T.C.
DOKUZ EYLUL UNIVERSITY
IZMIR INTERNATIONAL BIOMEDICINE AND GENOME INSTITUTE

DEVELOPMENT OF ANTI-CANCER SINGLE-DOMAIN ANTIBODY TOOLBOX

NAZLI EDA ESKİER

DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

DOCTOR OF PHILOSOPHY THESIS
IZMIR-2024

THESIS CODE: MBG 6199-2017850053

T.C.
DOKUZ EYLUL UNIVERSITY
IZMIR INTERNATIONAL BIOMEDICINE AND GENOME INSTITUTE

DEVELOPMENT OF ANTI-CANCER SINGLE-DOMAIN ANTIBODY TOOLBOX

DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

DOCTOR OF PHILOSOPHY THESIS

NAZLI EDA ESKİER

Advisor: Prof. Dr. Gökhan Karakülah

Co-advisor: Dr. Sibel Kalyoncu Uzunlar

(This thesis was supported by YÖK 100/2000, TÜBİTAK 2211-C, TÜBİTAK 2214-A and EMBO Scientific Exchange Grant)

THESIS CODE: MBG 6199-2017850053

Dokuz Eylül University Izmir International Biomedicine and Genome Institute Department of Genomics and Molecular Biotechnology, Molecular Biology and Genetics graduate program Doctor of Philosophy student Nazlı Eda Eskier has successfully completed her Philosophy thesis titled **'DEVELOPMENT OF ANTI-CANCER SINGLE-DOMAIN ANTIBODY TOOLBOX'** on the date of **18 / 01 / 2024**.

Prof. Dr. Gökhan Karakülah
Dokuz Eylul University / İzmir International Biomedicine and Genome Institute / Department of Genomics and Molecular Biotechnology
HEAD OF THE COMMITTEE

Prof. Dr. Mehmet İnan
Akdeniz University / Faculty of Engineering / Department of Food Engineering
MEMBER

Prof. Dr. Şerif Şentürk
Dokuz Eylul University / İzmir International Biomedicine and Genome Institute / Department of Genomics and Molecular Biotechnology
MEMBER

Prof. Dr. Esin Firuzan
Dokuz Eylul University / Faculty of Sciences / Department of Statistics
MEMBER

Assoc. Prof. Burcu Kaplan Türköz
Ege University / Faculty of Engineering / Department of Food Engineering
MEMBER

TABLE OF CONTENTS

TABLE OF CONTENTS ..... **i**

ABBREVIATIONS ..... **x**

LIST OF TABLES ..... **xiv**

LIST OF FIGURES ..... **xx**

LIST OF EQUATIONS ..... **xxv**

ACKNOWLEDGEMENTS ..... **xxvi**

ABSTRACT ..... **xxvi**

ÖZET **2**

1. INTRODUCTION AND AIM **3**

1.1. Statement and importance of the problem **3**

1.2. Aim of the study **3**

2. BACKGROUND INFROMATION & LITERATURE **5**

2.1. Antibodies **5**

2.1.1 Immunoglobulin Gs *6*

2.1.2. Alternative antibodies in nature *8*

2.1.2.1. The emergence of alternative antibodies *8*

2.1.2.2. Heavy-chain only antibodies *9*

2.1.3. Antibody fragments *11*

2.1.3.1. Nanobodies *14*

2.1.3.2. Single-domain antibodies *16*

2.1.4. Engineering approaches for functional sdAb generation *17*

2.1.4.1. Camelisation *17*

2.1.4.2. Humanisation *17*

2.1.5. Antibody discovery *18*

2.1.5.1. Computational approaches in antibody discovery *19*

2.2. Cancer **20**
2.2.1. Hallmarks of cancer *20*
2.2.2. Metastasis *22*
2.2.2.1. Angiogenesis *23*
2.2.2.1.1. Vascular endothelial growth factor *23*
2.2.2.1.1.1. VEGF in the body *25*
2.2.2.1.1.2. VEGF in cancer therapy and diagnosis *25*
2.3. Bevacizumab **25**
2.3.1. Bevacizumab working mechanism *26*
2.3.2. Molecular imaging of tumour angiogenesis *26*
2.3.2.1. Bevacizumab in angiogenesis labelling *27*
2.4. Antibody labelling **27**
3. MATERIALS AND METHODS **30**
3.1. Type of research **30**
3.2. Location and time frame of research **30**
3.3. Research population, sampling and experimental group **30**
3.4. Research materials **31**
3.4.1. General *31*
3.4.2. Human cell line *45*
3.4.2. Expression vectors *46*
3.4.3. Oligonucleotides *47*
3.4.4. Primers *48*
3.4.5. In vivo experimental animal models *50*
3.5. Research variables **50**
3.6. Data collection tools **50**
3.6.1. Plasmid design *50*

3.6.2. In Silico approaches for prediction of aggregation prone residues *51*

3.6.2.1. PDBePISA *51*

3.6.2.2. Pasta 2.0 ..... *52*

3.6.2.3. Tango 2.2 ..... *52*

3.6.2.4. Waltz 2.0 ..... *52*

3.6.2.5. Aggrescan 3D 2.0 ..... *52*

3.6.2.6. CamSol Intrinsic v2.2 ..... *52*

3.6.2.7. CamSol structurally corrected v2.1 ..... *53*

3.6.3. The databases *53*

3.6.3.1. abYsis: integrated antibody sequence database *53*

3.6.3.2. sdAb-DB: the single domain antibody database *53*

3.6.4. Site-directed mutagenesis (SDM) *53*

3.6.4.1. SDM with QuikChange lightning kit *54*

3.6.4.2. SDM with Q5 high-fidelity kit *55*

3.6.4.3. SDM with KOD hot start kit *56*

3.6.5. DpnI enzyme digestion *57*

3.6.5.1. NEB DpnI kit *57*

3.6.5.2. Thermo DpnI kit *57*

3.6.6. PCR clean up - QIAquick gel extraction kit *58*

3.6.7. Agarose gel electrophoresis *59*

3.6.7.1. Gel clean up- QIAquick gel extraction kit *60*

3.6.8. Transformation *61*

3.6.8.1. Electroporation *61*

3.6.8.1.1. Electrocompotent cell making-BioRad *63*

3.5.7.1.2. Electrocompotent cell making-Custom *64*

3.6.8.2. Heat-shock transformation- BL21(DE3) pLysS / DH5α *65*

3.6.8.2.1. Competent cell making- DH5α *66*

3.6.9. Plasmid isolation *67*

3.6.9.1. MN plasmid DNA purification kit *67*

3.6.9.2. EZ-10 spin column plasmid mini-prep kit *68*

3.6.9.3. Qiagen QIAprep® spin miniprep kit *69*

3.6.10. Protein production and purification *70*

3.6.10.1. Protein production and purification- BL21(DE3) pLysS *70*

3.6.11. Redesign for new sdAbs *81*

3.6.12. Cloning *81*

3.6.12.1. Plasmid recovery *82*

3.6.12.2. Preparation of the primers *82*

3.6.12.3. Amplification of Bevacizumab V_H sequences by PCR *83*

3.5.12.4. Digestion of vector pHen6 ..... *84*

3.5.12.5. Ligation of vector pHen6 with modified Bevacizumab V_H fragments *84*

3.6.12.6. Heat-shock transformation after ligation with WK6 cells *85*

3.6.12.7. Plasmid isolation - Genelute plasmid miniprep kit *86*

3.6.12.8. Confirmation of ligation into expression vector pHen6 by sequencing *87*

3.6.13. Protein production and purification – WK6 ..... *87*

3.6.13.1. Unfold/refold before purification *95*

3.6.13.2. Unfold/refold after purification *97*

3.6.13.3. Fast protein liquid chromatography (FPLC)-Protein A *97*

3.6.14. sdAb characterization *99*

3.6.14.1. SDS-PAGE *99*

3.6.14.2. Western Blot *102*

3.6.14.3. Enzyme-Linked Immunosorbent Assay (ELISA) *105*

3.6.14.4. Tm determination *109*

3.6.14.4.1. Thermal denaturation *109*

3.6.14.4.2. Chemical denaturation *111*

3.6.14.5. Thermal stability determination assay *112*

3.6.14.6. Isothermal Titration Calorimetry (ITC) *113*

3.6.14.7. In vivo zebrafish studies *114*

3.6.15. In vivo mice studies *117*

3.6.15.1. HT-29 cell line and culture conditions *119*

3.6.15.2. HT-29 cell counting *120*

3.6.16. Fluorescent labelling *121*

3.6.16.1. Reduction and buffer exchange *121*

3.6.16.2. Labelling and Purification *122*

3.6.16.3. Quality control *123*

3.6.16.3.1. Gel filtration for QC *123*

3.6.16.3.2. SDS-PAGE for QC *124*

3.6.17. Radioactive labelling *124*

3.6.17.1. Random coupling with DOTA *124*

3.6.17.2. ^{111}*In labelling of Bevacizumab* *127*

3.6.17.3. Instant Thin Layer Chromatography *127*

3.6.17.4. ^{99m}*Tc-labelling of sdAbs* *127*

3.6.18. In silico approaches for determination of differences between antibodies and nanobodies *129*

3.6.18.1. Datasets *129*

3.6.187.2. Multiple sequence alignment *130*

3.6.18.3. Paratope binding prediction *130*

3.6.18.4. Sequence conservation motif analysis *130*

3.6.18.5. Disulphide bond analysis *131*

3.6.18.6. Salt bridge analysis *131*

3.6.18.7. Analysis of physicochemical properties *131*

3.6.18.8. Statistical analysis *132*

3.6.18.8.1. The binary logistic regression model *132*

3.6.18.8.2. The Boruta algorithm *133*

3.6.18.8.3. Statistical significance calculation *133*

3.6.18.9. Tools and plots *133*

3.7. Research plan and timeline **133**

3.8. Data analysis **139**

3.9. Limitations of research **139**

3.10. Ethical committee approval **139**

3.10.1. Ethical committee approval from Animal Experiments Local Ethics Committee at IBG *139*

3.10.2. Ethical committee approval from Ethische Commissie Dierproeven at VUB *139*

4. RESULTS **140**

4.1. *In silico* approaches for prediction of aggregation prone residues **140**

4.2. Site-directed mutagenesis for generation of sdAb variants **154**

4.3. Optimizing the recombinant protein production yield **161**

4.3.1. Designing the mutations based on E.coli codon usage *161*

4.3.2. The choice of recombinant protein production host *162*

4.3.3. Transformation methods to the host cell *165*

4.3.4. Inoculation for starting culture *166*

4.3.5. The production vessel *166*

4.3.6. The production volume *167*

4.3.7. The boosting the production by altering temperature *169*

4.3.8. The production pH *170*

4.3.9. The production shaking rate *172*

4.3.10. The induction for target gene expression *173*

4.3.11. Boosting the target protein production yield by sucrose addition *174*

4.3.12. Harvesting recombinant protein from different cellular compartments *177*

4.3.13. Different agents for IMAC *179*

4.3.14. Choice of purification method *181*

4.4. Optimizing the recombinant protein production purity 181

4.4.1. Eliminating the non-specific interactions *181*

4.4.2. Unfold/refold *183*

4.4.3. Elimination of contaminant through chromatography techniques *186*

4.4.3.1. Purification with Protein A *186*

4.4.3.2. Purification by gel filtration *189*

4.4.4. Different reducing conditions *192*

4.5. Recombinant protein characterization 194

4.5.1. ELISA *194*

4.5.2. Thermal stability determination by protein thermal shift assay *196*

4.5.3. Chemical stability determination by GdmCl denaturation *201*

4.5.4. Aggregation kinetics *203*

4.5.5. High-performance liquid chromatography (HPLC) *206*

4.5.6. Isothermal titration calorimetry *211*

4.5.7. In vivo testing of Bevacizumab sdAb variants in zebrafish *213*

4.6. Computational methods to uncover distinctions between antibodies and nanobodies 222

4.6.1. Length variation in HCDRs *222*

4.6.2. Antigen binding scores of HCDRs *224*

4.6.3. Differences between residues at the same location *226*

4.6.4. Non-canonical cysteine bond profiles *231*

4.6.5. Stability enhancing salt bridge profiles around HCDR3 ..... *233*

4.6.6. Statistical analysis on physicochemical properties of FW residues *234*

4.6.7. Physicochemical characteristics of individual residues *237*

4.7. The host lab studies on abroad **240**

4.7.1. Preparation of functional camelised anti-VEGF sdAbs at home lab *241*

4.7.2. Ethical committee approval *245*

4.7.3. Re-design of anti-VEGF single domain antibodies *246*

4.7.3.1. Insert design *246*

4.7.3.2. Cloning and ligation *250*

4.7.3.3. Transformation and sequencing *251*

4.7.4. Redesigned recombinant sdAb production and purification by IMAC *251*

4.7.5. Optimisation of redesigned recombinant sdAb production and purification *252*

4.7.5.1. Isolation of proteins from different cellular compartments *252*

4.7.5.2. Utilization of different auto-induction media compositions *255*

4.7.5.3. Exploration of different E. coli strain types *256*

4.7.6. SPECT-CT imaging to assess whole body distribution at the macroscopic level *259*

4.7.6.1. Modification of sdAbs and Bevacizumab with a DOTA chelator *259*

4.6.6.2. Radiolabelling with radioisotope ^{111}In *260*

5. DISCUSSION **261**

6. CONCLUSION & FUTURE PERSPECTIVES **270**

7. REFERENCES **273**

8. ADDENDA **290**

8.1. Ethical committee approval report from Animal Experiments Local Ethics Committee (IBG-HAYDEK) at IBG **290**

8.2. Ethical Committee approval report from Ethische Commissie Dierproeven at VUB 291

8.3. CV in ARBIS format 292

ABBREVIATIONS

^{111}In Indium-111

^{125}I Iodine-125

2-MEA ..... 2-Mercaptoethylamine hydrochloride

^{99m}Tc Technetium-99m

Ab ..... Antibody

ADCC ..... Antibody-Dependent Cellular Cytotoxicity

AMP ..... Ampicillin

Amp^R ..... Ampicillin resistance

ATCC ..... American Type Culture Collection

ATP ..... Adenosine 5'-Triphosphate

AUC ..... Area Under the Curve

BSA ..... bovine serum albumin

BSA ..... Buried Surface Area

CAM ..... Chloramphenicol

CDR ..... Complementarity-Determining Region

C_H ..... Heavy chain constant domain

C_L ..... Light chain constant domain

CMIM ..... Cellular and Molecular Immunology

DMSO ..... Dimethyl sulfoxide

dNTP ..... deoxynucleotide

DOL ..... Degree of labelling

DOTA ..... 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid

dpf ..... days post-fertilization

DTT 1,4-Dithiothreitol

E. coli: ..... *Escherichia coli*

ECD ..... Ethische Commissie Dierproeven

ECM ..... Extracellular matrix

EDC ..... 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide

EDTA ..... Ethylenediaminetetraacetic acid

EGFP ..... Enhanced Green Fluorescent Protein

ELISA ..... Enzyme-Linked Immunosorbent Assay

EMBO.....European Molecular Biology Organization
Fab ..... Fragment, antigen-binding
FBS ..... Foetal Bovine Serum
Fc ..... Fragment, crystallizable
FcγR..... neonatal Fc receptors
FITC ..... Fluorescein isothiocyanate
Flt-1 ..... fms related receptor tyrosine kinase 1
FPLC..... Fast Protein Liquid Chromatography
Fv ..... Fragment, variable
FW ..... Framework
GdmCl ..... Guanidine hydrochloride
GFP..... Green Fluorescence Protein
GST..... Glutathione S-Transferase
HBSS ..... Hank's Balanced Salt Solution
HCDR ..... Heavy Chain Complementarity-Determining Region
hcIgGs..... heavy-chain-only IgGs
Hif-1α ..... Hypoxia-inducible factor-1 alpha
His..... Histidine
HPLC:..... High-Performance Liquid Chromatography
HRP ..... Horseradish Peroxidase
HT-29 ..... Human colourectal adenocarcinoma cells
HV ..... Hypervariable Region
IBG ..... İzmir Biomedicine and Genome Institute
IBG-BIP..... IBG Bioinformatics Platform
ICMI ..... *In vivo* Cellular and Molecular Imaging
Ig..... Immunoglobulin
IgNAR ..... İmmunoglobulin New Antigen Receptors
IMAC..... Immobilized Metal Affinity Chromatography
IPTG ..... Isopropyl β- d-1-thiogalactopyranoside
ITC..... Isothermal titration calorimetry
iTLC ..... instant Thin Layer Chromatography
IV ..... Intravenous

KDR ..... Kinase insert domain receptor
kV ..... kilovolt
LCDR ..... Light Chain Complementarity Determining Region
mAb ..... monoclonal antibody
mBq ..... mega Becquerel
MCS ..... Multiple Cloning Site
MIMA ..... Medical Imaging Core Facility
MMPs ..... Matrix metalloproteinases
ms ..... Millisecond
MW ..... molecular weight
NHS ..... N-hydroxysuccinimide
NRP ..... neuropilin
NSCLC ..... non-small cell lung cancer
OD ..... optical density
ori ..... origin of replication
PBS ..... Phosphate Buffered Saline
PDB ..... Protein Data Bank
pI ..... isoelectric point
PIGF ..... placenta growth factor
PVDF ..... polyvinylidene fluoride
QC ..... Quality control
RF ..... Random Forest
rpm ..... round per minute
RT ..... RT
scFv ..... single-chain fragment variable
sdAb ..... single-domain antibody
SDM ..... Site-Directed Mutagenesis
SDS ..... Sodium dodecyl sulfate
SEC ..... Size Exclusion Chromatography
SIV ..... Subintestinal Vessel/Vein
TAE ..... Tris Acetate
TB ..... Terrific Broth

TES .....Tris-EDTA-Sucrose

Tm..... Thermal melting

VEGF.....Vascular Endothelial Growth Factor

VEGFR .....Vascular Endothelial Growth Factor Receptor

VEGFR1 .....Flt1

VEGFR2 .....Flk1 or KDR

V_H .....Heavy chain variable domain

V_HH ..... Variable domain of heavy-chain antibody

VIF.....Variance Inflation Factor

V_L.....Light chain variable domain

VUB..... Vrije University Brussels

WB.....Western Blot

WT.....Wild Type

βME 2-Mercaptoethanol

LIST OF TABLES

Table 1. Antibodies. 31
Table 2. Bacterial strains. 31
Table 3. Bacterial media. 31
Table 4. Chemicals. 32
Table 5. Kits. 36
Table 6. PCR Reagents. 36
Table 7. Plasmids. 37
Table 8. Reagents and solutions. 37
Table 9. Resins. 38
Table 10. Other materials. 38
Table 11. Equipment. 39
Table 12. Instruments. 42
Table 13. Software. 44
Table 14. HT-29 cell properties. 45
Table 15. Cell culture reagents for HT-29 cells. 46
Table 16. List of forward and reverse primers used in the mutations introduced on Bevacizumab V_H sequence at the home lab. 48
Table 17. The forward and reverse primers used at the host lab for the targeting of Bevacizumab V_H sequences (WT and its variants) 49
Table 18. The reverse primer targeting MP57 region 50
Table 19. The details regarding in vivo studies of animal models 50
Table 20. Reagents for SDM reaction using QuikChange Lightning kit. 54
Table 21. SDM reaction steps using QuikChange Lightning kit. 54
Table 22. Reagents for SDM reaction using Q5 High-Fidelity kit. 55
Table 23. SDM reaction steps using Q5 High-Fidelity kit. 55
Table 24. Reagents for SDM reaction using KOD Hot Start kit. 56
Table 25. SDM reaction steps using KOD Hot Start kit. 56
Table 26. Reagents for DpnI digestion reaction using NEB DpnI kit. 57
Table 27. Reagents for DpnI digestion reaction using Thermo 10X Fast Digest kit. 58
Table 28. Preparation of 50X stock TAE Buffer and 1X TAE Buffer. 59
Table 29. Preparation of Agarose gel, 1% (w/v) with SYBR Safe. 59

Table 30. Preparation of SOC media and preceding SOB media. 62
Table 31. Preparation of LB Agar with AMP and CAM for electrocompetent cells. 63
Table 32. Preparation of LB broth. 63
Table 33. Preparation of 10% (v/v) glycerol. 64
Table 34. Preparation of 0.14 M β-ME. 66
Table 35. Preparation of Transformation Broth media 67
Table 36. Preparation of home-made auto-induction media. 71
Table 37. Preparation of Buffer A, Buffer B and stock solutions of imidazole and PBS. 73
Table 38. Preparation of 20X and 1XPBS. 74
Table 39. Preparation of Buffer A and Buffer A with DTT for sonication. 77
Table 40. Preparation of Buffer A with 5 mM MgATP and denaturated E.coli lysate. 79
Table 41. Reagents for amplification of the Bevacizumab V_H fragments by PCR. 83
Table 42. Reaction steps for amplification of the Bevacizumab V_H fragments by PCR. 84
Table 43. Reagents for pHen6 expression vector digestion using NcoI and BstEII enzymes. 84
Table 44. Reagents for ligation of vector pHen6 with Bevacizumab V_H fragments. 85
Table 45. LB Agar with 2% glucose and 100 µg/ml AMP for WK6 cells. 89
Table 46. Preparation of Terrific Broth (TB) production media for WK6 cells. 90
Table 47. Preparation of IPTG. 91
Table 48. Preparation of Tris-EDTA-Sucrose (TES) buffer. 91
Table 49. Preparation of TES/4 buffer. 92
Table 50. Detailed information on Bevacizumab V_H host-lab sdAb variants 94
Table 51. Preparation of 10X and 1X Purification Buffer. 95
Table 52. Preparation of Buffer A with 25 mM imidazolen and 6M GdmCl. 96
Table 53. Preparation of Buffer B with 6M GdmCl. 96
Table 54. Preparation of buffers for Protein A purification by FPLC. 98
Table 55. Preparation of APS for SDS-PAGE. 99
Table 56. Preparation of 12% resolving gel for SDS-PAGE. 99
Table 57. Preparation of 8% stacking gel for SDS-PAGE. 100
Table 58. Preparation of 10X and 1X running buffer for SDS-PAGE. 100
Table 59. Preparation of 10X and 1X transfer buffer for WB. 102
Table 60. Preparation of PBS-T and 5% blocking buffer for WB. 103
Table 61. Preparation of anti-Flag tag mouse antibody for WB. 104

Table 62. Preparation of HRP-conjugated antibody for WB. 104
Table 63. Preparation of mouse or human VEGF for coating of ELISA well plate. 105
Table 64. Preparation of 10% blocking buffer for ELISA. 106
Table 65. Preparation of PBS-T for ELISA. 107
Table 66. Preparation of Anti-Flag mouse antibody for ELISA. 108
Table 67. Preparation of HRP-Conjugated Goat Anti-mouse antibody for ELISA. 108
Table 68. Preparation of 2M H_2SO_4 solution for ELISA..... 109
Table 69. Preparation of protein thermal shift reactions for T_m determination. 110
Table 70. The set up details of thermal melt reaction 110
Table 71. Preparation of 8M GdmCl for chemical denaturation assay. 111
Table 72. Preparation of GdmCl ascending concentration range. 111
. Table 73. The run conditions of the plate reader for chemical denaturation assay. 112
Table 74. The experimental parameters of ITC. 113
Table 75. Preparation of 60X stock and 1X E3 medium for zebrafish studies. 114
Table 76. Preparation of anaesthetics for zebrafish studies..... 115
Table 77. Preparation of decolourization solution for zebrafish studies. 116
Table 78. Preparation of basic and low melting agarose for zebrafish studies. 116
Table 79. Preparation of paraformaldehyde for zebrafish studies. 117
Table 80. The groups indicated for the comparison and the total number of mice needed per group for whole body distribution studies based on the power analysis. 118
Table 81. Total number of mice bred for the investigation of tumour penetration of Bevacizumab-dervied sdAbs. 119
Table 82. Preparation of metal-free 10X and 1X Carbonate buffer. 124
Table 83. Preparation of metal-free 1M NaOH. 125
Table 84. Preparation of metal-free 1M HCl. 125
Table 85. Preparation of metal-free 0.2M Na_2CO_3. 125
Table 86. Preparation of metal-free 0.2M NH_4OAc, pH 5. 125
Table 87. Preparation of metal-free 1M NH_4OAc pH 7. 126
Table 88. IsoLink commercial kit contents for ^{99m}Tc-labelling. 128
Table 89. Definition of the constant and variable regions in V_H and V_HH sequences based on the Chothia numbering scheme 130

Table 90. Classification of physicochemical properties shown as single letter amino acid codes defined by EMBOSS Pepstats. 131
Table 91. Amino acid hydropathy classes in single letter amino acid codes. 132
Table 92. The work packages executed in accordance with the timeline of the thesis progression in the home lab. 134
Table 93. The work packages executed in accordance with the timeline of the project progression in the host lab. 135
Table 94. The list of V_H-V_L interface residues of Bevacizumab 140
Table 95. The predicted aggregation scores for the V_H sequence of Bevacizumab by Tango 2.2. 142
Table 96. The results of the Aggrescan 3D 2.0. 143
Table 97. The results of CamSol Intrinsic v2.2 predictions on aggregation propensity. 143
Table 98. CamSol Structurally Corrected predictions on aggregation prone residues. 145
Table 99. The aggregation-prone residues in Bevacizumab V_H sequence based on Aggrescan. 145
Table 100. List of mutational hotspots for aggregation prone residues by every *in silico* tool 148
Table 101. The suggested mutation candidates on Bevacizumab V_H based on general strategy to humanise nanobody research study 149
Table 102. The specific amino acid residues found within the FW-3 region of Bevacizumab and other antibodies from different sources at the same positions. 152
Table 103. The total list of possible sdAb variants of Bevacizumab V_H 153
Table 104. The total list of all sdAb variants of Bevacizumab V_H. 153
Table 105. The codon-optimised V_H peptide sequence of Bevacizumab was utilized, along with engineered mutations to create its associated variants. 154
Table 106. The codon-optimised V_H nucleotide sequences of Bevacizumab 156
Table 107. Bevacizumab V_H home-lab codon-optimised variants, their molecular weight, theoretical isoelectric points and extinction coefficient values calculated by ExPASy ProtParam 162
Table 108. The production concentration of every sdAb produced through BL21 pLysS E.coli cells. 162

Table 109. The effect of different type of Erlenmeyer flasks on recombinant protein production concentration 166
Table 110. The effect of recombinant protein production volume 167
Table 111. Promoting bacterial growth through different temperature conditions 169
Table 112. The influence of varying pH values on protein production concentrations 170
Table 113. The impact of varying the shaking rate on the concentrations 172
Table 114. The influence of distinct induction methods on recombinant protein production 174
Table 115. The impact of incorporating sucrose into the media 174
Table 116. The variation in concentrations of recombinant proteins sourced from different cellular compartments 177
Table 117. Various IMAC agents 179
Table 118. The effect of using different purification strategies 181
Table 119. The effect of different additives introduced to the media 182
Table 120. The compatibility analysis of Protein A application on the V_H of Bevacizumab sequence 186
Table 121. The fractions collected after FPLC and their corresponding pH and concentration values 188
Table 122. The fractions collected after SEC and their concentrations for sdAb WT. 190
Table 123. The fractions collected after SEC and their concentration values for sdAb V37Y/G44E/L45R/W47G. 191
Table 124. Various concentrations of β-mercaptoethanol and DTT were examined 193
Table 125. The apparent dissociation constant (K_d) for binding of Bevacizumab sdAb-WT and its variants to VEGF was calculated using the Hill1 function in the OriginPro 8.5.0 software 195
Table 126. The chemical unfolding determination of Bevacizumab WT and its variant sdAbs 202
Table 127. The aggregation kinetic value apparent rate constant (k_{app}) of sdAb WT and its variants 204
Table 128. The aggregation kinetic value apparent rate constant of Bevacizumab sdAb-WT and its variants 205
Table 129. The mean length of in V_Hs categorized by species 223
Table 130. The mean length of in V_HHs, categorized by species 223

Table 131. The total count of canonical and non-canonical disulphide bridges in V_H dataset 232
Table 132. The overall count of canonical and non-canonical disulphide bridges in V_HH dataset 232
Table 133. The total count of canonical salt bridges (R94 – D101) and non-canonical salt bridges in V_HS 233
Table 134. The total count of canonical and non-canonical salt bridges in V_HHs 234
Table 135. The logistic regression analysis, utilizing the EMBOSS Pepstats output, included all FW regions (FW1-4) collectively 235
Table 136. The concentration of the V37Y sdAb variant fractions collected after SEC 242
Table 137. List of host lab codon optimised Bevacizumab V_H fragment nucleotide sequences of redesigned sdAbs 247
Table 138. List of host lab codon optimised Bevacizumab V_H fragment peptide sequences of redesigned sdAbs 249
Table 139. The production concentration of every sdAb produced using WK6 E.coli cells. 252
Table 140. The optimisation studies on varying production media composition and production conditions 256
Table 141. Various *E. coli* cell lines transformed with different sdAb plasmids 256
Table 142. The protein concentration results obtained after optimisation studies on utilizing different E.coli strains 257

LIST OF FIGURES

Figure 1. Different types of immunoglobulins. 5
Figure 2. A basic structure of Immunoglobulin G. 6
Figure 3. The genetic configuration of an antibody structure. 8
Figure 4. Different antibody types found in Camelidae. 10
Figure 5. Various antibody types found in nature and engineered antibody fragments. 13
Figure 6. The 3D structure of a typical V_HH. 15
Figure 7. The comparison between conventional drug discovery and AI based automated discovery. 20
Figure 8. The hallmarks of cancer and associated therapeutic strategies. 21
Figure 9. Multi-stages of metastasis. 22
Figure 10. Different VEGF family ligands and their corresponding receptors. 24
Figure 11. The working mechanism of VEGF in tumour cells. 24
Figure 12. The target receptors of VEGF on endothelial cells. 26
Figure 13. The schematic representation of ^{111}In incorporated into DOTA chelator 28
Figure 14. HT-29 cells. 45
Figure 15. The plasmid map of pET-17b expression vector. 46
Figure 16. The map of pHen6 expression vector. 47
Figure 17. Variable heavy chain of codon optimised Bevacizumab amino acid and nucleotide sequences. 48
Figure 18. Schematic representation of the gene sequence of Bevacizumab V_H in expression plasmid pET-17b plasmid. 51
Figure 19. The workflow diagram of the sdAb production and purification applied in the home lab. 70
Figure 20. The sequence and the properties of the Bevacizumab V_H sequence targeting forward and reverse primers designed in the host lab. 83
Figure 21. The workflow of the host lab nanobody/sdAb expression and purification 88
Figure 22. The assembly of transfer sandwich for WB. 102
Figure 23. An ELISA plate alignment. 106
Figure 24. ELISA plate configuration for antibody sample concentration range. 107
Figure 25. A power analysis with α set at 0.05:11 (11 comparisons) = 0.0045. 118
Figure 26. Flowchart for work packages 5 and 6 137

Figure 27. The interfacing residues of Bevacizumab at the V_H-V_L fragments.....141
Figure 28. The predicted aggregation propensity score of V_H sequence of Bevacizumab by Pasta 2.0.....141
Figure 29. An evaluation of the overall aggregation tendency for the Bevacizumab V_H sequence based on Waltz 2.0 predictions. 142
Figure 30. V_H structure of Bevacizumab. 150
Figure 31. The former interfacing residues of Bevacizumab at the FW-2 of V_H. 151
Figure 32. The evolutionary trajectory of Bevacizumab V_H sequences during its discovery.152
Figure 33. 1% concentration agarose gel of sdAb L45R.....160
Figure 34. The SDS-PAGE results of sdAbs produced using BL21 pLysS E. coli cells. 164
Figure 35. The immunoblotting results of sdAbs produced using BL21 pLysS E. coli cells.164
Figure 36. Agar plates containing AMP+ and CAM+ antibiotics were inoculated with an equal quantity of sdAb plasmids. 165
Figure 37. SDS-PAGE and WB results of the different production flask comparisons of F70I/L72R/T74N/S77N/A79L sdAb sample. 168
Figure 38. The effect of change in pH on recombinant protein production quality. 171
Figure 39. The impact of sucrose incorporation into the production medium 176
Figure 40. The comparison of sdAb isolated from distinct cellular compartments. 178
Figure 41. The effect of using different IMAC agents 180
Figure 42. The impact of MgATP and E.coli cell lysate supplementation on the purity of leaky expression 183
Figure 43. The impact of dialysis on the unfolding and refolding 185
Figure 44. The results of FPLC using Protein A as resin 187
Figure 45. Protein A trial on F70I/L72R/T74N/S77N/A79L sdAb.....188
Figure 46. The SEC results of WT sdAb.....190
Figure 47. The SEC results for the V37Y/G44E/L45R/W47G sdAb,.....191
Figure 48. The fractions obtained after SEC separation analysed by SDS-PAGE.....192
Figure 49. The SDS-PAGE analysis performed by using different reducing agents (β-mercaptoethanol and DTT) at varying concentrations 193
Figure 50. The apparent dissociation constant (K_d) values of WT sdAb and its variants calculated based on ELISA.....195

Figure 51. The derivative melt curves of WT sdAb under varying conditions for T_m determination. 197
Figure 52. The derivative melt curve profiles of G44E sdAb under varying conditions for T_m determination. 197
Figure 53. The derivative melt curve profiles of L45R sdAb under varying conditions for T_m determination. 198
Figure 54. The derivative melt curve profiles of W47G sdAb 199
Figure 55. The melt curve profiles of V37Y sdAb 200
Figure 56. The melt curve profiles of V37F sdAb under varying conditions for T_m determination. 200
Figure 57. The Chemical stability assessment of Bevacizumab WT and its variant sdAbs by GdmCl denaturation. 202
Figure 58. The aggregation kinetics of Bevacizumab sdAb WT and its variants, L45R and W47G under heat stress conditions at A) 60°C and B) 65°C 203
Figure 59. The aggregation kinetics of Bevacizumab sdAb WT and its variants 204
Figure 60. The aggregation kinetics analysis of Bevacizumab sdAb WT and its variants 205
Figure 61. HPLC chromatogram of the purity analysis of G44E/L45R sdAb. 206
Figure 62. Chromatogram of the purity analysis of G44E/W47G sdAb assessed using HPLC. 207
Figure 63. Chromatogram of the purity analysis of G44E/L45R/W47G sdAb via HPLC..... 207
Figure 64. Chromatogram depicting the purity analysis of V37F/G44E/L45R/W47G sdAb assessed via HPLC. 208
Figure 65. Chromatogram of the purity analysis of V37Y/G44E/L45R/W47G sdAb assessed by HPLC. 208
Figure 66. SDS-PAGE results of sdAb 9 and sdAb 10. 208
Figure 67. Chromatogram showing the purity analysis of V37F/G44E/L45R/W47G and G44E/L45R/W47G sdAbs assessed using HPLC. 209
Figure 68. Chromatographic analysis of V37Y sdAb aggregation kinetics via HPLC. 210
Figure 69. Chromatogram showing the aggregation behaviour of sdAb 9 assessed using HPLC. 210
Figure 70. Chromatogram of the aggregation behaviour of sdAb 10 assessed using HPLC. 211
Figure 71. ITC thermogram plot of V37Y. 212

Figure 72. ITC titration plot of V37Y. 212
Figure 73. Microinjection, visualization and imaging of zebrafish Tg(fli1:EGFP)^{y1} strain embryos. 213
Figure 74. Optimization studies for zebrafish. 214
Figure 75. Confocal microscopy images depicting the GFP-labelled zebrafish eye after a 48-h treatment with Bevacizumab in E3 media in the fish swimming water, captured at 25X magnification. 215
Figure 76. Zebrafish embryos treated for 48 h with Bevacizumab in E3 media in the fish swimming water were analysed for SIV development. 216
Figure 77. The Bevacizumab treatment administered through SIV injection, utilizing varying concentrations of Bevacizumab. 218
Figure 78. The confocal microscopy images displaying zebrafish injected with A) PBS, B) sdAb WT and C) sdAb V37Y at SIV area. 220
Figure 79. Analysis of zebrafish embryonic vasculature alterations following microinjection with the tested sdAbs. 221
Figure 80. The variation in HCDR lengths for HCDR1-3 in both V_Hs and V_HHs 223
Figure 81. The heatmaps of the antigen binding prediction scores of HCDR1-3 regions 225
Figure 82. The graphical logo representation of the multiple sequence alignment of complete sequence datasets 227
Figure 83. Graphical logo representations for multiple sequences of complete V_H dataset 229
Figure 84. Graphical representation for multiple sequences of the V_HH datasets 230
Figure 85. Canonical and non-canonical disulphide bond profiles 231
Figure 86. A box-and-whiskers plot to show the Random Forest – Boruta algorithm ranking of the variable importance scores 236
Figure 87. The box-and-whiskers plot of the ranking of variable importance scores for each amino acid–region association based on the Random Forest – Boruta algorithm 237
Figure 88. Heatmaps of variations in physicochemical characteristics among residues in V_Hs and V_HHs 240
Figure 89. The SDS-PAGE and WB analysis of V37Y sdAb after individually optimizing it in terms of purity and yield. 241
Figure 90. SEC analysis of V37Y sdAb variant performed in the home lab at IBG, Turkey. 242
Figure 91. SDS-PAGE analysis of V37Y sdAb after SEC performed in home lab at IBG. 243

Figure 92. The SEC run of the V37Y sdAb was performed in the host laboratory at VUB. . 244
Figure 93. The results of SDS-PAGE and WB analyses of the V37Y sdAb subsequent to an additional SEC procedure performed in the host lab at VUB 245
Figure 94. Insert design of Bevacizumab-derived, codon optimised WT V_H fragment..... 246
Figure 95. The pairwise sequence alignment and comparison of the new and prior WT sdAb designs 247
Figure 96. 1% agarose gel electrophoresis to visualize the sdAbs after PCR amplification.. 251
Figure 97. The effect of isolating redesigned sdAbs from different cellular compartments.. 253
Figure 98. The SDS-PAGE results of the redesigned sdAb 2 (upper) and redesigned sdAb 8 (lower) in the gel..... 254
Figure 99. The WB results of A) sdAb 2 (upper) and B) sdAb 8 (lower)..... 255
Figure 100. The SDS-PAGE results of the redesigned sdAb 1, sdAb 3 and sdAb 4 isolated from different cellular compartments..... 255
Figure 101. The effect of different induction media and different E.coli strains on redesigned sdAb yield and purity 258
Figure 102. The analysis on the Lysine residues in the heavy and light chains of Bevacizumab 260
Figure 103. The evaluation of the incorporation of ^{111}In into the Bevacizumab-DOTA by iTLC analysis..... 260

LIST OF EQUATIONS

Equation 1. Fluorescent dye concentration calculation based on Lambert-Beer law.....122
Equation 2. sdAb concentration calculation based on Lambert-Beer law.....123
Equation 3. Calculation of the degree of labelling estimation. 123
Equation 4. The Binary Logistic Regression model.....132

ACKNOWLEDGEMENTS

For their academic and moral contributions for the completion of this thesis, I would like to extend my sincerest gratitude to the following individuals and organizations:

To my supervisor, Prof. Dr. Gökhan Karakülah for his guidance and supportive mentorship regarding the field of computational biology.

To the co-advisor Dr. Sibel Kalyoncu-Uzunlar for her mentorship regarding the field of antibody engineering,

To Prof. Dr. Esin Firuzan for her valuable insights in stastistical analysis and her contributions throughout the publication progression.

To the all members of my thesis defence jury for their contributions during the discussion sessions to improve quality of this thesis.

I am grateful to my former supervisor, Asst. Prof. Ralph Meuwissen, for his invaluable teachings in experimental animal models, in which the insights I gained guided me during the design of my experimental animal xenograft model.

All members of the İzmir Biomedicine and Genome Center and İzmir International Biomedicine and Genome Institute, past and present, for giving me the opportunity and the environment to conduct my studies.

My past interns who helped with the side work, easing the burden of preparing the solutions and especially to Derya Göksu Helvacı for his assistance during my experiments. To the members of İnan Lab for their generosity of exchanging materials. I appreciate Emine Çolak for always finding a way to fit my media in autoclave and therefore speeding up my experimental timeline.

With the expertise in the field, Prof. Dr. Sophie Hernot, for her continuous support and contribution to abroad studies. She presented me many opportunities at Vrije Universiteit Brussel (VUB), Brussel, Belgium. Thanks to her invaluable help, I participate in a wide array of projects, and I acquired a diverse range of new techniques and skills. I had the honour of collaborating with Prof. Dr. Nick Devoogdt, an expert in the field of nanobody design, engineering, and production, who provided me instrumental guidance and expertise in his field. Presenting at the same conference with Prof. Dr. Serge Muyldermans (who pioneered the discovery of nanobodies at VUB in ‘93) and having the privilege of sharing the same lab where nanobodies all unfolded were an exceptional honour and an unforgettable experience.

The Fancy Fluorescence team members Lukasz Mateusiak, Noemi Declerck and Dora Mugoli Chigoho for their companionship both in experiments and friendship outside the lab. The post-docs at VUB, Dr. Timo De Groof and Dr. Jessica Bridoux for their consultation whenever I needed it, and Jessica's purple stuffed alpaca, Schaschlik, to cheer up everybody. Lab manager Dr. Janik Puttermans for his help to arrange everything. Research assistances Dr. Isabel Remory and Dr. Sonja van den Block for their interpretation of the experimental results. Nanobody core facility technician Jan De Jonge for his great support and help in preparing everything in advance for me for nanobody the production studies. Kevin de Jonghe for his extreme talent and speed in animal dissection and help in handling radioactivity studies. And all members of Laboratory for *In Vivo* Cellular and Molecular Imaging, ICMI-BEFY at VUB.

I would like to thank to YÖK 100/2000 (Translational Medicine), EMBO (Short Term Fellowship), TUBITAK 2211-C (BIDEB), TUBITAK 2214-A (BIDEB), TUBITAK 3501 (Project No: 119Z161), TUBITAK 1004 (Project No: 20AG006) and TUBITAK 2250 (BIDEB) programs for supporting me throughout my PhD studies.

My large and loving family -my parents, and extended relatives- who have been supporting and encouraging me throughout this journey. Their belief in my abilities and their sacrifices played an important role in my academic success. I am deeply grateful for their endless love, patience, and motivation, which have helped me overcome challenges and achieve my goals.

Last but certainly not least, I would like to express my heartfelt gratitude to my spouse, Doğa Eskier. Throughout this challenging period, he has been my source of support, understanding, and strength. His boundless belief in my abilities has been very precious. His unwavering support has made this thesis possible. I am deeply grateful for his love and understanding along the way and standing by my side and being my greatest source of inspiration. Finally, the shining star of my life, our baby boy, I am so glad to be with you during writing of this thesis. I am truly blessed to have you in my life.

Hereby, I dedicate this thesis to my beloved father who can watch over me all the time from the brightest skies above.

Nazlı Eda Eskier

2024

ABSTRACT

DEVELOPMENT OF ANTI-CANCER SINGLE-DOMAIN ANTIBODY TOOLBOX

Nazlı Eda Eskier, İzmir International Biomedicine and Genome Institute, Dokuz Eylul University Health Campus, Balcova 35340 - Izmir / TURKEY

nazli.eda.kaleli@gmail.com

Most antibody therapeutics are based on full-length Immunoglobulin G (IgG) format, but antibody fragments have gained attention over the last decade due to their smaller size, better tissue penetration, augmented affinity and stability, enhanced target specificity and ease of engineering. Variable heavy chain (V_H) is the smallest functional antibody fragment. Although V_H does not exist without variable light chain (V_L) in human, *Camelidae* naturally have such antibodies. Removal of V_L from human IgG scaffold would cause exposure of hydrophobic residues on V_H-V_L interface, resulting in undesired biophysical properties. By exploiting the camelid V_H features through camelisation approach, it would be possible to obtain single-domain Antibodies (sdAbs) with improved properties using traditional IgGs as scaffold. Our aims are engineering an anti- Vascular Endothelial Growth Factor (VEGF) commercial IgG, Bevacizumab (Avastin®), based on camelid V_H (V_HH) at framework-2 (FW-2) residues and also FW-3 residues using humanisation approaches and determining physico-biochemical properties of the resulting sdAbs.

Aggregation-prone, FW-2 residues were mutated into corresponding camelid residues through camelisation approaches. Additionally, FW-3 residues were altered via humanisation to eliminate possible immunogenicity problems caused by mouse-originated residues. The recombinant protein expression was done in bacteria under novel auto-induction conditions. Anti-VEGF sdAbs were purified through affinity chromatography and biophysically characterized *in vivo* and *in vitro*. The camelisation and humanisation based mutations have showed better biophysical properties in terms of yield, stability and aggregation/solubility. Interestingly, affinity of engineered sdAbs has also substantially increased.

This work shows that sdAb format of a humanised antibody can be generated successfully through engineering approaches to be used as next-generation therapeutics/diagnostics.

Keywords: single-domain antibody, antibody engineering, camelisation, humanisation, framework-2, framework-3

ÖZET

ANTİ-KANSER TEK DOMAİNLİ ANTİKOR TAKIMLARININ GELİŞTİRİLMESİ

Nazlı Eda Eskier, İzmir Uluslararası Biyotıp ve Genom Enstitüsü, Dokuz Eylül Üniversitesi Sağlık Kampüsü, Balçova 35340 - İzmir / Türkiye

nazli.eda.kaleli@gmail.com

Çoğu antikor terapötikleri tam boy Immunoglobülin G (IgG) formatına dayanırken, son on yılda antikor parçacıkları, daha küçük boyutları, daha iyi doku penetrasyonu, artırılmış bağlanma gücü ve stabilitesi, geliştirilmiş hedef spesifikliği ve kolay mühendisliği nedeniyle dikkat çekmektedir. Değişken ağır zincir (V_H), en küçük işlevsel antikor parçacığıdır. İnsanlarda V_H, değişken hafif zincir (V_L) olmadan var olmazken, *Camelidae* türleri doğal olarak bu tür antikorlara sahiptir. V_L'nin insan IgG yapısından çıkarılması, V_H-V_L ara yüzünde hidrofobik rezidülerin açığa çıkmasına ve istenmeyen biyofiziksel özelliklere yol açar. Devegillerin V_H özelliklerinden yararlanarak kamelizasyon yaklaşımıyla, geleneksel IgG'yi, iyileştirilmiş özelliklere sahip tek alan antikorlara (sdAb) dönüştürmek mümkün olabilir. Amaçlarımız, antikor mühendisliğiyle devegil V_H'sine (V_HH) dayalı olarak anti-Vasküler Endotelyal Büyüme Faktörü (VEGF) ticari IgG'si olan Bevacizumab (Avastin®) antikorunu, çerçeve-2 (FW-2) rezidülerini ve çerçeve-3 (FW-3) insanlaştırma yaklaşımlarını kullanarak mühendislik yapmak ve elde edilen sdAb'lerin fiziko-biyokimyasal özelliklerini belirlemektir.

Agregasyon eğilimli FW-2 rezidüleri, kamelization yaklaşımlarıyla karşılık gelen devegil rezidülerine dönüştürülmüştür. Ek olarak, fare kökenli rezidüler nedeniyle olası immünojenisite sorunlarını ortadan kaldırmak için FW-3 rezidüleri humanisasyon yoluyla değiştirilmiştir. Rekombinant protein üretimi, özgün oto-indüksiyon koşulları altında bakterilerde gerçekleştirilmiştir. Anti-VEGF sdAb'ler, afinite kromatografisi ile saflaştırılmış ve *in vivo* ve *in vitro* olarak karakterize edilmiştir. Kamelization ve humanisasyona dayalı mutasyonlar, verimlilik, stabilite ve kümeleşme/aggregasyon açısından daha iyi biyofiziksel özellikler göstermiştir. İlginç bir şekilde, antikor mühendisliği uygulanan sdAb'lerinin afinitesi de önemli ölçüde artmıştır.

Bu çalışma, insanlaştırılmış bir antikorun sdAb formatının mühendislik yaklaşımlarıyla başarıyla oluşturulabileceğini ve yeni nesil terapötikler/diyagnostikler olarak kullanılabileceğini göstermektedir.

Anahtar kelimeler: tek alan antikoru, antikor mühendisliği, kamelization, humanisasyon, çerçeve-2, çerçeve-3

1. INTRODUCTION AND AIM

1.1. Statement and importance of the problem

Most antibody therapeutics are based on full-length IgG (Immunoglobulin G) format, but antibody fragments have gained attention over the last decade due to their smaller size, better tissue penetration, augmented affinity and stability, enhanced target specificity and ease of engineering. Variable heavy chain (V_H) is the smallest functional antibody fragment. Although V_H does not exist without variable light chain (V_L) in human, *Camelidae* naturally have such antibodies.

Removal of V_L from human IgG scaffold would cause exposure of hydrophobic residues on V_H-V_L interface, resulting in undesired biophysical properties. By exploiting the camelid V_H properties through camelisation approach, it may be possible to obtain single-domain antibodies (sdAbs) with improved physico-chemical properties.

The utilization of hybridoma technology plays a role, in the development of novel antibody therapeutics. By combining B lymphocytes that produce antibodies with immortalized myeloma cells hybridoma technology enables the production of monoclonal antibodies (mAbs) that exhibit specificity and potency. These mAbs have revolutionized disease treatment by offering targeted therapy with side effects and improved effectiveness. However one limitation is that mAbs derived from hybridomas with mouse origins can cause immunogenicity concerns, in humans. Despite this obstacle ongoing research endeavors to address immunogenicity issues and advance the field of antibody therapeutics towards medicine and tailored treatments.

1.2. Aim of the study

In this study, the main objective was to create sdAb utilizing Bevacizumab V_H (Avastin®) with improved properties via antibody engineering techniques. We aimed to engineer the Bevacizumab V_H by incorporating V_HH hallmark residues into the Framework 2 (FW 2) region of the antibody through camelisation.

Initially the V_H sequence was derived from Avastin®, a well characterized commercial drug, targeting VEGF. It was originally developed in mice, and later underwent CDR grafting, where certain residues were converted back to humanised form. Subsequently some of these humanised residues were also changed back, to their original mouse residues. The final step involved transforming Bevacizumab into a humanised Fab format by incorporating CDR grafting.

To address immunogenicity concerns associated with mouse derived residues we also focused on modifying the FW-3 region through a humanisation approach. Ultimately our goal was to evaluate the physico-chemical properties of the resulting sdAbs, whether they were subdue their original state.

The methods applied in this study for the development, can be generalized for antibodies with different targets.

2. BACKGROUND INFROMATION & LITERATURE

2.1. Antibodies

Antibodies, also known as immunoglobulins (Igs), are pivotal players in our adaptive immune system, with remarkable specificity and versatility in recognizing and neutralizing foreign materials. These proteins are produced by B lymphocytes and can be classified into subtypes: IgG, IgM, IgA, IgD, and IgE, where each has distinct structure and functions (Schroeder and Cavacini 2010) (**Figure 1**).

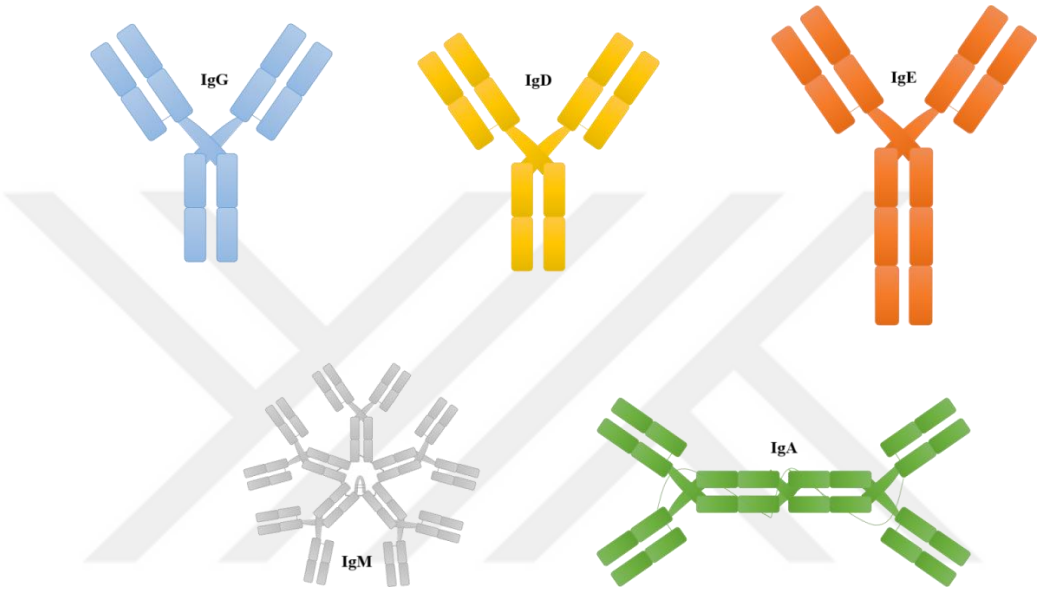


Figure 1. Different types of immunoglobulins. Naturally, IgG, IgD, and IgE are found as monomers, where IgM assembles as pentamer and IgA has two monomers joined by a secretory protein linker.

IgG, the most abundant antibody in human, provides long-term protection by recognizing and binding to a diverse array of pathogens by adaptive immune system. The abundance of IgG makes it a pivotal component of the human immune response, allowing for the recognition and neutralization of various agents. There are four IgG subclasses (IgG1, IgG2, IgG3, and IgG4) having diverse functions in immune defence, such as opsonization (marking pathogens for destruction), activation of complement system, and mediation of antibody-dependent cellular cytotoxicity (ADCC) (Vidarsson, Dekkers, and Rispens 2014a).

IgD is a membrane-bound antibody primarily found on the surface of mature B-cells. IgD acts as a receptor for antigens, helping to initiate the immune response by recognizing and binding to foreign substances. Eventhough its low levels in human sera, it is an important player in B-cell activation and immune surveillance (Gutzeit, Chen, and Cerutti 2018).

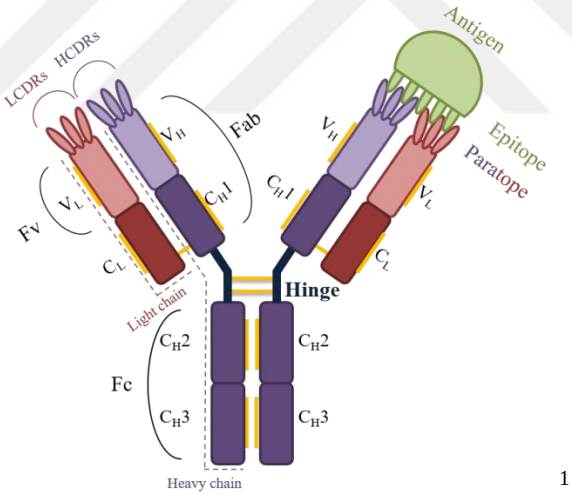
IgE plays role in allergic responses and defence against parasitic infections. It is present in trace amounts in the sera. IgE triggers the release of inflammatory mediators, such as histamines, in response to allergens. In case of parasitic infections, IgE helps to initiate an immune response against them (Galli and Tsai 2012).

IgM serves as an early line of defence, and produced during the initial stages of an immune response. Its pentameric arrangement enhances its binding capacity, whereby facilitating efficient pathogen recognition (Miyazaki et al. 2018).

IgA is primarily found in mucosal tissues and secretions and defends against pathogens at entry points of the respiratory and gastrointestinal tracts. IgA are divided into two subclasses, IgA1 and IgA2 (de Sousa-Pereira and Woof 2019).

2.1.1 Immunoglobulin Gs

IgGs consist of two heavy chains and two light hains, with the light chain existing in two variants: κ or λ chains (Vidarsson, Dekkers, and Rispens 2014b) (**Figure 2**). The fineger-like extensions of these chains forms the antigen-binding sites, allowing antibodies to recognize and bind to their target antigens (Schroeder and Cavacini 2010).



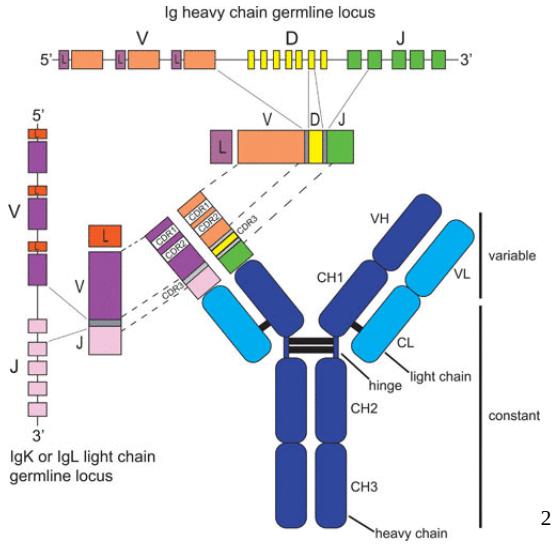
[1]

Figure 2. A basic structure of Immunoglobulin G. Hetero-dimeric Y-shaped scaffold is made up of two light (25 kDa each) and two heavy chains (50 kDa each) which are joined by a total of four inter-chain disulphide bonds (yellow lines): one in between each heavy (purple) and light chain (pink) and two in between flexible hinge region (dark blue lines) connecting two monovalent structures together. The intra-chain disulphide bonds are present in each domain to rigidify overall structure. Light chain consists of one constant (C_L) and one variable domain

[1] (Karadag et al. 2020)

(V_L). Heavy chain has one variable (V_H) and three constant regions (C_H1-3), where C_H1 is joined to C_H2 via hinge region. Fv (fragment, variable) stands for the antigen engaging portion of heavy and light chain, where Fab (fragment, antigen binding) also includes constant region of V_L and first constant region of V_H in addition to Fv. Fc (fragment, crystallizable) constitutes the bottom part of the typical IgG structure, where C_H2 and C_H3 are located. The finger-like protruding extensions are target antigen binding portions of heavy and light chains. These N-terminal variable domain protrusions are called as paratopes made out of hypervariable polypeptide segments of complementarity determining regions (CDRs) found on an antibody. There are particularly three distinct CDRs on the tip of each light and heavy chain forming loops to fit into crevices on the antigen, meanwhile determining the specificity of the antibody.

In a bivalent IgG structure, there are six LCDRs on light chain and six HCDRs on heavy chain in total. Epitopes are the corresponding regions of paratopes on the antigen that are responsible for antibody binding (Chiu et al. 2019; Huber 1980). The generation of the both heavy chain and light chain regions of IgG involves the contribution of multiple genes; the variable (V), the diversity (D), and the joining (J) genes, where the process known as known as V(D)J recombination (**Figure 3**).The heavy chain regions tend to exhibit greater diversity with an expanded repertoire compared to the light chain regions due to differences in gene segments usage (Ye et al. 2013). This rearrangement introduces variability particularly in the CDR3 loop, which greatly influences antigen specificity. In the light chain, the variable domain is encoded only by a combination of the V and J genes, which results in a limited repertoire (Collins and Watson 2018).



[2]

Figure 3. The genetic configuration of an antibody structure. The IgG heavy chain locus and κ or λ light chain locus are shown. V(D)J genes are presented on germline locus. Disulphide bridges in the full antibody structure are represented by black line segments.*2.1.2. Alternative antibodies in nature*

Amongst antibody types, IgGs are the most commonly found ones in humans (Townsend et al. 2016). However, there are certain diversities in the nature; different species possess different antibody varieties to gain benefit for survival. The *Camelidae* family (camels, llamas, alpacas, and vicunas) possesses the ability to produce diverse antibody varieties, including both conventional IgGs similar to those found in humans and heavy-chain-only IgGs (hcIgGs), achieved through somatic V(D)J hyper-mutations (Muyldermans et al. 1994). The *Elasmobranchii* family (sharks and cartiligious fishes) also can naturally produce a different type of non-canonical antibodies known as immunoglobulin new antigen receptors (IgNARs) (Greenberg et al. 1995).

2.1.2.1. The emergence of alternative antibodies

While various theories have been proposed to elucidate the exact underlying reason of presence of alternating antibodies in nature, it was postulated that historically critical events play crutial role in the emergence of those antibodies. These events are still not unfolded, yet the camelids and cartilaginous fishes have come up with a solution to cope with prior extreme pressure with adaptive changes, and finally such incidents have naturally paved a way to fully functional alternative antibodies; hcIgGs and IgNARs.

[2] Adapted from (Boyd and Joshi 2014)

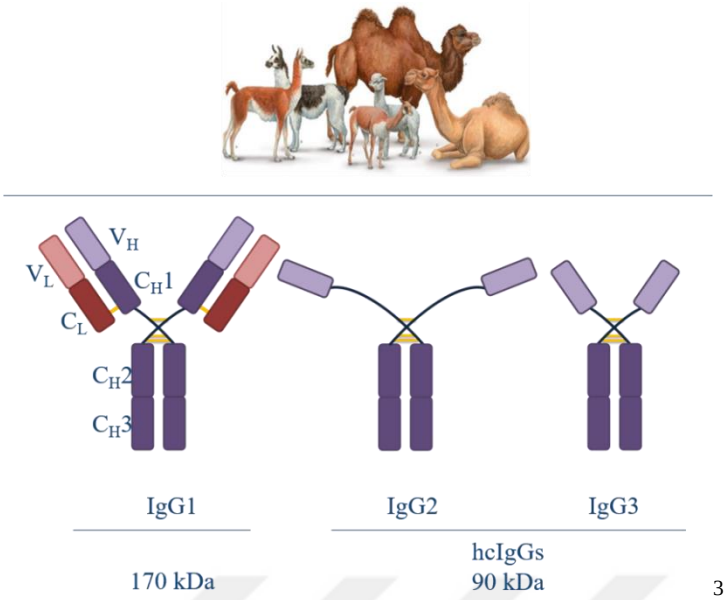
The emergence of the alternative antibodies is a relatively recent event in Camelids compared to the Elasmobranchs. IgNARs have arose more than 220 million years ago (Lee et al. 2000), whereas hcIgGs genes have diverged about 25 million years ago (Conrath et al. 2003). These alternative antibodies have been developed throughout the evolution and their presence is attributed to the adaptations of the certain changes in the environment (Bouaouda et al. 2014). Although the exact driving force behind these alternative antibodies creation is still unknown, there are some theories that explain the loss of light chain. It was speculated that, during their evolution of these animals, they may have encountered severe threats to their survival. The reason could be a microorganism, maybe against those stress factors, light chain was more susceptible to impairment. In that case, the loss of light chain is a reasonable compromise of immune system to benefit for the organism (Flajnik, Deschacht, and Muyldermans 2011).

Sharks have elevated concentrations of urea in their blood to preclude water loss caused by osmotic pressure. Urea is a well-known denaturant, therefore urea enrichment could be an environmental pressing factor for sharks to have such alternative antibodies as an evolutionary advantage (Feige et al. 2014).

Animals living in desert environments possess unique adaptations to the conditions with extreme heat and water deprivation (Bouaouda et al. 2014). Heat stress is the key issue affecting the thermo-stability of antibody structures. Under these harsh conditions, heavy chain antibodies may have emerged from need of adaptation for smaller, heat-resistant antibodies that could also compensate the heterodimeric binding nature of IgGs with better binding properties reaching into smaller epitope crevices (De Genst, Silence, et al. 2006).

2.1.2.2. Heavy-chain only antibodies

Camelid-derived hcIgGs differ from conventional antibodies as they lack light chains and constant regions in their heavy chains. Instead, they have an extended hinge region (**Figure 4**), which provides them with physical flexibility when binding to antigens. Both hcIgGs and heterodimeric canonical IgGs are present in their antibody repertoire. During B-cell maturation, they even share the same germline H-chain genes -(D)J-. However, the relative abundance of IgG1 is higher in the serum of *Camelidae* (De Genst, Saerens, et al. 2006).



[3]

Figure 4. Different antibody types found in *Camelidae*. IgG1 (170 kDa) shares significant homology to human counterparts in IgGs (150 kDa), based on structure and sequence. hcIgGs (IgG2 and IgG3) (90 kDa), have longer hinge regions than IgG1 (Hamers-Casterman et al. 1993). (light orange: variable light chain (V_L), light blue: variable heavy chain (V_H), dark orange: constant region of light chain (C_L), dark blue: constant regions of heavy chain (C_H1-3), yellow lines: disulphide bonds, black arcs: hinge regions).

Camelid hcIgGs (90 kDa) exhibit a closer resemblance to human IgGs (150 kDa) compared to IgNARs (175 kDa) in cartilaginous fishes. In their homo-dimeric scaffold, the heavy chains of hcIgGs (V_HHs) display 80% sequence homology and structural similarity to human IgGs V_H counterparts compared to IgNARs (Conrath et al. 2001), whereas vNAR show about 30% sequence identity to those found in humans (Kovalenko et al. 2013).

Several other modifications are necessary to adopt a homo-dimeric scaffold. One striking difference is the removal of the constant region from the H-chain (C_H1), which formerly served as a binding site for the L-chain. The elongated hinge region confers flexibility to the overall structure upon antigen binding (Nguyen et al. 2002). The framework-2 (FW-2) region, located at the former interaction site at the V_H-V_L, harbours unorthodox charged and polar residues to prevent association with other hcIgG molecules. This hydrophilic framework renders hcIgGs with superior stability and solubility compared to their

[3] Adapted from Eskier, NE; Eskier, D; Firuzan, E; and Kalyoncu, S (2023) "Physicochemical differences between camelid single-domain antibodies and mammalian antibodies," *Turkish Journal of Biology*: Vol. 47: No. 6, Article 6.

human IgG counterparts (Goldman et al. 2017). Notably, these FW-2 regions exhibit a conserved motif in both IgGs and hcIgGs in terms of structure and sequence. The solubility-enhancing residues in FW-2 ensure structural stability even in the absence of the V_L.

hcIgGs are assembled through recombination of V-(D)-J gene segments, and further somatic hyper-mutations increase their variability, resulting in an enriched antigen binding repertoire (Li et al. 2016). There are unique adaptations of hcIgGs to compensate for the absence of the light chain, such as improved solubility and resistance to aggregation (Muyldermans 2001). Despite the hetero-dimeric structure of IgGs featuring six different CDRs involved in antigen association (HCDR1-3, LCDR1-3), hcIgGs exclusively exhibit HCDRs only. The conformation of HCDR1 and HCDR2 in hcIgGs does not always align with the canonical IgG HCDR1 formation, as certain residues in these regions influence the hydrophobic core packing of the overall structure. To compensate for the absence of light chain epitope recognition, hcIgGs recruit additional residues in the HCDRs 1 and 3 regions, resulting in longer loops compared to IgGs (Muyldermans et al. 1994). To maintain the stability of these elongated loops, inner disulphide bonds are introduced between HCDRs, as further structural adaptation to the hcIgG format. These non-canonical disulphide bonds enhance stability and reduce entropic folding penalties (Mendoza et al. 2020). Previous study by Chothia *et al*. (1989) had demonstrated the superior stability of V_H over V_H-V_L by comparing the hypervariable regions of IgGs (Chothia et al. 1989). Furthermore, hcIgGs have shown the ability to target epitopes that are minimally antigenic to conventional IgGs (Lauwereys et al. 1998). Such comprehensive studies on IgNARs have been relatively limited compared to hcIgGs (Zielonka et al. 2015). IgNARs exhibit unique epitopes, characterized by the presence of two HCDRs 1 and 3, and two hypervariable regions (HV2, HV4) in their structure (Kovaleva et al. 2014). The diversity of HCDR3 subtypes and genetic rearrangements (V-D1-D2-D3-J) contributes to their heterogeneity, distinguishing IgNARs from IgGs (Diaz et al. 2002). Additionally, IgNARs display response kinetics that more closely resemble IgM rather than IgGs (Sharrock and Sun 2020). In contrast, hcIgGs have gained considerable attention in biomedical research, diagnostics, and therapeutics with their homology and compatibility with human IgGs (Belanger et al. 2019).

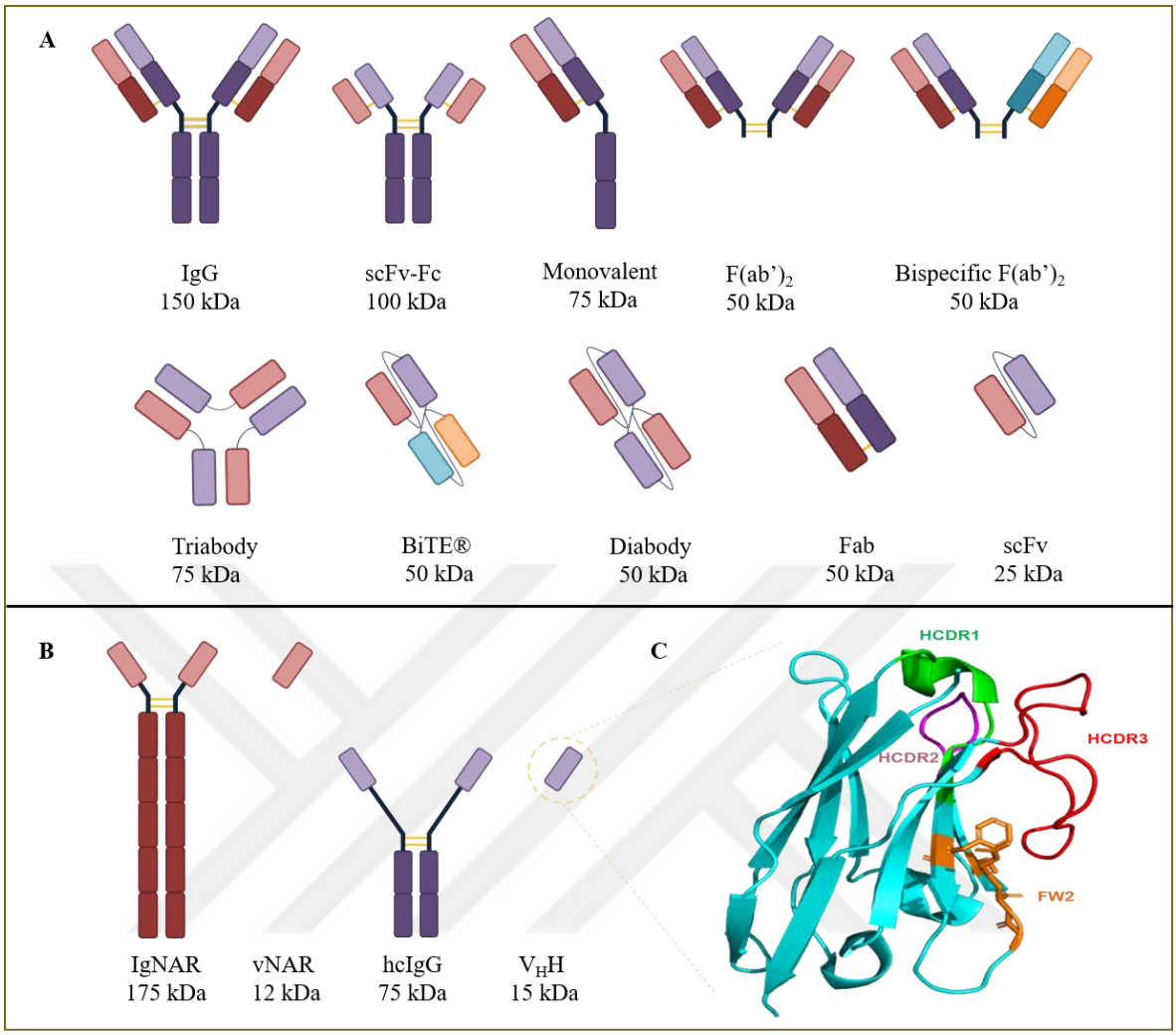
2.1.3. Antibody fragments

While antibodies are commonly used for the design of targeted therapeutic or diagnostic agents, antibody-derived fragments are considered to possess improved

pharmacokinetic proporties. There are many different strategies have been established by many successful teams, all around the world to optimize the biophysical and biochemical characteristics of antibodies to retain specificity while remedying the other properties such as solubility, stability and aggregation-resistance, while converting them into smaller fragments (Tiller and Tessier 2015). Even though a significant progression has been made on the immediate technologies that increase the developability profile of the candidate antibodies at the earlier stages, yet the current work concentrates on mainly monoclonal antibodies (mAbs).

Using antibody engineering, fragments are designed by selectively eliminating some or all of the constant regions while retaining the essential antigen binding portion. These fragments exhibit similar binding properties to their full-length counterparts, often with enhanced biophysical characteristics (Bradbury et al. 2011). Advantages of antibody fragments include: (i) reduced immunogenicity stemming from the absence of constant regions, (ii) enhanced tumour penetration into cryptic regions inaccessible to IgGs, (iii) cost-effective and scalable production using microbial systems, and (iv) the availability of diverse *in vitro* screening technologies for optimisations (Nelson 2010; Bannas, Hambach, and Koch-Nolte 2017). Furthermore, protein engineering and fusion techniques enable the diversification of antibody fragments, leading to the generation of novel formats (**Figure 5**). Among these next-generation fragments, Nanobodies®[4] have emerged as a particularly prominent format in antibody discovery and engineering. Antibody-based imaging offers a sensitive approach to visualize, characterize, and quantify molecular and cellular processes in a non-invasive manner (Warram et al. 2014). Antibody fragments also possess remarkable capabilities for recognizing specific molecules, making them valuable in cancer imaging. Overcoming the challenge of heterogeneous antibody distribution within tumours following systemic administration has led to increased exploration of smaller antibody-derived constructs to optimize pharmacokinetic profiles (Thurber, Schmidt, and Wittrup 2008).

[4] Sanofi's affiliate Ablynx N.V. is the holder worldwide of the NANOBODY® trademark. (https://www.sanofi.com/en/magazine/our-science/nanobody-technology-platform)



[5]

Figure 5. Various antibody types found in nature and engineered antibody fragments. **A)** Full-length IgG, its associated modified antibody fragments and assorted engineered antibody fragments by fusion techniques augmented through antibody engineering techniques (disulphide bonds: yellow, linker: blue) and their approximate molecular weights. **B)** Heavy-chain only antibodies of *Elasmobranchii* (pink) and *Camelidae* (purple) and their associated single-domain antigen-binding fragments. **C)** 3-D structure of a typical camel nanobody (PDB ID: 5U65). Hypervariable HCDR loops and hallmark residues were highlighted; HCDR1 in green, HCDR2 in purple, HCDR3 in red, and FW-2 in yellow, with hallmark residues shown in stick format.

Numerous studies have elucidated the diagnostic and therapeutic potential of single-domain antibody fragments (Wesolowski et al. 2009) originated from different species; sharks (Cheong et al. 2020; English, Hong, and Ho 2020) and members of *Camelidae* (De

[5] (Karadag et al. 2020)

Meyer, Muyldermans, and Depicker 2014; Yang and Shah 2020). The implications of these findings suggest a potential future wherein these engineered antibody fragments may supplant conventional mAbs. Further research is required to explore deeper regarding these promising antibody fragments (Yang and Shah 2020). By delving deeper into their unique characteristics and refining their engineering, their full potential can be unlocked and pave the way for a new era in antibody-based diagnostics and therapeutics.

2.1.3.1. Nanobodies

In last decade, studies demonstrated that sdAbs derived by engineering from hcIgGs found in *Camelidae* (Hamers-Casterman et al. 1993), also known as nanobodies, possess ideal characteristics to act as molecular probes, both for diagnostics and therapeutics (Vaneycken et al. 2011). Nanobodies, derived from the antigen binding portion of hcIgGs known as V_HH, represent a breakthrough in the field of antibody-based therapeutics (Arbabi-Ghahroudi 2017). These molecules retain their full functionality despite their small size (15 kDa, 4 nm x 2.5 nm size), in stark contrast to conventional hcIgGs weighing approximately 90 kDa (Van Audenhove and Gettemans 2016). This compact size enables nanobodies to penetrate tissues more rapidly and homogeneously, enhancing their potential as molecular probes for diagnostics and therapeutics.

The framework of V_HH is completely hydrophilic, in contrast to human which has a hydrophobic side that faces V_L. In V_HH, the corresponding interface region is hydrophilic, which contributes to its exceptional stability and solubility compared to V_H (Goldman et al. 2017). One of the notable structural distinctions between V_H and V_HH lies in the hallmark residues found at positions 42nd, 49th, 50th, and 52nd according to the Kabat numbering (**Figure 6**). Particularly, residues at 49th and 50th positions consistently consist of charged residues in V_HH, which results in a reduction of the exposed hydrophobic surface area and imparts a hydrophilic nature to the overall structure. These differences contribute to the unique properties of V_HH (Vincke et al. 2009).

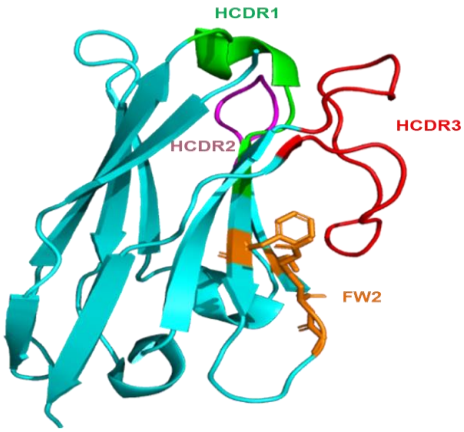


Figure 6. The 3D structure of a typical V_HH. (PDB ID: 5U75)[6]. The solvent exposed hydrophobic V_H-V_L interface residues are shown. The highlighted FW-2 hallmark are exhibiting the fundamental difference in between a single-domained V_HH and their counterpart human V_H at FW-2 region. Hypervariable HCDR loops and hallmark residues were highlighted; HCDR1 in green, HCDR2 in purple, HCDR3 in red, and FW-2 in yellow, with hallmark residues shown in stick format.

Nanobodies have demonstrated significant potential in many different fields including research and the biomedical science. They have shown promise in therapeutics, serving as tools for targeted therapies (Yang and Shah 2020). For diagnostics, nanobodies have emerged as powerful tools, enabling the development of highly sensitive and specific diagnostic assays (De Meyer, Muyldermans, and Depicker 2014). Their unique properties also make them well-suited for applications in imaging, facilitating the visualization of specific targets with high precision (Schumacher et al. 2018; Warram et al. 2014). Moreover, nanobodies are used in agriculture, in crop protection and improvement (Warram et al. 2014). The versatility and broad applicability of nanobodies reflects their potential in diverse scientific disciplines.

Antibodies are typically too large to be eliminated through urinal excretion, as they cannot pass through the filtration system of the kidneys. Instead, their elimination pathway primarily involves proteolytic degradation mediated by endocytosis (Waldmann, Strober, and Mogielnicki 1972). This process is facilitated by the interaction between antibodies and neonatal Fc receptors (FcγR) found on effector immune cells. The binding of the constant portions of antibodies to FcγR triggers cellular internalization, followed by lysosomal degradation. This mechanism influences the circulation time of antibodies in the bloodstream, which is attributed to their half-life (Pyzik et al. 2015). In contrast, nanobodies, with their

[6] (Langley et al. 2017)

smaller size, can undergo renal clearance, being excreted through the kidneys, where this aspect makes them good candidates in diagnostics field. Although total accumulation in tumour tissue is lower than mAbs, significantly higher tumour-to-background ratios are obtained. This implies that high contrast images can be obtained early after injection for diagnosis (D'Huyvetter et al. 2014) and low nonspecific toxicity is expected when nanobodies are labeled with cytotoxic drugs or therapeutic radioisotopes (Xavier et al. 2013). Because of their smaller size, nanobodies are able to penetrate more rapidly and homogenously within tumour tissue (Debie et al. 2020).

A notable example of their therapeutic efficacy is Caplacizumab, a nanobody-based drug approved for the treatment of acquired thrombotic thrombocytopenic purpura (Peyvandi et al. 2016). The exceptional characteristics and versatile applications of nanobodies make them a promising avenue for next-generation antibody-based therapies.

2.1.3.2. Single-domain antibodies

sdAbs are the smallest functional antigen binding fragments that can be also derived from antibodies (Li et al. 2020). By applying a camelisation approach based on the mutation of framework residues found on former V_H-V_L interface (Tanha et al. 2001), it is possible to generate sdAbs with improved characteristics for *in vivo* applications, by using traditional IgGs as scaffold. sdAbs are distinguished from nanobodies: Nanobodies are derived from the hcIgGs found in camelids like llamas and camels, in contrast, sdAbs are modified versions of conventional mAbs, typically comprising a single protein domain that retains the antigen-binding capability (Arbabi-Ghahroudi 2022).

One notable feature of sdAbs is their enhanced stability compared to IgGs. The single-domain format eliminates the need for inter-chain disulphide bonds, reducing the likelihood of protein misfolding or degradation. Consequently, sdAbs exhibit robust resistance to harsh conditions, including extremes of pH, high temperatures, and proteolytic enzymes. This stability aspects render them highly suitable for diagnostic assays and imaging techniques, and prolonged shelf-life (Wu, Jiang, and Ying 2017). Moreover, sdAbs possess exceptional tissue-penetration properties due to their small size and ability to access cryptic epitopes on protein surfaces (Liu et al. 2022). By leveraging these advantages, sdAbs have potential success in targeting challenging intracellular or membrane-bound proteins.

The versatility of sdAbs is further underscored by their potential on engineering. Through phage display libraries, researchers can generate diverse sets of sdAbs with

improved binding specificities and affinities (Verheesen and Laeremans 2012). Additionally, by harnessing this potential, it is possible to construct bi-specific or multi-specific antibodies, where single-domains can be fused to different antibody fragments or functional moieties (Krah et al. 2016). Taking into account to all advantages offered by sdAbs, they surpass from nanobodies and conventional antibodies, providing distinct biochemical advantages and expanding possibilities in diverse fields (Joseph et al. 2023).

2.1.4. Engineering approaches for functional sdAb generation

2.1.4.1. Camelisation

The generation of novel nanobodies typically involves immunizing a camelid with the target antigen. However, to expedite and bypass this process, an alternative strategy known as "camelising" sdAbs derived from human antibodies can be employed to obtain functional sdAbs. Additionally, this strategy is in line with EU-directive 2010/63/EU on the protection of animals used for scientific purposes by being an animal-free method to generate new and/or improved targeting molecules (Marinou and Dontas 2023), as it offers an animal-free method for generating new and improved molecules.

The camelisation procedure aims mutating four hydrophobic FW-2 residues that typically interact with the V_L domain. This deliberate modification enhances water solubility and mitigates the aggregation tendencies of conventional IgG-derived sdAbs, leading to more stable and functional molecules. By harnessing "camelisation" technique, the production of sdAbs can be accelerated while adhering to ethical considerations and expanding the repertoire of target-specific molecules for various diagnostic and therapeutic applications (Vincke et al. 2009).

2.1.4.2. Humanisation

mAbs have revolutionized the field of therapeutics with their specificity and efficacy in targeting molecules. Humanisation strategies have been employed to modify the non-human components of IgGs, ensuring their compatibility with the human immune system. These strategies are particularly crucial for antibodies produced through hybridoma technology, where antibodies are initially generated in mice (Quinteros et al. 2017).

Hybridoma technology, a common method for antibody production, involves fusing murine B-cells with immortalized myeloma cells. This technique allows for the generation of high-affinity antibodies; however, it introduces murine-based amino acid residues into the antibody (Mitra and Tomar 2021). When hybridoma-derived antibodies are administered to

humans, they may exhibit immunogenic responses. To address this limitation, humanisation techniques are applied to convert the murine residues back to their human counterparts, thereby attaining lower immunogenicity (Hansel et al. 2010).

Humanisation can be achieved through various techniques, including framework shuffling and CDR grafting ('6 - Variable chain engineering – humanization and optimization approaches' 2012; Choi et al. 2015). Framework shuffling involves exchanging murine framework regions with their human counterparts (Dall'Acqua et al. 2005), while CDR grafting focuses on transferring the CDRs responsible for antigen recognition from the murine antibody to a human framework (Kim and Hong 2012).

Humanisation of antibodies produced through hybridoma technology is crucial for expanding the therapeutic potential of mAbs. By lowering the immunogenicity, humanised antibodies exhibit improved safety profiles and increased efficacy. One notable example of humanisation via framework shuffling is Bevacizumab, a therapeutic antibody used in cancer treatment. Bevacizumab initially derived from a murine antibody, and then underwent engineering processing in the Fab region, to replace murine residues with human counterparts (Ferrara, Hillan, and Novotny 2005). These modifications have significantly reduced immunogenicity profile of the antibody, making it more suitable for human use.

Advancements in antibody engineering using computational approaches and structural biology hold promise for further enhancing the humanisation for optimisation of therapeutic outcomes (Lu et al. 2020).

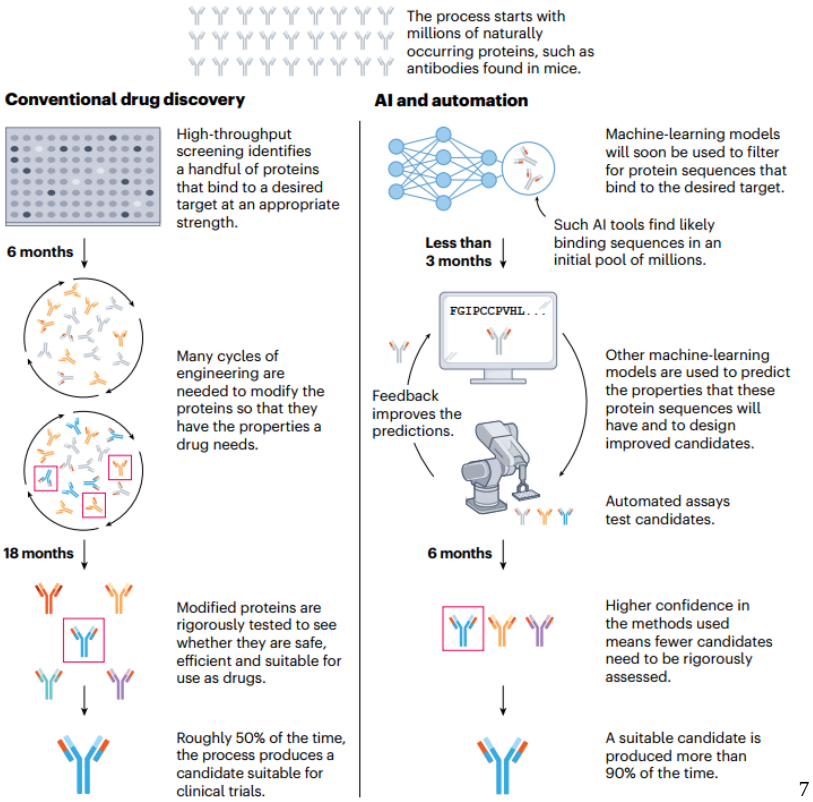
2.1.5. Antibody discovery

Antibody discovery has undergone a remarkable progression, by technological advancements that significantly impact the therapeutic and diagnostic agent development. Traditionally, antibody discovery through hybridoma technology could take several months to over a year, including the immunization of animals, cell fusion, and screening steps (Mitra and Tomar 2021). Modern antibody discovery encompasses high-throughput screening techniques, phage display libraries, and advanced *in vitro* selection methods (Lu et al. 2020). The integration of recombinant DNA technology and synthetic biology has enabled the creation of diverse antibody repertoires with improved affinies (Frenzel, Schirrmann, and Hust 2016). These advancements in antibody discovery speed up the identification of novel therapeutic targets.

2.1.5.1. Computational approaches in antibody discovery

Considering the challenges of experimental approaches, the integration of *in silico* approaches to antibody discovery has emerged as a revolutionary improvement. This change in computational analyses allows for a comprehensive exploration of antibody repertoires involving structural differences (Sormanni, Aprile, and Vendruscolo 2018). By using advanced bioinformatics tools and predictive homology modelling, *in silico* approaches empower researchers to unravel significant differences between traditional IgGs and nanobodies. This methodical exploration not only provides a detailed understanding of key residues at specific locations but also aids the design and optimisation of sdAbs (Hoey, Eom, and Horn 2019). In essence, the *in silico* approaches become a powerful tool for uncovering novel insights, facilitating the strategic engineering of antibodies with improved functionality and specificity.

Recently, advancements in computational approaches have expedited the antibody discovery timeline. Some *in silico* tools even boast timelines on the order of weeks, especially when they are used during antibody design, synthetic biology, and computational modelling (**Figure 7**) (Hummer, Abanades, and Deane 2022). However, other variable factors are playing role in the determination of this timeline duration, such as the chosen methodology, target complexity, and the specific requirements of the intended application (Shepard et al. 2017).



[7]

Figure 7. The comparison between conventional drug discovery and AI based automated discovery.

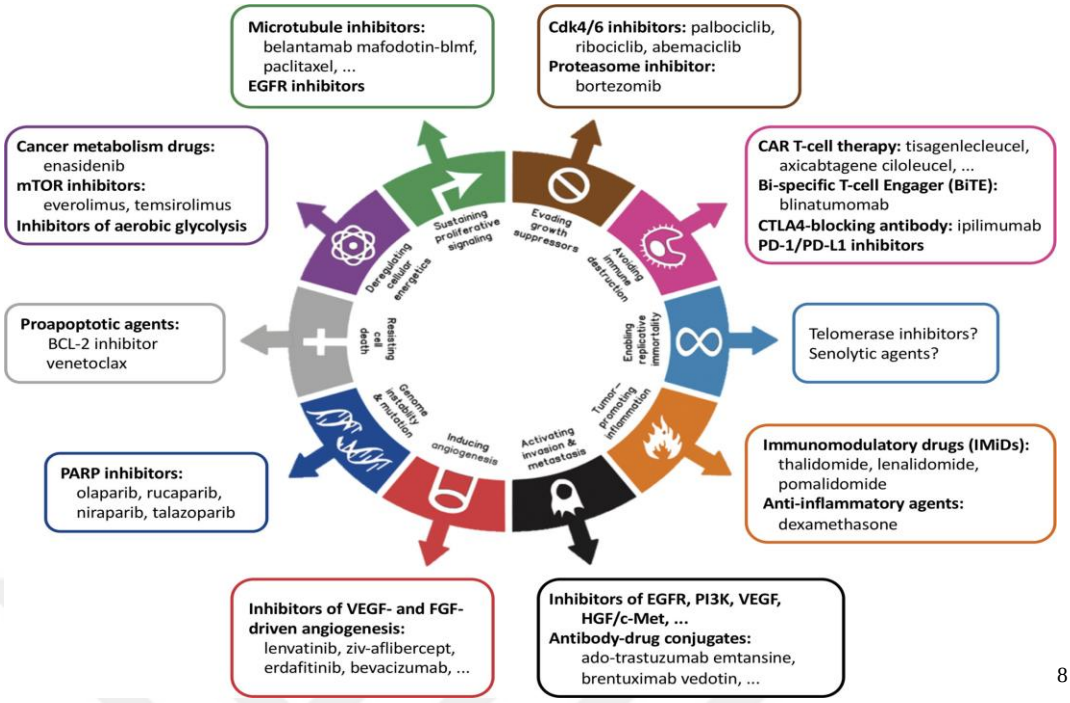
2.2. Cancer

Cancer is a complex disease characterized by the accumulation of genetic and epigenetic alterations that confer distinct capabilities to tumour cells (Hanahan and Weinberg 2000).

2.2.1. Hallmarks of cancer

The concept of hallmarks of cancer (**Figure 8**), as proposed by Weinberg, revolutionized the understanding of the fundamental biological changes that occur in cancer cells and drive tumour progression (Hanahan and Weinberg 2000). Understanding these hallmarks is crucial for developing effective diagnostic and therapeutic strategies.

[7]: Adapted from (Mock et al. 2023)



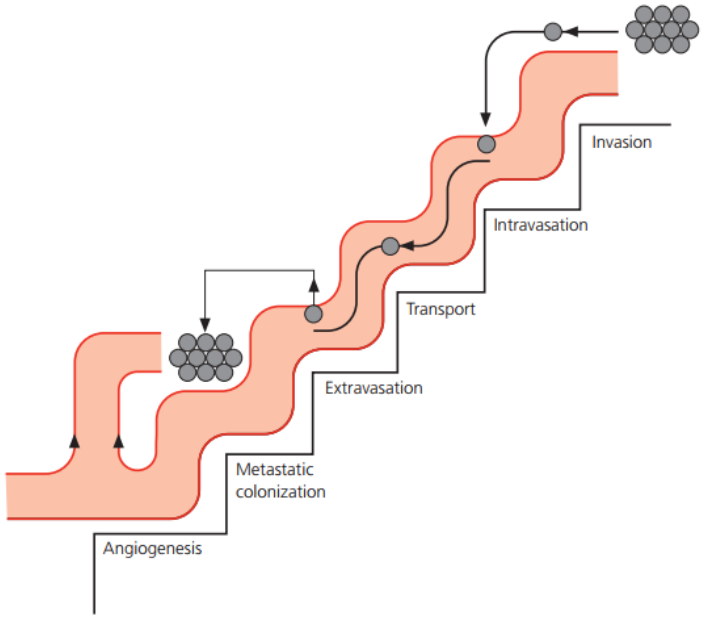
[8]

Figure 8. The hallmarks of cancer and associated therapeutic strategies. 1. Restoring growth suppressors: cyclin-dependent kinase inhibitors are used to restore or activate genes/pathways that suppress excessive cell growth whereby inhibiting cancer cell proliferation. 2. Enhancing immune response: immunotherapies or immune checkpoint inhibitors are used to activate or boost the immune system to recognize and destroy cancer cells. 3. Limiting replicative potential: telomerase or telomere pathway inhibitors are targeted to prevent replicative immortality of cancer cells. 4. Anti-inflammatory treatments: anti-inflammatory drugs or immunomodulatory agents are used in reducing tumour-promoting inflammation within the tumour microenvironment. 5. Preventing invasion and metastasis: agents targeting associated signalling pathways are used for inhibiting the invasive properties of cancer cells. 6. Inhibiting angiogenesis: vasculature inhibitory agents to block the formation of new blood vessels are utilized. 7. DNA damage and repair: DNA-damaging agents or inhibitors of DNA repair mechanisms are used to sustain genomic stability of cancer cells. 8. Promoting cell death: apoptosis inducing agents are used to eliminate cancer cells. 9. Modulating cellular metabolism: aerobic glycolysis inhibitors are used in altered metabolic pathways of cancer cells, to disrupt their energy production and hinder tumour growth. 10. Targeting proliferative signalling: microtubule inhibitors are used for inhibiting aberrant signalling pathways leading uncontrolled cell division and proliferation in cancer cells (Hanahan and Weinberg 2011).

[8] Adapted from (Dupont et al. 2021).

2.2.2. Metastasis

Metastasis, the spread of cancer cells from the primary tumour to distant sites, is a complex and multi-stage process involving a series of molecular players. It is a critical aspect of cancer progression that significantly contributes to morbidity and mortality (Guan 2015). Metastasis involves a sequence of steps, each governed by a distinct network of molecular interactions and signalling pathways (**Figure 9**) (Seyfried and Huysentruyt 2013; van Zijl, Krupitza, and Mikulits 2011).



[9]

Figure 9. Multi-stages of metastasis. These steps include the detachment of cancer cells from the primary tumour, invasion into surrounding tissues, intravasation into blood or lymphatic vessels, survival during transport through the circulation, extravasation into distant organs, and ultimately, the establishment and growth of secondary tumour. To maintain the successful establishment of a metastatic colony on a distant site of the body, the metastatic cancer cells initiate the formation of new blood vessels to supply nutrients and oxygen to the newly established colony by process called sprouting from the pre-existing blood vessels.

Metastatic colonization of cells from the primary tumour can occur individually or in clusters, often facilitated by interactions with blood components such as platelets. These cells can travel either in blood stream (Paget 1889) or lymphatic system (Brown et al. 2018) circulating to reach distant sites in the body. However, for successful metastatic colonization to take place, the establishment of neovascularization through the sprouting of new blood

[9] Adapted from (Pecorino 2012)

vessels from pre-existing ones is essential (Massagué and Obenauf 2016). This angiogenic process is crucial for ensuring the of oxygen and nutrients supply for the sustenance and growth of newly formed distant colonies (Saman et al. 2020).

2.2.2.1. Angiogenesis

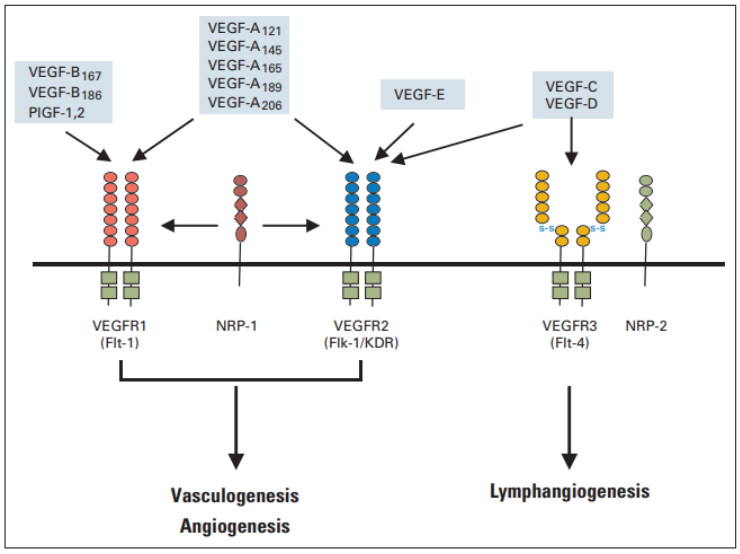
Angiogenesis is the rate determining stage of metastatic colonisation (Bielenberg and Zetter 2015). Without angiogenesis, tumours are unable to grow beyond a few millimetres in size (Nishida et al. 2006). The term "angiogenesis" was coined by Judah Folkman in 1971 to explain that tumour growth relies on the formation of new blood vessels from pre-existing ones (Saman et al. 2020). The transition from a non-angiogenic state to an angiogenic state, known as the angiogenic switch induces metastasis. This switch occurs due to an imbalance between positive and negative regulators of angiogenesis, resulting in the up-regulation of endothelial growth factors and down-regulation of angiogenesis inhibitors (Lugano, Ramachandran, and Dimberg 2020). Among these factors, vascular endothelial growth factor (VEGF) plays a key role as a mitogenic agent.

2.2.2.1.1. Vascular endothelial growth factor

The VEGF family consists of six growth factors (VEGF-A to E, and placental growth factor- (PlGF)) and three receptors (VEGFR-1 to 3) (**Figure 10**) (Ferrara 2002). VEGF-A is a homo-dimeric, disulphide-bound glycoprotein, which binds to VEGFR-1 and VEGFR-2 tyrosine-kinase receptors causing increased permeability, migration, and proliferation of endothelial cells (Peach et al. 2018).

VEGF-A exhibits multiple isoforms generated via alternative gene splicing. Among the various isoforms of VEGF, $VEGF_{165}$ has been primarily associated with the formation of new blood vessels (Shinkaruk et al. 2003). VEGF plays a crucial role in various biological processes, including cell survival, migration, and differentiation. Its role in promoting angiogenesis has been extensively studied and characterized (Platania et al. 2015).

VEGF promotes the development of tumour blood vessels, which can contribute to increased vascular permeability in malignant tumours (**Figure 11**). The expression level of VEGF has been found to correlate with poor cancer prognosis and diagnosis (Presta et al. 1997).



[10]

Figure 10. Different VEGF family ligands and their corresponding receptors. The VEGF family comprises seven ligands; VEGF-A, -B, -C, -D, -E, PlGF-1, and PlGF-2. NRP: neuropilin) and three receptors; VEGFR 1-3. Only two of these receptors are related with initiation of angiogenesis.

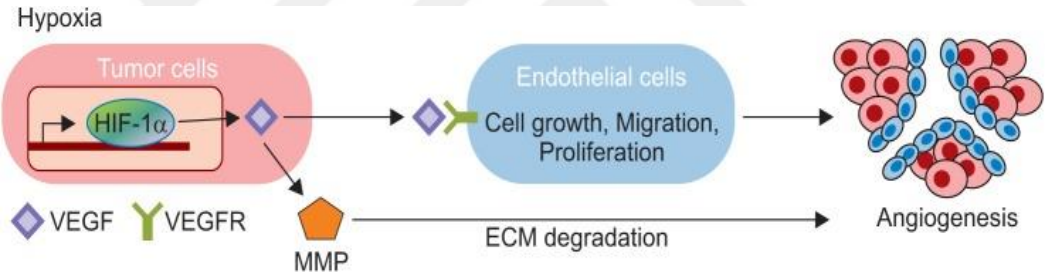


Figure 11. The working mechanism of VEGF in tumour cells. The hypoxic conditions within metastatic tumour environment trigger the induction of Hif-1α (Hypoxia-Inducible Factor-1 alpha), which is a transcription factor that detects low oxygen levels. In return, Hif-α induces the expression of VEGF. VEGF exerts its effects through various mechanisms. Firstly, it regulates matrix metalloproteinases (MMPs), enzymes responsible for degrading the extracellular matrix (ECM) surrounding cells. This degradation of the ECM facilitates the remodelling of tissues and enables blood vessel growth into the tumour. Secondly, VEGF directly influences the behaviour of endothelial cells lining the inner surface of blood vessels. It promotes their growth, migration, and proliferation, thereby facilitating the formation of new blood vessels (Lee et al. 2015).

[10] Adapted from (Hicklin and Ellis 2005)

2.2.2.1.1.1. VEGF in the body

VEGF can be found in different components of the blood: plasma, platelets, and leukocytes. Although tumours have high VEGF concentrations, their size is relatively small compared to other tissues that contain VEGF (Shibuya 2011), such as skeletal muscle, harbour a significant reservoir of VEGF (Kut, Mac Gabhann, and Popel 2007). Therefore, tumours are not the primary source of VEGF in the body, even in the case of cancer. This has important implications for the development of cancer therapeutics, to selectively target cancer-cell associated VEGF, considering the presence of other VEGF reservoir of the body. The previous findings indicate that, although myocytes produce the primary source of VEGF in muscle, this VEGF proves ineffective in promoting angiogenesis (Wagner 2011).

2.2.2.1.1.2. VEGF in cancer therapy and diagnosis

The recent studies have shown the role of VEGF in the mobilization of endothelial cells from the bone marrow to distant sites and promoting neovascularization (Chung, Lee, and Ferrara 2010). Understanding the intricate mechanisms of the VEGF pathway in metastasis holds great potential for the development of targeted therapeutic strategies.

Targeting VEGF alone may not be adequate to effectively target all types and stages of tumours. The complexity of the tumour environment also contributes to neovascularization (Ranieri et al. 2006). Therefore, the diverse factors and signalling pathways involved in tumour angiogenesis must be comprehensively considered to effectively inhibit tumour growth and metastasis. Combination therapies targeting multiple angiogenic pathways can achieve better outcomes in cancer treatment (Lugano, Ramachandran, and Dimberg 2020).

2.3. Bevacizumab

Bevacizumab (Avastin®; Genentech) is a humanised mAb that targets VEGF-A (Kazazi-Hyseni, Beijnen, and Schellens 2010). It has been widely used in the treatment of various cancers, including anal carcinoma (Kusumanto et al. 2003), lymphoma (Salven, Orpana, and Joensuu 1999), lung cancer (Kishiro et al. 2002), gastric carcinoma (Yoshikawa et al. 2000), ovarian cancer (Kraft et al. 1999), renal cell (Ljungberg et al. 2003), brain tumour (Stockhammer et al. 2000), hepatocellular carcinoma (Poon et al. 2001), breast cancer (Bando et al. 2005), prostate cancer (Li et al. 2005) and colorectal cancer (Haraguchi et al. 2002). Bevacizumab inhibits neovascularization in angiogenesis by binding to VEGF-A. Its generally administrated combined with other chemotherapy agents to enhance its therapeutic effects (Kazazi-Hyseni, Beijnen, and Schellens 2010).

2.3.1. Bevacizumab working mechanism

Bevacizumab functions by binding to VEGF-A and subsequently impairing its interaction with its receptors, Flt-1 (fms related receptor tyrosine kinase 1/ VEGFR-1) and KDR (Kinase insert domain receptor/ VEGFR-2), located on the surface of endothelial cells (**Figure 12**). In a preclinical study, it was shown that the administration of Bevacizumab resulted in a reduction of vascular growth and inhibition of metastatic colon cancer progression (Kanno et al. 2000). In clinical trials with non-small cell lung cancer (NSCLC), Bevacizumab demonstrated significant therapeutic impact when combined with standard chemotherapy, prolonging overall survival and progression-free survival (Russo et al. 2017).

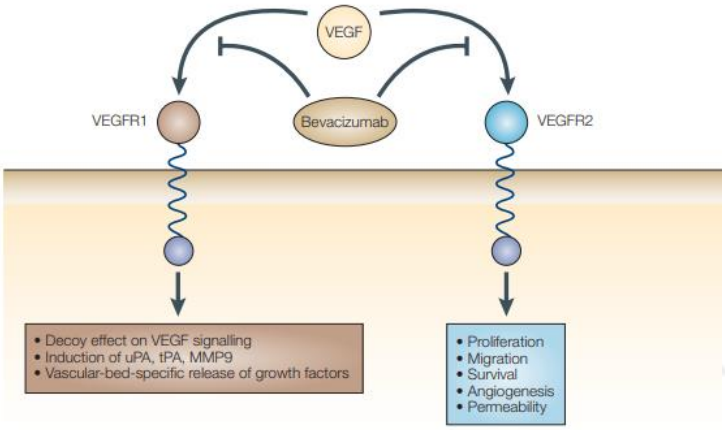


Figure 12. The target receptors of VEGF on endothelial cells. VEGF receptors, VEGFR-1, and -2, are found on the surface of blood endothelial cells.

Combination approaches in cancer therapy take advantage of the potential synergy between cytotoxic agents and antiangiogenic compounds. One proposed mechanism for this synergy is the normalization of tumour blood vessels, which are often tortuous due to the imbalance of proangiogenic and antiangiogenic factors. Anti-VEGF therapy is thought to promote vascular normalization, resulting in a reduction of interstitial hypertension within the tumour microenvironment. This normalization process improves the delivery of large therapeutic molecules to solid tumours, enhancing the efficacy of treatment (Gerber and Ferrara 2005).

2.3.2. Molecular imaging of tumour angiogenesis

Cancer cells are able to produce isforms via alternative splicing. The most commonly encountered VEGF-A isoforms are $VEGF_{121}$ and $VEGF_{165}$, whereas $VEGF_{121}$ is freely soluble, and $VEGF_{165}$ is partly secreted but also remains localized within the ECM and cell

surface. These isoforms present attractive imaging targets for *in vivo* VEGF imaging, offering insights into the measurement of VEGF levels in a non-invasive way and therefore give information about the eligibity of a patient to anti-angiogenic therapy (Cai and Chen 2008).

2.3.2.1. Bevacizumab in angiogenesis labelling

VEGF/VEGFR-targeted molecular imaging represents a novel approach to evaluate the effectiveness of antiangiogenic cancer therapy. It has the potential to enhance cancer management by providing valuable insights into the role and modulation of VEGF signalling in cancer development and intervention. This imaging technique offers a new paradigm for assessing the efficacy of antiangiogenic therapy and holds promise in advancing the understanding of VEGF pathway-related processes in cancer (Cai and Chen 2008).

The use of Bevacizumab is widespread in the treatment of metastatic colorectal cancer, NSCLC, and glioblastoma (Garcia et al. 2020). However, this mAb has limitations, including limited tumour penetration and significant side effects, primarily due to its large size. When it comes to diagnostic purposes, radiolabelled derivatives of Bevacizumab are inadequate for non-invasive monitoring of antiangiogenic treatment because they have prolonged blood circulation time and low tumour penetration (Niu and Chen 2011). Recognizing these challenges, there is a growing emphasis on the development of smaller antibody fragments, which hold promise for improved tumour penetration and reduced systemic side effects.

2.4. Antibody labelling

Antibody labelling is a fundamental technique used for visualization of specific molecules. By attaching a label to antibodies, molecules of interest can be targeted and identified with precision. Various antibody labelling methods have been employed using fluorescent dyes, enzymes, biotin-streptavidin systems, and radioisotopes (Zhang, Park, et al. 2018), where each approach offers unique advantages. Fluorescence labelling involves attaching a fluorescent dye to the antibody, allowing for detection and visualization of the labelled target. Its advantages include high sensitivity, real-time imaging, and the ability to simultaneously detect multiple targets using different coloured dyes (Vira et al. 2010). One commonly used fluorescent dye is fluorescein isothiocyanate (FITC), which emits green fluorescence upon excitation (Fang et al. 2019). The covalent attachment of dyes like FITC to antibodies enables specific and stable labelling, while microscopy techniques such as fluorescence and confocal microscopy offer various levels of resolution and imaging

capabilities, allowing for precise visualization and analysis of labelled targets, which grant detailed insights into cellular processes, molecular interactions, and disease mechanisms (Ishikawa-Ankerhold, Ankerhold, and Drummen 2012).

On the other hand, radioisotope labelling involves the incorporation of a radioactive isotope, typically Technetium-99m (^{99m}Tc) or Indium-111 (^{111}In) (Kręcisz et al. 2021), into the antibody. The emitted radiation can be detected using imaging systems, facilitating precise quantification and localization of the target molecule (Vaneycken et al. 2011). Radioactive labelling of antibodies is usually a two-step process. Initially, the antibody is modified to introduce sulfhydryl groups by reduction of disulphide bonds (Dewulf et al. 2020). This is typically done by activating the DOTA (1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid) and conjugating it with the antibody under controlled conditions. The resulting DOTA-conjugated antibody possesses a high-affinity binding site for radioisotopes, such as ^{111}In (**Figure 13**). The chelate prevents dissociation of the radioisotope, ensuring minimal release of ^{111}In *in vivo* (Holik et al. 2022). This specificity enables accurate and sensitive imaging and detection of targeted molecules in various applications, such as cancer diagnostics and targeted therapy (Lu et al. 2020; Debie et al. 2021; Debie et al. 2020). While radioisotope labelling offers exceptional sensitivity and can be used for *in vivo* studies, its use of radioactive materials necessitates strict safety protocols.

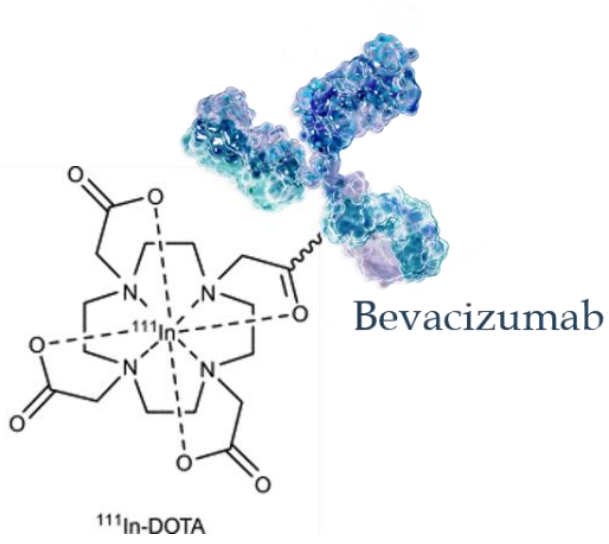


Figure 13. The schematic representation of ^{111}In incorporated into DOTA chelator which is randomly conjugated into Bevacizumab at its lysine residue side chains via amide bond formation.

Common fluorescence conjugation methods include targeting lysine residues using amide bond formation or cysteine residues using maleimide chemistry. The maleimide group selectively reacts with sulfhydryl groups, ensuring site-specific conjugation (Morais and Ma 2018). The conjugation site on the antibody may vary depending on the specific antibody and conjugation strategy used.

3. MATERIALS AND METHODS

3.1. Type of research

The study pursues an analytical and computational approach as well as experimental methods.

3.2. Location and time frame of research

This study was conducted at multiple institutions and different campuses indicated as follows: From February 2018 to May 2021 the experimental research was carried out at İzmir Biomedicine and Genome (IBG) Institute (9 Eylul University, Izmir, Turkey) with some analytical studies being conducted remotely due to COVID-19 curfew restrictions. The radiochemistry and medical imaging studies were performed at the *In vivo* Cellular and Molecular Imaging (ICMI) facility at Jette Health Campus (Laarbeeklaan 103, 1090 Jette, Brussels, Belgium) from May 2021 to July 2021, as part of the research conducted at Vrije University Brussels (VUB) in Brussels, Belgium. Furthermore, the sdAb production studies were predominantly carried out at the Cellular and Molecular Immunology (CMIM) laboratory of Medical Imaging Core Facility (MIMA), located at the VUB Pleinlaan campus (Pleinlaan 2, 1050 Elsene, Brussels, Belgium) from July 2021 to November 2021. Depending on the specific requirements of each study, the laboratory facilities at VUB were interchanged as needed. Throughout this thesis, the laboratories engaged in research activities overseas will be designated as the host lab, while studies conducted at IBG will be referred to as the home lab. Finally, from November 2021 to November 2023, the remaining experimental and computational studies were conducted at the Kalyoncu Lab at the Antibody Engineering and the IBG Bioinformatics Platform (IBG-BIP) at the IBG Institute, excluding that there was a period of pregnancy and maternity leave including the surgical recovery period from November 2022 to May 2023, during which the research activities were temporarily halted.

3.3. Research population, sampling and experimental group

In this study, the zebrafish model was employed as an *in vivo* test system. The zebrafish used in the experiments were housed and bred in the IBG-Vivarium Zebrafish Facility.

Furthermore, for abroad studies athymic nude mice, which were intended for use in the radiochemistry and fluorescence studies, were bred at the Universitair Ziekenhuis Brussel in Belgium and athymic nude mice bearing subcutaneous tumours were subsequently cared for and studied at the ICMI, VUB.

For the approved ethical committee reports please refer to the Addenda section.

3.4. Research materials

3.4.1. General

The antibodies (**Table 1**), bacterial strains and media (**Table 2 and 3**), chemicals (**Table 4**), kits (**Table 5**), PCR reagents and plasmids (**Table 6 and 7**), common reagents and solutions (**Table 8**), resins (**Table 9**), unclassified materials (**Table 10**), equipment (**Table 11**), instruments (**Table 12**), software (**Table 13**), human cell lines (**Figure 14**, **Table 14 -15**), expression vectors (**Figure 15 and 16**), oligonucleotides (**Figure 17**), primers (**Table 16 -18**) and animal models (**Table 19**) are listed in details above tables and figures.

Table 1. Antibodies.

Reagent Name	***Source***	***Catalogue #***	***Location***
Antibodies			
Anti-Flag M2 Antibody	Sigma-Aldrich	SLBT7654	IBG
Anti-Mouse IgG (H+L), Peroxidase Conjugate	Sigma-Aldrich	SAB3701066	IBG
Bevacizumab (Altuzan®/Avastin® 100 mg/4 ml)	Roche	N/A	IBG/VUB

Table 2. Bacterial strains.

Reagent Name	***Source***	***Catalogue #***	***Location***
Bacterial Strains			
DH5α Competent *E.Coli*	Invitrogen	18265017	IBG
One Shot™ BL21(DE3) Chemically Competent *E. Coli*	Agilent	200132	IBG
XL10-Gold Ultracompetent *E.Coli*	Agilent	200317	IBG
WK6 Competent *E. Coli*	ATCC	47078	VUB

Table 3. Bacterial media.

Reagent Name	***Source***	***Catalogue #***	***Location***
Bacterial Media			
LB Agar, Miller	BioBasic	SD7003	IBG
LB Agar, Miller	BioShop Canada	LBA408.500	IBG
LB Broth	Sangon Biotech	A507002	IBG
LB Broth, High Salt	Duchefa Biochemie	L1704	VUB

LB Broth, Miller	BioBasic	SD7002	IBG
Luria Broth Base, Miller	Sigma-Aldrich	L3522	IBG
Micro Agar	Duchefa Biochemie	M1002	VUB

Table 4. Chemicals.

Reagent Name	***Source***	***Catalogue #***	***Location***
Chemicals			
α-Lactose monohydrate	Sigma-Aldrich	L2643	IBG
1,4-Dithiothreitol (DTT)	GoldBio	DTT10	IBG
DTT	BioRad	1610611	VUB
2-Mercaptoethylamine hydrochloride (2-MEA)	ACROS Organics	AC153770050	VUB
2-Mercaptoethanol ≥99.0% (βME)	Aldrich	M6250	IBG
Acetic acid (Glacial) 100 %	Merck	100056	IBG
Adenosine 5'-Triphosphate (ATP) disodium salt	BioBasic	AB0020	IBG
Agarose, for molecular biology	Sigma-Aldrich	A4718	IBG
Agarose, for molecular biology, low eeo	Sigma-Aldrich	A9539	IBG
Ammonium acetate	BioUltra	09689	VUB
Ammonium persulfate	BioRad	1610700	IBG
Ammonium sulfate	Merck Milipore	101217	IBG
Ammonium sulfate	Sigma-Aldrich	31119	IBG
Ampicillin, sodium salt	BioShop	AMP201.100	IBG
Ampicillin	Duchefa Biochemie	A0104	VUB
Bovine serum albumin, lyophilized powder	Sigma-Aldrich	A9418	IBG
Calcium chloride	Merck	102378	IBG
Calcium chloride dihydrate	Isolab	909.026	IBG
Chloramphenicol	BioShop Canada	CLR201	IBG
Citric acid monohydrate	Sigma-Aldrich	C1909	IBG
D-glucose	Duchefa	G0802	VUB

	Biochemie		
D-glucose, anhydrous, reagent grade	BioShop Canada	GLU501	IBG
D-glucose, monohydrate	BioBasic	GB0218	IBG
Dimethyl Sulfoxide (DMSO)	EMD Milipore	20-139	IBG
DMSO	Sigma-Aldrich	D8418	VUB
DMSO	Tekkim	TK.201790	IBG
Di-sodium hydrogen phosphate anhydrous	Sigma-Aldrich	1.06585	IBG
Di-sodium hydrogen phosphate dihydrate	Merck	106580	VUB
Di-potassium hydrogen phosphate trihydrate	Merck	105099	VUB
Ethanol	Merck	1.00971	IBG
Ethanol 96% (v/v), for analysis	Scharlab	ET00042500	IBG
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich	E134	IBG
EDTA	Duchefa Biochemie	E0511	VUB
EDTA	BioUltra	EDS	VUB
Glycerine 99.5 %	Tekkim	TK070190	IBG
Glycerol	Duchefa Biochemie	G1345	VUB
Glycerol 99.7%	BioShop Canada	GLY001	IBG
Glycine	BioBasic	GB0235-25	IBG
Glycine 99%	BioShop Canada	GLN001.500	IBG
Guanidine hydrochloride 98%	Fisher Scientific	AAA1354330	IBG
Hydrochloric acid fuming 37%	Isolab	932.103	IBG/VUB
Hydrogen peroxide 35%	Isolab	932.122.2500	IBG
Hydrogen peroxide 50%	Tekkim	TK.080220	IBG
Imidazole reagent plus 99%	Sigma-Aldrich	1202-100g	IBG
Imidazole	Sigma	I2399	VUB
Isopropyl–b–D–thiogalactoside (IPTG)	Duchefa Biochemie	I1401	VUB

L-histidine hydrochloride	BioBasic	HB0330-500	IBG
Magnesium chloride ≥99 %	Carl Rot	Kk36.1	IBG
Magnesium chloride	Merck	814733	VUB
Magnesium chloride hexahydrate	Merck	105833	IBG
Magnesium sulfate anhydrous	Merck	106067	IBG
Magnesium sulfate anhydrous	AFG Scientific	147676	IBG
Magnesium sulfate heptahydrate ≥99.0%	Sigma-Aldrich	M2773	IBG
Mes free acid, monohydrate	BioBasic	MB0341-500	IBG
Mes hydrate	Sigma-Aldrich	M2933	IBG
Methanol	Tekkim	TK.120320	IBG
Methanol ≥99.7%	Sigma-Aldrich	24229	IBG
Methylene blue	Merck	1592700010	IBG
Nickel (II) chloride hexahydrate	AFG Bioscience	7791-20-0 250	IBG
Nickel(II) chloride	Santa Cruz	sc-394490	IBG
Peptone-A	BioBasic	G213	IBG
Pipes	Sigma	P6757	IBG
Potassium chloride	Tekkim	TK.150440	IBG
Potassium chloride	Merck	104936	IBG/VUB
Potassium chloride ≥ 99%	Isolab	960.036	IBG
Potassium dihydrogen phosphate	BioBasic	PB0445	IBG
Potassium dihydrogen phosphate	Merck	104873	IBG/VUB
Potassium hydroxide	Merck	105033	IBG
Potassium phosphate monobasic	Tekkim	TK200990	IBG
Potassium phosphate monobasic	Sigma-Aldrich	PHR1330	IBG
Potassium phosphate monobasic, anhydrous	AFG Scientific	292188	IBG
Potassium phosphate monobasic, anhydrous	BioBasic	PB0445	IBG
SDS	BioBasic	SB0485	IBG
Skim milk	BioShop	SKI400	IBG
SOC medium	BioShop	SOC001	IBG

Sodium acetate anhydrous	AFG Scientific	218137	IBG
Sodium acetate trihydrate	Isolab	969.015	IBG
Sodium azide	Sigma	822335	VUB
Sodium bicarbonate	Sigma-Aldrich	S6014	IBG
Sodium chloride	Merck	106404	IBG
Sodium chloride	Fisher Scientific	S271-500	VUB
Sodium chloride ≥99.5%	Isolab	969.036	IBG
Sodium dodecyl sulfate (SDS)	Sigma-Aldrich	L3771	IBG
Sodium hydrogen carbonate	Isolab	969.103	IBG
Sodium hydroxide	Merck	106462	VUB
Sodium hydroxide 0.1 n	BioShop	SHY999	IBG
Sodium hydroxide beads	BioBasic	SB6789	IBG
Sodium phosphate dibasic dihydrate	Sigma-Aldrich	04272	IBG
Sodium phosphate dibasic anhydrous	BioBasic	S0404	IBG
Sucrose	EMD Millipore	100892	IBG
Sucrose	BioBasic	SB0498	IBG
Sucrose	Duchefa Biochemie	S0809	VUB
Sulfuric acid 95-98%	Isolab	9700.23	IBG
Tris	BioBasic	TB0195	IBG
Tris	Sigma	GE17-1321-01	VUB
Tris HCl	Goldbio	T-095	IBG
Tris-HCl	Sangon Biotech	B548127	IBG
Triton X-100	Sigma-Aldrich	T8787	IBG
Trizma® Base ≥99.9%	Sigma-Aldrich	T1503	IBG
Tryptone	Duchefa Biochemie	T1332	VUB
Tryptone	Sigma-Aldrich	T7293	IBG
Tryptone	Oxoid	LP0042	IBG
Tween-20	Merck	822184	IBG
Tween-20	Thermo	P1379	IBG

	Scientific		
Yeast Extract	BioShop Canada	YEX555	IBG
Yeast Extract	BioBasic	G0961	IBG
Yeast Extract	Duchefa Biochemie	Y1333	VUB
Yeast Extract	Thermo Scientific	211929	IBG
Zinc Chloride	ZAG	400.ZA.CB3401	IBG

Table 5. Kits.

Reagent Name	***Source***	***Catalogue #***	***Location***
Kits			
EZ-10 spin column DNA cleanup miniprep kit	BioBasic	BS367	IBG
EZ-10 spin column plasmid DNA miniprep kit	BioBasic	BS614	IBG
Genelute™ plasmid miniprep kit	Sigma Aldrich	PLN70	VUB
IsoLink kit	Mallinckrodt	Covidien, Petten	VUB
Protein thermal shift dye kit	Applied Biosystems	4461146	IBG
Protein thermal shift starter kit	Applied Biosystems	4462263	IBG
QiaPrep spin miniprep kit	Qiagen	27104	IBG
QiaQuick gel extraction kit	Qiagen	28704	IBG
Tgx™ Fastcast™ acrylamide kit, 12%	BioRad	1610175	IBG

Table 6. PCR Reagents.

Reagent name	***Source***	***Catalogue #***	***Location***
PCR reagents			
10X Tango buffer with BSA	Thermo Scientific	00630570	VUB
5X Colourless GoTaq® reaction buffer	Promega	0000452089	VUB
BstEII	Thermo Scientific	00732586	VUB
Deoxynucleotide (dNTP) solution set	NEB Biolabs	N0446S	IBG

DpnI digestion kit	NEB Biolabs	R0176S	IBG
FastAp thermosensitive alkaline phosphatase (1 U/µl)	Thermo Scientific	EF0651	VUB
FastDigest DpnI	Thermo Scientific	FD1704	IBG
GoTaq® polymerase (5 u/ml)	Promega	0000356814	VUB
KOD hot start DNA polymerase	Merck Millipore	71086-3	IBG
NcoI	Thermo Scientific	0079944	VUB
Nucleospin gel and PCR clean-up	Macherey-Nagel	740609	IBG
Q5® Hot start high-fidelity 2X master mix	NEB Biolabs	M0494S	IBG
Quikchange lightning multi site-directed mutagenesis kit	Agilent	210513	IBG

Table 7. Plasmids.

Reagent name	***Source***	***Catalogue #***	***Location***
Plasmids			
pUC18	Thermo Scientific	SD0051	IBG
pET-17b	GenScript	pET-17b	IBG
pHen6 vector (CC13 or CC1168)	Custom made	-	VUB

Table 8. Reagents and solutions.

Reagent name	***Source***	***Catalogue #***	***Location***
Reagents and solutions			
1-Step ultra TMB-ELISA substrate solution	Thermo	34028	IBG
2X Laemmli sample buffer	BioRad	1610737	IBG/VUB
4X Laemmli sample buffer	BioRad	1610747	IBG
Bio-Safe coomassie stain	BioRad	1610786	IBG
Coomassie brilliant blue	ThermoFisher	R-250	VUB
IRDye800CW maleimide	LI-COR Biotechnology	929-80020	VUB
Phosphate buffered saline (PBS) pH 7.4	Calbiochem	P3813	VUB
Phenol red solution	Sigma-Aldrich	P0290	IBG

SYBR™ safe DNA gel stain	Thermo	S33102	IBG
SYPRO™ orange protein stain	Sigma-Aldrich	S5692	IBG
Tris/Glycine/SDS (TGS) running buffer (10X)	BioRad	1610732	VUB

Table 9. Resins.

Reagent name	*Source*	*Catalogue #*	*Location*
Resins			
Chelex® 100	Sigma-Aldrich	C7901	VUB
His-Pur™ Ni-NTA resin	Thermo	88221	IBG
His-Select® Nickel affinity gel	Sigma	P6611	VUB
Nickel NTA agarose beads	GoldBio	H-350	IBG
Ni-NTA sepharose resin	GE Life Sciences	17531802	IBG
Protein A resin	Genscript	LP00210	IBG

Table 10. Other materials.

Reagent name	*Source*	*Catalogue #*	*Location*
Other materials			
12% mini-protean® TGX™ precast protein gels, 10-well	Sigma-Aldrich	4561043	IBG
DNA ladder (1kb)	Invitrogen	10787018	IBG
HiLoad® Superdex 75 column	Cytiva	16/600, 10/300 GL	VUB
HisTrap HP, 5 ml	GE Healthcare	71-5027-68 AH	IBG
Human VEGF-165 recombinant protein	Thermo Scientific	RVEGFI	IBG
Novex™ Wedgewell™ 16 %, tris-glycine, 1.0mm, mini protein gel, 10-well	Life Technologies	XP08160BOX	VUB
Pageruler™ Pre-stained protein ladder	Thermo Scientific	26616	IBG/VUB
PD-10 desalting column	GE Healthcare	GE17-0851-01	VUB
Precision plus protein dual colour	BioRad	1610374	IBG

standards			
Pre-stained SDS-PAGE standards, broad range	BioRad	1610318	IBG
Sephadex G25 disposable column	GE Healthcare	NAP-5	VUB

Table 11. Equipment.

Reagent name	*Source*	*Catalogue #*	*Location*
Equipment			
-20°C freezer	Bosch	GSN51AW30	IBG
4°C fridge	Bosch	KSV36AI31	IBG
7500 Fast real-time PCR system	Applied Biosystems	4406985	IBG
-80°C freezer	New Brunswick	R207-14	IBG
99Mo/^{99m}Tc generator	GE Healthcare	Drytec	VUB
Äktaexplorer	Cytiva	ÄKTAexplorer 100	VUB
Automated cell counter	Orflo	MOXI Z	VUB
Automated liquid chromatography system	BioRad	NGS	VUB
Avanti® ultracentrifuge	Beckman Coulter	J-30I	IBG
Benchtop centrifuge	Beckman Coulter	Allegra X-15R	IBG
Benchtop centrifuge	Beckman Coulter	Allegra X-22R	IBG
Benchtop centrifuge	Eppendorf	5810 R	IBG/VUB
Bioruptor® Pico sonication device	Drogade	B01060010	IBG
Bransonic ultrasonic cleaner	Branso	1210	VUB
ChemiDoc MP imaging system	BioRad	12003154	IBG
Cimarec+™ stirring hotplates	Thermo Scientific	SP131320-33Q, SP88854100	IBG
Compact digital mini rotator	Thermo Scientific	88880025	IBG
Compact digital mini rotator with orbital motion	Thermo Scientific	88880026	IBG

Compact digital rocking shaker	Thermo Scientific	88880020	IBG
Edge®blu bluetooth® smart pH electrode and meter	Hanna	HI2202-01	IBG
Entris® II essential line precision balance	Sartorius	BCE822-1S	IBG
Forma™ high-performance auto defrost lab freezers	Thermo Scientific	FFGL430A	IBG
Fume cupboard	Kötterman	Standard	IBG
Galaxy® 170 s high-capacity CO_2 incubator	New Brunswick	Galaxy 170S	IBG
High performance liquid chromatograph	Shimadzu	SIC20AC, LC20AD, DG02A, CTD-10AS VP, CBM-20A, RCIOAXL, SDD-20A, RID-10A	IBG
Incubator SI500 with orbital shaker	Stuart	9951612	IBG
iTLC reader	Raytest	Elysia	VUB
KNF laboport® mini-pump	Sigma-Aldrich	Z288292	
MicroCL 17R microcentrifuge	Thermo Scientific	75002499	IBG
Micropulser electroporator	BioRad	1652100	IBG
Microwave oven	Beko	MWB 2000 EXN	IBG
Mini-PROTEAN tetra vertical electrophoresis cell	BioRad	1658004	IBG
mySPIN™ 6 mini centrifuge	Thermo scientific	75004061	IBG
NanoDrop spectrophotometer	Thermo Scientific	ND-2000	IBG
NGC Medium-Pressure Liquid Chromatography System	BioRad	NGC 10	VUB
Odyssey® CLx imaging	LI-COR	Odyssey CLx Imager	IBG/VUB

system			
Owl™ D2 wide-gel electrophoresis system	Thermo Scientific	D2-BP	IBG
Owl™ EasyCast™ B1A mini gel electrophoresis systems	Thermo Scientific	B1A-BP	IBG
Peristaltic pump, 120V	Watson Marlow	500000	IBG
PowerPac™ universal power supply	BioRad	1645070	IBG
PowerPac™ basic power supply	BioRad	1645050	IBG
Precision weighing balance	Sartorius	GC 1603 OCE	IBG
Shaker incubator	Barnstead	BS-SHKE5000-1 CE, BS-SHKE4000-8 CE	IBG
Shaker incubator	Thermo Scientific	Max Q4000, Max Q5000	IBG
Shaking bacterial incubator	Shel-Lab	5I6-2	IBG
SimpliAmp™ thermal cycler	Applied Biosystems	A24811	IBG
Sl 8 small benchtop centrifuge	Thermo scientific	75007220	IBG
Spectrofluorophotometer	Shimadzu	RF-6000	VUB
Stackable incubator shaker	New Brunswick	Innova® 44/44R	VUB
Stuart dry block heater	Cole-Parmer	STSBH130D	IBG
Synergy H1 multimode reader	BioTek	Synergy H1	IBG
Table top centrifuge	Eppendorf	5920	VUB
Thermomixer comfort heating block	Eppendorf	5382000015	VUB
Tube revolver rotator	Thermo Scientific	88881002	IBG
UV transilluminator	Vilber Lourmat	ECX-F20L	IBG
Vacuum filter set	Isolab	S.043.03.001	IBG
Varioskan flash multi detection microplate reader	Thermo Scientific	14044	IBG

Vortex	Scientific Industries	Genie2	IBG
Vortex mixer	Thermo scientific	88880018	
Water bath	Nüve	NB9	IBG
XCell SureLock mini-cell electrophoresis system	Thermo Fisher Scientific	EI0001	VUB
Zoom dual power supply	Thermo Fisher Scientific	ZP10001	VUB

Table 12. Instruments.

Reagent name	***Source***	***Catalogue #***	***Location***
Instruments			
Amicon® ultra centrifugal filters, 0.5, 4, 15 ml	Merck Millipore	UFC5010, UFC8010, UFC9010	IBG
Baffled flask, 1000 ml	Isolab	027.06.901	IBG
Baffled flask, Pyrex® flask, 1000 ml	Corning	4450-1000	IBG/ VUB
Cell culture flasks, T-25, T-75	Corning	430168, 430641	VUB
Cell spreader	Isolab	S.082.03.001	IBG
Centrifuge tubes, micro, flat cap, sterile, 1.5, 2 ml	Isolab	S.078.03.022 S.078.03.021	IBG
Costar® cell culture well-plate, 24-well	Corning	3524	VUB
Cryovials	Corning	430659	VUB
Filter paper, Whatman® 3mm	Merck	WHA3030917	IBG
Gene Pulser/MicroPulser Electroporation Cuvettes, 1, 2 mm	BioRad	1652083, 1652082	IBG
Glass bottle, 250, 500, 1000, 2000 ml	Isolab	I.061.01.250, I.061.01.500, I.061.01.901, I.061.01.902	IBG
GP Express plus filter	Millipore	GPWP09050	VUB
Graduated cylinders, 100, 1000 ml	Isolab	016.06.100,	IBG

		016.06.901	
Gyrodisc 30 mm, 0.2 µm	Orange Scientific	CA-PC	VUB
iTLC paper strips	Agilent	SGI0001	VUB
Membra-cel ® dialysis tubing, MWCO 7000RC, 16 mm	Serva Electrophoresis	44313.01	IBG
MicroAmp™ clear adhesive film	Applied Biosystems	4306311	IBG
MicroAmp™ fast optical 96-well reaction plate	Thermo Scientific	4346906	IBG
Micropipette kits	Gilson	FB10202, F167920	IBG
Micropipettes, 10, 200, 1000 µl	Isolab	011.20.910, 011.20.200, 011.20.905	IBG
Multichannel micropipette 200 µl	Axygen	AP-12-200-ALT	IBG
Nalgene™ PPCO centrifuge bottles, 250 ml	Thermo Scientific	3120-0250	IBG
Nitrocellulose membranes	Thermo Scientific	88018	IBG
PCR plates, 96-well	Nest	402301	IBG
PCR tubes, 200 µl	Isolab	I.123.01.028	IBG
Petri dishes	Isolab	120.13.035	IBG
Pierce™ centrifuge columns, 10ml	Thermo Scientific	89898	IBG
Pipet-lite multi pipette	Rainin	17013805	IBG
Pipette tips, 10, 200, 1000 µl	Isolab	LB.IS.005.01.001, LB.IS.005.01.002, LB.IS.005.01.003	IBG
PVDF (polyvinylidene fluoride) transfer membrane	Thermo Scientific	88520	IBG
Ratiolab™ 96-well polystyrene micro	Fisher	RL6018717	IBG

test plates	Scientific		
Safe-lock tubes, 1,5 ml, amber	Eppendorf	EPPR4502	VUB
Safe-lock tubes, 1,5 ml, colourless	Eppendorf	EPPR4501	VUB
Serological pipettes, 5, 10, 25 ml	Isolab	7695552, 7695553, 7695554	IBG
Sterile falcon tubes, 15, 50 ml	Isolab	S.078.02.007, S.078.02.008	IBG
Sterile syringe filter, PVDF, 20, 45 µm	Isolab	I.094.07.003, I.094.06.006	IBG
Sterile syringe, 5, 20, 50 ml	Isolab	S.094.91.005, S.094.91.020, S.094.91.050	IBG
Tube rack	Isolab	I.079.04.001	IBG
Tube stand	Isolab	S.080.01.005	IBG
Vivaspin® 2, 6	Sartorius	VS02V1, VS0611	VUB

Table 13. Software.

Software	***Source***	***Location***
Aggrescan3D (version 2.0)	Universitat Autònoma de Barcelona (Spain)	IBG
ANARCI	Oxford Protein Informatics Group (OPIG), Oxford University (UK)	IBG
CamSol	Vendruscolo Lab., University of Cambridge (UK)	IBG
Clustal Omega	EMBL European Bioinformatics Institute (UK)	IBG
EMBOSS Pepstats	EMBL-EBI, Hinxton (UK)	IBG
Eurofins Oligo Analysis Tool	Eurofins Genomics (Germany)	IBG
GelDoc XR+ Image Lab	BioRad (USA)	IBG
GraphPad Prism (version 7.05)	GraphPad (USA)	IBG
ImageJ	Rasband W.S., National Institutes of Health,Bethesda, Maryland (USA)	IBG

OrginPro 8.5.0	OriginLab Northampton, MA (USA)	IBG
Parapred	Yusuf Hamied Department of Chemistry, University of Cambridge (UK)	IBG
PASTA (version 2.0)	Biocomputing UP (Italy)	IBG
ProtParam	Expasy, SIB Swiss Institute of Bioinformatics (Switzerland)	IBG
PyMOL (version 2.5.5)	Schrödinger, Inc. (USA)	IBG
QuikChange® primer design	Stratagene (La Jolla, CA)	IBG
Tango (version 2.2)	VIB Switch Laboratory, Leuven (Belgium)	IBG
Waltz	VIB Switch Laboratory, Leuven (Belgium)	IBG
WebLogo (version 2.8.2)	Department of Plant and Microbial Biology, University of California, Berkeley	IBG

3.4.2. Human cell line

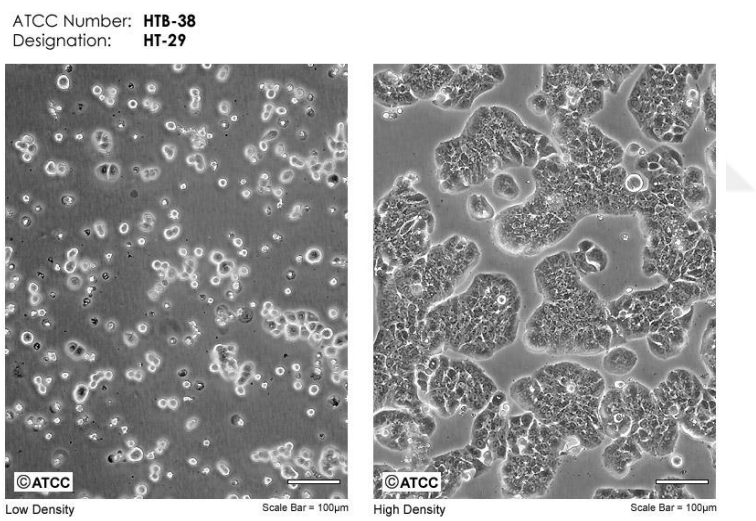


Figure 14. HT-29 cells.Human immortalized colorectal adenocarcinoma cells (HT-29) were originally obtained from ATCC.

Table 14. HT-29 cell properties.

Product name	***HT-29***
Product category	Human cells
Organism	*Homo sapiens*, Human
Disease	Adenocarcinoma; Colourectal
Cell type	Epithelial
Tissue	Colon

Table 15. Cell culture reagents for HT-29 cells.

Reagent Name	*Source*	*Catalogue #*	*Location*
McCoy's 5A (Modified) Medium	Gibco	16600082	VUB
Fetal Bovine Sera (FBS)	Tico Europe	FBSEU500	VUB
Penicilline/Streptomicine (10,000 U/ml)	Gibco	15140122	VUB
L-Glutamine (200mM, 100X)	Gibco	25030024	VUB
TrypLE Express	Gibco	12605028	VUB

3.4.2. Expression vectors

The pET system is the most powerful system for the expression of recombinant proteins in E. coli and is widely used in prokaryotic expression.

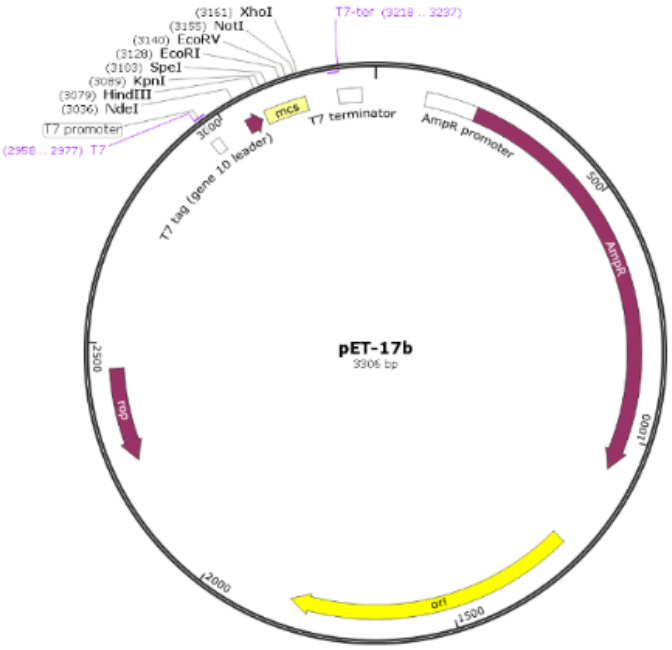


Figure 15. The plasmid map of pET-17b expression vector. The plasmid is 3306 bp long without any insert. pET-17b is a prokaryotic expression vector that contains an N-terminal T7 tag. The multiple cloning site (MCS) is indicated in the plasmid map. The plasmid includes restriction sites for cloning of different genes that are shown in the plasmid map. The ampicillin resistance (Amp^R) gene is included in the construct, followed by an origin of replication (ori). Expression of the desired genes is induced by T7 RNA polymerase provided by the host cells.

pHen6-derived vector (CC13 or CC1168) has been extensively used in nanobody production using *E.coli* as host cell.

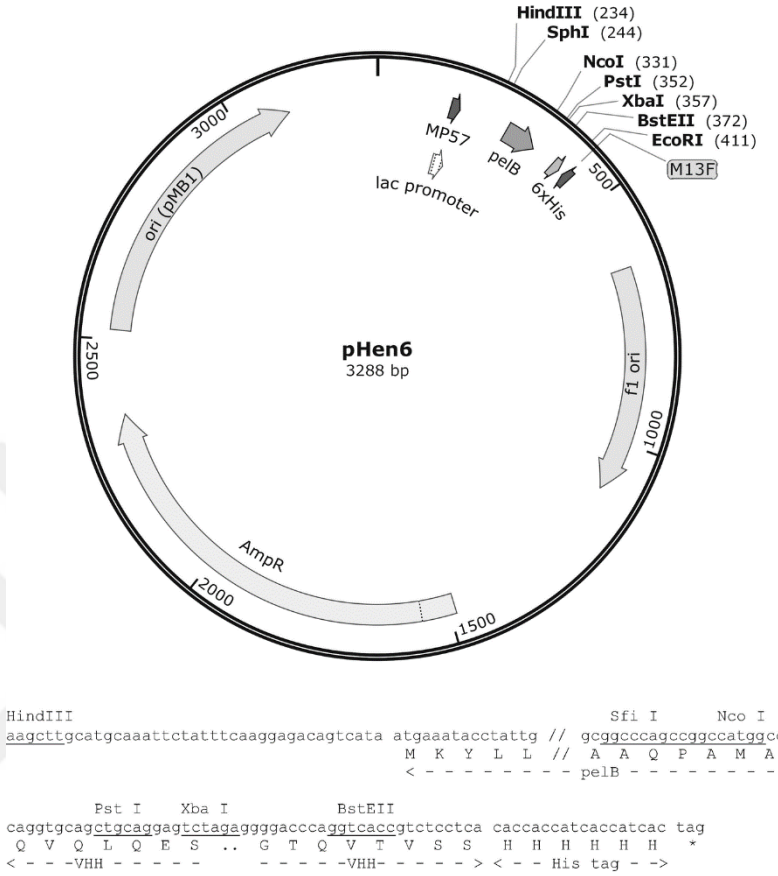


Figure 16. The map of pHen6 expression vector. The plasmid is 3288 bp long without any insert. The pHen6 plasmid is an engineered construct, specifically created to facilitate the effective expression of the nanobody genes. Lac promoter is followed by protein secretion signal sequence (PelB), MCS for insertion of the gene of interest and His_6 tag, respectively. The pHen6 plasmid includes essential components, of MCS, an Amp^R gene, an ori, an inducible LacZ promoter, a PelB signal, and a His_6 tag. MP57 serves as an annealing site for associated primer.

3.4.3. Oligonucleotides

Bevacizumab V_H WT

HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYT FTNYGMNW*V*RQAPGK*GLEW*VGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQM NSLRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVD**IEGR**DYKDDDD KAADYKDDDDKAADYKDDDDKAA*HHHHHHH*-LE

Bevacizumab V_H WT Codon optimised

CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG GTC CGC CAA GCG CCA GGG AAA GGA TTA GAG TGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG

Figure 17. Variable heavy chain of codon optimised Bevacizumab amino acid and nucleotide sequences. The V_H sequence of codon optimised Bevacizumab was designed for recombinant protein production in *E. coli* cells. The sequence is derived from the non-mutated wild-type (WT) Bevacizumab V_H scaffold. The codon optimisation was based on the codon usage frequency of the host *E. coli* cells. The following regions were highlighted in the sequences. NdeI sequence and first nucleotide of the start codon were highlighted in cyan, PelB sequence and the last two nucleotides of the start codon in green, NheI in grey, Codon optimised V_H sequence, Mutation, SalI in dark gold, Factor Xa protease in bold text, $(Flag\ Tag)_3$ in underlined text, His_7 tag in italicised text, Stop codon in red, and XhoI in pink.

3.4.4. Primers

All the primers were ordered from Eurofins for both home and the host lab studies.

Table 16. List of forward and reverse primers used in the mutations introduced on Bevacizumab V_H sequence at the home lab. The mutations are listed alphabetically and not correspond to the order in which they were introduced. These mutations were performed by substituting the amino acid codons with desired ones, taking into account the codon usage frequency of the host *E. coli* cell. The mutated codons were highlighted in red. The mutations

were written in the following order [original residue][residue location][mutated residue] using single letter amino acid codes.

Mutation	Primer	Primer sequence (5'→3')
V37F	Reverse	C TGG CGC TTG GCG GAA CCA GTT CAT ACC ATA ATT TG
	Forward	CA AAT TAT GGT ATG AAC TGG TTC CGC CAA GCG CCA G
V37Y	Reverse	C TGG CGC TTG GCG GTA CCA GTT CAT ACC ATA ATT TG
	Forward	CA AAT TAT GGT ATG AAC TGG TAC CGC CAA GCG CCA G
G44E	Reverse	CC TAC CCA CTC TAA TTC TTT CCC TGG CGC TTG
	Forward	CAA GCG CCA GGG AAA GAA TTA GAG TGG GTA GG
L45R	Reverse	CCA CCC TAC CCA CTC TCT TCC TTT CCC TGG CGC
	Forward	GCG CCA GGG AAA GGA AGA GAG TGG GTA GGG TGG
W47G	Reverse	ATT GAT CCA CCC TAC GCC CTC TAA TCC TTT CCC TG
	Forward	CA GGG AAA GGA TTA GAG GGC GTA GGG TGG ATC AAT
G44E L45R	Reverse	CCC TAC CCA CTC TCT TTC TTT CCC TGG CGC TTG G
	Forward	C CAA GCG CCA GGG AAA GAA AGA GAG TGG GTA GGG
G44E W47G	Reverse	GAT CCA CCC TAC CCC CTC TAA TTC TTT CCC TGG
	Forward	CCA GGG AAA GAA TTA GAG GGG GTA GGG TGG ATC
G44E L45R W47G	Reverse	GT ATT GAT CCA CCC TAC CCC CTC TCT TTC TTT CCC TGG CGC TTG GC
	Forward	GC CAA GCG CCA GGG AAA GAA AGA GAG GGG GTA GGG TGG ATC AAT AC
F70I L72R T74N S77N A79L	Reverse	CAT TTG TAA GTA GAC TGT GTT TTT GGA ATT GTC ACG TGA AAT GGT AAA ACG GCG C
	Forward	G CGC CGT TTT ACC ATT TCA CGT GAC AAT TCC AAA AAC ACA CTG TAC TTA CAA ATG

Table 17. The forward and reverse primers used at the host lab for the targeting of Bevacizumab V_H sequences (WT and its variants) to retrieve them from the home lab plasmid construct.

Target	Primer	Primer sequence (5'→3')
Bevacizumab VH Sequence	Reverse	GGA AGA GAC GGT GAC CAA TGT CCC
	Forward	CCC GCC ATG GCG GAT GTC CAG TTG GTT GAA AGT GG

Table 18. The reverse primer targeting MP57 region located upstream of the insert on the pHen6 plasmid.

Target	Primer	Primer sequence (5'→3')
MP57	Reverse	TTATGCTTCCGGCTCGTATG

This primer was only used for sequencing purposes; therefore there was no need to order a complementary forward primer.

3.4.5. In vivo experimental animal models

Table 19. The details regarding *in vivo* studies of animal models including information on the species, strain/genotype, age, sex, number of animals, source and their location.

Species	Strain	Age/# animals	Gender	Source	Location
Danio rerio	Tg(fli1:EGFP)^{y1}	2 days post-fertilization (dpf)	Female/Male	IBG	IBG
Athymic nude mouse	Crl:NuFoxn1^{NU}	6-8 weeks old x 96 mice	Female	Charles River	VUB

3.5. Research variables

In the following section (3.6. Data collection tools), the research variables were individually mentioned in detail and categorized according to each procedure.

3.6. Data collection tools

3.6.1. Plasmid design

The insertion of a V_H sequence of Bevacizumab (PDB ID: 1BJ1) (Muller et al. 1998) into MCS of the pET17b plasmid is enabling the expression of a desired protein. To accomplish this, the NdeI and XhoI restriction enzymes, located strategically on the plasmid, were employed. NdeI serves as the upstream restriction enzyme site, while XhoI acts as the downstream restriction enzyme site. These enzymes recognize and cleave specific DNA sequences, generating cohesive ends for accommodation of the insert (**Figure 18**).

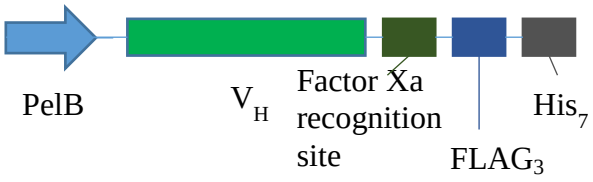


Figure 18. Schematic representation of the gene sequence of Bevacizumab V_H in expression plasmid pET-17b plasmid. PelB leader sequence is followed by the insert for the expression of the gene of interest. Protease cleavage site and the following tags were included for purification or detection purposes.

The pelB leader sequence is composed of a series of amino acids that, when added to a protein, directs it towards the bacterial periplasm found in gram-negative bacteries (such as *E.coli*), which is a favourable environment for proteins (Zarschler et al. 2013). Once the protein reaches to periplasmic space, the pelB sequence is cleaved by a signal peptidase. Following the PelB, Bevacizumab V_H gene (WT and its variants) comes for target amino acid production. Factor Xa cleavage site (IEGR) serves for precise removal of the extra tags from the purified protein of interest, meanwhile allowing the generation of a tag-free protein for the utilization in the downstream studies. $FLAG_3$ tag (DYKDDDDKAA), which is repeated three times incorporated into the plasmid design, for detection of the expressed protein in the characterization studies. For purification purposes, a His_7 tag (HHHHHHH) is appended for detection by IMAC (immobilized metal affinity chromatography). It is worth noting that instead of the more common hexa-His tag, our system employs a hepta-His tag for this purpose.

The pET17b plasmid was purchased from GenScript. The designed construct was transformed into BL21(DE3) *E. coli*, using appropriate transformation technique for its expression.

3.6.2. In Silico approaches for prediction of aggregation prone residues

When the V_L is dissociated from the Bevacizumab V_H, these residues located at the V_H-V_L interface become exposed to the surrounding solvent environment. Consequently, they tend to revert to their original state by interacting with any available surface in the physiological environment. To maintain V_H in a sdAb format, it becomes necessary to modify these exposed residues. The following *in silico* approaches were employed for the identification of regions that are likely to form aggregates due to exposed hydrophobic residues froun on the former V_H-V_L interaction site:

3.6.2.1. PDBePISA

The prediction of interacting residues at the V_H-V_L interface was carried out using PDBePISA (Krissinel and Henrick 2007). To achieve this, the PDB entry of Bevacizumab (1BJ1) was used as input and analyzed utilizing the "Interfaces" option. Amongst the listed

interfacing surfaces, the heavy and light chains were specifically selected and their corresponding details were listed. Residues with a higher Buried Surface Area (BSA) are more likely to be associated with the interface. The V_H-V_L pairing residues exhibiting the highest BSA scores are considered as potential candidates for introducing mutations to prevent their natural inclination to interact with the complementary residues on V_L.

3.6.2.2. Pasta 2.0

PASTA 2.0 (Walsh et al. 2014) assesses hydrophobicity, secondary structure propensity, and amino acid composition of given protein seqeuences. It can detect the regions on the submitted protein sequences that are likely to be potential aggregation hotspots. The FASTA format of V_H sequence of Bevacizumab was submitted to the server using the default settings.

3.6.2.3. Tango 2.2

Tango 2.2 (Fernandez-Escamilla et al. 2004; Rousseau, Schymkowitz, and Serrano 2006; Linding et al. 2004) is an algorithm for predicting aggregating regions within unfolded polypeptide chains. It can detect potential aggregation propensity of specific regions in the given sequence. The propensity calculations were performed by inputting Bevacizumab amino acid sequence, in FASTA format using the default settings.

3.6.2.4. Waltz 2.0

Waltz 2.0 (Louros et al. 2020) software is used for predicting of regions that are prone to forming amyloid fibrils within the given protein sequences. The V_H sequence of Bevacizumab was submitted to the server using the default settings in FASTA format.

3.6.2.5. Aggrescan 3D 2.0

Aggrescan3D 2.0 (Conchillo-Solé et al. 2007; de Groot et al. 2012) can predict and analyze protein aggregation propensity in three-dimensional structures. By taking into account of both sequence and structural information, it can predict the aggregation prone regions of the given protein. The PDB ID of V_H of Bevacizumab was submitted to the server using the default settings.

3.6.2.6. CamSol Intrinsic v2.2

CamSol Intrinsic v2.2 (Sormanni et al. 2017; Sormanni, Aprile, and Vendruscolo 2015), as the name implies, predicts the intrinsic solubility of proteins. It uses the protein sequence as input to determine the solubility behavior of a protein. The V_H sequence of Bevacizumab was submitted to the server using the default settings in FASTA format.

3.6.2.7. CamSol structurally corrected v2.1

CamSol Structurally Corrected v2.1 (Sormanni, Aprile, and Vendruscolo 2015) also uses combinatory approach as Aggrescan. It takes both the amino acid sequence and structural information of a protein into account when doing predictions. The PDB file of Bevacizumab was submitted to the server using default settings.

3.6.3. The databases

3.6.3.1. abYsis: integrated antibody sequence database

abYsis (Swindells et al. 2017) is a comprehensive database that houses an extensive collection of antibody sequences from diverse species. In our study, abYsis platform was harnessed as tool to examine the distributions of amino acids within a collection of mammalian V_Hs and camelid V_HHs for a comprehensive understanding of the natural distributions.

3.6.3.2. sdAb-DB: the single domain antibody database

In our project, the Single Domain Antibody Database (sdAb-DB) (Wilton et al. 2018) was used for comparison purposes, providing a data on collection of V_HHs derived from camelids. The diverse collection of sdAb-DB contains PDB ID, whenever it is available, as well as sequence information of the proteins. It is worth noting that most of the information is sourcing from the research publication therefore missing structural information.

3.6.4. Site-directed mutagenesis (SDM)

SDM was employed to introduce specific mutations at desired locations within Bevacizumab V_H sequence to create variants of our WT sdAb. Based on the consensus outcomes of the preceding section (3.6.2.), the mutations were systematically designed and engineered at various levels, including single, double, triple, quadruple, quintuple, and nonuple mutations. To create those variants, a two-fold approach was applied: (i) introducing mutations individually (to observe the individual effects of introduced mutations) and (ii) introducing multiple mutations simultaneously (to observe the synergetic effect of combined mutations). When introducing multiple mutations to create sdAb variants, the location and proximity of the target residues to each other were crucial. Therefore when designing the associated primers, this information was taking into consideration.

For mutation introductions via SDM, the QuikChange Lightning SDM kit, Q5 High-Fidelity 2X Master Mix, or KOD Hot Start DNA Polymerase were employed as appropriate depending on the availability of the products. Based on the specific polymerase utilized,

variations in the reaction mixture composition and reaction conditions were altered accordingly.

3.6.4.1. SDM with QuikChange lightning kit

50 ng/µl of dsDNA templates (pET17-b bacterial expression plasmids) were amplified by using QuikChange Lightning kit (10x QuikChange Lightning Multi reaction buffer, Quick Solution, QuikChange Lightning Multi enzyme blend), dNTP mix, forward and reverse primers (50 ng/ml) and DMSO. For the conventional PCR amplification, reaction mixture was prepared using the reagents as on the **Table 20**, then gently and homogenously mixed. To the each PCR tube, 24 µl of reaction mixture and 1µl of 50 ng/µl dsDNA template were added.

Table 20. Reagents for SDM reaction using QuikChange Lightning kit.

Reaction reagents (templates <5 kb)	Volume (µL)
10× Quikchange lightning multi reaction buffer	2.5
Nuclease free ddH_2O	13.5
Quick solution	0.5
dsDNA template (100 ng)	2
Forward primer (100 ng)	2
Reverse primer (100 ng)	2
dNTP mix	1
Quikchange lightning multi enzyme blend	1
DMSO	0.5
Total volume	**25**

The cover temperature of the thermal cycler was set to 99°C. Thermal-cycler was programmed to initial denaturation at 95°C for 2 min, followed by 30 cycles of denaturation at 95℃ for 20 sec, annealing at 55°C for 30 sec (depending on annealing temperature of primers and product size), elongation at 65°C for 2.5 min and 5 min of final extension at 65°C (**Table 21**).

Table 21. SDM reaction steps using QuikChange Lightning kit.

Temperature	Time	Cycles
95°C	2 min	1
95°C	20 sec	30
55°C	30 sec	

*30 sec/kb of plasmid length;
pET-17b vector size: 3300bp
V_H insert size: 588bp
Total: 3888bp~4000bp → 2.5 min

65°C	2.5 min*	
65°C	5 min	1
4°C	Hold	∞

The reaction mixture was kept on hold at 4°C indefinitely (**Table 21**). All the steps of the procedure were carried out under nuclease-free conditions.

3.6.4.2. SDM with Q5 high-fidelity kit

50 ng/µl of dsDNA templates (pET17-b bacterial expression plasmids) were amplified by using Q5 High-Fidelity 2X Master Mix (Q5 High-Fidelity reaction buffer, Q5 High-Fidelity DNA polymerase, dNTPs, Mg^{+2}), forward and reverse primers (50 ng/ml) and DMSO. For the conventional PCR amplification, reaction mixture was prepared using the reagents as on the **Table 22**, then gently and homogenously mixed. To the each PCR tube, 24 µl of reaction mixture and 1µl of 50 ng/µl dsDNA template were added.

Table 22. Reagents for SDM reaction using Q5 High-Fidelity kit.

Reaction reagents (templates <5 kb)	**Volume (µl)**
Q5 High-Fidelity 2X Master Mix	12.5
ddH_2O	7
dsDNA template (50 ng/µl)	1
Forward primer (50 ng/µl)	2
Reverse primer (50 ng/µl)	2
DMSO (2%)	0.5
Total volume	**25**

The cover temperature of the thermal cycler was set to 102°C. Thermal-cycler was programmed to initial denaturation at 98°C for 2 min, followed by 30 cycles of denaturation at 98°C for 10 sec, annealing at 55°C for 30 sec (depending on annealing temperature of primers and product size), elongation at 72°C for 2.5 min and 5 min of final extension at 72°C (**Table 23**).

Table 23. SDM reaction steps using Q5 High-Fidelity kit.

Temperature	**Time**	**Cycles**
98	2 min	1
98	10 sec	30
55	30 sec	

*30 sec/kb of plasmid length;
pET-17b vector size: 3300bp
V_H insert size: 588bp
Total: 3888bp~4000bp → 2.5 min

72	2.5 min*	
72	5 min	1
4	Hold	∞

The reaction mixture was kept on hold at 4°C indefinitely (**Table 23**). All the steps of the procedure were carried out under nuclease-free conditions.

3.6.4.3. SDM with KOD hot start kit

50 ng/µl of dsDNA templates (pET17-b bacterial expression plasmids) were amplified by using KOD Hot Start Kit (KOD Hot Start DNA polymerase, dNTPs, reaction buffer with $MgSO_4$), forward and reverse primers (50 ng/ml) and DMSO. For the conventional PCR amplification, reaction mixture was prepared using the reagents as on the **Table 24**, then gently and homogenously mixed. To the each PCR tube, 24 µl of reaction mixture and 1µl of 50 ng/µl dsDNA template were added.

Table 24. Reagents for SDM reaction using KOD Hot Start kit.

Reaction Reagents	Volume (µl)
10X Buffer for KOD Hot Start DNA Polymerase	2.5
Sterile dH_2O (PCR grade)	15.5
dNTPs (2 mM each)	2.5
$MgSO_4$ (25 mM)	1.5
Forward primer (10 µM)	0.75
Reverse primer (10 µM)	0.75
KOD Hot Start DNA Polymerase (1 U/µl)	0.5
dsDNA template (20 ng)	1
Total volume	**25 µl**

The cover temperature of the thermal cycler was set to 99°C. Thermal-cycler was programmed to initial denaturation at 95°C for 2 min, followed by 18 cycles of denaturation at 95℃ for 20 sec, annealing at 60°C for 30 sec (depending on annealing temperature of primers), elongation at 70°C for 3 min and 5 min of final extension at 72°C (**Table 25**).

Table 25. SDM reaction steps using KOD Hot Start kit.

Temperature	Time	Cycles
95	2 min	1
95	20 sec	18

60	30 sec	
70	3 min	
72	4 min	1
4	Hold	∞

The reaction mixture was kept on hold at 4°C indefinitely. All the steps of the procedure were carried out under nuclease-free conditions. Most of the mutations introduced were generated by using KOD enzyme.

3.6.5. DpnI enzyme digestion

After performing SDM, it is crucial to digest the reaction mixture with DpnI enzyme. It specifically recognizes and cleaves methylated DNA sequences. It is used to eliminate the parental, methylated dsDNA template that remains in the mixture after the SDM reaction. By digesting the methylated dsDNA template with DpnI, it will be selectively cleaved, while the newly synthesized, unmethylated (mutated) ones remain intact. This digestion step effectively eliminates the dsDNA templates used in the SDM reactions. Either of the following kits was used for DpnI digestion:

3.6.5.1. NEB DpnI kit

In each reaction tube, the components from **Table 26** were added directly. To ensure thorough mixing, the solution should be gently pipetted up and down several times. Afterwards, the reaction mix was briefly spun down before placing the samples in the thermal cycler. The cover temperature of the thermal cycler was set to 42°C. The samples were incubated at 37°C for 30 min to allow for DpnI digestion of the parental (non-mutated) dsDNA.

Table 26. Reagents for DpnI digestion reaction using NEB DpnI kit.

Reaction Reagents	Volume (µl)
PCR reaction	~25-30
10X rCutSmart™ Buffer	3
NEB DpnI restriction enzyme	0.6

3.6.5.2. Thermo DpnI kit

In each reaction tube, the components from **Table 27** were added directly. The reaction mixtures were gently mixed by pipetting the solution up and down several times. Subsequently, the reaction mix is briefly spun down before placing the samples in the PCR

apparatus. The cover temperature of the thermal cycler was set to 42°C. The reaction volume was adjusted to 30 µl. The samples were then incubated at 37°C for 10 min to facilitate DpnI digestion of the parental dsDNA.

Table 27. Reagents for DpnI digestion reaction using Thermo 10X Fast Digest kit.

Reaction Reagents	Volume (µl)
PCR reaction	~25-30
Thermo 10X Fast Digest Buffer	3
Thermo Fast DpnI	1.5

3.6.6. PCR clean up - QIAquick gel extraction kit

The clean-up procedure following SDM is important step for the removal of DMSO remaining in the SDM mix. DMSO is commonly used in SDM reactions to enhance primer annealing and therefore improving mutagenesis efficiency. However, the presence of DMSO can negatively affect downstream applications, such as DNA sequencing. Therefore, to ensure the purity of the mutated dsDNA, DMSO and other contaminants (SDM reagents) were selectively removed by using a special kit to clean up the SDM reaction mixture.

QIAquick Gel Extraction Kit for cleaning DNA after PCR reactions was used. The following steps were carried out based on the instructions of the manufacturer. Buffer QG was directly added to the PCR reaction at a ratio of 5 volumes to 1 volume and mixed. The total reaction volume was determined to be approximately 30 µl, including the DpnI digestion reaction mix, with a 5X reaction volume of 150 µl. If the mixture exhibited an orange or violet colour, 10 µl of 3 M sodium acetate with a pH of 5.0 was added and mixed. Consequently, the colour of the mixture would turn yellow. Subsequently, a QIAquick spin column was placed in the provided 2 ml collection tube. The sample was applied to the QIAquick column to allow for DNA binding. Centrifugation was then performed for 1 min at room temperature (RT) at 13,000 rpm, in a conventional table-top microcentrifuge until all samples had passed through the column. For sample volumes exceeding 800 µl, an additional loading and spin step were conducted. The flow-through was discarded, and the QIAquick column was returned to the same tube. For the washing step, 750 µl of Buffer PE was added to the QIAquick column. In cases where the DNA would be used for salt-sensitive applications, such as sequencing, the column was allowed to stand for 2-5 min after the addition of Buffer PE. The flow-through was again discarded, and the QIAquick column was transferred to a clean 1.5 ml microcentrifuge tube. To elute the DNA, 30 µl of Buffer EB (10

mM TrisCl, pH 8.5) was added to the center of the QIAquick membrane. Alternatively, dH_2O could be used instead of Buffer EB. The column was left to stand for 3 min to enhance the yield of purified DNA, followed by centrifugation for 1 min at RT and at 13,000 rpm. The concentration of the purified DNA plasmid was measured using the NanoDrop spectrophotometer with calculated molecular weight and extinction coefficient, with elution buffer used as a blank. A portion of 5 µl of purified mutant DNA was reserved and stored at 4°C for the following transformation experiment, while the remaining portion of approximately 25 µl was stored at -20°C as a backup.

3.6.7. Agarose gel electrophoresis

Agarose gel electrophoresis is a widely used technique for validating the accuracy of the SDM process by visualizing DNA fragments present in the SDM mix. After the SDM reaction, the DNA samples, (the mutated dsDNA and any remaining parental dsDNA), were loaded onto an agarose gel. Visualization of the gel using was done under UV light to ensure the absence of any unintended DNA products, such as primer dimers or DNA super-coils.

The electrophoresis system was constructed by assembling the casting tray and holders. To prepare the 1XTAE buffer, the following steps were followed at **Table 28**:

Table 28. Preparation of 50X stock TAE Buffer and 1X TAE Buffer.

50X TAE Buffer, 100 ml

Tris base, 24.2 g	Tris base weighing 24.2 g was mixed with acetic acid and DMSO. The mixture was dissolved in deionized water (ddH_2O). To ensure sterility, the solution was filtered. The final solution was then stored at RT.
%100 Acetic acid, 5.71 ml	
0.5M DMSO, 10 ml	
ddH_2O, ~85 ml	

1X TAE Buffer, 1000 ml

50X TAE Buffer, 20 ml	The 50XTAE Buffer was diluted with ddH_2O.
ddH_2O, 980 ml	

To prepare the agarose gel, the following steps were followed at **Table 29**:

Table 29. Preparation of Agarose gel, 1% (w/v) with SYBR Safe.

Agarose gel, 1% (w/v), 50 ml

Agarose powder 1% (w/v), 0.5 g	In a flask, 0.5 g of agarose powder was weighed and mixed with 50 ml of 1X TAE buffer. From 4°C storage, 10,000X SYBR Safe was added to the mixture. To
1X TAE Buffer, 50ml	
10.000X SYBR Safe, 5 µl	

	ensure thorough incorporation of the dye, it was carefully mixed.

Agarose powder and 1XTAE buffer were mixed in a glass flask. The flask was then heated in the microwave for 1-3 min, with intermittent interruptions to shake and ensure complete dissolution of the powder. Subsequently, 5 µl of 10,000X SYBR Safe was added and carefully swirled into the solution. The gel was poured into the casting tray, taking care to avoid bubbles and removing any present towards the bottom. The gel was left to solidify for approximately 30 min. The comb was gently removed. The casting tray was then placed into the electrophoresis chamber. The tank was filled with 1X TAE buffer, submerging the gel with the buffer level approximately 0.5 cm above the gel. DNA samples were prepared for loading by mixing them with 5X Dye, dH_2O, and DNA in a microcentrifuge tube. The samples were then loaded into the wells along with the ladder. The gel was run at a constant voltage of 80-100V for 70-90 min or until the bands reached the bottom end. The gel was visualized under UV light, and the image was captured. After visualizing the gel and removing the upper part containing the super-coiled DNA plasmid, the gel clean up procedure was continued.

3.6.7.1. Gel clean up- QIAquick gel extraction kit

Gel cleanup is a necessary step after agarose gel electrophoresis to purify dsDNA fragments from the contaminants coming from the agarose gel, to ensure the purity of the sample for downstream applications.

QIAquick Gel Extraction Kit for cleaning DNA after PCR reactions was used. The following steps were carried out based on the instructions of the manufacturer. The DNA fragment excised from the agarose gel was carefully cut with a clean, sharp scalpel. An empty sterile Eppendorf tube was weighed. The gel piece containing the mutant DNA plasmid was transferred to the Eppendorf tube and re-weighed. To this, 3 volumes of Buffer QG were added to 1 volume of gel, ensuring not to exceed 400 µl of gel per spin column. The Eppendorf tube was incubated at 50°C for 10 min, with intermittent vortexing to aid gel dissolution. After complete dissolution, the mixture was checked for a yellow colour. If the colour changed to orange or violet, 10 µl of 3 M sodium acetate, pH 5.0, was added and mixed. Following this, 1 gel volume of isopropanol was added, and the sample was mixed. The processed sample was applied to a QIAquick spin column, and centrifuged at 13,000 rpm for 1 min at RT. The flow-through was discarded, and the column was washed with 500 µl of

Buffer QG. For subsequent applications like sequencing, an additional step involved the addition of 500 µl Buffer QG and subsequent centrifugation. After washing with 750 µl Buffer PE, the QIAquick column was placed in a clean microcentrifuge tube. To elute DNA, 30 µl of Buffer EB was added to the center of the QIAquick membrane, followed by centrifugation at 13,000 rpm for 1 min. The concentration of the purified DNA plasmid was measured using the NanoDrop spectrophotometer. For further use, 5 µl of purified DNA was stored at 4°C, while the remaining portion was stored at -20°C as a backup.

3.6.8. Transformation

Bacterial transformation is a pivotal step for recombinant protein production. Via transformation, SDM products were introduced into bacterial hosts for their expression recombinantly. Several methods are commonly used for bacterial transformation. These include:

(i) Chemical Transformation: Bacterial cells are treated with chemicals that make the cell membrane more permeable, allowing the uptake of foreign DNA.

(ii) Electroporation: A brief electric pulse is applied to the bacterial cells, creating transient pores in the host cell membrane where foreign DNA can enter.

(iii) Heat Shock Transformation: A brief heat shock is applied to increase the permeability of bacterial cells, so that foreign DNA can pass through the enlarged bacterial pores.

Both electroporation and heat shock transformation techniques were employed in our procedures. The selection criteria depend on the following downstream application.

3.6.8.1. Electroporation

Bacterial transformation using electroporation was initiated by taking electro-competent XL-10 Gold *E.coli* cells from -80°C storage and thawing them on ice. For each sample to be electroporated, a 1.5 ml micro-centrifuge tube and a 0.1 cm BioRad *E. coli* Pulser Cuvette were placed on ice. In a cold Eppendorf tube, the cell suspension was mixed with 5 µl of plasmid for the SDM procedure or 1 µl of plasmid for the standard transformation procedure with XL-10 Gold *E.coli* cells. The tube was gently flicked to mix the contents without pipetting up and down. The cell mix was incubated on ice for 1 min. The electroporator was turned on, and the BioRad MicroPulser is set to "Ec1" when using 0.1 cm cuvettes. The cell mixture was transferred to a cold electroporation cuvette, and the cuvette was tapped gently to move the suspension to the bottom, ensuring contact with both electrodes. The lid of the cuvette is closed, and it was placed on ice for 1 min. The cuvette

was pushed into the device by sliding it into the electroporation chamber. The pulse was triggered using the device, and immediately afterward, the lid was removed, and 1 ml of SOC recovery media (**Table 30**) was added to the cells at the bottom of the cuvette.

Table 30. Preparation of SOC media and preceding SOB media.

To be able to make SOC media, SOB media should be prepared first;

SOB medium for SOC media, 1000 ml

Tryptone, 20 g **Yeast extract, 5 g** **NaCl, 0.5 g** **dH_2O, 950 ml**	The ingredients were weighed. They were added into the glass beaker. They were dissolved in 950 ml of dH_2O.

10 ml of 250 mM KCl solution was added to SOB medium.

250 mM KCl for SOC media, 100 ml

KCl, 1.86 g	The ingredient was weighed. It was added into the glass beaker. It was dissolved in 100 ml of dH_2O.
dH_2O, 100 ml	

The pH of the medium was adjusted to 7.0 with 5 N NaOH (~0.2 ml). The volume of the solution was adjusted to 1 L with dH_2O. The solution was sterilized by autoclaving. Just before use, 5 ml of a sterile solution of 2 M $MgCl_2$ was added.

2 M $MgCl_2$ for SOC media, 100 ml

$MgCl_2$, 19 g	The ingredient was weighed. It was added into the glass beaker. It was dissolved in 100 ml of dH_2O. The solution was sterilized by autoclaving.
dH_2O, 100 ml	

20 ml of sterile 1 M glucose was added to the sterilized solution.

1 M glucose, 100 ml

Glucose, 18 g	The ingredient was weighed. It was added into the glass beaker. It was dissolved in 100 ml of dH_2O. The solution was filtered using a 0.22 µm filter for sterilization.
dH_2O, 100 ml	

The cell suspension was gently resuspended within the media by pipetting up and down 1-2 times. The cell suspension was then transferred into a sterile Eppendorf tube. The pulse parameters (kV and ms) were checked from the device screen by pressing the "Measurements" button, and the applied pulse time should be close to 5 ms with a voltage of around 1.8 V. If the voltage or time value is critically lower than the reference values, the procedure is repeated using a new cuvette. The cells were incubated at 37°C with shaking at

225 rpm for 1 h. Meanwhile, the previously prepared agar plates with the appropriate antibiotic (**Table 31**) were warmed up. The cells are diluted as appreciated. The transformed cells were then plated on pre-warmed selective agar plates for growth in a 37°C incubator, using a volume ranging from 50 µl to 500 µl.

Table 31. Preparation of LB Agar with AMP and CAM for electrocompetent cells.

LB Agar with 100 µg/ml AMP and 25 µg/ml CAM (100 ml)

LB Agar, 4 gr	LB Agar powder was weighed and added to dH_2O in a glass bottle. The mixture was then sterilized in an autoclave. Before pouring the agar into petri dishes, AMP/CAM antibiotics were added. The petri dishes were stored in the fridge.
dH_2O, 100 ml	

100 mg/ml AMP (10 ml) for LB Agar

AMP, 1 g	The powder was weighed and dissolved in dH_2O in a glass bottle. To ensure sterilization, a filter was used. The solution was then aliquoted into Eppendorf tubes in 1 ml volumes. Finally, the aliquots were stored at -20°C.
dH_2O, 10 ml	

25 mg/ml CAM (10 ml) for LB Agar

CAM, 0.25 g	The powder was weighed and dissolved in ethanol in a glass bottle. The mixture was not filter. The solution was then aliquoted into Eppendorf tubes in 1 ml volumes. The aliquots were stored at -20°C.
Ethanol, 10 ml	

3.6.8.1.1. Electrocompotent cell making-BioRad

On day 1: The frozen glycerol stock of bacterial cells was streaked out onto an LB Agar plate with CAM (prepared as in **Table 31**). The plate was then grown overnight at 37°C to allow for colony formation.

On Day 2: A starter culture of cells was prepared by selecting a single colony of *E. coli* from a fresh LB plate and inoculating a 10 ml starter culture of LB broth (**Table 32**). The culture was grown overnight in a shaker at 37°C.

Table 32. Preparation of LB broth.

LB Broth (1000 ml)

LB Broth, 15.5 g	The powder was weighed and dissolved in dH_2O in a glass bottle. The

dH_2O, 1000 ml	solution was sterilized in an autoclave. AMP/CAM antibiotics were added before use.

On Day 3: 500 ml of LB Broth was inoculated with a 1/100 volume of a fresh overnight *E. coli* culture. The cells were grown at 37°C with shaking at 300 rpm until reaching an Optical desnsity at 600 nm wavelength (OD_{600}) of approximately 0.5-0.7. To maintain low temperature, the cells were chilled on ice for approximately 20 min. All containers were also chilled in ice before adding cells. To harvest, the cells were transferred to a cold centrifuge bottle and spun at 4,000xg for 15 min at 4°C. The supernatant was carefully poured off and discarded. The pellet was gently resuspended in 500 ml of ice-cold 10% glycerol (**Table 33**). It was then centrifuged at 4,000xg for 15 min at 4°C, and the supernatant was poured off and discarded. This resuspension and centrifugation process was repeated with 250 ml of ice-cold 10% glycerol.

Table 33. Preparation of 10% (v/v) glycerol.

10% (v/v) glycerol (1000 ml)	
Glycerol, 100 ml **dH_2O, 900 ml**	The liquid was measured and dissolved in dH_2O in a glass bottle. The solution was sterilized in an autoclave.

Next, the pellet was resuspended in approximately 20 ml of ice-cold 10% glycerol and transferred to a 30 ml sterile Falcon tube. Centrifugation was performed at 4,000xg for 15 min at 4°C, and the supernatant was carefully poured off and discarded. Finally, the cell pellet was resuspended in a final volume of 1-2 ml of ice-cold 10% glycerol, aiming for a cell concentration of about 1-3 x 10^{10} cells/ml. This suspension could be aliquoted and frozen on dry ice for long-term storage at -80°C. The cells remained stable for at least 6 months under these conditions.

3.5.7.1.2. Electrocompotent cell making-Custom

On Day 1: The frozen glycerol stock of bacterial cells was streaked out onto an LB Agar plate with CAM, and it was incubated overnight at 37°C to allow for colony formation.

On Day 2: The following materials and media/solutions were autoclaved: 2 L of ddH_2O, 100 ml of 10% glycerol (**Table 33**), 1 L LB media (**Table 32**), and 4 centrifuge bottles and caps. Additionally, the following materials and media/solutions were chilled overnight at 4°C: ddH_2O, 10% glycerol, and the centrifuge rotor. A starter culture of cells was

prepared by selecting a single colony of *E. coli* from a fresh LB plate and inoculating a 10 ml starter culture of LB, which was grown overnight at 37°C in a shaker.

On Day 3: 1 L of LB media was inoculated with 10 ml of the starter culture and grown in a 37°C shaker. The OD_{600} was measured regularly, especially when it exceeded 0.2. When the OD_{600} reached 0.35-0.4, the cells were immediately chilled on ice for 20-30 min, with occasional swirling for even cooling. Centrifuge bottles were placed on ice, and the 1 L culture was divided into four parts by pouring approximately 250 ml into ice-cold centrifuge bottles. The cells were harvested by centrifugation at 1,000xg for 20 min at 4°C. The supernatant was decanted, and each pellet was resuspended in 200 ml of ice-cold ddH_2O. A second round of centrifugation at 1,000xg for 20 min at 4°C was performed. The supernatant was decanted again, and each pellet was resuspended in 100 ml of ice-cold ddH_2O. The re-suspensions were combined into two centrifuge bottles, each containing about 200 ml of cell suspension. The cells were harvested again by centrifugation at 1,000xg for 20 min at 4°C. Two 50 ml Falcon tubes were chilled on ice and rinsed with ddH_2O. The supernatant was decanted, and each pellet was resuspended in 40 ml of ice-cold 10% glycerol. Each suspension was transferred to a 50 ml Falcon tube. A third round of centrifugation at 1,000xg for 20 min at 4°C was performed. Microfuge tubes (1.5 ml) were chilled on ice. The supernatant was carefully aspirated, and each pellet was resuspended in 1 ml of ice-cold 10% glycerol by gently swirling. The final OD_{600} of the resuspended cells should be approximately 200-250. The cells were aliquoted into sterile 1.5 ml microfuge tubes and flash-frozen with liquid nitrogen. Finally, the frozen cells were stored in the -80°C freezer.

3.6.8.2. Heat-shock transformation- BL21(DE3) pLysS / DH5α

The BL21(DE3) pLysS or DH5α competent *E. coli* cells stored at -80°C were used for the transformation procedure. The competent cells were thawed on ice for 5-10 min and were kept on ice throughout the procedure. It was always ensured to store the competent cells at -80°C, and they were not subjected to more than two freeze-thaw cycles. For each tube of transformation, 50 µl of cells were required. Meanwhile, the Eppendorf tubes to be used for transformation were chilled on ice prior to the experiment. After thawing, the competent cells were gently mixed, and 50 µl was taken for each Eppendorf tube. To each tube containing the competent cells, 2.1 µl of 0.142 M 2-ME (**Table 34**) from the fridge at 4°C was added. The tube contents were pipetted up and down and gently swirled.

Table 34. Preparation of 0.14 M β-ME.

0.14 M β-ME (1:1000 diluted)	
Stock β-ME (14.2 M), 100 µl	The liquid was measured and diluted in dH_2O in a falcon tube. The solution was filtered for sterilization.
dH_2O, 9.9 ml	

The reactions were incubated on ice for 5-10 min, with gentle swirling every 2 min. Meanwhile, the water bath was pre-heated to 42°C for following step. To each transformation reaction tube, 3 µl of 50 ng/µl of the appropriate plasmid DNA were added, followed by gentle swirling. For the control, only 1 µl of pUC18 control plasmid was added. The reaction mixtures were incubated on ice for 20-30 min. Meanwhile, the SOC medium (RT) was pre-heated in a 42°C water bath. Additionally, the heat block was warmed up to 37°C. Each transformation mixture was heat-pulsed in the 42°C water bath for exactly 45 sec. Immediately after heat-pulsing, the Eppendorf tubes were transferred to ice. The reaction mixtures were incubated on ice for 2 min. To each tube, 500 µl of SOC medium was added. The reaction mixtures were incubated at 37°C in a heat block for 1 h, shaking at 225 rpm. Meanwhile, the LB agar plates prepared in advance and stored in the fridge at 4°C were pre-warmed to RT. After the incubation, 250 µl of the transformation mixture was transferred onto LB Agar plates containing CAM and AMP for the protein coding plasmids and control plasmid. In case the transformation efficiency was not achieved, transformation mixtures could be stored in the fridge at 4°C until colony observation was achieved on day 2 or up to a maximum of 3 days. After this, the transformation mixtures were discarded. The LB Agar plates were incubated at 37°C overnight (18-22 h) or until the colonies became visible.

3.6.8.2.1. Competent cell making- DH5α

Day 1: The frozen glycerol stock of bacterial cells was streaked out onto an LB Agar plate with CAM. The plate was incubated overnight at 37°C, allowing colony formation.

Day 2: The following materials and media/solutions were autoclaved: 100 ml Transformation broth (**Table 35**), and 4 centrifuge bottles and caps. Additionally, Transformation broth media was chilled overnight at 4°C after autoclave. A starter culture of cells was prepared by selecting a single colony of *E. coli* from a fresh LB plate and inoculating a 10 ml starter culture of LB. The culture was grown overnight at 37°C in a shaker.

Table 35. Preparation of Transformation Broth media.

Transformation Broth media, 100 ml

10 mM Pipes, 0.302 g **15 mM $CaCl_2.2H_20$, 0.221 g** **250 mM KCl, 1.864 g** **90 ml dH_20** **55 mM $MnCl_2.4H_20$, 1.71 g**	The powders were weighted and dissolved in in a glass flask using dH_2O. The pH was set to 6.7 adding NaOH. $MnCl_2$ was added to adjust the volume to 100 ml. The solution filtered using 0.22 µm filter for sterilization.

Day 3: In 250 ml flasks, 2x50 ml of LB broth with CAM were inoculated with DH5α at OD_{600}:0.01 and OD_{600}:0.05. The cells were grown at 16°C and 150 rpm overnight.

Day 4: When the OD_{600} reached 0.4-0.5, the cells were immediately put on ice and chilled for 20-30 min, with occasional swirling to ensure even cooling. The cells were then placed in centrifuge bottles on ice. The stored cells in the fridge were transferred to a 50 ml Falcon flask. The cells were centrifuged at 2,500xg for 10 min at 4°C, and the supernatant was removed, keeping the cells. The pellet was washed with 10 ml of Transformation broth media.

The Falcon tube was put on ice for 5 min, followed by centrifugation at 2500xg for 10 min at 4°C. The supernatant was removed, and the cells were kept. The cells were resuspended in 4 ml of Transformation broth. Drop by drop, 0.3 ml of DMSO was added to the cells while gently shaking. The cells were aliquoted in 100 and 200 µl on ice. The cells were flash frozen and stored at -80°C for a couple of months until the competency of the cells decreased.

3.6.9. Plasmid isolation

Plasmid isolation involves extracting circular DNA from bacterial cells for its downstream use. Various commercial kits have been extensively used to simplify this process.

3.6.9.1. MN plasmid DNA purification kit

The following steps were carried out based on the instructions of the manufacturer. Bacterial cells were cultivated and harvested by utilizing 10 ml of a confluent *E. coli* LB culture, followed by pelleting cells in a standard bench-top microcentrifuge for 30 sec at 10,000 rpm. The supernatant was then discarded, by ensuring to remove as much liquid as possible.

The subsequent step involved cell lysis: 250 µl of Buffer A1 was added to completely resuspend the cell pellet through vortexing or pipetting. 250 µl of Buffer A2, was warmed up

and added gently to the mixture. The tube was mixed gently by inverting it 6–8 times. The tube was incubated at RT for up to 5 min or until clear lysate appears. 300 µl of Buffer A3 was added to the mixture. The tube was mixed gently by inverting it 6–8 times without vortexing. To ensure complete neutralization, LyseControl should turn completely colorless without any traces of blue. The mixture was centrifuged for 5 min at 10,000 rpm at RT. This step was repeated if the supernatant remained unclear.

DNA binding was carried out by placing a NucleoSpin® Plasmid/Plasmid Column in a Collection Tube provided by the kit. The supernatant was decanted onto the column, followed by centrifugation for 1 min at 10,000 rpm. The flow-through was discarded and the tube was placed back into the collection tube.

The silica membrane was washed with 500 µl of pre-heated Buffer AW to 50°C. The tube was centrifuged for 1 min at 10,000 rpm. The flow-through was discarded and the tube was placed back into the collection tube. The membrane was further washed with 600 µl of Buffer A4. The tube was centrifuged for 1 min at 10,000 rpm. The flow-through was discarded and the tube was placed back into the collection tube.

The silica membrane was dried through by centrifugation for 2 min at 10,000 rpm to remove any residual ethanolic wash buffer.

Finally, DNA elution was achieved by placing the NucleoSpin® Plasmid/Plasmid Column in a fresh 1.5 ml microcentrifuge tube and adding 30 µl Buffer AE. The tube was incubated at RT for 3 min. The tube was centrifuged for 1 min at 10,000 rpm. Alternative elution buffers, such as TE buffer or water, could be used depending on the downstream application. The DNA concentration was subsequently measured using NanoDrop spectrophotometer.

3.6.9.2. EZ-10 spin column plasmid mini-prep kit

The following steps were carried out based on the instructions of the manufacturer. The 10 ml of overnight-grown bacterial cell culture was centrifuged at 3,900 rpm for 15 min. The supernatant was discarded, leaving behind the cell pellet. 200 µl of Solution I was added to the pellet, and the mixture was transferred to an Eppendorf tube. The tube was left incubation for 1 min. Subsequently, 400 µl of Solution II was introduced to the Eppendorf tube, followed by gentle mixing. 350 µl of Solution III was then added and gently mixed. The Eppendorf tubes were centrifuged at 13,000xg for 15 min.

The resulting supernatant was carefully transferred to an EZ-1 column provided by the kit, and the column was centrifuged at 11,000xg for 2 min. The flow-through was discarded. 750 µl of Washing Solution was added to the column. The tube was centrifuged at 11,000xg for 2 min. This step was repeated twice. The flow-through was discarded again, and the column was centrifuged at 11,000xg for 2 min.

The column was transferred to a fresh 1.5 ml Eppendorf tube, and pre-heated 30 µl of Elution Buffer was added to the column. After incubating the column for 5 min at RT, the tube was centrigued at 11,000xg for 5 min. The DNA concentration was subsequently measured using NanoDrop spectrophotometer.

3.6.9.3. Qiagen QIAprep® spin miniprep kit

The following steps were carried out based on the instructions of the manufacturer. The 10 ml of bacterial overnight culture was pelleted through centrifugation >8,000 rpm for 3 min at RT (15–25°C). The pelleted bacterial cells were resuspended in 500 µl of Buffer P1 and then transferred to a microcentrifuge tube. Subsequently, 250 µl of Buffer P2 was added, and thorough mixing was accomplished by inverting the tube 4–6 times until the solution became clear. The lysis reaction should not exceed 5 min. If LyseBlue reagent was used, the solution turned blue at this stage. The 350 µl of Buffer N3 was added to the mixture and it was mixed by inverting the tube 4–6 times, resulting in the solution turning colorless if LyseBlue reagent was used. The tube was centrifuged at 13,000 rpm for 10 min in a table-top microcentrifuge.

The 800 µl of the supernatant from the previous step was applied to the QIAprep 2.0 spin column provided by the manufacturer by pipetting. The tube was centrifuged for 1 min at 13,000 rpm. The flow-through from the QIAprep 2.0 spin column was discarded. Washing steps were carried out by adding 500 µl of Buffer PB. The tube was centrifuged for 1 min at 13,000 rpm. The 750 µl of Buffer PE was added than the centrifugation step was repeated. The QIAprep 2.0 spin column was transferred to the collection tube. To remove residual wash buffer, the tube was centrifuged at 13,000 rpm for 1 min. The QIAprep 2.0 column was placed in a clean 1.5 ml microcentrifuge tube.

To elute DNA, 30 µl of Buffer EB (10 mM TrisCl, pH 8.5) or water was added to the center of the QIAprep 2.0 spin column, allowing it to stand for 3 min (1-5 min). The tube was centrifuged for 1 min at 6700 rpm. The DNA values were then measured using NanoDrop spectrophotometer.

3.6.10. Protein production and purification

Expression is started from a single colony of *E. coli* BL21(DE3) or WK6 either freshly transformed with expression vector or plated out on LB agar plates with AMP/CAM or only AMP containing Bevacizumab V_H gene cloned into pET17b or pHen6-derived vector (CC13 or CC1168). Upon expression, the V_H is fused to a PelB signal sequence and a His_6-tail at the N- and C-terminus, respectively when pHen6 plasmid is used. Once target protein expression is initiated in the pET17b plasmid, the PelB sequence, His_7-tag, and $FLAG_3$-tag are simultaneously incorporated. The Bevacizumab V_H is translocate to the periplasmic space of the *E. coli* BL21(DE3) or WK6 cells with subsequent cleavage of the PelB signal sequence. The periplasmic extract containing the His_6-tagged V_H is prepared via osmotic shock using *E. coli* WK6 cells. Either the leaky expression or periplasmic extract products collected from supernatant of *E. coli* BL21(DE3) cells. Using IMAC, the His_6 or His_7 tagged V_Hs are captured from the periplasmic extract and/or supernatant (**Figure 19**).

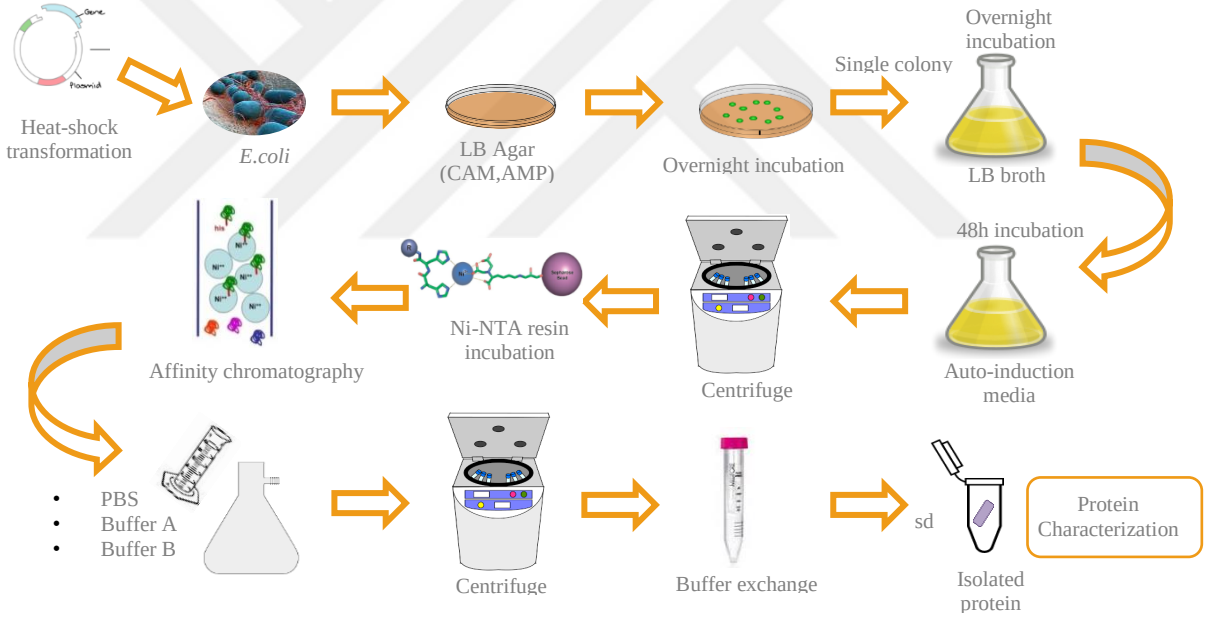


Figure 19. The workflow diagram of the sdAb production and purification applied in the home lab.

3.6.10.1. Protein production and purification- BL21(DE3) pLysS

Day 1: Starter culture preparation

The LB Agar plates were checked for colony formation of transformed cells. The LB Broth was prepared for autoclaving at least a day before the experiment. To complete the LB Broth recipe, CAM and AMP were added to the sterilized portion of the broth.

The LB Broth was pre-warmed and the incubator was set up to 37°C to be used in the following step, before the experiment was started. A volume of 50 ml of LB Broth was added into a 250 ml baffled flask. CAM and AMP antibiotics were added before inoculation. A single colony or ten bacterial colonies were picked from a plate using a pipette tip or a loop. Alternatively, a loopfull of frozen bacterial media of a single colony in 50% glycerol stock can be used. The colonies were transferred into the LB Broth-containing baffled flask or the pipette tip was left within the flask. The culture was incubated shaking at 130-225 rpm at 37°C overnight until the broth became opaque, indicating colony formation. The auto-induction media (Studier 2005) was prepared for sterilization on Day 2.

Day 2: Expression of sdAbs in E. coli BL21(DE3) pLysS periplasm

The turbidity of LB Broth bacterial culture was checked as an indication of cell growth. The baffled flask was taken from the incubator. A sample of 200 µl was taken from the flask and transferred into a 96-well plate. The absorbance values at OD_{600} were measured in a spectrophotometer. Using the absorbance values and the correction factor (calculated empirically before), the volume of cells required to reach 0.05 OD_{600} in the auto-induction medium was calculated. 300 ml of complete auto-induction media was prepared by adding the remaining reagents into a readily sterilized portion within a 1000 ml flask (**Table 36**).

Table 36. Preparation of home-made auto-induction media.

Auto-induction media (300 ml)

Tryptone, 3 g **Yeast Extract, 1.5 g** **dH_2O, 277.5 ml**	Tryptone and yeast extract powders were weighed and dissolved in dH_2O in a glass flask. The mixture was sterilized in an autoclave. Other remaining ingredients were added before use.
20X NPS, 15 ml **50X 50/52, 6 ml** **1M $MgSO_4$, 300 µl** **CAM, 300 µl** **AMP, 300 µl**	After the sterilized portion of the auto-induction media cooled down, the liquid reagents were added to fill up the total volume to 300 ml. The auto-induction media was prepared just before use.

The following solutions were prepared prior to use in complete auto-induction media.

20X NPS (pH 6.8) for auto-induction media (100 ml)

$(NH_4)_2SO_4$, 6.6 g **KH_2PO_4, 13.6 g** **Na_2HPO_4, 14.2 g**	The powders were weighed and dissolved in dH_2O in a glass bottle. A filter was used for sterilization of the solution.

dH_2O, 100 ml	

20X NPS (pH 7.8) for Auto-induction media (100 ml)

$(NH_4)_2SO_4$, 6.6 g	The powders were weighed and dissolved in dH_2O in a glass bottle. A filter was used for sterilization of the solution.
KH_2PO_4, 2.04 g	
Na_2HPO_4, 25.99 g	
dH_2O, 100 ml	

50X 50/52 for Auto-induction media (100 ml)

Glycerol, 25 ml	The powders were weighed and dissolved in dH_2O in a glass bottle. A filter was used for sterilization of the solution.
Glucose, 2.5 g	
α-Lactose monohydrate, 10 g	
dH_2O, 100 ml	

1 M $MgSO_4$ for Auto-induction media (50 ml)

$MgSO_4$, 6.02 g	The powders were weighed and dissolved in dH_2O in a glass bottle. A filter was used for sterilization of the solution.
dH_2O, 50 ml	

The appropriate amount of bacterial culture was inoculated into the complete auto-induction media as calculated previously. Immediately after inoculation, samples of 3-4 ml were taken and transferred to a falcon tube to measure absorbance and pH values at the time of inoculation (0th h). The values were recorded as data. From the falcon tube, a 200 µl sample was taken and transferred into a 96-well plate. The absorbance values of OD_{600} were then measured in a spectrophotometer. The bacterial culture was incubated shaking at 200-225 rpm at 18°C for 48 h in an incubator with a cooling system, allowing the cells to produce the protein of interest.

Day 3: Expression of sdAbs in *E. coli* BL21(DE3) pLysS periplasm

Samples of 3-4 ml were taken from the culture and transferred to a falcon tube to measure absorbance and pH values at the 24th h of inoculation. The values obtained were recorded as data. From the falcon tube, a 200 µl sample was taken and transferred into a 96-well plate. The absorbance values were measured in a spectrophotometer at OD_{600}. The bacterial culture was continued to be incubated shaking at 200-225 rpm at 18°C for a total of 48 h in an incubator with a cooling system, allowing the cells to continue producing the

protein of interest. In preparation for the purification on the following day, Buffer A and Buffer B solutions (**Table 37**) were prepared and chilled in advance.

Table 37. Preparation of Buffer A, Buffer B and stock solutions of imidazole and PBS.

Buffer A (250 ml)

20X PBS, 12.5 ml **2M Imidazole, 3.75 ml** **dH_2O, 233.75 ml**	The reagents were mixed in a glass bottle. It was not necessary to filter them for sterilization since imidazole was already filtered and 20X PBS was already sterilized by autoclaving.

Buffer B (50 ml)

20X PBS, 2.5 ml **2M Imidazole, 12.5 ml** **dH_2O, 35 ml**	The reagents were mixed in a glass bottle. It was not necessary to filter them for sterilization since imidazole was already filtered and 20X PBS was already sterilized by autoclaving.

2M Imidazole for Buffer A and Buffer B (250 ml)

Imidazole, 34.04 g **dH_2O, 250 ml**	The imidazole powder was weighed and dissolved in dH_2O in a glass bottle. A filter was used for sterilization of the solution.

20X PBS for Buffer A and Buffer B (1000 ml)

NaCI, 160 g **KCI, 4 g** **KH_2PO_4, 4.8 g** **Na_2HPO_4, 28.8 g** **dH_2O, 1000 ml**	The powders were weighed and dissolved in dH_2O in a glass bottle. The pH of the solution was adjusted to 7.4. The sterilization of the solution could be done using a filter or by autoclaving. The sterilized solution was then stored at RT.

Day 4: Leaky expression extract isolation using vacuum purification method with His-Pur (Thermo) Ni-NTA beads

The confluence of the cultures was checked in the morning, and the centrifuge was set to 4°C. After the 48 h of incubation within the auto-induction media, the culture medium was transferred into 250 ml ultra-centrifuge tubes. These tubes were then centrifuged at 4°C and 6500 rpm for 15-20 min. The resulting supernatant was carefully transferred back into the same flask used previously, while keeping the cells at the bottom of the tubes for disposal. If

necessary, a 1 ml sample was taken into a 1.5 ml Eppendorf tube for further SDS-PAGE analysis and stored in the fridge at 4°C. The protein slurry was divided into 50 ml falcon tubes and centrifuged at 3,900-4,000 rpm (maximum speed) for 20 min. This step was repeated three times, each time using fresh falcon tubes. Meanwhile, another centrifuge was set to 4°C in order to prepare the Ni-NTA beads. For each 100 ml of protein slurry, 1 ml of Ni-NTA beads was added to 50 ml falcon tubes. The volume was topped up to 50 ml by adding 1X PBS into the tube (**Table 38**). The tubes were then centrifuged at 4°C and 3,750 rpm for 2 min to wash the beads before use. The supernatant was discarded, and the beads were retained. For the Ni-NTA beads to work on their best affinity the pH of the environment was set to pH 7.0.

Table 38. Preparation of 20X and 1XPBS.

20X PBS (1000 ml)

NaCI, 160 g **KCI, 4 g** **KH_2PO_4, 4.8 g** **Na_2HPO_4, 28.8 g** **dH_2O, 1000 ml**	The powders were weighted and dissolved in dH_2O in a glass bottle. The pH was set to 7.4 0.22µm filter or autoclave was used for sterilization. The solution was stored at RT.

1X PBS (1000 ml)

20X PBS, 50 ml **dH_2O, 950 ml**	The powders were dissolved in dH_2O in a glass bottle. There was no need to filter the solution. The pH was adjusted to 7.4 before use. The solution was then stored at RT.

The supernatant media slurry was mixed with the Ni-NTA beads in the protein slurry flask. This mixture was then placed on a magnetic stirrer in the cold room or within the fridge at 4°C for 2 h, allowing the proteins of interest to bind to the Ni-NTA beads with high affinity. If necessary, a 1 ml sample was taken from the protein-beads mix after incubation for further analysis using SDS-PAGE and stored in the fridge at 4°C.

An alternative option was to extend the bead incubation time to overnight incubation in the fridge at 4°C. After the Ni-NTA incubation, a 10 ml ÄKTA Column was placed into a vacuum filtration tap, and the twist-off bottom closure was removed. The tap was then placed into a 1000 ml vacuum flask. The entire bead slurry was added to the 10 ml ÄKTA Column, and vacuum filtration was initiated by starting the pump. After the complete mixture passed

through the column, the washing stage was initiated. To wash the column, 100 or 200 ml of 1X PBS was added to the column. Once the PBS passed through the column by filtration, 100 or 200 ml of 25 mM Buffer A was added to the column.

The column was then removed from the tap and placed into a 50 ml falcon tube containing 1X PBS at the bottom. 1 ml of Buffer B was added for each 1 ml of beads into the column. The samples were either slowly passed through the column at 4°C fridge or centrifuged at 4°C and 3,750 rpm for 2-3 min. At the same time, the centrifuge was set to 4°C. After the Ni-NTA bead incubation, the samples were eluted into 50 ml falcon tubes containing 1X PBS. The eluent at the bottom of the tubes was then transferred to 15 ml Amicon tubes. The final volume was filled till the upper level of the tube with 1X PBS, and the tubes were centrifuged at 4°C and 3,750 rpm for approximately 12 min. The filter part was removed, and the solution at the bottom was discarded, while the top part, which contained the protein, was retained. This step was repeated three times to remove imidazole by exchanging the buffer solution with 1X PBS. The final volume at the top of the filter was reduced to 250-300 µl, and then the volume was transferred into Eppendorf tubes. The protein concentration was measured using NanoDrop spectrophotometer. Depending on the downstream workflow, the volume was divided into fractions, and flash-frozen for storage at -80°C. Additionally, 15 µl of the protein sample was mixed with 15 µl of Leammli buffer and stored at -20°C for SDS-PAGE analysis. The collected protein concentration was measured using NanoDrop spectrophotometer, and 15 µl of protein samples were stored at -20°C for SDS-PAGE analysis.

Day 4: Leaky expression extracts isolation using batch purification method with His-Pur (Thermo) Ni-NTA beads

The confluence of the cultures was checked, and the centrifuge was set to 4°C. All of the culture medium was transferred into 50 ml falcon tubes and centrifuged at 4°C, 3,750 rpm for 20 min. 1 ml of the supernatant was taken and stored in 1.5 ml Eppendorf tubes for SDS-PAGE experiments. The 500 µl or 1 ml of His-Pur Ni-NTA beads solution was added into 50 ml falcon tubes. The final volume was completed to 50 ml with 1X PBS, and the tubes were centrifuged at 4°C, 3,750 rpm for 2 min to wash the His-Pur Ni-NTA beads. The supernatant was then discarded. The remaining 49 ml of culture supernatant was added onto the beads washed with PBS. The tubes were mixed on an end-over-end rotator for 1-2 h in a cold room (4°C). The tubes were centrifuged for 2 min at 4°C, 3,750 rpm. 1 ml of the residual

supernatant was stored at 4°C for SDS-PAGE analysis. The beads were washed with 50 ml of 1X PBS. The tubes were centrifuged for 2 min at 4°C, 3,750 rpm. The supernatant was poured off, and the pellets or beads were washed with 20-50 ml of Buffer A. The tubes were gently mixed and then centrifuged for 2 min at 4°C, 3,750 rpm. 1 ml of the supernatant was stored in 1.5 ml Eppendorf tubes for SDS-PAGE analysis. The residual supernatant was discarded. This step was repeated twice. 1-2 ml of Buffer B was added to the tubes, and the solution was mixed. The tubes were centrifuged for 2 min at 4°C, 3,750 rpm. 100 µl of the supernatant was stored in 1.5 ml Eppendorf tubes for SDS-PAGE analysis. The rest of the supernatant was saved in conical tubes for downstream analysis, as it contained the protein. This step was repeated twice, and another 100 µl of supernatant was taken from the second wash and added to the 1.5 ml Eppendorf tube for SDS-PAGE analysis. The remaining supernatant was added to the 15 ml conical tube. The beads were kept in 1X PBS at 4°C until regeneration. The amicon filter tube was washed by adding 1X PBS up to the upper line of the filter and centrifuging for 5 min at 4°C, 3,750 rpm. The solution was then discarded. The elution sample was filtered into the amicon tube using a syringe filter (0.22 µm hydrophilic PVDF Syringe Filter). The final volume was filled up to the upper level of the tube with 1X PBS, and the tube was centrifuged at 4°C, 3,750 rpm for 12 min. The filter part was removed, and the solution at the bottom was discarded as the top part contained the protein. This step was repeated three times to remove imidazole by exchanging the buffer solution with 1X PBS. The protein concentration was measured using a NanoDrop spectrophotometer. The eluted protein could be directly analyzed by SDS-PAGE.

Day 4: Leaky expressions extract isolation using centrifuge purification method using GoldBio Ni-NTA beads

In the morning, the confluence of the cultures was checked, and the centrifuge was set to 4°C. After the 48 h of incubation within the auto-induction media, the culture medium was transferred into 250 ml ultra-centrifuge tubes. The tubes were centrifuged at 4°C, 8,000 rpm for 15-20 min. The supernatant was then transferred back into the same flask used previously, while keeping the cells at the bottom of the tubes for disposal. If necessary, a 1 ml sample was taken into a 1.5 ml Eppendorf tube for further SDS-PAGE analysis and stored in the fridge at 4°C. The protein slurry was divided into 50 ml Falcon tubes and centrifuged at 3,900-4,000 rpm (maximum speed) for 20 min. This step was repeated three times, each time using fresh Falcon tubes. Meanwhile, another centrifuge was set to 4°C to prepare the Ni-NTA beads. 1

ml of GoldBio Ni-NTA beads was added into 50 ml Falcon tubes for each 100 ml of protein slurry. The volume was topped up to 50 ml by adding 1X PBS into the tube. The tubes were then centrifuged at 14°C, 1,400 rpm for 7 min to wash the beads before use. The supernatant was discarded, and the beads were retained. The washing procedure described previously was repeated one more time using 1X PBS. The supernatant media slurry was mixed with the Ni-NTA beads in the protein slurry flask. The flask was placed on a magnetic stirrer in the cold room or within the fridge at 4°C for 2 h, allowing the proteins of interest to bind to the Ni-NTA beads with high affinity. If necessary, a 1 ml sample was taken from the protein-beads mix after incubation for further analysis using SDS-PAGE and stored in the fridge at 4°C. Alternatively, the bead incubation time could be extended to overnight incubation in the fridge at 4°C. After the Ni-NTA incubation, a 10 ml ÄKTA Column was placed into a vacuum filtration tap, and the twist-off bottom closure was removed. Then, it was placed in a 50 ml Falcon flask. The entire bead slurry was added to the 10 ml ÄKTA Column, and centrifugation was initiated at 3,500 rpm at 4°C for 1 min for each 10 ml loaded column. The flask containing the sample was shaken every time to resuspend the beads while transferring the contents into the column. After the whole mixture passed through the column, the washing stage was initiated. 100 or 200 ml of 1X PBS was added to the column.

Cell lysis by sonication to prepare denaturated *E.coli* proteins:

The cell lysis process involved taking an overnight grown 200 ml BL21 *E.coli* culture from the incubator and dividing the volume into 50 ml Falcon tubes. The tubes were then centrifuged at 3,220xg for 20 min at 4°C. After centrifugation, the supernatant was gently discarded by removing the media from the cells. The tubes were placed back on ice to keep the bacterial pellet cold. Next, 10 ml of Buffer A (**Table 39**) was added to each cell pellet, and the pellets were dissolved in Buffer A.

Table 39. Preparation of Buffer A and Buffer A with DTT for sonication.

Buffer A for sonication, 1 L	
1M Tris acetate (pH 8.2), 10 ml	The preparation process involved making 1M Tris acetate, 1M $MgCl_2.6H_2O$, and 1M KCl solutions in advance. The stock solutions were mixed together in a
1M $MgCl_2.6H_2O$, 14 ml [11]	

[11] The original Buffer A recipe uses 1M stock solutions of tris acetate, magnesium glutamate and potassium glutamate. In this protocol alternative solutions were prepared based on the availability of the chemicals in the lab, where magnesium glutamate -$Mg(C_5H_8NO_4)_2$- is replace by $MgCl_2.6H_2O$ and potassium glutamate -$KC_5H_8NO_4$- is replaced by KCl.

1M KCl, 60 ml	glass bottle. To reach a total volume of 1L, dH_2O was added. The resulting mixture was then stored at 4°C.
dH_2O, 1000 ml	

Buffer A with DTT for sonication, 1 L

1M Tris acetate (pH 8.2), 10 ml	The preparation process involved making 1M Tris acetate, 1M $MgCl_2.6H_2O$, 1M KCl, and 1M DTT solutions in advance. The stock solutions were mixed together in a glass bottle. To reach a total volume of 1L, dH_2O was added. The resulting mixture was thoroughly mixed and then stored at 4°C.
1M $MgCl_2.6H_2O$, 14 ml	
1M KCl, 60 ml	
1 M DTT, 2 µl	
dH_2O, 1000 ml	

1 M KCl for Buffer A for sonication, 100 ml

KCl, 7.45 g	The powder was weighed, and then it was dissolved in 50 ml of dH_2O. The solution was mixed well until the powder was completely dissolved. Next, the total volume was topped up to 100 ml using additional dH_2O. The solution was stored at RT.
dH_2O, 100 ml	

1M $MgCl_2.6H_2O$ Buffer A for sonication, 100 ml

$MgCl_2.6H_2O$, 20.33 g	The powder was weighed, and then it was dissolved in 50 ml of dH_2O. The solution was mixed well until the powder was completely dissolved. Next, the total volume was topped up to 100 ml using additional dH_2O. The solution was stored at 4°C.
dH_2O, 100 ml	

1 M Tris Acetate (TAE) Buffer A for sonication, 1 L

Tris base, 12.11 g	The powder was weighed. Then, 50 ml of dH_2O was added to the powder to dissolve. The pH of the solution was adjusted to 8.2 using acetic acid. Afterward, the total volume was topped up to 1L by adding additional dH_2O. The solution was thoroughly mixed to ensure homogeneity. The solution was stored at RT.
dH_2O, 50 ml	

The dissolved pellets were then pooled into a single Falcon tube. The pooled cell suspension was centrifuged at 3,220xg for 10 min at 4°C, and the supernatant was gently

discarded. This washing step was repeated by adding 10 ml of Buffer A to the cell pellet, dissolving the pellet in Buffer A, and centrifuging at 3,220xg for 10 min at 4°C. The supernatant was discarded, and the washing step was repeated once more. After the final centrifugation, the supernatant was discarded by gently removing the media from the cells. The weight of the tube containing the pellet was measured. Finally, the lysed cells were flash frozen using liquid nitrogen and stored at -80°C.

Alternating Buffer A solutions were prepared in advance, including Buffer A (25 mM imidazole), Buffer A with 5 mM MgATP, and Buffer A with 5 mM MgATP + denatured *E.coli* proteins (**Table 40**). The purification process using the GoldBio Ni-NTA beads continued with these Buffer A solutions and subsequent steps.

Table 40. Preparation of Buffer A with 5 mM MgATP and denaturated *E.coli* lysate.

Buffer A with 5 mM MgATP (240 ml)

Buffer A, 240 ml	The reagents were mixed in a glass bottle. There was no need for filtering for sterilization since imidazole in Buffer A was already filtered, and 20X PBS in Buffer A had already been sterilized by autoclaving.
200 mM MgATP stock, 6 ml	

200 mM MgATP, 200 ml

Na_2ATP, 2.21 g	The powder was weighed and dissolved in 10 ml of ultrapure dH_2O in a glass bottle. Then, 4 ml of 1M $MgCl_2$ was added to the solution. The pH was adjusted to 7.0 by adding 1.7 M Tris. The total volume was adjusted to 20 ml. The solution was stored at -20°C in 1 ml aliquots in 1.5 ml Eppendorf tubes.
1M $MgCl_2$, 4 ml	
dH_2O, 20 ml	

Buffer A + 5 mM MgATP + denaturated *E.coli* lysate (250 ml)

Buffer A, 240 ml	The reagents were mixed in a glass bottle. The mixture was then heated at 65°C for 10 min for denaturation. After heating, the mixture was centrifuged for 10 min at 12,000xg.
0.6 mg/ml (BL21(DE3) pLysS) *E.coli* cell lysate, 4 ml	
6 ml 200 mM MgATP stock	MgATP stock was added the mixture before use.

For BL21 *E. coli* cell lysis, the following steps were taken: The cells were taken from the -80°C freezer and thawed on ice, ensuring proper handling to maintain their integrity. The cells were taken from the -80°C freezer and thawed on ice. Buffer A with DTT was prepared for sonication, with a volume of 1 L and a composition of 2 mM DTT, 10 mM TAE, 14 mM $MgCl_2.6H_2O$, and 60 mM KCl. This buffer should ideally be made fresh each time, but if short-term storage is required, it can be stored at -20°C.

Keeping the cell pellet on ice, Buffer A with DTT was added to the cells at a ratio of 1 ml buffer per 1 g wet mass as determined earlier. The cells were then vortexed thoroughly to dissolve the pellet and ensure proper mixing with the buffer.

Next, 6x1 ml of the cell suspensions were transferred into 1.5 ml Eppendorf tubes. It was crucial to maintain the 1 ml volume for each tube. The sonicator was set up at 4°C, and all 6 tubes were placed inside the cooled sonicator. The cells were sonicated with 10-second pulses followed by 30-second pauses for a total of 10 cycles. The suggested sonication parameters were at 400 J using 50% amplitude and pulses of 10 sec on and 10 sec off. At the 5th cycle, the sonicator was stopped, and vortexing was applied to all the samples. The tubes were then returned to the sonicator, and sonication process was continued for 5 more cycles.

After the sonication was completed, the samples were placed on ice. The table top centrifuge was pre-cooled to 4°C, and the tubes were centrifuged at 12,000xg for 10 min at 4°C. The supernatant from each tube was carefully transferred into fresh Eppendorf tubes, taking care not to transfer any bacterial debris. To degrade any remaining DNA and RNA products, the lysate was placed in an incubator at 37°C, with shaking at 200 rpm for 1 h and 30 min. This step, known as the run-off reaction, helped to improve the purity of the lysate. After the incubation, the samples were centrifuged again at 12,000xg for 10 min at 4°C. The supernatants from the tubes were collected, and all the samples were pooled into the same tube. To estimate the total protein concentration collected, the concentration was measured using a NanoDrop spectrophotometer.

Buffer A with DTT (**Table 39**) for sonication was prepared in advance. Once the PBS had passed through the filtration, 100 or 200 ml of Buffer A were added to the column, dividing it into maximum 15 ml portions. The column was then removed from the tap and placed into a 50 ml Falcon tube containing 1X PBS at the bottom. Next, 1 ml of Buffer B was added to the column for each 1 ml of beads. The samples were slowly passed through the column either by keeping it in a 4°C fridge or by centrifuging the column at 4°C, 3,750 rpm

for 2-3 min. Meanwhile, the centrifuge was set to 4°C. The samples were eluted into 1X PBS (6 ml sample + 6 ml 1X PBS; 1:1) and collected in 50 ml Falcon tubes within the fridge. The eluent at the bottom of the 50 ml Falcon tube was transferred to a 15 ml Amicon tube. The final volume was filled up to the upper level of the tube with 1X PBS, and the tube was centrifuged at 4°C, 3,750 rpm for approximately 12 min. The filter part was removed, and the solution at the bottom (which does not contain the protein) was discarded. This process was repeated three times to remove imidazole by exchanging the buffer solution with 1X PBS. The final volume at the top of the filter was reduced to 250-300 µl, and then the volume was transferred into Eppendorf tubes. The protein concentration was measured using a NanoDrop spectrophotometer.

Depending on the downstream workflow, the volume was divided into fractions, and the samples were flash frozen and stored at -80°C. Additionally, a 15 µl protein sample was mixed with 15 µl of Leammli and stored at -20°C for SDS-PAGE analysis. The collected protein concentration was measured using NanoDrop spectrophotometer.

3.6.11. Redesign for new sdAbs

In the host lab, a redesign of the sdAbs from the home lab was carried out to enhance compatibility of the sdAbs in animal studies meanwhile minimizing the potential immunogenic effects stemmed for extra tags that home design have. With this new design, additional tags at the C-terminus, including FLAG tag, His tag, and the Factor Xa recognition site, would be selectively eliminated. Additionally, a new mutation was introduced by converting the 1st "E" residue into "D" to which is mostly in nanobodies at this position. By removing these extra tags, the purification process by IMAC would not be hampered as the cloning vector also changes with insert design. The plasmid was switched to pHen6, which inherently has His_6 tag for both purification and detection purposes for downstream applications. The resulting plasmid would yield into the expression of newly designed sdAbs with more compact size and compatible with downstream animal studies.

3.6.12. Cloning

The cloning process initiated with the isolation of send plasmids by the home lab by retrieving them from filter paper that they being send on. Cloning involves the replication of send home lab-design sdAbs sequence and its subsequent insertion into new expression vector for its amplification.

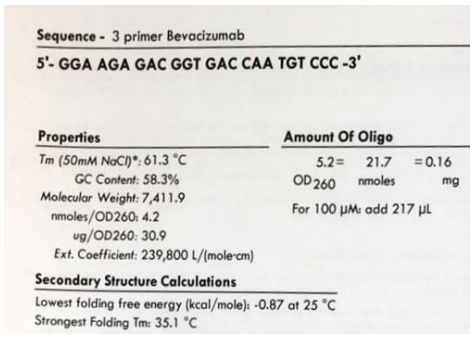
3.6.12.1. Plasmid recovery

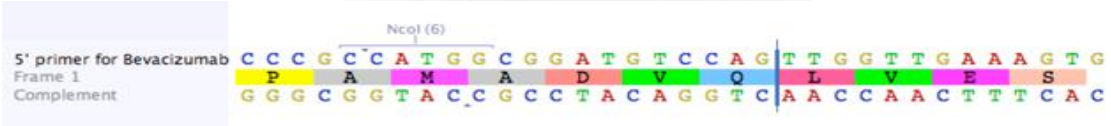
The home lab sent the original sdAb variant plasmids by placing a drop of them on a filter paper. In the elution of plasmids from filter paper, the associated plasmid circles were carefully cut without any physical contact to the inside of the marked are which constitutes the associated sdAb plasmid sample. Subsequently, the cut papers were placed in individual 1.5 ml Eppendorf tubes. To ensure proper retrieval, the paper pieces were gently pushed to the bottom of the tube. 100 µl of dH_2O was added to each tube. The tubes were gently tapped to facilitate elution. The tubes were then left for incubation at RT for 30 min, with occasional tapping. The plasmid concentration was measured using a NanoDrop spectrophotometer.

3.6.12.2. Preparation of the primers

The newly designed upstream and downstream primers were ordered and received in lyophilized formulation (**Figure 20**) to facilitate the selective targeting of home lab sdAb sequence from pET17b plasmid, meanwhile amplifying them in PCR and introducing the intended mutation on the first residue position. To prepare them for use, the lyophilized primers were centrifugation for 2 min and subsequently, an appropriate amount of dH_2O as recommended by manufacturer was added to each tube to dissolve the primers in. Thorough mixing was performed, and the tubes were vortexed to homogeneously dissolve the primers.







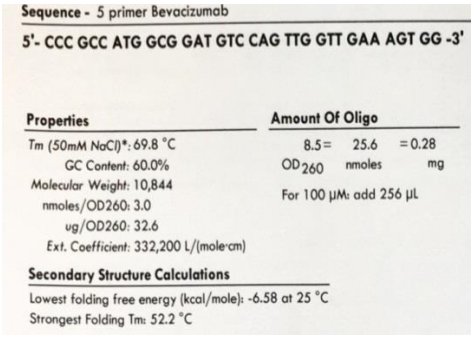


Figure 20. The sequence and the properties of the Bevacizumab V_H sequence targeting forward and reverse primers designed in the host lab. The primers are compatible with all home lab-designed Bevacizumab V_H variants including all the sdAb variants and unaltered WT version as well.

The primers were carefully designed to incorporate recognition sites for the restriction enzymes BstEII and NcoI to create compatible sticky ends that match with the corresponding sites on the expression vector pHen6. Prior to PCR, the expression vector was digested using the same enzymes, BstEII and NcoI, to generate complementary sticky ends. By flanking the sdAbs sequences, the primers specifically exclude the tags and enzyme recognition site.

3.6.12.3. Amplification of Bevacizumab V_H sequences by PCR

The following PCR process involved several steps. First, the primers were prepared to dissolve as described above. Next, for each sample, a master mix of PCR reagents was prepared, which also included a negative control reaction without any DNA template.

The PCR reaction reagents (**Table 41**) and run conditions (**Table 42**) were as follows:

Table 41. Reagents for amplification of the Bevacizumab V_H fragments by PCR.

Reaction Reagents	Volume (µl)	Final concentration
5X Colourless GoTaq Reaction Buffer	20	1X
Sterile dH_2O (PCR grade)	64.5	-
dNTPs (10 mM)	2	0.2 mM (each)
Forward primer (10 µM)	4	0.4 µM
Reverse primer (10 µM)	4	0.4 µM
GoTaq Polymerase (5 U/µl)	0.5	0.025 U/µl
dsDNA template (1-5 ng)	5	0.001-0.05 ng
Total volume	**100**	**-**

Table 42. Reaction steps for amplification of the Bevacizumab V_H fragments by PCR.

Temperature (°C)	Time	Cycles
95	2 min	1
94	1 min	
50	1 min	34
72	1 min	
72	10 min	1
4	Hold	∞

3.5.12.4. Digestion of vector pHen6

The stock pHen6 plasmid was diluted using dH_2O prior to use. The vector pHen6 was digested with NcoI and BstEII restriction enzymes. A master mix was prepared without adding a DNA template, following these steps at **Table 43**, where as the last reagent, 5 µl of pHen6 vector was added to the mix.

Table 43. Reagents for pHen6 expression vector digestion using NcoI and BstEII enzymes.

Reaction Reagents	Volume (µl)	Final concentration
pHen6 vector (600 ng/µl)	5	10 ng/µl
10X Tango buffer	20	2X
NCoI	1	-
BstEII	1	-
dH_2O	73	-
Total volume	**100**	-

The digestion was initiated by incubating the tubes in a water bath at 37°C for 3 h. After the 3 h of digestion, the samples were removed from the water bath. To prevent self-ligation of the digested vector, 1 µl of FastAP (Alkaline Phosphatase, 1 U/µl) was added to each tube. The tubes were then placed in a water bath at 37°C for 10 min. To inactivate the alkaline phosphatase enzyme, the tubes were transferred to a heat block set at 75°C for 5 min.

3.5.12.5. Ligation of vector pHen6 with modified Bevacizumab V_H fragments

On Day 1, the water bath was turned on and set to a temperature of 16°C prior to the ligation reaction. A total of 11 reactions were prepared for ligation, comprising 10 sdAbs and 1 negative control (without any insert). A master mix was prepared without the addition of DNA template (**Table 44**). The master mix included the already digested vector pHen6 (NcoI,

BstEII, and phosphatase) (N/B/F)) as described previously, which was compatible with the insert.

Table 44. Reagents for ligation of vector pHen6 with Bevacizumab V_H fragments.

Reaction Reagents	Volume (µl)	Final concentration
Vector pHen6 (N/B/F) (73 ng/µl)	0.6	-
Insert (25 ng)	2	-
10X T4 DNA Ligase Buffer	2	1X
T4 DNA Ligase	1	-
dH_2O	14.4	-
Total volume	**20**	-

The master mix was divided into separate Eppendorf tubes, with each tube corresponding to a specific sample. The inserts were added to their respective Eppendorf tubes. The Eppendorf tubes were then placed in the floating block in the water bath. The samples were left for overnight ligation at a temperature of 16°C, allowing the ligase enzyme to work on the joining of the inserts with the digested pHen6 vector.

On Day 2, the samples were retrieved from the 16°C water bath in the afternoon. Subsequently, the samples were kept at the bench to equilibrate them to RT before proceeding with the transformation.

3.6.12.6. Heat-shock transformation after ligation with WK6 cells

The ligation mix was spun down to pool everything at the bottom because evaporation may occur during overnight incubation. The competent WK6 cells were taken from -80°C and put on ice to thaw. For each plasmid 50 µl of competent cells were used. For each plasmid, 5 µl of ligation mix was added. The tubes were incubated on ice for 30 min. Meanwhile, the water bath was warmed up to 42°C. The tubes were heat-pulsed at 42°C for 45 sec. The samples were put on ice for a couple of min. 900 µl of LB broth at RT was added to the tubes and mixed well with the transformation mix.

The tubes were put in a 37°C incubator for 15 min. 2xAMP resistance agar plates were used for each plasmid. The sterile cabin was used to inoculate the cells into agar plates. 200 µl of transformation mix was added to one of the agar plates of each sample. For the remaining volume of the transformation mix, the Eppendorf tubes were centrifuged at 5,000 rpm for 2 min. The LB broth on the top was removed until about 200 µl was left in the tube.

The centrifuged cells were dissolved in the remaining LB. The 200 µl was added to the second plate. Glass beads were added to each plate, and they were gently shaken until the transformation mix dried up. The plates were incubated either in a 37°C incubator overnight or at RT on the bench over the weekend. All the plates were covered with aluminium foil to prevent possible mutation introduction caused by UV light present in the environment.

3.6.12.7. Plasmid isolation - Genelute plasmid miniprep kit

By extracting plasmids from bacterial cultures, purified DNA sequences would be prepared for sequencing to determine if the desired sequences have been successfully incorporated into pHen6 vector, simultaneously confirming the presence of the alteration of amino acid residue at the first position and the effectiveness of the cloning process.

Genelute™ plasmid miniprep kit was utilized. The kit procedure was carried out based on instructions of the manufacturer. Overnight grown cells in LB broth were retrieved from the incubator. The falcon tubes containing the cells were centrifuged at 14°C and 2,500 rpm for 15 min, resulting in the formation of a pellet after removal of the supernatant. A re-suspension solution containing RNase was added and mixed by pipetting up and down or vortexing. The resulting suspension was transferred into Eppendorf tubes, followed by the addition of 200 µl of lysis solution to each tube. The tubes were inverted 4-6 times to gently mix them, and they were left aside for 5 min allowing the clearing of the reaction. Next, 350 µl of neutralizing solution was added and mixed gently by inverting the tubes. The debris was then centrifuged at 13,000xg for 10 min. A binding column was placed into a collection tube provided by the kit, and 500 µl of column preparation solution was added to the binding column, followed by centrifugation at 13,000xg for 1 min. The flow-through was discarded, and the cleaned lysate was transferred into the prepared binding column, which was centrifuged again at 13,000xg for 1 min. The flow-through was discarded. Subsequently, 500 µl of optional washing solution was added to the binding column, which was then centrifuged at 13,000xg for 1 min, and the flow-through was discarded. This washing step was repeated using 730 µl of washing solution, followed by centrifugation at 13,000xg for 1 min and discarding the flow-through. To remove any remaining ethanol in the column, another centrifugation step was performed at 13,000xg for 2 min. The column was then transferred to a fresh collection tube, and 75 µl of elution solution was added to the centre of the collection tube. After 1-min incubation at RT, the mixture was centrifuged at 13,000xg for 1 min.

Finally, the column was removed from the collection tube, and the plasmid concentration was measured, using the elution buffer as the blank.

3.6.12.8. Confirmation of ligation into expression vector pHen6 by sequencing

The Eurofins over-night sequencing service was used. The tubes were pre-labelled and already provided by Eurofins. 2 µl of MP-57 primers (10 µM) were added to each tube that targets the first codon of the plasmid of interest. Specifically, 600 ng of each sdAb, 15 µl of 40 ng/µl sdAb sample was diluted if required using appropriate amount of dH_2O, ensuring the appropriate concentration for sequencing.

3.6.13. Protein production and purification – WK6

After the sequence confirmation, the sdAb plasmids were ready for expression. The recombinant protein production and purification using *E. coli* strain WK6 involves the series of stages for protein expression. After harvesting the protein extracts and the samples were purified using chromatographic techniques (**Figure 21**).

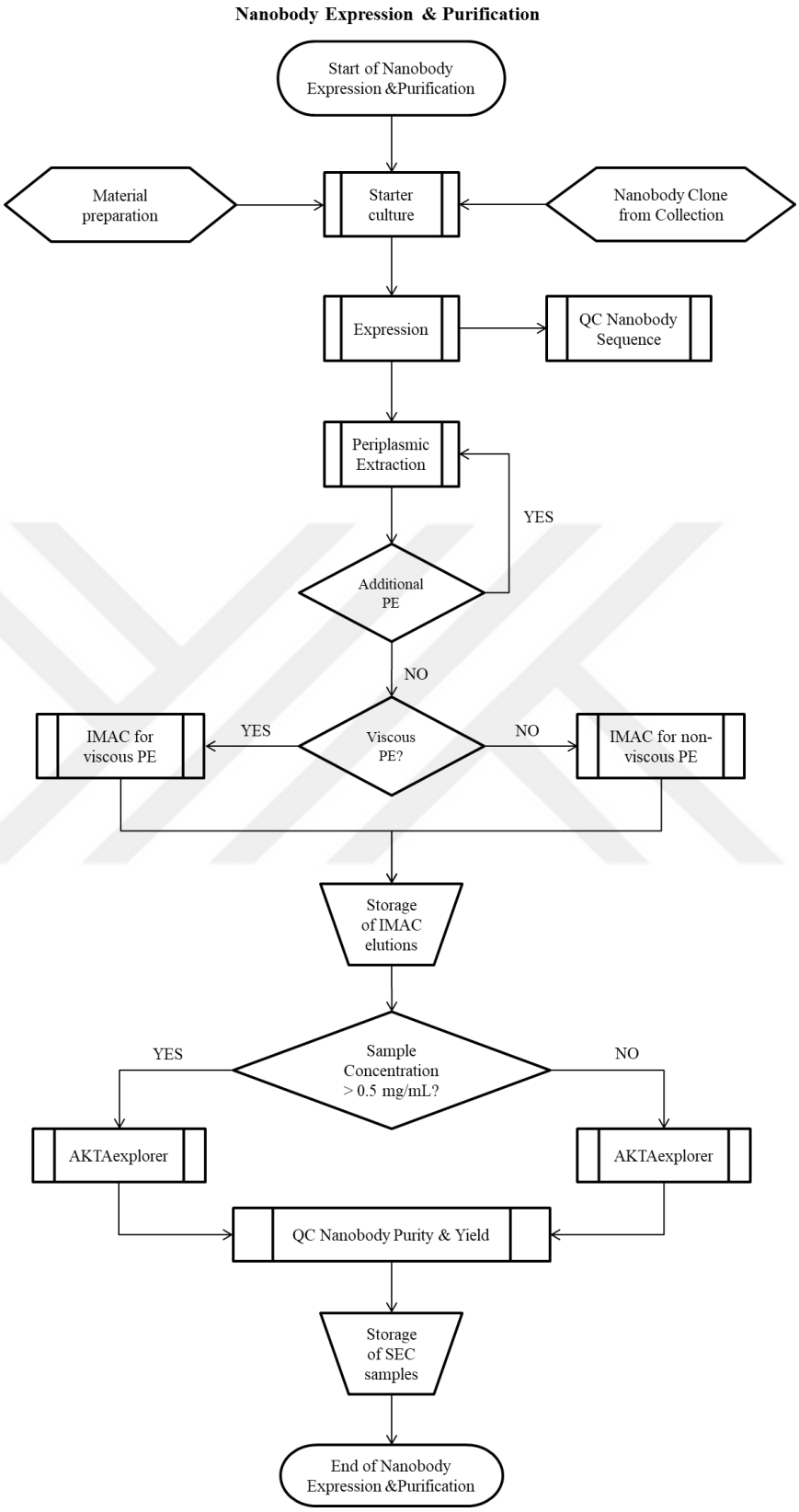


Figure 21. The workflow of the host lab nanobody/sdAb expression and purification progression.

Day1: Starter culture preparation

The following steps were performed under sterile conditions in a laminar flow: First, a 50 ml Falcon tube was selected, and 15 ml of LB was dispensed into it. Next, 15 µl of 1000x AMP stock (final concentration of 100 µg ml-1 AMP) was added to the tube. Then, a single colony from an LB agar plate (containing AMP + glucose) (**Table 45**) was inoculated into the 15 ml LB + AMP Falcon tube using a sterile tip. The culture was grown overnight at 37°C and 200 rpm in a New Brunswick Incubator Shaker.

Table 45. LB Agar with 2% glucose and 100 µg/ml AMP for WK6 cells.

LB Agar with 2% glucose and 100 µg/ml AMP for WK6 cells

LB high salt, 25 g **Micro agar, 15 g** **milliqH_2O, 900 ml** **1000X AMP stock solution (100 mg/ml), 1 ml** **20% sterile glucose, 100 ml**	The dry powders were measured. The milliqH_2O was added first, and the powders were added to prevent floating of the dusts. They were mixed in 900 ml milliqH_2O in a plastic flask on a magnetic stirrer. The liquid was transferred into a glass bottle and autoclaved. After autoclaving, the bottle was allowed to cool down to 56°C in an incubator. Under the sterile hood, 1 ml of 1000X AMP stock solution (50 ml at 100 mg/ml) and 100 ml of 20% sterile glucose were added to the bottle. The mixture was well mixed with the agar by rolling the bottle on the bench. The lids of the petri dishes were removed, and approximately 20 ml of agar was poured into each petri dish under the hood. The plates were cooled down for 20 min to let them solidify. Subsequently, the plate lids were closed, and the related details were written down at the bottom of each plate. The plates were stored in the cold room until their use.

1000X AMP stock (50 ml at 100 mg/ml) for LB Agar

AMP, 5 g **dH_2O, 50 ml**	The powder was weighed. It was dissolved in dH_2O in a glass bottle. A 0.22 µm filter was used for sterilization. The solution was aliquoted into Eppendorf tubes in 1 ml each. The tubes were stored at -20°C.

20% (w/v) D-Glucose (100 ml) for LB Agar

D-glucose, 20 g **dH_2O, 100 ml**	The powder was weighed. It was dissolved in dH_2O in a glass bottle. The bottle was autoclaved and then stored at RT.

Day 1: Production media preparation

For production media prepapration, a master mixture of Terriffic Broth (TB) was prepared in a big flask (**Table 46**), and then it was divided into baffled flasks. First, milliqH_2O was added, followed by the addition of the solid parts to prevent dusting. The glycerol was added last. Each flask would contain a volume of 0.3333 L and was placed in a 1L baffled flask. The flasks were closed with tops and covered with aluminum foil and then auto-claved for sterilization. The autoclaved sterilized flasks were kept at RT until further use.

Table 46. Preparation of Terrific Broth (TB) production media for WK6 cells.

Terrific Broth for WK6 cells

Tryptone **Yeast extract** **KH_2PO_4** **$K_2HPO_4.3H_2O$** **Glycerol 100%** **dH_2O**	The dry powders were measured. The milliqH_2O was added first, and the powders were added to prevent floating of the dusts. They were mixed in milliqH_2O in a plastic flask on a magnetic stirrer. The glycerol was added last. The liquid was transferred into a glass bottle and divided into baffled flasks. The flasks were closed with tops and covered with aluminium foil. Subsequently, the flasks were autoclaved for sterilization. After autoclaving, the sterilized flasks were kept at RT until further use.

Reagent	**Volume (L)**					
	1	**2**	**3**	**3.5**	**4**	**5**
Tryptone	12 g	24 g	36 g	42 g	48 g	60 g
Yeast extract	24 g	48 g	72 g	84 g	96 g	120 g
KH_2PO_4	2.3 g	4.6 g	6.9 g	8.05 g	9.2 g	11.5 g
$K_2HPO_4.3H_2O$	16.4 g	32.8 g	49.2 g	57.4 g	65.6 g	82 g
Glycerol 100%	4 ml	8 ml	12 ml	14 ml	16 ml	20 ml
dH_2O	0.95 L	1.9 L	2.85 L	3.32 L	3.8 L	4.75 L

Day 2: Expression of nanobodies in *E. coli* WK6 periplasm

Under sterile conditions in a laminar flow, each 330 ml TB baffled shaker flask was added with the following: 330 µl of 1000x AMP stock (100 µg ml-1 AMP final concentration), 1.5 ml of 20% glucose (0.1% glucose final concentration), and 330 µl of 2 M $MgCl_2$ (2 mM $MgCl_2$ final concentration). Subsequently, 1 ml of the starter culture was inoculated into the 330 ml TB baffled shaker flask. The flask was grown at 37°C and 200 rpm

in a New Brunswick Incubator Shaker until turbidity (OD_{600} nm) reached a range of 0.6 to 0.9, which typically takes approximately 2-4 h. An excess of the starter culture was spared for a miniprep isolation, which were then sent out for sequencing for quality control (QC) measures. Finally, each 330 ml TB baffled shaker flask was added 330 µl of 1000x IPTG stock, resulting in a final concentration of 1 mM IPTG (**Table 47**), and was further incubated overnight at 28°C and 200 rpm in an incubator shaker.

Table 47. Preparation of IPTG.

1000X IPTG stock solution (50 ml, 1M)

IPTG, 11.92 g	The powder was weighed and dissolved in dH_2O in a glass bottle. For sterilization, a 0.22 µm Gyrodisc CA-PC 30 mm filter was used. The solution was aliquoted into 10 to 15 ml portions and stored at -20°C.
milliqH_2O, 50 ml	

Day 3-4: Periplasmic extract preparation

The following steps were taken to harvest the *E. coli* cells. A 330 ml culture from the shaker flask was added to a 500 ml centrifuge bottle. The centrifuge bottle was centrifuged for 8 min at 8,000 rpm and 14°C using a Beckman Coulter Avanti J-E centrifuge with rotor JA-10. The supernatant was decanted. If the sdAb clone had been expressed in multiple 330 ml shaker flasks, the next 330 ml was reloaded into the corresponding centrifuge bottle. The centrifugation step was repeated until all the supernatants were completed. The cellular pellet was resuspended in Tris-EDTA-Sucrose (TES) buffer (**Table 48**) using the following guidelines:

- If the *E. coli* culture had OD_{600}: 20-30, 4 ml of TES buffer per pellet from a 330 ml culture was used.
- If the *E. coli* culture had OD_{600} < 20, 3 ml of TES buffer per pellet from a 330 ml culture was used.

The suspension was pipetted up and down rigorously to ensure it was free of cellular clumps. The mixture was incubated in the cold room while shaking at 200 rpm on a tabletop shaker for 1 to 6 h.

Table 48. Preparation of Tris-EDTA-Sucrose (TES) buffer.

TES Buffer (Tris 0.2 M, EDTA 0.5 mM, sucrose 0.5 M)

1M Tris (pH 8.0), 200 ml	The sucrose was measured. Tris and EDTA were added to the

0.5M EDTA disodium dehydrate, 1 ml **Sucrose, 171.2 g** **milliqH_2O, 1000 ml**	sucrose. They were mixed well and brought to a total volume of 1000 ml with milliqH_2O in a plastic flask on a magnetic stirrer. The bottle containing the solution was then stored in the cold room.

Osmotic shock was performed as follows:

If the two combined bottles of *E. coli* culture had OD_{600}: 20 to 30, 8 ml of TES/4 buffer (**Table 49**) per pellet from a 330 ml culture was used. If the *E. coli* culture had OD_{600} < 20, 6 ml of TES/4 buffer per pellet from a 330 ml culture was used. The suspension was pipetted up and down rigorously. The mixture was further incubated in the cold room while shaking at 200 rpm for 2 h (or overnight) on a tabletop shaker.

Table 49. Preparation of TES/4 buffer.

TES/4 Buffer

TES Buffer, 250 ml **milliqH_2O, 750 ml**	The solution was brought to a total volume of 1000 ml with milliqH_2O in a plastic flask on a magnetic stirrer. Subsequently, the bottle containing the solution was stored in the cold room.

100 µl of 2M $MgCl_2$ per pellet from a 330 ml culture (10-15 mM $MgCl_2$ final concentration) was added. The mixture was centrifuged for 30 min at 8,000 rpm using a Beckman Coulter Avanti J-E centrifuge with rotor JA-10. The supernatant, which was the periplasmic extract, was pipetted into 50 ml Falcon tubes without disturbing the pellet. If a large amount of protein was required for experiments, a second osmotic shock could be performed by repeating the previously explained osmotic shock procedure and collecting the second extraction in Falcon tubes (the yield of the second extraction varied between 30% to 100% of the first extraction).

Day 3-4: IMAC for non-viscous periplasmic extract

To begin, the first clean white frit was inserted into an empty PD-10 column and pushed to the bottom. The His-Select solution was then rigorously swirled to solubilize the beads in the suspension, and the resulting solution was loaded onto the column. 1 ml of His-Select solution was sufficient for loading the periplasmic extract derived from one liter of the original *E. coli* culture, with a capturing capacity of approximately 15 mg Nb/ml of His-Select solution. A second frit was placed on top of the His-Select bed. The His-Select column was washed by pipetting 10 ml of PBS per ml of His-Select solution onto the column, allowing the

gravitational force to drain the PBS buffer. Subsequently, the periplasmic extract was loaded onto the column, again allowing the gravity to drain the extract. The His-Select column was washed further by pipetting 20 ml of PBS per ml of His-Select solution into the column and allowing the PBS buffer to drain by gravitational force. Fraction elution was then performed using 5 ml of PBS + 0.5 M imidazole. This elution process was repeated twice, and the OD_{280} nm of the imidazole elutions was measured using a NanoDrop spectrophotometer. If the third elution fraction had an OD_{280} nm greater than 0.2, subsequent elutions were performed until the OD_{280} nm was less than 0.2. It was crucial to store the imidazole elutions at 4°C, as storing them in imidazole at -20°C or for more than one week at 4°C could lead to irreversible protein precipitation.

Day 3-4: IMAC for viscous periplasmic extract

For the viscous periplasmic extract, the His-Select solution was vigorously mixed to solubilize the slurry. The His-Select solution was then directly added to the Falcon tube containing the periplasmic extract and incubated for 30 min to 1 h with shaking. 1 ml of His-Select solution was adequate for loading the periplasmic extract derived from one liter of the original *E. coli* culture, with a capturing capacity of approximately 15 mg Nb/ml of His-Select solution. The mixture was then centrifuged for 7 min at 1,400 rpm to settle the resin, and the supernatant was removed, leaving approximately 5 ml of slurry. Next, 45 ml of PBS was added to the tube, and shaking was performed for 30 min. The tube was centrifuged again at 1,400 rpm, and the supernatant was removed, leaving approximately 5 ml of slurry on top of the resin. An empty PD-10 column was prepared by inserting the first frit and pushing it to the bottom. The periplasmic extract/His-Select solution was rigorously swirled, and the resulting mixture was loaded onto the column. The His-Select column was washed by pipetting 20 ml of PBS per ml of His-Select solution into the column, and the PBS buffer was allowed to drain via gravitational force. Elution fractions were obtained using 1 column volume of PBS + 0.5 M imidazole. The fraction elution process was repeated 2-3 times, and the OD_{280} nm of the imidazole elutions was measured using a NanoDrop spectrophotometer. If the third elution fraction had an OD_{280} nm greater than 0.2, subsequent elutions could be performed until the OD_{280} nm was less than 0.2. Similar to the non-viscous extract, the imidazole elutions from the viscous extract had to be stored at 4°C, and it was important to avoid storing them in imidazole at -20°C or for more than one week at 4°C to prevent irreversible protein precipitation.

Day 4-5: Preparation for SEC (Size Exclusion Chromatography) runs

The amount of protein in the elution fractions was calculated using the extinction coefficient (**Table 50**) of each clone, which was determined using the ProtParam tool from ExPASy (Duvaud et al. 2021). The pI (isoelectric point) and MW (molecular weight) parameters calculated by the ProtParam tool were necessary for performing SEC on the AKTAxpress. The choice of machine for SEC depended on the expression and purification yield of the elution fractions: If the elution fractions contained less than 0.5 mg of sdAb in 5 ml, SEC was performed on the S75 (10/300 GL) column using the AKTAexplorer. Suitable fractions were concentrated to a volume of 0.5 ml using Vivaspin 5,000 MW PES by centrifuging at 2,500 rpm and 14°C on an Eppendorf centrifuge 5810R. The OD_{280} nm of the concentrated fractions was checked using a NanoDrop spectrophotometer. For samples containing more than 0.5 mg of sdAbs in 5 ml, SEC was performed on the HiLoad S75 (16/60) column using the AKTAxpress.

Table 50. Detailed information on Bevacizumab V_H host-lab sdAb variants, their molecular weight, theoretical isoelectric points and extinction coefficient values calculated by ExPASy ProtParam tool.

sdAb VUB #	**sdAb name**	**MW (Da)**	**Theoretical pI**	**Extinction coefficient ($M^{-1}cm^{-1}$)**
1	**WT**	14555.13	7.98	44015
2	**V37F**	14584.13	7.25	44015
3	**V37Y**	14600.13	7.25	45505
4	**G44E**	14608.15	6.76	44015
5	**L45R**	14579.11	7.99	44015
6	**W47G**	14406.92	7.25	38515
7	**G44E/L45R**	14651.18	7.25	44015
8	**G44E/W47G**	14478.99	6.76	38515
9	**G44E/L45R/W47G**	14522.01	7.25	38515
10	**V37F/G44E/L45R/W47G**	14570.06	7.25	38515
11	**V37Y/G44E/L45R/W47G**	14586.06	7.25	40005

Day 4-5: SEC runs

SEC was performed on the AKTAexplorer using the S75 (10/300 GL) column (<0.5 mg of sdAb in 5 ml). To ensure the IMAC eluted sample was free of precipitates or

aggregates, two methods were employed. The eluted sample was either centrifuged for 10 min at 14,000 rpm using an Eppendorf centrifuge 5417C or filtered through a 0.2 µm Gyrodisc CA-PC 30 mm filter. Before use, a fresh PBS SEC buffer was prepared. The AKTAexplorer had specific buffer inlet positions: A1.1: PBS SEC buffer, A1.2: 1 M NaOH + 1 M NaCl SEC buffer, A1.3: 200 mM Tris pH 8.0 SEC buffer, A1.4 and A1.8: 20% Ethanol and B1: 20% Ethanol. The Unicorn 5.10 program was initiated, and then SEC procedure was commenced.

3.6.13.1. Unfold/refold before purification

The purification process initiated by transferring the entire culture medium into 300 ml ultra-centrifuge tubes, which were then centrifuged at 4°C and 6,500 rpm for 20 min during two consecutive cycles, using fresh tubes each time. The resulting supernatant was collected, while the pellet was retained.

Next, Ni-NTA beads were washed by adding 0.5 ml or 1 ml of GE Ni-NTA beads solution into 50 ml Falcon tubes for every 100 ml of protein slurry. The final volume was adjusted to 50 ml with 1X PBS, followed by centrifugation at 4°C and 3,750 rpm for 2 min. The ensuing supernatant was discarded.

The culture supernatant was combined with the PBS-washed beads and mixed within a flask. Subsequently, the flask was placed on a magnetic stirrer and kept at 4°C in a cold room for 2 h, resulting in the binding of proteins to the beads. The protein-bead mixture was purified by centrifugation at 4°C and 3,750 rpm. The resin was washed with 100 ml of 1X Purification Buffer (**Table 51**). Thereafter, the resin was dissolved in 100 ml of Buffer A containing 20 mM imidazole and 6 M GdmCl (**Table 52**), followed by incubation at 4°C on a magnetic stirrer for 30 min.

Table 51. Preparation of 10X and 1X Purification Buffer.

10X Purification Buffer (1000 ml)

Na_2HPO_4, 23.015 g	The powders were weighed, dissolved in dH_2O in a glass flask, and sterilized by filtering. They were then stored at RT or in the fridge.
KH_2PO_4, 4.56 g	
NaCl, 292.2 g	
dH_2O, 1000 ml	

1X Purification Buffer (1000 ml)

10X Purification Buffer, 100 ml	The stock buffer was diluted with dH_2O by bringing to a total volume of 1000 ml in a plastic flask on a magnetic stirrer. Subsequently, the

dH_2O, 900 ml	bottle containing the solution was stored at RT or in the fridge.

Table 52. Preparation of Buffer A with 25 mM imidazolen and 6M GdmCl.

Buffer A with 6M GdmCl (400 ml)

10X Purification Buffer, 40 ml	The reagents were combined within a glass bottle. The powder was weighed and subsequently dissolved in Buffer A within another glass bottle. The resultant solution was stored at RT and was safeguarded against exposure to light. Filtration for sterilization was deemed unnecessary, given that the 10X Purification Buffers were pre-sterilized through the filtration process.
2M Imidazole, 5 ml	
GdmCl, 229.272 g	
dH_2O, 400 ml	

Centrifuge purification was again applied to the mixture as explained previously, followed by consecutive washing with 100 ml of purification buffer and 100 ml of Buffer A with 20 mM imidazole + 6 M GdmCl. Then 3 ml of Buffer B with 250 mM imidazole + 6 M GdmCl (**Table 53**) was added to the tubes, after adding 3 ml of PBS at the tube bottom.

Table 53. Preparation of Buffer B with 6M GdmCl.

Buffer B with 6M GdmCl (50 ml)

10X Purification Buffer, 5 ml	The reagents were combined in a glass bottle, and the powder was weighed before being dissolved in Buffer B within another glass bottle. The resulting mixture was filled up to 50 ml and stored at RT, with precautions taken to shield it from light exposure. Notably, there was no requirement for filtration to achieve sterilization, as both imidazole and the 10X Purification Buffer were previously sterilized through the filtration process.
2M Imidazole, 6.25 ml	
GdmCl, 28.659 g	
dH_2O, 50 ml	

The dialysis was initiated by diluting 3 ml of the eluted sample with 10 ml of Buffer B, containing 250 mM imidazole + 6 M GdmCl. The resulting solution was loaded into a 20 cm membrane with sealed ends. A dialysis container was filled with 3 L of 1X PBS for each sample, and the membrane was submerged within the PBS solution. Over the course of up to 18 h, an overnight dialysis procedure was conducted within a cold room, with gentle stirring of the buffer. Following dialysis, the membrane contents were transferred into 50 ml Falcon

tubes and filtered into 15 ml Falcon tubes. Subsequently, each solution was transferred into Amicon tubes.

3.6.13.2. Unfold/refold after purification

The previously isolated proteins were retrieved from -80°C storage and thawed on ice. All protein samples of the same sdAb were pooled into a single tube and the volume was adjusted to 100 ml using 1X PBS. Subsequently, 100 ml of Buffer A containing 20 mM imidazole + 6M GdmCl was added.

The complete 200 ml protein mixture was combined with 1 ml of Ni-NTA beads and placed in a flask for incubation. The flask was positioned on a magnetic stirrer and mixed for 2.5 h within a cold room at 4°C. During this phase, the proteins underwent binding to the beads. Following incubation, centrifuge purification of the protein-bead mixture was performed at 4°C and 3,750 rpm to gather the beads within a vacuum column. The resin was then washed with 100 ml of 1X Purification Buffer. Subsequently, the resin was washed with 75 ml of Buffer A containing 25 mM imidazole + 6M GdmCl.

For the subsequent Buffer B application, 10 ml of Buffer B with 250 mM imidazole + 6M GdmCl was added to the tubes. Notably, PBS was not used at the bottom of the tubes when Buffer B was introduced.

Dialysis was initiated by using 10 ml of eluted sample in Buffer B with 250 mM imidazole + 6M GdmCl. The resulting sample mixture was loaded into a 20 cm membrane with 10 kDa cut off, and the membrane ends were sealed with latches. Dialysis container (a big glass jar) was filled with 3 L of 1X PBS for each sample, and the membrane was immersed in the PBS solution. The procedure involved up to 18 h of overnight dialysis within a cold room, with gentle buffer stirring at a minimal rate.

Subsequent to dialysis, the contents of the membrane were transferred into 50 ml Falcon tubes and filtered into 15 ml Falcon tubes. Each solution was then transferred into Amicon tubes. Notably, extensive washing of the sample was not necessary after this point.

3.6.13.3. Fast protein liquid chromatography (FPLC)-Protein A

FPLC is a widely used technique in protein purification that utilizes high-performance liquid chromatography to separate and purify proteins based on their size, charge, or affinity.

The purification procedure was performed based on the instructions of the manufacturer, and it was optimised for a column of 0.5 ml bed volume.

Firstly, sdAb sample was thawed on ice, and 2 µl of the sample was set aside for further SDS-PAGE analysis. Next, the Protein A resin was taken from the fridge, and 500 µl of resin slurry was transferred to a fresh Eppendorf tube. Then, 500 µl of binding/wash buffer (**Table 54**) prepared was added to the resin. The resin was centrifuged for 1 min at 5,000xg at 4°C, and the washing step was repeated once more to remove the preservative ethanol from the resin.

Table 54. Preparation of buffers for Protein A purification by FPLC.

Binding/Wash Buffer (200 ml)

NaCI, 1.753 g **Na_2HPO_4, 0.5678 g** **dH_2O, 200 ml**	The powders were weighed. They were dissolved in dH_2O in a glass bottle. The pH was set to 7.0. The solution was filtered with a 0.45 µm filter for sterilization. It was then stored at RT.

Elution Buffer (100 ml)

Glycine, 0.750 g **dH_2O, 100 ml**	The powders were weighed. They were dissolved in dH_2O in a glass bottle. The pH was set to 3.0. The solution was filtered with a 0.45 µm filter for sterilization. It was then stored at RT.

Neutralization buffer (50 ml)

Tris HCl, 6.057 g **dH_2O, 50 ml**	The powders were weighed. They were dissolved in dH_2O in a glass bottle. The pH was set to 8.5. The solution was filtered with a 0.45 µm filter for sterilization. It was then stored at RT.

For resin incubation, the washed Protein A resin was mixed with the thawed sdAb sample. The incubation was carried out at RT for 1 h and 30 min, or alternatively, it was incubated overnight in an Eppendorf tube placed on an orbital shaker at a rate of 10 in the fridge.

For batch purification, the 1.5 ml Eppendorf tube containing the incubated sample and resin was centrifuged at 5,000xg at 4°C for 1 min. The supernatant was carefully removed without disturbing the resin, and it was kept for further SDS-PAGE analysis as the flow-through. The resin was then washed with 500 µl of binding/wash buffer (pH 7.0), and the sample was centrifuged at 5,000xg at 4°C for 1 min. The supernatant was carefully removed, leaving the resin at the bottom of the Eppendorf tube. Samples from the supernatant after each

washing step were collected for further SDS-PAGE analysis. This washing process was repeated ten times with 500 µl of washing buffer. Afterward, 450 µl of Elution buffer (pH 3.0) was added to the resin, and it was incubated for a few minutes at RT. The sample was then centrifuged at 5,000 rpm at 4°C for 1 min, and the eluate was collected in a fresh Eppendorf tube already containing 50 µl of neutralization buffer (pH 8.5).

The elution stage was repeated five times to collect five different elution fractions. The concentration of the collected samples was measured using NanoDrop spectrophotometer for further analysis. The blank for elution was prepared by combining 450 µl of elution buffer with 50 µl of neutralization buffer. Similarly, the blank for the wash was simply 500 µl of wash buffer. After the purification process, the collected samples were stored at 4°C until SDS-PAGE results would be obtained.

3.6.14. sdAb characterization

After recombinant protein production in bacterial host, sdAb characterization studies encompassed both qualitative and quantitative analyses. Qualitative characterization involves evaluating the purity and integrity of the antibody fragment, assessing the presence of any potential aggregation or degradation products. SDS-PAGE and Western Blotting (WB) techniques were extensively mployed for these purposes. Quantitative characterization aimed to assess various characteristics of sdAbs. For example apparent binding affinity of the sdAbs was determined using traditional ELISA assay. The following *in vitro* characterization studies were performed on the isolated sdAb samples:

3.6.14.1. SDS-PAGE

The SDS-PAGE system was constructed using one large plate and one short plate and the plates were securely attached to the stand. For the resolving (running) buffer, the TGX™ FastCast™ Acrylamide Kit, 12% from BioRad, was used. The 12% resolving buffer was prepared according to the provided instructions by manufacturer, and the 10% APS (**Table 55**) and TEMED polymerizing agents were not added until the remaining components of the solution were added (**Table 56**).

Table 55. Preparation of APS for SDS-PAGE.

10% APS, 10 ml	
Ammonium per sulphate, 1 g	The powder was dissolved in dH_2O. The aliquots of 1 ml were prepared and stored at -20℃.
dH_2O, 10 ml	

Table 56. Preparation of 12% resolving gel for SDS-PAGE.

12% Resolving (Separating) Buffer	1 gel (µl)	2 gels (µl)
Resolver A	2500	5000
Resolver B	2500	5000
10% APS	25	50
TEMED	2.5	5

10% APS (-20°C) and TEMED were added to the resolving gel. Immediately, the gel was poured into the system until it reached the first green level in the gasket. The resolving gel was allowed to polymerize for 10-15 min at RT. Meanwhile, 8% stacking buffer was prepared according to the **Table 57**.

Table 57. Preparation of 8% stacking gel for SDS-PAGE.

8% Stacking Buffer	1 gel (µl)	2 gels (µl)
Stacker A	750	1500
Stacker B	750	1500
10% APS	75	150
TEMED	1.5	3

The separating gel was prepared by adding 10% APS and TEMED. It was immediately poured on top of the resolving gel. The combs were immersed into the stacking gel right after pouring the gel. While waiting for the gel to polymerize, the sdAb samples were taken from -20°C and kept on ice until they thawed. Once they were thawed, the samples mixed with Laemmli (1:1, 15 µl Laemmli + 15 µl sample 0.1 mg/ml).

The stacking gel was allowed to polymerize for 15-20 min at RT. Lastly, a cold running buffer was prepared according to the **Table 58**.

Table 58. Preparation of 10X and 1X running buffer for SDS-PAGE.

10X Running Buffer, 1000 ml	
0.25 M Tris, 30.2 g	The powders were weighed, and Tris and Glycine were mixed together. Approximately 700 ml of dH_2O was added to a beaker. The powders were then dissolved in the beaker using a magnetic stirrer. SDS was added and mixed until it dissolved completely. The volume was topped up to reach a total of 1000 ml. The pH of the
1.92 M Glycine, 144 g	
1% SDS, 10 g	

dH_2O, 900 ml	solution was measured, and it should be around 8.1-8.5. The pH should not be adjusted using acid or base.

1X Running Buffer (pH 8.3)

10 X Running buffer, 100 ml	The 10X running buffer from the fridge was diluted using dH_2O.
dH_2O, 900 ml	

The gel was placed into the electrophoretic chamber of the Mini-Protean Tetra Cell kit. The vessel was filled with cold running buffer. After polymerization, the combs were carefully removed. Before loading into the gel, the samples were briefly centrifuged.

To prepare the samples, they are diluted into the desired concentration by using 1X PBS. Each sample was added 4X Leammli loading buffer. The samples with 2-ME were boiled 95°C for 5 min. If the loading dye includes DTT instead of 2-ME, then the samples were heated at 70°C for 10 min before running. Loading started with the ladder, followed by the samples using a micropipette considering the total volume per well depending on the total well number of the gel. More running buffer was added to the main vessel after sample loading, ensuring that both the vessel and inner chamber were fully loaded with buffer until it reached the 2 gel filling level. The electrophoretic chamber was connected to the power supply, and the gel was run at RT at a constant voltage of 80 V for the first 10-15 min. During this time, the samples should reach the end of the stacking gel or become aligned with each other. The amperage should not exceed 30 mA per gel. Alternatively, the gel could be run within the fridge using RT buffer. Bubble formation and movement of bubbles upward should be confirmed before increasing the voltage to 100 V. The voltage was kept constant during the running for about 2 h. The running of the samples was checked regularly to see if the dye front reached the bottom of the gel. The gel run was terminated when the sample bands reached the bottom of the gel. The gel was then removed from the plates and placed in a box containing dH_2O.

Protein staining after SDS was performed by transferring the gel to a box and quickly rinsing it with some dH_2O. The gel was washed with dH_2O for 15 min while being shaken at a low rpm on a shaker. Three washes with dH_2O were given for 15 min each. The dH_2O was removed, and Coomassie Brilliant Blue stain was added to the gel until it was covered completely. The gel was left to stain overnight with its lid closed tightly to prevent evaporation. On the following day, the dye was removed and poured into a collection bottle

for potential reuse. The gel was quickly rinsed with some dH_2O, and three more washes were performed with dH_2O for 15 min each while shaking vigorously. If needed, the destaining of the gel could be extended to overnight. Finally, the stained gel was stored within a box filled with dH_2O, and the box was closed with a lid until it was ready for visualization.

3.6.14.2. Western Blot

After SDS-PAGE was terminated, the gel was put into the box and left there for 1-2 min. The PVDF or nitrocellulose membrane and filter papers were cut to a size appropriate for the gel (7cm×9cm). The PVDF membrane was equilibrated in methanol for 1 min and then soaked in dH_2O for 2 min. A large tray filled with cold transfer buffer was used, and everything needed for the blotting was soaked in the buffer (sponge pad×2, filter paper×4, PVDF membrane, and the gel). The transfer sandwich was assembled according to the diagram at **Figure 22**:

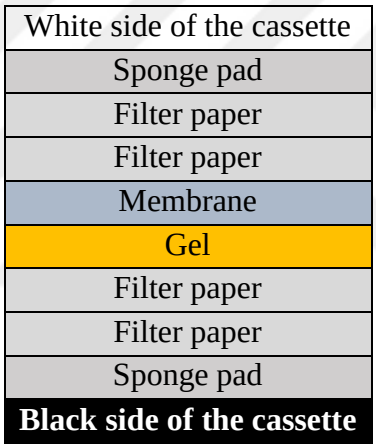


Figure 22. The assembly of transfer sandwich for WB.

The cassette was locked down and placed into the blotting vessel. An ice pack was placed into the blotting chamber. The tank was added the remaining 1X transfer buffer (**Table 59**) until the tank was fully filled.

Table 59. Preparation of 10X and 1X transfer buffer for WB.

10X Transfer buffer, 1000 ml	
25 mM Tris base , 30.2 g **192 mM Glycine, 144.1 g**	The powders were weighed and then Tris and Glycine were combined in a beaker. Approximately 900 ml of dH_2O was added to the mixture. The beaker was placed on a magnetic stirrer, ensuring proper dissolution of the powders. The volume

dH_2O, 900 ml	was then brought up to 1000 ml. pH was expected to be within the range of 8.1 to 8.5.

1X Transfer Buffer, 1000 ml

10X Transfer buffer, 100 ml	Before use, 10X transfer buffer from the fridge was taken. Methanol and dH_2O was prepared just before use. It was kept cold to maintain its freshness.
Methanol (or isopropanol), 100 ml	
dH_2O, 800 ml	

The blotting initiated with the connection of the vessel to the power supply, setting constant amperage of 180-350 mA for an estimated duration of 60-70 min. During the transfer, the TBST buffer and the blocking buffer were prepared (**Table 60**).

Table 60. Preparation of PBS-T and 5% blocking buffer for WB.

PBS-T, 1000 ml

10 mM Sodium phosphate (pH 7.2), 10 ml	The powders were dissolved in approximately 700 ml of dH_2O on a magnetic stirrer. Subsequently, the pH was adjusted to 7.6.
NaCl, 9 g	
Tween-20, 1 ml	Once the pH adjustment was completed, Tween 20 was added to the solution. Following this, the volume was increased to reach a total of 1L. The solution was stored either at RT or in the fridge.
filtered dH_2O, 1000 ml	

5% Blocking Buffer

fat-free milk powder, 12.5 g	The powder was dissolved in a glass bottle using a magnetic stirrer. Once the solution was prepared, it was stored in a +4°C refrigerator.
1X PBS-T, 250 ml	

After the power supply was terminated, the cassette was carefully removed from the tank. The cassette was then opened to remove the membrane inside. The membrane was carefully transferred to a box containing blocking buffer and it was then incubated with blocking buffer for 1 h on a shaker at 65 rpm and at RT. Throughout the incubation, the box lid was kept closed to prevent evaporation. Alternatively, the blocking step could be performed overnight at 4°C. After the incubation, the blocking buffer was removed by discarding the liquid into the sink. Subsequently, the membrane was washed with PBST

buffer, ensuring enough volume to cover its surface. A quick wash was followed by placing the membrane on the shaker at approximately 65 rpm for 15 min at RT. This washing step was repeated three times to ensure thorough and efficient removal of excess blocking buffer.

For the next step, 40 ml of blocking buffer was added to a 50 ml falcon tube. The primary antibody, Anti-Flag Mouse Antibody (1:10000) (**Table 61**), was taken from the -20°C storage and dissolved in 1% PBST buffer for use.

Table 61. Preparation of anti-Flag tag mouse antibody for WB.

Anti-Flag Mouse Antibody (1:10000), 40ml
Anti-Flag Mouse Antibody, 4 µl
5% blocking buffer (1% final concentration), 8 ml
PBST, 31.96 ml

The membrane was incubated with the primary antibody overnight in the cold room at 4°C shaking at 55 rpm. Alternatively, the antibody buffer can be stored at -20°C for re-probing in the future experiments. Depending on the primary antibody, an alternative incubation method of 1 h at RT can be considered. After the primary antibody incubation, the membrane was washed with PBST buffer, ensuring sufficient volume to cover its surface. A quick wash with the buffer was followed by placing the membrane on the shaker at approximately 65 rpm for 15 min at RT. This washing step was repeated three times to achieve thorough removal of any unbound primary antibody and to prepare the membrane for subsequent stages. Subsequently, the secondary HRP-conjugated antibody (1:10000) was taken from the -20°C and dissolved in 1% PBST buffer (**Table 62**).

Table 62. Preparation of HRP-conjugated antibody for WB.

HRP-conjugated antibody (1:10000), 40ml
HRP-conjugated antibody, 4 µl
5% milk (in PBST) (1% final concentration), 8 ml
PBST, 31.96 ml

The membrane was incubated with the secondary antibody at RT on the shaker at 65 rpm, for 1-2 h. Following the incubation, the blocking buffer was removed from the membrane. Next, the membrane was washed with PBS, using sufficient volume to cover its surface. A quick wash with the buffer was performed, followed by placing the membrane on the shaker at 65 rpm for 15 min at RT. This washing process was repeated three times to

ensure effective removal of any remaining unbound antibody from the membrane for its visualization.

For membrane imaging, the EZ-ECL Chemi-luminescence Detection Kit was used. Equal volumes of Reagent A (peroxidase solution) and Reagent B (luminol enhancer solution) were added and mixed in a falcon tube (1500 µl +1500 µl). It was important to prepare this mixture just before use and protect it from light to prevent from bleaching. The membrane was taken out of the box, and any excess moisture was drained from it wih the help of a tissue paper. The membrane was then carefully placed on the black platform using forceps. Next, the detection mixture (ECL) was added to the membrane, ensuring it was covering the all surfaces of the membrane. The chemi-luminescence mode was selected for membrane visualization. The exposure time was adjusted accordingly (auto or manual). For the ladder visualization, picture of the membrane was captured.

3.6.14.3. Enzyme-Linked Immunosorbent Assay (ELISA)

On Day 1 of the experiment, the first step involved measuring the protein concentration using NanoDrop spectrophotometer before proceeding with the experiment. Based on the results obtained, the protein concentration range to be tested was planned. The required amount of protein for the desired range of concentrations was calculated to ensure the presence of triplicate for each concentration range. Next, a 96-well plate was prepared for the experiment. The plate was coated with 1.5 ng/ml VEGF (**Table 63**). 1 µg/ml VEGF stock was diluted in 1XPBS. The final concentration of VEGF coating in the wells was set at 1.5 ng/ml.

Table 63. Preparation of mouse or human VEGF for coating of ELISA well plate.

1.5 ng/ml mouse or human VEGF, 9.9 ml	
1 µg/ml (1000 ng/ml) VEGF, 14.85 µl	PBS was added into a 15 ml Falcon tube. Next, 1 µg/ml (1000 ng/ml) VEGF stock was directly added into the PBS in the tube.
PBS, 9.9 ml	

For each sample well, 100 µl of the 1.5 ng/ml VEGF solution was added (**Figure 23**). As for the control wells, 1XPBS was added instead of the VEGF solution. Later in the afternoon, the well plate was set up for overnight incubation in the fridge. The plate lid was closed, and it was covered with aluminium foil to ensure a controlled and dark environment during incubation. To prevent possible evaporation during the incubation period, the plate was placed in a sealed plastic bag.

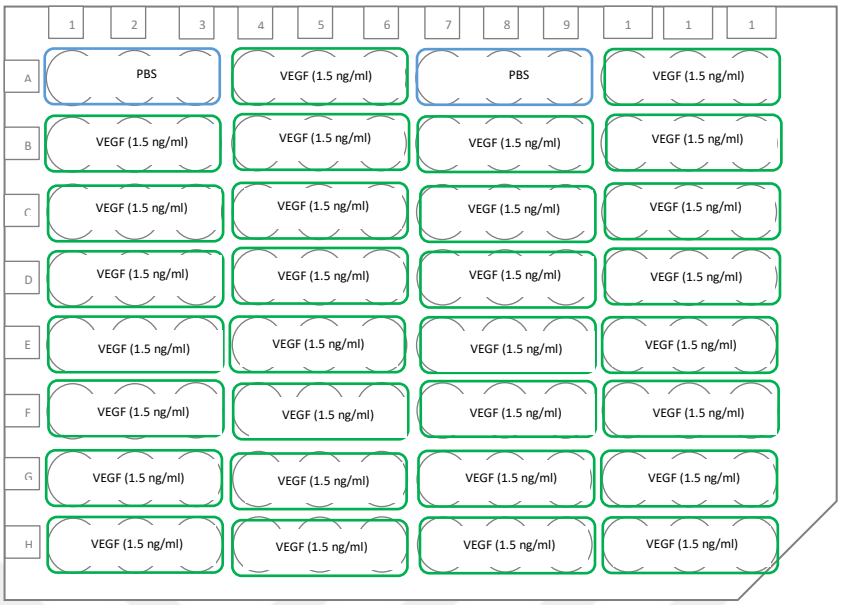


Figure 23. An ELISA plate alignment.

On Day 2, in the morning, the blocking buffer (**Table 64**) was prepared using skimmed milk and 1XPBS according to the table below:

Table 64. Preparation of 10% blocking buffer for ELISA.

10% (w/v) Blocking buffer, 11 ml

Powdered skimmed milk, 1.1 g **1X PBS, 11 ml**	Cold 1XPBS was poured into a 15 ml falcon tube. The required amount of milk powder was weighed and then added into the falcon tube containing 1XPBS. To ensure thorough mixing, a vortex was applied.

The well plate was blocked using a 10% (w/v) skimmed milk solution in 1XPBS. Each well was added 100 µl of the blocking buffer without removing the excess VEGF. The plate was then closed and covered with aluminium foil. It was sealed in a bag and left for 6-8 hours of incubation at RT. Alternatively, overnight incubation could be applied in the fridge for blocking. During the incubation period, the samples were prepared in triplicates by diluting them from the main protein stock with 1XPBS through serial dilution. These samples were stored in the fridge until they were ready to be added to the well plate.

After the incubation with the blocking buffer, the contents in the wells were discarded. The wells were then washed three times with 1X PBS. The washing process was generous, almost filling the wells to the top. After washing with 1XPBS, the contents of each well were

discarded into the sink with a swift movement, and the wells were briefly dried using tissue paper. Following the washing step, varying concentrations of protein samples ranging from 0-1000 nM were added to the wells accordingly (**Figure 24**). Control wells were added 1XPBS instead of protein samples for comparison.

After adding the varying concentrations of protein samples and 1XPBS to the respective wells, the 96-well plate was left for overnight incubation. The plate lid was closed, and it was covered with aluminium foil for the incubation. The plate was then incubated overnight at 4°C in the fridge.

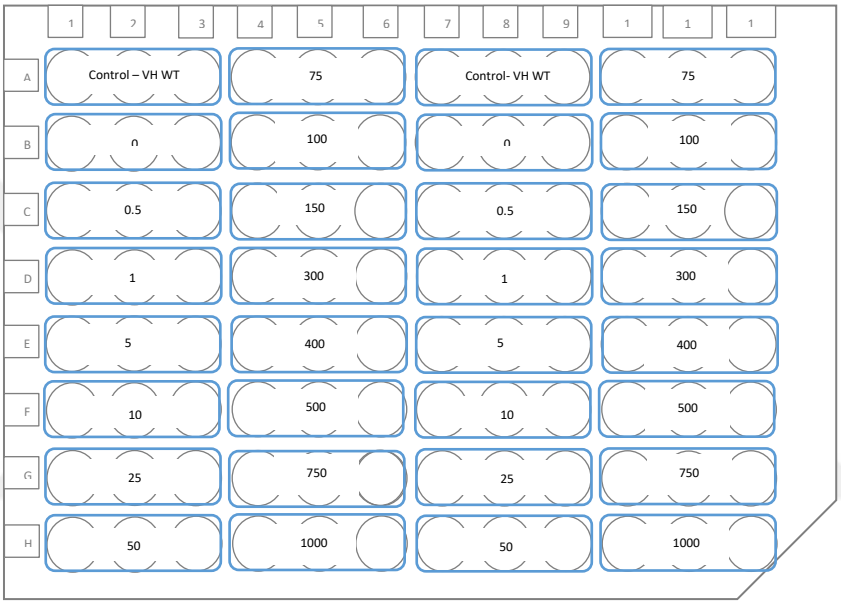


Figure 24. ELISA plate configuration for antibody sample concentration range.

On Day 3, PBS-T was prepared according to **Table 65** to dissolve the antibodies in.

Table 65. Preparation of PBS-T for ELISA.

0.1% (v/v) PBST, 11 ml	
1X PBS (filtered), 11 ml	1X PBS was added into a 15 ml falcon tube. Triton X-100 was then added to the 1X PBS in the falcon tube, and the mixture was thoroughly mixed to achieve a homogeneous solution. It was stored in a +4℃ fridge.
0.1% (v/v) Triton-X 100, 11 µl	

Following the previous washing step with 1X PBS, each well was incubated with 100 µl of primary antibody in PBS-T. The primary antibody used was mouse anti-Flag-M2 IgG (**Table 66**), and it was diluted at a ratio of 1:5000.

Table 66. Preparation of Anti-Flag mouse antibody for ELISA.

1:5000 Primary Antibody, 10.6 ml	
Anti-Flag mouse IgG, 1.06 µl	To prepare PBS-T was added into a 15 ml falcon tube. The antibody was directly added to the PBS-T in the falcon tube.
PBS-T, 10.6 ml	

The incubation was performed at RT for 1 h, during which the plate lid was kept closed, covered with aluminium foil, and placed in a sealed bag. These measures ensured a controlled environment for the antibody incubation, promoting specific binding to the target proteins immobilized on the plate surface. After the incubation period, the plate was ready for further analysis or additional steps in the experiment.

Following the previous washing step with 1X PBS, each well was treated with 100 µl of anti-mouse IgG HRP-conjugated secondary antibody in PBS-T (**Table 67**). The secondary antibody was diluted at a ratio of 1:10000.

Table 67. Preparation of HRP-Conjugated Goat Anti-mouse antibody for ELISA.

1:10000 HRP-Conjugated Goat Anti-mouse Secondary Antibody, 10.6 ml	
HRP-Conjugated Goat Anti-mouse IgG, 1.06 µl	To prepare PBS-T, PBS was added into a 15 ml falcon tube. Then, the antibody was directly added to the PBST in the falcon tube.
PBS-T, 3 ml	

The incubation with the secondary antibody was performed at RT for 1 h. During this incubation period, the plate lid was kept closed, covered with aluminium foil, and placed in a sealed bag. After the incubation, the plate was ready for further processing or detection steps in the experiment.

After the incubation with the secondary antibody, the wells were washed three times with 1X PBS, following the procedure described earlier. To determine peroxidase activity, 100 µl of the associated substrate, 1-Step Ultra TMB-ELISA, was added to each well. This reagent is light-sensitive, so an appropriate amount of TMB was taken into an aluminium foil-covered falcon tube, and the wells were treated in a dark environment to prevent degradation of the reagent. Next, the wells were incubated with the HRP substrate TMB at RT for at least 30 min. This incubation allowed the enzymatic reaction to take place, leading to the development of a colourimetric signal. After the incubation, the reactions were quenched by adding 2M H_2SO_4 to each well (**Table 68**).

Table 68. Preparation of 2M H_2SO_4 solution for ELISA.

2M H_2SO_4 Solution, 12 ml	
18.4 M of H_2SO_4, (w/v) %98, 1.304 ml	The falcon tube was filled with dH_2O. Using a glass Pasteur pipette, H_2SO_4 was carefully transferred into the falcon tube until the volume reached 12 ml. A 2M H_2SO_4 solution was prepared in the falcon tube.
dH_2O, 10.696 ml	

This step effectively quenched the enzymatic reaction. The ELISA plate was read using a plate reader at 450 nm wavelength. The absorbance measurements obtained from the reader were then the values were further used in the calculations.

To calculate the results of the ELISA experiment, the mean of the triplicates was obtained for each concentration of the protein sample. The background signal was calculated by subtracting the absorbance value of the 0 nM concentration (control row) from the absorbance values of each sample. The control row, which contained 1XPBS instead of antigen coating, was excluded from the background calculation. Next, the ratio (Sample-Background) / Background was calculated for each concentration by dividing the subtracted value from the previous step by the background value. After the calculations, a graph was plotted with concentration on the x-axis and absorbance at 450 nm on the y-axis. The linear equation was added to the plot for data interpretation.

3.6.14.4. Tm determination

Thermal melting (T_m) determination is used to assess the stability characteristics of proteins. It involves subjecting a sample to increasing temperatures and monitoring the changes in its structure and function. Two methods used in thermal melting studies are heat denaturation and chemical denaturation. Heat denaturation involves gradually increasing the temperature of the sample, causing the proteins to change their three-dimensional structure. Chemical denaturation relies on the use of denaturants, such as urea or GdmCl, to induce unfolding of the molecule. The denaturant concentration is gradually increased, and the unfolding process is observed through measurement of absorbance or fluorescence.

3.6.14.4.1. Thermal denaturation

The protein melt reactions were prepared by following these steps:

Four replicates of each reaction were prepared. A fresh dilution of Protein Thermal Shift™ Dye was prepared from the 1000X stock to an 8X working volume. The reaction plate

was placed on ice, and then the reaction components were added to the plate in the order listed at **Table 69**.

Table 69. Preparation of protein thermal shift reactions for T_m determination.

Reagent	Volume (µl)
Protein Thermal Shift™ Buffer	5.0
1X PBS + protein (0.1 to 1mg/ml stock) + buffer or ligand	12.5
Diluted Protein Thermal Shift™ Dye (8X)	2.5
Total volume for each reaction	**20.0**

The protein melt reactions were mixed thoroughly by pipetting up and down 10 times. Next, 20 µl of the reaction volume was added into each well of a 96-well plate. The plate was sealed with film and centrifuged at a very low speed (300 rpm) for 1 min. For screening analyses using the ABI 7500 Fast Systems, the real-time PCR instrument was set up and run accordingly. The run conditions were as listed **Table 70**.

Table 70. The set up details of thermal melt reaction using ABI 7500 Fast Systems instrument.

Setup	Setting
Experiment properties	• Experiment type: Melt Curve • Reagents: Other • Ramp speed: Fast or Standard
Target properties	• Reporter: ROX • Quencher: None
Play layout	• Assign targets to all wells in use • Passive reference: None

Run method	• Reaction volume per well: 20 µl
	• Ramp mode: Continuous
	• Thermal profile:
	  o Step 1, Temp: 25°C, Time: 2 minutes
	  o Step 2, Temp: 99°C, Time: 2 minutes
	• Ramp rate:
	  o Step 1: 1.6°C/s, 100%
	  o Step 2: 0.05°C/s, 0.5%
	• Optical filters:

- Excitation filter: x4(580±10)
- Emission filter: m4(623±14)

The system run of Real-Time PCR was initiated by loading the reaction plate into the instrument. Melt curves on the ABI 7500 Fast System were created using the Real-Time PCR System Software. In the Well Results screen of the melt curve plots, fluorescence data were displayed as fluorescence melt curves and derivative melt curves. The melt curve plots included information about the regions of analysis, Boltzmann T_m values, and derivative T_m values. The fluorescence melt curves showed noise-reduced fluorescence data obtained during the protein melt curve run.

3.6.14.4.2. Chemical denaturation

GdmCl is commonly employed as a denaturant in chemical denaturation studies. The proteins to be tested were thawed or fresh protein samples were used. A stock solution of 8M GdmCl was prepared (**Table71**). The following reactions were then prepared, each containing different concentrations of GdmCl as specified in the **Table 72** below.

Table 71. Preparation of 8M GdmCl for chemical denaturation assay.

8M GdmCl (10 ml)	
GdmCl, 7.64 g	The powder was weighed and added to a glass beaker, and then the dH_2O was gradually added while stirring, as it dissolves rapidly. The light sensitive solution was stored at RT, covered with aluminium foil.
dH_2O, 10 ml	

Table 72. Preparation of GdmCl ascending concentration range.

GdmCl concentration (M)	**Protein concentration in PBS (mg/ml)**	**Total Volume (µl)**
0	0.075	110
0.5	0.075	110
1.0	0.075	110
1.5	0.075	110
2.0	0.075	110
2.5	0.075	110
3.0	0.075	110
3.5	0.075	110
4.0	0.075	110

4.5	0.075	110
5.0	0.075	110
5.5	0.075	110
6.0	0.075	110

The samples with GdmCl were placed in PCR tubes and incubated overnight at 10°C in a PCR machine. The thermal cycler was set to a total volume of approximately 100 µl, with the cover temperature option turned off. After the overnight incubation, the protein samples were transferred from the PCR tubes into black plates. The run conditions on the plate reader were set as indicated at **Table 73**:

. **Table 73.** The run conditions of the plate reader for chemical denaturation assay.

Setup	Setting
Excitation	280 nm
Emission	300-400 nm
Range	Auto-range
Measurement time	500 ms

Following the run, the results were exported as both an Excel file and a Notepad file. The raw data was used in the Excel file, and based on the average maximum and minimum readings at the corresponding wavelengths; the final graphs were plotted using the fluorescence reading ratios at F_{345} nm / F_{325} nm.

3.6.14.5. Thermal stability determination assay

This method is employed for monitoring the global propensity for protein aggregation. It uses the soluble fraction of total proteins, which is incubated *in vitro* for several hours.

Protein samples are diluted to 20 µM using 1X PBS. However, depending on the sample, the concentration may be optimised within the range of 20-30 µM.

The total number of data points for sample collection was determined. To begin the process, 30 µl of a 20 µM protein sample was added to each PCR tube. An additional tube was set aside for the 0 min time point as the first sample. Subsequently, the PCR tubes containing the protein samples were incubated at 65°C in a thermal cycler. Alternatively, another physical insult, shaking at 225 rpm, was included in the setting of the experiment. Samples were then collected at different time points. The second sample was taken at 2 min, and the third sample was taken at 10 min after the start of the incubation. After taking the 10-

min sample, subsequent samples are collected at 30-min intervals until the desired data points are reached.

Right after incubation, the contents of the collected PCR tubes were transferred to 1.5 µl Eppendorf tubes, with their lids facing downward. The samples are then centrifuged at a maximum of 17,000xg at RT for 10 min. The aggregated proteins were expected to accumulate at the bottom of the tubes, tilted upwards direction. To retrieve the samples, 6 µl of the contents were carefully taken from the downward direction of the bottom of the tubes. The concentration changes were monitored with sequential measurements at the designated time intervals to track the aggregation of proteins over time.

For the protein concentration measurement, 6 µl of the collected samples from each tube were carefully transferred into PCR tubes. Next, the concentration value of each sample was measured at least twice using a NanoDrop spectrophotometer using the appropriate calculated molecular weight and extinction coefficient values. The average of the protein concentration measurements was obtained for each time point. To calculate the aggregation fraction, the concentration value at time 0 min was subtracted from each time point value (t_0 - t_x). Then, the resulting value was divided by the concentration value at time 0 min to obtain the aggregation fraction, expressed as (t_0 - t_x)/t_0. This fraction represents the aggregation of proteins over time compared to their initial concentration at time 0 min. The values for each protein sample were plotted in graph, with the x-axis representing time (in min) and the y-axis representing the aggregation fraction.

3.6.14.6. Isothermal Titration Calorimetry (ITC)

The ITC assay was performed as service from IBG Pharma department. All the buffers and samples were de-gassed before use. The run conditions were as per the instructions of the device manufacturer. The details of the run are outlined in **Table 74** as follows:

Table 74. The experimental parameters of ITC.

Experimental parameters	Value
Syringe Concentration (mM)	0.147
Sample Concentration (mM)	0.019
Initial Volume (µl)	182
Temperature (°C)	25
Stir Rate (rpm)	75

3.6.14.7. In vivo zebrafish studies

The ideal living conditions for zebrafish, which thrive at temperatures ranging from 27-28°C. However, they are remarkably adaptable and can survive in extreme conditions, tolerating temperatures as low as 17°C and as high as 36°C, even reaching up to 41°C. As adults, they reach a size of 3-5 cm. These fully-boned fish share approximately 84% gene homology with human (Avdesh et al. 2012).

During the first 1 to 3 dpf, the zebrafish larvae exhibit transparency, which proves advantageous for vasculature studies as it allows direct visualization. The larvae become fully adult and capable of producing their own offspring after approximately 3 months. Being competitive and carnivorous, zebrafish depend on external fertilization, which takes around 5 min to complete. Within just 3 h after fertilization, the embryos develop to a 1000-cell stage by dividing every 15 min. Within the first 24 h post-fertilization, a prototype organism with all organs can be obtained (Bauer, Mally, and Liedtke 2021).

For the initial 5 days of life, young zebrafish feed on their yolk sack. During the first 2 dpf, zebrafish are surrounded by a protective chorionic membrane. This membrane must be manually removed or treated with the enzyme pronase to facilitate successful injection into the yolk sack and avoid interference with anaesthetics.

The zebrafish utilized in these experiments were supplied by the zebrafish core facility at the IBG. All procedures involving zebrafish were approved by the ethical committee (HADYEK) of the IBG. Adult zebrafish were housed at the IBG Zebrafish Facility, maintained under standard conditions with a temperature of 28°C and a 14/10-h light/dark cycle. To obtain zebrafish embryos, the transgenic Tg(fli1:EGFP)^{y1} strain was crossed with wild type ones. Embryos were collected within 30 min of fertilization and incubated at 28°C in a dark incubator using E3 embryo medium (**Table 75**).

Table 75. Preparation of 60X stock and 1X E3 medium for zebrafish studies.

60X E3 medium

NaCl, 17.2 g	To make 60X stock solution, the powders were measured dissolved in 950 ml ddH_2O. The pH was adjusted to 7.2 using NaOH and HCl The final volume was adjusted to 1 L using ddH_2O. Medium was sterilized by autoclaving.
KCl , 0.76 g	
$CaCl_2.2H_2O$, 2.9 g	
$MgCl_2.6H_2O$, 4.9 g	
ddH_2O, 1000 ml	

1X E3 medium

60X stock E3 Medium, 150 ml **0.3 % Methylene Blue, 180 µl** **ddH_2O, 9 L**	60X stock medium was diluted in ddH_2O. 0.3 % Methylene Blue was added to the medium.

Embryos obtained from the Fli1a:GFP transgenic zebrafish line were cultured in E3 medium at 28°C for 2 days. Zebrafish larvae hatched from their chorions were anesthetized with 0.04% tricaine and arranged in microinjection moulds prepared with 2% agarose. Antibodies were prepared at a concentration of 1.8 mg/ml in PBS, aliquoted, and stored at -80°C in liquid nitrogen (in previous instances, it was noted that the effectiveness of antibodies diminished when thawed following freezing). Phenol red at 0.01% was added to the antibody solution and loaded into the capillary injection needle. The microinjection volume was measured by injecting it into mineral oil on a specialized microscope slide. The yolk sac of each larva was injected with 20 nl of antibody solution. PBS was injected into the control group. It is essential to inject all peer fishes consistently, using the same direction, the same location, and the same amount of the antibody. Larvae were removed from anaesthesia and incubated in E3 medium in 6-well plates for 1 day.

The larvae were anesthetized with tricaine (**Table 76**) for live imaging. A mounting solution was prepared by dissolving 1% low-melting agarose in E3 medium, maintained at 37°C to prevent solidification. Larvae were transferred into the agarose using a glass Pasteur pipette, then transferred to glass-bottom imaging dishes. Larvae were oriented using a fine needle until the agarose solidified. For microinjection and fixation, 40 ng of Bevazicumab WT sdAb and Bevazicumab V37Y sdAb, and either 125, 250 or 500 µM of Bevacizumab were kept on ice until injection in PBS supplemented with 0.5% phenol red. A stage micrometer was employed to estimate the injection volume, and each anesthetized zebrafish embryo at 2 dpf with tricaine received 50 nL of sdAbs and Bevacizumab injected into the yolk sac.

Table 76. Preparation of anaesthetics for zebrafish studies.

50 mg/ml Stock Tricaine Solution

Tricaine methanesulfonate, 2.5 g **ddH_2O, 50 ml**	The powder was measured and dissolved in ddH_2O. The solution was stored at RT.

More than 15 embryos were used for each condition, and the microinjected larvae were kept at 28°C, with the development of subintestinal vessels (SIV) examined at one day post-injection. The larvae were then fixed in 4% formaldehyde and rinsed with PBS, followed by melanocyte decolourization using a pigment discolouration solution (**Table 77**).

Table 77. Preparation of decolourization solution for zebrafish studies.

Pigment Discolouration Solution

Hydrogen peroxide (30%), 1 ml **Potassium hydroxide (5%), 1 ml** **ddH_2O, 8 ml**	The liquids were mixed and diluted in ddH_2O. The solution was stored at RT.

Z-stack images of the entire SIV vasculature and a portion of the dorsal embryo vessels were captured using a 25X water immersion objective on the LSM 880 confocal microscope. Due to the time-consuming nature of imaging, fixation was necessary to capture all samples. For this purpose, fixation was tested with 4% PFA for 2-4 h at RT and 8-16 h at 4°C. When the fluorescent signal of the samples was tested with a stereomicroscope, it was determined that 2-h RT fixation preserved the samples and fluorescent signal best. This method was chosen for subsequent experiments.

For imaging and quantitative analysis, the larvae were laterally mounted in 1% low melting agarose (**Table 78**) in an imaging dish and imaged using a Zeiss LSM880 confocal microscope. When performing embryo moulding, 3% basic agarose (**Table 78**) is used. Live moulding yields better results; however, fish can only remain in the mould for a maximum of 1.5 h due to potential harm from the laser used in confocal microscopy. For fixation purposes, 4% paraformaldehyde (**Table 79**) was prepared fresh.

Table 78. Preparation of basic and low melting agarose for zebrafish studies.

Low melting agarose gel, 1% (w/v), 50 ml

Agarose powder 1% (w/v), 0.5 g **1X TAE Buffer, 50ml**	In a flask, 0.5 g of agarose powder was weighed and mixed with 50 ml of 1X TAE buffer.

Agarose gel, 3% (w/v), 50 ml

Agarose powder 3% (w/v), 1.5 g **1X TAE Buffer, 50ml**	In a flask, 1.5 g of agarose powder was weighed and mixed with 50 ml of 1X TAE buffer.

Table 79. Preparation of paraformaldehyde for zebrafish studies.

Paraformaldehyde, 4% (w/v), 100 ml

Paraformaldehyde, 4 g	The powder was weighted and dissolved in 1XPBS.
1X PBS, 100 ml	
1N NaOH	The pH was adjusted to pH 6.9 to 7.4 using NaOH.

The area of the SIV was determined using the ImageJ tool, and statistical analysis was performed in Excel using a one-tail t-test with a 95% confidence interval. The statistical significance was indicated using stars (*p≤0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001) to represent a range of p-values. The results are presented as average ± standard error.

3.6.15. In vivo mice studies

The *in vivo* behavior of camelised sdAbs derived from the commercial full-format mAb Bevacizumab, targeting VEGF, was planned to be studied in athymic nude mice (Crl:NuFoxn1^{nu}). This project was written in collaboration with Prof Dr. Sophie Hernot at VUB (Belgium) as part of the application for European Molecular Biology Organization (EMBO) short-term project (STF8943). The second half of the project was further funded by TUBITAK 2214A program (1059B142100119).

Following a successful application, EMBO awarded funding for the application, enabling the initiation of the project. An application for mice studies was made for the approval of the local ethical committee of the host lab institute. The total number of animals required for the planned experiments was detailed in the **Figure 25**, **Table 80** and **Table 81**. Upon obtaining approval from the ethical committee at VUB, the initiation of animal breeding and studies on human cancer cell growth commenced.

The aim was to assess whether camelised sdAbs derived from the clinically approved antibody Bevacizumab exhibit improved *in vivo* pharmacokinetics, tumour targeting, and penetration compared to Bevacizumab itself. The study also sought to confirm that the biodistribution of these camelised sdAbs is equivalent to that of traditional nanobodies. To achieve this, 10 mutants of the Bevacizumab-derived sdAb were engineered in the home lab with one to four mutations in their framework to enhance solubility. Following the generation and *in vitro* characterization of up to 10 anti-VEGF camelised sdAbs derived from Bevacizumab, their whole-body biodistribution and tumour targeting potential were planned to be investigated via non-invasive nuclear imaging in mice bearing a subcutaneous tumour. The data would be compared with the biodistribution of the non-camelised WT sdAb and a

non-targeting traditional nanobody. The traditional nanobody served as both a negative control for tumour uptake and a reference for a proper general biodistribution profile, characterized by fast blood clearance, low liver uptake, and high kidney accumulation. Additionally, the biodistribution and tumour uptake of the camelised sdAbs obtained at 1 h post-injection were planned to be compared with that of Bevacizumab at both 1 h and 24 h post-injection. The 1-h time point was chosen as it is typically optimal for sdAbs/nanobodies, while the 24-h time point is commonly used for imaging with full antibodies.

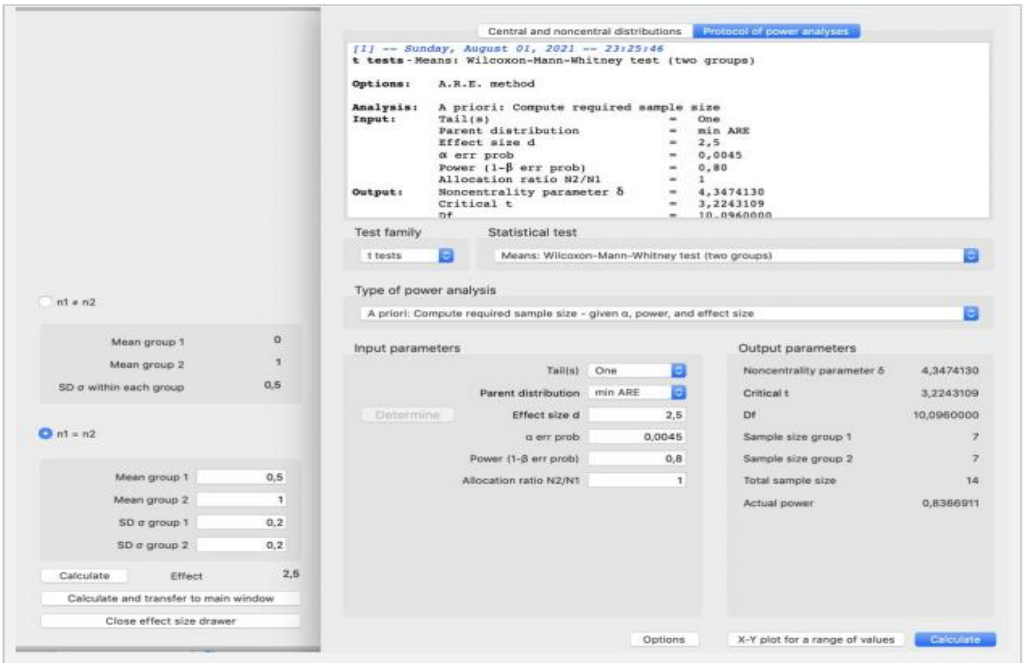


Figure 25. A power analysis with α set at 0.05:11 (11 comparisons) = 0.0045. A power of 0.80, indicates that minimally 7 mice per group are needed to reach statistical significance.

The application sought approval from the local Ethische Commissie Dierproeven (ECD) at VUB.

Table 80. The groups indicated for the comparison and the total number of mice needed per group for whole body distribution studies based on the power analysis.

Mice group	# animal per group
10x camelised anti-VEGF sdAbs	10x7 mice
1x WT anti-VEGF sdAb	1x7 mice
1x control non-targeting traditional nanobody	1x7 mice
2x Bevacizumab (at 1h and 24h)	2x3 mice
Total #	**90 mice**

The evaluation of the microscopic localization of the two tracers (Bevacizumab and the anti-VEGF sdAb) would be conducted in tumour tissue. Due to the qualitative character of the results, the least possible number of animals were requested to assess this variation (n=3 per group) (**Table 81**).

In the second part of the project, a more detailed investigation was planned to be conducted on the tumour penetration of the most promising anti-VEGF camelised sdAb selected from the first part of the planned experiment. Both the selected sdAb and Bevacizumab would be fluorescently labelled, intravenously injected, and tumour samples subsequently resected for microscopic analysis.

Table 81. Total number of mice bred for the investigation of tumour penetration of Bevacizumab-dervied sdAbs.

Mice group	# animal per group
1x camelised anti-VEGF sdAb	3 mice
1x Bevacizumab	3mice
Total #	**6 mice**

The welfare, food, and water supply of the mice were monitored daily, along with observations of animal behavior, physical appearance, and tumour growth as outlined in the in the scoring system within the ethical committee application of the host lab. Measurements of body weight and tumour size were conducted every 2-3 days.

3.6.15.1. HT-29 cell line and culture conditions

The HT-29 cell line (HTB-38, ATCC) was available at the host lab at VUB (**Figure 14**). The cell culturing and growth protocol was based on the instructions of the manufacturer. The complete growth media includes 10 % (v/v) FBS and 100 U/ml penicillin and 100 µg/ml streptomycin. In addition to these reagents, the cells were cultured in McCoy's5A media and 2 mM L-glutamine (**Table 14**).

The flasks were taken from the cell culture incubator. The confluence of the cell culture flask was checked under a microscope. Next, the media was removed from the flask, and, if needed, the cells were rinsed with PBS. Subsequently, trypsin was added to the cells, and the cells were incubated for 5 min in a 37°C incubator where the culture flasks were kept. The cells were then dislodged by tapping the flask and the detachment was confirmed under the microscope. After that, complete media was added to the flasks to eliminate the trypsin effect. The cells were resuspended and transferred into a 50 ml pre-labelled falcon tube.

Following this, the cells were centrifuged at 1,500xg for 5 min, and the supernatant was removed. The cells were then resuspended in 3 ml of media. The resuspended cells were re-seeded into the flask containing complete media. On the flask, the cell line, medium information, passage number, date of passaging, dilution ratio of the cells, and the name of the person handling the procedure were recorded. Finally, the flask was placed back into the incubator for further cultivation.

The cells were grown until they reach max $3x10^6$ cells in a maximal volume of 150 µl and subsequently would be injected subcutaneously on the right flank of the animal.

3.6.15.2. HT-29 cell counting

The cell growth was checked under a microscope. The media was then removed from the flask. Next, trypsin was added to the cells and the cells were incubated for 5 min in a 37°C incubator. The cells were dislodged by tapping the flask, and the detachment was confirmed under the microscope. Complete media was added to the flasks to neutralize the trypsin effect. Subsequently, the cells were resuspended and transferred into a 50 ml pre-labelled falcon tube. In the case of multiple flasks containing the same kind of cells, they were combined into a single falcon tube. The cells were then centrifuged at 1,500 rpm for 5 min and the supernatant was removed without disturbing the cells.

The cells were resuspended in 10 ml of fresh media until no clusters or clumped cells remained. At least 2x 1:10 (10 µl cells + 90 µl of medium) dilution of the cells were prepared in an Eppendorf tube. If the 1:10 ratio was not sufficient, a 1:100 (1 µl cells + 99 µl of medium) dilution was prepared.

The automated cell counter was turned on, and a new cell counting chip was inserted by pushing down the front lever. A 75 µl sample of cells from the Eppendorf tube was pipetted, and one of the available two sides of the chip was used to load the sample. The counting process was initiated by tapping the screen, and the results appeared on the screen without using gating, but the curve option was used. The cell counting process was repeated up to 4 times, and the average of the cell counts was calculated. This average result was multiplied by the dilution factor (1:10 or 1:100) to determine the number of cells in 1 ml.

The HT-29 cells were successfully cultured for the intended animal studies involving subcutaneous injections into mice flanks.

3.6.16. Fluorescent labelling

Fluorescent labelling involves the reduction of cysteine bonds within the sdAb samples and followed by the conjugation with a maleimide derivative of fluroscent dyes (IRDye800CW and Alexa Fluor-647).

3.6.16.1. Reduction and buffer exchange

The working quantity of sdAb-Cys tagged (3 mg) was found sufficient for sample preparation, *in vitro* characterization of the end product, and for a small-scale *in vivo* study in mice, typically requiring 2 nmol of fluorescent-labeled sdAb per animal. A test run was conducted before perfoming the planned experiments.

2-MEA was employed in a molar excess as a mild reducing agent, applied to reduce the oxidized sulfhydryl group of the cysteine tag on a monomeric sdAb or to cleave the disulphide bond formed between the cysteine tags of two sdAbs, without disturbing their native structure. The use of EDTA prevented the re-oxidation due to chelation of divalent metal cations present in the reaction mix, keeping the sulfhydryl groups ready for the conjugation reaction.

To begin the protocol, 3 mg of pure sdAb with a carboxy-terminal cysteine tag was prepared in a maximum volume of 2 ml of PBS. Next, a 90-fold molar excess of 2-MEA was calculated for monomeric sdAb, based on the molar mass of the sdAb to be labelled. The 2-MEA was then weighed and dissolved in 100 µl of PBS and transferred to the reaction mix containing the sdAb in a 5 ml Eppendorf tube. Following this step, EDTA was added to a final concentration of 5 mM, and the solution was gently pipetted up and down a few times to thoroughly mix the reagents. The volume was then topped up to 2.5 ml with PBS.

Once the mixture was ready, it was immediately incubated for 90 min at 37°C in a sealed tube. Subsequently, a PD-10 desalting column was equilibrated with 25 ml of 0.2 M NH_4OAc buffer at pH 6.0. After the sample incubation, buffer exchange was performed using 0.2 M NH_4OAc buffer (pH 6.0), on the PD-10 desalting column, following the gravity protocol. The reduced sdAb sample was applied to the PD-10 column, and elution from the packed bed was carried out with 3.5 ml of 0.2 M NH_4OAc buffer. The first 0.5 ml was discarded, and the remaining 3 ml was collected as 3x1 ml fractions in Eppendorf tubes.

To assess the protein concentration in each fraction, a NanoDrop spectrophotometer was used, and fractions with significant concentrations of 0.2 mg/ml and higher (usually the first 2 ml) were pooled. The collected eluate was concentrated at 4,000xg or 8,000xg for a

swing-out and 25° fixed angle rotors, respectively, for approximately 40 min at 8°C to a final volume of 0.5 ml using the Vivaspin 2 concentration tubes in a table top centrifuge. After the concentration step, the final concentration of the sample was determined, and the labelling protocol was continued accordingly.

3.6.16.2. Labelling and Purification

Each step in this section needed to be performed under minimal light exposure to prevent bleaching and prolong the lifetime of the dye.

A 20 mg/ml stock solution of the functional maleimide derivative of IRDye800CW was prepared in DMSO in an amber Eppendorf tube. The previously reduced sdAb was incubated with a 5-fold molar excess of pre-dissolved dye for a test run. This was calculated based on the concentration of reduced sdAb measured previously and was done in a 5 mM EDTA solution with a maximum total volume of 1.0 ml for 2 h at 37°C.

The Superdex 75 Increase 10/300 GL column was equilibrated with 2 column volumes of PBS buffer using the NGC chromatography system with a flow rate of 0.8 ml/min. Immediately after the conjugation reaction, all of the sample was injected over the column. Using PBS as running buffer, a method was run to detect the OD_{280} for sdAb and at 774 nm for IRDye800CW. Once the absorbance intensity of 80 mAU was reached for the 280 nm wavelength, 0.25 ml fractions were collected in amber tubes. All the fractions were collected until the absorbance at 280 nm dropped below 80 mAU. The fractions corresponding to a sharp peak eluting at around 17-18 min were selected. The concentration of the dye was calculated using the Lambert-Beer law based on the mean absorbance measured in triplicate at 774 nm, according to the formula at **Equation 1**:

$$C_d = \frac{A_{774}}{\varepsilon_d * \iota} \qquad (1)$$

where:

C_d is the molar concentration of the dye [M];

A_{774} is the measured absorbance at 774 nm;

ε_d is the molar attenuation coefficient of the dye, 240,000 in PBS for IRDye800CW [$M^{-1}\,cm^{-1}$]

ι is the optical path length (cm).

Equation 1. Fluorescent dye concentration calculation based on Lambert-Beer law.

The sdAb concentration was calculated following the same rule (**Equation 2**). The absorbance values obtained at 280 nm were adjusted using a correction factor of 3% for the absorbance at 774 nm.

$$C_{sdAb} = \frac{A_{280} - (c_f * A_{774})}{\varepsilon_{sdAb} * \iota} \quad (2)$$

where:

C_{sdAb} is the molar concentration of the sdAb [M];

A_{280}, A_{774} are the measured absorbances at 280 and 774 nm, respectively;

c_f is the correction factor;

ε_{sdAb} is the molar attenuation coefficient of the sdAb [M^{-1} cm^{-1}]

ι is the optical path length (cm).

Equation 2. sdAb concentration calculation based on Lambert-Beer law.

The degree of labelling (DOL) was estimated by calculating the dye to sdAb concentration ratio, according to the formula at **Equation 3**.

$$DOL = \frac{C_d}{C_{sdAb}} \quad (3)$$

Equation 3. Calculation of the degree of labelling estimation.

The collected product was stored at 4°C and protected from light for a maximum of two weeks to ensure stability and maintain the integrity of the conjugate. For longer-term storage, the conjugate was aliquoted and stored at -20°C, which helps to preserve its quality over an extended period.

3.6.16.3. Quality control

For QC of the sdAb-fluorophore conjugate prepared in previously, gel filtration, SDS-PAGE, and fluorescence spectroscopy were conducted.

3.6.16.3.1. Gel filtration for QC

If the same gel-filtration column utilized for conjugate purification is also used for QC, it is essential the column is thoroughly cleaned before the analysis. The column was flushed with the appropriate cleaning agents and followed by a blank run. To perform the blank run, 1 ml of the running buffer should be injected into the cleaned column, and the chromatogram should be monitored for any remaining contaminants.

At least 80 µg of the purified sdAb-conjugate was injected to the previously decontaminated Superdex 75 Increase 10/300 GL column, using PBS as the running buffer at a flow rate of 0.8 ml/min. To analyse the conjugate, detection of the absorbance at 280 nm

and 774 nm for sdAb and IRDye800CW, were performed respectively. The elution volume (retention time) of the conjugate was checked, and the chromatogram was thoroughly inspected for any suspicious peaks. However, no fractions were collected during this run.

3.6.16.3.2. SDS-PAGE for QC

The samples were prepared using 2 µg of the purified sdAb-conjugate in reducing conditions by calculating the volume of Laemmli buffer and reducing buffer corresponding to each sample for gel electrophoresis. Next, the sample mixtures were added to Eppendorf tubes and boiled for 4 min at 95°C on the heat block. Then, the samples and a molecular weight marker were loaded onto the SDS-PAGE gel. The gel was run in pre-made 1X TGS running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS at pH 8.3) at 80 V for 10 min initially to line up the samples, and then at 150 V until the dye front left the mini gel approximately for 70 min. The SDS-PAGE process was protected from light throughout.

Before proceeding with any staining, the gel was scanned in Odyssey (Licor) at 800 nm at low intensity. Subsequently, the gel was stained using the staining solution following the standard host lab protocol. Next, the de-staining procedure was performed, and then the migration distance of the stained proteins was compared with the Odyssey image to validate the accurate labelling and purification of the sdAb-conjugate.

3.6.17. Radioactive labelling

Radiolabelling of sdAbs involves the attachment of radioactive isotopes to the sdAbs for their further use in imaging applications.

3.6.17.1. Random coupling with DOTA

The following solutions were prepared a day before to make them metal-free for coupling experiments commencing with DOTA (**Table 82-87**).

Table 82. Preparation of metal-free 10X and 1X Carbonate buffer.

10X Carbonate buffer (0.5M), 50 ml

$NaHCO_3$, 2.016g	The powders were added and dH_2O added. Magnetic stirrer was used to homogenize the solution. pH was adjusted to 8.5-8.7.
Na_2CO_3, 0.106g	
NaCl, 4.383g	
dH_2O, 50 ml	
Chelex (2 g/L), 0.1 g	Chelex was added to the solution and kept overnight at RT. It was filtered with 0.22 µm filter and stored in the fridge.

1X Carbonate buffer (0.5M), 50 ml

10X Carbonate buffer, 20 ml **dH_2O, 180 ml**	The stock buffer was diluted with dH_2O.
Chelex (2 g/L), 0.4 g	Chelex was added to the solution and kept overnight at RT. It was filtered with 0.22 µm filter and stored in the fridge.

Table 83. Preparation of metal-free 1M NaOH.

1 M NaOH, 50 ml

NaOH, 2g **dH_2O, 50 ml**	The powders were added and dissolved in dH_2O.
Chelex (2 g/L), 0.1 g	Chelex was added to the solution and kept overnight at RT. It was filtered with 0.22 µm filter and stored in the fridge.

Table 84. Preparation of metal-free 1M HCl.

1 M HCl, 50 ml

30% HCl, 6 ml **dH_2O, 44 ml**	The liquid was dissolved in dH_2O.
Chelex (2 g/L), 0.1 g	Chelex was added to the solution and kept overnight at RT. It was filtered with 0.22 µm filter and stored in the fridge.

Table 85. Preparation of metal-free 0.2M Na_2CO_3.

0.2 M Na_2CO_3, (pH 12-14), 15 ml

Na_2CO_3, 0.3197g **dH_2O, 15 ml**	The powder was added and dissolved in dH_2O.
Chelex (2 g/L), 0.03 g	Chelex was added to the solution and kept overnight at RT. It was filtered with 0.22 µm filter and stored in the fridge.

Table 86. Preparation of metal-free 0.2M NH_4OAc, pH 5.

0.2 M NH_4OAc, pH 5, 500 ml

NH_4OAc, 7.708 g **dH_2O, 500 ml**	The powder was added and dissolved in dH_2O. The pH was adjusted to 5.
Chelex (2 g/L), 0.1 g	Chelex was added to the solution and kept overnight at RT. It was filtered with 0.22 µm filter and stored in the fridge.

Table 87. Preparation of metal-free 1M NH_4OAc pH 7.

0.1M NH_4OAc, pH 7, 500 ml

NH_4OAc, 3.854 g **dH_2O, 500 ml**	The powder was added and dissolved in dH_2O. The pH was adjusted to 7.
Chelex (2 g/L), 0.1 g	Chelex was added to the solution and kept overnight at RT. It was filtered with 0.22 µm filter and stored in the fridge.

Before the buffer exchange with PD-10 desalting columns, all buffers were transformed into metal-free formulations. Equilibration of the columns was carried out with 20 ml of 1X sodium carbonate buffer (metal-free) for each column. During the sample application step, 2.5 ml of the buffer was poured down the column, and the flow-through was discarded to prevent cross-contamination. To collect the fractions, ten Eppendorf tubes were placed under each column.

For subsequent purification and analysis of the test sdAb, 3.5 ml of sodium carbonate buffer (pH 8-8.5) was added to each column, and the resulting eluate was collected in fractions of 0.5 ml. NanoDrop spectrophotometer was used to measure the different fractions. Fractions containing the targeted sdAbs were pooled for further experiments.

Next, DOTA was incorporated into the samples. 10 mg of DOTA was dissolved in 2 ml of 1X sodium carbonate buffer. The samples were vortexed to completely dissolve the chelator. For each sample type, sdAb or mAb, 20 or 140-fold molar excesses of DOTA were added, respectively, to maintain exact stoichiometry.

The pH was measured after the addition of DOTA, to adjust the pH to 8-8.5 using 0.2 M sodium carbonate buffer (pH 14). The samples were incubted at RT for 2 h or overnight. The the reactions were terminated by lowering the pH to 7 using metal-free 1M HCl. The concentrations of the samples were measured using NanoDrop spectrophotometer.

A Vivaspin tube (0.5 ml) was used to effectively concentrate the samples. For high-performance liquid chromatography (HPLC) analysis 0.1M NH_4OAc (pH 7) buffer and Superdex75 10/300 column were used. Labelling yields were verified by doing further assessments. The MW of DOTA was corrected by accounting for its HCl content. The MW of Ab was further adjusted with a factor of 1.5 DOTA. The extinction coefficient of sdAb was also adjusted with a factor of 1.5 DOTAs. The concentrations were adjusted to reflect the actual number of DOTA moieties in the samples.

3.6.17.2. ^{111}In labelling of Bevacizumab

First, a preliminary trial was conducted with a test nanobody to test its effective interaction with ^{111}In. Afterwards, coupling of Bevacizumab was commenced. Specific amounts of ^{111}In -Ab (1.70 mBq (megabecquerel) in 29 µl), and ^{111}In -Nb (1.23 mBq in 21 µl) were used for this test labelling. To determine the necessary quantity of Ab for 21 µl of ^{111}In, the reference paper (Khorami-Moghadam et al. 2013) recommended a combination of 700 µl of NH_4OAc buffer and 500 µl of Bevacizumab (2.11 mg/ml). Based on this information, 24 µl of Bevacizumab (50 mg) and 35.6 µl of buffer (0.5 M NaOAc, pH 5, metal-free) were mixed with 21 µl of ^{111}In, resulting in a final volume of 80.6 µl.

The labelling process involved thawing DOTA-conjugated Ab on ice that was taken from -20°C, while simultaneously pre-warming the heater to 40°C. The Bevacizumab+DOTA and ^{111}In samples were pre-heated in separate tubes for 5 min and subsequently combined in a single Eppendorf tube, followed by a 90-min incubation at 40°C. The incorporation efficiency of the samples was evaluated via Instant Thin Layer Chromatography (iTLC).

3.6.17.3. Instant Thin Layer Chromatography

During the incubation step, iTLC strips were prepared, and a line (approximately 1 cm) was drawn at the bottom and a dot was put at the top of the strip with a pencil. Subsequently, 5 µl of sample solution was dropped on the line, and an additional 2 µl (pH 4.5 – 4.7) was added. To aid in sample mobility, 0.1 M sodium citrate, functioning as a chelator for collecting free indium, was also added. The iTLC strips were then carefully dipped into a buffer-filled glass beaker. The level of the buffer in the beaker should not pass the line level on the strip. The samples were allowed to run for 5 min before analysing them in iTLC reader. iTLC is used for separating and quantifying radioactive compounds whether they are successfully incorporated into the intended sample or they remain in the solution freely.

3.6.17.4. ^{99m}Tc-labelling of sdAbs

The half-life of ^{99m}Tc is 6 h. Everything was prepared within the safety cabinet designated for the particular radioactive element. Before starting, the activity of ^{99m}Tc was quickly measured using the dose calibrator.

Synthesis of $[^{99m}Tc(CO)_3(H_2O)_3]^+$:

The radiopharmaceutical was prepared using a commercial kit (IsoLink) containing the following lyophilized formulation (**Table 88**):

Table 88. IsoLink commercial kit contents for ^{99m}Tc-labelling.

Reagent	Formula	Mass (mg)
Sodium tartrate	$C_4H_4Na_2O_6$	8.50
Sodium tetraborate	$Na_2[B_4O_5(OH)_4]\cdot 8H_2O$	2.85
Sodium carbonate	Na_2CO_3	7.15
Sodium boranocarbonate	$CH_3BNa_2O_2$	4.50

1.5 ml of fresh $^{99m}TcO_4$-eluate from a ^{99}Mo/^{99m}Tc generator containing approximately 740 MBq – 3.7 GBq was added to the commercial kit vial. The container was placed in a secure submerging container before immersing in water. Next, the kit was placed in a boiling water bath for 20 min, meanwhile remaining the vial within the lead container throughout the process. After cooling down to RT, 200 µl of 1N HCl was added to neutralize the solution and eliminate any residual boranocarbonate, by setting pH to 6.5. Before adding HCl, the container was unscrewed, and a hole was made on the vial lid using a syringe needle allowing the escape of air pressure accumulated during the heating up process of the reaction. The needle was kept on the vial until transferring the contents into the heat block. Subsequently, the mixture was allowed for 5-10 min to react, and the pH was measured using pH paper. For certain experiments that did not require a high degree of radioactivity, sterile NaCl was used to increase the total volume if a higher injection volume is needed.

In the first step, 50 µl of the 1 mg/ml test nanobody was taken, followed by the addition of 500 µl of the ^{99m}Tc-tricarbonyl solution from the commercial kit. The mixture was then heated to 50°C for 90 min on a heat block placed within a safety cabinet. During this time, preparations were made for the subsequent steps. Three tubes were set up for each sample, labelled as PBS, Elution, and Filtered elution. After 60 min, a GE Naps column was prepared for each sample and placed over the sink, and 10 ml of PBS with 0.1% Tween was used to rinse them out using a syringe. Subsequently, 500 µl of labelled nanobody was added to the column, and after everything passed through, 1 ml of PBS with 0.1% Tween was added to the column, and the eluate was collected in a new tube. The radioactivity of elutes and the radioactivity level of the column were measured in the gamma counters. To remove the aggregates, the purified nanobody were transferred using a 1 ml syringe without a needle, and a Millex filter (0.2 µm) was placed on the syringe, pressing the labelled nanobody through and collecting the filtrate in a new tube. The radioactivity of the filtrate and the filter were measured in the gamma counter.

Meanwhile, paper strips were prepared iTLC QC of each sample, by drawing lines at the top of the strips to observe the movement of the liquid until it reached the top. About 5 ml of 100% acetone was added to a glass vial for migration of the samples. For QC of the labelling yield, a drop of the labelled nanobody (approximately 2-5 µl) was placed on the bottom, unlabelled end of an iTLC paper strip within the safety cabinet. The strip was then placed in a glass vial containing 100% acetone outside of the safety cabinet, and the movement of the liquid was monitored until it almost reached the top, indicated by the dissolving of the previously drawn ink line. Afterward, the strip was removed from the glass vial and placed in an iTLC reader to calculate the percentage of labelled vs. unlabelled sdAbs. The labelling yield was once again controlled using iTLC by placing a drop of labelled nanobody (approximately 2 µl) on the bottom unlabelled end of an iTLC paper strip, and the procedure was resumed as explained above. Afterward, the strip was removed from the glass vial and placed in the gamma counter, and the software calculated the percentage of labelled vs. unlabelled nanobody. Following these steps, the unbound ^{99m}Tc-tricarbonyl in the equilibrated Sephadex G25 disposable column was removed with 10 ml of PBS with 0.1% Tween.

3.6.18. In silico approaches for determination of differences between antibodies and nanobodies

During the redesign phase at the host lab, collaborative brainstorming sessions with Prof. Dr. Nick Devoogdt at VUB yielded valuable insights aimed at enhancing sdAb design to reflect the properties of naturally occurring nanobodies. The distinctions between the V_Hs of Abs and V_HHs of nanobodies were highlighted using *in silico* approaches with multiple sequence alignment and an array of structural analyses. The sequential and/or structural analyses included paratope length and binding predictions, motif analysis, disulphide bridge examination, salt bridge profiling, and logistic regression analysis combined with the Boruta–Random Forest machine learning algorithm. The resulting data focused on conserved FWs, uncovering the structural differences important for the development of V_H sdAbs with minimal immunogenicity, while preserving their function and specificity.

3.6.18.1. Datasets

The V_H sequences of 5030 mammalian IgGs (*Cricetulus migratorius*; chimeric *Homo sapiens/Mus musculus*, chimeric *Mus musculus/Homo sapiens*; *Homo sapiens*; *Mus musculus*; *Pan troglodytes*; *Rattus norvegicus*) were obtained from the abYsis database (Swindells et al.

2017). The V_H sequences of the full-length IgGs were isolated by eliminating non-V_H domains from the provided sequences, ensuring a focused analysis of the variable heavy chains within the mammalian IgGs. Additionally, V_HH sequences of 1260 nanobodies (*Camelidae*, *Camelus bactrianus*, *Camelus dromedaries*, *Lama glama*, *Vicugna pacos*) were obtained from the Single Domain Antibody Database (Wilton et al. 2018).

3.6.187.2. Multiple sequence alignment

The FASTA sequences were aligned according to the Chothia numbering scheme (**Table 89**). The HCDRs and FW regions were identified using a local installation of the Antigen Receptor Numbering And Receptor Classification "ANARCI" tool (Dunbar et al. 2016).

Table 89. Definition of the constant and variable regions in V_H and V_HH sequences based on the Chothia numbering scheme

VH/ V_HH Region	Chothia Residue
FW1	1-25
HCDR1	26-32
FW2	33-51
HCDR2	52-56
FW3	57-94
HCDR3	95-102
FW4	103-113

3.6.18.3. Paratope binding prediction

The paratope binding regions and corresponding binding scores, including the two flanking residues of HCDRs, were predicted with the "Parapred: Antibody Paratope Prediction using Convolutional and Recurrent Neural Networks" tool (Liberis et al. 2018), which is directly compatible with the output of the ANARCI tool.

3.6.18.4. Sequence conservation motif analysis

To identify amino acids conserved at specific positions, amino acid motifs within the V_H and V_HH sequences were visualized with WebLogo (version 2.8.2), an online application designed for the visualization of protein motifs (Crooks et al. 2004), with the amino acid frequency scoring method.

3.6.18.5. Disulphide bond analysis

To investigate V_H and V_HH sequences featuring double or multiples of two cysteine residues, previously aligned sequences were to identify and characterize canonical and non-canonical disulphide bridges. These bridges were confirmed both sequentially and structurally, using the PyMOL software (version 2.5.5) (Deac, VeliČković, and Sormanni 2019). As numerous V_HHs lacked PDB IDs, the NanobodyBuilder2 (Abanades et al. 2023) online tool was used, facilitating the homology modelling of V_HHs with known sequence information and the generation of their three-dimensional structures.

3.6.18.6. Salt bridge analysis

Afterwards, the salt bridges within the Chothia-numbered aligned sequences were identified with a targeted filtering approach. V_H and V_HH sequences which contained both arginine (R), lysine (K), or histidine (H) residues at the 94th position, and an aspartic acid (D) or glutamic acid (E) residues at the 101st position were considered to have a salt bridge.

3.6.18.7. Analysis of physicochemical properties

EMBOSS Pepstats (Madeira et al. 2022) was used to assess the physicochemical characteristics for both the HCDRs and the FW regions of the sequences. The output data contained information regarding the amino acid composition, hydrophobicity, charge, and other physicochemical properties as outlined at **Table 90** and **Table 91** within the specified regions.

Table 90. Classification of physicochemical properties shown as single letter amino acid codes defined by EMBOSS Pepstats.

Property	Amino Acid Residues
Tiny	A+C+G+S+T
Small	A+B+C+D+G+N+P+S+T+V
Aliphatic	A+I+L+V
Aromatic	F+H+W+Y
Non-Polar	A+C+F+G+I+L+M+P+V+W+Y
Polar	D+E+H+K+N+Q+R+S+T+Z
Charged	B+D+E+H+K+R+Z
Basic	H+K+R
Acidic	B+D+E+Z

Table 91. Amino acid hydropathy classes in single letter amino acid codes.

Property	Amino Acid Residues
Hydrophilic	D+E+K+N+R+Q
Neutral	G+H+P+T+S+Y
Hydrophobic	A+C+F+I+L+M+V+W

3.6.18.8. Statistical analysis

A logistic regression modelling approach coupled with the Boruta algorithm was employed for feature selection based on machine learning to evaluate the significance of the identified physicochemical variables. The statistical significance of physicochemical variables (predictors) was measured, according to their ability to predict the classification of V_Hs (0) and V_HHs (1). Simultaneously, the Boruta algorithm, a supervised machine learning method chosen for feature selection in classification models, was applied.

3.6.18.8.1. The binary logistic regression model

The Stepwise Regression Backward Elimination technique was used with the Likelihood Ratio test. This technique eliminated non-significant predictors from the model in a stepwise manner, enhancing its precision in predicting the classification of V_Hs and V_HHs. The Binary Logistic Regression model, represented by **Equation 4**, explains the relationship between the predictors and the probability of the V_H (0) or V_HH (1) classification.

$$ln(\frac{p}{1-p}) = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + \cdots + \beta_k.X_k \qquad (4)$$

Equation 4. The Binary Logistic Regression model.

In the equation (p:V_HH) denotes the probability of V_HH classification, $X_1, X_2, \ldots, X_n$ represent the physicochemical properties as predictors, and $\beta_0, \beta_1, \ldots, \beta_n$ are the corresponding coefficients determined through the Stepwise Regression Backward Elimination process. The model output provides insights into the relative influence of each physicochemical variable on the likelihood of V_HH classification.

In adherence to the assumptions of independence and linearity within the binary logistic regression model, particular attention was devoted to addressing multicollinearity among the predictors. To mitigate the impact of multicollinearity, the Variance Inflation Factor (VIF) was computed for each predictor in the model. The VIF quantified the extent to which the variance of an estimated regression coefficient is increased due to collinearity with

other predictors. Consequently, predictors with high VIF values (>10) were systematically removed from the model to enhance the precision and accuracy of the model.

3.6.18.8.2. The Boruta algorithm

The Boruta algorithm, a wrapper built around the Random Forest (RF) classification algorithm, was used to identify key features within the dataset concerning amino acids for all regions of V_Hs and V_HHs. Operating it as learning method, Boruta aimed to capture all relevant and impactful features with respect to the outcome variable, in this case, the classification of V_Hs and V_HHs. During the fitting process, the algorithm collectively considered amino acid data for all regions. The responses were systematically ordered based on their importance for classification, where the importance measure defined as the Z-score of the mean decrease in accuracy.

3.6.18.8.3. Statistical significance calculation

McFadden's R^2, also known as McFadden's pseudo-R^2, serves as a metric for evaluating the fit in binary logistic regression models. It assesses the effectiveness of the model in predicting the outcome, offering insights into the variance of the predictors.

The Area Under the Curve (AUC) is a widely used metric that provides a summary of the diagnostic accuracy of a binary logistic regression model. It is particularly useful in assessing the ability of the model to correctly classify antibody sequences as V_Hs and V_HHs. An AUC>0.5 signifies that the model outperforms random chance in discriminating the difference between V_Hs and V_HHs, meanwhile validating its efficacy in classification.

3.6.18.9. Tools and plots

All Unix-compatible tools were executed on a Windows operating system, facilitated through access within a Docker Container. Length variation bar plots and heatmaps depicting antigen binding prediction scores were generated in the R statistical computing environment (4.0.0). The ggplot2 graphics package (3.4.2) was used for the creation of length variation bar plots, while the pheatmap package (1.10.12) generated the antigen binding prediction score heatmaps. The outcomes of the Boruta algorithm were visualized using the ROCR R package.

3.7. Research plan and timeline

The following research plans provided a roadmap for accomplishing the study goals and foreseen timeline for both home lab and the host lab studies (**Table 92 and 93**).

Table 92. The work packages executed in accordance with the timeline of the thesis progression in the home lab.

Work package no	Work packages and objectives	Time (months)
1	Plasmid design of sdAb variants	1-18
1.1	Ordering WT sdAb plasmid and purchasing consumables	1-4
1.2	Design of sdAb mutants by site-targeted-mutagenesis	4-18
2	Production and purification of sdAb variants in bacterial system	5-18
3	Characterization of sdAb variants	12-30
3.1	Biophysical characterization of sdAb variants	12-26
3.2	Binding characterization of sdAb variants to VEGF	12-30
4	Testing sdAb variants *in vivo* in zebrafish model	12-30

Work Package-1: We pursued the design of antibody fragments, including FLAG and His tags, specifically bind to VEGF and possess favourable biophysical properties. These fragments were formatted in the sdAb structure using protein engineering techniques. The V_H sequences of Bevacizumab (WT sdAb) were cloned into the pET-17b plasmid, serving as a foundation for designing the proceeding antibody variants through SDM. Computational tools were employed to estimate regions prone to aggregation, which guided the selection of mutation sites. The identified regions, distinct from the HCDRs, were subjected to single or multiple mutations, substituting hydrophobic residues with hydrophilic and charged amino acids to increase overall solubility. The sequences of these engineered antibody variants were confirmed via Sanger sequencing and confirmed for bacterial production.

Work Package-2*:* In this work package, every designed antibody fragment within the plasmid was expressed as recombinant proteins using a bacterial host. To facilitate characterization studies, the purification of the expressed antibody variants was isolated from the bacterial culture. Affinity purification, utilizing the His tag, was employed for this purpose. Subsequently, SEC would be employed to further refine the purification process, ensuring the removal of any remaining impurities and obtaining highly purified antibody variants suitable for downstream characterization analyses.

Work Package-3: The obtained antibody variants, along with the commercial Bevacizumab, would be subjected to comparative characterization studies in terms of functionality. ELISA and/or Surface Plasmon Resonance (SPR) methods would be employed, to determine the binding affinities and kinetics. *In vitro* solubility tests would be conducted on the antibody variants to evaluate their propensity for aggregation meanwhile elucidating their stability under physiological conditions. Solubility tests of antibody variants would be tested by *in vitro* methods. The thermal stability of the antibody variants would be calculated using the rtPCR, following the methodology outlined previously (Huynh and Partch 2015).

Work Package-4: The visualizing and quantifying the anti-angiogenesis process in mouse models poses significant challenges. To overcome this limitation, recent research has employed transgenic zebrafish models to visualize and quantify the effects of Bevacizumab on angiogenesis (Zhang, Gao, et al. 2018). We aimed to apply both the obtained antibody variants and Bevacizumab to the transgenic zebrafish model, enabling a comparative observation of their effects on angiogenesis along with Bevacizumab.

Table 93. The work packages executed in accordance with the timeline of the project progression in the host lab.

Work package no	Work packages and objectives	Time (months)
0	Preparation of functional camelised anti-VEGF sdAbs at home laboratory	0
1	Ethical committee application	0-1
2	Orientation to the host lab	0-1
2.1	Safety training for radioactivity studies	0-1
2.2	Training for application of common procedures	0-1
3	Initiation of animal studies	2-4
3.1	Cell culture studies for mice tumour inoculation	2-4
3.2	Breeding mice for pharmacokinetic analysis of anti-VEGF sdAbs	2-4
4	Re-design of anti-VEGF sdAbs	3-4
4.1	Cloning of redesigned sdAbs	3-4

4.2	Expression and purification of redesigned sdAbs	4-5
5	Intravital imaging to assess tumour targeting at the microscopic level	5-6
6	Whole body biodistribution and tumour targeting via fluorescent imaging at the macroscopic level	5-6

Work Package-0: Our camelised recombinant anti-VEGF sdAbs were predesigned and characterized *in vitro* biophysical/biochemical tests (affinity, purity, aggregation, thermal stability) at the home lab in IBG. Under auto-induction conditions, sufficient anti-VEGF sdAbs was produced by periplasmic expression in *E. coli* cultures and purification was done by affinity chromatography. Already produced and purified sdAbs as well as Bevacizumab were sent from Turkey to host lab in Belgium through cold chain cargo. To show the low interaction of mouse VEGF with our sdAbs, an ELISA was conducted before pursuing animal studies.

Work Package-1: The application to the ethical committee for *in vivo* mouse experiments was conducted subsequent to determining the type of subcutaneous tumour and the required number of mice for each group of sdAb intended for the planned experiments (**Figure 25**, **Table 80** and **81**). Bevacizumab, which was commercially purchased, would be included in the planned studies for comparison.

Work Package-2: Upon joining the host lab, all researchers must participate to orientation program. It aims familiarization with the lab facilities, safety protocols, research goals, and the team members. The most important part is safety training when working with radioactive chemicals was explained, especially radiation safety protocols, radioactive material handling, radioactivity contamination prevention and decontamination procedures and emergency procedures.

Work Package-3: Once the home lab sdAbs have been obtained, the cell culture studies for mice tumour inoculation would be commenced as well. The HT-29 tumour cells were cultured in appropriate cell culture media, maintaining optimal conditions for their growth and viability. The cells passaged and expanded for tumour inoculation. The cultured cells then were harvested, and their viability was verified to ensure their successful engraftment in mice prior to inoculation process. The requested mice were reserved upon successful ECD application. For the pharmacokinetic analysis of anti-VEGF sdAbs, the mice were bred in-

house, and selected based on their genetic background, health status, and suitability for the specific pharmacokinetic analysis. The breeding process involved controlled mating of specific strains to ensure the required genetic traits are present in the offspring. During the breeding process, the animal facility would adhere to ethical guidelines and regulations to ensure the well-being and humane treatment of the mice. Once the mice reach the appropriate age and are ready for the pharmacokinetic analysis, they would be included in the study. The mice were acclimatized to the laboratory environment before the start of the studies.

Work Package-4: A new design strategy was pursued for the sdAbs due to insufficient quantity and suboptimal purity of the sdAbs. The plasmids of the home lab sdAb designs would be utilized to create new Bevacizumab-derived sdAb variants based on a traditional nanobody template. The redesign of the sdAbs would be carried out in collaboration with Prof. Dr. Nick Devoogdt at VUB. The cloning, transformation, and production studies would be conducted at VUB, Ixelles campus. The expression vector was changed to pHen6, and the expression cell line strain was also changed to *E. coli* WK6 competent cells for recombinant protein production. Moreover, the novel auto-induction approach was replaced with the traditional IPTG induction.

The requested mice were reserved upon successful ECD application. Upon achieving successful production of the new sdAb designs, the animal studies would proceed as outlined in the original project proposal and in accordance with the ethical committee application (**Figure 26**).

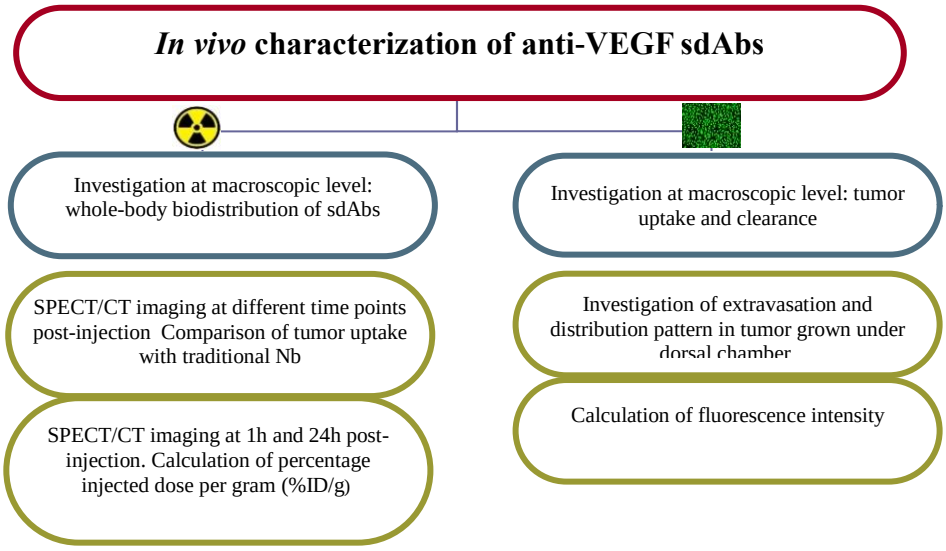


Figure 26. Flowchart for work packages 5 and 6 – "in vivo characterization of anti-VEGF sdAbs".

Work Package-5: This part of the study aimed to determine if the Bevacizumab-derived sdAbs demonstrated significantly higher tumour uptake than a non-targeting traditional nanobody by studying their whole-body bio-distribution and tumour targeting potential using non-invasive nuclear imaging in mice with subcutaneous tumours. For tumour inoculation, the mice were gently anesthetized using isoflurane via inhalation. Subsequently, HT-29 tumour cells were subcutaneously injected into the right flank of each animal. The radiolabeled sdAbs, nanobody and Bevacizumab would be intravenously injected in immunodeficient mice bearing a subcutaneous high VEGF-expression. Prior to imaging, the mice would be briefly anesthetized with isoflurane via inhalation. At different time points post-injection, the mice would be subjected to *in vivo* SPECT/CT imaging. In addition, organs and tissues of interest would be resected, weighed and counted for radioactivity against a standard of known activity to accurately calculate the tracer uptake expressed as percentage of injected dose per gram (%ID/g). The data were compared with the bio-distribution of the non-camelised WT sdAb and a non-targeting traditional nanobody, serving as a negative control for tumour uptake and a reference for general bio-distribution profile.

Work Package-6: Based on the results of the previous workpackage, the tumour targeting capacity using non-invasive nuclear imaging in a subcutaneous mouse model at microscopic level of the lead compound, fluorescently labeled, were investigated and compared to Bevacizumab through microscopic analysis of tumour sections. HT-29 tumour cells would be subcutaneously injected into the right flank of each animal. Tumour cells would be cultured under a dorsal window on athymic nude mice. Once the tumours reach the desired size, the mice would be positioned under a fluorescent stereomicroscope. The mice would be briefly anesthetized with isoflurane through inhalation. Subsequently, they would be intravenously injected with fluorescently labeled (IRDye800CW and Alexa Fluor-647) sdAbs or Bevacizumab via the tail vein. Extravasation and distribution of the fluorescent constructs from the vasculature into the tumour tissue, would be assessed continuously at multiple time points over 72 h. For image analysis, fluorescence intensity of a fixed region of interest within the tumour would be used to estimate rate of uptake and clearance of tracers from the tumour over time. Co-localization of the sdAbs and VEGF-expression would be confirmed *ex vivo* via confocal microscopy.

3.8. Data analysis

In the methods section, specific details regarding the data analysis for each experiment were outlined whenever it was applicable. Overall, in this thesis, the data analysis may involve statistical methods, software or online tools tailored to the specific requirements of the experiment design.

3.9. Limitations of research

The main limitation of the research is the use of a zebrafish developmental model to test our designed sdAbs, which are modified from a chemotherapeutic antibody drug indicated for advanced cancers in humans. Due to the lack of recapitulation of the complex tumour microenvironment, the effects of our sdAbs on zebrafish development might not translate to their human theranostic use.

During our collaboration with the host lab, we planned to test our sdAbs on a more appropriate mouse cancer model for more accurate characterisation of their pharmacokinetic properties; however, due to logistics and optimisation concerns during the handling of our plasmid samples and the production of sdAbs, the utilisation of the mice model could not be successfully completed. The reserved animals for the experiments were already bred and cared for and the cell culture studies were also initiated, in addition to the completed precursor test studies with a control nanobody and Bevacizumab. The allocated nude mice were all left without benefit, which eventually impacted the planned research outcomes the collaboration studies.

3.10. Ethical committee approval

3.10.1. Ethical committee approval from Animal Experiments Local Ethics Committee at IBG

Approval date: 11.03.2020

Document number: 2020-014

3.10.2. Ethical committee approval from Ethische Commissie Dierproeven at VUB

Approval date: 30.08.2021

Document number: 21-272-19

4. RESULTS

4.1. *In silico* approaches for prediction of aggregation prone residues

One of the key challenges in the engineering of therapeutic mAbs is amino acid residues that are prone aggregation. These can affect the functionality when converting mAb into sdAb format. To address this issue, *in silico* computational tools were used to predict aggregation-prone residues within the V_H of Bevacizumab. The outcome of the each tool was individually explained in the following tables (**Table 94-100**) and figures (**Figure 27 and 28**).

PDBePISA results (**Table 94, Figure 28**) are showing the interface resdiues between H and L chains. The exposed residues are aggregation prone once V_L is removed from V_H.

Table 94. The list of V_H-V_L interface residues of Bevacizumab determined by PDBePISA. First column shows the location and the name of the V_H residue. BSA: buried surface area % (Å²), the location and the name of the V_L residue, the distance between the residues of V_H and V_L (Å) and the type of interaction, respectively. The lowest scores of BSA were eliminated as they are not significantly contributing the interfacing residues.

List of interactions between V_H and V_L of Bevacizumab

Residue (V_H)	% BSA (Å²)	Residue (V_L)	Distance (Å)	Interaction
V 37	92	F 98	4.5	Interface
Q 39	58	Y 87, Q 38	3.3, 3.0	H-bond
G 44	44	Y 87	-	Interface
L 45	86	F 98	3.9-4.3	Interface
W 47	92	V 94, P 95, W 96	3.9, 3.7, 3.7	Interface
A 61	75	P 95	4.9	Interface
Y 95	75	K 42, Q 38	3.8, 3.7	H-bond
H 107	82	Y 91, F 50	3.2, 2.6	H-bond, Interface
W 108	41	N 34, Q 89, Y 91	3.1, 3.4, 3.5	H-bond, Interface
Y 109	88	W 96	2.9	Interface
F 110	99	Y 36, Q 89	3.1, 2.7	H-bond
W 113	71	P 44	4.1	Interface
G 114	64	A 43	3.8	Interface

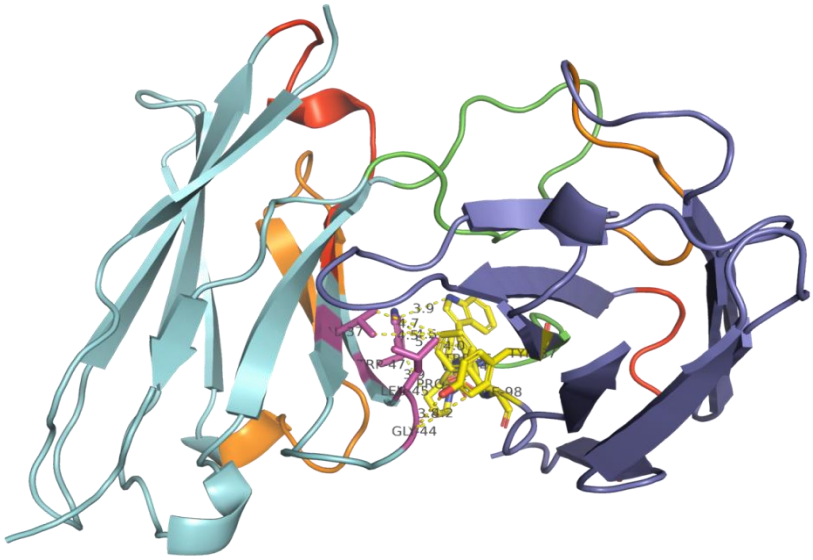
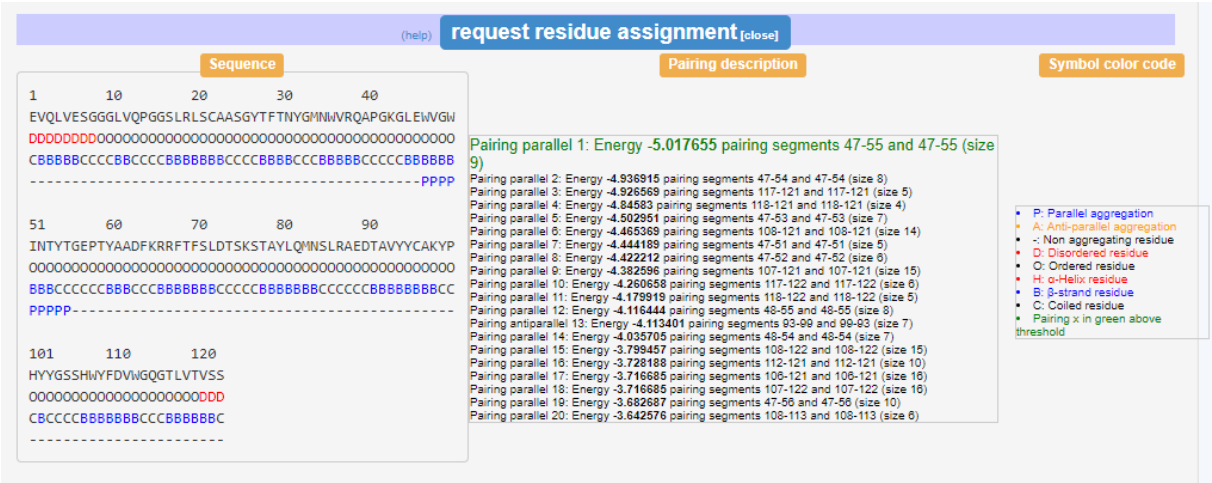


Figure 27. The interfacing residues of Bevacizumab at the V_H-V_L fragments. The 3D structure of Bevacizumab F_v was retrieved from the PDB repository using ID: 1BJ1. Only the distance between the V_H FW-2 interface residues and corresponding V_L interaction partners was measured. The selection of interface residues was based on the analysis results of PDBePISA (**Table 94**). The V_H chains are shown in light blue, while V_L chains are shown in dark blue. CDR1 is shown in red, CDR2 is shown in orange, and CDR3 is shown in green. The interface residues were depicted in stick representation, with V_H interface shown in purple, and V_L interface in yellow). The interaction partners at V_H to V_L respectively are paired as follows: V37-F98, G44-Y87, L45-F98, W47-V94, W47-P95, W47-W96.

Pasta score (**Figure 28**) shows the probability of the sequence to form aggregates. The following regions are potential hotspots for aggregation.

Figure 28. The predicted aggregation propensity score of V_H sequence of Bevacizumab by Pasta 2.0.



Tango scores (**Table 95**) indicate the propensity of the sequence to form aggregates. The residues with higher aggregation score define the potential sites for aggregation.

Table 95. The predicted aggregation scores for the V_H sequence of Bevacizumab by Tango 2.2.

Position	Amino acid	Beta sheet	Beta turn	Alpha helix	Beta aggregation	Helix aggregation
27	Y	11.5	0.1	0	0.176	0
28	T	28.8	0	0	0.176	0
29	F	30.1	0.1	0	0.176	0
30	T	20.3	0.1	0	0.176	0
53	T	23.2	0	0	2.725	0.004
54	Y	22.7	0.1	0	2.576	0.004
55	T	20.1	0.1	0	0.49	0.001
68	F	11.7	0	0.869	4.285	0.006
69	T	37.6	0	0	4.285	0.006
70	F	38.5	0	0	4.285	0.006
71	S	31.4	0.4	0.135	4.285	0.006
119	V	18.7	0	0	14.878	0.023
120	T	32.7	0	0	14.878	0.023
121	V	33.8	0	0	14.754	0.023
122	S	18.5	0	0	6.506	0.01

Waltz scores (**Figure 29**) indicates the aggregation propensity of the Bevacizumab V_H sequence.

WaltzJob_1548155547

Positions	Sequence	Average score per residue
46-54	WVGWINTYT	96.8784836666667
90-94	TAVYY	98.662207
96-122	AKYPHYYGSSHWYFDVWGQGTLVTVSS	96.6555183333333

Figure 29. An evaluation of the overall aggregation tendency for the Bevacizumab V_H sequence based on Waltz 2.0 predictions. The amino acid positions and corresponding sequence were given along with average score per residue.

Based on Aggrescan scores (**Table 96**), specific regions listed below within the Bevacizumab V_H sequence are identified as potential hotspots for aggregation.

Table 96. The results of the Aggrescan 3D 2.0.

Chain	Position	Residue	Score
H	5	V	0.5407
H	10	G	0.6266
H	11	L	1.1902
H	45	L	0.4943
H	47	W	0.3377
H	54	Y	0.6364
H	93	V	0.6181
H	95	Y	0.3963
H	102	Y	1.1891
H	103	Y	1.3053
H	104	G	0.2621
H	106	S	0.0163
H	107	H	0.0761
H	108	W	1.3627
H	109	Y	1.4932
H	110	F	0.888
H	112	V	0.1109
H	113	W	0.2129
H	117	T	0.2276
H	118	L	1.0923
H	120	T	0.1398

Table 97. The results of CamSol Intrinsic v2.2 predictions on aggregation propensity. The negative score indicates the aggregation prone residues.

Position	Amino acid	Amino acid intrinsic solubility
2	V	-1.07646
26	G	-1.113125
27	Y	-1.192652
29	F	-1.447556
30	T	-1.447556

34	M	-1.236968
47	W	-1.171758
48	V	-1.747311
49	G	-1.228806
50	W	-1.925576
51	I	-1.809743
52	N	-1.40763
53	T	-1.396483
69	T	-1.078086
70	F	-1.255598
71	S	-1.926146
72	L	-1.214331
80	Y	-1.08335
93	V	-1.563592
94	Y	-2.167094
95	Y	-1.542698
96	C	-1.873579
97	A	-1.091451
105	S	-1.205558
106	S	-1.193173
107	H	-1.328667
109	Y	-1.548865
110	F	-2.197917
112	V	-1.421969
115	Q	-1.272964
116	G	-1.28956
118	L	-1.555253
119	V	-1.504119
120	T	-1.54591
121	V	-1.632937
122	S	-1.501584

Table 98. CamSol Structurally Corrected predictions on aggregation prone residues. Structurally corrected score: confidence level of the prediction. Intrinsic score: the predicted aggregation propensity of the given residue. Solvent exposure: The scores greater than 1 indicates high solubility, where the negative scores show the probability of aggregation. Lower values point out aggregation prone residues. Contacts: the corresponding interaction residues associated with aggregation risk.

Residue	Structurally corrected score	Intrinsic score	Solvent exposure	Contacts
Y102	-1.095170449	-0.916398024	0.534379062	105, 106, 107
S105	-1.008905511	-1.205558271	0.510030355	100
H107	-1.257945872	-1.328666708	0.548676179	98, 99, 100, 101, 102
Y109	-1.330055779	-1.548864505	0.521336437	96, 97, 98, 99
D111	-1.10626003	-2.029795232	0.348069802	96, 97, 98, 99, 100
Q115	-1.242438129	-1.272963765	0.53399121	2, 3, 4, 5, 6, 21, 92, 93, 94, 95, 96

Table 99. The aggregation-prone residues in Bevacizumab V_H sequence based on Aggrescan. A^4V: average residues on total sequence length, HSA: Hot spots average, NHSA: number of HSA, a^4vAHS: a4v average in a given hot spot.

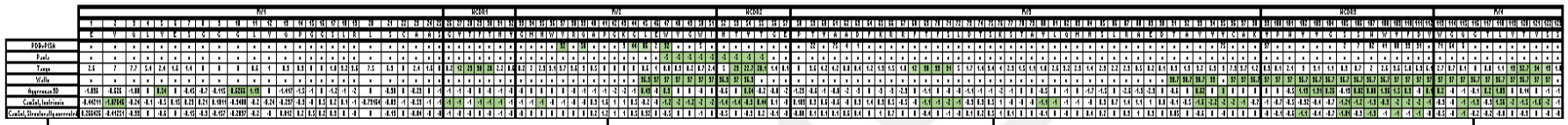
Residue	Amino acid	A^4V	HSA	NHSA	A^4VAHS
18	L	-0.02	2.634	0.176	0.156
19	R	0.143	2.634	0.176	0.156
20	L	0.214	2.634	0.176	0.156
21	S	0.251	2.634	0.176	0.156
22	C	0.012	2.634	0.176	0.156
23	A	0.113	2.634	0.176	0.156
24	A	0.081	2.634	0.176	0.156
25	S	0.1	2.634	0.176	0.156
26	G	0.265	2.634	0.176	0.156
27	Y	0.247	2.634	0.176	0.156
28	T	0.066	2.634	0.176	0.156
29	F	0.274	2.634	0.176	0.156

30	T	0.274	2.634	0.176	0.156
31	N	0.238	2.634	0.176	0.156
32	Y	0.075	2.634	0.176	0.156
45	L	0.085	3.289	0.365	0.345
46	E	0.085	3.289	0.365	0.345
47	W	0.367	3.289	0.365	0.345
48	V	0.703	3.289	0.365	0.345
49	G	0.32	3.289	0.365	0.345
50	W	0.499	3.289	0.365	0.345
51	I	0.517	3.289	0.365	0.345
52	N	0.266	3.289	0.365	0.345
53	T	0.266	3.289	0.365	0.345
78	T	0.118	0.942	0.157	0.137
79	A	-0.016	0.942	0.157	0.137
80	Y	0.247	0.942	0.157	0.137
81	L	0.103	0.942	0.157	0.137
82	Q	0.084	0.942	0.157	0.137
83	M	0.286	0.942	0.157	0.137
92	A	0.067	2.934	0.367	0.347
93	V	0.355	2.934	0.367	0.347
94	Y	0.612	2.934	0.367	0.347
95	Y	0.502	2.934	0.367	0.347
96	C	0.673	2.934	0.367	0.347
97	A	0.397	2.934	0.367	0.347
98	K	0.084	2.934	0.367	0.347
99	Y	0.084	2.934	0.367	0.347
105	S	0.171	2.634	0.293	0.273
106	S	0.171	2.634	0.293	0.273
107	H	0.256	2.634	0.293	0.273
108	W	0.07	2.634	0.293	0.273
109	Y	0.34	2.634	0.293	0.273
110	F	0.53	2.634	0.293	0.273

111	D	0.601	2.634	0.293	0.273
112	V	0.277	2.634	0.293	0.273
113	W	0.035	2.634	0.293	0.273
115	Q	0.222	3.213	0.357	0.337
116	G	0.222	3.213	0.357	0.337
117	T	0.051	3.213	0.357	0.337
118	L	0.355	3.213	0.357	0.337
119	V	0.489	3.213	0.357	0.337
120	T	0.523	3.213	0.357	0.337
121	V	0.391	3.213	0.357	0.337
122	S	0.391	3.213	0.357	0.337
123	S	0.391	3.213	0.357	0.337

The consensus aggregation-prone hot spots, as defined by the following tools—PDB-ePISA, Pasta, Tango, Waltz, Aggrescan 3D, CamSol intrinsic, and structurally corrected CamSol—were combined, and the results were compiled into a single table (**Table 100**). The residues with the highest likelihood of mutation were highlighted in green. To maintain antigen specificity, the CDRs were excluded as potential mutational targets. Additionally, FW1 was disregarded since it did not exhibit consensus hotspots according to multiple aggregation determining tools. FW2, FW3 and FW4 were focussed in the table below.

Table 100. List of mutational hotspots for aggregation prone residues by every *in silico* tool mentioned earlier



PDBePISA
Pasta
Tango
Waltz
Aggrescan 3D
CamSol, Instrinsic
CamSol, Structurally corrected

FW2																				
33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
G	M	N	W	V	R	Q	A	P	G	K	G	L	E	W	V	G	W	I
x	x	x	x	92	x	58	x	x	x	3	44	86	2	92	x	x	5	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	-5	-5	-5	-5	-5
0.2	2	2.3	3.1	3.7	3.6	3	0.5	0	0	0	0	0.6	1	0.8	0.3	0.2	0.7	2.4
x	x	x	x	x	x	x	x	x	x	x	x	96.9	97	97	97	97	97	97
-0	0	0	0	0	-0	-1	-1.1	-1	-2	-2	-1	0.49	-0	0.3	0	0	-0	0
-1	-1	-1	-0	-1	-0	-0	0.9	1.6	1	1	0.5	-0.2	-0	-1.2	-2	-1.2	-2	-1.8
-0	0	0	0	0	0	0	0.2	1.2	1	1	0.5	0.32	0	-0.5	0	0	-1	0

FW4											
113	114	115	116	117	118	119	120	121	122	123
W	G	Q	G	T	L	V	T	V	S	S
71	64	6	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x
2.7	0.7	0.1	0	0.8	1.1	19	32.7	34	19	1.8
97	97	97	96.7	97	96.7	97	96.7	97	97	x
0.2	-0.2	-1	-0.1	0.2	1.09	0	0.14	0	-1	-1
-0.9	-0.4	-1	-1.3	-0.9	1.56	-2	-1.5	-1.6	-2	-1
-0.5	-0	-1	-0.2	-0.2	-0.8	0	-0.3	0	-0	-0

FW3																																							
58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98
P	T	Y	A	A	D	F	K	R	R	F	T	F	S	L	D	T	S	K	S	T	A	Y	L	Q	M	N	S	L	R	A	E	D	T	A	V	Y	Y	C	A	K
x	22	x	75	4	1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	75	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
0	3.6	4.2	4.2	0.8	0.4	1.2	1.3	1.5	1.8	12	38	39	31	5	1.7	1.4	3.4	4	2.9	1.5	1.1	1.8	2.6	3.2	2.9	1.4	2.3	2.2	2.3	0.5	0.2	0.1	0.9	1.3	3.2	6.9	9	7.9	3.7	1.2
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	98.7	98.7	98.7	99	x	97	97	96.7
-1.23	-0.6	-1	-0.8	-1.6	-2.6	0	-3	-3	-2.3	0	-1.1	0	-0	-0	-1.1	-1	-1	-2	-1.1	0	0	-0.5	0	-1	0	-1.7	-1.5	0	-2.6	-1.9	-2.3	0	-0.6	0	0.62	0	0	0	0	0
0.189	0.3	0.6	-0.6	-0.4	0.3	1.37	0.9	0.5	-0.5	-0	-1.1	-1	-2	-1	-0.3	0.3	0.5	1	-0.1	-0	-0	-1.1	-1	-1	-1	-0	0.25	0.67	1.4	1.1	1	0.8	-0.1	-0.5	-1.6	-2.2	-2	-2	-1	-0.7
-0.08	0.1	0.1	0.05	0.6	0.4	0	1	0.7	0	0	-0.4	0	-1	-0	0.1	0.2	0.5	1	0.1	0	0	-0.1	0	-0	0	0.4	0.24	0	0.9	1	0.9	0	0.05	0	-0.6	0	-0	0	0	0

Additionally, a literature search was conducted to identify previously applied successful strategies for defining aggregation-prone residues (**Table 101**).

Table 101. The suggested mutation candidates on Bevacizumab V_H based on general strategy to humanise nanobody research study (Vincke et al. 2009).

Residue	Property	Prone to mutation?	Camelisation
V37	Hydrophobic	Yes	V → F
G44	Small	Yes	G → E
L45	Hydrophobic	Yes	L → R
W47	Hydrophobic	Yes	W → G

Table 101 presents the identified mutations with their physicochemical properties. Several potential mutations were identified within the V_H sequences of Bevacizumab by impacting the overall protein scaffold. Position 37 could be a double mutation candidate, by substituting it V to F (hydrophobic-to-hydrophobic). At position 44, an small-to-charged mutation G to E could increase the stability of the molecule. Additionally, position 45 a hydrophobic-to-charged mutation, L to R, may lead to a trade-off between higher stability and lower solubility. Amongst all the identified mutations, position 47 displayed a hydrophobic-to-small mutation by replacing W to G and emerged as the most promising candidate for rational design. On the other hand, positions W108, F110, and W113 are located in the HCDR3. Therefore they were not considered as potential aggregation hotspots.

The vast sequential difference between V_H and V_HH is observed at the FW2 region, which harbours hallmark residues (**Figure 30**).

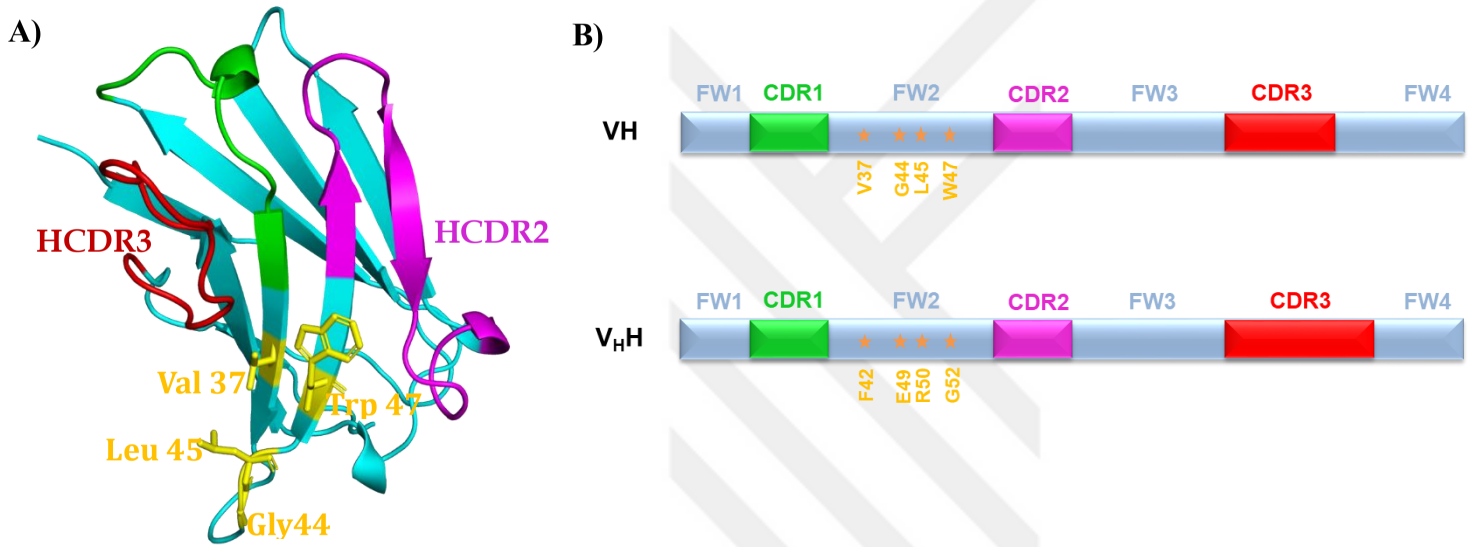


Figure 30. V_H structure of Bevacizumab.**A)** The 3D structure of Bevacizumab (PDB ID: 1BJ1) was shown. **B)** The comparison of aligned V_H and Camelid V_HH structures. A linear representation of the aligned V_H and camelid V_HH structures was shown. The HCDRs and FW regions were displayed in their respective alignments. In the V_HH structure, the hallmark residues at FW-2 were marked by yellow stars and corresponding V_H residues that align with these hallmark residues were also shown.

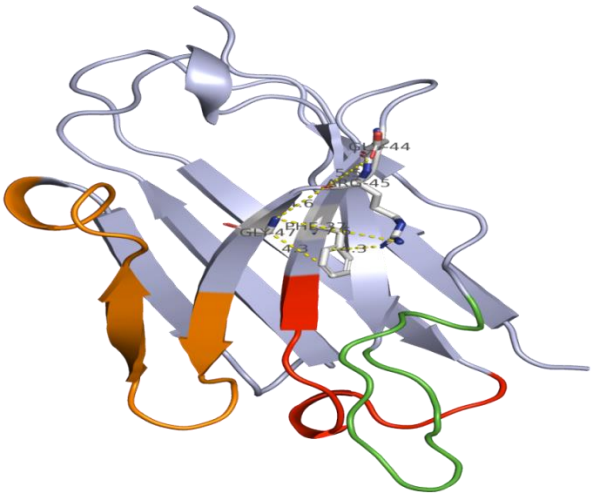


Figure 31. The former interfacing residues of Bevacizumab at the FW-2 of V_H. The 3D structure of Bevacizumab V_H was retrieved from the PDB repository (PDB ID: 1BJ1). The distance between the V_H FW-2 interface residues was measured. The V_H chain is shown in grey. HCDR1 is shown in red, HCDR2 is shown in orange, while HCDR3 is shown in green. The FW-2 hallmark residues were depicted in stick representation, with the V_H interface colored light grey.

The original V_H sequence was derived from Bevacizumab (WT sdAb) (**Figure 31**). The development process of Bevacizumab was done using hybridoma technology. First Fab format of Bevacizumab was generated in mice. Then, CDR-grafting was employed, by converting some mouse residues to human to improve compatibility for therapeutic use. In the next stage of the Bevacizumab design, some of the humanised residues were converted back to their original mouse counterparts. Finally, the optimised Bevacizumab sequence was obtained as humanised Fab format, (**Figure 32**) which contains substitutions of the human framework outside of the CDRs (Presta et al. 1997).

```
                        FW                                  FW
Mouse Fab       EIQLVQSGPELKQPGETVRISCKASGYTFTNYGMNWVKQAPGKGLKWMGWINTYTGEPTY    60
Humanized Fab   EVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGKGLEWVGWINTYTGEPTY    60
Humanized-12 Fab EVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGKGLEWVGWINTYTGEPTY    60
Final Fab       EVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGKGLEWVGWINTYTGEPTY    60
                *:***:**  * *** ::*:** **************:*******:*:************

                                    FW                                  FW
Mouse Fab       AADFKRRFTFSLETSASTAYLQISNLKNDDTATYFCAKYPHYYGSSHWYFDVWGAGTTVT    120
Humanized Fab   AADFKRRFTISRDNSKNTLYLQMNSLRAEDTAVYYCARYPHYYGSSHWYFDVWGQGTLVT    120
Humanized-12 Fab AADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVT    120
Final Fab       AADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVT    120
                *********:* :.* .* ***:..*: :***.*:**:**************** ** **

Mouse Fab       VSS       123
Humanized Fab   VSS       123
Humanized-12 Fab VSS      123
Final Fab       VSS       123                                                  12
                ***
```

[12] Adapted from IMGT Repertoire, Humanized antibodies, Protein display: bevacizumab and ranibizumab.

Figure 32. The evolutionary trajectory of Bevacizumab V_H sequences during its discovery. Mouse Fab: preliminary sequence discovered in mice through hybridoma technology, CDR-grafted Fab: CDR-grafted version, Humanised Fab: the humanised version, Final Fab: commercial version. The HCDRs were underlined. CDR1 sequences are showin in green text, CDR2 in yellow text, and CDR3 in red text. The residues modified to transform the mouse-derived Bevacizumab into a human-compatible form are highlighted in yellow. Dual transformed residues that were changed to human and converted back to mouse, were highlighted in cyan. The hallmark residues at FW2 were highlighted in gray.

The FW-3 region of the final Bevacizumab Fab exhibits differences compared to the original Bevacizumab at positions 70 and 72. However, when compared to V_Hs of antibodies from different sources (*Mus musculus*, *Homo sapiens*, and *Camelidae*), curated from abYsis database distributions, the amino acids in this region remain conserved across all antibodies at the positions 72, 74, 77, and 79 (**Table 102**).

Table 102. The specific amino acid residues found within the FW-3 region of Bevacizumab and other antibodies from different sources at the same positions. The table lists the amino acid sequences at positions 70, 72, 74, 77, and 79 for each respective antibody source. In the final Bevacizumab Fab form, the FW-3 region exhibited different amino acids compared to the original Bevacizumab. The corresponding amino acid sequences for *Mus musculus*, *Homo sapiens*, and the *Camelidae*-derived nanobody were retrieved from abYsis database.

FW-3 Residues[13]	**Bevacizumab**	**abYsis**[14]			**sdAb-DB**
	Final Fab	***Mus musculus***	***Homo sapiens***	***Camelidae***	**Nanobody**
70	F	L	I	I	I
72	L	V	R	R	R
74	T	K	N	N	N
77	S	S	N	N	N
79	A	A	L	V	V

The identified substitutions at each position contribute to the unique combinations, resulting in distinct sdAb variants of Bevacizumab V_H sequences (**Table 103**).

[13] The amino acid positions are aligned based on the numbering scheme of Bevacizumab, considering the presence of an additional residue in its CDR2.

[14] For some positions, there is more than 1 alternative residue available in the database/dataset, but the ones with highest frequency were written in the table.

Table 103. The total list of possible sdAb variants of Bevacizumab V_H based on the consensus results of the analysis mentioned above tables.

Residue	Mutation	Amino acid property[15]
37	Val→Tyr	Hydrophobic/Aliphatic→Less Hydrophobic/Aromatic
	Val→Phe	Hydrophobic/Aliphatic→Less Hydrophobic/Aromatic
44	Gly→Glu	Hydrophobic/Small→Hydrophilic/Acidic
45	Leu→Arg	Hydrophobic/Aliphatic→Hydrophilic/Basic
47	Trp→Gly	Hydrophobic/Aromatic→Hydrophilic/Aliphatic
70	Phe→Ile	Less Hydrophobic/Aromatic→ More Hydrophobic/Aliphatic
72	Leu→Arg	Hydrophobic/Aliphatic→Hydrophilic/Basic
74	Thr→Asn	Neutral→ Hydrophilic
77	Ser→Asn	Neutral→ Hydrophilic
79	Ala→Leu	Hydrophobic/Aliphatic→ More Hydrophobic/Aliphatic

The **Table 104** shows various mutations at specific residues of Bevacizumab, their locations within the FW-2 and FW-3 regions, and the nature of the mutation (Camelisation, Humanisation, or both). These mutations can potentially improve the structural and functional properties of Bevacizumab to generate biobetter sdAb format of it.

Table 104. The total list of all sdAb variants of Bevacizumab V_H.

Residue #	Mutations	Location	Nature of mutation
37	Val→Tyr	FW2	Camelisation
44	Val→Phe		
45	Gly→Glu		
47	Leu→Arg		
44,45	Gly→Glu, Leu→Arg	FW2	Camelisation
44,47	Gly→Glu, Trp→Gly		
44,45,47	Gly→Glu, Leu→Arg, Trp→Gly		
37,44,45,47	Val→Tyr, Gly→Glu, Leu→Arg, Trp→Gly		

[15] The hydrophobicity index was based on Kyte-Doolitle scale (Kyte and Doolittle 1982)

37,44,45,47	Val→Phe, Gly→Glu, Leu→Arg, Trp→Gly		
70,72,74,77,79	Phe→Ile, Leu→Arg, Thr→Asn, Ser→Asn, Ala→Leu	FW3	Humanisation
37,44,45,47,70,72,74,77,79	Val→Tyr, Gly→Glu, Leu→Arg, Trp→Gly, Phe→Ile, Leu→Arg, Thr→Asn, Ser→Asn, Ala→Leu	FW2+ FW3	Camzelization + Humanisation

When transforming Bevacizumab into the V_H-only format, certain residues tend to interact with each other to compensate the absence of V_L in the original design. To simulate this natural evolution of camelids, particularly FW-2 residues were focussed on. The hallmark residues at FW-2 were mutated into corresponding camelid residues, through the camelisation approach. On the other hand, humanisation mutations were designed for the FW-3 region based on the residues that were previously converted back to their original mouse counterparts (**Table 104**).

By incorporating camelisation and humanisation strategies, the V_H format of Bevacizumab was optimised to possible enhance its stability, specificity, and binding properties, meanwhile making it closer to naturally found camelid hcIgGs.

4.2. Site-directed mutagenesis for generation of sdAb variants

The Bevacizumab sdAb variants were generated through performing sequential SDM reaction. The generation of variant was strategically planned to further generate sdAb variants with multiple mutations (double, triple, etc.). The amino acid (**Table 105**) and nucleotide (**Table 106**) sequences of sdAbs designed through SDM were listed as follows;

Table 105. The codon-optimised V_H peptide sequence of Bevacizumab was utilized, along with engineered mutations to create its associated variants. The protein sequences were given for each sdAb variant. If any mutations were introduced to the WT sequence, they are indicated using the following format: Mutated residue as single letter amino acid code: [Original amino acid][Kabat numbering][Mutated amino acid]. The following residues and regions were highlighted in the sequences as follows: NdeI sequence and first nucleotide of the start codon were highlighted in cyan, PelB sequence and the last two nucleotides of the start codon in green, NheI in grey, Codon optimised V_H sequence, Mutation, SalI in dark gold, Factor Xa protease in bold text, $(Flag\ Tag)_3$ in underlined text, His_7 tag in italicised text, Stop codon in red, and XhoI in pink.

sdAb	sdAb name	Designed peptide sequence
1	WT	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWVRQAPGKGLEWVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
2	L45R	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWVRQAPGKGREWVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
3	W47G	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWVRQAPGKGLEGVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
4	G44E	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWVRQAPGKELEWVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
5	V37F	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWFRQAPGKGLEWVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
6	V37Y	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWYRQAPGKGLEWVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
7	G44E/L45R	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWVRQAPGKERWVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
8	G44E/W47G	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWVRQAPGKELEGVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
9	G44E/L45R/W47G	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWVRQAPGKERWVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
10	V37F/G44E/L45R/W47G	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWFRQAPGKERWVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
11	V37Y/G44E/L45R/W47G	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWYRQAPGKERWVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
12	F70I/L72R/T74N/S77N/A79L	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWVRQAPGKGLEWVGWINTYTGEPTYAADFKRRFTISRDNSKNTLYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE
13	V37Y/G44E/L45R/W47G+F70I/L72R/T74N/S77N/A79L	HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTF TNYGMNWYRQAPGKERWVGWINTYTGEPTYAADFKRRFTISRDNSKNTLYLQMNS LRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADY KDDDDKAADYKDDDDKAAHHHHHHH-LE

Table 106. The codon-optimised V_H nucleotide sequences of Bevacizumab, were utilized along with designed mutations to create its variants. For each sdAb variant, the used template and associated primer were listed. The codon-optimised nucleotide sequences for each sdAb variant were presented in the 5' to 3' direction. If any mutations were introduced to the WT sequence, they are indicated using the following format: Mutated residue: [Original amino acid][Kabat numbering][Mutated amino acid]. The following residues were highlighted as follows: The following regions were highlighted in the sequences. NdeI sequence and first nucleotide of the start codon were highlighted in cyan, PelB sequence and the last two nucleotides of the start codon in green, NheI in grey, Codon optimised V_H sequence, Mutation, SalI in dark gold, Factor Xa protease in bold text, (Flag Tag)3 in underlined text, His7 tag in italicised text, Stop codon in red, and XhoI in pink.

sdAb	sdAb name/ template/ primer	Codon optimised nucleotide sequence
1	WT Template: Codon optimised Bevacizumab VH fragment	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG GTC CGC CAA GCG CCA GGG AAA GGA TTA GAG TGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
2	L45R Template: WT Primer: L45R	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG GTC CGC CAA GCG CCA GGG AAA GGA AGA GAG TGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
3	W47G Template: WT	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT

	Primer: W47G	GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG GTC CGC CAA GCG CCA GGG AAA GGA TTA GAG GGC GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
4	G44E Template: WT Primer: G44E	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG GTC CGC CAA GCG CCA GGG AAA GAA TTA GAG TGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
5	V37F Template: WT Primer: V37F	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG TTC CGC CAA GCG CCA GGG AAA GGA TTA GAG TGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
6	V37Y Template: WT Primer: V37Y	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG TAC CGC CAA GCG CCA GGG AAA GGA TTA GAG TGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
7	G44E/L45R Template: G44E Primer:	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG GTC CGC CAA GCG CCA GGG AAA GAA AGA GAG TGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG

	L45R	CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
8	G44E/W47G Template: G44E Primer: W47G	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG GTC CGC CAA GCG CCA GGG AAA GAA TTA GAG GGC GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
9	G44E/L45R/W47G Template: WT Primer: G44E/L45R/W47G	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG GTC CGC CAA GCG CCA GGG AAA GAA AGA GAG GGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
10	V37F/G44E/L45R/W47G Template: G44E/L45R/W47G Primer: V37F	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG TTC CGC CAA GCG CCA GGG AAA GAA AGA GAG GGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
11	V37Y/G44E/L45R/W47G Template: G44E/L45R/W47G Primer: V37Y	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG TAC CGC CAA GCG CCA GGG AAA GAA AGA GAG GGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC TTT TCA CTT GAC ACT TCC AAA AGC ACA GCG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC

		CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
12	F70I/L72R/T7 4N/S77N/A79 L Template: WT Primer: F70I/L72R/T7 4N/S77N/A79 L	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG GTC CGC CAA GCG CCA GGG AAA GGA TTA GAG TGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC ATT TCA CGT GAC AAT TCC AAA AAC ACA CTG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG
13	V37Y/G44E/L 45R/W47G+ F70I/L72R/T7 4N/S77N/A79 L Template: V37F/G44E/L 45R/W47G Primer: F70I/L72R/T7 4N/S77N/A79 L	CAT ATG AAG TAT TTA CTG CCT ACA GCT GCC GCT GGA CTG CTG CTG TTG GCC GCC CAG CCC GCT ATG GCG ATG AGT GCT AGC GAG GTC CAG TTG GTT GAA AGT GGT GGA GGT CTG GTT CAG CCT GGT GGC TCG CTG CGT TTA AGC TGT GCA GCA TCA GGG TAC ACA TTT ACA AAT TAT GGT ATG AAC TGG TTC CGC CAA GCG CCA GGG AAA GAA AGA GAG GGG GTA GGG TGG ATC AAT ACT TAT ACC GGC GAG CCC ACC TAC GCT GCA GAT TTC AAG CGC CGT TTT ACC ATT TCA CGT GAC AAT TCC AAA AAC ACA CTG TAC TTA CAA ATG AAC TCC TTG CGC GCG GAG GAT ACA GCT GTG TAT TAT TGT GCT AAG TAC CCG CAC TAT TAC GGT TCA TCC CAT TGG TAT TTC GAT GTG TGG GGT CAG GGG ACA TTG GTA ACC GTC TCT TCC GTC GAC **ATC GAG** **GGA CGC** GAT TAT AAG GAT GAT GAT GAT AAG GCC GCT GAT TAT AAA GAC GAC GAT GAC AAA GCA GCT GAT TAT AAA GAC GAC GAC GAT AAA GCT GCG *CAC CAT CAC CAT CAT CAT CAC* TAA CTC GAG

Following the SDM procedure, DpnI digestion was employed to selectively eliminate the DNA template and/or unmutated parental DNA remaining in the reaction mixture.

The outcomes of the SDM experiment, was analysed through agarose gel electrophoresis to assess the success of the mutagenesis process. The gel was visiualized under UV light (**Figure 33**), using a UV trans-illuminator plate.

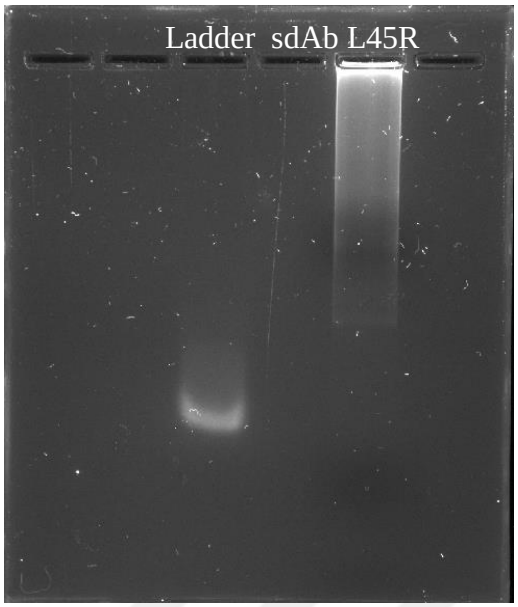


Figure 33. 1% concentration agarose gel of sdAb L45R, was prepared and incorporated with SYBR Safe dye (1:10000) for visualization. 1 kb DNA ladder was run in parallel with L45R variant. The sample was combined with 5X loading dye (1:4). A total volume of 25 µl of the sample+dye mixture was loaded into a well within the gel, and the gel was run as described earlier.

Upon visualization, the super-coiled DNA plasmid material manifested as a smear in the upper portions of the gel, as expected. With the aid of a surgical scalpel, the diffused portion was excised. In this way, any DNA fragment, other than super-coiled plasmid DNA, would be selectively eliminated, whereby the removal of the contaminants, such as primer pairs, which are also small DNA fragments, would be facilitated. Following the agarose gel electrophoresis, DNA fragments trapped within the excised portion of the agarose gel was purified for downstream transformation into recombinant expression cells in *E. coli*.

Prior to the transformation stage, the outcome of the SDM reaction should be confirmed by Sanger sequencing. Once, the sequencing results align with the expected mutation on the amino acid level, (sometimes the nucleotide comes with a different alteration than the intended change, however the resulting codon yields the same amino acid), the spared and sequence confirmed plasmids are ready to be transformed into *E. coli* cells (BL21(DE3) pLysS) for recombinant protein production. There are two options for this transformation: heat-shock and electroporation. Electroporation was preffered right after this transformation confirmation step specifically, because of the low DNA yields achived after SDM reaction. Besides, the corresponding electrocompetent cells (XL10-Gold

Ultracompetent *E. coli* cells) are not suitable hosts for recombinant protein expression. After this electroporation, there could be a second round of transformation to amplify the sequence confirmed plasmids for long term storage. This is achieved by utilizing DH5α *E. coli* strain for the heat-shock transformation procedure as detailed in the methods section.

4.3. Optimizing the recombinant protein production yield

When optimizing the recombinant protein production, we considered variying parameters, ranging from different inoculation methods to purification techniques from the start of the whole process till the end. Major changes were also explored, such as changing the production cell line from BL21(DE3) pLysS to WK6 to compare expression efficiency. Mostly the production conditions were altered to observe the impact of these changing parameters, such as production pH, temperature, induction temperature, and shaking rates. Each modification was individually experimented and their impact on the protein quality and the quantity was confirmed with relevant measures, depending on the nature of the optimisation. The other optimizing strategies includes introducing boosting agents in the production media, utilization of different types of production flasks, different induction methodologies, different IMAC agent use were assessed to improve overall recombinant protein production yield and purity.

4.3.1. Designing the mutations based on E.coli codon usage

Bacterial production systems offer unique advantages such as rapid growth rate, easier genetic manipulation via engineering techniques, scalability of the production and being cost effective. The BL21(DE3) pLysS strain of *E. coli* cells, serve as a hosting platform for the efficient production of our recombinant sdAbs. This bacterial expression system utilizes T7 RNA polymerase, allowing the controlled expression of the target gene under the control of an inducible promoter.

In the design of primers for the expression of sdAbs in hosting *E. coli* BL21(DE3) pLysS strain, codon preference frequency of the bacteria significantly impact the success rate of the SDM reaction. When designing the primers for the generation of the new sdAb variants, this issue was detailly considered to prevent possible translational mistakes. Each primer was tailored based on the codon usage frequency of our bacterial strain.

The following sdAb variants were designed carefully and their respective calculated properties were listed in the table below (**Table 107**). These data are required for protein concentration measurement when using a spectrophotometer.

Table 107. Bevacizumab V_H home-lab codon-optimised variants, their molecular weight, theoretical isoelectric points and extinction coefficient values calculated by ExPASy ProtParam (Duvaud et al. 2021).

sdAb IBG #	Bevacizumab sdAb variant	MW (Da)	Theoretical pI	Extinction coefficient (M^{-1} cm^{-1})
1	WT	19010.64	5.09	48485
2	L45R	19058.68	5.09	48485
3	W47G	19074.68	5.09	49975
4	G44E	19082.70	5.00	48485
5	V37F	19053.66	5.22	48485
6	V37Y	18881.48	5.09	42985
7	G44E/L45R	19125.73	5.11	48485
8	G44E/W47G	18953.54	5.00	42985
9	G44E/L45R/W47G	18996.57	5.11	42985
10	V37F/G44E/L45R/W47G	19044.61	5.11	42985
11	V37Y/G44E/L45R/W47G	19060.61	5.11	44475
12	F70I/L72R/T74N/S77N/A79L	18859.48	5.24	48485
13	V37Y/G44E/L45R/W47G+ F70I/L72R/T74N/S77N/A79L	18909.45	5.26	44475

4.3.2. The choice of recombinant protein production host

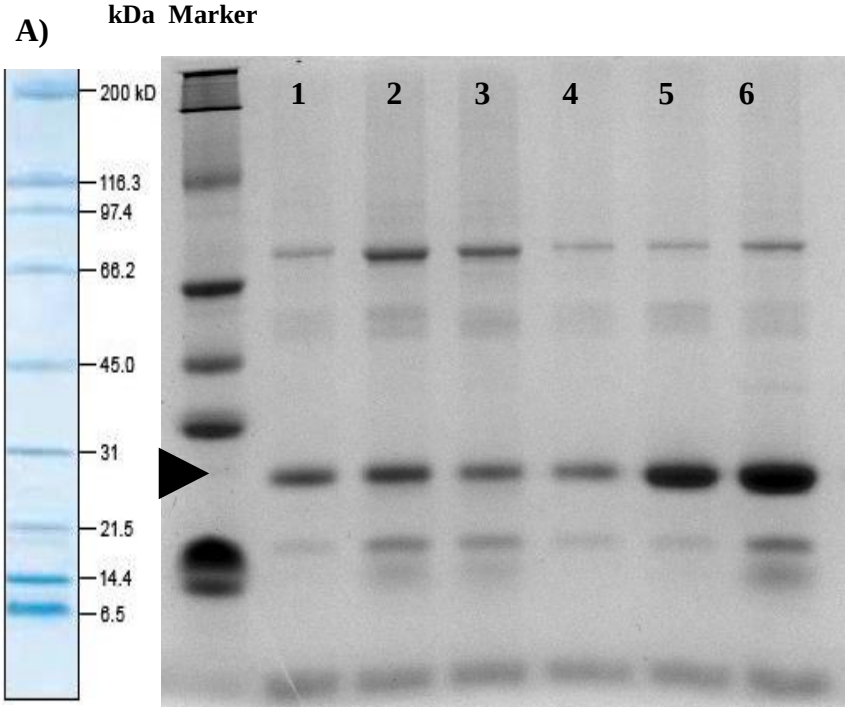
The well-characterized and widely used BL21 pLysS *E.coli* cell line offers distinct advantages but also disadvantages to consider. One of its key advantages is harbouring chloramphenicol resistance gene enabling to exert selection pressure on transformation process, however this feature may also cause false positives or negatives.

The protein production concentration (mg/L culture) is presented in **Table 108,** while the assessment of purity was verified through SDS-PAGE (**Figure 34**) and WB analysis (**Figure 35**) using BL21 pLysS cells.

Table 108. The production concentration of every sdAb produced through BL21 pLysS *E.coli* cells. For each sdAb sample, the production was repeated at least three different times, using the same production conditions. The average concentration of productions was pooled and the

concentration was calculated as mg/L culture and presented with standard deviation of three separate batches.

sdAb IBG #	sdAb	Concentration (mg/L culture)
1	WT	0.68 ± 0.05
2	L45R	2.42 ± 0.38
3	W47G	1.67 ± 0.86
4	G44E	1.76 ± 0.25
5	V37F	7.72 ± 0.85
6	V37Y	8.51 ± 0.63
7	G44E/L45R	1.99 ± 0.05
8	G44E/W47G	2.80 ± 0.23
9	G44E/L45R/W47G	1.18 ± 0.27
10	V37F/G44E/L45R/W47G	1.08 ± 0.09
11	V37Y/G44E/L45R/W47G	3.56 ± 0.33
12	F70I/L72R/T74N/S77N/A79L	0.56 ± 0.14
13	V37Y/G44E/L45R/W47G/F70I/L72R/T74N/S77N/A79L	0.96 ± 0.24



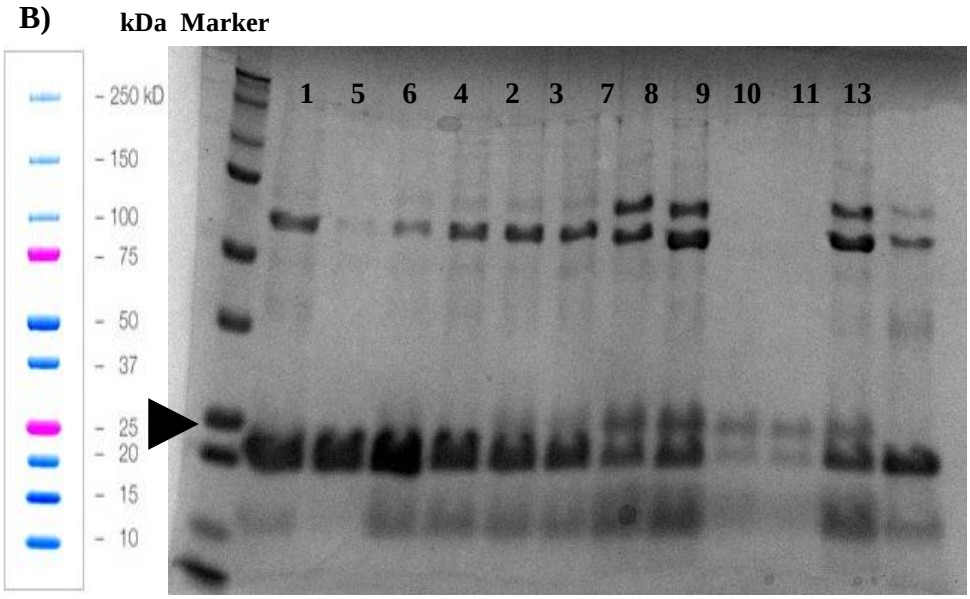


Figure 34. The SDS-PAGE results of sdAbs produced using BL21 pLysS *E. coli* cells. **A)** The isolated protein samples (750 ng, if the sample concentration is lesser than this value, than it was loaded into the gel as it was) were run using 12% acrylamide gel as explained in the methods section. **B)** The isolated protein samples (1000 ng, if the sample concentration is lesser than this value, than it was loaded into the gel as it was). The lane numbers represent the sdAb #. Please refer to **Table 107** for the full nomenclature of the introduced mutations.

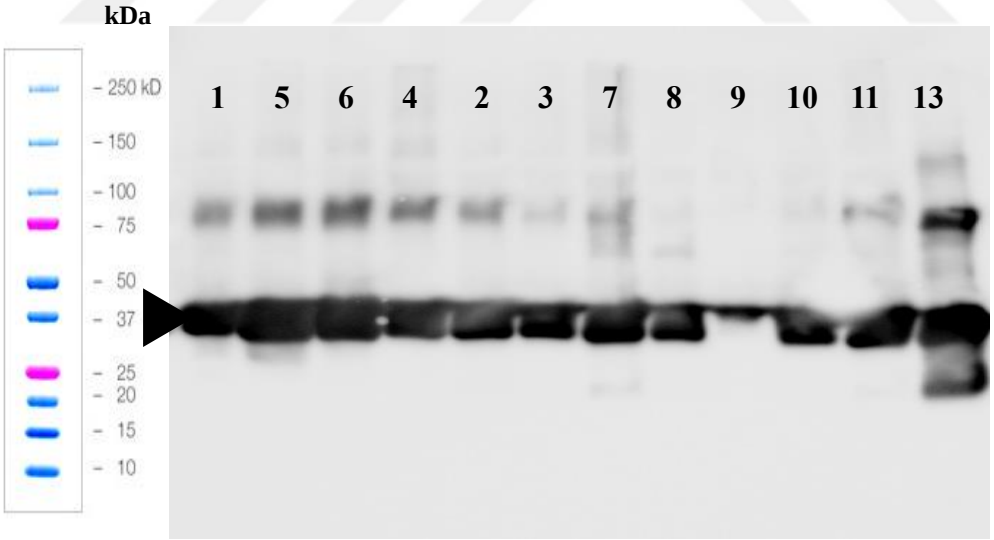


Figure 35. The immunoblotting results of sdAbs produced using BL21 pLysS *E. coli* cells. The purified protein samples (100 ng) were probed with monoclonal Anti-Flag M2 IgG produced in mice (1:5000) and polyclonal HRP-Conjugated secondary Goat Anti-mouse IgG (1:10000). The lane numbers indicate the sdAb #. Please refer to **Table 107** for the full

nomenclature of the introduced mutations. The samples loaded into the gel were identical to the SDS-PAGE samples presented in the previous figure, except having 1:10 dilution.

4.3.3. Transformation methods to the host cell

For introducing foreign DNA into host cells, two approaches were utilized: heat shock and electroporation. They exhibit distinct advantages and limitations. Heat shock transformation is relatively simple and widely applicable, where electroporation offers enhanced transformation efficiency, but demands more optimisation. Special pulser instrurment and compatible cuvette are required for electroporation takes place, however the instrument gave "Arc" error frequently, causing the loss of both electrocompotent cells and the insert plasmid. These methods are chosen based on the following factors: the type of experiment (protein expression or plasmid amplification), cell type and resources availability.

The transformation efficacy can be monitored by observing the total number of viable colonies cultivated on selective growth media with antibiotic resistance. The higher the number of colonies indicates successful transformation process (**Figure 36**).

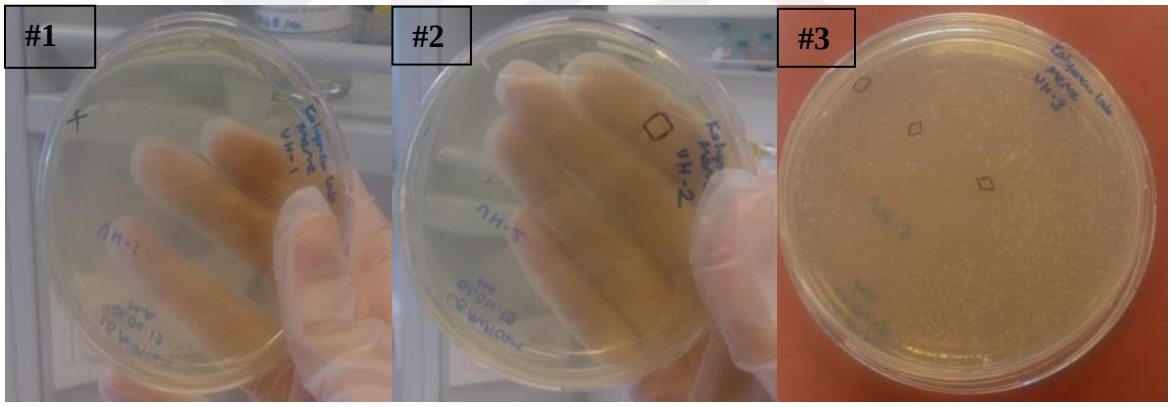


Figure 36. Agar plates containing AMP+ and CAM+ antibiotics were inoculated with an equal quantity of sdAb plasmids. The plates were then left to incubate at 37°C overnight to allowing colony formation. The plates **#1** and **#2** have the transformation mixture after heat-shock transformation, whereas the plate **#3** displayed the colonies derived from electroporation. Each transformation was carried out using the corresponding competent cell line strain associated with the specific method.

Based on the observations, the electroporation method demonstrated higher success rates. However electrocompetent XL10-Gold Ultracompetent *E. coli* cells, are not compatible with recombinant protein production. Therefore, for downstream processes, especially during the initiation of production, the heat shock procedure was preferred. Nevertheless,

electroporation was employed specifically for plasmid multiplication purposes after SDM. An empirical observation suggests that BL21 pLysS *E. coli* cells transformed with plasmids cultivated using XL10-Gold Ultracompetent *E. coli* cells yield more protein (mg/culture L) compared to production under the same conditions but with a different transformation method (heat-shock).

4.3.4. Inoculation for starting culture

Another optimisation endeavour was on inoculation for starting culture. Two distinct methodologies were used for inoculum: fresh single colony and frozen stock of a single colony. As the name implies, using a fresh single colony means utilization of the newly generated colony following the transformation. However, for each production batch, performing another transformation increases the likelihood of introducing mutations and/or false positive or negative colony formation on the selective media. In contrast, when utilizing a frozen stock of inoculum (frozen aliquots of starter culture in 50% glycerol + 50% cell suspension), this risk is eliminated by ensuring the addition of sequence confirmed plasmid all the time. Additionally, if the same frozen stock is used all the time, the reproducibility of the production cycles will be increased, hence all the bacteria is transformed with the sequence confirmed plasmid coming from the same single-colony. Considering these facts, utilization of a frozen stock is a better choice over using a fresh colony as an inoculum.

4.3.5. The production vessel

Another important consideration for recombinant protein production optimisation is the selection of the appropriate production vessel. Two types of flasks, namely the standard Erlenmayer flask and the baffled flask, were investigated in our studies. The baffled flasks are further divided into two options: Isolab and Pyrex brands. The Isolab baffled flask, crafted with four deeply-integrated baffles, whereas the Pyrex baffled flask contains four indents. These crevices provide optimal and uniform agitation, promoting homogenous growth of suspension cultures, especially in in aerobic bacterial. The baffles provide more gas exchange area for oxygenation by expanding the surface area. However, it is worth noting that *E. coli*, classified as a facultative anaerobe. The comparison in the production yields on the same sdAb sample was investigated utilizing different Erlenmeyer flasks (**Table 109**).

Table 109. The effect of different type of Erlenmeyer flasks on recombinant protein production concentration. The average yield of protein production using different types of Erlenmayer flasks was shown based on the protein yield. The yields were represented as the

means of at least three different flasks used in the same batch of the production and the experiments were repeated at least three times and presented with standard deviation of three separate flasks.

sdAb sample	Type of flask	Concentration (mg/L culture)
F70I/L72R/T74N/S77N/A79L	Erlenmayer Flask	0.17 ± 0.03
	Baffled Flask, Isolab	1.48 ± 0.14
	Baffled Flask, Pyrex®	2.67 ± 0.27

Based on the protein production results, the baffled type of Erlenmeyer flask leads the competition over regular ones, however Pyrex brand surpass the Isolab in terms of yield.

4.3.6. The production volume

The effect of different production volumes in Erlenmeyer flasks were evaluated next for optimization purposes. The optimisation stage includes assessment of different size of Erlenmeyer flasks constituting different volumes. Following the adoption of baffled flasks, a notable improvement in both the quality (**Figure 37**) in comparison to regular Erlenmayer flasks was observed. Additionally, the quantity (**Table 110**) of sdAb was increased, when using 1:3 ratio of production medium to production flask.

Table 110. The effect of recombinant protein production volume to Erlenmayer flask ratio. The protein production yield of different volumes in different sized of baffled Erlenmayer flasks were shown. The concentration were represented as the means of at least three different flasks used in the same batch of the production and the experiments were repeated at least three times and presented with standard deviation of three separate flasks.

sdAb Sample	Production volume	Flask volume	Concentration (mg/L culture)
F70I/L72R/T74N/S77N/A79L	85	250	2.13 ± 0.15
	165	500	0.04 ± 0.01
	300	1000	2.58 ± 0.39
	330	1000	2.62 ± 0.35
	600	2000	0.96 ± 0.11

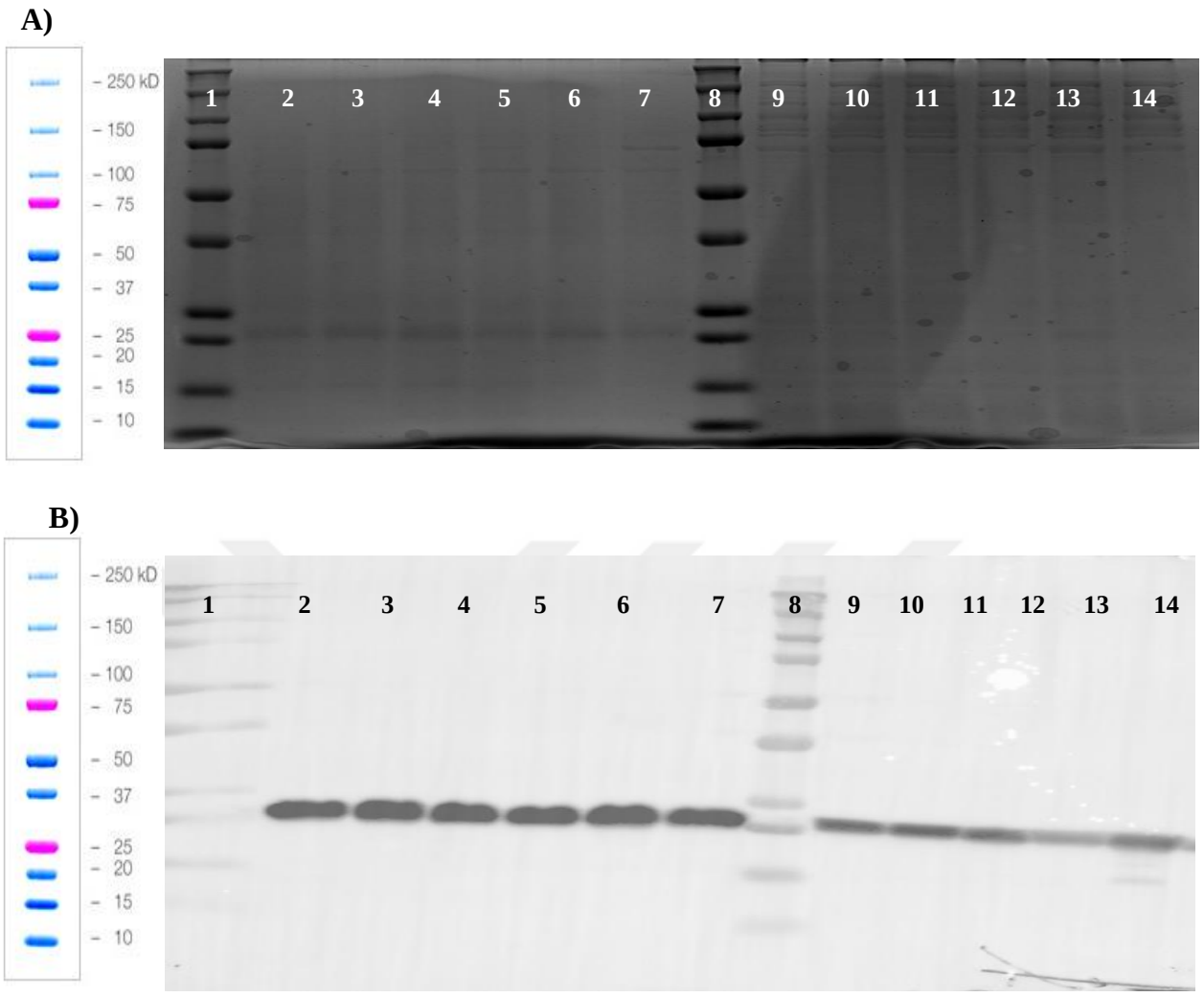


Figure 37. SDS-PAGE and WB results of the different production flask comparisons of F70I/L72R/T74N/S77N/A79L sdAb sample. The samples were isolated directly through the supernatant via leaky expression (the first set: lanes 2-7) and also collected from the flow through (the second set: lanes 9-14). The flask numbered 1-3 have less baffled cavities (Pyrex) and flasks 4-6 have deeper cavities (Isolab). **A)** For the SDS-PAGE analysis, the samples were loaded as 1000 ng and flow through samples were loaded as they were, whereas **B)** WB contains 100 ng of each sample. The WB membrane was probed against mouse anti-flag primary antibody (1:10000), followed by secondary HRP-conjugated antibody (1:10000).

It has been known that, the volume of the suspension culture and the surface are associated with increased cell growth, therefore increased production rates. By assessing these differences in production, the optimal ratio that yields the highest protein production is 1:3.

4.3.7. The boosting the production by altering temperature

One other approach for the optimisation of recombinant protein production is utilizing boosting temperatures to promote the growth of *E.coli* cells, especially during the lag phase. For that purpose, a two-step strategy could be used. Starting with an optimal temperature for *E.coli*, like 30°C, can result in more rapid cell growth. When inducing the recombinant protein production with auto-induction media, a lower temperature, 18°C, is required so that the expressed proteins can form properly folding, meanwhile preventing the inclusion body formation and promoting optimal protein solubility. Alternatively, the direct start of the *E. coli* culture at 18°C was investigated. In case of IPTG induction, another two-step temperature strategy was used. The *E. coli* growth was boosted at 37°C, and following IPTG addition, the temperature was lowered to 28°C.

Table 111. Promoting bacterial growth through different temperature conditions involved applying the boosting temperature within the 8 to 18-h timeframe. The duration of this temperature elevation was determined by referencing the OD_{600} value. The resulting protein production concentrations for different induction methods, with varying boosting temperatures, were presented. These concentration values were calculated as means derived from either a minimum of three separate flasks utilized within the same production batch or from a different production batch and the experiments was repeated no less than three times presented with standard deviation of three separate batches.

sdAb Sample	Induction type	Boosting Temperature (°C)	Induction Temperature (°C)	Concentration (mg/L culture)
V37F/G44E/L45R/W47G	Auto induction	-	18	2.38 ± 0.67
		30	18	0.31 ± 0.14
		37	18	1.13 ± 0.01
V37Y/G44E/L45R/W47G/ F70I/L72R/T74N/S77N/A79L	IPTG	-	18	0.52 ± 0.09
		37	28	0.34 ± 0.1

As per the results (**Table 111**), the most favourable protein yield outcomes were attained by maintaining a lower temperature throughout the bacterial growth process. For the upcoming production attempts, 18°C was preferred and this temperature was remained the same throughout the production.

4.3.8. The production pH

When optimizing recombinant protein production, another strategy is trying different production pH conditions, such as pH 6.8 and pH 7.8, attained by altering in the composition of 20X NPS buffer (**Table 36**).

At pH 6.8 could provide a suitable environment for the enhanced stability and proper protein folding, meanwhile minimizing the formation of inclusion bodies and improving solubility. In contrast, at pH 7.8, different proteins might exhibit improved production yields compared to pH 6.8.

Table 112. The influence of varying pH values on protein production concentrations was examined. Recombinant protein production was carried out under distinct pH conditions. For each sdAb, both pH conditions were tested by altering the 20X NPS buffer composition. The resulting concentration values were calculated as averages from either a minimum of three separate flasks utilized within the same production batch or obtained from different production batches and presented with standard deviation. The experimental process involved a minimum of three repetitions to ensure reproducibility.

sdAb Sample	pH	Concentration (mg/L culture)
WT	6.8	0.14 ± 0.02
	7.8	0.14 ± 0.04
L45R	6.8	2.42 ± 0.25
	7.8	1.55 ± 0.24
W47G	6.8	2.48 ± 0.28
	7.8	1.67 ± 0.16
G44E	6.8	2.27 ± 0.17
	7.8	1.48 ± 0.21
V37F	6.8	7.72 ± 0.93
	7.8	6.87 ± 0.91
V37Y	6.8	8.51 ± 0.79
	7.8	6.01 ± 0.58
G44E/L45R	6.8	2.04 ± 0.16
	7.8	0.99 ± 0.13

G44E/W47G	6.8	3.03 ± 0.15
	7.8	0.58 ± 0.14
G44E/L45R/W47G	6.8	1.19 ± 0.53
	7.8	1.06 ± 0.09
V37F/G44E/L45R/W47G	6.8	2.38 ± 0.14
	7.8	0.98 ± 0.13
V37Y/G44E/L45R/W47G	6.8	0.44 ± 0.07
	7.8	0.35 ± 0.08
F70I/L72R/T74N/S77N/A79L	6.8	0.52 ± 0.06
	7.8	0.47 ± 0.08
V37Y/G44E/L45R/W47G/F70I/L72R/T74N/S77N/A79L	6.8	0.88 ± 0.12
	7.8	0.76 ± 0.23

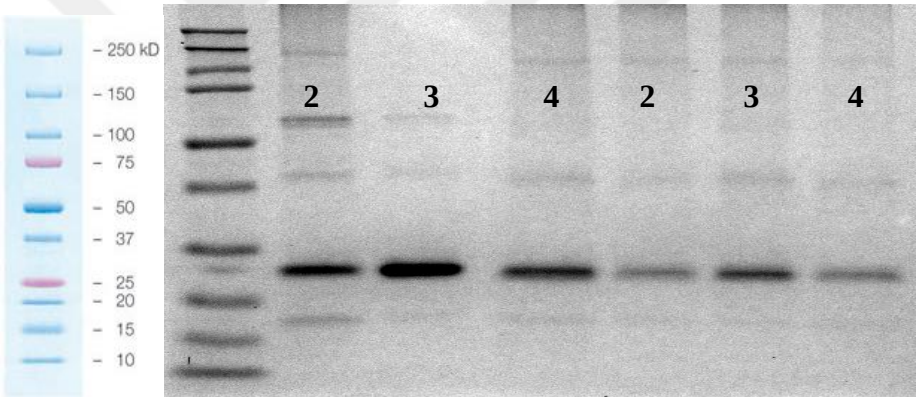


Figure 38. The effect of change in pH on recombinant protein production quality. The first three lanes of SDS-PAGE gel were loaded with 1000 ng of sdAbs 2:L45R, 3:W47G and 4: G44E grown at pH 6.8, where the second triple set constitutes the same samples but incubated at pH 7.8. The specific pH values were adjusted by using 20X NPS recipes. Please refer to **Table 107** for the full nomenclature of the introduced mutations.

Based on the protein production concentrations, it was observed that in 50% of the instances, utilizing a production media with a pH of 6.8 led to a two fold increase in yield (**Table 112**). Although for some sdAb variants, using media with pH of 7.8 yielded better results. In summary, a pH of 6.8 emerged as the preferred choice for subsequent production trials based on the overall results.

4.3.9. The production shaking rate

Another influencing factor in optimisation of recombinant protein production is production shaking rates, which can eventually affect the growth kinetics of the host cells and therefore protein expression levels. There were two options being tested in our trials: 225 and 250 rpm. At 225 rpm, sufficient oxygen levels and less shear stress can be achieved compared to 250 rpm. In contrast, 250 can provide increased oxygen supply and agitation rates, meanwhile contributing to the foaming. When shaking the cultures at higher rates, the probability of the shear stress should be also considered. To decide which shaking rate is the best for our production, the production concentrations of every sdAb was evaluated.

Table 113. The impact of varying the shaking rate on the concentrations of recombinant protein production was examined. Recombinant protein synthesis was carried out at different shaking rates using baffled flasks. The resultant concentration measurements were calculated as means derived from either at least three separate flasks employed within the same production run or acquired from different production batches and presented with standard deviation. To ensure the reproducibility of the results, the experimental procedure was repeated a minimum of three times.

sdAb Sample	Shaking rate (rpm)	Concentration (mg/L culture)
WT	225	0.68 ± 0.08
	250	4.42 ± 0.20
L45R	225	2.42 ± 0.67
	250	2.71 ± 0.65
W47G	225	1.67 ± 0.42
	250	2.48 ± 0.18
G44E	225	1.75 ± 0.39
	250	4.22 ± 0.51
V37F	225	7.72 ± 1.51
	250	11.32 ± 1.36
V37Y	225	8.51 ± 0.36
	250	9.24 ± 0.41
G44E/L45R	225	1.99 ± 0.67

	250	2.99 ± 0.57
G44E/W47G	225	2.80 ± 0.32
	250	3.03 ±0.45
G44E/L45R/W47G	225	1.18 ± 0.21
	250	7.78 ± 0.98
V37F/G44E/L45R/W47G	225	1.65 ± 0.38
	250	6.11 ± 1.24
V37Y/G44E/L45R/W47G	225	3.81 ± 0.45
	250	6.31 ± 1.07
F70I/L72R/T74N/S77N/A79L	225	0.56 ± 0.16
	250	2.97 ± 0.43
V37Y/G44E/L45R/W47G/F70I/L72R/T74N/S77N/A79L	225	0.96 ± 0.15
	250	9.56 ± 1.34

In conclusion, the analysis of yield results pointed out 250 rpm is the optimal shaking rate for recombinant protein production in this comparison (**Table 113**). The data demonstrates that cultures subjected to this shaking rate consistently yield higher concentrations of protein. The decision was made to pursue 250 rpm in subsequent production studies to attain improved yields.

4.3.10. The induction for target gene expression

The auto-induction method (Studier 2005), is a novel strategy for recombinant protein production using BL21(DE3) pLysS cells. This approach utilizes the intricate interplay between lac promoter and the T7 RNA polymerase. The auto-induction media contains the mixture of glucose, lactose and galactose, where each component plays a distinct role. In the initial phase of the production, the presence of glucose exerts a dual effect on suppression of T7 RNA polymerase activity, meanwhile offers a rich carbon source for rapid bacterial duplication. When glucose is consumed, lactose kicks in triggering the T7 promoter that initiates the expression of gene of interest. Galactose supplements the cell to maintain basal expression levels as a secondary carbon source, once glucose levels are depleted. This inherent machinery does not require an inducing agent from the outside, which renders an autonomous induction. Auto-induction surpasses the traditional IPTG induction that requires manual intervention, with its many advantages.

Table 114. The influence of distinct induction methods on recombinant protein production was investigated. Protein synthesis was induced either with auto-induction media or by addition of IPTG once a specific level of BL21(DE3) pLysS cells confluency was attained. The resulting concentration measurements were calculated as averages obtained from either a minimum of three separate flasks used within the same production sequence or collected from diverse production sets and presented with standard deviation. To guarantee result consistency, the experimental process was replicated at least three times.

sdAb Sample	Induction type	Boosting Temperature	Induction Temperature	Concentration (mg/L culture)
V37Y/G44E/L45R/W47G/ F70I/L72R/T74N/S77N/A79L	Auto induction	-	18	1.02 ± 0.09
	IPTG	37	28	0.34 ± 0.10

In conclusion, the evaluation of yield results (**Table 114**) favours the use of auto-induction methods over traditional IPTG induction for recombinant protein production in BL21(DE3) pLysS cells. The data reflects the superior performance of the auto-induction approach in terms of protein yield per liter of culture.

4.3.11. Boosting the target protein production yield by sucrose addition

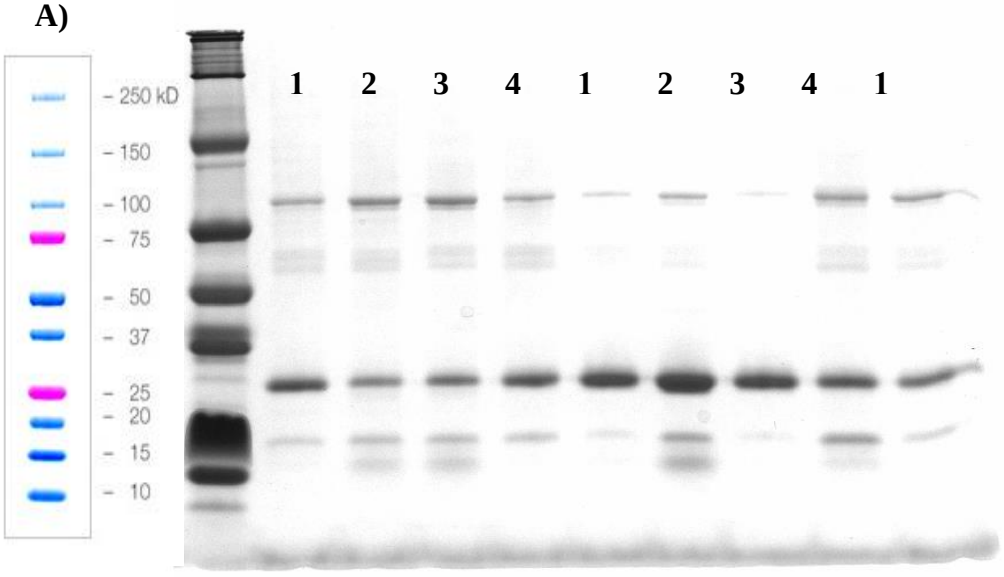
In order to increase the production yield, the addition of sucrose to the production media is another optimisation strategy. The introduction of sucrose can serve multiple purposes: (i) being an additional carbon to supplement energy supply, (ii) stimulating cell proliferation, therefore increased cell growth (Bao et al. 2016), (iii) creating favourable environment for increased protein expression, (iv) exerting osmo-protective effect on reducing cellular stress, (v) enhancing protein folding and stability. By exerting its positive effects on multiple magnitudes, addition of sucrose to the growth media hold a great potential as an optimisation strategy on increasing production yields.

Table 115. The impact of incorporating sucrose into the media for recombinant protein production to enhance the concentration was investigated. 2M of sucrose was introduced at the 24th h of the production cycle; by considering the total production volume not exceed the previously defined limit. The resulting concentration measurements were calculated as mean values derived from a minimum of three distinct flasks employed within the same production batch or gathered from various other batches and presented with standard deviation. In order

to ensure the reliability of the results, the experimental procedure was repeated a minimum of three times.

sdAb Sample	Sucrose (2M)	Concentration (mg/L culture)
WT	No	0.68 ± 0.04
	Yes	4.03 ± 1.51
L45R	No	2.42 ± 0.53
	Yes	2.71 ± 0.60
W47G	No	1.67 ± 0.43
	Yes	2.18 ± 0.41
G44E	No	1.75 ± 0.29
	Yes	4.22 ± 1.39
V37F	No	7.72 ± 1.11
	Yes	8.87 ± 1.80
V37Y	No	8.51 ± 0.51
	Yes	9.24 ± 0.42
G44E/L45R/W47G	No	1.18 ± 0.27
	Yes	2.13 ± 0.39
V37F/G44E/L45R/W47G	No	1.65 ± 0.71
	Yes	2.38 ± 0.60

A)



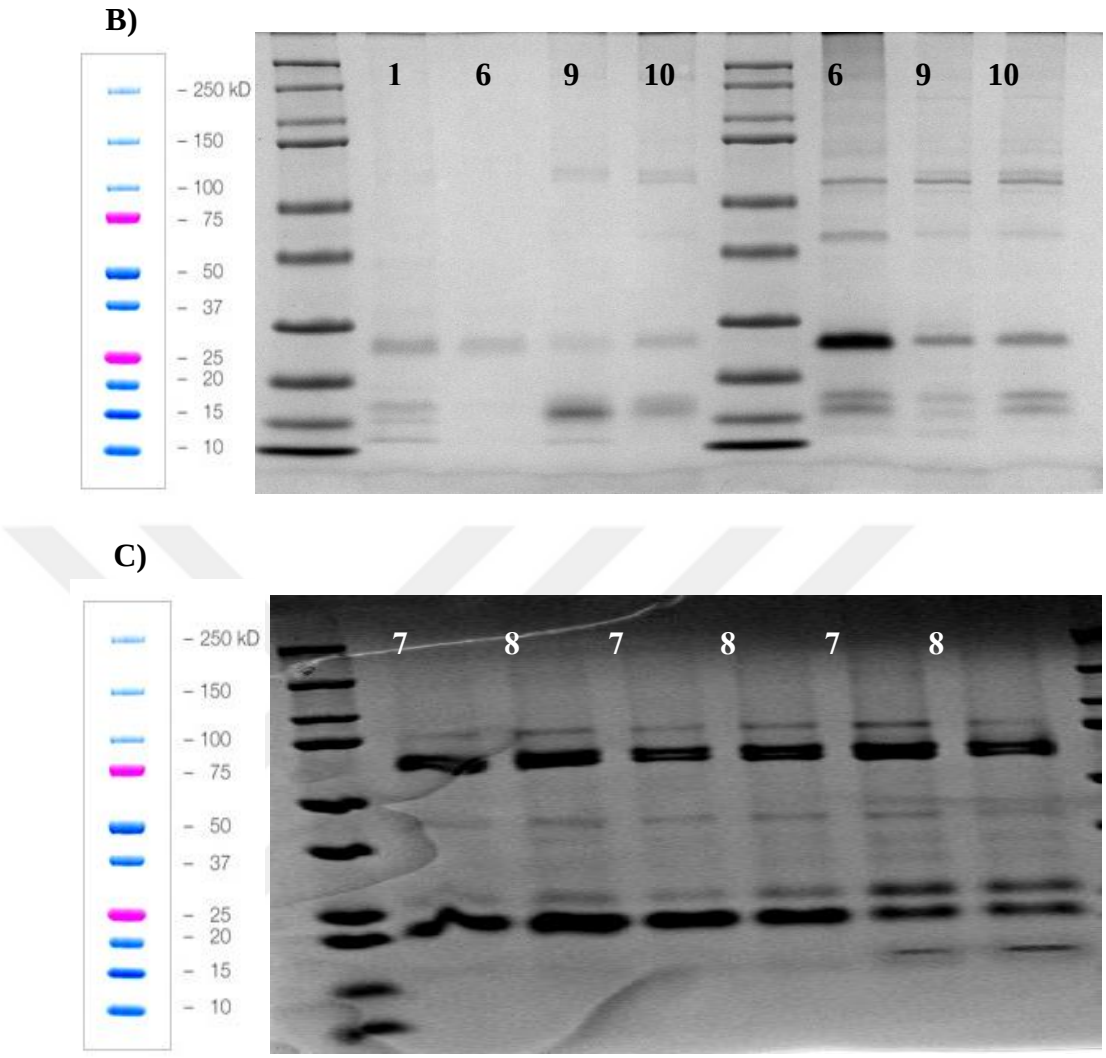


Figure 39. The impact of sucrose incorporation into the production medium was examined. The SDS-PAGE results of different sdAbs (**A**, **B**, **C**) were assessed. SDS-PAGE gels were loaded with 1000 ng of sdAbs. The gels were stained with Coomasie Blue dye. The numbers on the lanes represents the sdAb code. The complete nomenclature of the introduced mutations is based on **Table 107**. (**A)** The set of sdAbs1-4 were cultured under standard conditions. The second following set of sdAbs 1-4 were exposed to 2M sucrose at the 24th h of production. **(B)** The first set of sdAbs 1, 6, 9, 10 ere cultured under standard conditions. The following set of sdAbs after the ladder represents sucrose added sdAbs 6, 9, 10. **(C)** The first set of sdAbs 7 and 8 were cultured under standard conditions, whereas the samples in the last two lanes were cultured in the presence of 2M sucrose into the production medium at the 24th h of the production.

While the addition of 2M sucrose to the production media resulted in a substantial increase in recombinant protein concentrations (**Table 115**), the accompanying compromise in the quality of sdAb variants (**Figure 39**) raised significant concerns. The elevated impurity rates observed in the SDS-PAGE results indicate that the improved yield may be attributed to an increase in the presence of contaminants within the production medium. This compromise in sdAb purity affected the overall quality of the sdAbs. Despite the initial advantages, this trade-off questions the incorporation of sucrose to the production media.

4.3.12. Harvesting recombinant protein from different cellular compartments

In the optimisation of recombinant protein production, the extraction of proteins from various cellular compartments is another strategy. One approach involves harvesting proteins from the supernatant, also known as leaky expression. Leaky expression refers to a phenomenon where some recombinant protein molecules are released into the culture medium when cells are dividing. This method takes advantage of the protein secretion and diffusion into the surrounding environment, allowing it to accumulate in the culture medium. And then, from the protein-rich supernatant the protein of interest can be isolated. This technique offers a relatively easier means of protein collection. The other approach involves the extraction of proteins from the periplasm of *E. coli*. The periplasm is the space between the outer and inner membranes of the bacterial cell envelope. It is known for its suitable environment for proper folding and post-translational modifications of the proteins. To extract proteins from the periplasm, an additional step is required. For this purpose, TES buffer is used to disrupt the bacterial cell membrane by exerting osmotic shock effect, meanwhile facilitating the release of the periplasmic contents.

Table 116. The variation in concentrations of recombinant proteins sourced from different cellular compartments is shown. Leaky expression is obtained from the supernatant, whereas periplasmic extract is acquired from the periplasmic space of *E. coli* by employing the TES buffer. TES buffer-based extraction involves two rounds of treatment to enhance the overall yield. The resulting concentration measurements were calculated as average values derived from at least three distinct flasks employed within the same production and presented with standard deviation.

sdAb Sample	Cellular compartment	Concentration (mg/L culture)
F70I/L72R/T74N/S77N/A79L	Leaky expression	0.12 ± 0.07

	Periplasmic extract	0.49 ± 0.30
V37Y/G44E/L45R/W47G/ F70I/L72R/T74N/S77N/A79L	Leaky expression	0.77 ± 0.07
	Periplasmic extract	0.43 ± 0.12

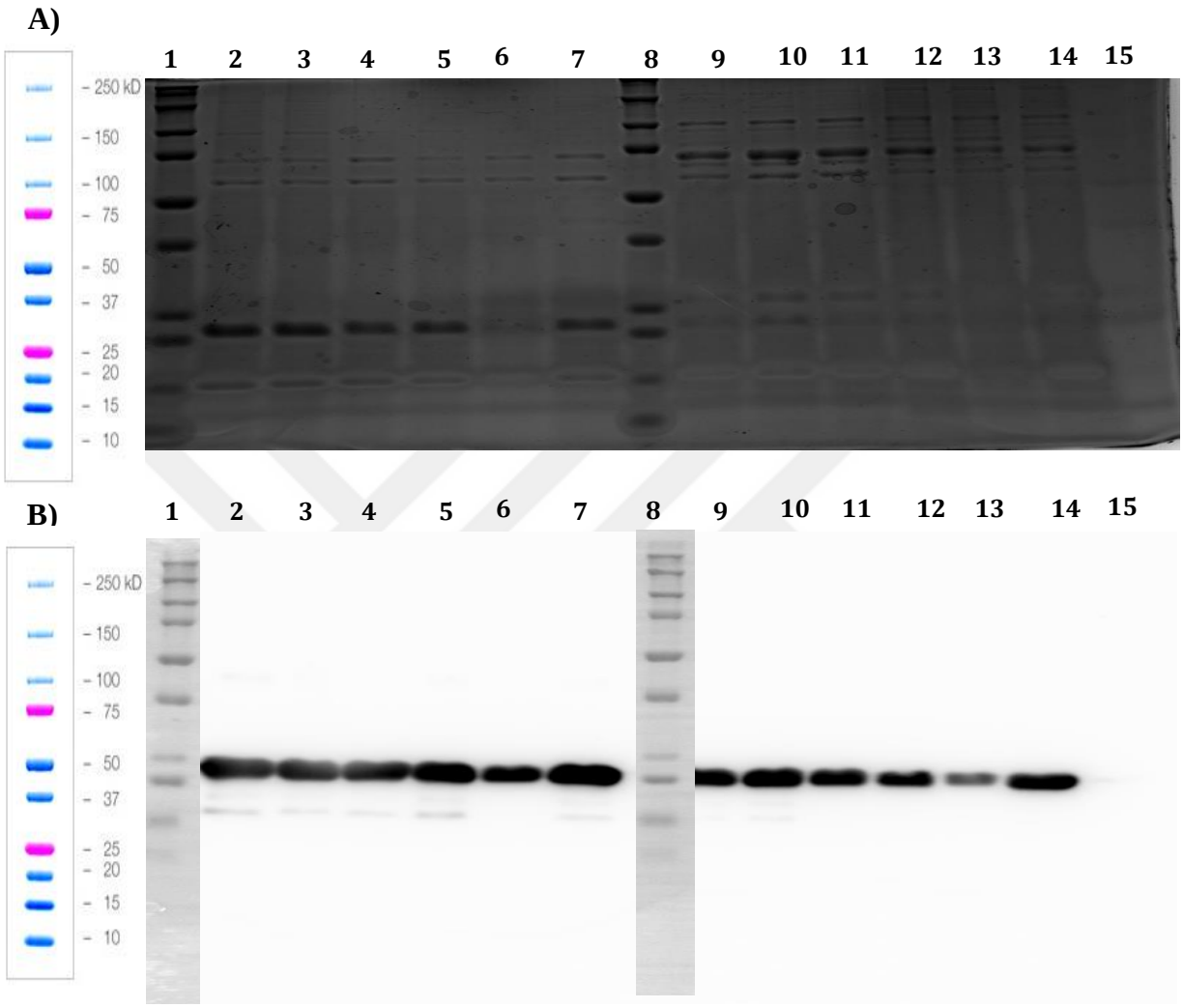


Figure 40. The comparison of sdAb isolated from distinct cellular compartments. V37Y/G44E/L45R/W47G/F70I/L72R/T74N/S77N/A79L sdAb was assessed in two parallel analysis. The first subset of samples, till the second ladder (2-7), originated from the supernatant of the protein production media (leaky expression). The second subset of samples (9-14) was collected from the periplasmic space of *E. coli* through osmotic shock facilitated by two rounds of TES buffer treatment. **A)** In SDS-PAGE, protein samples were loaded at a concentration of 1000 ng, and the isolated proteins were visualized using Coomassie Blue dye staining. **B)** For the WB, loaded samples were 100 ng and they were probed with monoclonal

Anti-Flag M2 IgG generated in mice (1:5000), followed by treatment with polyclonal HRP-Conjugated secondary Goat Anti-mouse IgG (1:10000).

Based on the evaluation of concentration results (**Table 116**), complemented by SDS-page and WB analyses (**Figure 40**), the advantages and trade-offs associated with different recombinant protein expression methods were evaluated. The utilization of the TES buffer-facilitated periplasmic extraction stands out as a preferable option, resulting in increased quantities of proteins. However, when considering leaky expression, which, in certain instances, showed better yield profiled for some sdAb variants. While leaky expression offers higher quantities of the proteins, the trade-off is the potential compromise of protein folding and quality.

4.3.13. Different agents for IMAC

One other approach for optimisation of recombinant protein production is releated with selection of appropriate IMAC agents. Thermo, GE and GoldBio branded Ni-NTA beads offer unique characteristics for this purpose. Each manufacturer recommends specific buffer formulations for IMAC. These buffers facilitate optimal interaction between the metal ions and the target protein. Additionally, GoldBio requires reducing a reducing agent in the buffer formulation, where we used DTT (1M).

Table 117. Various IMAC agents were utilized in the process of isolating recombinant target proteins. In each case, the suitable buffer was used according to the specific brand used. The resulting concentration measurements were determined as average values, obtained from a minimum of three separate flasks employed within the identical production batch, or sourced from different batches and presented with standard deviation. For result consistency, the experimental procedure was repeated a minimum of three times.

sdAb Sample	IMAC agent	Concentration (mg/L culture)
F70I/L72R/T74N/S77N/A79L	GE	0.40 ± 0.13
	Thermo	0.84 ± 0.06
	GoldBio	1.48 ± 0.15
V37Y/G44E/L45R/W47G/ F70I/L72R/T74N/S77N/A79L	GE	0.07 ± 0.02
	Thermo	0.87 ± 0.13
	GoldBio	0.99 ± 0.19

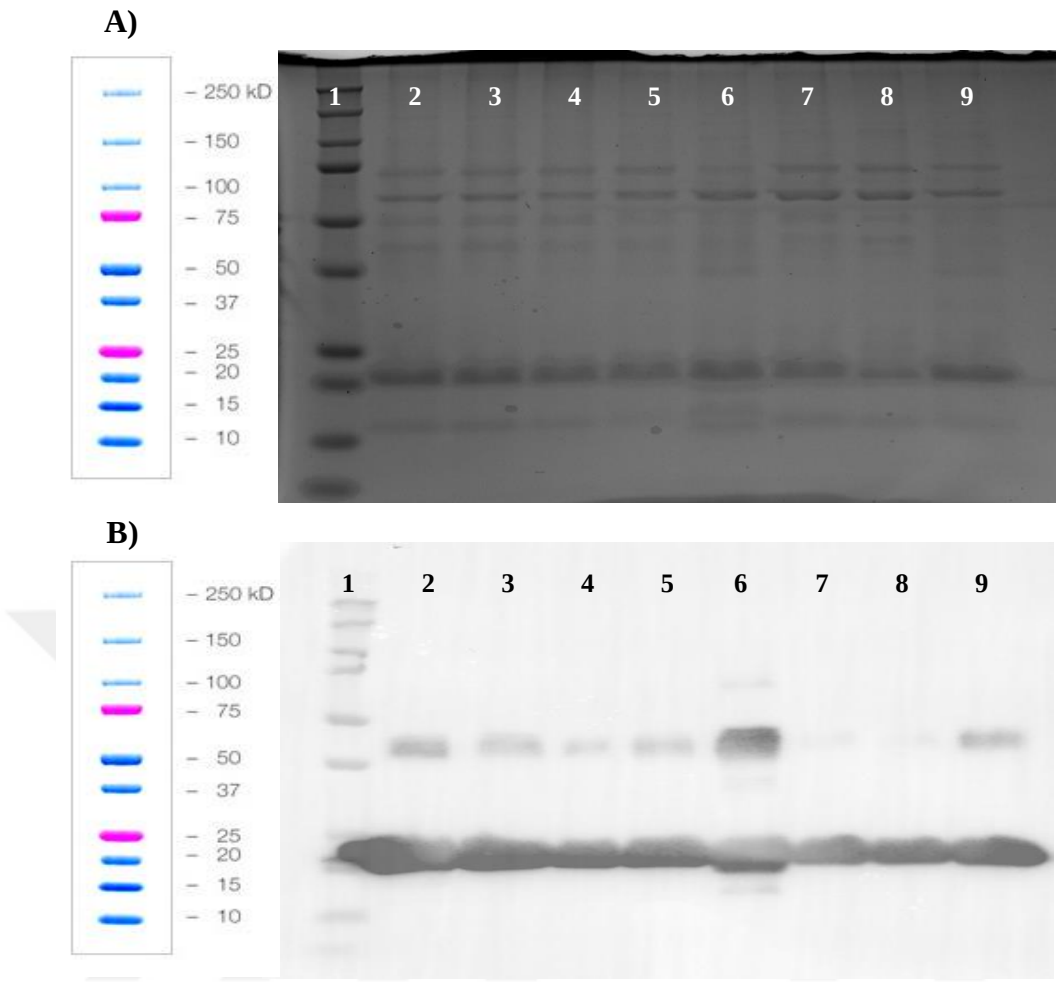


Figure 41. The effect of using different IMAC agents on F70I/L72R/T74N/S77N/A79L sdAb sample assessed by **A)** SDS-PAGE and **B)** WB analysis. All the samples were the cultured in the same environmental conditions. The sample on the 6th lane was treated with Goldbio beads, where the other flasks were added equal amount of Thermo Ni-NTA beads. For SDS-PAGE sample concentration 1000 ng. For the WB analysis, loaded samples were at 100 ng and they were probed with monoclonal Mouse Anti-Flag M2 IgG (1:5000), followed by treatment with polyclonal HRP-Conjugated secondary Goat Anti-mouse IgG (1:10000).

Based on the results, Goldbio beads demonstrate optimal performance under the reducing conditions provided by DTT (**Table 117**). In the WB analysis, the upper possible dimer band appeared thicker when compared to the other lanes (**Figure 41**).

During the trials, different IMAC bead incubation durations and environments were also explored: overnight shaking in the cold room, overnight shaking in the incubator, overnight shaking in the fridge and shaking at RT for a few hours. These incubation times exhibited minimal impact on the resulting yields (data not shown). Since the results are not

affected by incubation duration, this feature was exploited when planning the overall experimental timeline.

Another consideration regarding the utilization of the IMAC beads is using fresh stock or regenerated beads. While fresh beads offer high binding capacities due to the pristine nature of their metal ions, regenerated beads, are cost-effective and may exhibit slightly diminished binding capacities due to potential metal ion bleaching.

4.3.14. Choice of purification method

Various purification methods (batch purification, vacuum filtration and centrifuge purification) were employed to optimize recombinant protein production, each offering distinct advantages and disadvantages. Batch purification uses relatively straightforward approach. However, this method can introduce variability in purity due to manual handling of wash steps. Vacuum filtration relies on negative pressure, removing impurities while retaining the target protein, but there is a risk of protein denaturation due to mechanical stress. Centrifuge purification, separates proteins based on density and offers high yield and purity levels. Although it requires longer processing times, ability of processing multiple samples in parallel proves an advantage.

Table 118. The effect of using different purification strategies on **WT** sdAb. The average concentration measurements were calculated from at least three separate flasks used in the same production batch or from other batches and presented with standard deviation. To ensure result consistency, the experimental procedure was replicated at least three times.

sdAb Sample	Purification method	Concentration (mg/L culture)
WT	Batch purification	0.05 ± 0.02
	Vacuum filtration	0.18 ± 0.11
	Centrifuge purification	1.33 ± 0.16

Overall, centrifuge purification is the best purification method compared in parallel experimentation (**Table 118**).

4.4. Optimizing the recombinant protein production purity

Achieving high levels of purity in recombinant protein production is a fundamental for their use in downstream applications.

4.4.1. Eliminating the non-specific interactions

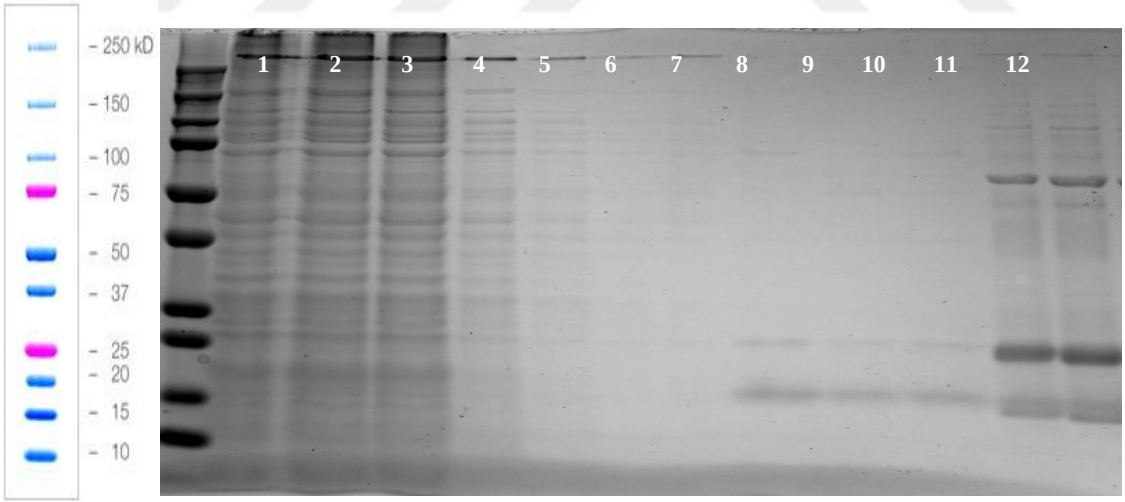
The recombinant protein production optimisation approach involves the introduction of specific additives during bead incubation, such as MgATP, *E. coli* cell lysate or a

combination of both (Kwon and Jewett 2015). These additions play role in preventing the attachment of contaminants to the IMAC beads. By incorporating MgATP, potential non-specific interactions would be disrupted. Similarly, addition of *E. coli* cell lysate, further enhances selectivity of the beads by competing with contaminants. The synergistic effect can be achieved through the combined use of MgATP and *E. coli* cell lysate to improve purity status of the sdAbs.

Table 119. The effect of different additives introduced to the media during bead incubation to enhance the purity of the F70I/L72R/T74N/S77N/A79L sdAb was investigated. Concentration measurements were calculated as mean values, derived from at least three distinct flasks used within the same production batch and presented with standard deviation.

sdAb Sample	Additions	Concentration (mg/L culture)
F70I/L72R/T74N/S77N/A79L	MgATP	0.99 ± 0.34
	E. coli cell lysate	1.26 ± 0.12
	MgATP + *E. coli* cell lysate	1.67 ± 0.53

A)



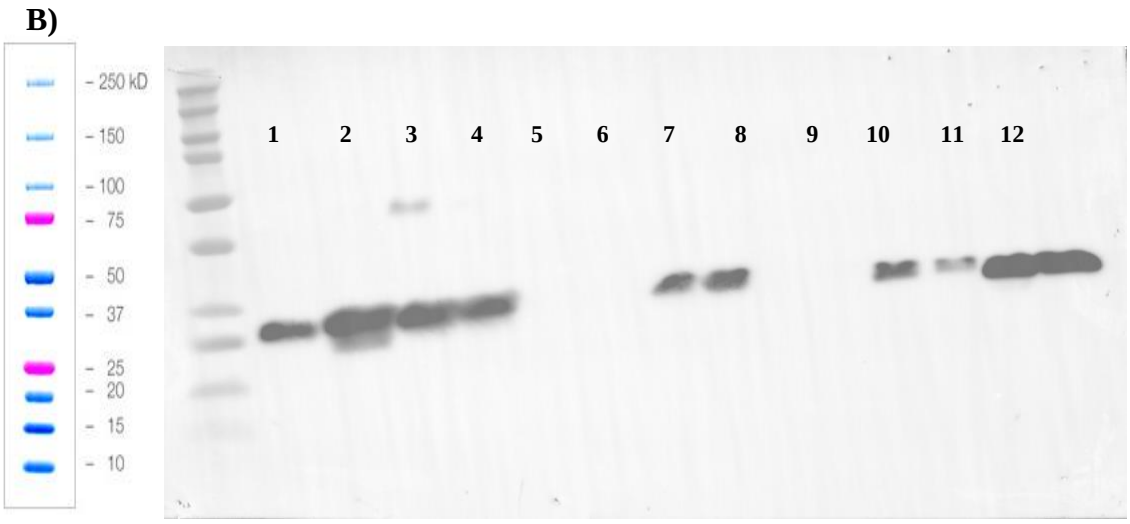


Figure 42. The impact of MgATP and *E.coli* cell lysate supplementation on the purity of leaky expression was investigated on F70I/L72R/T74N/S77N/A79L sdAb. Two parallel gels were run, with the only distinction was protein amounts: **A)** SDS-PAGE samples were at 1000 ng, whereas **B)** WB samples were at 100 ng. Coomassie Blue dye was used for protein staining in SDS-PAGE, while WB, the sdAbs were probed with monoclonal Anti-Flag M2 IgG generated in mice (1:5000) and polyclonal HRP-Conjugated secondary Goat Anti-mouse IgG (1:10000). The initial three sample lanes (1-3) were loaded with the flow-through. The next set of triple samples (4-6) have the discarded liquid collected from the Eppendorf tubes' bottom after treatment with wash buffer, (7-9) loaded with remainder of secondary wash performed using supplemented Buffer A (25 mM imidazole + 5 mM MgATP + 0.6 mg/ml denatured *E.coli* lysate). The last three lanes (10-12) included elution samples from each individual flask.

The utilization of MgATP, *E. coli* cell lysate, or a combination of both did not yield significant improvements in the results (**Table 119** and **Figure 42**). Considering the *E. coli* cell lysate preparation process and the overall increase in time and effort associated with this additional step, was not fruitful to pursue further.

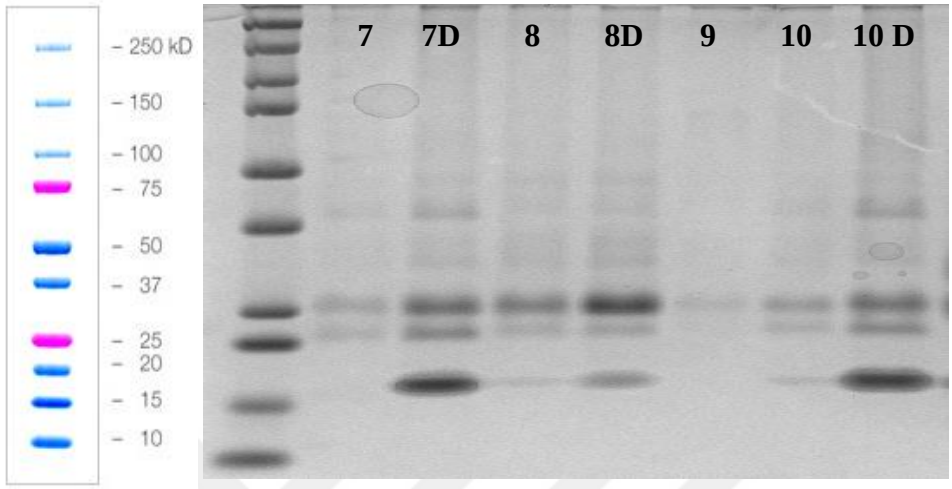
4.4.2. Unfold/refold

The optimisation of recombinant protein production involves the utilization of dialysis for the unfolding and refolding of proteins, whether in their IMAC bead-bound state or after the completion of the purification procedure.

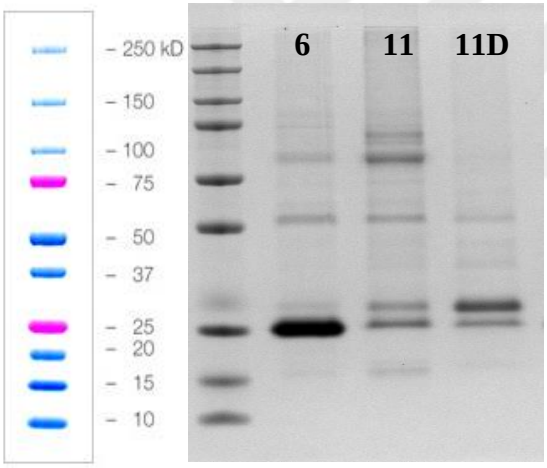
In this technique, the unfolded protein samples, are treated with imidazole and GdmCl, and then introduced into a dialysis membrane, submerging it in a buffered solution. The gradual exchange of the surrounding solution with fresh buffer allows for controlled

removal of denaturants, facilitating the refolding of the protein. The dialysis applied overnight, while gentle stirring the membrane. After the overnight buffer exchange, the membrane contents were filtered and further purified.

A)



B)



C)

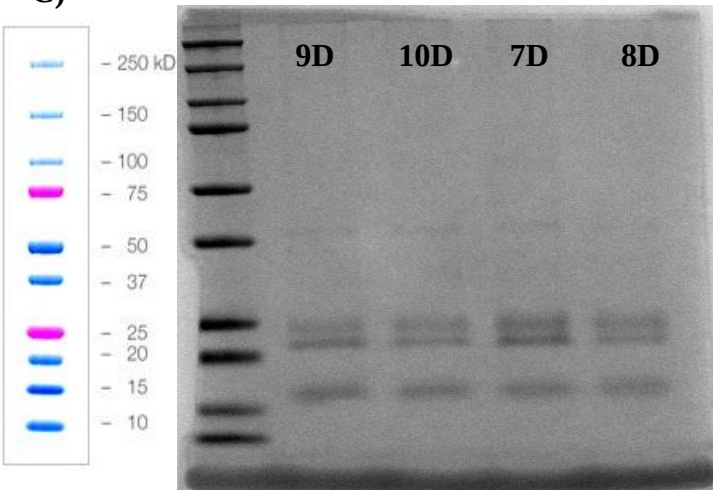


Figure 43. The impact of dialysis on the unfolding and refolding processes during the purification of sdAbs was investigated. SDS-PAGE analysis was performed on different sdAb samples (**A**, **B**, **C**), which were loaded as 1000 ng and stained using Coomassie blue dye. The sdAbs were isolated using the conventional purification method and a dialysis-based purification procedure. (D: sdAb sample after dialysis treatment). Refer to **Table 107** for full nomenclature of the sdAbs.

Upon evaluating the quality through SDS-PAGE analysis (**Figure 43**), it was observed that the introduction of the dialysis step intensifies the thickness of the lower-tier contaminants, potentially representing truncated versions of sdAbs. This issue may arise due to the selective removal of molecules during dialysis, exaggerating the presence of lower weight impurities trapped within the dialysis membrane with a 10 kDa cut-off. However, there was a reduction in the upper-tier contaminants on the gel, possibly indicating the effective removal of larger molecules during the dialysis.

The shift in the balance between lower and upper contaminants, and the increased band thickness at the expected molecular weight of sdAb, can be explained by the principles of the dialysis. If the upper contaminants are present (chaperons like GroEL/ES) (Ryabova et al. 2013), which frequently play a role in unfolding and refolding processes, the treatment with denaturing agent GdmCl disrupts the 3-D structure of the chaperon-smaller molecule complex. Consequently, these two molecules are separated in the denaturing environment. Given the 10 kDa cut-off of the membrane, any molecule larger than this is trapped within the membrane along with sdAbs, while smaller molecules below 10 kDa are released into the surrounding buffer (PBS). However, further purification is necessary for the selective extraction of sdAbs in the purest form possible. During this re-purification step, the mixture of denatured proteins trapped within the membrane washed rigorously. During this stage, there may be a tendency for the previously separated molecule couples to reunite as refolding occurs. It is important to note that the upper and lower band contaminants do not completely disappear due to the refolding process.

As an additional step, re-incubation with IMAC beads could enhance the quality of dialysis outcomes. However, the practicality of this approach is limited by small quantities of sdAbs. This limitation is caused from both the loss of sdAb volume using dialysis and the manual dexterity required for the procedure. Despite the potential benefits, the current results prevent the further exploration of this optimisation strategy. Overall, the dialysis treatment is

a laborious step, demanding a lot of manual adjustments due to the delicate nature of working with the membrane. Taking into account these considerations, dialysis would not be preferred to improve sample quality of the future experiments.

4.4.3. Elimination of contaminant through chromatography techniques

The recombinant protein production was resumed with the optimisation of chromatography techniques to improve purity.

4.4.3.1. Purification with Protein A

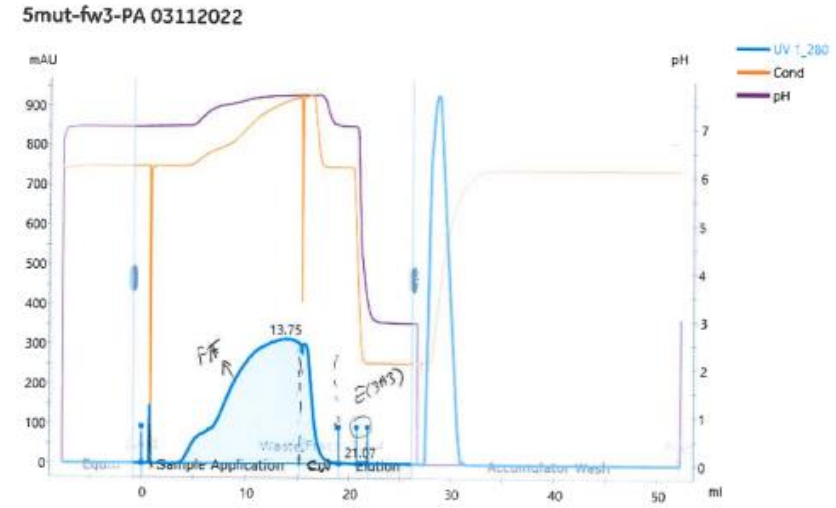
The utilization of FPLC is an effective separation method of contaminants. Before using Protein A in FPLC, compatibility of the resin with Bevacizumab V_H sequence was investigated (**Table 120**).

Table 120. The compatibility analysis of Protein A application on the V_H of Bevacizumab sequence. The important residues in binding of Protein A to its target were listed. The Bevacizumab V_H sequence (sdAb row) and the essential and non-essential residues required for Protein A binding (V_H row) were aligned based on both IMGT and Kabat numbering. The essential residues were shown in **bold**. The six non-essential residues for Protein A binding were underlined. The matching essential residues were highlighted in yellow, while the non-essential matches were highlighted in grey. The CDR2 region residues were marked with asterisks. This analysis was based on a publication by Crauwels et al. (Crauwels et al. 2020).

IMGT	**16**	**18**	**20**	**65**	**67**	**72**	**74**	**75**	**77**	**79**	**90**	**92**	**93**
V_H	G	S	**R**	T	Y	K	**G**	**R**	**T**	**S**	**Q**	**N**	S
sdAb	G	S	R	P*	Y*	K*	R*	R	T	S	Q	N	S
Kabat	**15**	**17**	**19**	**57**	**59**	**64**	**65**	**66**	**68**	**70**	**81**	**82A**	**82B**

Among these residues, seven essential amino acids were identified as important for Protein A binding, while six non-essential amino acids were also relevant. The comparison of these residues with the Bevacizumab V_H structure revealed that 6/7 essential residues and 5/6 non-essential residues for Protein A binding were present in the Bevacizumab V_H sequence. However, it was specified in the manufacturer manual that recombinant Protein A structure was modified to remove non-essential residues.

Protein A purification was commenced using the Protein A resin from Genscript. The procedure was applied based on manufacturer instructions and was employed prior to the standard purification process, replacing with the purification step using Ni-NTA beads.



Peak Table

Peak	Retention (ml)	Area (mAU*ml)	Area %	Ext.coeff. (mg ml⁻¹ cm⁻¹)	Fraction(s)	Volume (ml)	Conductivity (mS/cm)
Peak A	13.745	2698	99.87		1.A.1 - 1.A.2	19.007	19.73
Peak B	21.068	3.595	0.13		3.A.3	1.064	8.05

Created 03-Nov-22 12:39:02 AM -07:00 by Default on AVANT150
Printed 03-Nov-22 1:31:56 AM -07:00 by Default
\Default\Home\Columns-Chromatography Methods and Results\Protein A MabSelectSureHiTrap1mL\

Figure 44. The results of FPLC using Protein A as resin. The method was performed for the purification of the F70I/L72R/T74N/S77N/A79L sdAb sample prior to doing standard purification with IMAC. This purification was conducted using a Hi-Trap 1 ml MabSelect Sure column on the GE AKTA Avant HPLC machine. The initial column calibration was done with phosphate buffer prepared diluting 10X PBS at pH 7.6. The elution buffers: 75 ml of 1N Tris-HCl (pH 8.0), 50 mM $NaNH_3$ (pH 3.0). The Protein A resin (Genscript) was used to capture the target sdAb. Retention Time (min), UV 280 is shown with the blue line, and corresponds to the elution of the target sdAb. Conductivity is shown with the red line, and reflects the elution of the target protein. pH is shown with the purple line, and shows the elution of the target protein. The initial sections of the graph represent the baseline correction.

Right after the FPLC run, the resulting graph was assessed (**Figure 44**). The fractions that were likely to constitute the purified version of the sdAb sample were separated and some were combined. Prior to the FPLC, a small portion of the sample was spared. The flow-through liquid from the waste and the liquid that passed through the column during the wash steps were collected for the measurement of remaining protein concentration (**Table 121**). These fractions were assessed by SDS-PAGE (**Figure 45**).

Table 121. The fractions collected after FPLC and their corresponding pH and concentration values for F70I/L72R/T74N/S77N/A79L sdAb. Each row in the first column corresponds to a fraction obtained during the FPLC process. Load: sdAb sample spared before analysis, Flow-through: collected from the waste, Wash: collected after wash step. The eluate fractions: collected during successive rounds of purification via FPLC. For concentration measurement, each fraction was blanked with its corresponding buffer.

Fraction	pH	Concentration (mg/ml)	Volume (µl)
3A3	7.6 (increased from 4.2)	0.021	500
3A2	8.0	0.005	500
3A1	8.3	0.012	500
Wash	-	0.202	500
Load	8.1	0.733	500
Flow through	7.9	0.261	500

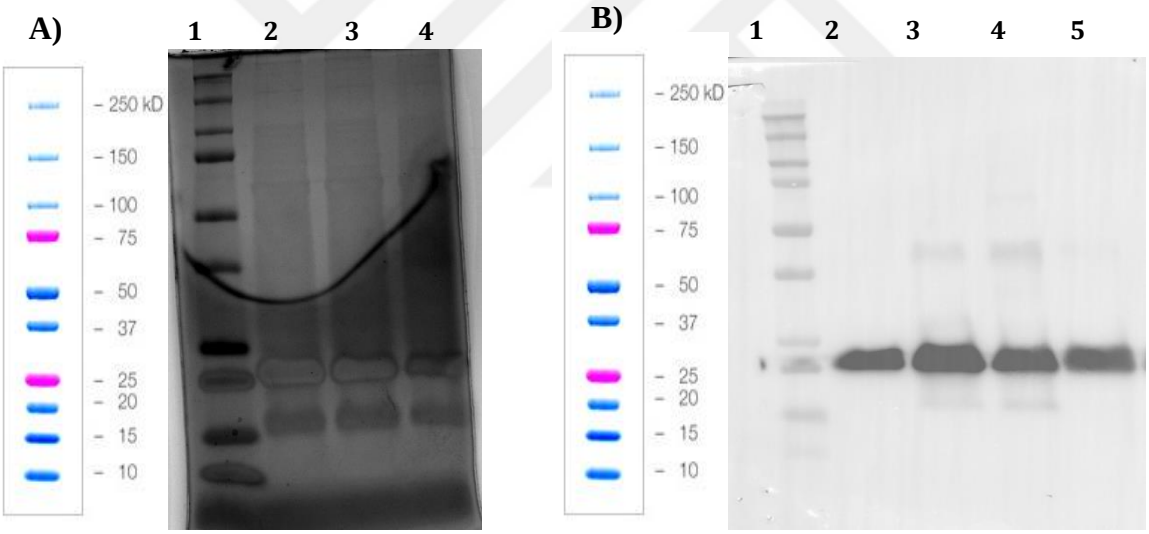


Figure 45. Protein A trial on F70I/L72R/T74N/S77N/A79L sdAb. **A)** The SDS-PAGE gel was run with samples were at 1000 ng, while **B)** WB samples were loaded at 100 ng. SDS-PAGE gel was stained with Coomassie Blue dye for visualization, whereas in WB, the sdAbs were probed using monoclonal Anti-Flag M2 IgG produced in mice (1:5000), followed by treatment with polyclonal HRP-Conjugated secondary Goat Anti-mouse IgG (1:10000). **A)** Following the ladder, the initial lane of SDS-PAGE gel was the Load. The following samples were the Flow-through and the Wash. **B)** The first lane was the Load, followed by the pooled the 3A1 and the 3A2 eluded fractions. The next lanes had the Flow-through and the Wash.

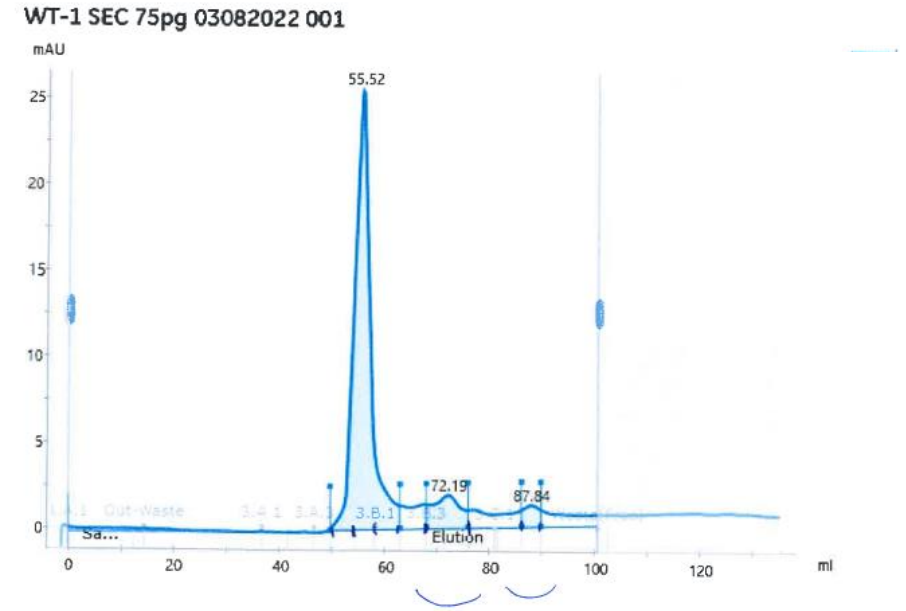
Following the FPLC procedure, an assessment of the quality and quantity of each fraction was performed (**Figure 45**). Despite observing optimal band intensity in both SDS-PAGE and WB analyses for the sdAb, the WB results suggest the presence of truncated sdAb fragments and dimers of sdAbs as reflected from the upper and lower extra bands beyond the expected molecular weight. Additionally, according to SDS-PAGE results, the lanes exhibited a smear pattern, indicative of suboptimal isolation quality. In conclusion, the utilization of Protein A via FPLC, as opposed to IMAC using Ni-NTA beads, does not emerge as a favourable alternative for enhancing purity.

4.4.3.2. Purification by gel filtration

Gel filtration is a chromatographic technique, which can be used in purification of proteins. This method, also known as size exclusion chromatography, based on the principle of differential molecular size to separate and purify protein samples.

For this purpose, the frozen sdAb WT samples were thawed on ice, preventing potential denaturation or aggregation. The concentrations of the pooled WT sdAbs were measured using NanoDrop spectrophotometer (**Table 122**) prior to the analysis (**Figure 46**).

A parallel procedure was executed for another sdAb sample, V37Y/G44E/L45R/W47G, following the same protocol (**Figure 47** and **Table 123**).



Peak Table

Peak	Retention (ml)	Area (ml*mAU)	Area %	Ext coeff. (mg ml⁻¹ cm⁻¹)	Fraction(s)	Volume (ml)
Peak A	55.524	87.64	84.75		3.A.4 - 3.B.1	13.332
Peak B	72.192	11.56	11.18		3.B.3 - 3.B.4	8.110
Peak C	87.842	4.213	4.07		3.C.3	3.832

Created 03-Aug-22 2:25:37 AM -07:00 by Default on AVANT150
Printed 03-Aug-22 5:49:55 AM -07:00 by Default
DefaultHome\Columns-Chromatography Methods and Results\SEC\Superdex 75 PG

Figure 46. The SEC results of WT sdAb, previously purified by IMAC. The sample was separated based on its molecular weight. The SEC assay was carried out as a service by the IBG Pharma department. The SEC assay utilized Superdex® 75 10/300 GL column. The run conditions during the assay were based on the instructions of the manufacturer.

Table 122. The fractions collected after SEC and their concentrations for sdAb WT. After completing the SEC process, the concentrated fractions of WT sdAb were measured using NanoDrop spectrophotometer.

Fraction	Concentration (mg/ml)	Volume (µl)
3A4	0.016	250
3A5	0.220	250
3B1	0.033	250
3B3	0.029	250
3C3	0.007	250

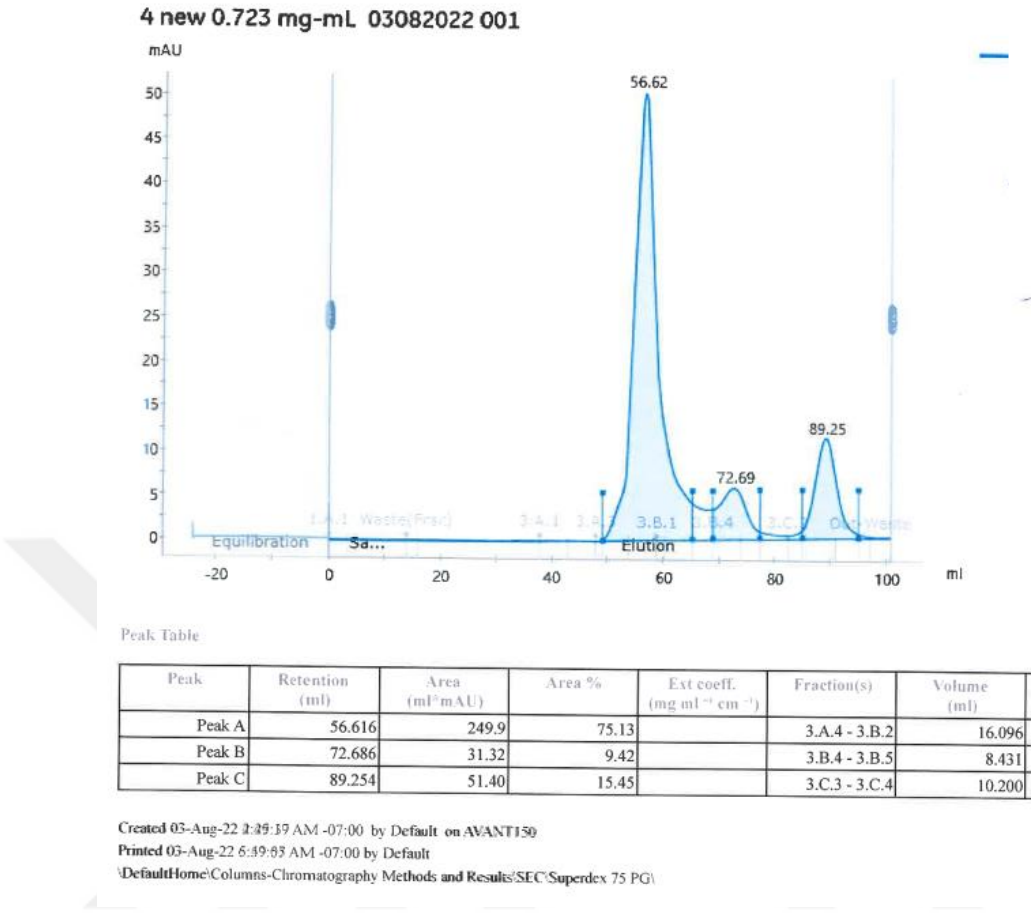


Figure 47. The SEC results for the V37Y/G44E/L45R/W47G sdAb, purified via IMAC,. The IBG Pharma department conducted the SEC assay as a service, for the separation of the sample according to molecular weight. Superdex® 75 10/300 GL column was employed for the assay. The run conditions followed the instructions of the manufacturer.

Table 123. The fractions collected after SEC and their concentration values for sdAb V37Y/G44E/L45R/W47G. Following the SEC procedure, the concentrated fractions of the V37Y/G44E/L45R/W47G sdAb concentration was measured using NanoDrop spectrophotometer. The individual concentrations of the fractions and the quantity of each collected fraction was listed.

Fraction	Concentration (mg/ml)	Volume (µl)
3A4	0.033	250
3A5	0.759	250
3B1	0.172	250
3B4	0.090	250
3C3	0.303	250

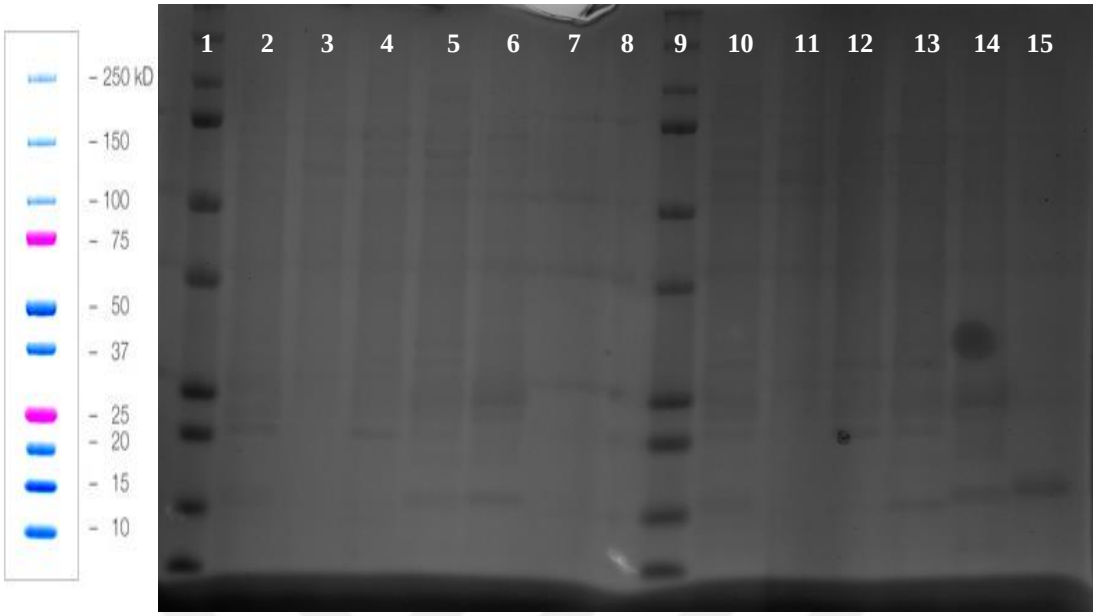


Figure 48. The fractions obtained after SEC separation analysed by SDS-PAGE. WT and V37Y/G44E/L45R/W47G sdAb samples wereloaded at 1000 ng. Staining was done using Coomassie blue dye. For the initial set of samples up to the second ladder, the arrangement reflects the order of separated WT sdAb fractions as indicated in **Table 122**. Exclusively, the first sdAb in the lane was WT sdAb spared prior to SEC analysis. The subsequent set of loaded sdAbs consisted of V37Y/G44E/L45R/W47G variants. Similarly, the first sample within this set was the analysed sdAb prior to SEC, with the subsequent samples were in the same order provided in **Table 123**.

Upon qualitatively evaluating the SEC results through SDS-PAGE *(***Figure 48***)*, it became apparent that the isolated fractions exhibited poor quality. Moreover, their concentration was found to be quite low, with only a 250 µl of volume. Consequently, this trial was inefficient in achieving a meaningful separation of sdAbs into purer fractions.

4.4.4. Different reducing conditions

In order to visualize the isolated sdAb samples on SDS-PAGE, different reducing conditions were tested: β-mercaptoethanol and DTT. Each of these chemicals has different treatment protocols for sample preparation for electrophoresis. β-mercaptoethanol operates as a thiol-based reducing agent, disrupting disulphide bonds and breaking tertiary structures. For β-mercaptoethanol the samples were boiled at 95°C for 5 min for denaturation. DTT, also targets disulphide bonds and helps in unfolding protein structures. For DTT, the samples were heated to 70°C for 10 min using a heat block.

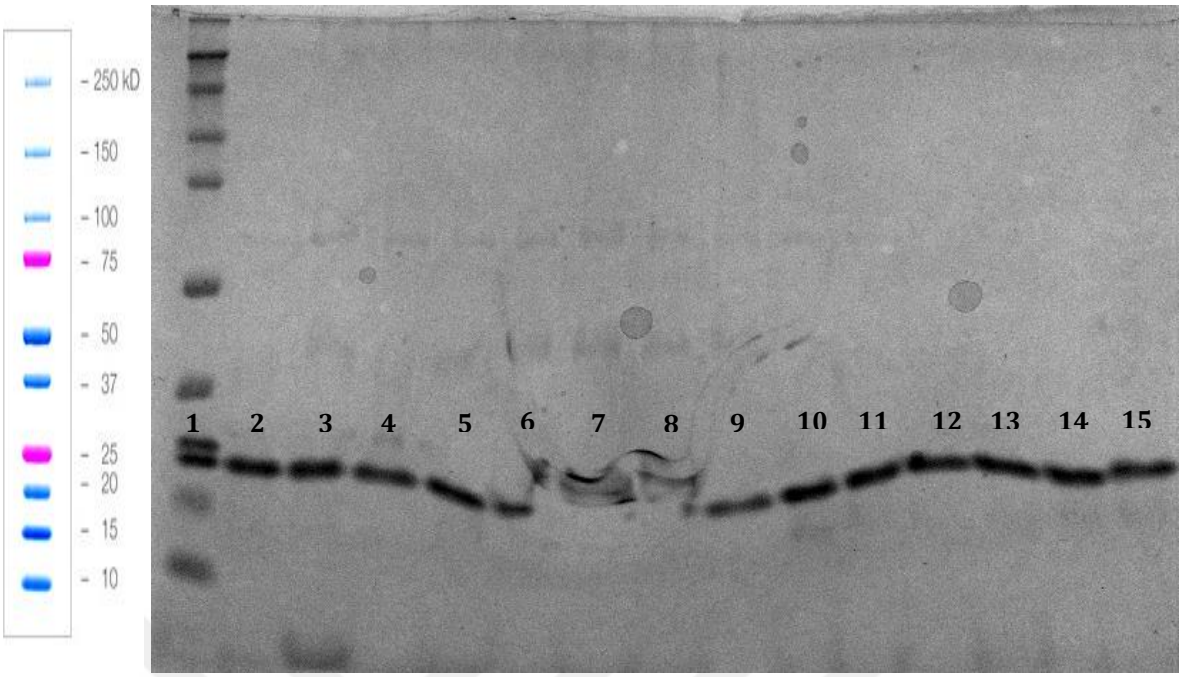


Figure 49. The SDS-PAGE analysis performed by using different reducing agents (β-mercaptoethanol and DTT) at varying concentrations, for the V37Y/G44E/L45R/W47G/F70I/L72R/T74N/S77N/A79L sdAb. The specific condition of each lane are detailed in the accompanying table provided below (**Table 124**).

Table 124. Various concentrations of β-mercaptoethanol and DTT were examined, through SDS-PAGE gel analysis, accompanied by corresponding physical treatments applied to the V37Y/G44E/L45R/W47G/F70I/L72R/T74N/S77N/A79L sdAb samples.

Sample lane	β-me concentration (mM)	DTT concentration (mM)	Boiling (°C/min)
1	-	-	-
2	355	-	95/5
3	710	-	95/5
4	1420	-	95/5
5	355	-	-
6	710	-	-
7	1420	-	-
8	-	-	95/5
9	-	-	-
10	-	50	70/10
11	-	100	70/10

12	-	200	70/10
13	-	50	-
14	-	100	-
15	-	200	-

This analytical approach allows for a comprehensive assessment of the sdAb under varying reducing conditions during electrophoresis. As anticipated, the absence of a reducing agent in the loading mix of the middle samples (lanes 8 and 9) during SDS-PAGE resulted in peculiar band patterns (**Figure 49**). The gel was divided into sections, with the left side containing β-mercaptoethanol and the right side containing DTT as the reducing agent, each with their specified boiling time and temperature (**Table 124**). The outcome of the SDS-PAGE run indicated that reducing the samples with DTT was a more effective strategy, even without boiling. The optimal condition was identified as a 100 mM DTT reduction, and this condition has been selected for future experiments.

4.5. Recombinant protein characterization

Both *in vitro* and *in vivo* recombinant protein characterization studies play a pivotal role in elucidating the properties and functionalities of engineered proteins. For the assessment of the following optimisation studies, SDS-PAGE and WB were employed as QC. In the *in vitro* characterization of sdAbs physicochemical features were investigated. The purest possible and the most promising sdAb candidates identified through *in vitro* assays were nominated fro *in vivo* studies using the zebrafish model to investigate their functionality.

4.5.1. ELISA

The indirect ELISA is a widely utilized technique. In this technique, a microplate is coated with the antigen of interest. Subsequently, a primary antibody is applied and binds to the immobilized testing sdAbs. Then, a secondary enzyme-conjugated antibody is added, which recognizes and binds to the primary antibody. The enzyme attached to the secondary antibody catalyses a colourimetric reaction when exposed to a substrate. The resulting signal intensity is directly proportional to the quantity of tested sdAbs.

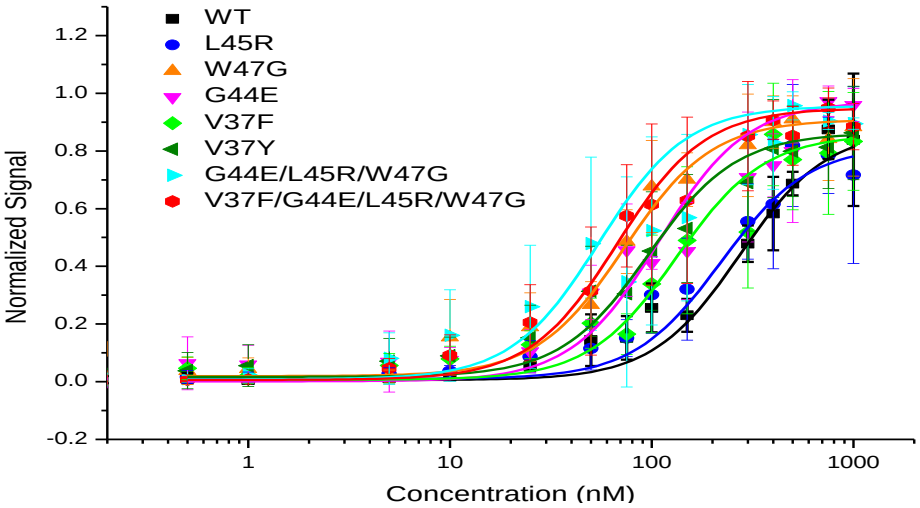


Figure 50. The apparent dissociation constant (K_d) values of WT sdAb and its variants calculated based on ELISA. Upon adding 100 µl of 2M H_2SO_4 to quench the reaction, peroxidase activity was measured at a wavelength of 450 nm. The obtained absorbance values ranging for 1 to 1000 nM concentration were background-subtracted, followed by normalization and the calculation of the IC_{50} value using a three-parameter logistic formula. This value provides the K_d values for the antibody variants. OriginPro 8.5.0 software was employed for this purpose. To ensure repeatability, experiments were conducted with a minimum of three replicates.

Table 125. The apparent dissociation constant (K_d) for binding of Bevacizumab sdAb-WT and its variants to VEGF was calculated using the Hill1 function in the OriginPro 8.5.0 software and standard errors were provided.

Bevacizumab VH variant	K_d (nM)
WT	273.39 ± 89
L45R	220.07 ± 80
W47G	73.14 ± 6
G44E	108.36 ± 10
V37F	132.84 ± 50
V37Y	96.01 ± 16
G44E/L45R/W47G	54.74 ± 24
V37F/G44E/L45R/W47G	67.29 ± 11

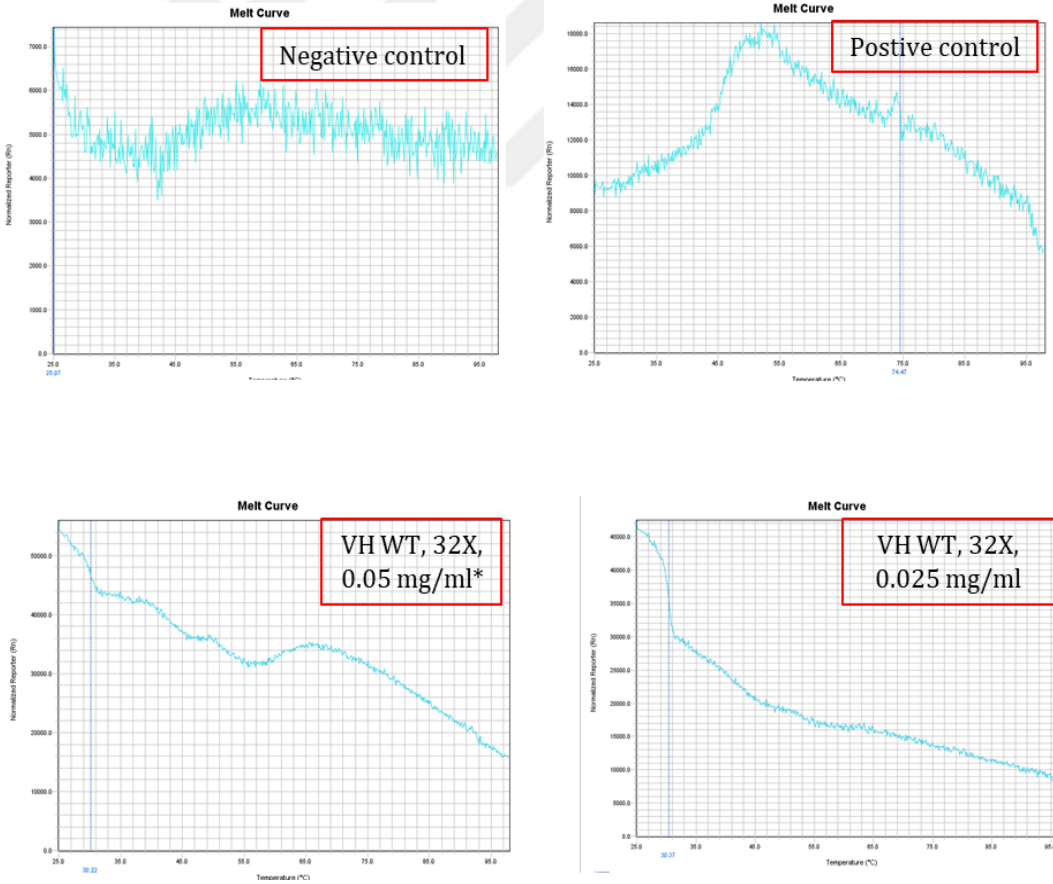
In antibody-antigen binding, a lower K_d indicates a stronger binding affinity. Hence the following order G44E/L45R/W47G > V37F/G44E/L45R/W47G > W47G > V37Y >

G44E > V37F > L45R > WT (**Figure 50** and **Table 125**). All the variants of Bevacizumab WT sdAb have lower K_d values compared to the WT. This implies that these variants exhibit improved binding characteristics and higher affinity for the target.

4.5.2. Thermal stability determination by protein thermal shift assay

Chemical stability assay investigates the integrity and functionality of the sdAbs. One widely applied method for chemical stability is the protein thermal shift assay using SYPRO™ orange protein dye.

The SYPRO™ Orange protein dye is a fluorescent dye. Its working principle is based on its ability to bind specifically to hydrophobic regions of proteins. When the dye binds to exposed hydrophobic patches on the protein surface, it undergoes a significant increase in fluorescence intensity. This increase in fluorescence can be monitored to observe protein unfolding or denaturation, as protein conformation changes in previously buried/hydrophobic regions.



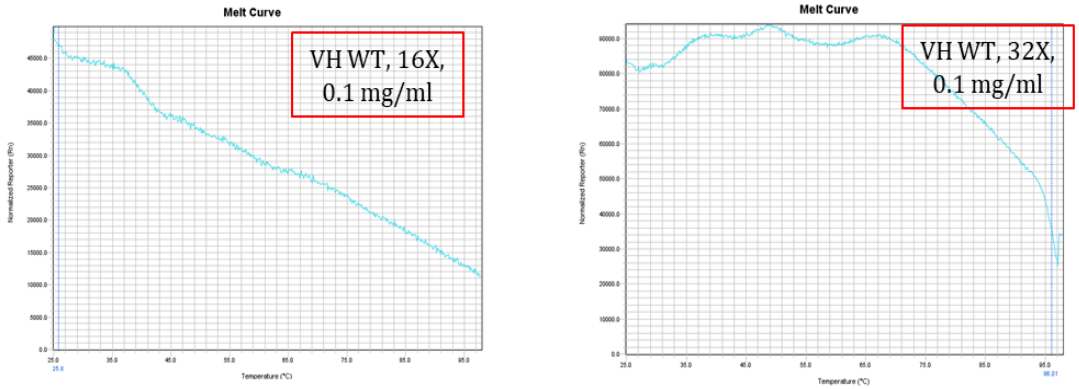


Figure 51. The derivative melt curves of WT sdAb under varying conditions for T_m determination. Different concentrations of SYPRO™ Orange dye (16X, 32X) as well as different protein concentrations were used to compare and optimize the experimental conditions. For each condition, triplicates of reactions mixtures were prepared, and all samples were run in parallel. Additionally, positive control (reaction mix with control ligand) and negative control samples (reaction mix with PBS) were included in triplicates as control reactions. The calculated melting temperatures were determined by the program software. The T_m (the temperature at which protein unfolding occurs) values are displayed along the y-axis in the middle point of unfolding.

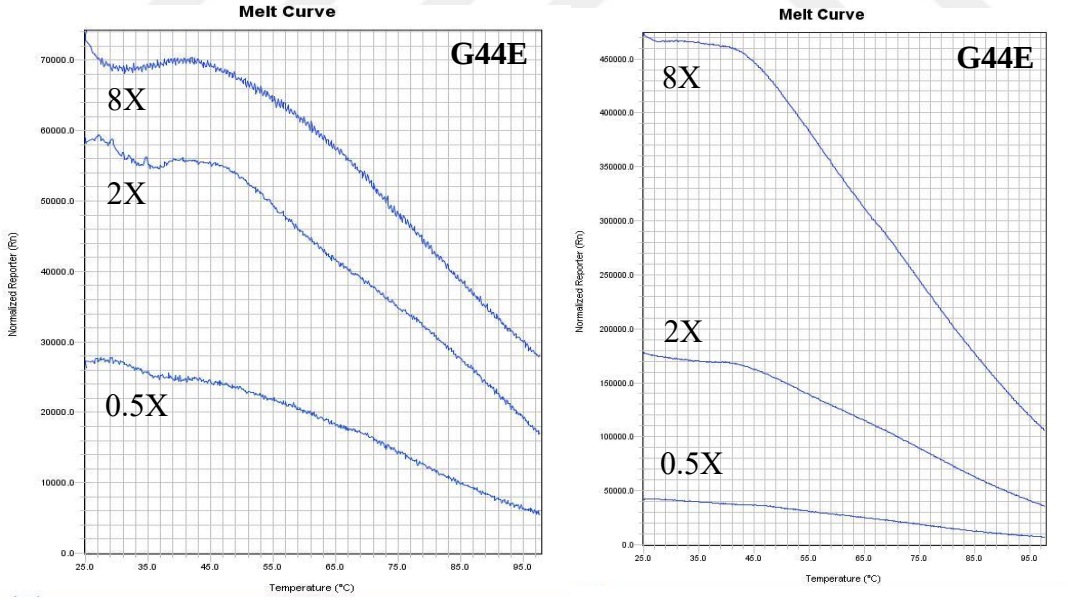


Figure 52. The derivative melt curve profiles of G44E sdAb under varying conditions for T_m determination. Different concentrations of SYPRO™ Orange dye (0.5X, 2X, 8X) were tested to optimize the experimental conditions. The first melt curve has a protein concentration of 2 µM, while the second melt curve protein concentration is 10 µM. For each condition, triplicates of reactions mixtures were prepared, and all samples were run in parallel.

Additionally, positive control and negative control samples were included in triplicates as controls (data not shown). The calculated melting temperatures were determined by the program software.

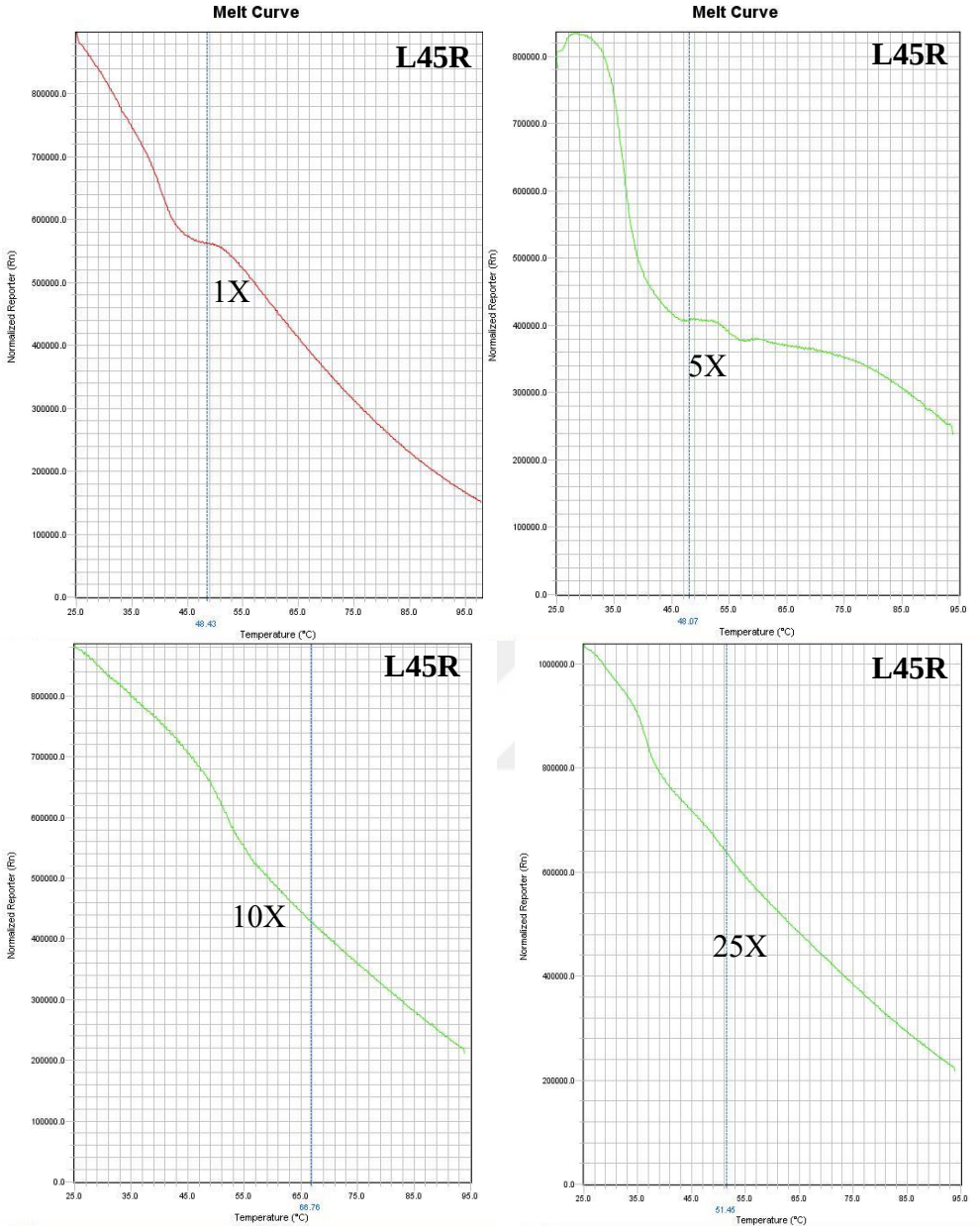


Figure 53. The derivative melt curve profiles of L45R sdAb under varying conditions for T_m determination. Different concentrations of SYPRO™ Orange dye (1X, 5X, 10X, 25X) were employed to optimize the conditions. The melt curves have 20 µM concentrations of samples. For each condition, triplicates of reactions mixtures were prepared, and all samples were run in parallel. Additionally, positive control and negative control samples were included in

triplicates to serve as controls (data not shown). The calculated melting temperatures were determined by the program software.

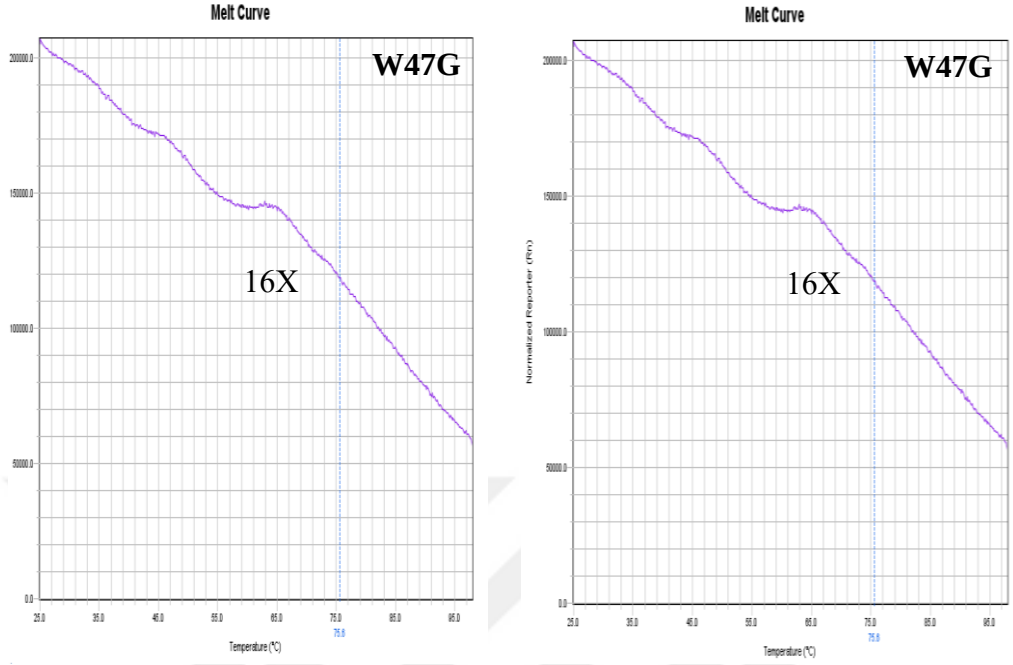


Figure 54. The derivative melt curve profiles of W47G sdAb under varying conditions for T_m determination. 16X of SYPRO™ Orange dye was employed for optimisation using variable sample protein concentrations: 2 µM (left) and 5 µM (right). For each condition, triplicates of reactions mixtures were prepared, and all samples were run in parallel. Additionally, positive control and negative control samples were included in triplicates to serve as controls (data not shown). The calculated melting temperatures were determined by the program software. The T_m values are indicated by blue dashed vertical lines.

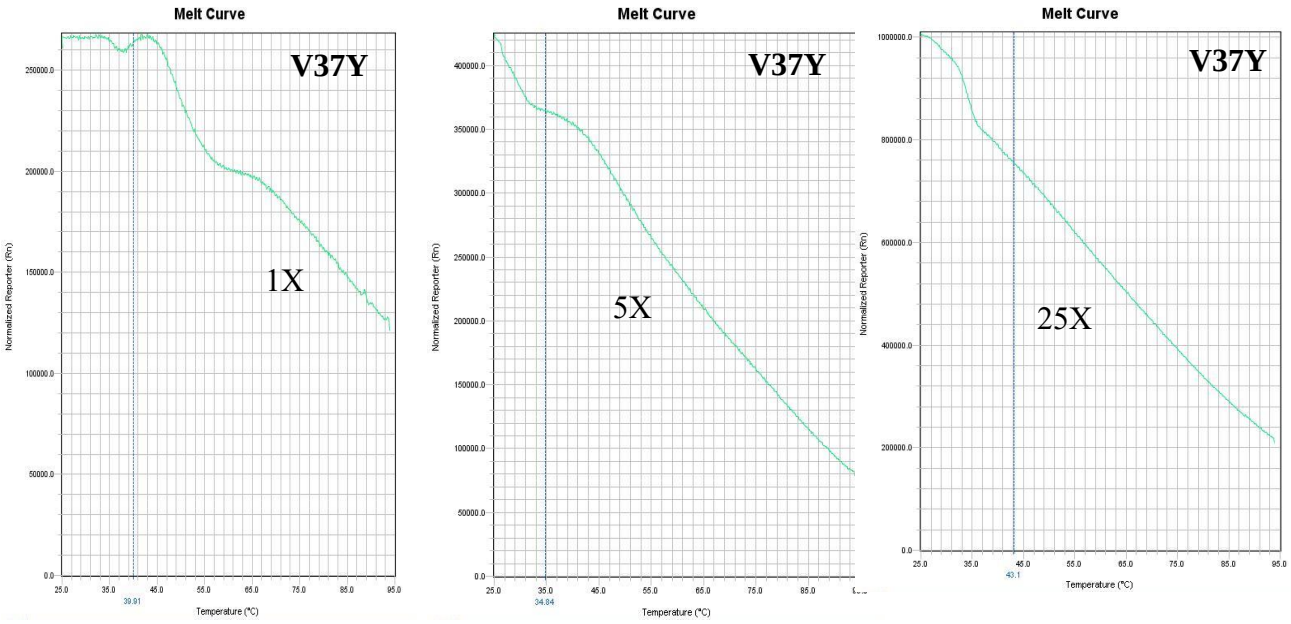


Figure 55. The melt curve profiles of V37Y sdAb under varying conditions for T_m determination. Different concentrations of SYPRO™ Orange dye (1X, 5X, 25X) were investigated for comparison using 20 µM of protein for each reaction. For each condition, triplicates of reactions mixtures were prepared, and all samples were run in parallel. Additionally, positive control and negative control samples were included in triplicates as controls (data not shown). The calculated melting temperatures were determined by the program software. The T_m values are indicated by blue dashed vertical lines.

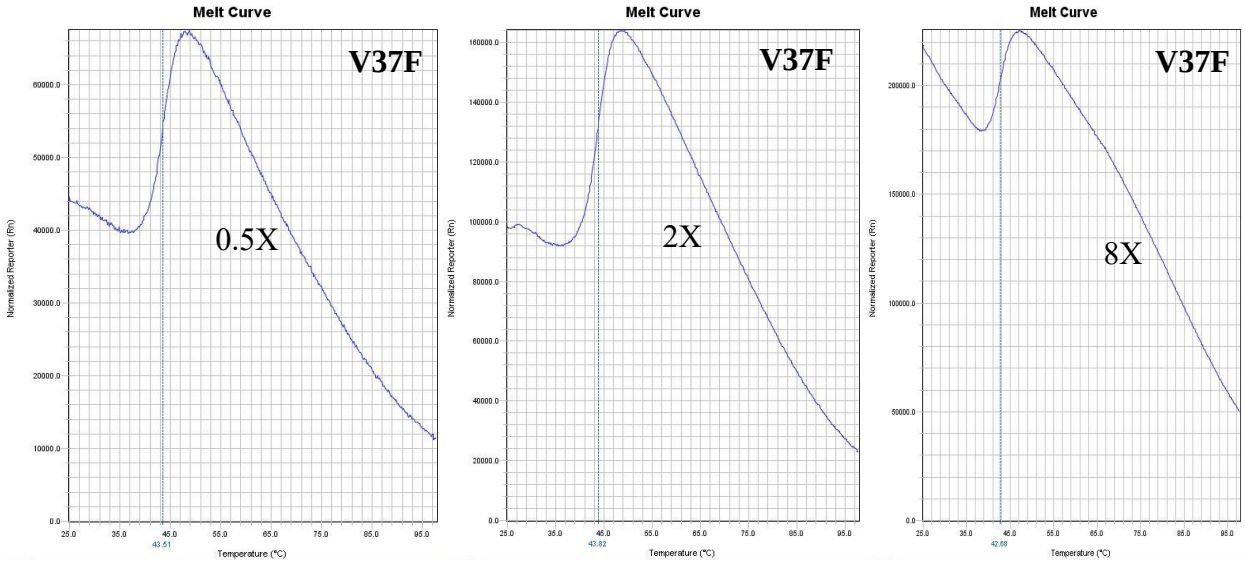


Figure 56. The melt curve profiles of V37F sdAb under varying conditions for T_m determination. Different concentrations of SYPRO™ Orange dye (0.5X, 2X, 8X) were investigated using 2 µM of protein samples. For each condition, triplicates of reactions mixtures were prepared, and all samples were run in parallel. Additionally, positive control and negative control samples were included in triplicates. The calculated melting temperatures were determined by the program software. The T_m values are indicated by blue dashed vertical lines.

The experiments were systematically conducted using various sdAb samples under different experimental conditions (**Figure 51**, **52**, **53**, **54**, **55, 56**), including variations in dye concentration and protein concentration. These trials were performed to optimize the overall procedure.

When optimizing the thermal shift assay, we continuously explored the available parameters within the reaction mixture (reaction buffer, the SYPRO™ Orange dye, and diluted protein sample in PBS). Given these limited number of variables, we can only focus on modifying the concentrations of the sdAbs and the dye. We systematically changed the

sdAb concentrations from 2 to 20 µM and the dye concentration from 0.5X to 50X. Even though we tried many different combinations of dye-protein modifications for optimisation purposes, the initial signal remained the same in most of our experiments. The interaction between the hydrophobic residues and the SYPRO™ Orange dye upon mixing created a consistent presence of baseline signal.

In the absence of the light chain, WT sdAb and its engineered variants possess inherent hydrophobicity. These residues are found at the former V_H-V_L interaction sites. As the temperature increases, previously buried residues also reveal, contributing to the initial hydrophobicity. This intrinsic feature of the sdAbs prevents the proper evaluation of the resulting melt curves

Exceptionally, amongst all tested sdAb samples, only V37F sdAb gave rise to a normal T_m profile, excluding the reactions with 8X dye (**Figure 56**). Unlike other samples, this particular sdAb did not display an initial fluorescence signal. To validate the reliability of V37F sdAb findings, the same experimental parameters were used, but this time other sdAb variants were used as protein samples. We failed to replicate the construct melting curves without an initial signal. This issue raised concerns about the consistency of this method and its applicability on the other sdAb samples.

To address this issue, we sought an alternative method for determination of T_m. We eventually changed our focus to Guanidinium chloride-induced chemical denaturation.

4.5.3. Chemical stability determination by GdmCl denaturation

When failed to obtain Tm values by using thermal stability assay, we moved to GdmCl-induced denaturation strategy by inducing chemical unfolding to explore stability features of the sdAb variants.

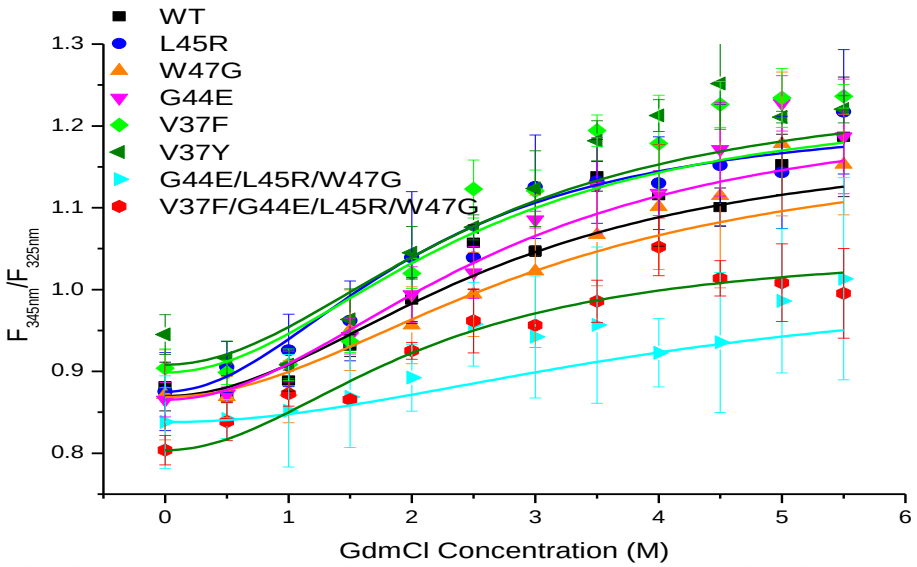


Figure 57. The Chemical stability assessment of Bevacizumab WT and its variant sdAbs by GdmCl denaturation. Protein- GdmCl mixtures, comprising a final protein concentration of 75 µg/ml and varying GdmCl concentrations ranging from 0 to 6 M were incubated overnight at 10°C. Intrinsic fluorescence was measured between 300 and 400 nm using a flash plate reader. The fluorescence ratio ($F_{345\ nm}/F_{325\ nm}$) was calculated to construct the unfolding curve and the chemical melting point (C_m) was calculated. Experimental repetitions were conducted and statistical analyses were included using the OriginPro software.

Table 126. The chemical unfolding determination of Bevacizumab WT and its variant sdAbs was calculated using the Hill1 function in the OriginPro 8.5.0 software and standard errors have been provided.

Bevacizumab V_H variant	**C_m (M)**
WT	2.68 ± 1.4
L45R	2.08 ± 0.5
W47G	3.01 ± 1.7
G44E	2.69 ± 1.0
V37F	2.47 ± 1.5
V37Y	2.14 ± 0.8
G44E/L45R/W47G	4.11 ± 1.2
V37F/G44E/L45R/W47G	2.08 ± 1.6

A higher C_m is indicative of a better stability profile in proteins. The alignment of decreasing order is G44E/L45R/W47G > W47G > G44E > WT > V37F > V37Y >

V37F/G44E/L45R/W47G > L45R (**Figure 57** and **Table 126**). In this context, it appears that the WT sdAb exhibits only moderate stability potential. These findings suggest that further modifications on several engineered sdAbs enhanced the stability of the WT sdAb to improve its overall performance.

Analyzing the C_m profile by chemical denaturant GdmCl helped to attain more reliable results and mitigates inconsistencies stemming from prior T_m results.

4.5.4. Aggregation kinetics

The aggregation kinetics aims to analyse behavioural chracteristics of sdAb variants. Aggregation kinetics were investigated under different stress conditions of thermal stress induced by heat exposure at 60°C and 65°C, and mechanical stress applied through shaking at 200 rpm. The sdAbs were subjected to either thermal stress alone, or a combination of both thermal and mechanical stress. The impact of these stressors on the aggregation kinetics of the sdAbs was analysed by calculating of the aggregation fraction.

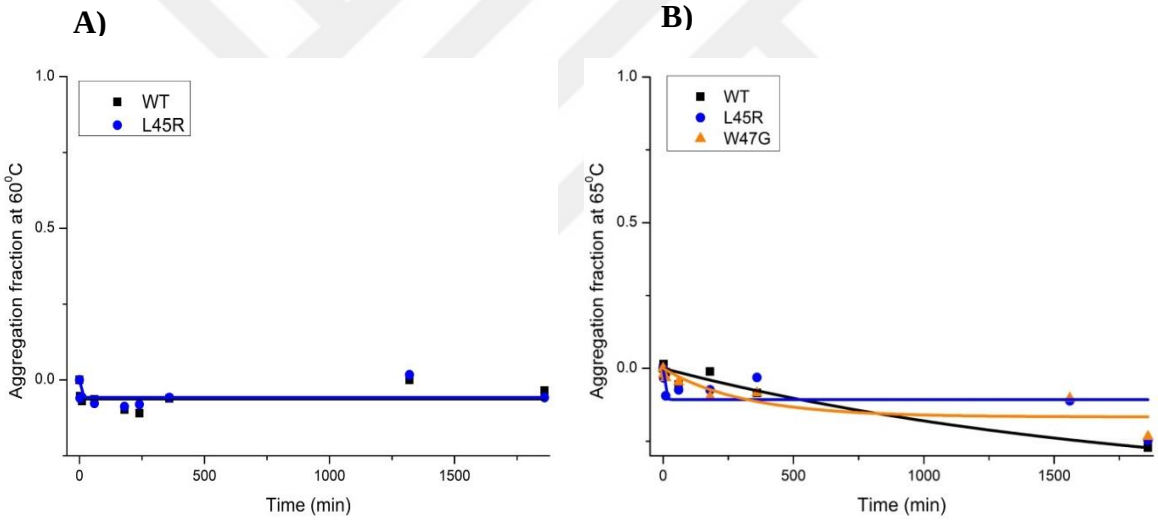


Figure 58. The aggregation kinetics of Bevacizumab sdAb WT and its variants, L45R and W47G under heat stress conditions at **A)** 60°C and **B)** 65°C were determined. The sdAb samples were prepared at a concentration of 0.5 mg/ml in PBS. These solutions were prepared in the tubes as aliquots of 25 µl. Subsequently, the sdAb solutions were incubated at 60°C and 65°C for the analysis. At specific time intervals following the start of incubation, samples were collected and centrifuged. The protein concentration was measured. The analysis continued until a steady decrease in protein concentration was observed. The aggregation fraction was calculated by taking the average protein concentration at each time point. The concentration value at 0 min (t_0) was subtracted from subsequent time points (t_0-t_x), and this difference was divided by the concentration at t_0. The calculated data were used to generate

aggregation fraction versus time graphs. Each experiment was performed in triplicate, and the reported values include standard deviations.

The initial test trial results with a limited number of data collection points of sdAb-WT and its two variants revealed that heat stress alone was not enough to exert significant stress on the selected samples (**Figure 58A**). Even when the temperature was increased from 60°C to 65°C, it did not efficiently mimic a physical stress condition (**Figure 58B**). As a response to this, an additional test trial was conducted where mechanical stress was also applied shaking at 200 rpm (**Figure 59** and **Table 127**). This combined approach allowed for a more comprehensive assessment., The use of both thermal and mechanical stress were necessary to induce notable changes in the aggregation behavior of the sdAbs.

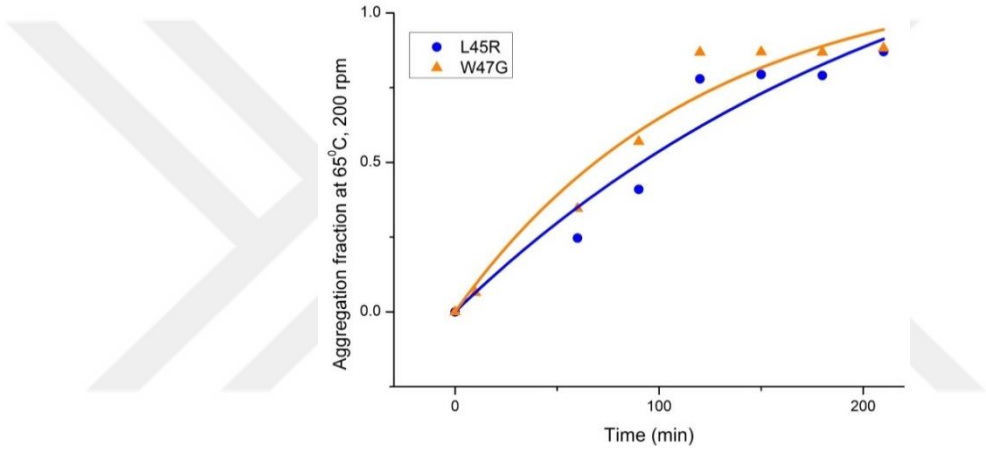


Figure 59. The aggregation kinetics of Bevacizumab sdAb WT and its variants , L45R and W47G were determined under heat and mechanical stress conditions at 65°C and shaking at 200 rpm. The protein concentrations and aggregation fraction calculation determined following the previously explained method, as described in the above figure legend.

Table 127. The aggregation kinetic value apparent rate constant (k_{app}) of sdAb WT and its variants, L45R and W47G were calculated using the BoxLucas1 function in the OriginPro 8.5.0 software including the standard errors.

Bevacizumab V_H variant	**k_{app} (s^{-1})**
WT	0.004
L45R	0.330
W47G	0.005

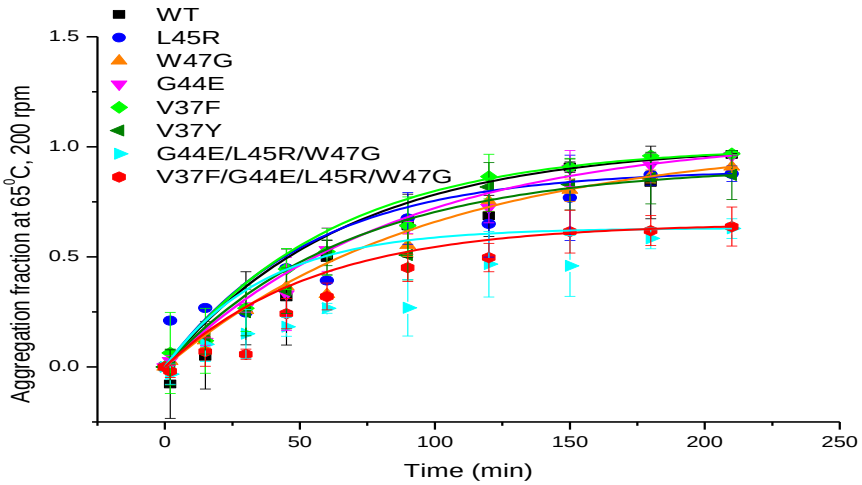


Figure 60. The aggregation kinetics analysis of Bevacizumab sdAb WT and its variants were employed to compare the propensity of sdAb variants against aggregation. To this end, sdAb solutions were prepared at a concentration of 0.5 mg/ml in PBS. The prepared antibody solutions were aliquoted into tubes as 30 µl with a minimum number of 7 tubes. These sdAb solutions were incubated at 65°C by shaking at 200 rpm. At specific time intervals from the initiation of incubation, samples were collected and centrifuged and their concentrations were measure in NanoDrop spectrophotometer. The analysis was completed once a steady decrease in protein concentration was observed. The obtained results were used to construct an aggregation fraction vs time graph.

Table 128. The aggregation kinetic value apparent rate constant of Bevacizumab sdAb-WT and its variants were calculated using the BoxLucas1 function in the OriginPro 8.5.0 software.

Bevacizumab V_H variant	k_{app} (s^{-1})
WT	0.015
L45R	0.018
W47G	0.010
G44E	0.011
V37F	0.016
V37Y	0.015
G44E/L45R/W47G	0.027
V37F/G44E/L45R/W47G	0.018

A lower k_{app} value is indicative of better stability against aggregation. This means that the sdAb being studied exhibits reduced propensity for aggregation, which is a favourable

characteristic. The Bevacizumab WT sdAb had mediocre stability profile compared to the other sdAb variants; W47G > G44E > V37Y > WT > V37F > V37F/G44E/L45R/W47G > L45R > G44E/L45R/W47G (**Figure 60** and **Table 128**).

4.5.5. High-performance liquid chromatography (HPLC)

HPLC analysis was used to validate the presence of impurities (dimers, truncated sdAb fragments) that keep appearing in many SDS-PAGE and WB analysis even after many trials.

In SDS-PAGE, the presence of any band deviating from the anticipated molecular weight indicates a contamination. However, when extra bands emerged on WB membranes, concerns were raised about potential dimers (particularly if they appeared at higher molecular weights) or truncated forms of the target protein (especially if they manifested at lower molecular weights) in addition to monomeric form of sdAbs. In WB, only specifically probed proteins became visible on the blotted membrane. In this case, only the molecules carrying the Flag-tag at their respective C-termini can be specifically targeted.

HPLC provided higher resolution results compared to SDS-PAGE, enabling the differentiation between monomeric sdAbs and their potential dimeric forms. If dimers were present, they were likely to exhibit doubled retention times compared the original format (**Figure 61-67**). Additionally, this analysis aided the identification of other contaminants that might interfere with sdAb purity and help ruling out.



Figure 61. HPLC chromatogram of the purity analysis of G44E/L45R sdAb.

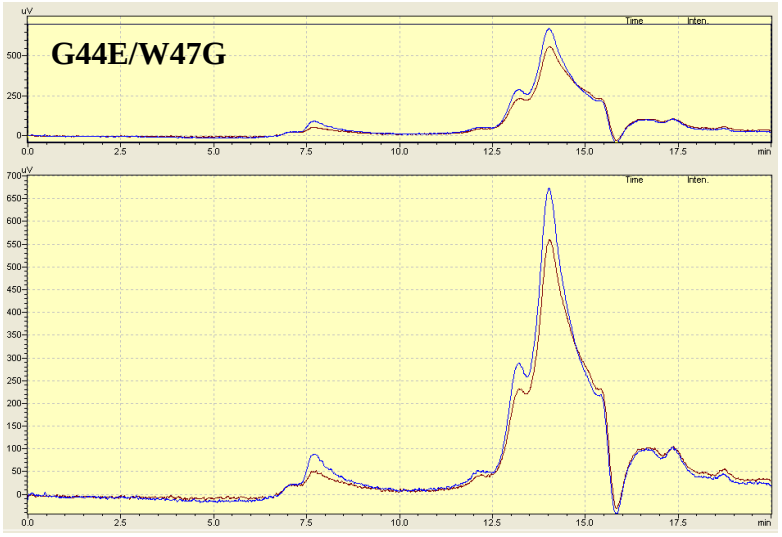


Figure 62. Chromatogram of the purity analysis of G44E/W47G sdAb assessed using HPLC.



Figure 63. Chromatogram of the purity analysis of G44E/L45R/W47G sdAb via HPLC.



Figure 64. Chromatogram depicting the purity analysis of V37F/G44E/L45R/W47G sdAb assessed via HPLC.

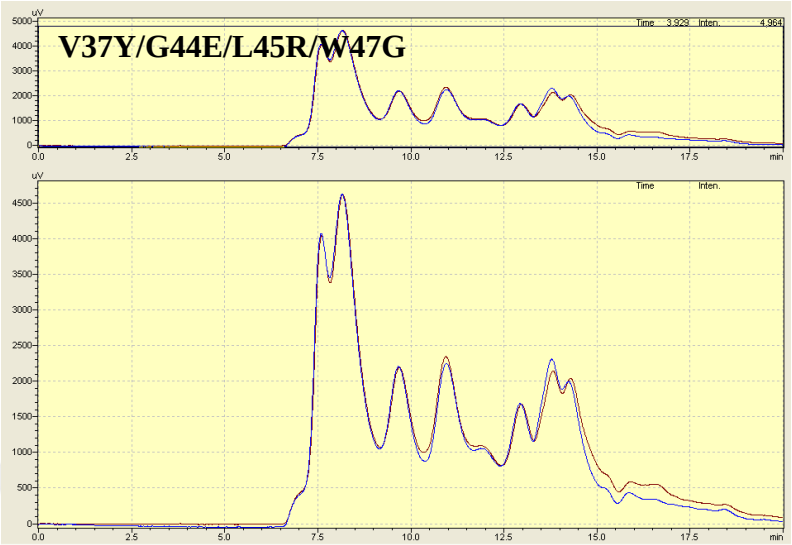


Figure 65. Chromatogram of the purity analysis of V37Y/G44E/L45R/W47G sdAb assessed by HPLC.

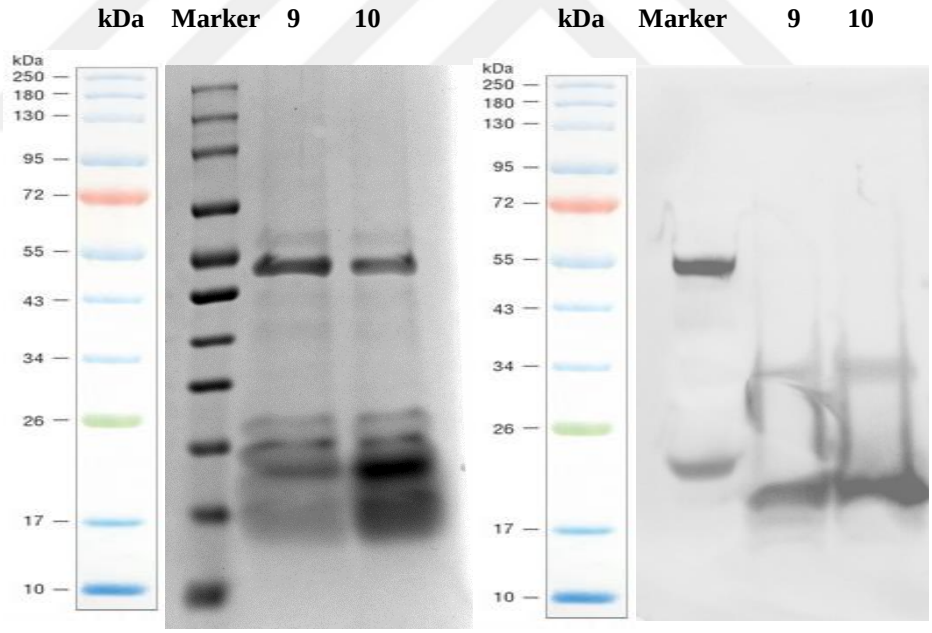


Figure 66. SDS-PAGE results of sdAb 9 and sdAb 10. Dimer formation is further confirmed by HPLC shown in **Figure 67**. Please refer to **Table 107** for the full nomenclature of the introduced mutations.

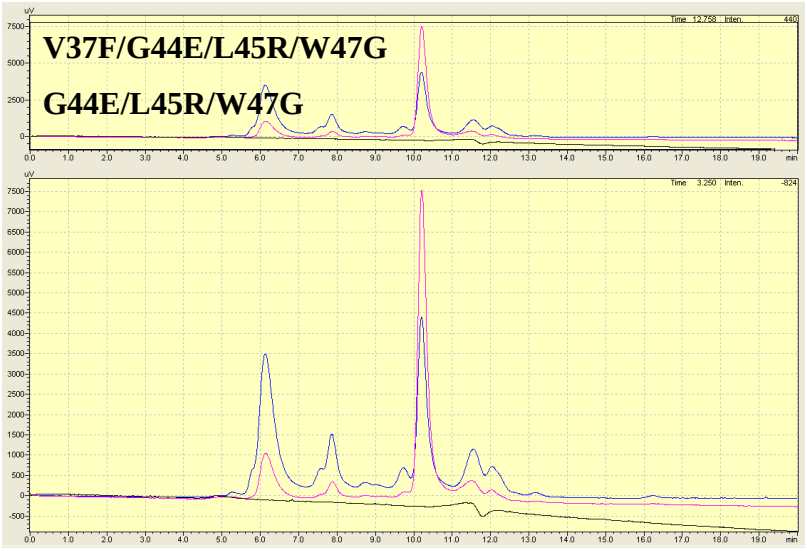
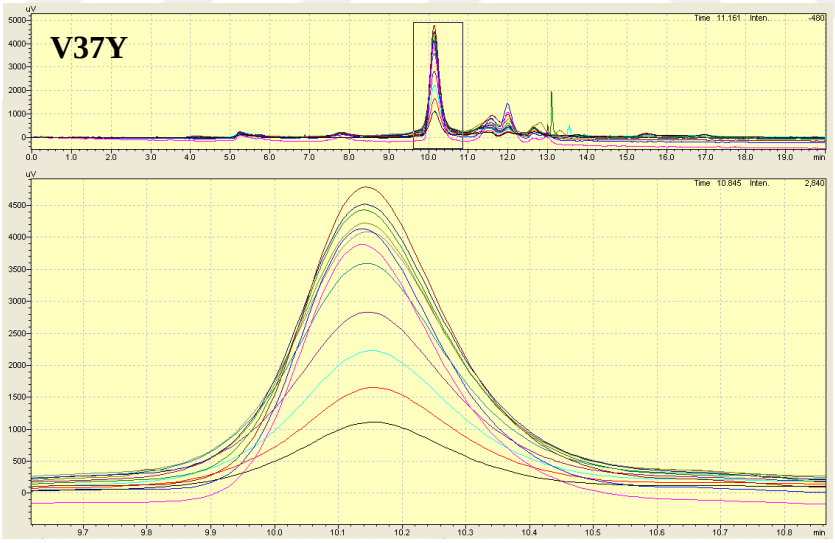


Figure 67. Chromatogram showing the purity analysis of V37F/G44E/L45R/W47G and G44E/L45R/W47G sdAbs assessed using HPLC.

HPLC was also used in validating our aggregation kinetics assay (**Figure 68-70**). The aggregation kinetics assay demanded analysis at multiple time points and takes longer period of time and attention. In contrast to the aggregation assay, HPLC allows parallel run of the samples and simultaneous analysis of multiple time points.

A)



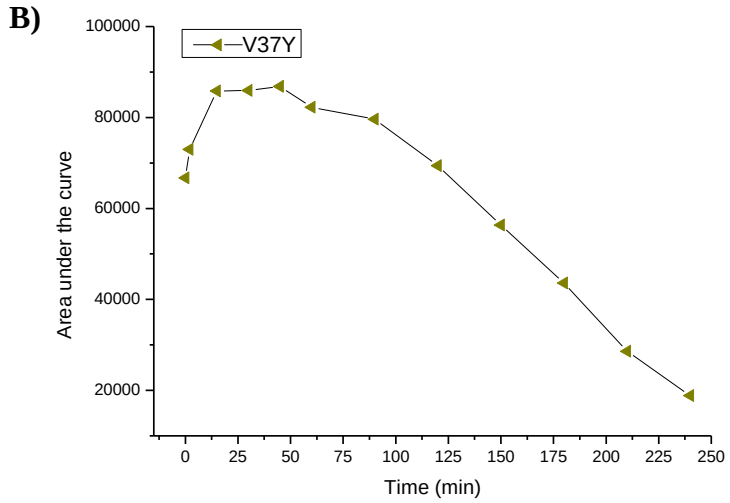


Figure 68. Chromatographic analysis of V37Y sdAb aggregation kinetics via HPLC. **A)** The chromatogram provides an overview of all peaks obtained following V37Y sdAb analysis. **B)** The graph reflects a closer inspection of the main peak of the chromatogram, aligning with the expected retention time of 10.2 min. The graph illustrates the area under the curve for each focused peak, corresponding to individual V37Y sdAb samples collected at specific time points. Over time, the decreasing area under the curve indicates diminishing sdAb peak intensities.

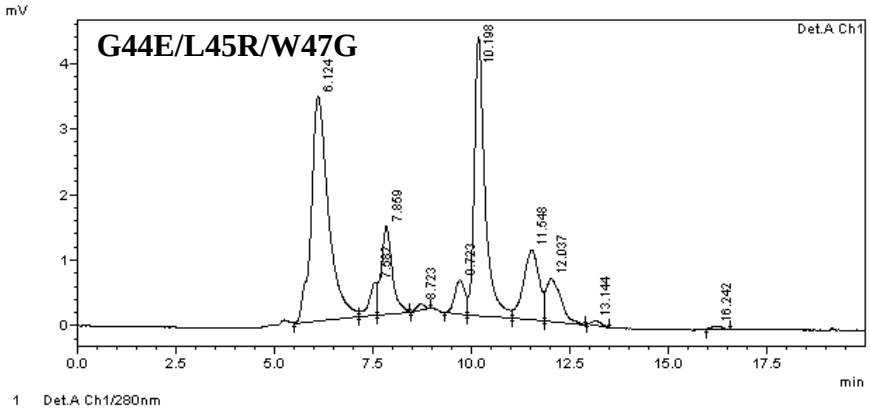


Figure 69. Chromatogram showing the aggregation behaviour of sdAb 9 assessed using HPLC.

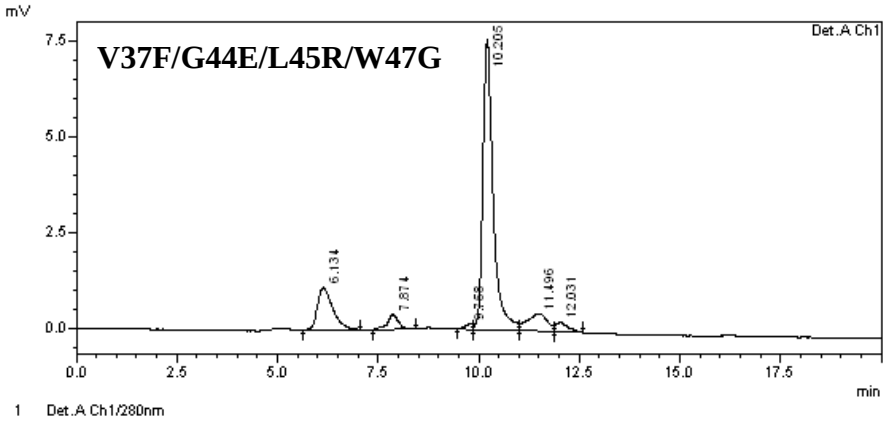


Figure 70. Chromatogram of the aggregation behaviour of sdAb 10 assessed using HPLC.

For our sdAbs, the expected retention time is 10.2 min based on their approximately calculated molecular weight (18.9 kDa) Based on this information, the results indicate the absence of our protein of interest in significant amounts (**Figures 61-64**) particularly for the sdAbs G44E/L45R, G44E/W47G, G44E/L45R/W47G and V37F/G44E/L45R/W47G. However, V37Y/G44E/L45R/W47G sdAb did appear at the expected retention time, yet followed by various other contaminants with higher molecular weights (**Figure 65**), which is also in agreement with SDS-PAGE results (**Figure 66**). Notably, in instances where dimers were present in the mixture, the expected retention time should be, but no such peaks were visible in any of the aforementioned cases. Eventually, we ruled out the risk of dimers as contaminants, but this explanation leads to presence of other unknown contaminants which their mitigation is harder due to their unidentified nature.

In conclusion, HPLC analysis proved to be used a useful alternative tool the context of aggregation kinetics in addition to determination of contaminants.

4.5.6. Isothermal titration calorimetry

ITC is an analytical technique to study molecular interactions of antibody-antigen binding (binding affinity: K_d). This method calculates the energetics of these interactions by directly measuring heat changes associated with the binding process. In addition to binding affinity, ITC also gives information about the other binding kineticks related information, such as stoichiometry, enthalpy, and entropy of the tested interactions.

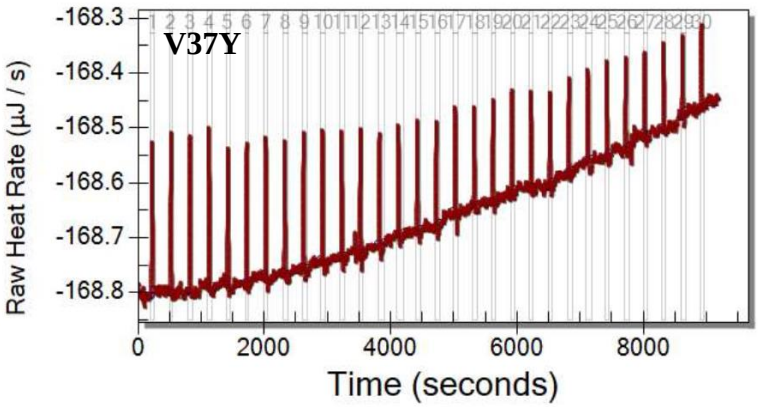


Figure 71. ITC thermogram plot of V37Y. The titration curve represents the raw data of heat changes (µJ/s) in time, showing the injection of a titrant solution into the sample cell containing the target sdAb.

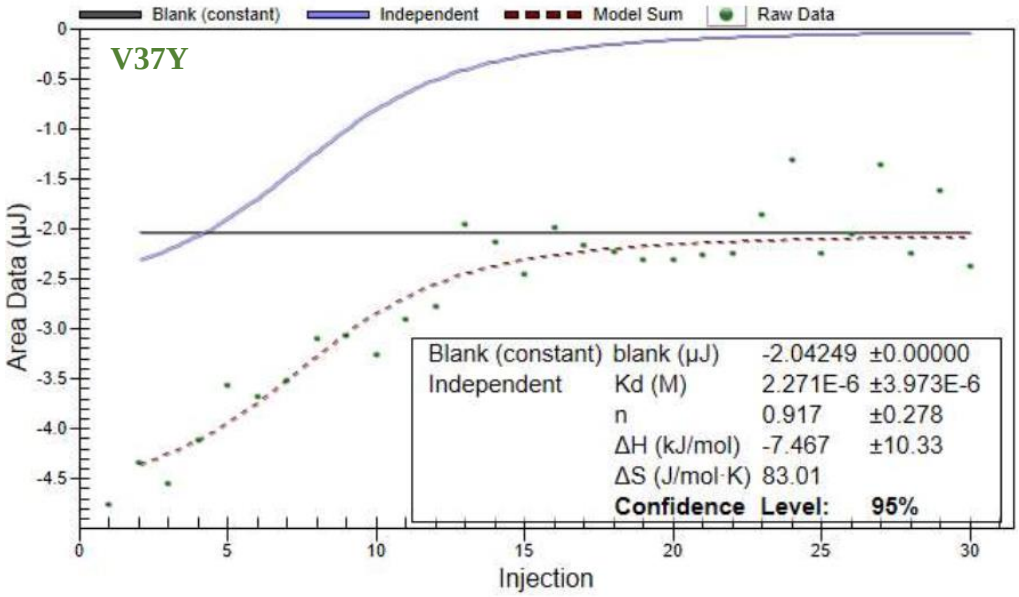


Figure 72. ITC titration plot of V37Y. Blank correction dashed line corresponds to the baseline correction, reflecting the heat changes observed in the absence of the binding. The integrated heat changes (∫ΔH) over time provide a cumulative view of the heat changes during the titration. These heat changes facilitate the determination of binding enthalpy (ΔH). The smooth line of fit curve represents the best fit of the experimental data to a binding sdAb. The other thermodynamic data calculated by ITC are binding affinity (K_d) and entropy change (ΔS).

The ITC results did not exhibit the typical and easily discernible S-shape (**Figure 71** and **72**). When comparing our findings with previously calculated reference values calculated in the earlier studies, the reported K_d by ITC for Bevacizumab is 2.3 nM (Khalili et al. 2018).

In contrast, the ITC results for V37Y sdAb revealed an unfavourable difference, with a determined K_d value of 2.3 mM. This discrepancy shows a substantial change from the expected binding affinity. It can be definitely concluded that the tested V37Y sdAb was not improved in term of binding kinetics.

4.5.7. In vivo testing of Bevacizumab sdAb variants in zebrafish

To test the *in vivo* efficacy of antibody variants in the zebrafish vascularization model, Tg(fli1:EGFP)^{y1} strain was used. SIV development in zebrafish larvae occurs between days 2-5 post-fertilization on the yolk sac (**Figure 73**).

First, the optimisation studies were conducted, followed by testing of antibody variants. Optimisation studies involved definition of injection time frame, effective dose range and the injection region into the transgenic zebrafish strain for imaging and image analysis method, before moving into the injections of sdAb-WT and its variant.

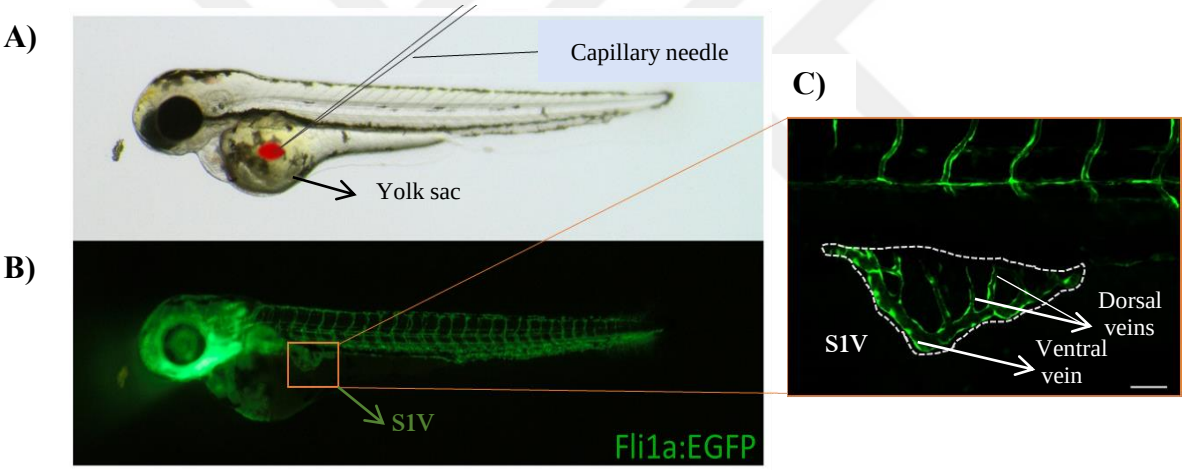


Figure 73. Microinjection, visualization and imaging of zebrafish Tg(fli1:EGFP)^{y1} strain embryos. **A**) Microinjection site in the yolk sac, **B**) All vessels in transgenic zebrafish strain are labelled with eGFP. Whole-body image of a 3-day-old larva was obtained by tile-scan imaging with a 10X objective. The SIV area is indicated with the orange box, **C**) Z-stack images were captured with a 25X objective for analysis, and the entire SIV area was recorded using maximum projection. The measured areas and vessels are marked for analysis.

ImageJ software was utilized for image analysis. Initially, Z-projections of the images were generated, and various parameters related to SIV vessel development were manually measured using the ImageJ software to quantify the results. The different parameters measured for optimisation included the total SIV area, ventral vein (main vessel), dorsal and ventral, and overall average vessel lengths.

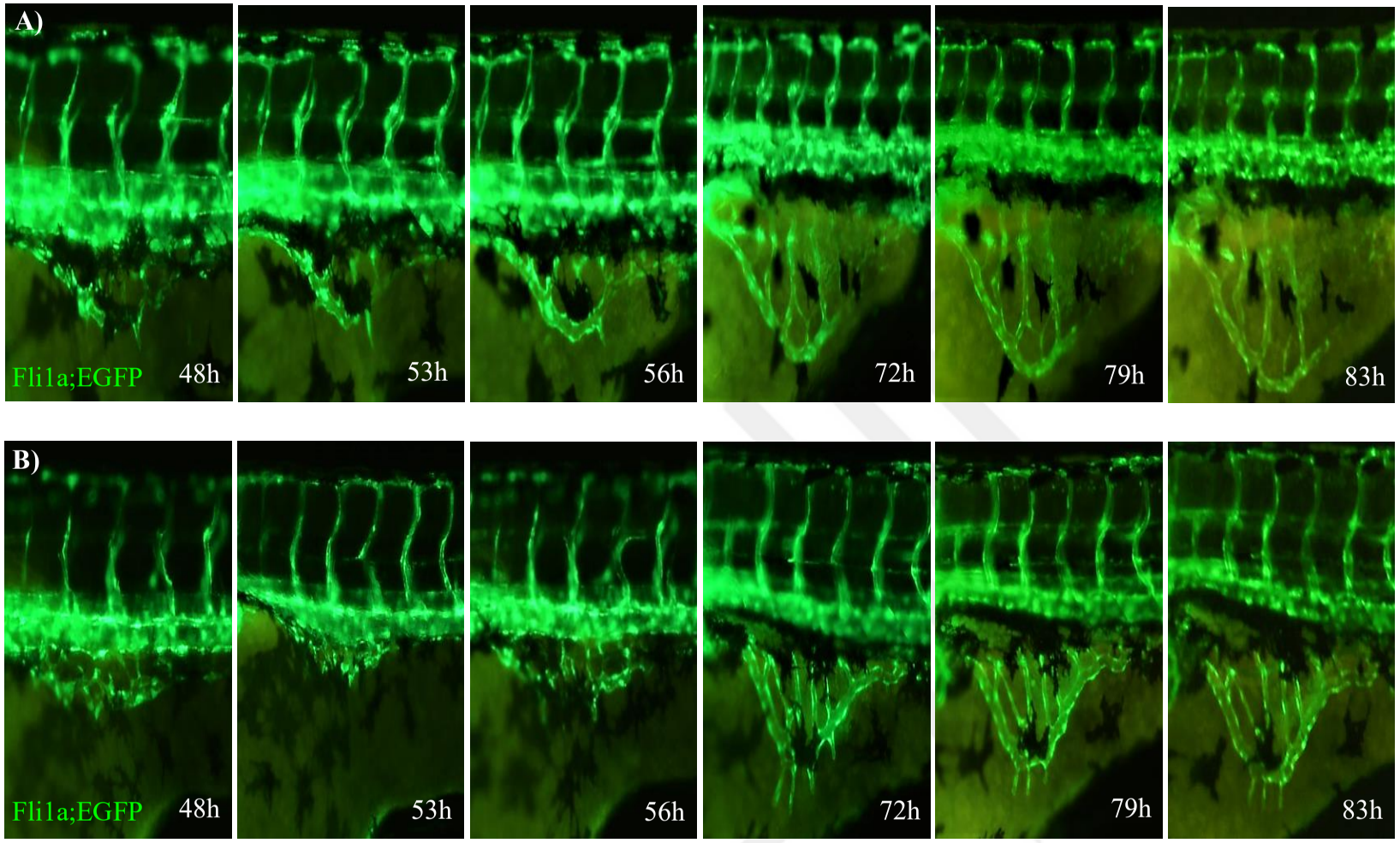


Figure 74. Optimization studies for zebrafish. Zebrafish development visualization between 48h-83h for injection time frame determination on different zebrafishes at SIV area (**A** and **B**).

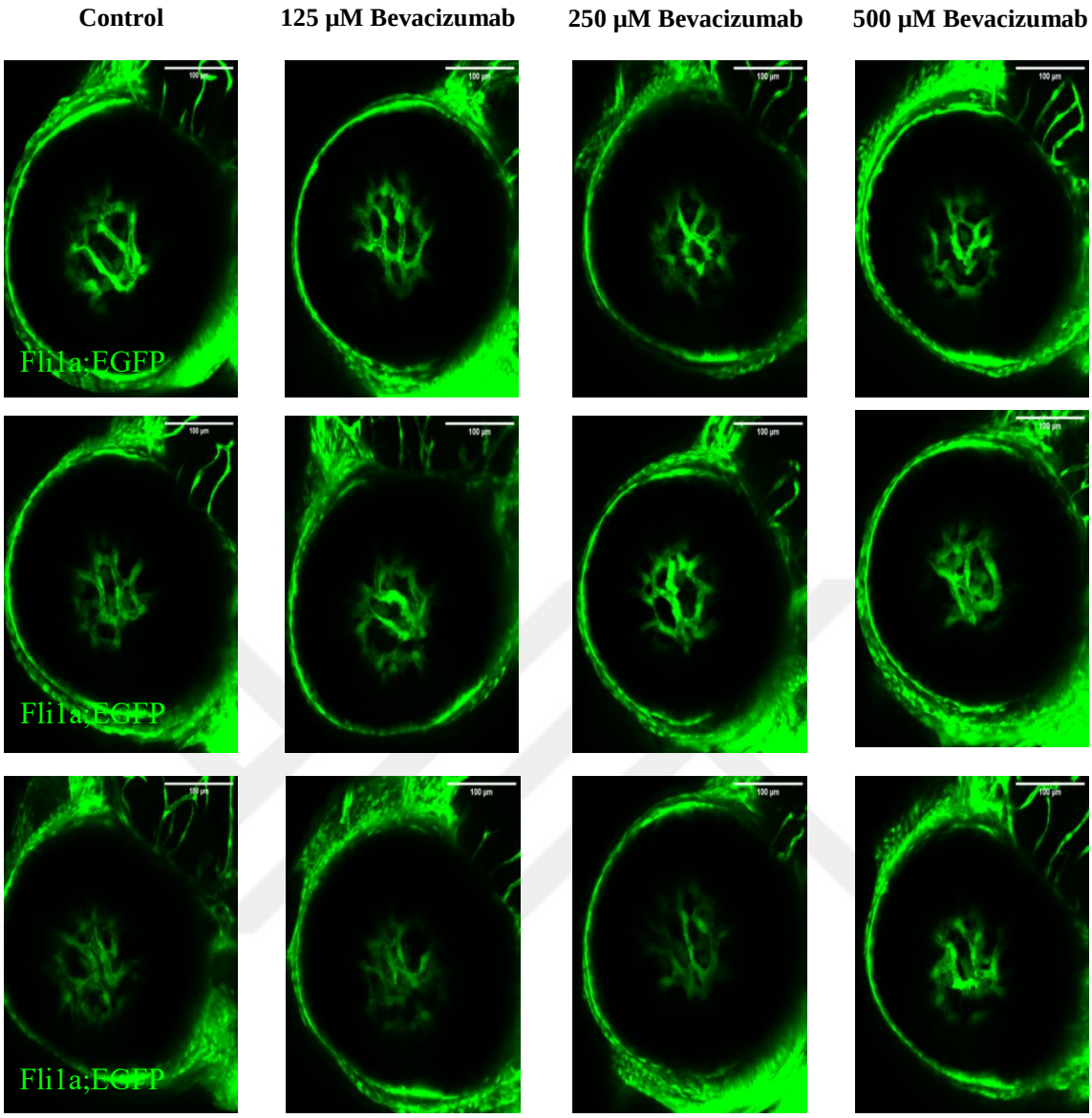


Figure 75. Confocal microscopy images depicting the GFP-labelled zebrafish eye after a 48-h treatment with Bevacizumab in E3 media in the fish swimming water, captured at 25X magnification. The zebrafishes were treated with varying concentrations of Bevacizumab. The control group was treated with PBS. The images illustrate the effects within the defined effective dose range of Bevacizumab.

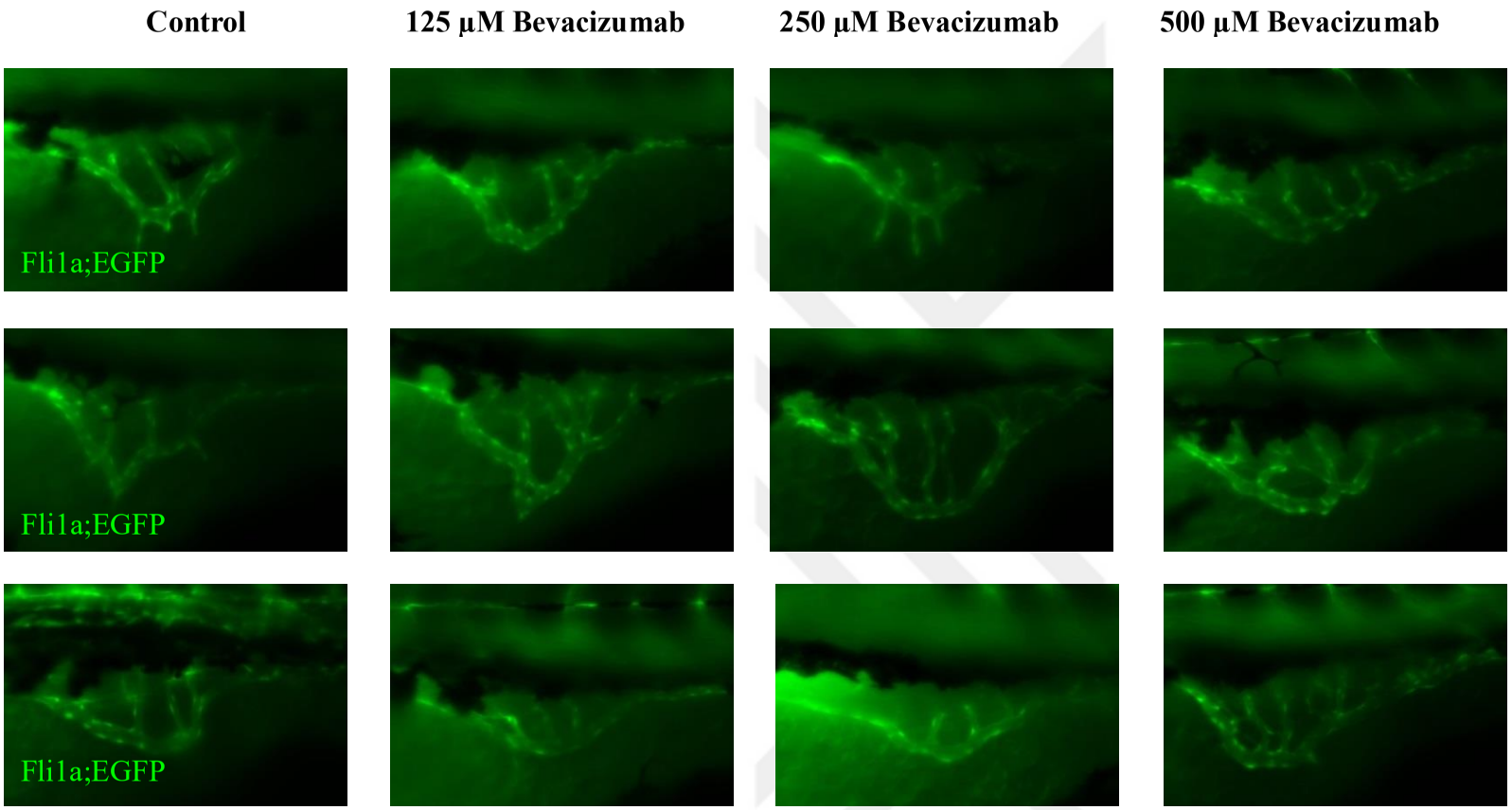
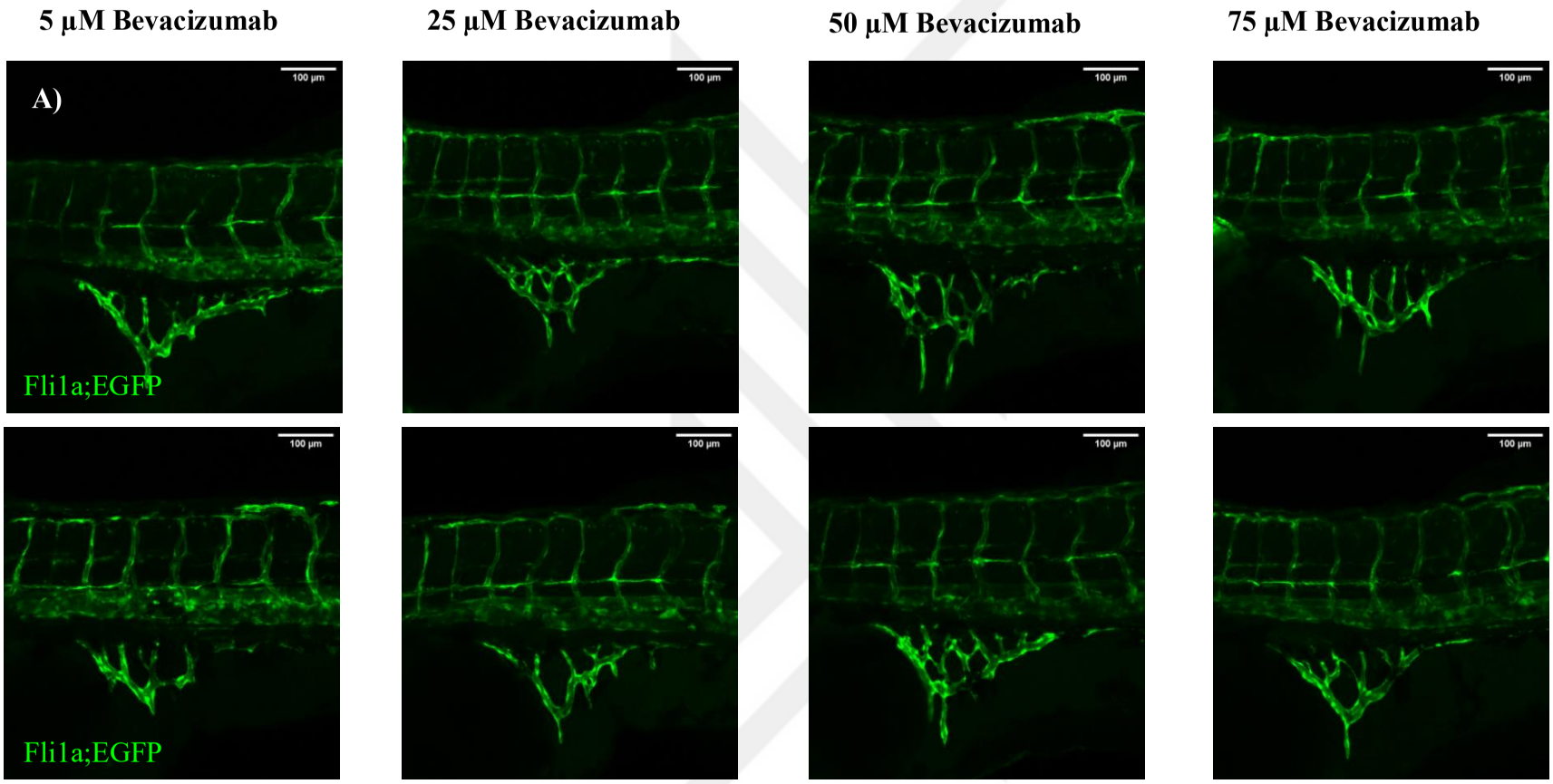


Figure 76. Zebrafish embryos treated for 48 h with Bevacizumab in E3 media in the fish swimming water were analysed for SIV development. Bevacizumab treatment with a ranging concentration was administered via swimming water, with PBS serving as the control group. Fluorescent images were captured at 25X magnification for detailed visualization of eGFP-labelled SIV area.


5 µM Bevacizumab
25 µM Bevacizumab
50 µM Bevacizumab
75 µM Bevacizumab
A)
100 µm
Fli1a;EGFP

B)

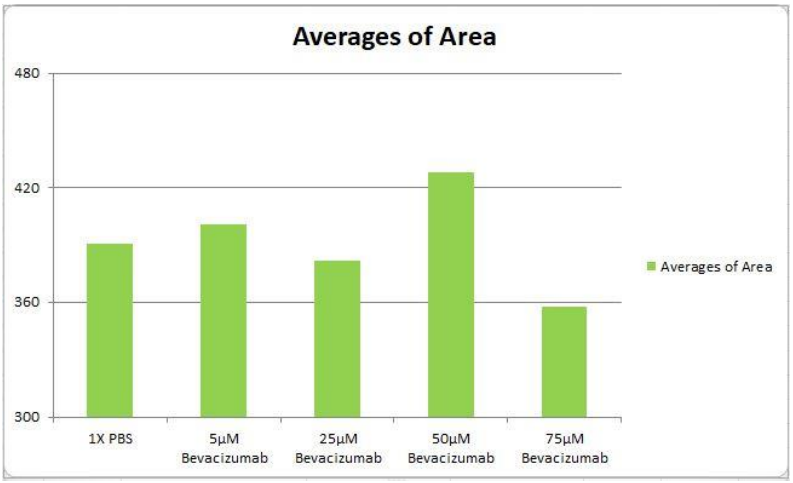
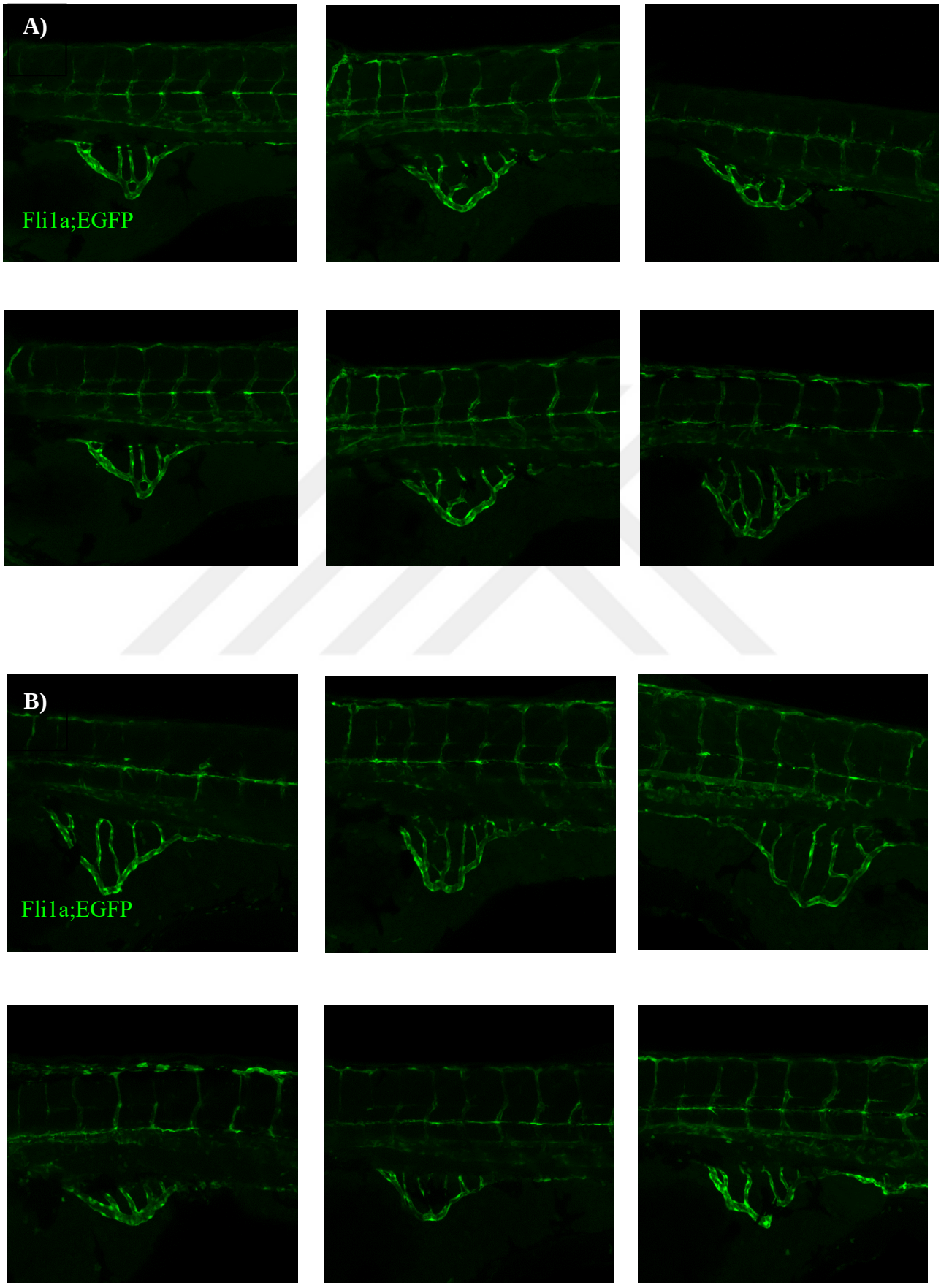


Figure 77. The Bevacizumab treatment administered through SIV injection, utilizing varying concentrations of Bevacizumab. **A**) Confocal images were acquired at 25X magnification to visualize eGFP-labelled SIV veins. **B**) The subsequent analysis focused on precise area measurements of the SIV region. At least 20 different zebrafish embryos were analysed for SIV area measurement.

Preliminary trials were conducted as optimisation studies on zebrafish. Initially, the injection timeframe was established (**Figure 74**). Subsequently, the injection concentration of Bevacizumab was determined for comparative purposes with the tested sdAbs, aiming to identify the minimum effective dose for inhibiting vascularization in the zebrafish eye (**Figure 75**). A similar dose determination study was also carried out in the SIV area of the zebrafish (**Figure 76**). Additionally, a comparison of efficiency of fluorescence and confocal microscopy visualizations was performed (**Figure 76** and **Figure 77**).


A)
Fli1a;EGFP
B)
Fli1a;EGFP

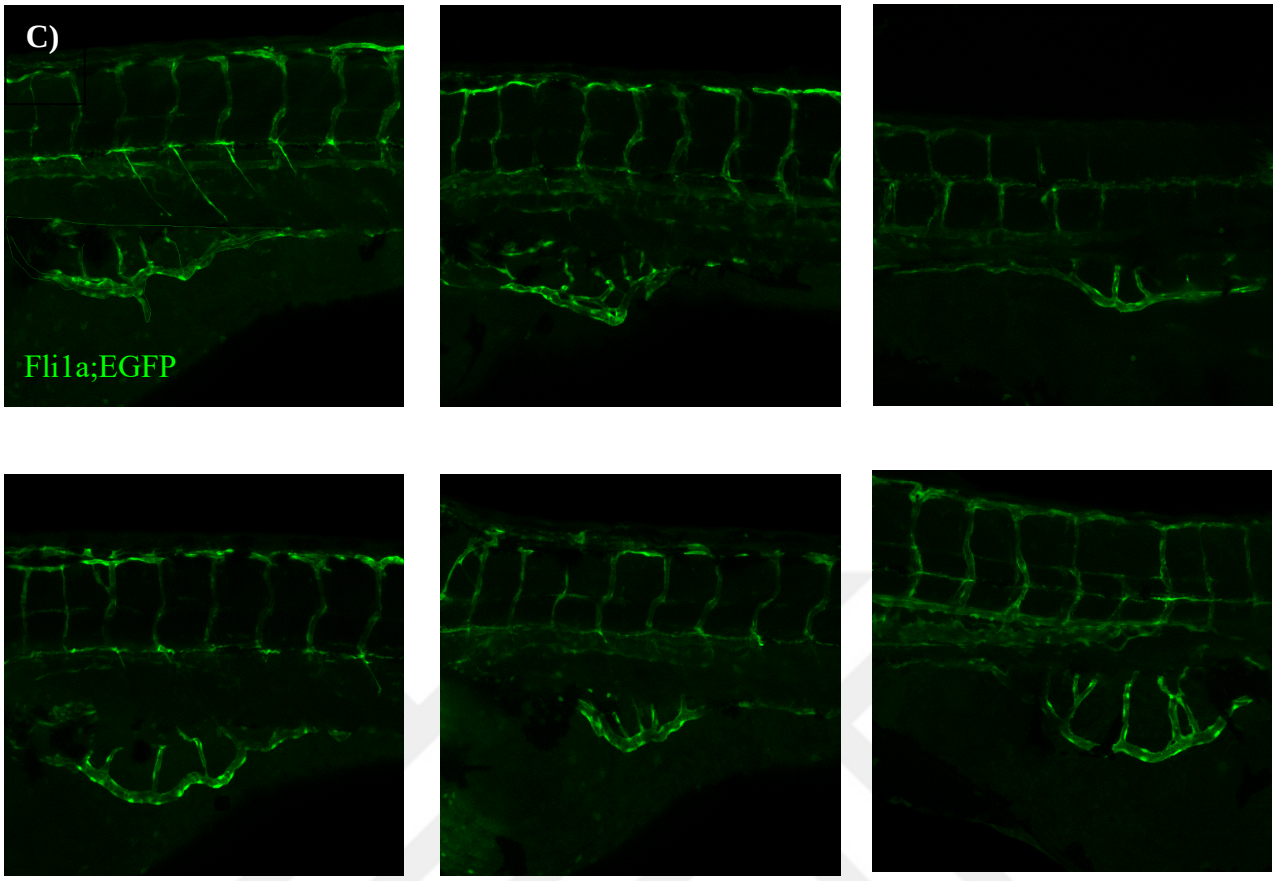


Figure 78. The confocal microscopy images displaying zebrafish injected with **A**) PBS, **B**) sdAb WT and **C**) sdAb V37Y at SIV area.

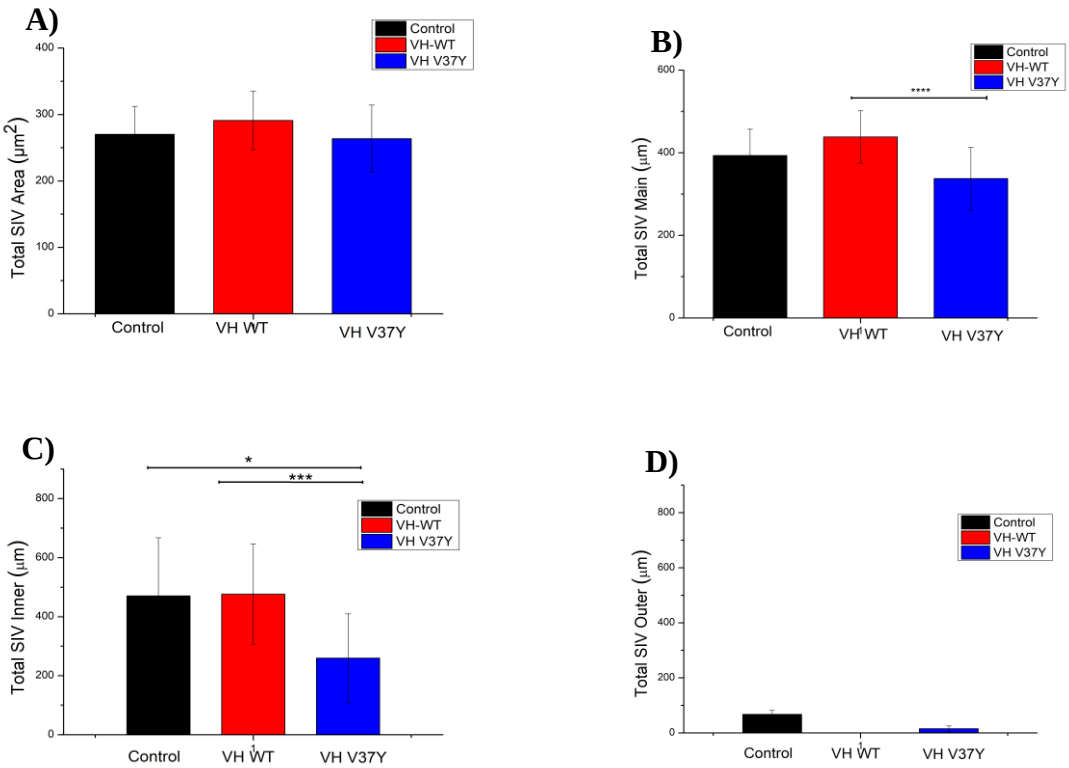


Figure 79. Analysis of zebrafish embryonic vasculature alterations following microinjection with the tested sdAbs. The figure represents the measurements of changes in vasculature, on zebrafish embryonic blood vessel development. Average changes in SIV area for PBS control, sdAb-WT, and sdAb-V37Y at different locations were investigated. **A**), The changes in ventral vein (main vessel) length **B**), the changes in dorsal vein length **C**), and the changes in ventral vein length and **D**) the outer SIV branch length were plotted. In the measurements obtained using the ImageJ program, the "Area" calculation was used for calculating the area, while the "Perimeter" calculation was used for measuring vessel lengths. Line thickness was also included in the area calculation; therefore, drawings were made along the inner boundary of the area. For vessel length calculations, the drawings were made through the middle of the vessels. For analysis, in each injected sample, total SIV area, ventral vein (main vessel), dorsal, ventral, and overall average vessel lengths were calculated, along with their means and standard deviations. The dataset obtained from protein injections was subjected to a two-sample t-test against the PBS control dataset, and significance was determined as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

After the completion of all optimisation studies, including the determination of the injection time frame, Bevacizumab injection concentration, and the zebrafish injection region, the *in vivo* testing of the sdAbs were commenced. The most promising sdAb variant, in terms of purity and quantity, sdAb V37Y, was selected as the testing candidate. Zebrafish were injected with sdAb WT, sdAb V37Y, and PBS as a control under the optimised conditions established earlier. The SIV area was visualized using a confocal microscope. Following visualization, the inhibited SIV area was determined through a detailed comparison via image analysis (**Figure 78**). Based on the comparative results obtained through confocal microscopy with optimised experimental parameters, the total SIV area of the zebrafish remained a similar size even after injections with either WT sdAb or V37Y sdAb. However, a significant change was observed in the SIV perimeter compared to WT sdAb, indicating that the V37Y exhibited a more effective inhibitory ability on the SIV region. Additoinally, inner part lengths of the SIV were measured, revealing substantial changes compared to both the non-treated PBS-injected control and when comparing sdAbs to each other (**Figure 79**). Overall, the therapeutic inhibitory effect of sdAbs was demonstrated *in vivo* in Tg(fli1:EGFP)^{y1} transparent zebrafish model. The altered version of Bevacizumab V_H sdAb, V37Y proved to be superior to its unaltered counterpart WT in terms of therapeutic effectiveness.

4.6. Computational methods to uncover distinctions between antibodies and nanobodies

Following unsuccessful attempts to purify sdAbs in our home lab and encountering suboptimal yields and purity with the redesigned sdAbs in the host lab as well, a novel alternative strategy was designed. An *in silico* approach was decided to be used for successful designs, focusing on the conversion of conventional IgGs into V_H sdAb formats for future efforts. An array of computational methods was applied to unveil significant distinctions between the antibody and nanobody sequences. Advanced bioinformatics tools were used to detailly investigate the physicochemical characteristics as well as sequential and structural differences of variable domains of these both groups. The application of diverse *in silico* approaches, including sequence alignment, structural modelling, and machine learning analyses, provided a comprehensive understanding of the differences between V_Hs and V_HHs.

First, we aligned multiple FASTA sequences of vast number of V_H and V_HH datasets according to the Chothia numbering scheme. By doing this so, the boarders of HCRs and FW regiongs were defined. After setting the sequences, we conducted numerous in-depth analyses, where HCDR length variation comparison was the first of those.

4.6.1. Length variation in HCDRs

The comparative analysis of HCDRs lengths revealed notable distinctions between V_Hs and V_HHs. The plot at **Figure 80** visually represents the distribution of HCDR lengths for both antibody types, offering a comparative view of their length encounter frequencies across the specified HCDRs.

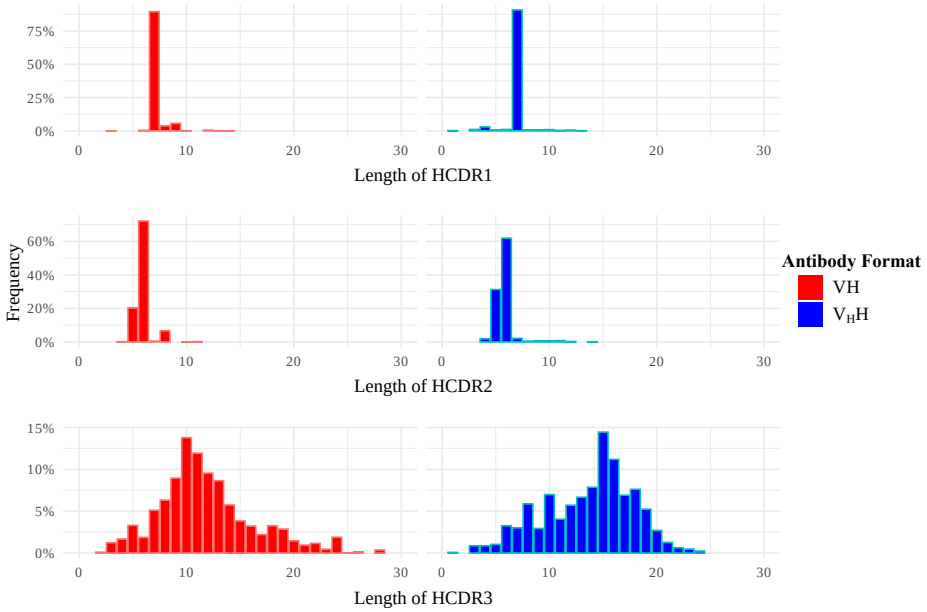


Figure 80. The variation in HCDR lengths for HCDR1-3 in both V_Hs and V_HHs was illustrated by bar plots. The y-axis represents the frequency of HCDR lengths based on their encounter rate in each dataset. The HCDR windows were defined according to the Chothia numbering scheme. V_Hs are represented in red, while V_HHs are depicted in blue.

Table 129. The mean length of in V_Hs categorized by species, was computed individually based on the each dataset.

Species	HCDR1 length	HCDR2 length	HCDR3 length
Cricetulus Migratorius	7	6	7.9
Homo Sapiens	7	6.4	10.5
Homo Sapiens/Mus Musculus	7.4	5.8	9.9
Mus Musculus	7.1	6	9.8
Mus Musculus/Homo Sapiens	7.1	5.7	10.1
Pan Troglodytes	5	7	15
Rattus Norvegicus	7	5.7	8.2
Rattus Norvegicus/Homo Sapiens	7	6	8
Average V_Hs length	**7.2**	**5.9**	**11.9**

Table 130. The mean length of in V_HHs, categorized by species, was computed individually based on the each dataset.

Species	HCDR1 length	HCDR2 length	HCDR3 length
Camelidae	7	5.8	14.5
Camelus bactrianus	7.8	6.1	15.6
Camelus dromedaries	6.9	5.7	13.2
Lama glama	6.9	5.8	13.9
Vicugna pacos	7.1	5.8	12.5
Average V_HHs length	**6.9**	**5.8**	**13.5**

While HCDR1 and HCDR2 lengths exhibited high similarity between the two classes, the most significant variation emerged in the most important in antigen binding, HCDR3. Yet, HCDR1 and HCDR2 are slightly longer in antibody V_Hs. V_Hs predominantly displayed shorter HCDR3 lengths with a mode of 10 residues, while V_HHs exhibited a right-skewed distribution, indicating longer HCDR3 loops with a mode of 15 residues, aligning with previous findings in the literature. Despite V_Hs having fewer outliers with extended HCDR3s, the overall count of HCDR3s longer than 15 was higher in V_HHs (**Figure 80**).

Longer HCDR3s in V_HHs served to fold over hallmark residues at FW2 to enhance antigen specificity/affinity. Species-specific analysis highlighted variations (**Table 129**), with *Vicugna pacos* exhibiting the shortest average V_HH HCDR3 length (**Table 130**), suggesting species-specific influences on the antibody repertoire, possibly rooted by the specific immune responses or evolutionary adaptations.

4.6.2. Antigen binding scores of HCDRs

In the investigation of antibody structure, prominent significance was attributed to its CDRs, given their pivotal role in dictating the specificity and affinity to target antigen binding. The identification of key residues within the HCDRs is crucial for target recognition. This led to the analysis of paratope scores indicative of antigen binding of HCDRs including the two residues of flanking FW regions.

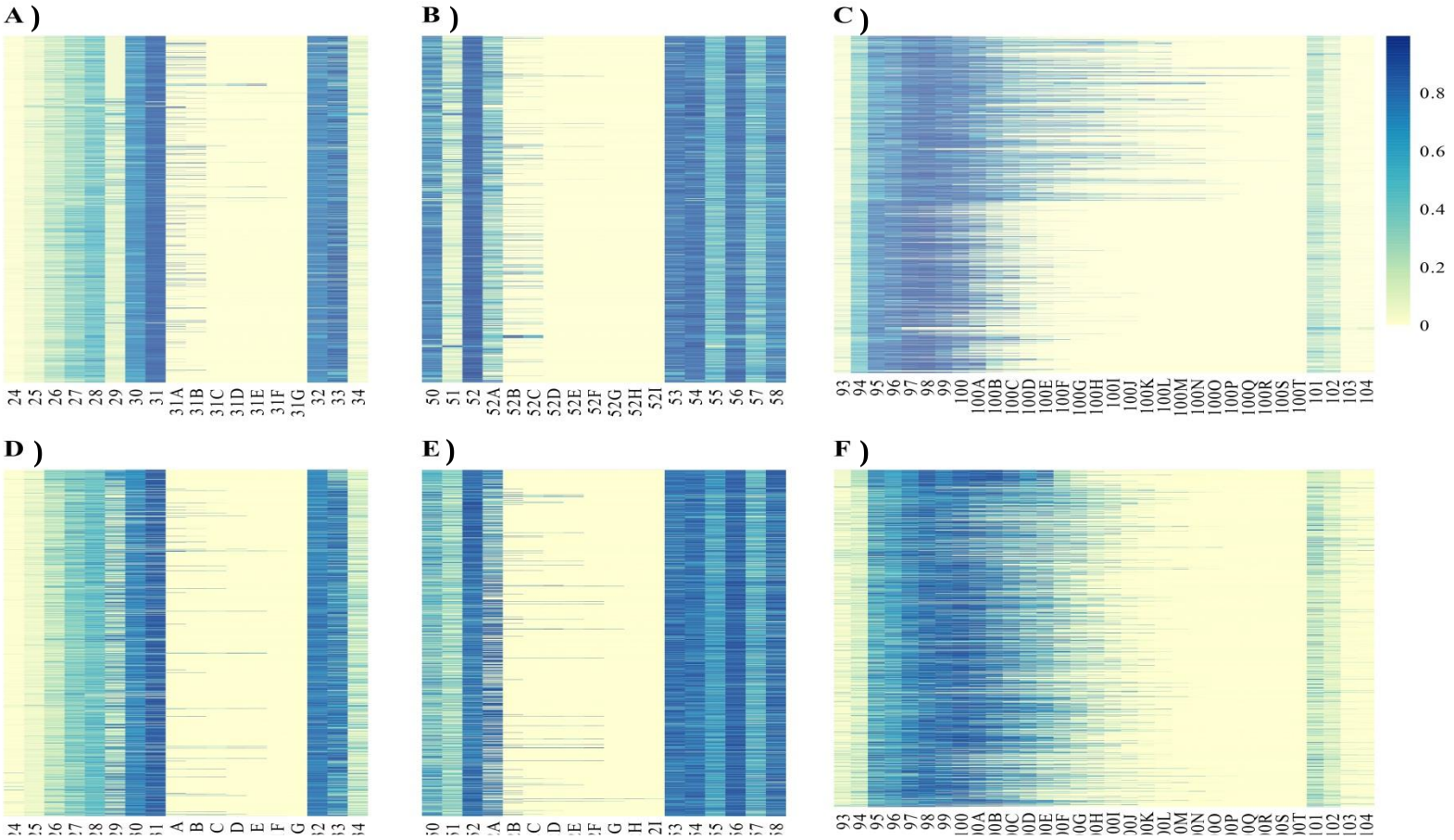


Figure 81. The heatmaps of the antigen binding prediction scores of HCDR1-3 regions for V_Hs (**A**-**C**) and V_HHs (**D**-**F**) were presented. The multiple sequences were aligned based on the Chothia numbering scheme. Blue cells: high binding scores, yellow cells: low binding scores. The two HCDR flanking columns indicate FW neighbouring FW residues. They served as predictors for the binding scores generated by the software. Empty columns (yellow) were intentionally inserted for the proper alignment of shorter lops, at alternating positions.

The visual representation by heatmaps allows for a comprehensive examination of the antigen binding prediction scores. The heatmaps revealed a comparable antigen binding profile for both V_Hs and V_HHs, showing shared characteristics in their binding patterns (**Figure 81**).

HCDR1 exhibited the lowest contribution to antigen binding ability, where the 31st position is the most influential residue. Additional residues at the 31st A-G positions with high binding scores are potentially significant in antigen binding within HCDR1. The 32nd and 33rd positions also have high binding scores.

In HCDR2, the 52nd position stands out with the highest binding score, at 52ndA-F positions. Similarly, in V_HHs, the incorporation of an additional residue between 52ndA-I positions consistently resulted in very high binding scores.

HCDR3 exhibited the greatest variability in paratope residues. Even locating at FW region, the 94th position in V_Hs displays high binding score, but it is less pronounced in V_HHs. In V_Hs, high-intensity binding scored residues was observed in the 95-100A range, while in V_HHs, this intensity shifted towards the C-terminus.

4.6.3. Differences between residues at the same location

Next, motif analysis was performed, providing a comprehensive view of the sequence conservation landscape in both V_Hs and V_HHs. Due to the high variability in HCDRs, the focus was shifted to conserved FW regions rather than the hypervariable HCDRs.

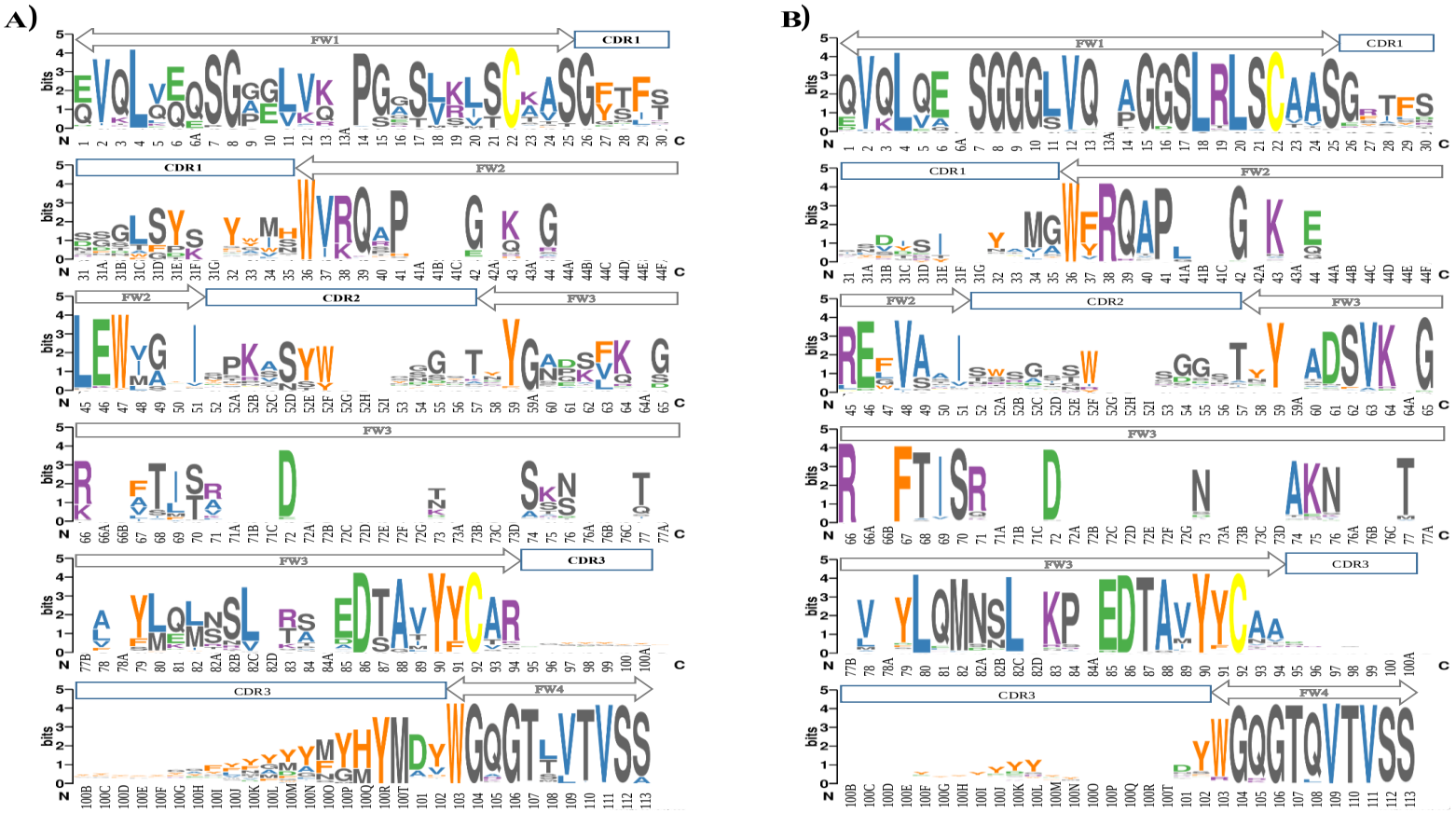


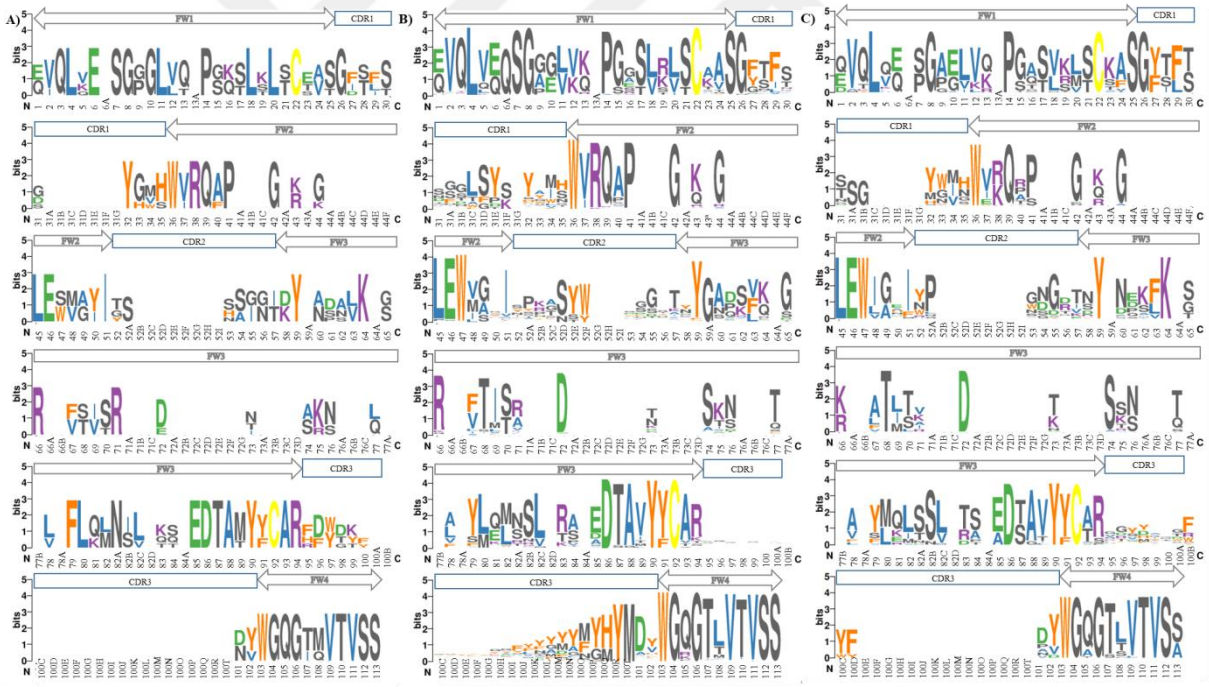
Figure 82. The graphical logo representation of the multiple sequence alignment of complete sequence datasets for **A)** V_Hs and **B)** V_HHs were shown. Alignment was based on the Chothia numbering scheme. The size of each logo corresponds to the relative frequency of amino acids, measured in bits. Colour codes are shown in coloured single letter amino acid codes: basic (K, R): purple, acidic (D, E): green, aliphatic (A, I, L, V): blue, aromatic (F, H, W, Y): orange, cysteine (C): yellow, and mostly small (G, M, N, P, Q, S, T): grey. Certain amino acids were categorized into multiple subclasses due to their overlapping physicochemical features. Grey and dark blue boxes on the

In FW1, a prominent distinction between V_Hs and V_HHs emerged in the initial amino acid residues, where V_HHs start with 'Q' instead of 'E,' present in V_H counterparts. V_Hs exhibited basic residues at the 9th, 10th, 12th, and 13th positions, rendering positive charge. 23rd position of V_Hs has positively charged 'K,' while V_HHs displayes 'A' in addition to 'K.'

A prominent divergence at FW2 hallmark residue positions 37th, 44th, 45th, and 47th, is observed. The 'VGLW' motif in V_Hs changes into 'FERF/YERF' in V_HHs. There are also minor differences at the 48th and 49th positions.

In FW3, numerous positions displayed minor differences, with the 94th position being a major one between V_Hs 'R' and V_HHs 'A'.

FW4 is a relatively shorter but highly conserved region compared to other FWs. In V_Hs recruits aromatic 'Y' or 'W' residues around the 100th position. Intriguingly, V_HHs have a substitution of 'W' at the 103rd position in FW4. Additionally, at the 108th position, 'Q' in V_HHs changes into 'L/T/S' in V_Hs (**Figure 82**).



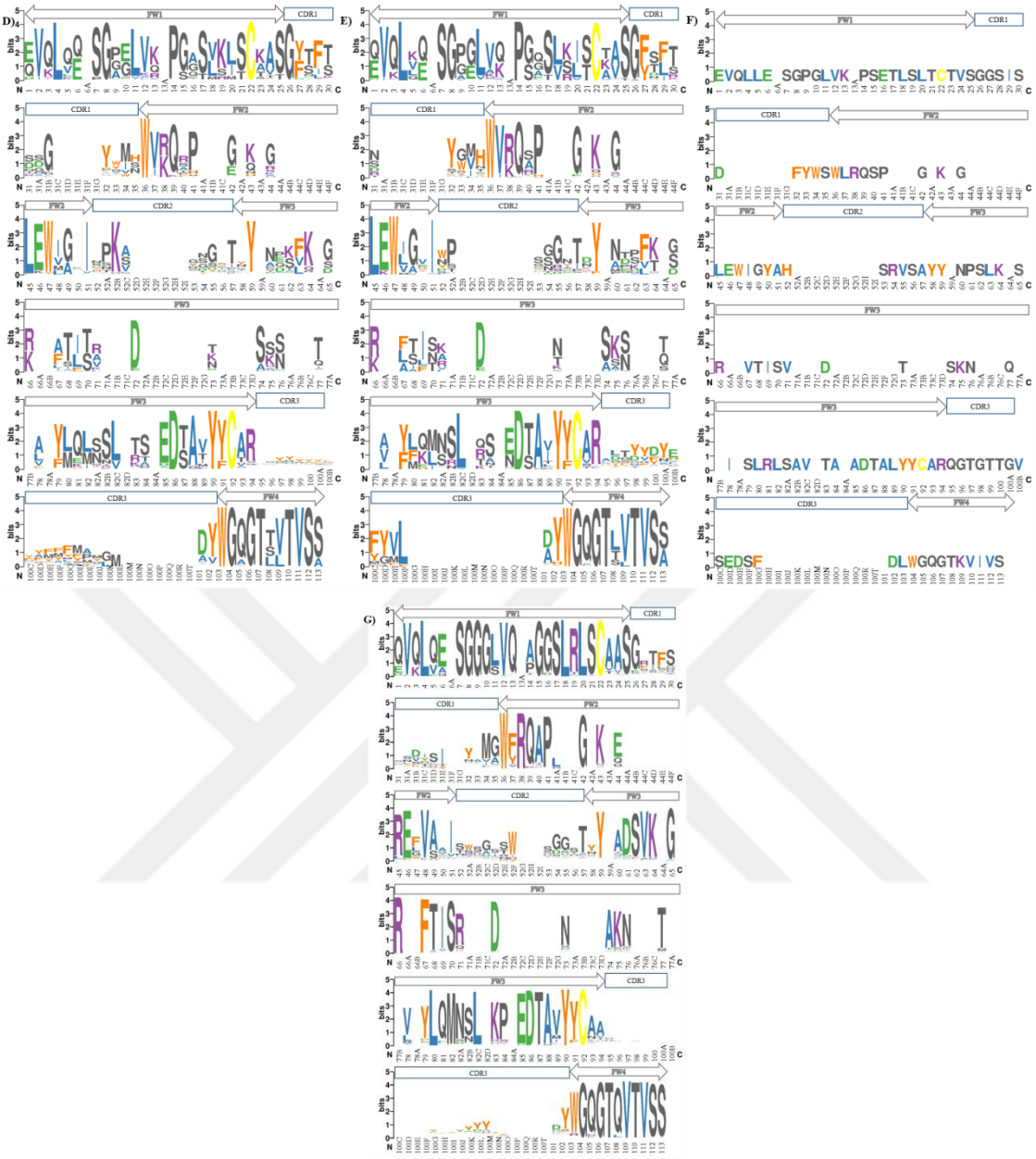


Figure 83. Graphical logo representations for multiple sequences of complete V_H dataset, based on species were generated. The sequences were aligned using the Chothia numbering scheme. The logos present the following species: **A**) *Cricetulus migratorius*, **B**) *Homo sapiens*, **C**) *Homo sapiens*/*Mus musculus*, **D**) *Mus musculus*, **E**) *Mus musculus*/*Homo sapiens*, **F**) *Pan troglodytes*, and **G**) *Rattus norvegicus*. The size of each logo corresponds to the relative frequency of amino acid positions, measured in bits. Single letter amino acids were colour-coded based on their biochemical properties, as explained in **Figure 82** legend. The variation in *Pan troglodytes* is low due to limited number of sequence information availability.

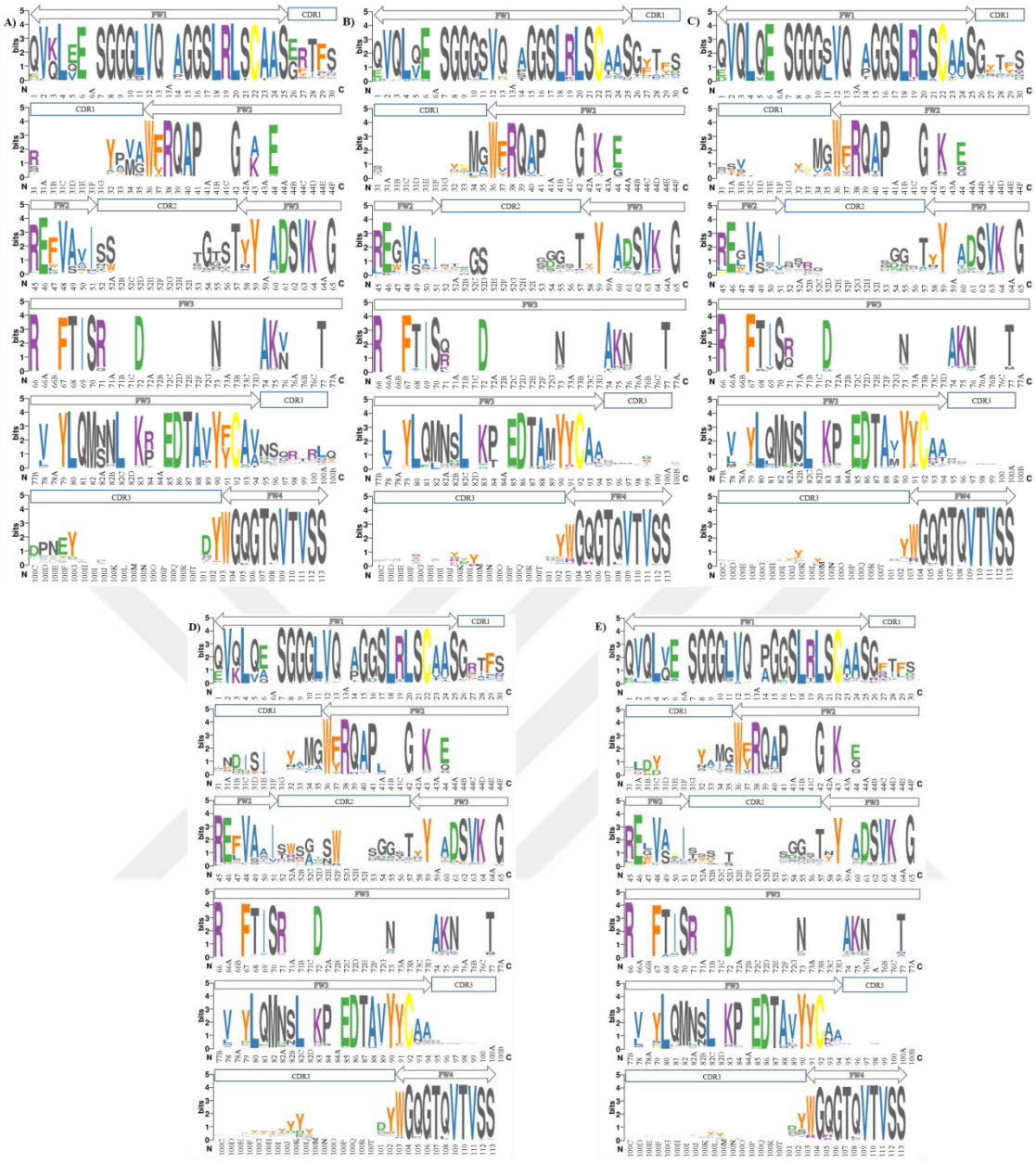


Figure 84. Graphical representation for multiple sequences of the V_HH datasets, represented by logos and organized by species was created. The sequence alignment conducted using the Chothia numbering scheme, is displayed for the following species: **A**) *Camelidae*, **B**) *Camelus bactrianus*, **C**) *Camelus dromedaries*, **D**) *Lama glama*, and **E**) *Vicugna pacos*. The size of each logo reflects the relative frequency of amino acid positions in bits. Single letter amino acids were colour-coded based on their biochemical properties, as explained in **Figure 82** legend.

Consistent patterns in V_H and V_HH datasets emerged when clustered into species-specific subsets (**Figure 83** and **84**). However, minor variations at the subset levels were observed due to a relatively smaller number of sequences in some species datasets.

4.6.4. Non-canonical cysteine bond profiles

The introduction of additional intra-disulphide bonds, act as structural reinforcements especially in antibodies with extended HCDR3 regions. To assess their contribution to antibody stability in different formats, a comprehensive analysis of non-canonical disulphide bridges was performed. Both amino acid sequences and 3D structural conformations were investigated.

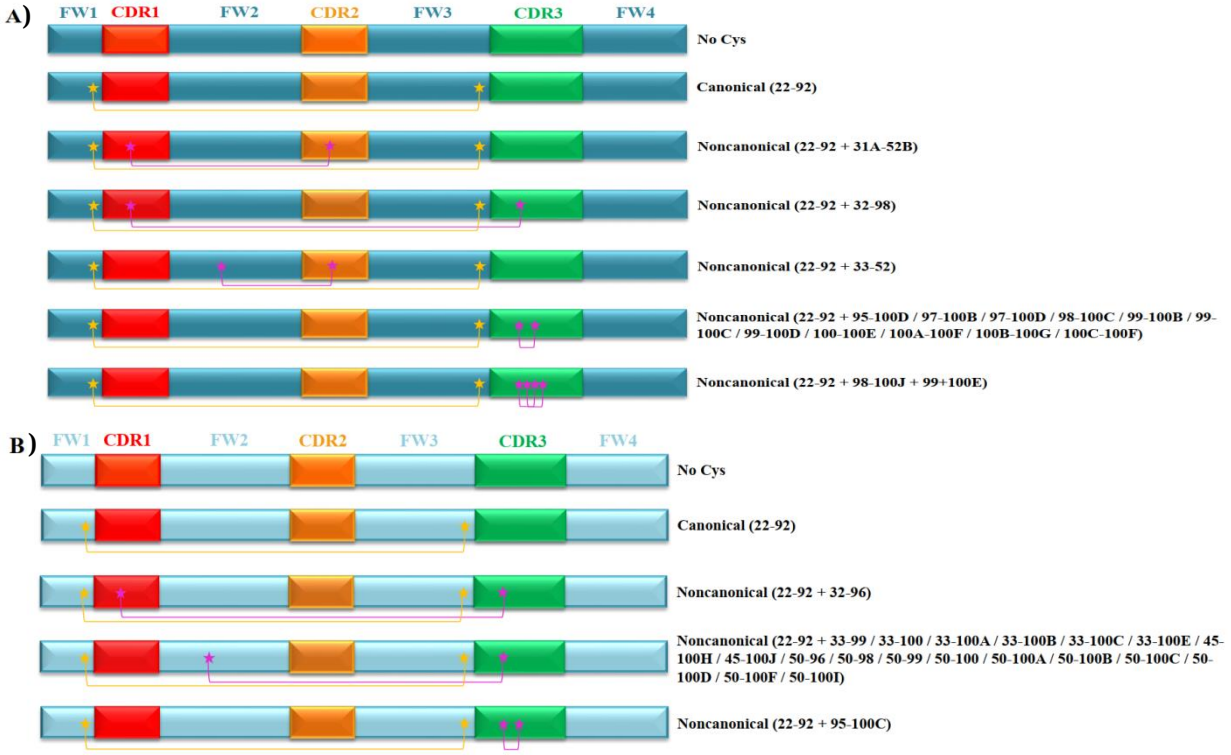


Figure 85. Canonical and non-canonical disulphide bond profiles of **A**) V_Hs and **B**) V_HHs were depicted. The following regions were highlighted: V_H FW regions were highlighted in dark blue, V_HH FW regions in light blue, HCDR1 in red, HCDR2 in orange, and HCDR3 in green. Canonical disulphide bonds formed between the 22^{nd}C-92^{nd}C were denoted by yellow stars connected by a line. Non-canonical cysteine bonds are represented with purple stars joined by a line. The shapes of FWs and HCDRs are not to scale. The positioning of the stars representing canonical and non-canonical cysteine residues serves as a putative visual representation and does not reflect their precise locations.

Results revealed that while V_Hs may have exhibited various non-canonical cysteine types, V_HHs tended to have a higher frequency of these additional bonds. Species specific cysteine residues forming non-canonical disulphide bridges in V_Hs and V_HHs were also investigated.

Table 131. The total count of canonical and non-canonical disulphide bridges in V_H dataset, based on species was listed.

Species	No Bond	Canonical Bonds	Non-Canonical Bonds	Total V_Hs
Cricetulus Migratorius	-	10	-	10
Homo Sapiens	1	2456	130	2457
Homo Sapiens/*Mus Musculus*	-	26	-	26
Mus Musculus	3	2287	2	2290
Mus Musculus/*Homo Sapiens*	-	188	-	188
Pan Troglodytes	-	4	-	4
Rattus Norvegicus	-	53	-	53
Rattus Norvegicus/*Homo Sapiens*	-	2	-	2
Total	**4**	**5026**	**132**	**5030**

Table 132. The overall count of canonical and non-canonical disulphide bridges in V_HH dataset, classified by species was listed.

Species	No Bond	Canonical Bonds	Non-Canonical Bonds	Total V_HHs
Camelidae	-	65	3	65
Camelus bactrianus	-	63	4	63
Camelus dromedaries	-	263	26	263
Lama glama	1	669	16	670
Vicugna pacos	-	199	20	199
Total	**1**	**1259**	**69**	**1260**

The encounter frequency of non-canonical bonds in V_Hs (2.6%) was significantly lower than that in V_HHs (5.5%). For V_Hs, four distinct types of non-canonical bonds were identified; between HCDR1-HCDR2, HCDR1-HCDR3, FW2-HCDR2, and HCDR3-HCDR3. In V_HHs, three categories of non-canonical cysteine pairs were identified, in addition to canonical bond. These formed between HCDR1-HCDR3, FW2-HCDR3, and HCDR3-

HCDR3, where HCDR1-HCDR3 and HCDR3-HCDR3 non-canonical disulphide bonds were found to be common in both formats (**Figure 85**). Although V_Hs exhibited a lower frequency of non-canonical bonds than V_HHs, the variety in their positions was higher than that of V_HHs (**Table 131** and **Table 132**).

Species-specific extra bonds were observed in both datasets. *Homo sapiens* exhibited a relatively high frequency of non-canonical bonds (**Table 131**). *Camelus dromedaries* displayed the highest diversity in terms of extra bonds compared to other species. Interestingly, *Vicugna pacos* and *Lama glama*, which had the shortest average V_HH HCDR3 length (**Table 130**), also had the highest occurrence of extra cysteine bonds (~10%) in their V_HHs (**Table 132**).

4.6.5. Stability enhancing salt bridge profiles around HCDR3

Salt bridges played a crucial role in providing stability of antibodies. Particularly at the 94th position, V_Hs exhibited conserved 'R' residues, while V_HHs featured 'A' residues (as previously revealed at **Figure 82**). The presence of a positively charged residue, such as 'R' or 'K', at the 94th position facilitated favourable electrostatic interactions with negatively charged regions on the antigen, enhancing binding specificity. To comprehensively compare salt bridges in V_Hs and V_HHs, a detailed examination was conducted.

Table 133. The total count of canonical salt bridges (R94 – D101) and non-canonical salt bridges in V_Hs (combinations of R/H/K 94 – D/E101 excluding the canonical one) were indicated on species level.

Species	Canonical Salt Bridges	Non-canonical Salt Bridges	No Salt Bridge	Total V_Hs
Cricetulus Migratorius	5	-	5	10
Homo Sapiens	1409	282	766	2457
Homo Sapiens/*Mus Musculus*	10	-	16	26
Mus Musculus	1291	44	955	2290
Mus Musculus/*Homo Sapiens*	80	5	103	188
Pan Troglodytes	4	-	-	4
Rattus Norvegicus	32	10	11	53
Rattus Norvegicus/*Homo Sapiens*	-	2	-	2
Total	**2831**	**343**	**1856**	**5030**

Table 134. The total count of canonical and non-canonical salt bridges in V_HHs was listed. The canonical and non-canonical salt bridge definition is as explained at **Table 133**.

Species	Canonical Salt Bridges	Non-canonical Salt Bridges	No Salt Bridge	Total V_HHs
Camelidae	-	-	65	65
Camelus bactrianus	-	1	62	63
Camelus dromedaries	2	-	261	263
Lama glama	12	13	645	670
Vicugna pacos	5	2	192	199
Total	**19**	**16**	**1225**	**1260**

In V_Hs, the canonical salt bridge primarily occurs between the 94th 'R' residue and the 101st 'D' residue. However, upon detailed examination, various combinations of acidic and basic residues were observed in V_Hs beyond this canonical pairing that contradicts the previously reported conserved pattern in the literature. Contrasting with V_Hs, V_HHs exhibited a lower incidence of this canonical salt bridge interaction. Specifically, 63.1% of V_Hs had salt bridges between 94th and 101st positions (**Table 133**), while V_HHs showed only a 2.7% of incidence (**Table 134**). The absence of this canonical salt bridge in many V_HHs resulted in structural diversity and adaptability, with 97% lacking any salt bridge at this position. The structural implications of these findings, especially in the context of humanisation/camelisation studies, underscore the significance of residues at the 94th and 101st positions in antibody stability and function.

4.6.6. Statistical analysis on physicochemical properties of FW residues

As an extended analysis, the distribution of physicochemical amino acid categories within the FW regions of V_H and V_HH was investigated. This investigation aims to pinpoint specific physicochemical characteristics that potentially influencing the folding, stability, or affinity of V_H and V_HH functional structures. An amino acid classification system (**Table 90** and **91**) and molecular charge and hydropathy properties was explored, and the percentage of residues in each sequence for each category was calculated.

For this purpose EMBOSS Pepstats online tool was used to calculate certain amino acid features of the sequences from the both datasets. The resulting values of this assessment were used in further analysis. We benefited the integration of statistical approaches to augment our findings. Both logistic regression and machine learning approaches were applied

together to assess these comparisons collectively for all FW regions (FW1-4) and individually. The logistic regression model focused on "Beta" and "Exp(Beta)" values.

Table 135. The logistic regression analysis, utilizing the EMBOSS Pepstats output, included all FW regions (FW1-4) collectively, along with various physicochemical properties. In the analysis, variables not displayed in the table were deemed statistically insignificant based on their VIF values.

		Logistic Regression			Confidence Interval	
FWs	**Physicochemical Property**	**Beta**	**Exp(Beta)**	**p-values**	**Lower Bound**	**Upper Bound**
	Constant	-47.781	0.000	$1.014.e^{-04}$		
FW1	Tiny	0.517	1.678	$1.823.e^{-08}$	1.4205	2.1739
	Small	-0.312	0.732	$9.455.e^{-04}$	0.5942	0.8749
	Aliphatic	0.206	1.228	$2.241.e^{-02}$	0.9629	1.4402
	Aromatic	0.745	2.107	$1.002.e^{-02}$	1.2448	4.0962
	Basic	-0.366	0.693	$1.569.e^{-05}$	0.5717	0.8254
	Acidic	0.162	1.176	$4.168.e^{-02}$	1.0103	1.4292
	Polar	0.146	1.157	$1.956.e^{-02}$	1.0096	1.2873
FW2	Tiny	0.635	1.887	$1.358.e^{-13}$	1.5956	2.4403
	Aromatic	0.247	1.279	$4.606.e^{-03}$	1.0703	1.5909
	Aliphatic	-0.328	0.720	$2.898.e^{-04}$	0.5298	0.8343
	Polar	-0.150	0.861	$1.939.e^{-02}$	0.7367	0.9975
	Basic	0.330	1.392	$1.279.e^{-04}$	1.1885	1.8180
	Hydrophobic	0.328	1.388	$4.172.e^{-06}$	1.2448	1.8178
	Hydrophilic	0.647	1.909	$1.676.e^{-10}$	1.5592	2.4595
FW3	Small	0.578	1.782	$8.100.e^{-10}$	1.5255	2.7124
	Aromatic	-0.360	0.698	$2.448.e^{-02}$	0.4978	0.9796
	Polar	-0.311	0.733	$8.081.e^{-03}$	0.5951	0.9799
	Basic	-0.342	0.710	$8.842.e^{-03}$	0.5525	0.9388
	Hydrophobic	-0.265	0.767	$3.919.e^{-03}$	0.5376	0.9429
FW4	Tiny	-0.153	0.858	$6.413.e^{-04}$	0.7049	0.9999
	Aliphatic	-0.170	0.844	$1.539.e^{-04}$	0.6246	0.9680
	Aromatic	-0.577	0.562	$2.182.e^{-09}$	0.4168	0.6628

Hydrophilic	0.867	2.380	$7.299.e^{-20}$	1.8508	3.069

The Exp(Beta) values exceeding 1 signify an overrepresentation of physicochemical characteristics (specifically, FW1-Tiny, Aliphatic, Aromatic, Acidic, and Polar) in V_HHs. Conversely, values below 1 indicate the opposite trend for features such as FW1-Small and Basic. According to the results, FW1 residues are found to be more hydrophilic, FW3 displayed specifically hydrophilic residues, and FW4 is characterized as hydrophilic compared to other framework regions (**Table 135**).

To validate feature selection, the Random Forest – Boruta algorithm, a wrapper around the Random Forest algorithm, was employed. This dual approach, combining statistical and machine learning methods, provided a comprehensive understanding on the analysed data. Feature importance scores were ranked for each amino acid property, validating findings through established statistical techniques.

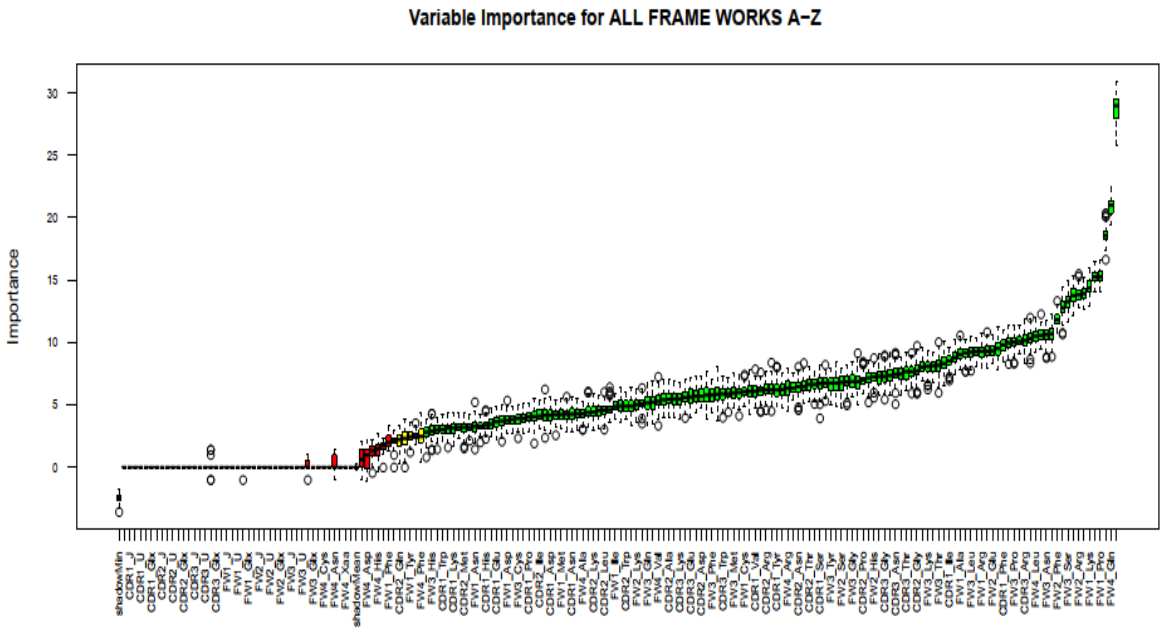


Figure 86. A box-and-whiskers plot to show the Random Forest – Boruta algorithm ranking of the variable importance scores of features utilized in the logistic regression modelling analysis was generated. The box represents the interquartile range, while the whiskers extended to show the full range of scores. The position of each feature in the plot indicates its importance, with higher-ranking features positioned towards the top.

The VI scores for each modeled feature were systematically ranked (**Figure 86**). The results show the accuracy of the logistic regression analysis model, emphasizing the

importance of specific features like hydrophilic FW4 residues and tiny FW1 residues. The ranking of amino acids in different regions highlights the significance of FW1 and FW3, supported by motif conservation analysis. Additionally, the shorter length of FW4, with a low constant value, may hinder its lower importance in humanisation/camelisation approaches or potential masking of individual features due to its small range.

4.6.7. Physicochemical characteristics of individual residues

Next, the physicochemical properties of V_Hs and V_HHs were examined at the residue level. The analysis revealed important variances in certain physicochemical characteristics between V_Hs and V_HHs at specific residue positions.

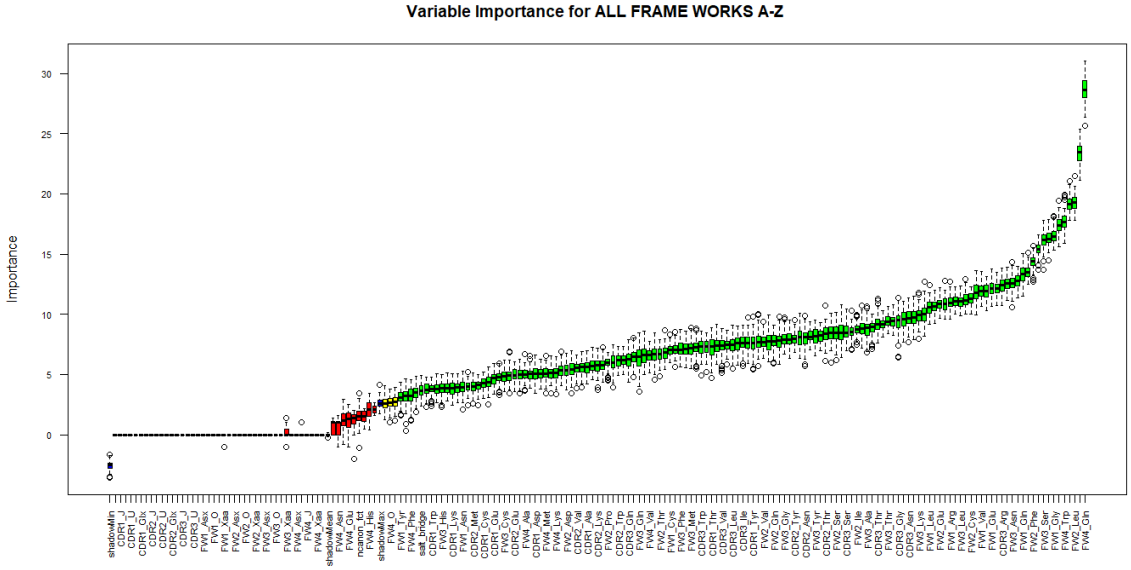
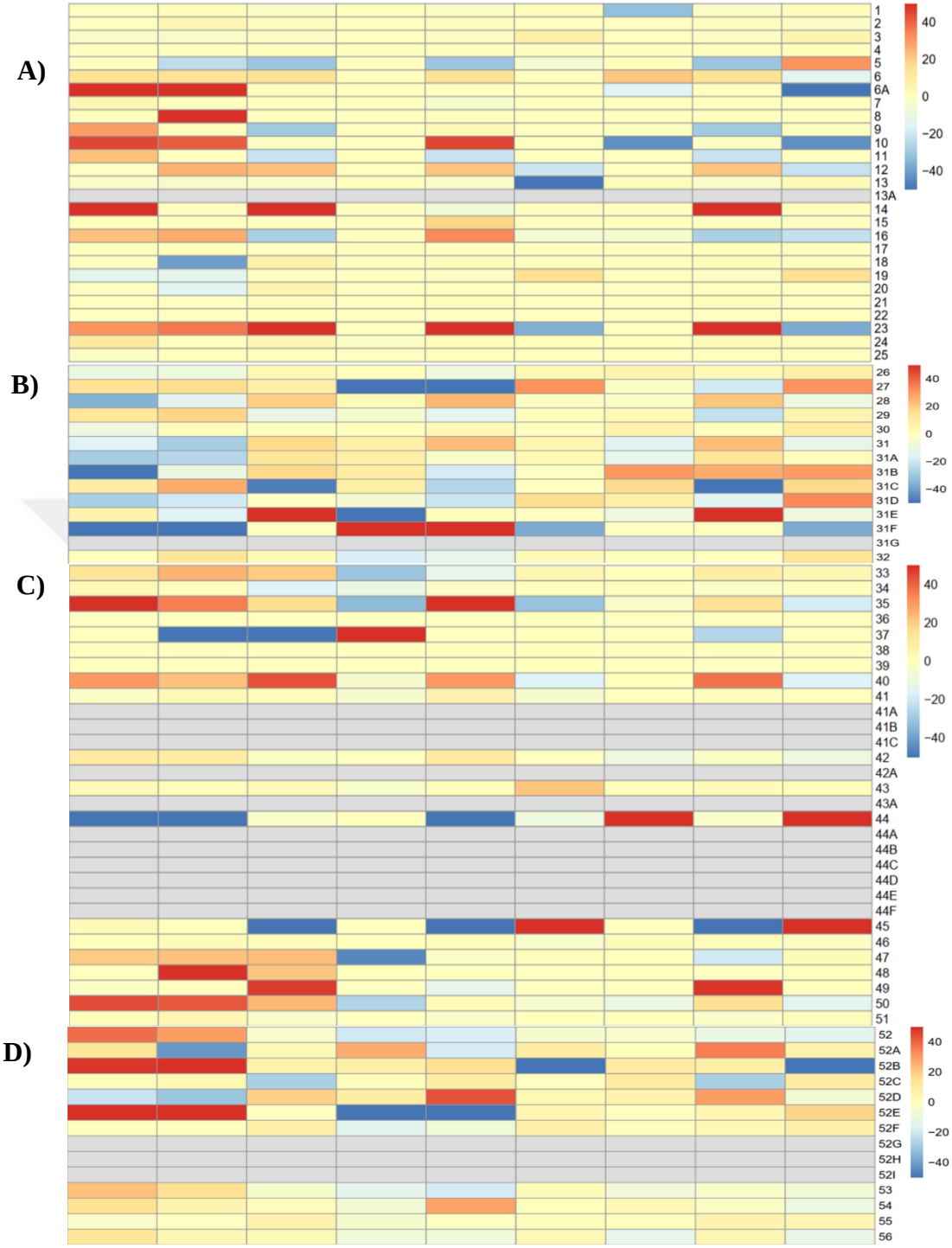
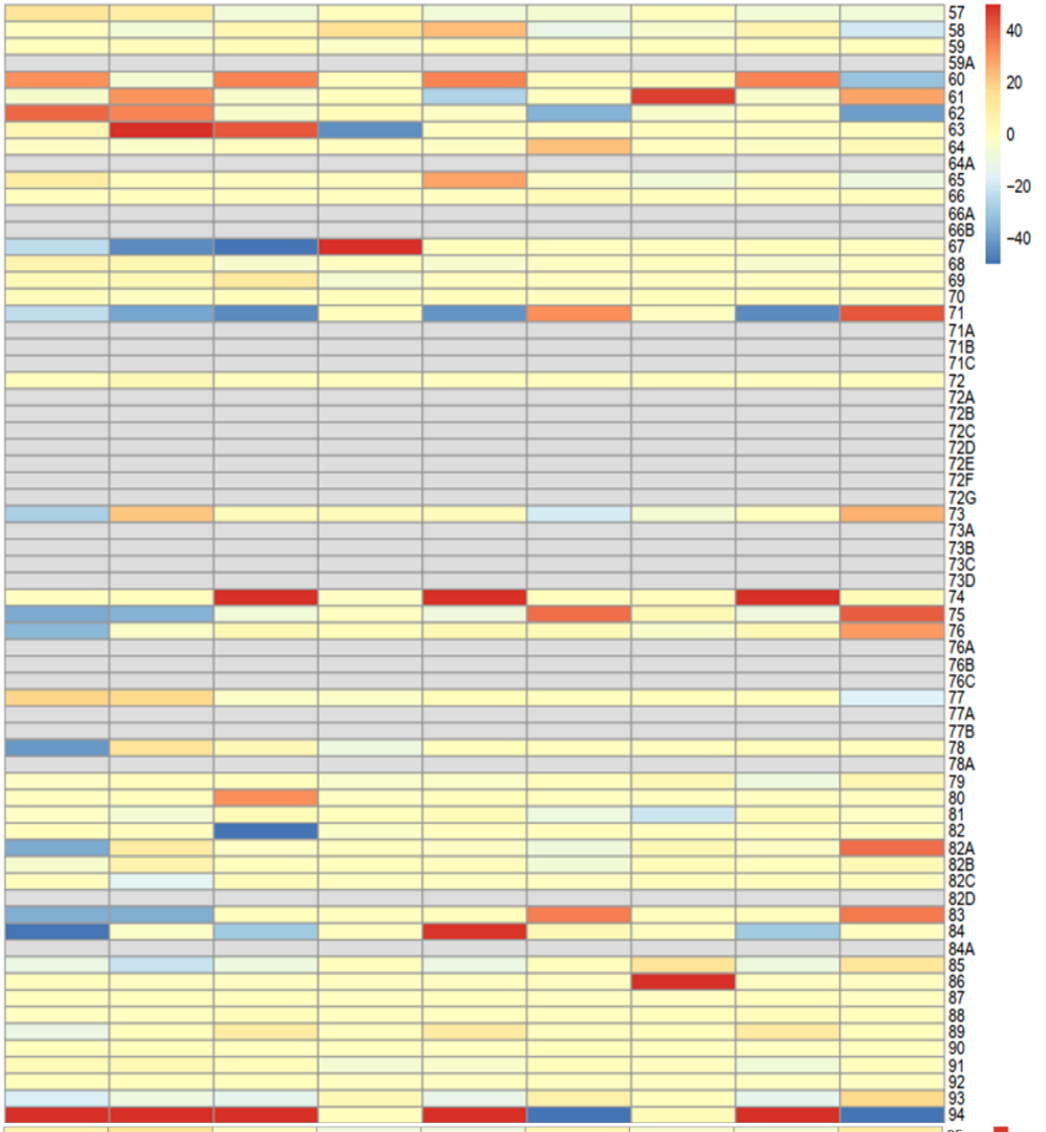


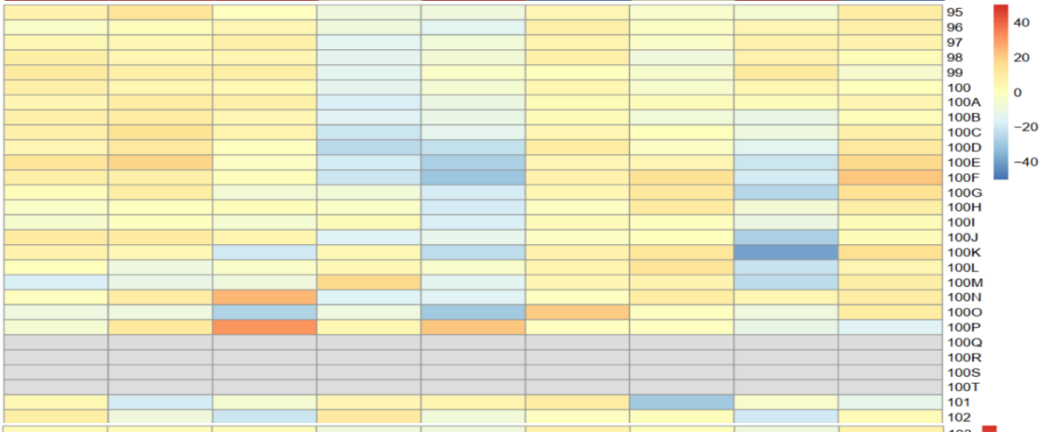
Figure 87. The box-and-whiskers plot of the ranking of variable importance scores for each amino acid–region association based on the Random Forest – Boruta algorithm was illustrated. Red boxes denote variables that did not significantly contribute to accurate classification. Green boxes highlight amino acids that significantly contributed to the classification and were subsequently selected for further evaluation. The centres of the boxes correspond to the median, while the lower and upper hinges represent the first and third quartiles of the data. The upper and lower whiskers extend from the hinges to the largest and smallest values, respectively, but no further than 1.5* times the interquartile range from the hinges. Any data beyond the end of the whiskers was plotted individually.


A)
B)
C)
D)
1
2
3
4
5
6
6A
7
8
9
10
11
12
13
13A
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
31A
31B
31C
31D
31E
31F
31G
32
33
34
35
36
37
38
39
40
41
41A
41B
41C
42
42A
43
43A
44
44A
44B
44C
44D
44E
44F
45
46
47
48
49
50
51
52
52A
52B
52C
52D
52E
52F
52G
52H
52I
53
54
55
56
40
20
0
−20
−40

E)


57
58
59
59A
60
61
62
63
64
64A
65
66
66A
66B
67
68
69
70
71
71A
71B
71C
72
72A
72B
72C
72D
72E
72F
72G
73
73A
73B
73C
73D
74
75
76
76A
76B
76C
77
77A
77B
78
78A
79
80
81
82
82A
82B
82C
82D
83
84
84A
85
86
87
88
89
90
91
92
93
94
40
20
0
−20
−40

F)


95
96
97
98
99
100
100A
100B
100C
100D
100E
100F
100G
100H
100I
100J
100K
100L
100M
100N
100O
100P
100Q
100R
100S
100T
101
102
40
20
0
−20
−40

G)

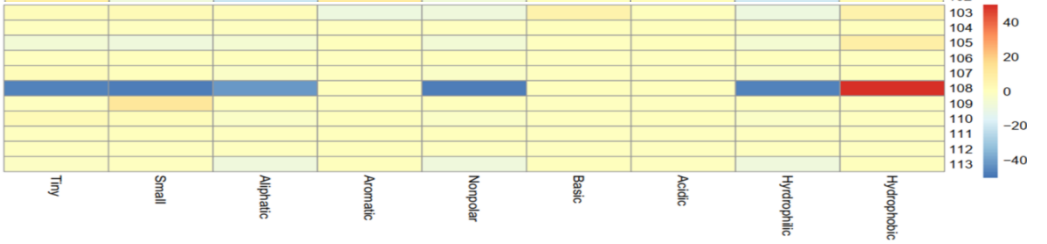

103
104
105
106
107
108
109
110
111
112
113
40
20
0
−20
−40
Tiny
Small
Aliphatic
Aromatic
Nonpolar
Basic
Acidic
Hydrophilic
Hydrophobic

Figure 88. Heatmaps of variations in physicochemical characteristics among residues in V_Hs and V_HHs were visually represented. Individual heatmaps show **A**) FW1, **B**) HCDR1, **C**) FW2, **D**) HCDR2, **E**) FW3, **F**) HCDR3, and **G**) FW4 residues. Rows represent individual residues, while columns denote specific physicochemical properties identified through logistic regression analysis. Cell colour signifies the antibody format with a higher percentage of the specified property in the residue. Red cells denote a higher prevalence in V_HHs, while blue cells signify a higher occurrence in V_Hs. Gray rows indicate alternating positions where either V_Hs or V_HHs lack the corresponding value, preventing a difference calculation. The colour scale ranges from 50 to -50, with values exceeding the set maximum or minimum being adjusted accordingly.

FW1, positions 6, 6A, 10 and 23, exhibits similar distinctions in terms of physicochemical characteristics. FW2, in addition to its hallmark residues, demonstrates distinct disparities at the 35th, 40th, and 50th positions. V_Hs tended to favour aromatic and hydrophobic residues, while V_HHs displayed a higher prevalence of smaller and non-polar residues at these positions. FW3 has similar trends at the 60th, 62nd, and 94th positions. Intriguingly, FW4 lacked such distinct residue patterns, except for the 108th position, favouring small, non-polar, and hydrophilic residues in V_Hs, and hydrophobic residues in V_HHs (**Figure 87** and **88**).

A particularly interesting aspect among the physicochemical characteristics was the distribution of hydrophobic and hydrophilic amino acids in the FWs of V_Hs and V_HHs, respectively. The absence of the V_H-V_L interaction in V_HHs led to a higher concentration of hydrophilic residues, aiding in proper domain folding and packaging. With few exceptions, such as the 27th position favouring aromatic, non-polar residues in V_Hs, and slightly favouring hydrophilic residues, the overall differences in non-alternating residues between V_Hs and V_HHs were minimal. This observation underscored that major distinctions in HCDR regions were primarily driven by target recognition rather than species origin. The preference for less polar and more hydrophilic residues in HCDR3s compared to those of V_HHs was evident across multiple residues, implicating the importance of HCDR3 in epitope recognition.

4.7. The host lab studies on abroad

The efforts dedicated to the refinement of recombinant protein production as described on the **Sections 4.3** and **4.4** of this thesis. As a result, a standout competitor was identified, surpassing other sdAbs in both quality and quantity (**Figure 89**). This optimal V37Y sdAb

variant was defined as prime candidate, and decided to be tested in mice studies planned on abroad in the context of EMBO fellowship. The primary objective was to unravel the intricate pharmacokinetic properties of the chosen sdAb variant, meanwhile testing it *in vivo* in a mouse model that is superior to zebrafish.

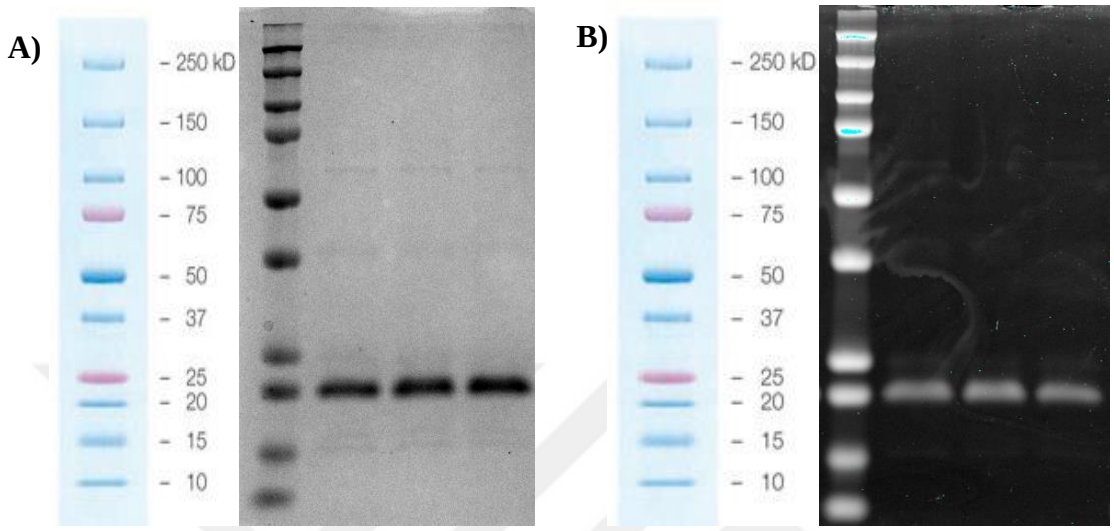


Figure 89. The SDS-PAGE and WB analysis of V37Y sdAb after individually optimizing it in terms of purity and yield. The samples were from three different flasks. **A**) In the SDS-PAGE analysis, the gel loaded with 1000 ng sdAb sample. **B**) For the WB analysis in, the samples were prepared as 100 ng. For SDS-PAGE, Coomassie Blue dye was used for protein visualization. In the WB, the sdAbs were probed using a monoclonal Anti-Flag M2 IgG antibody (1:5000). Then as a secondary antibody, polyclonal HRP-Conjugated secondary Goat Anti-mouse IgG (1:10000) was used for visualization.

4.7.1. Preparation of functional camelised anti-VEGF sdAbs at home lab

The V37Y sdAb samples previously generated in the home laboratory were thawed and combined into a single sample within a 5 ml tube. Subsequently, the total volume was decreased to 4.5 ml utilizing Amicon tubes with a 10 kDa cut-off. The V37Y sdAb was filtrated and subsequently subjected to SEC analysis (**Figure 90**). The concentration of starting protein sample was at 3.660 mg/ml. At the end of SEC procedure, 15 different fractions were obtained, where each fraction approximately contains 2 ml of sample (**Table 136**). The concentration each fraction was measured using NanoDrop spectrophotometer.

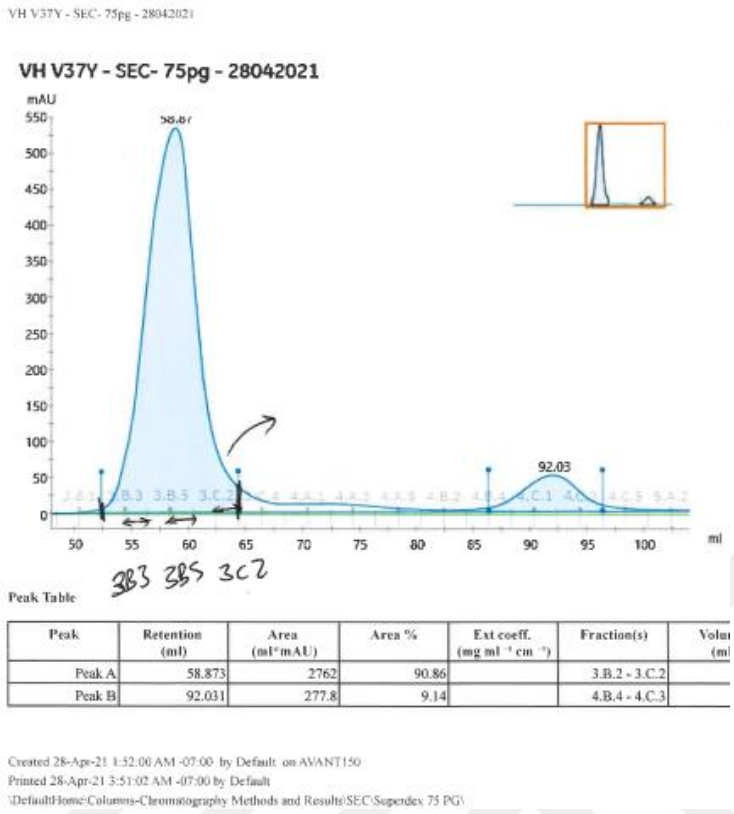


Figure 90. SEC analysis of V37Y sdAb variant performed in the home lab at IBG, Turkey.

Table 136. The concentration of the V37Y sdAb variant fractions collected after SEC analysis were measured using NanoDrop spectrophotometer.

Fraction	Protein concentration (mg/ml)	A280	A260/A280
3B2	-0.018	-0.046	4.85
3B3	0.247	0.68	0.36
3B4	0.796	0.64	1.04
3B5	0.771	2.02	0.46
3C1	0.308	0.822	0.29
3C2	0.027	0.187	-0.67
3C3	0.013	-0.037	6.68

After the SEC run (**Figure 90**), the collected fractions were stored in the fridge until their further purity evaluation by SDS-PAGE. The SDS-PAGE indicated that the fractions did not meet the necessary purity threshold for advancing into animal studies (**Figure 91**). As a result, only relatively pure fractions were pooled and then transported to VUB using a temperature-controlled system.

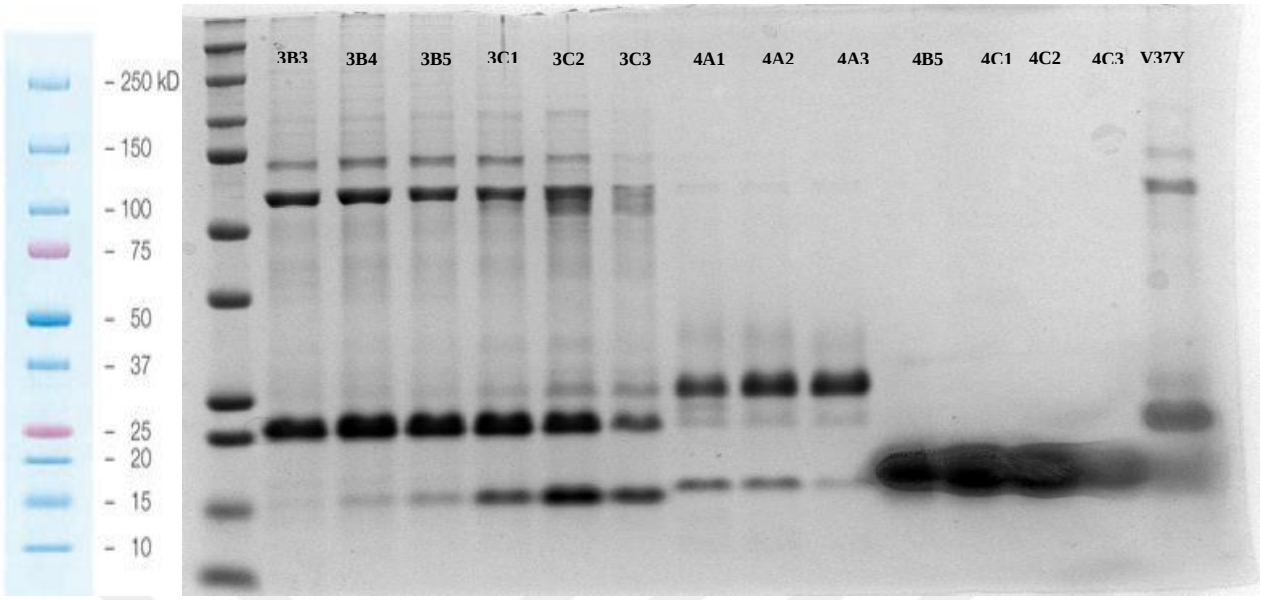


Figure 91. SDS-PAGE analysis of V37Y sdAb after SEC performed in home lab at IBG. The collected fractions (**Table 136**) were loaded into the gel at 1000 ng. V37Y lane indicates the protein sample being analysed by SEC before loading into the column.

When the samples were received, the total volume was inadequate for our purposes, as some of the samples would be consumed during the test trials, which had not been accounted for. This necessitated an additional request for the transportation of V37Y sdAb samples. However, the remaining stock at the home lab was not suitable for immediate use, as reflected in the results of SEC scan (**Figure 92**) performed at the host lab.

While the samples were on route, the type of tumour model for animal studies were defined based on the availability of the cell lines in the host institute and considering the choice of a human tumour with relatively high VEGF expression levels to test our sdAbs *in vivo*. Simultaneously, an ethical committee application was written, where the planned animal studies were explained in detail.

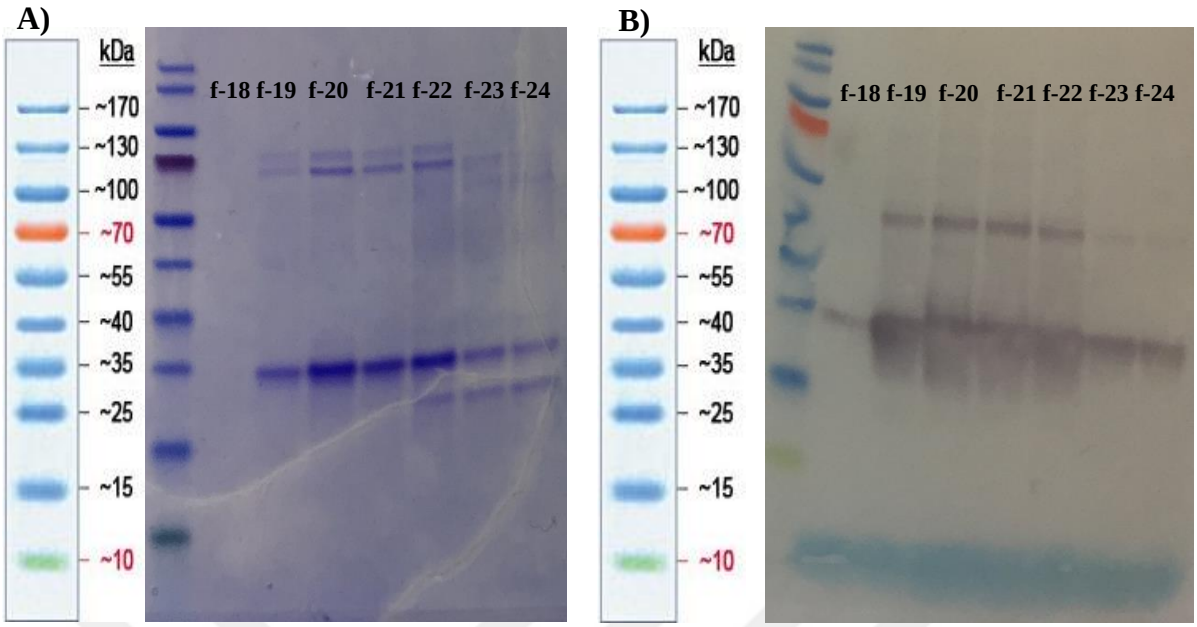


Figure 93. The results of SDS-PAGE and WB analyses of the V37Y sdAb subsequent to an additional SEC procedure performed in the host lab at VUB, were shown. The protein concentrations were 1000 ng for SDS-PAGE gel **A**), where WB gel was loaded with 100 ng of samples (if the sample concentration is lesser than this value, than it was loaded into the gel as it was) **B**). SDS-PAGE gel was stained with Coomasie blue dye, where WB membrane was probed with anti-His mouse antibody (1:2000), followed by an incubation with anti-mouse HRP antibody (1:2000).

The anticipated molecular weight of our V_H fragment was approximately 18.90 kDa (**Table 107)**, but according to the WB result in **Figure 93**, the dimer forms were also present in addition to monomeric sdAbs, which is not also a favourable feature in animal studies. As a result, all the delivered V37Y sdAb samples were discarded and the planned animal studies were put on pause as we pursued an alternative strategy for the production of sdAbs with a novel design without the additional tags that could potentially cause immunogenicity in mice.

When the V37Y sdAb was found to be unuseful for the animal studies, we planned an alternative strategy so that the preparations made were not wasted. Collaboration was made with an expert in antibody design and engineering from VUB. To ensure both the quality and quantity of the sdAb variants, our samples decided to be redesigned. The following studies were initated and preliminary preparations were completed for animal studies.

4.7.2. Ethical committee approval

The submission to the "Ethical Committee for Animal Experiments" ECD at VUB resulted in the approval of the requested allocation of mice: a total of 96 athymic nude mice

(Crl:NuFoxn1^{NU}), detailed in both **Table 80** and **Table 81**. Upon this approval, the reserved animals were started to taken care, meanwhile the cell culture studies were initiated on HT-29 cell line.

4.7.3. Re-design of anti-VEGF single domain antibodies

After the collaborative redesign of the sdAbs, the required molecular tools (primers, plasmid templates) were obtained. The cloning, transformation, and production studies were conducted at the VUB, Ixelles campus.

In the home lab, the sdAb expression utilized using pET-17b plasmid as the vector. However, with these new designs, an alternative expression vector pHen6 was employed. This vector has been extensively used in various nanobody expression studies in the host lab, in fact it was first discovered in there (Conrath et al. 2001).

4.7.3.1. Insert design

The studies commenced by isolating the plasmids send from the home lab by retrieving them from filter paper. Then the sdAbs sequences were specifically targeted and amplified using forward and reverse primers (**Table 17**). These primers were also designed *de novo* for this project to obtain new sdAb designs (**Figure 94**).

VH WT from home lab

HMKYLLPTAAAGLLLLAAQPAMAMSASEVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGM
NWVRQAPGKGLEWVGWINTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYY
CAKYPHYYGSSHWYFDVWGQGTLVTVSSVD**IEGR**DYKDDDDKAADYKDDDDKAADYKDDDD
KAA*HHHHHHH*–LE

VH WT re-desgin of the host lab

DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGKGLEWVGWINTYTGEPTYA
ADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYFDVWGQGTLVTVS
S*HHHHHH*

Figure 94. Insert design of Bevacizumab-derived, codon optimised WT V_H fragment. For a fuctional protein production, the sdAbs were redesigned. The first non-mutated WT Bevacizumab V_H sequence design (including the tags) belongs to the home lab, where the second fragment was redesigned considering the nanobody scaffold in the host lab based on the home lab template. The restriction enzymes, the leader sequence, tags and protease recognition sites were highlighted as follows: NdeI sequence and first nucleotide of the start codon were highlighted in cyan, PelB sequence and the last two nucleotides of the start codon in green, NheI in grey, Codon optimised V_H sequence in navy, SalI in dark gold, Factor Xa

protease in bold text, $(Flag\ Tag)_3$ in underlined text, His_7 tag in italicised text, Stop codon in red, and XhoI in pink.

```
beva          1 EVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGKGLEWVGW     50
                :||||||||||||||||||||||||||||||||||||||||||||||||||
vub           1 DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGKGLEWVGW     50

beva         51 INTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYP    100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
vub          51 INTYTGEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYP    100

beva        101 HYYGSSHWYFDVWGQGTLVTVSSVDIEGRDYKDDDDKAADYKDDDDKAAD    150
                ||||||||||||||||
vub         101 HYYGSSHWYFDVWGQGT---------------------------------    117

beva        151 YKDDDDKAAHHHHHHH    166
```

Figure 95. The pairwise sequence alignment and comparison of the new and prior WT sdAb designs using EMBOSS Needle (Needleman and Wunsch 1970) tool. beva: home lab V_H fragment design, vub: host lab V_H fragment design. The HCDRs were underlined: HCDR1 was shown in green, HCDR2 in yellow, and HCDR3 in red.

A pairwise sequence alignment and comparison between the new and previously design WT sdAb revealed the modified residues in the V_H sequence (**Figure 95**). Additionally, during the PCR amplification extra nucleotides were appended at both ends of the nucleotide sequence, to make them compatible with pHen6 vector (**Table 137**). The translated amino acid sequences for each redesigned sdAb were listed in **Table 138**.

Table 137. List of host lab codon optimised Bevacizumab V_H fragment nucleotide sequences of redesigned sdAbs The introduced mutations were indicated using the following format: Mutated residue: [Original amino acid][Kabat numbering][Mutated amino acid]. The restriction enzyme recognition sites for NcoI and BstEII were underlined.

sdAb	Name	Sequence (5' – 3')
1	WT	CCATGGCGGATGTCCAGTTGGTTGAAAGTGGAGGAGGTCTGGTTCAGCCTGGTGGCT CGCTGCGTTTAAGCTGTGCAGCATCAGGGTACACATTTACAAATTATGGTATGAACT GGGTCCGCCAAGCGCCAGGGAAAGGATTAGAGTGGGTAGGGTGGATCAATACTTATA CCGGCGAGCCCACCTACGCTGCAGATTTCAAGCGCCGTTTTACCTTTTCACTTGACAC TTCCAAAAGCACAGCGTACTTACAAATGAACTCCTTGCGCGCGGAGGATACAGCTGT GTATTATTGTGCTAAGTACCCGCACTATTACGGTTCATCCCATTGGTATTTCGATGT GTGGGGTCAGGGGACATTGGTCACC
2	V37F	CCATGGCGGATGTCCAGTTGGTTGAAAGTGGTGGAGGTCTGGTTCAGCCTGGTGGCT CGCTGCGTTTAAGCTGTGCAGCATCAGGGTACACATTTACAAATTATGGTATGAACT GGTTCCGCCAAGCGCCAGGGAAAGGATTAGAGTGGGTAGGGTGGATCAATACTTATA CCGGCGAGCCCACCTACGCTGCAGATTTCAAGCGCCGTTTTACCTTTTCACTTGACAC TTCCAAAAGCACAGCGTACTTACAAATGAACTCCTTGCGCGCGGAGGATACAGCTGT

		GTATTATTGTGCTAAGTACCCGCACTATTACGGTTCATCCCATTGGTATTTCGATGT GTGGGGTCAGGGGACATTGGTCACC
3	**V37Y**	CCATGGCGGATGTCCAGTTGGTTGAAAGTGGTGGAGGTCTGGTTCAGCCTGGTGGCT CGCTGCGTTTAAGCTGTGCAGCATCAGGGTACACATTTACAAATTATGGTATGAACT GGTACCGCCAAGCGCCAGGGAAAGGATTAGAGTGGGTAGGGTGGATCAATACTTATA CCGGCGAGCCCACCTACGCTGCAGATTTCAAGCGCCGTTTTACCTTTTCACTTGACAC TTCCAAAAGCACAGCGTACTTACAAATGAACTCCTTGCGCGCGGAGGATACAGCTGT GTATTATTGTGCTAAGTACCCGCACTATTACGGTTCATCCCATTGGTATTTCGATGT GTGGGGTCAGGGGACATTGGTCACC
4	**G44E**	CCATGGCGGATGTCCAGTTGGTTGAAAGTGGAGGAGGTCTGGTTCAGCCTGGTGGCT CGCTGCGTTTAAGCTGTGCAGCATCAGGGTACACATTTACAAATTATGGTATGAACT GGGTCCGCCAAGCGCCAGGGAAAGAATTAGAGTGGGTAGGGTGGATCAATACTTATA CCGGCGAGCCCACCTACGCTGCAGATTTCAAGCGCCGTTTTACCTTTTCACTTGACAC TTCCAAAAGCACAGCGTACTTACAAATGAACTCCTTGCGCGCGGAGGATACAGCTGT GTATTATTGTGCTAAGTACCCGCACTATTACGGTTCATCCCATTGGTATTTCGATGT GTGGGGTCAGGGGACATTGGTCACC
5	**L45R**	CCATGGCGGATGTCCAGTTGGTTGAAAGTGGAGGAGGTCTGGTTCAGCCTGGTGGCT CGCTGCGTTTAAGCTGTGCAGCATCAGGGTACACATTTACAAATTATGGTATGAACT GGGTCCGCCAAGCGCCAGGGAAAGGAAGAGAGTGGGTAGGGTGGATCAATACTTAT A CCGGCGAGCCCACCTACGCTGCAGATTTCAAGCGCCGTTTTACCTTTTCACTTGACAC TTCCAAAAGCACAGCGTACTTACAAATGAACTCCTTGCGCGCGGAGGATACAGCTGT GTATTATTGTGCTAAGTACCCGCACTATTACGGTTCATCCCATTGGTATTTCGATGT GTGGGGTCAGGGGACATTGGTCACC
6	**W47G**	CCATGGCGGATGTCCAGTTGGTTGAAAGTGGTGGAGGTCTGGTTCAGCCTGGTGGCT CGCTGCGTTTAAGCTGTGCAGCATCAGGGTACACATTTACAAATTATGGTATGAACT GGGTCCGCCAAGCGCCAGGGAAAGGATTAGAGGGGGTAGGGTGGATCAATACTTATA CCGGCGAGCCCACCTACGCTGCAGATTTCAAGCGCCGTTTTACCTTTTCACTTGACAC TTCCAAAAGCACAGCGTACTTACAAATGAACTCCTTGCGCGCGGAGGATACAGCTGT GTATTATTGTGCTAAGTACCCGCACTATTACGGTTCATCCCATTGGTATTTCGATGT GTGGGGTCAGGGGACATTGGTCACC
7	**G44E** **L45R**	CCATGGCGGATGTCCAGTTGGTTGAAAGTGGTGGAGGTCTGGTTCAGCCTGGTGGCT CGCTGCGTTTAAGCTGTGCAGCATCAGGGTACACATTTACAAATTATGGTATGAACT GGGTCCGCCAAGCGCCAGGGAAAGAAAGAGAGTGGGTAGGGTGGATCAATACTTAT A CCGGCGAGCCCACCTACGCTGCAGATTTCAAGCGCCGTTTTACCTTTTCACTTGACAC TTCCAAAAGCACAGCGTACTTACAAATGAACTCCTTGCGCGCGGAGGATACAGCTGT GTATTATTGTGCTAAGTACCCGCACTATTACGGTTCATCCCATTGGTATTTCGATGT GTGGGGTCAGGGGACATTGGTCACC
8	**G44E** **W47G**	CCATGGCGGATGTCCAGTTGGTTGAAAGTGGTGGAGGTCTGGTTCAGCCTGGTGGCT CGCTGCGTTTAAGCTGTGCAGCATCAGGGTACACATTTACAAATTATGGTATGAACT GGGTCCGCCAAGCGCCAGGGAAAGAATTAGAGGGGGTAGGGTGGATCAATACTTATA CCGGCGAGCCCACCTACGCTGCAGATTTCAAGCGCCGTTTTACCTTTTCACTTGACAC

		TTCCAAAAGCACAGCGTACTTACAAATGAACTCCTTGCGCGCGGAGGATACAGCTGT GTATTATTGTGCTAAGTACCCGCACTATTACGGTTCATCCCATTGGTATTTCGATGT GTGGGGTCAGGGGACATTGGTCACC
9	**G44E** **L45R** **W47G**	CCATGGCGGATGTCCAGTTGGTTGAAAGTGGTGGAGGTCTGGTTCAGCCTGGTGGCT CGCTGCGTTTAAGCTGTGCAGCATCAGGGTACACATTTACAAATTATGGTATGAACT GGGTCCGCCAAGCGCCAGGGAAAGAAAGAGAGGGGGTAGGGTGGATCAATACTTAT ACCGGCGAGCCCACCTACGCTGCAGATTTCAAGCGCCGTTTTACCTTTTCACTTGACA CTTCCAAAAGCACAGCGTACTTACAAATGAACTCCTTGCGCGCGGAGGATACAGCTG TGTATTATTGTGCTAAGTACCCGCACTATTACGGTTCATCCCATTGGTATTTCGATGT GTGGGGTCAGGGGACATTGGTCACC
10	**V37F** **G44E** **L45R** **W47G**	CCATGGCGGATGTCCAGTTGGTTGAAAGTGGTGGAGGTCTGGTTCAGCCTGGTGGCT CGCTGCGTTTAAGCTGTGCAGCATCAGGGTACACATTTACAAATTATGGTATGAACT GGTTCCGCCAAGCGCCAGGGAAAGAAAGAGAGGGGGTAGGGTGGATCAATACTTAT ACCGGCGAGCCCACCTACGCTGCAGATTTCAAGCGCCGTTTTACCTTTTCACTTGACA CTTCCAAAAGCACAGCGTACTTACAAATGAACTCCTTGCGCGCGGAGGATACAGCTG TGTATTATTGTGCTAAGTACCCGCACTATTACGGTTCATCCCATTGGTATTTCGATGT GTGGGGTCAGGGGACATTGGTCACC
11	**V37Y** **G44E** **L45R** **W47G**	CCATGGCGGATGTCCAGTTGGTTGAAAGTGGTGGAGGTCTGGTTCAGCCTGGTGGCT CGCTGCGTTTAAGCTGTGCAGCATCAGGGTACACATTTACAAATTATGGTATGAACT GGTACCGCCAAGCGCCAGGGAAAGAAAGAGAGGGGGTAGGGTGGATCAATACTTAT ACCGGCGAGCCCACCTACGCTGCAGATTTCAAGCGCCGTTTTACCTTTTCACTTGACA CTTCCAAAAGCACAGCGTACTTACAAATGAACTCCTTGCGCGCGGAGGATACAGCTG TGTATTATTGTGCTAAGTACCCGCACTATTACGGTTCATCCCATTGGTATTTCGATGT GTGGGGTCAGGGGACATTGGTCACC

Table 138. List of host lab codon optimised Bevacizumab V_H fragment peptide sequences of redesigned sdAbs The introduced mutations were indicated using the following format: Mutated residue: [Original amino acid][Kabat numbering][Mutated amino acid]. The following residues were highlighted as follows: Codon optimised V_H peptide sequence, Mutation (in red) and *His_6* were shown.

sdAb	VUB sdAb name	Redesigned peptide sequence
1	**WT**	DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGKGLEWVGWINTYT GEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYF DVWGQGTLVTVSS*HHHHHH*
2	**V37F**	DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNW*F*RQAPGKGLEWVGWINTYT GEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYF DVWGQGTLVTVSS*HHHHHH*
3	**V37Y**	DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNW*Y*RQAPGKGLEWVGWINTYT GEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYF DVWGQGTLVTVSS*HHHHHH*

4	**G44E**	DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGK*E*LEWVGWINTYT GEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYF DVWGQGTLVTVSS*HHHHHH*
5	**L45R**	DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGKG*R*EWVGWINTYT GEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYF DVWGQGTLVTVSS*HHHHHH*
6	**W47G**	DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGKGLE*G*VGWINTYT GEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYF DVWGQGTLVTVSS*HHHHHH*
7	**G44E/L45R**	DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGK*ER*EWVGWINTYT GEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYF DVWGQGTLVTVSS*HHHHHH*
8	**G44E/W47G**	DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGK*E*LE*G*VGWINTYT GEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYF DVWGQGTLVTVSS*HHHHHH*
9	**G44E/L45R/ W47G**	DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGK*ER*E*G*VGWINTYT GEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYF DVWGQGTLVTVSS*HHHHHH*
10	**V37F/G44E/ L45R/W47G**	DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNW*F*RQAPGK*ER*E*G*VGWINTYT GEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYF DVWGQGTLVTVSS*HHHHHH*
11	**V37Y/G44E/ L45R/W47G**	DVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNW*Y*RQAPGK*ER*E*G*VGWINTYT GEPTYAADFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSRWYF DVWGQGTLVTVSS*HHHHHH*

4.7.3.2. Cloning and ligation

The cloning procedure was resumed by the PCR Clean Up. To incorporate the new sdAb designs into the pHen6 plasmid, both ends of the inserts (the purified and amplified PCR products) were digested using NcoI and BstEII restriction enzymes. The same enzymes were employed for the vector pHen6 digestion to generate compatible ends. To prevent the plasmid from undergoing self-ligation, the restiricted plasmid was treated with the Alkaline Phosphatase enzyme before ligation.

After the PCR amplification, an agarose gel (1%) electrophoresis was conducted as QC analysis (**Figure 96**).

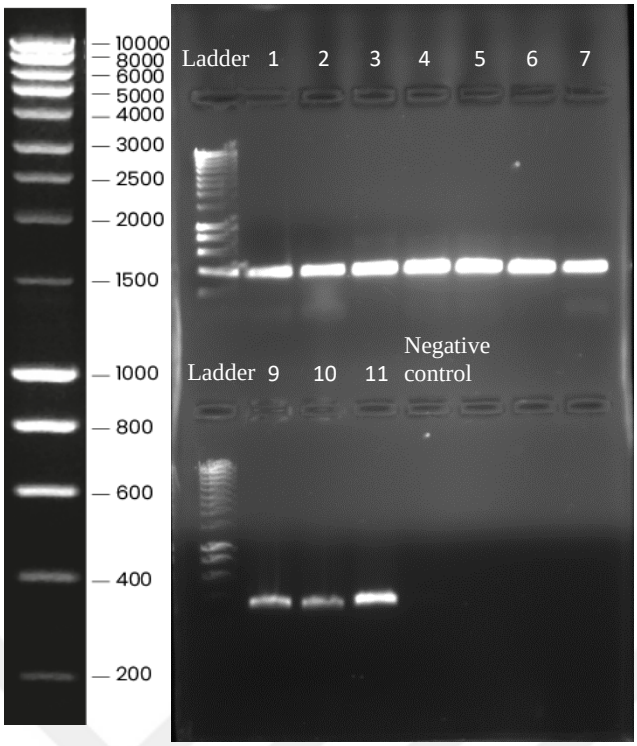


Figure 96. 1% agarose gel electrophoresis to visualize the sdAbs after PCR amplification. As a negative control, a sample without of plasmid template (dH_2O) was was also included in the run.

Both the digested vector and the sdAbs were joined at the same tubes to facilitate an overnight ligation reaction at 16°C.

4.7.3.3. Transformation and sequencing

Next, the ligated vector with the insert was transformed into competent WK6 cells through a heat-shock transformation (42°C for 45 sec). The transformed cells were incubated overnight allowing for colony formation. The next day, plasmids wre extracted using Miniprep kit to prepare them for sequencing to validate the correct orientation of the insert into the plasmid. For the sequencing, 40 ng/µl of sdAbs plasmid mixed with 2 µl of 10 µM MP-57 primers (pHen6 plasmid is already carrying a MP7 recognition site).

4.7.4. Redesigned recombinant sdAb production and purification by IMAC

Following the confirmation of each sdAb sequence through the Sanger sequencing, production studies were commenced by transforming the sequence confirmed plasmids into expression cell line *E. coli* WK6 by heat-shock transformation. Pre-cultures were prepared by using single colonies from each redesigned sdAb. The pre-culture was grown overnight at 37°C by shaking at 200 rpm.

The production procedure initiated at 37°C, 220 rpm for approximately 3 h, or continued until the cells reached at certain confluency. Upon completing the primary log

phase, 1M of IPTG was added to the shaking cultures. The production was resumed at 28°C and 220 rpm for overnight induction of the gene of interest.

The redesigned sdAbs were purified using IMAC. Following this step, further the samples were further purified via SEC. The concentrations of produced redesigned sdAbs were measured using a spectrophotometer (**Table 139**).

Table 139. The production concentration of every sdAb produced using WK6 *E.coli* cells. For each sdAb sample, the production was repeated at least two times, using the same production conditions. The average concentration of productions was pooled and the concentration was calculated as mg/L culture and presented with standard deviation.

Redesigned sdAb #	Redesigned sdAb	Concentration (mg/L culture)
1	WT	0.29 ± 0.08
2	V37F	0.31 ± 0.34
3	V37Y	0.19 ± 0.07
4	G44E	0.25 ± 0.19
5	L45R	0.21 ± 0.07
6	W47G	0.18 ± 0.19
7	G44E/ L45R	0.20 ± 0.13
8	G44E/ W47G	0.36 ± 0.12
9	G44E/L45R/W47G	0.16 ± 0.45
10	V37F/G44E/L45R/W47G	0.12 ± 0.17
11	V37Y/G44E/L45R/W47G	0.19 ± 0.16

4.7.5. Optimisation of redesigned recombinant sdAb production and purification

For enhancing the quality and quantity of redesigned sdAbs, key parameters were selectively optimised by implementing the following strategies.

4.7.5.1. Isolation of proteins from different cellular compartments

The sdAbs were isolated from diverse cellular compartments, including the periplasmic space and the cytoplasm (**Figure 97-99**) to improve the production yield. A control nanobody was included alongside the sdAbs.

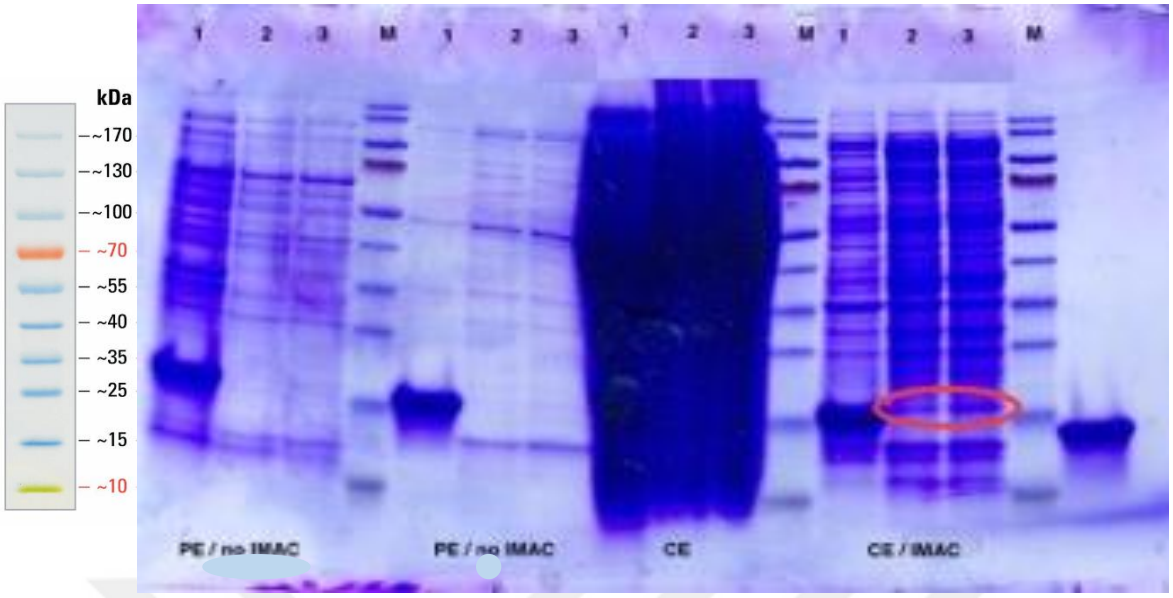
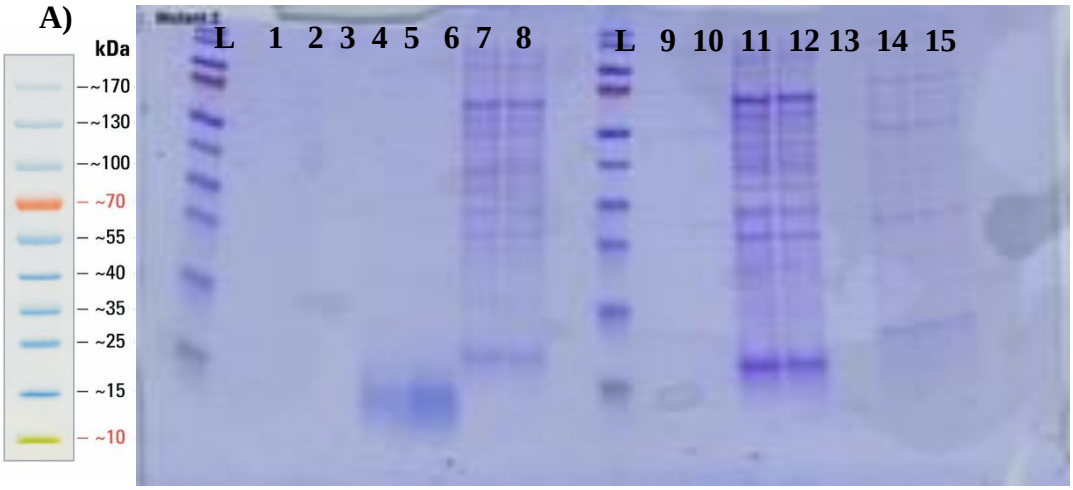


Figure 97. The effect of isolating redesigned sdAbs from different cellular compartments assessed by SDS-Page. The sdAb samples were randomly selected amongst the redesigned sdAbs. SDS-Page was performed on sdAb 1, sdAb 11 and a control nanobody. The gel lane configuration was as follows: 1: the control nanobody (R3b23-His_6), 2: sdAb 11 and 3: sdAb 1 (Please refer to **Table 50** for full nomenclature). The first and the second sets of samples from the periplasmic extracts was labelled as PE which were obtained through two rounds of TES buffer application. The difference is first set was the samples taken before applying IMAC. The third and fourth triple sets of samples represent cytoplasmic extracts (CE) obtained by supernatant as leaky expression. The third set of the triplicates includes load samples before incubating with IMAC beads. The last set of triple samples is the control nanobody before IMAC and followed by sdAb 1 and sdAb11 purified by IMAC. The last lane includes purified control nanobody. M: protein ladder (marker). The expected molecular weight range of sdAbs was encircled on the gel image.



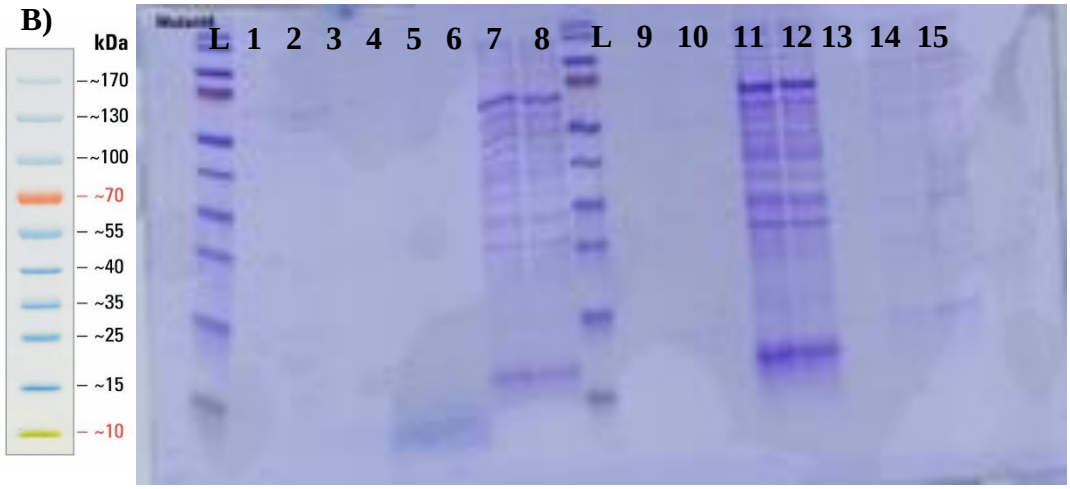


Figure 98. The SDS-PAGE results of the redesigned sdAb 2 (upper) and redesigned sdAb 8 (lower) in the gel were illustrated. Each lane corresponds to the specific sample or fraction on the gel. L: Ladder.The lanes are as follows: Supernatant Elution Fraction 1-2: SN E1-2, Supernatant Elution Fraction 3-4: SN E3-4, Supernatant Elution Fraction 5-6: SN E5-6, Periplasmic Extract 1: PE1, PE1 Flow Through: PE1 FT, PE1 Elution Fraction 1-2: PE1 E1-2, PE1 Elution Fraction 3-4: PE1 E3-4, PE1 Elution Fraction 5-6: PE1 E5-6, Periplasmic Extract 2: PE2, PE2 Flow Through: PE2 FT, PE2 Elution Fraction 1-2: PE2 E1-2, PE2 Elution Fraction 3-4: PE2 E3-4, PE2 Elution Fraction 5-6: PE2 E5-6, Positive Control: U6594 uPAR nanobody uPAR 92-His_6 (0.98 mg/ml). Supernatant: SN, Supernatant flow through: FT, PE: Periplasmic extract.

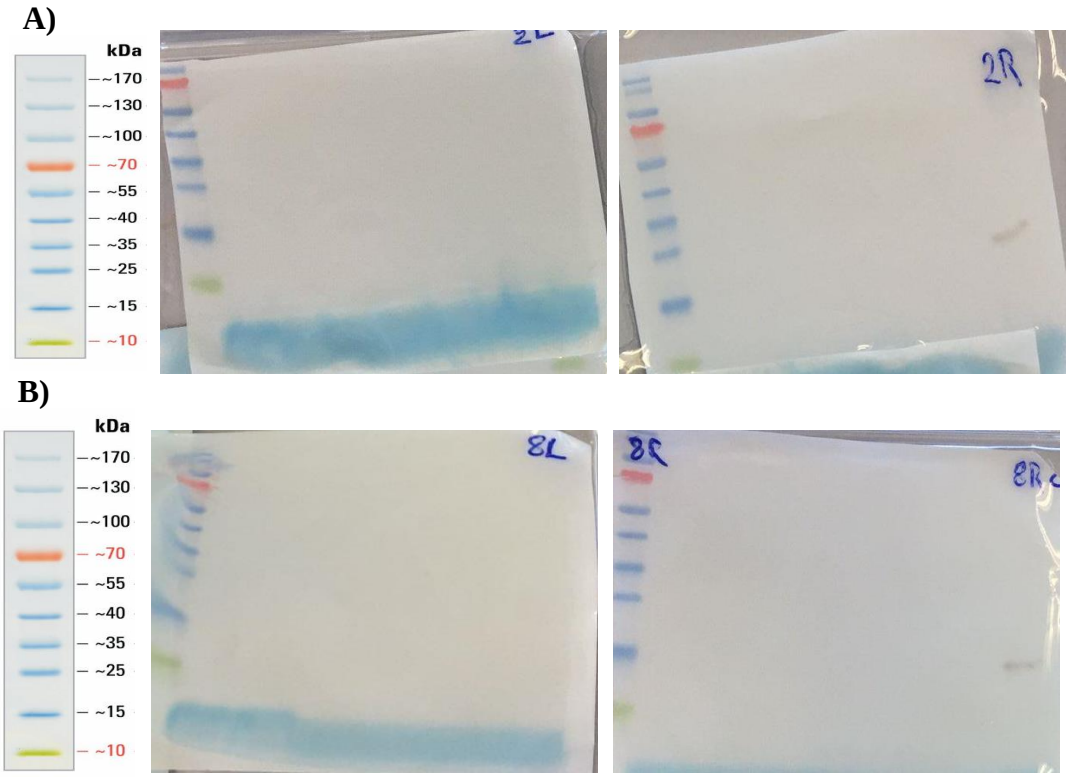


Figure 99. The WB results of **A)** sdAb 2 (upper) and **B)** sdAb 8 (lower) were displayed. The membranes were cut in half to identify leaky and periplasmic extracts for the convenience of visualization.The samples were not marked on the membrane as it was very hard to distinguish the lanes from each other. The sample labels for the lanes are as follows from left to right order; SN E1, SN E2, SN E3, SN E4, SN E5, SN E6, PE1 FT, PE1 E1, PE1 E2, PE1 E3, PE1 E4, PE1 E5, PE1 E6, PE2, PE2 FT, PE2 E1, PE2 E2, PE2 E3, PE2 E4, PE2 E5, PE2 E6, Positive Control: U6594 uPAR nb uPAR92-His_6 (0.98 mg/ml). (Supernatant, leaky expression: SN, Supernatant flow through: FT, E: Elution Fraction, PE1: Periplasmic Extract obtained after one round of TES buffer treatment, PE2: Periplasmic Extract obtained after two rounds of TES buffer treatment).

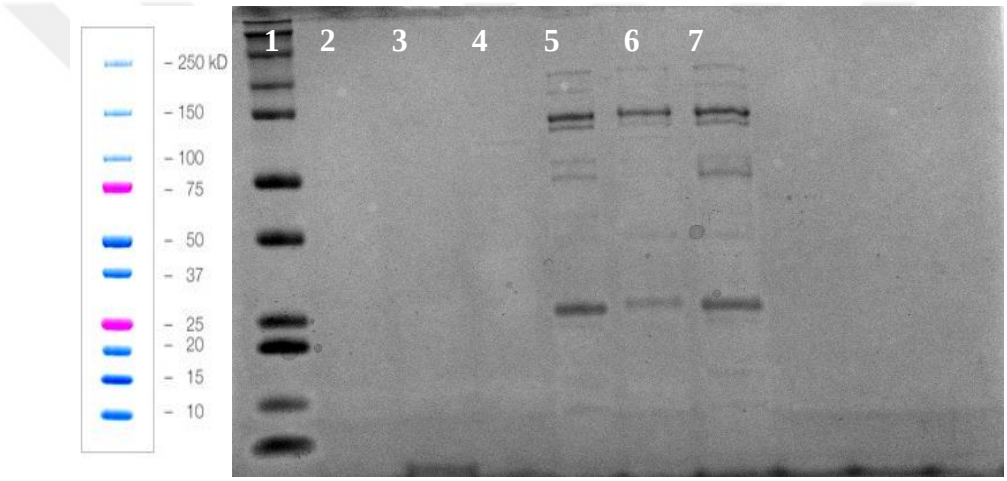


Figure 100. The SDS-PAGE results of the redesigned sdAb 1, sdAb 3 and sdAb 4 isolated from different cellular compartments were illustrated in the gel. The sample labels for the lanes are as follows; 1: Ladder, 2: sdAb 1 SN, 3: sdAb 3 SN, 4: sdAb 4 SN, 5: sdAb 1 PE1, 6: sdAb 3 PE1, 7: sdAb 4 PE1. (SN: Supernatant, leaky expression, PE1: periplasmic extract obtained after one round of TES and TES/4 buffer treatment).

Regrettably, successful yield of production results could not be attained even after the optimisation studies, as evidenced by **Figure 97-100**.

4.7.5.2. Utilization of different auto-induction media compositions

Various auto-induction media, including home-brew and commercial Terrific broth media, as well as home-lab recipe of auto-induction media based on Studier's method, were utilized to culture and induce recombinant sdAbs. When cells reach to certain confluency, Terrific Broth, requires the addition of IPTG from the outside. In contrast, the novel auto-

induction media inherently possesses such induction mechanism for the expression of the protein of interest.

4.7.5.3. Exploration of different E. coli strain types

As another optimisation parameter, different *E. coli* strain types were utilized. The effects of changes in media were assessed also along with different production cell lines. Each *E.coli* strain possesses unique characteristics that could potentially influence protein expression.

For this study, one of the most promising redesigned sdAbs was chosen for protein production. Varying cell lines and induction conditions were employed to facilitate protein production, as detailed in the following tables (**Table 140** and **141**).

Table 140. The optimisation studies on varying production media composition and production conditions were assessed. (TB: Terrific Broth).

	TB Media	**Auto-Induction Media**
Reagents	55.85 g TB	10 g tryptone
	1000 ml dH_2O	5 g yeast extract
		928 ml dH_2O
Supplements	330 µl 1000X AMP	330 µl 1000X AMP
		330 µl 1M $MgSO_4$
		7 ml of 50X 50/52
		17ml of 20X NPS
Production conditions	24°C, 220 rpm, 48 h	18°C, 250 rpm, 48 h
Initial inoculation	3 ml	3 ml

Table 141. Various *E. coli* cell lines transformed with different sdAb plasmids an cultured in two different auto-induction media for the optimisation of protein production, The production settings were organized as follows:

Terrific Broth	**Auto-induction media**
1L X sdAb 3 (WK6)	1L X sdAb 3 (WK6)
1L X sdAb 3 (BL21)	1L X sdAb 3 (BL21)
1L X sdAb 4 (WK6)	1L X sdAb 4 (WK6)
1L X sdAb 4 (BL21)	1L X sdAb 4 (BL21)
4L (12 bottles)	**4L (12 bottles)**

Table 142. The protein concentration results obtained after optimisation studies on utilizing different *E.coli* strains. The optimisation was based on assessing the effect of different *E.coli* strains and different auto-induction conditions. Auto-induction media was prepared using a custom recipe based on (Studier 2005) method. Commercially available TB, was prepared in accordance to the instructions of the manufacturer. TB was induced through the addition of IPTG, as previously described. The results reflected in series of tables for different sdAbs. (PE1: Periplasmic extract after 1 round of TES buffer treatment, PE2: Periplasmic extract after two rounds of TES buffer treatment, SN: supernatant after two rounds of TES buffer treatment, Concentration 1 to 5: indicates the collected fraction number)

Sample Name	**Sample ID**	**Sample**	**Concentration (mg/ml)**				
Compartment	**PE1**		**1**	**2**	**3**	**4**	**5**
V37Y (sdAb 3)	3_1	WK6 Auto induction	-0.44	-0.24	-0.14	-0.06	-0.03
	3_2	BL21 Auto induction	-0.43	-0.15	-0.18	-0.26	-0.01
	3_3	WK6 TB	-0.43	-0.21	0.00	0.06	0.05
	3_4	BL21 TB	-0.38	-0.32	-0.03	0.03	0.06
G44E/W47G (sdAb 8)	8_1	WK6 Auto induction	-0.39	-0.24	0.20	0.16	0.15
	8_2	BL21 Auto induction	-0.36	-0.25	-0.14	-0.09	-0.06
	8_3	WK6 TB	-0.21	-0.56	0.02	0.02	0.00
	8_4	BL21 TB	-0.42	-0.34	0.07	0.13	0.08

Sample Name	**Sample ID**	**Sample**	**Concentration (mg/ml)**				
Compartment	**PE2**		**1**	**2**	**3**	**4**	**5**
V37Y (sdAb 3)	3_1	WK6 Auto induction	-0.44	-0.27	-0.11	-0.06	-0.04
	3_2	BL21 Auto induction	-0.27	-0.03	-0.44	-0.03	0.05
	3_3	WK6 TB	-0.24	-0.05	0.03	0.08	0.11
	3_4	BL21 TB	-0.25	-0.47	0.36	0.34	0.16
G44E/W47G (sdAb 8)	8_1	WK6 Auto induction	0.03	0.21	0.13	0.11	0.12
	8_2	BL21 Auto induction	-0.39	-0.21	-0.03	0.01	0.02
	8_3	WK6 TB	-0.37	-0.15	0.00	0.05	0.05
	8_4	BL21 TB	-0.32	-0.14	-0.03	0.03	0.05

Sample Name	Sample ID	Sample	Concentration (mg/ml)				
Compartment	**SN**		**1**	**2**	**3**	**4**	**5**
V37Y (sdAb 3)	3_1	WK6 Auto induction	-0.31	-0.10	0.02	0.06	0.09
	3_2	BL21 Auto induction	-0.37	-0.07	0.03	0.07	0.08
	3_3	WK6 TB	-0.18	-0.02	0.09	0.12	0.13
	3_4	BL21 TB	-0.14	0.04	0.10	0.11	0.11
G44E/W47G (sdAb 8)	8_1	WK6 Auto induction	-0.34	-0.06	0.04	0.08	0.08
	8_2	BL21 Auto induction	-0.31	-0.21	-0.03	0.01	0.02
	8_3	WK6 TB	-0.24	-0.10	-0.02	0.04	0.08
	8_4	BL21 TB	0.12	-0.09	0.09	0.11	0.11

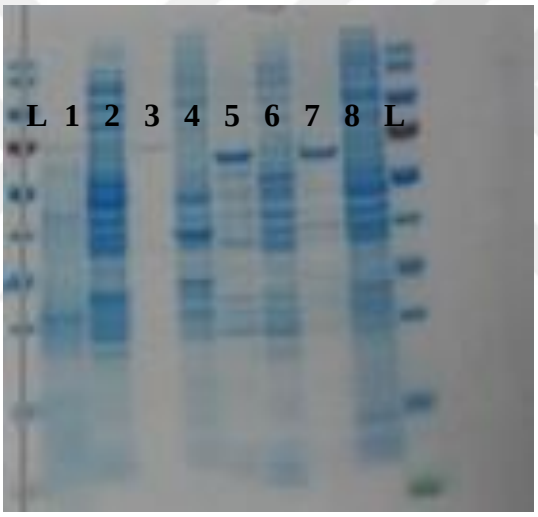


Figure 101. The effect of different induction media and different *E.coli* strains on redesigned sdAb yield and purity was investigated by WB analysis. The sdAb 3 (V37Y) and sdAb 8 (G44E/W47G) were isolated from Supernatant (SN) of the leaky expression. Two different *E.coli* strains were utilized. The membrane was probed with anti-His mouse antibody (1:2000), followed by an anti-mouse HRP antibody (1:2000) incubation. Terrific Broth (TB), Auto-induction (AI). The samples have the following conditions: 1: sdAb3 SN TB WK6, 2: sdAb3 SN TB BL21, 3: sdAb3 SN AI WK6, 4: sdAb3 SN AI BL21, 5: sdAb4 SN TB WK6, 6: sdAb8 SN TB BL21, 7: sdAb8 SN AI WK6, 8: sdAb8 SN AI BL21.

Based on the results of concentration measurements of individual fractions collected after IMAC purification *(***Table 142***)* and WB analysis *(***Figure 101***)*, optimisation attempts using different *E. coli* strains (WK6 or BL21) and varied media compositions (auto-induction

media or terrific broth) did not successfully yield. The concentration of each collected fraction exhibited consistently poor yields, which was also reflected from the associated SDS-PAGE and WB results. Especially, optimisations with the BL21 strain presented smear-like results on SDS-PAGE which is an indicator of poor purification. In contrast, the utilization of the WK6 cell line for recombinant protein extraction from the supernatant seems like more promising strategy irrespective of auto-induction or TB usage. However, when testing these conditions with two randomly selected sdAbs, the presence of contaminants were still there. Overall, further efforts were required to improve current optimisation conditions.

4.7.6. SPECT-CT imaging to assess whole body distribution at the macroscopic level

For SPECT-CT imaging, the following experiments were performed:

4.7.6.1. Modification of sdAbs and Bevacizumab with a DOTA chelator

Before Radiolabelling with a radioisotope, in this case radioactive indium isotope - ^{111}In, both sdAbs and Bevacizumab should be incorporated with a DOTA chelator.

The enzymatic conjugation of the DOTA chelator to Bevacizumab, conducted in the presence of a 0.5 M carbonate buffer. Once treated with this buffer, DOTA moieties are speficially attach to lysine side chains present in Bevacizumab/sdAb structure via ester chemistry.

Hence the conjugation of DOTA is based on random attachment to Lysine residues; the presence of these residues within both in the heavy and light chains of the Bevacizumab antibody should be detailly investigated (**Figure 102**) before the reaction takes place.

DIQMTQSPSSLSASVGDRVTITCSASQDISNYLNWYQQKPGKAPKVLIYFTSSLHSGVPSRF
SGSGSGTDFTLTISSLQPEDFATYYCQQYSTVPWTFGQGTKVEIKRTVAAPSVFIFPPSDEQ
LKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADY
EKHKVYACEVTHQGLSSPVTKSFNRGEC

EVQLVESGGGLVQPGGSLRLSCAASGYTFTNYGMNWVRQAPGKGLEWVGWINTYTGEPTYAA
DFKRRFTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYYGSSHWYFDVWGQGTLVTVSSA
STKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLY
SLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHT*CPPCPAPELLGGPSVFL*
FPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVS
VLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLT
CLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVM
HEALHNHYTQKSLSLSPGK

Figure 102. The analysis on the Lysine residues in the heavy and light chains of Bevacizumab was investigated to check its suitability for random labelling using DOTA prior to radiolabelling with ^{111}In. The sequence information for both the heavy and light chains was obtained using PDB ID: 1BJ1. The upper sequence belongs to light chain of Bevacizumab, where the below one is heavy chain. The CDRs were highlighted as follows based on Kabat numbering scheme; CDR1 was highlighted in red, CDR2 in yellow, and CDR3 in green. Lysine residues are italicised and shown in cyan text. The underlined letters showing the constant regions of Bevacizumab Fab fragment.

Upon conducting this analysis, it was revealed that only two Lysine residues are present within the LCDR1 and 2 of Bevacizumab. These Lysine residues can be neglected, considering the most important part of an antibody is its HCDR3 in target antigen binding. Therefore, the antigen binding capacity of Bevacizumab would not be significantly affected as the conjugation of DOTA takes place in these particular Lysines (**Figure 102**).

4.6.6.2. Radiolabelling with radioisotope ^{111}In

Right after the chelation step, the modified antibody would be ready for radiolabelling with radioisotope ^{111}In. For this purpose, DOTA was previously randomly conjugated to the antibody and subsequently stored at -20°C for future use in the ^{111}In conjugation process. 21 µl of ^{111}In was prepared in 35.6 µl of 0.5 M NaOAc buffer (pH 5, metal-free). For this reaction, 29 µl of 1.70 mBq ^{111}In solution was mixed with 24 µl of DOTA-conjugated Bevacizumab (50 mg) and incubated at 40°C for 90 min.

The efficiency of this test run reaction was validated by iTLC (**Figure 103**). The results revealed that approximately 98% of the free ^{111}In had been successfully incorporated into the Bevacizumab-DOTA compound, whereby the 2% was remaining free.

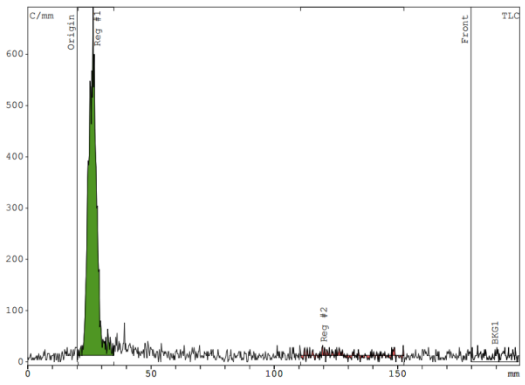


Figure 103. The evaluation of the incorporation of ^{111}In into the Bevacizumab-DOTA by iTLC analysis.

5. DISCUSSION

While monoclonal antibodies have been the central focus of optimisation efforts, there has been a growing interest in the refinement of alternative antibody formats, such as sdAbs. sdAbs have distinctive advantages, such as reduced size, improved tissue penetration, and enhanced manufacturing efficiency. Various engineering strategies are employed to convert conventional antibodies into sdAbs. Camelisation, which involves the use of monomeric human V_H as a scaffold, introduces key residues of hcIgGs into this framework (Riechmann and Muyldermans 1999). Given the sequence homology between camelid and human V_H domains, hcIgGs emerge as optimal candidates for the development of functional sdAbs (Bélanger et al. 2019; Vaneycken et al. 2011).

It is important to consider the distribution of amino acids with different physicochemical properties throughout the FWs, which provide scaffolding for the CDRs, when developing strategies to improve antibody characteristics. By carefully analysing the amino acid distribution, regions that may benefit from the introduction of amino acids with specific properties can be identified. For example, introducing amino acids with solubility or stability enhancing characteristics in regions prone to aggregation can improve the overall performance of the antibody (Tiller and Tessier 2015).

At the beginning of this thesis, a diverse array of online tools was employed to identify aggregation-prone residues within the V_H sequence of the Bevacizumab scaffold. The V_H region was analysed with these computational tools to pinpoint the residues with higher propensity for aggregation. The consensus of these comprehensive analyses guided the transformation of the V_H Bevacizumab sequence into a functional sdAb format.

For the use of *in silico* tools, the V_H sequence of Bevacizumab was exclusively used as the input for online platforms. Despite their possible predisposition to aggregation, the additional tags (Flag, His) and restriction enzyme recognition site (Factor Xa) in the sequence design were deemed non-interchangeable and left unaltered. During these analyses, certain online tools only permitted PDB IDs as input, rendering them inapplicable to use with the whole designed sequence including the extra residues in manual FASTA format.

These *in silico* analyses were followed by camelisation efforts. The residues chosen for camelisation were carefully selected based on their alignment with the aggregation-prone residues identified by computational predictions. Initially, the focus was on hallmark residues of Bevacizumab V_H at FW-2 region to embrace the sdAb format. This conversion was

followed by the alteration of residues at FW-3 by humanisation approach to eliminate mouse residues (93% human, 7% murine sequences) originated during the discovery of Bevacizumab through hybridoma technology (https://www.ema.europa.eu/en/medicines/human/EPAR/avastin). Subsequently, mutations in both FW-2 and FW-3 were combined to observe their synergistic effects on adaptation to sdAb format, as well as lower immunogenicity for animal studies.

A total of 10 distinct Bevacizumab-derived sdAb mutants were engineered, with one to four mutations in their FW-2 region, in addition to WT-sdAb without any mutations. These mutations were anticipated to compensate for the absence of the V_L-domain, thereby enhancing their solubility. To begin, these 10 mutants were characterized *in vitro*, focusing on production yield, affinity, and stability profile assessment of those sdAbs.

When modifying the Bevacizumab V_H sequence for conversion into the sdAb format using camelisation, specific residues were chosen based on their critical roles and interactions. Namely, the V37F substitution in FW-2 was chosen due to its possible impact on expression yield, binding capacity, and stability on the overall molecule. However, the nature of this substitution was not straightforward, as its effects were also influenced by its interactions with the neighbouring residues at 44th, 45^{h}, and 47th positions. The substitution was expected to stabilise the former V_L interface and compensate for the stability loss caused by other substitutions designed at FW-2. Thus, the individual and the synergetic effects of mutations were investigated by designing single and combinatory alternations on FW-2 resdiues.

Humanisation is an antibody engineering approach to decrease the immunogenicity of antibodies derived from xenogeneic sources – in our case, the murine residues. Humanisation involves a two-step strategy when converting a V_H into V_HH-like sdAb format. In the first step, the amino acids in FW-1, -3, and -4 were humanised, followed by FW-2 amino acids in the second step. The humanisation of FW regions 1, 3, and 4 was expected to have minimal effects on overall antibody fragment strucutre. For FW-3 region, five alterations were made simultaneously, since the individual mutations might introduce potentially negligible affects on the antibody moiety.

Only two of the V_HH hallmark residues, at 44th and 45th positions, are humanised without compromising solubility, monomeric behaviour, and functionality of V_HH (Davies and Riechmann 1994; Conrath et al. 2005). The remaining two FW-2 V_HH-specific amino acids (37, 47) could not be humanised, as they played critical roles in maintaining the

structural integrity of the single domain, especially interacting with the antigen-binding HCDR3 loop (Saerens et al. 2005; Vincke et al. 2009). The conversion of 44th and 45th residues was prime candidates for alteration, as suggested by the available literature. Therefore, the amino acids at positions 44 and 45 (G and L) in FW-2 were initially chosen for camelisation. Based on these findings, we decided the residues at 37th and 47th position should remain in their humanised form. Amongst our double mutation designs, we also utilised the alteration of FW-2 44th and 45th residues together to test this hypothesis. However, our G44E/L45R sdAb variant exhibited very low purity levels, making it impossible to further characterise to validate prior findings.

To reach the project objectives, the individual and combined synergistic effects of the designed mutations on the Bevacizumab V_H scaffold were systematically explored. The characterisation studies unveiled the interplay of advantages and disadvantages associated with each sdAb variant. The majority of sdAb variants demonstrated improved features over their unaltered version, WT sdAb, as evidenced by the physicochemical characterisation results. Yet, more effort should be spent on the improvement of both the quality and the quantity of each sdAb, in order to achieve a biobetter state.

Despite notable advancements, challenges in maintaining the purity have been encountered repeatedly throughout the whole process. Unfortunately, such obstacles prevented the successful production of pure sdAbs, while lowering the confidence levels of the characterisation studies. To address this purity challenge, a series of optimisation studies were designed. SDS-PAGE and WB analyses were frequently used as a QC mechanism to evaluate the outcomes of the optimisation efforts on sdAb quality after every trial.

Each optimisation was discussed in detail, with their respective pros and cons in the corresponding results sections. Some of the selected optimisation strategies and their results can be summarized as follows:

(i) Unfold/Refold (the application of dialysis):

-Aim: eliminating the high molecular weight contaminants, while enhancing the yield and purity of the protein of interest.

-Result: increase in the lower fragmentation with molecular weight of more than dialysis memebrane cut-off (10 kDa) but lesser than the expected molecular weight of the expressed sdAbs (~18.9 kDa) at the bottom of the SDS-PAGE gel during QC check.

(ii) FPLC with Protein A:

-Aim: alternative resin utilisation for sdAb isolation to obtain improved sample purity and yields.

-Result: significant sample loss in the collected flow-through fractions and subsequent fractions collected after washing stages.

(iii) Extraction from different cellular compartments:

-Aim: increasing yield by searcing for alternative pools of protein of interest in the host cell for recombinant protein production.

-Result: improved yields were obtained but lower purity levels were attained when isolating from periplasmic space. This increase in the yields may be attributed to an increase in upper contaminants due to the presence of chaperones accompanying the protein of interest for the facilitation of their proper folding.

In the original thesis proposal, contingency plans were detailed to address various difficulties that might arise during the experiments. These plans were as follows; for any resolution and stability issues, Dr. Seyit Kale, an expert in protein molecular modelling, would be engaged to investigate whether mutations led to lower resolution or increased aggregation. Molecular modelling would be employed to understand the reasons behind the suboptimal results of rational designs. For expression level-related problems, a range of methods (such as sonication, osmotic shock, etc.) would be applied to assess the expression in different cellular compartments. If necessary, IPTG induction may also be considered as an alternative strategy. For the challenges which arise during protein purification, adjustments to the purification method would be implemented based on when the protein loss occurs. In this case, HPLC would be employed. Alternative affinity chromatography methods, such as anti-FLAG resin or other tags like GST, may be explored. It should be noted that in our design, the FLAG tag (DYKDDDDK) sequence includes two additional amino acids (AA) repeated three times at the end of each traditional FLAG tag frame, and the effects of this addition remain unknown. Therefore, anti-FLAG resin may not be a plausible alternative. In case of aggregate formation and the inability to obtain protein after purification, adjustments to the purification method would be made based on the identified location of protein loss.

Among the alternative strategies described above, those intended to mitigate aggregate formation were partially executed. Alternative resin (Protein A) was assessed for capturing the target protein, which resulted in lower protein yields due to poor interaction of sdAb residues responsible for resin recognition in its spatial structure. However, to further solve

this issue, the alteration in the tag mentioned in the backup plan of the project proposal remained unexplored. Additionally, the backup strategy regarding the expression level challenges was executed by the host institute by applying osmotic shock to isolate protein of interest from the different compartments as well as utilization of IPTG, an analog of allolactose, for recombinant protein induction purposes. Several other contingency plans could have been explored further, considering that all the necessary equipment and experts were within the same research centre.

A diverse range of optimisation studies were conducted on both the yield and purity of recombinant protein production. It takes considerable time to conclude a single parameter change for optimisation. Finalising an optimisation study is a time consuming process, because only one parameter could be altered at a time, and each trial had to be repeated at least three times to ensure the reproducibility of the results. Taking into account the one-week timeframe for each batch of recombinant protein production, each optimisation should be evaluated carefully before its application. It should be also noted that there is no guarantee of success.

In addition to the problems which arose during protein production, challenges were also encountered during the *in vitro* characterisation studies. During the determination of the melting temperature using the SYPRO™ Orange protein dye, there was a problem stemming from the initial baseline signalling. Unlike typical proteins, sdAb variants exhibited detectable signals from the very beginning of the thermal shift assay due to its unpaired nature with V_L. This resulted from the immediate hydrophobic interactions of sdAbs at the former V_H-V_L interaction sites upon mixing with the dye. As a result, it was found that thermal shift assays were not suitable to determine the T_m values. Therefore, a transition was made to chemical denaturation using GdmCl to obtain reliable results.

Apart from the challenges mentioned earlier in this section, the detailed interpretation of each remaining challenge was thoroughly explained and discussed primarily in the results section of the associated subsections accordingly.

In the concluding phases of the characterization studies, the V37Y variant emerged as a promising candidate, with its superior yield and purity characteristics compared to the other *in vitro* characterized sdAbs. It was selected as the primary molecule for testing in animal studies abroad. For comparative purposes, all available variants were also intended to be included in the injection of the designed animal model for *in vivo* characterisation. This study

aimed to investigate how VEGF-targeted sdAbs behave within human tumours with different VEGF expression levels in comparison to Bevacizumab and traditional nanobodies. The tumour-penetrating capabilities of the sdAbs would be investigated by utilising various advanced imaging techniques (SPECT-CT, fluorescence microscopy, confocal microscopy and intravital microscopy) and quantitative analysis.

Unfortunately, the *in vivo* studies did not yield fruitful results. Upon receiving the detailed information on the sdAbs intended to be used for the project, it was realised that the designed sequences contained additional tags, such as Flag tag_3 and His_7 tag. These tags raised concerns about the direct comparability of the sdAbs with Bevacizumab, and the potential immunogenic effects once injected in mice. As a solution, a collaborative effort was initiated with Prof. Dr. Nick Devoogdt at VUB, an esteemed expert in the field. The Bevacizumab-derived sdAbs were redesigned together, while removing the extra tags, the first residue of the V_H sequence was also altered to one that is mostly encountered in camelid counterparts to increase the adaptability potential of sdAbs into V_HH-like format. To facilitate the cloning studies to create these new designs, plasmid samples of the Bevacizumab sdAb variants were required. However, this redesign process faced significant delays due to the logistic-related problems regarding the dispatch of plasmid samples, which resulted in hampering on adherence to the planned timeline.

With the new design of our sdAbs, the change in the expression vector was deemed essential for the removal of the extra tags. pHen6, which inherently possesses His_6-tag for purification purposes, was selected for its compatibility for our purposes. Another change was on the the first amino acid of the sdAbs. It was substituted from E to D, because most nanobodies start with this residue in their sequence. The production cell line of the host lab was *E. coli* WK6 strain. For the target protein induction, the traditional IPTG induction method, which is an established method for generation of nanobodies in the host lab, was used (nanobodies were first discovered in 1993 in this particular lab)(Hamers-Casterman et al. 1993). Each production cycle was initiated by a use of a freshly transformed colony. After every transformation, the isolated plasmids were sent to sequencing to eliminate the risk of commencing the production with a false colony (Pardon et al. 2014).

In contrast, in the home lab, the auto-induction method had been utilised (Studier 2005). This novel method offers many advantages over IPTG, including the reducing the risk of over-expression, toxicity, and metabolic burden on the host cells (Dvorak et al. 2015). The

auto-induction approach also has potential to yield higher protein levels due to the extended period of production time. Despite the advantages of the auto-induction method, the IPTG method was selected due to its compatibility with the specific requirements of the redesigned sdAbs. In order to utilize IPTG machinery, the production process was further optimised by changing the production cell line to WK6 *E. coli* at the host institute. The disadvantage of this strain is being devoid of selection pressure like the one in BL21 (DE) pLysS *E. coli* cells, which is the presence of chloramphenicol resistance gene. However the ampicillin antibiotic resistance of pHen6 plasmid can compensate for this absence (**Figure 16**).

As an optimisation strategy, WK6 *E. coli* cells were also grown under novel auto-induction conditions (both commercial and home-made auto-induction media), instead of using TB media and IPTG induction. Based on the outcomes of this trial, the overall yield was improved, while the impurities were diminished compared to recombinant production done in parallel with BL21 (DE) pLysS *E. coli* cells under the same conditions. It was shown that the aforementioned lacks of W6K *E. coli* cells was not a great disadvantage, as the overall performance of this strain compensates for these disadvantages. As such, many parameters would prove to be irrelevant as a source of the general problems related with the quality and the quantity of the sdAbs. Despite many different optimisation attempts, regrettably, none of the sdAb productions yielded pure enough protein to proceed with the planned animal studies. During these trials, the production of established control nanobodies was also included to eliminate the technical problems caused by the performer or the methodological executions. The purity of these control nanobodies was confirmed with SDS-PAGE results, which eliminates the recombinant protein production related technical difficulties.

Due to unforeseen problems, the planned studies abroad were left incomplete. In response to these challenges, I volunteered in several other projects at the host lab in order to compensate for the time, effort, and costs they invested in the original project. Until this time, I maintained contact with the host lab in Belgium. They accepted to conduct the remaining planned animal studies (WP5 and WP6 at **Table 93**) as outlined in the original fellowship application, if we could successfully produce the newly designed sdAbs in the host lab and ship them to Belgium.

Upon returning to the home institute, efforts were made to salvage the situation by focusing on purification optimisation studies to successfully purify our sdAbs, as well as the host lab designed sdAbs. The goal was to produce purified samples and send them to the host

lab again, allowing them to complete the studies using the reserved mice (validity extended until August 2023). Despite attempting various strategies to improve the purification of sdAbs, we could not achieve the desired outcome. Although we had initially successfully purified sdAb samples and sent them to the host lab, further attempts at reproducing them were unsuccessful, even with the new designs of sdAbs. During this time, the VUB ethical committee approval has expired (**Addenda**, **Section 8.2.**). Therefore, the reserved animals were no longer eligible for the host institute. As a consequence, this part of the project was left incomplete. This was the main setback for the abroad studies, impacting the expected research outcomes. Considering the work packages that were completed, over 40% of the originally planned project was executed. Although the studies did not go as planned, the overall experience gained during this fellowship was invaluable. During this fellowship, I gained a diverse set of techniques and great knowledge. It was a great honour and an extraordinary experience to be at the VUB, where the pioneering discovery of nanobodies first occurred.

In addition to the mice studies conducted abroad, zebrafish was also utilized for *in vivo* testing of the Bevacizumab sdAb at the home lab. The Tg(fli1:EGFP)^{y1} zebrafish strain was used during these studies. This transgenic zebrafish model expresses eGFP under the control of the fli1 gene promoter, allowing real-time visualisation of blood vessels via fluorescence. According to the results, significant changes in the vascularisation of the SIV were not observed as expected. This suggests that the tested sdAb variants and Bevacizumab did not have a significant impact on the vascularisation in the zebrafish under the examined conditions. Optimisation studies were also inititated on zebrafish studies to define the optimal injection time, concentration, and injection site. The visualisation method was also changed between fluorescent and confocal microscopy. Despite the higher quality images produced in confocal microscopy, fluorescent microscopy also has the advantage of faster visualisation of the fish immediately after the microinjection. Previously, there were variations in the incubation conditions that may also have caused discrepancies in the experimental outcomes. The zebrafish maintenance temperature was originally set at 34°C by the zebrafish facility, but it was later lowered to an optimal temperature of 28°C. We encountered problems regarding the measurements caused by the drawing tool of the imaging software. Previously a pixel-based calculation system was used to measure the SIV area. To address these various challenges, optimisation studies were performed. The altered parameters included the fixation

liquid, injection time frame (between 72nd to 80th h) to prevent outer branching at the SIV area. The mating of the fish was also altered to generate better transgenic models. Previously, wild-type fish were mated with Fli strain ones. Later WT fish was changed to TL strain to crossbreeed with Fli that produced progeny with reduced outer branch formation.

Despite the optimisation efforts, the main concern was the relevancy of the selected zebrafish model for testing Bevacizumab. Human VEGF-A gene can produce six alternatively spliced isoforms: VEGF-A_{121}, VEGF-A_{145}, VEGF-A_{165}, VEGF-A_{183}, VEGF-A_{189}, and VEGF-A_{206} (Enomoto et al. 2003). Among these mRNA isoforms, zebrafish produces VEGF-A_{121} as the predominant isoform in adults and VEGF-A_{165} isoform in early embryos via alternative splicing (Gong et al. 2004). Based on comparative sequence analysis of zebrafish $VEGF_{165}$ to the human $VEGF_{165}$ isoform, it was discovered that there is only a 52% amino acid identity (Liang et al. 1998; Holmes and Zachary 2005). That raised a concern regarding the compatibility of zebrafish for the testing of Bevacizumab and sdAbs as *in vivo* model. There is a possibility that the specificity of Bevacizumab for human VEGF-A may limit its effectiveness in zebrafish. The differences originate at the key residue positions. The 87th and 88th residues of Bevacizumab create a tightly packed interface with the human VEGF-A (Yu et al. 2008), which are not present in the zebrafish counterparts. Additionally, the varying abundance of VEGF-A isoforms in zebrafish ($VEGF_{165}$ to $VEGF_{121}$ transition) during its rapid development stage complicates the model organism structure. In light of these findings, the choice of the zebrafish model for testing Bevacizumab may not be suitable, as it does not reflect the complexity of the metastatic tumour environment.

In response to unsuccessful experimental outcomes, the focus was shifted towards *in silico* approaches. These studies aimed to determine any significant differences between traditional IgGs and nanobodies. By undertaking a detailed analysis of these differences, we can identify key residues at specific locations to facilitate the successful conversion of full-IgG to sdAb format. For that purpose, both the sequential and the structural features of antibodies and nanobodies were investigated via several in-depth analyses using diverse sets of data. Drawing upon the lessons learned from this study paves a way for a more refined and advanced design of new and improved sdAbs in the future.

6. CONCLUSION & FUTURE PERSPECTIVES

Our endeavors in the production of sdAb as a recombinant protein have been met with several challenges, but also presented learning opportunities. The indecipherable purity obstacle promotes the idea of formation of inclusion bodies during the protein expression process, which suggests the design of our sdAb may require further optimization.

The remaining prospective avenue for the future is the establishment of a feasible and reliable purification method for the successful isolation of sdAbs through recombinant protein production. In the design of our sdAbs, additional residues were incorporated at the C-terminus, including a FLAG tag, a His tag, and a Factor Xa recognition site. Given the relatively small size of the V_H sdAbs, the addition of these extra residues is a point of consideration, as it may influence their overall folding and consequently, their stability. This redesign resulted in the sdAbs having ~14.5 kDa molecular weight where the originally designed sdAbs at ~18.9 kDa. This is a substantial increase in molecular weight, approximately 30%. It is worth noting that the His tag could have been utilized both for purification and detection purposes, rendering the FLAG tag potentially redundant in the original design meanwhile significantly lowering the total molecular weight of the molcules. This alternative requires a further evaluation, especially considering the potential impact on the biophysical properties and functional efficacy of the sdAbs. Alternative protein isolation can be also achieved by extracting proteins either from the supernatant as leaky expression, or from the bacterial periplasmic space. For this purpose, the backup plans described in this thesis can serve as a future initiative.

The overall research outcomes have led us to consider the necessity of molecular docking simulations to better understand the interaction dynamics and improve the design. Furthermore, although the binding affinity of our sdAb does not currently match that of Bevacizumab, rendering it unsuitable for therapeutic use, there is a promising avenue in its application as a tool for companion diagnostics.

The elevated levels of VEGF have been frequently reported in various cancers types; therefore, monitoring VEGF levels has become a valuable tool in cancer diagnosis and progression (Niu and Chen 2010). Specifically, it holds potential for patient stratification, offering valuable insights in personalized medicine as a quick, easy, and indivually tailored method, which gives information about the eligibity of a patient to anti-angiogenic therapy.

The utilisation of a smaller sdAb presents a niche as a diagnostic agent. This limited application arises from its inherent characteristic of a shortened half-life, while preventing its use as a therapeutic agent. This limitation explains the need for further research and optimisation to explore its therapeutic potential.

Considering the half-life of sdAbs, an additional study was designed to extend their molecular weight, and consequently their half-life, so that the therapeutic potential of sdAbs can be investigated. This approach involves doubling their size and fusing them together using a specifically designed linker, inspired by the hinge region found in camelid hcIgGs. The problem is that fine-tuning of the linker engineering is required to balance the structural flexibility, while retaining target binding specificity. This project was designed in great detail, but it could not be executed due to budget concerns. Regrettably, this led to the eventual cancellation of the project. However, the potential extension of sdAb half-life through molecular modification remains as a prospective avenue for future research.

In addition to half-life extention, *in vitro* testing of sdAbs was planned in detail, utilising HUVEC cells. The planned experiments aim to assess the effects of engineered sdAbs on HUVEC cells. The experiments involved *in vitro* testing on HUVEC cells to determine: activity of engineered sdAbs, inhibition of VEGF-induced proliferation using the MTT assay and effect on migration ability based on a wound healing experiment to understand how VEGF influences the migratory behaviour of HUVEC cells and how sdAbs may modulate this effect. Although the cells were obtained and successfully cultured, further studies were unfortunately not conducted due to unforeseen circumstances. The potential insights from these experiments remain unexplored as it presents an important avenue for future research.

The paradigm shift towards *in silico* methodologies can address the drawbacks of experimental problems, and open new avenues in the field of antibody design and development. Additionally, the insights gained from the *in silico* study results can be leveraged for better designs of sdAbs in the future.

The optimisation experiments in this study can significantly advance the understanding of antibody engineering. They can provide a foundation to the design, optimisation, and development of innovative antibody formats with enhanced characteristics and functionalities. The examination of the physicochemical features and structural

adaptations in V_HHs compared to conventional V_Hs provides insights into the adaptive strategies employed in nature to optimise antibody function across diverse contexts.

This research contributes to the optimisation efforts on antibody engineering by addressing numerous possible strategies in the improvement of recombinant protein production yield and purity. Additionally, our findings from the *in silico* study shed light on the engineering approaches for the efficient conversion of humanised V_H into smaller sdAb formats through camelisation, while addressing the immunogenicity concerns by drawing inspiration from the ingenuity of nature.

Overall, while presenting some obstacles, this journey has opened up new pathways for research and application.

7. REFERENCES

'6 - Variable chain engineering – humanization and optimization approaches.' in. 2012. William R. Strohl and Lila M. Strohl (eds.), *Therapeutic Antibody Engineering* (Woodhead Publishing).

Abanades, B., W. K. Wong, F. Boyles, G. Georges, A. Bujotzek, and C. M. Deane. 2023. 'ImmuneBuilder: Deep-Learning models for predicting the structures of immune proteins', *Communications Biololgy*, 6: 575.

Arbabi-Ghahroudi, M. 2022. 'Camelid Single-Domain Antibodies: Promises and Challenges as Lifesaving Treatments', *Int J Mol Sci*, 23.

Arbabi-Ghahroudi, Mehdi. 2017. 'Camelid Single-Domain Antibodies: Historical Perspective and Future Outlook', *Frontiers in Immunology*, 8.

Avdesh, A., M. Chen, M. T. Martin-Iverson, A. Mondal, D. Ong, S. Rainey-Smith, K. Taddei, M. Lardelli, D. M. Groth, G. Verdile, and R. N. Martins. 2012. 'Regular care and maintenance of a zebrafish (Danio rerio) laboratory: an introduction', *J Vis Exp*: e4196.

Bando, H., H. A. Weich, M. Brokelmann, S. Horiguchi, N. Funata, T. Ogawa, and M. Toi. 2005. 'Association between intratumoral free and total VEGF, soluble VEGFR-1, VEGFR-2 and prognosis in breast cancer', *Br J Cancer*, 92: 553-61.

Bannas, P., J. Hambach, and F. Koch-Nolte. 2017. 'Nanobodies and Nanobody-Based Human Heavy Chain Antibodies As Antitumor Therapeutics', *Front Immunol*, 8: 1603.

Bao, Ru-Meng, Hong-Ming Yang, Chang-Mei Yu, Wei-Fen Zhang, and Jin-Bao Tang. 2016. 'An efficient protocol to enhance the extracellular production of recombinant protein from Escherichia coli by the synergistic effects of sucrose, glycine, and Triton X-100', *Protein Expr Purif*, 126: 9-15.

Bauer, B., A. Mally, and D. Liedtke. 2021. 'Zebrafish Embryos and Larvae as Alternative Animal Models for Toxicity Testing', *Int J Mol Sci*, 22.

Belanger, K., U. Iqbal, J. Tanha, R. MacKenzie, M. Moreno, and D. Stanimirovic. 2019. 'Single-Domain Antibodies as Therapeutic and Imaging Agents for the Treatment of CNS Diseases', *Antibodies (Basel)*, 8.

Bélanger, K., U. Iqbal, J. Tanha, R. MacKenzie, M. Moreno, and D. Stanimirovic. 2019. 'Single-Domain Antibodies as Therapeutic and Imaging Agents for the Treatment of CNS Diseases', *Antibodies (Basel)*, 8: 27-52.

Bielenberg, D. R., and B. R. Zetter. 2015. 'The Contribution of Angiogenesis to the Process of Metastasis', *Cancer J*, 21: 267-73.

Bouaouda, H., M. R. Achaaban, M. Ouassat, M. Oukassou, M. Piro, E. Challet, K. El Allali, and P. Pevet. 2014. 'Daily regulation of body temperature rhythm in the camel (Camelus dromedarius) exposed to experimental desert conditions', *Physiol Rep*, 2.

Boyd, S. D., and S. A. Joshi. 2014. 'High-Throughput DNA Sequencing Analysis of Antibody Repertoires', *Microbiol Spectr*, 2.

Bradbury, Andrew R. M., Sachdev Sidhu, Stefan Dübel, and John McCafferty. 2011. 'Beyond natural antibodies: the power of in vitro display technologies', *Nat Biotechnol*, 29: 245-54.

Brown, M., F. P. Assen, A. Leithner, J. Abe, H. Schachner, G. Asfour, Z. Bago-Horvath, J. V. Stein, P. Uhrin, M. Sixt, and D. Kerjaschki. 2018. 'Lymph node blood vessels provide exit routes for metastatic tumor cell dissemination in mice', *Science*, 359: 1408-11.

Cai, W., and X. Chen. 2008. 'Multimodality molecular imaging of tumor angiogenesis', *J Nucl Med*, 49 Suppl 2: 113s-28s.

Cheong, W. S., C. Y. Leow, A. B. Abdul Majeed, and C. H. Leow. 2020. 'Diagnostic and therapeutic potential of shark variable new antigen receptor (VNAR) single domain antibody', *Int J Biol Macromol*, 147: 369-75.

Chiu, M. L., D. R. Goulet, A. Teplyakov, and G. L. Gilliland. 2019. 'Antibody Structure and Function: The Basis for Engineering Therapeutics', *Antibodies (Basel)*, 8.

Choi, Y., C. Hua, C. L. Sentman, M. E. Ackerman, and C. Bailey-Kellogg. 2015. 'Antibody humanization by structure-based computational protein design', *MAbs*, 7: 1045-57.

Chothia, Cyrus, Arthur M. Lesk, Anna Tramontano, Michael Levitt, Sandra J. Smith-Gill, Gillian Air, Steven Sheriff, Eduardo A. Padlan, David Davies, William R. Tulip, Peter M. Colman, Silvia Spinelli, Pedro M. Alzari, and Roberto J. Poljak. 1989. 'Conformations of immunoglobulin hypervariable regions', *Nature*, 342: 877-83.

Chung, A. S., J. Lee, and N. Ferrara. 2010. 'Targeting the tumour vasculature: insights from physiological angiogenesis', *Nat Rev Cancer*, 10: 505-14.

Collins, A. M., and C. T. Watson. 2018. 'Immunoglobulin Light Chain Gene Rearrangements, Receptor Editing and the Development of a Self-Tolerant Antibody Repertoire', *Front Immunol*, 9: 2249.

Conchillo-Solé, Oscar, Natalia S. de Groot, Francesc X. Avilés, Josep Vendrell, Xavier Daura, and Salvador Ventura. 2007. 'AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides', *BMC Bioinformatics*, 8: 65.

Conrath, K. E., M. Lauwereys, M. Galleni, A. Matagne, J. M. Frère, J. Kinne, L. Wyns, and S. Muyldermans. 2001. 'Beta-lactamase inhibitors derived from single-domain antibody fragments elicited in the camelidae', *Antimicrob Agents Chemother*, 45: 2807-12.

Conrath, K. E., U. Wernery, S. Muyldermans, and V. K. Nguyen. 2003. 'Emergence and evolution of functional heavy-chain antibodies in Camelidae', *Dev Comp Immunol*, 27: 87-103.

Conrath, K., C. Vincke, B. Stijlemans, J. Schymkowitz, K. Decanniere, L. Wyns, S. Muyldermans, and R. Loris. 2005. 'Antigen binding and solubility effects upon the veneering of a camel VHH in framework-2 to mimic a VH', *J Mol Biol*, 350: 112-25.

Crauwels, M., N. Van Vaerenbergh, N. B. Kulaya, C. Vincke, M. D'Huyvetter, N. Devoogdt, S. Muyldermans, and C. Xavier. 2020. 'Reshaping nanobodies for affinity purification on protein a', *N Biotechnol*, 57: 20-28.

Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner. 2004. 'WebLogo: a sequence logo generator', *Genome Res*, 14: 1188-90.

D'Huyvetter, M., C. Xavier, V. Caveliers, T. Lahoutte, S. Muyldermans, and N. Devoogdt. 2014. 'Radiolabeled nanobodies as theranostic tools in targeted radionuclide therapy of cancer', *Expert Opin Drug Deliv*, 11: 1939-54.

Dall'Acqua, W. F., M. M. Damschroder, J. Zhang, R. M. Woods, L. Widjaja, J. Yu, and H. Wu. 2005. 'Antibody humanization by framework shuffling', *Methods*, 36: 43-60.

Davies, J., and L. Riechmann. 1994. "Camelising' human antibody fragments: NMR studies on VH domains', *FEBS Lett*, 339: 285-90.

De Genst, E., D. Saerens, S. Muyldermans, and K. Conrath. 2006. 'Antibody repertoire development in camelids', *Dev Comp Immunol*, 30: 187-98.

De Genst, Erwin, Karen Silence, Klaas Decanniere, Katja Conrath, Remy Loris, Jörg Kinne, Serge Muyldermans, and Lode Wyns. 2006. 'Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies', *Proceedings of the National Academy of Sciences of the United States of America*, 103: 4586-91.

de Groot, N. S., V. Castillo, R. Graña-Montes, and S. Ventura. 2012. 'AGGRESCAN: method, application, and perspectives for drug design', *Methods Mol Biol*, 819: 199-220.

De Meyer, T., S. Muyldermans, and A. Depicker. 2014. 'Nanobody-based products as research and diagnostic tools', *Trends Biotechnol*, 32: 263-70.

de Sousa-Pereira, P., and J. M. Woof. 2019. 'IgA: Structure, Function, and Developability', *Antibodies (Basel)*, 8.

Deac, A., P. VeliČković, and P. Sormanni. 2019. 'Attentive Cross-Modal Paratope Prediction', *Journal of computational biology*, 26: 536-45.

Debie, P., N. B. Declerck, D. van Willigen, C. M. Huygen, B. De Sloovere, L. Mateusiak, J. Bridoux, J. Puttemans, N. Devoogdt, F. W. B. van Leeuwen, and S. Hernot. 2021. 'The Design and Preclinical Evaluation of a Single-Label Bimodal Nanobody Tracer for Image-Guided Surgery', *Biomolecules*, 11.

Debie, P., C. Lafont, M. Defrise, I. Hansen, D. M. van Willigen, F. W. B. van Leeuwen, R. Gijsbers, M. D'Huyvetter, N. Devoogdt, T. Lahoutte, P. Mollard, and S. Hernot. 2020. 'Size and affinity kinetics of nanobodies influence targeting and penetration of solid tumours', *J Control Release*, 317: 34-42.

Dewulf, J., K. Adhikari, C. Vangestel, T. V. D. Wyngaert, and F. Elvas. 2020. 'Development of Antibody Immuno-PET/SPECT Radiopharmaceuticals for Imaging of Oncological Disorders-An Update', *Cancers (Basel)*, 12.

Diaz, M., R. L. Stanfield, A. S. Greenberg, and M. F. Flajnik. 2002. 'Structural analysis, selection, and ontogeny of the shark new antigen receptor (IgNAR): identification of a new locus preferentially expressed in early development', *Immunogenetics*, 54: 501-12.

Dunbar, James, Konrad Krawczyk, Jinwoo Leem, Claire Marks, Jaroslaw Nowak, Cristian Regep, Guy Georges, Sebastian Kelm, Bojana Popovic, and Charlotte M. Deane. 2016. 'SAbPred: a structure-based antibody prediction server', *Nucleic Acids Research*, 44: W474-W78.

Dupont, Camille Amandine, Kristina Riegel, Malvika Pompaiah, Hartmut Juhl, and Krishnaraj Rajalingam. 2021. 'Druggable genome and precision medicine in cancer: current challenges', *Febs j*, 288: 6142-58.

Duvaud, Séverine, Chiara Gabella, Frédérique Lisacek, Heinz Stockinger, Vassilios Ioannidis, and Christine Durinx. 2021. 'Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users', *Nucleic Acids Res*, 49: W216-W27.

Dvorak, P., L. Chrast, P. I. Nikel, R. Fedr, K. Soucek, M. Sedlackova, R. Chaloupkova, V. de Lorenzo, Z. Prokop, and J. Damborsky. 2015. 'Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway', *Microb Cell Fact*, 14: 201.

English, H., J. Hong, and M. Ho. 2020. 'Ancient species offers contemporary therapeutics: an update on shark VNAR single domain antibody sequences, phage libraries and potential clinical applications', *Antib Ther*, 3: 1-9.

Enomoto, H., I. Inoki, K. Komiya, T. Shiomi, E. Ikeda, K. Obata, H. Matsumoto, Y. Toyama, and Y. Okada. 2003. 'Vascular endothelial growth factor isoforms and their receptors are expressed in human osteoarthritic cartilage', *Am J Pathol*, 162: 171-81.

Fang, X., Y. Zheng, Y. Duan, Y. Liu, and W. Zhong. 2019. 'Recent Advances in Design of Fluorescence-Based Assays for High-Throughput Screening', *Anal Chem*, 91: 482-504.

Feige, M. J., M. A. Grawert, M. Marcinowski, J. Hennig, J. Behnke, D. Auslander, E. M. Herold, J. Peschek, C. D. Castro, M. Flajnik, L. M. Hendershot, M. Sattler, M. Groll, and J. Buchner. 2014. 'The structural analysis of shark IgNAR antibodies reveals evolutionary principles of immunoglobulins', *Proc Natl Acad Sci U S A*, 111: 8155-60.

Fernandez-Escamilla, A. M., F. Rousseau, J. Schymkowitz, and L. Serrano. 2004. 'Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins', *Nat Biotechnol*, 22: 1302-6.

Ferrara, N. 2002. 'Role of vascular endothelial growth factor in physiologic and pathologic angiogenesis: therapeutic implications', *Semin Oncol*, 29: 10-4.

Ferrara, N., K. J. Hillan, and W. Novotny. 2005. 'Bevacizumab (Avastin), a humanized anti-VEGF monoclonal antibody for cancer therapy', *Biochem Biophys Res Commun*, 333: 328-35.

Flajnik, M. F., N. Deschacht, and S. Muyldermans. 2011. 'A case of convergence: why did a simple alternative to canonical antibodies arise in sharks and camels?', *PLoS Biol*, 9: e1001120.

Frenzel, A., T. Schirrmann, and M. Hust. 2016. 'Phage display-derived human antibodies in clinical development and therapy', *MAbs*, 8: 1177-94.

Galli, S. J., and M. Tsai. 2012. 'IgE and mast cells in allergic disease', *Nat Med*, 18: 693-704.

Garcia, Josep, Herbert I. Hurwitz, Alan B. Sandler, David Miles, Robert L. Coleman, Regula Deurloo, and Olivier L. Chinot. 2020. 'Bevacizumab (Avastin®) in cancer treatment: A review of 15 years of clinical experience and future outlook', *Cancer Treatment Reviews*, 86: 102017.

Gerber, H. P., and N. Ferrara. 2005. 'Pharmacology and pharmacodynamics of bevacizumab as monotherapy or in combination with cytotoxic therapy in preclinical studies', *Cancer Res*, 65: 671-80.

Goldman, E. R., J. L. Liu, D. Zabetakis, and G. P. Anderson. 2017. 'Enhancing Stability of Camelid and Shark Single Domain Antibodies: An Overview', *Frontiers in Immunology*, 8: 865.

Gong, B., D. Liang, T. G. Chew, and R. Ge. 2004. 'Characterization of the zebrafish vascular endothelial growth factor A gene: comparison with vegf-A genes in mammals and Fugu', *Biochim Biophys Acta*, 1676: 33-40.

Greenberg, A. S., D. Avila, M. Hughes, A. Hughes, E. C. McKinney, and M. F. Flajnik. 1995. 'A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks', *Nature*, 374: 168-73.

Guan, X. 2015. 'Cancer metastases: challenges and opportunities', *Acta Pharm Sin B*, 5: 402-18.

Gutzeit, C., K. Chen, and A. Cerutti. 2018. 'The enigmatic function of IgD: some answers at last', *Eur J Immunol*, 48: 1101-13.

Hamers-Casterman, C., T. Atarhouch, S. Muyldermans, G. Robinson, C. Hammers, E. Bajyana Songa, N. Bendahman, and R. Hammers. 1993. 'Naturally occurring antibodies devoid of light chains', *Nature*, 363: 446-48.

Hanahan, D., and R. A. Weinberg. 2000. 'The hallmarks of cancer', *Cell*, 100: 57-70.

———. 2011. 'Hallmarks of cancer: the next generation', *Cell*, 144: 646-74.

Hansel, Trevor T., Harald Kropshofer, Thomas Singer, Jane A. Mitchell, and Andrew J. T. George. 2010. 'The safety and side effects of monoclonal antibodies', *Nature Reviews Drug Discovery*, 9: 325-38.

Haraguchi, M., K. Komuta, A. Akashi, S. Matsuzaki, J. Furui, and T. Kanematsu. 2002. 'Elevated IL-8 levels in the drainage vein of resectable Dukes' C colorectal cancer indicate high risk for developing hepatic metastasis', *Oncol Rep*, 9: 159-65.

Hicklin, D. J., and L. M. Ellis. 2005. 'Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis', *J Clin Oncol*, 23: 1011-27.

Hoey, R. J., H. Eom, and J. R. Horn. 2019. 'Structure and development of single domain antibodies as modules for therapeutics and diagnostics', *Exp Biol Med (Maywood)*, 244: 1568-76.

Holik, H. A., F. M. Ibrahim, A. A. Elaine, B. D. Putra, A. Achmad, and A. H. S. Kartamihardja. 2022. 'The Chemical Scaffold of Theranostic Radiopharmaceuticals: Radionuclide, Bifunctional Chelator, and Pharmacokinetics Modifying Linker', *Molecules*, 27.

Holmes, D. I., and I. Zachary. 2005. 'The vascular endothelial growth factor (VEGF) family: angiogenic factors in health and disease', *Genome Biol*, 6: 209.

Huber, R. 1980. 'Spatial structure of immunoglobulin molecules', *Klin Wochenschr*, 58: 1217-31.

Hummer, A. M., B. Abanades, and C. M. Deane. 2022. 'Advances in computational structure-based antibody design', *Curr Opin Struct Biol*, 74: 102379.

Huynh, K., and C. L. Partch. 2015. 'Analysis of protein stability and ligand interactions by thermal shift assay', *Curr Protoc Protein Sci*, 79: 28.9.1-28.9.14.

Ishikawa-Ankerhold, H. C., R. Ankerhold, and G. P. Drummen. 2012. 'Advanced fluorescence microscopy techniques--FRAP, FLIP, FLAP, FRET and FLIM', *Molecules*, 17: 4047-132.

Joseph, T. M., D. Kar Mahapatra, A. Esmaeili, Ł Piszczyk, M. S. Hasanin, M. Kattali, J. Haponiuk, and S. Thomas. 2023. 'Nanoparticles: Taking a Unique Position in Medicine', *Nanomaterials (Basel)*, 13.

Kanno, S., N. Oda, M. Abe, Y. Terai, M. Ito, K. Shitara, K. Tabayashi, M. Shibuya, and Y. Sato. 2000. 'Roles of two VEGF receptors, Flt-1 and KDR, in the signal transduction of VEGF effects in human vascular endothelial cells', *Oncogene*, 19: 2138-46.

Karadag, M., M. Arslan, N. E. Kaleli, and S. Kalyoncu. 2020. 'Physicochemical determinants of antibody-protein interactions', *Adv Protein Chem Struct Biol*, 121: 85-114.

Kazazi-Hyseni, F., J. H. Beijnen, and J. H. Schellens. 2010. 'Bevacizumab', *Oncologist*, 15: 819-25.

Khalili, H., S. Brocchini, P. T. Khaw, and S. K. Filippov. 2018. 'Comparative thermodynamic analysis in solution of a next generation antibody mimetic to VEGF', *RSC Adv*, 8: 35787-93.

Khorami-Moghadam, A., A. R. Jalilian, K. Yavari, B. Alirezapour, M. Mazidi, and M. Mirzaii. 2013. 'Synthesis, radiolabeling and quality control of 111In-DOTA-bevacizumab for radioimmunoscintigraphy of VEGF receptors', *Radiochimca Acta*, 101: 577-84.

Kim, J. H., and H. J. Hong. 2012. 'Humanization by CDR grafting and specificity-determining residue grafting', *Methods Mol Biol*, 907: 237-45.

Kishiro, I., S. Kato, D. Fuse, T. Yoshida, S. Machida, and N. Kaneko. 2002. 'Clinical significance of vascular endothelial growth factor in patients with primary lung cancer', *Respirology*, 7: 93-8.

Kovalenko, O. V., A. Olland, N. Piché-Nicholas, A. Godbole, D. King, K. Svenson, V. Calabro, M. R. Müller, C. J. Barelle, W. Somers, D. S. Gill, L. Mosyak, and L. Tchistiakova. 2013. 'Atypical antigen recognition mode of a shark immunoglobulin new antigen receptor (IgNAR) variable domain characterized by humanization and structural analysis', *J Biol Chem*, 288: 17408-19.

Kovaleva, Marina, Laura Ferguson, John Steven, Andrew Porter, and Caroline Barelle. 2014. 'Shark variable new antigen receptor biologics – a novel technology platform for therapeutic drug development', *Expert Opinion on Biological Therapy*, 14: 1527-39.

Kraft, A., K. Weindel, A. Ochs, C. Marth, J. Zmija, P. Schumacher, C. Unger, D. Marmé, and G. Gastl. 1999. 'Vascular endothelial growth factor in the sera and effusions of patients with malignant and nonmalignant disease', *Cancer*, 85: 178-87.

Krah, Simon, Christian Schröter, Stefan Zielonka, Martin Empting, Bernhard Valldorf, and Harald Kolmar. 2016. 'Single-domain antibodies for biomedical applications', *Immunopharmacology and Immunotoxicology*, 38: 21-28.

Kręcisz, Paweł, Kamila Czarnecka, Leszek Królicki, Elżbieta Mikiciuk-Olasik, and Paweł Szymański. 2021. 'Radiolabeled Peptides and Antibodies in Medicine', *Bioconjugate Chemistry*, 32: 25-42.

Krissinel, E., and K. Henrick. 2007. 'Inference of macromolecular assemblies from crystalline state', *J Mol Biol*, 372: 774-97.

Kusumanto, Y. H., W. A. Dam, G. A. Hospers, C. Meijer, and N. H. Mulder. 2003. 'Platelets and granulocytes, in particular the neutrophils, form important compartments for circulating vascular endothelial growth factor', *Angiogenesis*, 6: 283-7.

Kut, C., F. Mac Gabhann, and A. S. Popel. 2007. 'Where is VEGF in the body? A meta-analysis of VEGF distribution in cancer', *Br J Cancer*, 97: 978-85.

Kwon, Y. C., and M. C. Jewett. 2015. 'High-throughput preparation methods of crude extract for robust cell-free protein synthesis', *Sci Rep*, 5: 8663.

Kyte, J., and R. F. Doolittle. 1982. 'A simple method for displaying the hydropathic character of a protein', *J Mol Biol*, 157: 105-32.

Langley, Ries J., Yi Tian Ting, Fiona Clow, Paul G. Young, Fiona J. Radcliff, Jeong Min Choi, Richard P. Sequeira, Silva Holtfreter, Heather Baker, and John D. Fraser. 2017. 'Staphylococcal enterotoxin-like X (SElX) is a unique superantigen with functional features of two major families of staphylococcal virulence factors', *PLOS Pathogens*, 13: e1006549.

Lauwereys, M., M. Arbabi Ghahroudi, A. Desmyter, J. Kinne, W. Hölzer, E. De Genst, L. Wyns, and S. Muyldermans. 1998. 'Potent enzyme inhibitors derived from dromedary heavy-chain antibodies', *Embo j*, 17: 3512-20.

Lee, S. H., D. Jeong, Y. S. Han, and M. J. Baek. 2015. 'Pivotal role of vascular endothelial growth factor pathway in tumor angiogenesis', *Ann Surg Treat Res*, 89: 1-8.

Lee, S. S., D. Fitch, M. F. Flajnik, and E. Hsu. 2000. 'Rearrangement of immunoglobulin genes in shark germ cells', *The Journal of experimental medicine*, 191: 1637-48.

Li, H., P. W. Kantoff, J. Ma, M. J. Stampfer, and D. J. George. 2005. 'Prediagnostic plasma vascular endothelial growth factor levels and risk of prostate cancer', *Cancer Epidemiol Biomarkers Prev*, 14: 1557-61.

Li, W., A. Schäfer, S. S. Kulkarni, X. Liu, D. R. Martinez, C. Chen, Z. Sun, S. R. Leist, A. Drelich, L. Zhang, M. L. Ura, A. Berezuk, S. Chittori, K. Leopold, D. Mannar, S. S. Srivastava, X. Zhu, E. C. Peterson, C. T. Tseng, J. W. Mellors, D. Falzarano, S. Subramaniam, R. S. Baric, and D. S. Dimitrov. 2020. 'High Potency of a Bivalent Human V(H) Domain in SARS-CoV-2 Animal Models', *Cell*, 183: 429-41.e16.

Li, X., X. Duan, K. Yang, W. Zhang, C. Zhang, L. Fu, Z. Ren, C. Wang, J. Wu, R. Lu, Y. Ye, M. He, C. Nie, N. Yang, J. Wang, H. Yang, X. Liu, and W. Tan. 2016. 'Comparative Analysis of Immune Repertoires between Bactrian Camel's Conventional and Heavy-Chain Antibodies', *PLoS One*, 11: e0161801.

Liang, Dong, Xingzhi Xu, Alvin J. Chin, Nataraja V. Balasubramaniyan, Mellissa A. L. Teo, T. J. Lam, Eric S. Weinberg, and Ruowen Ge. 1998. 'Cloning and characterization of vascular endothelial growth factor (VEGF) from zebrafish, Danio rerio', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1397: 14-20.

Liberis, E., P. Velickovic, P. Sormanni, M. Vendruscolo, and P. Liò. 2018. 'Parapred: antibody paratope prediction using convolutional and recurrent neural networks', *Bioinformatics*, 34: 2944-50.

Linding, R., J. Schymkowitz, F. Rousseau, F. Diella, and L. Serrano. 2004. 'A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins', *J Mol Biol*, 342: 345-53.

Liu, C., H. Lin, L. Cao, K. Wang, and J. Sui. 2022. 'Research progress on unique paratope structure, antigen binding modes, and systematic mutagenesis strategies of single-domain antibodies', *Front Immunol*, 13: 1059771.

Ljungberg, B., J. Jacobsen, S. Häggström-Rudolfssson, T. Rasmuson, G. Lindh, and K. Grankvist. 2003. 'Tumour vascular endothelial growth factor (VEGF) mRNA in relation to serum VEGF protein levels and tumour progression in human renal cell carcinoma', *Urol Res*, 31: 335-40.

Louros, N., K. Konstantoulea, M. De Vleeschouwer, M. Ramakers, J. Schymkowitz, and F. Rousseau. 2020. 'WALTZ-DB 2.0: an updated database containing structural information of experimentally determined amyloid-forming peptides', *Nucleic Acids Res*, 48: D389-d93.

Lu, R. M., Y. C. Hwang, I. J. Liu, C. C. Lee, H. Z. Tsai, H. J. Li, and H. C. Wu. 2020. 'Development of therapeutic antibodies for the treatment of diseases', *J Biomed Sci*, 27: 1.

Lugano, R., M. Ramachandran, and A. Dimberg. 2020. 'Tumor angiogenesis: causes, consequences, challenges and opportunities', *Cell Mol Life Sci*, 77: 1745-70.

Madeira, F., M. Pearce, A. R. N. Tivey, P. Basutkar, J. Lee, O. Edbali, N. Madhusoodanan, A. Kolesnikov, and R. Lopez. 2022. 'Search and sequence analysis tools services from EMBL-EBI in 2022', *Nucleic Acids Res*, 50: W276-w79.

Marinou, Katerina A., and Ismene A. Dontas. 2023. 'European Union Legislation for the Welfare of Animals Used for Scientific Purposes: Areas Identified for Further Discussion', *Animals*, 13: 2367.

Massagué, J., and A. C. Obenauf. 2016. 'Metastatic colonization by circulating tumour cells', *Nature*, 529: 298-306.

Mendoza, Matthew N., Mike Jian, Moeko T. King, and Cory L. Brooks. 2020. 'Role of a noncanonical disulfide bond in the stability, affinity, and flexibility of a VHH specific for the Listeria virulence factor InlB', *Protein science : a publication of the Protein Society*, 29: 1004-17.

Mitra, S., and P. C. Tomar. 2021. 'Hybridoma technology; advancements, clinical significance, and future aspects', *J Genet Eng Biotechnol*, 19: 159.

Miyazaki, T., T. Yamazaki, R. Sugisawa, M. E. Gershwin, and S. Arai. 2018. 'AIM associated with the IgM pentamer: attackers on stand-by at aircraft carrier', *Cell Mol Immunol*, 15: 563-74.

Mock, M., S. Edavettal, C. Langmead, and A. Russell. 2023. 'AI can help to speed up drug discovery - but only if we give it the right data', *Nature*, 621: 467-70.

Morais, M., and M. T. Ma. 2018. 'Site-specific chelator-antibody conjugation for PET and SPECT imaging with radiometals', *Drug Discov Today Technol*, 30: 91-104.

Muller, Y. A., Y. Chen, H. W. Christinger, B. Li, B. C. Cunningham, H. B. Lowman, and A. M. de Vos. 1998. 'VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 A resolution and mutational analysis of the interface', *Structure*, 6: 1153-67.

Muyldermans, S. 2001. 'Single domain camel antibodies: current status', *J Biotechnol*, 74: 277-302.

Muyldermans, S., T. Atarhouch, J. Saldanha, J. A. Barbosa, and R. Hamers. 1994. 'Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains', *Protein Eng*, 7: 1129-35.

Needleman, Saul B., and Christian D. Wunsch. 1970. 'A general method applicable to the search for similarities in the amino acid sequence of two proteins', *J Mol Biol*, 48: 443-53.

Nelson, Aaron L. 2010. 'Antibody fragments', *MAbs*, 2: 77-83.

Nguyen, V. K., C. Su, S. Muyldermans, and W. van der Loo. 2002. 'Heavy-chain antibodies in Camelidae; a case of evolutionary innovation', *Immunogenetics*, 54: 39-47.

Nishida, N., H. Yano, T. Nishida, T. Kamura, and M. Kojiro. 2006. 'Angiogenesis in cancer', *Vasc Health Risk Manag*, 2: 213-9.

Niu, G., and X. Chen. 2010. 'Vascular endothelial growth factor as an anti-angiogenic target for cancer therapy', *Curr Drug Targets*, 11: 1000-17.

———. 2011. 'Size effect in molecular imaging of vascular endothelial growth factor', *Chem Biol*, 18: 819-20.

Paget, Stephen. 1889. 'THE DISTRIBUTION OF SECONDARY GROWTHS IN CANCER OF THE BREAST', *The Lancet*, 133: 571-73.

Pardon, E., T. Laeremans, S. Triest, S. G. Rasmussen, A. Wohlkönig, A. Ruf, S. Muyldermans, W. G. Hol, B. K. Kobilka, and J. Steyaert. 2014. 'A general protocol for the generation of Nanobodies for structural biology', *Nat Protoc*, 9: 674-93.

Peach, C. J., V. W. Mignone, M. A. Arruda, D. C. Alcobia, S. J. Hill, L. E. Kilpatrick, and J. Woolard. 2018. 'Molecular Pharmacology of VEGF-A Isoforms: Binding and Signalling at VEGFR2', *Int J Mol Sci*, 19.

Pecorino, Lauren. 2012. *Molecular biology of cancer : mechanisms, targets, and therapeutics* (Oxford University Press: Oxford).

Peyvandi, Flora, Marie Scully, Johanna A. Kremer Hovinga, Spero Cataland, Paul Knöbl, Haifeng Wu, Andrea Artoni, John-Paul Westwood, Magnus Mansouri Taleghani, Bernd Jilma, Filip Callewaert, Hans Ulrichts, Christian Duby, and Dominique Tersago. 2016. 'Caplacizumab for Acquired Thrombotic Thrombocytopenic Purpura', *New England Journal of Medicine*, 374: 511-22.

Platania, C. B., L. Di Paola, G. M. Leggio, G. L. Romano, F. Drago, S. Salomone, and C. Bucolo. 2015. 'Molecular features of interaction between VEGFA and anti-angiogenic drugs used in retinal diseases: a computational approach', *Front Pharmacol*, 6: 248.

Poon, R. T., I. O. Ng, C. Lau, L. X. Zhu, W. C. Yu, C. M. Lo, S. T. Fan, and J. Wong. 2001. 'Serum vascular endothelial growth factor predicts venous invasion in hepatocellular carcinoma: a prospective study', *Ann Surg*, 233: 227-35.

Presta, L. G., H. Chen, S. J. O'Connor, V. Chisholm, Y. G. Meng, L. Krummen, M. Winkler, and N. Ferrara. 1997. 'Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders', *Cancer Res*, 57: 4593-9.

Pyzik, Michal, Timo Rath, Wayne I. Lencer, Kristi Baker, and Richard S. Blumberg. 2015. 'FcRn: The Architect Behind the Immune and Nonimmune Functions of IgG and Albumin', *The Journal of Immunology*, 194: 4595-603.

Quinteros, Daniela A., José M. Bermúdez, Soledad Ravetti, Alicia Cid, Daniel A. Allemandi, and Santiago D. Palma. 2017. 'Chapter 25 - Therapeutic use of monoclonal antibodies: general aspects and challenges for drug delivery.' in Ecaterina Andronescu and Alexandru Mihai Grumezescu (eds.), *Nanostructures for Drug Delivery* (Elsevier).

Ranieri, G., R. Patruno, E. Ruggieri, S. Montemurro, P. Valerio, and D. Ribatti. 2006. 'Vascular endothelial growth factor (VEGF) as a target of bevacizumab in cancer: from the biology to the clinic', *Curr Med Chem*, 13: 1845-57.

Riechmann, L., and S. Muyldermans. 1999. 'Single domain antibodies: comparison of camel VH and camelised human VH domains', *J Immunol Methods*, 231: 25-38.

Rousseau, F., J. Schymkowitz, and L. Serrano. 2006. 'Protein aggregation and amyloidosis: confusion of the kinds?', *Curr Opin Struct Biol*, 16: 118-26.

Russo, A. E., D. Priolo, G. Antonelli, M. Libra, J. A. McCubrey, and F. Ferraù. 2017. 'Bevacizumab in the treatment of NSCLC: patient selection and perspectives', *Lung Cancer (Auckl)*, 8: 259-69.

Ryabova, N. A., V. V. Marchenkov, S. Y. Marchenkova, N. V. Kotova, and G. V. Semisotnov. 2013. 'Molecular chaperone GroEL/ES: unfolding and refolding processes', *Biochemistry (Mosc)*, 78: 1405-14.

Saerens, D., M. Pellis, R. Loris, E. Pardon, M. Dumoulin, A. Matagne, L. Wyns, S. Muyldermans, and K. Conrath. 2005. 'Identification of a universal VHH framework to graft non-canonical antigen-binding loops of camel single-domain antibodies', *J Mol Biol*, 352: 597-607.

Salven, P., A. Orpana, and H. Joensuu. 1999. 'Leukocytes and platelets of patients with cancer contain high levels of vascular endothelial growth factor', *Clin Cancer Res*, 5: 487-91.

Saman, H., S. S. Raza, S. Uddin, and K. Rasul. 2020. 'Inducing Angiogenesis, a Key Step in Cancer Vascularization, and Treatment Approaches', *Cancers (Basel)*, 12.

Schroeder, H. W., Jr., and L. Cavacini. 2010. 'Structure and function of immunoglobulins', *J Allergy Clin Immunol*, 125: S41-52.

Schumacher, D., J. Helma, A. F. L. Schneider, H. Leonhardt, and C. P. R. Hackenberger. 2018. 'Nanobodies: Chemical Functionalization Strategies and Intracellular Applications', *Angew Chem Int Ed Engl*, 57: 2314-33.

Seyfried, T. N., and L. C. Huysentruyt. 2013. 'On the origin of cancer metastasis', *Crit Rev Oncog*, 18: 43-73.

Sharrock, J., and J. C. Sun. 2020. 'Innate immunological memory: from plants to animals', *Curr Opin Immunol*, 62: 69-78.

Shepard, H. M., G. L. Phillips, D. Thanos C, and M. Feldmann. 2017. 'Developments in therapy with monoclonal antibodies and related proteins', *Clin Med (Lond)*, 17: 220-32.

Shibuya, M. 2011. 'Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies', *Genes Cancer*, 2: 1097-105.

Shinkaruk, S., M. Bayle, G. Laïn, and G. Déléris. 2003. 'Vascular endothelial cell growth factor (VEGF), an emerging target for cancer chemotherapy', *Curr Med Chem Anticancer Agents*, 3: 95-117.

Sormanni, P., F. A. Aprile, and M. Vendruscolo. 2018. 'Third generation antibody discovery methods: in silico rational design', *Chem Soc Rev*, 47: 9137-57.

Sormanni, Pietro, Leanne Amery, Sofia Ekizoglou, Michele Vendruscolo, and Bojana Popovic. 2017. 'Rapid and accurate in silico solubility screening of a monoclonal antibody library', *Scientific Reports*, 7: 8200.

Sormanni, Pietro, Francesco A. Aprile, and Michele Vendruscolo. 2015. 'The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility', *J Mol Biol*, 427: 478-90.

Stockhammer, G., A. Obwegeser, H. Kostron, P. Schumacher, A. Muigg, S. Felber, H. Maier, I. Slavc, E. Gunsilius, and G. Gastl. 2000. 'Vascular endothelial growth factor (VEGF)

is elevated in brain tumor cysts and correlates with tumor progression', *Acta Neuropathol*, 100: 101-5.

Studier, F. W. 2005. 'Protein production by auto-induction in high density shaking cultures', *Protein Expr Purif*, 41: 207-34.

Swindells, M. B., C. T. Porter, M. Couch, J. Hurst, K. R. Abhinandan, J. H. Nielsen, G. Macindoe, J. Hetherington, and A. C. Martin. 2017. 'abYsis: Integrated Antibody Sequence and Structure-Management, Analysis, and Prediction', *J Mol Biol*, 429: 356-64.

Tanha, Jamshid, Ping Xu, Zhigang Chen, Feng Ni, Howard Kaplan, Saran A. Narang, and C. Roger MacKenzie. 2001. 'Optimal Design Features of Camelized Human Single-domain Antibody Libraries*', *Journal of Biological Chemistry*, 276: 24774-80.

Thurber, G. M., M. M. Schmidt, and K. D. Wittrup. 2008. 'Antibody tumor penetration: transport opposed by systemic and antigen-mediated clearance', *Adv Drug Deliv Rev*, 60: 1421-34.

Tiller, K. E., and P. M. Tessier. 2015. 'Advances in Antibody Design', *Annual Review of Biomedical Engineering*, 17: 191-216.

Townsend, Catherine L., Julie M. J. Laffy, Yu-Chang Bryan Wu, Joselli Silva O'Hare, Victoria Martin, David Kipling, Franca Fraternali, and Deborah K. Dunn-Walters. 2016. 'Significant Differences in Physicochemical Properties of Human Immunoglobulin Kappa and Lambda CDR3 Regions', *Frontiers in Immunology*, 7.

Van Audenhove, I., and J. Gettemans. 2016. 'Nanobodies as Versatile Tools to Understand, Diagnose, Visualize and Treat Cancer', *EBioMedicine*, 8: 40-48.

van Zijl, F., G. Krupitza, and W. Mikulits. 2011. 'Initial steps of metastasis: cell invasion and endothelial transmigration', *Mutat Res*, 728: 23-34.

Vaneycken, I., M. D'Huyvetter, S. Hernot, J. De Vos, C. Xavier, N. Devoogdt, V. Caveliers, and T. Lahoutte. 2011. 'Immuno-imaging using nanobodies', *Curr Opin Biotechnol*, 22: 877-81.

Verheesen, P., and T. Laeremans. 2012. 'Selection by phage display of single domain antibodies specific to antigens in their native conformation', *Methods Mol Biol*, 911: 81-104.

Vidarsson, G., G. Dekkers, and T. Rispens. 2014a. 'IgG subclasses and allotypes: from structure to effector functions', *Frontiers in Immunology*, 5: 520.

———. 2014b. 'IgG subclasses and allotypes: from structure to effector functions', *Front Immunol*, 5: 520.

Vincke, C., R. Loris, D. Saerens, S. Martinez-Rodriguez, S. Muyldermans, and K. Conrath. 2009. 'General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold', *J Biol Chem*, 284: 3273-84.

Vira, S., E. Mekhedov, G. Humphrey, and P. S. Blank. 2010. 'Fluorescent-labeled antibodies: Balancing functionality and degree of labeling', *Anal Biochem*, 402: 146-50.

Wagner, P. D. 2011. 'The critical role of VEGF in skeletal muscle angiogenesis and blood flow', *Biochem Soc Trans*, 39: 1556-9.

Waldmann, T. A., W. Strober, and R. P. Mogielnicki. 1972. 'The renal handling of low molecular weight proteins. II. Disorders of serum protein catabolism in patients with tubular proteinuria, the nephrotic syndrome, or uremia', *The Journal of clinical investigation*, 51: 2162-74.

Walsh, I., F. Seno, S. C. Tosatto, and A. Trovato. 2014. 'PASTA 2.0: an improved server for protein aggregation prediction', *Nucleic Acids Res*, 42: W301-7.

Warram, J. M., E. de Boer, A. G. Sorace, T. K. Chung, H. Kim, R. G. Pleijhuis, G. M. van Dam, and E. L. Rosenthal. 2014. 'Antibody-based imaging strategies for cancer', *Cancer Metastasis Rev*, 33: 809-22.

Wesolowski, J., V. Alzogaray, J. Reyelt, M. Unger, K. Juarez, M. Urrutia, A. Cauerhff, W. Danquah, B. Rissiek, F. Scheuplein, N. Schwarz, S. Adriouch, O. Boyer, M. Seman, A. Licea, D. V. Serreze, F. A. Goldbaum, F. Haag, and F. Koch-Nolte. 2009. 'Single domain antibodies: promising experimental and therapeutic tools in infection and immunity', *Med Microbiol Immunol*, 198: 157-74.

Wilton, Emily E., Michael P. Opyr, Senthilkumar Kailasam, Ronja F. Kothe, and Hans-Joachim Wieden. 2018. 'sdAb-DB: The Single Domain Antibody Database', *ACS Synthetic Biology*, 7: 2480-84.

Wu, Y., S. Jiang, and T. Ying. 2017. 'Single-Domain Antibodies As Therapeutics against Human Viral Diseases', *Front Immunol*, 8: 1802.

Xavier, C., I. Vaneycken, M. D'Huyvetter, J. Heemskerk, M. Keyaerts, C. Vincke, N. Devoogdt, S. Muyldermans, T. Lahoutte, and V. Caveliers. 2013. 'Synthesis,

preclinical validation, dosimetry, and toxicity of 68Ga-NOTA-anti-HER2 Nanobodies for iPET imaging of HER2 receptor expression in cancer', *J Nucl Med*, 54: 776-84.

Yang, E. Y., and K. Shah. 2020. 'Nanobodies: Next Generation of Cancer Diagnostics and Therapeutics', *Front Oncol*, 10: 1182.

Ye, J., N. Ma, T. L. Madden, and J. M. Ostell. 2013. 'IgBLAST: an immunoglobulin variable domain sequence analysis tool', *Nucleic Acids Res*, 41: W34-40.

Yoshikawa, T., A. Tsuburaya, O. Kobayashi, M. Sairenji, H. Motohashi, S. Yanoma, and Y. Noguchi. 2000. 'Plasma concentrations of VEGF and bFGF in patients with gastric carcinoma', *Cancer Lett*, 153: 7-12.

Yu, L., X. Wu, Z. Cheng, C. V. Lee, J. LeCouter, C. Campa, G. Fuh, H. Lowman, and N. Ferrara. 2008. 'Interaction between bevacizumab and murine VEGF-A: a reassessment', *Invest Ophthalmol Vis Sci*, 49: 522-7.

Zarschler, K., S. Witecy, F. Kapplusch, C. Foerster, and H. Stephan. 2013. 'High-yield production of functional soluble single-domain antibodies in the cytoplasm of Escherichia coli', *Microb Cell Fact*, 12: 97.

Zhang, J., B. Gao, W. Zhang, Z. Qian, and Y. Xiang. 2018. 'Monitoring antiangiogenesis of bevacizumab in zebrafish', *Drug Des Devel Ther*, 12: 2423-30.

Zhang, Y., K. Y. Park, K. F. Suazo, and M. D. Distefano. 2018. 'Recent progress in enzymatic protein labelling techniques and their applications', *Chem Soc Rev*, 47: 9106-36.

Zielonka, S., M. Empting, J. Grzeschik, D. Konning, C. J. Barelle, and H. Kolmar. 2015. 'Structural insights and biomedical potential of IgNAR scaffolds from sharks', *MAbs*, 7: 15-25.

8. ADDENDA

8.1. Ethical committee approval report from Animal Experiments Local Ethics Committee (IBG-HAYDEK) at IBG

İZMİR BİYOTIP VE GENOM MERKEZİ
HAYVAN DENEYLERİ YEREL ETİK KURULU (İBG-HADYEK)
KARARI

Toplantı Tarihi	: 11/03/2020	**Toplantı Günü**	: Çarşamba
Toplantı Sayısı	: 4	**Toplantı Saati**	: 15:30

Sayın Doç.Dr Gökhan KARAKÜLAH,

2020-014 Protokol No'lu; yürütücüsü olduğunuz "Anti-Kanser Tek Domainli Antikor Takımlarının Geliştirilmesi" başlıklı projenin uygulanmasında etik açıdan sakınca olmadığına oy birliği ile karar verilmiştir.

Bilgilerinizi ve gereğini rica ederiz.

Doç. Dr. Güneş ÖZHAN BAYKAN **Başkan**	**Vet. Hek. Kerem ESMEN** **Başkan Yardımcısı**
Prof. Dr. H. Alper BAĞRIYANIK **Üye** (Katılmadı)	**Prof. Dr. Ensari GÜNELİ** **Üye**
Doç. Dr. Kasım DİRİL **Üye** (Katılmadı)	**Dr. Öğr. Üyesi Ayşe Banu DEMİR** **Üye**
(Projede araştırmacı) **Dr. Gülçin ÇAKAN AKDOĞAN** **Üye**	**Vet. Hek. Umur KELEŞ** **Üye**
H. Muammer KARAGÖZ **Üye**	

8.2. Ethical Committee approval report from Ethische Commissie Dierproeven at VUB

VRIJE UNIVERSITEIT BRUSSEL

Ethische Commissie Dierproeven
Ethical Committee for Animal Experiments

VERGUNNING - *LICENCE*

De Ethische Commissie Dierproeven geeft een gunstig advies voor onderstaand project en verleent hierbij de vergunning om het project te starten met inachtneming van onderstaande voorwaarden.

Based on a positive evaluation, the Ethical Committee for Animal Experiments hereby grants permission for this project to start, under the following conditions.

PROJECTNUMMER / *PROJECT NUMBER*: 21-272-19

TITEL PROJECT / *TITLE*: Pharmacokinetic analysis of camelized single domain antibodies targeting a growth factor involved in angiogenesis

DUUR / DURATION

Startdatum / *start date:*	Einddatum / *end date:*
01-09-2021	31-08-2023

GEBRUIKERS / *USERS*

Verantwoordelijke laboratoriumdirecteur / *Responsible laboratory director:*	Prof Tony Lahoutte
Proefleider / *Experiment manager:*	Sophie Hernot
Overige uitvoerende personen / *Other researcher(s):*	Kevin De Jonghe Fleur De Boe Lukasz Mateusisk Noemi Declercq Maxime Deladrière Sofie Pollenus
Verantwoordelijke van het laboratorium waar het project (deels) plaatsvindt, indien verschillend van hierboven / *Responsible of the laboratory where the experiments are (partially) conducted:*	

AANTAL DIEREN / NUMBER OF ANIMALS: 96

INSCHATTING VAN ERNSTKLASSE DOOR DE COMMISSIE / *ASSESSMENT OF PAIN AND SUFFERING BY THE COMMITTEE*

☒ Licht / *Mild*
☐ Matig / *Moderate*
☐ Ernstig / *Severe (duur/duration:* between 7 and 30 dagen/days)

RETROSPECTIEVE ANALYSE / *RETROSPECTIVE ANALYSIS*
Gelieve binnen de 2 maanden na het beëindigen van het project de retrospectieve analyse en aangepaste NTS op te sturen naar ecd-secretariaat@minf.be.
A retrospective analysis with an adjusted NTS should be submitted within 2 months after conclusion of the project. Please send it to ecd-secretariaat@minf.be.

DE VOORZITTER / THE *CHAIR*:
Prof. dr. Eline Menu

DATUM / *DATE*:
30-08-2021



8.3. CV in ARBIS format



NAZLI EDA ESKİER

Personal Information

Contact Information

Contact Address	Yalı Bu. [illegible] No:1[illegible] [illegible] -ıyaka, İzmir/TÜRKİYE
Telephone	1537 [illegible]
Email	n[illegible] [illegible] ail.com
Web Page	

Education Information

01 February 2018 - Now (6 Year)
Doktora, Doktora, DOKUZ EYLÜL ÜNİVERSİTESİ, TÜRKİYE
İZMİR ULUSLARARASI BİYOTIP VE GENOM ENSTİTÜSÜ, MOLEKÜLER BİYOLOJİ VE GENETİK (DR) (İNGİLİZCE)
Diploma Number: -
Cumulative Grade Point Average: 4.0 / 4.0

30 September 2015 - 14 January 2019 (3 Year 4 Month)
Lisans, Açıköğretim, ANADOLU ÜNİVERSİTESİ, TÜRKİYE
İŞLETME FAKÜLTESİ, İŞLETME PR. (AÇIKÖĞRETİM)
Diploma Number: 2019-12576
Cumulative Grade Point Average: 2.57 / 4.0

01 January 2014 - 01 July 2015 (1 Year 7 Month)
Yüksek Lisans, Tezli Program, NOTHINGHAM TRENT UNIVERSITY, İNGİLTERE
NOTHINGHAM TRENT UNIVERSITY, NOTHINGHAM TRENT UNIVERSITY
Thesis Title: EXPRESSION OF TISSUE TRANSGLUTAMINASE (TG2) ISOFORMS IN PROSTATE CANCER CELLS UNDER HYPOXIC CONDITIONS
Thesis Subject: Cancer Biology
Date: 2014

01 September 2012 - 01 January 2014 (1 Year 5 Month)
Yüksek Lisans, Tezli Program, YEDİTEPE ÜNİVERSİTESİ, TÜRKİYE
SAĞLIK BİLİMLERİ ENSTİTÜSÜ, MOLEKÜLER TIP ANABİLİM DALI
Thesis Title: Apolipoprotein E Geni Polimorfizimi ve İleri Dönem Alzheimer Hastalığı Arasındaki İlişkinin Araştırılması
Thesis Subject: Apoliporprotein geni polimorfizimi ve İleri Dönem Alzheimer Hastalığı arasındaki ilişkini araştırılması
Date: 2013
Thesis Advisor: TURGAY İSBİR

12 November 2010 - 03 June 2013 (2 Year 7 Month)
Ön Lisans, Açıköğretim, ANADOLU ÜNİVERSİTESİ, TÜRKİYE
AÇIKÖĞRETİM FAKÜLTESİ, SAĞLIK KURUMLARI İŞLETMECLİĞİ PR. (AÇIKÖĞRETİM)
Diploma Number: 2013-38682
Cumulative Grade Point Average: 2.39 / 4.0

27 August 2007 - 14 June 2012 (4 Year 10 Month)
Lisans, Anadal/Normal Öğretim, YEDİTEPE ÜNİVERSİTESİ, TÜRKİYE
MÜHENDİSLİK FAKÜLTESİ, GENETİK VE BİYOMÜHENDİSLİK PR. (İNGİLİZCE) (ÜCRETLİ)

Diploma Number: 522
Cumulative Grade Point Average: 2.62 / 4.0

27 August 2007 - 14 June 2012 (4 Year 10 Month)
Lisans, Anadal/Normal Öğretim, YEDİTEPE ÜNİVERSİTESİ, TÜRKİYE
MÜHENDİSLİK-MİMARLIK FAKÜLTESİ, GENETİK VE BİYOMÜHENDİSLİK PR. (İNGİLİZCE) (ÜCRETLİ)
Diploma Number: 522
Cumulative Grade Point Average: 2.62 / 4.0

Experience Information

01 June 2021 - 01 December 2021 (7 Month) (Tam Zamanlı)
ARAŞTIRMACI, DOKTORA ÖĞRENCİSİ, VRIJE UNIVERSITEIT BRUSSEL

01 February 2018 - Now (6 Year) (Tam Zamanlı)
MİSAFİR ARAŞTIRMACI, DOKTORA ÖĞRENCİSİ, İZMİR BİYOTIP VE GENOM MERKEZİ

01 July 2012 - 01 December 2013 (1 Year 6 Month) (Yarı Zamanlı)
ARAŞTIRMA GÖREVLİSİ, YÜKSEK LİSANS ÖĞRENCİSİ, YEDİTEPE ÜNİVERSİTESİ TIP FAKÜLTESİ

01 January 2010 - 01 January 2011 (1 Year 1 Month) (Yarı Zamanlı)
ARAŞTIRMA GÖREVLİSİ, YEDİTEPE ÜNİVERSİTESİ MÜHENDİSLİK-MİMARLIK FAKÜLTESİ

Scientific Technological Activity Fields

Scientific Technological Activity Field Information

Temel Bilimler -- Yaşam Bilimleri -- Moleküler Biyoloji ve Genetik -- Kanser Moleküler Biyolojisi

Temel Bilimler -- Yaşam Bilimleri -- Moleküler Biyoloji ve Genetik -- Protein Mühendisliği

R&D Competency

Books

M. ARSLAN, H. ALOTAİBİ, N. E. KALELİ, M. KARADAĞ, S. KALYONCU UZUNLAR & M. C. YARIMÇAM, KODON OPTİMİZASYONU VE REKOMBİNANT PROTEİN ÜRETİMİNDEKİ YERİ, S. E. ACUNER ZORLUUYSAL [Editors], PROTEİN: YAPISI, MÜHENDİSLİĞİ, ETKİLEŞİMLERİ, DİNAMİĞİ VE İLAÇ TASARIMINDAKİ YERİ(115 - 134), ISBN: 978-625-7564-03-8, TÜRKİYE: Ankara Nobel Tıp Kitapevleri, 01 October 2021, Kitapta Bölüm.

Articles

N. E. ESKİER, D. ESKİER, E. FİRUZAN & S. KALYONCU UZUNLAR, Physicochemical differences between camelid single-domain antibodies and mammalian antibodies, Turkish Journal of Biology, 2023, 1300-0152, 47, 6, 423.

M. KARADAG, M. ARSLAN, N. E. KALELİ & S. KALYONCU, Physicochemical determinants of antibody-protein interactions, ADVANCES IN PROTEIN CHEMISTRY AND STRUCTURAL BIOLOGY, VOL 121, 2020, 1876-1623, 121, 30.

N. E. KALELİ, M. KARADAG & S. KALYONCU, Phage display derived therapeutic antibodies have enriched aliphatic content: Insights for developability issues, PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS, 2019, 0887-3585, 87, 7, 607-618.

Conference Papers	N. E. KALELİ & S. KALYONCU UZUNLAR, Camelization of an anti-VEGF nanobody for cancer treatment, Tam Metin Bildiri, 2nd International Cancer & Ion Channels Congress, 22 September 2019, 24 October 2019.

TÜBİTAK Scholarships and Fundings

Project Info	20AG006, Hedefe Özgü Pan-Kanser Terapiler (PAN-TER), BİDEB Burslusu, Yürürlükte, KAMAG - Kamu Araştırmaları Destek Grubu, 1004 - Mükemmeliyet Merkezi Destek Programı, ARDEB, Project Participation/Leave Dates: 8/1/22 - 5/6/23, Project Start/End Dates: 10/15/21 - 10/15/25. 119Z161, Kanser Tedavisi İçin VEGF'e Bağlanan Yeni Nesil Rekombinant Antikor Parçacıklarının Geliştirilmesi, Burslu, Sonuçlandı, KBAG - Kimya Biyoloji Araştırma Destek Grubu, 3501 - Kariyer, ARDEB, Project Participation/Leave Dates: 10/15/19 - 10/31/20, Project Start/End Dates: 10/15/19 - 6/15/22.
BIDEB Supports	NAZLI EDA ESKİER, Etkinlik Destekleri ve Eğitim Bursları Müdürlüğü, 2250-Lisansüstü Burslan Performans Programı, Destek Sona Erdi, 2023 - 1. NAZLI EDA ESKİER, Etkinlik Destekleri ve Eğitim Bursları Müdürlüğü, 2224-A Yurt Dışı Bilimsel Etkinliklere Katılma Desteği Programı, BİDEB İade, 2021 - 3, 9/8/21 - 9/10/21. NAZLI EDA ESKİER, Araştırma Burs ve Destekleri Müdürlüğü, 2214-A Yurt Dışı Araştırma Burs Programı (Doktora öğrencileri için), Türkiye'ye Döndü, 2021 - 1, 8/13/21 - 11/15/21. NAZLI EDA ESKİER, Etkinlik Destekleri ve Eğitim Bursları Müdürlüğü, 2211-C Öncelikli Alanlara Yönelik Yurt İçi Doktora Burs Programı, Bursu Kesildi, 2020 - 1, 3/1/20 - 2/28/23. NAZLI EDA ESKİER, Etkinlik Destekleri ve Eğitim Bursları Müdürlüğü, 2211-A Genel Yurt İçi Doktora Burs Programı, Başvurusu Reddedildi, 2019 - 3, 10/1/19 - 9/30/23.

Number of Panelist/Observer/Reporter Jobs

Refereeing/Panelist/Outer Consultant Count	ARDEB/BİDEB 0	TEYDEB 0	Total 0
Number of Observer/Consultant Jobs	ARDEB/BİDEB 0	TEYDEB 0	Total 0
Number of Reporter Jobs	ARDEB/BİDEB 0	TEYDEB 0	Total 0

Turkish Journal of Biology

Volume 47 | Number 6 Article 6

12-28-2023

Physicochemical differences between camelid single-domain antibodies and mammalian antibodies

NAZLI EDA ESKİER

DOĞA ESKİER

ESİN FİRUZAN

SİBEL KALYONCU

Follow this and additional works at: https://journals.tubitak.gov.tr/biology

 Part of the Biology Commons

Recommended Citation

ESKİER, NAZLI EDA; ESKİER, DOĞA; FİRUZAN, ESİN; and KALYONCU, SİBEL (2023) "Physicochemical differences between camelid single-domain antibodies and mammalian antibodies," *Turkish Journal of Biology*: Vol. 47: No. 6, Article 6. https://doi.org/10.55730/1300-0152.2674
Available at: https://journals.tubitak.gov.tr/biology/vol47/iss6/6

This Article is brought to you for free and open access by TÜBİTAK Academic Journals. It has been accepted for inclusion in Turkish Journal of Biology by an authorized editor of TÜBİTAK Academic Journals. For more information, please contact academic.publications@tubitak.gov.tr.

TÜBİTAK

Turkish Journal of Biology

http://journals.tubitak.gov.tr/biology/

Research Article

Turk J Biol
(2023) 47: 423-436
© TÜBİTAK
doi:10.55730/1300-0152.2676

Physicochemical differences between camelid single-domain antibodies and mammalian antibodies

Nazlı Eda ESKİER[1,2], Doğa ESKİER[1,2], Esin FİRUZAN[3], Sibel KALYONCU UZUNLAR[2,*]

[1]İzmir International Biomedicine and Genome Institute, Dokuz Eylül University, İzmir, Turkiye
[2]İzmir Biomedicine and Genome Center, İzmir, Turkiye
[3]Department of Statistics, Faculty of Sciences, Dokuz Eylül University, İzmir, Turkiye

Received: 01.11.2023 • **Accepted/Published Online:** 07.12.2023 • **Final Version:** 28.12.2023

Background/aim: In recent years, single-domain antibodies, also known as nanobodies, have emerged as an alternative to full immunoglobulin Gs (IgGs), due to their various advantages, including increased solubility, faster clearance, and cheaper production. Nanobodies are generally derived from the variable domain of the camelid heavy-chain-only immunoglobulin Gs (hcIgGs). Due to the high sequence homology between variable heavy chains of camelids (V_HHs) and humans (V_Hs), hcIgGs are ideal candidates for nanobody development. However, further examination is needed to understand the structural differences between V_Hs and V_HHs. This analysis is essential for nanobody engineering to mitigate potential immunogenicity, while preserving stability, functionality, and antigen specificity.

Materials and methods: We obtained the V_H and V_HH sequences of various camelid and non-camelid mammalian antibodies from public databases and used multiple sequence alignment based on the Chothia numbering scheme. Aligned sequences were subjected to diverse analyses encompassing paratope length, binding prediction, motif, disulfide bridge, salt bridge profiling, and physicochemical characteristic distribution. Logistic Regression coupled with the Boruta - Random Forest algorithm facilitated the comprehensive examination of physicochemical properties.

Results: Our findings revealed longer, less variable paratope sequences in V_HHs, along with specific antigen binding residues with increased binding potential compared to V_Hs. Although the V_Hs showed more heterogeneous noncanonical disulfide bond patterns, the V_HHs had a higher number of noncanonical disulfide bridges. Intriguingly, a typical salt bridge between the 94th and 101st positions in the V_Hs had a very low encounter rate in the V_HHs. Surprisingly, we also identified notable differences in the physicochemical patterns of mostly conserved frameworks (FWs), especially the FW2 and FW3 regions, between V_Hs and V_HHs.

Conclusion: Our findings point to possible key sites in V_HHs as candidate residues for nanobody engineering efforts.

Key words: Antibody, nanobody, camelid, physicochemical properties, logistic regression, machine learning

1. Introduction

Antibodies play a vital role in safeguarding our immune system. Among these immune defenders, immunoglobulin Gs (IgGs) are the most commonly found ones in humans and mammalians (Vidarsson et al., 2014). However, different species possess different antibody formats with novel benefits that contribute to an increased survival rate. Camelids have evolved a solution to cope with extreme environments through adaptive changes (Flajnik et al., 2011), and such environmental circumstances eventually paved the way for fully functional alternative antibodies: heavy-chain-only IgGs (hcIgGs) (Hamers Casterman et al., 1993).

To generate diversity in variable domains, members of the Camelidae (camels, llamas, alpacas, and vicunas) can encode both conventional IgGs and hcIgGs (Figure 1) (Muyldermans et al., 1994). Approximately the same amount of heterodimeric canonical IgGs (IgG1) and homodimeric hcIgGs (IgG2 and IgG3) coexist in the camelid repertoire, even sharing the same germline (D) J-genes during B-cell maturation (De Genst et al., 2006). hcIgG molecules lack the light chain (L) and the first constant region of the heavy chain (H). To compensate, hcIgGs developed unique adaptations, such as a longer hinge region, which confers physical flexibility upon antigen binding (Muyldermans, 2013). There are unconventionally charged and polar residues at the former variable H–variable L (V_H–V_L) interface, which correspond to the framework (FW)-2 region in hcIgGs. These preclude association with other hcIgG molecules

* Correspondence: sibel.kalyoncu@ibg.edu.tr

This work is licensed under a Creative Commons Attribution 4.0 International License.

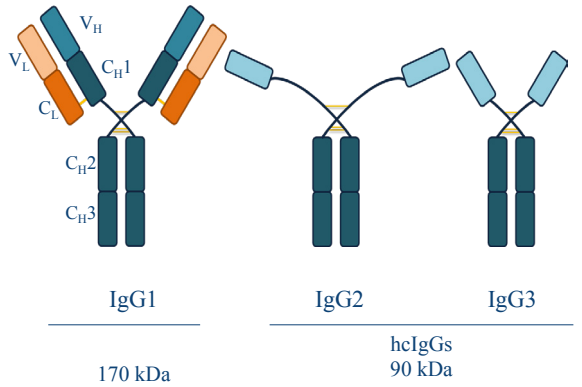


Figure 1. Different antibody types found in camelids. Light orange sections show the variable light chain (V_L) domain, while the light blue sections represent the variable heavy chain (V_H). Dark orange sections and dark blue regions indicate the constant region of the light chain (C_L) and the constant regions of the heavy chain (C_H1–3), respectively. Yellow lines indicate intradomain disulfide bridges.

and increase solubility and stability (Goldman et al., 2017). These FW2 residues are extremely conserved in both IgGs and hcIgGs (Mitchell and Colwell, 2018). Although there are six different complementarity-determining regions (CDRs) (HCDR1–3 and LCDR1–3) responsible for antigen binding in the IgGs, there are only HCDR1–3 in hcIgGs (Vu et al., 1997). Instead of relying on the LCDRs for epitope recognition, hcIgGs have circumvented their absence by recruiting more residues into HCDR1 and HCDR3 (Muyldermans et al., 1994). The HCDR3 loop in the variable heavy chains of camelids (V_HHs) shields the FW2 residues to preclude the interaction of hydrophilic residues found at the former V_H-V_L association site (Bond et al., 2003). To maintain the stability of such loops, the introduction of internal disulfide bonds between HCDRs is another adaptation in the format of hcIgGs. The presence of noncanonical disulfide bonds further improves the stability by allowing reversible refolding of the structure as well as reducing the entropic folding penalty, while providing protection against proteases under physiological conditions (Mendoza et al., 2020). Intriguingly, through all these adaptations, hcIgGs retain the same level of antigen binding capability as IgGs and might surpass conventional IgGs with superior affinity and avidity (Nguyen et al., 2002).

The field of antibody research has witnessed a paradigm shift with the emergence of stand-alone V_HHs of hcIgGs, also referred to as nanobodies or single-domain antibodies. The term "nanobody" was first coined because they are stable and soluble binders, retaining their nanomolar affinity to their target even at the smallest possible size. Nanobodies possess several advantages, such as rapid clearance, robustness, cheaper and faster production, resistance to aggregation, improved stability and solubility, recognition of novel/hidden epitopes, faster extravasation, and better tumor/tissue penetration compared to IgGs (Pillay and Muyldermans, 2021). Due to these benefits, potential applications of nanobodies have spread across many different fields, such as therapeutics, diagnostics, in vivo imaging, and even agriculture (Muyldermans, 2021).

While monoclonal antibodies have been the primary focus of optimization efforts, there is growing interest in optimizing other antibody formats. There are engineering approaches to convert traditional IgGs into nanobody formats. Camelization is a process where the monomeric human variable heavy chain (V_H) is used as a scaffold and key residues of V_HHs are introduced into this moiety (Riechmann and Muyldermans, 1999). The V_HHs of hcIgGs display 80% sequence homology and structural similarity to human V_Hs (Conrath et al., 2001). Due to this homology between camelids and humans, V_HHs are ideal candidates for nanobody development (Bélanger et al., 2019).

Although prior studies have acknowledged differences between V_Hs and V_HHs, the precise and comprehensive structural and physicochemical characterization of these variances remains insufficiently explored. While it is widely recognized that V_HHs possess longer CDR3 regions, the exact lengths and their functional implications have yet to be precisely determined. Additionally, the physicochemical nature of conserved FW regions and the prevalence and precise locations of noncanonical cysteine bonds in V_HHs remain elusive.

2. Materials and methods

2.1. Datasets

In this study, we analyzed the V_H sequences of 5030 mammalian IgGs (*Cricetulus migratorius*; chimeric *Homo sapiens*/*Mus musculus*; chimeric *Mus musculus*/*Homo sapiens*; *Homo sapiens*; *Mus musculus*; *Pan troglodytes*; *Rattus norvegicus*) acquired from the abYsis database (Swindells et al., 2017) and the V_HH sequences of 1260 nanobodies (camelid otherwise unspecified; *Camelus bactrianus*; *Camelus dromedarius*; *Lama glama*; *Vicugna pacos*) publicly available in the Single Domain Antibody Database (Wilton et al., 2018). We obtained the V_H sequences by trimming non-V_H domains from the given sequences. The list of V_Hs and V_HH IDs is available in Supplementary File 1. After obtaining the sequences, we conducted numerous in-depth analyses (Figure 2).

2.2. Multiple sequence alignment

To ensure a precise and standardized annotation of HCDRs and FW regions, and to facilitate a comprehensive and uniform analysis of the aligned sequences, we aligned

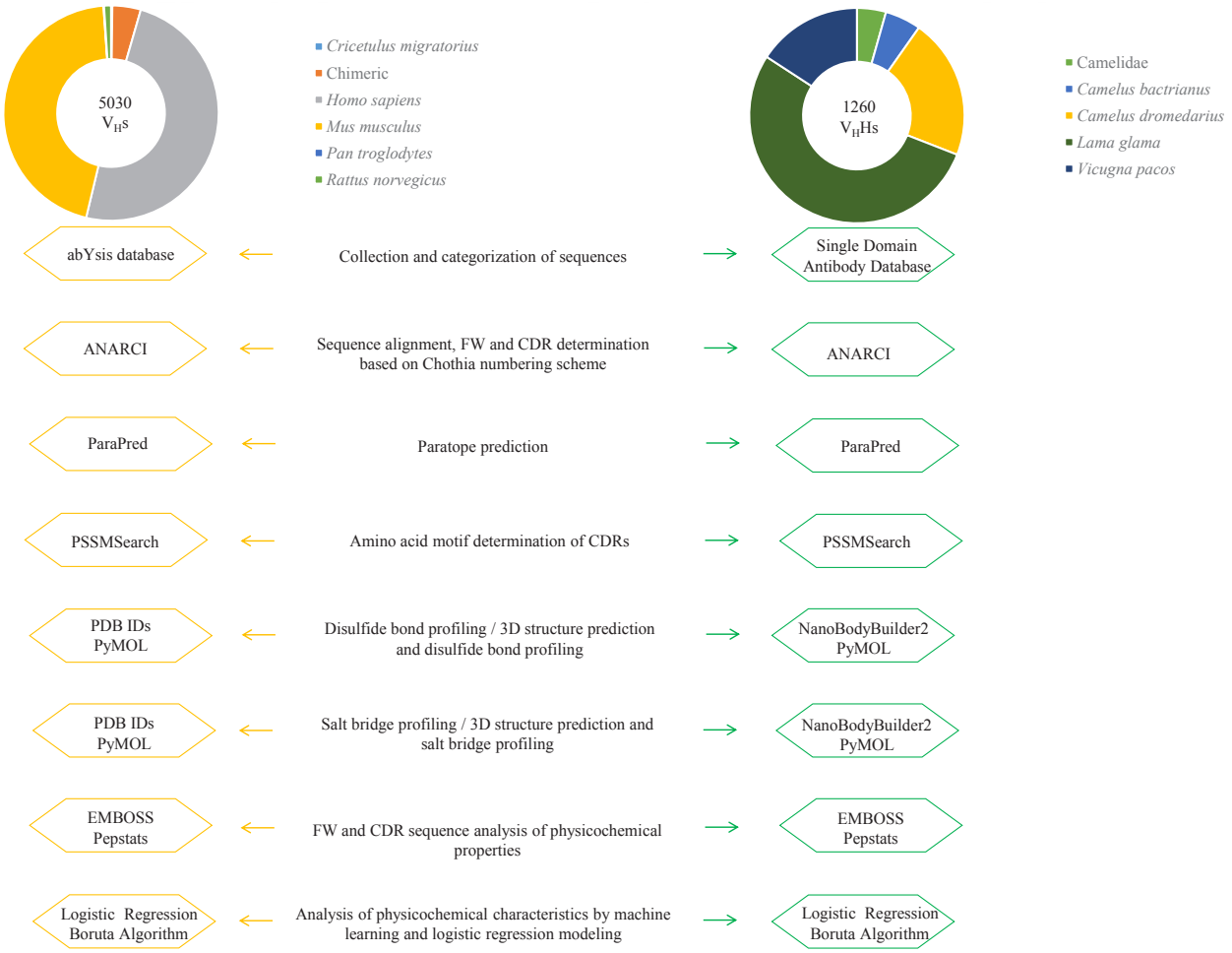


Figure 2. Flowchart showing the analyses conducted in this study and the tools used for each analysis. The pie charts indicate the fraction of each species within the obtained sequences, with the total number of sequences shown inside the circle.

multiple FASTA sequences according to the Chothia numbering scheme using a locally installed version of the ANARCI Antigen Receptor Numbering and Receptor Classification tool (Dunbar et al., 2016) to define HCDRs and FW regions (Table S1).

2.3. Paratope binding prediction

After the determination of HCDRs, to categorize the binding potential differences between the CDRs of V_Hs and V_HHs, we predicted the paratope binding regions and binding scores including those for the two flanking residues of HCDRs using "Parapred: Antibody Paratope Prediction Using Convolutional and Recurrent Neural Networks" (Liberis et al., 2018), which is compatible with the output of ANARCI.

2.4. Sequence conservation motif analysis

To elucidate conserved amino acids at specific positions within the sequences in a comprehensive and visually intuitive manner, we determined the amino acid motifs of V_H and V_HH sequences using WebLogo (version 2.8.2) (Crooks et al., 2004), an online application for visualization of protein motifs, to show which amino acids are conserved at certain positions. FW and HCDR motif visualization is based on an amino acid frequency scoring method.

2.5. Disulfide bond analysis

We aligned all V_H and V_HH sequences that harbored double or multiples of two cysteine residues according to the Chothia numbering system. We highlighted canonical/noncanonical disulfide bridges and confirmed them both sequentially and structurally using PyMOL software (version 2.5.5) (Deac et al., 2019). Many V_HHs lack PDB IDs, as most of the sequence information was obtained from publications. To allow for the structural modeling of V_HHs even in instances where PDB IDs were unavailable, we homology-modeled the 3D structures of the V_HHs using NanoBodyBuilder2 (Abanades et al., 2023) to produce the structural modeling of the V_HHs with known sequence information.

2.6. Salt bridge analysis

After the previous analyses, to better understand the electrostatic interactions that contribute to the structural stability and functionality of the aligned V_H and V_HH sequences, we detected salt bridges in sequences aligned according to Chothia numbering. Specifically, we filtered the V_H and V_HH sequences to identify the presence of R, K, and H residues at the 94th position, as well as the D and E amino acids at the 101st position.

2.7. Physicochemical properties determination

We used the EMBOSS Pepstats tool (Madeira et al., 2022) to facilitate the calculation of physicochemical contents for both HCDRs and FW regions and to obtain information regarding the amino acid composition, hydrophobicity, charge, and other physicochemical properties as listed in Tables S2 and S3.

2.8. Statistical analysis

To better determine the power of the physicochemical variables to discern between V_Hs and V_HHs, we applied a logistic regression modeling approach and the Boruta algorithm based on supervised machine learning to data produced by EMBOSS Pepstats (Supplementary File 2).

2.9. Tools and plots

We ran all Unix-compatible tools on the Windows operating system and accessed them through a Docker Container. We generated length variation bar plots and antigen binding prediction score heatmaps using the R statistical computing environment (version 4.0.0), using the *ggplot2* graphics package (version 3.4.2) and the *pheatmap* (version 1.10.12) heatmap generation package, respectively. We conducted binary logistic regression using the *pscl*, *caret*, and *car* R packages to measure the statistical significance of physicochemical variables (predictors) in the classification of V_Hs (0) and V_HHs (1). We visualized the Boruta algorithm via the Boruta R package results using the *ROCR* R package. Further methods can be found in the Supplementary Files.

3. Results and discussion

3.1. Length variation in HCDRs

Initially, we performed a comparative analysis of the differences in the lengths of HCDRs (Figure 3). The distribution of HCDR1 and HCDR2 length profiles were highly similar between V_Hs and V_HHs. As anticipated, the most significant variation in length was encountered in HCDR3, which is the region primarily responsible for target binding. The length distribution of V_Hs was skewed to the left, with a mode length of 10 residues. Conversely, the V_HHs exhibited a length distribution that was skewed to the right, with a mode of 15 residues, consistent with other studies that highlighted the prevalence of longer HCDR3 loops in V_HHs. Interestingly, while V_Hs had a low number of outlier sequences with longer HCDR3s, the overall count of HCDR3s longer than 15 was higher in V_HHs than in V_Hs. There are two main functions of typically longer HCDR3s in V_HHs: (i) to fold over the hallmark residues residing at FW2, shielding those hydrophilic residues to preclude them from any other interactions, and (ii) to increase the antigen specificity/affinity by enlarging the surface to compensate for the absence of LCDRs (Conrath et al., 2005). The presence of longer HCDR3s became apparent when calculating average HCDR lengths based on species (Table S4 and Table S5). When examining the species-specific HCDR lengths, we found that *Vicugna pacos* exhibited the shortest average V_HH HCDR3 length, followed by *Camelus dromedarius* (Table S5). These species-specific variations in

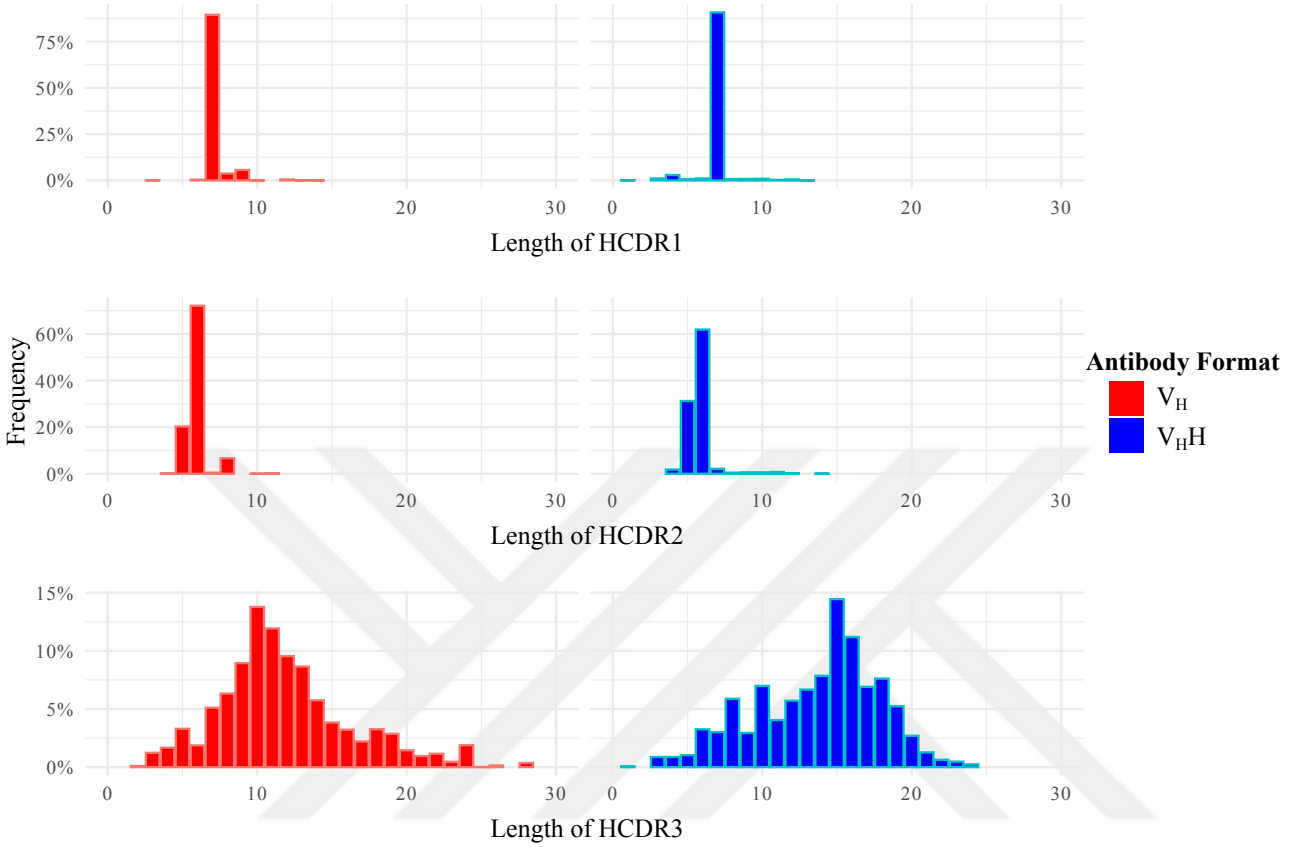


Figure 3. Bar plot displaying the HCDR length variation of VHs and V_HHs for HCDR1–3 and their encounter frequencies in each dataset of VHs and V_HHs. VHs are shown in red, while V_HHs are shown in blue. The y-axis represents the frequency of HCDR length based on their encounter rates. The HCDR windows were defined according to the Chothia numbering scheme.

the antibody repertoire can stem from the specific immune responses they generate. These differences can arise from evolutionary adaptations, such as variations in germline gene usage, or selection pressures acting on the antibody repertoire.

3.2. Antigen binding scores

CDRs are the most important parts in the antibody structure, as they mostly define the specificity/affinity to the target antigen. To identify which HCDR residues are the most important in target recognition, we proceeded to analyze the paratope scores indicating antigen binding within the HCDRs and the flanking FW regions (two residues from each side of the HCDR loops). According to the heatmaps, the V_Hs and V_HHs exhibited similar antigen binding profiles (Figure 4). Notably, HCDR1 exhibited the lowest contribution to antigen binding ability (Figures 4A and 4D). In each case, the 31st position consistently yielded the highest sum of scores, underscoring its pivotal role as the most important residue within HCDR1. The extra residues featured at the 31st A–G positions consistently exhibited high binding scores, suggesting their significant role in antigen binding whenever they were incorporated into the HCDR1 sequence. The 32nd position, as the last residue of HCDR1, and the 33rd position, as the first residue of FW2, also showed very high binding scores.

In HCDR2, the 52nd position had the highest binding score, closely followed by the 54th and 56th residues. Notably, the 52nd A residues were prevalent in numerous sequences within both datasets. In V_Hs, the presence of extra residues at 52A–F was associated with moderately to highly elevated binding scores (Figure 4B). For V_HHs, HCDR2 sequences incorporating a residue between the 52A–I positions consistently had very high binding scores (Figure 4E). These observations suggest that the inclusion of an additional residue within the frame of the 52nd position plays a significant role in antigen binding.

As anticipated, HCDR3 exhibited the greatest variability in paratope residues. For V_Hs, the 94th position made a considerable contribution (Figure 4C). However, in V_HHs, the functionality of the 94th residue was less pronounced (Figure 4F). Notably, in V_Hs, there was a high-intensity concentration of binding scores within the range of 95–100A. Conversely, when examining the V_HHs, this intensity concentration shifted towards the C-terminus, with the

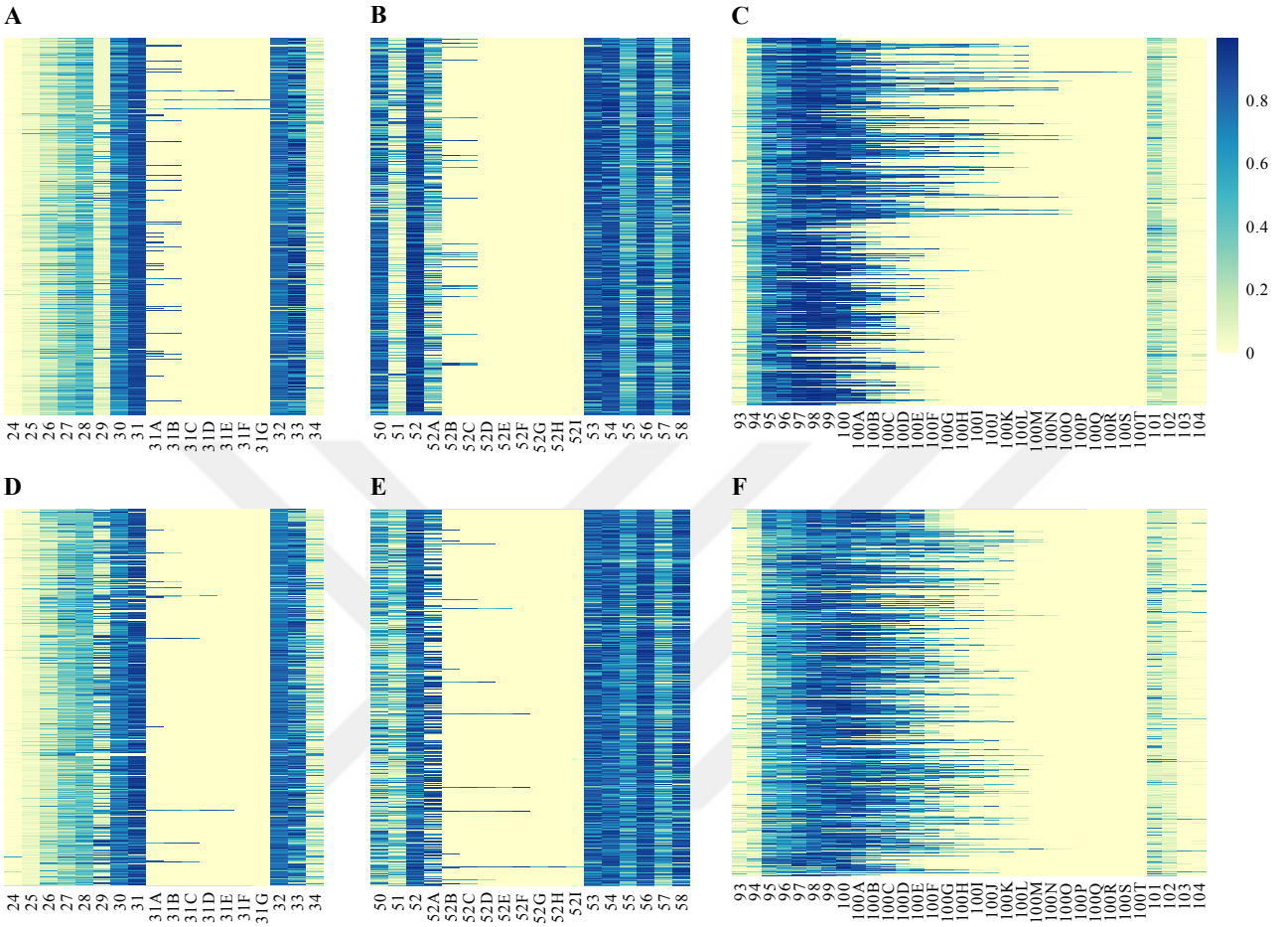


Figure 4. Heatmaps displaying the antigen binding prediction scores of HCDR1–3 regions of VHs (A–C) and V_HHs (D–F). Multiple sequences were aligned based on the Chothia numbering scheme. High binding scores are denoted by blue cells, whereas low binding scores are represented by yellow cells. The flanking framework residues on the two leftmost and rightmost columns serve as predictors for the binding score generated by the software. To facilitate proper alignment between VHs and V_HHs, empty columns were intentionally inserted for the alignment of shorter loops, all of which are indicated by yellow cells at alternating positions.

highest-scoring residues clustered between the 97 and 100B positions.

3.3. Differences between conserved motifs

Next, we extended our investigation to conservation motif analysis, which involves systematic examination of FWs and HCDRs, providing a holistic view of the sequence conservation landscape for both V_H and V_HH sequences (Figure 5). As CDRs exhibit a high degree of variability depending on their target specificity, direct comparisons between them were often not meaningful. While interpreting the results of this analysis, we concentrated on the comparatively more conserved FW regions rather than the hypervariable HCDRs.

In FW1, the remarkable distinction between V_Hs and V_HHs lay primarily in their initial amino acid residues. Most V_HHs started with 'Q' instead of 'E' compared to their V_H counterparts. In FW1, V_Hs exhibited standout basic residues at the 9th, 10th, 12th, and 13th positions, resulting in a positive charge. Specifically, the 23rd position in the V_Hs presented a positively charged 'K,' while V_HHs displayed more conserved 'A' in addition to 'K.'

A prominent divergence occurred in FW2 at the 37th, 44th, 45th, and 47th positions, and these transitions are well-known hallmark residues in camelids. The 'VGLW' motif at these hallmark residue positions in V_Hs underwent a transformation into 'FERF/YERF' in V_HHs. There were other minor differences at the 48th and 49th positions, although not as obvious as the hallmark residues.

The number of positions with minor differences between V_Hs and V_HHs notably increased in FW3, namely at the 63rd, 71st, 73rd, 74th, 78th, 81st, 82nd, 83rd, 84th, and 94th positions. The 94th position majorly differentiated between V_Hs and V_HHs, with V_Hs featuring 'R,' a positively charged amino acid, in contrast to 'A' in V_HHs.

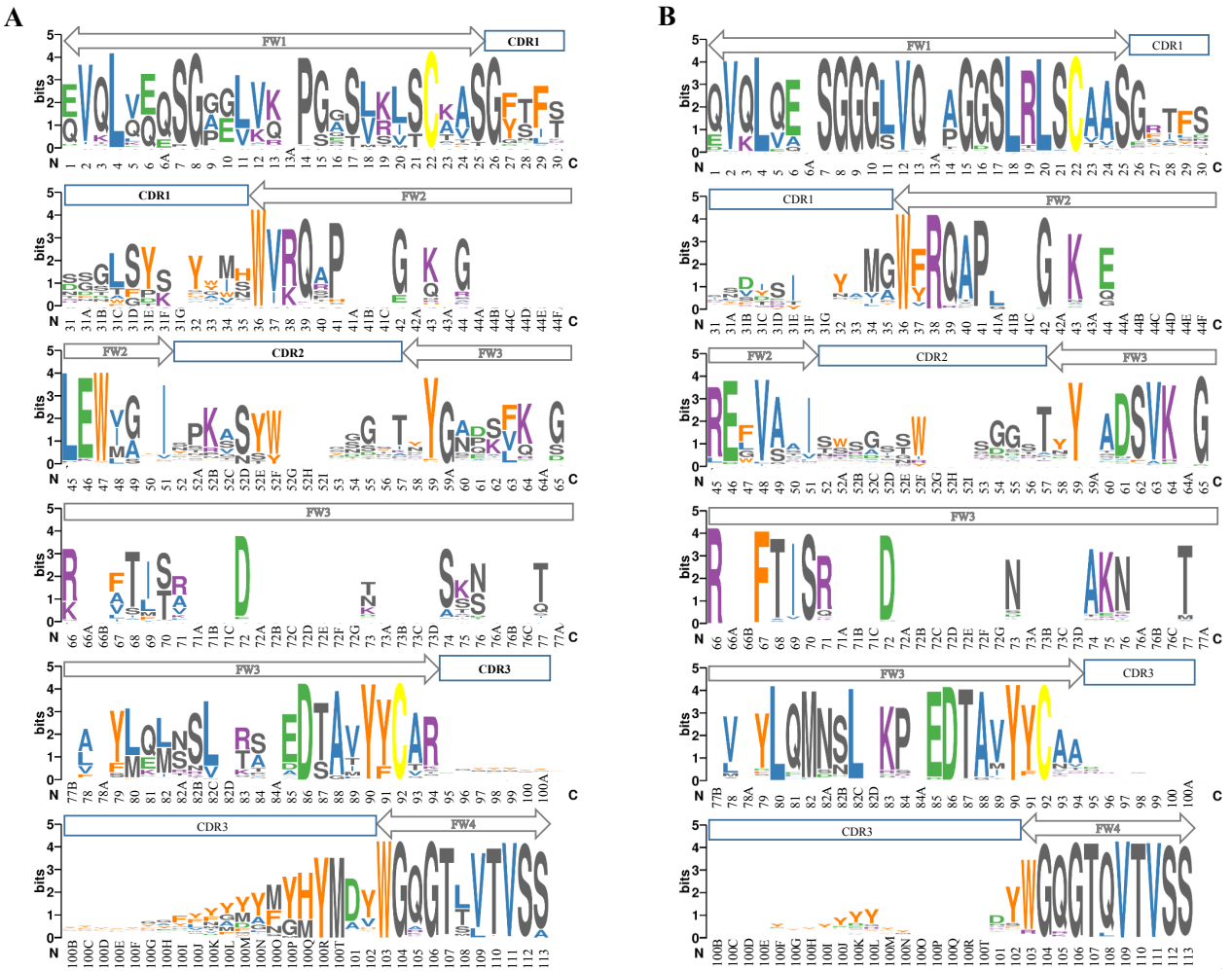


Figure 5. Graphical logo representations of multiple sequence alignment of the complete sequence datasets of (A) VHs and (B) V_HHs. The multiple sequences were aligned based on the Chothia numbering scheme. The size of each logo reflects the relative frequency of amino acids at each position, measured in bits. Color codes were applied based on single-letter-code amino acid biochemical properties; basic: K, R (purple); acidic: D, E (green); aliphatic: A, I, L, V (blue); aromatic: F, H, W, Y (orange); cysteine: C (yellow); mostly small: G, M, N, P, Q, S, T (dark gray). Certain amino acids were categorized in multiple subclasses. The grey and dark blue boxes indicate the borders of the FW and HCDR regions, respectively.

Hydrophobic residues have a natural tendency to avoid contact with solvents and tend to fold inward, contributing to the protein's structural stability. By minimizing the hydrophobic content in the constant regions, V_HHs promote proper folding and overall stability of the antibody structure. Considering the results of the physicochemical properties of FW2 and FW3, the prevalence of hydrophilic residues in V_HHs compared to V_HHs suggests a potential indicator for the higher stability and solubility of V_HHs, even in the absence of a V_L (Mitchell et al., 2018). In addition to FW2, our findings highlight the critical role of these FW3 residues, underscoring their potential impact on the overall structure and function of the antibody. As a prospective avenue, researchers can explore the differences identified in this study, especially in these regions, to guide future investigations.

FW4 represents the relatively shorter region, and it is the most conserved segment within the entire variable domain in terms of sequence. However, there were notable differences in our dataset. V_HHs predominantly recruited aromatic 'Y' or 'W' residues around the 100th position, highlighting a consistent feature in their HCDR3 region. This preference for aromatic residues in V_HHs may play a crucial role in their antigen-binding specificity and affinity. Another intriguing observation within our dataset was the substitution of the 'W' at the 103rd position in the FW4 of

V_HHs. At the 108th position, the 'Q' in V_HHs transitioned into 'L/T/S' in V_Hs.

Moreover, similarly consistent patterns emerged in both the V_H and V_HH datasets once we clustered them into species-specific subsets (Figure S1 and S2). However, we observed conserved residues at the subset levels, which varied marginally from the overall consensus, since some species datasets contained relatively smaller numbers of sequences.

Our observations on the prevalence of differences among all FWs stand out, particularly given the emphasis in previous studies on the FW2 region as the primary locus for engineering the V_H–V_L interface of antibodies to convert them into V_HH format.

3.4. Encounter rate of noncanonical cysteine bonds

The introduction of a noncanonical disulfide bridge in addition to the canonical 22^{nd}C–92^{nd}C bond can further enhance the thermal stability of V_HHs. This additional disulfide bond can increase the thermal denaturation temperature, making the antibody more resistant to unfolding and improving its overall stability while helping to reduce the entropic folding penalty (D'Huyvetter et al., 2017). The presence of such extra bonds remedies the destabilizing effect caused by hallmark residues at FW2 and rigidifies the inherently longer HCDR3 loop by rendering a stronger, more specific interaction with their target (Govaert et al., 2012). These extra bonds could serve as structural reinforcements that contribute to the maintenance of the specific conformation required for effective antigen recognition and binding, particularly in antibodies with extended HCDR3 regions.

To analyze their contributions to antibody stability in different formats, we conducted a comprehensive exploration of the noncanonical disulfide bridges within antibody structures. We performed a dual analysis, encompassing both the examination of amino acid sequences and the inspection of 3D structural conformations. This combined approach allowed us to pinpoint the specific cysteine residues involved in the formation of noncanonical disulfide bridges in V_Hs. By combining the insights gained from sequence and structural analyses, we attempted to unravel the presence and significance of noncanonical disulfide bridges in both V_Hs and V_HHs.

Our results showed that while V_Hs may have a variety of noncanonical cysteine types, V_HHs tended to have a higher frequency of these additional bonds, highlighting a distinctive structural characteristic in V_HHs (Figure 6). The encounter frequency of noncanonical bonds of V_Hs (2.6%) (Table S6) was significantly lower than that of V_HHs (5.5%) (Table S7). Among the observed noncanonical bonds in the V_Hs, four distinct types were identified, connecting various regions. These included bonds between HCDR1 and HCDR2, HCDR1 and HCDR3, FW2 and HCDR2, and HCDR3 and HCDR3 (Figure 6A). Our analysis of the V_HHs unveiled three distinct categories of noncanonical cysteine pairs coexisting alongside the canonical bond. These were formed between HCDR1 and HCDR3, FW2 and HCDR3, and HCDR3 and HCDR3 (Figure 6B). Although the frequency of V_H noncanonical bonds was lower than that of the V_HHs, the variety in their positions was higher than that of the V_HHs. For both formats, HCDR1–HCDR3 and HCDR3–HCDR3 noncanonical disulfide bonds were common. Bonds involving HCDR3s likely enhance the stabilization of HCDR3 loops, thereby contributing significantly to the structural integrity of the antibody's antigen-binding site.

Notably, species-specific extra bonds were observed in both datasets (Tables S6 and S7). *Camelus dromedarius* stood out by demonstrating the highest diversity in terms of extra bonds occurring at various locations compared to other species in the dataset. In this study, 2.4% of *Lama glama* V_HHs were found to have noncanonical extra cysteine bonds (Table S7). *Homo sapiens* exhibited a relatively high frequency of noncanonical bonds, albeit at a rate of 5.3% (Table S6). *Vicugna pacos* and *Camelus dromedarius* exhibited the shortest average V_HH HCDR3 length (Table S4), and these two species also had the highest occurrence of extra cysteine bonds (~10%) in their V_HHs (Table S7).

While some antibodies possessed remarkably long HCDR3 regions, extending up to 20 amino acids, it is essential to emphasize that length alone did not guarantee the presence of additional stabilizing bonds. The formation of disulfide bridges relies on the correct pairing of cysteine residues and their proper alignment in the protein's three-dimensional structure. Failure to form disulfide bonds at these specific positions could be due to structural constraints, steric hindrance, or the absence of suitable cysteine pairing partners. For future prospective studies, researchers should delve deeper into the structural determinants influencing disulfide bridge formation, offering avenues to enhance antibody stability and functionality.

3.5. Stability-enhancing salt bridge profiles of V_Hs and V_HHs around HCDR3

Salt bridges are another mechanism used to regulate the folding and stability of antibodies. The 94th position had a high number of conserved 'R' residues and a positively charged amino acid in V_Hs, compared to the large number of 'A' residues in V_HHs as seen from motif conservation analysis (Figure 5). The presence of a positively charged residue ('R' or 'K') at this position, which serves as an N-flanking residue for HCDR3, enables the antibody to engage in favorable electrostatic interactions with negatively charged regions on the antigen, enhancing binding specificity. This formation not only influences the overall structure but also plays a pivotal role in mediating the interaction between antibodies and antigens (Tsumoto et al., 1996). As such,

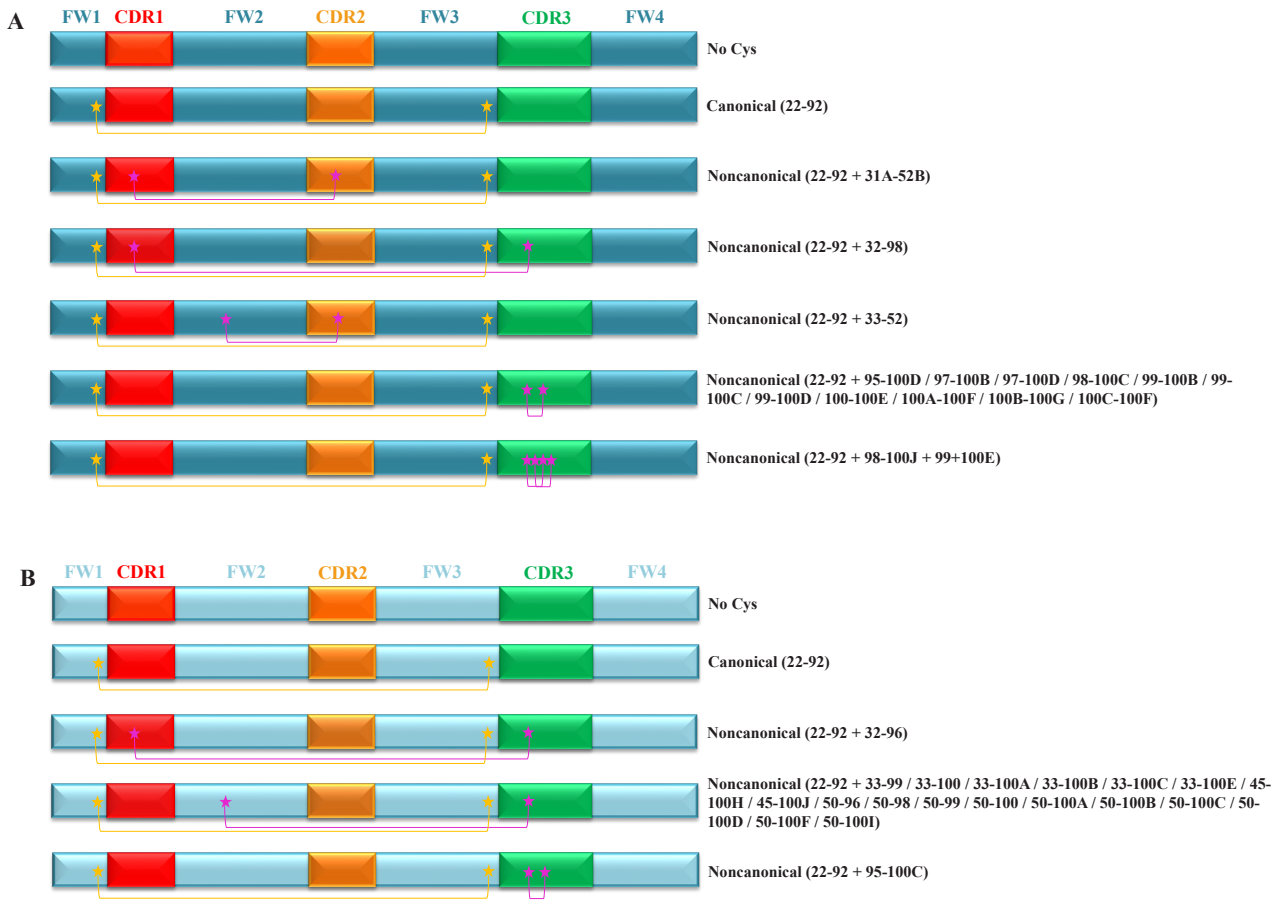


Figure 6. Canonical and noncanonical disulfide bond profiles of (A) VHs and (B) V_HHs. The FW regions (dark blue: VHs, light blue: V_HHs) and HCDRs (HCDR1: red, HCDR2: orange, HCDR3: green) are represented by boxes. Canonical disulfide bonds formed between 22ndC and 92ndC are denoted by yellow stars connected by a yellow line. Noncanonical cysteine bonds are represented with magenta stars joined by a magenta line. The shapes of FWs and HCDRs are not to scale. The positioning of the stars representing canonical and noncanonical cysteine residues serves as a putative visual representation and does not reflect their precise locations within the marked regions.

we performed a comprehensive comparison of the salt bridges in V_Hs and V_HHs. The presence of a salt bridge was a prevalent characteristic in V_Hs, typically observed with an 'R' residue at the 94th position and a 'D' residue at the 101st position (hereafter referred to as the canonical salt bridge, while other combinations of 'R/H/K' at the 94th position with 'D/E' at the 101st position, excluding the canonical pair, are considered noncanonical salt bridges).

There are several rule-determination studies suggesting the existence of a salt bridge between the 94th R and 101st D, yet they all used a narrow number of samples in their datasets, which is insufficient to claim that these particular residues are conserved at those positions. They specifically reported that the presence of a negatively charged 101st D residue consistently forms a salt bridge with the basic 94th R residue (Shirai et al., 1999; Weitzner et al., 2015). So far, these published studies have strictly focused on the 94th R–101st D interaction; however, a salt bridge in proteins is formed when two residues with opposite charges come together. In this context, the presence of a salt bridge is determined by the proximity of certain atoms. If the carboxylate of the 'D/E' side-chain is within 4.0 Å of an ammonium of an 'R/K/H' side-chain, these residues form a salt bridge (Kumar and Nussinov, 2002). Additionally, D/E at the 101st position complements this interaction by forming hydrogen bonds or salt bridges with antigen surface residues, where they significantly contribute to the ability of the antibody to recognize and tightly bind to its target, influencing its specificity and affinity (North et al., 2011).

Intriguingly, within our dataset, various combinations of acidic and basic residues existed beyond the 94th R–101st D pairing, as well as instances where no salt bridge-forming residues were present at all, which contradicts the previously reported conserved pattern of the 94th R–101st D salt bridge (Tables S8 and S9). An interesting contrast also emerged in V_HHs, as most of them lacked this canonical salt bridge interaction (Table S9). While 63.1% of V_Hs had one of these salt bridges connecting the residues at the 94th and 101st positions (Table S8), V_HHs only had a 2.7% incidence (Table S9). Specifically, in 56% of V_Hs, the 'R-D' pairing was observed, which results in the strongest interaction, leading to the shortest interatomic distance compared to other combinations of salt bridge-forming residues (Musafia et al., 1995). Therefore, the residues at the 94th and 101st positions are likely to be of importance in humanization/camelization studies.

A distinctive conformation is achieved through the interaction between the basic 94th residue's hydrophobic part and the HCDR1, along with a hydrogen bond linking the conserved 103rd 'W' to the preceding residue at the 101st position. When a negatively charged residue is at the 101st position without a corresponding positively charged residue at the 94th position, the HCDR3 torso takes on a nonbulged form. In this scenario, the acidic side chain, lacking a salt bridge partner, forms a hydrogen bond with the side chain of the conserved 103rd 'W,' creating a notable kink in the structure (Morea et al., 1997). The absence of a charged amino acid at the 94th position of camel V_HHs was reported before (Muyldermans et al., 1994). When no salt bridge was observed between the 94th and 101st positions, due to the absence of the 101st residue, the presence of the 103rd 'W' residue was not always consistent to confer a bulged kink for favorable additional interactions. Overall, 37% of the V_Hs did not harbor any salt bridge at this position, while this value reached up to 97% in V_HHs. This highlights the structural diversity and adaptability of different antibody structures. In light of the diverse structural manifestations observed, particularly regarding the intricate interplay between residues at the 94th, 101st, and 103rd positions, future prospective studies should delve into the functional implications arising from variations in these key structural elements.

3.6. Statistical analysis of physicochemical properties of FW residues of V_Hs and V_HHs

Next, our analysis extended to the distribution of physicochemical amino acid categories within the FW regions of V_Hs and V_HHs. The primary objective was to identify specific physicochemical characteristics that may contribute to the distinctions between these two antibody classes, potentially influencing the folding, stability, or affinity of their functional structures. To accomplish this, we employed an amino acid classification system that encompassed a total of nine distinct categories, as outlined in Table S2, based on EMBOSS Pepstats. The percentage of residues in each sequence for each category was calculated using the EMBOSS Pepstats online tool (Supplementary File 2). It is worth noting that certain amino acids were categorized in multiple subclasses due to their overlapping physicochemical characteristics. As part of our analysis, in addition to the physicochemical properties, we also evaluated the molecular charge and hydropathy (Table S3) properties of the V_Hs and V_HHs.

We employed both a logistic regression model and a machine learning approach to evaluate and analyze these comparisons. The logistic regression modeling process was performed for each FW region (FW1–4) collectively (Table; Table S10) and individually (Table S11). For this analysis, we specifically focused on the FW regions, as the hypervariability of CDRs created many confounding factors, making it difficult to ensure that the results were due to the differences between the antibody formats. When we applied our chosen statistical model with all physicochemical properties, the contribution of the variables could potentially affect the overall results, leading to otherwise significant findings being ruled out due to multicollinearity. Furthermore, the outcomes of the analysis were overly intricate, making it challenging to grasp the implications.

The collective statistical analysis provided valuable insights into the outcomes of the logistic regression, primarily focusing on the Beta and Exp(Beta) values (Table). Exp(Beta) values higher than 1 indicate that the physicochemical characteristics were overrepresented in V_HHs (e.g., FW1-tiny and aromatic), while lower values indicate the reverse (e.g., FW1-small and basic). The most significant differences occurred in tiny/small and/or hydrophobic/hydrophilic contents in almost all FW regions. The statistical significance of the collective model is further detailed in Table S10.

We also conducted logistic regression analysis of individual FWs, with the outputs being evaluated based on the previously mentioned scoring criteria (Table S11). Most of the results confirmed the whole V_H sequence model. Upon performing the individual analysis, distinct characteristics were observed for each FW region. In addition to the previous findings, FW1 residues were more hydrophilic. FW3 displayed specifically hydrophilic residues compared to the nonhydrophobic residues identified in the V_H-wide analysis. Notably, FW4 was characterized as hydrophilic and was found to be the shortest among all other framework regions. It also had a very low constant value, leading to lower variable importance (VI). This feature in the FW4 structure underscored its specific role, emphasizing the significance of individual analysis in capturing nuanced details that might be overlooked in collective assessments.

Table. Logistic regression analysis results based on EMBOSS Pepstats output incorporating all FW1–4 regions, along with the various physicochemical properties.

		Logistic Regression			Confidence interval	
FWs	Physicochemical property	Beta	Exp(Beta)	p-values	Lower bound	Upper bound
	Constant	–47.781	0.000	$1.014.e^{-04}$		
FW1	Tiny	0.517	1.678	$1.823.e^{-08}$	1.4205	2.1739
	Small	–0.312	0.732	$9.455.e^{-04}$	0.5942	0.8749
	Aliphatic	0.206	1.228	$2.241.e^{-02}$	0.9629	1.4402
	Aromatic	0.745	2.107	$1.002.e^{-02}$	1.2448	4.0962
	Basic	–0.366	0.693	$1.569.e^{-05}$	0.5717	0.8254
	Acidic	0.162	1.176	$4.168.e^{-02}$	1.0103	1.4292
	Polar	0.146	1.157	$1.956.e^{-02}$	1.0096	1.2873
FW2	Tiny	0.635	1.887	$1.358.e^{-13}$	1.5956	2.4403
	Aromatic	0.247	1.279	$4.606.e^{-03}$	1.0703	1.5909
	Aliphatic	–0.328	0.720	$2.898.e^{-04}$	0.5298	0.8343
	Polar	–0.150	0.861	$1.939.e^{-02}$	0.7367	0.9975
	Basic	0.330	1.392	$1.279.e^{-04}$	1.1885	1.8180
	Hydrophobic	0.328	1.388	$4.172.e^{-06}$	1.2448	1.8178
	Hydrophilic	0.647	1.909	$1.676.e^{-10}$	1.5592	2.4595
FW3	Small	0.578	1.782	$8.100.e^{-10}$	1.5255	2.7124
	Aromatic	–0.360	0.698	$2.448.e^{-02}$	0.4978	0.9796
	Polar	–0.311	0.733	$8.081.e^{-03}$	0.5951	0.9799
	Basic	–0.342	0.710	$8.842.e^{-03}$	0.5525	0.9388
	Hydrophobic	–0.265	0.767	$3.919.e^{-03}$	0.5376	0.9429
FW4	Tiny	–0.153	0.858	$6.413.e^{-04}$	0.7049	0.9999
	Aliphatic	–0.170	0.844	$1.539.e^{-04}$	0.6246	0.9680
	Aromatic	–0.577	0.562	$2.182.e^{-09}$	0.4168	0.6628
	Hydrophilic	0.867	2.380	$7.299.e^{-20}$	1.8508	3.069

To further validate the accuracy of our feature selection, we also used the Boruta Algorithm-Random Forest. The Boruta algorithm, as a wrapper around the random forest algorithm, played a pivotal role in highlighting the physicochemical variables with the highest VI scores. Combining the Boruta algorithm with logistic regression modeling yielded compelling results in our study. This dual approach involved the integration of a statistical method and a machine learning technique, providing a comprehensive and reliable perspective on the analyzed data, as well as validating our findings through established statistical techniques, ensuring the credibility and accuracy of our study outcomes. We ranked the VI scores for each modeled feature (Figure S3) and the frequency of amino acids in each region (HCDRs and FWs) (Figure S4). The results indicated the accuracy of the logistic regression analysis model, as features with very high and very low Exp(Beta) scores, such as hydrophilic FW4 residues and tiny FW1 residues, were ranked highly compared to other features. It is important to note that while the hydrophilic residues of FW2 are well known, other features, such as the tiny residues in FW1 and the small residues of FW3, were ranked higher in comparison, both supported by our motif conservation analysis. This highlights the importance of FW1 and FW3, which might contain elements of higher impact for the structure of the V_H domain than the hydrophilic FW2 residues. When examining the individual amino acid contributions in different regions, we noted that the residues found among the hallmark residues of FW2, especially 'FW2_Leu' and 'FW2_Phe,' were ranked near the top of all amino acids, both of which contribute to the more hydrophilic interface of V_HHs. Similarly, the high 'G' content of FW1 likely contributes to the importance of tiny residues in the FW1 of V_HHs. Finally, the highly conserved

'S' residues in FW3 in V_HHs are often substituted for other residues in V_Hs, which may explain its high ranking.

FW4 was found to be the shortest among all other framework regions and it had a very low constant value, leading to lower VI scores. This might indicate that FW4 is not of high importance for humanization/camelization approaches. Alternatively, the short length of FW4 might mask the importance of individual features when comparing V_Hs and V_HHs, due to the small range of each feature.

3.7. Physicochemical characteristics of individual residues in V_Hs and V_HHs

Finally, we examined the physicochemical properties of V_Hs and V_HHs on the residue level (Figure S5). The results showed that several residues have higher variance in certain physicochemical characteristics between V_Hs and V_HHs compared to the overall sequence. In FW1, residues with similar physicochemical characteristic differences included the 6th and 6Ath positions, as well as the 10th and 23rd positions. In addition to the hallmark residues, where each residue displayed the appropriate physiochemical characteristics, the 35th, 40th, and 50th positions of FW2 also showed high differences, with V_Hs favoring aromatic and hydrophobic residues while V_HHs showed a higher degree of smaller and nonpolar residues at the same positions. FW3 showed similar residue traits in the 60th, 62nd, and 94th positions. Interestingly, FW4 was entirely devoid of such residues, with the 108th residue instead favoring small, nonpolar, hydrophilic residues in V_Hs and hydrophobic residues in V_HHs.

One feature of interest among the physicochemical characteristics was the distribution of hydrophobic and hydrophilic amino acids among the FWs of V_Hs and V_HHs, respectively. While the absence of the V_H–V_L promotes a higher degree of hydrophilic residues in V_HHs, these residues assist in the proper folding and packing of the domain into its appropriate shape. Other than the 27th position, which strongly favored aromatic, nonpolar residues and slightly favored hydrophilic residues in V_Hs, very few nonalternating residues had strong differences between V_Hs and V_HHs. This confirmed our previous statement that the major differences in HCDR regions were caused by target recognition rather than species of origin. However, there was a weak preference across multiple residues for less polar and more hydrophilic residues in HCDR3s compared to those of V_HHs. The absence of such differences in HCDR3 otherwise further points to its importance in epitope recognition.

This study has revealed significant distinctions between antibodies and nanobodies, emphasizing the longer and less variable HCDR3s, specific binding-enhancing residues (at the 50th and 94th positions in antibodies and the 29th and 51st positions in nanobodies), and distinct motif conservation patterns in the latter. Analyzing the antigen binding scores allows for the identification of crucial residue locations on CDRs, providing valuable insights for studies on CDR grafting. Nanobodies exhibit a higher prevalence of species-specific noncanonical disulfide bonds. Additionally, a prominent salt bridge concentration was highlighted between the 94th and 101st positions in the antibodies, which is missing in the nanobodies. Unique physicochemical characteristics of nanobodies' FWs were revealed compared to IgGs: the residues in FW1 were predominantly tiny, hydrophilic, and noncharged; those in FW2 featured a smaller size and higher hydrophilicity; the FW3 region's residues were small, nonpolar, and nonhydrophobic; and FW4 residues exhibited small, nonaromatic, and polar traits. The logistic regression-based model employed feature selection based on p-value scores and information criteria. In contrast, the Boruta algorithm, a decision tree-based method, randomly explores features using Monte Carlo techniques (Sanchez-Pinto et al., 2018). Logistic regression follows a minimal optimal approach, relying on a small subset of features that minimize error on a chosen classifier. On the other hand, the Boruta algorithm, by virtue of its random exploration, can discover solutions closer to the global optimum. This characteristic provides the Boruta algorithm with a significant advantage in effectively handling interactions between physicochemical variables. The insights of this study are pivotal for converting conventional antibodies into single-domain V_H formats and providing valuable considerations for camelization processes. These findings contribute to the advancement of antibody engineering strategies, drawing inspiration from natural evolution.

It is important to consider potential limitations and sources of error in the data analysis of this study, such as partial missing sequence information. In some cases, the sample sequences may have unresolved residues due to limitations in experimental techniques like crystallography. These missing residues were more commonly found within the constant regions of the antibodies, which are generally well conserved in both sequence and structure. The total percentage of nonresolved residues was within acceptable levels and did not have any major impact on our results. Additionally, the accuracy of structural predictions for noncanonical bonds may introduce a margin of error.

This study has enhanced our knowledge of antibody engineering and provides a foundation for the rational design and development of novel antibody formats with improved characteristics and functionalities. By studying the physicochemical characteristics and structural adaptations of V_HHs compared to conventional V_Hs, we gain insights into how nature has adapted antibody formats to optimize their function in different contexts. In particular, our findings contribute to the understanding of engineering approaches for the efficient conversion of full-length IgG

into smaller V_H format through camelization and assist in the mitigation of immunogenicity concerns, inspired by nature's ingenuity.

Acknowledgement/Disclaimers/Conflict of interest
We thank Alirıza Arıbaş for his contributions to the computational analysis and sequence retrieval.

Informed consent
N/A.

References

Abanades B, Wong WK, Boyles F, Georges G, Bujotzek A et al. (2023). ImmuneBuilder: Deep-learning models for predicting the structures of immune proteins. Communications Biology 6: 575. https://doi.org/10.1038/s42003-023-04927-7

Bélanger K, Iqbal U, Tanha J, MacKenzie R, Moreno M et al. (2019). Single-domain antibodies as therapeutic and imaging agents for the treatment of CNS diseases. Antibodies 8 (2): 27-52. https://doi.org/10.3390/antib8020027

Bond CJ, Marsters JC, Sidhu SS (2003). Contributions of CDR3 to VHH domain stability and the design of monobody scaffolds for naive antibody libraries. Journal of Molecular Biology 332 (3): 643-655. https://doi.org/10.1016/S0022-2836(03)00967-7

Conrath K, Vincke C, Stijlemans B, Schymkowitz J, Decanniere K et al. (2005). Antigen binding and solubility effects upon the veneering of a camel VHH in framework-2 to mimic a VH. Journal of Molecular Biology 350 (1): 112-125. https://doi.org/10.1016/j.jmb.2005.04.050

Conrath KE, Lauwereys M, Galleni M, Matagne A, Frère JM et al. (2001). Beta-lactamase inhibitors derived from single-domain antibody fragments elicited in the Camelidae. Antimicrobial Agents and Chemotherapy 45 (10): 2807-2812. https://doi.org/10.1128/AAC.45.10.2807-2812.2001

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004). Weblogo: A sequence logo generator. Genome Research 14: 1188-1190. https://doi.org/10.1101/gr.849004

Deac A, Veličković P, Sormanni P (2019). Attentive cross-modal paratope prediction. Journal of Computational Biology 26 (6): 536-545. https://doi.org/10.1089/cmb.2018.0175

De Genst E, Saerens D, Muyldermans S, Conrath K (2006). Antibody repertoire development in camelids. Developmental and Comparative Immunology 30 (1-2): 187-198. https://doi.org/10.1016/j.dci.2005.06.010

D'Huyvetter M, De Vos J, Xavier C, Pruszynski M, Sterckx YGJ et al. (2017). 131I-labeled anti-HER2 camelid sdAb as a theranostic tool in cancer treatment. Clinical Cancer Research 23 (21): 6616-6628. https://doi.org/10.1158/1078-0432.CCR-17-0310

Dunbar J, Krawczyk K, Leem J, Marks C, Nowak J et al. (2016). SAbPred: A structure-based antibody prediction server. Nucleic Acids Research 44 (W1): W474-W478. https://doi.org/10.1093/nar/gkw361

Flajnik MF, Deschacht N, Muyldermans S (2011). A case of convergence: Why did a simple alternative to canonical antibodies arise in sharks and camels? PLoS Biology 9 (8): e1001120. https://doi.org/10.1371/journal.pbio.1001120

Goldman ER, Liu JL, Zabetakis D, Anderson GP (2017). Enhancing stability of camelid and shark single domain antibodies: an overview. Frontiers in Immunology 8: 865. https://doi.org/10.3389/fimmu.2017.00865

Govaert J, Pellis M, Deschacht N, Vincke C, Conrath K et al. (2012). Dual beneficial effect of interloop disulfide bond for single domain antibody fragments. Journal of Biological Chemistry 287 (3): 1970-1979. https://doi.org/10.1074/jbc.M111.242818

Hamers Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hammers C et al. (1993). Naturally occurring antibodies devoid of light chains. Nature 363: 446-448. https://doi.org/10.1038/363446a0

Kumar S, Nussinov R (2002). Relationship between ion pair geometries and electrostatic strengths in proteins. Biophysical Journal 83 (3): 1595-1612. https://doi.org/10.1016/S0006-3495(02)73929-5

Liberis E, Velickovic P, Sormanni P, Vendruscolo M, Liò P (2018). Parapred: Antibody paratope prediction using convolutional and recurrent neural networks. Bioinformatics 34 (17): 2944-2950. https://doi.org/10.1093/bioinformatics/bty305

Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J et al. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Research 50 (W1): W276-W279. https://doi.org/10.1093/nar/gkac240

Mendoza MN, Jian M, King MT, Brooks CL (2020). Role of a noncanonical disulfide bond in the stability, affinity, and flexibility of a VHH specific for the *Listeria* virulence factor InlB. Protein Science 29 (4): 1004-1017. https://doi.org/10.1002/pro.3831

Mitchell LS, Colwell LJ (2018). Analysis of nanobody paratopes reveals greater diversity than classical antibodies. Protein Engineering, Design and Selection 31 (7-8): 267-275. https://doi.org/10.1093/protein/gzy017

Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM (1997). Antibody structure, prediction and redesign. Biophysical Chemistry 68 (1-3): 9-16. https://doi.org/10.1016/S0301-4622(96)02266-1

Musafia B, Buchner V, Arad D (1995). Complex salt bridges in proteins: Statistical analysis of structure and function. Journal of Molecular Biology 254 (4): 761-770. https://doi.org/10.1006/jmbi.1995.0653

Muyldermans S (2013). Nanobodies: Natural single-domain antibodies. Annual Review of Biochemistry 82 (1): 775-797. https://doi.org/10.1146/annurev-biochem-063011-092449

Muyldermans S (2021). Applications of nanobodies. Annual Review of Animal Biosciences 9 (1): 401-421. https://doi.org/10.1146/annurev-animal-021419-083831

Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R (1994). Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. Protein Engineering, Design and Selection 7 (9): 1129-1135. https://doi.org/10.1093/protein/7.9.1129

Nguyen VK, Su C, Muyldermans S, van der Loo W (2002). Heavy-chain antibodies in Camelidae; a case of evolutionary innovation. Immunogenetics 54: 39-47. https://doi.org/10.1007/s00251-002-0433-0

North B, Lehmann A, Dunbrack RL Jr (2011). A new clustering of antibody CDR loop conformations. Journal of Molecular Biology 406 (2): 228-256. https://doi.org/10.1016/j.jmb.2010.10.030

Pillay TS, Muyldermans S (2021). Application of single-domain antibodies ("nanobodies") to laboratory diagnosis. Annals of Laboratory Medicine 41 (6): 549-558. https://doi.org/10.3343/alm.2021.41.6.549

Riechmann L, Muyldermans S (1999). Single domain antibodies: comparison of camel VH and camelised human VH domains. Journal of Immunological Methods 231 (1-2): 25-38. https://doi.org/10.1016/S0022-1759(99)00138-6

Sanchez Pinto LN, Venable LR, Fahrenbach J, Churpek MM (2018). Comparison of variable selection methods for clinical predictive modeling. International Journal of Medical Informatics 116: 10-17. https://doi.org/10.1016/j.ijmedinf.2018.05.006

Shirai H, Kidera A, Nakamura H (1999). H3-rules: Identification of CDR-H3 structures in antibodies. FEBS Letters 455 (1-2): 188-197. https://doi.org/10.1016/S0014-5793(99)00821-2

Swindells MB, Porter CT, Couch M, Hurst J, Abhinandan KR et al. (2017). abYsis: Integrated antibody sequence and structure-management, analysis, and prediction. Journal of Molecular Biology 429 (3): 356-364. https://doi.org/10.1016/j.jmb.2016.08.019

Tsumoto K, Ogasahara K, Ueda Y, Watanabe K, Yutani K et al. (1996). Role of salt bridge formation in antigen-antibody interaction. Entropic contribution to the complex between hen egg white lysozyme and its monoclonal antibody HyHEL10. Journal of Biological Chemistry 271 (51): 32612-32616. https://doi.org/10.1074/jbc.271.51.32612

Vidarsson G, Dekkers G, Rispens T (2014). IgG subclasses and allotypes: from structure to effector functions. Frontiers in Immunology 5: 520. https://doi.org/10.3389/fimmu.2014.00520

Vu KB, Ghahroudi MA, Wyns L, Muyldermans S (1997). Comparison of llama VH sequences from conventional and heavy chain antibodies. Molecular Immunology 34 (16-17): 1121-1131. https://doi.org/10.1016/S0161-5890(97)00146-6

Weitzner BD, Dunbrack RL Jr, Gray JJ (2015). The origin of CDR H3 structural diversity. Structure 23 (2): 302-311. https://doi.org/10.1016/j.str.2014.11.010

Wilton EE, Opyr MP, Kailasam S, Kothe RF, Wieden HJ (2018). sdAb-DB: The single domain antibody database. ACS Synthetic Biology 7 (11): 2480-2484. https://doi.org/10.1021/acssynbio.8b00407

Supplementary Files

1. Supplementary File 1: Statistical Methods

1.1. Binary logistic regression model

As the strategy for selecting important physicochemical predictors, stepwise regression backward elimination with the likelihood ratio test was used.

The binary logistic regression model is shown in Eq. (1).

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + \cdots + \beta_k.X_k \quad (1)$$

Here, p stands for the probability of the presence of V_HHs.

The binary logistic regression model relied on the assumption of independence (multicollinearity) among predictors, along with linearity. To control the multicollinearity for each predictor in the model, the variance inflation factor (VIF) was calculated. Removing predictors with high VIF values (>10) helped to reduce the multicollinearity and improve the accuracy of the binary logistic regression model. This model, refined through stepwise elimination, provides a succinct representation of the influential factors in the classification process, thereby offering a robust foundation for understanding the underlying physicochemical determinants governing the distinctions between VHs and V_HHs. By implementing this VIF-based predictor selection strategy, we ensured a more robust and reliable assessment of the relationship between VHs and V_HHs.

1.2. Boruta Algorithm

The Boruta algorithm is a wrapper built around the Random Forest (RF) classification algorithm. It tries to capture all the important, interesting features in a dataset with respect to an outcome variable. The Boruta algorithm was fitted jointly to data for amino acids for all regions of VHs and V_HHs. The responses were ordered from left to right by their importance for classification. The importance measure was defined as the Z-score of the mean decrease in accuracy (normalized permutation importance). Operating as an ensemble learning method, the Boruta algorithm was used to capture all relevant and impactful features with respect to the outcome variable, which, in this case, pertained to the classification of VHs and V_HHs.

1.3. Statistical significance calculations

McFadden's R^2 or McFadden's pseudo-R^2 is a measure of the goodness of fit of the binary logistic regression model. It measures how well the model predicts the outcome. The area under the curve (AUC) is an overall summary of the diagnostic accuracy of the binary logistic regression model. If the AUC is greater than 0.5, the model performs well at classifying antibody sequences as VHs or V_HHs.

McFadden's R^2 assesses the effectiveness of the model in predicting the outcome, offering insights into the proportion of variance explained by the predictors.

The AUC is particularly useful in assessing the model's ability to correctly classify antibody sequences as VHs and V_HHs.

2. Supplementary File 2: Tables and Figures

Supplementary data available at: https://data.mendeley.com/datasets/4g5njh2ct2/1

Table S1. Definition of constant and variable regions in VH and V_HH sequences based on the Chothia numbering scheme (FW1–4: framework regions, HCDR1–3: complementarity determining regions).

VH/ V_HH region	Chothia residue
FW1	1–25
HCDR1	26–32
FW2	33–51
HCDR2	52–56
FW3	57–94
HCDR3	95–102
FW4	103–113

Table S2. Classification of physicochemical properties shown as single-letter amino acid codes defined by EMBOSS Pepstats.

Property	Amino acid residues
Tiny	A+C+G+S+T
Small	A+B+C+D+G+N+P+S+T+V
Aliphatic	A+I+L+V
Aromatic	F+H+W+Y
Non-polar	A+C+F+G+I+L+M+P+V+W+Y
Polar	D+E+H+K+N+Q+R+S+T+Z
Charged	B+D+E+H+K+R+Z
Basic	H+K+R
Acidic	B+D+E+Z

Table S3. Amino acid hydropathy classes in single-letter amino acid codes.

Property	Amino acid residues
Hydrophilic	D+E+K+N+R+Q
Neutral	G+H+P+T+S+Y
Hydrophobic	A+C+F+I+L+M+V+W

Table S4. Average length of HCDRs in VHs categorized by species, based on our dataset.

Species	HCDR1 length	HCDR2 length	HCDR3 length
Cricetulus migratorius	7	6	7.9
Homo sapiens	7	6.4	10.5
Homo sapiens/Mus musculus	7.4	5.8	9.9
Mus musculus	7.1	6	9.8
Mus musculus/Homo sapiens	7.1	5.7	10.1
Pan troglodytes	5	7	15
Rattus norvegicus	7	5.7	8.2
Rattus norvegicus/Homo sapiens	7	6	8
Average VHs length	7.2	5.9	11.9

Table S5. Average length of HCDRs in V_HHs categorized by species, based on our dataset.

Species	HCDR1 length	HCDR2 length	HCDR3 length
Camelidae	7	5.8	14.5
Camelus bactrianus	7.8	6.1	15.6
Camelus dromedarius	6.9	5.7	13.2
Lama glama	6.9	5.8	13.9
Vicugna pacos	7.1	5.8	12.5
Average V_HHs length	6.9	5.8	13.5

Table S6. Total number of canonical and noncanonical disulfide bridges in VHs for each species from our dataset.

Species	No bond	Canonical bonds	Noncanonical bonds	Total VHs
Cricetulus migratorius	-	10	-	10
Homo sapiens	1	2456	130	2457
Homo sapiens/Mus musculus	-	26	-	26
Mus musculus	3	2287	2	2290
Mus musculus/Homo sapiens	-	188	-	188
Pan troglodytes	-	4	-	4
Rattus norvegicus	-	53	-	53
Rattus norvegicus/Homo sapiens	-	2	-	2
Total	4	5026	132	5030

Table S7. Total number of canonical and noncanonical disulfide bridges in V_HHs for each species from our dataset.

Species	No bond	Canonical bonds	Noncanonical bonds	Total V_HHs
Camelidae	-	65	3	65
Camelus bactrianus	-	63	4	63
Camelus dromedarius	-	263	26	263
Lama glama	1	669	16	670
Vicugna pacos	-	199	20	199
Total	1	1259	69	1260

Table S8. Total number of canonical (R94–D101) and noncanonical salt bridges (combinations of R/H/K 94–D/E101 excluding canonical ones) in VHs on species basis.

Species	Canonical salt bridges	Noncanonical salt bridges	No salt bridge	Total VHs
Cricetulus migratorius	5	-	5	10
Homo sapiens	1409	282	766	2457
Homo sapiens/Mus musculus	10	-	16	26
Mus musculus	1291	44	955	2290
Mus musculus/Homo sapiens	80	5	103	188
Pan troglodytes	4	-	-	4
Rattus norvegicus	32	10	11	53
Rattus norvegicus/Homo sapiens	-	2	-	2
Total	2831	343	1856	5030

Table S9. Total number of canonical (R94–D101) and noncanonical salt bridges (combinations of R/H/K 94–D/E101 excluding canonical ones) in V_HHs on species basis.

Species	Canonical salt bridges	Noncanonical salt bridges	No salt bridge	Total V_HHs
Camelidae	-	-	65	65
Camelus bactrianus	-	1	62	63
Camelus dromedarius	2	-	261	263
Lama glama	12	13	645	670
Vicugna pacos	5	2	192	199
Total	19	16	1225	1260

Table S10. Logistic regression analysis results incorporating all four FW regions, FW1–4, collectively along with the various physicochemical properties. *: The variables not shown in the table were statistically insignificant based on their VIF values.

FWs	Physicochemical property	Variable importance	VIF*	McFadden R^2	AUC
FW1	Tiny	5.628	3.453	0.9730	0.999
	Small	3.306	2.891		
	Aliphatic	2.283	5.979		
	Aromatic	2.575	1.315		
	Basic	4.319	3.612		
	Acidic	2.037	2.544		
	Polar	2.336	6.038		
FW2	Tiny	7.400	6.938		
	Aromatic	2.833	3.587		
	Aliphatic	3.624	5.596		
	Polar	2.338	5.049		
	Basic	3.830	2.382		
	Hydrophobic	4.603	7.424		
	Hydrophilic	6.388	8.388		
FW3	Small	6.143	3.103		
	Aromatic	2.249	2.578		
	Polar	2.649	5.999		
	Basic	2.618	1.724		
	Hydrophobic	2.885	6.332		
FW4	Tiny	3.414	1.880		
	Aliphatic	3.785	1.823		
	Aromatic	5.984	3.002		
	Hydrophilic	9.123	4.596		

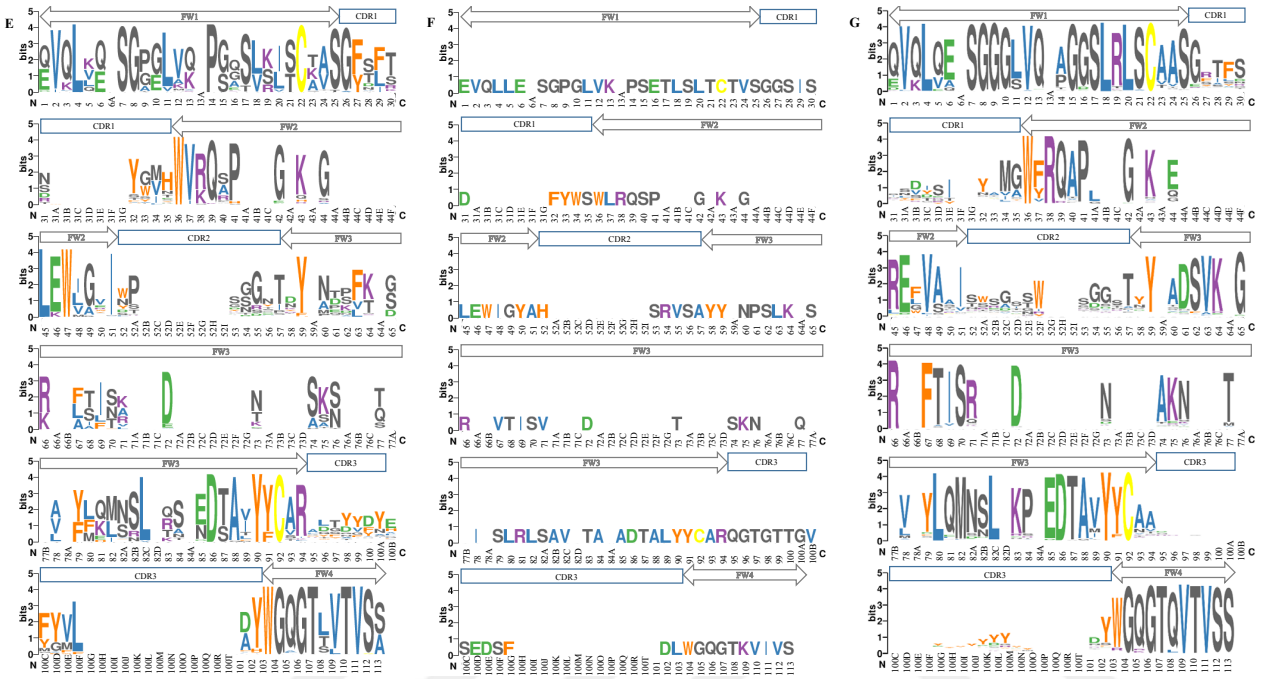
Table S11. Logistic regression analysis incorporating all four FW regions, FW1–4, individually along with various physicochemical properties.

FWs	Physicochemical property	Beta	Exp(Beta)	Variable importance	VIF	McFadden R^2	AUC
FW1	Constant	–39.139	0.0000	-	-	0.688	0.97
	Tiny	0.725	2.065	24.964	3.088		
	Aliphatic	–0.065	0.937	1.968	4.237		
	Aromatic	0.643	1.902	11.503	1.423		
	Polar	–0.257	0.773	10.533	4.069		
	Basic	–0.533	0.587	18.738	2.016		
	Acidic	–0.331	0.718	14.686	1.131		
	Hydrophobic	0.059	1.062	2.639	3.806		
	Hydrophilic	0.806	2.238	19.970	6.629		
FW2	Constant	–39.591	0.00	-	-	0.677	0.9539
	Tiny	0.501	1.65	26.48	4.29		
	Aliphatic	–0.139	0.87	7.75	3.65		
	Aromatic	0.202	1.22	9.43	2.34		
	Polar	–0.123	0.89	10.15	2.26		
	Basic	0.182	1.19	10.73	1.44		
	Hydrophobic	0.207	1.23	16.43	3.48		
	Hydrophilic	0.649	1.91	27.39	5.25		
FW3	Constant	15.233	4126747	-	-	0.535	0.9432
	Tiny	0.295	1.343	13.706	5.26		
	Small	0.552	1.736	26.822	2.75		
	Aliphatic	–0.083	0.920	3.307	3.48		
	Aromatic	–0.082	0.921	2.782	1.75		
	Polar	–0.699	0.497	27.249	3.65		
	Hydrophobic	–0.225	0.798	11.716	4.12		
	Hydrophilic	0.318	1.374	12.437	4.21		
FW4	Constant	–1.120	0.326	-	-	0.799	0.9711
	Tiny	–0.093	0.911	8.729	1.090		
	Small	0.129	1.138	5.046	1.861		
	Aromatic	–0.119	0.888	2.582	1.172		
	Acidic	–0.553	0.575	4.116	1.029		
	Hydrophobic	–0.110	0.895	8.000	1.887		
	Hydrophilic	0.656	1.927	32.363	1.097		



Figure S1. Graphical logo representation of multiple sequences of the complete VH dataset based on species. A) *Cricetulus migratorius*, B) *Homo sapiens*, C) *Homo sapiens*/*Mus musculus*, D) *Mus musculus*, E) *Mus musculus*/*Homo sapiens*, F) *Pan troglodytes*, and G) *Rattus norvegicus*. Multiple sequences were aligned based on the Chothia numbering scheme. The relative frequency of each amino acid position is quantified by logo size in terms of bits. Color codes are based on amino acid biochemical properties: basic: K, R (purple); acidic: D, E (green); aliphatic: A, I, L, V (blue); aromatic: F, H, W, Y (orange); Cys bonds: C (yellow); mostly small: G, M, N, P, Q, S, T (dark gray). Gray boxes show constant FW regions and dark blue boxes show variable HCDR zones.

Figure S1. Continued



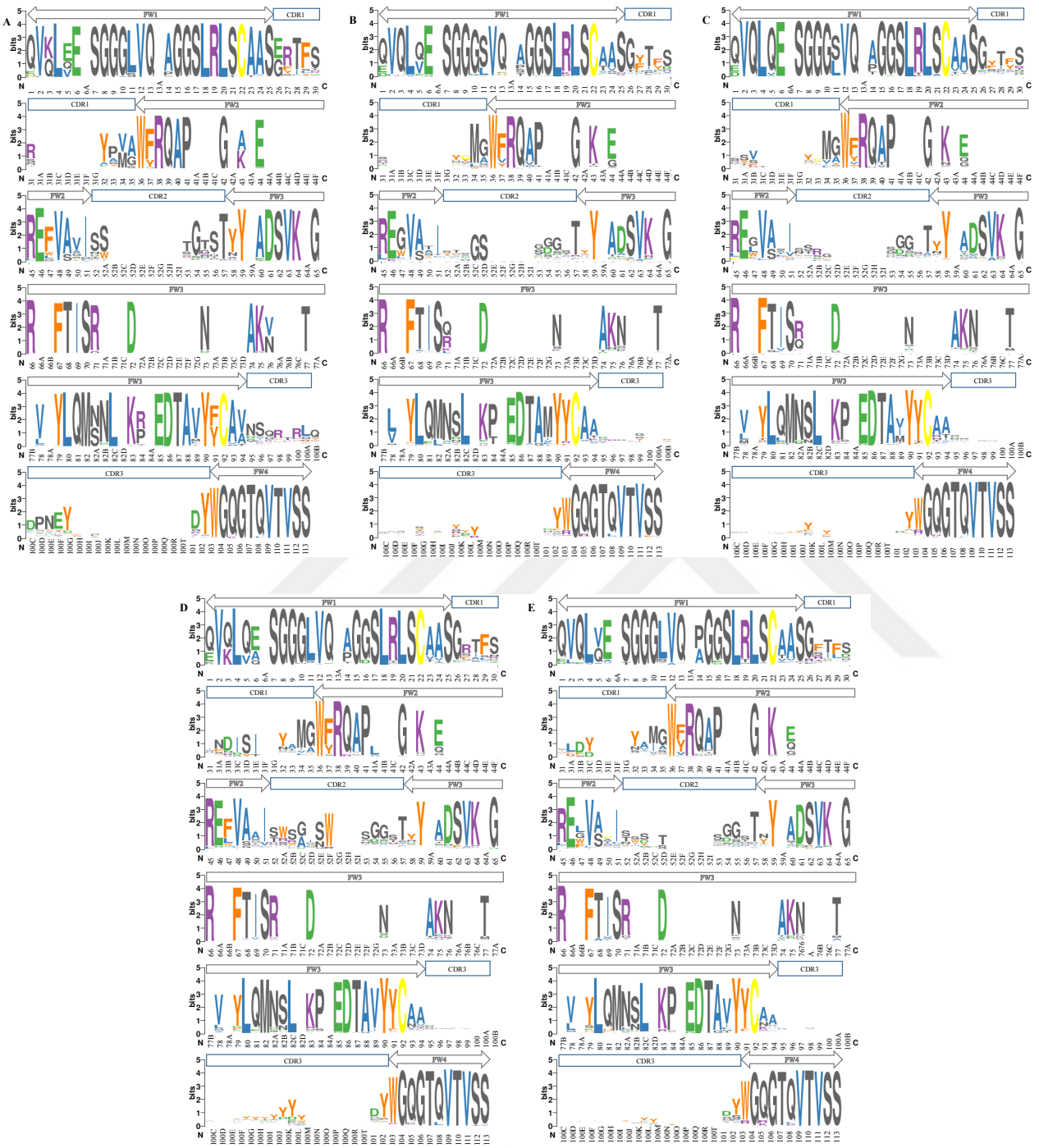


Figure S2. Graphical representation of multiple sequences of the V_HH datasets by logos based on species. A) Camelidae, B) *Camelus bactrianus*, C) *Camelus dromedarius*, D) *Lama glama*, and E) *Vicugna pacos*. Multiple sequences were aligned based on the Chothia numbering scheme. The relative frequency of each amino acid position was quantified by logo size in terms of bits. The color codes were based on amino acid biochemical properties: basic: K, R (purple); acidic: D, E (green); aliphatic: A, I, L, V (blue); aromatic: F, H, W, Y (orange); Cys bonds: C (yellow); mostly small: G, M, N, P, Q, S, T (dark gray). Gray boxes show constant FW regions and dark blue boxes show variable HCDR zones.

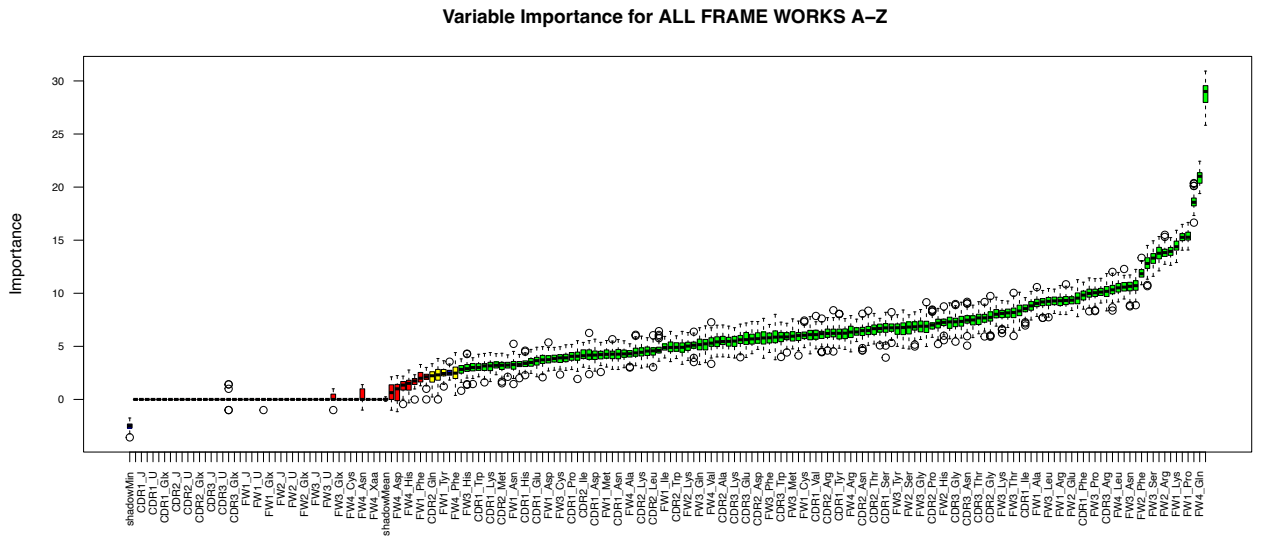


Figure S3. A box-and-whiskers plot showing the Boruta Algorithm-Random Forest ranking of the variable importance scores of features used for logistic regression modeling analysis.

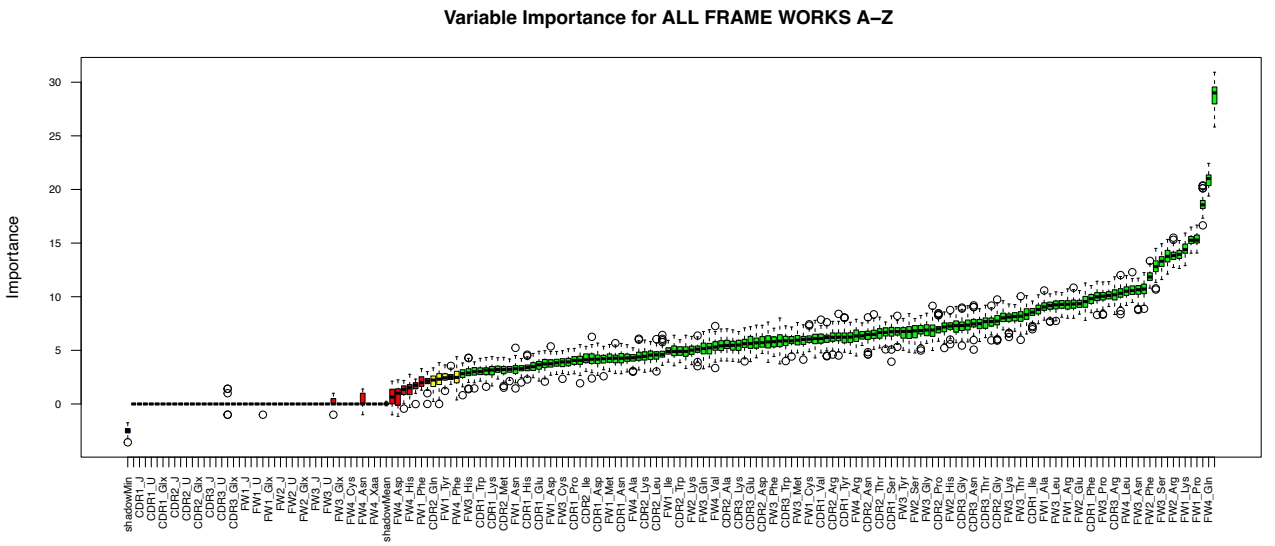


Figure S4. A box-and-whiskers plot showing the Boruta Algorithm-Random Forest ranking of the variable importance scores of each amino acid–region association. Blue boxes correspond to the minimal, average, and maximum Z-scores of shadow features in Figure S3. Red boxes indicate variables not contributing significantly to accurate classification. Green boxes indicate the amino acids contributing significantly to the classification that were selected for further evaluation. The centers of the boxes correspond to the median. The lower and upper hinges of the boxes correspond to the first and third quartiles of the data. The upper and lower whiskers extend from the hinges to the largest and smallest values, respectively, no further than 1.5 times the interquartile range from the hinges. Data beyond the end of the whiskers were plotted individually.

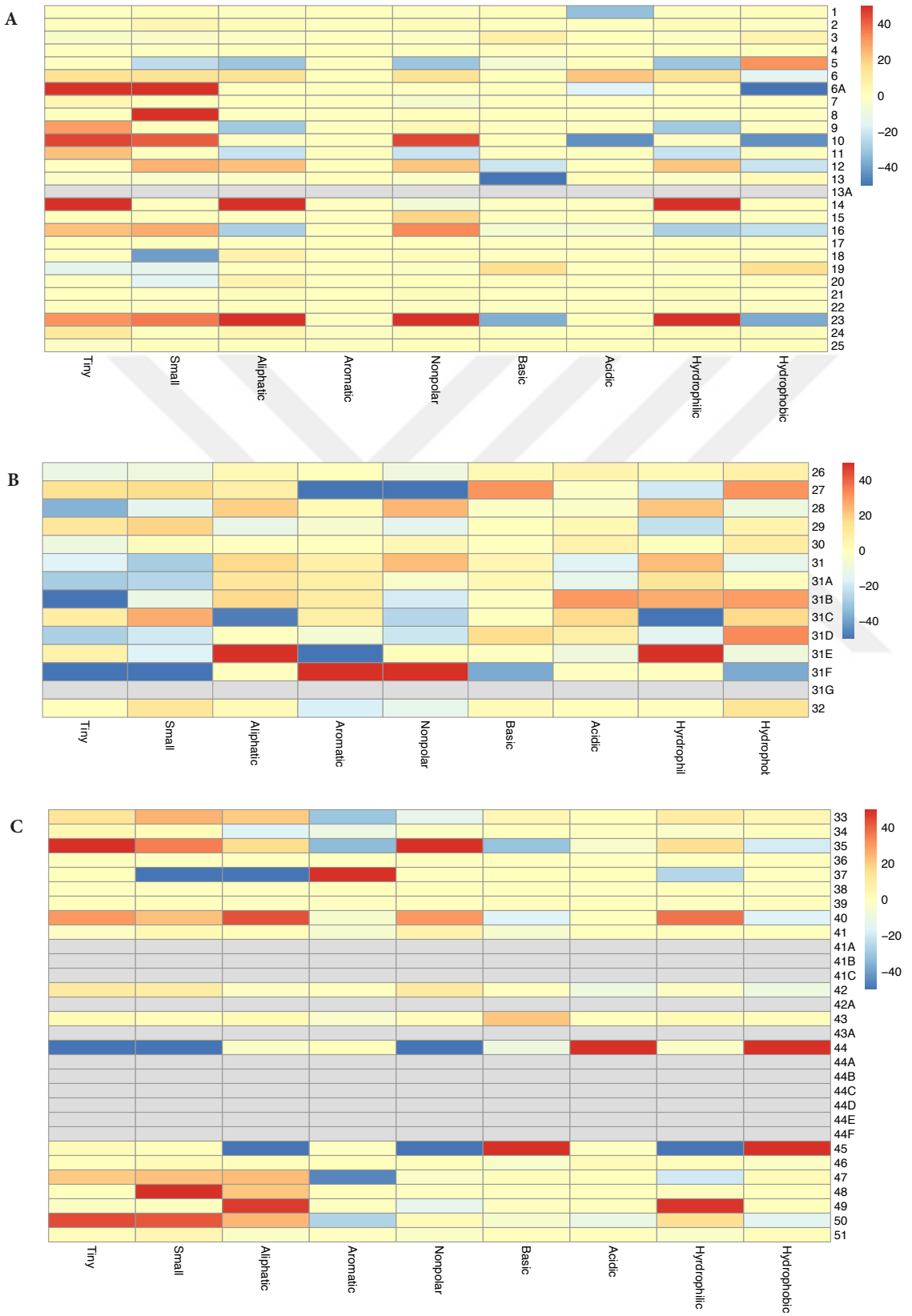

A
1
2
3
4
5
6
6A
7
8
9
10
11
12
13
13A
14
15
16
17
18
19
20
21
22
23
24
25
40
20
0
−20
−40
Tiny
Small
Aliphatic
Aromatic
Nonpolar
Basic
Acidic
Hyrdrophilic
Hydrophobic
B
26
27
28
29
30
31
31A
31B
31C
31D
31E
31F
31G
32
40
20
0
−20
−40
Tiny
Small
Aliphatic
Aromatic
Nonpolar
Basic
Acidic
Hyrdrophil
Hydrophot
C
33
34
35
36
37
38
39
40
41
41A
41B
41C
42
42A
43
43A
44
44A
44B
44C
44D
44E
44F
45
46
47
48
49
50
51
40
20
0
−20
−40
Tiny
Small
Aliphatic
Aromatic
Nonpolar
Basic
Acidic
Hyrdrophilic
Hydrophobic

Figure S5.

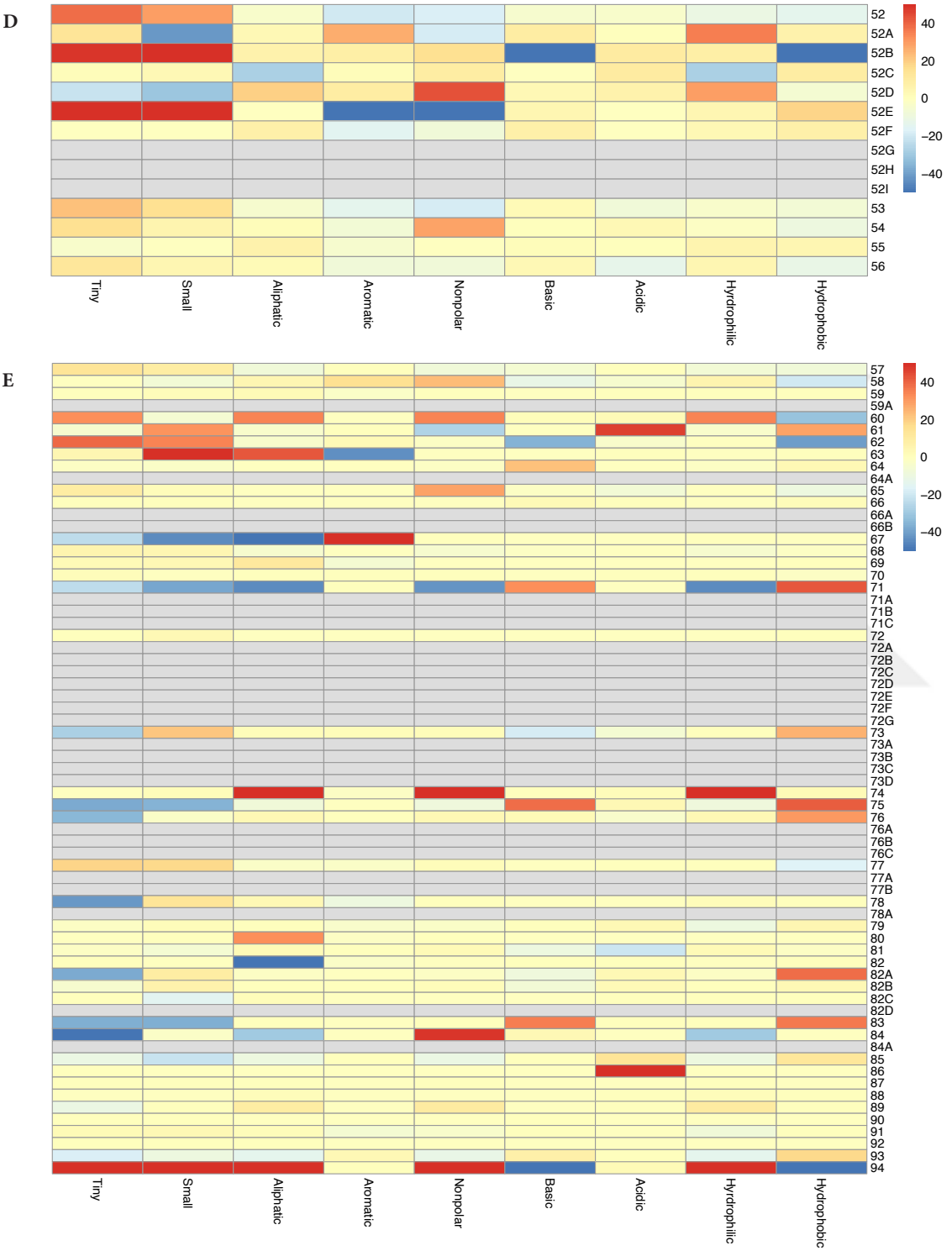


Figure S5. (Continued)

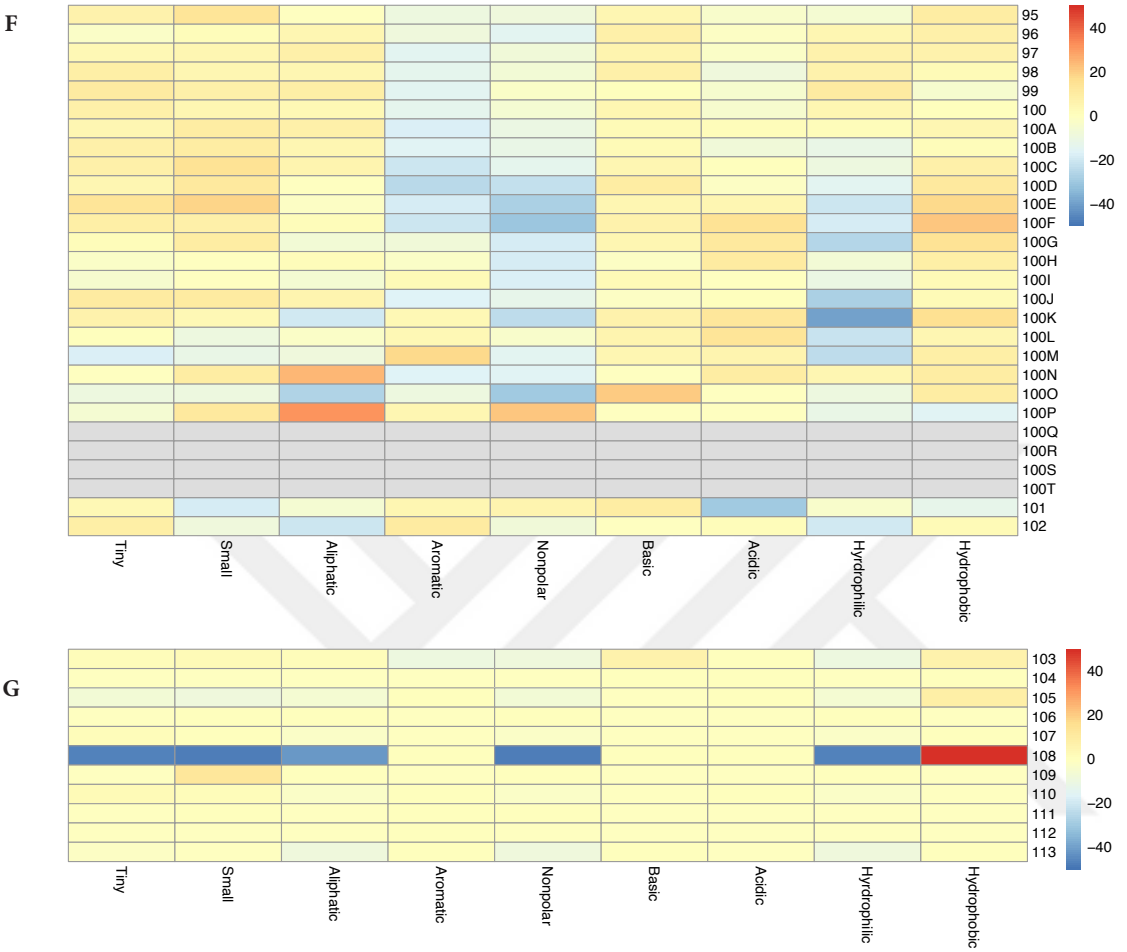


Figure S5. Heatmaps display the differences in physicochemical characteristics of residues between VHs and V_HHs. Rows represent individual residues, while columns indicate physicochemical properties of interest as determined in logistic regression analysis. Cell color indicates which antibody format contains a higher percentage of the indicated property in the residue. Red cells indicate a higher property in V_HHs, while blue cells indicate a higher property in VHs. Rows in gray indicate alternating positions, where either VHs or V_HHs lack the corresponding value, preventing a difference from being calculated. Individual heatmaps show (A) FW1, (B) HCDR1, (C) FW2, (D) HCDR2, (E) FW3, (F) HCDR3, and (G) FW4 residues. The color scale was set to 50 to –50, with higher or lower difference values being set to the set maximum or minimum, respectively.