mixed. The results are reported as the first derivative of fluorescence versus temperature. In these derivative plots, the T_u represents the minimum of the curves. Also raw data polt, representing fluorescence versus temperature, is shown.

4.1.4 Lysozyme binding of VHH_129 and VHH_135 occurs at the enzyme active site

A significant advantage of the computational design of nanobodies lies in the possibility of precisely defining their binding site on the target, thus reducing the risk of cross-reactivity and obtaining nanobodies that are highly specific against protein isoforms differing for a limited number of residues. This result, hardly achieved through immunization methods, is extremely valuable in contexts spanning from medical applications to crystallographic studies aimed at a specific protein conformation.

In the present proof-of-concept study, the selected region of the lysozyme target involves a surface close to the active site of the enzyme. At first, a structural approach was attempted to validate the computational predictions regarding VHH-lysozyme binding. Crystallization experiments of the VHH-lysozyme complex were set up using vapor diffusion techniques. However, for both VHH_129 and VHH_135, the result was the same: only lysozyme crystal could be obtained from a wide range of crystallization conditions. It is likely that the nucleation and growth kinetics of lysozyme crystals, faster than those of the complex, lead to a gradual removal of lysozyme from the solution. This, in turn, moves the equilibrium of the VHH-HEWL binding reaction towards the unbound form, decreasing the concentration of the complex in solution.

Thus, a biochemical approach was considered. Based on computational models, the VHH binding site on the enzyme is expected to cause a decrease in enzyme activity due to the steric hindrance of the VHH that hampers the access of the substrate to the catalytic pocket. To verify this prediction, we combined two different methods described in the literature: binding of the VHH is detected by measuring the residual activity of lysozyme as previously reported in the case of carbonic anhydrase and α-amylase[130] , and the activity assay is conducted on a bacterial suspension, a method reported by Salton *et al.* for the assessment of lysozyme activity[131].

Lysozyme, a common enzyme with antimicrobial proprieties, catalyzes the hydrolysis of the β-(1,4) linkage between the N-acetylglucosamine and the N-acetylmuramic acid leading to the destruction of the bacterial cell wall. Its antibacterial activity is particularly strong against Gram-positive bacteria. The turbidimetric assay exploits *M. luteus*, a Gram-positive bacterium, as a lysozyme substrate. The enzyme is active over a broad pH range, with a maximal activity at pH 6.2. In the activity assays, VHH-lysozyme solutions were added to small amounts of *M. luteus* cultures, and the decrease in turbidity of the culture was followed using a spectrophotometer for 5 minutes after the addition of the

enzyme solution. For each test, lysozyme activity per mL of culture was estimated as Enzyme Units (EU), considering that one EU is the amount of enzyme that causes a 0.001 decrease per minute in the optical density of the cell culture, compared with a cell culture without lysozyme (black curves in Figure 24). A lysozyme solution without VHHs was used as a negative control. The VHH D2-L29[120] was used as a positive control (CTRL+) since his interaction with HEWL, biophysically and structurally thoroughly characterized, involves the enzyme active site. However, assays demonstrating its inhibitory effect on HEWL have not been previously reported in the literature.

At pH 8.0, lysozyme activity is moderate (6.7 ± 0.4 kEU/mL, light blue curve in the upper panel of Figure 24). When the *M. luteus* cultures and the lysozyme-containing solutions were buffered at pH 8, all examined VHHs demonstrated a negative impact on enzyme activity, albeit with different intensities, indicating different degrees of inhibition of the enzymatic function (Figure 24, upper panel). While VHH_129 and VHH_135 had a comparable outcome (green and yellow curves in the upper panel of Figure 24, respectively; 3.7 ± 0.2 kEU/mL and 3.9 ± 0.4 kEU/mL, respectively), VHH_0 showed a lower inhibitory effect (red curve in the upper panel of Figure 24; 5.5 ± 0.3 kEU/mL), suggesting that it recognizes the active site of HEWL, but the binding is weak. The addition of the strong binder D2-L29 has the effect of suppressing completely the activity of lysozyme (blue curve in the upper panel of Figure 24).

To confirm the effect of pH on lysozyme binding, the same assay was repeated in a solution buffered at pH 5.8, where lysozyme activity is significantly higher (light blue curve in the lower panel of Figure 24, 59 ± 2 kEU/mL) compared to pH 8.0. Although in this condition VHH_129 and VHH_135 are more stable, as demonstrated by TSA experiments, their presence does not influence the action of the lysozyme on the substrate, as evidenced by the green and yellow curves, respectively, in the lower panel of Figure 24 and by the calculated activities of 59 ± 1 kEU/mL and 61 ± 1 kEU/mL. As expected, the weaker binder VHH_0 is also ineffective in inhibiting the HEWL (red curve in the lower panel of Figure 24, 62 ± 1 kEU/mL). On the contrary, the inhibitory effect of D2-L29 is still detectable at pH 5.8 (blue curve in the upper panel of Figure 24, 39 ± 2 kEU/mL), even though to a lower extent than at pH 8.0. Having ruled out the possibility of lower stability of the nanobodies in this buffer condition, we propose that the lack of inhibition observed with VHH_129 and VHH_135 is an effect of the absence of binding under these specific conditions, confirming results from the TSA assays.

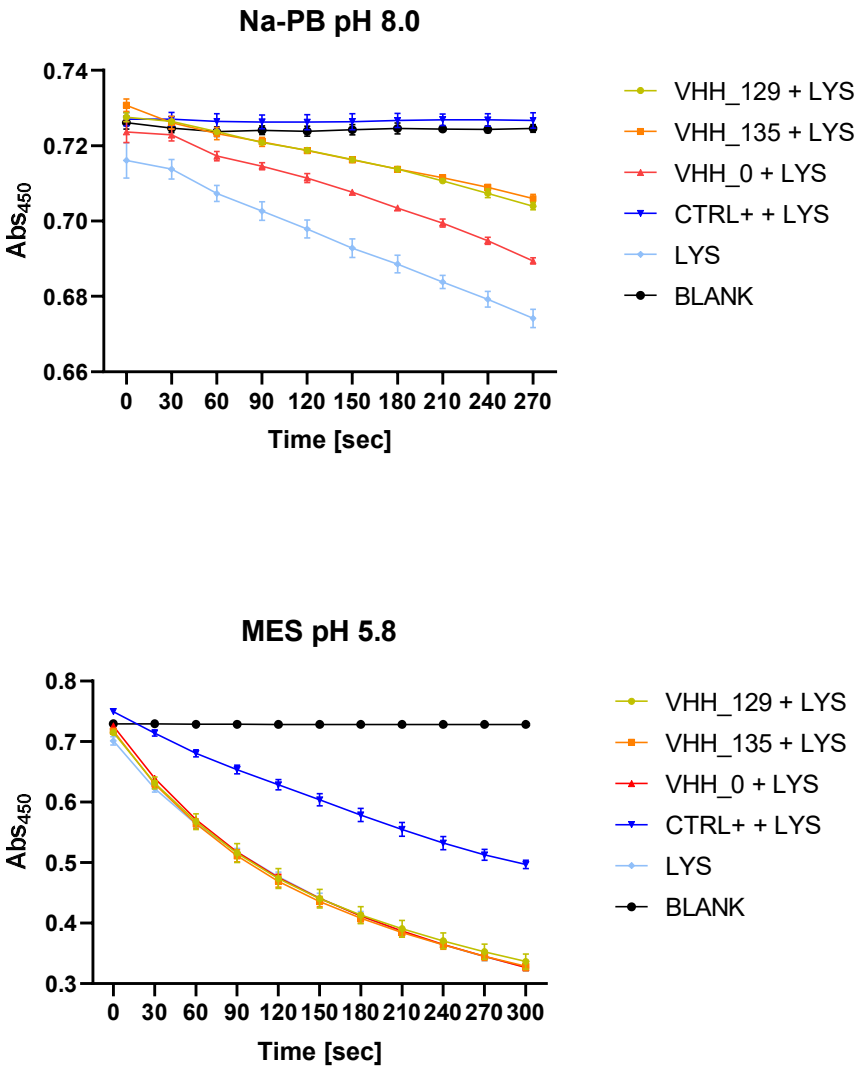


Figure 24: **Effect of the VHHs on the lysozyme activity**. Lysozyme activity assays in the presence and absence of VHHs at pH 8 (upper panel) and 5.8 (lower panel). The curves represent the optical density of the culture after the addition of lysozyme (t = 0). The black curves represent blanks, in which only the buffer was added to the cell suspension. Cyan curves represent the negative control, obtained by adding lysozyme to the culture. Blue curves were used as positive control and recorded in experiments in which the VHH D2-L29 was added to the culture. Curves green, yellow, and red show the inhibitory effect of VHH_129, VHH_135, and VHH_0, respectively.

4.2 Application: recognition of the C-terminal region of P105

In the former section, a case study demonstrated that epitope-specific VHHs can be generated de novo by the PARCE algorithm. This section explores the potential application of this computational approach within the biomedical field.

A nanobody library whose sequences were identified by the BINDEsignER/PARCE algorithm as selectively targeting the P105 fragment of HER2. This fragment, produced *in vivo* through proteolytic cleavage from the whole-length receptor, differs from the P100 isoform of HER2 only by 14 residues located at the C-terminus of the proteins. Such a small difference in protein sequence is difficult to target with the traditional methods for VHH production but has a significant impact on tumor development in HER2+ breast cancers. By precisely defining the binding surface on the protein target, the computational approach may be ideal for developing VHHs binding selectively to P105 without cross-reactivity with P100.

The VHH library targeting P105 was produced via heterologous expression in an *E. coli* bacterial system. A minimalistic version of the selected target was synthesized through Solid Phase Peptide Synthesis (SPPS). Although the work is still in progress, the preliminary essential steps for the VHH library characterization are here reported.

4.2.1 Theoretical assessment of VHH solubility and stability

In order to have a starting point for the VHH optimization, a model of the initial VHH-HER2 complex was generated computationally through a Molecular Dynamics (MD) approach. From the MD simulation, 3 different poses were selected, at 100 ns, 200 ns, and 300 ns, respectively. Starting from these poses, the PARCE protocol generated a library of 18 optimized VHH sequences in three separate runs. After another 500 ns MD simulation performed on the VHH-P105 complex and followed by contact analysis, 10 VHHs were chosen for experimental validation based on their higher affinity for the C-terminal region of P105 compared to the others. Two additional VHHs were selected for their affinity towards P100, as a comparison.

The PARCE protocol used to produce optimized VHHs focuses on enhancing the affinity of the nanobody for the selected target and the stability of the nanobody-target complex, but does not account for the solubility and stability of the single VHH proteins, a crucial factor for successful experimental validation. Low protein solubility and stability

significantly hamper bacterial expression and, even for VHH successfully expressed, can lead to aggregation phenomena under typical storage conditions. Three interconnected factors influence the formation of insoluble and non-functional protein aggregates: the primary sequence of the polypeptide chain, particularly as regards biophysical properties of the residues such as hydrophobicity, charge, and preferences for specific secondary structures; the native conformation of the protein and the exposure to the solvent of hydrophobic or uncharged amino acid residues on its surface; and the thermodynamic stability to denaturation of the native state of the protein. For poorly stable proteins, conformational fluctuations increase the probability of exposure of hydrophobic residues on the protein surface, thus facilitating aggregation. Although the scaffold of all mutants generated by the algorithm is that of a stable nanobody previously used in other studies, the large number of mutations on solvent-exposed regions, the CDRs, is likely to alter significantly the properties of the VHHs.

A preliminary analysis of the solubility of the optimized VHHs was conducted using the ExPASy ProtParam tool (https://web.expasy.org/protparam/)[132]. This software analyzes the protein primary sequences and provides assessments of the hydrophobicity and hydrophilicity of the residues and overall protein stability (Table 8).

In this analysis, the Instability Index (II) provides a quantitative assessment of the stability of the protein under examination. The term "instability" describes the propensity to degrade through both unfolding or the introduction of mutations in critical residues during protein recombinant expression. A protein with a predicted instability index lower than 40 can be considered stable, while when the instability index is higher than 40 the protein is considered unstable. The ProtParam software provides also the hydrophobicity index of the protein sequence, or GRAVY (Grand Average of Hydropathy) index, calculated as the average of the hydropathy values of all residues. All the VHHs exhibit negative values, according to the GRAVY index, indicating a predominance of hydrophilic residues. On the other hand, the instability index indicates that the VHHs from the 100 group, are unstable, whereas the 200 and 300 groups show stability comparable or superior to C8WT.

***Table 8**: Results of the analysis on the VHH protein sequences conducted with the ExPASy ProtParam tool.*

VHH	GRAVY	*Instability Index (II)*
C8_WT	- 0.269	37.75
100_53	- 0.281	41.06
100_72	- 0.413	40.77
100_85	- 0.441	42.29
100_90	-0.409	46.18
100_178	- 0.339	40.18
200_108	- 0.359	34.99
200_200	- 0.298	38.28
300_93	- 0.355	34.23
300_121	- 0.237	26.55

A more thorough analysis of the optimized VHHs employed the CamSol Structurally Correlated software[125] This tool allows for an estimation of the protein solubility considering not only the primary amino acid sequence but also the three-dimensional structure and the pH of the solution to which the protein is exposed. This approach provides a more comprehensive solubility assessment compared to methods that rely only on the sequence, such as the ExPASy ProtParam tool.

A comparison of the predicted solubility profiles of the VHH structures (Figure 9) reveals that C8WT contains a small number of poorly soluble residues, primarily located in the N-terminal region, which is a feature shared by all mutants. Overall, starting from the C8WT progenitor, the computational evolution of VHHs favored mutations that introduce soluble or highly soluble residues in the CDR3. However, for the other CDRs differences are encountered in VHH sequences produced by the three independent runs of the PARCE algorithm. In the first set of VHHs (designated with the code 100), the first two mutants, 100_53 and 100_72 (a VHH selected as the anti-P100 control and an underperforming P105 binder, respectively) exhibit poorly soluble residues in the CDR1, whereas for the other VHHs of this group highly soluble residues are present in the same region. A mix of moderate-to-poorly soluble and highly soluble residues is present in the CDR2 for all VHHs generated in the first run. In the 200 series, a common trait is the presence of a poorly soluble CDR2. VHH 200_108 has a relatively soluble CDR1, while 200_200 shows poor solubility in this region. The 300 mutants display poor solubility in both CDR1 and CDR2.

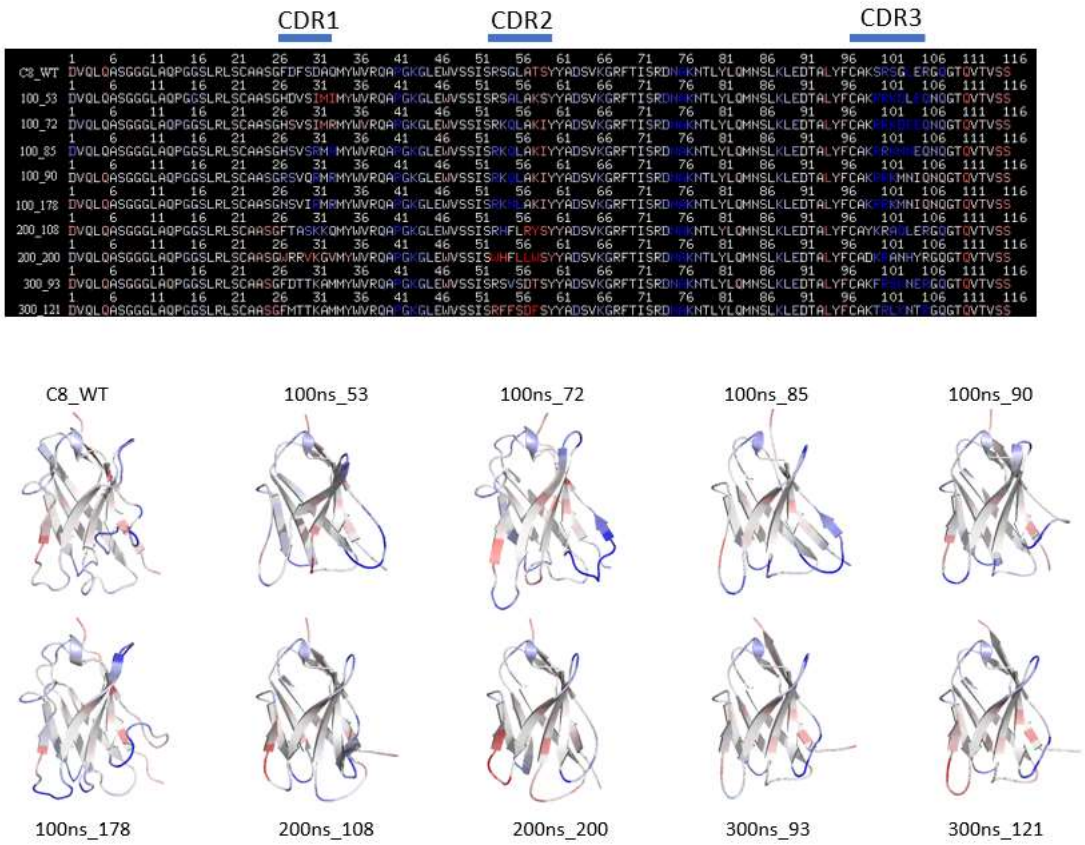


Figure 25: Prediction of the solubility of the C8WT VHH, the anti-P100 (100_53 and 100_72), and the anti-P105 VHHs (100_85, 100_90, 100_178, 200_108, 200_200, 300_93, and 300_121) provided by the CamSol Structurally Corrected software. The color coding follows a palette in which residues predicted to decrease VHH solubility are colored in red, while residues predicted to be highly soluble are colored in blue. *(Upper panel)* Primary sequence alignment. *(Lower panel)* Structural models used by the software for solubility predictions.

Preliminary analyses conducted on either the primary sequence or the three-dimensional models generated during the VHH optimization process do not account for the kinetics and thermodynamics of protein folding in a cellular environment. Additionally, in the following experimental work, all VHHs were expressed as fusion proteins with an eGFP tag. The eGFP domain, with a molecular weight approximately twice that of the VHH, is expected to play a significant role in enhancing protein solubility.

4.2.2 Gene cloning

During the gene cloning step, nine of the ten optimized VHHs were successfully inserted into the expression vector (100_85, 100_90, 100_167, 100_178, 200_108, 200_173, 200_200, 300_93, 300_121 Figure 26), along with one putative P100 binder (100_53). In addition, a control was obtained by cloning a VHH sequence discarded from the pool of top binders (100_72). The original nanobody used as the starting point for PARCE optimization,

C8WT, was cloned as well. Each optimized VHH contains 12-18 mutations distributed across all CDRs, compared to C8_WT.

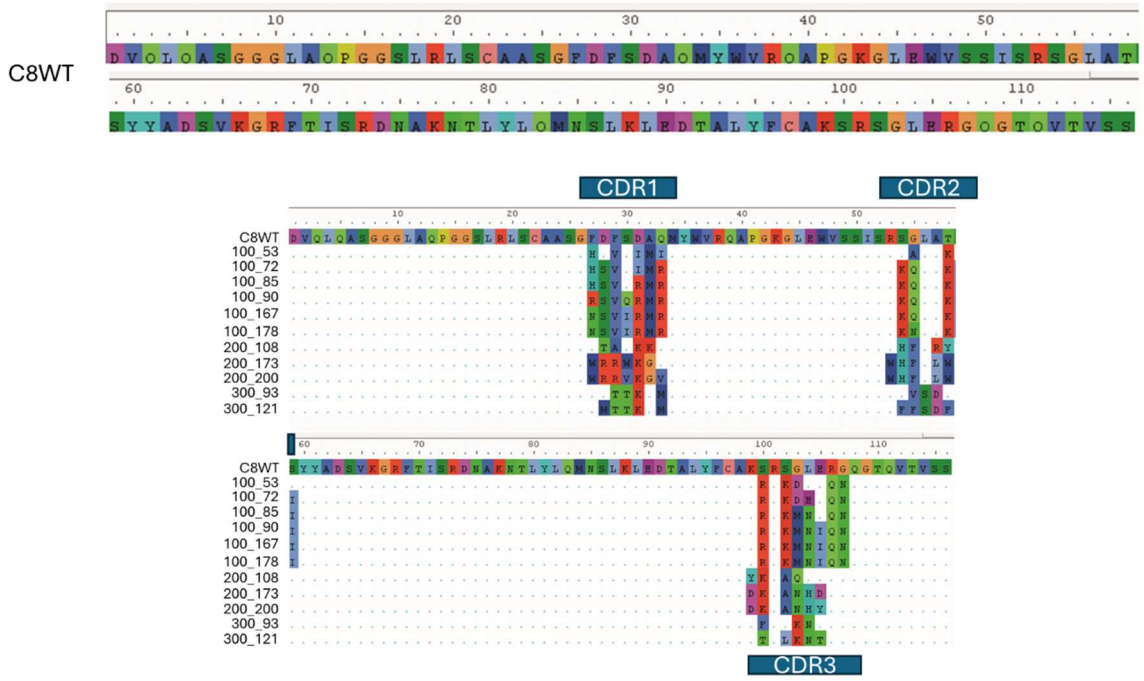


Figure 26: (Upper panel) Sequence of the nanobody used as input for the PARCE algorithm, C8WT. *(Lower panel)* Alignment of the library of anti-P105 VHHs used in this study, in comparison with C8WT. Mutations in the CDR segments are highlighted.

4.2.3 Differences in *E. coli* expression and Inclusion Body formation among the VHHs

Small-scale protein expression experiments were conducted in parallel for all the VHH successfully cloned. Several factors were considered to evaluate the results of these experiments. The cell growth assessment reveals notable differences in the expression behavior between the construct of the progenitor C8WT and the computationally optimized mutants.

Considering that the nanobodies were expressed as fusion proteins with a GFP domain, measuring the fluorescence of the cultures allows the estimation of the expression yield for each nanobody. The fluorescence intensity of cell cultures expressing C8WT was consistently higher than that of cultures expressing any of the mutants (Figure 27, left panel, bright green bar). However, cultures expressing the mutants show fluorescence signals that were still higher than the signal registered for the culture of the untransformed negative

control (Figure 27 A, light yellow bar), indicating a successful expression of the fluorescent protein in all cases.

Additionally, the optical density of the culture at the end of the expression period (16h) was considered. In recombinant protein expression, after the addition of the inducing agent, the optical density of the culture is expected to increase at a lower rate than normal, or even to decrease in the case of toxic proteins. Measurements taken at the end of the expression phase show that the non-induced control has a higher density than the transformed cultures (Figure 27 B, light yellow bar). This suggests that the recombinant protein expression takes up a considerable part of the cell's resources and slows cell division. Small differences can be observed in the optical density of the cultures expressing the VHHs at the end of the expression phase.

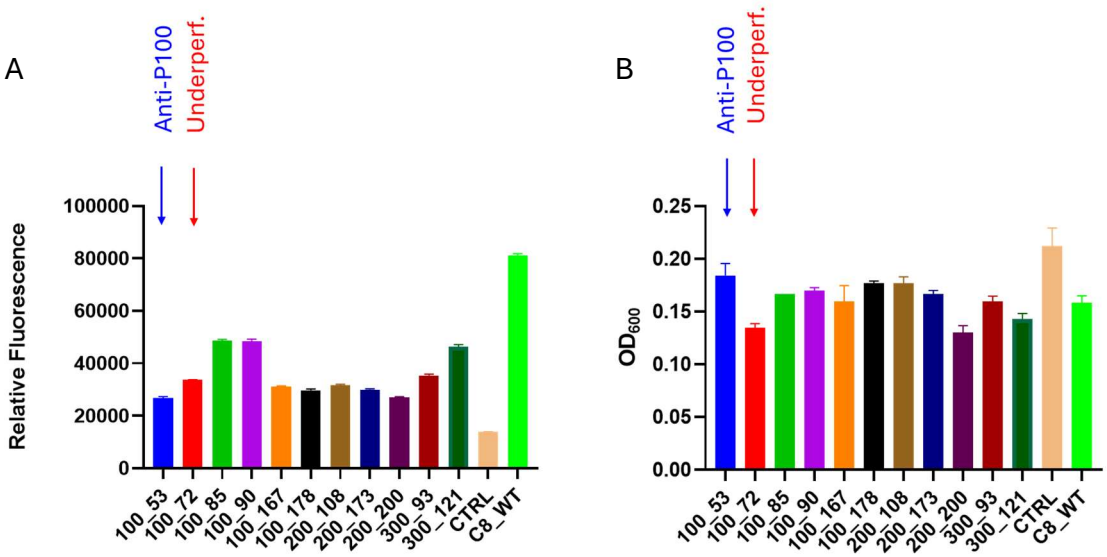


Figure 27: **Pre-lysis analyses of small-scale cultures**: A) Relative fluorescence of the cell culture measured at the end of the expression phase; B) Optical density at λ=600 nm, measured at the end of the expression phase on 10x diluted samples of cell cultures. The negative control (CTRL) represents a non-transformed BL21(DE3) SOX cell culture induced with 0.5% arabinose.

After cell lysis, the resulting solution, known as the total extract, was centrifuged to separate the fraction containing soluble proteins from the insoluble fraction. The insoluble fraction was treated with urea, a chaotropic agent known to aid the solubilization of insoluble proteins or protein aggregates. The insoluble fraction of the cell lysate contains many different cell components, including membranes and membrane proteins. When the expression of a poorly soluble or toxic heterologous protein is induced in bacteria, it often results in aggregates of

unfolded protein, known as Inclusion Bodies (IBs). From the cell lysate, IBs are separated from the soluble fraction and can be recovered in the insoluble fraction by solubilizing it with a chaotropic agent. The soluble and urea-solubilized fractions were analyzed by SDS-PAGE and Western Blot analyses to assess protein expression and solubility (Figure 28).

Profound differences in expression yield and solubility between C8WT and the computationally optimized VHHs are visible in the electrophoretic analyses, even though the small-scale tests were conducted with identical expression hosts, growth conditions, and induction protocols for all VHHs. In particular, SDS-PAGE analysis of the soluble fraction shows a band at the expected molecular weight (42 kDa) for all the samples (Figure 28, panel A), but the subsequent Western Blot analysis with anti-His antibody reveals that only in the case of C8WT this band corresponds to the VHH-GFP-6His construct (Figure 28, panel C). However, analysis of the urea-solubilized fraction shows bands at 42 kDa for all samples except C8WT, albeit with different intensities (Figure 28, panel B). The anti-His antibody identification in the Western Blot analysis confirms that all constructs were expressed as insoluble proteins (Figure 28, panel D). While the fluorescence of the eGFP tag can be used to track its presence on the gel, the sample processing temperature for SDS-PAGE may affect fluorescence intensity. Conversely, employing lower temperatures risks improper protein unfolding, potentially altering band positions on the gel. To address these limitations, a more sensitive and reliable technique, such as Western blotting, was preferred, as it provides a stable and enhanced signal.

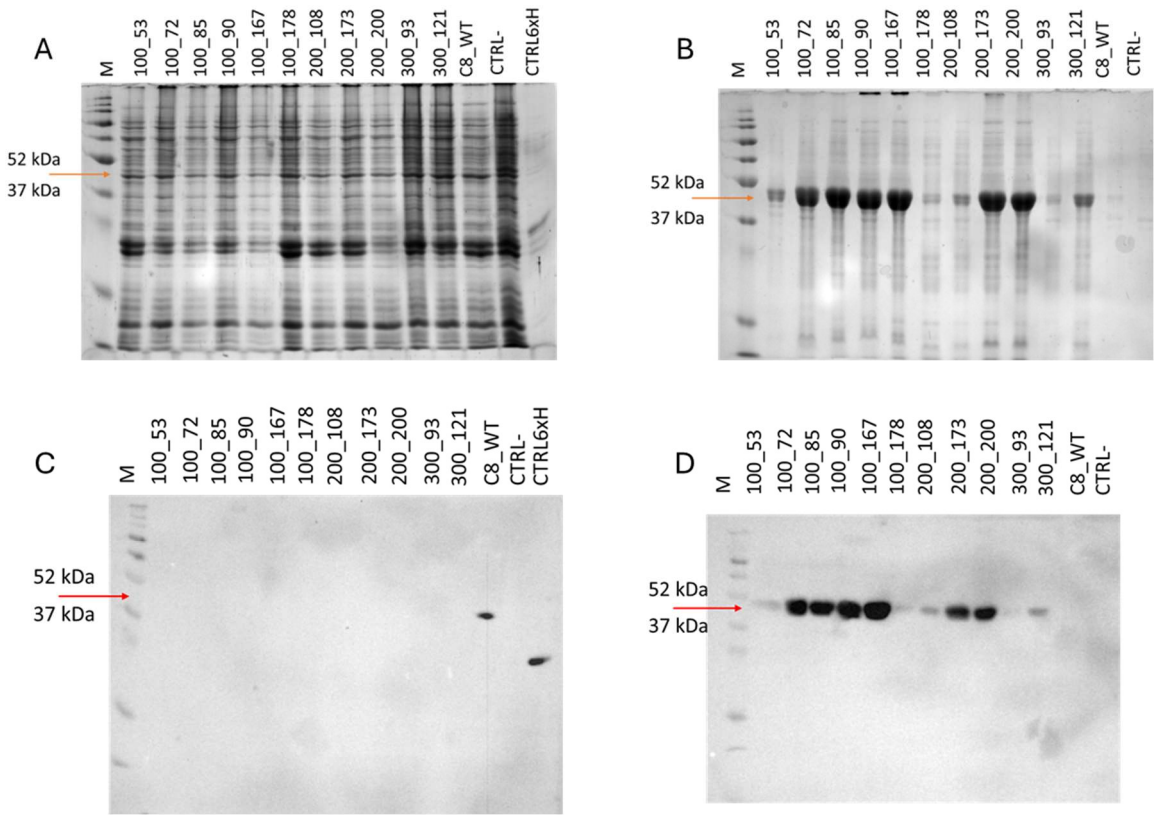


Figure 28: **Evaluation of VHH expression.** In the upper panels, SDS-PAGE of the lysate soluble fraction (A) and urea-solubilized fraction (B). The lower panels report the corresponding Western Blot analyses of the soluble fraction (C) and the urea-solubilized fraction (D). The arrows indicate the expected molecular weight for the fusion proteins VHH::GFP (42 kDa). In soluble fraction analyses, a GFP protein sample with a 6xHis-tag (CTRL6xH) was included as a positive control for the Western Blot.

While the progenitor C8WT is expressed to a good yield as a soluble protein and is easily recovered from the cell lysate supernatant after centrifugation, all the mutants accumulate in IBs with different yields. Proteins with point mutations frequently demonstrate variable expression outcomes, with some accumulating in IBs while others remain soluble in the cytosol[133]. The computationally optimized VHHs exhibit mutations across all three CDRs, affecting 10.3% to 15.5% of the residues in the nanobody domain. The extent of the primary sequence differences significantly heightens the likelihood of the occurrence of mutations critical for maintaining proper protein folding. Moreover, the VHH::eGFP fusion proteins contain four cysteine residues, two in the VHH portion and two in the eGFP domain. In the native structure, disulfide bonds are expected to form exclusively within each domain. However, the potential for improper intramolecular and intermolecular disulfide bond formation cannot be excluded. Notably, even though the anti-HEWL VHHs discussed in the case study were also extensively mutated from the wild type and fused to an eGFP tag, none accumulated in IBs during expression.

The expression of proteins in inclusion bodies offers several advantages[134]. For instance, inclusion bodies can facilitate high protein expression levels, overcoming the limitations of protein solubility. Within these aggregates, proteins tend to be more stable and resistant to proteolytic degradation than their soluble counterparts [135]. Additionally, the presence of inclusion bodies simplifies purification processes, as they can typically be separated from soluble fractions through centrifugation or filtration, with the aggregates consisting of up to 95% recombinant protein. Lastly, inclusion bodies can be stored for extended periods with minimal loss of protein integrity, making them suitable for long-term use[136].

Nevertheless, several limitations exist when expressing proteins in inclusion bodies[137]. Proteins are usually aggregated and denatured, requiring a refolding procedure that frequently yields inconsistent recovery. Moreover, the correct refolding conditions must be identified to produce functional and monodisperse proteins. This includes ensuring the proper formation of structural features such as disulfide bonds. These challenges can complicate the downstream process.

The scale-up of protein production, purification, and characterization of the VHHs begins with the progenitor, C8WT, to establish a baseline for comparison of the computationally designed mutants. Based on the results from small-scale expression, the five anti-P105 VHHs with the highest yields were selected for scale-up (100_85, 100_90, 100_167, 200_173, and 200_200). Initial expression optimization and purification trials were conducted on VHH 100_85.

4.2.4 C8WT purification and characterization

The purification of the C8WT::eGFP fusion protein was performed in two steps (Figure 29). The first affinity chromatography was used to separate the fusion protein from the other components of the soluble fraction of the lysate. SDS-PAGE analysis revealed the presence of contaminants in the sample (Figure 29, left panel, lanes E1-E3). Through a second polishing step of Size Exclusion Chromatography (Figure 29, right panel), most of the contaminants still present in the sample were removed. Upon completion, SDS-PAGE analysis confirmed that the sample had a degree of purity suitable for MST assays, despite small amounts of low-molecular weight contaminants (Figure 29, left panel, lanes marked as SEC). A total yield of 8 mg/L of culture was obtained.

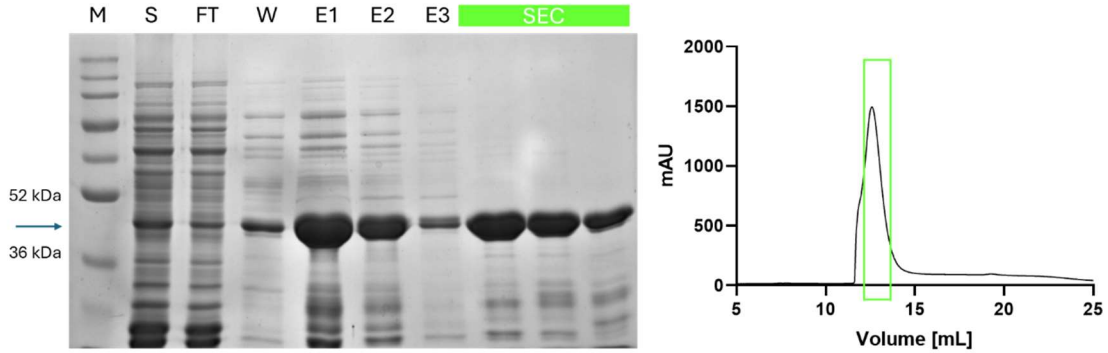


Figure 29: **C8WT::eGFP purification**. *(Left panel)* SDS-PAGE analysis of the samples from different stages of the purification protocol. Lanes: M, protein molecular weight marker; S, cell lysate soluble fraction; FT, flow-through from affinity chromatography; E1-E3, affinity chromatography elution fractions; Size Exclusion Chromatography (SEC) fractions are highlighted in green. *(Right panel)* SEC chromatogram. The green box indicates the selected fractions.

To obtain the C8WT VHH protein without tags, the eGFP tag was removed through proteolytic cleavage with TEV protease. The enzyme used in this reaction is routinely produced in a recombinant expression system, with a histidine tag to facilitate the following purification. The reaction lasted 16h and the results were analyzed by SDS-PAGE (Figure 30, panel A). Despite a significant amount of uncleaved protein (band at 42 kDa), the C8WT protein is visible at the correct molecular weight (~15 kDa). At the end of the reaction, the VHH was isolated in the flow-through of a reverse IMAC protocol (Figure 30, panel B), while the TEV protease, the cleaved eGFP domain, and the uncleaved fusion protein were retained by the resin. The weak intensity of the bands visible in the SDS-PAGE gel for the flow-through samples is due to protein dilution. A SEC polishing step was performed (Figure 30, panel C), yielding the pure VHH protein (Figure 30, panel D).

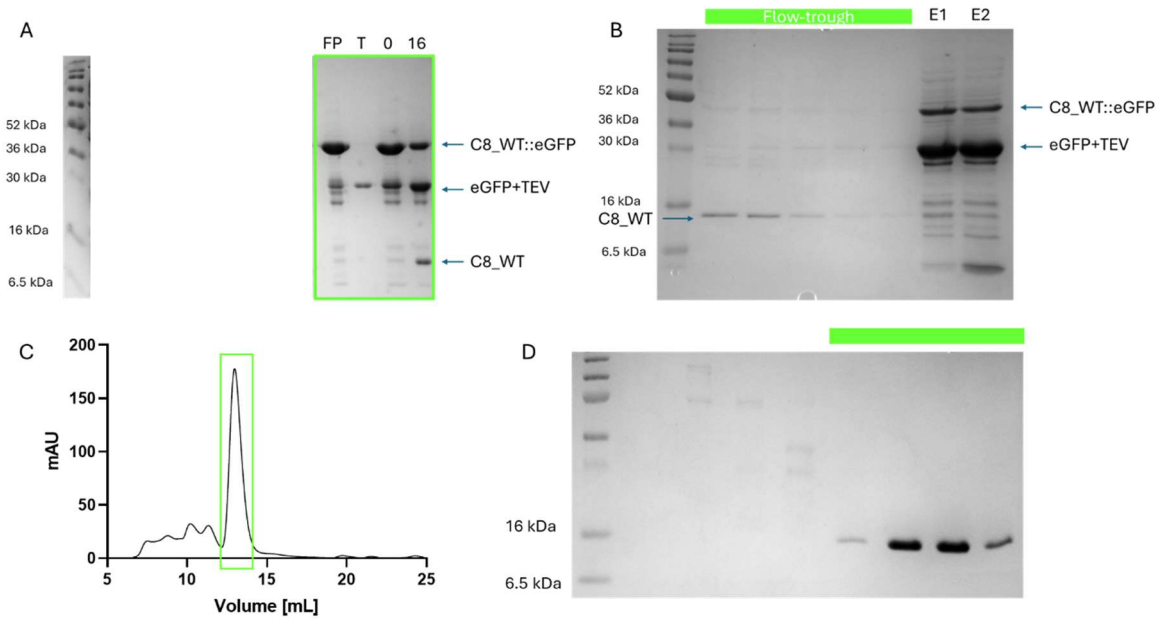
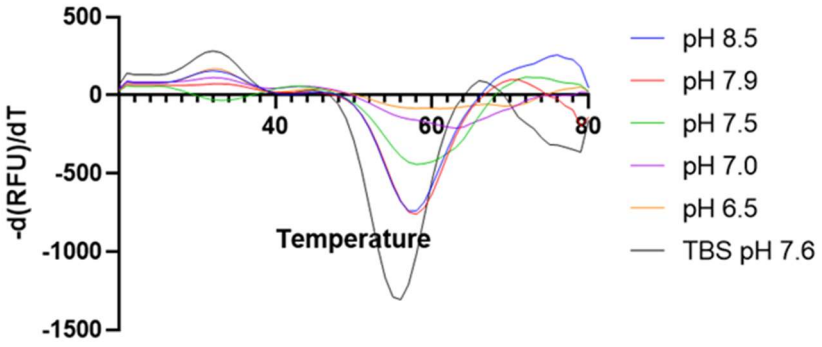


Figure 30:: **eGFP tag removal.**: A) SDS-PAGE analysis of the TEV cleavage reaction: FP, C8WT::eGFP fusion protein; T, recombinant TEV protease added to the reaction mixture; 0, initial reaction mixture; 16: reaction mixture after 16 hours. The arrows highlight the species present in the last lane. B) SDS-PAGE analysis of the reverse IMAC purification: the flow-through samples are marked in green, while lanes E1-E2 correspond to the elution fractions. C) Size exclusion chromatography performed after the eGFP-tag proteolytic removal; the green box marks the fractions containing the pure VHH protein. D) SDS-PAGE analysis of the SEC; fractions containing VHH are marked in green.

A sample of the purified VHH was used to characterize the thermal stability of the protein through TSA analysis. As expected for a cysteine-containing globular protein, a well-packed hydrophobic core and the presence of disulfide bridges contribute to the thermal stabilization of the VHH. However, VHH unfolding temperatures can vary significantly[138]. In particular, the conformation of CDR3, the longest and most variable of the three variable regions, and the interactions it forms with the nanobody framework are believed to affect significantly the unfolding temperature [139].

TSA analysis of the C8WT was conducted in different buffer conditions to analyze the stability of the VHH at different pH values (Figure 31) In the buffer used for the SEC purification (Tris buffer at pH 7.6), the VHH unfolds relatively early in the analysis, with a calculated unfolding temperature of 56 °C (Figure 31, grey curve). However, other conditions without NaCl and at higher pH values (pH 7.9 and pH 8.5) provided higher protein stability, with a unfolding temperature of 58°C (Figure 31, blue and red curves). The lower fluorescence signal in experiments conducted with buffer at pH values of 7.5 or lower, along with the presence of multiple transitions (Figure 31, green, purple, and yellow curves),

suggests a more complex dynamic in neutral or mildly acidic conditions, than that expected for a simple two-state native-denatured system, warranting further investigation.



Buffer	C_f	pH	T_u [°C]
Bicine	0.1 M	8.5	58
Na Phosphate	0.1 M	7.9	58
HEPES	0.1 M	7.5	N.A.
HEPES	0.1 M	7,0	N.A.
MES	0.1 M	6.5	N.A.
TBS	1x	7.6	56

Figure 31: **C8WT Thermal Shift Assay**: The top panel displays the first derivative of the fluorescence signal as a function of temperature, with the minimum of each curve representing the T_u for the corresponding buffer condition. The bottom panel reports the calculated unfolding temperatures for each condition.

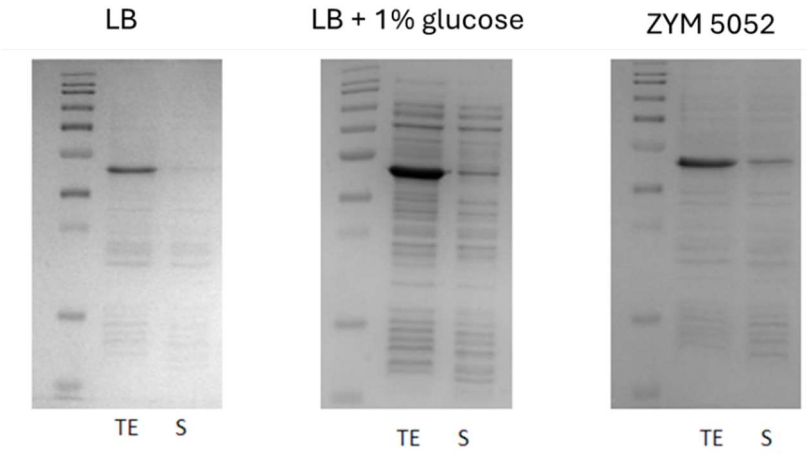
4.2.5 Anti-P105 VHH expression optimization

Several attempts were made to optimize the expression of computationally designed VHHs targeting the P105 fragment. Experiments performed on the expression of the VHH 100_85, used as a model, analyzed the effect of different induction temperatures and times, and inducer concentrations. However, expression yields remained low, at the same levels obtained with the initial protocol. The use of a different growth medium was also considered. Expression tests in LB were compared with those performed by adding 1% glucose to the LB medium used for the pre-inoculum culture, and with expression tests in the autoinduction medium ZYM5052[116]. SDS-PAGE analysis of these experiments revealed an improvement in the protein yield when 1% glucose was added to the pre-inoculum medium (Figure 32). Although the protein is still absent from the soluble fraction of the lysate, the amount of protein that can be recovered from IBs is higher.

When heterologous expression is performed in BL21(DE3) strains using the pET expression system, glucose acts as a repressor of the basal expression levels. Glucose significantly reduces the basal transcription of T7 RNA polymerase, which is crucial for

initiating protein expression from the pET vector. This repression minimizes the risk of toxic protein accumulation, especially during the stationary phase of culture growth. Moreover, the presence of glucose can enhance initial growth rates, resulting in a higher cell density in the pre-inoculum culture compared to cultures grown without glucose.

Considering the results of tests on the VHH 100_85, for all the computationally designed VHHs, the scale-up of protein production was performed with the optimized protocol.



Figyre 32: **Effect of different cell growth conditions on VHH 100_85 expression**. SDS-PAGE analyses of samples of the total cellular extracts (TE) and the soluble fractions (S) in the different culture conditions.

4.2.6 Protein refolding: a trial and error process

Protein refolding is crucial in biotechnology and in the production of biopharmaceuticals, processes in which expression in recombinant systems often results in unfolded or misfolded proteins that require refolding to regain their native structure and function. During the refolding process, folded states compete with aggregation forces that lead to unfolded and aggregated proteins. Usually, a lower starting protein concentration and a slower change in the solution conditions help reach a higher yield of refolded protein[140]. Except for a few notable cases, among which the early work on refolding of ribonuclease A by Anfinsen and co-workers[141] , protein refolding is achieved through a trial-and-error process and requires often multiple attempts. In this study, three distinct methods were employed: a dialysis method (Method A) and two dilution methods (Methods B and C). Method A closely resembles the approach used by Hamidi *et al.* for the refolding of a single chain variable fragment (scFv), which, however, yielded suboptimal results [142]. Method B is analogous to the technique used by Xu *et al.* to refold a VHH targeting EGFR[143]; this

approach is notably rapid, but the authors recommend using very low protein concentrations for effective results. Method C was designed to reproduce the method employed by Moghadam *et al.* to refold a cysteine-rich protein [144].

The results of the three different approaches were evaluated considering not only the yield of refolded protein but also the formation of intramolecular rather than intermolecular disulfide bridges. Figure 33 shows the SDS-PAGE analysis of refolding experiments with the three methods described above. Following the refolding process, the precipitate was removed, and the samples were subsequently loaded onto an SDS-PAGE gel using two distinct loading buffers: the first containing a reducing agent (lanes marked as R in Figure 33) and the second in non-reducing conditions (lanes marked as NR in Figure 33). This analysis allows the detection of protein samples in which intermolecular disulfide bonds are formed. The SDS-PAGE analysis shows that method A is ineffective due to the formation of covalent aggregates of the VHH held together by disulfide bonds, as highlighted by the smear at molecular weights above 42 kDa that can be observed in the NR lane corresponding to the sample refolded through method A in Figure 33. Method B, as well, proved inefficient, considering that the protein sample treated with the non-reducing loading buffer has a very high molecular weight, suggesting the formation of intermolecular cystines. Method C, on the other hand, provides a monodisperse sample, with a small amount of high-molecular weight aggregates. This method was, therefore, selected for the refolding of all computationally designed VHHs.

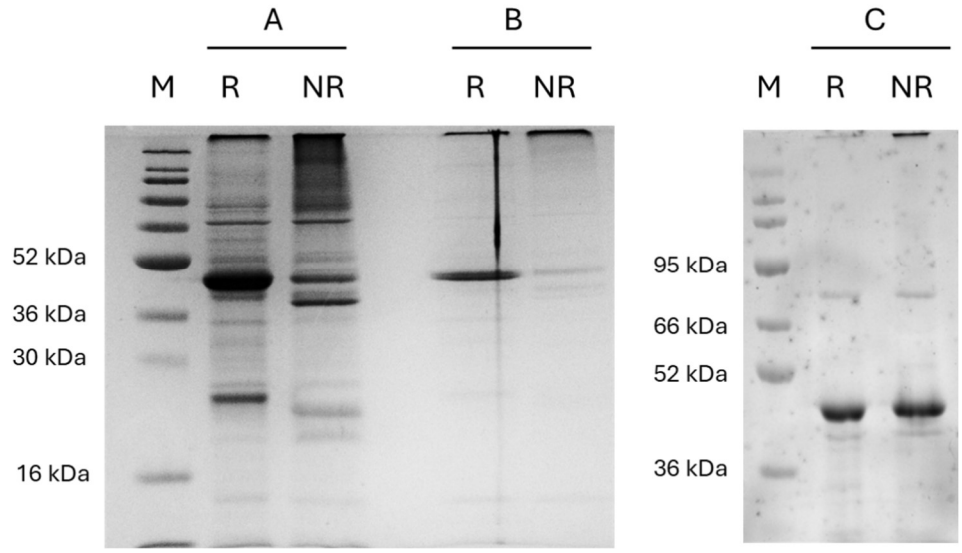


Figure 33: **Comparison of protein refolding protocols**: SDS-PAGE analysis of VHH 100_85 samples obtained in IBs and refolded with different methods (A, B, and C, respectively). Lanes: M, protein marker; R, sample treated with a loading buffer containing β-mercaptoethanol as reducing agent; NR: sample treated with a non-reducing loading buffer.

Using method C, the final yield of the functional VHH 100_85 protein was ~20 mg/250 mL of cell culture. Notably, this value is higher than the yield obtained for the C8WT protein expressed as a soluble protein. However, while both the denatured protein solution containing 8 M urea and the precipitate appear green due to the fluorescence of the eGFP tag, all samples obtained after refolding with the described methods lead to the loss of eGFP fluorescence.

The method resulting in a successful refolding of the VHH 100_85 protein was applied to other VHHs computationally optimized for P105 recognition, followed by a dialysis step against TBS with 10% glycerol and 2 mM DTT. After the refolding step and the buffer exchange, the recovery yield was between 10 and 20% of the protein expressed in inclusion bodies. Panel A in Figure 34 shows the results of the SDS-PAGE analyses of the final samples containing these VHHs (results for VHH 200_173 are not shown). Lacking a confirmation of the protein identity from the eGFP-tag fluorescence, a dot blot analysis was performed using an anti-His antibody. A sample of Bovine Serum Albumin (BSA) was added as the negative control, while a sample containing a His-tagged GFP protein provided a positive control. The results confirmed the presence of histidine-tagged proteins in all the VHH samples (Figure 34.B).

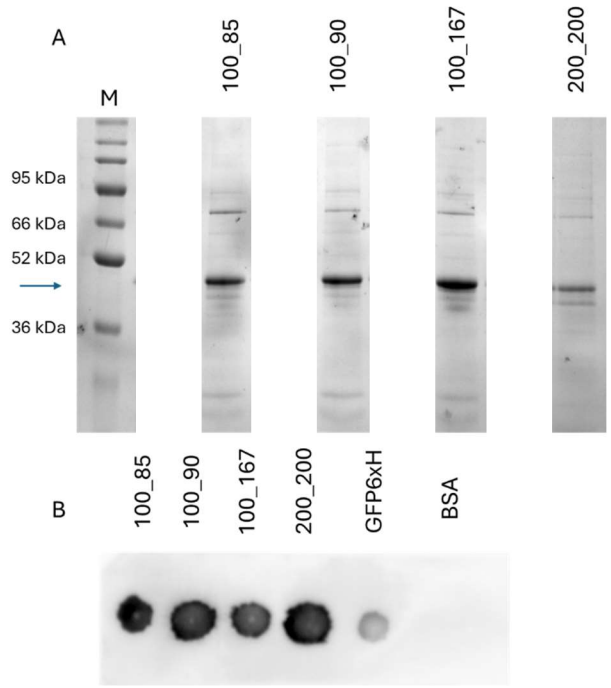


Figure. 34: **Refolded anti-P105 VHH::eGFP protein samples**. A) SDS-PAGE analysis of the purified and refolded VHH::eGFP proteins, after buffer exchange to TBS, 10% glycerol, and 2 mM DTT. In lane M, protein marker. B) Dot blot analysis with an anti-His-tag antibody to confirm the identity of the VHH::eGFP proteins. A His-tagged eGFP was used as the positive control and a sample of bovine serum albumin (BSA) as the negative control.

4.2.7 eGFP tag removal attempts

Considering the high yields obtained for VHHs 100_85 and 200_173 after refolding, the proteolytic cleavage reaction using the TEV protease was performed on these samples, with the same protocol applied for C8WT. SDS-PAGE analyses performed after the proteolytic cleavage show that the reaction was partially successful (Figure 35): while bands at about 15 kDa are present in both experiments, a large amount of the initial construct is still present (bands at 42 kDa). In addition, affinity chromatography with metal ions does not work well in the presence of reducing agents (DTT, present in the refolded and buffer-exchanged samples). Therefore, size-exclusion chromatography was chosen as a technique to isolate the untagged VHHs. However, this method did not allow a successful separation of the VHHs proteins from the uncleaved fusion protein, the eGFP fragment, and TEV protease. Therefore, an alternative purification strategy will be required in future experiments to obtain tag-free VHHs for thermal stability, target binding or crystallization studies.

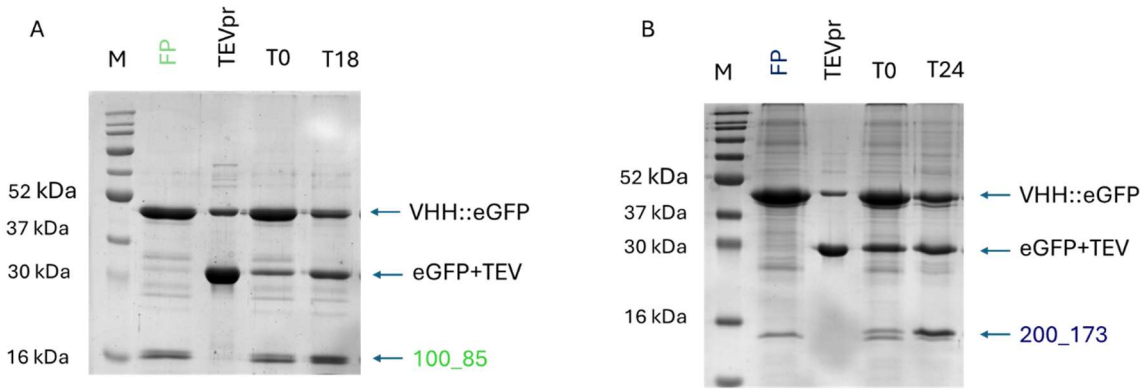


Figure 35: **eGFP tag removal with TEV protease**. SDS-PAGE analysis of the proteolytic reactions on VHH 100_85 (A) and VHH 200_173 (B). Lanes: M, protein markers; FP, VHH::eGFP fusion protein; TEVpr, TEV protease; T0, samples of the reaction mixtures at the beginning of the reaction (time = 0); T18 and T24, samples of the reaction mixtures after 18 and 24 hours of reaction, respectively.

4.2.8 Target production: a reductionistic approach

To experimentally validate the binding between the optimized VHHs and their molecular target, P105 needs to be produced. The region involved in binding is the C-terminal end of P105 (Figure 36), lying between domain IV of the protein, which is extremely rich in cysteines, and the juxtamembrane region. Due to the cytoplasm-reducing environment, *E. coli* expression systems cannot efficiently handle proteins with many disulfide bonds. Expression in mammalian or yeast systems is suitable but, on the other hand, requires more time for optimization and the protein production process. To address this challenge, a reductionist approach was chosen: the region of P105 against which the antibodies were developed was produced by solid-phase peptide synthesis to experimentally validate the binding between the optimized VHHs and their molecular target.

The sequence between residues 627 and 647 of HER2 (UniProt: P04626) was selected and carefully modified in order to obtain a peptide with structural elements similar to the natural region of the protein (peptide P105) as well as control peptides for binding assays (peptides P105 RC, P105 NC, and P100). The sequences, peptide synthesis results, and experimental masses are shown in Table 9.

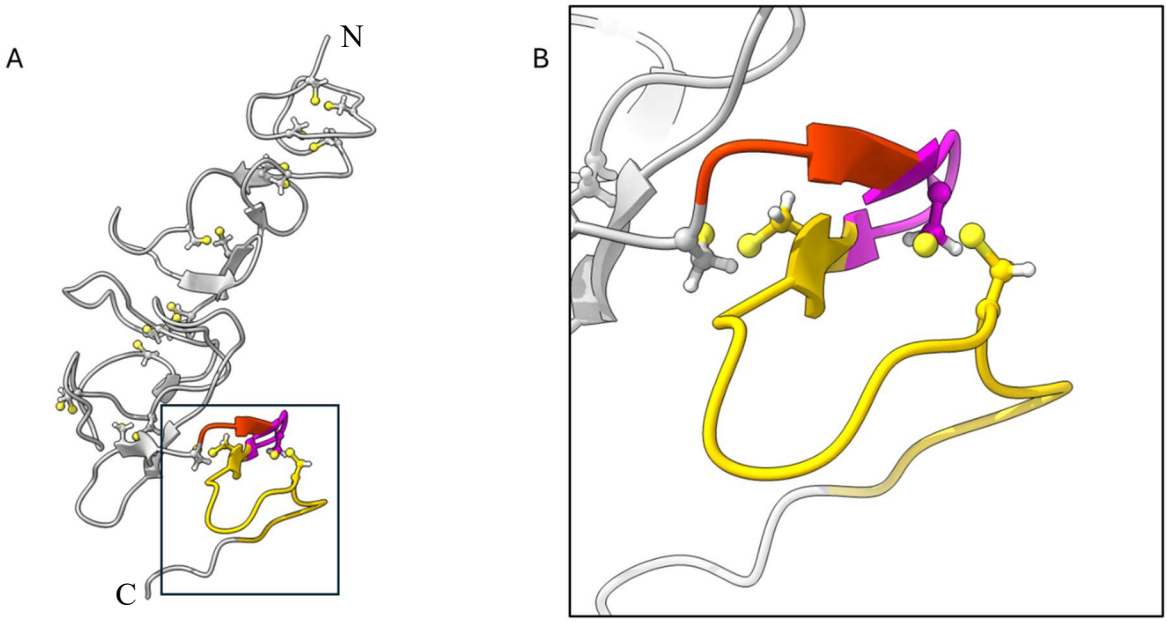


Figure. 36: **P105 C-terminal region.**:A) The structure of the C-terminal region of P105, specifically from residues 510 to 652, showcases a high concentration of cysteine residues, which are depicted using the ball-and-stick model for clarity. B) A detailed view of the 627-647 region, which was a focus during VHH optimization and peptide design: The **PIN** sequence is highlighted in red; The **CTHS** sequence is shown in magenta, with P100 terminating at **S;** The **AVDLDDKGCPAEQR** sequence is highlighted in yellow, representing the final residues of P105 after HER2 cleavage at the major site.

*Table 9: sequence and the experimental mass observed are reported***.**

Name	Sequence	Experimental [MW+H^{+}] [Da]	Final yield mg / 100 mg of HypoGel® 200 RAM resin
Peptide P105	Ac-PINCTHS**A**VDLDDKGCPAEQR-NH2	2308.45	1.8
Peptide P105 RC (reverse charge)	Ac-PINCTHS**A**VRLRRDGCPARQD-NH2	2405.01	1.4
Peptide P105 NC (negative control)	Ac-PINCTHS**A**XXXXXXGCPXXXA-NH2	1644.76	1.2
Peptide P100	Ac-PIN**A**THS-NH2	779.51	18.8

After synthesis, crude materials were purified by HPLC (Figure 37) to isolate the peptides from the impurities, truncated or internal deleted sequences and checked by mass spectrometry to confirm the identity.

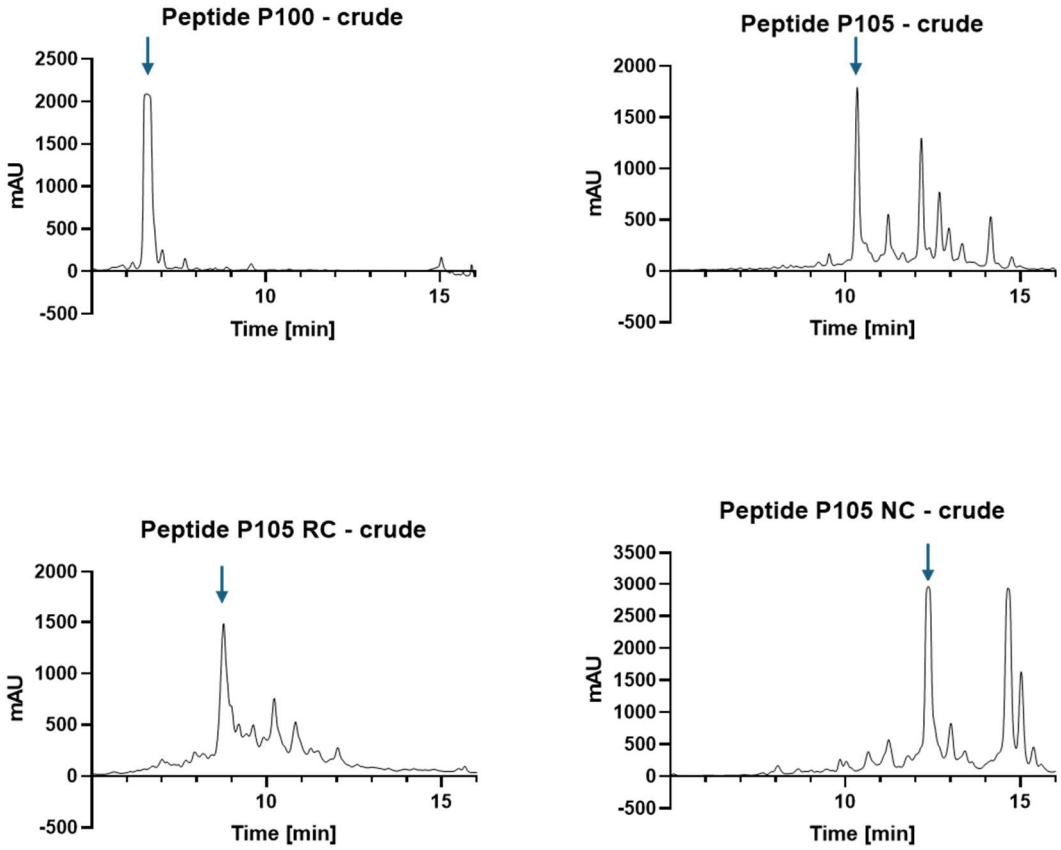


Figure 37: **Crude peptide purification**. The chromatograms from reversed-phase HPLC purification are displayed, with arrows marking the peaks corresponding to the correct linear peptide in each purification process.

P100 showed a good puritiy in the crude. Its experimental mass was determined as shown in Figure 38.

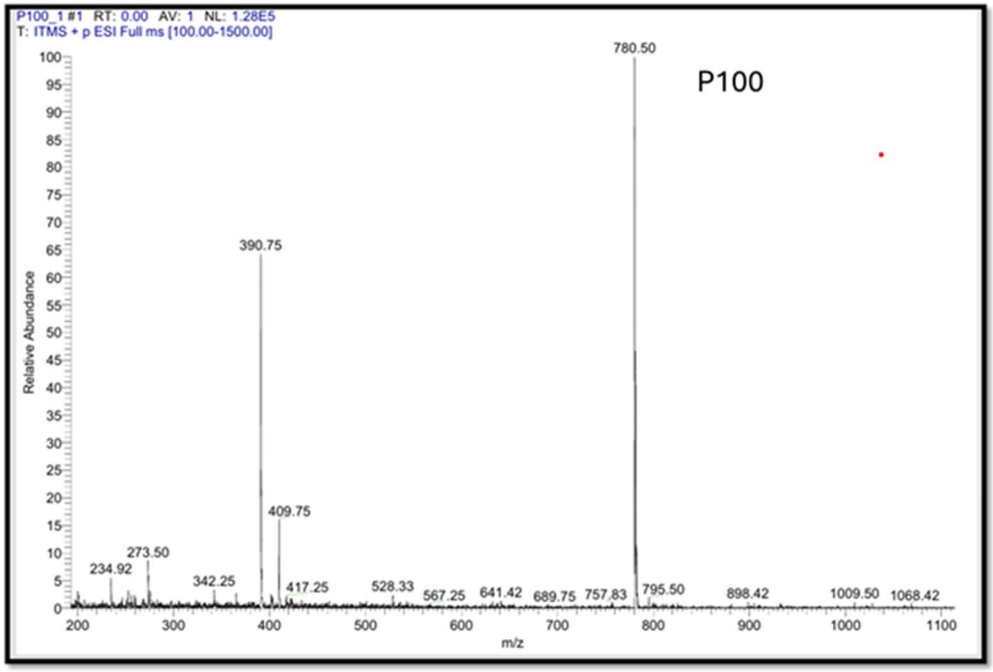


Figure 38: **P100 experimental mass**. Mass spectre with ESI source are shown. Multipy charged ions can be detected (z = 1, 2).

The purified linear peptides were cyclized in solution , using 200 mM ammonium acetate solution at pH 9.1 and 2% v/v DMSO to obtain the disulfide bridge . As H^+ ions are released during the reaction, decreasing the pH of the solution the reaction seems to stall. It has been observed that the pH has to be adjusted back to 9.1, when the pH reaches ~ 8.7. The reaction can be easily monitored by HPLC in the cyclization of peptides P105 and P105 RC because the cyclized and linear forms have different retention times (Figure 39). After two overnight cycles, an increase in the cyclized species (first peak) relative to the linear species (second peak) is observed. However, while a similar pH drop occurs during the cyclization of peptide P105 NC, distinguishing between the two forms using HPLC is not possible because the linear and cyclic peptides have the same retention time. Anyway, mass spectrometry revealed a difference before and after the cyclization process.

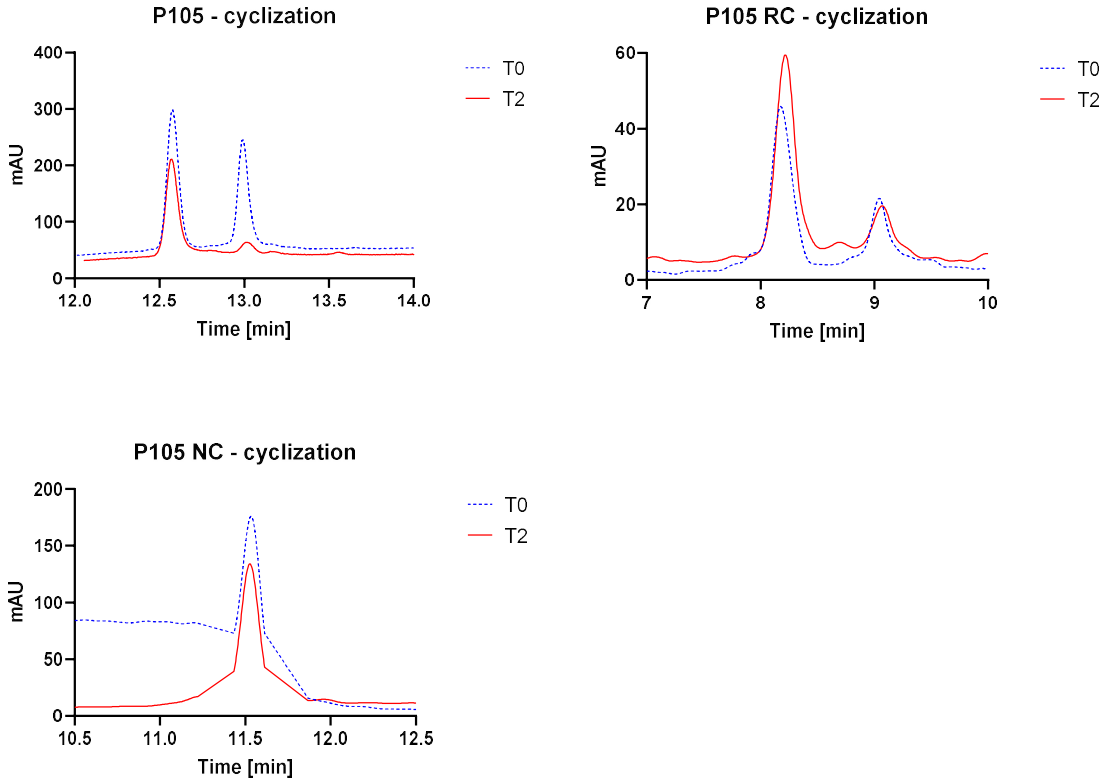


Figure 39: **Cyclization analysis**. HPLC results are reported, detection at λ = 214. T0: reaction at time 0; T2: reaction after two overnight.

To have the cyclic peptides, another purification into HPLC allows is necessary for the separation and purification of the final cyclic product.

The retention times of peptides in reversed-phase HPLC are influenced by their differences in charge and polarity (Figure 40). Peptide P105 for example contains 4 negatively charged and 2 positively charged amino acids, while P105 RC contains 4 positively charged and 2 negatively charged residues, making it more basic. For these reasons, the acidic conditions are chosen for preparation of the mobile phase to reduce ionic interactions between molecules. In this environment, protonated acidic residues become less polar than their ionized counterparts, while protonated basic residues increase in polarity. Consequently, we can observe different retention times. The peptide P105, being more hydrophobic, has greater affinity for the column and elutes later, whereas the more hydrophilic P105 RC elutes before. Peptide P105 NC, on the other hand, contains the non-natural amino acid O2Oc, which lacks side chains. As a result, this peptide has significantly more affinity towards the stationary phase of the column than the others, leading to a stronger

interaction with the non-polar stationary phase and a longer retention time in reversed-phase HPLC.

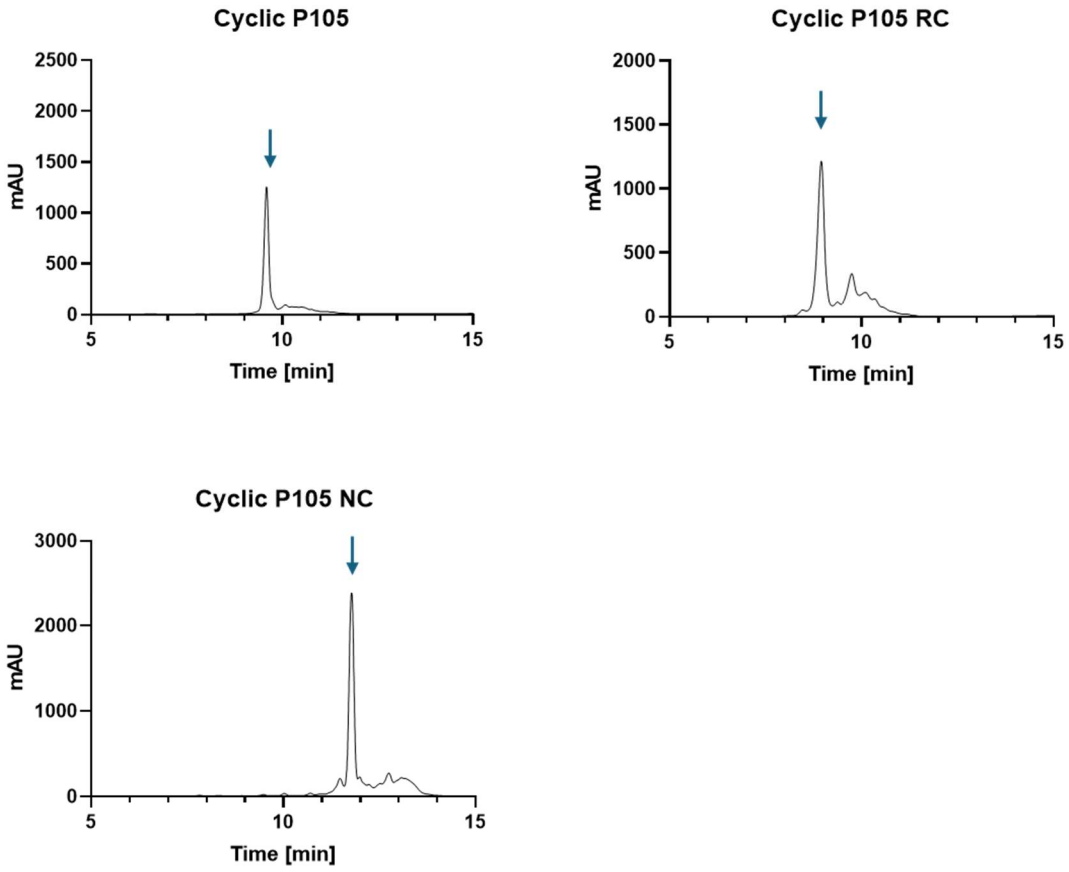


Figure. 40: **Cyclic peptide purification**: The reversed-phase HPLC results are shown, with arrows indicating the peaks representing the correct linear peptide for each purification.

To confirm the identity of the samples, a final mass analysis was conducted on the purified samples. The mass of the cyclized P105 NC peptide was found to be 32 Da higher than expected (30 Da more than the cyclized form), a difference consistent with the mass of CH_3OH, which was used in the mass spectrometer analysis to wash the system and to dissolve the samples (Figure 41).

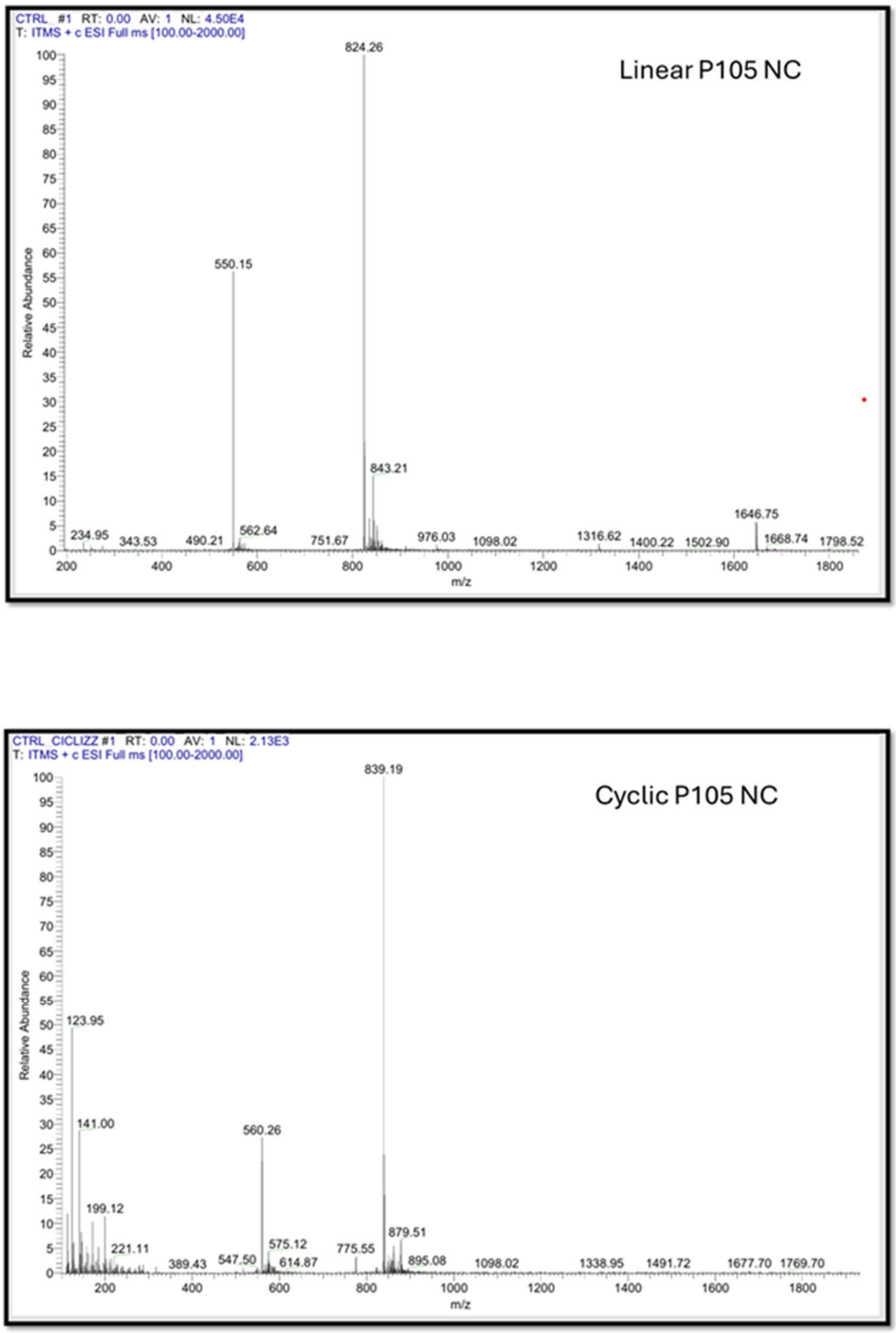


Figure. 41: **Comparison between linear and cyclic P105 experimental masses**. Mass spectres with ESI source are reported. Multipy charged ions can be detected (z = 1, 2, 3).

In Figure 42 are reported the mass spectra of peptides P105 and P105 RC. ESI frequently generates multiply charged ions. This process typically results in a series of peaks representing different charged states of the same molecule, with the spacing between these peaks determined by the number of charges (z) associated with the ions. Notably, if

the predominant charge state observed is odd, such as z = 3, it may suggest that the molecule is predominantly in a monomeric state.

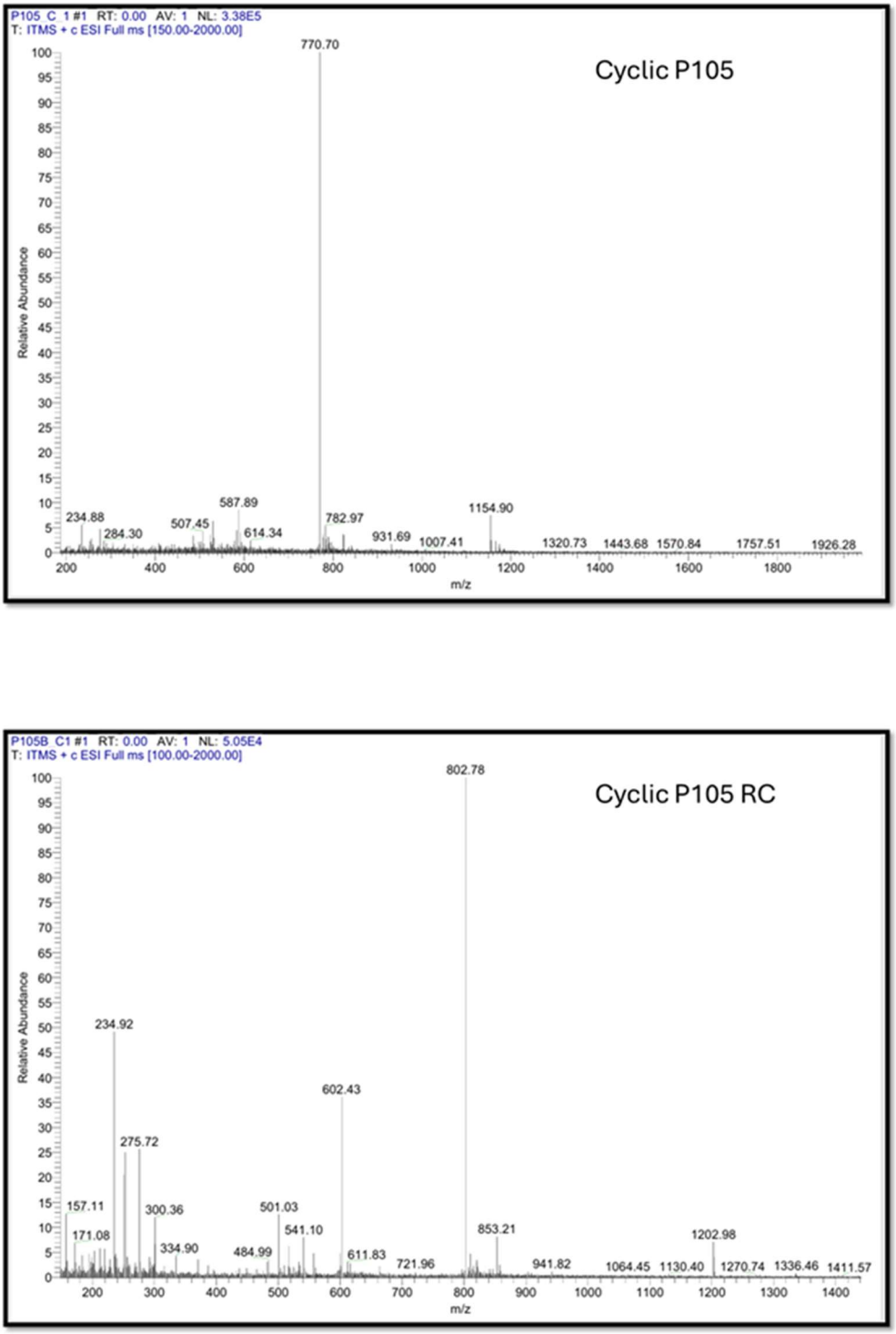


Figure 42: **Cyclic P105 and P105 RC masses.** Mass spectres with ESI source are reported. Multipy charged ions can be detected (z = 2, 3, 4).

5. CONCLUSIONS

The BINDesignER/PARCE protocol is a new tool for the design of nanobody libraries with high affinity for selected targets. Starting from a nanobody model, the evolutionary algorithm of this protocol introduces mutations that enhance epitope recognition.

In this work, two different applications of the PARCE algorithm were considered. In the first proof-of-concept study, the ability of the algorithm to divert the selectivity of a nanobody was tested. A small library of VHH sequences was designed by the software to target the active site of the Hen Egg White Lysozyme starting from a nanobody against a completely different target. The corresponding VHHs were recombinantly expressed in a bacterial host and their affinity for the selected epitope was tested. Binding assays followed by microscale thermophoresis (MST) demonstrated that VHHs optimized against the lysozyme active site exhibit considerable affinity, achieving nanomolar K_D values.

Unfortunately, an experimental crystal structure of the VHH-target complex could not be obtained. However, protein-protein contacts predicted by the PARCE software to drive the VHH binding to the active site of lysozyme were analyzed. The involvement of a histidine residue in a key position of the interaction surface suggested the investigation of the pH-dependent behavior of the nanobodies, as the predicted interaction should be disrupted by the protonation of the imidazole nitrogen of the histidine side chain. The thermal stability of the optimized VHHs was assessed at pH 8.0, where histidine residues are deprotonated, and at pH 5.8, where histidine residues are expected to be protonated. Results showed a decreased stability of the nanobodies at basic pH compared to the acidic solutions. However, when lysozyme was added to the VHH solutions at pH 8, a thermal shift to lower values was observed, suggesting the occurrence of a destabilizing interaction between the VHHs and lysozyme. On the contrary, the stability of the nanobodies at pH 5.8 remained unaffected by the addition of the putative target. Inhibition assays demonstrated that optimized VHHs reduce the enzymatic activity of lysozyme at basic pH, whereas no inhibition was observed at acidic pH. This evidence suggests that the epitope recognized by the VHHs may be located close to the active site pocket of the enzyme, as expected from the PARCE models of the VHH-lysozyme interactions, and that target recognition occurs only when the key histidine residue is deprotonated, aligning with computational expectations.

In the second phase of this project, a set of VHH sequences produced by the PARCE algorithm to target the HER2 proteolytic fragment P105 was cloned into a plasmid suitable

for recombinant expression in bacteria. Interestingly, the *E. coli* expression of the software-designed nanobodies exhibited a markedly different behavior compared to that of their progenitor. While the original VHH was easily purified from the soluble fraction of the cell lysate, the PARCE-generated mutants accumulated in inclusion bodies, though with varying yields. Five of these mutants were expressed at high yields and could be purified. The advantage of protein production in inclusion bodies lies in the high purity of the final sample, as these protein aggregates predominantly contain the recombinant construct, with a minimal amount of contaminants. However, proteins recovered from inclusion bodies are not properly folded and require refolding, a process for which no universal strategy is available and that usually has to be optimized through a trial-and-error approach. Ultimately, the refolding strategy applied to the VHH samples succeeded and yielded a considerable amount of soluble protein, higher than the expression levels of the soluble form of the progenitor VHH. However, during refolding, the fluorescence of eGFP-fused VHHs was lost. Efforts to remove the eGFP tag to obtain tag-free VHHs were successful, but further refinement of the purification process is still needed to obtain samples of the nanobodies suitable for affinity binding analysis.

In parallel, a set of peptides was designed to mimic the selected binding site of the nanobodies on the C-terminal segment of the P105 protein. This library of peptides was synthesized via solid-phase synthesis and analysis of crude samples indicated that peptides with the correct molecular weight were the predominant species. Cyclization of the longer peptides was achieved, albeit requiring two days for the reaction be completed. Purification of these cyclized peptides provided suitable targets for VHHs to be tested in future binding assays.

The experimental validation of the computational results obtained with anti-lysozyme VHHs strengthens the credibility of the PARCE algorithm, paving the way for its application to challenging biomedical targets. One such challenge was undertaken in the second part of this work, where the PARCE algorithm was applied for the selection of nanobodies against P105. The successful expression and purification of VHHs against P105 provide a starting point for the following binding analysis, which could lead to the selection of a nanobody with high selectively for a specific soluble form of HER2, increasing the number of available diagnostic tool and, potentially, laying the foundations for the development of new therapeutic tools.

6. REFERENCES

(1) Hamers-Casterman, C.; Atarhouch, T.; Muyldermans, S.; Robinson, G.; Hamers, C.; Songa, E. B.; Bendahman, N.; Hamers, R. Naturally Occurring Antibodies Devoid of Light Chains. *Nature* **1993**, *363* (6428), 446–448. https://doi.org/10.1038/363446a0.

(2) Muyldermans, S. Nanobodies: Natural Single-Domain Antibodies. *Annu. Rev. Biochem.* **2013**, *82* (1), 775–797. https://doi.org/10.1146/annurev-biochem-063011-092449.

(3) Ingram, J. R.; Schmidt, F. I.; Ploegh, H. L. Exploiting Nanobodies' Singular Traits. *Annual Review of Immunology* **2018**, *36* (Volume 36, 2018), 695–715. https://doi.org/10.1146/annurev-immunol-042617-053327.

(4) Muyldermans, S. A Guide to: Generation and Design of Nanobodies. *FEBS J* **2021**, *288* (7), 2084–2102. https://doi.org/10.1111/febs.15515.

(5) de Marco, A. Recombinant Expression of Nanobodies and Nanobody-Derived Immunoreagents. *Protein Expr Purif* **2020**, *172*, 105645. https://doi.org/10.1016/j.pep.2020.105645.

(6) Muyldermans, S.; Baral, T. N.; Retamozzo, V. C.; De Baetselier, P.; De Genst, E.; Kinne, J.; Leonhardt, H.; Magez, S.; Nguyen, V. K.; Revets, H.; Rothbauer, U.; Stijlemans, B.; Tillib, S.; Wernery, U.; Wyns, L.; Hassanzadeh-Ghassabeh, Gh.; Saerens, D. Camelid Immunoglobulins and Nanobody Technology. *Veterinary Immunology and Immunopathology* **2009**, *128* (1–3), 178–183. https://doi.org/10.1016/j.vetimm.2008.10.299.

(7) Zavrtanik, U.; Lukan, J.; Loris, R.; Lah, J.; Hadži, S. Structural Basis of Epitope Recognition by Heavy-Chain Camelid Antibodies. *Journal of Molecular Biology* **2018**, *430* (21), 4369–4386. https://doi.org/10.1016/j.jmb.2018.09.002.

(8) Salvador, J.-P.; Vilaplana, L.; Marco, M.-P. Nanobody: Outstanding Features for Diagnostic and Therapeutic Applications. *Anal Bioanal Chem* **2019**, *411* (9), 1703–1713. https://doi.org/10.1007/s00216-019-01633-4.

(9) Soler, M. A.; Medagli, B.; Wang, J.; Oloketuyi, S.; Bajc, G.; Huang, H.; Fortuna, S.; de Marco, A. Effect of Humanizing Mutations on the Stability of the Llama Single-Domain Variable Region. *Biomolecules* **2021**, *11* (2), 163. https://doi.org/10.3390/biom11020163.

(10) Veggiani, G.; de Marco, A. Improved Quantitative and Qualitative Production of Single-Domain Intrabodies Mediated by the Co-Expression of Erv1p Sulfhydryl Oxidase. *Protein Expression and Purification* **2011**, *79* (1), 111–114. https://doi.org/10.1016/j.pep.2011.03.005.

(11) Djender, S.; Schneider, A.; Beugnet, A.; Crepin, R.; Desrumeaux, K. E.; Romani, C.; Moutel, S.; Perez, F.; de Marco, A. Bacterial Cytoplasm as an Effective Cell Compartment for Producing Functional VHH-Based Affinity Reagents and Camelidae IgG-like Recombinant Antibodies. *Microbial Cell Factories* **2014**, *13* (1), 140. https://doi.org/10.1186/s12934-014-0140-1.

(12) Veggiani, G.; Giabbai, B.; Semrau, M. S.; Medagli, B.; Riccio, V.; Bajc, G.; Storici, P.; de Marco, A. Comparative Analysis of Fusion Tags Used to Functionalize Recombinant Antibodies. *Protein Expression and Purification* **2020**, *166*, 105505. https://doi.org/10.1016/j.pep.2019.105505.

(13) Greenberg, A. S.; Avila, D.; Hughes, M.; Hughes, A.; McKinney, E. C.; Flajnik, M. F. A New Antigen Receptor Gene Family That Undergoes Rearrangement and Extensive Somatic Diversification in Sharks. *Nature* **1995**, *374* (6518), 168–173. https://doi.org/10.1038/374168a0.
(14) Juma, S. N.; Gong, X.; Hu, S.; Lv, Z.; Shao, J.; Liu, L.; Chen, G. Shark New Antigen Receptor (IgNAR): Structure, Characteristics and Potential Biomedical Applications. *Cells* **2021**, *10* (5), 1140. https://doi.org/10.3390/cells10051140.
(15) Krah, S.; Schröter, C.; Zielonka, S.; Empting, M.; Valldorf, B.; Kolmar, H. Single-Domain Antibodies for Biomedical Applications. *Immunopharmacology and Immunotoxicology* **2016**, *38* (1), 21–28. https://doi.org/10.3109/08923973.2015.1102934.
(16) Yang, E.; Liu, Q.; Huang, G.; Liu, J.; Wei, W. Engineering Nanobodies for Next-Generation Molecular Imaging. *Drug Discovery Today* **2022**, *27* (6), 1622–1638. https://doi.org/10.1016/j.drudis.2022.03.013.
(17) Zhang, Q.; Zhang, N.; Xiao, H.; Wang, C.; He, L. Small Antibodies with Big Applications: Nanobody-Based Cancer Diagnostics and Therapeutics. *Cancers* **2023**, *15* (23), 5639. https://doi.org/10.3390/cancers15235639.
(18) Massa, S.; Vikani, N.; Betti, C.; Ballet, S.; Vanderhaegen, S.; Steyaert, J.; Descamps, B.; Vanhove, C.; Bunschoten, A.; van Leeuwen, F. W. B.; Hernot, S.; Caveliers, V.; Lahoutte, T.; Muyldermans, S.; Xavier, C.; Devoogdt, N. Sortase A-Mediated Site-Specific Labeling of Camelid Single-Domain Antibody-Fragments: A Versatile Strategy for Multiple Molecular Imaging Modalities. *Contrast Media Mol Imaging* **2016**, *11* (5), 328–339. https://doi.org/10.1002/cmmi.1696.
(19) Najjar, M. K.; Manore, S. G.; Regua, A. T.; Lo, H.-W. Antibody-Drug Conjugates for the Treatment of HER2-Positive Breast Cancer. *Genes* **2022**, *13* (11), 2065. https://doi.org/10.3390/genes13112065.
(20) Pankhurst, T.; Adu, D. Antibodies in the Prevention of Renal Allograft Rejection. *Expert Opinion on Biological Therapy* **2004**, *4* (2), 243–252. https://doi.org/10.1517/14712598.4.2.243.
(21) Scully, M.; Cataland, S. R.; Peyvandi, F.; Coppo, P.; Knöbl, P.; Kremer Hovinga, J. A.; Metjian, A.; de la Rubia, J.; Pavenski, K.; Callewaert, F.; Biswas, D.; De Winter, H.; Zeldin, R. K.; HERCULES Investigators. Caplacizumab Treatment for Acquired Thrombotic Thrombocytopenic Purpura. *N Engl J Med* **2019**, *380* (4), 335–346. https://doi.org/10.1056/NEJMoa1806311.
(22) Hmila, I.; Saerens, D.; Ben Abderrazek, R.; Vincke, C.; Abidi, N.; Benlasfar, Z.; Govaert, J.; El Ayeb, M.; Bouhaouala-Zahar, B.; Muyldermans, S. A Bispecific Nanobody to Provide Full Protection against Lethal Scorpion Envenoming. *FASEB J* **2010**, *24* (9), 3479–3489. https://doi.org/10.1096/fj.09-148213.
(23) Pagneux, Q.; Roussel, A.; Saada, H.; Cambillau, C.; Amigues, B.; Delauzun, V.; Engelmann, I.; Alidjinou, E. K.; Ogiez, J.; Rolland, A. S.; Faure, E.; Poissy, J.; Duhamel, A.; Boukherroub, R.; Devos, D.; Szunerits, S. SARS-CoV-2 Detection Using a Nanobody-Functionalized Voltammetric Device. *Commun Med* **2022**, *2* (1), 1–11. https://doi.org/10.1038/s43856-022-00113-8.
(24) Aksu, M.; Kumar, P.; Güttler, T.; Taxer, W.; Gregor, K.; Mußil, B.; Rymarenko, O.; Stegmann, K. M.; Dickmanns, A.; Gerber, S.; Reineking, W.; Schulz, C.; Henneck, T.; Mohamed, A.; Pohlmann, G.; Ramazanoglu, M.; Mese, K.; Groß, U.;

Ben-Yedidia, T.; Ovadia, O.; Fischer, D. W.; Kamensky, M.; Reichman, A.; Baumgärtner, W.; von Köckritz-Blickwede, M.; Dobbelstein, M.; Görlich, D. Nanobodies to Multiple Spike Variants and Inhalation of Nanobody-Containing Aerosols Neutralize SARS-CoV-2 in Cell Culture and Hamsters. *Antiviral Research* **2024**, *221*, 105778. https://doi.org/10.1016/j.antiviral.2023.105778.

(25) Su, Q.; Shi, W.; Huang, X.; Wan, Y.; Li, G.; Xing, B.; Xu, Z. P.; Liu, H.; Hammock, B. D.; Yang, X.; Yin, S.; Lu, X. Screening, Expression, and Identification of Nanobody against SARS-CoV-2 Spike Protein. *Cells* **2022**, *11* (21), 3355. https://doi.org/10.3390/cells11213355.

(26) Huo, J.; Mikolajek, H.; Le Bas, A.; Clark, J. J.; Sharma, P.; Kipar, A.; Dormon, J.; Norman, C.; Weckener, M.; Clare, D. K.; Harrison, P. J.; Tree, J. A.; Buttigieg, K. R.; Salguero, F. J.; Watson, R.; Knott, D.; Carnell, O.; Ngabo, D.; Elmore, M. J.; Fotheringham, S.; Harding, A.; Moynié, L.; Ward, P. N.; Dumoux, M.; Prince, T.; Hall, Y.; Hiscox, J. A.; Owen, A.; James, W.; Carroll, M. W.; Stewart, J. P.; Naismith, J. H.; Owens, R. J. A Potent SARS-CoV-2 Neutralising Nanobody Shows Therapeutic Efficacy in the Syrian Golden Hamster Model of COVID-19. *Nat Commun* **2021**, *12* (1), 5469. https://doi.org/10.1038/s41467-021-25480-z.

(27) Oshima, H. S.; Sano, F. K.; Akasaka, H.; Iwama, A.; Shihoya, W.; Nureki, O. Optimizing Cryo-EM Structural Analysis of Gi-Coupling Receptors via Engineered Gt and Nb35 Application. *Biochemical and Biophysical Research Communications* **2024**, *693*, 149361. https://doi.org/10.1016/j.bbrc.2023.149361.

(28) Schlimgen, R. R.; Peterson, F. C.; Heukers, R.; Smit, M. J.; McCorvy, J. D.; Volkman, B. F. Structural Basis for Selectivity and Antagonism in Extracellular GPCR-Nanobodies. *Nat Commun* **2024**, *15* (1), 4611. https://doi.org/10.1038/s41467-024-49000-x.

(29) Manglik, A.; Kobilka, B. K.; Steyaert, J. Nanobodies to Study G Protein-Coupled Receptor Structure and Function. *Annual review of pharmacology and toxicology* **2016**, *57*, 19. https://doi.org/10.1146/annurev-pharmtox-010716-104710.

(30) Rasmussen, S. G. F.; Choi, H.-J.; Fung, J. J.; Pardon, E.; Casarosa, P.; Chae, P. S.; DeVree, B. T.; Rosenbaum, D. M.; Thian, F. S.; Kobilka, T. S.; Schnapp, A.; Konetzki, I.; Sunahara, R. K.; Gellman, S. H.; Pautsch, A.; Steyaert, J.; Weis, W. I.; Kobilka, B. K. Structure of a Nanobody-Stabilized Active State of the B2 Adrenoceptor. *Nature* **2011**, *469* (7329), 175–180. https://doi.org/10.1038/nature09648.

(31) Verheesen, P.; ten Haaft, M. R.; Lindner, N.; Verrips, C. T.; de Haard, J. J. W. Beneficial Properties of Single-Domain Antibody Fragments for Application in Immunoaffinity Purification and Immuno-Perfusion Chromatography. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2003**, *1624* (1), 21–28. https://doi.org/10.1016/j.bbagen.2003.09.006.

(32) Zhang, L.; Zang, B.; Huang, C.; Ren, J.; Jia, L. One-Step Preparation of a VHH-Based Immunoadsorbent for the Extracorporeal Removal of B2-Microglobulin. *Molecules* **2019**, *24* (11), 2119. https://doi.org/10.3390/molecules24112119.

(33) De Coninck, B.; Verheesen, P.; Vos, C. M.; Van Daele, I.; De Bolle, M. F.; Vieira, J. V.; Peferoen, M.; Cammue, B. P. A.; Thevissen, K. Fungal Glucosylceramide-Specific Camelid Single Domain Antibodies Are Characterized by Broad

Spectrum Antifungal Activity. *Front. Microbiol.* **2017**, *8*. https://doi.org/10.3389/fmicb.2017.01059.

(34) Ghannam, A.; Kumari, S.; Muyldermans, S.; Abbady, A. Q. Camelid Nanobodies with High Affinity for Broad Bean Mottle Virus: A Possible Promising Tool to Immunomodulate Plant Resistance against Viruses. *Plant Mol Biol* **2015**, *87* (4), 355–369. https://doi.org/10.1007/s11103-015-0282-5.

(35) Mahdavi, S. Z. B.; Oroojalian, F.; Eyvazi, S.; Hejazi, M.; Baradaran, B.; Pouladi, N.; Tohidkia, M. R.; Mokhtarzadeh, A.; Muyldermans, S. An Overview on Display Systems (Phage, Bacterial, and Yeast Display) for Production of Anticancer Antibodies; Advantages and Disadvantages. *International Journal of Biological Macromolecules* **2022**, *208*, 421–442. https://doi.org/10.1016/j.ijbiomac.2022.03.113.

(36) McCafferty, J.; Griffiths, A. D.; Winter, G.; Chiswell, D. J. Phage Antibodies: Filamentous Phage Displaying Antibody Variable Domains. *Nature* **1990**, *348* (6301), 552–554. https://doi.org/10.1038/348552a0.

(37) Arbabi Ghahroudi, M.; Desmyter, A.; Wyns, L.; Hamers, R.; Muyldermans, S. Selection and Identification of Single Domain Antibody Fragments from Camel Heavy-Chain Antibodies. *FEBS Letters* **1997**, *414* (3), 521–526. https://doi.org/10.1016/S0014-5793(97)01062-4.

(38) Ren, Z.; Black, L. W. Phage T4 SOC and HOC Display of Biologically Active, Full-Length Proteins on the Viral Capsid. *Gene* **1998**, *215* (2), 439–444. https://doi.org/10.1016/S0378-1119(98)00298-4.

(39) Deng, X.; Wang, L.; You, X.; Dai, P.; Zeng, Y. Advances in the T7 Phage Display System (Review). *Molecular Medicine Reports* **2018**, *17* (1), 714–720. https://doi.org/10.3892/mmr.2017.7994.

(40) Lankes, H. A.; Zanghi, C. N.; Santos, K.; Capella, C.; Duke, C. M. P.; Dewhurst, S. In Vivo Gene Delivery and Expression by Bacteriophage Lambda Vectors. *J Appl Microbiol* **2007**, *102* (5), 1337–1349. https://doi.org/10.1111/j.1365-2672.2006.03182.x.

(41) Boder, E. T.; Wittrup, K. D. Yeast Surface Display for Screening Combinatorial Polypeptide Libraries. *Nature Biotechnology* **1997**, *15* (6), 553–557. https://doi.org/10.1038/nbt0697-553.

(42) Uchański, T.; Zögg, T.; Yin, J.; Yuan, D.; Wohlkönig, A.; Fischer, B.; Rosenbaum, D. M.; Kobilka, B. K.; Pardon, E.; Steyaert, J. An Improved Yeast Surface Display Platform for the Screening of Nanobody Immune Libraries. *Sci Rep* **2019**, *9* (1), 382. https://doi.org/10.1038/s41598-018-37212-3.

(43) Benatuil, L.; Perez, J. M.; Belk, J.; Hsieh, C.-M. An Improved Yeast Transformation Method for the Generation of Very Large Human Antibody Libraries. *Protein Engineering, Design and Selection* **2010**, *23* (4), 155–159. https://doi.org/10.1093/protein/gzq002.

(44) Salema, V.; Marín, E.; Martínez-Arteaga, R.; Ruano-Gallego, D.; Fraile, S.; Margolles, Y.; Teira, X.; Gutierrez, C.; Bodelón, G.; Fernández, L. Á. Selection of Single Domain Antibodies from Immune Libraries Displayed on the Surface of E. Coli Cells with Two β-Domains of Opposite Topologies. *PLOS ONE* **2013**, *8* (9), e75126. https://doi.org/10.1371/journal.pone.0075126.

(45) Schneewind, O.; Mihaylova-Petkov, D.; Model, P. Cell Wall Sorting Signals in Surface Proteins of Gram-positive Bacteria. *The EMBO Journal* **1993**, *12* (12), 4803–4811. https://doi.org/10.1002/j.1460-2075.1993.tb06169.x.
(46) Lång, H. Outer Membrane Proteins as Surface Display Systems. *International Journal of Medical Microbiology* **2000**, *290* (7), 579–585. https://doi.org/10.1016/S1438-4221(00)80004-1.
(47) Majander, K.; Korhonen, T. K.; Westerlund-Wikström, B. Simultaneous Display of Multiple Foreign Peptides in the FliD Capping and FliC Filament Proteins of the Escherichia Coli Flagellum. *Applied and Environmental Microbiology* **2005**, *71* (8), 4263–4268. https://doi.org/10.1128/AEM.71.8.4263-4268.2005.
(48) Nicolay, T.; Vanderleyden, J.; Spaepen, S. Autotransporter-Based Cell Surface Display in Gram-Negative Bacteria. *Critical Reviews in Microbiology* **2015**, *41* (1), 109–123. https://doi.org/10.3109/1040841X.2013.804032.
(49) Chen, X.; Gentili, M.; Hacohen, N.; Regev, A. A Cell-Free Nanobody Engineering Platform Rapidly Generates SARS-CoV-2 Neutralizing Nanobodies. *Nat Commun* **2021**, *12* (1), 5506. https://doi.org/10.1038/s41467-021-25777-z.
(50) Wagner, H. J.; Wehrle, S.; Weiss, E.; Cavallari, M.; Weber, W. A Two-Step Approach for the Design and Generation of Nanobodies. *International Journal of Molecular Sciences* **2018**, *19* (11), 3444. https://doi.org/10.3390/ijms19113444.
(51) Sormanni, P.; Aprile, F. A.; Vendruscolo, M. Rational Design of Antibodies Targeting Specific Epitopes within Intrinsically Disordered Proteins. *Proceedings of the National Academy of Sciences* **2015**, *112* (32), 9902–9907. https://doi.org/10.1073/pnas.1422401112.
(52) Bertoline, L. M. F.; Lima, A. N.; Krieger, J. E.; Teixeira, S. K. Before and after AlphaFold2: An Overview of Protein Structure Prediction. *Frontiers in Bioinformatics* **2023**, *3*, 1120370. https://doi.org/10.3389/fbinf.2023.1120370.
(53) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.
(54) Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Hanikel, N.; Pellock, S. J.; Courbet, A.; Sheffler, W.; Wang, J.; Venkatesh, P.; Sappington, I.; Torres, S. V.; Lauko, A.; De Bortoli, V.; Mathieu, E.; Ovchinnikov, S.; Barzilay, R.; Jaakkola, T. S.; DiMaio, F.; Baek, M.; Baker, D. De Novo Design of Protein Structure and Function with RFdiffusion. *Nature* **2023**, *620* (7976), 1089–1100. https://doi.org/10.1038/s41586-023-06415-8.
(55) Bennett, N. R.; Watson, J. L.; Ragotte, R. J.; Borst, A. J.; See, D. L.; Weidle, C.; Biswas, R.; Shrock, E. L.; Leung, P. J. Y.; Huang, B.; Goreshnik, I.; Ault, R.; Carr, K. D.; Singer, B.; Criswell, C.; Vafeados, D.; Sanchez, M. G.; Kim, H. M.; Torres, S. V.; Chan, S.; Baker, D. Atomically Accurate de Novo Design of Single-Domain

Antibodies. bioRxiv March 18, 2024, p 2024.03.14.585103. https://doi.org/10.1101/2024.03.14.585103.
(56) Soler, M. A.; Medagli, B.; Semrau, M. S.; Storici, P.; Bajc, G.; Marco, A. de; Laio, A.; Fortuna, S. A Consensus Protocol for the in Silico Optimisation of Antibody Fragments. *Chem. Commun.* **2019**, *55* (93), 14043–14046. https://doi.org/10.1039/C9CC06182G.
(57) Ochoa, R.; Soler, M. A.; Laio, A.; Cossio, P. PARCE: Protocol for Amino Acid Refinement through Computational Evolution. *Computer Physics Communications* **2021**, *260*, 107716. https://doi.org/10.1016/j.cpc.2020.107716.
(58) de Vries, S. J.; van Dijk, M.; Bonvin, A. M. J. J. The HADDOCK Web Server for Data-Driven Biomolecular Docking. *Nat Protoc* **2010**, *5* (5), 883–897. https://doi.org/10.1038/nprot.2010.32.
(59) Andrusier, N.; Nussinov, R.; Wolfson, H. J. FireDock: Fast Interaction Refinement in Molecular Docking. *Proteins: Structure, Function, and Bioinformatics* **2007**, *69* (1), 139–159. https://doi.org/10.1002/prot.21495.
(60) Fogolari, F.; Corazza, A.; Yarra, V.; Jalaru, A.; Viglino, P.; Esposito, G. Bluues: A Program for the Analysis of the Electrostatic Properties of Proteins Based on Generalized Born Radii. *BMC Bioinformatics* **2012**, *13 Suppl 4* (Suppl 4), S18. https://doi.org/10.1186/1471-2105-13-S4-S18.
(61) Soler, M. A.; Minovski, N.; Rocchia, W.; Fortuna, S. Replica-Exchange Optimization of Antibody Fragments. *Computational Biology and Chemistry* **2023**, *103*, 107819. https://doi.org/10.1016/j.compbiolchem.2023.107819.
(62) Barletta, G. P.; Tandiana, R.; Soler, M.; Fortuna, S.; Rocchia, W. Locuaz: An in Silico Platform for Protein Binders Optimization. *Bioinformatics* **2024**, *40* (8), btae492. https://doi.org/10.1093/bioinformatics/btae492.
(63) Cantarutti, C.; Vargas, M. C.; Dongmo Foumthuim, C. J.; Dumoulin, M.; La Manna, S.; Marasco, D.; Santambrogio, C.; Grandori, R.; Scoles, G.; Soler, M. A.; Corazza, A.; Fortuna, S. Insights on Peptide Topology in the Computational Design of Protein Ligands: The Example of Lysozyme Binding Peptides. *Phys Chem Chem Phys* **2021**, *23* (40), 23158–23172. https://doi.org/10.1039/d1cp02536h.
(64) Ferguson, K. M. Structure-Based View of Epidermal Growth Factor Receptor Regulation. *Annual Review of Biophysics* **2008**, *37* (Volume 37, 2008), 353–373. https://doi.org/10.1146/annurev.biophys.37.032807.125829.
(65) Hart, V.; Gautrey, H.; Kirby, J.; Tyson-Capper, A. HER2 Splice Variants in Breast Cancer: Investigating Their Impact on Diagnosis and Treatment Outcomes. *Oncotarget* **2020**, *11* (46), 4338–4357. https://doi.org/10.18632/oncotarget.27789.
(66) Jones, J. T.; Akita, R. W.; Sliwkowski, M. X. Binding Specificities and Affinities of Egf Domains for ErbB Receptors. *FEBS Letters* **1999**, *447* (2–3), 227–231. https://doi.org/10.1016/S0014-5793(99)00283-5.
(67) Burgess, A. W.; Cho, H.-S.; Eigenbrot, C.; Ferguson, K. M.; Garrett, T. P. J.; Leahy, D. J.; Lemmon, M. A.; Sliwkowski, M. X.; Ward, C. W.; Yokoyama, S. An Open-and-Shut Case? Recent Insights into the Activation of EGF/ErbB Receptors. *Molecular Cell* **2003**, *12* (3), 541–552. https://doi.org/10.1016/S1097-2765(03)00350-2.

(68) Lemmon, M. A. Ligand-Induced ErbB Receptor Dimerization. *Experimental Cell Research* **2009**, *315* (4), 638–648. https://doi.org/10.1016/j.yexcr.2008.10.024.
(69) Cho, H.-S.; Mason, K.; Ramyar, K. X.; Stanley, A. M.; Gabelli, S. B.; Denney, D. W.; Leahy, D. J. Structure of the Extracellular Region of HER2 Alone and in Complex with the Herceptin Fab. *Nature* **2003**, *421* (6924), 756–760. https://doi.org/10.1038/nature01392.
(70) Ahmad, A. Breast Cancer Statistics: Recent Trends. In *Breast Cancer Metastasis and Drug Resistance: Challenges and Progress*; Ahmad, A., Ed.; Springer International Publishing: Cham, 2019; pp 1–7. https://doi.org/10.1007/978-3-030-20301-6_1.
(71) Tsé, C.; Gauchez, A.-S.; Jacot, W.; Lamy, P.-J. HER2 Shedding and Serum HER2 Extracellular Domain: Biology and Clinical Utility in Breast Cancer. *Cancer Treatment Reviews* **2012**, *38* (2), 133–142. https://doi.org/10.1016/j.ctrv.2011.03.008.
(72) Loibl, S.; Gianni, L. HER2-Positive Breast Cancer. *The Lancet* **2017**, *389* (10087), 2415–2429. https://doi.org/10.1016/S0140-6736(16)32417-5.
(73) Niikura, N.; Liu, J.; Hayashi, N.; Mittendorf, E. A.; Gong, Y.; Palla, S. L.; Tokuda, Y.; Gonzalez-Angulo, A. M.; Hortobagyi, G. N.; Ueno, N. T. Loss of Human Epidermal Growth Factor Receptor 2 (HER2) Expression in Metastatic Sites of HER2-Overexpressing Primary Breast Tumors. *JCO* **2012**, *30* (6), 593–599. https://doi.org/10.1200/JCO.2010.33.8889.
(74) Niikura, N.; Tomotaki, A.; Miyata, H.; Iwamoto, T.; Kawai, M.; Anan, K.; Hayashi, N.; Aogi, K.; Ishida, T.; Masuoka, H.; Iijima, K.; Masuda, S.; Tsugawa, K.; Kinoshita, T.; Nakamura, S.; Tokuda, Y. Changes in Tumor Expression of HER2 and Hormone Receptors Status after Neoadjuvant Chemotherapy in 21 755 Patients from the Japanese Breast Cancer Registry. *Annals of Oncology* **2016**, *27* (3), 480–487. https://doi.org/10.1093/annonc/mdv611.
(75) Aksamitiene, E.; Kiyatkin, A.; Kholodenko, B. N. Cross-Talk between Mitogenic Ras/MAPK and Survival PI3K/Akt Pathways: A Fine Balance. *Biochem Soc Trans* **2012**, *40* (1), 139–146. https://doi.org/10.1042/BST20110609.
(76) Yamaguchi, H.; Chang, S.-S.; Hsu, J. L.; Hung, M.-C. Signaling Cross-Talk in the Resistance to HER Family Receptor Targeted Therapy. *Oncogene* **2014**, *33* (9), 1073–1081. https://doi.org/10.1038/onc.2013.74.
(77) Gao, X.; Nawaz, Z. Progesterone Receptors - Animal Models and Cell Signaling in Breast Cancer: Role of Steroid Receptor Coactivators and Corepressors of Progesterone Receptors in Breast Cancer. *Breast Cancer Research* **2002**, *4* (5), 182. https://doi.org/10.1186/bcr449.
(78) Scaltriti, M.; Rojo, F.; Ocaña, A.; Anido, J.; Guzman, M.; Cortes, J.; Di Cosimo, S.; Matias-Guiu, X.; Ramon y Cajal, S.; Arribas, J.; Baselga, J. Expression of p95HER2, a Truncated Form of the HER2 Receptor, and Response to Anti-HER2 Therapies in Breast Cancer. *JNCI: Journal of the National Cancer Institute* **2007**, *99* (8), 628–638. https://doi.org/10.1093/jnci/djk134.
(79) Lin, Y. Z.; Clinton, G. M. A Soluble Protein Related to the HER-2 Proto-Oncogene Product Is Released from Human Breast Carcinoma Cells. *Oncogene* **1991**, *6* (4), 639–643.

(80) Sanderson, M. P.; Dempsey, P. J.; Dunbar, A. J. Control of ErbB Signaling through Metalloprotease Mediated Ectodomain Shedding of EGF-like Factors. *Growth Factors* **2006**, *24* (2), 121–136. https://doi.org/10.1080/08977190600634373.
(81) Reiss, K.; Saftig, P. The "A Disintegrin And Metalloprotease" (ADAM) Family of Sheddases: Physiological and Cellular Functions. *Seminars in Cell & Developmental Biology* **2009**, *20* (2), 126–137. https://doi.org/10.1016/j.semcdb.2008.11.002.
(82) Segatto, O.; King, C. R.; Pierce, J. H.; Di Fiore, P. P.; Aaronson, S. A. Different Structural Alterations Upregulate In Vitro Tyrosine Kinase Activity and Transforming Potency of the erbB-2 Gene. *Molecular and Cellular Biology* **1988**, *8* (12), 5570–5574. https://doi.org/10.1128/mcb.8.12.5570-5574.1988.
(83) D'Huyvetter, M.; De Vos, J.; Xavier, C.; Pruszynski, M.; Sterckx, Y. G. J.; Massa, S.; Raes, G.; Caveliers, V.; Zalutsky, M. R.; Lahoutte, T.; Devoogdt, N. 131I-Labeled Anti-HER2 Camelid sdAb as a Theranostic Tool in Cancer Treatment. *Clinical Cancer Research* **2017**, *23* (21), 6616–6628. https://doi.org/10.1158/1078-0432.CCR-17-0310.
(84) Elizabeth Fitzpatrick, B. J. C. Interference-Free HER2 ECD as a Serum Biomarker in Breast Cancer. *J Mol Biomark Diagn* **2013**, *04* (03). https://doi.org/10.4172/2155-9929.1000151.
(85) Rigakos, G.; Razis, E.; Koliou, G.-A.; Oikonomopoulos, G.; Tsolaki, E.; Sperinde, J.; Chrisafi, S.; Zarkavelis, G.; Pazarli, E.; Batistatou, A.; Kourea, H. P.; Papakostas, P.; Bafaloukos, D.; Asimakopoulou, N. I.; Res, E.; Kotsakis, A.; Pectasides, D.; Koutras, A.; Christodoulou, C.; Fountzilas, G. Evaluation of the Role of P95 HER2 Isoform in Trastuzumab Efficacy in Metastatic Breast Cancer. *Anticancer Research* **2021**, *41* (4), 1793–1802. https://doi.org/10.21873/anticanres.14945.
(86) Carney, W. P.; Neumann, R.; Lipton, A.; Leitzel, K.; Ali, S.; Price, C. P. Potential Clinical Utility of Serum HER-2/Neu Oncoprotein Concentrations in Patients with Breast Cancer. *Clin Chem* **2003**, *49* (10), 1579–1598. https://doi.org/10.1373/49.10.1579.
(87) Jackson, C.; Browell, D.; Gautrey, H.; Tyson-Capper, A. Clinical Significance of HER-2 Splice Variants in Breast Cancer Progression and Drug Resistance. *International Journal of Cell Biology* **2013**, *2013*, e973584. https://doi.org/10.1155/2013/973584.
(88) Pedersen, K.; Angelini, P.-D.; Laos, S.; Bach-Faig, A.; Cunningham, M. P.; Ferrer-Ramón, C.; Luque-García, A.; García-Castillo, J.; Parra-Palau, J. L.; Scaltriti, M.; y Cajal, S. R.; Baselga, J.; Arribas, J. A Naturally Occurring HER2 Carboxy-Terminal Fragment Promotes Mammary Tumor Growth and Metastasis. *Molecular and Cellular Biology* **2009**, *29* (12), 3319–3331. https://doi.org/10.1128/MCB.01803-08.
(89) Alajati, A.; Sausgruber, N.; Aceto, N.; Duss, S.; Sarret, S.; Voshol, H.; Bonenfant, D.; Bentires-Alj, M. Mammary Tumor Formation and Metastasis Evoked by a HER2 Splice Variant. *Cancer Research* **2013**, *73* (17), 5320–5327. https://doi.org/10.1158/0008-5472.CAN-12-3186.
(90) Castiglioni, F.; Tagliabue, E.; Campiglio, M.; Pupa, S. M.; Balsari, A.; Ménard, S. Role of Exon-16-Deleted HER2 in Breast Carcinomas. **2006**. https://doi.org/10.1677/erc.1.01047.

(91) Marchini, C.; Gabrielli, F.; Iezzi, M.; Zenobi, S.; Montani, M.; Pietrella, L.; Kalogris, C.; Rossini, A.; Ciravolo, V.; Castagnoli, L.; Tagliabue, E.; Pupa, S. M.; Musiani, P.; Monaci, P.; Menard, S.; Amici, A. The Human Splice Variant Δ16HER2 Induces Rapid Tumor Onset in a Reporter Transgenic Mouse. *PLOS ONE* **2011**, *6* (4), e18727. https://doi.org/10.1371/journal.pone.0018727.
(92) Shamieh, L. S.; Evans, A. J.; Denton, M. C.; Clinton, G. M. Receptor Binding Specificities of Herstatin and Its Intron 8-Encoded Domain. *FEBS Letters* **2004**, *568* (1–3), 163–166. https://doi.org/10.1016/j.febslet.2004.05.027.
(93) Doherty, J. K.; Bond, C.; Jardim, A.; Adelman, J. P.; Clinton, G. M. The HER-2/Neu Receptor Tyrosine Kinase Gene Encodes a Secreted Autoinhibitor. *Proceedings of the National Academy of Sciences* **1999**, *96* (19), 10869–10874. https://doi.org/10.1073/pnas.96.19.10869.
(94) Aigner, A.; Juhl, H.; Malerczyk, C.; Tkybusch, A.; Benz, C. C.; Czubayko, F. Expression of a Truncated 100 kDa HER2 Splice Variant Acts as an Endogenous Inhibitor of Tumour Cell Proliferation. *Oncogene* **2001**, *20* (17), 2101–2111. https://doi.org/10.1038/sj.onc.1204305.
(95) Scott, G. K.; Robles, R.; Park, J. W.; Montgomery, P. A.; Daniel, J.; Holmes, W. E.; Lee, J.; Keller, G. A.; Li, W. L.; Fendly, B. M. A Truncated Intracellular HER2/Neu Receptor Produced by Alternative RNA Processing Affects Growth of Human Carcinoma Cells. *Molecular and Cellular Biology* **1993**, *13* (4), 2247–2257. https://doi.org/10.1128/mcb.13.4.2247-2257.1993.
(96) Kumar Pal, S.; Pegram, M. Targeting HER2 Epitopes. *Seminars in Oncology* **2006**, *33* (4), 386–391. https://doi.org/10.1053/j.seminoncol.2006.04.004.
(97) Garrett, J. T.; Rawale, S.; Allen, S. D.; Phillips, G.; Forni, G.; Morris, J. C.; Kaumaya, P. T. P. Novel Engineered Trastuzumab Conformational Epitopes Demonstrate In Vitro and In Vivo Antitumor Properties against HER-2/Neu1. *The Journal of Immunology* **2007**, *178* (11), 7120–7131. https://doi.org/10.4049/jimmunol.178.11.7120.
(98) Hao, Y.; Yu, X.; Bai, Y.; McBride, H. J.; Huang, X. Cryo-EM Structure of HER2-Trastuzumab-Pertuzumab Complex. *PLOS ONE* **2019**, *14* (5), e0216095. https://doi.org/10.1371/journal.pone.0216095.
(99) Fabi, A.; Mottolese, M.; Segatto, O. Therapeutic Targeting of ERBB2 in Breast Cancer: Understanding Resistance in the Laboratory and Combating It in the Clinic. *J Mol Med* **2014**, *92* (7), 681–695. https://doi.org/10.1007/s00109-014-1169-7.
(100) Ubbiali, D.; Orlando, M.; Kovačič, M.; Iacobucci, C.; Semrau, M. S.; Bajc, G.; Fortuna, S.; Ilc, G.; Medagli, B.; Oloketuyi, S.; Storici, P.; Sinz, A.; Grandori, R.; de Marco, A. An Anti-HER2 Nanobody Binds to Its Antigen HER2 via Two Independent Paratopes. *International Journal of Biological Macromolecules* **2021**, *182*, 502–511. https://doi.org/10.1016/j.ijbiomac.2021.04.032.
(101) Fanning, S. W.; Walter, R.; Horn, J. R. Structural Basis of an Engineered Dual-Specific Antibody: Conformational Diversity Leads to a Hypervariable Loop Metal-Binding Site. *Protein Eng Des Sel* **2014**, *27* (10), 391–397. https://doi.org/10.1093/protein/gzu033.
(102) Genst, E. D.; Silence, K.; Decanniere, K.; Conrath, K.; Loris, R.; Kinne, J.; Muyldermans, S.; Wyns, L. Molecular Basis for the Preferential Cleft

Recognition by Dromedary Heavy-Chain Antibodies. *PNAS* **2006**, *103* (12), 4586–4591. https://doi.org/10.1073/pnas.0505379103.
(103) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78* (8), 1950–1958. https://doi.org/10.1002/prot.22711.
(104) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *The Journal of Chemical Physics* **2007**, *126* (1), 014101. https://doi.org/10.1063/1.2408420.
(105) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.
(106) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.
(107) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
(108) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L. Improved Prediction of Protein Side-Chain Conformations with SCWRL4. *Proteins* **2009**, *77* (4), 778–795. https://doi.org/10.1002/prot.22488.
(109) Vreven, T.; Hwang, H.; Weng, Z. Integrating Atom-Based and Residue-Based Scoring Functions for Protein–Protein Docking. *Protein Science* **2011**, *20* (9), 1576–1586. https://doi.org/10.1002/pro.687.
(110) Viswanath, S.; Ravikant, D. V. S.; Elber, R. Improving Ranking of Models for Protein Complexes with Side Chain Modeling and Atomic Potentials. *Proteins: Structure, Function, and Bioinformatics* **2013**, *81* (4), 592–606. https://doi.org/10.1002/prot.24214.
(111) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J Am Chem Soc* **2003**, *125* (7), 1731–1737. https://doi.org/10.1021/ja026939x.
(112) Sarti, E.; Granata, D.; Seno, F.; Trovato, A.; Laio, A. Native Fold and Docking Pose Discrimination by the Same Residue-Based Scoring Function. *Proteins* **2015**, *83* (4), 621–630. https://doi.org/10.1002/prot.24764.
(113) Veggiani, G.; de Marco, A. Improved Quantitative and Qualitative Production of Single-Domain Intrabodies Mediated by the Co-Expression of Erv1p Sulfhydryl Oxidase. *Protein Expression and Purification* **2011**, *79* (1), 111–114. https://doi.org/10.1016/j.pep.2011.03.005.
(114) Nguyen, V. D.; Hatahet, F.; Salo, K. E.; Enlund, E.; Zhang, C.; Ruddock, L. W. Pre-Expression of a Sulfhydryl Oxidase Significantly Increases the Yields of Eukaryotic Disulfide Bond Containing Proteins Expressed in the Cytoplasm of E.Coli. *Microbial Cell Factories* **2011**, *10* (1), 1. https://doi.org/10.1186/1475-2859-10-1.
(115) Djender, S.; Schneider, A.; Beugnet, A.; Crepin, R.; Desrumeaux, K. E.; Romani, C.; Moutel, S.; Perez, F.; de Marco, A. Bacterial Cytoplasm as an Effective Cell Compartment for Producing Functional VHH-Based Affinity Reagents and

Camelidae IgG-like Recombinant Antibodies. *Microbial Cell Factories* **2014**, *13* (1), 140. https://doi.org/10.1186/s12934-014-0140-1.
(116) Studier, F. W. Protein Production by Auto-Induction in High-Density Shaking Cultures. *Protein Expression and Purification* **2005**, *41* (1), 207–234. https://doi.org/10.1016/j.pep.2005.01.016.
(117) La Manna, S.; Lopez-Sanz, L.; Mercurio, F. A.; Fortuna, S.; Leone, M.; Gomez-Guerrero, C.; Marasco, D. Chimeric Peptidomimetics of SOCS 3 Able to Interact with JAK2 as Anti-Inflammatory Compounds. *ACS Med. Chem. Lett.* **2020**, *11* (5), 615–623. https://doi.org/10.1021/acsmedchemlett.9b00664.
(118) Huynh, K.; Partch, C. L. Analysis of Protein Stability and Ligand Interactions by Thermal Shift Assay. *Current Protocols in Protein Science* **2015**, *79* (1), 28.9.1-28.9.14. https://doi.org/10.1002/0471140864.ps2809s79.
(119) Stura, E. A.; Nemerow, G. R.; Wilson, I. A. Strategies in the Crystallization of Glycoproteins and Protein Complexes. *Journal of Crystal Growth* **1992**, *122* (1–4), 273–285. https://doi.org/10.1016/0022-0248(92)90256-I.
(120) De Genst, E.; Silence, K.; Decanniere, K.; Conrath, K.; Loris, R.; Kinne, J.; Muyldermans, S.; Wyns, L. Molecular Basis for the Preferential Cleft Recognition by Dromedary Heavy-Chain Antibodies. *Proceedings of the National Academy of Sciences* **2006**, *103* (12), 4586–4591. https://doi.org/10.1073/pnas.0505379103.
(121) Salton, M. R. J. Cell Wall of Micrococcus Lysodeikticus as the Substrate of Lysozyme. *Nature* **1952**, *170* (4331), 746–747. https://doi.org/10.1038/170746a0.
(122) De Genst, E.; Silence, K.; Decanniere, K.; Conrath, K.; Loris, R.; Kinne, J.; Muyldermans, S.; Wyns, L. Molecular Basis for the Preferential Cleft Recognition by Dromedary Heavy-Chain Antibodies. *Proc Natl Acad Sci U S A* **2006**, *103* (12), 4586–4591. https://doi.org/10.1073/pnas.0505379103.
(123) Ashrafi, N.; Shareghi, B.; Farhadian, S.; Hosseini-Koupaei, M. The Effect of Putrescine on the Lysozyme Activity and Structure: Spectroscopic Approaches and Molecular Dynamic Simulation. *Colloids and Surfaces B: Biointerfaces* **2022**, *213*, 112402. https://doi.org/10.1016/j.colsurfb.2022.112402.
(124) Monegal, A.; Ami, D.; Martinelli, C.; Huang, H.; Aliprandi, M.; Capasso, P.; Francavilla, C.; Ossolengo, G.; de Marco, A. Immunological Applications of Single-Domain Llama Recombinant Antibodies Isolated from a Naïve Library. *Protein Engineering, Design and Selection* **2009**, *22* (4), 273–280. https://doi.org/10.1093/protein/gzp002.
(125) Sormanni, P.; Aprile, F. A.; Vendruscolo, M. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *Journal of Molecular Biology* **2015**, *427* (2), 478–490. https://doi.org/10.1016/j.jmb.2014.09.026.
(126) Caporale, A.; Doti, N.; Monti, A.; Sandomenico, A.; Ruvo, M. Automatic Procedures for the Synthesis of Difficult Peptides Using Oxyma as Activating Reagent: A Comparative Study on the Use of Bases and on Different Deprotection and Agitation Conditions. *Peptides* **2018**, *102*, 38–46. https://doi.org/10.1016/j.peptides.2018.02.006.
(127) Reinwarth, M.; Glotzbach, B.; Tomaszowski, M.; Fabritz, S.; Avrutina, O.; Kolmar, H. Oxidative Folding of Peptides with Cystine-Knot Architectures:

Kinetic Studies and Optimization of Folding Conditions. *ChemBioChem* **2013**, *14* (1), 137–146. https://doi.org/10.1002/cbic.201200604.
(128) Zimmermann, I.; Egloff, P.; Hutter, C. A. J.; Kuhn, B. T.; Bräuer, P.; Newstead, S.; Dawson, R. J. P.; Geertsma, E. R.; Seeger, M. A. Generation of Synthetic Nanobodies against Delicate Proteins. *Nat Protoc* **2020**, *15* (5), 1707–1741. https://doi.org/10.1038/s41596-020-0304-x.
(129) Dai, R.; Geders, T. W.; Liu, F.; Park, S. W.; Schnappinger, D.; Aldrich, C. C.; Finzel, B. C. Fragment-Based Exploration of Binding Site Flexibility in Mycobacterium Tuberculosis BioA. *J Med Chem* **2015**, *58* (13), 5208–5217. https://doi.org/10.1021/acs.jmedchem.5b00092.
(130) Lauwereys, M.; Arbabi Ghahroudi, M.; Desmyter, A.; Kinne, J.; Hölzer, W.; De Genst, E.; Wyns, L.; Muyldermans, S. Potent Enzyme Inhibitors Derived from Dromedary Heavy-Chain Antibodies. *EMBO J* **1998**, *17* (13), 3512–3520. https://doi.org/10.1093/emboj/17.13.3512.
(131) Salton, M. R. J. Cell Wall of Micrococcus Lysodeikticus as the Substrate of Lysozyme. *Nature* **1952**, *170* (4331), 746–747. https://doi.org/10.1038/170746a0.
(132) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 2005; pp 571–607. https://doi.org/10.1385/1-59259-890-0:571.
(133) Wang, P.; Qin, W.; Xu, J.; Yan, Y.; Tian, J.; Wu, N.; Yao, B. Enhancing the Soluble Expression of an Amylase in *Escherichia Coli* by the Mutations Related to Its Domain Interactions. *Protein Expression and Purification* **2016**, *120*, 35–41. https://doi.org/10.1016/j.pep.2015.12.010.
(134) Ramon, A.; Señorale, M.; Marin, M. Inclusion Bodies: Not That Bad.... *Front. Microbiol.* **2014**, *5*. https://doi.org/10.3389/fmicb.2014.00056.
(135) Jürgen, B.; Breitenstein, A.; Urlacher, V.; Büttner, K.; Lin, H.; Hecker, M.; Schweder, T.; Neubauer, P. Quality Control of Inclusion Bodies in Escherichia Coli. *Microb Cell Fact* **2010**, *9* (1), 41. https://doi.org/10.1186/1475-2859-9-41.
(136) Palmer, I.; Wingfield, P. T. Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from Escherichia Coli. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **2004**, *CHAPTER*, Unit. https://doi.org/10.1002/0471140864.ps0603s38.
(137) Bhatwa, A.; Wang, W.; Hassan, Y. I.; Abraham, N.; Li, X.-Z.; Zhou, T. Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia Coli and Strategies to Address Them for Industrial Applications. *Front. Bioeng. Biotechnol.* **2021**, *9*. https://doi.org/10.3389/fbioe.2021.630551.
(138) Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Molina-Abad, E.; Moreno, E. NbThermo: A New Thermostability Database for Nanobodies. *Database* **2023**, *2023*, baad021. https://doi.org/10.1093/database/baad021.
(139) Kinoshita, S.; Nakakido, M.; Mori, C.; Kuroda, D.; Caaveiro, J. M.; Tsumoto, K. Molecular Basis for Thermal Stability and Affinity in a VHH: Contribution of the Framework Region and Its Influence in the Conformation of the CDR3. *Protein Science : A Publication of the Protein Society* **2022**, *31* (11), e4450. https://doi.org/10.1002/pro.4450.

(140) Buscajoni, L.; Martinetz, M. C.; Berkemeyer, M.; Brocard, C. Refolding in the Modern Biopharmaceutical Industry. *Biotechnology Advances* **2022**, *61*, 108050. https://doi.org/10.1016/j.biotechadv.2022.108050.
(141) Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H. The Kinetics of Formation of Native Ribonuclease during Oxidation of the Reduced Polypeptide Chain. *Proceedings of the National Academy of Sciences* **1961**, *47* (9), 1309–1314. https://doi.org/10.1073/pnas.47.9.1309.
(142) Hamidi, S. R.; Safdari, Y.; Sheikh Arabi, M. Test Bacterial Inclusion Body for Activity Prior to Start Denaturing and Refolding Processes to Obtain Active Eukaryotic Proteins. *Protein Expression and Purification* **2019**, *154*, 147–151. https://doi.org/10.1016/j.pep.2018.10.013.
(143) Xu, L.; Song, X.; Jia, L. A Camelid Nanobody against EGFR Was Easily Obtained through Refolding of Inclusion Body Expressed in Escherichia Coli. *Biotechnology and Applied Biochemistry* **2017**, *64* (6), 895–901. https://doi.org/10.1002/bab.1544.
(144) Moghadam, M.; Ganji, A.; Varasteh, A.; Falak, R.; Sankian, M. Refolding Process of Cysteine-Rich Proteins:Chitinase as a Model. *Rep Biochem Mol Biol* **2015**, *4* (1), 19–24.

PART B

1. INTRODUCTION

The Human Immunodeficiency Virus (HIV) weakens the immune system, making individuals more susceptible to infections and diseases. HIV-1 is the prevalent strain, responsible for most global HIV infections. Without treatment, HIV-1 can progress to AIDS, the most advanced stage, marked by a CD4+ cell count below 200 cells/μL or the emergence of specific infections or cancers. AIDS significantly reduces life expectancy if untreated[1]. HIV has been a global health crisis since the 1980s, claiming approximately 42.3 million lives. By the end of 2023, an estimated 39.9 million people were living with HIV, with 630,000 deaths and 1.3 million new infections that year[2].

While no cure is yet available, antiretroviral therapy (ART) helps manage HIV as a chronic condition, allowing those affected to live long, healthy lives by inhibiting viral replication.

1.1 HIV-1 description and replication cycle

HIV is a single-stranded RNA (ssRNA) retrovirus belonging to Retroviridae family, specifically within the Orthoretrovirinae subfamily and the Lentivirus genus. The HIV1 strain is divided into four groups (Major, Outlier, Non-m/non-o, and P), each resulting from unique cross-species transmissions. With 19 recombinant forms reported for the Major group, the HIV mutation rate is notably high.

HIV-1 particles measure ~80-100 nm and are enveloped in a phospholipid bilayer, containing the ENV spike glycoprotein[3]. Within this envelope, the truncated cone-shaped capsid core is surrounded by a matrix protein layer. The capsid contains two copies of the RNA-protein complex and three key viral enzymes: aspartic protease, reverse transcriptase, and integrase[4].

This virus infects CD4+ T cells, a type of white blood cells essential for the immune response. Once the virus enters the host organism, it binds to CD4+ T cells, integrates its genetic material into the host DNA, and replicates, destroying infected cells. As the number of CD4+ cells decreases, the infected person becomes increasingly susceptible to opportunistic infections and cancers, which would be effectively fought off by a healthy immune system.

The HIV-1 replication cycle[5,6] begins with the attachment of the virus to the host CD4+ T cell. The glycoprotein gp120 present on the virus surface binds specifically to the host CD4 receptor. This interaction induces a conformational change of gp120, allowing it to interact with a coreceptor, CCR5 or CXCR4, present on the surface of the host cell. When

this double binding occurs, the fusion process with the cell membrane begins and the viral capsid enters the cytoplasm. Once inside the cell, the capsid disassembles, releasing the single-stranded viral RNA and its enzymes.

The viral reverse transcriptase (RT) synthesizes double-stranded DNA complementary to the viral RNA (cDNA)[7]. The cDNA is thus transported to the nucleus and integrated into a random position of the host genome by the enzyme integrase. The provirus (the integrated viral genome) is transcribed into mRNA and then exported to the cytoplasm. Here, host ribosomes translate HIV-1 mRNA into Gag (Group-specific antigen) and Gag-Pol polyproteins in a 20:1 ratio (Figure 1).

The viral aspartic protease is formed through autoprotolysis of GagPol. This enzyme is key for the following maturation of the polyprotein and the production of functional HIV-1 proteins. The GagPol cleavage process produces the essential components of the virion: the Gag protein (forming the viral core) and the enzymes reverse transcriptase and integrase. The envelope glycoproteins gp120 and gp41 are produced from *env* gene and translocated on the cell membrane (Figure 1).

The new virion is assembled near the surface of the host cell membrane. The Gag protein forms a protective shell around the viral RNA and the enzymes forming the capsid core. The new viral particle is then coated with an envelope incorporating *env*-encoded proteins.

The maturation of GagPol is essential for producing infectious HIV particles, with aspartic protease playing a critical role in this process. Targeting this enzyme can effectively prevent virions from maturing into infectious HIV particles. Since the late 20th century, protease inhibitors have become a central focus in antiretroviral therapy (ART) due to the enzyme pivotal role in HIV replication, leading pharmaceutical companies to prioritize these drugs in HIV treatment development.

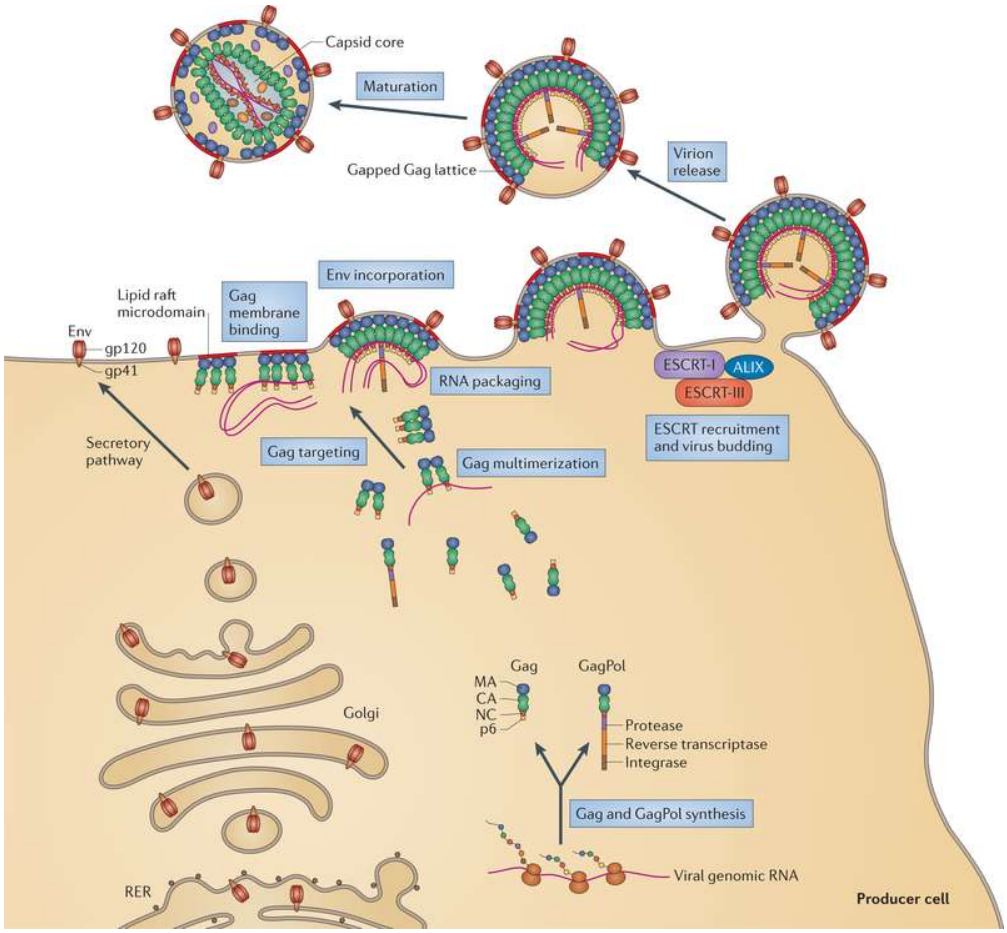


Figure 1: **Schematic HIV-1 replication cycle**. The polyproteins Gag and GagPol are produced from the viral RNA and their proteolytic cleavage produces three enzymes (protease, reverse transcriptase and integrase). The envelope glycoproteins gp120 and gp41 are produced from the *env* gene. Once matured through proteolysis, the Gag protein assembles into a coating of the viral RNA and forms the capsid core. The *env* proteins are incorporated into the viral envelope before the new viral particle is release from the cell. (From Freed E.O., 2015.[5])

1.2 Anti-retroviral pharmacological targets in HIV-1 therapy

Currently, antiviral drugs against HIV approved by the Food and Drug Administration (FDA) act on different stages of the viral replication cycle and are classified according to their therapeutic target[8–10].

Nucleoside and non-nucleoside reverse transcriptase inhibitors (NRTIs and NNRTIs) bind to the viral reverse transcriptase, the enzyme used by HIV to convert its RNA into DNA, inhibiting its activity and preventing the virus from using cellular mechanisms to replicate. NRTIs became the first HIV treatment available in 1987, followed by four additional FDA-approved options. These inhibitors act as nucleoside analogs and must undergo triphosphorylation in the host cell to integrate into the viral DNA. Lacking a 3' hydroxyl group, they cause the termination of the growing DNA chain, disrupting HIV replication. However, NRTIs carry a risk of mitochondrial toxicity, potentially causing lactic acidosis,

hepatic steatosis, myopathy, neuropathy, and lipoatrophy. Resistance can develop as HIV reduces NRTI recognition and incorporation, or the rate of excision of integrated NRTIs by the proof-reading function of the DNA polymerase increases[11]. NNRTIs, on the other hand, are small allosteric inhibitors designed to precisely target HIV-1 reverse transcriptase, decreasing off-target and side effects compared to NRTIs. By binding to a pocket near the active site of the reverse transcriptase, NNRTIs allosterically restrict the movement of active residues or block access to the catalytic pocket, effectively inhibiting viral replication. However, their high specificity leads to the rapid development of resistance, particularly if treatment is inconsistent or used as monotherapy. Resistant HIV variants often present mutations in the NNRTI binding pocket, diminishing the effectiveness of the inhibitors[12].

Inhibitors of virus attachment to the cell surface interfere with the early interaction between the gp120 protein on the outer surface of HIV and the CD4 receptor on the host cell, preventing the virus from binding to and entering CD4 cells, while post-attachment inhibitors prevent the virus from binding to CCR5 and CXCR4 coreceptors. Fusion inhibitors prevent fusion between the HIV envelope and the host CD4 cell membrane[13]; enfuvirtide[14] is the only approved drug belonging to this class. Similar in function, but different in mechanisms, CCR5 antagonists block the CCR5 coreceptor, preventing the virus from entering the cell. Among these, the most popular CCR5 antagonist is Maraviroc[15].

Capsid inhibitors interfere with capsid assembly or cause premature uncoating, blocking proper virus replication. This is the most recently discovered class of drugs targeting HIV-1. Lenacapavir[16], so far its only member, binds to the interface between two capsid protein subunits, CA1 and CA2, specifically at the junction between the N-terminal and the C-terminal domains. By stabilizing the interaction between CA subunits, this drug prevents proper capsid disassembly necessary for reverse transcription. This results in malformed capsids that cannot effectively facilitate viral replication.

Other drugs target the viral enzymes. Besides the mentioned reverse transcriptase inhibitors, drugs targeting the integrase hinder the insertion of viral cDNA into the host cell DNA. At last, protease inhibitors block proteolytic activity, preventing the cleavage of proteins needed for viral particle maturation. Drugs belonging to the latter class are competitive inhibitors that mimic the Gag/GagPol polyprotein, the natural protease substrate, but have a stronger affinity for the catalytic pocket than the substrate.

1.3 HIV.1 protease

HIV-1 protease is a member of the aspartic protease family, a class of proteolytic enzymes active at acidic pH[17–19]. The active enzyme is a dimer composed of two identical subunits, each consisting of 99 amino acid residues[20].

The β-sheet are the predominant secondary structure of HIV-1 protease, apart from a single α-helix in the *h′* segment comprising residues 86-94 (Figure 2, left panel). In the substrate-free form (*apo*), the interface between the two monomers consists of 4 β-sheets, two belonging to each dimer: the N-terminal β-strand, referred to as the sheet *a* (residues 1-4) forms the outer part of the dimer interface and interacts with the sheet *q* of the opposite monomer, while the C-terminal β-strand *q* (residues 95-99) forms the inner part, interacting with the sheets *a* and *q* of the opposite monomer (Figure 2, right panel). Segments *b* and *c* (residues 9-15 and 18-24, respectively) align to form a β-sheet that play a crucial structural role. Segment *c* contains the enzyme active site triad at its end, a highly conserved sequence comprising Asp25-Thr26-Gly27 (Figure 2, inset). In this triad, the Asp25 residue is particularly significant, as it is the catalytic site essential for the enzyme activity. These residues are located in a loop whose structure is stabilized by hydrogen bonds with both monomers[21]. The active site loop terminates in segment *d* (residues 30-35), followed by segment *h* (residues 36-42), with an undefined secondary structure. Segment *h* is quite distorted compared to the other aspartic proteases and forms a large loop, followed by three consecutive β-strands, segments *a′* (residues 43-49), *b′* (residues 52-66), and *c′* (residues 69-78); the first two strands are involved in a β-hairpin structure, known as flap, that stabilizes the interaction with the substrate. The peptide chain continues with a loop (residues 79-82) and segment *d′* (residues 83-85), to the helix *h′* (residues 86-94), ending with the C-terminal leaflet β-strand *q*[18,22].

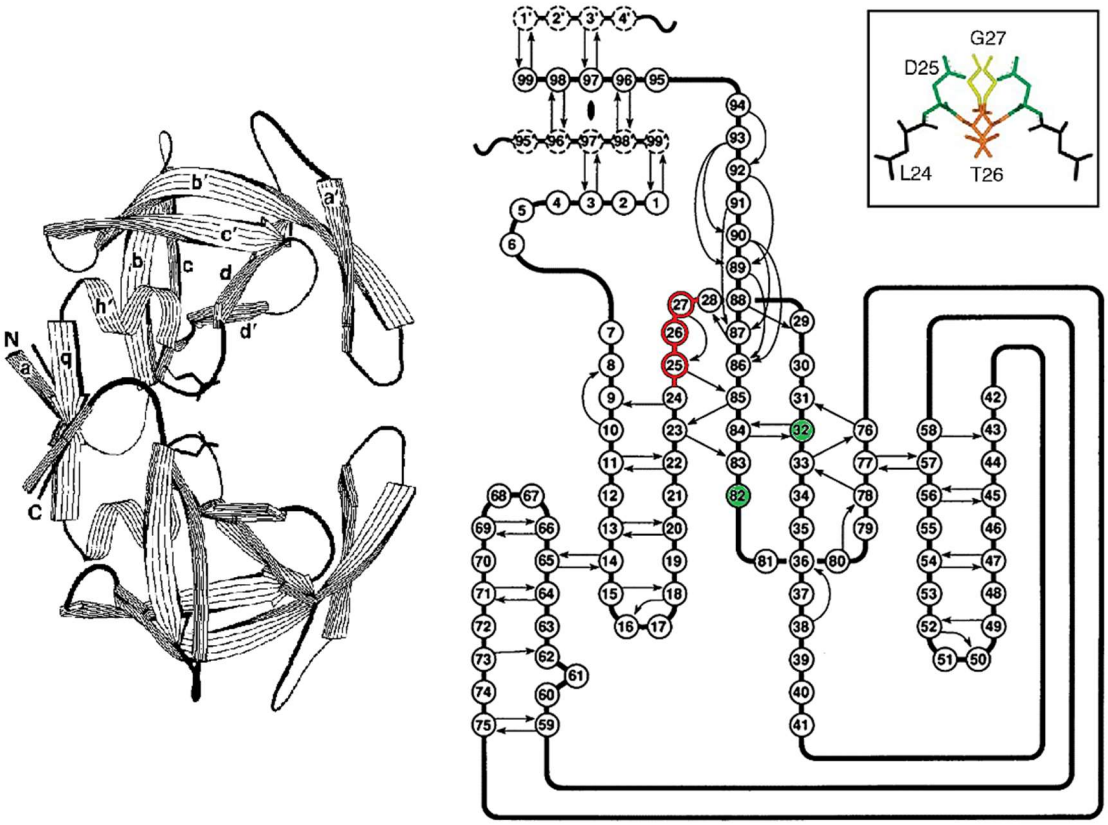


Figure 2: **HIV-1 quaternary structure**. Left panel: HIV-1 protease in the open *apo* form; letters indicate the secondary structure elements. Right panel: HIV-1 protease secondary structure with the catalytic triad highlighted in red and the residues whose mutations are the subject of this study in green. Arrows indicate hydrogen bonds and dashed circles correspond to the residues of the opposite monomer. Inset: the catalytic triad composed of Asp25-Thr26-Gly27, in the fireman's grip conformation. (From Wlodawer A. *et al.*, 1993;[22] Wlodawer A. et al., 1989;[23] and Louis, J. M. et al., 2007.[24])

The active site is located at the dimer interface, stabilized by the "fireman's grip" conformation, i.e., a network of hydrogen bonds involving the two subunits. The oxygen atom of the side chain of threonine 26 accepts a hydrogen bond from the amide group of the main chain of the corresponding residue (Thr 26') of the opposite monomer, while the amide nitrogen of the glycine 27 residue donates a hydrogen bond to a carbonyl oxygen atom of the side chain of the catalytic aspartate 25, in both monomers (Figure 3)[23]. The two long flaps, formed by flexible β-hairpins and located on the opposite side of the active pocket, regulate substrate access to the catalytic site[22].

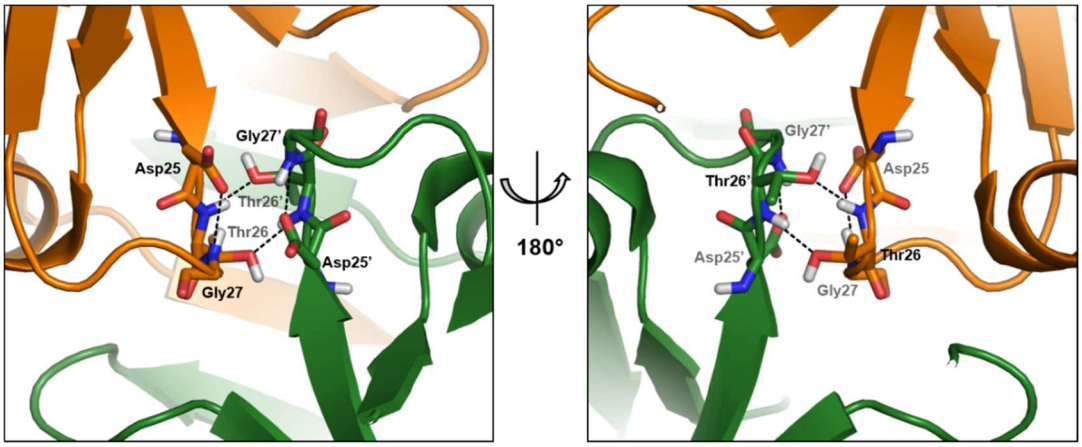


Figure 3: **The fireman's grip**. Crystal structure of the active site of HIV-1 protease (PDB 1HVR[25]), showing the hydrogen bond interactions (dashed lines) that characterized the conformation known as "fireman's grip". Residues 25-27 of the two monomers (orange and green) are represented as sticks. The right panel shows a view generated by rotating the model in the left panel by 180° around the vertical axis.

In the *apo* form (i.e., without substrate or inhibitor), the HIV-1 protease flap region can adopt either an "open" or "curled" conformation. In the open conformation (Figure 4, left panel), the flaps are 7.7 Å apart, with no hydrogen bonds or other stabilizing interactions between the two chains. In the curled conformation, the ends of the two flaps engage in hydrophobic interactions, involving specifically the residues phenylalanine 53 on one chain and isoleucine 50' on the opposing chain, with each residue interacting across the chains, stabilizing the flap region[26].

Upon binding of a substrate to the protein, a significant conformational change occurs in which the flap tips of each monomer close the active site (*holo* form) (Figure 4, right panel). The major flap movement involves residues Lys45, Met46, Gly52, and Phe53 on each monomer[27].

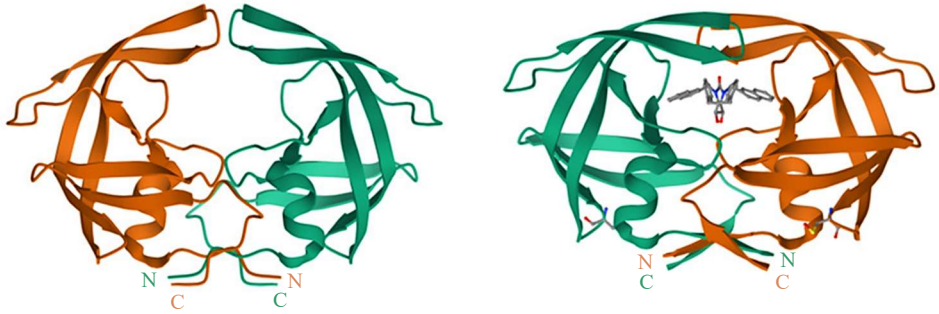


Figure 4: ***Apo* and *holo* HIV-1 protease structures**. Left panel: crystal structure of the *apo* conformation of the HIV-1 protease (PDB: 1HHP)[28]. Right panel: crystal structure of the HIV-1 proteasi in *holo* conformation, in complex with an inhibitor (PDB: 1HVR)[29]. Monomers are shown with different colors.

The function of HIV-1 protease is to hydrolyze the viral polyproteins Gag and GagPol, thereby forming functional proteins. The active protease is formed from the Gag-Pol polyprotein by autoprotolysis (Figure 5).

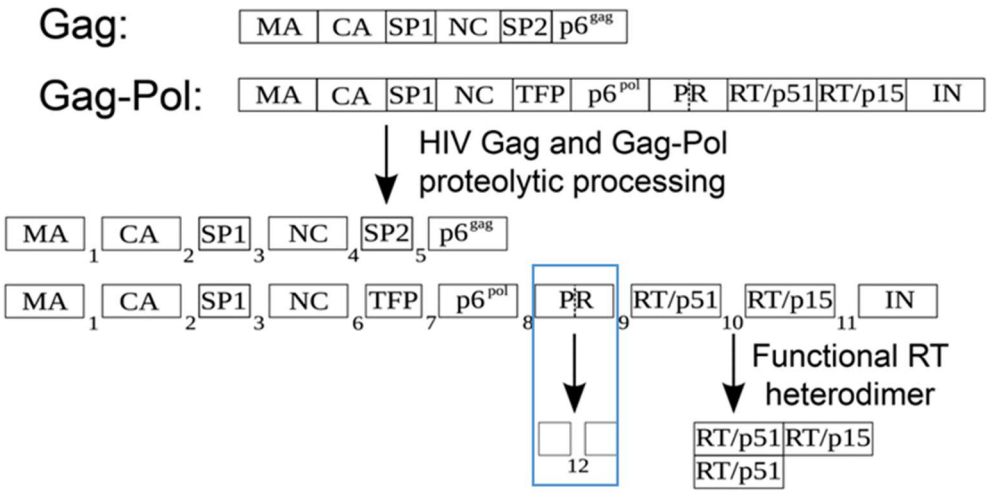


Figure 5: **Proteolytic cleavage sites of the HIV-1 protease**. The Gag and GagPol polyproteins contain 5 and 9 specific sites recognized by the protease, respectively. In addition, GagPol contains another site within the protease sequence whose hydrolytic cleavage leads to the inactivation of the enzyme (blue box). Protein sequences are marked as follows: MA, matrix protein; CA, capsid protein; SP1 and SP2, spacer peptides; NC, nucleocapsid; TFP, transframe peptide; PR, protease; RT, reverse transcriptase; IN, integrase. (From Könnyű B. *et al.*, 2013.[30])

The protease reaction mechanism involves an oxyanionic tetrahedral intermediate, stabilized by the structure of the active site and whose existence has been experimentally confirmed by neutron crystallography (NC) data[31] . In the resting state of the enzyme, a single aspartate residue (Asp25) in the catalytic site is deprotonated (Figure 6, apo form). When the catalytic mechanism starts, this residue is thought to be responsible for the activation of a water molecule, that becomes a base after deprotonation. The activated water molecule performs a nucleophilic attack on the carbonyl carbon atom of the substrate peptide bond, generating an oxoanionic tetrahedral intermediate (Figure 6), which is stabilized by hydrogen bonds formed with the catalytic site. The next reaction step consists in the cleavage of the peptide bond, generating the products of the catalytic reaction (Figure 6, on the right)[32].

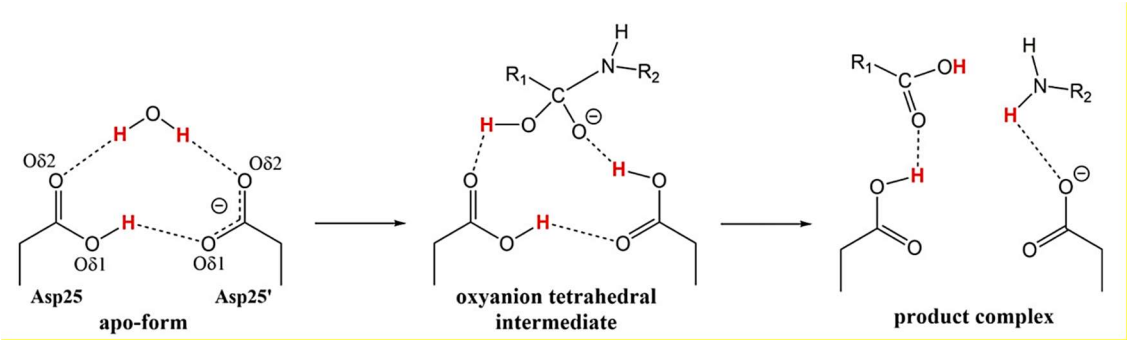


Figure 6: **HIV-1 protease catalytic mechanism**. Hydrogen atoms involved in the reaction are highlighted in red, hydrogen bonds are represented as dashed lines[32].

1.4 Protease inhibitors (PIs)

HIV-1 protease is a crucial enzyme for viral maturation and infectivity, making it an ideal target for antiviral drug development. Most protease inhibitors (PIs) are peptidomimetics designed to mimic the enzyme natural substrate, closely resembling the tetrahedral intermediate formed during peptide cleavage (Figure 7, Table 1). Unlike natural substrates, these inhibitors feature a hydroxyethylene structure, which is resistant to cleavage, allowing them to effectively block the active site. An exception is the drug Tipranavir, which targets flap residues, stabilizing the closed state and reducing the flexibility required for the enzyme activity[33].

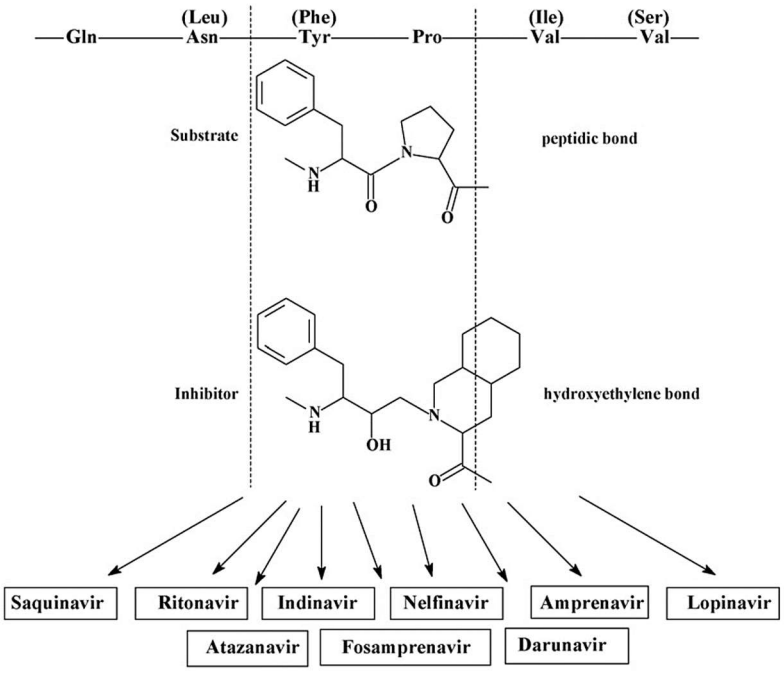
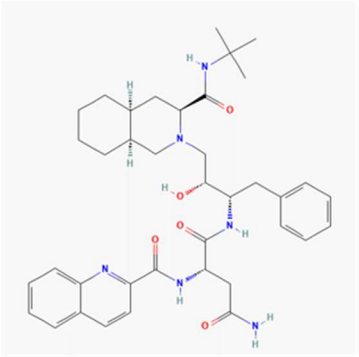
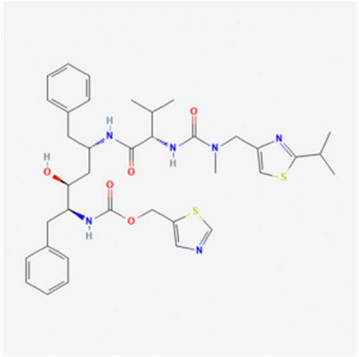


Figure 7: **Peptidomimetic HIV-1 protease inhibitors**[34]. The scaffold of these inhibitors (lower structure) resembles that of the natural substrate of the enzyme (upper structure), except for a hydroxyethylenic moiety in place of the carbonyl carbon.

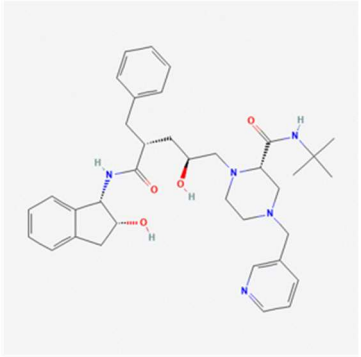
***Table 1**: Protease Inhibitors. FDA approved HIV-1 protease inhibitors[34].*



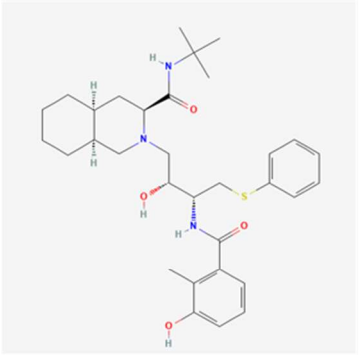
Saquinavir (SQV)
Commercial brand: Invirase
1995



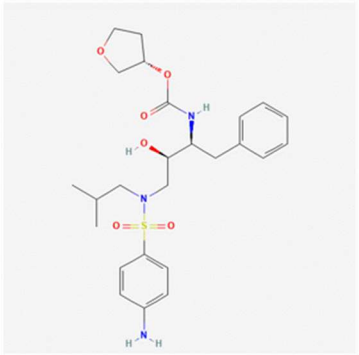
Ritonavir (RTV)
Commercial brand: Norvir
1996



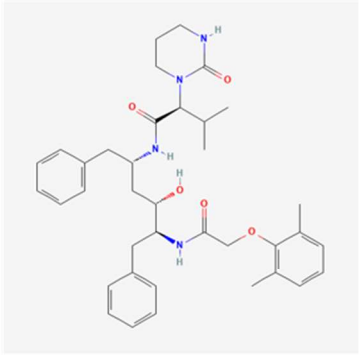
Indinavir (IDV)
Commercial brand: Crixivan
1996



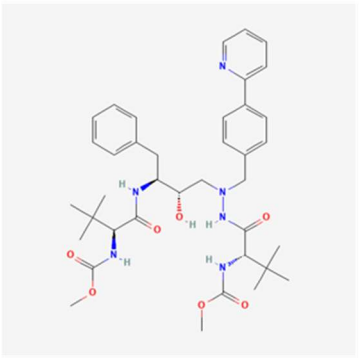
Nelfinavir (NFV)
Commercial brand: Viracept
1997



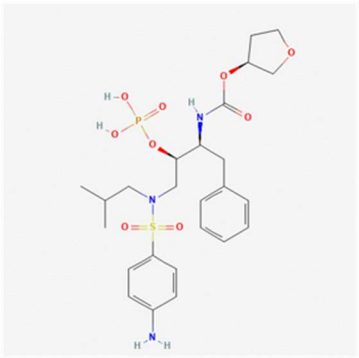
Amprenavir (APV)
Commercial brand: Agenerase
1999



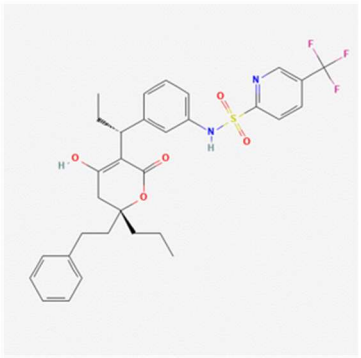
Lopinavir (LPV)
Commercial brand: Kaletra
2000



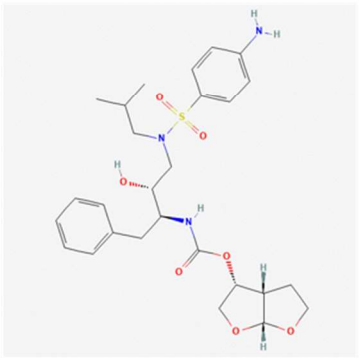
Atazanavir (ATV)
Commercial brand: Evotaz, Reyataz
2003



Fosamprenavir (APV)
Commercial brand: Lexiva, Telzir
2003



Tipranavir (TPV)
Commercial brand: Aptivus
2005



Darunavir (DRV)
Commercial brand: Prezista
2006

All currently available drugs are competitive inhibitors, reversibly binding to the enzyme active site and blocking the access of the Gag-Pol polyprotein. This inhibition prevents the formation of mature, infectious viral particles, thereby disrupting the viral life cycle[34]. The binding affinity of protease inhibitors (PIs) to the enzyme active site is largely attributed to the flexibility of their main chains. This flexibility enables the inhibitors to establish both polar and predominantly nonpolar interactions between their side chains and the enzyme catalytic pocket subsites (Figure 8). The catalytic pocket in the enzyme dimer is divided into six subsites (S1, S1', S2, S2', S3, and S3'), plus two additional, low-specificity subsites (S4 and S4'). Subsites with the same number are structurally equivalent in the *apo* form but the structure becomes asymmetrical upon binding of an asymmetric substrate or inhibitor and the subsites lose their equivalence. The central subsites, S1 and S1', include the Asp25 residues from each monomer, which are key to the catalytic function. The side chains of the substrate residues adjacent to the cleaved peptide bond (P1 and P1') interact specifically with the S1 and S1' subsites, ensuring precise binding and orientation within the catalytic channel.

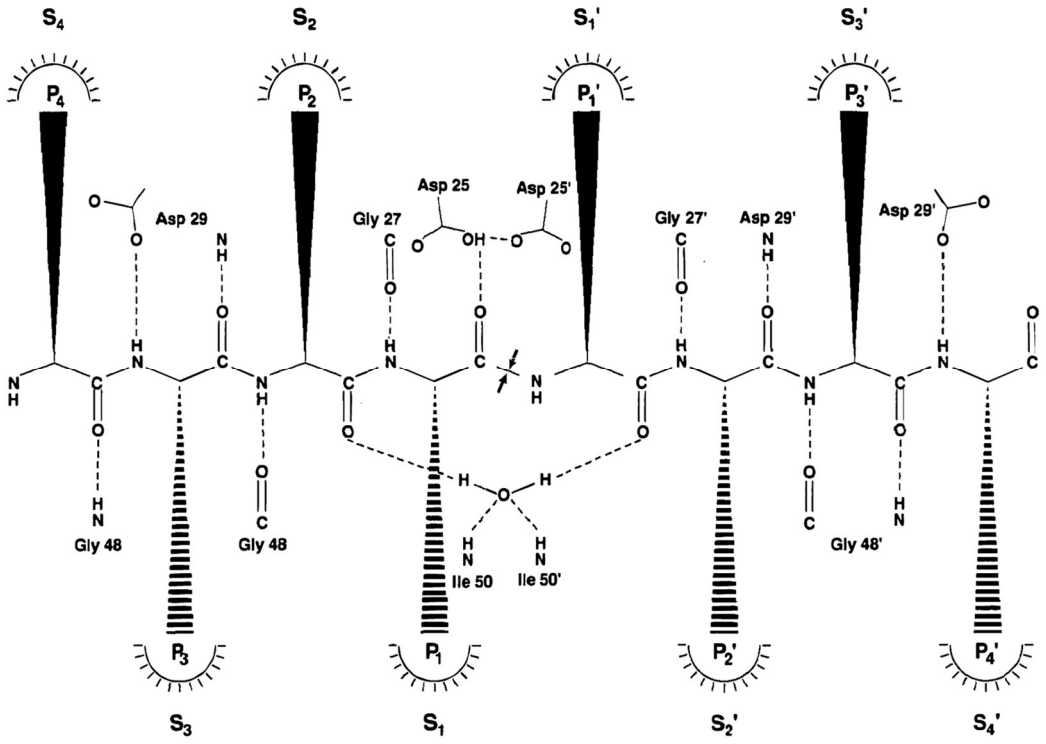


Figure 8: **Interaction between the substrate and the HIV-1 protease binding pocket**. Schematic representation of a polypeptide chain within the binding site of the HIV-1 protease, highlighting the peptide bond subject to hydrolysis (black arrows) and the main hydrogen bond interactions between the substrate and the HIV protease. The enzyme subsites are labeled S1, S1', S2, S2', S3, and S3', while the substrate residues with which they interact are labeled P1, P1', P2, P2', P3, P3', respectively.

PIs have relatively similar chemical structures (Table 1), with some common features. In most inhibitors, the presence of a secondary hydroxyl group allows the formation of stabilizing interactions with the aspartate residues of the protease catalytic site (Asp25 and Asp25'). Early studies on Saquinavir found that the R configuration of the inhibitor hydroxyethylenic carbon establishes more effective binding interactions with the enzyme compared to the S configuration. However, this finding cannot be extended to all compounds, as further studies on Ritonavir and Indinavir identified that the optimal configuration for these two inhibitors actually has an S configuration at this stereocenter[18].

In all crystallographic studies of complexes between the HIV-1 protease and peptidomimetic inhibitors, a water molecule is present in the active site of the enzyme, forming 4 hydrogen bonds that bridge oxygen atoms of the inhibitor with the amide groups of residues of isoleucine 50 of the protease (Ile50)[35].

By the end of December 2022, 29.8 million people had access to antiretroviral therapy, significantly more than the 7.7 million of 2010. Unfortunately, the increased use of antiretroviral therapy has been accompanied by the emergence of anti-HIV drug-resistant variants of the virus. These variants are often characterized by point mutations in the viral enzymes that impair the inhibitory activity of available drugs. All antiretroviral drugs, including those in new and emerging classes, are at risk of becoming partially or completely inactive due to the emergence of viral variants that exhibit drug resistance[2]. In particular, due to the similarity of the protease inhibitors, it is unfortunately common for viral variants to exhibit resistance to more than one inhibitor (multidrug resistance). In addition, multidrug resistance virions may emerge as a result of homologous recombination (Figure 9)[36].

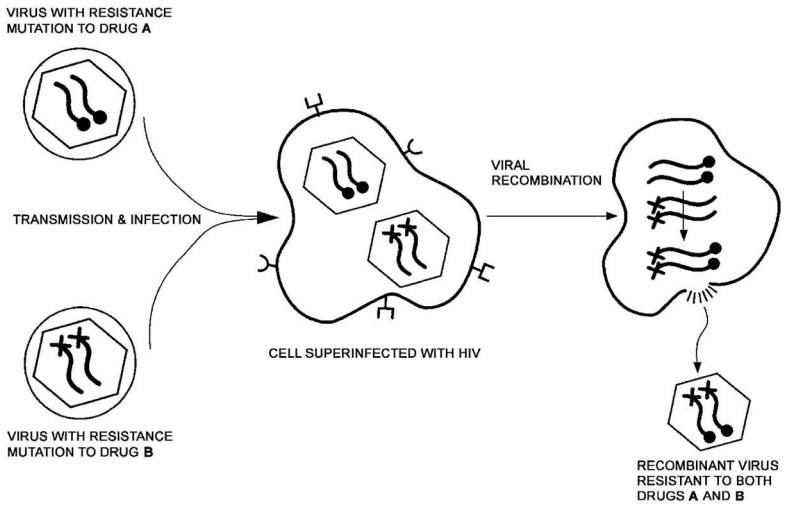
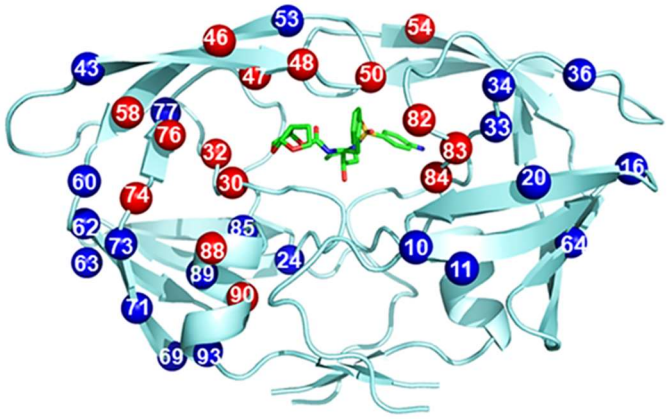


Figure 9: **Development of multidrug resistant variants.** Homologous recombination drives the insurgence of multidrug-resistant recombinant viral variants. (From Blackard J.T. *et al.*, 2002.[36])

1.5 HIV-1 protease mutations

The emergence of drug resistance is a significant challenge to the effectiveness of PI therapy. This resistance develops as a result of random mutations introduced into the viral genome. Some viral variants acquire mutations that do not compromise the enzyme activity but diminish its affinity for inhibitors[37]. Mutations that induce resistance lead to an increase in the inhibition constant (K_i), indicating reduced performance of the inhibitors (Figure 10 and Table 2)[35]. This resistance can accumulate through an evolutionary mechanism, resulting in the emergence of new viral variants with enhanced survival capabilities. Notably, mutations affecting more than half of the residues in the protease have been identified, highlighting the extraordinary enzyme adaptability and the challenges faced in developing effective therapies[38].



Major Protease Inhibitor (PI) Resistance Mutations											
	32	46	47	48	50	54	76	82	84	88	90
Cons	V	M	I	G	I	I	L	V	I	N	L
ATV/r	I	IL	V	VM	L	VLM		ATFS	V	DS	M
DRV/r	I		VA		V	LM	V	F	V		
LPV/r	I	IL	VA	VM	V	VLM	V	AFTS	V		M

Figure 10: **HIV-1 PR mutations**. Left panel: Representation of the crystal structure of the HIV-1 protease-Darunavir complex highlighting the position of mutations conferring pharmacoresistance. Red color denotes a mutation inducing significant drug resistance, while minor mutations are in blue. Right panel: mutations that confer to the protease resistance to some of the FDA-approved inhibitors[39].

***Table 2**: Affinity constants of FDA-approved protease inhibitors. For each inhibitor, the main mutation conferring resistance are considered.*

Inhibitor	*K*i (nM)			
	WT/Q7K	L10I, G48V, I54V, L63P, V82A	D30N, L63P, N88D	L10I, L63P, A71V, G73S, I84V, L90M
Saquinavir	0.065	90	1.0	78
Indinavir	0.18	34	0.73	21
Ritonavir	0.055	3.0	0.46	2.8
Nelfinavir	0.28	15	3.5	19
Amprenavir	0.10	0.15	0.21	1.40
Lopinavir	0.005	6.1	0.04	0.90
Atazanavir	0.046	0.33	0.009	0.49
Tipranavir	0.088	0.014	0.001	0.032
Darunavir	0.008	0.005	0.041	0.025

Point mutations V32I, V82A, and V82SA are among the major protease mutations causing resistance to FDA-approved drugs (Figure 10). Specifically, the V32I mutant causes resistance to IDV, RTV, and APV and, when concomitant mutations are present, also to DRV, ATV and LPV. Mutations V82A and V82S cause resistance to IDV, RTV and LPV and, if other mutations are present, also to NFV, APV, SQV, ATV, and DRV [40–42]. Interestingly, the frequency of mutations at positions 32 and 82 in the enzyme correlates with the number of inhibitors administered in the treatment (Table 3). This suggests that these specific mutations may play a crucial role in the development of drug resistance, reflecting the selective pressure exerted by ongoing PI therapy [38].

***Table 3**: Mutations and therapies. Mutation frequency in positions 32 and 82 correlates to the number of PIs administered.*

Position[a]	aa[b]	Mutation frequency at no. of PIs[c]					Association with PIs (*P*)[d]	Association with no. of PIs[e]		Common substitutions[f]
		0	1	2	3	≥4		Coefficient	*P*	
32	V	0.0	4.2	3.9	3.8	6.3	**<1.0e-9**	0.11	3.2e-1	I
82	V	1.2	25.0	33.3	46.2	42.3	**<1.0e-9**	0.30	**1.6e-9**	ATFI

1.6 ALES inhibitors: a novel PI library

The ongoing challenge of HIV-1 drug resistance requires the continuous development of new drug generations. A new library of potential HIV protease inhibitors has been synthesized at the University of Basilicata, under the supervision of Prof. Maria Funicello and Prof. Lucia Chiummiento (Figure 11 and Figure 12). These inhibitors are designed as non-peptide molecules, featuring a common scaffold resembling Darunavir. Key characteristics of these compounds include a sulfonamide group bound to an aromatic ring, an additional aromatic ring, an amide group, and a trifluoromethyl moiety.

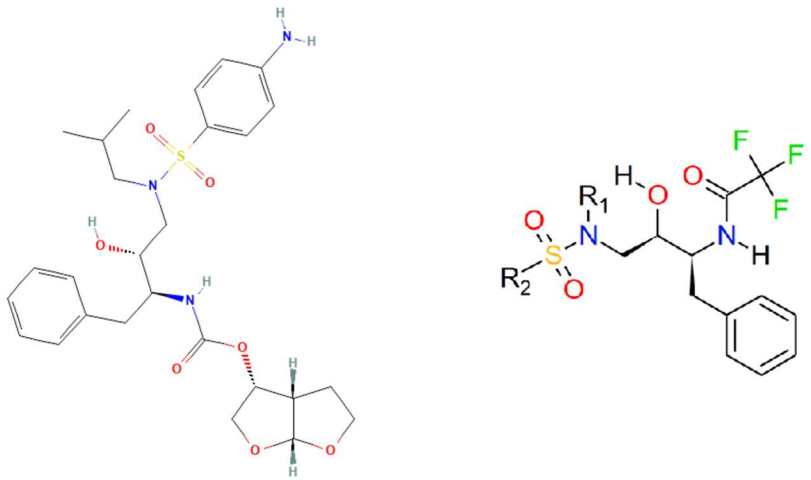


Figure 1: **The ALES scaffold.** Comparison between Darunavir (left panel) and ALES inhibitor scaffold (right panel).

Compounds of the ALES library differ for the substitutions on the sulphonamide nitrogen and the aryl-sulfonic ring. The two chiral centers have R and S configurations, respectively for the carbon attached to the hydroxyl group and for the carbon linked to the benzyl group.

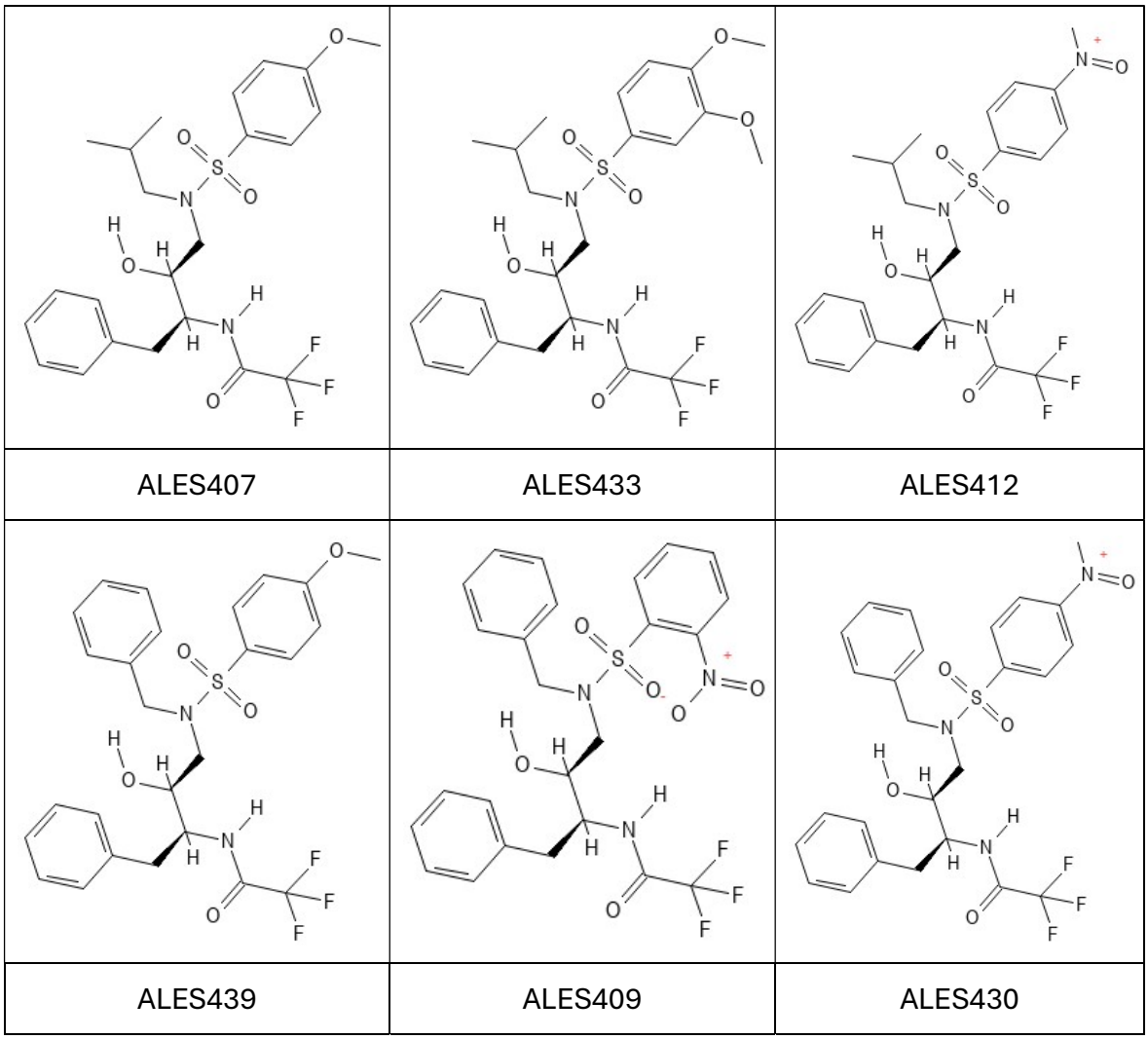


Figure 2: **The ALES library.** Chemical structures of the synthesized ALES compounds.

In an unpublished study, the compound ALES 407 was identified as the lead compound in the library, with an IC50 value of lower than 6 nM when used against the wild type form of the HIV-1 protease. This inhibitor is structurally very similar to the commercial inhibitor Darunavir: they share not only the scaffold but also the isobutyl substitution on the nitrogen atom of the sulfonamide group. ALES407 has a molecular weight of 432.17 g/mol, slightly lower compared to the 547.7 g/mol of Darunavir. A Thermal Shift Assay (TSA) showed an increase in the melting temperature (T_m) of the wild-type protein when ALES407

is added, confirming the binding to the protease. In the same assay, a higher thermal shift was observed for the V82A mutant, suggesting that the efficacy of this inhibitor extends to HIV-1 variants that contain this mutation.

2. AIM

Despite nearly 40 years of research leading to the development of potent drugs against the human immunodeficiency virus (HIV), HIV infections remain a public health issue, particularly in underdeveloped countries. In these regions, a significant problem is the lack of continuity in therapy, which contributes to the emergence of drug-resistant viral strains. Moreover, the use of combination therapies has also resulted in the development of HIV variants with multiple resistances.

Among the antiviral drugs developed since the 1990s, viral protease inhibitors have achieved considerable success, mostly thanks to the availability of crystal structures of enzyme-inhibitor complexes that drove the design of new drugs using a structure-based approach. This protease is crucial for virus replication, as it is responsible for the proteolytic cleavage of viral polyproteins. To date, the Food and Drug Administration has approved 10 peptidomimetic drugs that target this protease. However, the high mutation rate of the viral genome leads to the development of resistance against the known inhibitors, prompting the continuous search for new inhibitors.

This study begins with a library of compounds designed and synthesized by Professors Lucia Chiummiento and Maria Funicello at the University of Basilicata and aims to identify compounds of the library that display inhibitory activity, particularly for enzymes that contain mutations known to confer resistance to currently FDA-approved drugs.

To elucidate the binding mode of these compounds and suggest possible modifications that might improve their inhibitory activity, in this work the wild-type HIV-1 protease and three of its drug-resistant mutants (V32I, V82A, and V82S) are produced in a recombinant expression system, according to previously optimized protocols. Enzyme inhibition assays are conducted to assess the inhibitory potency of the most promising compounds in the library. The structural characterization of the protein-inhibitor complexes allow the identification of the stabilizing interactions that inhibitors form with the active site of the wild-type protein and with the drug-resistant mutants.

3. MATERIALS AND METHODS

This experimental project focuses on the crystallographic analysis of pharmacoresistant mutants of the HIV-1 protease in complex with members of the ALES library of inhibitors developed by the University of Basilicata.

The work includes three experimental steps: (1) expression, purification, and refolding of the proteases; (2) evaluation of the inhibitory activity of ALES compounds on the wild-type protein; (3) structural characterization of protein-inhibitor complexes using X-ray crystallography.

3.1 HIV-1 protease expression

pET-11a plasmids containing the protease gene sequences (wild-type and V32I, V82A, and V82S mutants, Figure 13) under the control of the T7 promoter were available at the Center of Excellence in Biocrystallography. The sequence of the HIV-1 protease used in this study contains stabilizing point mutations to reduce autoproteolytic activity and facilitate protein folding by removing a disulfide bridge. This form (Figure 13), also known as PR5, has a similar enzymatic activity to the native HIV-1 protease: K_M = 164 µM and k_{cat} = 26 min^{-1} for the stabilized form, compared to K_M = 130 µM and k_{cat} = 114 min^{-1} for the native protease[43,44].

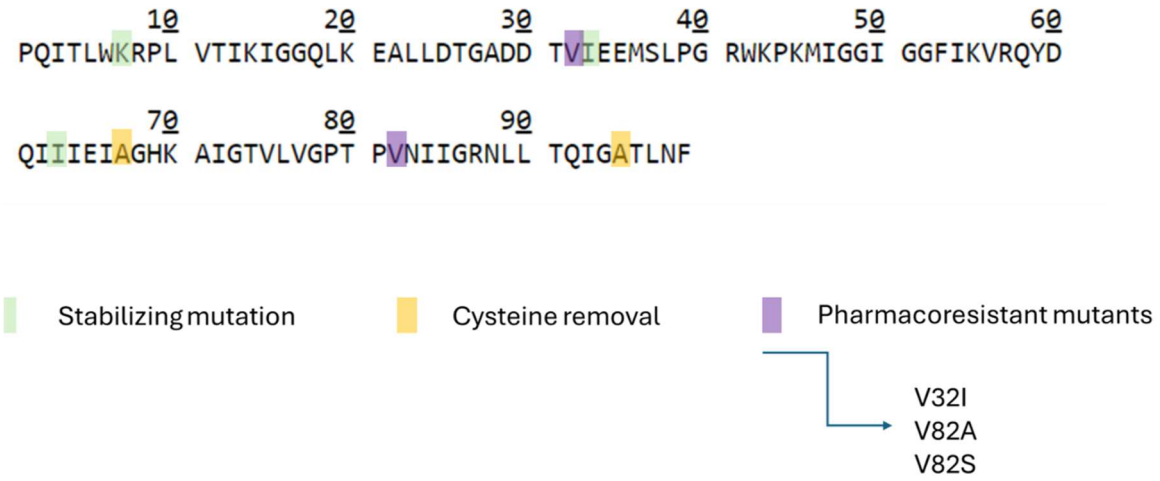


Figure 13: **Stabilized form of the HIV-1 protease.** In the PR5 form, point mutations highlighted in green stabilize the protein against autoproteolytic cleavage, mutations in yellow were introduced to remove a disulphide bridge and facilitate folding of the protein. Mutations involved in pharmacoresistance that have been considered in this thesis are highlighted in violet.

Proteins were expressed in the *E. coli* BL21(DE3) LEMO21 strain (New England Biolabs). This system contains a pLemo plasmid, that codes for the T7 lysozyme enzyme

under the control of a rhamnose-inducible promoter. T7 lysozyme is the natural inhibitor of T7 RNA polymerase, so its action blocks the unwanted basal expression of genes under the control of T7 promoters. Transformed cells were pre-inoculated in LB in the presence of 100 µg/mL ampicillin (AMP) and 0.5 mM L-rhamnose, and the cell culture was incubated overnight at 37 °C under stirring. The following day, a fresh LB medium with ampicillin was inoculated with a 1:100 dilution of the pre-culture and incubated at 37 °C until the optical density at λ=600 nm (OD600) reached approximately 0.5. At this point, IPTG was added to induce recombinant protein expression, and the culture was incubated for additional 3 hours at 37 °C before the cells were harvested by centrifugation at 6000×g for 15 minutes. The cell pellet was washed with TBS, centrifuged again to remove the supernatant and stored at -20 °C until further use.

3.2 HIV-1 purification

HIV-1 protease proteins are expressed by the bacterial host in inclusion bodies (IBs) in their unfolded form. The pellet resulting from 500 mL of bacterial culture was resuspended in 150 mL of Lysis Buffer (Tris 20 mM pH 8, NaCl 250 mM, PMSF 1 mM), yielding a turbid white suspension. Cell lysis was performed using an Emulsiflex-C5 mechanical homogenizer at a pressure of 120 MPa at 4 °C. The lysate was centrifuged at 4450×g for 1 hour at 4 °C. The cellular debris was rinsed twice with the Inclusion Body Wash Buffer (2 M urea, 20 mM Tris pH 8, 250 mM NaCl, 1% v/v Triton X-100), resuspending the IBs in a Potter-Elvehjem homogenizer with approximately 20 mL of this solution at each step. At the end, IBs were resuspended in Urea Buffer, consisting of 8 M urea, 50 mM MES pH 6.5, and 1 mM DTT and incubated overnight at 4 °C while stirring.

After IBs solubilization, the sample was centrifuged at 4450×g for 10 minutes at 4 °C to remove the insoluble debris. This process was repeated three times. Then, the sample was purified by size exclusion chromatography using a Superdex^{TM} 75 10/300 GL column (Cytiva) equilibrated with the Urea Buffer.

3.3 Protein refolding

After protein purification, a refolding step was necessary to restore the native folding and activity of the protease. Refolding was performed in two different ways: using the templating effect of the inhibitor when the goal was to crystallize the protease-inhibitor complex; or by removing the chaotropic agent in the absence of any inhibitor for samples used in enzymatic activity assays.

3.3.1 Inhibitor-free refolding

The unfolded protein solution was dialyzed against buffers containing 10 mM Na_2HPO_4 at pH 6.5, 1 mM EDTA, 100 mM NaCl, 10% v/v glycerol, and 1 mM β-mercaptoethanol, with progressively decreasing urea concentrations (6, 4, 2, and 0 M) to remove the chaotropic agent. Each dialysis step was incubated for 2 hours, except for the final one, which was incubated overnight. After dialysis, the protein solution was centrifuged at 13,000×g at 4 °C for 10 minutes to remove any precipitate. The protein solution was stored in 5-µL aliquots at a temperature of -80 °C, for the activity assays. The concentration of the stabilized HIV-1 protease in the aliquots was 0.3 mg/mL (14 µM).

3.3.2 Refolding with inhibitors

To renature 500 µL of a sample in 8 M urea, 3.9 mL of Refolding Buffer was prepared. This solution contained 50 mM sodium acetate at pH 4.5 and the chosen inhibitor at a final molar concentration 2 to 5 times greater than the estimated final concentration of the enzyme. In this protocol, four aliquots of equal volume were added to the Refolding Buffer with 1.5 hour intervals between them, keeping the system under stirring at a temperature of 18 °C. After the last addition, the sample was incubated overnight while stirring. The following day, the sample was centrifuged at 13000×g for 15 minutes at 4 °C to remove the precipitate. The final protein concentration was assessed through absorption spectroscopy at 280 nm, assuming the theoretical extinction coefficients calculated using the ProtParam tool (https://web.expasy.org/protparam/)[45] from the amino acid sequences. This approach is both efficient and effective, producing higher yields of refolded protein compared to the previous method. The inclusion of a protease-binding compound enhances protein folding and stability[46]. Moreover, this robust method has been successfully employed to evaluate both research compound[47,48] and available drug[49] PIs through crystallographic approaches.

3.4 Wild-type HIV-1 protease activity assays

To estimate the inhibition constant of the ALES inhibitors, enzyme activity assays were performed using the wild-type, stabilized form of HIV-1 protease. Enzyme activity was measured by following the fluorescence of the modified peptide Abz-NF*-6 (Bachem) utilized as the protease substrate. This compound (Figure 14), also known as the Anthranilyl-HIV Protease Substrate, is a hexapeptide with sequence Abz-Thr-Ile-Nle-p-nitro-Phe-Gln-Arg-NH_2, where Abz is a 2-amino-benzoic acid residue and Nle is the unnatural amino acid residue norleucine.

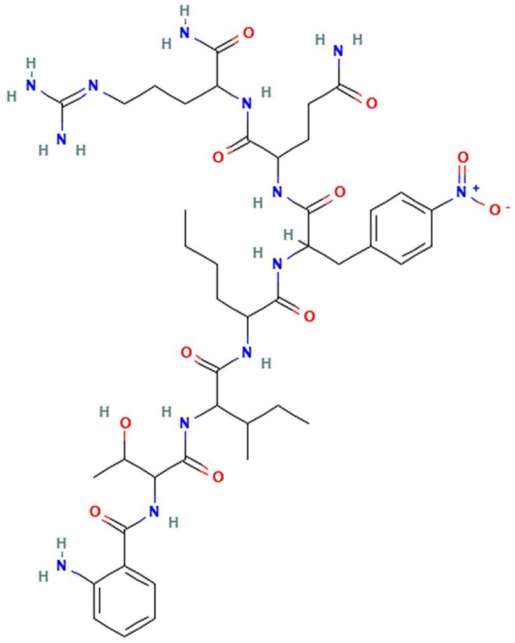


Figure 14: ***Anthranilyl-HIV Protease Substrate****. Chemical structure of the substrate used in HIV-1 protease activity assays, with sequence (Abz-Thr-Ile-Nle-p-nitro-Phe-Gln-Arg-*NH_2*).*

Before enzymatic cleavage, the Anthranilyl-HIV Protease Substrate does not display any fluorescence because the signal from its 2-amino-benzoic end is quenched by the p-nitro phenyl group. Substrate cleavage by the HIV-1 protease results in the release of the fluorescent tripeptide, allowing signal detection. The rate of increase in the fluorescence signal is directly related to enzyme activity.

Fluorescence emission was measured using the Perkin Elmer LS50B luminescence spectrometer, employing the FL WinLab software with the following setup parameters: excitation wavelength at 325 nm; emission wavelength at 420 nm; excitation slit width at 7; emission slit width at 7; run of 30 minutes; data collection interval of 0.2 minutes.

3.4.1 Preparation of Buffer Solutions for Enzyme Activity Assays

To set up enzyme activity assay and inhibition assays, the following solutions were prepared:

- Buffer solution: 0.1 mM MES pH 5.5, 400 mM NaCl, 1 mM EDTA, and 1 mM DTT;
- BSA solution: 1 mg/mL bovine serum albumin (BSA) in Buffer solution;
- Abz-NF*-6 solutions: the substrate was diluted in Buffer solution;
- HIV-1 protease solutions: HIV-1 protease aliquots were resuspended in BSA solution to obtain 0.003 mg/mL (0.14 μM) protease solutions;
- Inhibitor solutions: serial dilutions (from 10^{-2} to 10^{-12} mol/L) of ALES compounds were performed using dimethyl sulfoxide (DMSO).

3.4.2 HIV-1 Protease Activity Assay: Experimental Procedure

Protease activity assays were performed at different substrate concentrations, measuring the initial reaction rates, to derive the Michaelis-Menten constant (K_M). This constant represents the substrate concentration at which the reaction rate is half the maximum rate (v_{max}). Except for the final substrate concentration, every assay was carried out with the same protocol. More precisely, 5, 10, 15, 25, 50, and 75 ng/μL concentrations (corresponding to 5.3, 10.6, 16.0, 26.6, 53.2, and 79.8 μM) of substrate were utilized.

189 μL of the diluted Abz-NF*-6 substrate solution at pH 5.5 was introduced into a quartz cuvette. The fluorescent signal was recorded for several minutes until stable. At this point, 11 μL of a solution of HIV-1 in BSA solution, pH to 5.5, was added to the cuvette, resulting in a final enzyme concentration of 0.165 ng/μL (7.7 nM). The acidic pH of this buffer allows enzyme dimerization and recovering of its functional state. After adding the protease, a continuous fluorescence measurement was performed for about 20 minutes. Fluorescence data for the first five minutes of the reaction were used to determine the v0 value.

3.4.3 HIV-1 Protease inhibition assays: Experimental Procedure

Enzyme activity inhibition assays were performed with ALES compounds to determine the inhibitor concentration (IC_{50}) required to halve the reaction rate compared with the rate of the uninhibited enzyme. The commercial drug Darunavir (DRV) was used as the reference inhibitor because its structure is similar to that of the ALES library compounds.

189 µL of the Abz-NF*-6 substrate solution at pH 5.5 was introduced into a quartz cuvette and the fluorescence was measured until a baseline was defined. Then, 11 uL of HIV-1 protease solution was added to obtain a final concentration of substrate of 50 ng/mL (53.2 µM) and a protease concentration of 0.165 ng/µL (7.7 nM). The inhibitors were tested by adding 2 µL aliquots of the compound in DMSO, progressively increasing the inhibitor concentration into the cuvette, one order of magnitude at each addition. After each addition, a decrease in the rate of substrate hydrolysis was observed. When the slope of the reaction curve reached zero, no further aliquots of the inhibitor were added.

3.5 Protein crystallization

For the crystallization experiments of HIV-1 protease and its mutants in complex with inhibitors of the ALES library, the vapor diffusion crystallization technique was used, with a hanging drop configuration. In these experiments, 1 µL of unsaturated protein solution and 1 µL of the reservoir were dispensed on a coverslip, which was turned upside-down and sealed on a well containing 1 mL of the reservoir, in 24-well plates. The concentration of the protein in the drop ranged from 5 to 7.5 mg/mL. Based on previous results obtained for the crystallization of the wild-type protein in the presence of the ALES407 inhibitor, a set of 8 different crystallization conditions from the the Stura FootPrint Screen #1 were used (Table 4).

Table 4**:** Crystallization of HIV-1 protease-inhibitor complexes**.** Crystallization experiments for the complexes of HIV-1 protease (wild-type and mutants) with inhibitors of the ALES library were set up by the hanging-drop vapour-diffusion method with the following reservoir solutions.

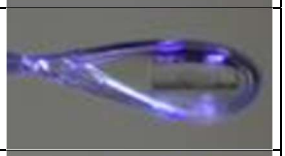
	Column 1	Column 2
A	10% PEG 4K, 0.2 M imidazole malate pH 7.0	7,5% PEG 10K, 0.2 M imidazole malate pH 8.5
B	15%PEG 4K, 0.2 M imidazole malate pH 7.0	12,5%PEG 10K, 0.2 M imidazole malate pH 8.5
C	20%PEG 4K, 0.2 M imidazole malate pH 7.0	17,5% PEG 10K, 0.2 M imidazole malate pH 8.5
D	25% PEG 4K, 0.2 M imidazole malate pH 7.0	22,5% PEG 10K, 0.2 M imidazole malate pH 8.5

3.6 Diffraction data collection and analysis

Single crystals were fished from their crystallization solutions using nylon loops and stored in liquid nitrogen until diffraction analysis. Data collection was performed at the XRD2 beamline of the Elettra synchrotron (Basovizza, Trieste, Italy).

Table 5 shows the parameters used to collect diffraction data from the crystals of the complexes between HIV-1 proteases and the ALES compounds. For each data collection, 720 images were acquired, with a rotation of 0.5° per image. In all data collections, X-ray radiation at a wavelength of 1.000 Å was used.

Table 5: *Diffraction data collections. Parameters used in the data collection of complexes of HIV-1 protease (wild-type, WT, and mutants V82S, V82A, and V32I) and some of the inhibitors of the ALES library.*

	Exposure time (s)	Beam size (µm)	Crystal-detector distance	Crystal
WT+ALES407	360	100x100	153.5	
V82S+ALES407	360	100x100	153.5	
V82S+ALES407	360	100x100	182.5	
V82A+ALES407	1080	50x50	153.5	
V82A+ALES430	360	50x50	153.5	
V82A+ALES430	720	50x50	153.5	
V32I+ALES407	360	75x75	153.5	
V32I+ALES407	720	50x50	153.5	
V32I+ALES407	360	100x100	153.5	

Diffraction data were collected for the complex between the wild-type stabilized protease (WT) and the ALES407 inhibitor, and for the following HIV-1 protease mutant-inhibitor complexes: V82S with ALES407, V32I with ALES407, V82A with ALES407, and V82A with ALES430.

Indexing of the reflections and data integration were performed using the XDS[50] software, while scaling and merging were conducted using the AIMLESS software[51]

(CCP4i2 suite[52]). The phase problem was solved by the MOLREP software [53](CCP4i2 suite) with the Molecular Replacement method. Dimeric HIV-1 protease (PDB: 2A1E[47]) was used as a model to solve the HIV-1 protease V82S-ALES407 structure. In the first step, refinement of the model was carried out using REFMAC5[54] (CCP4i2 suite). Later, given the high resolution of the structures, the SHELXL[55] software was employed for the final refinement cycles and anisotropic thermal parameters for non-hydrogen atoms were refined. The COOT[56] software was used to visualize models and electron density maps, and adjust atom positions.

Since all the protein-inhibitor crystals were isomorphic, the refined structure of HIV-1 protease V82S-ALES407 was used as a starting model to refine the other structures. Residues that differed from the starting model were mutated accordingly. Restrained files for the inhibitors were obtained using the ACEDRG software[57] (CCP4i2 suite). Some side chains of the protein residues showed disorder and were, hence, modeled with alternate conformations; the occupancy factor of each conformation was determined using the difference map and the B-factor analysis. Atoms with occupancy factors less than 1.0 were refined with isotropic thermal parameters.

Water molecules were manually added to the model. Water molecules with a B-factor greater than 80.00 Å, with a R.M.S.D. level less than 1.00 electrons/Å, or with contacts closer than 2.30 Å or further than 3.50 Å to atoms of the protein or the inhibitor were monitored during refinement cycles. For some water molecules, the occupancy factor was adjusted based on the analysis of thermal factors (high B-factor values and anisotropy) and electron density.

Throughout the refinement phase, the R and R_{free} parameters were monitored.

4. RESULTS AND DISCUSSION

4.1 HIV-1 protease purification

Purification by Size Exclusion Chromatography (SEC) was performed on the protein samples in presence of the chaotropic agent (urea). Chromatograms of the SEC purification step of all HIV-1 protease mutants (Figure 15, upper panels) are consistent, showing an isolated peak at about 10.5 mL. This value is lower than the elution volume expected considering the molecular weight of the protease (11 kDa). However, as the elution from a SEC column depends on the hydrodynamic radius of the isolated species, a lower retention volume can be anticipated for unfolded proteins in urea.

SDS-PAGE analysis confirms the presence of a protein with dimensions consistent with the protease monomer (Figure 15, lower panel). Since no contaminants are present, no further purification steps are required.

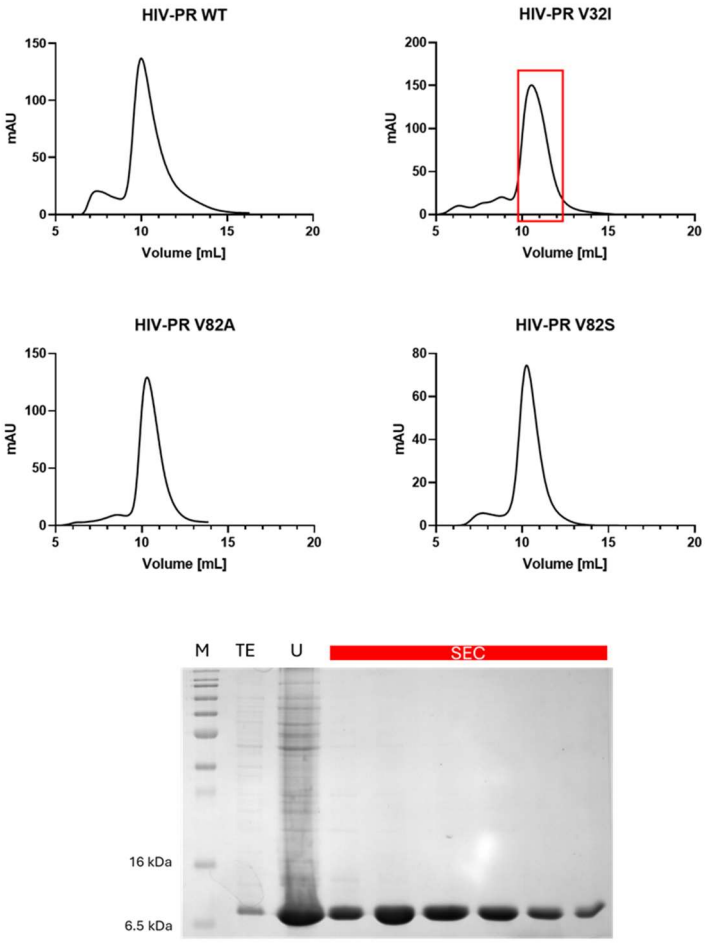


Figure 15: **Purification of HIV-1 protease mutants**. Upper panels: SEC chromatograms; the fraction corresponding to the peak observed in the chromatogram of the V32I mutant purification are highlighted in red. Lower panel: SDS_PAGE analysis of the purification of the V32I mutant (M: protein marker; TE: cellular total extract; U: IBs in urea; red box: SEC peak fractions).

4.2 HIV-1 WT activity assay

To assess the activity of the wild-type HIV-1 protease without inhibitors, enzyme activity assays were conducted using the fluorescent peptide Abz-NF*-6. These assays measured enzyme kinetics across varying substrate concentrations, allowing calculation of the Michaelis-Menten constant (K_M) and maximum rate (v_{max}). Initial reaction rates (v_0) were determined for each substrate concentration based on the plot of the fluorescence intensity over time. Before the enzymatic reaction, no fluorescence is observed due to quenching interactions between N- and C-terminal chromophores of the substrate. Following enzymatic cleavage, fluorescence increases as quenching interactions are lost. The fluorescence signal is, therefore, directly proportional to the concentration of the reaction product.

The calculated reaction rates were reported on a Michaelis-Menten plot, showing v_0 as a function of substrate concentration (Figure 16). At high substrate concentrations, where the enzyme active site is expected to be saturated, the initial reaction rate v_0 measured by the assay approaches the maximum reaction rate (v_{max}). The Michaelis-Menten constant (K_M) is determined as the substrate concentration at which the reaction rate is half of v_{max}. Data analysis performed with GraphPad Prism 8.0 determined a v_{max} of 70 min^{-1} and a K_M of 65 μM, in line with the reported values for the stabilized form of the HIV-1 protease (v_{max} = 26 min^{-1} and K_M = 164 μM) [58].

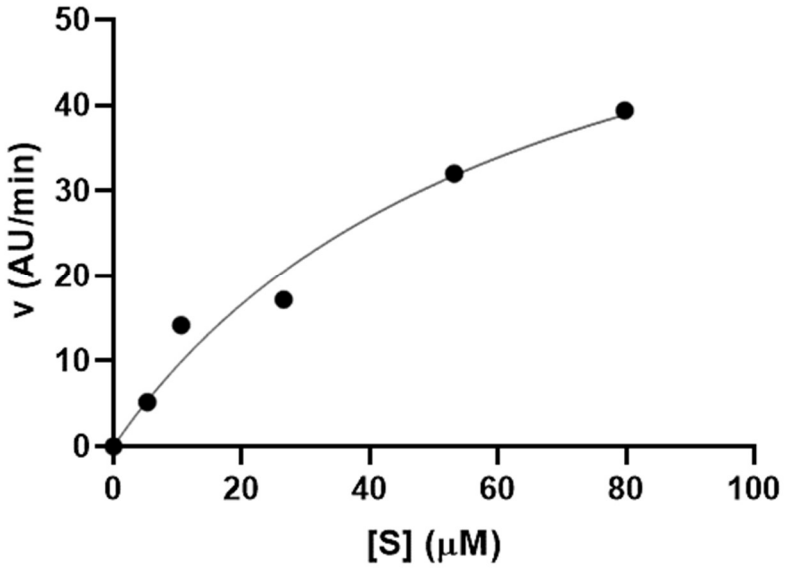


Figure 16: **Wild-type HIV-1 protease activity assay.** Micahelis Menten plot of the activity assays performed on the wild-type protein with different substrate concentrations.

4.3 Assessment of the IC_{50} of ALES inhibitors

Enzyme activity inhibition assays were conducted with four compounds from the ALES library and with Darunavir (DRV) to determine the inhibitor concentration required to halve the reaction rate (IC_{50}), compared with the rate of the uninhibited HIV-1 protease enzyme. Tests were conducted by adding increasing concentrations of the inhibitor to the reaction mixture. As the inhibitor concentration increased, a decrease in the slope of the fluorescence emission curve was observed. For the inhibition assays, the residual activity, defined as the ratio between the rate of the inhibited reaction and the reaction rate in the absence of inhibitor, was reported as a function of the logarithm of the inhibitor concentration (Figure 17). A sigmoid shape is expected for the resulting curves, with the inhibitor concentration at the inflection point representing IC50. Table 6 shows the IC50 values obtained for each inhibitor. Reaction rates were hard to determine for low concentrations of the inhibitors, due to the considerable drop in enzyme activity upon additions of nanomolar concentrations of the inhibitors. A preliminary test (data not shown) indicated the strong potency of ALES407. Thus, this compound, along with the PI Darunavir, was evaluated at lower concentrations. Data were analyzed using the GraphPad Prism 8.0 software, which allowed to obtain IC50 values by fitting sigmoidal curves on the experimental data through the following equation:

$$Y = Bottom + \frac{Top - Bottom}{1 + 10^{(LogIC50 - X) * HillSlope}}$$

Where the Top and Bottom values were set to 1 and 0, respectively. For each experimental curve, the GraphPad software calculates also the Hill slope, a useful parameter to evaluate allosteric contributions to the inhibition. For a competitive non allosteric inhibitor, the expected Hill slope is close to -1.0, as in the case of the competition curve for the inhibitor ALES430 (Table 6). For other inhibitors, Hill slopes are greater than -1.0. In particular, the competition curve for the ALES409 inhibitor deviates significantly from the value of -1.0, which could indicate a poor availability of the inhibitor in solution (e.g., due to solubility problems)[59]. For a deeper investigation of the Hill slopes more inhibitor concentrations should be tested. These analyses confirmed the inhibition efficacy of these compounds against the wild-type HIV-1 protease, confirming the ALES407 inhibitor as the most potent. The quantification of the IC_{50} value is crucial for characterizing a new inhibitor. However, experimental IC_{50} values are only meaningful if they exceed half the protein concentration used in the assay. Given that the lower protein concentration used in this assay was 7.7 nM, the IC50 for the potent ALES407 inhibitor can be confidently estimated as

under 4 nM, a result that also applies to the Darunavir inhibitor used as control. These preliminary tests provide an estimated potency range for the ALES407 inhibitor. More sensitive testing could offer a precise evaluation of its inhibitory effectiveness, exploring values near the maximum residual activity, while additional experimental replicates would help refine this range further. This initial analysis indicates that ALES407 has a comparable potency to Darunavir (< 4 nM), with ALES430 being the next most potent compound.

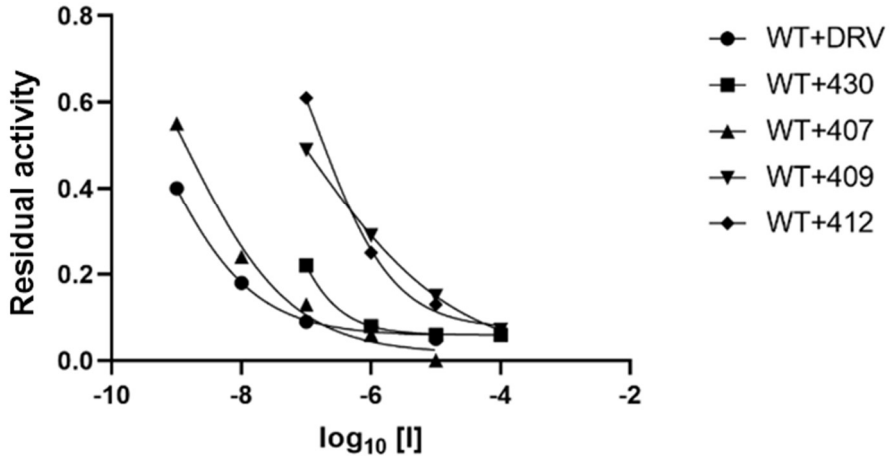


Figure 17: **Wild-type HIV-1 protease inhibition assay with ALES compounds.** The plot shows the residual activity (ratio between the rate of the inhibited reaction and the reaction rate without inhibitor) against the logarithm of the inhibitor concentration. Experiments were performed using inhibitors 407, 409, 412, and 430 from the ALES library, and with the commercial inhibitor Darunavir as control.

***Table 6**: IC50 values for the ALES inhibitors. Values were obtained by fitting sigmoid curves on the experimental curves of the residual activity of the wild-type HIV-1 protease.*

	WT+DRV	WT+ALES430	WT+ALES407	WT+ALES409	WT+ALES412
Log IC_{50}	-9,421	-7,715	-8,891	-7,052	-6,819
Hill Slope	-0,5782	-0,9575	-0,5295	-0,3709	-0,7432
IC_{50} (M)	3,791E-10	1,927E-08	1,287E-09	8,872E-08	1,52E-07

4.4 Structure analysis of HIV PR mutants in complex with ALES compounds

Crystals from the crystallization experiments of the protein-inhibitor complex were analyzed using synchrotron radiation at the XRD-2 beamline of the Elettra Synchrotron facility in Trieste. The resulting data were used to determine the unit cell dimensions and assess crystal quality. Data collection statistics are provided in Table 7.

***Table 7**: Data collection statistics. Statistics of the crystallographic data collections performed for crystals of the protease-inhibitor complexes .*

	WT + ALES407	**V82S + ALES407**	**V82A + ALES407**	**V82A + ALES430**	**V32I + ALES407**
Wavelength (Å)	1,000	1,000	1,000	1,000	1,000
Space group	P 2_1 2_1 2	P 2_1 2_1 2	P 2_1 2_1 2	P 2_1 2_1 2	P 2_1 2_1 2
a (Å)	58,28(5)	57,65(8)	58,31(2)	58,40(2)	58,37(6)
b (Å)	86,46(5)	86,01(11)	85,99(2)	85,86(2)	86,16(4)
c (Å)	46,47(2)	46,17(5)	46,41(1)	46,61(1)	46,31(2)
Matthews coefficient ($Å^3$ Da^{-1})	2,66	2,60	2,28	2,66	2,65
Fraction of the unit cell occupied by solvent	0,54	0,53	0,46	0,54	0,54
Polypeptide chains in the asymmetric unit	2	2	2	2	2
Resolution range (Å)	10,00-1,07 (1,13-1,07)	10,00-1,20 (1,27-1,20)	10,00-1,20 (1,27-1,20)	10,00-1,33 (1,41-1,33)	10,00-1,10 (1,17-1,10)
Average B-factor (Å)	18	20	19	23	17
Total No. of reflection measured	1190144 (17190)	877256 (14643)	873880 (15378)	676140 (11921)	1127354 (17289)
No. unique reflections	94558 (9375)	72321 (10664)	73429 (11270)	54320 (6578)	92614 (10880)
Completeness (%)	98% (56.2%)	100% (47.4%)	100% (49.4%)	100% (75.5%)	97% (71.4%)
Redundancy	12,6 (1,8)	12,1 (1,4)	11,9 (1,4)	12,4 (1,8)	12,2 (1,6)
<I/σ(I)>	18,40 (3,55)	15,90 (1,60)	17,36 (1,86)	19,18 (1,94)	20,58 (1,48)
R_{merge} (%)	8,7 (403,0)	5,3 (57,9)	5,4 (67,3)	8,6 (76,1)	5,8 (107,5)
CC(1/2) (%)	99,9 (78,0)	99,9 (70,8)	100 (72,9)	99,9 (72,3)	100 (72)

The structure of the complex between the V82S mutant and the ALES407 inhibitor was determined by Molecular Replacement, using the polypeptide chain template from the PDB structure 2A1E[47]. Since all protein-inhibitor crystals obtained in this study are isomorphic, the structure of the V82S-ALES407 complex served as a reference for modeling all other complexes.

The quality of the protein-inhibitor complex models was assessed based on the values reported in Table 8. For validation, 5% reflections were used as reference. Refined structures have not yet been deposited in the Protein Data Bank.

***Table 8**: Refinement statistics. Statistics of the refinement of the crystallographic structures of HIV-1 protease-inhibitor complexes.*

	WT + ALES407	**V82S + ALES407**	**V82A + ALES407**	**V82A + ALES430**	**V32I + ALES407**
Reflections with I>2σ(I) used in refinement	122990	90365	96391	65355	104975
Total reflections used in refinement	186868	131448	133504	98424	169249
R factor (I>2σ(I))	0,1203	0,1137	0,1211	0,1201	0,1196
R factor	0,1588	0,1353	0,1413	0,1495	0,1522
R_{free} (I>2σ(I))	0,1536	0,1480	0,1557	0,1755	0,1523
R_{free}	0,1916	0,1686	0,1781	0,206	0,1853
No. of constraints	16241	16369	16493	16389	16447
No. of restraints	30363	30615	30352	30359	19711

4.4.1 Analysis of the secondary structure

The HIV-1 protease structure remains consistent across all available protease-inhibitor complexes, even for proteins with point mutations in their sequence[22]. The secondary structure of the HIV-1 protease-inhibitor complexes was analysed with the software Procheck[60] (CCP4i2 suite) (Figure 18). The secondary structure elements present in all the HIV-1 protease complexes determined in this work align well with the previously published structures. All structures are mainly formed by β-sheet segments, with a single helix formed by residues 87-90. Several regions lack a defined secondary structure, though they are well-resolved in electron density maps. Figure 18 shows also the solvent accessibility of each residue. The mutation at position 32 occurs within a less accessible β-

sheet region, while the mutation at position 82 is in a more solvent-exposed segment, with an undefined secondary structure.

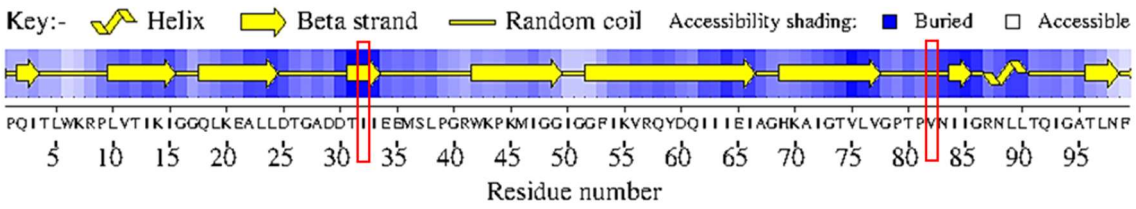


Figure 18: **Secondary structure and solvent accessibility estimation.** Analysis for the mutated HIV-1 protease protein V32I in complex with the inhibitor ALES407 is shown. Residues at positions 32 and 82, mutation sites of the structures considered in this study, are highlighted in red.

4.4.2 Considerations on the quaternary structures

The final electron density maps generated after refinement was highly detailed, enabling a clear structural interpretation. The main chain exhibited continuous electron density with no breaks, except in the segment corresponding to residues 66-70 of one polypeptide chain, which will be further examined in detail. Nearly all amino acid side chains were visible, except for a few terminal atoms of lysine and arginine residues on the surface of the protein.

The quaternary structure of HIV-1 protease (Figure 19) consists in a dimer of two identical subunits. Typical regions of HIV-1 protease were identified: the active site pocket with the catalytic triad and the flaps. The conformation of this flexible regions is closed, as expected considering the presence of the inhibitor in the active site.

4.4.3 Ramachandran Plot

The Ramachandran plot was used to validate the structures, assessing the torsional angles φ and ψ of the peptide bonds. The Ramachandran plots for all the structures of the protein-inhibitor complexes were similar. As an example, Figure 20 displays the plot for the V82S mutant in complex with the ALES407 inhibitor. No residues were found in non-allowed regions: the majority (98.45%) were in energetically favored regions, with a few in allowed regions. The latter group includes glycine residues and residues 66-70 of the poorly defined loop of one polypeptide chain.

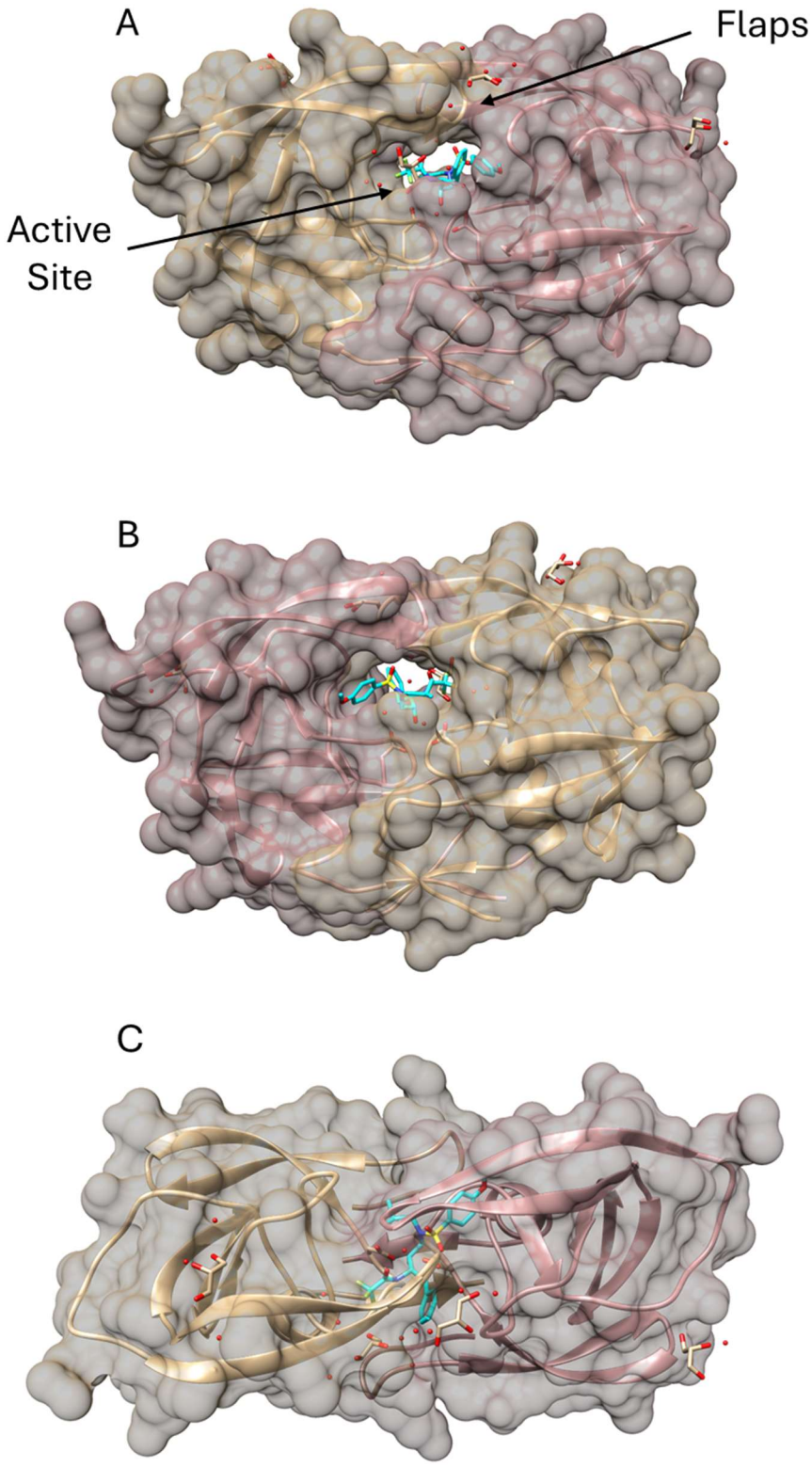


Figure 19: **Crystal structure of the wild-type HIV-1 protease complex with the ALES407 inhibitor.** Representation of the crystallographic structure of the wild-type HIV-1 protease complex with the ALES407 inhibitor. The two monomers are shown in different colors (yellow and pink), while the inhibitor is in cyan. The van der Waals surface is shown in grey. The structure in B was obtained after 180° rotation along the vertical axis of the structure in A, while the representation in C was obtained after 90° rotation along the horizontal axis of the complex in A.

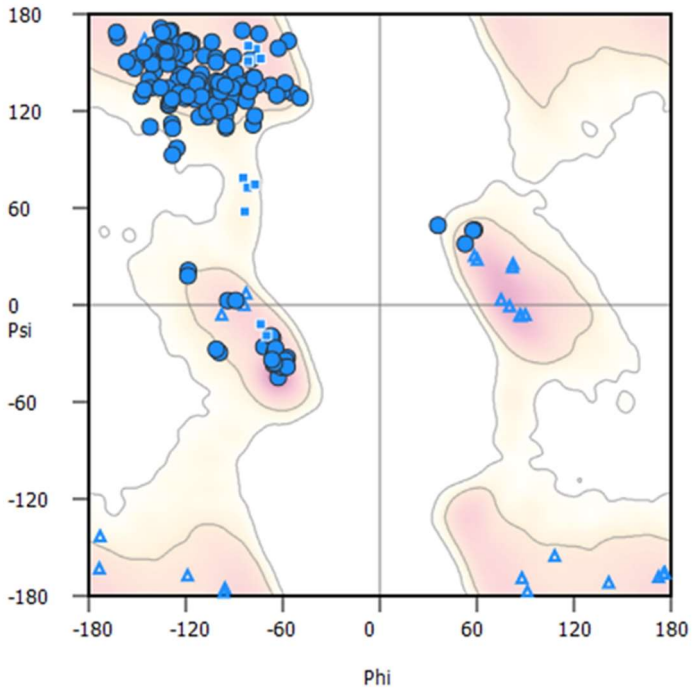


Figure 20: **Ramachandran Plot of the V82S-ALES407 complex.** Regions colored in pink represent favourable conformations of the peptide main chain torsion angles; regions in yellow represent allowed conformation of the main chain. The plot was obtained with the Coot [56]software.

4.4.4 Analysis of crystallographic Packing Contacts

The disorder in the loop of one of the polypeptide chains comprising residues 66-70 affects both the side chains and the main chain of the protein, as indicated by the electron density maps of all the crystal structures of protease-inhibitor complexes, which show low electron density values alongside high thermal factors in this region.

To understand the reasons for this structural disorder, the crystal packing was analyzed. Notably, there is a significant difference in the intermolecular interactions involved in crystal packing in the two polypeptide chains. Figure 21 highlights with spheres of different colors the positions of the alpha carbon atoms of the glycine residue 68 in the two crystallographically independent polypeptide chains. The disorder observed in the electron density for only one of the two chains is probably connected to its position on the surface of the protein and the lack of crystal contact interactions (blue spheres), while in the other polypeptide chain the same loop interacts with symmetrical molecules (magenta spheres). In the first case, the lack of interactions can lead to an increased mobility of the loop that, in turn, makes the localization of the atoms in the electron density harder.

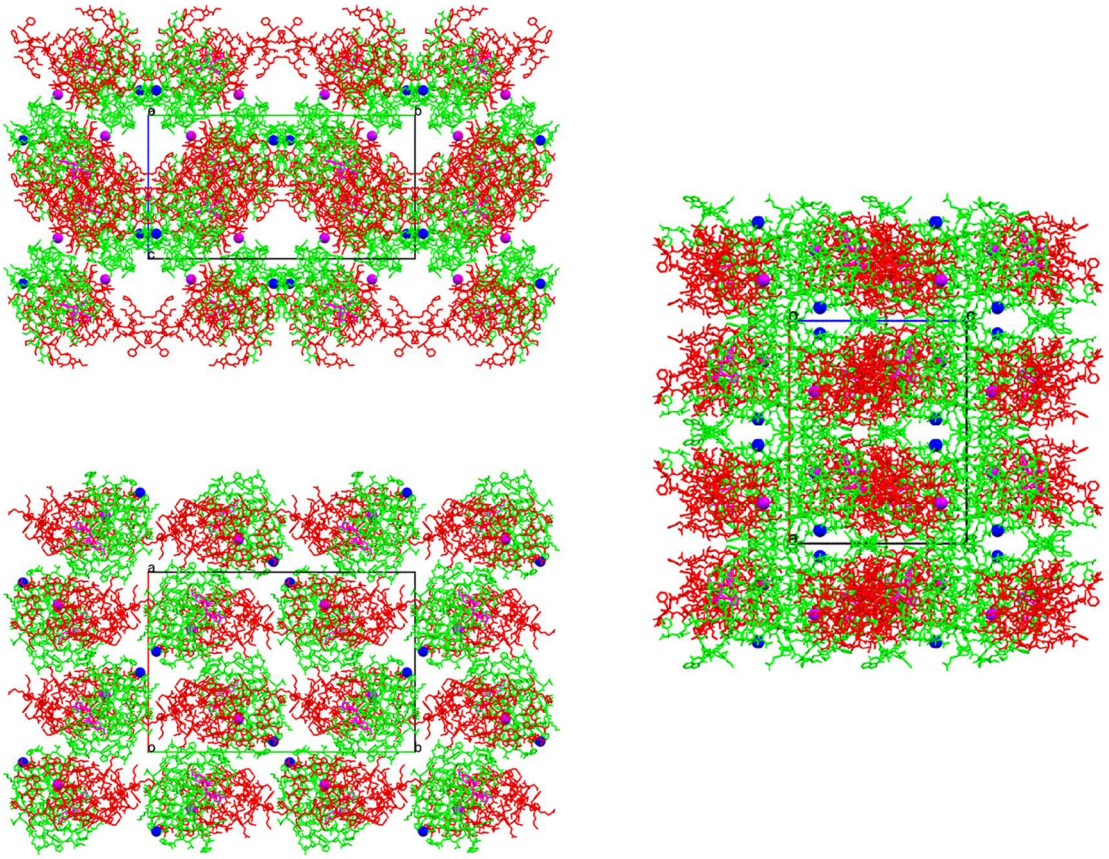


Figure 21: **Crystal packing in the V82S-ALES407 structure.** One of the two polypeptide chains is colored green with a blue sphere corresponding to the α-carbon of residue 68; the other chain is shown in red, with the α-carbon of residue 68 as a magenta sphere. Three different orientations of the crystal lattice are shown: (top left) along the *a* crystallographic axis, (right) along the *b* crystallographic axis, and (bottom left) along the *c* crystallographic axis.

4.4.5 Effect of point mutations

Complexes involving mutants of HIV-1 protease at two specific positions in the sequence, residues 32 and 82, were analyzed. These mutations were observed in variants of the virus showing drug resistance *in vivo*. Although not directly involved in enzyme catalysis, these residues are part of the hydrophobic pocket that houses the substrate or inhibitor.

Interestingly, the structure of the wild-type HIV-1 protease-ALES407 complex shows that the side chain of residue at position 32, a valine, is statistically disordered in two conformations in one of the polypeptide chains of the enzyme. Moreover, in the same structure the residue at position 82 appears even more flexible, as its side chain is disordered in both polypeptide chains (Figure 22). The high-resolution electron density map allowed to determine both partial occupancy conformations. In the case of Val32 and Val82, the electron density corresponding to three methyl groups was identified, with one of them showing a

higher value. This suggests that the residues assume two conformations, 120° apart from each other. The structural disorder observed for these two side chains leads to the hypothesis that, despite their proximity to the active site, these residues are not directly involved in contacts with the inhibitor.

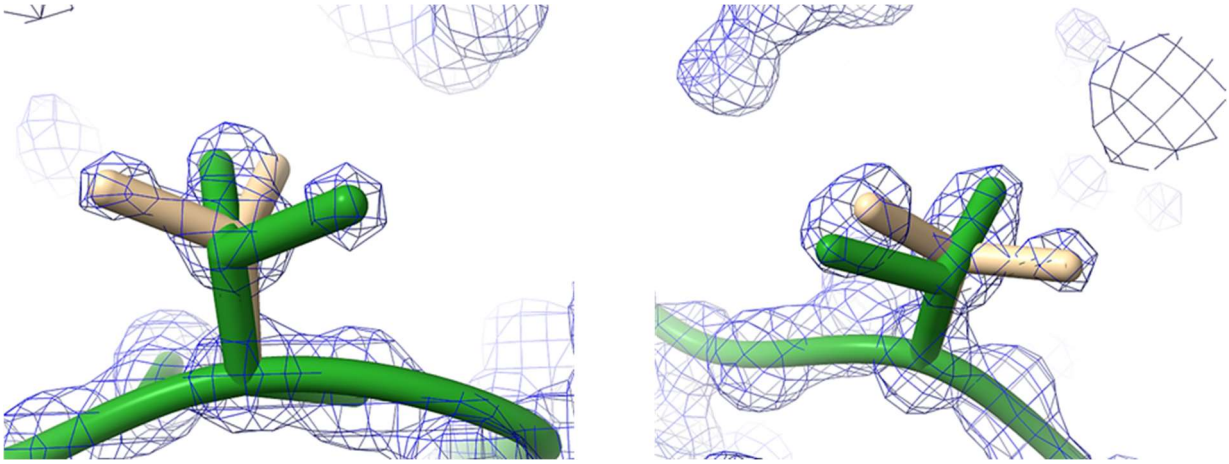


Figure 22: **Residues 32 and 82 in wild-type HIV-1 protease-ALES407 complex.** The side chain of valine 32 (left) shows a statistical disordered in two positions in one polypeptide chains. The side chain of valine 82 (right) has a similar disorder in both the chains.

4.4.6 Comparison between ALES407 and ALES430

In all the crystal structures obtained in this study, an electron density corresponding to the expected structure of the inhibitor was detected in the active site of the protease (Figure 23 The ALES407 inhibitor features two aromatic rings, with well-defined electron density at the para position of one ring, corresponding to the methoxy group. In contrast, the ALES430 inhibitor is characterized by a third aromatic ring bound to the sulfamidic nitrogen atom, replacing the isobutyl group found in ALES407.

Considering the high resolution of the available data that allowed anisotropic refinement of the thermal displacement parameters of full-occupancy atoms in the structures of all complexes, the thermal parameters of the inhibitors were analysed to evaluate their flexibility within the active site of the protein. Lower thermal parameters for some atoms of the inhibitor may suggest their involvement in interactions with the protease residues.

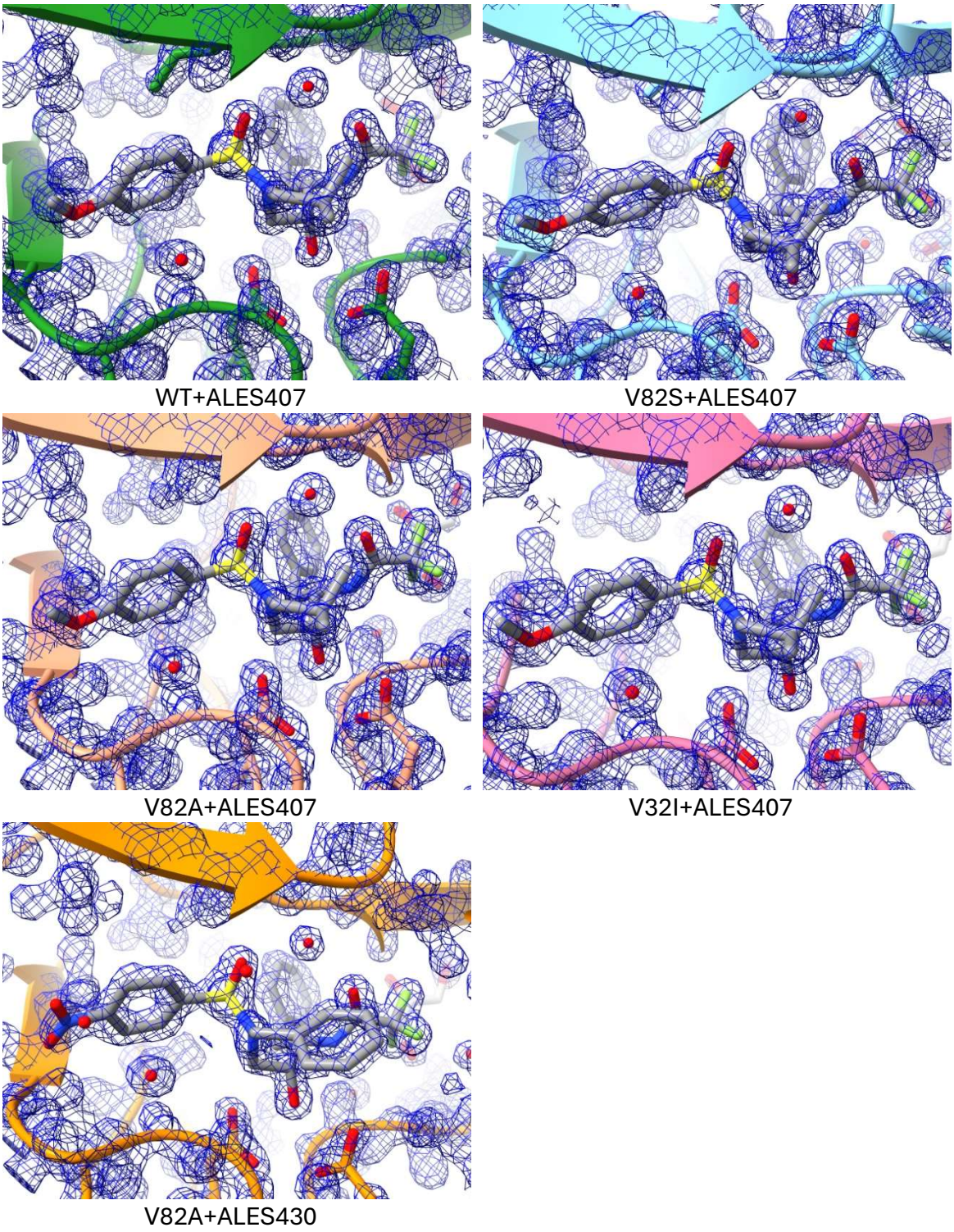


Figure 23: **Structures of the inhibitors in the protease active site.** Representation of the active site of the protein-inhibitor complexes. Green: wild-type protease-ALES407 complex; cyan: V82S protease-ALES407 complex; light salmon: V82A protease-ALES407 complex; pink: V32I protease-ALES407 complex; orange: V82A protease-ALES430 complex. The electron density is shown as a blue contour level mesh.

In all structures containing the ALES407 inhibitor, a similar pattern was observed for the thermal parameters of the atoms of the inhibitor. Figure 24 shows a graphical representation of the anisotropic thermal parameters of ALES407 as rotation ellipsoids in

the four crystallographic structures containing this inhibitor. In all the complexes of ALES407, the trifluoromethyl group of the inhibitor exhibits greater thermal displacement compared to the other atoms of the scaffold. Interestingly, the oxygen atom of the hydroxyl group present in the scaffold, whose role in stabilizing the inhibitor in the catalytic site is well-documented in the literature[34], has consistently slightly higher parameters compared to the neighbouring atoms. A higher thermal displacement was observed also for the atoms of the substituent groups that characterize the ALES407 inhibitor in the library: the methoxy and the isobutyl groups.

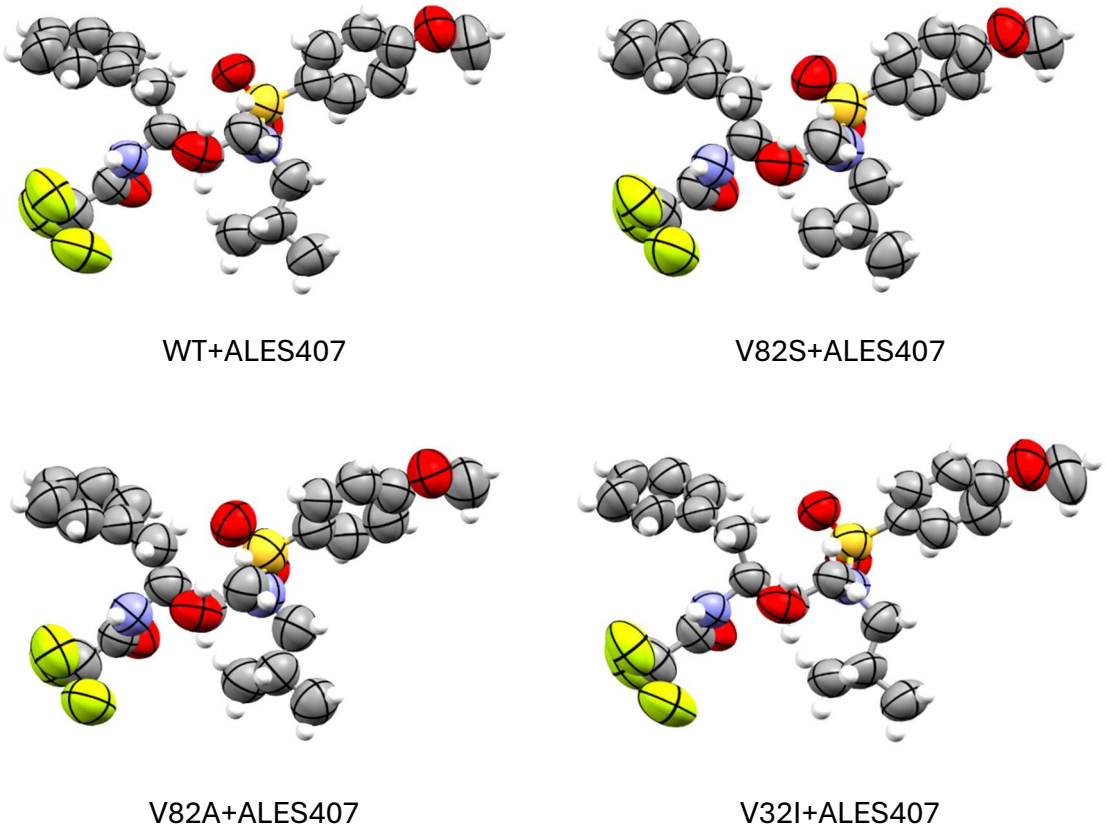


Figure 24: **Anisotropic thermal displacement parameters of ALES407.** Representation of the thermal displacement parameters of the atoms of the ALES407 inhibitor in complexes with different HIV-1 protease mutants. The anisotropic parameters are represented as rotation ellipsoids. Oxygen atoms are red, carbon atoms grey, nitrogen atoms blue, sulphur atoms yellow, fluorine atoms green-yellow. The thermal parameters of the hydrogen atoms (white) were not refined.

The anisotropic thermal parameters of the atoms of the ALES430 inhibitor in the V82A mutant HIV-1 protease complex are significantly higher than those of the ALES407 inhibitor (Figure 25). As with the ALES407 inhibitor, higher values were obtained from refinement for the trifluoromethyl group and the hydroxyl oxygen atom compared to those

of neighbouring atoms, but comparable to those of the ALES407 inhibitor. The substituents that characterize the ALES430 inhibitor (benzyl group and nitro group), on the other hand, exhibit significantly greater thermal parameters than the corresponding substituents in ALES407 (isobutyl group and methoxyl group). In particular, in the case of the nitro group, a significant difference is observed between the two oxygen atoms: the oxygen atom of this group that is located in the same position of the methyl group of the methoxyl substituent of ALES407 has lower thermal parameters compared to the other oxygen atom of the nitro group, suggesting the presence of stabilizing interactions between the protein and the atom located in a position corresponding to the methoxyl group of ALES407.

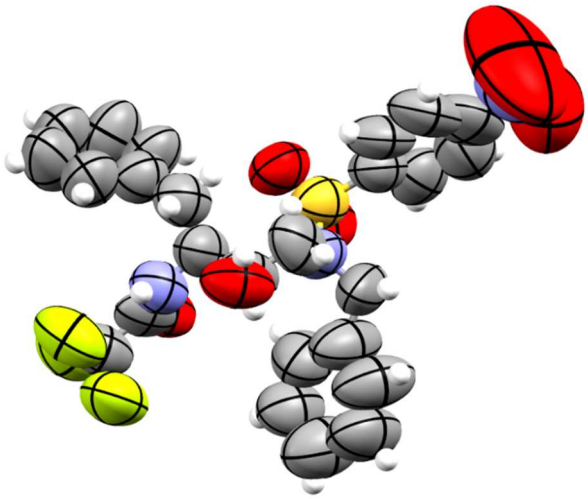


Figure 25: **Anisotropic thermal displacement parameters of ALES430.** Representation of the thermal displacement parameters of the atoms of the ALES430 inhibitor in complexes with the V82A protease mutant. The anisotropic parameters are represented as rotation ellipsoids. Oxygen atoms are red, carbon atoms grey, nitrogen atoms blue, sulphur atoms yellow, fluorine atoms green-yellow. The thermal parameters of the hydrogen atoms (white) were not refined.

4.4.7 Comparison between the structures of protease-inhibitor complexes

The crystallographic structures of the protein-inhibitor complexes were compared by superimposing the refined models (Figure 26.A and Figure 27). The analysis showed a clear similarity in both overall protein conformation and inhibitor binding in the active site. The inhibitors have the same conformation within the active site as their common scaffold is almost perfectly overlapped. Interestingly, some solvent and cryoprotectant molecules (water and glycerol, respectively) are conserved in the same positions in all the protein-inhibitor structures.

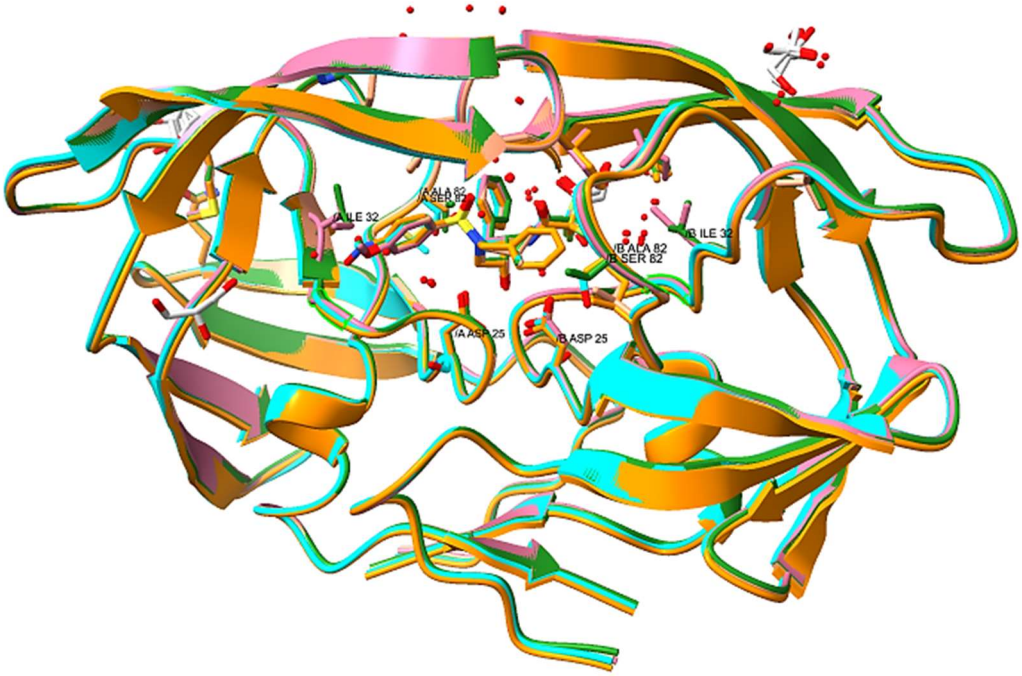


Figure 26 **Comparison between the overall structures of the HIV-1 protease-inhibitor complexes.** Crystal structures of the wild-type protease-ALES407 complex (green), mutant V82S-ALES407 complex (cyan), mutant V82A-ALES407 complex (light salmon), mutant V32I-ALES407 complex (pink), and mutant V82A-ALES430 complex (orange).

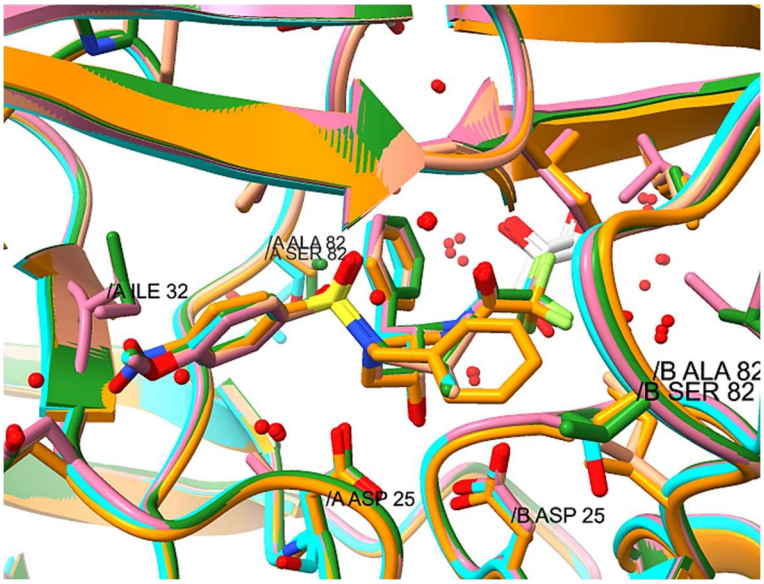


Figure 27: **Comparison between the active site of the HIV-1 protease-inhibitor complexes.** Crystal structures of the wild-type protease-ALES407 complex (green), mutant V82S-ALES407 complex (cyan), mutant V82A-ALES407 complex (light salmon), mutant V32I-ALES407 complex (pink), and mutant V82A-ALES430 complex (orange).

4.4.8 Interactions between the HIV-1 protease mutants and inhibitor ALES407

For each structure, the interactions of the inhibitor with the active site residues of the protein were analyzed. This analysis is crucial for a structure-based drug design approach. The pattern of hydrogen bond interactions is similar for all the mutants of the protease, involving residues that are known for stabilizing the natural substrate of protease (Table 9 and Figure 28). In particular, the inhibitor shows hydrophilic interactions with residues that characterize the surface of S2 to S2' subsites: Gly27, Asp25, Ile50 and Asp30 of the two polypeptide chains.

Most of the hydrogen bonds involving the ALES407 inhibitor in the active site are strong, with some exceptions. The hydrogen bond between the nitrogen atom of the main chain of the residue Asp30 and the oxygen atom of the methoxy group in the inhibitor is weak, as indicated by a bond length (donor-acceptor) of 3.2-3.3 Å in all the structures. Similarly, the hydrogen bond between the oxygen atom of the main chain of residue Gly27' and the amino nitrogen atom of the inhibitor is also weak, with a bond length of about 3.3 Å in all the structures.

On the contrary, a hydrogen bond interaction was observed between a fluorine atom of the trifluoromethyl group of the inhibitor and an oxygen atom from a glycerol molecule (used as a cryoprotectant during data collection) appears to be of moderate strength. This finding supports the relative mobility of the fluorinated group of the inhibitor, as suggested

by the thermal motion analysis, and suggest that the inhibitor could be modified by elongating the chain further than the trifluoromethyl group.

To further analyze the differences in the interactions of the ALES407 inhibitor with the active site of the protein mutants, interaction maps ware generated using the LigPlot[61] software, comparing the wild-type protein structure with those of the mutants. The software highlights not only the hydrogen bonding interactions but also those involving hydrophobic residues.

***Table 9**: Hydrogen bond interactions in the complexes of ALES407. Donor-acceptor distances of the hydrogen bond interactions involving the ALES407 inhibitor in the crystal structures of the complexes are compared.*

Donor atom	**Acceptor atom**	**Hydrogen-bond lengths (Å)**			
		WT-ALES407	**V82S-ALES407**	**V82A-ALES407**	**V32I-ALES407**
ALES407, hydroxyl oxygen atom	ASP25, OD2	2,58	2,60	2,60	2,59
ASP25', OD1	ALES407, hydroxyl oxygen atom	2,68	2,57	2,62	2,61
ASP30, N	ALES407, oxygen atom of the methoxy group	3,33	3,30	3,31	3,21
ALES407, amine nitrogen atom	GLY27', carbonyl oxigen atom	3,35	3,30	3,31	3,26
Oxigen atom, H_2O 10	ALES407, oxygen atom of the carbonyl group	2,74	2,76	2,72	2,74
Oxygen atom, H_2O 10	ALES407, oxygen atom of the sulfonic group	2,67	2,73	2,73	2,72
ILE50, N	Oxigen atom, H_2O 10	2,96	2,89	2,91	2,92
ILE50' N	Oxigen atom, H_2O 10	2,96	2,91	2,91	2,92
Terminal oxygen atom of a glycerol molecule	ALES407, fluorine atom	2,88	2,83	2,89	2,76

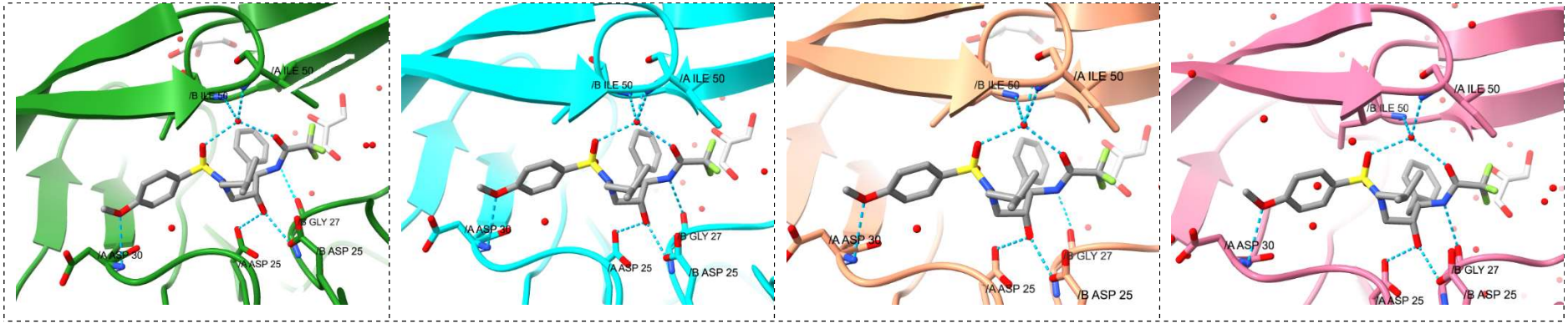


Figure 28: **Hydrogen bond interactions involving ALES407.** The panels show the hydrogen bond interactions between the inhibitor ALES407, the HIV-1 protease residues, and solvent and cryoprotectant molecules in the active site of the proteases. The wild-type protease is in green, the V82S mutant in cyan, the V82A mutant in light salmon, and the V32I mutant in pink.

The comparison between interactions formed by the wild-type protein and by the V82S mutant with the ALES407 inhibitor reveals few significant differences (Figure 29). Notably, however, there is a change in the interaction network involving the phenyl group of the inhibitor. In the wild-type complex, this group forms hydrophobic interactions with residues Pro81 and Val82. In the V82S mutant complex, interactions with both Pro81 and the residue at position 82 (a serine) are lost. On the other hand, a new interaction with Leu23 is observed in the latter but was not present for the wild-type protein. The crystallographic data do not allow a definitive conclusion on whether the interaction network between the protein and the inhibitor is weakened by the mutation at position 82. However, based on the residues involved, it is likely that the hydrophobic interactions of the inhibitor phenyl group are reduced in the V82S mutant protein compared to the wild-type protein.

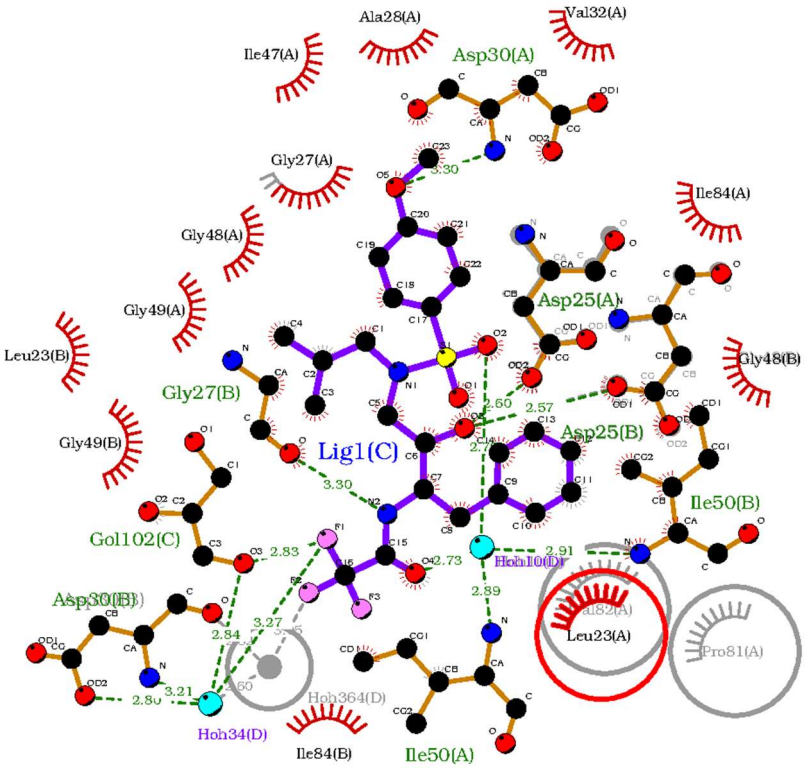


Figure 29: **LigPlot diagram comparing the wild-type protease-ALES407 complex and the V82S mutant-ALES407 complex.** In the foreground, the interactions with the V82S mutant protein are depicted with colors; in the background, the interactions of the wild-type protein are shown in grey.

The LigPlot comparison between the structures of the complexes wild-type protease-ALES407 and V82A mutant-ALES407 (Figure 30) shows that the mutation causes minimal changes in the interaction mode of the inhibitor, primarily affecting the proline 81 residues of both chains, albeit with opposite effects. In one chain, this proline residue loses a hydrophobic interaction, while in the other, its interaction with the inhibitor is strengthened

compared to the WT structure. The hydrophobic interaction observed for the wild-type protein between the valine 82 residue, the site of the mutation, and the phenyl ring of the inhibitor is preserved despite the shorter side chain of the alanine residue at this position, in the mutant complex. Crystallographic data alone cannot determine whether this change have a stabilizing or destabilizing effect and, therefore, the performance of the inhibitor is enhanced or decreased for this mutant.

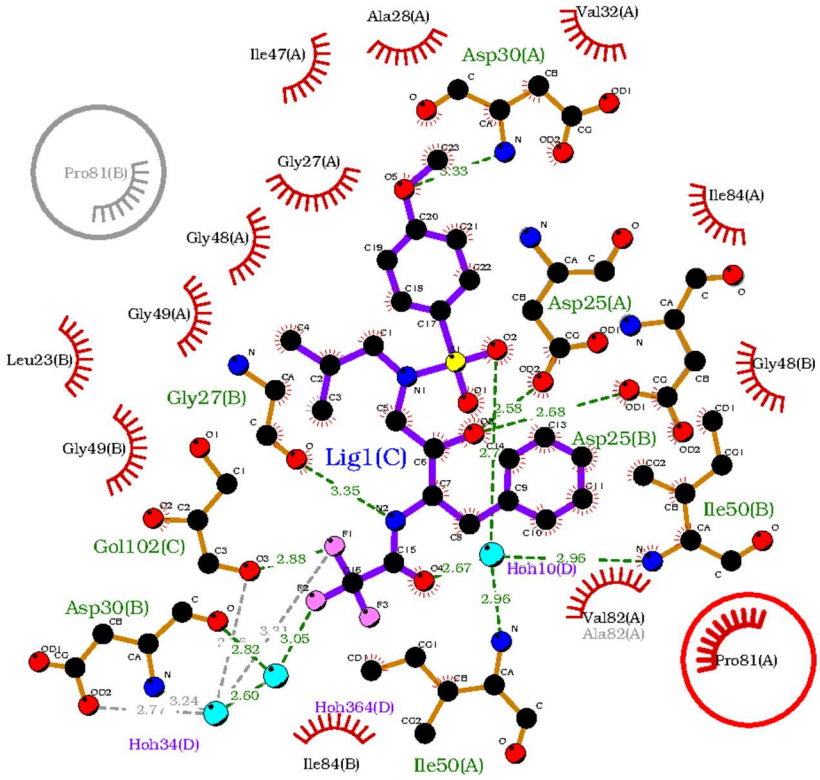


Figure 30: **LigPlot diagram comparing the wild-type protease-ALES407 complex and the V82A mutant-ALES407 complex.** In the foreground, the interactions of the ligand with the wild-type protein are depicted with colors; in the background, the interactions with the V82A mutant are shown in grey.

The analysis of the hydrophobic contacts between the V32I mutant and ALES407 shows that the mutation at position 32 does not prevent this residue from interacting with the ligand (Figure 31). However, a closer analysis of the electron density map of the mutant complex shows that the statistical disorder observed for the side chain of Val32 in the wild-type protein is absent for both the polypeptide chains when the aliphatic group is longer (Ile32) (Figure 32). In this case, the hydrophobic interactions of the Ile32 residue with the side chains of Ile84 and Leu76 contribute to the stabilization of a single conformer for the mutated side chain. The LigPlot comparison highlights also the loss of two hydrophobic

interactions in the structure of the V32I mutant protein, namely those involving residues Val82 and Ile47, likely due to minor conformational changes in the protein.

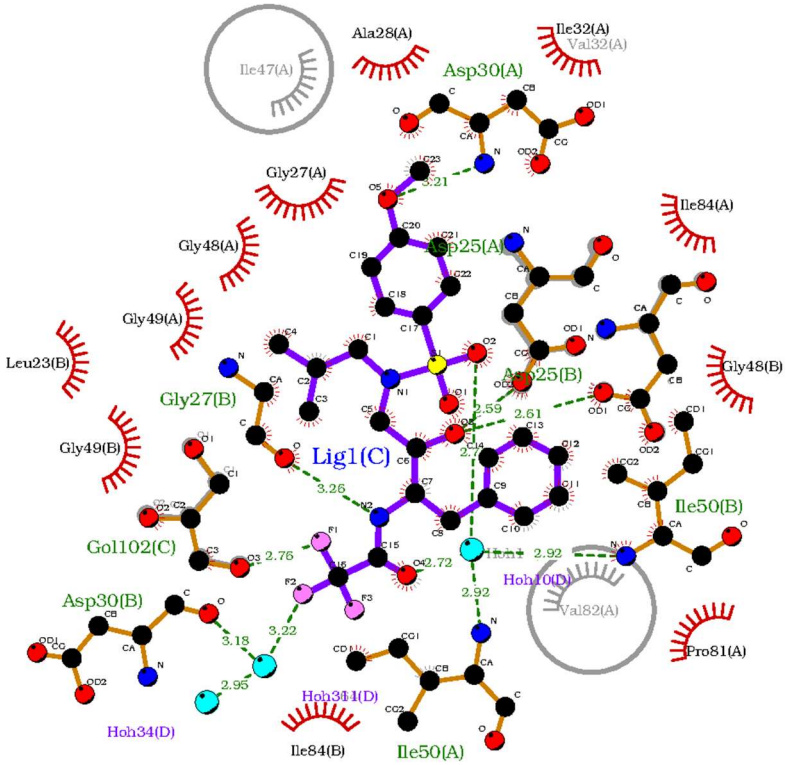


Figure 31: **LigPlot diagram comparing the wild-type protease-ALES407 complex and the V32I mutant-ALES407 complex.** In the foreground, the interactions of the ligand with the V32I mutant protein are depicted with colors; in the background, the interactions with the wild-type protein are shown in grey.

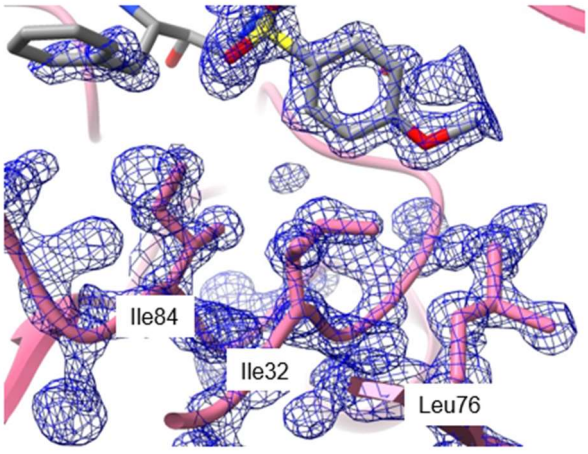


Figure 32: **Side chain of Ile32 in the V32I mutant-ALES407 complex.** Representation of the electron density map of the V32I mutant-ALES407 complex at the mutation position. The model is represented in pink, the electron density is shown as a blue contour level mesh.

4.4.9 Interactions between the V82A protease mutant and the inhibitor ALES430

The analysis of binding interactions between the ALES430 inhibitor and the V82A protease mutant shows that the hydrogen bond patters is similar to that observed for the ALES407 inhibitor in this work and, more generally, of the protease reported structures (Figure 33). In particular, hydrogen bonds involving the catalytic aspartate residues (25 and 25') are still present (Table 10), albeit with longer interaction distances (2.8 Å, compared to 2.6 Å in the structures of the ALES407 complex, Table 9). Similarly, the interaction involving the glycine 27' residue is weaker for this inhibitor compared to the lead compound of the ALES library (3.4 Å, compared to 3.3 Å for ALES407). As for most of the reported structures, including those involving the inhibitor ALES407 (Figure 28) a water molecule forms 4 hydrogen bonds, bridging the sulfonic and carbonyl groups of the inhibitor with the main chain of isoleucine residues 50 of both chains (Figure 33).

The most significant difference in the hydrophilic interactions of the ALES430 inhibitor compared to ALES407 lies in the hydrogen bonds involving the oxygen atoms of the nitro group, two water molecules, and the amide nitrogen atom of aspartate residue 30 (Table 10). Interestingly, this nitrogen atom is also involved in a hydrogen bond with the oxygen atom of the methoxy group in the ALES407 inhibitor complexes (Table 9). However, in the case of the complex with ALES430, the bond is significantly shorter, indicating a stronger and more stabilizing interaction. Additionally, three other hydrogen bonds involve two water molecules at statistical partial occupation. Despite the relatively short bond distances between the nitro group and these water molecules, the partial occupation of these positions suggest a weaker interaction, as confirmed by the high thermal displacement parameters observed for the oxygen atoms of the nitro group.

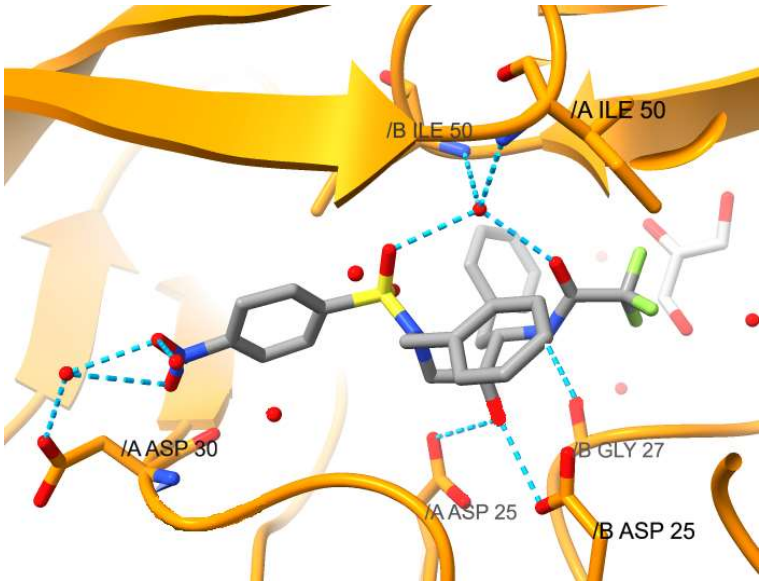


Figure 33: **Hydrogen bond interactions involving ALES430 in the active site of the V82A mutant.** Representation of the hydrogen bond interactions between the inhibitor ALES430, the HIV-1 protease residues, solvent and cryoprotectant molecules in the active site of the V82A protease mutant.

***Table 10**: Hydrogen bond interactions between the ALES430 inhibitor and the V82A mutant. Donor-acceptor distances of the hydrogen bond interactions involving the ALES430 inhibitor and the active site of the V82A mutant are reported.*

Donor	Acceptor	Bond length (Å)
ALES430, hydroxyl oxygen	ASP25, OD2	2,82
ASP25', OD1	ALES430, hydroxyl oxygen	2,77
ALES430, amine nitrogen	GLY27', O	3,40
Oxigen form H_2O 10	ALES430, carbonyl oxygen	2,81
Oxigen form H_2O 10	ALES430oxygen of the sulfonic group	2,70
ILE50, N	Oxigen form H_2O 10	2,90
ILE50' N	Oxigen form H_2O 10	2,94
Terminal oxygen of a glycerol molecule	ALES430, fluoro	2,96
ASP30, N	ALES430, O1 of the nitro group	2,65
Oxigen form H_2O 349	ALES430, O1 of the nitro group	3,27
Oxigen form H_2O 349	ALES430, O2 of the nitro group	2,60
Oxigen form H_2O 361	ALES430, O2 of the nitro group	2,40

The LigPlot software was used to analyze the differences in the hydrophobic interactions between the structure of the V82A protease-inhibitor complex with ALES430 and with ALES407 (Figure 434). The ALES430 inhibitor forms hydrophobic interactions with a leucine residue (Leu76) and an alanine residue at the mutation site (Ala82'). These interactions were not observed in the complex with the ALES407 inhibitor. Conversely, the

structure with ALES430 lacks hydrophobic interactions with a proline residue (Pro81') and a leucine residue (Leu23'), which were present in the complex with ALES407.

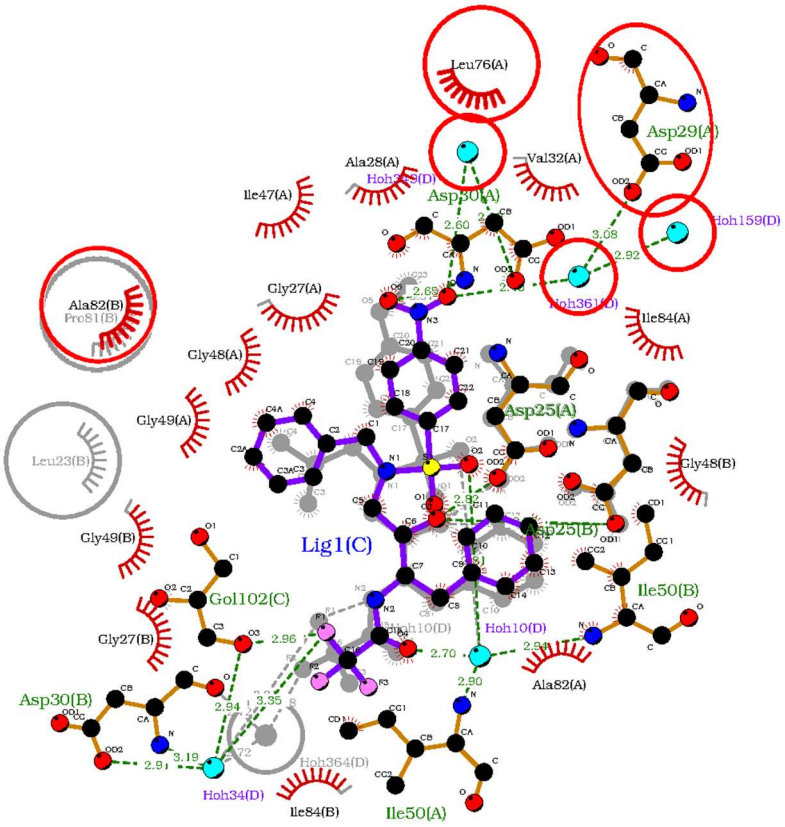


Figure 34: **LigPlot diagram comparing the complexes between the V82A protease and inhibitors ALES407 and ALES430.** In the foreground, the interactions of the ALES430 ligand with the protease mutant are depicted with colors; in the background, the interactions of the ALES407 inhibitor with the same protein are shown in grey.

5. CONCLUSIONS

As of today, numerous antiretroviral drugs are available to combat HIV-1 infection, a result of relentless research that has spanned the last 40 years. However, this virus still represents a significant public health issue in the world and, particularly, in underdeveloped countries, as new viral variants exhibiting multiple drug resistance emerge. Therefore, the development of new compounds able to inhibit viral replication is urgent and research efforts are still focused on this goal.

Recently, a compound library was synthesized at the University of Basilicata using a scaffold closely resembling the commercial drug Darunavir. Preliminary data showed that some of the compounds of this library are effective inhibitors not only for the wild-type HIV-1 protease but also for some of its medically relevant variants. The aim of this thesis was to characterize the complexes formed between various mutants of the HIV-1 protease and the proposed inhibitors, to aid the design of new antiretroviral drugs active against drug-resistant variants.

The study employs a structural approach to uncover the reasons behind the effectiveness of some of the compounds synthesized at the University of Basilicata. Notably, an initial analysis of the inhibitory activity on the wild-type protease of the library compounds revealed that one of them, ALES407, exhibits significantly higher inhibition compared to the others, achieving levels comparable to Darunavir.

Following this discovery, crystallization experiments were conducted with ALES407 and four HIV-1 protease variants: the wild-type protein and three mutants, V82S, V82A, and V32I, known for conferring resistance to available protease inhibitors. Crystallization trials were also performed using the compound ALES430, which exhibited the second-highest inhibitory activity on the wild-type protease. Crystals were obtained for complexes of all the protease variants and ALES407 and for the complex of ALES430 and the V82A protease. Diffraction experiments performed at the Elettra Synchrotron resulted in high-resolution structures (<1.33 Å), providing valuable details about the interactions between the inhibitors and the protease variants.

The crystallographic analysis highlight several key interactions that the inhibitor ALES407 forms with residues in the active site, which are crucial for understanding the compound potency. Notably, some interactions, such as the hydrogen bonds between the hydroxyl group of the inhibitor and the aspartate residues in the catalytic site, and between the carbonyl and sulfonic groups of the inhibitor and residues Ile50 of the protease dimer, bridged by a water molecule, have been previously observed with other protease inhibitors,

including Darunavir. However, the stable interaction between the methoxy group of ALES407 and an aspartate residue is novel and appears to contribute significantly to the binding efficacy of the compound.

6. REFERENCES

(1) *CD4 Count*. Testing.com. https://www.testing.com/tests/cd4-count/ (accessed 2024-10-28).

(2) *HIV and AIDS*. https://www.who.int/news-room/fact-sheets/detail/hiv-aids (accessed 2024-10-28).

(3) Engelman, A.; Cherepanov, P. The Structural Biology of HIV-1: Mechanistic and Therapeutic Insights. *Nat Rev Microbiol* **2012**, *10* (4), 279–290. https://doi.org/10.1038/nrmicro2747.

(4) Kalinichenko, S.; Komkov, D.; Mazurov, D. HIV-1 and HTLV-1 Transmission Modes: Mechanisms and Importance for Virus Spread. *Viruses* **2022**, *14* (1), 152. https://doi.org/10.3390/v14010152.

(5) Freed, E. O. HIV-1 Assembly, Release and Maturation. *Nature reviews. Microbiology* **2015**, *13* (8), 484. https://doi.org/10.1038/nrmicro3490.

(6) Chan, D. C.; Kim, P. S. HIV Entry and Its Inhibition. *Cell* **1998**, *93* (5), 681–684. https://doi.org/10.1016/s0092-8674(00)81430-0.

(7) Hu, W.-S.; Hughes, S. H. HIV-1 Reverse Transcription. *Cold Spring Harbor Perspectives in Medicine* **2012**, *2* (10), a006882. https://doi.org/10.1101/cshperspect.a006882.

(8) Peng, Y.; Zong, Y.; Wang, D.; Chen, J.; Chen, Z.-S.; Peng, F.; Liu, Z. Current Drugs for HIV-1: From Challenges to Potential in HIV/AIDS. *Frontiers in Pharmacology* **2023**, *14*, 1294966. https://doi.org/10.3389/fphar.2023.1294966.

(9) Arts, E. J.; Hazuda, D. J. HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harbor Perspectives in Medicine* **2012**, *2* (4), a007161–a007161. https://doi.org/10.1101/cshperspect.a007161.

(10) *FDA-Approved HIV Medicines | NIH*. https://hivinfo.nih.gov/understanding-hiv/fact-sheets/fda-approved-hiv-medicines (accessed 2024-10-28).

(11) Tressler, R.; Godfrey, C. NRTI Backbone in HIV Treatment. *Drugs* **2012**, *72* (16), 2051–2062. https://doi.org/10.2165/11640830-000000000-00000.

(12) Martins, S.; Ramos, M. J.; Fernandes, P. A. The Current Status of the NNRTI Family of Antiretrovirals Used Against HIV Infection. *Current Medicinal Chemistry 15* (11), 1083–1095. https://doi.org/10.2174/092986708784221467.

(13) Doms, R. W.; Moore, J. P. HIV-1 Membrane Fusion: Targets of Opportunity. *The Journal of Cell Biology* **2000**, *151* (2), f9. https://doi.org/10.1083/jcb.151.2.f9.

(14) Fung, H. B.; Guo, Y. Enfuvirtide: A Fusion Inhibitor for the Treatment of HIV Infection. *Clinical Therapeutics* **2004**, *26* (3), 352–378. https://doi.org/10.1016/S0149-2918(04)90032-X.

(15) Hunt, J. S.; Romanelli, F. Maraviroc, a CCR5 Coreceptor Antagonist That Blocks Entry of Human Immunodeficiency Virus Type 1. *Pharmacotherapy* **2009**, *29* (3), 295–304. https://doi.org/10.1592/phco.29.3.295.

(16) Hitchcock, A. M.; Kufel, W. D.; Dwyer, K. A. M.; Sidman, E. F. Lenacapavir: A Novel Injectable HIV-1 Capsid Inhibitor. *International Journal of Antimicrobial Agents* **2024**, *63* (1), 107009. https://doi.org/10.1016/j.ijantimicag.2023.107009.

(17) Pearl, L. H. The Catalytic Mechanism of Aspartic Proteinases. *FEBS Letters* **1987**, *214* (1), 8–12. https://doi.org/10.1016/0014-5793(87)80003-0.

(18) Wlodawer, A.; Vondrasek, J. INHIBITORS OF HIV-1 PROTEASE: A Major Success of Structure-Assisted Drug Design1. *Annual Review of Biophysics* **1998**, *27* (Volume 27, 1998), 249–284. https://doi.org/10.1146/annurev.biophys.27.1.249.

(19) DEBOUCK, C. The HIV-1 Protease as a Therapeutic Target for AIDS. *AIDS Research and Human Retroviruses* **1992**, *8* (2), 153–164. https://doi.org/10.1089/aid.1992.8.153.
(20) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C.-T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. Three-Dimensional Structure of Aspartyl Protease from Human Immunodeficiency Virus HIV-1. *Nature* **1989**, *337* (6208), 615–620. https://doi.org/10.1038/337615a0.
(21) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A.; Scolnick, E. M.; Sigal, I. S. Active Human Immunodeficiency Virus Protease Is Required for Viral Infectivity. *Proceedings of the National Academy of Sciences* **1988**, *85* (13), 4686–4690. https://doi.org/10.1073/pnas.85.13.4686.
(22) Wlodawer, A.; Erickson, J. W. STRUCTURE-BASED INHIBITORS OF HIV-1 PROTEASE. *Annual Review of Biochemistry* **1993**, *62* (Volume 62, 1993), 543–585. https://doi.org/10.1146/annurev.bi.62.070193.002551.
(23) Wlodawer, A.; Miller, M.; Jaskólski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. H. Conserved Folding in Retroviral Proteases: Crystal Structure of Synthetic HIV-1 Protease. *Science* **1989**, *245* (4918), 616–621. https://doi.org/10.1126/science.2548279.
(24) Louis, J. M.; Ishima, R.; Torchia, D. A.; Weber, I. T. HIV-1 Protease: Structure, Dynamics, and Inhibition. In *Advances in Pharmacology*; HIV-1: Molecular Biology and Pathogenesis Viral Mechanisms, Second Edition; Academic Press, 2007; Vol. 55, pp 261–298. https://doi.org/10.1016/S1054-3589(07)55008-8.
(25) Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N. Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas as HIV Protease Inhibitors. *Science* **1994**, *263* (5145), 380–384. https://doi.org/10.1126/science.8278812.
(26) Heaslet, H.; Rosenfeld, R.; Giffin, M.; Lin, Y.-C.; Tam, K.; Torbett, B. E.; Elder, J. H.; McRee, D. E.; Stout, C. D. Conformational Flexibility in the Flap Domains of Ligand-Free HIV Protease. *Acta Cryst D* **2007**, *63* (8), 866–875. https://doi.org/10.1107/S0907444907029125.
(27) Collins, J. R.; Burt, S. K.; Erickson, J. W. Flap Opening in HIV-1 Protease Simulated by "activated" Molecular Dynamics. *Nat Struct Biol* **1995**, *2* (4), 334–338. https://doi.org/10.1038/nsb0495-334.
(28) Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J. The Three-Dimensional Structure of the Aspartyl Protease from the HIV-1 Isolate BRU. *Biochimie* **1991**, *73* (11), 1391–1396. https://doi.org/10.1016/0300-9084(91)90169-2.
(29) Roberts, N. A.; Martin, J. A.; Kinchington, D.; Broadhurst, A. V.; Craig, J. C.; Duncan, I. B.; Galpin, S. A.; Handa, B. K.; Kay, J.; Kröhn, A. Rational Design of Peptide-Based HIV Proteinase Inhibitors. *Science* **1990**, *248* (4953), 358–361. https://doi.org/10.1126/science.2183354.
(30) Könnyű, B.; Sadiq, S. K.; Turányi, T.; Hírmondó, R.; Müller, B.; Kräusslich, H.-G.; Coveney, P. V.; Müller, V. Gag-Pol Processing during HIV-1 Virion Maturation: A Systems Biology Approach. *PLoS Comput Biol* **2013**, *9* (6), e1003103. https://doi.org/10.1371/journal.pcbi.1003103.
(31) Kumar, M.; Mandal, K.; Blakeley, M. P.; Wymore, T.; Kent, S. B. H.; Louis, J. M.; Das, A.; Kovalevsky, A. Visualizing Tetrahedral Oxyanion Bound in HIV-1 Protease

Using Neutrons: Implications for the Catalytic Mechanism and Drug Design. *ACS Omega* **2020**, *5* (20), 11605–11617. https://doi.org/10.1021/acsomega.0c00835.
(32) Kumar, M.; Mandal, K.; Blakeley, M. P.; Wymore, T.; Kent, S. B. H.; Louis, J. M.; Das, A.; Kovalevsky, A. Visualizing Tetrahedral Oxyanion Bound in HIV-1 Protease Using Neutrons: Implications for the Catalytic Mechanism and Drug Design. *ACS Omega* **2020**, *5* (20), 11605–11617. https://doi.org/10.1021/acsomega.0c00835.
(33) Wang, Y.; Liu, Z.; Brunzelle, J. S.; Kovari, I. A.; Dewdney, T. G.; Reiter, S. J.; Kovari, L. C. The Higher Barrier of Darunavir and Tipranavir Resistance for HIV-1 Protease. *Biochemical and Biophysical Research Communications* **2011**, *412* (4), 737–742. https://doi.org/10.1016/j.bbrc.2011.08.045.
(34) De Clercq, E. Anti-HIV Drugs: 25 Compounds Approved within 25 Years after the Discovery of HIV. *International Journal of Antimicrobial Agents* **2009**, *33* (4), 307–320. https://doi.org/10.1016/j.ijantimicag.2008.10.010.
(35) Ali, A.; Bandaranayake, R. M.; Cai, Y.; King, N. M.; Kolli, M.; Mittal, S.; Murzycki, J. F.; Nalam, M. N. L.; Nalivaika, E. A.; Özen, A.; Prabu-Jeyabalan, M. M.; Thayer, K.; Schiffer, C. A. Molecular Basis for Drug Resistance in HIV-1 Protease. *Viruses* **2010**, *2* (11), 2509–2535. https://doi.org/10.3390/v2112509.
(36) Blackard, J. T.; Cohen, D. E.; Mayer, K. H. Human Immunodeficiency Virus Superinfection and Recombination: Current State of Knowledge and Potential Clinical Consequences. *Clin Infect Dis* **2002**, *34* (8), 1108–1114. https://doi.org/10.1086/339547.
(37) Prabu-Jeyabalan, M.; Nalivaika, E. A.; King, N. M.; Schiffer, C. A. Viability of a Drug-Resistant Human Immunodeficiency Virus Type 1 Protease Variant: Structural Insights for Better Antiviral Therapy. *Journal of Virology* **2003**, *77* (2), 1306–1315. https://doi.org/10.1128/jvi.77.2.1306-1315.2003.
(38) Wu, T. D.; Schiffer, C. A.; Gonzales, M. J.; Taylor, J.; Kantor, R.; Chou, S.; Israelski, D.; Zolopa, A. R.; Fessel, W. J.; Shafer, R. W. Mutation Patterns and Structural Correlates in Human Immunodeficiency Virus Type 1 Protease Following Different Protease Inhibitor Treatments. *Journal of Virology* **2003**, *77* (8), 4836–4847. https://doi.org/10.1128/jvi.77.8.4836-4847.2003.
(39) *HIV Drug Resistance Database*. https://hivdb.stanford.edu/ (accessed 2024-10-29).
(40) Weber, I. T.; Wang, Y.-F.; Harrison, R. W. HIV Protease: Historical Perspective and Current Research. *Viruses* **2021**, *13* (5), 839. https://doi.org/10.3390/v13050839.
(41) Shafer, R. W.; Dupnik, K.; Winters, M. A.; Eshleman, S. H. A Guide to HIV-1 Reverse Transcriptase and Protease Sequencing for Drug Resistance Studies. *HIV sequence compendium* **2001**, *2001*, 1.
(42) Aoki, M.; Das, D.; Hayashi, H.; Aoki-Ogata, H.; Takamatsu, Y.; Ghosh, A. K.; Mitsuya, H. Mechanism of Darunavir (DRV)'s High Genetic Barrier to HIV-1 Resistance: A Key V32I Substitution in Protease Rarely Occurs, but Once It Occurs, It Predisposes HIV-1 To Develop DRV Resistance. *mBio* **2018**, *9* (2), 10.1128/mbio.02425-17. https://doi.org/10.1128/mbio.02425-17.
(43) Wondrak, E. M.; Louis, J. M. Influence of Flanking Sequences on the Dimer Stability of Human Immunodeficiency Virus Type 1 Protease. *Biochemistry* **1996**, *35* (39), 12957–12962. https://doi.org/10.1021/bi960984y.
(44) Mahalingam, B.; Louis, J. M.; Reed, C. C.; Adomat, J. M.; Krouse, J.; Wang, Y. F.; Harrison, R. W.; Weber, I. T. Structural and Kinetic Analysis of Drug Resistant

Mutants of HIV-1 Protease. *Eur J Biochem* **1999**, *263* (1), 238–245. https://doi.org/10.1046/j.1432-1327.1999.00514.x.
(45) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 2005; pp 571–607. https://doi.org/10.1385/1-59259-890-0:571.
(46) Tsumoto, K.; Ejima, D.; Kumagai, I.; Arakawa, T. Practical Considerations in Refolding Proteins from Inclusion Bodies. *Protein Expression and Purification* **2003**, *28* (1), 1–8. https://doi.org/10.1016/S1046-5928(02)00641-1.
(47) Geremia, S.; Demitri, N.; Wuerges, J.; Benedetti, F.; Berti, F.; Tell, G.; Randaccio, L. A Potent HIV Protease Inhibitor Identified in an Epimeric Mixture by High-Resolution Protein Crystallography. *ChemMedChem* **2006**, *1* (2), 186–188. https://doi.org/10.1002/cmdc.200500063.
(48) Olajuyigbe, F. M.; Demitri, N.; De Zorzi, R.; Geremia, S. Developing HIV-1 Protease Inhibitors through Stereospecific Reactions in Protein Crystals. *Molecules* **2016**, *21* (11), 1458. https://doi.org/10.3390/molecules21111458.
(49) Olajuyigbe, F.; Demitri, N.; Geremia, S. Investigation of 2-Fold Disorder of Inhibitors and Relative Potency by Crystallizations of HIV-1 Protease in Ritonavir and Saquinavir Mixtures. *Crystal Growth & Design* **2011**, *11* (10), 4378–4385. https://doi.org/10.1021/cg200514z.
(50) Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 2), 125–132. https://doi.org/10.1107/S0907444909047337.
(51) Evans, P. R.; Murshudov, G. N. How Good Are My Data and What Is the Resolution? *Acta Crystallogr D Biol Crystallogr* **2013**, *69* (Pt 7), 1204–1214. https://doi.org/10.1107/S0907444913000061.
(52) Potterton, L.; Agirre, J.; Ballard, C.; Cowtan, K.; Dodson, E.; Evans, P. R.; Jenkins, H. T.; Keegan, R.; Krissinel, E.; Stevenson, K.; Lebedev, A.; McNicholas, S. J.; Nicholls, R. A.; Noble, M.; Pannu, N. S.; Roth, C.; Sheldrick, G.; Skubak, P.; Turkenburg, J.; Uski, V.; von Delft, F.; Waterman, D.; Wilson, K.; Winn, M.; Wojdyr, M. CCP4i2: The New Graphical User Interface to the CCP4 Program Suite. *Acta Crystallogr D Struct Biol* **2018**, *74* (Pt 2), 68–84. https://doi.org/10.1107/S2059798317016035.
(53) Vagin, A.; Teplyakov, A. Molecular Replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 1), 22–25. https://doi.org/10.1107/S0907444909042589.
(54) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr D Biol Crystallogr* **2011**, *67* (Pt 4), 355–367. https://doi.org/10.1107/S0907444911001314.
(55) Sheldrick, G. M. Crystal Structure Refinement with SHELXL. *Acta Crystallogr C Struct Chem* **2015**, *71* (Pt 1), 3–8. https://doi.org/10.1107/S2053229614024218.
(56) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 4), 486–501. https://doi.org/10.1107/S0907444910007493.
(57) Long, F.; Nicholls, R. A.; Emsley, P.; Graæulis, S.; Merkys, A.; Vaitkus, A.; Murshudov, G. N. AceDRG: A Stereochemical Description Generator for Ligands.

Acta Crystallogr D Struct Biol **2017**, *73* (Pt 2), 112–122. https://doi.org/10.1107/S2059798317000067.
(58) Mahalingam, B.; Louis, J. M.; Reed, C. C.; Adomat, J. M.; Krouse, J.; Wang, Y. F.; Harrison, R. W.; Weber, I. T. Structural and Kinetic Analysis of Drug Resistant Mutants of HIV-1 Protease. *Eur J Biochem* **1999**, *263* (1), 238–245. https://doi.org/10.1046/j.1432-1327.1999.00514.x.
(59) Auld, D. S.; Farmen, M. W.; Kahl, S. D.; Kriauciunas, A.; McKnight, K. L.; Montrose, C.; Weidner, J. R. Receptor Binding Assays for HTS and Drug Discovery. In *Assay Guidance Manual*; Markossian, S., Grossman, A., Arkin, M., Auld, D., Austin, C., Baell, J., Brimacombe, K., Chung, T. D. Y., Coussens, N. P., Dahlin, J. L., Devanarayan, V., Foley, T. L., Glicksman, M., Gorshkov, K., Haas, J. V., Hall, M. D., Hoare, S., Inglese, J., Iversen, P. W., Lal-Nag, M., Li, Z., Manro, J. R., McGee, J., McManus, O., Pearson, M., Riss, T., Saradjian, P., Sittampalam, G. S., Tarselli, M., Trask, O. J., Weidner, J. R., Wildey, M. J., Wilson, K., Xia, M., Xu, X., Eds.; Eli Lilly & Company and the National Center for Advancing Translational Sciences: Bethesda (MD), 2004.
(60) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J Appl Cryst* **1993**, *26* (2), 283–291. https://doi.org/10.1107/S0021889892009944.
(61) Laskowski, R. A.; Swindells, M. B. LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* **2011**, *51* (10), 2778–2786. https://doi.org/10.1021/ci200227u.

FINAL REMARKS

This thesis explores two complementary biochemical strategies aimed at addressing key biomedical challenges.

Firstly, it introduces an innovative approach to overcoming the persistent difficulty of discovering new binders for diagnostic and theragnostic applications. By leveraging new insights and computational technologies, traditional drug discovery methods have been enhanced, leading to more efficient and viable solutions. A notable example of this progress is the use of computational algorithms to develop antibody fragments with optimized binding properties, streamlining the discovery process.

Secondly, the thesis highlights the enduring significance of classical drug discovery techniques, particularly those relying on protein structural data. These methods have been the cornerstone of biomedical advancements for over fifty years, resulting in the successful development of numerous therapeutic molecules currently available on the market. Despite the rise of newer technologies, traditional strategies, such as those used in the development of HIV protease inhibitors, remain powerful and essential in applied biomedical research.
Together, both novel and established approaches are indispensable for driving progress in biomedical research, fostering innovation, and ensuring the continued advancement of therapeutic and diagnostic tools.