PRODUCTION OF MONOCLONAL ANTIBODIES AGAINST ACTIVIN A RECEPTOR TYPE II LIKE KINASE 1 (ALK-1) BY PHAGE DISPLAY

NUR ALIA

UNIVERSITI SAINS MALAYSIA

2025

PRODUCTION OF MONOCLONAL ANTIBODIES AGAINST ACTIVIN A RECEPTOR TYPE II LIKE KINASE 1 (ALK-1) BY PHAGE DISPLAY

by

NUR ALIA

**Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science**

March 2025

ACKNOWLEDGEMENT

First and foremost, I wish to express my deepest gratitude to my esteemed supervisor, Associate Professor Dr. Lim Theam Soon. His unwavering support, insightful feedback, and constructive criticism have been truly invaluable throughout this academic journey. Dr. Lim's guidance has not only refined the direction of my research but also inspired me to strive for the highest standards of excellence.

I would also like to extend my sincere thanks to my co-supervisors, Associate Professor Dr. Yee Siew Choong and Dr. Wan Yus Haniff Wan Isa, for their involvement as part of my supervisory team that has provided essential support and reassurance, contributing significantly to the progress of my research.

My heartfelt appreciation goes to the entire AE team, especially my seniors Jing Yi and Angela. Their guidance and assistance from the very beginning of my research have been instrumental in ensuring its smooth and successful progression. A special note of thanks to my dear friend and steadfast supporter, Brenda Song. Her unwavering support, friendship, and encouragement have been a guiding light throughout this journey, and I am profoundly grateful for her constant presence.

I gratefully acknowledge the financial support of the Universiti Sains Malaysia, Special (Matching) Short-Term Grant with Project No: 304.CIPPM.6315708. The support has not only facilitated the necessary resources but also played a pivotal role in advancing my work. Additionally, I wish to express my appreciation to all INFORMM staff for their kind assistance in facilitating the completion of my studies.

Lastly, I owe my deepest thanks to my family members. Their prayers, steadfast belief in me, and unending support have been the foundation of my academic pursuit, providing me with the strength and motivation to succeed.

TABLE OF CONTENTS

ACKNOWLEDGEMENT .... **ii**

TABLE OF CONTENTS .... **iii**

LIST OF TABLES .... **vii**

LIST OF FIGURES .... **viii**

LIST OF SYMBOLS .... **x**

LIST OF ABBREVIATIONS .... **xi**

LIST OF APPENDICES .... **xiii**

ABSTRAK .... **xiv**

ABSTRACT .... **xvi**

CHAPTER 1 INTRODUCTION **1**

1.1 Background Study 1

1.2 Problem Statement 3

1.3 Research Objectives 5

1.3.1 Specific Objectives 5

CHAPTER 2 LITERATURE REVIEW **6**

2.1 Overview of Atherosclerosis 6

2.1.1 Role of LDL in Atherosclerosis 8

2.1.2 Role of ALK-1 in Atherosclerosis 9

2.2 Antibody 10

2.2.1 Antibody Structure and Function 10

2.2.2 MAbs 10

2.3 Phage Display Technology 11

2.3.1 Overview of Phage Display Technology 11

2.3.2 Bacteriophages 12

2.3.3 Mechanism of Infection by M13 Bacteriophages 13

2.3.4 Antibody Phage Display 13

2.3.5 Biopanning Process for Antibody Discovery 15

2.4 Antibody Therapy for CVDs 17

2.5 Bioinformatics Tools for Solubility and Sequence-Based Analysis in *E. coli* 21

2.5.1 IMGT/V-QUEST 21

2.5.2 SoluProt 22

2.5.3 DeepSoluE 23

2.5.4 SCRATCH 24

CHAPTER 3 METHODOLOGY 25

3.1 Experimental design 25

3.2 Materials 28

3.2.1 Preparation of media and buffer 28

3.2.2 Bacterial cells and helper phage 30

3.2.3 Additives and enzymes 31

3.2.4 Antibody conjugates 31

3.2.5 DNA and protein ladders 32

3.2.6 Primers 32

3.2.7 Kits and commercial protein 32

3.2.8 Equipment and software 33

3.3 Methods 33

3.3.1 ALK-1 antigen preparation 33

3.3.1(a) Expression of ALK-1 34

3.3.1(b) ALK-1 antigen extraction and purification 35

3.3.1(c) SDS-PAGE and Western blot analysis 36

3.3.1(d) Bradford’s assay 37

3.3.2 BMP-9 protein preparation 38

3.3.2(a) BMP-9 protein expression 38

3.3.2(b) BMP-9 protein extraction and purification 39

3.3.3 ALK-1 antigen functionality assay by ELISA 40

3.3.3(a) ALK-1 and BMP-9 binding assay 40

3.3.3(b) ALK-1 and LDL binding assay 40

3.3.4 Phage display 41

3.3.4(a) Preparation of M13KO7 helper phage 41

3.3.4(b) Packaging of human naïve scFv antibody phage library 43

3.3.4(c) Biopanning of phage library against ALK-1 44

3.3.4(d) Polyclonal phage ELISA 46

3.3.4(e) Monoclonal screening 47

3.3.4(f) Positive antibody clones verification with colony PCR 48

3.3.4(g) Solubility prediction of scFv clones 49

3.3.5 Expression and purification of positive antibody clones 50

3.3.6 Characterization of mAbs 51

3.3.6(a) Antibody titration assay by ELISA 51

3.3.6(b) Cross-reactivity assay by ELISA 52

3.3.6(c) LDL inhibition assay by ELISA 53

CHAPTER 4 RESULTS 54

4.1 Production of ALK-1 and BMP-9 protein 54

4.1.1 Expression and purification of ALK-1 protein 54

4.1.2 Expression and purification of BMP-9 protein 55

4.2 ALK-1 antigen functionality assay by ELISA 58

4.2.1 ALK-1 binding assay against BMP-9 protein and LDL 58

4.3 Isolation of monoclonal antibody 59

4.4 Antibody sequence analysis 62

4.5 Preparation of recombinant mAbs 64

4.6 Antibody titration and cross-reactivity test ... 67

4.7 LDL inhibition assay by ELISA ... 69

CHAPTER 5 DISCUSSION ... 72

5.1 Production and functional validation of ALK-1 ... 72

5.2 Monoclonal antibody selection and sequence analysis ... 73

5.3 Antibody titration, cross-reactivity, and LDL inhibition ... 75

CHAPTER 6 CONCLUSION ... 77

6.1 Conclusion ... 77

6.2 Future studies ... 78

REFERENCES ... 79

APPENDICES

LIST OF PUBLICATIONS

LIST OF TABLES

Page

Table 3.1 Preparation of media 28

Table 3.2 Preparation of buffer and reagents. 28

Table 3.3 List of bacterial cells and helper phage. 30

Table 3.4 List of additives 31

Table 3.5 List of enzymes. 31

Table 3.6 List of antibody conjugates. 32

Table 3.7 List of DNA and protein ladders. 32

Table 3.8 List of primers and their respective sequences. 32

Table 3.9 List of kits and commercial protein. 32

Table 3.10 List of laboratory equipment. 33

Table 3.11 List of software for protein analysis and visualization. 33

Table 3.12 List of reagents utilized for the preparation of the resolving and stacking gels in SDS-PAGE 36

Table 3.13 Biopanning conditions overview 46

Table 3.14 List of reagents used in the casting 12% SDS-PAGE resolving and stacking gel 51

Table 4.1 The VDJ gene family and CDR length of scFv clones. 63

Table 4.2 One-way ANOVA results for anti-ALK1 clone 4E/ALK-1 against LDL 71

LIST OF FIGURES

Page

Figure 2.1 A schematic representation of phage display biopanning procedure. 16

Figure 3.1 Flow chart of the experimental design of the study 27

Figure 4.1 SDS-PAGE and Western blot analysis of ALK-1 protein expression. 55

Figure 4.2 SDS-PAGE and Western blot analysis of ALK-1 protein purification. 55

Figure 4.3 SDS-PAGE analysis of BMP-9 protein expression and purification. 57

Figure 4.4 Western blot analysis of BMP-9 protein expression and purification. 57

Figure 4.5 Binding assay of ALK-1 against BMP-9 and LDL. 59

Figure 4.6 Biopanning of naïve phage display library against ALK-1. 60

Figure 4.7 A heatmap illustrating the results of monoclonal phage selection of an antibody against ALK-1, as indicated by the signal over noise ratio (S/N) ratio. 61

Figure 4.8 Agarose gel electrophoresis analysis of scFv mAbs. 62

Figure 4.9 Solubility prediction of soluble expression of anti-ALK1 clone 4E, 9F, and 7C in *E. Coli* using solubility prediction software SoluProt, DeepSoluE, and SCRATCH. 64

Figure 4.10 Western blot analysis of anti-ALK1 clones' expression in *E. Coli*. 65

Figure 4.11 SDS-PAGE analysis of purified anti-ALK1 clones. 66

Figure 4.12 Western blot analysis of purified anti-ALK1 clones. 67

Figure 4.13 Antibody titration assay of scFv clones via ELISA. 68

Figure 4.14 Antibody cross-reactivity assay of scFv clones via ELISA. 69

Figure 4.15 Inhibition Binding assay of anti-ALK1 clone 4E/ALK-1 against LDL via ELISA...70

LIST OF SYMBOLS

º C	Degree Celsius
g	Gram
kDa	Kilo Dalton
kV	Kilo Volt
%	Percent
µg	Microgram
µL	Microliter
nM	Nanomolar
mg	Milligram
mL	Milliliter
M	Molar
mM	Millimolar
rpm	Revolutions per minute
×g	Gravity force
v/v	Volume/volume
V	Volume
w/v	Weight/Volume

LIST OF ABBREVIATIONS

aa	Amino acid
ABTS	2,2'-azino-bis (3-ethylbenzothiazoline-6-sulfonic acid)
APS	Ammonium Persulfate
apoB	Apolipoprotein B
apoC-III	Apolipoprotein C-III
ALK-1	Activin receptor-like kinase 1
ANGPTL	Angiopoietin-like proteins
β-ME	Beta-mercaptoethanol
BMP-9	Bone morphogenetic 9
bp	Base pair
BSA	Bovine serum albumin
CAV-1	Caveolin 1
CDR	Complementarity-determining region
CV	Column volume
CVD	Cardiovascular disease
dH_20	Distilled water
DNA	Deoxyribonucleic acid
DNM2	Dynamin 2
dNTP	Deoxynucleotide triphosphate
D genes	Diversity genes
E. Coli	*Escherichia coli*
ELISA	Enzyme-linked immunosorbent assay
Fab	Antigen-binding fragment
HC	Heavy chain
HHT2	Hereditary hemorrhagic telangiectasia type 2
Ig	Immunoglobulin
J genes	Joining genes
LC	Light chain
LDL	Low density lipoprotein
LDL-C	LDL cholesterol
LDLR	LDL receptor

LPL	Lipoprotein lipase
mAb	Monoclonal antibody
M-PBS	Milk-phosphate buffer saline
mRNA	Messenger ribonucleic acid
Ni-NTA	Nickel-nitrilotriacetic acid
PBS	Phosphate buffer saline
PBS-T	Phosphate buffer saline-Tween20
PCSK9	Proprotein convertase subtilisin/kexin type 9
PLIN2	Perilipin 2
RE	Restriction enzyme
RF	Replicative form
RNA	Ribonucleic acid
scFv	Single chain variable fragment
ssDNA	Single-stranded deoxyribonucleic acid
TEMED	Tetramethyl ethylenediamine
TGF-β1	Transforming growth factor beta 1
VH	Heavy chain variable region
VL	Light chain variable region
V genes	Variable gene

LIST OF APPENDICES

Appendix A Standard curve for Bradford's assay

PENGHASILAN ANTIBODI MONOKLON TERHADAP RESEPTOR JENIS II SEPERTI KINASE 1 (ALK-1) MENGGUNAKAN PAPARAN FAJ

ABSTRAK

Aterosklerosis, penyakit keradangan kronik dinding arteri, adalah punca utama penyakit kardiovaskular. Lipoprotein berketumpatan rendah (LDL) memainkan peranan penting dalam perkembangan aterosklerosis. ALK-1 ialah reseptor jenis I dari superfamili TGF-β, telah dikenalpasti sebagai sasaran yang berpotensi untuk merawat aterosklerosis kerana penglibatannya dalam pengawalan tahap LDL. Kajian ini bertujuan untuk menghasilkan antibodi monoklonal terhadap ALK-1 menggunakan teknologi paparan faj untuk aplikasi terapeutik. Antigen ALK-1 telah berjaya diekspresikan dan ditulenkan, dan kefungsiannya telah disahkan melalui ujian pengikatan dengan ligan, *bone morphogenetic protein 9* (BMP-9), dan LDL. Biopanning perpustakaan faj antibodi naïve terhadap ALK-1 telah dilakukan, dan tiga klon scFv unik (4E, 9F, dan 7C) telah diperolehi. Analisis sikuen mendedahkan bahawa klon ini tergolong dalam keluarga gen VDJ yang berbeza dan mempunyai kepelbagaian panjang CDR yang berbeza. Keterlarutan klon scFv telah diramalkan menggunakan alat bioinformatik, dan klon telah diekspresikan dan ditulenkan daripada *Escherichia coli* (*E. coli*). Pencirian klon scFv yang telah ditulenkan melalui Ujian Immunosorbent Enzyme Terikat (ELISA) menunjukkan pengikatan khusus mereka kepada ALK-1, dengan klon 4E mempamerkan pertalian pengikatan tertinggi. Ujian reaktiviti silang mengesahkan kekhususan klon 4E kepada ALK-1. Terutama, klon 4E menunjukkan keupayaan untuk menghalang pengikatan LDL kepada ALK-1 dalam cara yang bergantung kepada kepekatan antibodi. Kesimpulannya, kajian ini berjaya menghasilkan antibodi monoklonal terhadap ALK-1 menggunakan teknologi

paparan faj. Klon scFv yang diperolehi, terutamanya klon 4E, menunjukkan hasil yang berpotensi dari segi kekhususan, pengikatan, dan keupayaan untuk menghalang pengikatan LDL kepada ALK-1. Oleh itu, kajian ini menekankan kepentingan dalam menangani cabaran yang ditimbulkan oleh aterosklerosis dan batasan rawatan semasa dengan membuka jalan untuk kemajuan masa depan dalam terapi kardiovaskular.

PRODUCTION OF MONOCLONAL ANTIBODIES AGAINST ACTIVIN A RECEPTOR TYPE II LIKE KINASE 1 (ALK-1) BY PHAGE DISPLAY

ABSTRACT

Atherosclerosis, a chronic inflammatory condition affecting the arterial wall, is a leading contributor to cardiovascular diseases (CVDs). Low-density lipoprotein (LDL) is a key factor in the pathogenesis of atherosclerosis by promoting plaque formation within arteries. Activate a receptor type II like kinase 1 (ALK-1), a type I receptor of the transforming growth factor (TGF)-β superfamily, has been identified as a promising target for atherosclerosis treatment due to its role in the regulation of LDL levels. This study aimed to generate monoclonal antibodies (mAbs) against ALK-1 using phage display technology for potential therapeutic applications. The ALK-1 antigen was successfully expressed and purified, and its functionality was confirmed through binding assays with its ligand, bone morphogenetic protein-9 (BMP-9), and LDL. Biopanning of a human naïve single-chain variable fragment (scFv) antibody phage library against ALK-1 was performed, resulting in the isolation of three unique scFv clones 4E, 9F, and 7C. Sequence analysis demonstrated that these clones pertained to distinct VDJ gene families and exhibited varying complementarity-determining region (CDR) lengths. Bioinformatics tools were used to predict the solubility of the scFv clones, which were subsequently expressed and purified from *E. coli.* Characterization of the purified scFv clones through enzyme-linked immunosorbent assay (ELISA) demonstrated their specific binding to ALK-1, with clone 4E exhibiting the highest binding affinity. Cross-reactivity tests confirmed the specificity of the clone 4E to ALK-1. Notably, clone 4E demonstrated the ability to

inhibit the binding of LDL to ALK-1, with its inhibitory effect increasing in a dose-dependent manner. In conclusion, this study successfully generated mAbs against ALK-1 using phage display technology. The isolated scFv clones, particularly clone 4E, showed promising results in terms of specificity, affinity, and the ability to inhibit LDL binding to ALK-1. Hence, this underscores the importance of this study in tackling the issues associated with atherosclerosis and the shortcomings of existing treatments, paving the way for future advancements in cardiovascular therapeutics.

CHAPTER 1

INTRODUCTION

1.1 Background Study

Atherosclerosis is a chronic inflammatory disease associated with an increased risk of mortality worldwide (Gisterå & Hansson, 2017; Młynarska et al., 2024). In 2021, CVDs were linked to an estimated 19.41 million deaths, underscoring the profound impact of atherosclerosis on public health (Our World in Data, 2021). This condition is triggered when there is an elevated level of LDL in the inner lining of the arteries which then accumulates and ultimately leads to the rupture of the blood vessel causing ischemic tissue damage (Das & Ingole, 2023; Gisterå & Hansson, 2017; Libby et al., 2019). ALK-1 which belongs to the TGF-β1 superfamily is a high-capacity receptor that can be utilized as a new therapeutic target for CVD treatment. ALK-1 has the ability to bind to LDL in a non-competitive manner promoting LDL transcytosis across the vascular endothelium (Kraehling et al., 2016; Lee et al., 2023). This high-capacity receptor is also able to bind to a dimeric TGF-β family cytokines such as BMP-9 at high affinity (Tao et al., 2020). The interactions between ALK-1 and BMP-9 initiate phosphorylation cascade of small mother against decapentaplegic (SMAD) which then relay signal from the membrane to the nucleus where the gene expression is regulated (González-Núñez et al., 2013). The binding of these proteins also triggers ALK-1 endocytosis through several pathways like caveolin 1 (CAV-1) and Dynamin 2 (DNM2). Interestingly, ALK-1 endocytosis mediated by BMP-9 reduces the uptake of LDL by ALK-1 (Kraehling et al., 2016).

Several medical interventions are available to address atherosclerosis, including the use of lipid-lowering medications such as statins and ezetimibe (Lusis, 2000). These medications are frequently administered as combination therapy and

have demonstrated effectiveness in reducing LDL levels and improving patient outcomes. However, the statin monotherapy has been observed to be challenging due to the development of intolerance in some patients towards the medication (Qiao et al., 2022). Meanwhile, the combination therapy of statins and ezetimibe has been shown to produce moderate improvements in lowering elevated cholesterol levels. Consequently, extensive research has been conducted to explore alternative non-statin therapeutic options (Bardolia et al., 2021). The presence of lipid accumulation on the walls of arteries is known to include antibodies suggesting that an immune response involving antibodies is essential in the development of atherosclerosis (Hansson, 2001; Ji & Lee, 2021). The Food and Drug Administration (FDA) approved antibody-based treatments, including Evolocumab and Alirocumab, are currently accessible to treat atherosclerosis by lowering LDL levels (Redberg & Prasad, 2017; Schwartz et al., 2018). These antibodies have the ability to impede the activity of Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9), which is also a target for therapeutic intervention in CVD.

Over the last decade, there has been significant research focusing on the development of mAbs owing to their potential for treating a wide range of diseases. The development of mAbs dates back to 1975, when Kohler and Milstein introduced Hybridoma technology. This process involved fusing B cells derived from immunized murine with myeloma cells, resulting in hybrid cells that are capable of producing antibodies with specific characteristics (Kohler and Milstein, 1975). Generating recombinant antibodies using Hybridoma technology is a complex and lengthy procedure, and when administered to humans, these antibodies often trigger immune responses that can lead to complications. Recombinant DNA technology has made it possible to produce recombinant antibodies using bacterial cells. Furthermore, the

advancement of recombinant human antibody technology has enabled a more robust and powerful method, such as phage display, to produce mAbs (Hamzeh-Mivehroud et al., 2013; Mustafa & Mohammed, 2024). Phage display is a technique that involves genetically engineering bacteriophages to display antibody fragments on their surface (Bazan et al., 2012; Ledsgaard et al., 2018). This technology enables the creation of an antibody library from which specific antibodies can be isolated (Bazan et al., 2012; Shukra et al., 2014). The potential application of mAbs to supplement the existing treatments for atherosclerosis via phage display technology, represents a promising approach for enhancing current therapeutic strategies. Hence, this research aims to explore the feasibility of producing mAbs against ALK-1 using phage display technology and to assess their potential as therapeutic agents to reduce the LDL accumulation.

1.2 Problem Statement

The Department of Statistics Malaysia's 2023 report revealed that CVD ranked among the top causes of death, contributing to 15.1% of deaths with medical certification (Department of Statistics Malaysia, 2023). Atherosclerosis, a common type of CVD are caused by the accumulation of LDL in vascular endothelium. Significant efforts have been dedicated to developing lipid-lowering drugs including statins and ezetimibe, which are employed to prevent the accumulation of cholesterol and decrease LDL levels in individuals with atherosclerosis. Although statins remain the primary treatment for CVD, numerous patients receiving less-intensive therapies continue to show LDL-cholesterol (LDL-C) concentrations exceeding the recommended target (Qiao et al., 2022; Yeoh et al., 2022). This may be attributed to factors such as genetic variations in cholesterol metabolism, inadequate LDL receptor

(LDLR) upregulation, or individual variations in drug metabolism. These factors contribute to the limited efficacy of statins (Yeoh et al., 2022). However, when statin is used in combination with ezetimibe, it only leads to moderate improvements in reducing elevated cholesterol levels (Kaddoura, Orabi, & Salam, 2020; Vavlukis & Vavlukis, 2018).

While newer therapies like the PCSK9 inhibitors (Evolocumab and Alirocumab) have demonstrated significant efficacy in lowering LDL-C by enhancing LDL receptor recycling, while Evinacumab, an antibody targeting angiopoietin-like protein 3 (ANGPTL3), has shown promise in reducing both LDL-C and triglyceride levels (Jeswani et al., 2024; Mohamed et al., 2022; Stoekenbroek et al., 2018). These approaches have a significant drawback as they mainly concentrate on decreasing LDL-C in the bloodstream rather than inhibiting LDL transcytosis across the endothelium, which is essential in the development of atherosclerotic plaques. Despite aggressive lipid-lowering treatments, atherosclerosis continues to progress in some patients due to ongoing LDL infiltration into the vascular intima. This underscores the necessity for novel therapeutic strategies to address this issue (Deng et al., 2024; W. Li et al., 2022). While commercial anti-ALK-1 mAbs are available, there remains a research gap in developing mAbs specifically aimed at inhibiting LDL transcytosis by targeting ALK-1. It is worth noting that the exploration of mAbs against ALK-1 for the specific purpose of inhibiting LDL transcytosis is not commonly reported, given that ALK-1 has additional biological functions and disease pathways, such as those related to cancer and inflammation. This highlights the significance of engaging in further research aimed at developing mAbs that specifically target ALK-1 in order to inhibit LDL transcytosis, while also enhancing our comprehension and potentially optimizing therapeutic interventions.

1.3 Research Objectives

The objective of this study was to generate and characterize mAbs against ALK-1 using phage display technology.

1.3.1 Specific Objectives

1. Truncated recombinant ALK-1 protein was expressed using a bacterial expression system.
2. MAbs against the expressed ALK-1 protein were developed using the phage display technique.
3. The developed anti-ALK-1 mAbs were studied and characterized.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview of Atherosclerosis

CVD is widely regarded as the one of the primary culprits that contributes to the increased rate of death in developing countries, including Malaysia. According to the 2023 data from the Department of Statistics Malaysia, CVD was identified as one of the leading causes of mortality, accounting for 15.1% of all medically certified deaths (Department of Statistics Malaysia, 2023). Ischemic heart disease refers to one of the clinical manifestations of atherosclerosis (Lahoz & Mostaza, 2007; Młynarska et al., 2024). The pathogenesis of atherosclerosis is initiated by the deposition of LDL-C on the inner linings of arteries, leading to plaque formation. With increased levels of LDL-C in the blood plasma, the blood vessel gradually becomes hardened and narrowed, potentially limiting blood flow to cardiac muscle tissue (Dai et al., 2024). If the coronary artery, which supplies blood to the heart, becomes obstructed by plaque, it may eventually rupture due to the force of blood flow, potentially leading to the formation of a blood clot. This may lead to various health issues, including a heart attack (medically termed myocardial infarction) and stroke (Hao & Friedman, 2014).

Multiple factors can contribute to the development of atherosclerosis, including poor lifestyle choices, dietary habits, and behaviours. These risk factors encompass tobacco use, excessive alcohol intake, lack of exercise, elevated stress levels, and the consumption of unhealthy foods (Beshir et al., 2021; Lechner et al., 2020). Additionally, pre-existing medical conditions and genetic hereditary can also contribute to the development of this condition (Firus Khan et al., 2022). Heart disease caused by high cholesterol levels has the lowest rate of diagnosis and can frequently go undiagnosed. The situation is compounded by the fact that this condition frequently does not present

any discernible symptoms, which leads to its underestimation. Consequently, some individuals are unaware of the elevation of cholesterol levels (Wan Ahmad et al., 2023).

In order to address this issue, medical interventions such as cholesterol-lowering medications have been administered to patients with elevated cholesterol levels. These medications include statins and ezetimibe. Statins are typically considered the first line of defence for lipid lowering, with their primary function being to inhibit cholesterol production in the liver. Some individuals have expressed difficulty tolerating the medication, which makes it challenging to reduce the LDL-C level (Guedeney et al., 2022; Robinson et al., 2015). It is evident that relying on statin monotherapy is insufficient in treating this condition. On the other hand, ezetimibe operates as a cholesterol absorption inhibitor that specifically targets cholesterol obtained from dietary sources, bile, or desquamated cells (Dembowski & Davidson, 2009; Lestari et al., 2023; Sun et al., 2021). Although combining statins with ezetimibe has been shown to lower cholesterol levels, the degree of reduction achieved is relatively modest (Kaddoura et al., 2020; Vavlukis & Vavlukis, 2018). A recent development in reducing lipid levels involves utilizing mAbs that target and suppress PCSK9. The FDA has approved evolocumab and alirocumab, two human immunoglobulin (Ig) G antibodies, for managing and preventing elevated cholesterol levels (Della Pepa et al., 2017; Jeswani et al., 2024; Sabatine et al., 2017). These antibodies function by binding to PCSK9 and inhibiting its ability to interact with the LDLR, which is a key receptor responsible for clearing circulating LDL through endocytosis. This has resulted in an accelerated clearance and reduction of LDL levels in the blood (Kaddoura, Orabi, Amar, et al., 2020).

2.1.1 Role of LDL in Atherosclerosis

Cholesterol is an essential element that plays a significant role in various cellular processes produced by the liver and absorbed from the intestines (Iqbal et al., 2017). LDL, commonly known as the primary carrier of cholesterol, is a large, spherical particle comprised of an oily core containing 1500 cholesterol esters molecules, enclosed in a layer of phospholipids and free cholesterol, with Apolipoprotein B (ApoB) 100 embedded within it (Duan et al., 2023; Goldstein & Brown, 1977; Ridker, 2014). The phospholipid layer consists of hydrophilic heads on the outside, enabling LDL to dissolve in the blood or intercellular fluid.

LDL internalization into peripheral tissue is regulated by a receptor called LDLR. LDL has high affinity to LDLR, and upon binding, it is internalized by the cell through a process akin to adsorptive endocytosis. Once a sufficient amount of cholesterol has been taken up by the cells, a mechanism exists that allows the cells to signal the LDLR to reduce the uptake of cholesterol, thus preventing the accumulation of cholesterol within the cells. The strong binding of LDL to its receptor enables cells to meet their cholesterol needs even when the LDL levels are low in both blood plasma and interstitial fluid (Goldstein & Brown, 1977; Shaya et al., 2022). Besides LDLR, LDL can also be internalized through an LDLR-independent pathway that is receptor-specific, such as the ALK-1 receptor, which functions to mediate LDL entry and transcytosis into the endothelial cell (Kraehling et al., 2016; X. Zhang et al., 2018). The pathogenesis of atherosclerosis is initiated by the binding of ALK-1 to LDL, which is followed by the transcytosis of LDL into the endothelial cell. This process contributes to the accumulation of plaques in the artery walls and is a significant contributor to atherosclerosis progression.

2.1.2 Role of ALK-1 in Atherosclerosis

ALK-1 has been recently recognized as a receptor that facilitates the transport of LDL into endothelial cells and plays a significant role in the development of atherosclerosis. It is primarily expressed in these cells (Kraehling et al., 2016; H. Wang et al., 2021). The structure of ALK-1 is composed of three primary components: the cysteine-rich extracellular domain (EC), the glycine-serine domain, and the serine-threonine kinase domain. The EC domain, which is primarily responsible for ligand binding, is the most crucial element of ALK-1 (González-Núñez et al., 2013). Loss-of-function mutations in ALK-1 are associated with the development of a medical condition known as Rendu-Osler-Weber syndrome, which is commonly referred to as Hereditary Hemorrhagic Telangiectasia (HHT). This condition results in arteriovenous fragility and malformations (Johnson et al., 1996; Salmon et al., 2020).

ALK-1 plays dual roles in atherosclerosis, one involving the direct binding of LDL to its extracellular domain to facilitate LDL uptake, and the other involving the BMP signalling (Akadam Teker et al., 2020). Briefly, the extracellular domain of ALK-1 that is positioned at the apex can interact with LDL containing ApoB-100 and facilitate the transport of LDL across the endothelial layer, ultimately leading to the formation of atherosclerotic lesions (Roman & Hinck, 2017). ALK-1 is widely recognized as the primary receptor for BMP9, displaying a strong affinity for the ligand (Tao et al., 2020). When BMP-9 binds to the extracellular domain of ALK-1, it initiates a series of events that ultimately lead to the phosphorylation of SMAD1/5/8 (Mahlawat et al., 2012; Ruiz et al., 2016). This event is critical for gene expression regulation. Additionally, the binding of BMP-9 to ALK-1 promotes the uptake of the receptor by the cell, which reduces LDL transcytosis (Tao et al., 2020). The potential of ALK-1 as a therapeutic target for diseases has been bolstered by its unique mechanism of direct

binding to LDL, which differs from the binding of LDLR. Furthermore, the absence of downregulation of this binding in response to sterols and PCSK9 distinguishes ALK-1 as a promising target for the treatment of atherosclerosis (Akadam Teker et al., 2020).

2.2 Antibody

2.2.1 Antibody Structure and Function

Ig, commonly referred to as antibodies, are produced by a single B cell clone and constitute a crucial element of the adaptive immune system. The main role of Ig is to defend the organism against harmful microorganisms like viruses and bacteria that attempt to invade the body (Megha & Mohanan, 2021). An antibody's basic structure is characterized by a "Y" configuration, consisting of two sets of identical chains: light chains (LCs) and heavy chains (HCs). These chains are interconnected through disulfide bonds to form the distinctive shape. Specifically, each HC is connected to a LC, while the two HCs are linked to one another (Chiu et al., 2019). Antibodies are classified into five distinct isotypes: IgD, IgA, IgM, IgG, and IgE. These isotypes are differentiated by their heavy chain (HC) sequences and exhibit unique functional properties (Goulet & Atkins, 2020). The most commonly used and preferred isotype for monoclonal antibody therapeutics is IgG due to its inherent stability, low immunogenicity, and efficient production of a large number of mAbs targeting various antigens with relative ease (Tang et al., 2021).

2.2.2 MAbs

MAbs are regarded as the "magic bullet" of medicine, capable of being utilized for both therapeutic and diagnostic purposes against a diverse array of antigens (Waldmann, 1991). The increase in demand for these specific and selective antibodies has been significant in recent years. These small molecules are derived of a single clone

of B cells (Nelson et al., 2000). The development of mAb production began in 1975 when Kohler and Milstein introduced hybridoma technology. Their research revealed that the fusion of B cells from immunized murine myeloma cells generated hybrid cells that produced antibodies with unique specificities (KÖHLER & MILSTEIN, 1975). The first therapeutic monoclonal antibody, Orthoclone OKT3, was developed using this technique. It targets the CD3 receptor on T lymphocytes and was later applied to prevent rejection in renal transplantation (Norman et al., 1988). The Hybridoma technology, which has been well-established for many years, continues to lead the way in monoclonal engineering. Nevertheless, this method is not without its drawbacks, including suboptimal efficiency, a lengthy duration for creation, and a substantial cost (Goding, 1980; Moraes et al., 2021). The emergence of hybridoma technology has prompted a significant transformation in the domain of antibody engineering, thereby opening doors for more sophisticated platforms, including display systems (D. T. Y. Chan & Groves, 2021; Chiu et al., 2019).

2.3 Phage Display Technology

2.3.1 Overview of Phage Display Technology

Phage display technology, well-known for its significance in antibody engineering, is a powerful method used to select peptides or proteins with targeted binding affinities from a vast array of variants. G. Smith first introduced this technique in 1985, showing that a protein from another organism could be displayed on the exterior of the M13 filamentous bacteriophage (Smith, 1985). In essence, phage display is a technique that display peptides, proteins, or antibody fragments on phage particles' surface. This is achieved by integrating the DNA sequence that encodes the protein intended for display into the phage or phagemid genome. The integration is done by

fusing this sequence to a gene that codes for a phage coat protein (Hoogenboom, 2002). The concept of phage display technology lies in the structural components that connect the expressed protein's phenotypic and genotypic characteristics to its ability to replicate (C. E. Z. Chan et al., 2014). This fusion ensures that mature phages exhibit the target protein on their surface, while simultaneously containing the genetic sequence encoding it within the same particle. The advantages of phage display technology include its cost-efficiency, ease of use, and time-saving nature, making it a preferred choice in the field of mAb development (Anand et al., 2021).

2.3.2 Bacteriophages

Phages, also referred to as bacteriophages, are viruses with single-stranded DNA that specifically infect Gram-negative bacteria, including *E. Coli*. Several types of phages are employed in phage display technology, such as M13, f1, fD, and ft (Aghebati-Maleki et al., 2016). Filamentous bacteriophages are widely employed in phage display technology due to their capacity to tolerate insertions in non-essential genomic regions without compromising phage packaging. Furthermore, filamentous bacteriophages exhibit stability across a range of potential conditions, such as various pH and temperature levels (Kehoe & Kay, 2005; Pande et al., 2010).

Among the available bacteriophages, M13 is the most commonly employed phage. This F positive family member measures approximately 6.5 nm in diameter and 900 nm in length (Ledsgaard et al., 2018). M13 is preferred over other phages due to its ease of purification and use. The M13 virion ssDNA genome, encased in a capsid composed of 2700 units of the primary coat protein p8, with both ends capped by specific minor coat proteins. Specifically, one terminus is capped with five copies each of p7 and p9, while the opposite terminus is capped with five copies of p3 and p6 (Kehoe & Kay, 2005).

2.3.3 Mechanism of Infection by M13 Bacteriophages

Filamentous phage undergoes a non-lytic life cycle, where infected bacteria continuously produce and release phage particles without undergoing cell lysis (Ledsgaard et al., 2018). The infection process begins with the attachment of phage p3 to the f pilus of male *E. Coli*, such as TG1 and insertion of the ssDNA into the host cytoplasm (Azzazy & Highsmith, 2002). Upon entry into a host cell, the coat protein disassembles, and its components are released into the cell envelope. The phage genome hijacks the host's cellular machinery to generate a complementary DNA strand, creating a double-stranded DNA known as the replicative form (RF). This RF serves as a template for the synthesis of RNA and proteins, as well as the production of RF progeny and single-stranded DNA (ssDNA) used in phage particle formation (Ebrahimizadeh & Rajabibazl, 2014).

2.3.4 Antibody Phage Display

There are two primary phage display vector systems for presenting antibodies: the M13 phage DNA or phage vector and the phagemid DNA vector (Aripov et al., 2024; Zhao et al., 2016). These systems use either minor enveloped protein p3 or the major envelope protein p8 to display of desired antibody fragments (Tohidkia et al., 2012). The phage vectors are derived from the M13 genome and incorporate all the crucial genes that are necessary for the processes of infection, replication, and assembly of filamentous phage (Paschke, 2006). The antibody genes to be displayed can be directly inserted into the phage genome, with the wild type M13 phage p3 gene serving as the fusion site. As a result, the antibody displayed by the phages will carry the recombinant genome encoding that antibody.

On the other hand, the phagemid vector system is a hybrid composed of *E. Coli* plasmid and M13 phage genome, which enables the cloning and expression of fusion proteins. These phagemid vectors are designed to incorporate the origin of replication for both the M13 phage and *E. Coli* plasmids, as well as an antibody p3 fusion gene, multiple cloning sites, and an antibiotic gene resistance gene (Ledsgaard et al., 2018). The phagemid vector, however, does not possess the requisite genes to assemble a fully functional phage (Azzazy & Highsmith, 2002). Hence, when employing a phagemid vector, co-infection with helper phage is indispensable for providing wild-type phage proteins that subsequently assemble into a phage particle (Soltes et al., 2007). Phagemid vectors are increasingly utilized in phage display due to their higher transformation efficiency, which is beneficial in constructing larger libraries (O'connell et al., 2002).

The phage display antibody library comprises various types, namely naïve, immune, semi-synthetic, and synthetic. These libraries are determined based on the origin of the antibody gene sequence (Kumar et al., 2019). The naive libraries are derived from the IgM mRNA of healthy donor B cells of non-immunized individuals. The most significant advantage of a naïve repertoire lies in its ability to screen antibodies against a diverse array of targets including toxins and self-antigens (Ponsel et al., 2011). The library, however, is limited by the non-specificity of the antibodies in the naive population. In order to increase the likelihood of obtaining high-affinity antibodies, a larger library size is necessary (S. K. Chan et al., 2017). Immune libraries, on the other hand, are obtained from immune donors extracted from peripheral lymphocytes and are typically smaller in size (Ponsel et al., 2011). In contrast to naïve libraries, immune libraries are not suitable for targeting a wide range of antigens and are more effective for isolating disease-specific binders (Lai & Lim, 2020). The primary advantage of this library is its ability to identify high affinity binders (Kumar et al.,

2019). As opposed to naïve and immune libraries, synthetic libraries are designed *in silico* involves the substitution of the CDRs with a diverse range of random sequences, resulting in an extensive variety of antigen binding sites (Guliy et al., 2023). Semi-synthetic libraries are created by combining natural and chemically synthesized sequences, resulting in a balance between preserving the natural framework and increasing diversity (C. C. Lim, Choong, et al., 2019).

2.3.5 Biopanning Process for Antibody Discovery

Biopanning is an *in vitro* process of selecting phage display antibody based on their affinity for a specific target in order to identify those with the desired binding properties following antibody library constructions (Panagides et al., 2022). The selection process of positive clones requires several rounds of biopanning of purified antigens against phage display antibody libraries. The biopanning process entails a series of steps: immobilization, phage binding, washing of unbound phages, elution and amplification (Guliy et al., 2023; Rami et al., 2017). This process is repeated several times, typically four to five rounds, to ensure the enrichment of phages with high affinity for the target antigen (C. C. Lim, Woo, et al., 2019).

The initial stage in the biopanning process involves the immobilization of the target antigen onto a solid support, such as the wells of a microtiter plate, a column matrix, or magnetic particles, particularly for biotinylated antigens (Bazan et al., 2012; França et al., 2023). After immobilization, antibody displaying phage is subsequently incubated with the target antigen, enabling the binding of specific binders to the antigen. The subsequent stage involves the elimination of unbound phages through washing, followed by the elution of phages bound to the target antigen via enzymatic treatment or pH modification (França et al., 2023; Krishnaswamy et al., 2009; Tulika et al., 2023). Eluted phages referred to as "rescued phages" are employed for the infection of *E. Coli*,

followed by co-infection with a helper phage to amplify the enriched phage (Azzazy & Highsmith, 2002; C. C. Lim et al., 2019; Thanongsaksrikul & Chaicumpa, 2011). Following multiple rounds of biopanning, the enriched pool of antibody populations from each round undergoes a detailed evaluation through polyclonal ELISA to assess the binding affinity and specificity. Following the last round of selection step, random colonies are chosen for monoclonal screening, and their unique antibody sequences are determined via sequencing. The biopanning approach can be tailored and refined by incorporating diverse competing conditions, including modifications to the antigen immobilization concentration, blocking agents, washing stringency, and elution conditions, to increase its efficacy and the likelihood of identifying candidates with desirable traits (Kumar et al., 2019; Weisser & Hall, 2009). Figure 2.1 illustrates the biopanning process utilizing phage display technology.

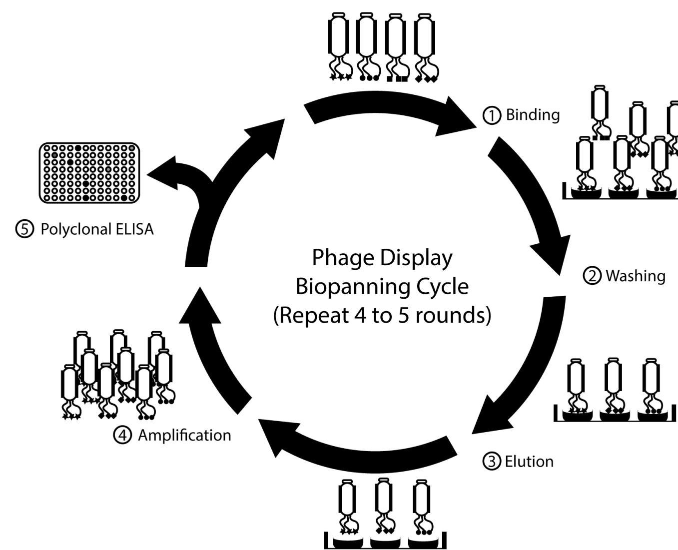

Figure 2.1 A schematic representation of phage display biopanning procedure.

The first step involves the binding of pre-blocked antibody phage particles with the target antigen coated on a solid phase. Following binding, unbound phages are removed through washing, and bound phages are eluted from the target. The eluted phages are then amplified by infecting *E. coli*, generating a new phage pool for subsequent selection rounds. Following 3-5 rounds, polyclonal ELISA is conducted to assess the enrichment and specificity of the selected phages.

2.4 Antibody Therapy for CVDs

Immunotherapy is becoming an increasingly promising strategy for treating CVDs, as evidenced by the Canakinumab Anti-inflammatory Thrombosis Outcome Study (CANTOS), which utilized an anti-IL1β antibody to improve CVD outcomes (Porsch & Binder, 2019). However, this approach demonstrated limited success in reducing mortality, emphasizing the need for alternative methods. An alternative approach involves targeting lipoprotein metabolism, particularly PCSK9, a protein that regulates LDL cholesterol levels (Stoekenbroek et al., 2018). The mAbs against PCSK9, such as Evolocumab and Alirocumab, have been developed and approved by the FDA and EMA. These antibodies prevent PCSK9 from binding to LDL receptors on liver cells, thereby enhancing LDL cholesterol clearance. These antibodies effectively lower LDL cholesterol, especially in patients with statin intolerance or genetic hypercholesterolemia, marking a significant advancement in CVD treatment (Seidah et al., 2019).

Another potential target associated with lipoprotein metabolism involves ANGPTLs, particularly ANGPTL3, 4, and 8. These proteins play a crucial role in lipid metabolism and high levels of triglycerides (TG) are a significant risk factor for CVD (Y. Li & Teng, 2014; Yang et al., 2021). The exploration of ANGPTLs as therapeutic targets to lower plasma TG levels is ongoing (Chen et al., 2021; Gugliucci, 2024; Thorin et al., 2023). ANGPTL3, primarily expressed in the liver, promotes the cleavage and degradation of lipoprotein lipase (LPL), inhibiting the hydrolysis of TG-rich lipoproteins into free fatty acids and glycerol, thereby increasing plasma TG levels (Morelli et al., 2020). The mAbs against ANGPTL3, such as Evinacumab (REGN1500), have shown promising results in reducing both TG and LDL cholesterol levels, consequently lowering CVD risk (Mohamed et al., 2022; Reiner, 2018). ANGPTL4 and

ANGPTL8 also inhibit LPL activity, but their roles are more complex due to their tissue-specific functions and interactions. Anti-ANGPTL4 antibodies, while effective in reducing TG levels, have shown adverse effects such as mesenteric lymphadenopathy, complicating their therapeutic use (Yang et al., 2021). ANGPTL8, lacking the C-terminal fibrinogen-like domain found in other ANGPTLs, indirectly regulates LPL by enhancing ANGPTL3's inhibitory effects (Mohamed et al., 2022; Quagliarini et al., 2012). The development of anti-ANGPTL8 antibodies, such as REGN3776, has shown potential in preclinical models, reducing TG levels and increasing high density lipoprotein (HDL) cholesterol (Gusarova et al., 2017). However, further investigation is needed to validate these targets for CVD therapy in humans and to fully understand the range of ANGPTL functions and potential side effects.

Apolipoprotein C-III (apoC-III) is a promising target in cardiovascular therapy due to its role in TG-rich lipoproteins, such as chylomicrons, VLDL, and HDL. It is produced by hepatocytes and its flexible structure allows it to adapt to various lipoprotein surfaces, impacting lipid metabolism. ApoC-III is believed to impede the binding of lipoproteins to LPL and endothelial cells, disrupting lipolysis and increasing hypertriglyceridemia (HTG) and CVD risk. Studies indicate that individuals with APOC3 mutations, such as the Amish, exhibit lower plasma TG levels and higher HDL cholesterol, suggesting a protective effect against CVD. Meanwhile, mAbs targeting apoC-III could prevent its integration with lipoproteins, enhancing LPL-mediated lipolysis and LDL clearance from the bloodstream. Targeting apoC-III's hydrophobic region may reduce CVD risk by facilitating lipoprotein remnant clearance.

Another potential therapeutic objective is Perilipin 2 (PLIN 2), a crucial lipid droplet-associated protein that plays a significant role in lipid storage and metabolism.

PLIN 2 is predominantly found in macrophages and atherosclerotic plaques, and it aids in lipid droplet formation and regulates intracellular cholesterol by binding to lipid droplets, protecting them from degradation, and affecting lipid accumulation. Targeting PLIN 2 with mAbs could disrupt its interaction with lipid droplets, facilitating lipolysis by enhancing access for lipases such as adipose TG lipase (ATGL). This strategy may reduce TG levels and alleviate atherosclerosis. Nonetheless, further research is necessary to fully understand PLIN 2's role in lipid metabolism and its therapeutic potential.

Alternatively, Lipoprotein Lipase (LPL) plays a pivotal role in lipid metabolism by binding to capillary surfaces and hydrolysing TGs in chylomicrons and VLDL (Ramasamy, 2014). LPL is secreted by adipose tissue, skeletal muscle, cardiac muscle, and macrophages. Activation of LPL results in a decrease in TG levels and an increase in HDL levels, thereby decreasing the risk of atherosclerosis (Packard et al., 2020). LPL influences atherosclerosis by promoting lesion formation through vessel wall activity while mitigating progression through plasma activity, highlighting its dual role dependent on localization (Stein & Stein, 2003). Therefore, targeting LPL with mAbs may prevent lipoprotein breakdown and reduce the occurrence of atherosclerotic lesions. Hence, it is important to carefully consider the dual effects of LPL. In addition, utilizing bispecific antibodies that target multiple lipid pathways may further enhance cardiovascular benefits, driven by advances in antibody engineering (Yeoh et al., 2022).

Another promising strategy involves the inhibition of ALK-1, a protein that plays a significant role in LDL transcytosis which is a process that contributes to the accumulation of LDL and the development of atherosclerosis (Kraehling et al., 2016; Tao et al., 2020). MAbs that target ALK-1's LDL binding sites have the potential to prevent LDL from entering the subendothelial space, thereby reducing the risk of plaque

formation. Consequently, this presents a promising therapeutic approach for managing LDL-related vascular diseases and atherosclerosis.

2.5 Bioinformatics Tools for Solubility and Sequence-Based Analysis in *E. coli*

2.5.1 IMGT/V-QUEST

The international ImMunoGeneTics information system (IMGT®) was established in 1989 by Marie-Paule Lefranc at Montpellier to systematically characterize the genes and alleles responsible for Ig and T cell receptor (TR) synthesis, thereby managing the complexity of adaptive immune responses in jawed vertebrates (Collins et al., 2023; Lefranc & Lefranc, 2019; Manso et al., 2022). Over time, IMGT® has evolved into an integrated repository that not only covers IG and TR but also major histocompatibility (MH) proteins and related molecules from the Ig and MH superfamilies across both vertebrates and invertebrates (Brochet et al., 2008; Manso et al., 2022). This resource has been foundational in establishing immunoinformatics as an interdisciplinary field that bridges immunogenetics and bioinformatics (Brochet et al., 2008; Lefranc & Lefranc, 2019).

A central component of this system is the IMGT/V-QUEST tool, which has been instrumental in analysing rearranged IG and TR nucleotide sequences (Garcia-Calvo et al., 2023; C. C. Liu et al., 2019). Using alignment-based methods, IMGT/V-QUEST identifies the variable (V), diversity (D), and joining (J) gene segments and quantifies nucleotide mutations and amino acid changes, including those from somatic hypermutation (Paulsen et al., 2022). The tool further refines its analysis by delineating V-D-J or V-J junctions with integrated modules such as IMGT/JunctionAnalysis and IMGT/Automat, ensuring comprehensive and standardized annotation of the variable domain (Brochet et al., 2008). Additionally, its capacity to detect insertions and deletions enhances the precision of immune receptor sequence analysis (Brochet et al., 2008; Nielsen & Boyd, 2018).

High-throughput sequencing (HTS) has underscored the necessity of tools like IMGT/V-QUEST for detailed repertoire analysis in both healthy and pathological conditions, such as infectious diseases, autoimmune disorders, and cancers. The tool's reference directories, which include sequences for functional genes, open reading frames, and in-frame pseudogenes, facilitate accurate sequence comparisons across species (Manso et al., 2022). By providing key parameters such as alignment scores, percentage identities, and CDR lengths, IMGT/V-QUEST enables researchers to assess the functionality and diversity of antigen receptor sequences effectively (Lefranc & Lefranc, 2019; Tausch et al., 2024). Its practical utility has been demonstrated in various studies, including antibody humanization and the quantification of somatic hypermutation levels (Defrancesco et al., 2021; Duggan et al., 2022; S. Li et al., 2025).

Overall, the integration of IMGT/V-QUEST within the broader IMGT® framework represents a significant advancement in immunogenetics. It offers a robust, standardized, and user-friendly platform for deciphering the complex mechanisms underlying antigen receptor diversity, thus supporting a wide range of immunological research and clinical applications (Brochet et al., 2008).

2.5.2 SoluProt

SoluProt is a novel, machine learning–based tool developed to predict soluble protein expression in *E. coli* from protein sequence data (Hon et al., 2021). SoluProt assesses both solubility and expressibility by incorporating up to 96 sequence-derived features, employing the gradient boosting machine method and trained using the TargetTrack database (Hon et al., 2021; B. Li & Ming, 2024). The system's architecture prioritizes user-friendliness and has demonstrated superior performance compared to earlier prediction techniques in both threshold-dependent and threshold-independent

evaluation criteria, thus improving the identification of proteins likely to exhibit soluble expression (Ghomi et al., 2020; Thumuluri et al., 2022).

SoluProt offers a practical solution by employing a threshold value of 0.5 and the sequences scoring above this value are predicted to be soluble when expressed in *E. coli* (Bashour et al., 2024; Zeidi et al., 2024). The tool's robust performance has been demonstrated in various studies, showing improved accuracy compared to other predictors, with reported accuracies ranging from 58.5% to 74% under different validation schemes (Elalouf & Yaniv-Rosenfeld, 2023; Kaur et al., 2024; Khan et al., 2023; Markus et al., 2023; Zaib et al., 2023). Overall, SoluProt stands out as a valuable resource in protein and antibody engineering research, enabling researchers to efficiently identify candidates with a higher likelihood of successful soluble expression in *E. coli* (Ghomi et al., 2020).

2.5.3 DeepSoluE

DeepSoluE is a recently introduced deep learning tool designed to predict protein solubility in *E. coli* using only sequence information. It combines physicochemical properties with distributed amino acid representations to capture key sequence patterns that influence solubility. A genetic algorithm is applied to identify the most informative feature subsets, and these selected features are integrated through long short-term memory networks for classification purposes (C. Wang & Zou, 2023; Wei et al., 2024).

The practical utility of DeepSoluE is further demonstrated in its application to vaccine development, where it has been used to predict the solubility of vaccine candidate proteins expressed in bacterial systems (Shey et al., 2025). Its ensemble learning strategy employs a soft voting approach with a threshold value of 0.4, which

helps ensure balanced and reliable prediction outcomes (C. Wang & Zou, 2023). Comparative analyses have shown that DeepSoluE outperforms other available solubility prediction servers, making it a valuable resource for prescreening targets in experimental workflows and reducing unnecessary laboratory costs (Rcheulishvili et al., 2023). Overall, the accessibility and robust performance of DeepSoluE support its use as an effective bioinformatics tool in the context of antibody engineering and related fields.

2.5.4 SCRATCH

The SCRATCH server is a user-friendly online platform that predicts key protein structural features from amino acid sequences, providing essential insights for vaccine design. It offers information on secondary structure, solvent exposure, and other important parameters by simply submitting a sequence and receiving the results by email (Cheng et al., 2005). In vaccine development studies, SCRATCH has been used to guide the selection of interferon-inducing $CD4^+$ T cell epitopes and to evaluate the overall properties of vaccine constructs (Kaushik et al., 2022). A notable component of the server is SOLpro, which estimates the solubility of proteins by assigning a probability score, with scores of 0.5 or higher suggesting a good chance of soluble expression (Atapour et al., 2022; Ullah et al., 2022). By integrating these predictive tools, the SCRATCH server streamlines the in-silico evaluation of protein candidates, thereby helping to reduce the need for extensive experimental testing (Sastry et al., 2017).

.