

DEVELOPMENT AND CHARACTERIZATION OF AFFINITY-MATURATED NANOBODIES TARGETING B-LYMPHOCYTE ANTIGEN CD19

**A Thesis Submitted to the Graduate School of Izmir Institute of Technology in Partial
Fulfillment of the Requirements for the Degree of MASTER OF SCIENCE**

in Molecular Biology & Genetics

by

Ghazi ALALI

IZMIR INSTITUTE OF TECHNOLOGY

DECEMBER, 2024

We approve the thesis of **Ghazi AL ALI**

Examining Committee Members:

Dr. Asst. Prof. Şerife Ayaz Güner

Department of Molecular Biology & Genetics, Izmir Institute of Technology

Dr. Asst. Prof. Hümeysra Taşkent Sezgin

Department of Bioengineering, Izmir Institute of Technology

Dr. Asst. Prof. Aslı Sade Memişoğlu

Department of Mathematics & Science Education, Dokuz Eylül University

12/12/2024

Dr. Asst. Prof. Şerife Ayaz Güner

Supervisor, Department of Molecular Biology & Genetics, Izmir Institute of Technology

Prof. Dr. Güneş Özhan

Head of the Department of Molecular
Biology & Genetics, Izmir Institute of
Technology

Prof. Dr. Mehtap EANES

Dean of the Graduate School

ACKNOWLEDGMENT

I would like to dedicate this thesis to everyone who has been a part of my academic journey here at Izmir Institute of Technology and throughout. I would like to thank my advisor Assistant Professor Şerife Ayaz-Güner for providing me with the opportunity and mentorship. I would also like to dedicate this thesis to my parents, friends, and loved ones.

The thesis has been part of the project titled “Development and Characterization of Affinity-maturated Nanobodies Targeting B-lymphocyte Antigen CD19” supported by Izmir Institute of Technology, Research University Support Program (ADEP project no. 2022IYTE-3-0040).

PREFACE

In this thesis the development of nanobodies specific to CD19 is investigated through multiple paradigms of reported literature on the *in silico* rational design and affinity maturation concerning protein science and engineering. Herein, the developed nanobodies are produced *in vitro* and momentarily reflected on their functional characterization to be continued as part of the general project which this thesis was based on.

Epistemically, I stand where Wittgenstein once said “The limits of my language means the limits of my world”. To arm oneself with the knowledge and the words to describe, it is a worthy cause that is noble to pursue.

ABSTRACT

DEVELOPMENT AND CHARACTERIZATION OF AFFINITY-MATURATED NANOBODIES TARGETING B-LYMPHOCYTE ANTIGEN CD19

B-Cell Acute Lymphoblastic Leukemia (B-ALL) is a subtype of leukemia characterized by the uncontrolled proliferation of B-lymphocytes. The diagnosis of B-ALL relies on blood counts, immunophenotyping, cytogenetic analysis, and molecular tests. Treatment strategies for B-ALL include chemotherapy, radiotherapy, allogeneic bone marrow transplantation, and stem cell transplantation. In the diagnosis and treatment of B-ALL, antibodies or antibody fragments specific to CD19, a protein known as a B-lymphocyte antigen, are commonly used. However, producing antibodies specific to targeted antigens is often cost-prohibitive. Nanobodies, which constitute the heavy chain variable region of heavy-chain only camelid antibodies have been elucidated to be more advantageous. Three nanobodies (Nbs), which were chosen from a natural Nb-expressing phage display library in the study of (H. Wang et al. 2021a), with affinity to CD19 were tested. The nanobodies, which are derived from the heavy-chain antibodies of camelids and constitute the single-domain heavy variable region of the antibodies, were matured in-silico. Two parent nanobodies, out of three, were chosen alongside four variants based on their in-silico performance. The purpose of this study is to verify the applicability of our in-house affinity maturation protocol (structure prediction, filtering, and scoring), by in-vitro production and functional characterization of affinity matured variants against their parents nanobodies. Moreover, it is reflected on multiple paradigms concerning in-silico affinity maturation of nanobodies, and how models fare under the reality of experimentation.

ÖZET

B-lenfosit Antijeni Hedefleyen Afinitesi Artırılmış Nanokorların Geliştirilmesi ve Karakterizasyonu

B-Hücre Akut Lenfoblastik Lösemi (B-ALL), B-lenfositlerin kontrolsüz çoğalması ile karakterize bir lösemi alt tipidir. B-ALL tanısı kan sayımları, immünofenotipleme, sitogenetik analiz ve moleküler testlere dayanır. B-ALL tedavi stratejileri arasında kemoterapi, radyoterapi, allojenik kemik iliği nakli ve kök hücre nakli bulunur. B-ALL'nin tanı ve tedavisinde, bir B-lenfosit antijeni olarak bilinen CD19 proteinine özgü antikorlar veya antikor parçaları yaygın olarak kullanılmaktadır. Ancak, hedef antijenlere özgü antikor üretimi genellikle yüksek maliyetlidir. Yalnızca ağır zincirli deve antikorlarının ağır zincir değişken bölgesini oluşturan nanokütlelerin (nanobody, Nb) daha avantajlı olduğu gösterilmiştir. (H. Wang ve ark., 2021a) çalışmasında, doğal bir Nb-ekspresyonlu faj gösterim kütüphanesinden seçilen ve CD19'a afinitesi olan üç nanokütle (Nb) test edilmiştir. Deve antikorlarının ağır zincirlerinden türetilen ve antikorların tek zincirli ağır değişken bölgesini oluşturan bu nanokütleler in-silico olarak olgunlaştırılmıştır. Üç nanokütleden ikisi, in-silico performanslarına dayanarak dört varyantla birlikte seçilmiştir. Bu çalışmanın amacı, kurum içinde geliştirilen afinite olgunlaştırma protokolünün (yapı tahmini, filtreleme ve derecelendirme) uygulanabilirliğini, afinitesi olgunlaştırılmış varyantların in-vitro üretimi ve fonksiyonel karakterizasyonu ile ana nanokütlelere karşı doğrulamaktır. Ayrıca, nanokütlelerin in-silico afinitesi olgunlaştırılmasıyla ilgili çeşitli paradigmlar ve modellerin deneysel gerçeklik karşısındaki durumu üzerine değerlendirmelerde bulunulmuştur.

TABLE OF CONTENTS

LIST OF FIGURES	ix
LIST OF TABLES	x
CHAPTER 1: INTRODUCTION & LITERATURE REVIEW	1
1.1 B-Acute Lymphoblastic Leukemia:	1
1.2 Clinical Relevance of CD19 as a Target in Leukemia:	2
1.3 Nanobodies as an Alternative to Antibody-based Therapies:	2
1.4 <i>In silico</i> Methods for Structural Investigation of Protein Space:	4
1.5 Rational Design and Affinity Maturation of Nanobodies:	5
1.6 Thesis Project:	6
CHAPTER 2: MATERIALS & METHODS	8
2. In-Silico Rational Mutation & Affinity Maturation of anti-CD19 Nanobodies	8
2.1 Sequence Allocation & Analysis:	8
2.2 Sequence-based Homology Modelling of CD19 and anti-CD19 Nanobodies:	8
2.3 Rigid-body Docking for the determination of protein-protein interactions:	9
2.4 Candidate Complex Selection for Rational Design & Affinity Maturation of anti-CD19 nanobodies:	9
2.5 Rational Design & Affinity Maturation of Anti-CD19 nanobodies:	10
2.6 Molecular Dynamics Simulations and Trajectory Analysis for Validation and Candidate Selection of Affinity Matured Anti-CD19 Nanobodies:	12
3. Recombinant Expression, Purification, & Functional Analysis of anti-CD19 Nanobodies:	13
3.1 Nanobody Plasmid Design:	13
3.2 Competent Cell Preparation and Transformation of Plasmids:	13
3.3 Recombinant Nanobody Expression:	15
3.3.1 First-phase Induction Tests:	15
3.3.2 Second-phase Induction tests:	16
3.3.3 Final-phase Induction tests:	17
3.3.4 Optimized Expression and Purification of Nanobodies:	18
3.4 Determination of Binding Affinity of anti-CD19 Nanobodies using Enzyme-linked Immunosorbent Assay (ELISA):	19

CHAPTER THREE: RESULTS & DISCUSSION.....	21
4. <i>In-Silico</i> Rational Mutation & Affinity Maturation of anti-CD19 Nanobodies.....	21
4.1 Anti-CD19 Nanobodies Sequence Allocation & Analysis:	21
4.2 Sequence-based Homology Modelling of CD19 and anti-CD19 Nanobodies and Structure Preparation for Rigid-body Docking:	22
4.3 Rigid-body Docking for the determination of protein-protein interactions:.....	24
4.4 Rational Design & Affinity Maturation of Anti-CD19 nanobodies:.....	28
4.5 Molecular Dynamics Simulations and Trajectory Analysis for Validation and Candidate Selection of Affinity Matured Anti-CD19 Nanobodies:	35
5. <i>In-vitro</i> Overexpression, Purification & Functional Analysis of anti-CD19 Nanobodies	38
5.1 Anti-CD19 Nanobody Plasmids, and Construct Design:.....	38
5.2 Recombinant Production and Purification of Nanobodies:	40
5.3 Determination of Binding Affinity of anti-CD19 Nanobodies using Enzyme-linked Immunosorbent Assay (ELISA):	44
CHAPTER 4: CONCLUSION & REMARKS.....	48
CHAPTER 5: SUPPLEMENTARY SECTION.....	49
6.1 Induction of Transformed anti-CD19 BL21 (DE3):	50
6.2 Rational Design & Affinity Maturation of Anti-CD19 nanobodies:.....	51
6.3 Optimization Steps for Overexpression and Purification of nanobodies:.....	52
REFERENCES	62

LIST OF FIGURES

Figure 1: The structure of Antibody vs. Nanobody.....	3
Figure 2: Implemented L25 orthogonal matrix design for 25 experiments for Nb1-3	11
Figure 3: Sequence analysis of anti-CD19 nanobodies	21
Figure 4: CD19 3D structure prediction and annotation.....	23
Figure 5: 901, 902, 903 (Nb1, 2, 3) 3D structure prediction and annotation.....	24
Figure 6: Proposed binding patches as per blind docking runs.	26
Figure 7: Binding pose for each of Nb1, Nb2, Nb3 to CD19, and superimposition of the binding states to the glycosylated form of CD19.	27
Figure 8: Nb1 Taguchi analysis results	31
Figure 9: Nb2 Taguchi analysis results	32
Figure 10: Consensus-based sequence design of Nb4.	34
Figure 11: Calculated binding free energy of matured anti-CD19 nanobodies vs. original anti-CD19 nanobodies.	37
Figure 12: pET-22b(+)construct for recombinant overexpression of anti-CD19 nanobodies.	39
Figure 13: Purification of Nb1 and validation via SDS-PAGE.	42
Figure 14: Purification of Nb1_T and validation via SDS-PAGE.	43
Figure 15: Dose-response curve of Nb1_T and Nb1 ELISA absorbance measurements using a primary histidine antibody and a secondary HRP conjugated secondary antibody.	45
Figure 16: Dose-response curve of Nb1_T and Nb1 ELISA absorbance measurements using Anti-Alpaca IgG HRP conjugated antibody.....	46

(The Supplementary figures are only related to the supplementary section maintaining a step-by-step report on the optimization of the recombinant expression and purification protocols.)

LIST OF TABLES

Table 1: ExPASy-ProtParam anti-CD19 nanobodies' protein sequence analysis.....	22
Table 2: Residue constraints introduced for CD19	28
Table 3: Response tables for SN ratios ranking mutations based on significance.....	31
Table 4: Response tables for SN ratios ranking mutations based on significance.....	32
Table 5: Nanobodies (original, and mutants) and their docking score of both Blues2 and PRODIGY.	35
Table 6: Binding free energy score at the end of each MD run calculated in KJ/mol	36



CHAPTER 1

INTRODUCTION & LITERATURE REVIEW

1.1 B-Acute Lymphoblastic Leukemia:

Alterations in both the proliferation and differentiation of precursor lymphoid cells which is a consequence of genetic alteration and chromosomal abnormalities are the hallmark of acute lymphoblastic leukemia (ALL) (Terwilliger and Abdul-Hay 2017). B-acute lymphoblastic leukemia (B-ALL), a subtype of leukemia, is a cell malignancy that is prevalent in the demographic space of children and adolescents as it has poorer prognosis in older age groups of the population (Zhang et al. 2017). B-ALL malignant cells arise from a single progenitor through the differentiation of hematopoietic stem cells (HSCs) which constitute the precursors to blood cells. This constitutes B-ALL malignancies as clonal in nature being derived from B-cell progenitors. B-ALL ranks as the top malignancies leading to mortality amongst the population (Podyma-Inoue et al. 2016). The incidence of B-ALL is estimated per year to occur in one to five per one-hundred thousand people (Dores et al. 2012). Symptoms of B-ALL patients include failure of bone marrow (BM) which accounts for the inability of BM to produce adequate amount of mature blood cells caused by the accumulation of immature lymphoblasts (Loghavi, Kutok, and Jorgensen 2015). Concurrently, failure in the function of BM leads to clinical manifestations such as cytopenia (low count of blood cells). Cytopenia can be accompanied with leukocytosis (high white blood cell counts), but not in all cases. Diagnosis of B-ALL is carried out through multiple standards including the study of cell morphology, cytogenic or genetic analysis, or immunophenotyping (Chiaretti, Zini, and Bassan 2014). Treatment of B-ALL which have led to the treatment of more than 80% of children patients and around 40% of adult patients have consisted of interventions in many forms: chemotherapy, radiotherapy, and stem cell transplantation, and allogeneic bone marrow transplantation (Pui and Jeha 2007). Moreover, both un-conjugated and conjugated monoclonal interventions for the treatment of B-ALL have been applied against B-ALL associated antigens. Monoclonal antibodies (mAbs) have

been conjugated with either immunotoxins, radioactive molecules, and anti-B-ALL drugs (Dillman 2006; Stasi et al. 2001).

1.2 Clinical Relevance of CD19 as a Target in Leukemia:

CD19, a pan-B-cell marker, and part of the immunoglobulin super-family is expressed throughout the development of B-cells from the early stages of B-cell differentiation up until the stage of plasma cell differentiation where it ends. This makes CD19 an attractive target for treatment interventions against B-ALL and other B-cell related malignancies (Gambella et al. 2022). CD19, with a molecular weight of 95 kD, is a type 1 transmembrane glycoproteins with a targetable extracellular N-terminal domain that has been used in the development of both monoclonal antibody and chimeric antigen receptor (CAR) T-cell therapies (Hammer 2012). Moreover, CD19 has a single transmembrane domain, and a C-terminal cytoplasmic signaling domain. CD19 has an integral functional part as a signaling block on the surface of B-cells (K. Wang, Wei, and Liu 2012) which adds to its relevance in B-cell cellular fate. As any signaling molecule, CD19 could be seen as a double-edged sword as its abnormal expression can lead to B-cell related diseases (X. Li et al. 2017). CD19 influences the fate of B-cells in that it is a positive regulator of its proliferation through its amplification of Src family protein tyrosine kinases (PTK) and mitogen-activated protein kinase (MAPK) activity. At the same time, CD19 has been implicated in negative roles in B-cells such as the regulation of calcium flux. CD19 link to ALL development has not been clearly elucidated however it stands as a well-defined tumor-specific antigen (Liu et al. 2020). Treatments targeting CD19 with conventional antibodies in comparison to bispecific, conjugated antibodies, or CAR-T cells have had limited reach on the modulation of malignancies.

1.3 Nanobodies as an Alternative to Antibody-based Therapies:

Conventional antibodies have been used in targeted therapies for different diseases. A conventional antibody is made up of four polypeptide chains: two heavy chains and two light chains which account for their large molecular weight of 150 kD (Kothari et al. 2024). Concurrently, three constant domains (CH1, 2, 3) and one variable domain (VH) make up each of the heavy chains of an antibody. One constant domain (CL) and one variable domain

(VL) comprise the light chain of an antibody. Animals of the *Camelidae* family express antibodies that are missing the light chain of a conventional antibody, these antibodies are referred to as heavy-chain only antibodies (HCAb). The single variable heavy domain of HCABs (VHH) has been expressed as a single domain antibody (sdAb) which has been trademarked with the term NANOBODY® (Hamers-Casterman et al. 1993). The structure of a nanobody is elucidated in (Figure 1) comparing it with the structure of an antibody.

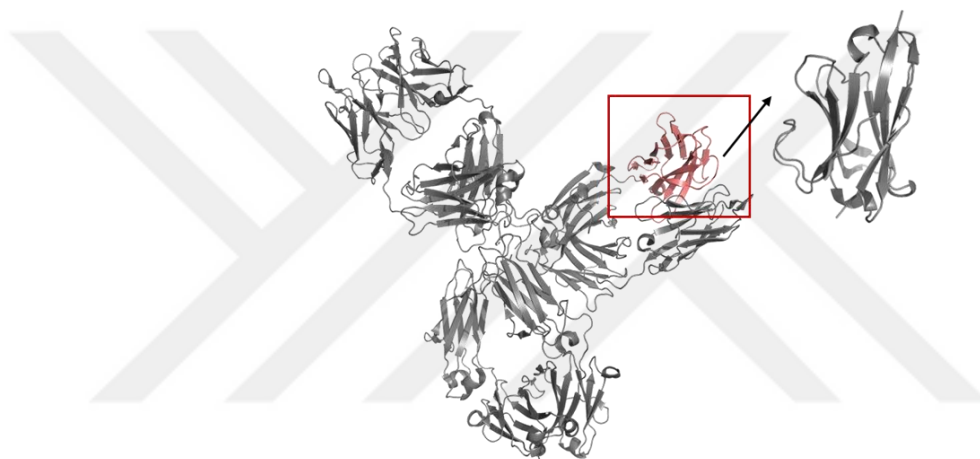


Figure 1: The structure of Antibody vs. Nanobody. Visualization of the structure of an Ab (to the left), with the structure of what would be its hypothetical Nb (VHH) portion highlighted in red and magnified (to the right).

Since their discovery, nanobodies have been considered as an alternative to antibodies for the development of diagnostics and therapeutic applications due to their favorable and net efficient properties (Muyldermans 2013; 2021c; 2021a). First, nanobodies can be well expressed in the periplasm of *Escherichia coli* which ensures adequate disulfide bond formation due to the periplasm's oxidizing environment and the assistance of chaperones in nanobodies' folding (Arbabi-Ghahroudi, Tanha, and MacKenzie 2005). The purification of nanobodies is considered simple, typically involving affinity chromatography, followed by size-exclusion chromatography to get rid of any dimerized or aggregated impurities. Nanobodies retain robust behavior: thermal stability, high specificity and affinity, small size

and unique paratope, versatility and modular design, and the methods by which they are generated. Their smaller sizes (~15 kD), and convex structure allows for nanobodies to penetrate and bind to cavities on their target that might otherwise be hidden. Nanobodies also diffuse readily into different cellular tissues and environments. There are multiple ways in which antigen specific nanobodies can be generated. Moreover, nanobodies can be incorporated into diverse designs (i.e., bispecific constructs, fusions with various protein actors like enzymes and immunotoxins). Finally, nanobodies retain homology with human VH family 3 which can be attributed to lower risk of immunogenicity. One of the ways nanobodies (Nbs) can be generated is through the immunization of Animals of the *Camelidae* family like camels, alpacas, or llamas. The immunization is carried out over a short duration of two to three months. Small blood samples allocated from the immunized camelids are then used for the cDNA synthesis from the purified lymphocytes. The amplified VHH repertoire of the purified lymphocytes is cloned in a phage display vector that is transformed into bacterial cells where proteomics methods are applied for the identification and creation of nanobody libraries after enrichment and application of binding assays like ELISA. This is followed by lead identification and drug design.

1.4 *In silico* Methods for Structural Investigation of Protein Space:

The prediction of the 3D structure of proteins have been transformed with the introduction of AlphaFold and other AI-powered algorithms (Efrimidis et al. 2023). The ability to accurately predict the structure of most proteins only from their amino acid sequences has constituted a major leap from the tiresome albeit fundamentally beneficial structure resolving techniques like X-ray diffraction and Cryo-Em. Accessibility to solutions like the AlphaFold database with more than 214 million structures, or ColabFold within Goggle Collaboratory environment allowed for the acceleration at which protein structures and function relations could be investigated. Other solutions for the accurate prediction of protein structure or interactions is available such as RoseTTAFold (Baek et al. 2021).

The *In silico* study of protein-protein interactions (PPIs) allows for the engineering and the elucidation on the function of proteins (Tsuchiya, Yamamori, and Tomii 2022). Physics based methods like ZDOCK (Pierce et al. 2014), HDOCK (Yan et al. 2020a), or HADDOCK (Dominguez, Boelens, and Bonvin 2003) allow for the sample of the conformational space

of proteins complexes. Physics-based modelling solutions such as HDock allow the prediction of protein-protein interactions in a reliable manner when compared with deep learning models such as AlphaFold for predicting interactions of complexes that are more diverse or unseen (Buttenschoen, Morris, and Deane 2024). The introduction of physical and chemical restraints by physics-based modelling solutions provides more plausibility. Many available solutions have been introduced and are actively researched in the area of structural biology.

1.5 Rational Design and Affinity Maturation of Nanobodies:

There has not been a single consensus on the rational design and affinity maturation of nanobodies in the literature as the field continued to grow (Chaudhuri et al. 2022; Cheng et al. 2019a; Hacisuleyman and Erman 2019; Hu et al. 2020; Muyldermans 2021b; Vivcharuk et al. 2017; Yau et al. 2005). One way where nanobodies have been tried to be improved in regard to their binding affinity to their antigen has been the introduction of point mutations to the CDRs of a nanobody. Single point mutations to any of the 18 amino acids with the exception to cysteine and proline are usually introduced to the CDRs with emphasis on CDR3 more than CDR1 and CDR2. This process is reiterated on where the best performing single point mutations are combined to generate an affinity matured nanobody. This process can involve the random mutation of amino acids within different regions of the nanobody (Ye et al. 2020). Moreover, one point mutation could prove more than apt in the development of better binding capacity to antigens. Mutation of FR residues has also been implicated in the improvement of affinity and stability of nanobodies. Another method that is simple but can be consider in the rational design of nanobodies is generating nanobody hybrids based on the sequence consensus of the top performing Nbs within a library (McMahon et al. 2018).

The nature of CDRs being mostly disordered, dictate that for the maturation of a nanobody, the choice of residues could be narrowed down to loop-forming residues or rather residues implicated in disordered structures (Gavenonis et al. 2014; Minuchehr and Goliaei 2005). On the other hand, statistical modes like the design of experiment (DOE) approach have been introduced to generate affinity matured nanobodies which save ample amount of time and computational cost as opposed to the generation of large sets of randomly mutated nanobody libraries. Methods to the rational design of nanobodies have been supplemented

with selection criteria (contact distances) of residue to be mutated expediting the process of development of *in silico* designed nanobodies which narrows down the search space while allowing for the potential discovery of improved nanobodies (Poustforoosh et al. 2022; 2023a).

1.6 Thesis Project:

A limited number of nanobodies have been developed against the CD19 protein (Banihahsemi et al. 2018; H. Wang et al. 2021b). In this thesis project, three anti-CD19 nanobodies were obtained from (H. Wang et al., 2021). These three anti-CD19 nanobodies were chosen from a natural Nb-expressing phage display library as they retained “high affinity and specificity” for CD19 and then were tested as part of a chimeric antigen receptor T cells (Nb CAR-T) constructs to investigate their cytotoxicity against lymphoma cells. The aim of the thesis project has been to develop and characterize rationally designed and affinity matured nanobodies with enhanced affinity for CD19 which could be used in both diagnostic and therapeutical interventions. To achieve this, nanobodies targeting CD19 were modelled in-silico using ColabFold, and the antigen-nanobody interface was predicted using HDock protein-protein docking. The binding affinities of the nanobodies were optimized using the available paradigms of in-silico rational design and affinity maturation which include statistical approach through design of experiment (DOE), iterative mutation cycles, and consensus-based sequence design. Two parent nanobodies out of the three anti-CD19 originals were chosen with four different mutant nanobodies. The selected mutant anti-CD19 nanobodies performed well in comparison to their parent nanobodies through both the scoring of their predicted complexes with CD19 and subsequent validation in binding free energy calculations of generated molecular dynamics simulations. The selected anti-CD19 nanobodies were produced and purified. Moreover, the production and purification methods have been optimized accordingly. Enzyme-linked Immunosorbent Assays have been applied to verify the binding energy and compare between the original and mutant anti-CD19 nanobodies.

The thesis has been part of the project titled “Development and Characterization of Affinity-matured Nanobodies Targeting B-lymphocyte Antigen CD19” supported by supported by Izmir Institute of Technology, Research University Support Program (ADEP

project no. 2022IYTE-3-0040). The project has aimed to identify at least one commercially viable nanobody with enhanced affinity for CD19, characterized by its biochemical and biophysical properties, and tested across various applications. The conclusion of the project aims to utilize said-developed anti-CD19 nanobodies in immunoassays for the diagnosis of the disease and in the development of CD19-targeted therapeutic methods. The discussed thesis herein reports on the progress, knowledge and experience that has been gained through this project facilitating the rapid and efficient development of nanobodies for other target proteins, directly contributing to the advancement of local and national biotechnological products.



CHAPTER 2

MATERIALS & METHODS

2. In-Silico Rational Mutation & Affinity Maturation of anti-CD19

Nanobodies

2.1 Sequence Allocation & Analysis:

The sequence of B-lymphocyte antigen CD19 was obtained from UniProt (accession number: P15391). The sequences of the three anti-CD19 nanobodies were obtained from (H. Wang et al., 2021). As per the article, the anti-CD19 nanobodies were chosen from a natural Nb-expressing phage display library as they retained “high affinity and specificity” for CD19 and then were tested as part of a chimeric antigen receptor T cells (Nb CAR-T) constructs to investigate their cytotoxicity against lymphoma cells.

The obtained sequences (Supplementary: Table 1) of the anti-CD19 nanobodies were analyzed using ExPASy-ProtParam Tool (Gasteiger et al., 2005) to predict the following physical and chemical parameters of the protein sequences: molecular weight (MW), isoelectric point (pI), estimated half-life, instability index (II), aliphatic index, and grand average of hydropathicity (GRAVY). The sequences of the anti-CD19 nanobodies were aligned using Clustal Omega (Sievers et al. 2011) and viewed using Jalview (Waterhouse et al. 2009).

2.2 Sequence-based Homology Modelling of CD19 and anti-CD19

Nanobodies

Following the allocation of CD19 and anti-CD19 nanobodies sequences, ColabFold (Mirdita et al., 2022) which is based on AlphaFold 2.1 (Jumper et al., 2021) was used for the sequence-based homology structural prediction of CD19 (Figure 2) and anti-CD19 nanobodies (Figure 3). The default parameters of ColabFold were used. Top ranked predicted structures based on pLDDT scores were used. Predicted CD19 3D structure was compared with available Protein Data Bank structures (7JIC, 6AL5). PDB2PQR (Dolinsky et al. 2004)

was used for the preparation of predicted structures of CD19 and anti-CD19 for protein-protein docking. For CD19 predicted structure, only the extra cellular domain (ECD) plus the signal peptide (SP) was used for protein-protein docking prediction.

2.3 Rigid-body Docking for the determination of protein-protein interactions:

Following the preparation of the structures of the three anti-CD19 nanobodies and CD19, HDOCK (Yan et al. 2020b) was used for protein-protein rigid-body docking. Two preliminary docking runs were done to cluster for viable protein-protein predicted interactions which would be used for the introduction of parameters for following targeted runs. First, a blind docking run was conducted which generates 2000 nanobody-CD19 predicted complex structure; no preferred parameters were input for the blind docking run. The blind docking run was followed by one that incorporated glycans into CD19 structure. N-linked glycans were generated using Glycam (<https://glycam.org/>) where the N-linked Glycan sites were obtained based on reported literature as per UniProt. Both the blind and glycosylated rigid-body docking runs were analyzed and used with available CD19 interaction literature (Susa et al. 2021; Teplyakov et al. 2018) to determine the binding patches on CD19. The CDRs of anti-CD19 nanobodies including residues from surrounding FRs were input as parameters for the docking runs based on IMGT numbering scheme using AbRSA (L. Li et al. 2019). A final docking run was conducted using knowledge derived from the preliminary runs, reported literature of CD19 interactions, and nanobody numbering.

2.4 Candidate Complex Selection for Rational Design & Affinity

Maturation of anti-CD19 nanobodies:

For all HDOCK runs, the generated complex structures of (Nb-CD19) were filtered using an in-house multi-step pipeline. First, “paratopic enforcement” was applied where the interaction of CDRs’ residues with CD19 was prioritized. Predicted complex structures that did not show interactions between CDRs of Nbs and CD19 were omitted from the candidate complex selection. Second, filtered complexes were ranked based on a combination of two scoring methods for rigid-body docked protein-protein complexes which are Bluues2 (Miguel Angel Soler et al. 2023), and PRODIGY (Xue et al. 2016). Bluues2 scores were

leveraged for the analysis of the electrostatic properties of the complex of Nb-CD19. Blueues2 was complemented by PRODIGY which predicts the binding affinity.

The scores were calculated following energy minimization with CHARMM forcefield and complex preparation using PDB2PQR (Dolinsky et al. 2007). Both Blueues2 and PRODIGY scores coupled with structural analysis of docked complexes were used to choose viable complexes that would be used for the maturation of the three anti-CD19 nanobodies. Top ranked filtered structures were manually analyzed emphasizing an almost perpendicular binding epitope of the nanobodies to CD19 which boosted complexes with maximal interface interaction at the CDR level common with observed X-ray crystallography modelled structures of nanobody interactions (Shi et al. 2022).

2.5 Rational Design & Affinity Maturation of Anti-CD19 nanobodies:

Protein-protein interactions of chosen candidate predicted structures for rational design and affinity maturation were analyzed using LigPlot+ (Laskowski and Swindells 2011). Nanobody residues were ranked based on the least contact distance of interaction, and most interactions formed at the interface of Nb-CD19 (electrostatic, hydrogen bond, etc.). Three paradigms were chosen for the rational design and affinity maturation of anti-CD19 nanobodies: Design of Experiment (DOE) using the Taguchi method, iterative introduction of mutations by increment of one residue (up to five), and consensus-based sequence design.

From the chosen candidate complexes, 5 candidate residues to be mutated were chosen based on interface residues with the least contact distances to CD19, and residues that formed multiple interactions (electrostatic, hydrogen bond, etc.). Based on these criteria, the candidate residues for mutation for each Nb were: Nb1 (Y103, R31, T28, Y54, L102); Nb2 (P106, G104, D54, T30, D62), Nb3 (Q32, R56, D54, G102, Y29). The candidate residues were chosen to represent each of the three CDRs and/or the surrounding FRs if applicable. The five residues were to be mutated to one of the loop-forming residues (W, F, H, D, Y).

To test the first paradigm of the generation of affinity matured nanobodies. A design of experiment approach based on the Taguchi method was applied to select for the best performing mutations. An L25 orthogonal matrix with five factors and five levels (5^5) was generated for each anti-CD19 nanobody. The factors are the chosen residue positions at which

the best mutations at each position would be tested. The variables of the experiment or the levels are the mutations tested (W, F, H, D, Y). Consequently, this would allow to predict the best performing mutation at each chosen residue position in 25 experiments or generated mutants as opposed to generating 5^5 possible mutated representation for each nanobody thus saving time and computation costs.

Moreover, other candidate residues and mutations were introduced based on structural considerations at the interface of interaction. Mutants were generated based on an iterative introduction of mutations by increments of one residue. For each nanobody, five mutants were generated whereby the first mutant has one residue change, the second having two residue changes, etc. Multiple runs for each nanobody were tested throughout for the sake of serendipity. This paradigm follows on conventional ration designs of affinity matured nanobodies reported in the literature where single point mutations are introduced, followed by a combination of best performing mutations at each residue position.

	res103	res31	res28	res54	res102
1	W	W	W	W	W
2	F	F	F	F	F
3	H	H	H	H	H
4	D	D	D	D	D
5	Y	Y	Y	Y	Y
6	W	F	H	D	Y
7	F	H	D	Y	W
8	H	D	Y	W	F
9	D	Y	W	F	H
10	Y	W	F	H	D
11	W	H	Y	F	D
12	F	D	W	H	Y
13	H	Y	F	D	W
14	D	W	H	Y	F
15	Y	F	D	W	H
16	W	D	F	Y	H
17	F	Y	H	W	D
18	H	W	D	F	Y
19	D	F	Y	H	W
20	Y	H	W	D	F
21	W	Y	D	H	F
22	F	W	Y	D	H
23	H	F	W	Y	D
24	D	H	F	W	Y
25	Y	D	H	F	W

	res106	res104	res54	res30	res62
1	W	W	W	W	W
2	F	F	F	F	F
3	H	H	H	H	H
4	D	D	D	D	D
5	Y	Y	Y	Y	Y
6	W	F	H	D	Y
7	F	H	D	Y	W
8	H	D	Y	W	F
9	D	Y	W	F	H
10	Y	W	F	H	D
11	W	H	Y	F	D
12	F	D	W	H	Y
13	H	Y	F	D	W
14	D	W	H	Y	F
15	Y	F	D	W	H
16	W	D	F	Y	H
17	F	Y	H	W	D
18	H	W	D	F	Y
19	D	F	Y	H	W
20	Y	H	W	D	F
21	W	Y	D	H	F
22	F	W	Y	D	H
23	H	F	W	Y	D
24	D	H	F	W	Y
25	Y	D	H	F	W

	res32	res56	res54	res102	res29
1	W	W	W	W	W
2	F	F	F	F	F
3	H	H	H	H	H
4	D	D	D	D	D
5	Y	Y	Y	Y	Y
6	W	F	H	D	Y
7	F	H	D	Y	W
8	H	D	Y	W	F
9	D	Y	W	F	H
10	Y	W	F	H	D
11	W	H	Y	F	D
12	F	D	W	H	Y
13	H	Y	F	D	W
14	D	W	H	Y	F
15	Y	F	D	W	H
16	W	D	F	Y	H
17	F	Y	H	W	D
18	H	W	D	F	Y
19	D	F	Y	H	W
20	Y	H	W	D	F
21	W	Y	D	H	F
22	F	W	Y	D	H
23	H	F	W	Y	D
24	D	H	F	W	Y
25	Y	D	H	F	W

Figure 2: Implemented L25 orthogonal matrix design for 25 experiments for Nb1-3, respectively. The same orthogonal matrix design was applied for Nb1-3, factors are residue positions chosen for mutation. Residues W, F, H, D, Y are the five levels implemented in the L25 matrix. Blues2 score was considered for each experiment and used in further analysis.

Lastly, one consensus-based sequence design approach was applied for Nb1 where its CDR3 region with surrounding FRs were reconstructed based on consensus of Nb1 and

Nb2 mainly introducing changes to the CDR3 regions of Nb. The new sequence termed Nb4 had noticeable improvement in stability as per ExPASy-ProtParam Tool. The paradigms tested for the rational design and affinity maturation of candidates allowed to cover multiple approaches whereby a review of current methods could be reviewed.

The generated mutated nanobodies were redocked to CD19 and filtered based the same criteria used for the selection of candidate complexes for the design of nanobodies. This allowed to reinforce the sampling of possible binding patches and the validation of observed patterns of docked poses throughout the discovery period to find better design of affinity matured nanobodies. For the Taguchi method, the best residue at each position was selected based on the generated main effect plot of the signal-to-noise (SN) ratios where the experimental results were measured for robustness based on *Larger-the-Better* scheme for maximizing the response based on Blueues2 score. Different combinations based on the highest-ranking residues for each nanobody were implemented into another round of maturation and compared with parent nanobodies after redocking. Only Blueues2 was considered in the ranking of each experiment guided by our observations throughout the discovery and test process.

2.6 Molecular Dynamics Simulations and Trajectory Analysis for Validation and Candidate Selection of Affinity Matured Anti-CD19 Nanobodies:

As MD simulations are computationally expensive and thus could not be part of the design process in our case, they were only used to validate top matured candidates and compare them to the performance of the original three anti-CD19 nanobodies. The complexes of the nanobodies in were used as the starting point for 50-150 ns MD simulations after preparing them for the MD run. GROMACS (Abraham et al. 2015) was used to carry out the molecular simulations. The atomic interactions within each complex were parametrized using OPLS-AA force field. The complexes were solvated in water using SPCE model. The complexes were neutralized at a salt concentration of 0.15 M where Na^+ and Cl^- ions are added to the system. Following energy minimization of the generated system for each complex, the molecular dynamics simulation was carried out based on the following parameters used for equilibration of the complexes: target temperature of 300K using the Langevin thermostat,

and pressure at 1 atm. MM/PBSA (E. Wang et al. 2019) was used for the calculation of binding free energy. Binding free energy plots were produced using xmgrace. All the matured Nbs performed better than the original anti-CD19 nanobodies except Nb3 where work on it was discontinued.

The binding free energy scores were taken from the end of each MD run. 150 ns MD runs were carried out to check if any structural imbalance could occur in the Nb-CD19 complex. The choice to carry out MD instead of counting on scoring approximation aided in the selection of the candidate matured Nbs and validated some of our choices as other matured candidates did not perform as per their initial scoring using either Blueues2 or PRODIGY.

3. Recombinant Expression, Purification, & Functional Analysis of anti-CD19 Nanobodies:

3.1 Nanobody Plasmid Design:

Anti-CD19 nanobody sequences were reverse-translated, codon-optimized for *E. coli* expression using Jcat (Grote et al. 2005) with the added options of avoid rho-independent transcription terminators, avoid prokaryotic ribosome binding sites, and avoid cleavage sites of restriction enzymes. Nanobody-pET-22b(+) expression vectors were ordered through GenScript and codon-further optimized through its proprietary software (GenSmart) for better expression in *E. coli*. A certificate of analysis from GenScript containing both the validation of pET-22b(+) vector and DNA sequencing of each designed construct cloned into the vector was provided by GenScript. The Nb constructs were designed according to the conventional periplasmic expression protocol which is to target the C-terminal 6X His-tagged single domain of the nanobody into the periplasm using an N-terminal periplasmic translocation signal (Pectate lyase B: PelB), (Nagy-Fazekas et al. 2023a). TEV protease site was introduced following the PelB sequence.

3.2 Competent Cell Preparation and Transformation of Plasmids:

DH5 α competent cells were prepared according to “The Inoue Method for Preparation and Transformation of Competent *E. coli*: “Ultra-competent Cells” provided in

(Sambrook 2001) in (Molecular Cloning – A laboratory Manual). Inoue transformation buffer was prepared and sterilized by filtration, aliquoted and stored at -20°C . A single DH5 α bacterial colony was picked from a plate that has been incubated for 16 hours at 37°C and transferred into 25 mL of SOB medium. The bacteria were incubated for 6-8 hours at 37°C at 250 rpm. The culture was then inoculated into SOB medium and incubated overnight at 18°C . OD₆₀₀ was measured in the following morning and the culture was transferred to an ice-water bath at OD₆₀₀ of 0.55 for 10 minutes. Cells were harvested by centrifugation at 2500 g at 4°C for 10 minutes. The supernatant was discarded, and the cells were gently resuspended in ice-cold Inoue transformation buffer. Another round of centrifugation was conducted. DH5 α Cells were then resuspended in ice-cold Inoue transformation buffer and glycerol stocks were prepared, aliquoted and frozen at -80°C . BL21 (DE3) competent cells were prepared using the calcium chloride method in a manner similar to the preparation of DH5 α competent cells. Aseptic techniques were followed for all steps involving the preparation of competent cells, and only chilled apparatus was used in each step after the first incubation to obtain the highest efficiency from the protocol.

The ordered plasmids stored at -20°C were thawed, diluted with autoclaved ddH₂O, and an aliquot of 10 ng/ μL was prepared for each of the six anti-CD19 nanobody vectors for transformation into DH5 α competent cells. The transformation was conducted according to the same procedure for the Inoue competent cell preparation protocol whereby 10 ng/ μL was added onto 50 μL of thawed competent DH5 α cells and rested on ice for 30 minutes. This was followed by heat-shock at 42°C for 90 seconds and the transfer of samples to an ice-bath for 2 minutes. SOC medium was added, and each transformation sample was incubated at 37°C shaking at 200 rpm for one hour. The transformation sample of each of the six anti-CD19 nanobodies' pET-22b(+) expression vectors was plated on Ampicillin positive (100 $\mu\text{g/mL}$) agar plates and incubated at 37°C overnight for 16 hours. pGEM-T vector was used as positive control. Anti-CD19 transformants exhibited growth comparable to that of pGEM-T transformation. Kanamycin positive (50 $\mu\text{g/mL}$) plates were used for negative control. The transformed DH5 α plates were then grown in overnight starter cultures at 37°C shaking at 200 rpm for subsequent plasmid isolation.

Plasmid isolation was conducted as provided in (Norgen Biotek Corp.) plasmid miniprep kit. The obtained plasmid concentration for each of the six anti-CD19 plasmids were measured by Nanodrop. Isolated plasmids were diluted to yield concentration in the range of 10-25 ng/ μ L as per recommended for the concentration of plasmids to be used in the transformation protocol of calcium chloride competent BL21 (DE3) cells (better transformation efficiency can be achieved for DH5 α competent cells if the need arises). The same procedure done for DH5 α was conducted for the transformation of isolated anti-CD19 into BL21 (DE3) cells apart from the duration of heat-shock which was 60 seconds. The difference in heat-shock duration comes from the fact that the Inoue transformation protocol yields competent cells with stronger membranes as compared to the chemically simple calcium chloride method. Already available calcium chloride competent BL21 (DE3) cells that had acceptable transformation efficiency were present and as such no Inoue competent BL21 (DE3) were prepared. Anti-CD19 BL21 (DE3) plates were stored at 4 °C. Glycerol stocks from transformed BL21 (DE3) were stored at -80 °C to be used for over-expression of anti-CD19 nanobodies.

3.3 Recombinant Nanobody Expression:

3.3.1 First-phase Induction Tests:

A protocol that combined knowledge from the induction and purification of nanobodies based on multiple literature articles was tried out (Chao et al. 2022; de Marco 2020; Kariuki and Magez 2021; Nagy-Fazekas et al. 2023b; Pardon et al. 2014; Salema and Fernández 2013) (Supplementary Table 2). The protocol carried out for the expression of a preliminary batch of anti-CD19 nanobodies was as follows: overnight starter cultures are incubated at 37 °C shaking at 200 rpm. 1% or 2% of the overnight culture is inoculated following overnight incubation (16-18 hours) into a certain volume of Lysogeny broth (LB). Bacterial cultures were induced at OD₆₀₀ of 0.6-0.8 with the addition of 0.2 mM Isopropyl--D-1-thiogalactopyranoside (IPTG) at 25 °C for 24 hours which was reported to result in the greatest yield of periplasmically expressed nanobodies. Samples were acquired at intervals for SDS-PAGE analysis. The cells were harvested at the around 24-hour mark and centrifuged at 5000 rpm at 4 °C for 30 minutes where the supernatant was discarded and pelleted cells were stored at -20 °C for later analysis.

To test the validity of the protocol in producing soluble nanobodies, a certain pelleted volume from each of the induced bacterial cultures was resuspended in Tris buffer (20 mM, pH= 7.4) to check for expression. The samples were centrifuged at 5000 rpm after sonication for 5 times with intervals (30 seconds of sonication: 30 seconds chilled on ice). The supernatant constituting the soluble fraction was obtained, and so for the pellet constituting the insoluble fraction. 100 μ L SDS-PAGE samples were prepared by mixing 4X loading dye (33 μ L) with samples collected. The SDS-PAGE samples were boiled at 95 °C for 10 minutes and ran on 15% polyacrylamide gel with a protein ladder (SERVA Unstained Protein Standard IV) at 90 V until the loading dye migrated out of the stacking gel followed by 120 V for the remainder of run for around one hour and 40 minutes. The gel was later stained with Coomassie blue R250 for ~30 minutes, then destained overnight. The gel picture is provided only as a reference to the insolubility of produced nanobodies (as it not of the highest quality) (Supplementary figure 1).

3.3.2 Second-phase Induction tests:

Induction's parameters (temperature, duration of induction) were tested using Nb2 as reference. Moreover, osmotic shock treatment was used for the extraction of periplasmic proteins where the protein is trafficked after its expression. Ni-NTA pull-down was conducted following osmotic shock treatment to check for the presence of eluted Nb2. The following parameters of induction were tested. Both induction temperature and duration of induction were tested as such: 37 °C for four hours, 30 °C at (4, 6, 8, and 12 hours), 25 °C at (4, 6, 8, 18, and 24 hours), 18 °C at (18 and 24 hours). Nb2 was chosen to conduct the optimization as it was not matured to act as a reference for the rest of anti-CD19 nanobodies assuming its nature to not be as recalcitrant.

Osmotic shock treatment was conducted according to (Nagy-Fazekas et al. 2023b) (adjusted). Following harvest of the cells by centrifugation, the pellet is resuspended in around 50 mL/L of ice-cold TES buffer (30 mM Tris, 20 w/v% sucrose, 1 mM EDTA, pH= 8) constituting the hypertonic treatment part of the osmotic shock protocol. The resuspended sample is incubated overnight at 4 °C on a shaker. The sample is centrifuged at 5000 rpm at 4 °C where the supernatant is obtained constituting the first hypertonic fraction. The cells are then resuspended in (30 mM Tris, 5 mM MgCl₂) and shaken for 2-4 hours on ice. The sample

is then centrifuged, and the supernatant is extracted constituting the first periplasmic extract (PE). The volumes of both the hypertonic and hypotonic solutions used are the same. The process can be repeated where the duration for the osmotic treatment is decreased. The samples are treated with the hypertonic solution for 2-4 hours, and the hypotonic solution for 1-2 hours. The osmotic shock treatment yields 4 different samples that can be analyzed (first hypertonic fraction, first periplasmic extract, second hypertonic fraction, and second periplasmic extract).

In some cases, purification of the nanobodies was tried by Ni-NTA pulldown as follows: The samples are incubated with binding buffer equilibrated Ni-NTA beads overnight at 4 °C in binding buffer (20 mM KH₂PO₄, 500 mM NaCl, and 20 mM imidazole). A fraction is then collected as flowthrough. The beads are washed three times with the binding buffer and fractions are collected for SDS-PAGE analysis. Finally, Elution buffer is added (20 mM KH₂PO₄, 500 mM NaCl, and 500 mM imidazole). The sample is shaken for 10 minutes on ice and the elution fraction is collected. The fractions are collected through centrifugation at 500 rpm for 2 minutes at 4 °C.

3.3.3 Final-phase Induction tests:

Adjusting for both temperature and duration of the induction did not yield any observed soluble band for Nb2 at 15 kD. The concentration of IPTG used to induce the samples were concurrent with the literature as for both the temperatures and durations investigated for the induction of soluble nanobodies. It was reasoned that the protein overexpression should be lowered to prevent aggregation in the periplasm. A survey of the literature on solving the insolubility problem in *E. coli* protein overexpression was carried out. Amongst the different strategies done to mitigate the problem of insolubility was the use of low concentration of IPTG, and inoculating the bacteria to in LB media containing up to 3% ethanol (Bhatwa et al. 2021; Chhetri, Kalita, and Tripathi 2015).

Two concentrations of 0.05 and 0.005 mM IPTG were tested for the induction of soluble nanobodies at both 30 °C and 25 °C harvesting cells at different intervals with 3% ethanol added to the inoculation media. The presence of up to 3% of ethanol in the inoculation media simulates the effect of heat-shock in bacteria (Chhetri, Kalita, and Tripathi 2015)

which would allow for the production of heat-shock proteins that might ensure proper folding of the expressed nanobody.

3.3.4 Optimized Expression and Purification of Nanobodies:

Transformed BL21 (DE3) glycerol stocks are inoculated into LB and an overnight starter culture of adequate volume is grown. The overnight grown culture is then inoculated into a medium that consists of ~2% pure ethanol and LB. The bacteria are then induced using 0.005 mM IPTG when OD₆₀₀ reaches 0.6-0.8. The induction is carried out overnight at 25 °C or for 24 hours at 16 °C. The induced samples are centrifuged at 5000 rpm at 4 °C for 30 minutes. The samples are resuspended with binding buffer (0.5% Tween-20) and centrifuged again at 5000 rpm at 4 °C for 30 minutes to wash any remaining LB. The samples are then stored in -20 °C for the later purification of nanobodies. The over-expression and purification are further discussed in the supplementary section where 25 kD impurity was investigated for possible hypothesized nanobody dimer formation.

The samples are thawed on ice and resuspended in binding buffer. The samples are sonicated up to 8 times at intervals of 30 seconds while resting on ice. The samples are then centrifuged twice at 20,000 rpm at 4 °C for 30 minutes. The supernatant is filtered with 0.45 µm filters. Fast protein liquid chromatography using ÄKTAprime plus and 1 mL Ni-NTA Cytiva column was used in the purification of nanobodies. The following Degassed buffers are used: Binding buffer (20 mM KH₂PO₄, 500 mM NaCl, and 50 mM imidazole, 0.5% Tween-20) and Elution Buffer (20 mM KH₂PO₄, 500 mM NaCl, and 250 mM imidazole, 0.5% Tween-20). Samples for SDS-PAGE analysis are taken (flowthrough, wash, and elution). The purification is carried out based on Cytiva instruction for affinity chromatography. One-step elution is done at 100% concentration of elution buffer. It should be noted that these values were adjusted based on multiple tests that involved the use of multi-step elution and gradient elution. The eluted samples are concentrated and buffer exchanged in sterile PBS with 0.05% Tween-20 using 10,000 MWCO concentrator tubes. Glycerol stocks at 25% are then prepared and samples are aliquoted as such and stored in either -20 °C or -80 °C based on designated use. For this report, Nb1 and Nb1_T have been tested after optimization with Nb2.

3.4 Determination of Binding Affinity of anti-CD19 Nanobodies using Enzyme-linked Immunosorbent Assay (ELISA):

Preliminary ELISA experiments were conducted where parameters were reiterated and optimized accordingly. The parameters involved antigen coating concentration, and the antibodies used for signal detection.

SPL Maxibinding Immunoplate (Cat. No. 32296) was utilized for the ELISA assay. Recombinant Human CD19-Fc Chimera (BioLegend; Cat. No 789004) was prepared at a 100 ng/mL concentration in coating buffer (PBS; 0.15 mol/L, pH 7.2) and a volume of 50 μ L was applied to wells in duplicates. The plate was covered and incubated on an orbital shaker overnight (~18 hours) at 4 °C. The antigen solution was aspirated, and later was decanted. The plates were inverted to dry on absorbent paper. Wells were washed three times with 200 μ L of PBS-T (0.15 mol/L, 0.05% Tween 20, pH 7.2) per well. The wells were blocked with the 5% skim milk prepared in PBS-T and incubated on an orbital shaker at room temperature for two hours. Following the blocking step, the wells were washed three-five times with 200 μ L PBS-T. Nb1 and Nb1_T samples were prepared in the blocking buffer and were applied to wells at the following concentrations diluted from glycerol stocks (1 μ g/mL, 500 ng/mL, 250 ng/mL, 100 ng/mL, 50 ng/mL, and 25 ng/mL). Two empty wells were blocked and used as negative controls. CD19-only coated wells were also used as positive controls. Following the addition of the nanobody samples, the plate was incubated and shaken for two hours. A washing step was applied in a similar fashion to the one preceding it. 100 μ L of Primary unconjugated anti-mouse his-tag antibody (Cat. No. bs-0287R) was applied after washing at 1:400 dilution and the plate was incubated for one hour then washed. 100 μ L of anti-mouse IgG secondary antibodies conjugated with horseradish peroxidase (HRP) was applied for one hour. After wash, 100 μ L of 3,3',5,5'-Tetramethylbenzidine substrate (TMB) was applied to each well and the plate was incubated on the orbital shaker for 15 minutes in the dark. The reaction was stopped with 1 M hydrochloric acid, and the absorbance was measured at 450 nm with a microplate reader. Data analysis was then followed. A dose-response curve was plotted.

The second run had the following changes: 100 μ L of recombinant CD19 was applied at a concentration of 1 μ g/mL. PBS-T was used instead of blocking buffer in the preparation

of the samples. The third run used conjugated Goat Anti-Alpaca IgG (Peroxidase AffiniPure™ Goat Anti-Alpaca IgG, VHH domain; Jackson ImmunoResearch) instead of the application of primary his-tagged antibody with secondary HRP conjugated one. The wells were blocked with 1% bovine serum albumin (BSA) prepared in PBS-T. The samples were prepared in the blocking buffer accordingly.



CHAPTER 3

RESULTS & DISCUSSION

4. *In-Silico* Rational Mutation & Affinity Maturation of anti-CD19 Nanobodies

4.1 Anti-CD19 Nanobodies Sequence Allocation & Analysis:

Following the sequence allocation of the anti-CD19 nanobodies, the sequences were analyzed. First, Clustal Omega was used for the generation of multiple sequence alignment (MSA) where conservation was highlighted using Jalview. The sequences showed conservation in the framework regions (FRs) while the complementarity-determining region 3 (CDR3) showed the most variation in terms of amino acids. Concurrently, CDR1 and CDR2 of the three anti-CD19 nanobodies showed slightly less variation (Figure 3).

```
901/1-127 1 QVQLQESGGGSGVQAGGSLRLSCAVRRPTDSRNCMAWFRQAPGEQREAVAGTIDIYR-ITGYAESVKGRFTISQDNAK 75
902/1-127 1 QVQLQESGGGSGVQTGGSLRLSCAASGFSTYTGHHMGWFRQVAAQEREWVGSITYDPGTTFYADSVKGRFTISQDVAK 76
903/1-127 1 QVQLQESGGGSGVQGGSLRLSCAVSGATYMIQYMGWFRQAPGKEREGVAVIHSDD-RTTYADSVKGRFTISQDNAN 75

901/1-127 76 NTLFLQMNSLKPEDSGTYCYCAARPCKYGSEWRRSASDFLYWGQGTQVTVSS 127
902/1-127 77 KVMVYLQMMNLKPTDTGIYFCAAEPVADGP-PLMSRDRCFGSWGQGTQVTVSS 127
903/1-127 76 HTLYLQMNSLKPEDTGMVVCASSLSKGGYYRGDLVPGQFTYWGRGTQVTVSS 127
```

Figure 3: Sequence analysis of anti-CD19 nanobodies. Multiple sequence alignment (MSA) of the anti-CD19 Nbs (901, 902, 903 for Nb1, Nb2, and Nb3, respectively). The sequence alignment was viewed using Jalview.

ProtParam prediction of the following physical and chemical parameters of the anti-CD19 nanobody sequences: molecular weight (MW), isoelectric point (pI), instability index (II), aliphatic index, and grand average of hydropathicity (GRAVY) is provided (Table 1).

Table 1: ExPASy-ProtParam anti-CD19 nanobodies' protein sequence analysis.

Nanobody	MW	Theoretical pI	Instability index (II)	Aliphatic index	GRAVY
Nb1 (901)	14kD	9.03	52.07	59.21	-0.553
Nb2 (902)	13.7kD	6.91	34.83	57.56	-0.277
Nb3 (903)	13.8kD	7.96	38.46	62.13	-0.443

The anti-CD19 nanobodies share similar molecular weight. The theoretical pI shows variability between the anti-CD19 nanobodies: 9.03, 6.91, and 7.96 for Nb1, Nb2, and Nb3, respectively. pH would be important in buffer preparation for each of the nanobodies as proteins tend to aggregate near their pI. The instability index prediction shows Nb1 to be unstable in comparison to Nb2, and Nb3. The instability index (II) predicts how much a protein is stable in a test tube. Nb1 is less likely to remain stable or might degrade quicker than other nanobodies based on its sequence. This has led to the decision to introduce a consensus-based sequence design of Nb1 based on the MSA of the three anti-CD19 nanobodies (refer to the in-silico rational design and affinity maturation section of the study). The aliphatic index showed positive thermostability for all nanobodies. The predicted GRAVY values show Nb1 to be most hydrophilic while Nb2 being the least hydrophilic. The negative value of GRAVY highlighting the general tendency for all nanobodies to be hydrophilic. This means that proteins would be more likely to interact in aqueous environments and are less likely to form aggregates. It should be noted that the Instability Index (II) predicted value was not highlighted in (H. Wang et al., 2021) which could be mainly argued to have been disregarded due to the nanobody being part of a much larger protein construct; thus, rendering it negligible.

4.2 Sequence-based Homology Modelling of CD19 and anti-CD19

Nanobodies and Structure Preparation for Rigid-body Docking:

The complete sequence of CD19 (UniProt: P15391) 3D structure was predicted using default parameter ColabFold (Figure 4). The signal peptide with the extracellular domain of CD19 was obtained from the generated ColabFold 3D structure prediction. The truncated

regions of CD19 would be used as the receptor to which the anti-CD19 nanobodies would be docked against. The ColabFold predicted 3D structure of CD19 was contrasted for validation by superimposing it to the X-ray diffraction-resolved structure of CD19 (PDB: 6AL5).

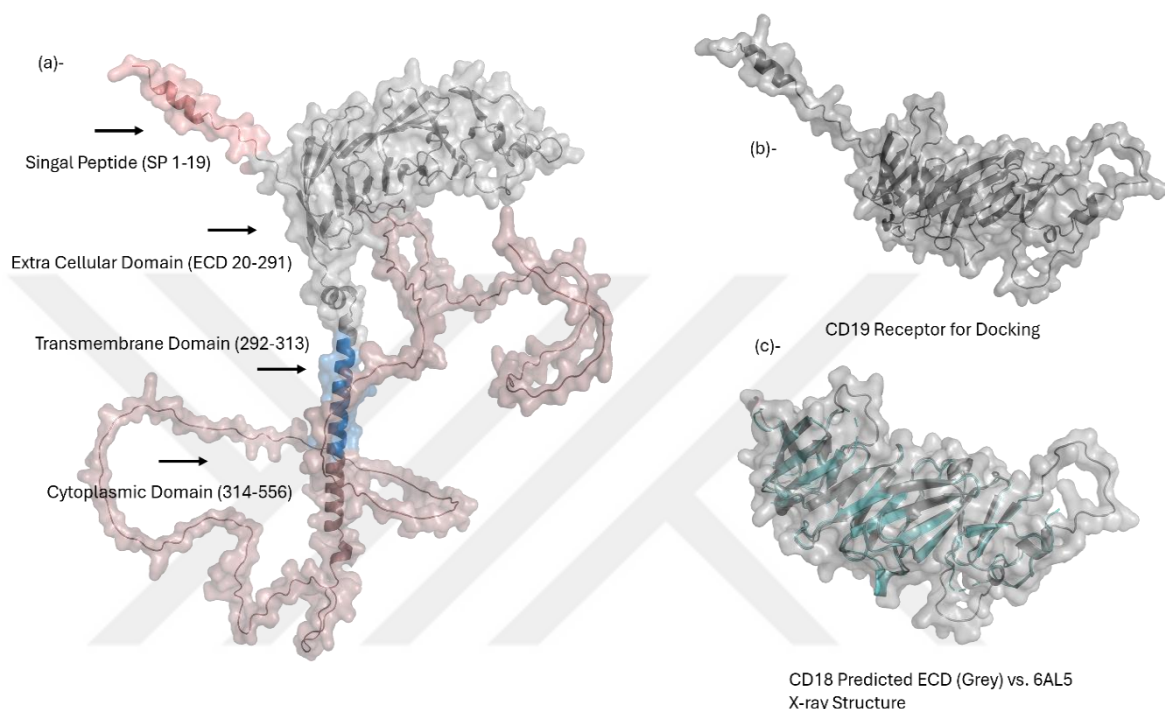


Figure 4: CD19 3D structure prediction and annotation. (a) Complete CD19 sequence 3D structure prediction using ColabFold with annotated structure (signal peptide in salmon, extra cellular domain in grey, transmembrane domain in blue, and cytoplasmic domain in dirty violet). (b) ECD regions of CD19, SP+ECD were used for protein-protein interaction prediction using HDOCK. (c) CD19 predicted ECD aligned to 6AL5 X-ray diffraction resolved CD19 structure; not complete. Structures were edited in Pymol.

The ColabFold predicted structure was used instead of the available structures in the Protein Data Bank as the X-ray modelled structures of CD19 were missing reasons and were not complete. The prediction using ColabFold was of high accuracy. Thus, it was elected not to use the available PDB structures of CD19 as it would require the extra steps of fixing them in preparation for the docking step. The steps involved in the preparation of such PDB structures would include adding missing hydrogen atoms, building missing loops, converting

non-standard residues to their standard equivalents, and deleting unwanted heterogens. The renditions might not rise to the quality of the ColabFold prediction.

In a similar fashion, the 3D structure of the three anti-CD19 nanobodies were predicted using default parameter ColabFold (Figure 5). CDR3 displays a non-loop structure in both Nb1, and Nb2 which can be either because of ColabFold or an unusual CDR3 structure (similar structures exist which show the extent of diversity of Nbs and their FR and CDR regions). The same is observed for Nb3 in the CDR1 region.

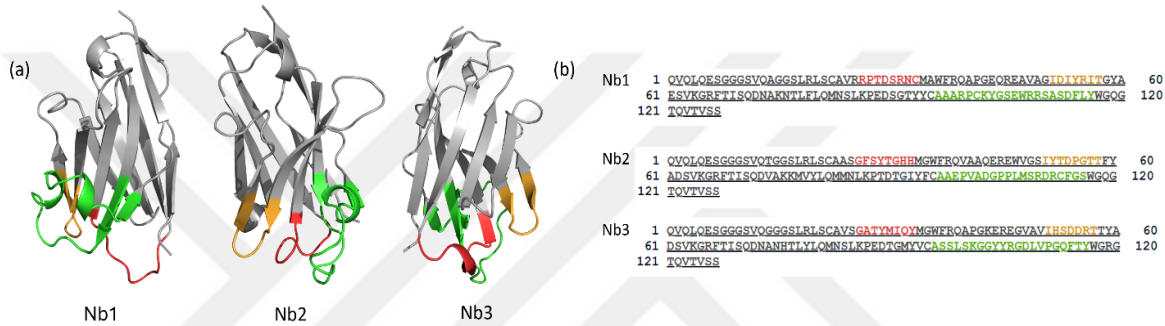


Figure 5: 901, 902, 903 (Nb1, 2, 3) 3D structure prediction and annotation. (a) Nanobody numbering and CDR delimiting based on IMGT scheme using AbRSA. CDR regions are annotated on the homology modelled (ColabFold) structure of Nb1-3 to match CDR1 (red), CDR2, (yellow), CDR3 (green). Structures were edited in Pymol.

The CDRs of each nanobody were numbered using IMGT scheme using AbRSA. The residues of CDRs with the surrounding residues of FRs for each nanobody would be used as parameters for the rigid-body docking to CD19 in the targeted runs.

4.3 Rigid-body Docking for the determination of protein-protein interactions:

HDOCK combines ab initio free docking with template-based modelling in a hybrid algorithm. One argument to the use of physics-based modelling solutions to predict protein-protein interactions is that it can be more reliable when compared with deep learning models for predicting interactions of complexes that are more diverse or unseen (Buttenschoen, Morris, and Deane 2024). The introduction of physical and chemical restraints by physics-based modelling solutions provides more plausibility. In all HDOCK dockings runs, both the

receptor (CD19) and the ligands (anti-CD19 nanobodies) were input in their ColabFold top-scoring PDB structure. HDOCK generates 2000 complexes and ranks the top 10 performing models for each docking prediction experiment. We have utilized an in-house script to further filter the generated 2000 complexes. The script allows for the filtration based on “paratopic enforcement” which favors interactions that are concurrent with available PDB structures of Nb-Target interactions. Mainly, predicted interaction interfaces between CD19 and anti-CD19 nanobodies are chosen to leverage interactions more concentrated within and close to the CDRs of nanobodies that are closer to a perpendicular homing of the nanobody to the structure. This step becomes more pronounced due to the lack of available structures of docked nanobodies to CD19 and helps with the clustering and identification of poses and binding patches on CD19 that are more structurally viable. A preliminary blind docking run which had no parameters for residue constraints was conducted. The blind docking run was then followed by one that tested for a glycosylated CD19 receptor.

The choice for the N-linked glycan trees were based on reported literature of glycosylation found in cancer (Pont et al. 2021; Reily et al. 2019; Ritchie et al. 2004). Glycan trees exhibit altered branching in disease (i.e., ALL) and fucosylation. Information on glycan trees were used to build N-linked glycosylation on the following residues on CD19 (86, 125, 138, 181, 265). The main reason for modelling glycans stemmed from the nature of the prediction of CD19 and anti-CD19 nanobodies where information was lacking, allowing for more probed discovery of possible binding behaviors. This constitutes a further step in substantiating the choice for parameterization and has not been used entirely in the main logic of our pipeline. Anti-CD19 nanobodies preferred to be docked in a similar manner when compared between the glycosylated and non-glycosylated CD19 structure in similar fashion (Figure 5). Moreover, from a look at the glycosylated state of CD19, it seems that not many options are viable for docking due to possible interference with glycan structures (Figure 6).

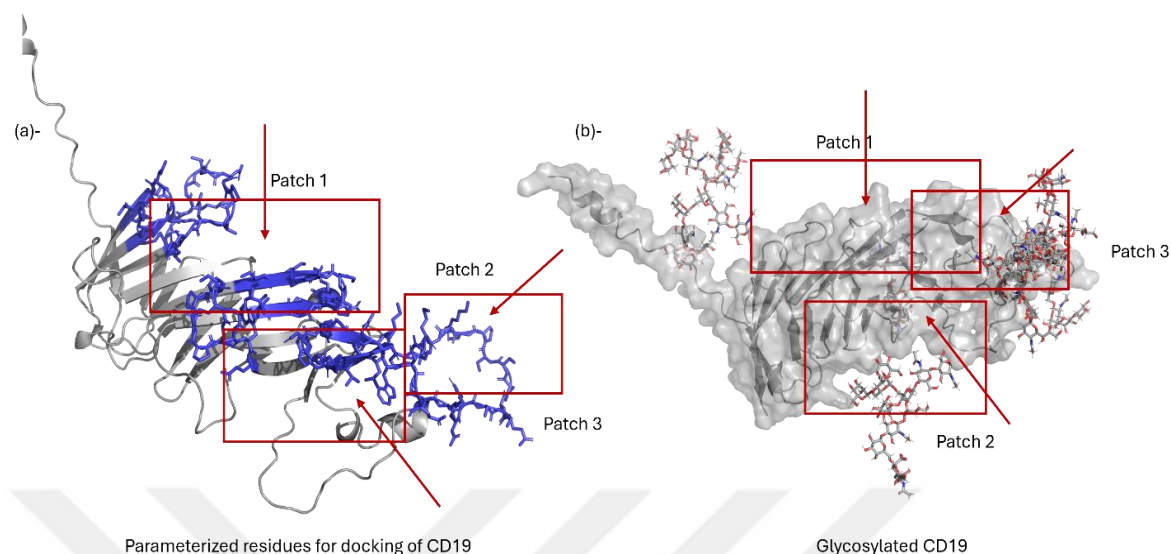


Figure 6: Proposed binding patches as per blind docking runs. (a) Non-glycosylated CD19, residues are highlighted in blue. The residues have been reported in the literature to interact with proteins/antibodies. Moreover, the patches are concurrent with the state of glycosylation of CD19 in (b). Figures were generated in Pymol; CD19 is in Grey, constrains for docking were chosen based on residues highlighted in blue.

Moreover, viable docking poses chosen did follow on the constraints of the glycosylated state of CD19 as elucidated in (Figure 7). The most viable binding patch for both Nb1, and Nb2 were chosen to follow on the binding patch 1 specified in the residue constraints for protein-protein interaction prediction. Patch 2 was more viable for Nb3 based on the recurrence of the prediction. Patch 3 had almost no viable predictions to necessitate its consideration. The introduced parameters for CD19 are visualized in (Figure 6) along with annotation and comparison with the glycosylated structure built by Glycam (<https://glycam.org/>).

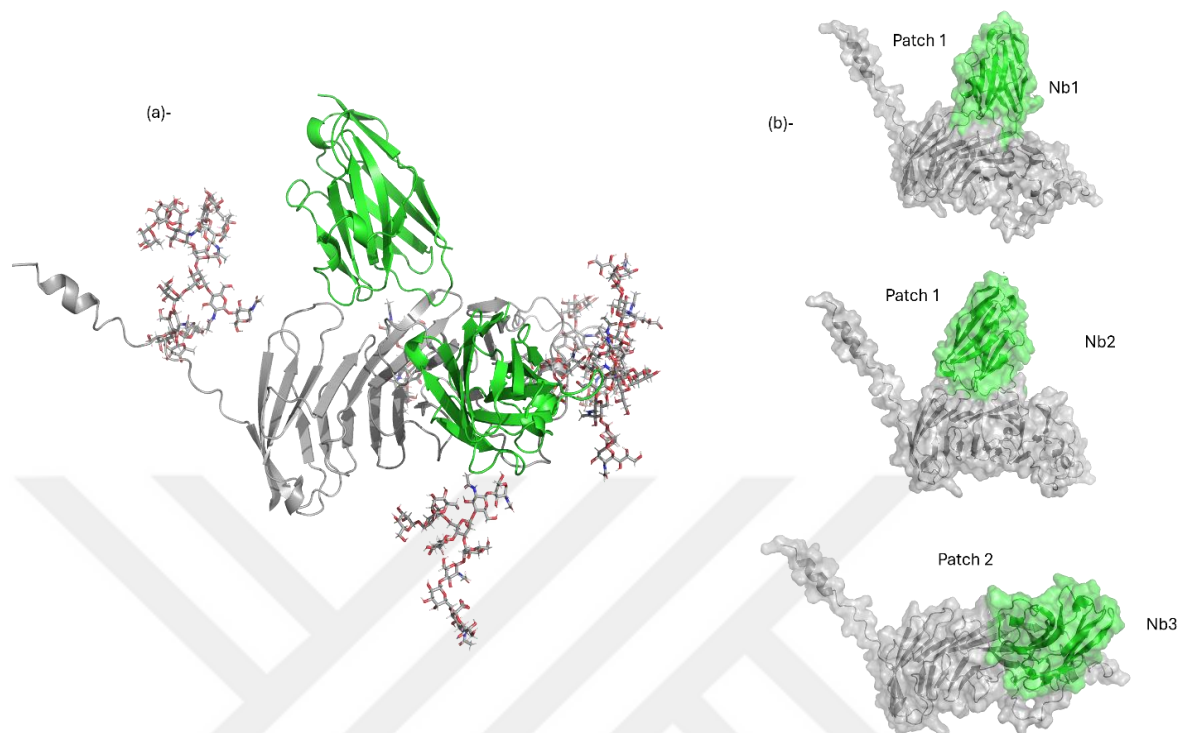


Figure 7: Binding pose for each of Nb1, Nb2, Nb3 to CD19, and superimposition of the binding states to the glycosylated form of CD19. (a) superimposition of binding states to the glycosylated form of CD19. (b) The most viable binding patch for both Nb1, and Nb2 (patch 1), and Nb3 (patch 2). Figures were generated in Pymol; CD19 is in Grey, nanobodies in TV green.

For the final docking run, the generated data of the previous runs were coupled with reported interactions of CD19 with other proteins/antibodies in the literature (Bradbury et al. 1992; Matsumoto et al. 1991), and available PDB structures (e.g., 7URV, 7JIC). The available PDB structures mainly constituted antibody interactions and other receptor interactions of CD19 which were part of the parameterization process. The residues constraints introduced for the docking run for the anti-CD19 nanobodies involved residues of their CDRs and neighboring residues of the FRs. For CD19, the residues chosen as to constrain the run involved the following patches of residues listed in (Table 2).

Table 2: Residue constraints introduced for CD19. Patches responding to the constraints are listed.

CD19 Receptor Patches	Residues for each Patch
1	39-50
	88-90
	215-225
	261-268
2	98-108
	261-268
	215-225
3	151-170
	136-150

Scoring the predicted complexes with both Blue2 and PRODIGY was applied at each step to gain more understanding on the general behavior of the predicted docking interfaces and docking poses. This is done after preparing the predicted complexes for scoring which is done by PDB2PQR. This constitutes a necessary step to introducing energy minimization to the structures using the CHARMM forcefield. The structures were analyzed for structural relevance to available crystal structures of nanobodies to antigens. Top ranked scoring structures that favor a perpendicular binding patch on CD19 were chosen for the rational design and affinity maturation of anti-CD19 nanobodies.

4.4 Rational Design & Affinity Maturation of Anti-CD19 nanobodies:

There has not been one consensus on the design of *in-silico* affinity matured nanobodies. After a thorough review of the related literature (Chaudhuri et al. 2022; Cheng et al. 2019a; Hacisuleyman and Erman 2019; Hu et al. 2020; Muyldermans 2021b; Vivcharuk et al. 2017; Yau et al. 2005), paradigms combining the latest reported methods and best practices have been chosen. For instance, the introduction of single-point mutations to the CDRs of the nanobody followed by the generation of double to triple mutants combining the top-scoring single-mutants is one method followed in the design of affinity matured nanobodies. This constitutes an iterative process whereby 18 different amino acids are

considered apart from cysteine and proline to replace chosen residues. Such method is not efficient as it requires more computation and is not mature in the way it considers the structure of the nanobodies where CDRs consist mainly of loop forming residues. Another method that is simple but can be consider in the rational design of nanobody is generating nanobody hybrids based on the sequence consensus of the top performing Nbs (McMahon et al. 2018).

Not all CDRs are necessarily mutated as CDR3 gets more attention compared to CDR1, and CDR2. Mutation of FR residues has also been implicated in the improvement of affinity and stability of Nbs. Moreover, the nature of CDRs being mostly loops, dictate that for the maturation of a nanobody, the choice of residues could be narrowed down to loop-forming residues or rather residues implicated in disordered structures (Gavenonis et al. 2014; Minucheir and Goliaei 2005). This permitted another look at the design of the rational mutation experiment and the paradigms used for the affinity maturation. Moreover, the nature of CDRs being mostly loops, dictate that for the maturation of a nanobody, the choice of residues could be narrowed down to loop-forming residues or rather residues implicated in disordered structures (Gavenonis et al. 2014; Minucheir and Goliaei 2005). The chosen residues to be mutated were: (W, F, H, D, Y).

Residue interactions at the interface of the predicted candidate CD19-nanobody complexes were analyzed by LigPlot+. The criteria implemented for the choice of the candidate residues for mutations on nanobodies were as follows. First, residues chosen should be part of the interface residues of the highest scoring filtered poses. The candidate residues were chosen based on the least contact distance (Cheng et al. 2019b; Poustforoosh et al. 2023b), having most interactions with CD19 residues as the metric of choice. Naturally, residues chosen were part of the CDRs of the nanobodies and emphasis was applied on most of the residues being part of CDR3. This part in the pipeline constitutes a more nuanced approach than that reported in previous literature. It has been reported that the selection of candidate residues for rational mutation did not involve an espoused criteria (Hacisuleyman and Erman 2020).

Five residues for each anti-CD19 nanobodies were chosen for mutation as follows: Nb1 (Y103, R31, T28, Y54, K102); Nb2 (P106, G104, D54, T30, D62), Nb3 (Q32, R56,

D54, G102, Y29). Taguchi design, or design of experiment (DOE) approach was implemented as a statistical model for the choice of the best performing model based on the introduced mutations were Blueue2 scoring was used for ranking of each experiment of the implemented L25 orthogonal matrix. Following redocking and analysis of the generated structures; 25 experiments for each nanobody, candidate residues were implemented into the design of multiple matured nanobodies to be tested and ranked accordingly. This constituted a double step in the validation of the statistical design. The result of the experiment is elucidated in (Figure 7, 8) and (Table 3, 4). The result of the analysis of Nb3 is omitted due to the low performance of all Nb3 mutants in both Taguchi recommended mutations and the reiterative rational design cycles.

MINITAB was used to conduct signal to noise ratio analysis of Taguchi design of experiment for all nanobodies based on pose-specific scoring of the 25 experiments using Blueues2 (Supplementary Figure T) under the supplementary section of “Rational Design & Affinity Maturation of Anti-CD19 nanobodies”. For Nb1, histidine performed the best for residue 103 (Y103H), tryptophan for residue 31 (R31W), tryptophan at residue 28 (T28W), and aspartic acid for residue 102 (K102D) (Figure 8). Tyrosine remained the best residue at position 54. Nb1_T (Y103H_R31W_T28W_K102D) performed the best out of all tested Nb1 mutants (Table 3, 4). The response table ranks mutations based on significance where the introduction of histidine at position 103 seems to have the most effect on the Nb1 followed by position 28 and position 102. The response table highlights one key aspect when affinity maturing nanobodies, which is to apply more attention to CDR3 in comparison to CDR1 and CDR2.

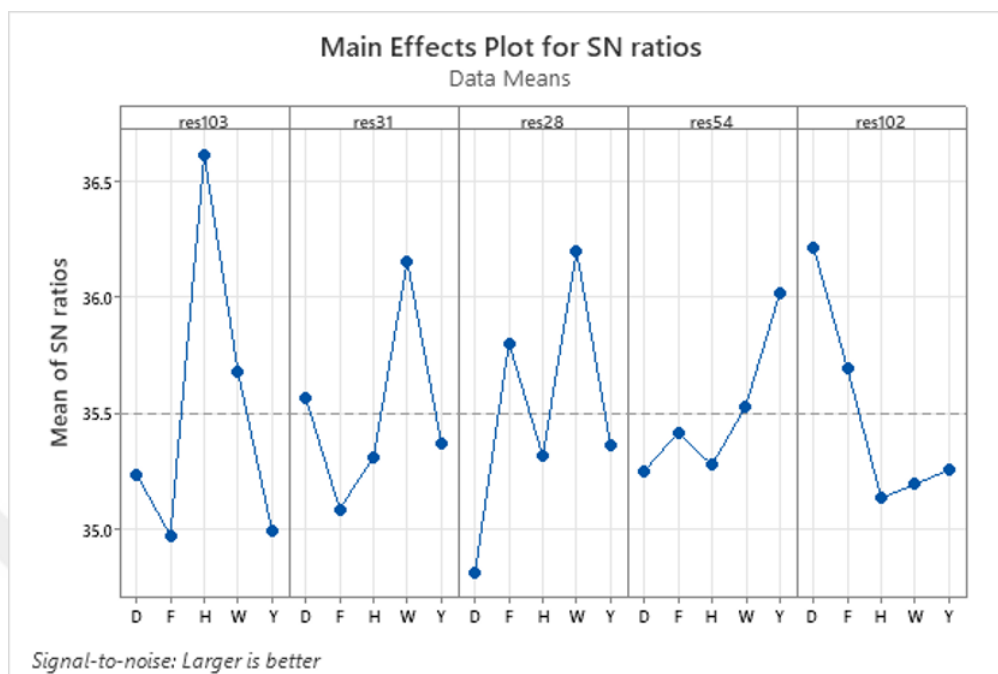


Figure 8: Nb1 Taguchi analysis results: Main effects plot for SN ratios of scores of each residue to each position. Best residues are as follows: Y103H, R31W, T28W, Y54Y, K102D.

Table 3: Response tables for SN ratios ranking mutations based on significance. The results were analyzed by MINITAB.

Response Table for Signal to Noise Ratios

Larger is better

Level	res103	res31	res28	res54	res102
1	35.24	35.57	34.81	35.25	36.22
2	34.97	35.09	35.80	35.42	35.69
3	36.61	35.31	35.31	35.28	35.13
4	35.68	36.16	36.20	35.53	35.19
5	34.99	35.37	35.36	36.02	35.25
Delta	1.64	1.07	1.39	0.77	1.08
Rank	1	4	2	5	3

For Nb2, aspartic acid seems to have been the name of the game as it performed the best out of all 25 experiments at the five residue positions chosen (Figure 9). However, two of the five residue positions already had aspartic acid (D62D, D54D). Nb2_T (P106D_G104D_T30D) performed the best out of all tested Nb1 mutants along with another mutant that was part of the reiterative design cycle reported later. The introduction of

mutations at residues 104 and 106 held the most significance which re-emphasizes the importance of CDR3 in the process of affinity maturation of nanobodies.

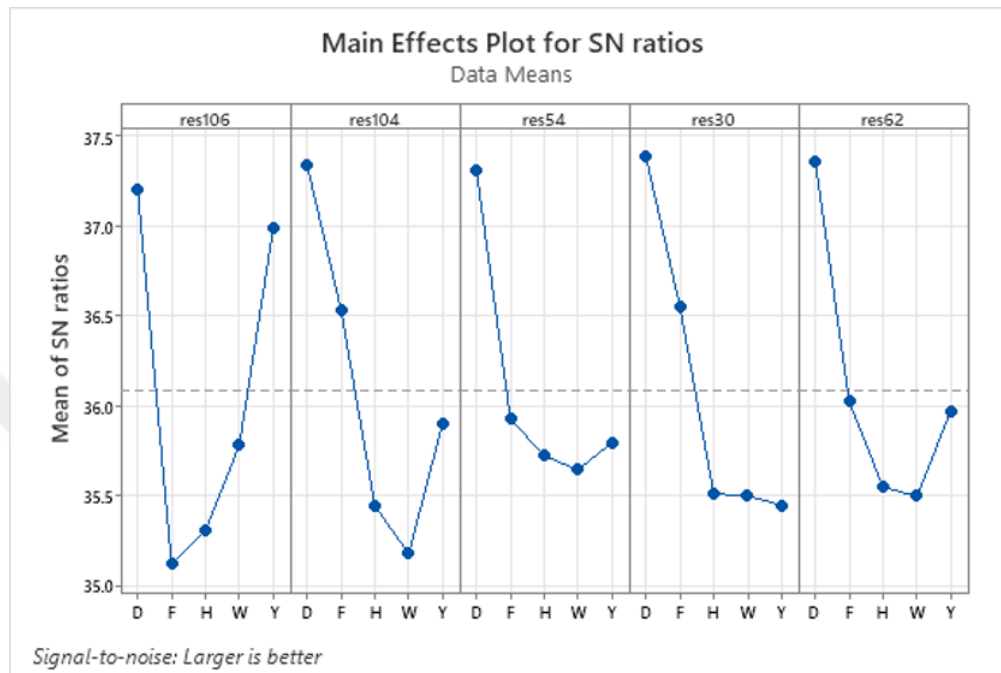


Figure 9: Nb2 Taguchi analysis results: (a) Main effects plot for SN ratios of scores of each residue to each position. Best residues are as follows: P106D, G104D, D54D, T30D, D62D.

Table 4: Response tables for SN ratios ranking mutations based on significance. The results were analyzed by MINITAB.

Response Table for Signal to Noise Ratios

Larger is better

Level	res106	res104	res54	res30	res62
1	37.20	37.34	37.31	37.38	37.36
2	35.12	36.53	35.93	36.55	36.02
3	35.31	35.44	35.72	35.51	35.55
4	35.78	35.18	35.64	35.50	35.50
5	36.99	35.91	35.80	35.45	35.97
Delta	2.08	2.16	1.66	1.94	1.86
Rank	2	1	5	3	4

The same five residues were mutated in the following fashion for serendipity and to explore the space of possible designs starting from introducing mutations to one residue for each nanobody and up to 5 residues incrementally. Nb2_S or Nb2_rational (T30H, D54F,

D62W, G104Y, P106Y) was one of the mutants of Nb2 that performed well. The sequence of Nb2_S constitutes a serendipitous aspect of the design process. Compared with the Taguchi design of Nb2 and the analysis of SN ratios, the following is observed: histidine at position 30 scores the third amongst candidate mutations, phenylalanine at position 54 scores the second, tryptophan at position 62 scores the least, tyrosine at position 104 scores third, and tyrosine at position 106 scores the second. This combination of introduced mutations to Nb2 has performed better than that recommended by the SN analysis in the molecular dynamics' simulations (Table 5).

Furthermore, a consensus-based sequence design was implemented for Nb1 using the CDR3 of Nb3 and the general FR of Nb2 after the generation of multiple sequences. The sequence of the consensus-based sequence design nanobody was termed (904 or Nb4) and yielded more favorable physiochemical properties in comparison with Nb1 (Figure 10, Table 5). Compared with an instability index (II) of 52.07 for Nb1, Nb4 scored 37.03 which are levels comparable with that of Nb2 and Nb3. One observation when comparing between the structures of Nb1, Nb3, and Nb4 is the more representative full loop structures in Nb4 when the 3D structure is predicted with ColabFold. Nb4 performed better than its parent nanobodies (Nb1, Nb2, and Nb3) in both Blues2 scores and the binding free energy calculations of the molecular dynamics' simulation runs.

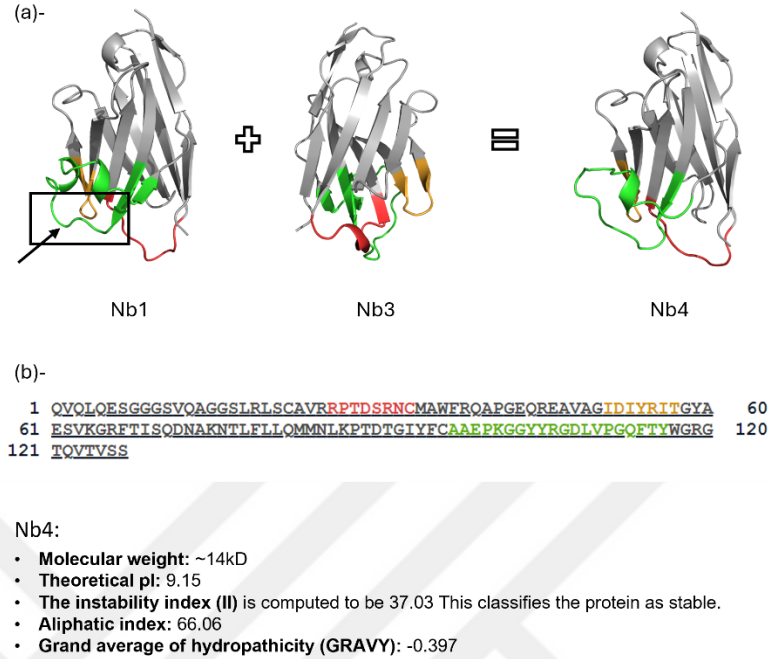


Figure 10: Consensus-based sequence design of Nb4. (a) Nb1, Nb3, and Nb4 visualized. (b) CDRs are annotated based on the homology modelled (ColabFold) structure of Nb1 to match CDR1 (red), CDR2, (yellow), CDR3 (green). (b) ProtParam sequence analysis results: improvement from Nb1 gaining stability; 52.07 to 37.03 which marks the nanobody sequence as stable. This was done by replacing 80-100 with that of Nb2 (minimal change in FR3), and 100-120 with that of Nb3 (CDR3).

In total, three paradigms were tested for the rational mutation and affinity maturation of the nanobodies. Moreover, multiple instances were considered for the creation of a more streamlined process accounting for both the design and the compute time: (criteria of candidate selection for mutation, residues used for mutation, statistical design).

Compared with the original Nbs, most of the matured Nbs overperformed, with Nb3 mutants performing the least. Throughout the in-silico maturation process all maturation techniques yielded positive results be it through statistical means, serendipity, or proven methods like consensus-based design in the case of Nb4 (Table 5). The sequences of the matured nanobodies are provided in (Supplementary Table 1).

Table 5: Nanobodies (original, and mutants) and their docking score of both Blues2 and PRODIGY. Not all data is presented and only ones significant to the process are reported in this table.

Nb	Blues2 score (kcal/mol)	PRODIGY score (dissociation constant (M))
Nb1	~ -67.54	-3.6e-07
Nb2	~ -60.70	-3.6e-06
Nb3	~ -48.13	-9.1e-08
Nb4	~ -73.4	-8.4e-07
Nb1_Taguchi	~ -55.23	N/A
Nb2_Taguchi	~ -149	-1.7e-09
Nb2_Rational	~ -45.67	-5.3e-10
Nb3_Taguchi	~ -50.73	-9.3e-08

All the nanobody complexes in (Table 2) as well as other matured nanobody complexes were further analyzed by molecular dynamics simulations. This served in both boosting the candidate selection and validate the scoring methods utilized.

4.5 Molecular Dynamics Simulations and Trajectory Analysis for Validation and Candidate Selection of Affinity Matured Anti-CD19 Nanobodies:

The binding affinity of nanobodies has been predicted using molecular dynamics (MD) simulations and used in the optimization of their features (Miguel A. Soler et al. 2018). However, MD simulations are computationally expensive and thus could not be part of the design process in our case and were only used to validate top matured candidates and compare them to the performance of the original three anti-CD19 nanobodies. The complexes of the nanobodies were used as the starting point for 50-150 ns MD simulations after preparing them for the MD run. GROMACS (Abraham et al. 2015) was used to carry out the molecular simulations. The atomic interactions within each complex were parametrized using OPLS-AA force field. The complexes were solvated in water using SPCE model. The complexes were neutralized at a salt concentration of 0.15 M where Na^+ and Cl^- ions are added to the system. Following energy minimization of the generated system for each complex, the

molecular dynamics simulation was carried out based on the following parameters used for equilibration of the complexes: target temperature of 300K using the Langvin thermostat, and pressure at 1 atm. MM/PBSA (E. Wang et al. 2019) was used for the calculation of binding free energy. All the matured Nbs performed better than the original anti-CD19 nanobodies except Nb3 where work on it was discontinued.

The binding free energy scores were taken from the end of each MD run. 150 ns MD runs were carried out to check if any structural imbalance could occur in the Nb-CD19 complex (Figure 11) and to validate the selection of candidate matured nanobodies. The choice to carry out MD instead of counting on scoring approximation aided in the selection of the candidate matured Nbs and validated some of our choices as other matured candidates did not perform as per their initial scoring using either Blues2 or PRODIGY.

All the matured Nbs performed better than the original anti-CD19 nanobodies except Nb3 where work on it was discontinued (Table 6). The binding free energy scores were taken from the end of each MD run.

Table 6: Binding free energy score at the end of each MD run calculated in KJ/mol.

Nb	50 ns (KJ/mol)	150 ns (KJ/mol)
Nb1	~ -155.1	N/A
Nb2	~ -215.2	N/A
Nb4	~ -349	~ -360.6
Nb1_Taguchi	~ -285	~ -414
Nb2_Taguchi	~ -259.8	~ -360
Nb2_Rational	~ -283.4	~ -392.1

Furthermore, even with the presence of the SP of CD19, the matured Nbs still overperformed as the SP of CD19 was reported to result in lower binding to structurally-specific antibodies as compared when it is not present (Laurent et al. 2021). As per (Figure 11), the interaction complexes arrived at a quick equilibrium indicating the stability of the original predicted pose.

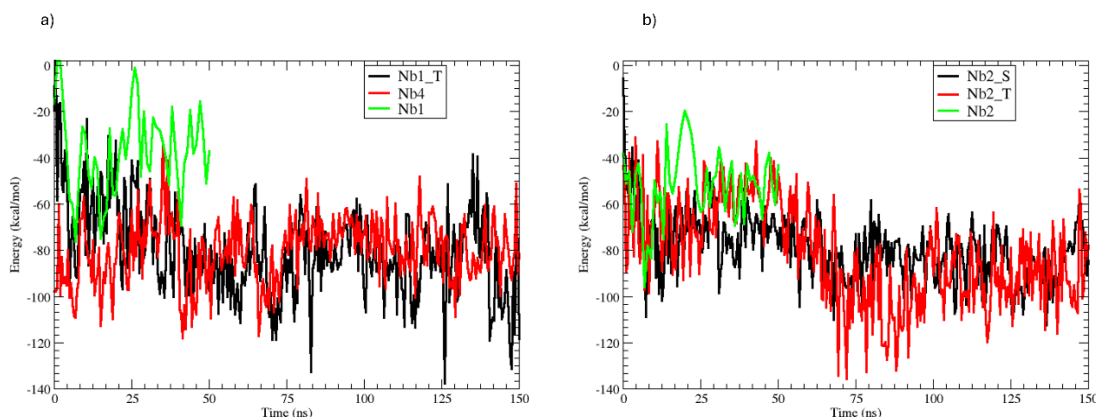


Figure 11: Calculated binding free energy of matured anti-CD19 nanobodies vs. original anti-CD19 nanobodies. (a) Calculated binding free energy of Nb1 (green) vs. Nb4 (red) and Nb1_T (black). Nb1 was run for 50 ns while Nb4 and Nb1_T are run for 150 ns. (b) Calculated binding free energy of Nb2 (green) vs. Nb2_T (red) and Nb2_S (black). Nb2 was run for 50 ns while Nb2_T and Nb2_S are run for 150 ns. All mutants reach a steep binding free energy minima performing better in the first 50 ns. Energy calculations were done through mm/PBSA and the graphs were plotted using XMGrace.

The binding free energy calculations were measured in Kcal/mol and plotted against time in nanoseconds. The reported matured anti-CD19 nanobodies in this thesis performed better than their parent nanobodies. The binding affinity calculated with mm/PBSA shows an improvement that is around twofold which permits further characterization of the matured nanobodies *in vitro*. When comparing the binding free energy calculations of the MD analysis of candidate complexes against their original counterparts, Blues2 scoring method manages to be similar in its prediction to the following complexes: Nb1-CD19, Nb2-CD19 and Nb2_T-CD19. The other candidate mutant complexes' Blues2 score prediction does not fit with that of their molecular dynamics' simulations. This emphasizes the need for the validation of initial predictions of candidate complexes as preliminary scoring only accounts for one snippet of the interactions. MDs allow to model the candidate complex and permit interactions to take place as a function of time in a simulated physical environment which gives an almost real moving picture instead of a still one.

5. *In-vitro* Overexpression, Purification & Functional Analysis of anti-CD19 Nanobodies

5.1 Anti-CD19 Nanobody Plasmids, and Construct Design:

Parent (Nb1, Nb2) and mutant (Nb1_T, Nb2_T, Nb2_S, Nb4) anti-CD19 nanobodies' sequences were built into a pET-22b(+) expression. The nanobody constructs were designed to contain an N-terminal periplasmic translocation signal peptide (Pectate lyase B: PelB), followed by a Recombinant Tobacco Etch Virus (TEV)-protease site, and a C-terminal 6X histidine-tag. PelB leader sequence of *Erwinia carotovora* CE is a 22 amino acid long (UniProt: Q04085; MKYLLPTAAAGLLLLAAQPAMA) signal peptide that is cleaved off by a signal peptidase once the constructs are trafficked into the bacterial periplasm (Lei et al. 1987). PelB utilizes the Sec-dependent pathway for the post-translational translocation of the constructs. The expressed constructs would translocate through the membrane of the cytoplasm in the unfolded state unlike translocation by the twin-arginine (Tat)-mediated pathway where the transported to the periplasmic space in its native form after folding in the cytoplasm (Thie et al. 2008). The folding state of each construct's translocation would have consequences on the recombinant overexpression in *E. coli* (DE3) which is discussed in its relevant sections of the thesis. The TEV-protease site (ENLYFQG) was introduced following the PelB signal. The addition of the TEV-protease site can be viewed as an oversight. The choice for the addition of TEV-protease site is to ensure that the constructs that would otherwise aggregate in the cytoplasm and can only be purified from their inclusion bodies are tested for functionality and analyzed without the presence of PelB attached to their N-terminal which can be a major hindrance. A 6X his-tag (HHHHHH) was added to the C-terminal of the construct to allow for the purification of the anti-CD19 nanobodies with Ni-NTA affinity chromatography/ Ni-NTA beads. A two amino-acid peptide (LE) connected the construct to the 6X his-tag. The designed constructs are illustrated in (Figure 12) where their designed sequences were introduced between the restriction cutting sites of NcoI and XhoI. The pET-22b(+) vector contains an ampicillin-resistance gene (AmpR) which would be utilized for the selection of transformed bacteria.

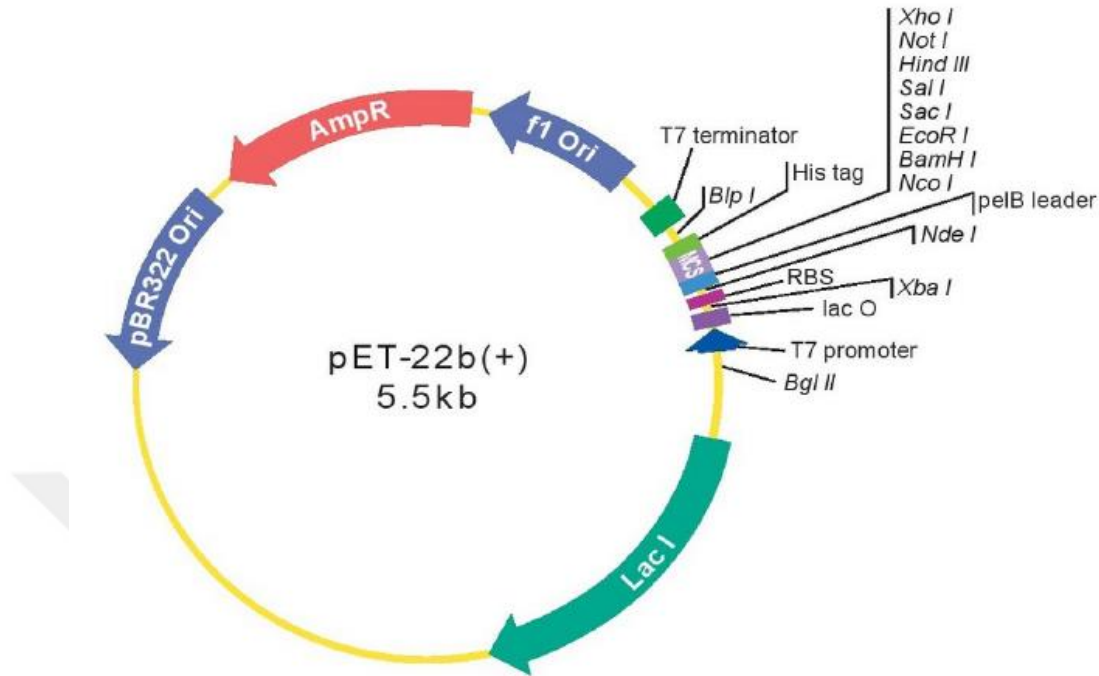


Figure 12: pET-22b(+) construct for recombinant overexpression of anti-CD19 nanobodies. The designed constructs were flanked by an N-terminal PelB signal peptide, followed by a TEV-protease site, and a C-terminal 6X his-tag. The constructs were ordered from GenScript and were cloned between the restriction cutting sites of NcoI and XhoI.

The N-terminal periplasmic translocation signal PelB would facilitate the proper disulfide bridge formation between the conserved cysteine residues of the nanobody in the oxidative environment of the periplasm as opposed to the reducing environment of the cytoplasm of *Escherichia coli* which would lead to the proper folding of the nanobodies (Nagy-Fazekas et al. 2023b). The lack of presence of a periplasmic translocation signal would result in the aggregation of expressed nanobodies in the reducing cytoplasm of conventional BL21 (DE3) which will be used in the recombinant overexpression of the anti-CD19 constructs. This would coincide with the need to apply complex purification and solubilization protocols where the native tertiary structure of the nanobody might not be preserved as opposed to it naturally occurring. Moreover, the PelB periplasmic translocation sequence has been reported to improve the solubility and therefore the yield of recombinant nanobodies expressed in *E. coli* (Chao et al. 2022).

Nanobody-pET-22b(+) expression vectors were ordered through GenScript and codon-optimized through their proprietary software (GenSmart) for better expression in BL21 (DE3) *E. coli*. A certificate of analysis from GenScript containing both the validation of Nanobody-pET-22b(+) vector and DNA sequencing of each designed construct cloned into the vector were provided by GenScript.

5.2 Recombinant Production and Purification of Nanobodies:

Herein, the optimized protocol for the recombinant expression and purification of anti-CD19 nanobodies is reported. Optimization phases with their respective results and discussion are reported in the supplementary section under “Optimization Steps for Overexpression and Purification of nanobodies” where Nb2 was chosen to optimize for the recombinant production and purification of nanobodies in their soluble form (Supplementary Figure 1-9). The reason for choosing Nb2 is twofold. First, Nb2 is based on one of the original sequences reported in the literature having been produced as a part of CAR T-cell construct. Concurrently, Nb2 sequence was used in the design of rationally designed and affinity matured anti-CD19 nanobodies that could be more unpredictable to handle *in vitro*. Following the optimization of the protocol for the recombinant production and purification of the anti-CD19 nanobodies, the rest of the anti-CD19 nanobodies were tested. Small volumes of induced bacterial BL21 (DE3) cultures in the range of 50-250 mL yielded low purity/solubility expressed anti-CD19 nanobodies. This was reflected in FPLC affinity chromatograms as the purification of the anti-CD19 nanobodies did not retain viable peaks in the elution step. Moreover, pH was one of the important factors in the purification of the Nbs as the pH of the buffers used played an important role in the purification of soluble nanobodies. Small deviations from the calculated pI of Nbs led to their aggregation and loss. Both Nb4 and Nb2_T were expressed and purified with the optimized protocol in their soluble form at very low concentration as most of the expression was observed in the insoluble fraction of the pellet (Supplementary Figure 10). It was decided to proceed with Nb1 and its designed mutant Nb1_T and halt the production of the rest of the anti-CD19 nanobodies till the conclusion of their functional analysis. Larger volumes of induced bacteria were pivoted to which resulted in the observation of better eluted peaks in affinity chromatograms and consequentially higher purity purification yields.

Nb1 and Nb1_T were less recalcitrant (disobedient) compared to the other anti-CD19 nanobodies (excluding Nb2). Around 1.6 L of LB broth (~2% ethanol) were inoculated with overnight grown cultures of Nb1 and Nb1_T transformed BL21 (DE3) and induced at OD₆₀₀ of 0.6-0.8. IPTG at 0.005 mM was used to induce the bacteria and the induction was carried out for 18-24 hours at either 25 or 16 °C. The bacteria were later harvested via centrifugation at 5000 rpm at 4 °C for 30 minutes. The supernatant was discarded, and the pellet was washed with Tris buffer and centrifuged again in a similar manner. The resuspended pellet was lysed by sonication at intervals of 30 seconds repeated 8 times. The supernatant constituting the soluble fraction was obtained after centrifugation at 20,000 rpm for one hour and filtered with 0.45 µm filters to get rid of cellular debris and other contaminants like DNA. Both Nb1 and Nb1_T samples were purified as reported in the method yielding similar concentrations of ~7 mg/mL.

A One-step elution (at 100% B concentration; 250 mM imidazole) was used for the purification of Nb1 and Nb1_T as per prior optimization. The results of Ni-NTA affinity chromatography for both Nb1 and Nb1_T with their respective SDS-PAGE are reported in (Figure 13, 14). 15% polyacrylamide SDS-PAGE gel was used for the visualization and validation of the observed Ni-NTA affinity chromatogram of both Nb1 and Nb1_T. The increased percentage of the SDS-PAGE gel used comes due to the fact that the nanobodies are small (at 15 kD) and would allow for better separation and visualization of the bands at low size. The eluted Samples were concentrated and buffer exchanged using a concentration column. PBS-T (0.05% Tween-20) was used for buffer exchange at each nanobody's respective pH (8 for Nb1_T, and 7.4 for Nb1). Both nanobodies precipitated upon concentration. Nb1 was easier to rescue compared to Nb1_T accounting for loss of yield following the buffer exchanging step. Glycerol stocks (25%) were made for each buffer-exchanged sample, aliquoted and stored at -80 °C and -20 °C for analysis.

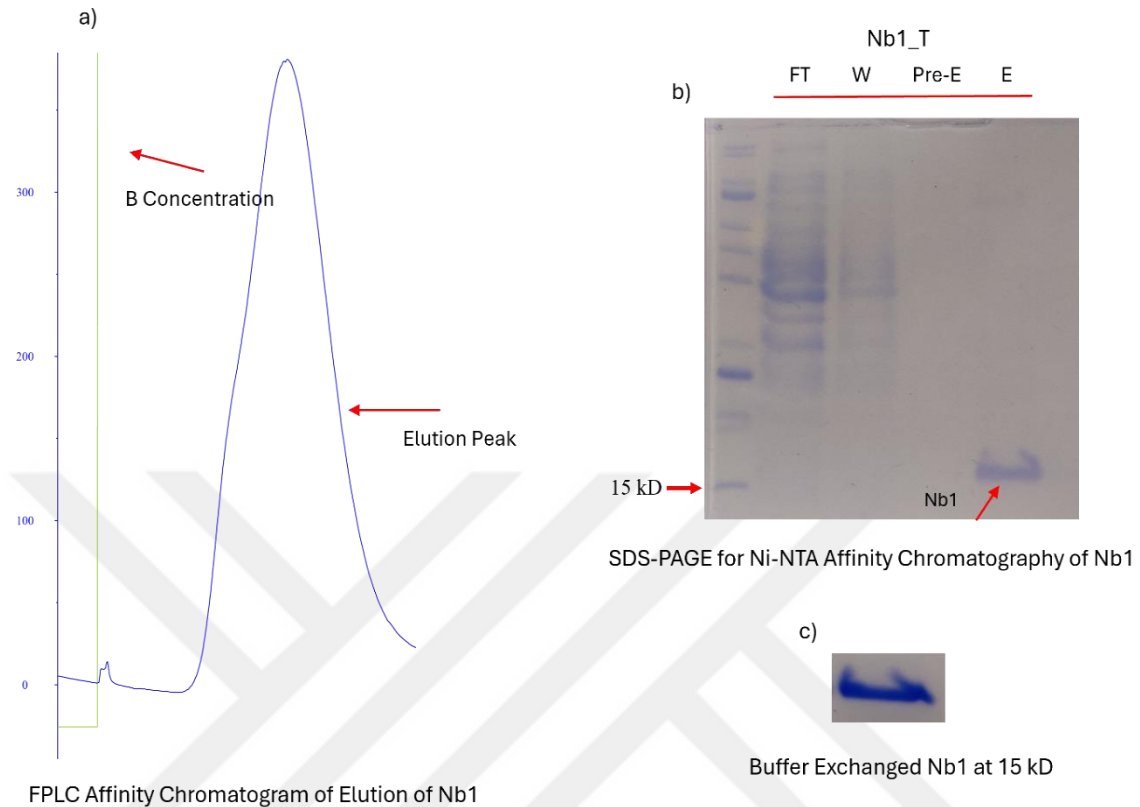


Figure 13: Purification of Nb1 and validation via SDS-PAGE. (a) FPLC affinity chromatogram of the elution step. 100% B (elution buffer) was used for the elution of Nb1. A sharp peak is observed in UV reading following the wash of the Ni-NTA affinity chromatography column. (b) SDS-PAGE of the affinity chromatography purification of Nb1_T. A noticeable band is observed in the elution fraction at 15 kD. (c) Nb1_T band after buffer exchange with PBS-T (0.05% Tween-20).

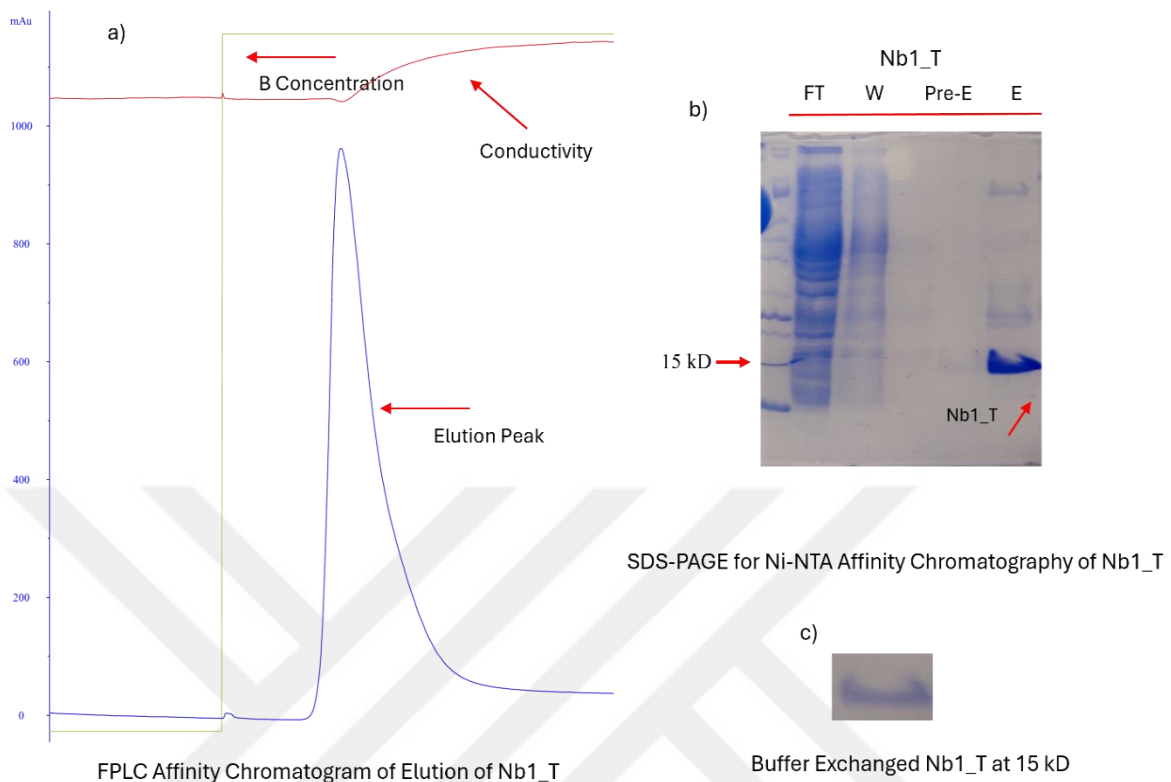


Figure 14: Purification of Nb1_T and validation via SDS-PAGE. (a) FPLC affinity chromatogram of the elution step. 100% B (elution buffer) was used for the elution of Nb1_T. A sharp peak is observed in UV reading following the wash of the Ni-NTA affinity chromatography column. (b) SDS-PAGE of the affinity chromatography purification of Nb1_T. A noticeable band is observed in the elution fraction at 15 kD. (c) Nb1_T band after buffer exchange with PBS-T (0.05% Tween-20).

Compared to Nb2 (Supplementary Figure 5), Nb1 and Nb1_T had been, upon increase of the volume of induced bacteria, better in terms of both the quality of the elution peak and the yield of the purified protein. It should be noted that Nb2 was purified after osmotically treating the bacterial cells and the equilibration of the sample with 2X binding buffer unlike Nb1 and Nb1_T that were directly subjected to lysis in their suspension via sonication. The lysis of the cells instead of a direct extraction of the periplasmic fraction using osmotic treatment could be argued as advantageous. The advantage of lysis over osmotic shock could be, in part, due to the lack of sugar presence in the sample before purification which could lead to interactions with the nanobodies that are unpredictable and unaccounted for. Consequently, the lysis of the cells as a method of protein extraction would

result in introducing impurities to the chromatography column which would decrease the purity at which the nanobodies are obtained. The purity of the obtained purified fractions could be further increased by the employment of follow up purification using either size-exclusion chromatography or ion-exchange chromatography.

5.3 Determination of Binding Affinity of anti-CD19 Nanobodies using Enzyme-linked Immunosorbent Assay (ELISA):

In the first ELISA test for the determination of the binding affinity of anti-CD19 nanobodies, a dose-response curve was plotted based on the absorbances measured for each well at 450 nm. The pattern displayed by both nanobodies was counter intuitive. Absorbance signals increased as the concentration of the applied binders (Nb1, Nb1_T) decreased. This is exemplified by the absorbance values measured against log₂ of applied concentrations of binders (Figure 15) to the wells that were coated by CD19 (1 µg, 500 ng, 250 ng, 100 ng, 50 ng, 25 ng per mL) corresponding to log₂ values of (10, 9, 8, 6.5, 5.5, and 4.5 respectively). Moreover, the blanks demonstrated absorbances similar to the one recorded at 1 µg for each nanobody. CD19 coated wells serving as a positive control showed absorbance values at 450 equal to around 0.2 which is greater than the absorbance values observed for both Nb1 and Nb1_T at their lowest administered concentration.

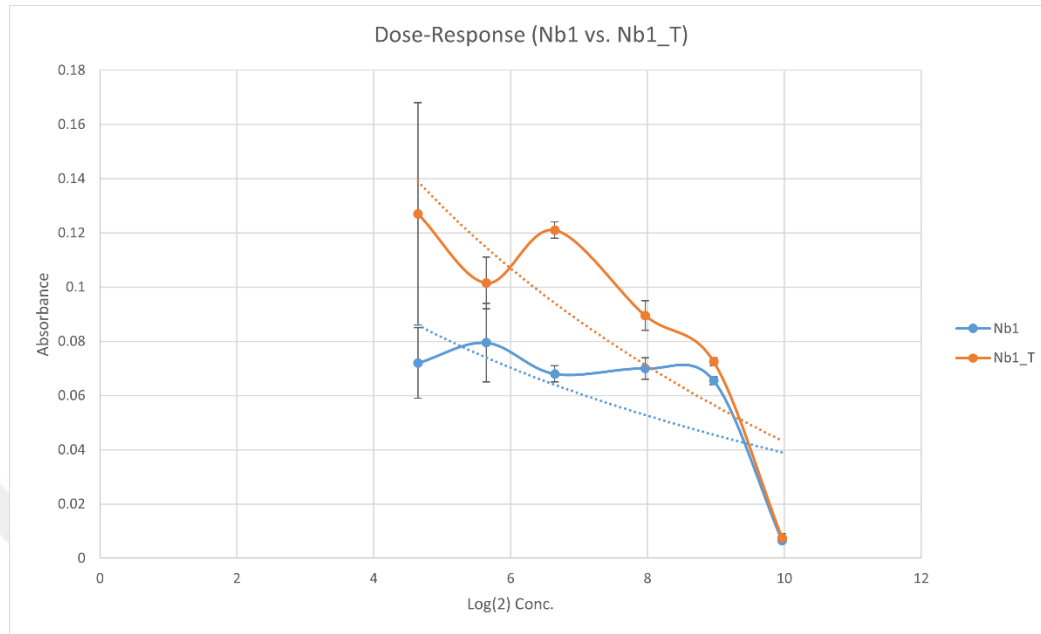


Figure 15: Dose-response curve of Nb1_T and Nb1 ELISA absorbance measurements using a primary histidine antibody and a secondary HRP conjugated secondary antibody. The dose-response plot shows absorbance against log2 concentration of applied binders. Nb1 is in blue, and Nb1_T is in blue. The error bars were calculated based on the standard deviation between measured signals at absorbance of 450.

Recombinant Human CD19-Fc Chimera used in coating the wells contains a histidine tag. One discussion of the dose-response plot is that the primary histidine antibodies might not have bound to the nanobodies and were rather bound to the CD19. This can highlight the absorbance measure for the positive control in contrast with the experimental measurements of Nb1 and Nb1_T. If the nanobodies did not bind properly or were washed away, one would argue for getting similar results to that of the positive control or close to it. One hypothesis based on the result of (Figure 15) is that the binding cannot be measured at the concentrations tested and is observed at lower concentrations starting from around 100 ng/mL. The corrected absorbance measurement against blank wells would exhibit a signal at around 250 ng/mL. For instance, nanobodies might bind differently at varying concentrations due to epitope competitions leading to antigen conformation change that would hinder further interactions. This would result in inaccurate representation of the binding capacity of each of the nanobodies. Steric clashes resulting from possible multivalent binding would lead to steric hinderances as well. In the first ELISA results discussed herein, Nb1_T can be argued to

perform better than Nb1 considering the increased signal at lower applied concentrations. At the same time, the opposite can be argued as the histidine tag is more accessible as less nanobody concentration is applied. The CD19 might not be coating the wells properly is one argument to be considered.

To verify the posited hypothesis based on the first round of ELISA. A second round, similarly, was conducted with the introduction of triplicates instead of duplicates, coating separate wells with only nanobodies, and applying a non-specific binder to CD19. conjugated Goat Anti-Alpaca IgG (Peroxidase AffiniPure™ Goat Anti-Alpaca IgG, VHH domain; Jackson ImmunoResearch) was applied instead which should hypothetically bind to Nb1 and Nb1_T of Bactrian camel origin due to conservation present between the camelid species in their FRs.

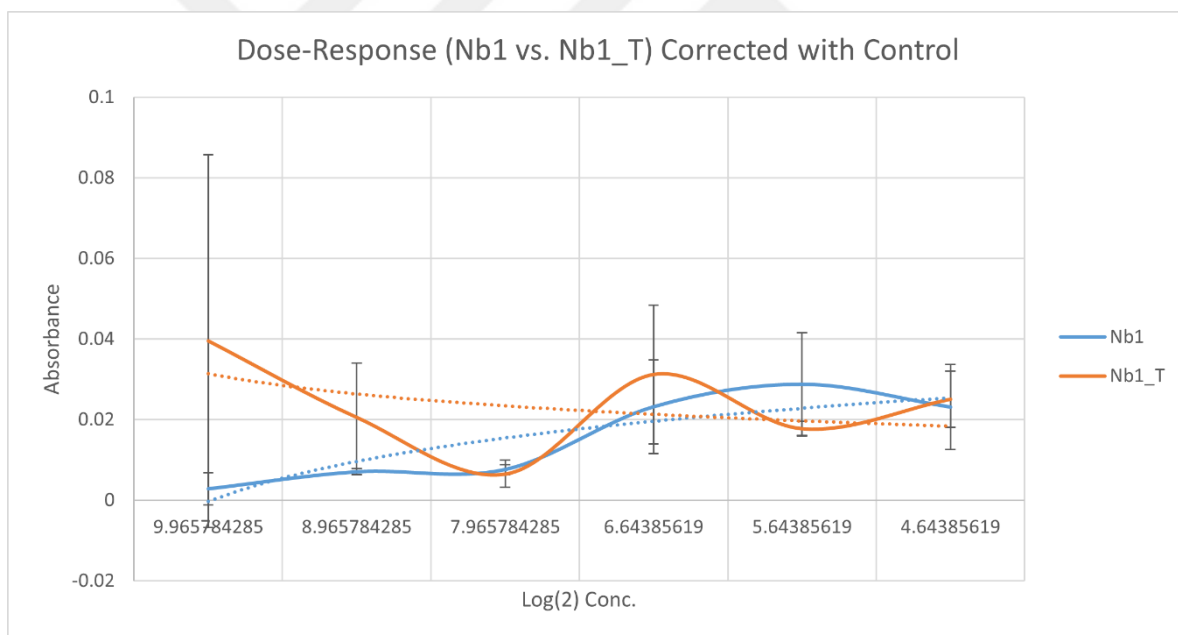


Figure 16: Dose-response curve of Nb1_T and Nb1 ELISA absorbance measurements using Anti-Alpaca IgG HRP conjugated antibody. The dose-response plot shows absorbance against log2 concentration of applied binders. Nb1 is in blue, and Nb1_T is in blue. The error bars were calculated based on the standard deviation between measured signals at absorbance of 450.

The results of the second round of ELISA utilizing an Anti-Alpaca IgG HRP conjugated antibody demonstrate inconclusive results which can be directly attributed to the lack of

binding of the antibody to the nanobodies (Figure 16). An almost negligible increase in the signal is observed for both nanobodies at lower concentrations similar to that of the first round of ELISA. The argument for improper coating could be presented to discuss the second round of ELISA. Nanobody-only coated wells demonstrated similar absorbances to the ones in (Figure 15).

Based on the first and second ELISA runs after the inconclusive run with 100 ng/mL recombinant CD19 coated wells, two directions could be viable. First, the concentration applied to the CD19-coated wells could be diluted serial by factors of 10 to test for better binding at lower concentrations where a sigmoidal curve could be observed; arguing for the concentrations used to display the lower end of the dose-response curve. Another route would be to coat the wells with the nanobodies followed by the applications of recombinant CD19 given the availability of an anti-CD19 antibody instead of an anti-histidine antibody. It stands that the binding affinity could not be predicted based on inconclusive but however intriguing results.

CHAPTER 4

CONCLUSION & REMARKS

In this thesis, a prevalent theme is observed that can be summed with a quotation from (Fraser and Murcko 2024) commentary titled “Structure is beauty, but not always truth”. “A structure is a model, not experimental reality. Representing wiggling and jiggling is hard. *In vitro* can be deceiving. Drugs mingle with many different receptors”.

Three anti-CD19 nanobodies were chosen from available literature for the development of rationally designed and affinity matured nanobodies targeting CD19. Two of the three nanobodies gave mutants that showed promise to be considered for an *in vitro* investigation. Against sequence predictions, the nanobodies were disobedient and were only able to be purified after considerable optimization and consultation of cellular wisdom. While the mutants performed better than their parents in both scores applied and molecular dynamic simulations, their *in vitro* manifestations seem to be elusive and require more apt attention. In this thesis, strides and pitfalls are discussed and many hypotheticals are introduced and some tested with the data allocated throughout the thesis work highlighting the importance of bridging the *in silico*, and the *in vitro*.

CHAPTER 5

SUPPLEMENTARY SECTION

Supplementary Table 1: The sequences of CD19 (UniProt accession: P15391), anti-CD19 nanobodies obtained from (H. Wang et al., 2021) and chosen affinity matured candidates (Nb4, Nb1_T, Nb2_T, and Nb2_S). The sequences of the anti-CD19 nanobodies were code named 901, 902, 903 for Nb1, Nb2, and Nb3, respectively for the in-silico analysis. The S stands for nanobodies that were a product of the rational design using iterative cycles of mutations while T stands for Taguchi as part of the nanobodies generated by statistical analysis of maturation experiments.

Nanobody Code	Nanobody Sequence
CD19	MPPPRLLFFLLFLTPMEVRPEEPLVVKVEEGDNAVLQCLKGTSDGPTQQLTW SRESPLKPFLKLSLGLPGLGIHMRPLAIWLFIFNVSQQMGGFYLCQPGPPSEK AWQPGWTVNVEGSGELFRWNVSDLGGLGCGLKNRSSEGPSSPSGKLMSPK LYVWAKDRPEIWEGEPPCLPPRDSL NQSLSQDLTMAPGSTLWLSCGVPPDS VSRGPLSWTHVHPKGP KSLLSLELKDDRPARDMWVMETGLLLPRATAQDA GKYYCHRG NLTMSFHLEITARPVLWHWLLRTGGWKVSAVTLAYLIFCLCSL VGILHLQRALVLRKRKRMTDPTRRFFKVTPPPGSGPQNQYGNVLSLPTPTS GLGRAQRWAAGLGGTAPSYGNPSSDVQADGALGSRSPPGVGPEEEEEGEGY EEDSEEDSEFYENDSNLGQDQLSQDGSYENPEDEPLGPEDEDSFSNAESY ENEDEELTQP VARTMDFLSPHGSAWDPSREATSLGSQSYEDMRGILYAAPQL RSIRGQPGPNHEEDADSYENMDNPDGPDPAWGGGGRMGTWSTR
Nb1 (901)	QVQLQESGGGSVQAGGSLRLSCAVRRPTDSRNCMAWFRQAPGEQREAVAG IDIYRITGYAESVKGRFTISQDNAKNTLFLQMNSLKPEDSGTY YCAAARPCK YGSEWRRSASDFLYWGQGTQVTVSS
Nb2 (902)	QVQLQESGGGSVQTGGSLRLSCAASGFSYTGHHMGWFRQVAAQEREWVG SIYTDPGTTFYADSVKGRFTISQDVAKKMVY LQMMNLKPTDTGIYFCAAEP VADGPPLMSRDRCFGSWGQGTQVTVSS

Nb3 (903)	QVQLQESGGGSVQGGGSLRLSCAVSGATYMIQYMGWFRQAPGKEREGVAV IHSDDRTTYADSVKGRFTISQDNANHTLYLQMNSLKPEDTGMYVCASSLSK GGYYRGDLVPGQFTYWGRGTQVTVSS
Nb4 (904)	QVQLQESGGGSVQAGGSLRLSCAVRRPTDSRNCMAWFRQAPGEQREAVAG IDIYRITGYAESVKGRFTISQDNAKNTLFLQMMNLKPTDTGIYFCAAEPKG GGYYRGDLVPGQFTYWGRGTQVTVSS
Nb1_T	QVQLQESGGGSVQAGGSLRLSCAVRRPWDSWNCMAWFRQAPGEQREAVA GIDIYRITGYAESVKGRFTISQDNAKNTLFLQMNSLKPEDSGTYTCAAARPC DHGSEWRRSASDFLYWGQGTQVTVSS
Nb2_T	QVQLQESGGGSVQTGGSLRLSCAASGFSYDGHMGMWFRQVAAQEREWVG SIYTDPGTTFYADSVKGRFTISQDVAKKMVYLQMMNLKPTDTGIYFCAAEP VADDPDLMSRDRCFGSWGQGTQVTVSS
Nb2_S	QVQLQESGGGSVQTGGSLRLSCAASGFSYHGHHMGMWFRQVAAQEREWVG SIYTFPGTTFYAWSVKGRFTISQDVAKKMVYLQMMNLKPTDTGIYFCAAEP VADYPYLMSRDRCFGSWGQGTQVTVSS

6.1 Induction of Transformed anti-CD19 BL21 (DE3):

Parameters that have not been reported in the method part regarding the induction of nanobodies involve the supplementation of LB media with MgCl₂, glycerol and the use of Terrific broth instead of LB. Media manipulation did not result in soluble overexpression of nanobodies at parameters that were outside of the optimized method that yielded soluble nanobodies. Further discussion to the overexpression of each nanobody is reported later in this section.

Supplementary Table 2: Overview of common induction parameters of nanobodies (not completely representative but common throughout).

	(Chao et al., 2022)	(Nagy-Fazekas et al., 2023)	(Salema & Fernández, 2013)
Culture Conditions	37 °C until OD600 reached 0.6–0.8	37 °C until OD600 reached 0.8	30 °C until OD600 reached 0.5
IPTG Concentration	0.2 mM	0.5 mM IPTG	0.3 mM IPTG
Induction Temperature	25°C	30 °C	30 °C
Induction RPM	220 rpm	180 rpm	250 rpm
Induction Duration	Intervals, Most retained at 24 hours	Overnight (~12 h)	3 hours

6.2 Rational Design & Affinity Maturation of Anti-CD19 nanobodies:

Implemented design array and scores of complexes following mutation are visualized here in (Supplementary Figure T).

(a)-

Data

Row	res103	res31	res28	res54	res102	score
1	W	W	W	W	W	71.83
2	F	F	F	F	F	58.48
3	H	H	H	H	H	63.22
4	D	D	D	D	D	59.21
5	Y	Y	Y	Y	Y	58.67
6	W	F	H	D	Y	54.92
7	F	H	D	Y	W	53.19
8	H	D	Y	W	F	70.65
9	D	Y	W	F	H	60.05
10	Y	W	F	H	D	68.04
11	W	H	Y	F	D	60.60
12	F	D	W	H	Y	55.83
13	H	Y	F	D	W	62.37
14	D	W	H	Y	F	63.70
15	Y	F	D	W	H	45.82
16	W	D	F	Y	H	60.04
17	F	Y	H	W	D	54.83
18	H	W	D	F	Y	60.43
19	D	F	Y	H	W	47.48
20	Y	H	W	D	F	55.11
21	W	Y	D	H	F	57.83
22	F	W	Y	D	H	58.13
23	H	F	W	Y	D	84.59
24	D	H	F	W	Y	59.90
25	Y	D	H	F	W	55.59

(b)-

Data

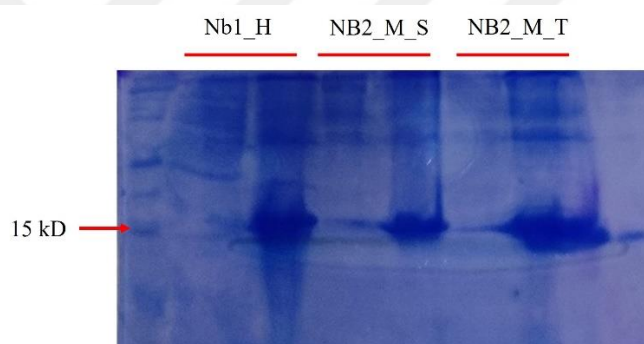
Row	res106	res104	res54	res30	res62	score
1	W	W	W	W	W	52.97
2	F	F	F	F	F	71.11
3	H	H	H	H	H	52.57
4	D	D	D	D	D	148.96
5	Y	Y	Y	Y	Y	70.71
6	W	F	H	D	Y	69.06
7	F	H	D	Y	W	51.33
8	H	D	Y	W	F	58.58
9	D	Y	W	F	H	64.85
10	Y	W	F	H	D	65.66
11	W	H	Y	F	D	73.49
12	F	D	W	H	Y	62.93
13	H	Y	F	D	W	66.17
14	D	W	H	Y	F	62.81
15	Y	F	D	W	H	81.89
16	W	D	F	Y	H	56.67
17	F	Y	H	W	D	53.89
18	H	W	D	F	Y	58.53
19	D	F	Y	H	W	60.02
20	Y	H	W	D	F	66.89
21	W	Y	D	H	F	57.91
22	F	W	Y	D	H	48.73
23	H	F	W	Y	D	56.37
24	D	H	F	W	Y	54.76
25	Y	D	H	F	W	69.38

Supplementary Figure T: Implemented design of experiment (DOE) for Nb1 and Nb2 with Blues2 scores in -KJ readjusted. (a) Nb1 DOE and scores for each of the 25 experiments used in the statistical design of mutants. (b) Nb1 DOE and scores for each of the 25 experiments used in the statistical design of mutants.

6.3 Optimization Steps for Overexpression and Purification of nanobodies:

A protocol that combined knowledge from the induction and purification of nanobodies based on multiple literature articles was tried out (Chao et al. 2022; de Marco 2020; Kariuki and Magez 2021; Nagy-Fazekas et al. 2023b; Pardon et al. 2014; Salema and Fernández 2013). From the literature review, a protocol was first tried out as followed: overnight starter cultures are incubated at 37°C shaking at 200 rpm. 1% or 2% of the overnight culture is inoculated following incubation on the shaker incubator of 16 hours into a certain volume of LB broth (40 mL). The bacterial cells were induced at OD₆₀₀ of 0.6-0.8 with the addition of 0.2 mM IPTG at 25 °C for 24 hours which was reported to result in the greatest yield of periplasmically expressed nanobodies. The cells were harvested at the around the 24-hour mark at 5000 rpm at 4 °C for 30 minutes where the supernatant was discarded. For this trial, 2 mL from each of the 40 mL induced cultures was resuspended in Tris 1 M (pH= 7.4) to check for expressed nanobodies. The sample was centrifuged at 5000 rpm after

sonication 5X (30 seconds breaks after each 30 seconds sonication on ice). The supernatant constituting the soluble fraction was obtained, the pellet constituting the insoluble fraction. 100 μ L SDS-PAGE samples were prepared by mixing 4X loading dye (33 μ L) with samples collected. The SDS-PAGE samples were boiled at 95 °C for 10 minutes and ran on 15% polyacrylamide gel with ladder at 120V for one hour and 40 minutes. The gel was later stained with Coomassie blue for ~30 minutes, then destained overnight. The gel showed nanobodies at their 15 kD band in the insoluble fraction (Supplementary Figure 1). The gel picture is provided only as a reference to the insolubility of produced nanobodies as it not of the highest quality. This would then be used to optimize for the production of soluble nanobodies.

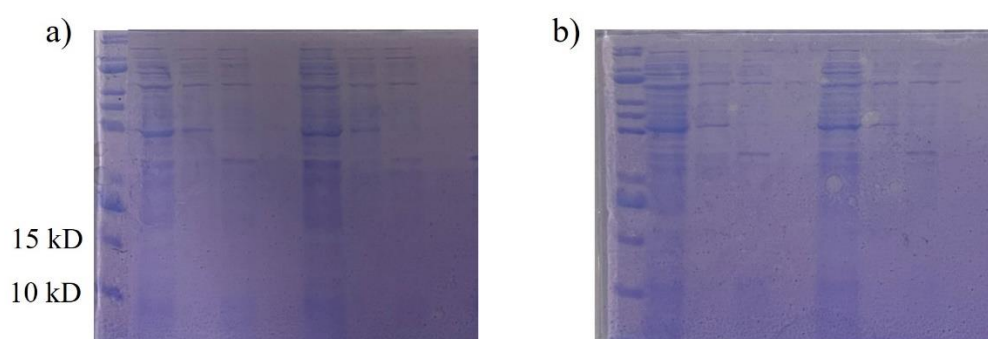


Supplementary Figure 1: 15% polyacrylamide SDS-PAGE of (from left to right): Nb1_H (Nb4), NB2_M_S, NB2_M_T (soluble, insoluble, respectively). This preliminary gel alludes to the insolubility of these nanobodies and possibly the rest at the induction parameters followed based on literature.

The preliminary expression experiment of Nb1_H (Nb4), NB2_M_S, NB2_M_T hinted towards the insolubility of the proteins. It was decided to optimize the expression of the rest of the nanobodies using Nb2. Following the transformation of the isolated anti-CD19 plasmids to BL21 (DE3). Induction's parameters (temperature, duration of induction) were tested using Nb2. Moreover, osmotic shock treatment (later optimized) was used for the extraction of periplasmic proteins where the protein is trafficked after its expression. Ni-NTA pull-down was conducted following osmotic shock treatment to check for the presence of eluted Nb2. As a first phase of optimization, both induction temperature and duration were tested as such: 37 °C for four hours, 30 °C at (4, 6, 8, and 12 hours), 25 °C at (4, 6, 8, 18, and

24 hours), 18 °C at (18 and 24 hours). Nb2 was chosen to conduct the optimization as it was not matured to act as a reference for the rest of anti-CD19 nanobodies.

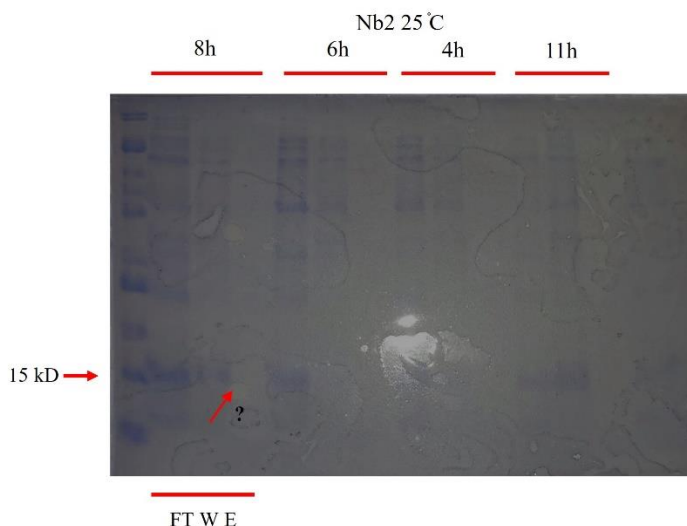
Osmotic shock treatment was conducted according to (Nagy-Fazekas et al. 2023b) (adjusted). Following harvest of the cells by centrifugation, the pellet is resuspended in around 50 mL/L of ice-cold TES buffer (30 mM Tris, 20 w/v% sucrose, 1 mM EDTA, pH=8) constituting the hypertonic treatment part of the osmotic shock protocol. The resuspended sample is incubated overnight at 4 °C on a shaker. The sample is centrifuged at 5000 rpm at 4 °C where the supernatant is obtained constituting the first hypertonic fraction. The cells are then resuspended in (30 mM Tris, 5 mM MgCl₂) and shaken for 2-4 hours on ice. The sample is then centrifuged, and the supernatant is extracted constituting the first periplasmic extract (PE). The volumes of both the hypertonic and hypotonic solutions used are the same. The process can be repeated where the duration for the osmotic treatment is decreased. The samples are treated with the hypertonic solution for 2-4 hours, and the hypotonic solution for 1-2 hours. The osmotic shock treatment yields 4 different samples that can be analyzed (first hypertonic fraction, first periplasmic extract, second hypertonic fraction, and second periplasmic extract) (Supplementary Figure 2).



Supplementary Figure 2: 15% polyacrylamide SDS-PAGE gel results of a) Nb2 sample induced at 25 °C harvested at 8 h (P1, 2, 3, 4), and ~11 h (P1, 2, 3, 4); where P stands for periplasmic. P1 would be the first periplasmic fraction collected (hypertonic treatment), P2 would be the second periplasmic fraction (hypotonic treatment). P3 would be the third periplasmic fraction (second hypertonic treatment), P4 would be the fourth periplasmic fraction (second hypotonic treatment). b) For the second gel, Ladder, 4 h (P1, 2, 3, 4), 6 h

(P1, 2, 3, 4). No band at 15 kD is observed for Nb2. Moreover, it seems that less protein is collected from hypotonic treatment.

Adjusting for both temperature and duration of the induction did not yield any observed soluble band for Nb2 at 15 kD. The concentration of IPTG used to induce the samples were concurrent with research as for both the temperatures and durations investigated for the induction of soluble Nb2 proteins. Pull-down of all samples investigated was done using Ni-NTA resin. Phosphate buffers were used for binding, washing, and elution buffers (pH= 8). The buffers contain the same concentration of potassium phosphate monobasic (20 mM), and sodium chloride (500 mM). Imidazole would be 20 mM for binding, 40 mM for wash, and 500 mM for elution buffer. Nb2 periplasmic extracts of all different induction experiments were collected and incubated with Ni-NTA resin overnight in the cold room. Samples were then washed, and elution buffer was added. Samples were collected at each step and run on 15% polyacrylamide gel; one example gel of the results is provided in (Supplementary Figure 3). No visible band at 15 kD was observed for Nb2 in all elutions.

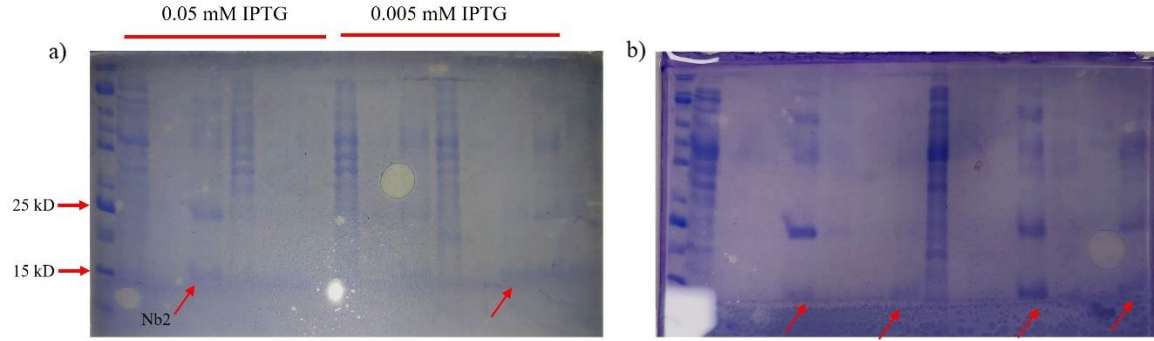


Supplementary Figure 3: Ladder, Nb2 osmotic shock treatment fractions of 25°C 8h (unbound, wash, elution), 25°C 6h (unbound, wash, elution 25°C 4h (elution, wash, unbound), and 25°C 11h (unbound, wash, elution).

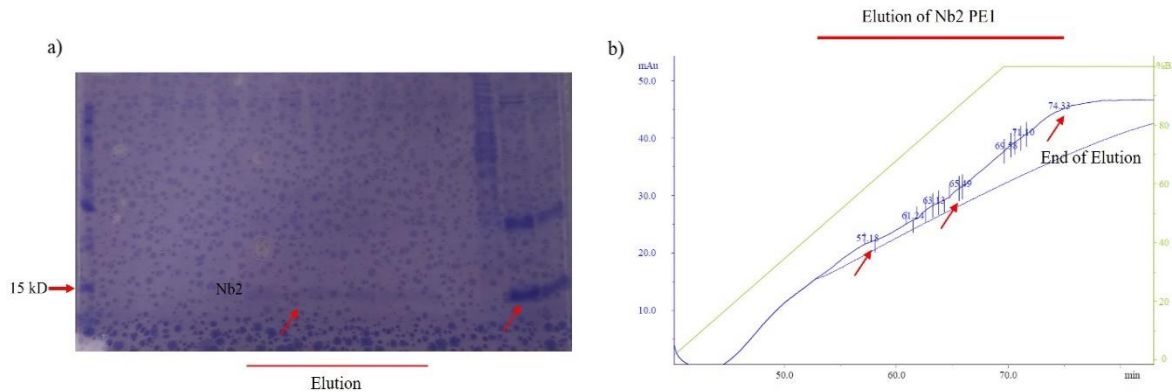
It was reasoned that aggregation is leading to the insolubility of Nb2 as the periplasmic space is small compared to the cytoplasm and as such, protein over-expression

should be lowered. A survey of the literature on solving the insolubility problem in *E. coli* protein over-expression was conducted. Amongst the different strategies done to mitigate the problem of insolubility was the use of low concentration of IPTG, and inoculating bacteria to be induced in LB containing up to 3% ethanol (Bhatwa et al. 2021; Chhetri, Kalita, and Tripathi 2015). Nb2 BL21 transformants were inoculated into LB containing 3% pure ethanol at 25 °C which did not yield any soluble expressed Nb2. It was decided to use 0.05 and 0.005 mM IPTG and test at both 30 °C and 25 °C harvesting cells at different intervals with 3% ethanol added to the inoculation medium. The presence of 3% of ethanol in the inoculation medium simulates the effect of heat-shock in bacteria (Chhetri, Kalita, and Tripathi 2015) which would allow for the production of heat-shock proteins that might ensure proper folding of the expressed nanobody. At the same time, the translational machinery would overexpress the protein less as there is a shift in cellular priority ultimately slowing down the expression of the nanobody. This coupled with lower concentration of IPTG would act as to further slowdown the overexpression of the nanobody. Ni-NTA pull-down was done, and 15% polyacrylamide SDS-PAGE was run (Supplementary Figure 4). Similar bands at 15 kD were observed for 0.05 mM IPTG at the 6h mark, and 0.005 mM IPTG at around the 24h mark. Affinity chromatography was later tried for Nb2 (Supplementary Figure 5). where pH was identified to be a player in the solubility of obtaining Nb2 in its soluble form as the buffers used for affinity chromatography had the pH of 7.4. This was later adjusted and Nb2 could be eluted with impurities at 25 kD. The buffers used for Ni-NTA affinity chromatography are the same as the ones used for Ni-NTA pull-down. Gradient elution was used to identify compatible working imidazole concentration as 500 mM imidazole led to precipitation of the eluted Nb2. Later, one step elution was used at 250 mM imidazole.

It should be noted that due to the SDS-PAGE polyacrylamide gel being at 15%, older pictures involved one round of staining/destaining as opposed to later pictures taken. Moreover, without the concentration of the nanobody's sample, bands seem to be difficult to spot as soluble Nb2 could only be obtained through lowering the rate of its overexpression.

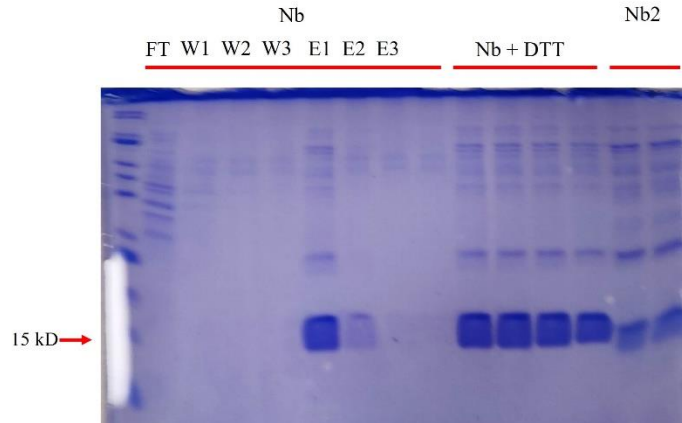


Supplementary Figure 4: a) Ladder, 0.05 mM IPTG 6h (FT, W, E), 0.05 mM IPTG o.n. (FT, W, E), 0.005 IPTG 6h (FT, W, E), 0.005 IPTG o.n. (FT, W, E), 0.005 IPTG 6h E, 0.005 IPTG o.n. E. b) Ladder, Nb2 0.005 mM IPTG 6h (FT, W1, W2-3, E1, E2, E3, old E), Nb2 0.005 mM IPTG o.n. (FT, W1, W2-3, E1, E2, E3, old E)



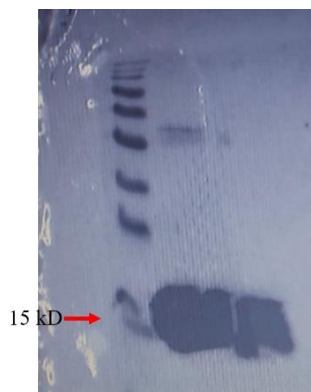
Supplementary Figure 5: a) Ladder, Nb2 PE2 elutions from samples collected at around the peak follow, followed by unbound (flowthrough) sample, followed by the two concentrated samples of Nb2 PE1; PE stands for periplasmic extract. b) FPLC Ni-NTA affinity chromatogram of Nb2 (gradient elution), similar chromatogram persists for anti-CD19 purification (same bands are observed with stepwise elution).

The impurity at 25 kD was first thought to be due to dimer formation (Goode, Dillon, and Millner 2016) given the fact that a concentration of around 0.7 M of β -mercaptoethanol was present in the loading dye. For further validation, fresh β -mercaptoethanol was obtained from a different lab, and different concentrations of DTT were added to Nb2 and another nanobody's sample, outside of the scope of this project, before SDS-PAGE gel runs (Supplementary Figure 6). Band at 25 kD persisted.



Supplementary Figure 6: (from left to right) Ladder, Nb (FT, W1-3, E1-3, resin), Nb (20 mM DTT + 1 mM EDTA, 20 mM DTT + 0.5 mM EDTA, 50 mM DTT + 1 mM EDTA, 100 mM DTT + 1 mM EDTA), Nb2 (no DTT, 100 mM DTT + 1 mM EDTA).

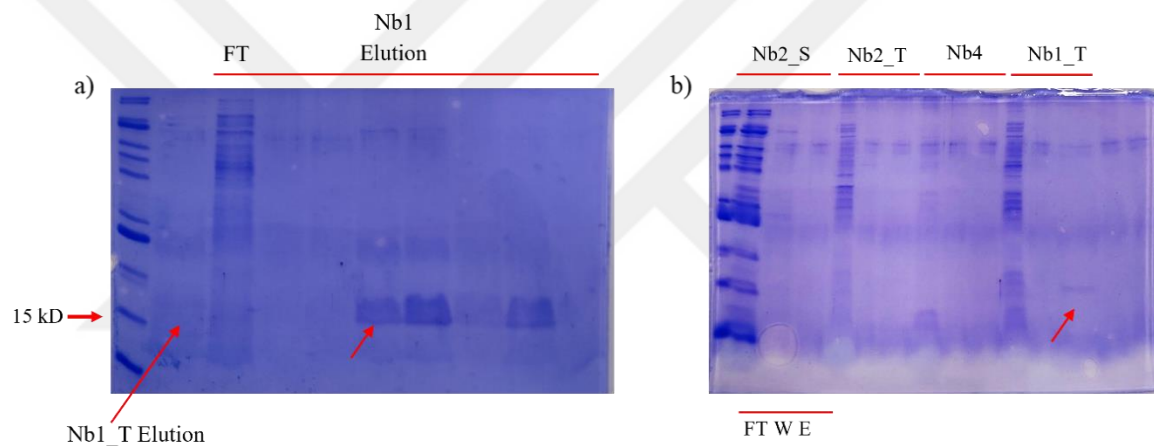
Smears persisted in the visualized gel, and upon renewal of buffers used in SDS-PAGE gel run, it was deemed that such smears are most likely due to voltage problems (as such smears did not persist before and were later solved through running the gel at 90V followed by 120V as samples migrated out of the stacking gel). To further validate that dimers were not forming, and that it was in fact an impurity eluted, a Western Blot (Wet transfer; Mouse His-tag primary and secondary antibodies) was done which reinforced that the band at 25 kD is not a nanobody dimer. Moreover, it validates the expression of his-tagged anti-CD19 nanobody (Supplementary Figure 7).



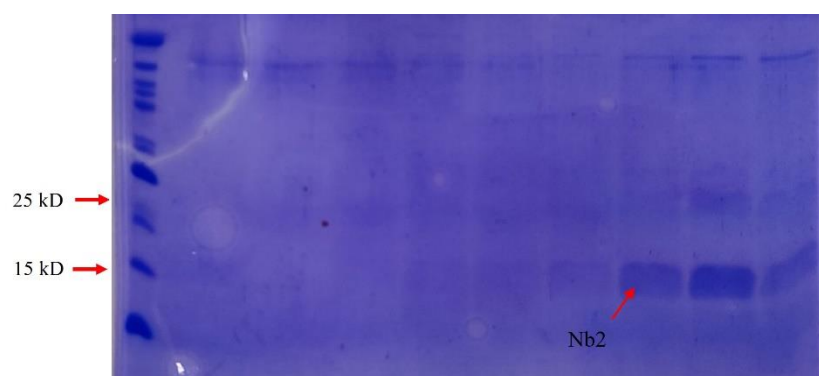
Supplementary Figure 7: (From left to right) Nb1, Nb2 at 15 kD bound to His-tag antibodies.

Induction of the rest of anti-CD19 nanobodies was conducted with accommodation to their pI when preparing the experiments. Nb1, and Nb1_M_S (Supplementary Figure 8).

seem to have worked (multiple times). However, more pull-down experiments were conducted to optimize for the rest of anti-CD19 nanobodies while trying to solve for the 25 kD impurity. Nb1 did not need optimization except for the buffers to be at pH= 7.4. Nb1_M_S was expressed in its soluble form and eluted at two instances (0.005 mM IPTG at 25 °C, 0.005 mM at 16 °C). Moreover, the protocols have been further optimized to accommodate for pH change of buffers with temperature, equilibration of periplasmic extracts with 2X binding buffer, increasing the imidazole concentration of the binding buffer to 50 mM, maintaining 0.5% Tween-20 concentration throughout the purification of nanobodies. This minimized the band observed at 25kD (Supplementary Figure 9).



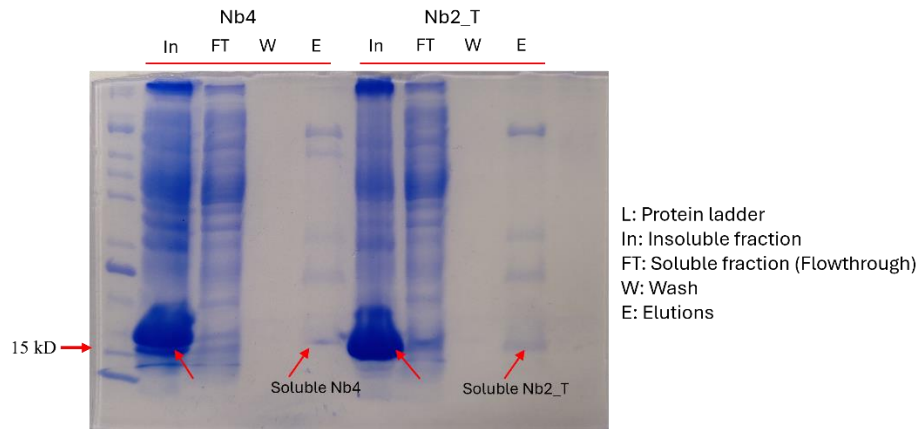
Supplementary Figure 8: a) (From right to left) Ladder, E6 of Nb1_T, Nb1 (FT, E(pre), E1-6); E(pre) stands for elution prior to acquiring of a peak in FPLC affinity chromatography. b) (from right to left): ladder, Nb2_M_S FT (with ladder; intern) followed by its wash and elution, Nb2_M_T (FT, W, E), Nb4 (FT, W, E), and Nb1_M_S (FT, W, E) followed by two random washes.



Supplementary Figure 9: Ladder, and elution of Nb2 sample via affinity chromatography. The band at 25kD is less pronounced. Smearing decreased as at that point in troubleshooting of SDS-PAGE results voltage was not accounted for.

The impurity at 25kD has been decreased but was not completely solved. Size exclusion chromatography might be one way to further purify the nanobody after affinity chromatography. Ion exchange chromatography (IEX) has been tried, however the nanobodies perform poorly when buffer exchanged and require further optimization. The nanobodies are negative at the pH they are processed in and should bind to an anion exchanger. However, it has come to our attention that the nanobody might not bind and subsequently be present in the flowthrough while the rest of impurities bind to the column. In retrospect, the need for buffer exchange might only aid in gaining a purer purified sample. The binding of the impurities to the column has been identified in a previous IEX experiment on affinity chromatography purified Nb2 as they were eluted together (not provided). The option of resolubilizing the nanobodies has not been tried as refolding the protein after denaturing usually does not yield optimal 3D native conformation and reduces the yield at which the nanobodies are obtained.

Nb4 and Nb2_T were purified at very low concentration using Ni-NTA bead pull-down (Supplementary Figure 10). Most of the nanobodies were present in the insoluble fraction of the pellet. The induced bacterial cultures were at a volume of 250 mL, and it is posited that an increase in the volume of the induced bacterial culture would result in better yields in purified Nbs in their soluble form.



Supplementary Figure 10: SDS-PAGE gel results of the purification of Nb4 and Nb2_T. A noticeable band at 15 kD in the insoluble fraction (In) is observed for both Nb4, and Nb2_T. Both the flowthrough (FT) constituting the soluble fraction, and wash (W) show no observed bands. A faint band is observed for both Nbs at 15 kD in the elution sample of Ni-NTA pull-down.

Following the successful elution of Nb2 and the production of soluble Nb1 and Nb1_T in considerable yields, there was a need to check the optimized protocol for the expression and purification as proof of concept applied to the rest of the nanobodies. The successful purification of soluble Nb4 and Nb2_T highlights the applicability of the protocol albeit at the expense of most of the fraction of expressed nanobodies being in the insoluble fraction. In terms of the behavior exhibited by all the anti-CD19 nanobodies, Nb4 and Nb2_T proved to be harder but not impossible to purify using a conventional BL21 (DE3) for their expression. The project, which this thesis is part of, could be resumed for the rest of the nanobodies where downstream functional characterization is possible.

REFERENCES

- Abraham, Mark James, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. 2015. "GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers." *SoftwareX* 1–2 (September):19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- Arbabi-Ghahroudi, Mehdi, Jamshid Tanha, and Roger MacKenzie. 2005. "Prokaryotic Expression of Antibodies." *Cancer Metastasis Reviews* 24 (4): 501–19. <https://doi.org/10.1007/s10555-005-6193-1>.
- Baek, Minkyung, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, et al. 2021. "Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network." *Science (New York, N.Y.)* 373 (6557): 871–76. <https://doi.org/10.1126/science.abj8754>.
- Banihahsemi, Seyed Reza, Ahmad Zavaran Hosseini, Fatemeh Rahbarizadeh, and Davoud Ahmadvand. 2018. "Development of Specific Nanobodies (VHH) for CD19 Immuno-Targeting of Human B-Lymphocytes." *Iranian Journal of Basic Medical Sciences* 21 (5): 455–64. <https://doi.org/10.22038/ijbms.2018.26778.6557>.
- Bhatwa, Arshpreet, Weijun Wang, Yousef I. Hassan, Nadine Abraham, Xiu-Zhen Li, and Ting Zhou. 2021. "Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia Coli and Strategies to Address Them for Industrial Applications." *Frontiers in Bioengineering and Biotechnology* 9 (February):630551. <https://doi.org/10.3389/fbioe.2021.630551>.
- Bradbury, L. E., G. S. Kansas, S. Levy, R. L. Evans, and T. F. Tedder. 1992. "The CD19/CD21 Signal Transducing Complex of Human B Lymphocytes Includes the Target of Antiproliferative Antibody-1 and Leu-13 Molecules." *Journal of Immunology (Baltimore, Md.: 1950)* 149 (9): 2841–50.
- Buttenschoen, Martin, Garrett M. Morris, and Charlotte M. Deane. 2024. "PoseBusters: AI-Based Docking Methods Fail to Generate Physically Valid Poses or Generalise to Novel Sequences." *Chemical Science* 15 (9): 3130–39. <https://doi.org/10.1039/D3SC04185A>.
- Chao, Shuangying, Yuhang Liu, Ning Ding, Yue Lin, Qian Wang, Junwen Tan, Wei Li, Yang Zheng, Xuejun Hu, and Junming Li. 2022. "Highly Expressed Soluble

- Recombinant Anti-GFP VHHs in Escherichia Coli via Optimized Signal Peptides, Strains, and Inducers.” *Frontiers in Molecular Biosciences* 9 (March):848829. <https://doi.org/10.3389/fmolb.2022.848829>.
- Chaudhuri, Dwaipayan, Satyabrata Majumder, Joyeeta Datta, and Kalyan Giri. 2022. “Designing of Nanobodies against Dengue Virus Capsid: A Computational Affinity Maturation Approach.” *Journal of Biomolecular Structure and Dynamics* 0 (0): 1–11. <https://doi.org/10.1080/07391102.2022.2029773>.
- Cheng, Xin, Jiewen Wang, Guangbo Kang, Min Hu, Bo Yuan, Yingtian Zhang, and He Huang. 2019a. “Homology Modeling-Based in Silico Affinity Maturation Improves the Affinity of a Nanobody.” *International Journal of Molecular Sciences* 20 (17). <https://doi.org/10.3390/ijms20174187>.
- Cheng, Xin, Jiewen Wang, Guangbo Kang, Min Hu, Bo Yuan, Yingtian Zhang, and He Huang. 2019b. “Homology Modeling-Based in Silico Affinity Maturation Improves the Affinity of a Nanobody.” *International Journal of Molecular Sciences* 20 (17): 4187. <https://doi.org/10.3390/ijms20174187>.
- Chhetri, Gaurav, Parismita Kalita, and Timir Tripathi. 2015. “An Efficient Protocol to Enhance Recombinant Protein Expression Using Ethanol in Escherichia Coli.” *MethodsX* 2 (October):385–91. <https://doi.org/10.1016/j.mex.2015.09.005>.
- Chiaretti, Sabina, Gina Zini, and Renato Bassan. 2014. “Diagnosis and Subclassification of Acute Lymphoblastic Leukemia.” *Mediterranean Journal of Hematology and Infectious Diseases* 6 (1): e2014073. <https://doi.org/10.4084/MJHID.2014.073>.
- Dillman, R. O. 2006. “Radioimmunotherapy of B-Cell Lymphoma with Radiolabelled Anti-CD20 Monoclonal Antibodies.” *Clinical and Experimental Medicine* 6 (1): 1–12. <https://doi.org/10.1007/s10238-006-0087-6>.
- Dolinsky, Todd J., Paul Czodrowski, Hui Li, Jens E. Nielsen, Jan H. Jensen, Gerhard Klebe, and Nathan A. Baker. 2007. “PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations.” *Nucleic Acids Research* 35 (Web Server issue): W522–525. <https://doi.org/10.1093/nar/gkm276>.
- Dolinsky, Todd J., Jens E. Nielsen, J. Andrew McCammon, and Nathan A. Baker. 2004. “PDB2PQR: An Automated Pipeline for the Setup of Poisson–Boltzmann Electrostatics Calculations.” *Nucleic Acids Research* 32 (suppl_2): W665–67. <https://doi.org/10.1093/nar/gkh381>.

- Dominguez, Cyril, Rolf Boelens, and Alexandre M. J. J. Bonvin. 2003. "HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information." *Journal of the American Chemical Society* 125 (7): 1731–37. <https://doi.org/10.1021/ja026939x>.
- Dores, Graça M., Susan S. Devesa, Rochelle E. Curtis, Martha S. Linet, and Lindsay M. Morton. 2012. "Acute Leukemia Incidence and Patient Survival among Children and Adults in the United States, 2001-2007." *Blood* 119 (1): 34–43. <https://doi.org/10.1182/blood-2011-04-347872>.
- Efraimidis, Evangelos, Marios G. Krokidis, Themis P. Exarchos, Tamas Lazar, and Panagiotis Vlamos. 2023. "In Silico Structural Analysis Exploring Conformational Folding of Protein Variants in Alzheimer's Disease." *International Journal of Molecular Sciences* 24 (17): 13543. <https://doi.org/10.3390/ijms241713543>.
- Fraser, James S., and Mark A. Murcko. 2024. "Structure Is Beauty, but Not Always Truth." *Cell* 187 (3): 517–20. <https://doi.org/10.1016/j.cell.2024.01.003>.
- Gambella, Massimiliano, Simona Carlomagno, Anna Maria Raiola, Livia Giannoni, Chiara Ghiggi, Chiara Setti, Chiara Giordano, et al. 2022. "CD19-Targeted Immunotherapies for Diffuse Large B-Cell Lymphoma." *Frontiers in Immunology* 13 (February). <https://doi.org/10.3389/fimmu.2022.837457>.
- Gavenonis, Jason, Bradley A. Sheneman, Timothy R. Siegert, Matthew R. Eshelman, and Joshua A. Kritzer. 2014. "Comprehensive Analysis of Loops at Protein-Protein Interfaces for Macrocyclic Design." *Nature Chemical Biology* 10 (9): 716–22. <https://doi.org/10.1038/nchembio.1580>.
- Goode, J., G. Dillon, and P. A. Millner. 2016. "The Development and Optimisation of Nanobody Based Electrochemical Immunosensors for IgG." *Sensors and Actuators B: Chemical* 234 (October): 478–84. <https://doi.org/10.1016/j.snb.2016.04.132>.
- Grote, Andreas, Karsten Hiller, Maurice Scheer, Richard Münch, Bernd Nörtemann, Dietmar C. Hempel, and Dieter Jahn. 2005. "JCat: A Novel Tool to Adapt Codon Usage of a Target Gene to Its Potential Expression Host." *Nucleic Acids Research* 33 (Web Server issue): W526-531. <https://doi.org/10.1093/nar/gki376>.

- Hacisuleyman, Aysima, and Burak Erman. 2019. “ModiBodies: A Computational Method for Modifying Nanobodies to Improve Their Antigen Binding Affinity and Specificity.” *bioRxiv*, October, 820373. <https://doi.org/10.1101/820373>.
- Hacisuleyman, Aysima, and Burak Erman. 2020. “ModiBodies: A Computational Method for Modifying Nanobodies in Nanobody-Antigen Complexes to Improve Binding Affinity and Specificity.” *Journal of Biological Physics* 46 (2): 189–208. <https://doi.org/10.1007/s10867-020-09548-3>.
- Hamers-Casterman, C., T. Atarhouch, S. Muyldermans, G. Robinson, C. Hamers, E. B. Songa, N. Bendahman, and R. Hamers. 1993. “Naturally Occurring Antibodies Devoid of Light Chains.” *Nature* 363 (6428): 446–48. <https://doi.org/10.1038/363446a0>.
- Hammer, Ohad. 2012. “CD19 as an Attractive Target for Antibody-Based Therapy.” *mAbs* 4 (5): 571–77. <https://doi.org/10.4161/mabs.21338>.
- Hu, Min, Guangbo Kang, Xin Cheng, Jiewen Wang, Ruowei Li, Zixuan Bai, Dong Yang, and He Huang. 2020. “A Sound Strategy for Homology Modeling-Based Affinity Maturation of a HIF-1 α Single-Domain Intrabody.” *bioRxiv*. <https://doi.org/10.1101/2020.05.14.096149>.
- Kariuki, Christopher K., and Stefan Magez. 2021. “Improving the Yield of Recalcitrant Nanobodies® by Simple Modifications to the Standard Protocol.” *Protein Expression and Purification* 185 (September):105906. <https://doi.org/10.1016/j.pep.2021.105906>.
- Kothari, Manjeet, Anil Wanjari, Sourya Acharya, Vineet Karwa, Roma Chavhan, Sunil Kumar, Ajinkya Kadu, and Rajvardhan Patil. 2024. “A Comprehensive Review of Monoclonal Antibodies in Modern Medicine: Tracing the Evolution of a Revolutionary Therapeutic Approach.” *Cureus* 16 (6): e61983. <https://doi.org/10.7759/cureus.61983>.
- Laskowski, Roman A., and Mark B. Swindells. 2011. “LigPlot+: Multiple Ligand-Protein Interaction Diagrams for Drug Discovery.” *Journal of Chemical Information and Modeling* 51 (10): 2778–86. <https://doi.org/10.1021/ci200227u>.
- Laurent, Elisabeth, Anna Sieber, Benjamin Salzer, Anna Wachernig, Jacqueline Seigner, Manfred Lehner, René Geyeregger, et al. 2021. “Directed Evolution of Stabilized Monomeric CD19 for Monovalent CAR Interaction Studies and Monitoring of

- CAR-T Cell Patients.” *ACS Synthetic Biology* 10 (5): 1184–98.
<https://doi.org/10.1021/acssynbio.1c00010>.
- Lei, S. P., H. C. Lin, S. S. Wang, J. Callaway, and G. Wilcox. 1987. “Characterization of the *Erwinia Carotovora* pelB Gene and Its Product Pectate Lyase.” *Journal of Bacteriology* 169 (9): 4379–83. <https://doi.org/10.1128/jb.169.9.4379-4383.1987>.
- Li, Lei, Shuang Chen, Zhichao Miao, Yang Liu, Xu Liu, Zhi-Xiong Xiao, and Yang Cao. 2019. “AbRSA: A Robust Tool for Antibody Numbering.” *Protein Science: A Publication of the Protein Society* 28 (8): 1524–31.
<https://doi.org/10.1002/pro.3633>.
- Li, Xinchun, Ying Ding, Mengting Zi, Li Sun, Wenjie Zhang, Shun Chen, and Yuekang Xu. 2017. “CD19, from Bench to Bedside.” *Immunology Letters* 183 (March):86–95. <https://doi.org/10.1016/j.imlet.2017.01.010>.
- Liu, Jia, Luyao Wang, Hongbo Yang, Meng Xing, Shihai Liu, Zhuang Yu, and Leina Ma. 2020. “A Narrative Review of Critical Factors for Better Efficacy of CD19 Chimeric Antigen Receptor T Cell Therapy in the Treatment of B Cell Malignancies.” *Translational Cancer Research* 9 (9). <https://doi.org/10.21037/tcr-20-1044>.
- Loghavi, Sanam, Jeffery L. Kutok, and Jeffrey L. Jorgensen. 2015. “B-Acute Lymphoblastic Leukemia/Lymphoblastic Lymphoma.” *American Journal of Clinical Pathology* 144 (3): 393–410.
<https://doi.org/10.1309/AJCPAN7BH5DNYWZB>.
- Marco, Ario de. 2020. “Recombinant Expression of Nanobodies and Nanobody-Derived Immunoreagents.” *Protein Expression and Purification* 172 (August):105645.
<https://doi.org/10.1016/j.pep.2020.105645>.
- Matsumoto, A. K., J. Kopicky-Burd, R. H. Carter, D. A. Tuveson, T. F. Tedder, and D. T. Fearon. 1991. “Intersection of the Complement and Immune Systems: A Signal Transduction Complex of the B Lymphocyte-Containing Complement Receptor Type 2 and CD19.” *The Journal of Experimental Medicine* 173 (1): 55–64.
<https://doi.org/10.1084/jem.173.1.55>.
- McMahon, Conor, Alexander S. Baier, Roberta Pascolutti, Marcin Wegrecki, Sanduo Zheng, Janice X. Ong, Sarah C. Erlandson, et al. 2018. “Yeast Surface Display Platform for Rapid Discovery of Conformationally Selective Nanobodies.”

- Nature Structural & Molecular Biology* 25 (3): 289–96.
<https://doi.org/10.1038/s41594-018-0028-6>.
- Minuchehr, Z., and B. Goliaei. 2005. “Propensity of Amino Acids in Loop Regions Connecting Beta-Strands.” *Protein and Peptide Letters* 12 (4): 379–82.
<https://doi.org/10.2174/0929866053765617>.
- Muyldermans, Serge. 2013. “Nanobodies: Natural Single-Domain Antibodies.” *Annual Review of Biochemistry* 82 (Volume 82, 2013): 775–97.
<https://doi.org/10.1146/annurev-biochem-063011-092449>.
- Muyldermans, Serge. 2021a. “A Guide to: Generation and Design of Nanobodies.” *The FEBS Journal* 288 (7): 2084–2102. <https://doi.org/10.1111/febs.15515>.
- Muyldermans, Serge. 2021b. “A Guide to: Generation and Design of Nanobodies.” *The FEBS Journal* 288 (7): 2084–2102. <https://doi.org/10.1111/febs.15515>.
- Muyldermans, Serge. 2021c. “Applications of Nanobodies.” *Annual Review of Animal Biosciences* 9 (February):401–21. <https://doi.org/10.1146/annurev-animal-021419-083831>.
- Nagy-Fazekas, Dóra, Pál Stráner, Péter Ecsédi, Nóra Taricska, Adina Borbély, László Nyitray, and András Perczel. 2023a. “A Novel Fusion Protein System for the Production of Nanobodies and the SARS-CoV-2 Spike RBD in a Bacterial System.” *Bioengineering* 10 (3): 389.
<https://doi.org/10.3390/bioengineering10030389>.
- Nagy-Fazekas, Dóra, Pál Stráner, Péter Ecsédi, Nóra Taricska, Adina Borbély, László Nyitray, and András Perczel. 2023b. “A Novel Fusion Protein System for the Production of Nanobodies and the SARS-CoV-2 Spike RBD in a Bacterial System.” *Bioengineering* 10 (3): 389.
<https://doi.org/10.3390/bioengineering10030389>.
- Pardon, Els, Toon Laeremans, Sarah Triest, Søren G. F. Rasmussen, Alexandre Wohlkönig, Armin Ruf, Serge Muyldermans, Wim G. J. Hol, Brian K. Kobilka, and Jan Steyaert. 2014. “A General Protocol for the Generation of Nanobodies for Structural Biology.” *Nature Protocols* 9 (3): 674–93.
<https://doi.org/10.1038/nprot.2014.039>.
- Pierce, Brian G., Kevin Wiehe, Howook Hwang, Bong-Hyun Kim, Thom Vreven, and Zhiping Weng. 2014. “ZDOCK Server: Interactive Docking Prediction of

- Protein–Protein Complexes and Symmetric Multimers.” *Bioinformatics* 30 (12): 1771–73. <https://doi.org/10.1093/bioinformatics/btu097>.
- Podyma-Inoue, Katarzyna A., Takuya Moriwaki, Anupama R. Rajapakshe, Kazue Terasawa, and Miki Hara-Yokoyama. 2016. “Characterization of Heparan Sulfate Proteoglycan-Positive Recycling Endosomes Isolated from Glioma Cells.” *Cancer Genomics & Proteomics* 13 (6): 443–52.
- Pont, Laura, Valeriia Kuzyk, Fernando Benavente, Victoria Sanz-Nebot, Oleg A. Mayboroda, Manfred Wuhrer, and Guinevere S. M. Lageveen-Kammeijer. 2021. “Site-Specific N-Linked Glycosylation Analysis of Human Carcinoembryonic Antigen by Sheathless Capillary Electrophoresis–Tandem Mass Spectrometry.” *Journal of Proteome Research* 20 (3): 1666–75. <https://doi.org/10.1021/acs.jproteome.0c00875>.
- Poustforoosh, Alireza, Sanaz Faramarz, Manica Negahdaripour, and Hassan Hashemipour. 2023a. “Modeling and Affinity Maturation of an Anti-CD20 Nanobody: A Comprehensive in-Silico Investigation.” *Scientific Reports* 13 (1): 582. <https://doi.org/10.1038/s41598-023-27926-4>.
- Poustforoosh, Alireza, Sanaz Faramarz, Manica Negahdaripour, and Hassan Hashemipour. 2023b. “Modeling and Affinity Maturation of an Anti-CD20 Nanobody: A Comprehensive in-Silico Investigation.” *Scientific Reports* 13 (1): 582. <https://doi.org/10.1038/s41598-023-27926-4>.
- Poustforoosh, Alireza, Sanaz Faramarz, Mohammad Hadi Nematollahi, Hassan Hashemipour, Manica Negahdaripour, and Abbas Pardakhty. 2022. “In Silico SELEX Screening and Statistical Analysis of Newly Designed 5mer Peptide-Aptamers as Bcl-Xl Inhibitors Using the Taguchi Method.” *Computers in Biology and Medicine* 146 (July):105632. <https://doi.org/10.1016/j.compbiomed.2022.105632>.
- Pui, Ching-Hon, and Sima Jeha. 2007. “New Therapeutic Strategies for the Treatment of Acute Lymphoblastic Leukaemia.” *Nature Reviews Drug Discovery* 6 (2): 149–65. <https://doi.org/10.1038/nrd2240>.
- Reily, Colin, Tyler J. Stewart, Matthew B. Renfrow, and Jan Novak. 2019. “Glycosylation in Health and Disease.” *Nature Reviews Nephrology* 15 (6): 346–66. <https://doi.org/10.1038/s41581-019-0129-4>.

- Ritchie, Mark, David Knight, Sara Kirkham, and David Thornton. 2004. "Site-Specific N-Linked Glycosylation Analysis on the Human Salivary Mucin MUC5B Using Precursor Ion Discovery on the CAPLC Q-TOF System." *International Journal of Experimental Pathology* 85 (4): A71–72. <https://doi.org/10.1111/j.0959-9673.2004.390aj.x>.
- Salema, Valencio, and Luis Ángel Fernández. 2013. "High Yield Purification of Nanobodies from the Periplasm of E. Coli as Fusions with the Maltose Binding Protein." *Protein Expression and Purification* 91 (1): 42–48. <https://doi.org/10.1016/j.pep.2013.07.001>.
- Sambrook, Joseph. 2001. *Molecular Cloning : A Laboratory Manual*. Third edition. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press, [2001] ©2001. <https://search.library.wisc.edu/catalog/999897924602121>.
- Shi, Zhenzhong, Xiyang Li, Lu Wang, Zengchao Sun, Haiwei Zhang, Xiaochen Chen, Qianqian Cui, et al. 2022. "Structural Basis of Nanobodies Neutralizing SARS-CoV-2 Variants." *Structure* 30 (5): 707-720.e5. <https://doi.org/10.1016/j.str.2022.02.011>.
- Sievers, Fabian, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, et al. 2011. "Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega." *Molecular Systems Biology* 7 (October):539. <https://doi.org/10.1038/msb.2011.75>.
- Soler, Miguel A., Sara Fortuna, Ario de Marco, and Alessandro Laio. 2018. "Binding Affinity Prediction of Nanobody–Protein Complexes by Scoring of Molecular Dynamics Trajectories." *Physical Chemistry Chemical Physics* 20 (5): 3438–44. <https://doi.org/10.1039/C7CP08116B>.
- Soler, Miguel Angel, Ozge Ozkilinc, Yamanappa Hunashal, Paolo Giannozzi, Gennaro Esposito, and Federico Fogolari. 2023. "Molecular Electrostatics and pKa Shifts Calculations with the Generalized Born Model. A Tutorial through Examples with Blues2." *Computer Physics Communications* 287 (June):108716. <https://doi.org/10.1016/j.cpc.2023.108716>.
- Stasi, R., A. Pagano, E. Stipa, and S. Amadori. 2001. "Rituximab Chimeric Anti-CD20 Monoclonal Antibody Treatment for Adults with Chronic Idiopathic Thrombocytopenic Purpura." *Blood* 98 (4): 952–57. <https://doi.org/10.1182/blood.v98.4.952>.

- Susa, Katherine J., Shaun Rawson, Andrew C. Kruse, and Stephen C. Blacklow. 2021. "Cryo-EM Structure of the B Cell Co-Receptor CD19 Bound to the Tetraspanin CD81." *Science* 371 (6526): 300–305. <https://doi.org/10.1126/science.abd9836>.
- Teplyakov, Alexey, Galina Obmolova, Jinqun Luo, and Gary L. Gilliland. 2018. "Crystal Structure of B-Cell Co-Receptor CD19 in Complex with Antibody B43 Reveals an Unexpected Fold." *Proteins* 86 (5): 495–500. <https://doi.org/10.1002/prot.25485>.
- Terwilliger, T, and M Abdul-Hay. 2017. "Acute Lymphoblastic Leukemia: A Comprehensive Review and 2017 Update." *Blood Cancer Journal* 7 (6): e577. <https://doi.org/10.1038/bcj.2017.53>.
- Thie, Holger, Thomas Schirrmann, Matthias Paschke, Stefan Dübel, and Michael Hust. 2008. "SRP and Sec Pathway Leader Peptides for Antibody Phage Display and Antibody Fragment Production in *E. Coli*." *New Biotechnology* 25 (1): 49–54. <https://doi.org/10.1016/j.nbt.2008.01.001>.
- Tsuchiya, Yuko, Yu Yamamori, and Kentaro Tomii. 2022. "Protein–Protein Interaction Prediction Methods: From Docking-Based to AI-Based Approaches." *Biophysical Reviews* 14 (6): 1341–48. <https://doi.org/10.1007/s12551-022-01032-7>.
- Vivcharuk, Victor, Jason Baardsnes, Christophe Deprez, Traian Sulea, Maria Jaramillo, Christopher R. Corbeil, Alaka Mullick, et al. 2017. "Assisted Design of Antibody and Protein Therapeutics (ADAPT)." *PLOS ONE* 12 (7): e0181490. <https://doi.org/10.1371/journal.pone.0181490>.
- Wang, Ercheng, Huiyong Sun, Junmei Wang, Zhe Wang, Hui Liu, John Z. H. Zhang, and Tingjun Hou. 2019. "End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design." *Chemical Reviews* 119 (16): 9478–9508. <https://doi.org/10.1021/acs.chemrev.9b00055>.
- Wang, Hongxia, Liyan Wang, Yanning Li, Guangqi Li, Xiaochun Zhang, Dan Jiang, Yanting Zhang, Liyuan Liu, Yuankui Chu, and Guangxian Xu. 2021a. "Nanobody-Armed T Cells Endow CAR-T Cells with Cytotoxicity against Lymphoma Cells." *Cancer Cell International* 21 (1): 450. <https://doi.org/10.1186/s12935-021-02151-z>.
- Wang, Hongxia, Liyan Wang, Yanning Li, Guangqi Li, Xiaochun Zhang, Dan Jiang, Yanting Zhang, Liyuan Liu, Yuankui Chu, and Guangxian Xu. 2021b. "Nanobody-Armed T Cells Endow CAR-T Cells with Cytotoxicity against

- Lymphoma Cells.” *Cancer Cell International* 21 (1): 450.
<https://doi.org/10.1186/s12935-021-02151-z>.
- Wang, Kemeng, Guoqing Wei, and Delong Liu. 2012. “CD19: A Biomarker for B Cell Development, Lymphoma Diagnosis and Therapy.” *Experimental Hematology & Oncology* 1 (1): 36. <https://doi.org/10.1186/2162-3619-1-36>.
- Waterhouse, Andrew M., James B. Procter, David M. A. Martin, Michèle Clamp, and Geoffrey J. Barton. 2009. “Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench.” *Bioinformatics* 25 (9): 1189–91.
<https://doi.org/10.1093/bioinformatics/btp033>.
- Xue, Li C., João Pglm Rodrigues, Panagiotis L. Kastitis, Alexandre Mjj Bonvin, and Anna Vangone. 2016. “PRODIGY: A Web Server for Predicting the Binding Affinity of Protein–Protein Complexes.” *Bioinformatics* 32 (23): 3676–78.
<https://doi.org/10.1093/bioinformatics/btw514>.
- Yan, Yumeng, Huanyu Tao, Jiahua He, and Sheng-You Huang. 2020a. “The HDock Server for Integrated Protein–Protein Docking.” *Nature Protocols* 15 (5): 1829–52. <https://doi.org/10.1038/s41596-020-0312-x>.
- Yan, Yumeng, Huanyu Tao, Jiahua He, and Sheng-You Huang. 2020b. “The HDock Server for Integrated Protein–Protein Docking.” *Nature Protocols* 15 (5): 1829–52. <https://doi.org/10.1038/s41596-020-0312-x>.
- Yau, Kerm, Ginette Dubuc, Shenghua li, Tomoko Hiram, Colin Mackenzie, Lutz Jermutus, J Hall, and Jamshid Tanha. 2005. “Affinity Maturation of a VHH by Mutational Hotspot Randomization.” *Journal of Immunological Methods* 297 (March):213–24. <https://doi.org/10.1016/j.jim.2004.12.005>.
- Ye, Gang, Joseph P. Gallant, Christopher Massey, Ke Shi, Wanbo Tai, Jian Zheng, Abby E. Odle, et al. 2020. “The Development of a Novel Nanobody Therapeutic for SARS-CoV-2.” *bioRxiv*. <https://doi.org/10.1101/2020.11.17.386532>.
- Zhang, Xiaohui, Perna Rastogi, Bijal Shah, and Ling Zhang. 2017. “B Lymphoblastic Leukemia/Lymphoma: New Insights into Genetics, Molecular Aberrations, Subclassification and Targeted Therapy.” *Oncotarget* 8 (39): 66728–41.
<https://doi.org/10.18632/oncotarget.19271>.