Doctoral Dissertation

Development of a binding protein from the beta-sandwich domain of a hyperthermophile

Chukwuebuka Maxwell Ononugbo

January 2025

Department of Biotechnology
Graduate School of Engineering,
Osaka University

Table of Contents

Table of Figures

Table of Tables

Abbreviations

AIC: aggregation initiation concentration, **APR**: aggregation-prone region, **ASA**: Accessible surface area, **BPB**: bromophenol blue, **BLAST**: Basic local alignment search tool, **BLASTp**: Basic local alignment search tool protein, **BLOSUM**: Blocks amino acid substitution matrices, **CBB**: Coomassie Brilliant Blue, **CCC-dsDNA**: covalently closed circular double-stranded deoxyribonucleic acid, **CD**: circular dichroism, **CFU**: Colony forming unit, **ClustalO**: Clustal Omega, **DNA**: Deoxyribonucleic acid, **dsDNA**: Double-stranded deoxyribonucleic acid, ***E. coli***: *Escherichia coli*, **ELISA**: Enzyme-linked immunosorbent assay, **EtOH**: Ethanol, **GKR**: Gatekeeper residue, **GUI**: graphical user interface, **IgG**: Immunoglobulin, **IPTG**: isopropyl-β-D-thiogalactopyranoside, **LB**: Luria Bertani, **LS**: long-term solubility, **MOI**: multiplicity of infection, **NaCl**: Sodium chloride, **NCBI**: National center for biotechnology information, **PBS**: Phosphate buffered saline, **PBST**: Phosphate buffered saline with tween 20, **PCR**: Polymerase chain reaction, **PDB**: protein database, **PEG**: Polyethylene glycol, **pH**: potentials of hydrogen, **Phage**: Bacteriophage, **pI**: Isoelectric point, **PPI**: protein-protein interaction, **PVDF**: Polyvinylidene fluoride, **RE**: Restriction enzyme, **RNA**: Ribonucleic acid, **SD1**: Sandwich domain 1, **SD1-M1**: Sandwich domain 1 mutant 1, **SDS-PAGE**: Sodium dodecyl-sulfate polyacrylamide gel electrophoresis, **ssDNA**: Single-stranded deoxyribonucleic acid, **SYNBIP**: Synthetic Binding Proteins for Research, Diagnosis, and Therapy, **TBS**: Tris-buffered saline, **TEMPURA**: TEMPeratures of Usual and RAre prokaryotes, **Tm**: Melting temperature, **UCSF**: University of California San Francisco, **UniProt**: Universal protein resource, **UniProtKB**: Universal protein resource knowledgebase, **Unipro UGENE**: Unified bioinformatics toolkit, **WT**: wild type, **ΔΔG**: change in free energy.

Chapter 1:

Introduction

1.1 Non-immunoglobulin binding proteins as antibody alternative: An overview

Proteins play a variety of vital roles in biological systems that significantly rely on their interactions with either one or several proteins, and this recognition and association with one another regulate a wide range of physiological, biological, and biochemical processes (Belvisi et al., 2021). Erroneous interactions have been linked to a number of illnesses, and the manipulation of protein-protein interactions (PPIs) has become crucial in the treatment of various diseases, drug discovery, diagnostics, development of diverse affinity reagents and molecular tools for basic science research and biomedicine (Lu et al., 2020; Belvisi et al., 2021). Non-immunoglobulin binding proteins are a group of proteins that have emerged as potent PPI modulators in recent years, with their applications expanding to include a broader range of protein types with desirable features.

Non-immunoglobulin binding proteins are a group of artificially designed affinity proteins engineered *in vitro* to have a similar or greater specificity, affinity, and potency as the monoclonal antibodies (Hober et al., 2019). They have recently gained popularity as a feasible alternative for protein-based biological molecules, serving as a universal binding framework and building block for proteins with targeted intended functions (Azhar et al., 2017). These groups of artificially designed binding proteins are often referred to as scaffold proteins. The successful engineering of scaffold proteins spanning several structural classes has enabled the rapid development of novel affinity reagents facilitated by rapidly evolving *in vitro* and *in vivo* techniques for protein library design, functional variant selection, and precise target recognition (Nygren and Skerra, 2004).

Several scaffold proteins have been described (Binz et al., 2005; Skerra, 2007; Weidle et al., 2013; Azhar et al., 2017; Simeon and Chen, 2018) and are generally divided into antibody-derived or non-antibody-derived groups. They are also classified into two categories; those that

bind ligands via amino acids on the exposed loops on a rigid protein structure and those that bind through amino acids on the secondary structure.

For several decades, the prototype for affinity binding to ligands with desired specificities has been provided by antibodies, of which the recombinant versions have entered into several applications with remarkable success (Gebauer and Skerra, 2009). However, with the increasing applications in research, biotechnology, and biomedicine, it became apparent that antibodies have several limitations. This propelled the development of protein scaffolds with binding functions using combinatorial protein design to address the several limitations associated with the use of monoclonal antibodies.

The full-length IgG has a molecular weight of about 150 kDa and consists of four polypeptide chains: two heavy chains and two light chains. The polypeptide chains are composed of well-folded domains, made up of twelve IgG domains, with the heavy and light chains connected by a disulfide bond (**Figure 1**). The complexity, size, and planar binding interfaces of IgG pose a limitation for use in several applications requiring the penetration of solid tumors and poorly vascularized tissues (Pandey and Mehrotra, 2024; Lai et al., 2018; Chames et al., 2009; Vazquez-Lombardi et al., 2015; Cruz and Kayser, 2019). The fragment crystallizable region (Fc)-mediated complement-dependent cytotoxicity (CDC) and the antibody-dependent cellular cytotoxicity (ADCC), which are rarely required in simple target protein neutralizations, can have undesirable effects, such as the antibody-dependent enhancement (ADE) of viral infections (Lee et al., 2020, Bournazos et al., 2020, and Thomas et al., 2022). Additionally, monoclonal antibodies (mAbs) require complex post-translational modifications, such as unique glycosylation patterns, necessitating their production in mammalian cells, which are more expensive to optimize and cultivate (Zheng et al., 2011; Jefferis, 2016) when compared to production in prokaryotic cells. This requirement for the production of mAbs in mammalian cells increases the cost of the products. The high cost of products and other limitations,

including the complex intellectual property involved with antibody engineering (Storz, 2011; 2015), propelled the extension of the concept of target ligand binding from antibodies to alternative protein frameworks.

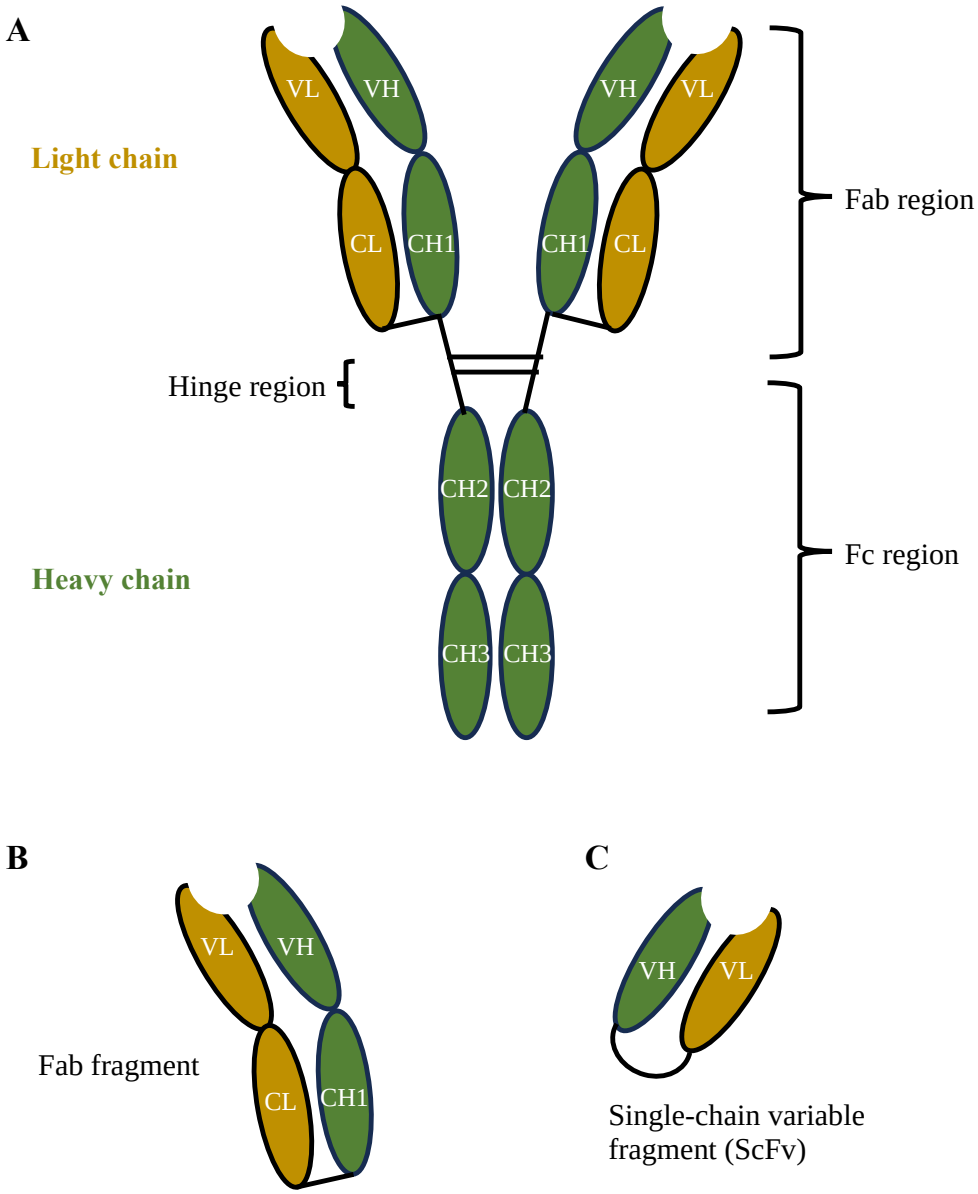

Figure 1: The structure of IgG. A: The structure of the full-length IgG molecule. **B:** The structure of the fragment antigen binding (Fab) fragment. **C:** The structure of the ScFv fragment

The idea to engineer alternative protein binders originated in the 20th century, a decade after the recombinant antibody technology was developed (Gebauer and Skerra, 2020), and has

continued to expand to include several proteins of different architectures and properties. The search for alternative protein binders began with using the different antibody fragments as a framework, mainly the fragment antigen-binding (Fab) region and the single chain fragment variable (scFv), the single variable domain on a heavy chain (VHH) antibodies (Munoz-Lopez et al., 2022, Nuttall et al., 2000, Tungekar and Ruddock, 2023, Bever et al., 2017, Hirao et al., 2023), down to including several unrelated proteins or protein fragments for similar applications simply by employing combinatorial engineering and molecular selection techniques to identify and select target binders with desired specificity (Gebauer and Skerra, 2009). The Fabs and scFvs, unlike other single-domain proteins, require hydrophobic residue-mediated interdomain interactions between the variable heavy (VH) and light (VL) domains, which limits thermodynamic stability and increases the risk of variable domain mispairing and the formation of aggregates (Pandey and Mehrotra, 2024).

Single-domain scaffold proteins are expected to complement the functions of IgG by providing the desired additional properties lacking in IgG. These properties include small size, high thermal stability, free of disulfide bonds, and easy production in bacteria cells without the need for denaturation and refolding, lowering the overall cost of production and the cost of the final product (Weidle et al., 2013). The production of small-sized folded single polypeptide chains in bacteria cells also provides a renewable supply of reagents, addressing the issues of batch-to-batch differences encountered in animal-produced antibodies (Bedford et al., 2017). Therefore, in addition to compensating for antibody size, stability, and cost constraints, protein scaffold design may also offer specific intellectual property protection (Gebauer and Skerra, 2009; Simeon and Chen, 2018). In addition, the simple design of scaffold proteins allows for extensive structural engineering to satisfy a range of requirements, including specificity, affinity, stability, solubility, and protease resistance, making it very attractive. It has opened avenues for applications in diverse fields in basic science, biotechnology, and medicine.

Non-immunoglobulin binding proteins are usually designed to exert their action by antagonizing receptors and inhibiting their ligand binding sites preventing the interaction with cognate receptors (targeted therapy), binding to targets enabling the detection, visualization, and the study of the molecular functions of target proteins, (imaging, biosensors, and immunoassays), preparation of immunoconjugates where an antigen binding moiety is genetically fused to reporter enzymes as seen in diagnostics and proteomics applications, capture probes in microarray detection, crystallization chaperones, affinity purification, fluorescence resonance electron transfer (FRET), and in functional protein knock-out based on protein interference and degradation (Nygren and Skerra, 2004, Ferrigno, 2016, Harmansa and Affolter, 2018, Limsakul et al., 2018, Zhao et al., 2023). These applications have expanded from simple target detection and neutralization (Jenkins et al., 2019; Qiao et al., 2023; Hannula et al., 2024) to include targeted drug delivery to molecular targets (Shipunova and Deyev, 2022; Serna et al., 2022), adapters for adenovirus delivery (Dreier et al., 2013, Schmid et al., 2018) and several new possibilities discussed briefly below and summarized in **Figure 2**.

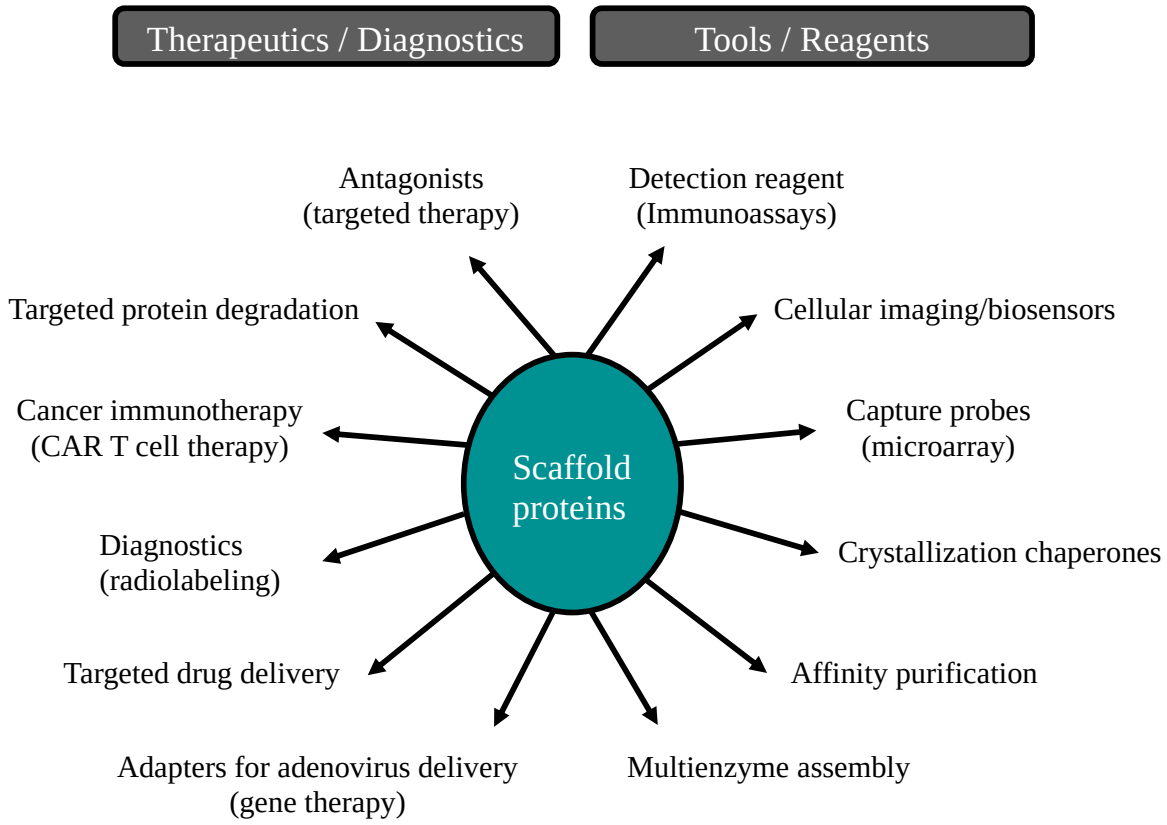

Figure 2: The summary of the several applications of non-immunoglobulin binding proteins.

The dysregulation of the normal cellular processes governed by the complex network of proteins has been implicated in oncogenic, degenerative, and autoimmune diseases. These proteins, which include modulators of gene expression, cell cycle progression, apoptosis, protein folding, and oncogenic proteins, among others (Miersch and Sidhu, 2016), offer rich potential targets for therapeutic interventions and the understanding of the dynamics of the protein-protein interactions (PPIs) involved in cell signaling providing the potential for the development of new therapies. Proteins that bind to and regulate biological processes by modulating the PPIs are often referred to as scaffold proteins. The natural scaffold proteins generally act as a framework to enable the efficient functioning of other biological components and thus lack enzymatic activity (Hata and Lida, 2009; DiRusso et al., 2022). However, with the significant advances in library design and selection methods, researchers could construct personalized affinity reagents to serve several purposes (Binz et al., 2005). In this study, the

term scaffold protein refers to protein frameworks that can allow a few or many amino acid modifications or sequence insertions to confer new functions on the protein variants, usually binding to desired targets. Scaffold proteins have shown tremendous success in modulating PPIs by targeting the cellular kinases, especially in targeting some of the groups of proteins labeled as undruggable (Oliver, 2017; Martin et al., 2018). However, this molecular targeted therapeutic approach where the target inhibitors exert their activity on the target only in the bound state is vulnerable to adaptive resistance, resulting in the requirement for high drug dose, which may lead to off-target toxicities (Kelm et al., 2023). This limitation led to the design of a new therapeutic method aimed at targeted protein degradation (TPD), which prevents the possibility of breakthrough signaling. Targeted protein degraders are bifunctional small molecules that recruit ubiquitin-proteasome systems to degrade target proteins (Bekes et al., 2022). Studies have shown that the use of scaffold proteins provides researchers with versatility by enabling the targeted degradation of challenging therapeutic targets (Shen and Dassama, 2023).

The invention of chimeric antigen receptor (CAR) T-cell therapy has transformed cancer immunotherapy, allowing for individualized treatments tailored to the specific needs of each patient. The CAR, a synthetic protein with an extracellular binding domain and intracellular signaling domains, directs and activates T lymphocytes to destroy cells expressing target antigens (Sterner and Sterner, 2021; Alnefaie et al., 2022; Mitra et al., 2023). CAR T cell immunotherapy enables target antigen recognition without presentation by the major histocompatibility complex (MHC), which prevents cancer cells from using MHC downregulation to evade the immune system (Keane and Posey, 2021; Kitching and Jaw, 2018). Commonly used CARs consist of the scFv directed against a tumor-associated antigen; however, the reliance on the correct paring between the linked VH and VL domains by the scFv and the tendency to oligomerize poses a significant drawback, causing tonic signaling, T cell

exhaustion, and poor in vivo performance (Long et al., 2015, Ajina and Maher, 2018, Frigault et al., 2015). Since the essential prerequisite for effective CAR T cell therapy is the specific interaction of the CAR molecule with the target antigen and the subsequent induction of T cell effector functions, the use of scaffold proteins completely fulfilled the requirements. The ever-advancing protein engineering techniques could eliminate the dependency on antibody-derived constructs for target recognition (Zajc et al., 2021).

Photo-regulation of protein functions aimed at improving cellular, molecular, and organismal level investigations and the development of novel tools and therapeutics (Kneuttinger, 2022) are becoming increasingly attractive. Optogenetics is revolutionizing the dynamic control of diverse cellular activities, allowing the non-invasive and remote application of light in biological samples (McCue and Kuhlman, 2022). The structural advantages provided by scaffold proteins and their ease of expression in many cell types, including remaining actively functional in and outside of the cell, are consistently being exploited by researchers to create new possibilities. The structural stability and flexibility provided by this group of proteins have expanded their applications to include optical control of binding to targets. Gil et al. (2020) and Carrasco-Lopez et al. (2020) described the development of light-sensitive protein binders designed by inserting the light-oxygen-voltage-sensing domain from *Avena Sativa* phototropin 1 (AsLOV2) into the solvent-exposed loop of nanobodies (OptoNBs) and monobodies (OptoMBs) respectively to reversely control the binding and release of untagged target proteins. Similarly, Woloschuk et al. (2021) designed photo-sensitive affinity reagents based on the Z-domain of affibodies by combining the Z-domain with a photoactive yellow protein. The controlled light-dependent binding to untagged targets opened up new avenues for modulating signaling pathway activities, reversible ligand neutralization, and real-time study of targets biological functions, as well as light-based protein purification from heterogenous mixtures without the need for affinity tags that could interfere with protein function.

The stability, substrate accessibility, catalytic performance, and prolonged availability of enzymes are improved by immobilization, and the techniques include adsorption, entrapment, encapsulation, covalent binding, and affinity immobilization (Datta et al., 2013; Maghraby et al., 2023). Scaffold proteins are becoming helpful in this field as they serve as tools for reversible multi-enzyme assembly. Scaffold protein based on the staphylococcal protein A was designed to mimic the architecture of cellulosomes and binds reversibly efficiently to the cellulose surface when fused with carbohydrate-binding protein (Eklund et al., 2004; Gad and Ayakar, 2021).

Scaffold proteins are expanding protein science and paving the way for the expansion of affinity reagents, and researchers are continuously extending these applications to several new possibilities. This is due to the architectural framework provided by this group of proteins, including their size, which enables easy genetic manipulations coupled with the ever-advancing protein engineering strategies. Therefore, in the selection and design of a scaffold protein, the intended application could play a significant role.

1.2 The design and selection of non-immunoglobulin binding proteins.

No doubt, scaffold proteins make up a diverse source of tools in several fields, as discussed above, and this is owing to the diversified origins, individual protein architectures, and properties that influence their suitability in different application contexts. The chances of obtaining a sufficiently robust affinity reagent depend on the features of the original protein framework. The protein framework is expected to provide a structurally compact and stable structure that can accommodate various side-chain substitutions to facilitate genetic engineering and functional expression for the development of binding functions without affecting the protein's structural integrity (Ahmadi et al., 2021; Nygren and Skerra, 2004). In general, scaffold proteins usually consist of a single polypeptide chain.

There are two significant ways to design and select a desired framework for the development of affinity reagents, and they include searching scaffolds from the protein database (PDB) or de novo protein design (Luo et al., 2022). For the selection of candidate scaffold from the PDB, several existing scaffolds are screened using a variety of criteria, which include small size, origin, presence or absence of disulfide bond, expression in *E. coli*, nontoxicity, absence of cofactor, solubility (available on literature data), and structural architecture (β-sheets, α-helix, number of loops, fold). Similarity alignments using Dali (Holm and Sander, 1995) or PDBeFold may be employed to refine the filtered results further. This is followed by experimentally verifying the final candidate to confirm their suitability for the desired applications. Then, the selected candidate is diversified by combinatorial protein engineering methods, and subsequently, binders are selected based on the affinity and specificity of the exposed targets.

The second approach is the de novo design of affinity binders using deep learning-based algorithms. In this approach, the need to prespecify the protein's secondary structure or fold may not be necessary (Wang et al., 2022). The common ways include first identifying the amino acid side chains that produce desired functions using fragment docking calculations. The target functional site can also be deduced from native proteins having desired binding functions. On identifying the functional amino acid side chains, the complete amino acid sequence of the protein is designed and optimized such that the predicted 3D structure bears the functional site (Anishchenko et al., 2021; Norn et al., 2021). This approach is called “constrained hallucination.” In “inpainting,” starting from the functional site, additional sequence and structure are filled in using a pre-trained structure-prediction network (RoseTTAFold), RoseTTAFold diffusion (Rfdiffusion) (Baek et al., 2021, Watson et al., 2023). Silva et al., 2019 and Cao et al., 2022, designed protein binders to specific sites on target proteins using only the information on the 3D structure of the target. These studies illustrate the unlimited possibilities

for the targeted design of binders to any target of interest. In the de novo design of protein binders, several thousand to millions of diversified new structures are usually generated. These are further streamlined to a few candidates based on several in silico evaluations, such as thermodynamic simulation calculations, among others (Luo et al., 2022). The best-performing candidates are then selected for experimental validations.

For this study, my discussion will mainly focus on the conventional protein design approach, which relies on identifying a protein framework (scaffold) based on the structural architecture and the biochemical properties of the protein.

1.3 The structural classes of non-immunoglobulin binding proteins

To design binders to any predetermined target, a clear structure of the starting framework is necessary to identify the positions on the scaffold protein that needs to be diversified before the construction of a combinatorial library. Although the structure of the scaffold protein often reveals the potential functions of the designed affinity binders, the structural integrity needed to withstand diverse mutations is a more critical factor to consider. However, considering the structures of designed affinity binders, each scaffold often shows distinct modes of recognition of its target, reflecting the architectural composition of the protein (Gilbreth and Koide, 2012). Therefore, it is essential to consider the different structural classes of scaffold proteins and how their architecture may affect their stability and potential binding functions. Based on the architecture of most proteins, scaffold proteins can be classified into three groups: those consisting predominantly of α-helices, those consisting predominantly of β-sheets, and lastly, those with a mixture of α-helices and β-sheets (Hosse et al., 2006). Positions for diversification could be on the surface exposed loops, α-helices, β-sheets, or a combination. The collection of designed synthetic binding proteins, the diversified positions, and the targets, are deposited in the Synthetic Binding Proteins for Research, Diagnosis, and Therapy (SYNBIP) database

(Wang et al., 2022). Some examples have been discussed to buttress further how the initial protein framework affects its design for binding and the kinds of targets (Gilbreth and Koide, 2012; Ramamurthy et al., 2012). DARPins developed from natural ankyrin repeat proteins (Stumpp et al., 2008) recognize out-curved surfaces due to their hollow binding sites. Anticalins a lipocalin-based scaffold (Skerra, 2001), tend to grip their targets just like DARPins because of their basket-like structure. Affibodies derived from the Z domain of protein A, a cell wall protein of *Staphylococcus aureus* (De et al., 2018) usually recognize and bind flat surfaces due to their flat binding site. The monobodies are derived from the domain of fibronectin type III and the nanobodies are derived from the variable domain of the Camelid antibodies (VHH) (Hantschel et al., 2021; Asaadi et al., 2021). They recognize groves and crevices and are suitable for binding the catalytic sites of enzymes because of the IgG-like fold structure and the presence of protruding surface-exposed loops. Additionally, monobodies designed to bind curved-out epitopes have been designed (**Figure 3**). These studies suggest that the structural architecture of the scaffolds determines the topography of the binding site, which directly affects the type of epitopes it recognizes with high affinity and specificity.

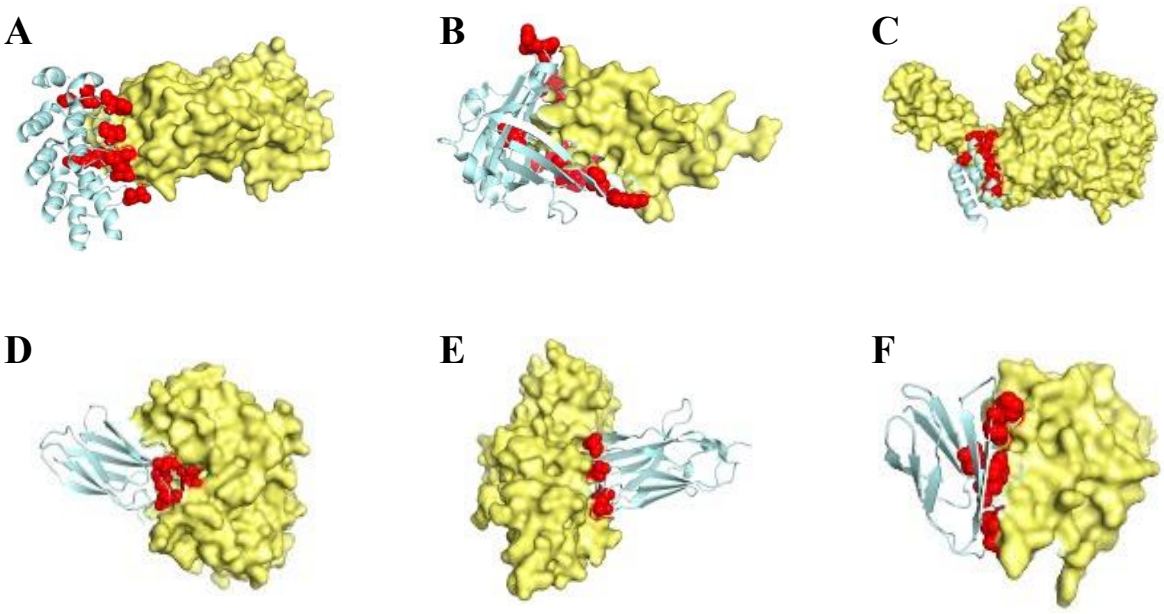

Figure 3: Non-immunoglobulin binding proteins and recognition targets. The blue color indicates the binding protein, the red color indicates the binding site and the red color indicates the target protein. **A:** DARPins bound to maltose binding protein (1SVX) **B:** Anticalins bound to the extracellular domain of human cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) (3BX7). **C:** Affibody bound to human epithelial growth factor 2 (HER2) (3MZW). **D:** Monobody bound to Klebsiella pneumoniae FtsZ (8GZV). **E:** Nanobody bound to Plasmodium falciparium Pf12p (7KJI). **F:** Monobody bound to Abelson tyrosine kinase 1 (Abl1)-SH2 domain via the amino acid residues on the secondary structure (3UYO).

1.4 The generation of non-immunoglobulin protein binders

Combinatorial protein libraries are created to design affinity binders by diversifying carefully selected surface-exposed residues on the protein. The generated libraries are then subjected to a selection process based on the specific binding to a predefined target, followed by amplification and identification of target binders (Sha et al., 2017). Different molecular selection systems for the identification of target binders have been described (Gronwall and Stahl, 2009; Binz et al., 2005; Simeon and Chen, 2018), and they have been categorized into cell-dependent display systems, cell-free display systems, and the non-display systems. The cell-dependent display systems include the phage display, bacterial surface display, and yeast

display. The cell-free display includes the ribosomal display and mRNA display, among others. The non-display systems include yeast-two-hybrid and protein complementary assays. These selection systems link phenotype to genotype, enabling sequences of target binders to be deduced and have been carefully described in detail (Gronwall and Stahl, 2009; Binz et al., 2005; Simeon and Chen, 2018; Zimmermann et al., 2020).

However, the choice of which molecular selection system to use would depend on the purpose, target, scaffold protein architecture, size, and available resources due to the various merits and demerits faced by each selection system. The phage display is the most widely used molecular display method for the screening and identification of binders to several targets due to the less complex nature of the design and construction of display libraries when compared with the other selection systems. Moreover, the need for the elution of binders from targets increases the possibility of losing the most robust binders due to harsh elution conditions. Except for the phage display, cell surface display systems like bacterial surface display and yeast display offer the possibility for the use of fluorescence labeling and flow cytometry sorting, enabling the selection of affinity binders without the need for elution from the target protein. Nevertheless, cell-dependent display systems are faced with the limitation of the library size's dependency on the transformation efficiency of the DNA. Alternatively, cell-free display systems offer the possibility of creating libraries of up to 10^{13} because of the in vitro transcription and translation used in library construction. Additionally, co-expressing the libraries together with the target proteins in vivo is possible in cell-free display systems. It could allow target recognition based on the expression of a correctly folded protein by the host organism. One major drawback of cell-free display is the requirement for a protein that is small enough for in vitro transcription. Unlike the display systems, the non-display system allows for selection based on direct screening for growth survival. However, they are faced with the limitations of unspecific

intracellular interactions and the need for the scaffold proteins to be well-suited for intracellular folding for a successful outcome.

The various molecular selection methods offer significant benefits as they provide researchers with a varied range of alternatives for developing target binders for an extensive range of applications in protein science, fundamental research, drug development, and affinity reagents. The use of several protein frameworks, taking advantage of their distinct characteristics, could provide numerous unique uses while also satisfying our unmet needs.

1.5 The β-sandwich domain 1 (SD1) of islandisin as a candidate scaffold protein

In this study, I introduce SD1, a domain of a multimeric protein (islandisin) and a structural homolog of fervidolysin (Kim et al., 2004) (**Figure 4**), as a candidate scaffold protein. Islandisin is a thermostable protease derived from the hyperthermophilic bacterium *Fervidobacterium islandicum*. Islandisin is active in a wide range of temperatures and pH levels (60-80 ºC and 6.0-8.5, respectively), with an optimum temperature and pH of 80 ºC and 8.0. Furthermore, islandisin has shown excellent thermal stability for an extended period (32 hours) at 80 ºC. Islandisin has also exhibited resistance to a wide range of solvents and detergents (Godde et al., 2005). These features make beta-sandwich domain 1 (SD1) of islandisin an appealing framework for the design of target binding for applications requiring harsh environmental conditions.

Figure 4: The structure of fervidolysin (1R6V); islandisin structural homolog.

SD1 is a single domain protein, predominantly of β-sheets with no disulfide bond, folded into a stable 3D structure with stability provided by the numerous hydrophobic residues present within the protein core (**Figure 5**). SD1 is a small-sized protein with a molecular weight of 11.7 kDa, structurally similar to the variable domain of IgG. SD1 is a thermostable protein with a melting temperature (Tm) of 73 ºC (**Figure 6**) and is easily expressed in *E. coli* (Ononugbo et al., 2024).

Figure 5: The structural model of SD1

SD1 has a beta-barrel architecture made up of closely packed anti-parallel sheets, similar to the structural conformations of the domains of heavy and light chains in IgG (Chiu et al., 2019). Contrasting the IgG domain, the immunoglobulin fold of SD1 is mainly stabilized by a number of hydrophobic residues within the hydrophobic core, resulting in structural stability. Moreover, Unlike the IgG domains that are stabilized by disulfide bonds, the absence of disulfide bonds

in the stabilization of the IgG-like fold of SD1 allows the protein to be expressed in a reducing environment such as the cytoplasm of bacteria, decreasing the difficulty of expression and the cost of production associated with using the eukaryotic expression systems. Therefore, SD1's IgG-like structural architecture, non-complexity, size and the presence of extended exposed surface loops, and thermal stability, including the lack of disulfide bonds, among other features, make it a suitable scaffold protein for the design of several target affinity binders. SD1 could be the first scaffold protein with a framework like that of an IgG fold isolated from a bacterium, a hyperthermophile known as *Fervidobacterium islandicum*. Furthermore, SD1's simple framework, which allows for extensive structural modifications, and its structural robustness to tolerate many side chain amino acid substitutions aimed at enhancing its properties for desired targeted functions, makes it an intriguing protein framework.

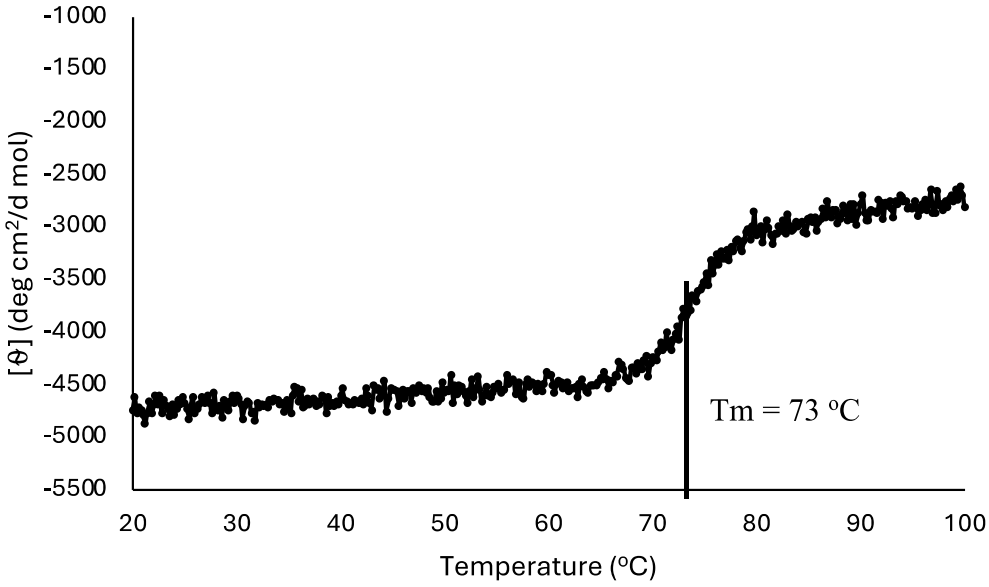

Figure 6: Thermal denaturation study of SD1. The denaturation curve was obtained by heating the sample from 20 °C to 100 °C at 1 °C/min at 222 nm. The Tm was determined at 73 °C.

Although SD1 is highly expressible in *E. coli*, it is expressed mainly in inclusion bodies, and the solubility of SD1 is necessary to serve as a framework for the design of binding functions.

Moreover, inclusion body formation is one of the significant challenges faced when a foreign gene is introduced into a host. This is because the new microenvironment may differ from the source in terms of osmolarity, pH, cofactors, and folding mechanisms (Rosano and Ceccarelli, 2014). Additionally, the presence of stretches of hydrophobic residues in the polypeptide chain, which may interact with similar residues at high expression levels due to high concentrations, might promote protein instability and aggregation, followed by the formation of inclusion bodies (Bhatwa et al., 2021, Rosano and Ceccarelli, 2014). To reduce the formation of protein inclusion bodies, several strategies have been discussed, including slowing down protein synthesis, inducing endogenous chaperone synthesis, introducing chemical chaperones and foldases into the culture medium, co-expression with chaperons, and modifying the protein in question (Bhatwa et al., 2021). The decision to use any of the strategies is based on the researcher's objective and suitability; in this study, because the goal is to improve SD1 solubility for target binding functions, modifying SD1 via amino acid substitutions with the goal of removing residues that contribute to inclusion body formation seem suitable. Moreover, a protein’s characteristics often depend on its primary sequence.

There are various methods for modifying a protein's primary sequence to improve desirable features, including directed evolution, semi-rational design, and rational design (Linse, 2018; Sachsenhauser and Bardwell, 2018). In directed evolution, the success of selecting variants with intended improved features is dependent on the efficacy of the designed screening system, as many mutants are typically generated. Though it offers the advantage of eliminating the need to know the structure of the protein of interest, it is a blind approach that relies solely on the screening system, making the selection process difficult and time-consuming. Semi-rational and rational design approaches may take less time. Still, they necessitate an in-depth knowledge of the protein's structure and the contributions of each of the 20 natural amino acids to protein folding and solubility. However, the considerable and ongoing progress in protein

research, including the development of new algorithms, increased computational resources, and characterization methodologies, has provided enhanced tools for protein solubility design that go beyond a shot in the dark.

1.6 Hypothesis and objectives

A scaffold made of thermophilic protein is renowned for its intrinsic sturdy thermostability, which is essential for creating new functions (Finch and Kim, 2018). SD1 has a structural framework similar to the IgG fold, with longer loops, and promises mutational robustness to varying side-chain modifications that could result in new functionalities. SD1 architecture and stability could be exploited to create new affinity tools with unique characteristics, thus expanding protein science research. However, the inclusion body formation of SD1 on expression in *E. coli* is a limiting factor for the design of affinity reagents. The following factors may contribute to the inclusion body formation of SD1. The protein's high level of expression in *E.coli*, coupled with the presence of surface-exposed hydrophobic residues or spans of hydrophobic regions within the protein's primary sequence, may interact during protein folding, resulting in instability and the formation of insoluble aggregates.

Therefore, the goal of this research is to increase SD1's soluble expression, folding yield, and solubility and design a target affinity reagent using SD1 as a structural framework. **In chapter two**, amino acid substitutions that could improve SD1 folding and solubility were studied, with a particular emphasis on how side chain modifications in the surface-exposed residues and aggregation-prone regions affect the protein solubility in vivo. **In chapter three**, the effect of synergistic interactions between SD1's inherent qualities, such as isoelectric point (pI), charge distributions, and thermal stability, as well as extrinsic factors, such as pH, on the protein solubility and storage stability, was studied. Then, the most suitable mutant for the design of target binders was decided based on the results obtained. **In chapter four**, the combinatorial

library of the improved soluble SD1 variant was constructed and evaluated for binding to TNF-α after verification of expression and monovalent display on the M13 phage pIII coat protein.

Chapter 2:

SD1 engineering for improved solubility and folding yield

2.1 Overview

In this chapter, I explored the effect of placing charged residues on the surface of SD1, including identifying stretches of hydrophobic regions and adding gatekeeper residues (GKRs) to reduce SD1 aggregation and inclusion body formation with the goal of increasing folding yield and solubility.

The number of aggregation-prone regions (APRs) in a protein sequence has been illustrated to be inversely proportional to the protein's solubility (Ganesan et al., 2016), implying that mutational suppression of such regions may increase the protein's solubility. Besides, in hydrophilic proteins, the surface hydrophilic residues and water molecules form hydrogen bonds while burying the hydrophobic residues, blocking their interactions with water molecules (Shaytan et al., 2009; Baumann et al., 1989), generally suggesting that solubility is determined by amino acid charge and polarity, as well as their ability to interact with surrounding water molecules. Charged residues increase the protein's net charge while also causing repulsive electrostatic interactions between similar charges. Allowing adequate time for proper protein folding, preventing aggregation, and the formation of the inclusion body (Lawrence et al., 2007; Paraskevopoulou and Falcone, 2018).

Therefore, this study hypothesizes that SDI surface charging and reduction of aggregation-prone regions would increase the soluble expression, folding yield, and solubility of the protein.

2.2 Materials and Methods

Bacterial strains: *Fervidobacterium islandicum*, a thermophilic bacterium, was obtained from the American Type Culture Collection (ATCC 49647, Manassas, USA) and grown in a complex medium (**Table 1**). The optimal growth temperature is 65 °C (Huber et al., 1990). Anaerobic incubation was carried out in a well-sealed vial at 60°C without shaking. Cloning and expression of the target genes were performed using E. *coli* JM109 and BL21(DE3), respectively. The *E. coli* cells were grown in LB broth (5g/l yeast extract, 10g/l tryptone, 5g/l NaCl) after the addition of 50 µg/ml ampicillin or kanamycin, the plasmid selection markers.

Table 1: ATCC Medium for Fervidobacterium islandicum

Solution 1	
NH_4Cl (g)	0.9
$MgCl_2$. $6H_2O$ (g)	0.2
KH_2PO_4 (g)	0.75
K_2HPO_4 (g)	1.5
Trace Element Solution (ml) (see below)	9.0
$FeSO_4$. $7H_2O$ (ml)	0.03
Resazurin (ml)	1.0
Wolfe's Vitamin Solution (ml) (see below)	5.0
Trypticase Peptone (g) (BD 211921)	10.0
Distilled water (ml)	850.0

Solution 2	
Yeast extract (g)	3.0
Distilled water (ml)	100.0

Solution 3	
Glucose (g)	5.0
Distilled water (ml)	50.0

Trace Element Solution:	
Nitrilotriacetic acid (g) (Adjusted to pH 6.5 with KOH)	12.5
$FeCl_3$. $4H_2O$ (g)	0.2
$MnCl_2$. $4H_2O$ (g)	0.1
$CoCl_2$. $6H_2O$ (g)	0.017
$CaCl_2$. $2H_2O$ (g)	0.1
$ZnCl_2$ (g)	0.1
$CuCl_2$ (g)	0.02
H_3BO_3 (g)	0.01
Na_2MoO_4. $2H_2O$ (g)	0.01
NaCl (g)	1.0
Na_2SeO_3 (g)	0.02
Distilled water (L)	1.0

Wolfe's Vitamin Solution:	
Biotin (mg)	2.0
Folic acid (mg)	2.0
Pyridoxine hydrochloride (mg)	10.0
Thiamine HCl (mg)	5.0
Riboflavin (mg)	5.0
Nicotinic acid (mg)	5.0
Calcium D- (+)-pantothenate (mg)	5.0
Vitamin B12 (mg)	0.1
p-Aminobenzoic acid (mg)	5.0
Thioctic acid (mg)	5.0
Distilled water (L)	1.0

Solution 1 and 2 were autoclaved separately in 20 ml vials (Malm, Osaka, Japan). Solution 3 was filter sterilized in a separate container, and all solutions were mixed in an anaerobic box after sterilization. 50 µl of 100 mM Na_2S solution was added to the mixed medium in the anaerobic box, allowed to stand for 1 hour, and when the color of the oxygen indicator (resazurin) changed, a rubber stopper was placed, and the mixture was sealed with an aluminum cap. Inoculation, dispensing, and collection of *Fervidobacterium islandicum* were performed using sterile syringes. After about a week of incubation, the medium began to become turbid, and the OD600 reached 1.0; the sample was subcultured twice.

Islandisin encoding gene extraction: The *Fervidobacterium islandicum* culture medium was centrifuged for 15 min at 15,000 rpm. The pellet was then resuspended in 15 µl of sterile MilliQ

water. An equal volume of Phenol: Chloroform: Isoamyl alcohol (PCI) at a ratio of 25:24:1 was added to the solution, vortexed vigorously, and centrifuged for 10 min at 15,000 rpm. The supernatant was added to a new microcentrifuge tube and was used as the genomic solution.

Islandisin encoding gene cloning: The islandisin encoding gene was cloned using the infusion cloning method. The DNA sequences of islandisin were retrieved from GenBank (accession number: AY190029). Oligonucleotides with sequences homologous to the cloning vector were designed, and the sequences were synthesized by Hokkaido System Science (Sapporo, Japan) (**Table 2**). The target genes were amplified using polymerase chain reaction (PCR), and the pET25b vector linearized for 2 h at 37 ºC with BamHI and NdeI restriction enzymes (TAKARA Bio Inc). Using the In-Fusion® HD cloning kit (TAKARA Bio Inc, Shiga, Japan), the linearized vector and amplified target gene were mixed and transformed with *E. coli* JM109. After overnight cultivation at 37 ºC, the colonies were analyzed by PCR using the Quick Taq HS dye mix (TOYOBO Co., LTD., Osaka, Japan). The positive clones with the correct insertion size were subcultured overnight with constant shaking in a 5 ml LB medium at 37 ºC. After cultivation for 16-20 h, the plasmid DNA from the samples was extracted (GenElute Plasmid Miniprep Kit, SIGMA Aldrich, St. Louis, MO), and sequences analyzed (Eurofins Scientific, Luxembourg) to validate the construction of the vector harboring SD1 coding fragment.

Table 2: Primer sequences used for cloning and site-directed mutagenesis.

		Primer sequence (5'→3')
Islandisin	Fwd	CAGGCCGCCCTGGGCGACTTGCCTTCAAGTGGA
	Rev	CGGAGCTCGAATTCTTATTCCGCTTTTCCAAAAAGC

		Primer sequence (5'→3')
SD1	Fwd	AACGAGCATATGTCAAGTGGAGGATTGGAC
	Rev	AAGTCAACTAGTCGATGCAGATGAGGAAAACC

		Primer sequence (5’→3’)
SD1_6K	Fwd.	AACGAGCATATGTCAAGTGGAGGATTGGAC
	Rev	TATTGTGGATCCTTACTTCTTCTTCTTCTTTCCTCCCGATGCAG ATGAGGAAAAC

		Primer sequence (5’→3’)
Y15K/W18R	Fwd	AAAAGTAGTCGGAGGGTGCCTTCCG
	Rev	AGCATCTGTGACGGTGAGTTG

		Primer sequence (5’→3’)
S106R	Fwd	CGATCTGCATCGTAAGGATCCG
	Rev	GGAAAACCTTACCGCATTGTTACG

		Primer sequence (5’→3’)
S2D	Fwd	GATGGAGGATTGGACTATCAAC
	Rev	TGACATATGGCTGCCGCG

		Primer sequence (5’→3’)
N73D	Fwd	GACGGATTAACGAGGGTAGCTTTC
	Rev	TGAGTTCACTTTCGTGTCTGGAC

		Primer sequence (5’→3’)
S107E	Fwd	GAAGCATCGTAAGGATCCGAATTC
	Rev	TGAGGAAAACCTTACCGCATTG

		Primer sequence (5’→3’)
S26K	Fwd	AAACTTCTTGGAATTTCAAGTAC
	Rev	GACACTAACGGAAGGCACC

		Primer sequence (5’→3’)
Q91P	Fwd	CCAGCGGCTTTGGTTGATAACCG
	Rev	GAAAATCACCGTTCGTTCCTCTGC

		Primer sequence (5’→3’)
I30R	Fwd	AGATCAAGTACTGGTCGCAATG
	Rev	TCCAAGAAGTGAGACACTAACG

		Primer sequence (5’→3’)
I30D	Fwd	GATTCAAGTACTGGTCGCAATG
	Rev	TCCAAGAAGTGAGACACTAACG

		Primer sequence (5'→3')
T87E	Fwd	GAGGTGATTTTCCAAGCGGC
	Rev	TCGTTCCTCTGCTTTTCTGAAAGC

		Primer sequence (5'→3')
A92R	Fwd	AGGGCTTTGGTTGATAACCG
	Rev	TTGGAAAATCACCGTTCGTTCC

		Primer sequence (5'→3')
S26L	Fwd	CTACTTCTTGGAATTTCAAGTAC
	Rev	GACACTAACGGAAGGCACC

		Primer sequence (5'→3')
D60Y	Fwd	TATGTAATTGTCAGTGGTCCAG
	Rev	GTACCTACCGGAATCTATACCG

		Primer sequence (5'→3')
D60Y	Fwd	TATTTCCAAGCGGCTTTGGTTG
	Rev	CACCGTTCGTTCCTCTGCTTTTC

		Primer sequence (5'→3')
A92K	Fwd	AAGGCTTTGGTTGATAACCG
	Rev	TTGGAAAATCACCGTTCGTTC

		Primer sequence (5'→3')
A93K	Fwd	AAATTGGTTGATAACCGTAAC
	Rev	CGCTTGGAAAATCACCGTTC

		Primer sequence (5'→3')
His10	Fwd	GTGATGATGATGATGATGGCTGC
	Rev	CATCACCATCACGGTGGAGGCGGTAGCAGCGGCCTGGTG

Construction of SD1 expression system: The SD1 gene was amplified from the pET25b vector, which carried the islandisin-encoding gene, and cloned into the multiple cloning sites of pET28a vector (pET28aSD1) using NdeI and BamHI restriction enzyme (TAKARA Bio Inc., Shiga, Japan). The pET28a vector contains a T7 promoter upstream of the sequence encoding hexahistidine (His6) and a kanamycin resistance marker. pET28aSD1 was transformed with

chemically competent *E. coli* BL21 (DE3) for SD1 expression and cultured on LB plates at 37 ^{o}C overnight. Single colonies of the overnight culture were analyzed using PCR for the presence of insertions, and positive clones were grown in an LB broth at 37 ^{o}C overnight with continuous shaking. Plasmid DNA from the liquid *E. coli* cultures was extracted, and sequences were analyzed by DNA sequencing (Eurofins Scientific, Luxembourg) using the T7 promoter and the T7 terminator primers to confirm correct SD1 sequence insertion.

Construction SD1 solubility enhancing peptide tag: Inverse PCR was used to add additional residues of hexa-lysine amino acids to the C-terminus of SD1 with designed primers harboring the additional sequences. Just before the hexa-lysin, two glycine residues were added to limit the interaction of the tag with SD1. The entire pET28aSD1 construct was amplified using the Q5 high fidelity 2X master mix (New England Biolabs, Ipswich, Massachusetts, United States), following the recommended reaction conditions. Single colonies of the constructs were analyzed using PCR, and the extracted plasmids of positive clones were verified for correct insertions by DNA sequencing.

Site-directed mutagenesis: Mutagenic primers with the desired mutation(s) (**Table 2**) were used to substitute specific amino acid residues in SD1. This was carried out by inverse PCR using the Q5 high fidelity 2X master mix (New England Biolabs, Ipswich, Massachusetts, United States) and following the recommended reaction conditions. The template DNA was removed from the PCR product by DpnI digestion (TAKARA Bio Inc, Shiga, Japan). The DpnI digested PCR product was ligated and transformed with *E. coli* BL21 (DE3) cells. Single colonies of the constructs were evaluated by PCR, and the recovered plasmids of positive clones were sequenced to ensure accurate insertions.

Multiple sequence alignment: To search for amino acids analogous to SD1, the default parameters of the UniProt BLAST (https://www.uniprot.org/blast) service (Target database =

UniProtKB, E-threshold = 10, Matrix = Auto-BLOSUM62, filter = none, gapped = yes, hits = 250, HSPs per hit = all), NCBI (blastp), and TEMPURA database (http://togodb.org/db/tempura (Sato et al., 2020)) were employed. All sequences were aligned and analyzed using ClustalO from the Unipro UGENE program (Okonechnikov et al., 2012) after removing similar sequences.

In silico mutagenesis: The solvent-accessible residues on SD1 were identified and substituted to charged amino acids (Arg (R), Lys (K), Glu (E), and Asp (D)). Mutations with the lowest predictions of the FoldX's ΔΔG values (Schymkowitz et al., 2005) for both the acidic and basic charged residues were considered stabilizing and were selected. To identify mutational hotspots in SD1, saturation mutagenesis on all SD1 residues was performed. All 20 amino acids were substituted at each position in SD1, and residues with the lowest energy values were deemed beneficial.

Identification of aggregation-prone regions and reduction of SD1 aggregation: The aggregation-prone regions (APRs) in SD1 were identified (Fernandez-Escamilla et al., 2004), and all were substituted to gatekeeper residues (GKRs) (R, K, E, D, and P). The most suitable residues on each position were selected based on the effect on the thermodynamic stability of SD1 based on the estimations of the FoldX's ΔΔG. Structural visualization was achieved using Yasara (Land and Humble, 2018).

Recombinant protein expression: LB medium supplemented with 50 ug/ml kanamycin was used to cultivate recombinant *E. coli* cells with constant shaking overnight at 37 ºC. An aliquot of the overnight culture was transferred into a fresh LB medium supplemented with kanamycin and grown at 37 ºC with constant shaking. When the cells' turbidity had reached an optical density at 600 nm (OD600) of 0.5 and 0.8, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to induce the expression of SD1. The incubation temperature was then reduced to 30 ºC

and cultured for about 18 h with continuous shaking. The *E. coli* culture was centrifuged at 12,000 xg for 2 mins at room temperature to harvest the cells. The cells were disrupted by sonication in a lysis buffer (20 mM Tris-HCl (pH 7.5)) and centrifuged at 4 ºC for 20 mins at 12,000 xg to separate the cell fractions. The supernatant became the soluble fraction, while the pellet became the insoluble fraction.

Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE): The separated protein samples were boiled for 5 mins in sample buffer (250 mM Tris-HCl (pH 6.8), 20% 2-mercaptoethanol, 8% SDS, 20% sucrose, and 0.2% bromophenol blue (BPB)), and separated on a 12% SDS-PAGE and a prestained triple-color protein ladder (Nacalai Tesque, Inc., Kyoto, Japan), was used as a molecular weight standard. The PAGE was stained with Coomassie Brilliant Blue (CBB) and destained with 5 % EtOH and 7 % acetic acid before visualization.

In silico predictions: The structure of the mutants was predicted with Alphafold2 (Jumper et al., 2021; Mirdita et al., 2022) using the UCSF ChimeraX graphical user interface (GUI) (Pettersen et al., 2021). The interacting pairs of amino-acid residues were identified using the Proteus online tool (Barroso et al., 2021). To create and display the structural model of SD1 showing the mutated residues that contribute to SD1 solubility, Pymol version 2.5.2 was used (https://www.pymol.org/). Mutations were incorporated into the model using the Pymol's mutagenesis wizard.

2.3 Results

2.3.1 Expression of SD1 in *E. coli.*

After SD1 induction, sonication, and separation, the protein was found mainly in the pellet fraction identified as aggregates in *E. coli* inclusion bodies (**Figure 7**). This is a limitation for use in the design of binding functions.

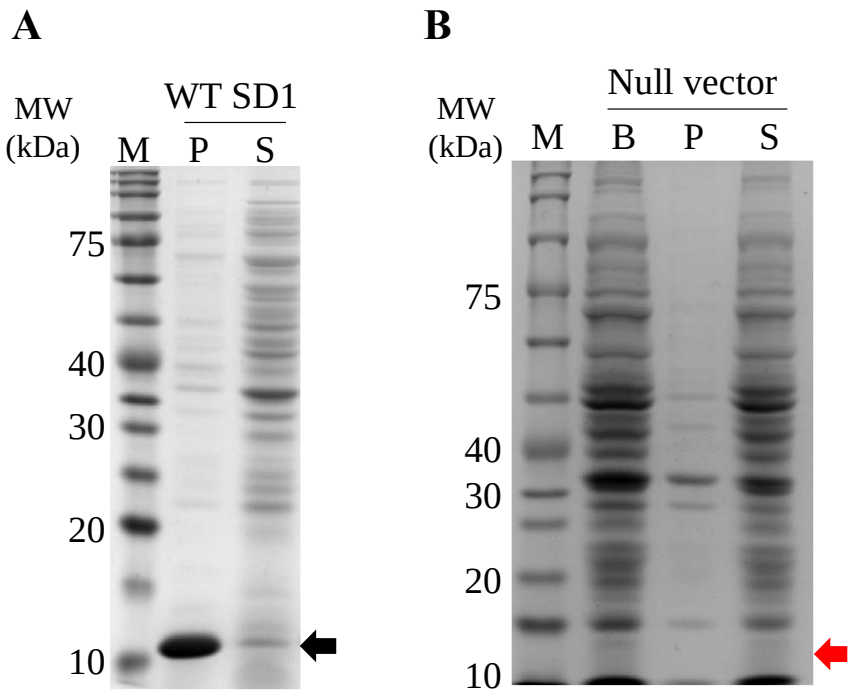

Figure 7: **A)** Expression of SD1 in *E. coli* BL21 (DE3). **B)** Expression of null pET28a vector (No SD1 insertion) in *E. coli* BL21 (DE3). Samples were separated by 12% SDS-PAGE. Target protein molecular weight (MW) = 11.7 kDa. M = marker, B = before cell lysis, P = pellet, S = supernatant. WT SD1 = wild type SD1. The black arrow indicates the target band. The red arrow indicates the expected target size in the case of SD1 insertion (no recombinant protein expression).

2.3.2 Expression of SD1 with a small peptide tag

The effect of a positively charged short peptide tag (hexa-lysin) on SD1 solubility was studied based on previous reports on the solubility-enhancing properties. Hexa-lysin was chosen to increase the pI of SD1 due to its charge similarity and to introduce intra and intermolecular electrostatic repulsive interactions. The C-terminal hexa-lysin tagged SD1 (SD1_6K) on expression in *E. coli* was significantly found in the supernatant fraction on SDS-PAGE (**Figure 8**), suggesting that the addition of the charged peptides significantly improved the solubility of the protein.

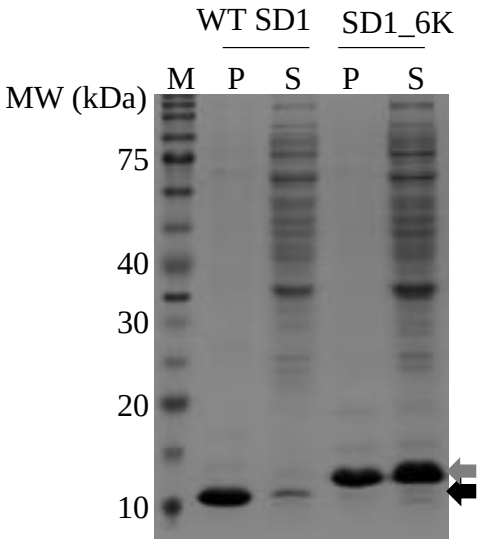

Figure 8: The effect of hexalysin fusion at the SD1 C-terminus on the soluble expression and folding yield. Samples were separated by 12% SDS-PAGE. Target protein molecular weight (MW) = 11.7 kDa. M = marker, P = pellet, S = supernatant. WT SD1 = wild type SD1. SD1_6K = hexalysin tagged SD1. The black arrow indicates the SD1 target band. The grey arrow indicates the SD1_6K target band.

2.3.3 In silico identification of mutational targets

Given that charged residues attached to the SD1 terminal could enhance the protein solubility, the overall protein charge distribution throughout the surface of SD1 was investigated. As a result, in silico methods were used to find appropriate sites for substituting solubility-enhancing residues while maintaining SD1 structural stability. Mutational hotspots were identified as those where the side chain substitutions were stabilizing ($\Delta\Delta G < 0$), considering that residues with the lowest energy values are desirable. The following positions, S22, S26, S64, N71, T76, N99, and A101, could allow for substitutions of 15 or more residues without affecting the structural integrity of SD1 based on the FoldX $\Delta\Delta G$.

2.3.3.1 Multi-sequence alignment

A total of 223 sequences made up of 49, 69, and 85 sequences were recovered from three databases, namely, the NCBI (blastp), UniProt BLAST, and Database of growth TEMperatures

of Usual and RAre prokaryotes (TEMPURA) (Sato et al., 2020), respectively. After removing identical sequences, a total of 154 homologous sequences were aligned with the SD1 sequence. Poorly conserved residues were mainly observed (**Figure 9**), making the selection of residues that could improve the solubility of SD1 more like random selection.

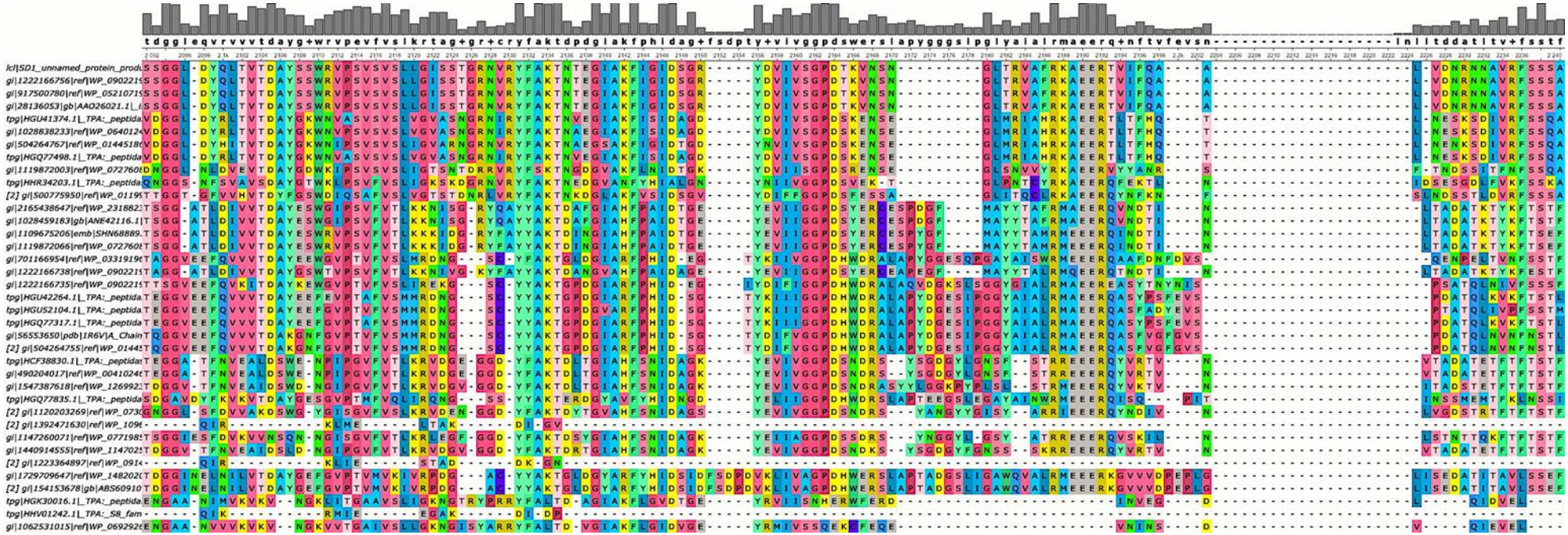

Figure 9: Multisequence alignment of SD1 with about 154 homologous sequences. Homologous sequences were recovered from NCBI (blastp), UniProt BLAST, and TEMPURA databases.

2.3.3.2 SD1 surface charging

The selected 50% solvent-accessible residues on SD1 were substituted to charged residues with the highest potential to increase stability at several locations in SD1 based on FoldX's predicted ΔΔG (**Table 3**). Mutants were assigned three sites for the basic and acidic residues, with ΔΔG values of < -0.6 and < -0.3, respectively. The residues on the surface-exposed residues were substituted with charged residues, and a Y15K/W18R/S106R combination, which displayed the highest stabilizing mutations for the positively charged surface charging was selected. For the negatively charged surface charging, the surface-exposed residues were substituted with basic amino acids, and an S2D/N73D/S107E combination, which had the highest stabilizing mutations, was selected.

Table 3: Calculated energy change by FoldX (ΔΔG) in Kcal/mol of the 50% surface exposed residue on SD1

AA position on SD1 sequence	AA seq.	Arg (R)	Lys (K)	Glu (E)	Asp (D)
001	S	-0.00	-0.25	-0.12	0.03^{+}
002	S	-0.18	0.09^{+}	0.23^{+}	-0.35**
015	Y	-0.57	-0.68*	0.16^{+}	0.43^{+}
018	W	-1.19*	-0.32	0.02^{+}	0.29^{+}
030	I	-0.14	-0.16	-0.02	-0.03
071	N	-0.13	-0.30	-0.21	-0.11
073	N	-0.11	-0.40	-0.28	-0.61**
075	L	0.29^{+}	0.40^{+}	0.87^{+}	0.85^{+}
091	Q	-0.01	-0.09	0.65^{+}	0.94^{+}
106	S	-0.61*	-0.51	-0.06	0.12^{+}
107	S	-0.56	-0.20	-0.40**	0.01^{+}
109	S	-0.12	-0.09	0.42^{+}	0.44^{+}
		Y15K/W18R/S106R		S2D/N73D/S107E	

AA = amino acid
* = Most stabilizing mutations on SD1 for basic AA
** = Most stabilizing mutations on SD1 for acidic AA
+ = Destabilizing mutations on SD1

2.3.3.3 Reduction of SD1 aggregation via the aggregation-prone regions

Two APRs were identified in SD1 (**Figure 10**), each having 8 and 9 residues of amino acid, making a total of 17 residues. All 17 residues were substituted to GKRs (P, R, K, E, and D) by in silico saturation mutagenesis. The predicted contributions of each of the gatekeeper residues to the reduction of aggregation (Tango) and thermodynamic stability (FoldX ΔΔG) of SDI were noted and shown in a mass plot (**Figure 11**). Tango is an algorithm for analyzing a protein's intrinsic aggregation tendency (Fernandez-Escamilla et al., 2004), and the changes in Tango caused by mutating each residue in the APR to GKR were shown as ΔTango. Additionally, The effect of each amino acid mutation on the protein's thermodynamic stability was calculated using the Foldx change in free energy (ΔΔG in Kcal/mol) (Schymkowitz et al., 2005). **Figure 11** makes it easier to identify substitutions with substantial negative values on both axes that could significantly enhance SD1 solubility by minimizing aggregation while increasing or maintaining SD1's structural integrity and thermodynamic stability. The results show that only a few residues could allow for the introduction of the GKRs, as many could negatively affect the stability of SD1 (**Figure 11**). Single mutations (S26K, Q91P, A92K, and A93K) were investigated for their effect on SD1 soluble expression. These residues were selected based on the predicted effect on increasing the thermodynamic stability of SD1 while reducing aggregation (S26K) and and the significant reduction reduction of the aggregation propensity of SD1 while not adversely affecting the thermodynamic stability of the protein (Q91P, A92K, and A93K). Each mutant was constructed and expressed in *E. coli* to confirm their solubility, but little or no expression was detected in the soluble fraction (**Figure 12**). Therefore, combination mutation was decided, and the positioning of charged residues on the APR flanks was investigated in SD1. The combinations were selected based on the following observations. I30R is a surface-exposed residue, so altering isoleucine (I) to arginine (R) or aspartic acid (D) at that site may increase SD1 soluble expression. The T87E and A92R mutations were

introduced because these positions are located near the edge of the beta-sheet that surrounds the APR. Based on the consideration that adding charged residues within the beta-sheet could jeopardize SD1 structural stability. When the combined mutations I30R/T87E/A92R and I30D/T87E/A92R were introduced into SD1, the structure did not read as aggregation-prone by Tango prediction. Furthermore, based on the saturation mutation scanning of SD1, S26 was identified as a hotspot and within the residues that make up the APR of SD1 as predicted by the Tango algorithm (Fernandez-Escamilla et al., 2004). Although S26 is positioned in the middle of SD1's beta-sheet, it was selected because of its hotspot status and the significant ΔΔG value of S26K generated by FoldX. For Q91P, this position was selected since Tango results reveal that this mutation contributes the most to the considerable reduction of SD1 aggregation. When the S26K/Q91P combination was inserted, the structure was no longer classified as aggregation-prone, according to the Tango prediction.

Figure 9: The structural model of SD1 showing the aggregation-prone regions (APR) in red. These regions consist of eight and nine stretches of amino acid residues respectively. Yasara was used for structural visualization.

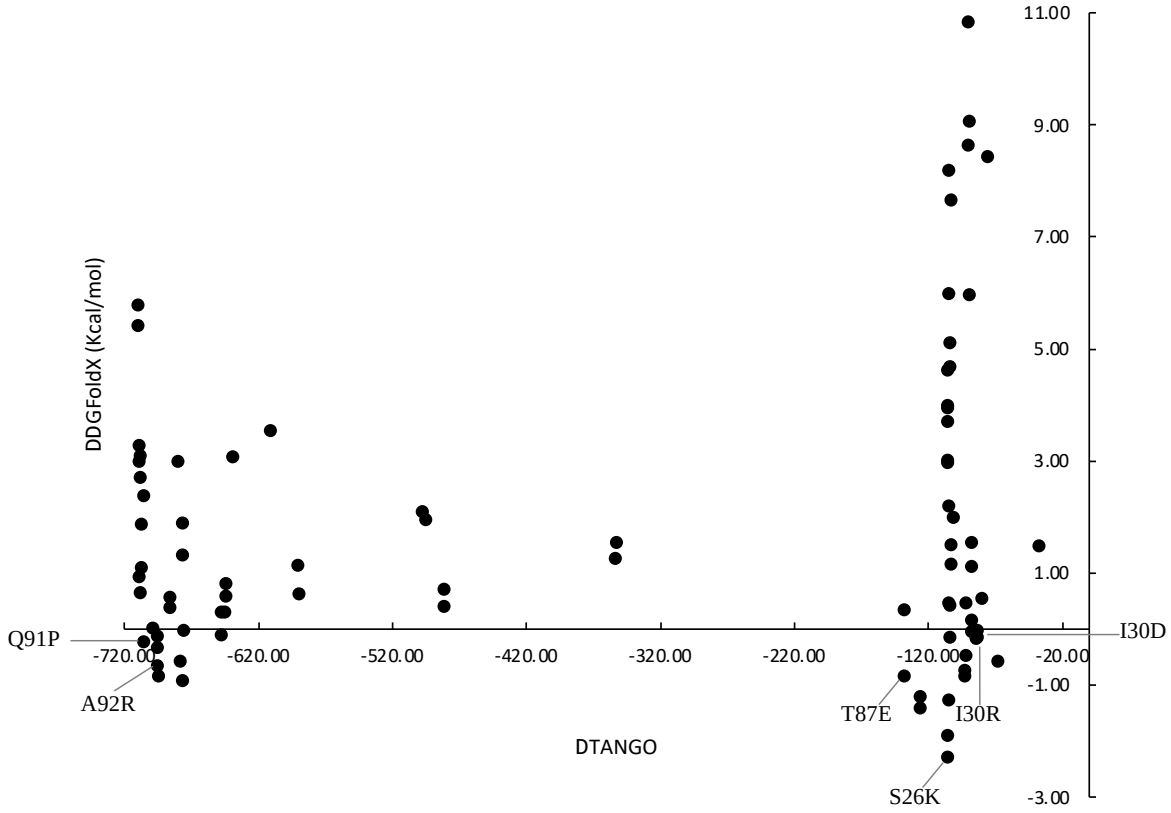

Figure 10: Scatter plot of the findings of computational gatekeeper residues scans for each of SD1's aggregation-prone regions. Aggregation gatekeeper residues were added to the 17 residues identified as long stretches of hydrophobic residues on SD1 using in-silico saturation mutagenesis with the Solubis suite, which combines Tango and FoldX. The Y-axis depicts the effect of each mutation on the protein's thermodynamic stability, while the X-axis represents the influence of each mutation on aggregation. Increasing negative values have a positive influence on thermodynamic stability and reduction of protein aggregation.

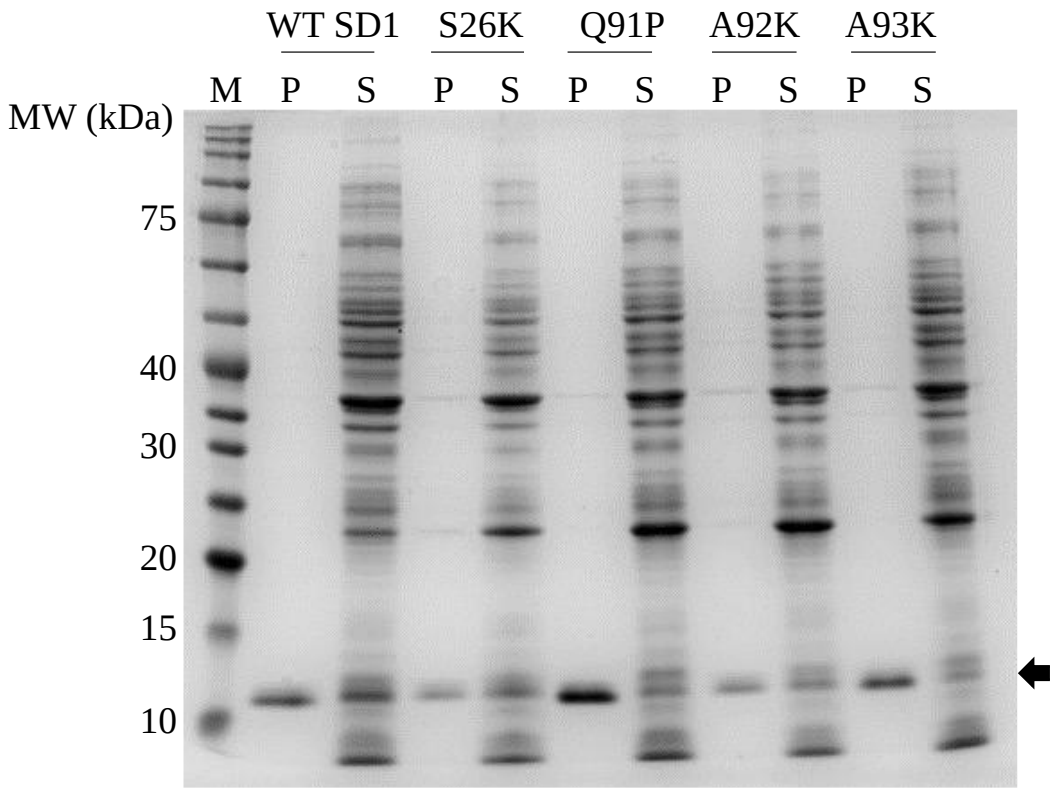

Figure 11: Investigations on the effect of single amino acid substitutions on the soluble expression of SD1. Samples were separated by 12% SDS-PAGE. Molecular weight: 11.7 kDa. MW = molecular weight. M = marker, P = pellet, S = supernatant. The black arrow indicates the expected size.

2.3.3.4 Stabilization of SD1 by introducing hydrophobic residues

Considering S26 is a hotspot, and S26L provided a considerably more stabilizing mutation (ΔΔG = -2.93 Kcal/mol), this mutation was hypothesized to provide much more stability. Also, the side chains are much simpler and can fit into specific areas during the SD1 folding process than S26K. Thus, I expected that the S26L mutation would increase stability. I added an S26L mutation into the surface-charged and APR reduction mutants, S2D/S26L/N73D/S107E and S26L/130D/T87E/A92R, respectively. Furthermore, tyrosine residues known to enhance IgG stability via the tyrosine corner were added to the sequence using the proteus algorithm to identify pairs of interacting partners. The D60Y/I89Y mutant pair, identified as interacting

pairs of amino-acid residues by the Proteus online tool with a ΔΔG value of -0.884, was considered and incorporated into one of the mutants (S2D/S26L/D60Y/N73D/I89Y/S107E).

2.3.4 Evaluation of SD1 soluble expression and folding yield

Compared to the WT SDI, which was predominantly present as insoluble aggregates in the pellet fraction on SDS-PAGE, all mutants show a significant increase in soluble expression (**Figure 13**). Both the SD1 surface charging and residue substitutions aimed at reducing the APR significantly increased the soluble expression of the protein. The folding yield was more than 80% compared to the WT SD1. The addition of hydrophobic residues and tyrosine showed a decrease in folding yield when compared to the other mutants (**Figure 14**). The folding yield of the SD1_6K was significantly higher than other mutants (**Figure 14**).

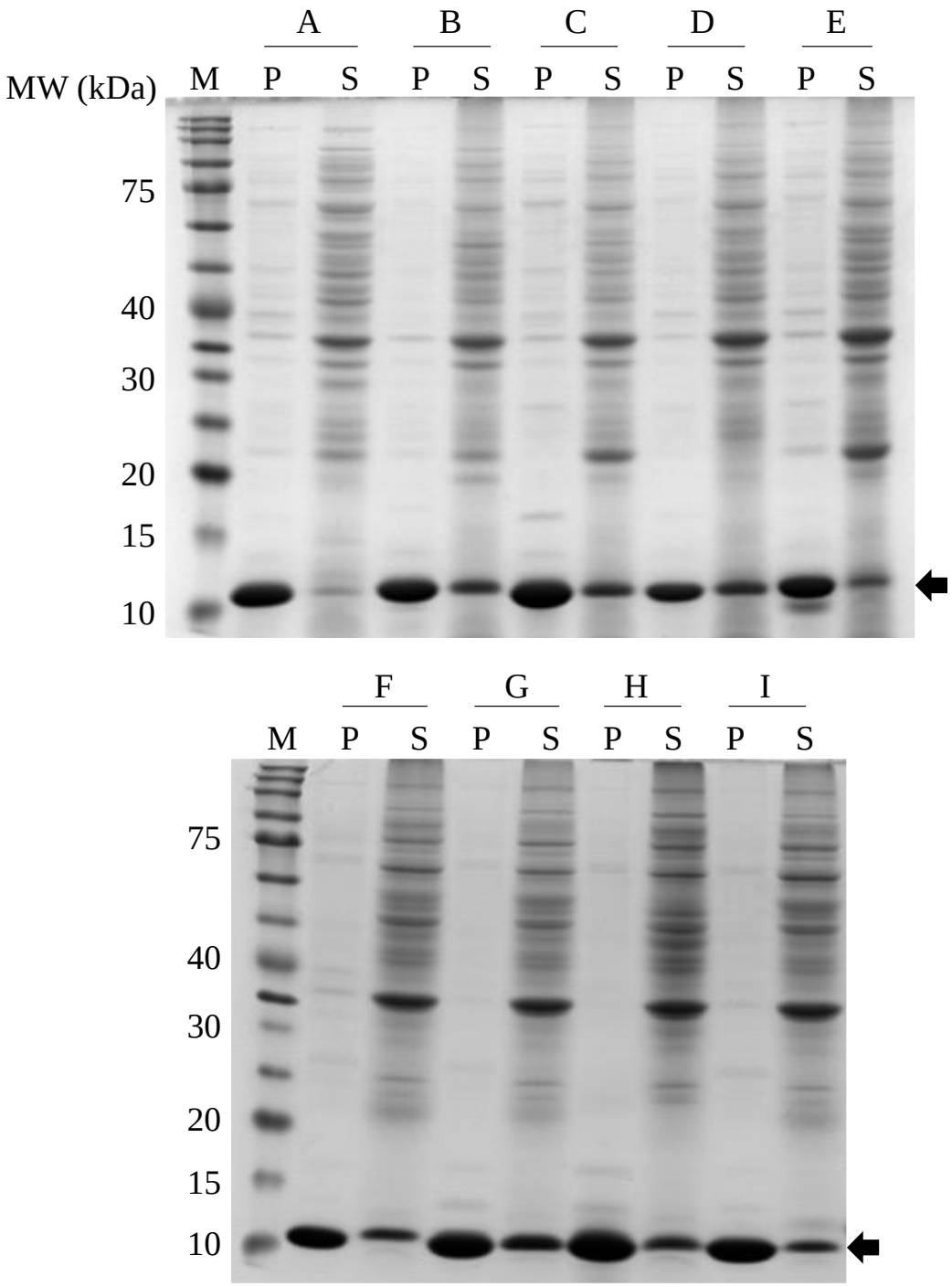

Figure 12: Evaluation of SD1 soluble expression and folding yield. Samples were separated by 12% SDS-PAGE. Target protein molecular weight = 11.7 kDa. MW = molecular weight. M = marker, P = pellet, S = supernatant. A =WT SD1, B = Y15K/W18R/S106R, C = S2D/N73D/S107E, D = S26K/Q91P, E = I30R/T87E/A92R, F = I30D/T87E/A92R, G = S2D/S26L/N73D/S107E, H = S2D/S26L/D60Y/N73D/I89Y/S107E, I = S26L/I30D/T87E/A92R. The black arrow indicates the target size.

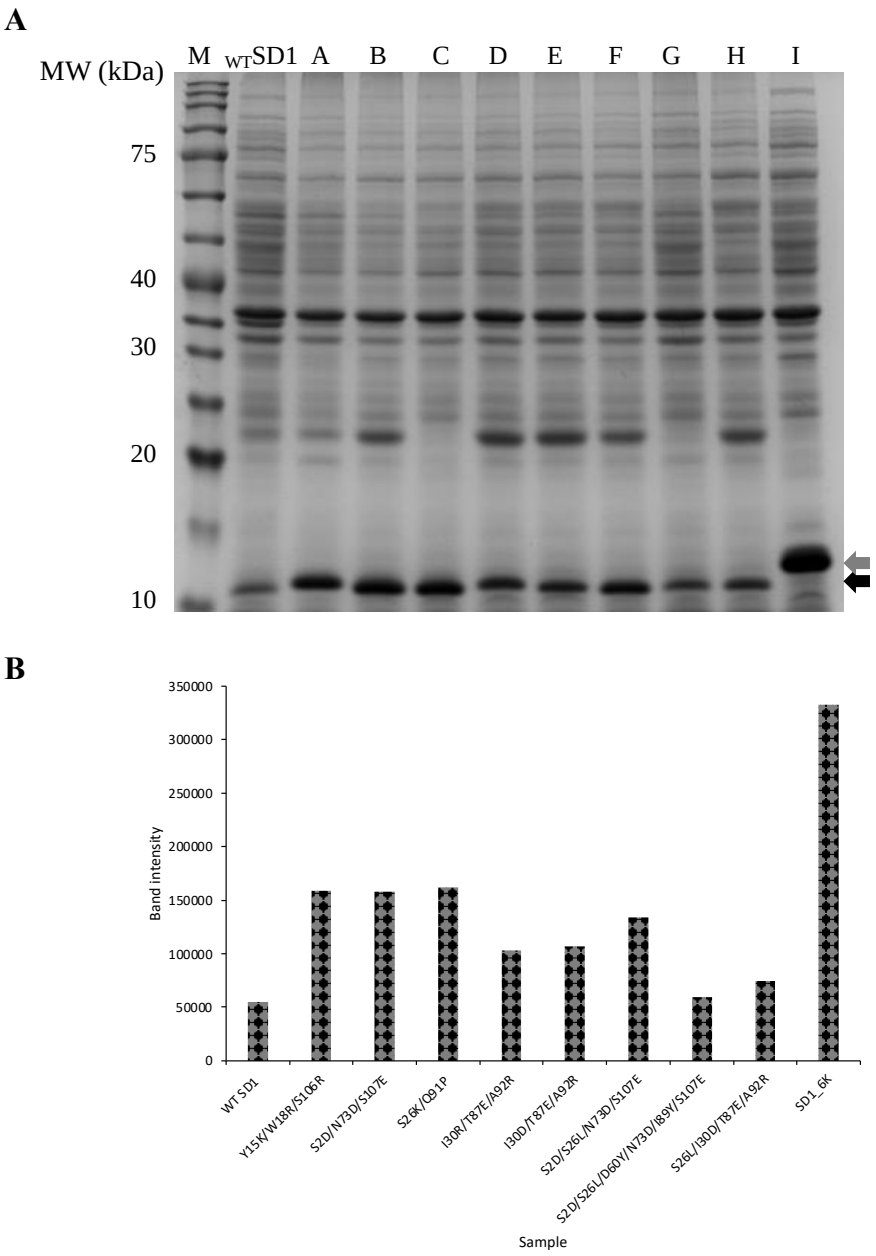

Figure 13: Evaluation of soluble expression and folding yield of WT SD1 and mutants. Supernatant fractions were separated by 12% SDS-PAGE. Target protein molecular weight = 11.7 kDa. MW = molecular weight. **A)** Analysis of supernatant fractions of the WT SD1 and mutants. Lane 1 = Marker, lane 2 = wild type (WT) SD1, lane A = Y15K/W18R/S106R, lane B = S2D/N73D/S107E, lane C = S26K/Q91P, lane D = I30R/T87E/A92R, lane E = I30D/T87E/A92R, lane F = S2D/S26L/N73D/S107E, lane G = S2D/S26L/D60Y/N73D/I89Y/S107E, lane H = S26L/I30D/T87E/A92R and lane I = SD1_6K mutants respectively. The black arrow indicates the target size. The grey arrow indicates the target size for SD1_6K. **B)** The band intensity of the target bands in **Figure A** was measured using CSAnalyser4.

2.4 Discussion

SD1 structural architecture and biochemical properties provide a suitable framework for the design of affinity reagents; the formation of insoluble aggregates in *E. coli* inclusion bodies during expression necessitated the improvement of the protein solubility in order to achieve excellent target affinity and specificity during the design and selection of target binders. SD1 solubility improvement was carried out by single surface residue substitutions and reduction of regions with the protein sequence that are prone to aggregation. Considering the factors that influence protein solubility, such as high protein concentration, the composition of the primary amino acid sequence, particularly the presence of long hydrophobic stretches, protein average net charge, defined by the pI at a given pH, and so on (Paraskevopoulou and Falcone, 2018, Rosano and Ceccarelli, 2014, Bhatwa et al., 2021), inspired the decisions of which residues and at what positions to substitute for the charged groups. The fusion of hexa-lysin to SD1 was explored to understand how the increase in the protein charges and pI could influence the solubility. The high soluble expression following induction of SD1_6K suggested that charged residues could enable appropriate folding of SD1 by providing electrostatic repulsion and disrupting the interactions of exposed hydrophobic groups during folding, allowing enough time for complete protein folding. Encouraged by the results of SD1_6K expression, the effect of charged groups on SDI solubility and stability was studied further.

In this chapter, the effect of SD1 surface charging and reduction of aggregation using the GKRs was investigated. Leveraging the developed in silico tools for the identification of surface residues (Guex and Peitsch, 1997), aggregation-prone regions (Fernanandez-Escamilla et al., 2004; Van Durme et al., 2016), and the effect of aggregation on the thermodynamic stability of proteins (Schymkowitz et al., 2005), different locations suitable for substitutions aimed at improving the

solubility of SD1 were identified and analyzed. Firstly, homologous sequences from multiple databases were analyzed to select mutation sites based on sequence conservation, but poorly conserved residues were generally observed. Therefore, the supercharging technique developed by Lawrence et al. (2007), based on the concept that proteins are least soluble at their isoelectric points, where the net charge is zero, was explored. Two mutation combinations (Y15K/W18R/S106R and S2D/N73D/S107E) designed by SD1 surface charging showed significantly improved soluble expression in *E. coli*. The result suggests that in addition to increasing the pI and protein net charge, surface charging of SD1 may have improved the protein's solubility by introducing electrostatic repulsive forces during the protein folding, preventing the interaction of transiently exposed hydrophobic side chains and allowing the protein to fold completely (Paraskevopoulou and Falcone, 2018).

Similarly, because the number of APRs in a protein sequence has a direct correlation to the protein's insoluble state (Ganesan et al., 2015), the SD1 sequence was searched for the presence of APR, and two APRs were detected. GKRs were incorporated into the APRs to increase SD1 solubility through the reduction of the APRs. GKRs are known to impede protein aggregation through charge repulsion (R, K, D, E), the entropic cost of aggregation due to lengthy and flexible side chains (R, K), or incompatibility with the aggregate β-structure due to the lack of a hydrocarbon in its primary chain (P) (Beerten et al., 2012, Sant'Anna et al., 2014). All mutants designed based on reducing the APRs in SD1 showed significantly higher soluble expression when compared with the wild type. The I30R/T87E/A92R and I30D/T87E/A92R mutants' solubility could be due to the introduced electrostatic repulsive forces by the charged residues limiting the interaction between transiently exposed hydrophobic residues allowing complete protein folding, similar to the protein surface charging approach. However, for S26K/Q91P, although on the introduction of the

combined mutation, the protein structure was no longer classified as aggregation-prone, according to the Tango prediction, likely, the substitution of Gln (Q) to Pro (P) might cause instability in SD1. Given that proline residues in a polypeptide chain have strictly restricted ϕ values of -60^{o} and no main chain amide hydrogen (Prajapati et al., 2007). Nonetheless, the FoldX's predicted ΔΔG revealed that this mutation had no effect on the thermodynamic stability of SD1 and is present in the variable loop region, thus the consideration for designing the mutant. The observed increase in soluble expression by the S26K/Q91P mutant could be attributed to a number of factors. These could be resulting from the combined contributions of the charged residue, and proline in the loop region may have caused a kink and rigidity, slowing folding kinetics and allowing partially folded conformations that are prone to aggregation to fold fully (Glatzova et al., 2021). In addition, similar to disulfide bond formation, proline residues have been reported to stabilize protein structure by lowering the conformational entropy of the denatured phase (Prajapati et al., 2007).

To summarize, the major obstacle to protein applications at concentrations higher than their natural levels is usually the formation of insoluble aggregates as *E. coli* inclusion bodies. This could be due to the microenvironment being different from the natural source of the protein. SD1's soluble expression in *E. coli* was enhanced by exploring approaches aimed at increasing the surface charge residues and suppressing aggregation by the introduction of gatekeeper residues, which resulted in more soluble proteins than the WT SD1.

Chapter 3:

Characterization of SD1 mutants for selection of suitable variant for the design of target binding

3.1 Overview

The previous chapter evaluated the impact of increasing surface-charged residues and reducing aggregation-prone regions in SD1 on the protein's soluble expression and folding yield. However, it is necessary to understand the influence of the protein intrinsic factors (pI, surface-exposed residues, charges) on solubility and stability to select suitable variants for the design of affinity reagents. Given that high protein concentrations in solutions are required for a range of applications, engineered proteins must remain folded under certain conditions while still performing their intended functions (Atsavapranee et al., 2021). Further, absolute solubility, which is defined as the maximum amount of proteins that can be dissolved in solvents (Qing et al., 2022), is imperative. Therefore, it is necessary to ascertain the absolute maximum solubility of WT SD1 and mutants by evaluating solubility under severe parameters such as pH, high-concentration denaturants, and others (Qing et al., 2022).

In this chapter, I investigated how changes in SD1's intrinsic properties, as well as the effects of surface charge distribution, pI, and electrostatics, affect protein conformation in solution (solubility) and stability under different pH and temperature levels. Understanding how different residue modifications in the SD1 sequence affect folding is vital for the anticipated binding applications. Furthermore, the combined effects of the fractions of surface polar, non-polar, and charged residues in SD1 and other physicochemical parameters on protein solubility and storage stability were investigated, taking into account the possibility that environmental circumstances such as pH (Bai and Warshel, 2019, Kougentakis et al., 2020), ionic strength (Huang et al., 2013, Bavishi et al., 2018), and molecule crowding (Dhar et al., 2010, Adams et al., 2019) could disrupt the equilibrium of protein populations. To provide relevant information on which mutants could remain folded under various environmental conditions in order to select the best variant for target

binder design, simplify protein physicochemical properties, and improve understanding of how positional amino acids affect protein solubility.

3.2 Materials and Methods

Production of recombinant proteins: To produce the proteins, the histidine tag in the pET28a vector was extended from His6 to His10 tagged proteins by adding four extra histidine and four glycine residues just before the vector's thrombin cleavage site. Following construction, the sequence was confirmed using DNA sequencing. *E. coli* BL21 (DE3) cells expressing the N-terminal His10 tagged WT SD1 and mutants were grown overnight at 37 ºC with continuous shaking. Then, aliquotes of the culture were transferred to a fresh LB medium and grown at 37 ºC until OD_{600} had reached between 0.5 and 0.8. The proteins were produced by induction with 0.1 mM IPTG and cultivated at 15 ºC for 36 h with continuous shaking. The culture was centrifuged at 12,000 xg for 2 min at 25 ºC to harvest the bacterial cells. The harvested cells were lysed in a lysis buffer (20 mM NaH_2PO_4, 150 mM NaCl, 50 mg/ml lysozyme, 1 tablet of protease inhibitor cocktail, pH 7.4) and sonication (50% amplitude, 60% duty cycle on ice using 6 mm probe for 20 - 30 min). The lysed cells were separated into two fractions (supernatant and pellet) by centrifugation at 12,000 xg for 20 mins at 4 ºC. The different fractions were stored at -20 ºC until use.

Purification of recombinant proteins: The lysed cell supernatant was precipitated (80% $(NH_4)_2SO_4$), reconstituted (20 mM NaH_2PO_4, 150 mM NaCl, pH 7.4), and dialyzed (20 mM NaH_2PO_4, 150 mM NaCl, pH 7.4). The recombinant proteins were purified with AKTA pure (Affinity chromatography) (GE Healthcare, Chicago, Illinois, United States). Before applying the sample, the column was equilibrated with 10 column volumes (CV) of equilibration buffer (20 mM NaH_2PO_4, 150 mM NaCl, pH 7.4). The samples were filtered through a 0.22 μM syringe-

driven filter before being applied to a 5 ml HisTrap HP column (Cytiva) prepacked with agarose beads charged with Ni^{2+} ions used to isolate recombinant proteins from the supernatant. The proteins were eluted with a 0–500 mM imidazole gradient after washing with 5 x CV of buffer (20 mM NaH_2PO_4, 150 mM NaCl, pH 7.4). Several fractions were collected on a 96-well fraction collection plate. Protein concentrations were measured using a BCA protein assay kit (Thermo Scientific, Waltham, Massachusetts, USA) on a 96-well microplate reader in line with the recommended procedure. The absorbance was measured at 560 nm with a PerkinElmer Espire™ 2300 multilabel plate reader. To create a standard curve, a BSA concentration range of 0-2000 µg/ml and a four-parameter logistic curve fit was used.

Lyophilization of purified proteins: The purified proteins were dialyzed, transferred to 15 ml polypropylene centrifuge tubes, and frozen overnight at -80 ºC. Lyophilization of the frozen samples was carried out on a benchtop Eyela FDU-1200 using the recommended setup procedure. The sample containers were covered with perforated parafilm, and the lyophilization condition was kept at 10 Pa and -47 ºC throughout the process. After freeze drying, the samples were weighed and stored in 1.5 ml microcentrifuge tubes at 4 ºC until needed.

Measurement of protein solubility: To determine the solubility of SD1 and mutants, the highest protein concentration in a supersaturated solution (Kato et al., 2007) was used. Solubility of the proteins was measured at 25 ºC, at pH 4.7 (50 mM acetate buffer), and 7.7 (50 mM Tris-HCl buffer). About 20 µL of the buffer solution was added to the lyophilized proteins and centrifuged at 17,000 xg for 30 min at 25 °C. The protein concentration in the supernatant was measured using a NanoDrop spectrometer (Thermo Scientific, Waltham, Massachusetts, United States) at 280 nm absorbance. The supernatant fraction was combined with another vial of lyophilized protein (500–10 mg), centrifuged, and the concentration was measured. This process was repeated until the

protein samples began to form precipitates in the solution, and the concentration began to decrease, indicating that maximal solubility was achieved. To assess storage stability and aggregation kinetics of the proteins, sample solutions were equilibrated at 25 °C for varying times (20 min, 6 h, 12 h, 24 h, 36 h, and 48 h) using the amorphous precipitation method (Khan et al., 2013) with 1.3 M ammonium sulfate. The sample solutions with varying initial concentrations were combined in equal volume with 2.6 M ammonium sulfate in 50 mM acetate buffer and 50 mM Tris-HCl at pH 4.7 and 7.7, respectively, and incubated for varying equilibration times at 25 ºC. Following incubation, the protein samples were centrifuged at 25 ºC for 30 mins at 17,000 xg. The protein concentration in the supernatant was measured using a NanoDrop spectrometer at an absorbance of 280 nm. To maintain constant equilibration throughout the process, separate stock solutions of 50 mM buffer and 2.6 M ammonium sulfate were prepared as previously described (Khan et al., 2013).

Measurement of thermal stability: The thermodynamic stability of the purified WT SD1 and mutants was measured using the circular dichroism (CD) spectrometer. Jasco 725 spectropolarimeter, Jasco PTC-348WI thermal control device set to 20 °C and a quartz cuvette with a 2 mm path length was used to acquire the CD spectra of all protein samples (0.1-0.2 mg/ml). The samples were heated from 20 °C to 100ºC at 1°C/min at 222 nm, and the observed signal changes during the heating process were recorded. The temperature at which half of the protein molecules unfold was recorded as the melting temperature (Tm) of the protein.

In silico calculations: The protein properties were estimated using ExPASy's ProtParam tool, which is based on the Kyte-Doolittle hydropathy value for each amino acid (Walker et al., 2005). The total accessible surface area (ASA) of the protein samples was determined using the Jpred4

online tool (Drozdetskiy et al., 2015). The charge of the WT SD1 and mutants was calculated using the PROTEIN CALCULATOR v3.4 (https://protcalc.sourceforge.net/).

Calculation of the fraction of accessible surface areas (ASAs) of WT SD1 and mutants: After identifying the residues on the ASAs of the protein samples using the Jpred4 online tool (Drozdetskiy et al., 2015). The fractions of ASA were calculated as follows;

i. Fractions of surface polar residues = $\frac{\text{Total number of surface polar residues}}{\text{Total number of surface residues}}$

ii. Fractions of surface non-polar residues = $\frac{\text{Total number of surface non}-\text{polar residues}}{\text{Total number of surface residues}}$

iii. Fractions of surface charged residues = $\frac{\text{Total number of surface charged residues}}{\text{Total number of surface residues}}$

iv. Fractions of surface positively charged residues = $\frac{\text{Total number of surface positively charged residues}}{\text{Total number of surface residues}}$

v. Fractions of surface negatively charged residues = $\frac{\text{Total number of surface negatively charged residues}}{\text{Total number of surface residues}}$

3.3 Results

3.3.1 Solubility of WT SD1 and mutants

The protein samples were freeze-dried before making a saturation solution and measuring the highest concentration in the solution, which was termed the solubility of the protein samples. To ensure that the proteins had remained folded after lyophilization and throughout the solubility experiments, the folding spectra of all samples in the buffer solutions were analyzed. All mutants had similar fold spectra, except for I30R/T87E/A92R and S2D/N73D/S107E, which may have had a slightly changed secondary structure following lyophilization (**Figure 15**). Precipitation occurred at concentrations above supersaturation, showing that the samples had reached their solubility limits.

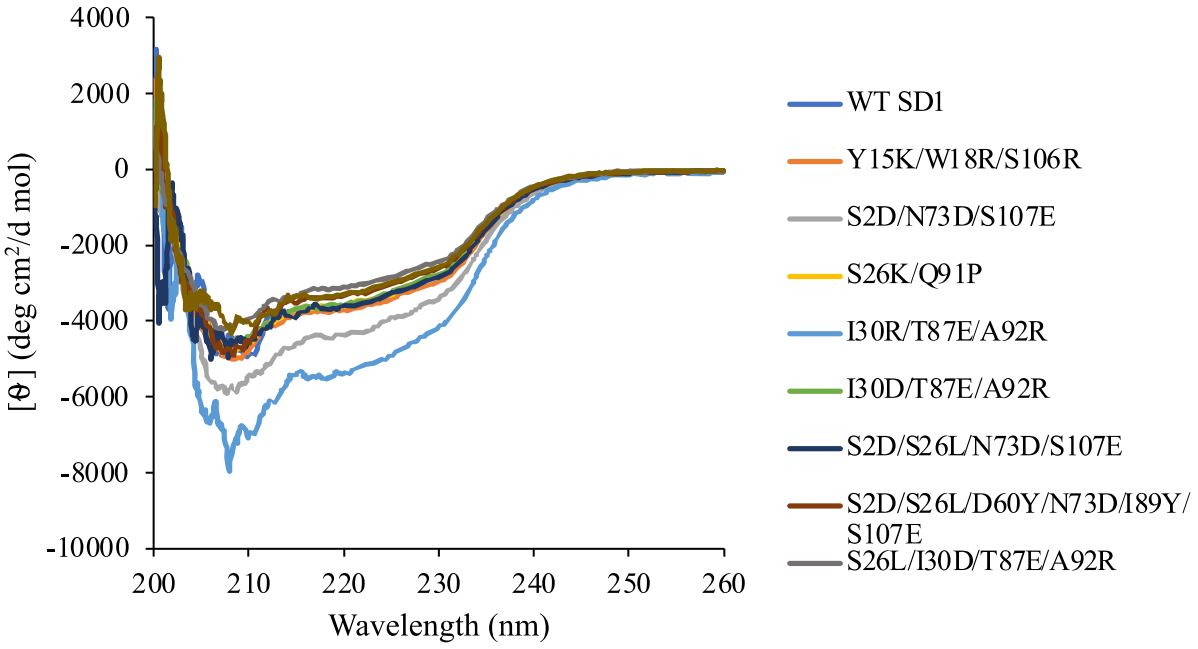

Figure 14: Circular dichroism spectra of WT SD1 and mutants after lyophilization. The spectra were obtained at 20 ºC using a quartz cuvette with a path length of 2 mm.

At pH 4.7 and 7.7, the solubility of the WT SD1 was 6.33 and 5.93 mg/ml, respectively (**Figure 16, Table 4**). For all the mutants, solubility was much higher than that of the WT SD1. The hexa-lysin tagged SD1 (SD1-6K) showed a very significant soluble expression and folding yield on SDS-PAGE; however, the solubility of the protein in vitro was not significantly higher than the WT SD1 in vitro. The approaches based on SD1 surface charging and the reduction of APRs through the insertion of GKRs significantly increased the solubility of the proteins at both pH measurements. SD1 surface charging with Y15K/W18R/S106R significantly increased the protein solubility to 60.3% at pH 4.7 and 97.4% at pH 7.7. Surface charging with S2D/N73D/S107E significantly enhanced the solubility of the protein to 90% at pH 4.7 and 84.1% at pH 7.7. Additionally, significant solubility improvements were observed for mutations aimed at reducing the APRs in SD1. S26K/Q91P mutant had about 182.3% increase at pH 4.7 and 191.9% increase at pH 7.7 when compared to the WT SD1. I30D/T87E/A92R mutant had an increase of about 176.1% at pH 4.7 and 184.6% at pH 7.7 when compared to the WT protein. Meanwhile, the I30R/T87E/A92R mutant showed a significant increase of 83.3% at pH 7.7 and 117.8% at pH 7.7 when compared to the WT SD1 (**Figure 16, Table 4**). The addition of hydrophobic residue (S26L) showed varying solubility improvement in different mutants. For S2D/S26L/N73D/S107E, an improvement in the protein solubility was observed, whereas for I30D/T87E/A92R, the solubility of the protein decreased. The tyrosine pairs (D60Y and I89Y) introduced in the mutant (S2D/S26L/D60Y/N73D/I89Y/S107E) decreased the solubility of the protein.

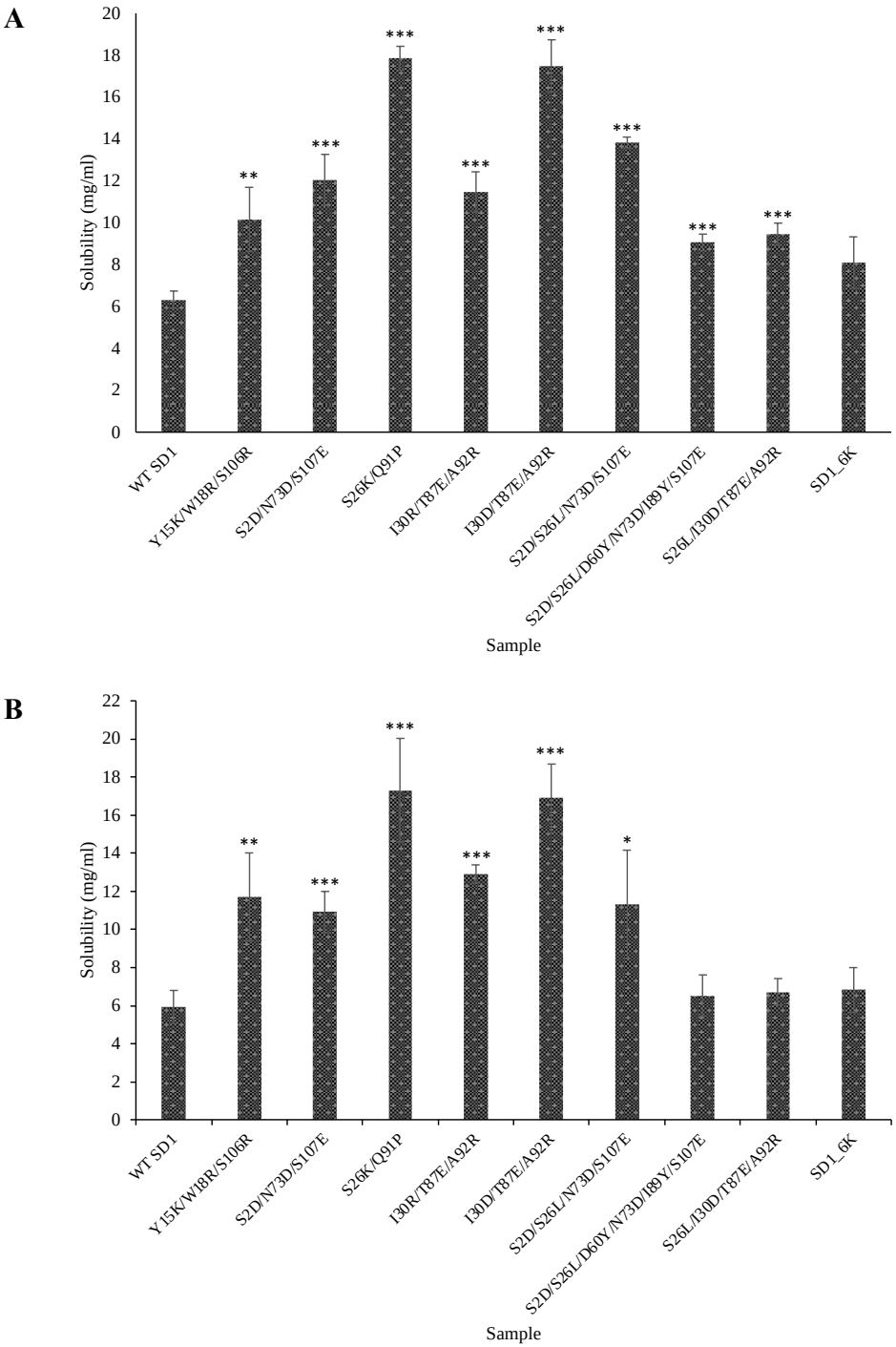

Figure 15: The effect of amino acid substitution on the solubility of SD1. Each solubility measurement was performed three times, and the error bars represent the standard deviation. *P ≤ 0.05, ** P ≤ 0.01, *** P ≤ 0.001. **A)** The solubility of SD1 was measured at 25 ºC in 50 mM acetate pH 4.7. **B)** The solubility of SD1 was measured at 25 ºC in 50 mM Tris-HCl pH 7.7.

Table 4: Solubility values of SD1 and mutants

Mutants	pH 4.7 Charge[a]	Protein solubility (mg/ml)[b]	AIC[c]	LS[c]	pH 7.7 Charge[a]	Protein solubility (mg/ml)[b]	AIC[c]	LS[c]
WT SD1	7.0	6.33 ± 0.41	2.70 ± 0.06	2.40 ± 0.09	3.6	5.93 ± 0.88	1.94 ± 0.07	1.70 ± 0.03
Y15K/W18R/S106R	10.0	10.15 ± 1.52	2.97 ± 0.07	2.95 ± 0.03	6.6	11.71 ± 2.29	4.10 ± 0.09	3.89 ± 0.20
S2D/N73D/S107E	5.0	12.03 ± 1.23	4.27 ± 0.41	3.85 ± 0.35	0.6	10.92 ± 1.09	3.69 ± 0.18	3.15 ± 0.09
S26K/Q91P	8.0	17.87 ± 0.57	4.12 ± 0.12	3.57 ± 0.10	4.6	17.31 ± 2.73	4.51 ± 0.17	4.19 ± 0.06
I30R/T87E/A92R	8.4	11.48 ± 0.97	4.80 ± 0.13	4.55 ± 0.06	4.6	12.92 ± 0.49	4.79 ± 0.42	3.29 ± 0.19
I30D/T87E/A92R	6.7	17.48 ± 1.25	4.76 ± 0.14	4.70 ± 0.14	2.6	16.88 ± 1.79	4.85 ± 0.01	4.04 ± 0.21
S2D/S26L/N73D/S107E	5.0	13.82 ± 0.26	5.92 ± 0.14	5.88 ± 0.32	0.6	11.32 ± 2.85	3.71 ± 0.13	3.17 ± 0.08
S2D/S26L/D60Y/N73D /I89Y/S107E	5.7	9.06 ± 0.40	3.34 ± 0.13	2.85 ± 0.29	1.6	6.52 ± 1.09	3.41 ± 0.12	2.03 ± 0.71
S26L/I30D/T87E/A92R	6.7	9.45 ± 0.54	3.06 ± 0.10	2.88 ± 0.11	2.6	6.72 ± 0.70	3.24 ± 0.05	2.79 ± 0.16
SD1_6K	13.0	8.11 ± 1.24	3.03 ± 0.02	2.82 ± 0.05	9.6	6.82 ± 1.20	2.92 ± 0.09	2.61 ± 0.10

[a]Calculated using PROTEIN CALCULATOR v3.4 (https://protcalc.sourceforge.net/)

[b]Solubility was measured as the maximum supernatant concentration of a supersaturated protein solution at 25°C in 50 mM acetate buffer (pH 4.7) and 50 mM Tris-HCl (pH 7.7), respectively. All values are presented as an average of three replicate experiments. The standard deviation represents error.

[c]The values were determined in the presence of 1.3 M ammonium sulfate in 50 mM acetate buffer (pH 4.7) and 50 mM Tris-HCl (pH 7.7). All data are the average of three replicate experiments. The standard deviation represents error. AIC is an abbreviation for aggregation initiation concentration, and LS is an abbreviation for long-term stability.

3.3.2 Storage stability of WT SD1 and mutants

The storage stability and aggregation kinetics of SD1 at 25°C were studied to evaluate how different residue modifications affected protein folding. Amorphous precipitation was induced in the protein samples using ammonium sulfate over a range of incubation periods to determine the storage stability of the proteins.

Protein solubility influences the aggregation start time in all the protein samples, and protein aggregation is proportional to total protein concentration (**Figures 17 and 18, Table 4**). Protein aggregation in solutions increased as protein concentrations approached saturation. The aggregation initiation concentrations (AIC) and the long-term solubility (LS) of all mutants were notably higher for all the mutants when compared with the WT SD1 at both pH values (**Figure 17 and 18, Table 4**). The AIC represents the minimal protein concentration required for aggregation to start (Khan et al., 2013). When compared to the WT SD1, which has an AIC of 2.70 mg/ml at pH 4.7 and 1.94 mg/ml at pH 7.7, the mutant AIC varied from 2.97 to 5.92 mg/ml at pH 4.7 and 2.92 to 4.85 mg/ml at pH 7.7 (**Table 4**). It was also observed that the protein concentration in the supernatant did not ultimately decrease after the 48 h incubation period, even for the WT SD1, suggesting the intrinsic stability of the protein.

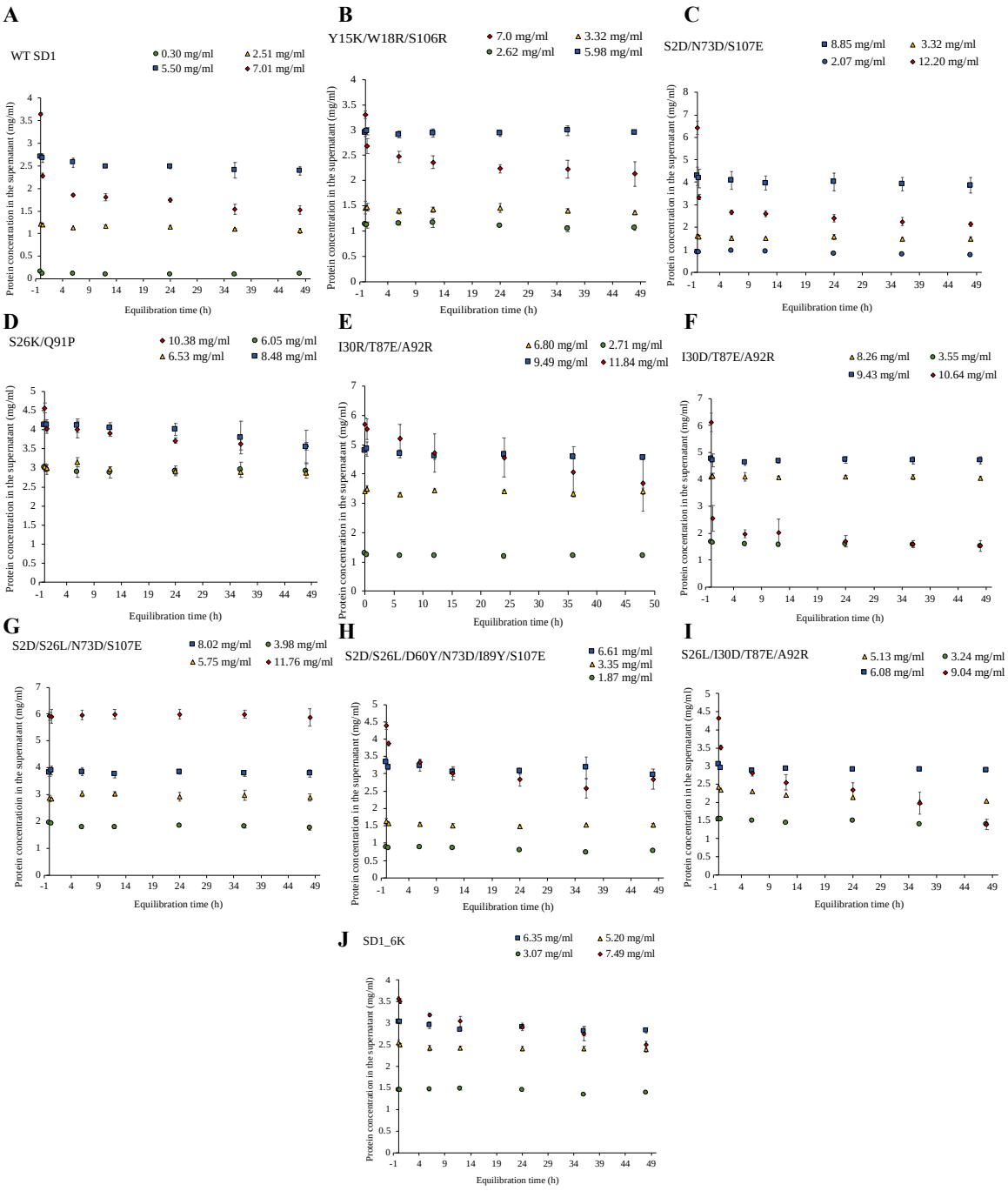

Figure 16: The storage stability and aggregation kinetics of WT SD1 and mutants at 25ºC, pH 4.7 (50 mM acetate buffer). Analyzed using the amorphous precipitation method with 1.3 M ammonium sulfate. The values listed in the legend were obtained prior to mixing with 2.6 M ammonium sulfate. Values at 0 h were established by mixing protein solutions with ammonium sulfate. A = wildtype SD1, B-J = Y15K/W18R/S106R, S2D/N73D/S107E, S26K/Q91P, I30R/T87E/A92R, I30D/T87E/A92R, S2D/S26L/N73D/S107E, S2D/S26L/D60Y/N73D/I89Y/S107E, S26L/I30D/T87E/A92R and SD1_6K mutants respectively.

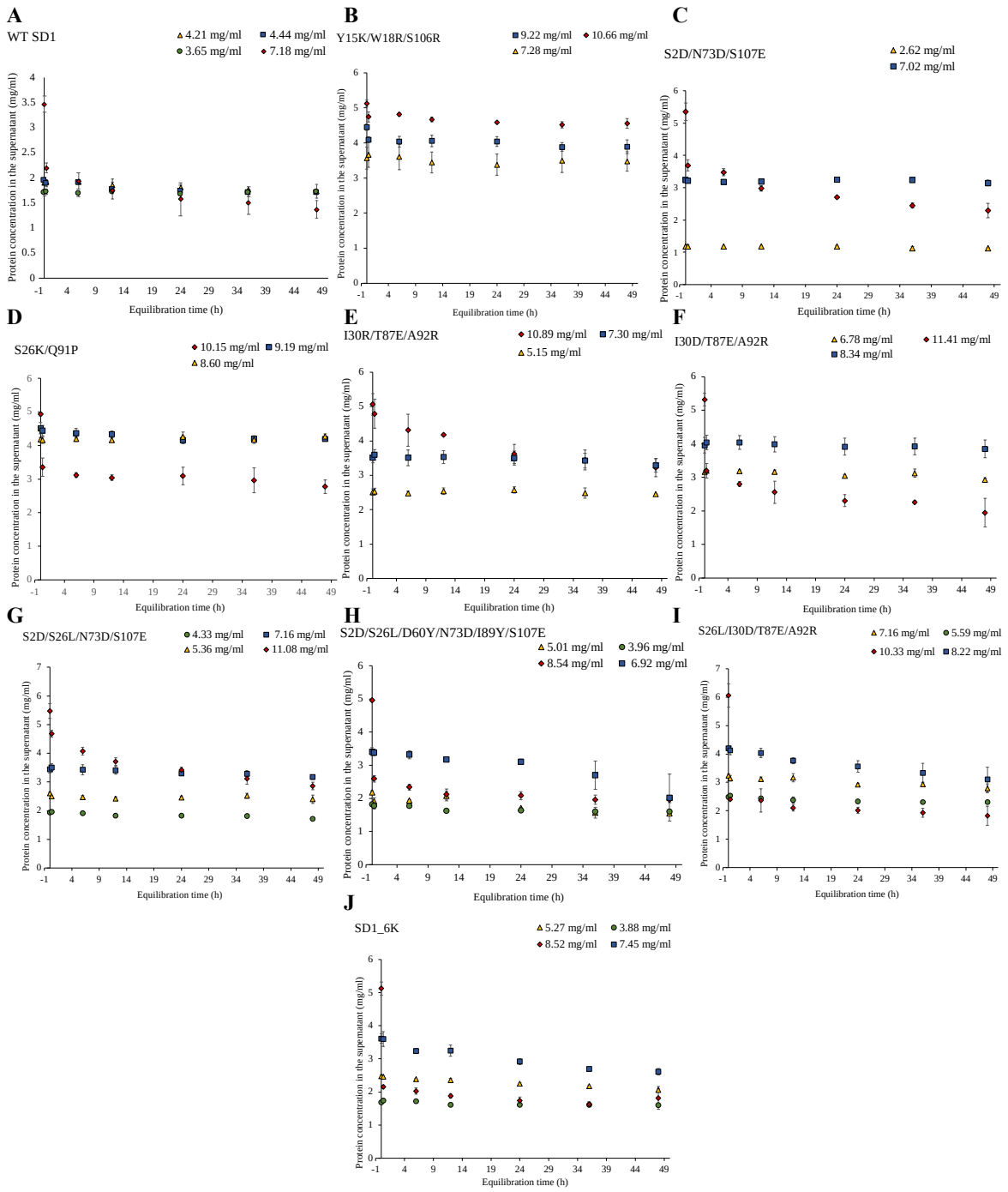

Figure 17: The storage stability and aggregation kinetics of WT SD1 and mutants at 25ºC, pH 7.7 (50 mM Tris-HCl buffer). Analyzed using the amorphous precipitation method with 1.3 M ammonium sulfate. The values listed in the legend were obtained prior to mixing with 2.6 M ammonium sulfate. Values at 0 h were established by mixing protein solutions with ammonium sulfate. A = wildtype SD1, B-J = Y15K/W18R/S106R, S2D/N73D/S107E, S26K/Q91P, I30R/T87E/A92R, I30D/T87E/A92R, S2D/S26L/N73D/S107E, S2D/S26L/D60Y/N73D/I89Y/S107E, S26L/I30D/T87E/A92R and SD1_6K mutants respectively.

3.3.3 The contributions of accessible surface residues to solubility

Protein charge distributions, number of charged and polar residues on the protein surface, pI, and pH in a given solution all affect protein solubility. Hence, their contributions to SD1 solubility improvements were investigated. The fractions of non-polar, polar, charged, positively, and negatively charged residues present on the accessible surface areas (ASAs) of WT SD1 and mutants were analyzed (**Table 5**). All the soluble mutants except for S26K/Q91P had a significant increase in the fraction of surface-charged residues when compared with the WT SD1. For the fraction of non-polar residues, a reduction in the fraction of surface non-polar residues was generally observed for most mutants, with the exception of S26K/Q91P, which had higher fractions of surface non-polar residues when compared with the WT SD1. The majority of mutants have a much lower fraction of surface polar residues than the WT SD1, with the exception of Y15K/W18R/S106R, S26K/Q91P, and S2D/S26L/D60Y/N73D/I89Y/S107E, which the fraction of surface polar residues was only marginally lower than the WT. All mutants have an uneven distribution of surface positively charged and negatively charged residues, while the WT SD1 has an equal amount of positively charged and negatively charged residues.

SD1's structural model is shown along with the substituted amino acid residues and the extended side chains (**Figure 19**). Surface residue to charged amino acid substitution shows as extended sticks sticking out from the structure to the surrounding. The majority of the designed mutations are in the variable loops of SD1, with the exception of S26K/Q91P, which had a single Ser to Lys substitution in the protein's β-sheet.

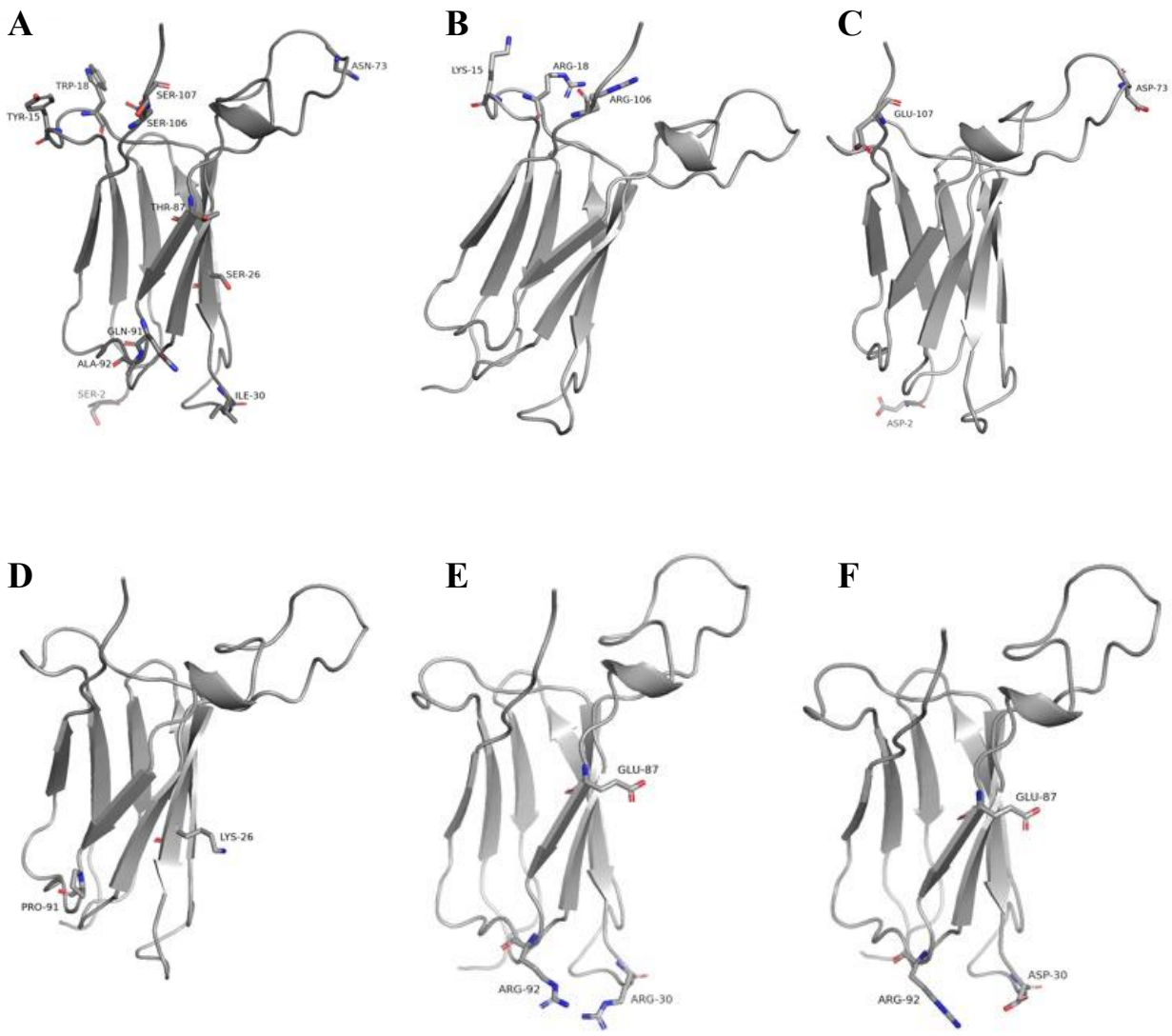

Figure 18: The structural model of SD1 and soluble mutants indicating mutated residues as sticks. Mutations were incorporated into the model using the mutagenesis wizard in Pymol. **A)** wild type SD1 showing mutated sites as sticks **B)** Y15K/W18R/S106R. **C)** S2D/N73D/S107E. **D)** S26K/Q91P. **E)** I30R/T87E/A92R. **F)** I30D/T87E/A92R.

Table 5: Physicochemical properties of SD1 and mutants

Mutants	Theoretical PI[a]	Aliphatic index[a]	GRAVY[a]	Fraction of ASA[b]					Stability Tm (°C)[c]	Positively charged residues[a]	Negatively charged residues[a]
				Non-polar	polar	Charged	Positive	negative			
WT SD1	9.77	83.12	-0.183	0.25	0.44	0.31	0.15	0.15	73.57 ± 0.08	13	9
Y15K/W18R/S106R	10.40	83.12	-0.274	0.20	0.43	0.37	0.24	0.12	73.23 ± 0.08	16	9
S2D/N73D/S107E	8.00	83.12	-0.233	0.25	0.38	0.36	0.20	0.16	73.47 ± 0.06	13	12
S26K/Q91P	9.89	83.12	-0.194	0.29	0.41	0.29	0.16	0.14	73.93 ± 0.08	14	9
I30R/T87E/A92R	9.98	78.62	-0.350	0.23	0.39	0.38	0.21	0.16	73.60 ± 0.12	15	10
I30D/T87E/A92R	9.51	78.62	-0.340	0.23	0.39	0.38	0.20	0.18	73.97 ± 0.12	14	11
S2D/S26L/N73D/S107E	8.00	86.70	-0.191	0.25	0.38	0.36	0.20	0.16	73.23 ± 0.11	13	12
S2D/S26L/D60Y/N73D /I89Y/S107E	9.04	83.12	-0.224	0.23	0.42	0.35	0.20	0.15	72.80 ± 0.19	13	11
S26L/I30D/T87E/A92R	9.51	82.20	-0.298	0.25	0.38	0.38	0.20	0.18	73.81 ± 3.1	14	11
SD1_6K	10.23	76.78	-0.358	0.20	0.39	0.41	0.27	0.14	72.51 ± 0.24	19	9

[a]Calculated using ProtParam tool (https://web.expasy.org/protparam/)

[b]Calculated using the Jpred 4 (http://www.compbio.dundee.ac.uk/jpred4/index.html). As fractions of total accessible surface area (ASA), non-polar, polar, charged, positively charged, and negatively charged residues were calculated.

[c]Thermal stabilities were determined by Circular Dichroism (CD). Tm stands for melting temperature.

3.3.4 The thermal stability of WT SD1 and mutants

The circular dichroism (CD) was used to study the structural conformations of WT SD1 and mutants on heat denaturation at varying temperatures. The spectra of the protein samples were analyzed in the UV range of 200 – 260 nm at 20 ºC. The samples displayed the characteristic spectra of β-sheet proteins), and most mutants showed an increase in helicity when compared with the WT (**Figure 20**). Additionally, the comparable spectral patterns for all samples indicate that the several mutations may not have adversely affected the primary structure of the protein.

The samples were heated from 20 ºC to 100 ºC at 1 ºC/min, and the signal changes upon heating at 222 nm were measured. Thermal-induced unfolding of the protein samples began at temperatures above 70 ºC, and all samples remained folded at temperatures below 70 ºC (**Figure 21**). This suggests that SD1 is a thermally stable protein, and the melting temperature (Tm) was determined at 73 ºC. The several amino acid substitutions in SD1 did not affect the thermal stability of the protein. After thermal denaturation, the protein samples did not refold when the temperature cooled to 20 ºC (**Figure 22**). Furthermore, no turbidity was seen upon heat denaturation, and the spectra were not identical to those of a random coil (**Figure 23**), indicating that the proteins may have retained some secondary structure and remained partially folded at 100 ºC.

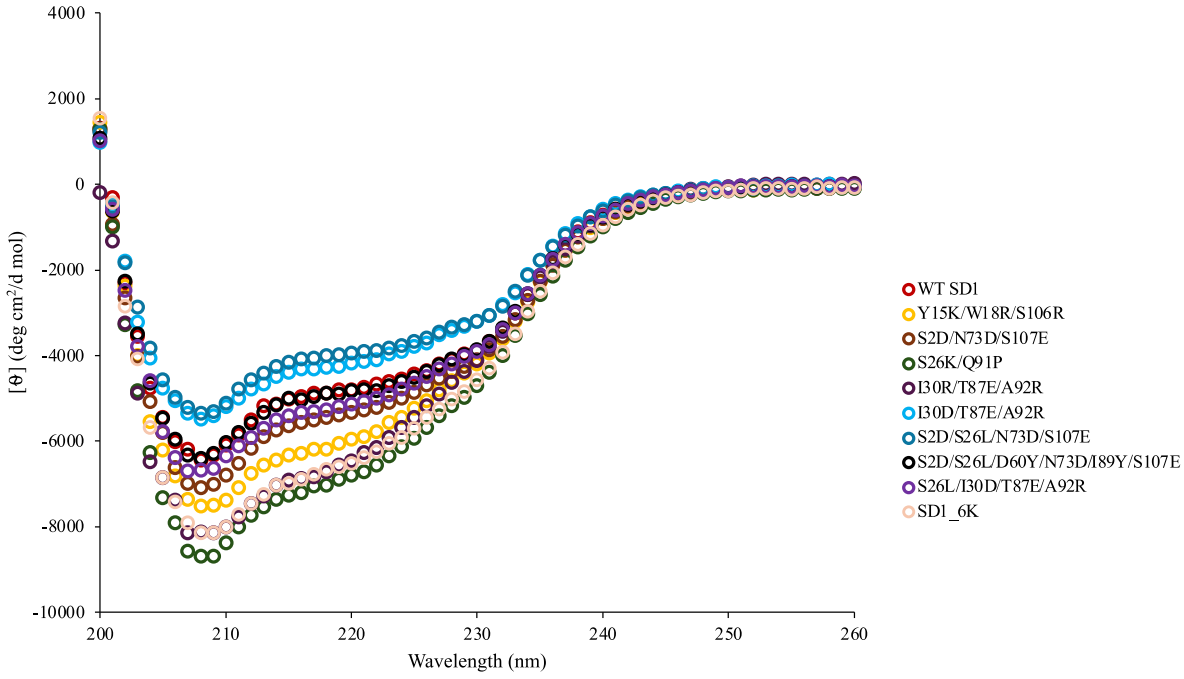

Figure 19: The CD spectra of WT SD1 and mutants. The spectra were obtained at 20 ºC using a quartz cuvette with a path length of 2 mm.

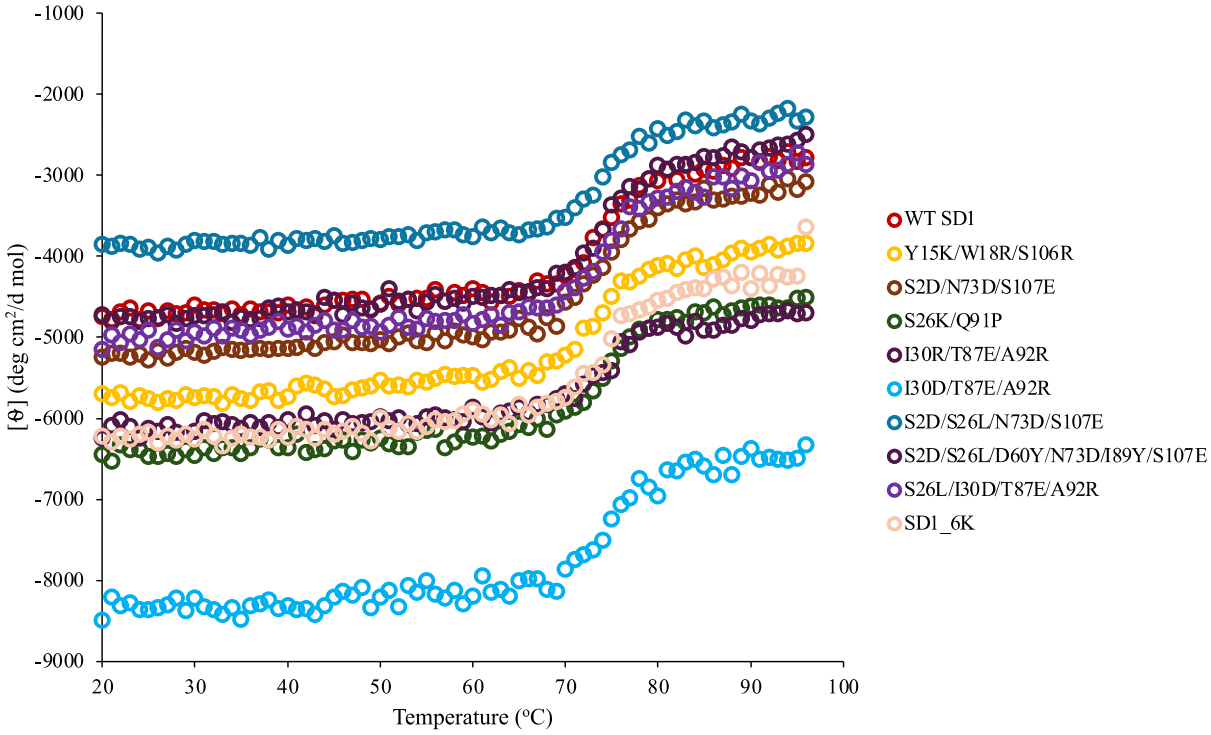

Figure 20: Thermal denaturation study of WT SD1 and mutants. The denaturation curves were obtained by heating the samples from 20 °C to 100 °C at 1 °C/min at 222 nm. The Tm was determined at 73 °C.

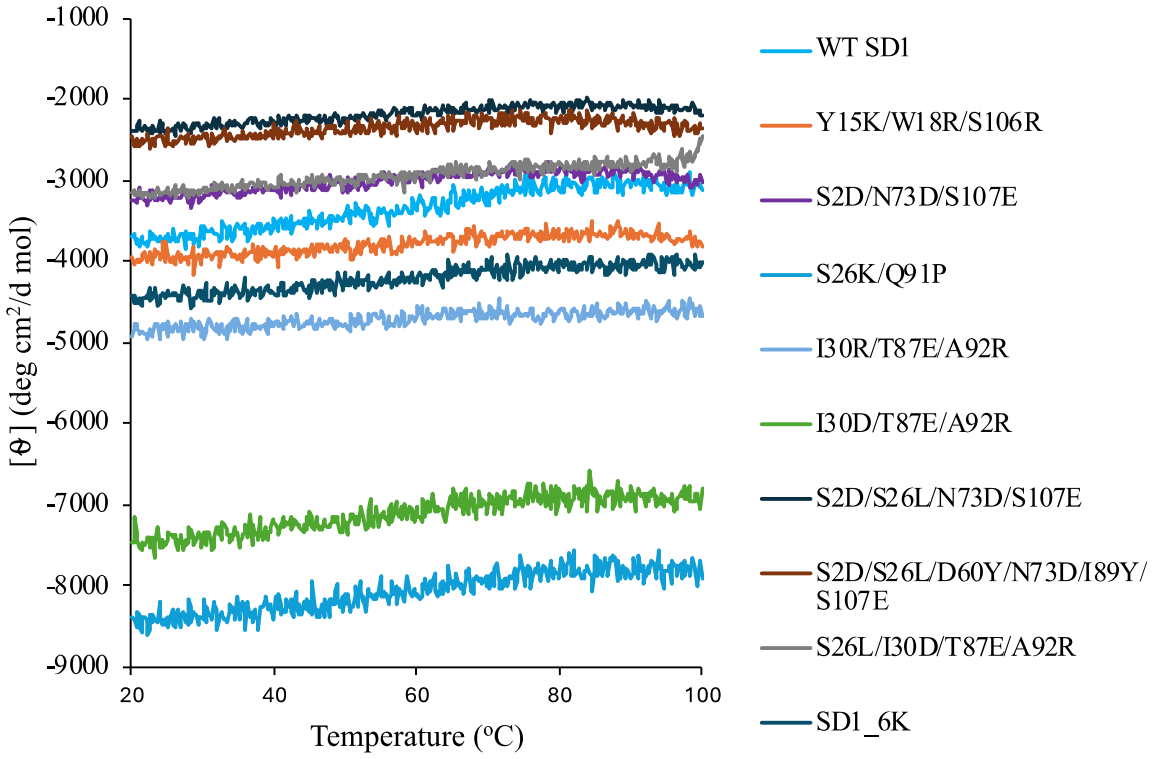

Figure 21: Thermal renaturation study of WT SD1 and mutants. The renaturation curves were obtained by cooling the samples from 100 °C to 20 °C at 1 °C/min at 222 nm.

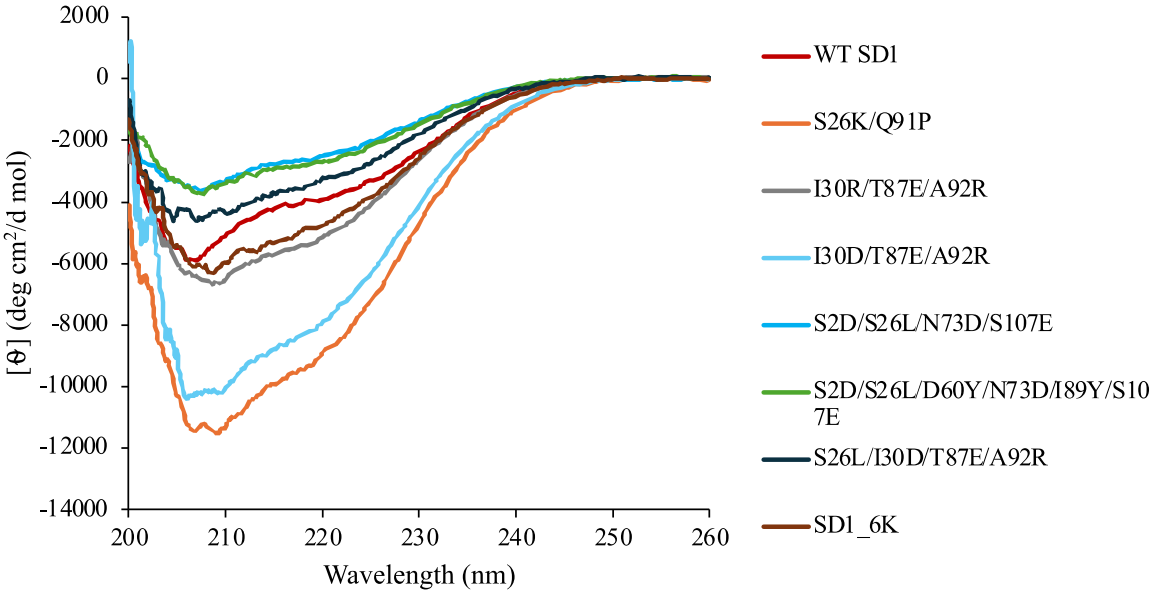

Figure 22: The CD spectra of WT SD1 and mutants after denaturation. The spectra were obtained at 20 °C using a quartz cuvette with a path length of 2 mm.

3.4 Discussion

Characterization of the WT SD1 and mutants was necessary to understand how the several mutations in SD1 affected the properties and structural integrity of the protein. Furthermore, the solubility of the various mutants needed to be investigated prior to selecting the most suitable mutant for binding applications. Understanding how the protein's intrinsic properties influence solubility, stability, and interactions in supersaturated solutions could inform future protein engineering efforts to improve the solubility of targeted proteins, especially since adequate protein concentrations in solution are required for a variety of applications, including structural and biochemical research.

In this study, the solubility of the several designed mutants and WT SD1 was studied at two different pH levels, as well as the stability of the proteins at increasing temperatures. Mutations were designed to increase the solubility of SD1 based on surface charging and the reduction of aggregation-prone regions (APRs) by substituting them with gatekeeper residues (GKRs). The highest solubility was observed for mutations involving the decrease in the APRs, and this could be due to the complete suppression of aggregation caused by the consecutive hydrophobic residues found in the SD1 sequence. APR refers to an uninterrupted sequence of five or more successive residues with Tango scores greater than 5% (Ganesan et al., 2016). GKRs are a group of amino acids that can inhibit protein aggregation via charge repulsion (R, K, D, E, P), entropy cost of aggregation due to their large and flexible side chains (R, K), or incompatibility with the β-structure of aggregation due to a lack of hydrocarbon in the primary chain (P) (Beerten et al., 2012, Sant'Anna et al., 2014). The solubility of the most soluble mutant (S26K/Q91P) may be related to the solubility of proline as an individual amino acid (David, 2014) and also due to the reduction of APR-induced aggregation. Proline at the 91st position in the SD1 sequence might have adopted a -60 ϕ angle at the i + 1 position of a type

II rotation to reduce conformational entropy while maintaining stability due to the favorable torsion angle (Hutchinson and Thornton, 1994).

Although Trevino et al., 2007 reported Ser, Glu, and Asp to contribute more to protein solubility than the other hydrophilic amino acids. However, based on the ΔΔG by FoldX, substituting Ser to Lys at the S26 position increases the thermostability of the protein. Therefore, the combined effects of proline and lysine in the mutant (S26K/Q91P) may have resulted in the mutants' high solubility in solution, for the surface-charged variants solubility of the proteins could be due to the electrostatic repulsion introduced by the charged residues and the charge distribution on the surface of the proteins preventing intermolecular interactions maintaining the fold of the proteins in solution. Surprisingly, the hexa-lysin tagged SD1 (SD1_6K), which was the most soluble protein in in vivo expression in *E. coli,* was not significantly soluble in the buffer solutions at both both pH 4.7 and 7.7. This could be because of the charge concentration at the termini of SD1 and not well distributed on the protein surface like the surface-charged mutants. Therefore, this suggests that surface charge distribution contributes to improving the solubility of proteins. The reduced solubility observed when hydrophobic and tyrosine residues were introduced to the mutants could be due to the additional mutations that led to intensifying the APR in the protein sequence. Noting the role of tyrosine in increasing protein intramolecular hydrogen bonding and the stabilization of the fibronectin type III (FN3) domains (Pace et al., 2014, Porebski et al., 2015), the addition of pair of tyrosine residues (S2D/S26L/D60Y/N73D/I89Y/S107E) decreased the solubility of SD1. The observed decrease in solubility could be attributed to an increase in the number of aromatic groups in the SD1 sequence. Moreover, Hou et al. (2018) reported that aromatic-rich proteins are less soluble than aromatic-depleted proteins. The S26L substitution in the two different mutants behaves differently with respect to protein solubility. In S2D/S26L/N73D/S107E, a surface-charged variant, the solubility of the protein significantly increased. Meanwhile, in

S26L/I30D/T87E/A92R, APR reduction mutant, the solubility of the protein decreased. There is no clear explanation for these differences. However, the amino acids in the different mutants might interact differently, resulting in different results.

Apart from ascertaining the solubility of the proteins, the storage stability and aggregation kinetics at 25 ^{o}C, at different pH (4.7 and 7.7) levels were studied to understand the influence of the surface residues, pI, and protein charges on solubility, storage stability, and aggregation. According to the findings, protein solubility has a significant impact on protein storage stability because more soluble proteins have a higher concentration in solution and a slower tendency to aggregate than less soluble proteins. The data also shows that the charges of proteins in solutions play a significant role in solubility and storage stability. These mutants, S2D/N73D/S107E and S2D/S26L/N73D/S107E showed a decrease in solubility and storage stability at pH 7.7 because, at this pH, the net charge of the proteins is close to zero. The observed decrease in storage stability could be reduced charge repulsion at pH 7.7, which is very close to the pI of the proteins. Therefore, this data supports the concept that increasing the net charge of proteins may help increase their solubility since protein solubility in solution is lowest when the solution pH is close to the protein's pI (Kramer et al., 2012). Likewise, the solubility and storage stability of the Y15K/W18R/S106R mutant were reduced at pH 4.7. The observed decrease in solubility and storage stability could be because of the increased basicity of the solution, which may have resulted in non-specific charge repulsion affecting the folding and stability of the protein in the solution. To support this assumption, Chi et al. (2003) argued that folded protein conformation could become unstable due to enhanced charge repulsion. Considering the charged residues, acidic residues contributed more to higher protein solubility than their charged counterparts. This could be because of their shorter side chains and charge-carrying carboxyl moieties, making the acidic residues more potent aggregation breakers than the basic residues (Chi et al., 2003). Additionally, a previous report by Trevino et al. (2007)

stated that Asp and Glu contributed more to protein solubility than other hydrophilic amino acids.

Considering that water molecule-protein surface interactions influence protein solubility, polar or charged amino acids are more inclined to form interactions with the surrounding water molecules, enhancing protein stability in an aqueous solution (Qing et al., 2022). The influence of the surface-exposed residues on the solubility of WT SD1 and mutants was analyzed. A significant relationship was observed between the mutant solubility and the fractions of the surface-charged groups. This observation was supported by the report of Kramer et al. (2012), who found a substantial correlation between solubility and charged ASA fractions of studied proteins. The fractions of non-polar and polar residues on the surface of the WT SD1 are higher than most of the mutants, suggesting the little contribution of protein surface polarity when compared to surface charges in protein solubility.

Additionally, an even number of positively and negatively charged surface exposed residues was observed in the WT SD1, which may have encouraged the instability and aggregation of the protein in solution through favorable attractive electrostatic interactions (Chi et al., 2003; Kim et al., 2012). The S26K/Q91P mutant, a significantly soluble mutant, had the highest fractions of surface non-polar groups and the lowest fractions of surface polar and charged groups. This result shows that the mutant's solubility may be attributable to a mechanism other than contributions from surface-charged groups. The following claims may support the assumption above: the proline residue may have assumed a favorable torsion angle, resulting in conformational stability and improved mutant solubility. The S26K mutation may have resulted in favorable coulomb interactions between the internally charged lysine residues and the surrounding negatively charged residues (Isom *et al*., 2011), stabilizing the protein structure and increasing the mutant's solubility. Additional studies may be required on the S26K/Q91P

mutant to confirm the mechanism of the mutant's folding, stability, and solubility, which is currently beyond the scope of the present study.

Given that the different amino acid side chains in a polypeptide preferentially interact to form a 3D architecture that minimizes the total free energy (Qing et al., 2022), the Y15K/W18R/S106R mutant may be the most suitable mutant for the design of target affinity reagents. This mutant had the highest fraction of surface polar and charged residues and the fewest fraction of surface non-polar groups when compared to the other soluble mutants. Proteins' structural stability in solution relies on the interaction with the surrounding water molecules, of which the polar and charged amino acids greatly enhance these interactions, thus burying the hydrophobic residues and limiting their interactions with water molecules in solution (Shaytan et al., 2009, Rose et al., 1985). This is the mechanism behind the solubility of hydrophilic soluble proteins. Therefore, the Y15K/W18R/S106R mutant which was the first soluble mutant designed in this study and was named SD1 mutant 1 (SD1-M1), was selected for the design of target affinity binders.

Furthermore, we present the structural model of SD1, displaying the substituted residues as sticks (**Figure 19**). The extended side chains of the charged residues on the protein surface could imply that the mutations increased SD1's local hydrophilicity. This improvement in the surface's hydrophilic qualities of SD1 may have enhanced the protein's ability to interact with surrounding water molecules while burying the hydrophobic core, consequently increasing its solubility. The side chains of non-polar residues (alanine (A), leucine (L), isoleucine (I), valine (V), and phenylalanine (F)) can't form hydrogen bonds with water and hence are hydrophobic. The polar residues glutamine (Q) and asparagine (N) are capable of forming four hydrogen bonds, whereas serine (S), threonine (T), and tyrosine (Y) can form three hydrogen bonds with water molecules. Aspartic acid (D) and glutamic acid (E) create four hydrogen bonds with water molecules, whereas arginine (R), lysine (K), and histidine (H) form five, three, and two

hydrogen bonds with water molecules, respectively, and get protonated at acidic or neutral pH levels (Liljas et al., 2016). Protonation makes it possible for them to generate strong electrostatic interactions, such as ionic bonds and salt bridges, which help to stabilize the proteins' 3D structural conformations in solutions (Qing et al., 2022). Therefore, the substitution of SD1 surface residues to charged residues and the reduction of APRs with the GKRs may have increased the inter and intra-electrostatic repulsion or coulomb interactions, enhancing the protein's ability to interact with the surrounding water molecules through effective electrostatic interactions that stabilized the mutants' structural and improved solubility.

CD spectroscopy is the most commonly utilized method for studying protein structures and temperature stability (Greenfield, 2007). CD spectroscopy was used to investigate how the different amino acid substitutions in SD1 affected the proteins' structural integrity and thermal stability. The spectra of the WT SD1 and mutants show that the mutations had not negatively affected the structural architecture of the protein, as similar helical patterns were observed for all proteins. Upon heating the protein samples, an irreversible denaturation at temperatures above 70 ^{o}C was observed for both the WT SD1 and mutants, suggesting the thermal stability of the protein. All samples recorded a Tm of about 73 ^{o}C, indicating that the several mutations in the protein had not affected the thermal stability. SD1's rich hydrophobic core residues could be one of the factors influencing protein thermal stability, as hydrophobic proteins have been shown to account for around 60% of protein stability (Pace et al., 2011). According to the analysis of all mutants' ASAs, the S26K/Q91P mutant had the highest fraction of non-polar residues and the lowest fraction of charged and polar residues, suggesting that the S26K mutation within the core residues could have a negative impact on the protein's thermal stability. Moreover, charges within the protein core have been reported to alter a protein's folded state because of its incompatibility with the hydrophobic environments (Isom et al., 2010). However, the CD spectra showed that the S26K/Q91P mutant's heat stability was the same as that of the

other mutants, and this implies that the internal charge mutation had not affected the structural integrity of SD1. Isom et al. (2011) suggested three possible protein reactions to internal group ionization. Firstly, placing the buried side chain in a polar or polarizable milieu may enable charge stabilization while conserving protein structure. Secondly, proteins may adopt a different fold conformation to fix the charge by interacting with internal or bulk water, whereas ionization can result in subglobular structural changes. Finally, in extreme situations, ionizing the core region can certainly unfold the protein. Based on the CD spectroscopic data, the S26K/Q91P mutant may have adopted the first two structural conformations.

To summarize, the several mutations targeted at enhancing SD1 solubility and the effect of the mutations on the protein stability solution and thermal stability were discussed. Improving SD1's solubility was driven by its unique property of intrinsic thermostability, which makes it an ideal scaffold protein for the development of novel affinity reagents for possible applications in fields that require thermostable proteins. The discussed data show that SD1 can tolerate various solubility-enhancing modifications. This supports the hypothesis that SD1 is mutationally robust and can tolerate and accommodate several side-chain alterations targeted at generating target binding functions.

Chapter 4:

Engineering SD1 for binding to TNF-α

4.1: Overview

The previous two chapters described the strategies used to increase the solubility of SD1, which was a limiting factor in the development of target affinity reagents. The study involved characterizing several generated soluble mutants in order to select the best variant for the construction and selection of target affinity reagents. The Y15K/W18R/S106R mutant named SD1 mutant 1 (SD1-M1) was selected because, in addition to the increased soluble expression and solubility, it had the highest fractions of surface polar and charged residues and the lowest fractions of surface non-polar residues, which is typical of soluble hydrophilic proteins. Surface polar and charged residues have been shown to interact with surrounding water molecules, preserving the protein's structural conformation in solution while maintaining folding and solubility (Shaytan et al., 2009; Qing et al., 2022).

In this chapter, SD1-M1 was used as a framework for the design of affinity binders to TNF-α. TNF-α is a proinflammatory cytokine involved in immune system maintenance, inflammation, and defense and has lately been linked to a variety of pathological states, including autoimmune, inflammatory, and malignant diseases (Balkwill, 2006; Josephs et al., 2018; Sethi and Hotamisligil, 2021; You et al., 2021; Jang et al., 2021). Thus, blocking TNF-α in these disease conditions is an effective means of treating autoimmune, inflammatory, and malignant diseases. The human TNF-α was chosen as a target protein for the purpose of evaluating the effectiveness of designing SD1 target affinity binders and because of the extensive study of protein-based TNF-α inhibitors.

After the construction of the SD1 phagemid vector, libraries of target binders were designed by combinatorial engineering based on the Kunkel mutagenesis, a site-directed mutagenesis technique that allows single or large insertions, deletions, or substitutions in a DNA sequence

(Huang et al., 2012; Liu et al., 2020). This was followed by screening and selection using the phage display system.

4.2: Materials and methods

Strains and vectors: *E. coli* strain XL1-blue *supE*44 *rel*A1 *gyr*A96 *hsd*R17 *rec*A1 *end*A1 *thi*-1 lac (F' [*pro*AB^{+} *lac*I^{q} *lac*ZΔM15 Tn10 (Tet^{r})], was used as the expression host during the construction of SD1 phagemid vector and for biopanning. *E. coli* strain CJ236 *dut*1, *ung*1, *thi*-1, *recA*1 / pCJ105 (F' *cam*^{r}) was used to synthesize uracilated ssDNA. *E. coli* SS320 [F' *pro*AB^{+} *lac*I^{q} *lac*ZΔM15 Tn10 (tet^{r})] *hsd*R *mcr*B *ara*D139 Δ(*ara*ABC-*leu*)7679 Δ*lac*X74 *gal*U *gal*K *rps*L *thi*, was used for the transformation, expression, and production of the phagemid library. The M13KO7 helper phage was used for the production of the filamentous phage display SD1 on the pIII coat protein. The phagemid pComb3xSS (Andris-Widhopf et al., 2000) was used for the cloning and construction of SD1 mutant libraries.

Construction of SD1 phagemid vector: The gene encoding SD1 was amplified from pET28aSD1 by PCR using designed primers (**Table 6**). Both the PCR product and pComb3xSS vector were digested using SpeI and SacI restriction enzymes (RE). For pComb3xSS, the SS stuffer was removed by agarose gel DNA clean-up following the manufacturer's recommended procedures (NucleoSpin® Gel and PCR clean-up; Macherey-Nagel GmbH & Co. KG, Germany). The purified RE-digested PCR product and vector were ligated using a ligation mix (Takara Bio). The ligated product (pComb3xSD1-M1) was transformed into *E. coli* XL1-blue competent cells and cultivated at 37 ^{o}C overnight. Single colonies were checked for the presence of insertions by colony PCR, and plasmids of positive clones were confirmed for correct insertions by DNA sequencing.

Table 6: Primer sequences used for construction of phagemid vector and phage display library

		Primer sequence (5'→3')
SD1-M1	Fwd	AAAACTGAGCTCATGTCAAGTGGAGGATTGG
	Rev	AAATCAACTAGTCGATGCAGATCGGGAAAACCT
Phage library		AGTGGTCCAGACACGAAAGTGNNCNNCNNCNNCNNCNNCNNCNNCN NCNNCAGAAAAGCAGAGGAACGA

Amplification and titration of M13KO7 helper phage: Single colony of *E. coli* XL1-Blue was inoculated into an LB broth supplemented with 10 µg/ml of tetracycline and incubated at 37 °C overnight with constant shaking. The following day, an aliquot of the overnight culture was inoculated into a 10 ml LB broth supplemented with 10 µg/ml of tetracycline in a 100 ml conical flask and incubated at 37 °C with constant shaking until OD_{600} had reached 0.3. The M13KO7 helper phage was added at a multiplicity of infection (MOI) of 20:1 (phage-to-cell-ratio) and grown at 37 °C for 30 min. Followed by the addition of kanamycin at a final concentration of 70 µg/ml, and the culture was continued for about 8 h at 30 °C. Following the 8 h incubation, the culture was then incubated for 15 min at 65 °C and centrifuged for 15 min at 4600 xg. The supernatant was transferred to a fresh tube, and the cell pellet was discarded into a phage destruction detergent. Aliquots of the supernatant were transferred into microcentrifuge tubes, and DMSO was added to a final concentration of 7% v/v and stored at -80 °C until use.

For the titration of the helper phage, serial dilutions of the amplified M13KO7 helper phage were prepared in tris-buffered saline (TBS). Then, 1 µl of each dilution was inoculated in 200 µl of exponentially growing *E. coli* XL1-Blue cells. The cells were incubated at 37 °C for 15 mins and spread on NZY agar plates (5 g/l NaCl, 2 g/l $MgSO_4$, 5 g/l yeast extract, 10 g/l NZ amine (casein hydrolate) pH 7.5, 15 g/l Agar) using pre-warmed NZY surface agar. The plates

were incubated at 37 ºC overnight, and the phage titer (pfu/ml) was determined the following day using the formula below.

$$\frac{\text{Number of plaques(pfu) x dilution factor}}{\text{volume plated(L)}} \text{ x } 1000\ \mu\text{L/ml}$$

The individual clones from the library were screened using colony PCR, and the library sequence and diversity were verified by DNA sequence analysis.

Production and preparation of phage particles displaying SD1: *E. coli* XL1-Blue cells harboring pComb3xSD1 were grown in 2YT medium (16 g/l Tryptone, 10 g/l yeast extract, 5 g/l NaCl) supplemented with 100 μg/ml of ampicillin, 10 μg/ml of tetracycline and 1% glucose to suppress the expression of the pComb3xSD1 and incubated at 37 ºC overnight with constant shaking. An aliquot of the overnight culture was added into a 5 ml 2YT medium in a 100 ml flask supplemented with 100 μg/ml of ampicillin, 10 μg/ml of tetracycline, and 1% glucose and grown until OD_{600} had reached between 0.5 and 0.6. This was followed by the addition of M13KO7 helper phage to the *E. coli* culture at an MOI of 20 and incubated at 37 ºC for 30 min without shaking and 30 min with constant shaking. The cell culture was then centrifuged at 3300 xg for 10 min at 25 ºC to harvest the cells. The supernatant was decanted, and the cell pellet was resuspended in a 50 ml 2YT medium supplemented with 100 μg/ml of ampicillin, 10 μg/ml of tetracycline, 50 μg/ml of kanamycin and incubated overnight at 30 ºC with constant shaking for phage production.

The overnight culture was centrifuged twice at 12000 xg for 10 min, and the supernatant was transferred into fresh centrifuge tubes. A 0.2 volume of PEG/NaCl solution was added, thoroughly mixed, and incubated overnight at 4 ºC for phage precipitation. The following day, the precipitated phage samples were centrifuged at 17,000 xg for 20 min at 4 ºC, and the supernatant was discarded. The pellet was resuspended in I ml TBS, transferred to a

microcentrifuge tube, and centrifuged for 2 min at 17,000 xg. The supernatant was transferred to a new microcentrifuge tube, and 150 µl of PEG/NaCl was added, thoroughly mixed, and incubated on ice for 30 min. After incubation, the sample was centrifuged at 17,000 xg for 10 min at 4 ºC, the supernatant was discarded, and the pellet was resuspended in 1 ml TBS containing 7% v/v DMSO and stored at -80 ºC until use.

Determination of infective titer of the phages: 100 µl of actively growing *E. coli* XL1-Blue culture (OD600 ~ 0.8) was inoculated with 10 µl of serial dilutions of the prepared phages and incubated for 15 mins at 37 ºC. Then, each dilution was plated on LB agar plates supplemented with 100 µg/ml of ampicillin and incubated overnight at 37 ºC. The infective titer (CFU/ml) was calculated using the formula below.

$$\frac{\text{Number of colonies(CFU)x dilution factor}}{\text{Volume plated(L)}} \text{ x } 1000\ \mu\text{L/ml}$$

Detection of SD1 display by western blotting: Phage proteins were separated by loading approximately 10^{12} recombinant phage particles into a freshly prepared 12% SDS-PAGE and transferred to a PVDF membrane. The membrane was blocked with 5% skim milk at room temperature for 1 h and incubated with primary anti-HA and anti-pIII primary antibody solution, respectively, overnight at 4 ºC. The membrane was washed 3 times with PBST and incubated with HRP-conjugated secondary antibody solution. Then, it was incubated at room temperature for 4 h. The membrane was washed 3 times, and luminescence was visualized under the luminograph after applying a mixture of luminol and peroxidase solution (1:1).

Library construction using Kunkel mutagenesis: *E. coli* CJ236 transformed with pComb3xSD1 was grown at 37 ºC in a 30 ml 2YT medium supplemented with 100 µg/ml of ampicillin, 10 µg/ml of chloramphenicol and 0.25 µg/ml of uridine until OD600 had reached about 0.5. M13KO7 helper phage was added to the culture at a final concentration of about

10^{10} pfu/ml and grown for 2 h at 37 ºC with constant shaking. Kanamycin at a final concentration of 70 μg/ml was added and grown overnight at 30 ºC with constant shaking. Uracilated ssDNA was isolated from the phage particles following the procedure described for Qiagen plasmid kits (Qiagen GmbH, Hilden, Germany). The library was generated with a mixture of 1 μg of the uracilated ssDNA, approximately 0.1 μg of 5' phosphorylated oligonucleotide (**Table 6**), and a temperature gradient at 70 ºC gradually decreased to 30 ºC at a rate of 1 ºC/min. This was followed by gap filling and ligation with T7 DNA polymerase and T4 DNA ligase. The mixture was dialyzed using a 0.025 μm membrane disk for 1 h, then transferred to a fresh microcentrifuge tube and stored at -20 ºC until use. The synthesized dsDNA was electroporated into freshly prepared electrocompetent *E. coli* SS320 cells, which were transferred to a pre-warmed 12 ml SOC and incubated at 37 ºC for 1 h. The titer was determined by plating and incubating 100 μl of 10-fold dilutions of the culture overnight at 37 ºC. After transformation, the cells were transferred to 200ml 2YT medium supplemented with 2% glucose and 100 μg/ml of ampicillin and incubated at 37 ºC with continous shaking until OD600 had reached 0.5. Then, the cells were infected with the M13KO7 helper phage and incubated for an additional 1 h (30 min with shaking and 30 min with constant shaking) at 37 ºC. The culture was centrifuged at 3000 xg for 20 min at 4 ºC. The cell pellet was resuspended in 500 ml 2YT medium in a 2 L flask supplemented with 100 μg/ml of ampicillin, 10 μg/ml of tetracycline, and 50 μg/ml of kanamycin and incubated overnight at 30 ºC with constant shaking for phage production. Phagemids were prepared through PEG/NaCl precipitation.

Calculation of library size: The library size was calculated after the determination of the transfection titer using the formula below.

$$\frac{\text{The number of mutated clones}}{\text{The total number of clones}} \text{ x} \frac{\text{number of colonies(CFU)x dilution factor}}{\text{volume plated(L)}} \text{ x } 1000\ \mu\text{L/ml}$$

Panning for selection of binders: About 2 μg/well of TNF-α (PeproTech, Inc., Cranbury, USA) was coated in a 96-well Maxisorp plate (Thermo Fisher Scientific, Roskilde, Denmark) using a coating buffer (0.1 M $NaHCO_3$, pH 9.6) and incubated overnight at 4 ^oC. Following the immobilization of TNF-α, the well was washed twice with TBS buffer and blocked with 1% bovine casein milk solution (Sigma-Aldrich, Missouri, United States) for 1 h at 37 ^oC. The phage pool was also blocked for 1 h by incubating at room temperature in 1% bovine casein milk solution. The blocked phage samples were introduced to the wells and incubated for 2 h at room temperature. After phage incubation and washing of unbound phages with TBST (0.1% tween 20), bound phages were eluted by incubating with 0.1 M HCl-glycine (pH 2.2) for 10 mins, followed by pipetting up and down to recover the eluate and transferred into a microcentrifuge tube containing 2 M Tris-base for neutralization. The acid elution was done twice before phage propagation. For the competitive elution (CE), about 6 μg/well of TNF-α was incubated with the target bound phages for 1 h, and the supernatant was collected and used for infecting freshly prepared young cultures of *E. coli* XL1-blue for phage production. Eluted-produced phages were purified using PEG/NaCl precipitation and used in subsequent rounds of biopanning.

Binding analysis of the polyclonal phages: Indirect ELISA of the purified phage library and the eluted phages from the rounds of biopanning were used to analyze binding to the target. The 96-well plates were coated with the target protein as described in the panning method and blocked for 1 h with 1% bovine casein milk solution. The purified phage pools were blocked for 1 h with 1% bovine casein milk solution added to each of the blocked wells and incubated at room temperature for 2 h. Unbound phages were washed three times with TBST (0.1% tween 20). To detect the bound phages, HRP-conjugated anti-M13 pIII antibody was added to the wells, incubated for 1 h at room temperature, and washed three times with TBST (0.1% tween 20) and once with TBS. Colorimetric assays were performed using Turbo-TMB, and 2M H_2SO_4

was used to stop the reaction after 20 mins incubation. The sample OD_{450} nm was measured using a plate reader.

4.3 Results

4.3.1 Phage display of SD1-M1

To determine whether SD1-M1 could be displayed on the filamentous phage, the gene encoding SD1-M1 was amplified using PCR and introduced into the pComb3xSS vector by removing the SS stuffer and replacing it with SD1-M1 via the SpeI and SacI restriction sites using REs. The resultant plasmid, named pComb3xSD1-M1, was transformed into *E. coli* XL1-Blue. Following superinfection with M13KO7 helper phage, the pComb3x pIII proteins and SD1-M1 were expressed. The engineered recombinant protein contains at the C-terminus a His6 tag, an HA-tag for purification and detection, and an amber stop codon preceding the pIII fragment. As shown in **Figure 24**, the recombinant SD1-M1 and phage protein were detected using anti-HA and anti-pIII monoclonal antibodies, respectively, showing that SD1-M1 was produced and displayed. SD1-M1 was produced as a fusion protein in the amber stop codon suppressor strain, *E. coli* XL1-Blue (**Figure 24**).

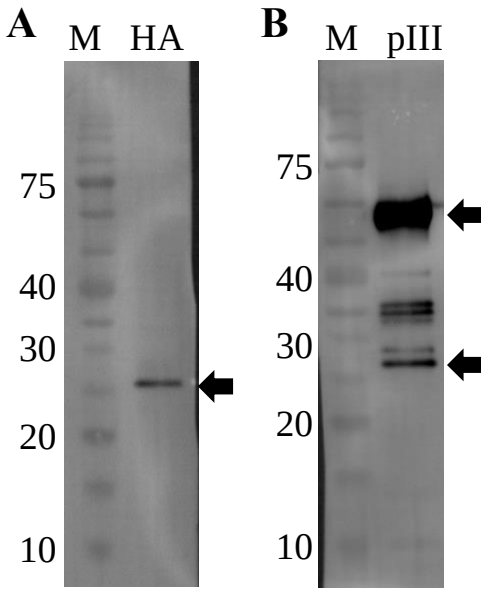

Figure 23: Detection of display of SD1-M1 recombinant protein on M13KO7 filamentous phage. Samples were separated by 12% SDS-PAGE. **A)** Detection of SD1-M1/pIII fusion expressed in an amber suppressor strain (E. coli XL1-Blue). Lane M = molecular weight marker and lane HA = detection of HA-tag anti-HA monoclonal primary antibody. **B)** Detection of pIII expression. Lane M = molecular weight marker and lane HA = detection of pIII with anti-pIII monoclonal primary antibody. HRP-conjugated mouse isotype IgG was used for the secondary detection and visualization of the primary targets. The black arrow indicates the target bands for SD1-M1/pIII fusion and pIII, respectively.

4.3.2 Construction, characterization, and expression of the SD1-MI combinatorial library in the pComb3xSD1-M1 plasmid.

Following confirmation that SD1-M1 was displayed on the filamentous phage pIII surface, the SD1-M1 phage display library was constructed. Ten amino acid residues from the exposed surfaces of the most extended loop region of SD1 (**Figure 19**) were selected. The library was generated using the NNC degenerate codons, which encode 15 amino acids with the exclusion of Lys (K), Met (M), Glu (E), Gln (Q), Trp (W), and stop codons. The theoretical library size is 16^{10} = 1.0 x 10^{12} DNA sequences, which encodes 15^{10} = 5.7 x 10^{11} protein sequences. SD1-M1 phage display library was constructed using Kunkel mutagenesis using ssDNA isolated from phages produced in an *E. coli* strain (CJ236) lacking dUTpase and uracil-N-glycosylase

(dut$^-$/ung$^-$) as a template (**Figure 25**). The ssDNA carries a high proportion of uracil (dU-ssDNA) instead of thymine and was used to generate covalently closed circular double-stranded DNA (CCC-dsDNA), verified on agarose gel electrophoresis (**Figure 26**). Electroporation of *E. coli* SS320 cells (dut+/ung+) deactivates the uracil-containing template, allowing for preferential replication of new mutagenic strands (Frei and Lai, 2016). Following electroporation of *E. coli* SS320 with the CCC-dsDNA, about 4.7 x 10^7 CFU/ml was obtained representing 4.7 x 10^7 unique transformants. A total of 48 clones from the library were screened using PCR, and the results showed that about 98% (47/48) of the clones had inserts of the expected length (**Figure 27**). Furthermore, 16 clones were randomly picked and sequenced for the analysis of library quality of heterogenicity (**Table 7**). The results show that the codons correlate to the library's degeneracy, in which the last nucleotide was intended to be only C and had sufficient diversity for target binder screening. The library size was calculated to be 3.5 x 10^7 CFU/ml. The display of the library was verified, and the amber stop codon between the HA-tag and the pIII fragment resulted in the expression of SD1-M1 free of the pIII protein in the non-suppressor strain *E. coli* SS320 (**Figure 28**).

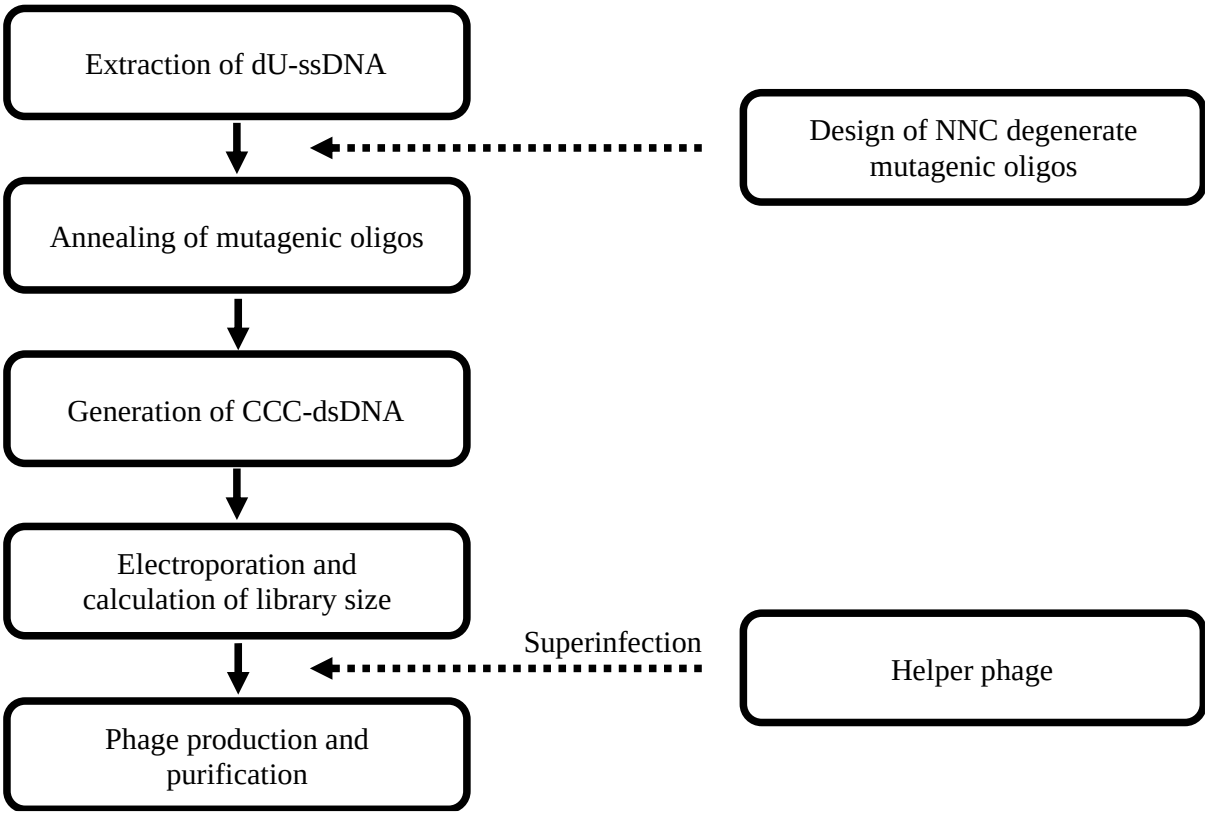

Figure 24: Illustration of the Kunkel mutagenesis method

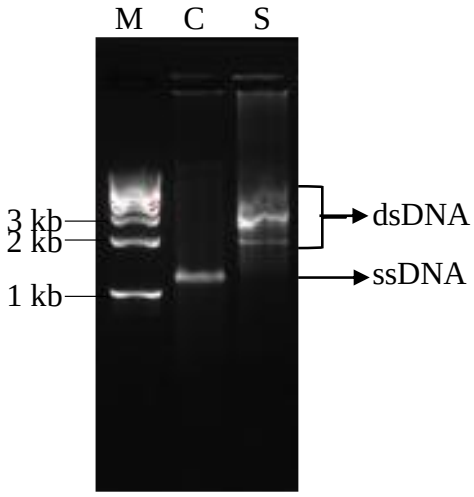

Figure 25: Agarose gel electrophoresis of the generated CCCdsDNA from dU-ssDNA. M = marker, C = control; dU-ssDNA template, S = sample; generated dsDNA

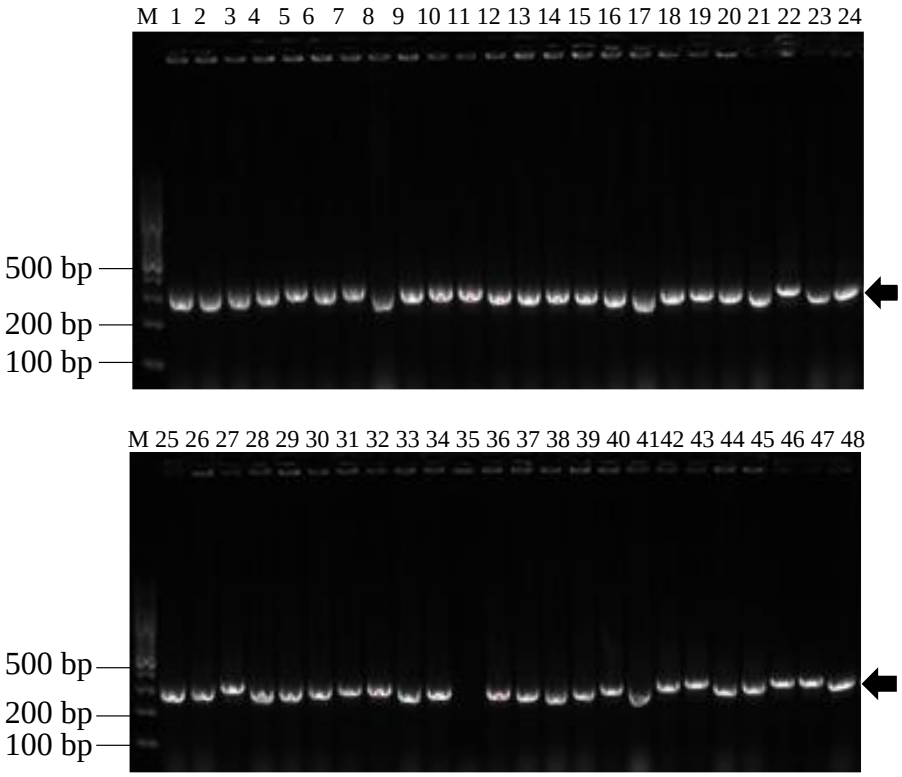

Figure 26: Agarose gel electrophoresis of library clones screened using colony PCR. M = marker, lanes 1- 48 indicates the individual clones. Black arrows indicate the expected target length.

Table 7: Sequence analysis of the designed phage display library

Name	Variegated loop (positions 72-81)	Mutation
Wild type	NSNGLTRVAF	-
L1	FNDVDSRLDF	Yes
L2	FGFCSVYTSF	Yes
L3	FFFCFPASDL	Yes
L4	NSHRLTRVAF	Yes
L5	DVVDGLADDL	Yes
L6	FPYAFFFDFG	Yes
L7	NSNGLTRVAF	No
L8	SGPFFASVVV	Yes
L9	NSNGLTRVAF	No
L10	NSNRLTRVAF	Yes
L11	LCISFSVSGV	Yes
L12	RCIIYNCAFH	Yes
L13	GFTLCSNGRS	Yes
L14	NSNGLTRVAF	No
L15	NIDFFSGRFF	Yes
L16	NSNGLTRVAF	No

L denotes library

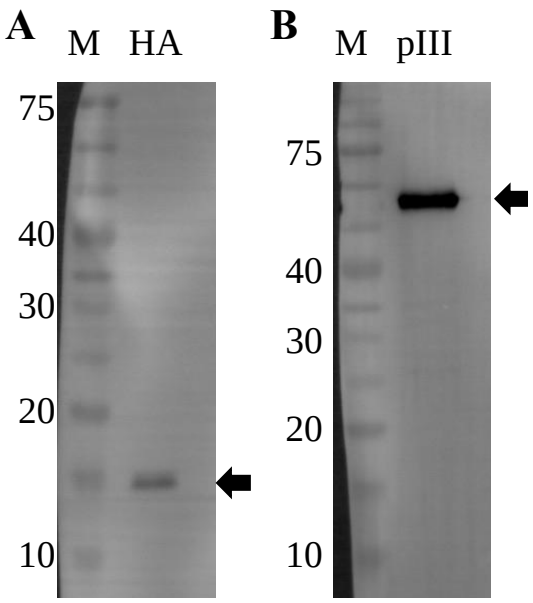

Figure 27: Detection of display of SD1-M1 library on M13KO7 filamentous phage. Samples were separated by 12% SDS-PAGE. **A)** Detection of SD1-M1 expressed in the amber non-suppressor strain (*E. coli* SS320). Lane M = molecular weight marker and lane HA = detection of HA-tag anti-HA monoclonal primary antibody. **B)** Detection of pIII expression. Lane M = molecular weight marker and lane HA = detection of pIII with anti-pIII monoclonal primary antibody. HRP-conjugated mouse isotype IgG was used for the secondary detection and visualization of the primary targets. The black arrow indicates the target bands for SD1-M1 and pIII respectively.

4.3.3 Screening for TNF-α binders

The SD1-M1 phage-displayed library was evaluated for affinity binding to TNF-α to identify binding variants. Throughout the screening process, an empty coated well was used as a negative control. Five rounds of biopanning with the SD1-M1 phage library were carried out against passively adsorbed TNF-α in 6-wells of a 96-well plate. The population of phages with an affinity for the target was determined by comparing amplified phage retention against the target to that of empty-coated wells. Five rounds of biopanning were carried out against TNF-α, with empty coated wells serving as negative controls. After five rounds of biopanning, the enrichments raised the ratio of eluted phages between the target and the negative control by

33-fold (**Figure 29**). Polyclonal phage ELISA was conducted with purified eluted phages from each round of biopanning, producing much higher signals than control wells, albeit the signal between enrichment rounds did not differ significantly (**Figure 30**). To optimize the selection of TNF-α target binders from the designed naïve SD1-M1 phage library, competitive elution (CE) of target binders with a high concentration of the target was explored. After 1 hour of incubation, the target binders were eluted, and a titer of roughly 1.5 x 10^3 CFU/ml was obtained after propagation of the eluted phages, representing a 0.5% decrease from the initial library. These findings suggest the possibility of selecting target binders from the library pool. The purified phages from the CE were analyzed using polyclonal phage ELISA, which revealed considerable enrichment from the first round of biopanning (**Figure 31**) compared to the first round of acid elution (**Figure 30**). In addition, the polyclonal phages demonstrated considerable specificity to the target and minimal observable binding to the empty-coated wells and BSA-coated wells, respectively (**Figure 31**), which served as negative controls.

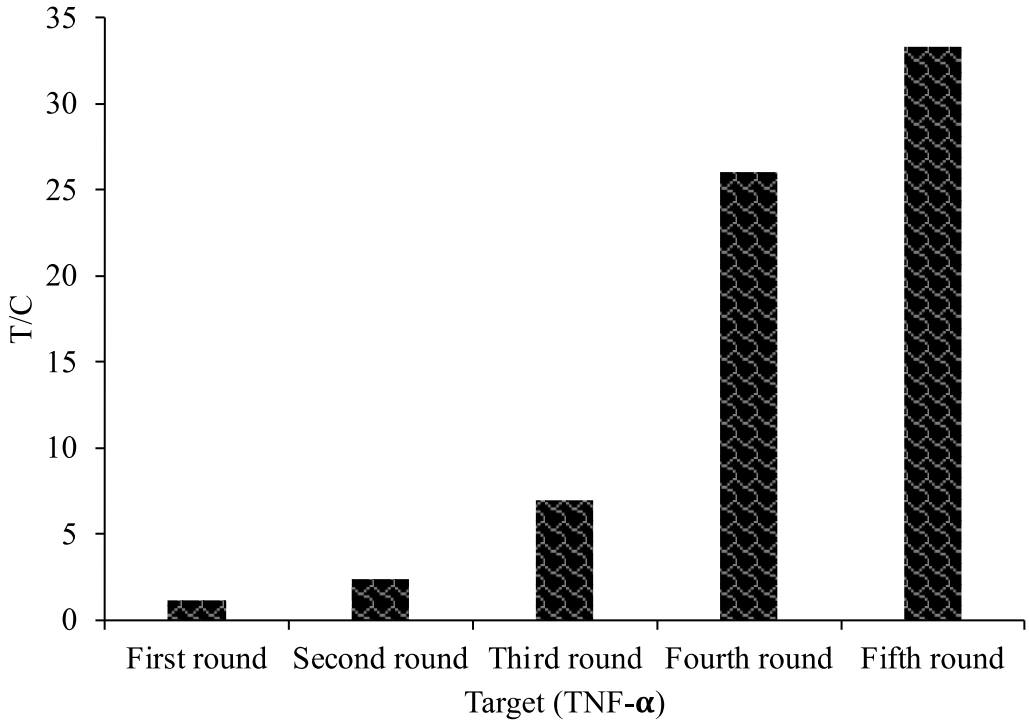

Figure 28: Enrichment of phages bound to TNF-α. The Y-axis represents enrichment, which was calculated as the ratio of eluted phages bound to target (T) coated wells to those bound to control (C) wells. The X-axis represents the various rounds of biopoanning performed against the target (TNF-α).

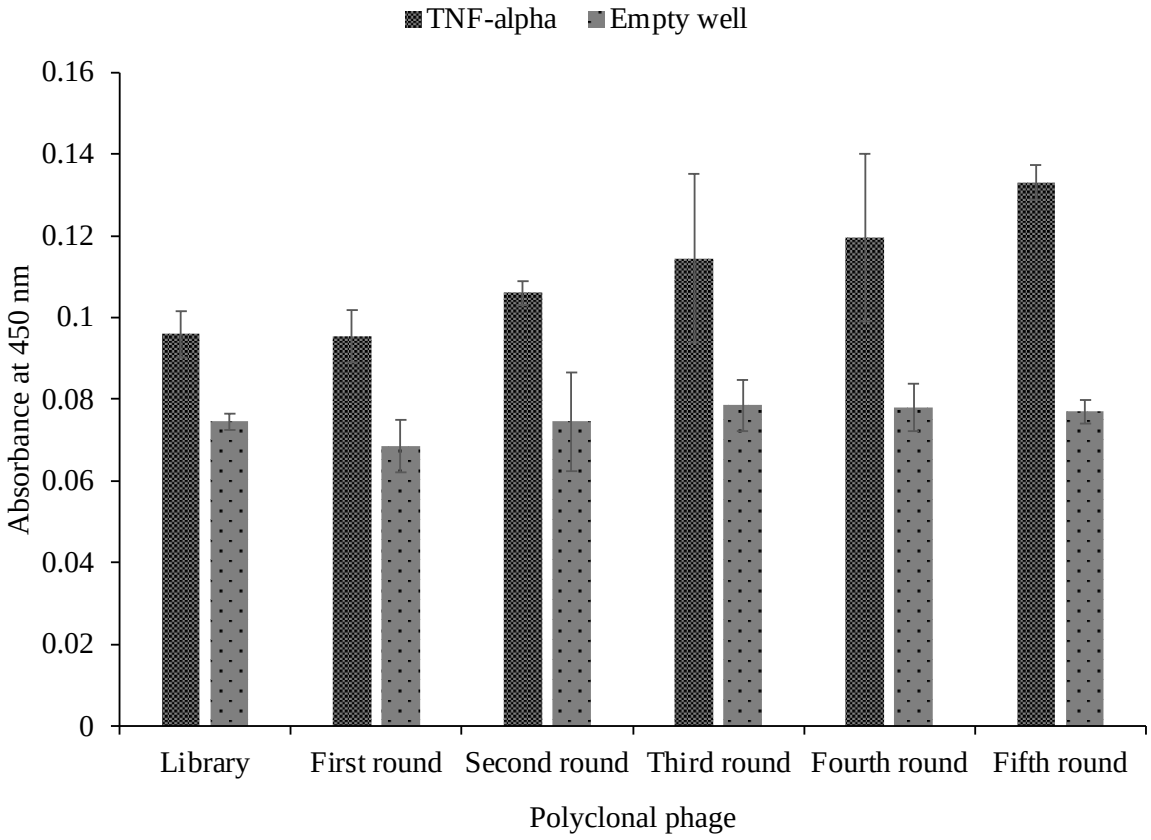

Figure 29: Polyclonal phage ELISA detection of TNF-α-coated wells. The binding of the library and the phages eluted from the several biopanning rounds was measured and empty coated wells were used as controls. Bound phages were detected using HRP-conjugated anti-M13 bacteriophage antibody. Error bars represent standard deviation (n = 2).

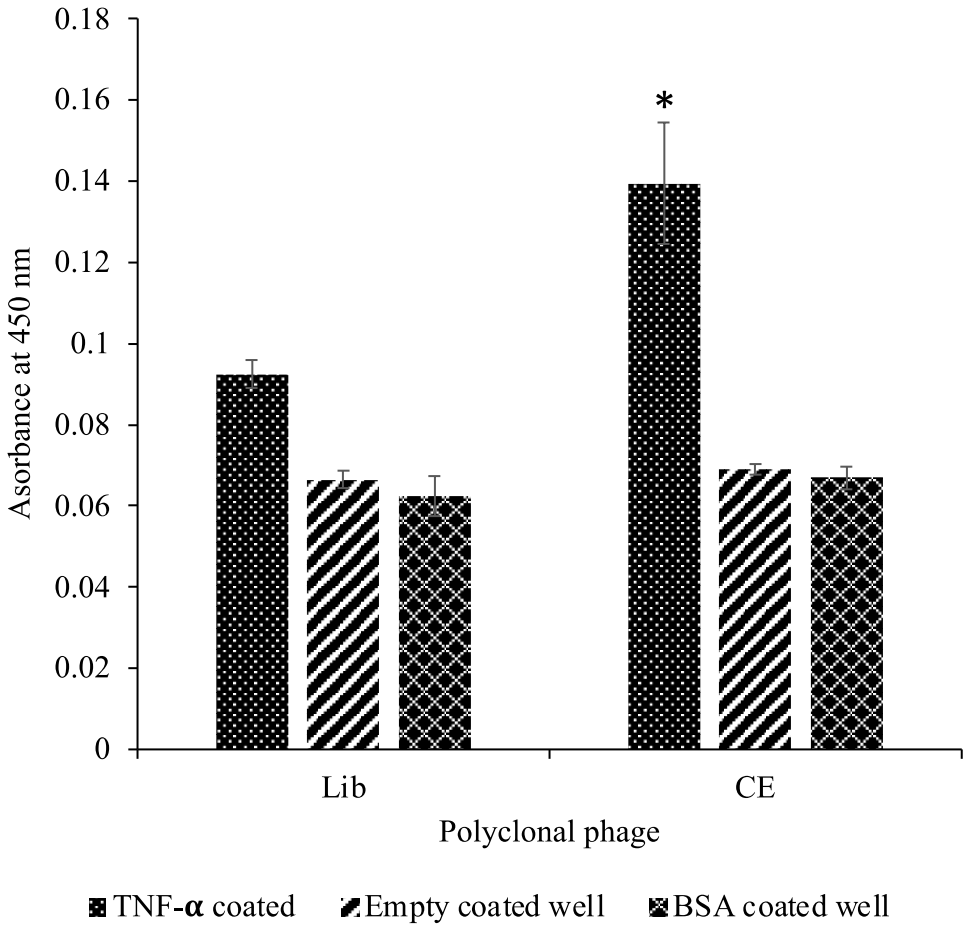

Figure 30: Polyclonal phage ELISA detection of TNF-α and BSA coated wells. The binding of the library and eluted phages, based on the incubated of high concentration of the target with bound phages for 1 h after washing off the weak and non-binders. Empty and BSA-coated wells were used as negative controls. Bound phages were detected using HRP-conjugated anti-M13 bacteriophage antibody. * $P \leq 0.05$. Error bars represent standard deviation (n = 2).

4.4 Discussion

TNF-α has been shown to be involved in inflammatory pathways, and blocking its interaction may reduce inflammation and disease progression (Esposito and Cuzzocrea, 2009; Sethi et al., 2009; Jarrot and Kaplanski, 2014; Smolen and Aletaha, 2015; Stoffer et al., 2016). This chapter describes the display of SD1-M1 on the M13 pIII surface protein, the creation of a phagemid library, and the screening for target binding variants. SD1 is a small, single-domain thermostable protein that can withstand several modifications to enable target-specific binding in a variety of applications, even in environments with minimal resources. Moreover, thermostable proteins have shown promise for several applications in basic science, diagnostics, and biotherapeutics, particularly in settings that necessitate extreme storage and deployment temperatures (Zhao et al., 2016). SD1 has a similar structural architecture to the variable domain of IgG, with extended surface exposed loops for designing target binding within a stable structural framework.

To evaluate the influence of conformational variations on SD1-M1's ability to recognize TNF-α, a well-defined combinatorial library was constructed and displayed on the surface of the M13 phage pIII protein for the selection of target binders using phage display technology. TNF-α was chosen as the target protein of choice owing to multiple studies of successful TNF-α-binder selection (Shingarova et al., 2018; Kronqvist et al., 2008; Petrovskaya et al., 2012; Nie et al., 2021; Hoffmann et al., 2012; Zhao et al., 2004; Byla et al., 2010). To construct the library, the longest extended surface loop was chosen, with the expectation that randomization would have little to no effect on the protein folding. The NNC degeneracy was utilized to lower the number of duplicated codons while enhancing screening efficiency by reducing the library's unbalanced amino acid distribution (Tang et al., 2012; Kille et al., 2012). Approximately 4.7 x 10^7 independent clones were obtained, which is just a fraction of the theoretical library size of 5.7 x 10^{11}. However, when individual clones were sampled, 98% had inserts of expected length,

and sequence analysis revealed significant heterogeneity and diversity, indicating that the library could be used for target binder screening. Furthermore, some reports of selecting clones with a high affinity for targets with a library size that represents only a fraction of the theoretical possible diversity have been described (Koide et al., 1998; Desiderio et al., 2001). The main drawback with cell-display techniques is their dependency on DNA transformation efficiency in the cells, which typically limits library size. The library was panned with TNF-α immobilized on a Maxisorp plate, and eluted phages were quantified by titering them after each round of biopanning and comparing them to eluted phages on an empty coated plate to evaluate binding enrichment. After five rounds of biopanning, enrichment increased by 33-fold, indicating that SD1-M1 could bind to the target. This shows that SD1-M1 may be well-folded, resulting in target binding. However, the observed enrichment was rather low in comparison to the 80-fold and 2815-fold enrichments reported by some studies for VHH library screening against TNF-α (Reiter et al., 1999; Nie et al., 2021). Polyclonal phage ELISA was utilized to validate the library enrichment, and the results showed little or no enrichment, implying that the enrichment from the biopanning results was insufficient to identify individual target binder clones.

TNF-α immobilization technique used in this study is based on passively unspecific adsorption to a maxisorp 96-well microtiter plate. Although frequently used and proven successful, it is uncertain what condition and form the immobilized protein takes, as the protein does not need to be denatured and the majority of its surfaces should remain accessible for target binding (Koide et al., 2009). Furthermore, elution based on the breaking of interactions between bound phages using extreme conditions may result in an enrichment of noise binders, lowering the effectiveness of detecting target-specific binders (Sidhu et al., 2000). To overcome these limitations, biotinylation of the target protein and immobilization on a surface-coated biotin-binding protein, which establishes an irreversible interaction and enables for selection even at

low concentrations of the target, have been proposed (Koide et al., 2009). It has also been shown that utilizing competitive elution to isolate scFvs that bind to small peptides decreases background binding while increasing target binder selection (Duan and Siegumfeldt, 2010). To optimize the selection and elution of TNF-alpha target binders from the SD1-M1 phage display library, bound phages were eluted at a high target concentration. The polyclonal phage ELISA data demonstrate a considerable increase in target binders from the phage pool after only one cycle of biopanning.

To summarize, the SD1-M1 phagemid vector was successfully constructed, and SD1-M1 displayed monovalently on the M13 pIII protein. SD1-M1 phagemid library was designed, and the library screening demonstrated binding to the target, as shown by 33-fold enrichment after the fifth round of biopanning. The polyclonal phage ELISA after competitive phage binder elution suggests that some clones from the phage pool have an affinity to the target, as evidenced by a considerable improvement in phage detection by the anti-M13 phage antibody when compared with the library.

Chapter 5:

Conclusion and future perspectives

5.1 Conclusion and Future Perspectives

The design of molecules capable of modulating protein-protein interactions (PPIs) creates new foundations for current drug development, diagnostics, and molecular tools in basic science research and medical treatments (Mabonga and Kappo, 2019; Belvisi et al., 2021). The use of monoclonal antibodies to modulate PPIs and other cellular processes was the first conception of protein-based modulators that required immunoglobulin humanization (Bakail and Ochsenbein, 2016). It was a significant success, with numerous drugs receiving FDA approval for therapeutic and diagnostic uses (Mullard, 2016). However, with the rising applications in research, biotechnology, and biomedicine, it became clear that antibodies have various limitations (Chames et al., 2009; Rodriguez-Nava et al., 2023; Samaranayake et al., 2009; Shepard et al., 2017). These limitations spurred the development of alternatives to potent PPI modulators, as well as molecular tools based on a protein framework that compensates for antibody size, stability, and cost constraints while maintaining adequate affinity and specificity (Lofblom et al., 2011; Olaleye et al., 2021; David et al., 2023). These groups of proteins, termed non-immunoglobin binding proteins or scaffold proteins, are advancing protein research and paving the way for the expansion of affinity reagents. The applications are constantly being broadened, and new possibilities are being established (McCue and Kuhlman, 2022; Zajc et al., 2021; Shen and Dassama, 2023; Zhao et al., 2023; Qiao et al., 2023; Hannula et al., 2024; Shipunova and Deyev, 2022; Serna et al., 2022; Schmid et al., 2018) taking advantage of the significant number of proteins deposited in the protein databases (https://www.rcsb.org/), as well as their distinct features. Increased computational resources (Luo et al., 2022; Wang et al., 2022) and characterization methodologies, as well as the rapidly expanding in vitro and in vivo techniques for protein library construction and precise target recognition and selection (Simeon and Chen, 2018; Zimmermann et al., 2020), are enabling fresh applications to satisfy the ever-growing unmet human needs.

Therefore, in this study, I presented SD1 as a viable scaffold protein that met all of the requirements for the generation of affinity binders, including the potential for additional benefits in this field. SDI has a fold similar to the IgG's variable domain and is isolated from *Fervidobacterium islandicum*, a hyperthermophilic bacterium, expressible in *E. coli*, as well as provides the structural strength to withstand numerous side chain amino acid substitutions aimed at improving its properties for desired targeted functions. This presents SD1 as a suitable protein framework for developing effective PPI modulators for applications in diverse fields, particularly those needing extreme storage and deployment temperatures. However, the requirement for SD1 denaturation and refolding, since it is primarily expressed as insoluble proteins in *E. coli* inclusion bodies, posed a significant challenge in the design of target affinity binders.

In chapter two, I proposed that SD1 inclusion body formation could be due to high-level protein expression and that transiently surface-exposed hydrophobic residues during protein folding caused protein instability and subsequent aggregation, resulting in the formation of inclusion bodies in *E. coli*. I investigated increasing SD1's net charge and decreasing aggregation-prone regions within the proteins by inserting charged residues on the protein surfaces and aggregation gatekeeper residues in the aggregation-prone regions.

To test the feasibility of this approach, hexalysin was fused at the C-terminus of SD1, and when expressed in *E. coli*, the majority of the proteins emerged in the soluble fraction on SDS-PAGE. Encouraged by this finding, surface-exposed residues (Guex and Peitsch, 1997) and aggregation-prone regions (Fernandez-Escamilla et al., 2004) were identified and replaced with charged residues and aggregation gatekeeper residues using in silico site saturation mutagenesis, with the effect of each mutation on protein solubility being constantly evaluated (Schymkowitz et al., 2005).

SD1's solubility was significantly improved by replacing surface-exposed residues with charged amino acids and introducing gatekeeper residues within the numerous hydrophobic residues found in the protein. These findings corroborated previous reports, which stated that increasing protein net charge and pI causes electrostatic repulsive forces during protein folding, which facilitates correct folding while reducing interactions with transiently exposed hydrophobic side chains (Paraskevopoulou and Falcone, 2018; Kohara et al., 2022).

The following mutations, S26K and Q91P, were believed to produce instability in SD1, but interestingly, the S26K/Q91P mutant was found to be the most soluble of all the mutants studied. The stabilization of this mutant may be attributed to the favorable torsion angle in the location, given that proline residues in a polypeptide chain have rigorously constrained ϕ values of -60^{o} and no main chain amide hydrogen (Prajapati et al., 2007). Therefore, the observed increase in soluble expression of the S26K/Q91P mutant could be attributed to proline activity within the loop region, which increases rigidity, slows folding kinetics, and allows partially folded conformations that are prone to aggregation to fold completely (Glatzova et al., 2021). Furthermore, proline residues can maintain protein structure by decreasing the conformational entropy of the denatured phase (Prajapati et al. 2007). The S26K mutation may have resulted in beneficial coulomb interactions between the internally charged lysine residues and the surrounding negatively charged residues (Isom et al., 2011), stabilizing the protein structure and increasing the mutant solubility.

Finally, I generated ten mutants that showed increased soluble expression when compared to the WT SD1, but mutations aimed at increasing protein stability by introducing hydrophobic residues within the protein core and providing stability via the tyrosine corner by introducing pairs of interacting tyrosine partners within the protein reduced soluble expression when compared to the other soluble mutants.

In chapter three, the several generated mutants were characterized to understand better how protein surface charging, surface charge-charge distributions, pI, and fractions of surface polar, non-polar, and charged residues at different pH levels affect protein solubility and stability. This chapter's findings informed the choice of a suitable mutant for the design of the target affinity reagent.

The solubility of the purified lyophilized proteins was investigated at pH 4.7 and 7.7. Mutations aimed at SD1 surface charging and those aimed at reducing aggregation-prone regions both showed significant solubility at the two pH levels tested; however, mutations aimed at decreasing aggregation-prone regions showed significantly higher solubility when compared to the surface charging approach. The in vitro solubility results confirmed the soluble enhancement observed in vivo by SDS-PAGE, with the exception of the hexa-lysin tagged SD1 (SD1_6K), which was the most soluble protein in vivo expression in *E. coli* but was not significantly soluble in buffer solutions at pH 4.7 and 7.7. This was assumed to be due to the charge concentration at the protein's termini, as charge distribution has been demonstrated to contribute to protein stability and resistance to amorphous aggregation in solution (Kohara et al., 2022). The most soluble mutant, as in the in vivo experiment, was S26K/Q91P, which could be attributed to contributions from proline, which at the 91st position may have a -60 ϕ angle at the i + 1 position of a type II rotation to reduce conformational entropy while maintaining stability due to the favorable torsion angle (Hutchinson and Thornton, 1994). Additionally, the S26K core mutation in the protein could increase coulomb interactions, enhancing solubility.

When evaluating the storage stability and aggregation kinetics of the various constructed soluble mutants, intrinsic protein solubility is critical in retaining a higher protein concentration in solutions for more extended periods while exhibiting a decreased tendency for quick aggregation. Mutants with solution storage pH near the protein's pI exhibited lower solubility. This discovery is reinforced by a study conducted by Kramer et al. (2012), who observed that

protein solubility in solution is lowest when the solution pH approaches the protein pI. Similarly, at pH 4.7, the Y15K/W18R/S106R mutant's solubility and storage stability were considerably affected, showing that the increased basicity of the solution may have caused non-specific charge repulsion, compromising protein folding and stability in the solution. This assumption is supported by the report of Chi et al. (2003), which states that under such conditions, folded protein structures can become unstable due to increased charge repulsion.

Fractions of surface polar, non-polar, and charged residues on the several mutants were studied to understand better the contributions of the variations to the different solubility observations among the mutants, taking into account the positional contributions of these residues to protein folding and solubility in solution. All other mutants exhibited an unequal distribution of charged residues on the protein surface. WT SD1 had an equal distribution of positively and negatively charged surface residues, which may have contributed to protein instability and the formation of insoluble aggregates via attractive electrostatic interactions (Chi et al., 2003; Kim et al., 2012). The Y15K/W18R/S106R mutant was chosen as the best candidate for the design of a target affinity reagent because it had the highest fractions of surface polar and charged residues and the lowest fraction of surface non-polar groups when compared to the other solubility mutants. Proteins' structural stability in solution is determined by interactions with surrounding water molecules, and polar and charged amino acids significantly enhance these interactions, burying hydrophobic residues and limiting their interactions with water molecules in solution (Shaytan et al., 2009; Rose et al., 1985), as is typical of soluble hydrophilic proteins. Unfortunately, the S26K/Q91P mutant, which was deemed the most soluble, had the highest fraction of surface non-polar groups and the fewest polar and charged groups. Therefore, additional studies are required to understand better the mutant's folding mechanism, which is currently beyond the scope of this work. Consequently, the first mutant, Y15K/W18R/S106R, was designated SD1-M1 for the design of binding functions in the subsequent chapter.

Thermal stability experiments revealed that all mutants were thermally stable, with a melting temperature (Tm) of approximately 73 ºC for all protein samples. This demonstrated that the numerous SD1 mutations did not affect the protein's thermal stability. This finding supported the hypothesis that SD1 is mutationally resilient and able to tolerate and accommodate a wide range of side-chain modifications aimed at generating target binding function.

In Chapter 4, I designed the SD1-M1 library and evaluated it for TNF-α target binding after confirming its monovalent display on the M13 pIII surface protein. The monovalent display of the protein was preferred because it has the potential advantage of allowing for the efficient selection of binders with high affinity to the targets. In contrast, a multivalent display could result in a more substantial avidity effect but is frequently confronted with challenges of little or no target binding upon the production of soluble binding proteins (Kwasnikowski et al., 2005; Song et al., 2024). After five rounds of biopanning, the constructed library exhibited binding to the target as well as a 33-fold enrichment, indicating that the selected soluble mutant could be folded correctly to enable target binding. This observation is relevant because previous work in this lab demonstrated that the SD1 phage library binds mainly to the negative control, and enrichment could not be established due to the lack of binding to targets (data not shown). Given that the observed 33-fold enrichment was significantly lower than previous reports of 80 and 2815-fold enrichments for the VHH library against TNF-alpha (Reiter et al., 1999; Nie et al., 2021) and that elution of targets passively immobilized on maxisorp microtitre plates could result in enrichment of background target binders. Competitive elution was explored, and it demonstrated a considerable increase in target binders from the phage pool after only one cycle of biopanning based on the polyclonal phage ELISA.

The binding analysis results showed significant specificity; however, the low affinity indicates that the library should be optimized to increase the chances of selecting high-affinity binders. One of the significant limitations of this study was the amplification biases generated mainly

from highly infectious mutants, which often reduces the diversity, resulting in poor enrichments and reduced ability to identify high target binders. Consequently, further efforts are needed to focus on devising strategies to reduce the impact of unwanted biases and improve the dependability of selecting clones that actually bind to the target. Therefore, I propose performing sequence evaluation on clones selected after the third or fourth round of biopanning to extract information regarding potential binders by identifying the amino acid sequence space where functional mutations are improved in order to construct an optimized library that will contain potential binding clones from the enhanced sequence after multiple rounds of the affinity maturation process.

From this perspective, SD1 has shown significant intrinsic thermostability and a high level of mutational resilience required for acquiring new functions. Owing to the paucity of proteins that function under extreme conditions, the use of thermostable proteins as biosensors for environmental monitoring is emerging. The necessity for stability across a wide range of environmental conditions requires the use of thermophilic proteins, and SD1's ability to withstand extreme conditions makes it worthwhile in the design of immuno-sensors for environmental monitoring applications.

References

Adams, L. M., Andrews, R. J., Hu, Q. H., Schmit, H. L., Hati, S. and Bhattacharyya, S. (2019). Crowder-induced conformational ensemble shift in *Escherichia coli* Prolyl-tRNA synthetase. Biophysical Journal, 117(7), 1269-1284

Ahmadi, M. K. B., Mohammadi, S. A., Makvandi, M., Momouei, M., Rahmati, M., Dehghani, H. and Wood, D. W. (2021). Recent advances in the scaffold engineering of protein binders. Current Pharmaceutical Biotechnology, 22, 878-891.

Ajina, A. and Maher, J. (2018). Strategies to address chimeric antigen receptor signaling. Molecular Cancer Therapeutics, 17(9), 1795-1815.

Alnefaie, A., Albogami, S., Asiri, Y., Ahmad, T., Alotaibi, S. S., Al-Sanea, M. M. and Althobaiti, H. (2022). Chimeric antigen receptor T-cells: An overview of concepts, applications, limitations, and proposed solutions. Frontiers in Bioengineering and Biotechnology, 10, 797440.

Andris-Widhopf, J., Rader, C., Steinberger, P., Fuller, R. and Barbas, C. F. (2000). Methods for the generation of chicken monoclonal antibody fragments by phage display. Journal of Immunological Methods, 242(1-2), 159-181.

Anishchenko, 1., Pellock, S. J., Chidyausiku, T. M., Ramelot, T. A., Ovchinnikov, S., Hao, J., Bafna, K., Norn, C., Kang, A., Bera, A. K., DiMaio, F., Carter, L., Chow, C. M., Montelione, G. T. and Baker, D. (2021). De novo protein design by deep network hallucination. Nature, 600, 547-552.

Asaadi, Y., Jouneghani, F. F., Janani, S. and Rahbarizadeh, F. (2021). A comprehensive comparison between camelid nanobodies and single chain variable fragments. Biomarker Research, 9, 87.

Atsavapranee, B., Stark, C. D., Sunden, F., Thompson, S. and Fordyce, P. M. (2021). Fundamentals to function: Quantitative and scalable approaches for measuring protein stability. Cell Systems, 12, 547-560.

Azhar, A., Ahmad, E., Zia, Q., Rauf, M. A., Owais, M. and Ashraf, G. M. (2017). Recent advances in the development of novel protein scaffolds-based therapeutics. International Journal of Biological Macromolecules, 102, 630-641.

Baek, M., DiMaio, F., Anishchenko, I., Daupara, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinc, L. N., Schaeffer, D. R., Millan, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, C. K., Grishin, N. V., Adams, P. D., Read, R. J. and Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science, 373, 871-876.

Bai, C. and Warshel, A. (2019). Revisiting the protomotive vectorial motion of F_0-ATPase. Proceedings of the National Academy of Sciences, 116, 19484-19489.

Bakail, M. and Ochsenbein, F. (2016). Targeting protein-protein interactions, a wide open field for drug design. Comptes Rendus Chimie, 19, 19-27.

Balkwill, F. (2006). TNF-α in promotion and progression of cancer. Cancer Metastasis Reviews, 25, 409-416.

Barroso, J. R. M. S., Mariano, D., Dias, S. R., Rocha, R. E. O., Santos, L. H., Nagem, R. A. P and De Melo-Minardi, R. C. (2020). Proteus: An algorithm for proposing stabilizing mutation pairs based on interactions observed in known protein 3D structures. BMC Bioinformatics, 21, 1-21.

Baumann, G., Froommel, C., Sander, C. (1989). Polarity as a criterion in protein design. Protein Engineering, 2, 329-334.

Bavishi, K., Li, D., Eiersholt, S., Hooley, E. N., Petersen, T. C., Moller, B. L., Hatzakis, N. S. and Laursen, T. (2018). Direct observation of multiple conformational staes of in chromosome P450 oxideoreductase and their modulation by membrane environment and ionic strength. Scientific Reports, 8, 6817.

Bedford, R., Tiede, C., Hughes, R., Curd, A., McPherson, M. J., Peckham, M. and Tomlinson, D. C. (2017). Biophys Rev, 9, 299-308.

Beerten, J., Jonckheere, W., Rudyak, S., Xu, J., Wilkinson, H., De Smet, F., Schymkowitz, J. and Rousseau, F. (2012). Aggregation gatekeepers modulate protein homeostasis of aggregating sequences and affect bacterial fitness. Protein Engineering, Design and Selection, 25, 357-366.

Bekes, M., Langley, D. R. and Crews, C. M. (2022). PROTAC targeted protein degraders: the past is prologue. Nature Reviews, 21, 181-200.

Belvisi, L., D'Andrea, L. D. and Jimenez, M. A. (2021). Editorial: peptides targeting protein-protein interactions: methods and applications. Frontiers in Molecular Biosciences, 8, 780106. doi: 10.3389/fmolb.2021.780106.

Bever, C. S., Dong, J., Vasylieva, N., Barnych, B., Cui, Y., Xu, Z., Hammock, B. D. and Gee, S. J. (2016). Analytical and Bioanalytical Chemistry, 408(22), 5985-6002.

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X. and Zhou, T. (2021). Challenges associated with the formation of recombinant protein inclusion bodies in *Escherichia coli* and strategies to address them for industrial applications. Frontiers in Bioengineering and Biotechnology, 9, 630551. Doi: 10.3389/fbioe.2021.630551.

Binz, H. K., Amstutz, P. and Pluckthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. Nature Biotechnology, 23, 1257-1268.

Bournazos, S., Gupta, A. and Ravetch, J. V. (2020). The role of IgG Fc receptors in antibody-dependent enhancement. Nature Reviews Immunology, 20, 633-643.

Byla, P., Andersen, M. H., Holtet, T. L., Jacobsen, H., Munch, M., Gad, H. H., Thogersen, H. C. and Hartmann, R. (2010). Selection of a novel and highly specific tumor necrosis factor α (TNFα) antagonist: Insight from the crystal structure of the antagonist-TNFα complex. The Journal of Biological Chemistry, 285, 12096-12100.

Cao, L., Coventry, B., Goreshnik, I., Huang, B., Sheffler, W., Park, J. S., Jude, K. M., Markovic, I., Kadam, R. U., Verschueren, K. H. G., Verstraete, K., Walsh, S. T. R., Bennett, N., Phal, A., Yang, A., kozodoy, L., DeWitt, M., Picton, L., Miller, L., Strauch, E., DeBouver, N. D., Pires, A., Bera, A. K., Halabiya, S., Hammerson, B., Yang, W., Bernard, S., Stewart, L., Wilson, I. A., Ruohola-Baker, H., Schlessinger, J., Lee, S., Savvides, S. N., Garcia, K. C. and Baker, D. (2022). Design of protein-binding proteins from the target structure alone. Nature, 605, 551-589.

Carrasco-Lopez, C., Zhao, E. M., Gil, A. A., Alam, N., Toettcher, J. E. and Avalos J. L. (2020). Development of light-responsive protein binding in the monobody non-immunoglobulin scaffold. Nature Communications, 11, 4045.

Chames, P., Van Regenmortel, M., Weiss, E. and Baty, D. (2009). Therapeutic antibodies: successes, limitations and hopes for the future. British Journal of Pharmacology, 157(2), 220-233.

Chi, E. Y., Krishnan, S., Randolph, T. W. and Carpenter, J. F. (2003). Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation, Pharmaceutical Research, 20, 1325–1336.

Chiu, M. L., Goulet, D. R., Teplyakov, A. and Gilliland, G. L. (2019). Antibody structure and function: the basis for engineering therapeutics. Antibodies, 8(4), 55.

Cruz, E. and Kayser, V. (2019). Monoclonal antibody therapy of solid tumors: clinical limitations and novel strategies to enhance treatment efficacy. Biologics: Targets and Therapy: 12, 33-51.

Datta, S., Christena, L. R., and Rajaram, Y. R. S. (2013). Enzyme immobilization: an overview on techniques and support materials. 3 Biotech, 3, 1-9.

David, I. T., Pestov, N. B., Korneenko, T. V. and Barlev, N. A. (2023). Non-immunoglobulin synthetic binding proteins for oncology. Biochemistry (Moscow), 88, 1232-1247.

David, R. L. (2014). Biochemistry; in Haynes, W. D. (ed), CRC Handbook of Chemistry and Physics, CRC Press, Boca Raton, 95th edition, 2014-2015, section 7-1.

De, A., Kuppusamy, G. and Karri, V. V. S. R. (2018). Affibody molecules for molecular imaging and targeted drug delivery in the management of breast cancer. International Journal of Biological Macromolecules, 107, 906-919.

Desiderio, A., Franconi, R., Lopez, M., Villani, M. E., Viti, F., Chiaraluce, R., Consalvi, V., Neri, D. and Benvenuto, E. (2001). A semi-synthetic repertoire of intrinsically stable antibody fragments derived from a single framwork scaffold. Journal of Molecular Biology, 310, 603-615.

Dhar, A., Samiotakis, A., Ebbinghaus, S., Nienhaus, L., Homouz, D., Gruebele, M. and Cheung, M. S. (2010). Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. Proceedings of the National Academy of Sciences, 107, 17586-17591.

Dias, A. M. G. C. and Roque, A. C. A. (2017). The future of protein scaffolds as affinity reagents for purification. Biotechnology and Bioengineering, 114(3), 481-491.

DiRusso, C. J., Dashtiahanger, M. and Gilmore, T. D. (2022). Scaffold proteins as dynamic integrators of biological processes. Journal of Biological Chemistry, 298 (12), 102628.

Dreier, B., Honegger, A., Hess, C., Nagy-Davidescu, G., Mittl, P. R. E., Grutter, M. G., Belousova, N., Mikheeva, G., Krasnykh, V. and Pluckthun, A. (2013). Development of a generic adenovirus delivery system based on structure-guided design of bispecific trimeric DARPin adapters. PNAS, 110(10), E869-E877.

Drozdetskiy, A., Cole, C., Procter, J. and Barton, G. J. (2015). Jpred4: a protein secondary structure prediction server. Nucleic Acids Research, 43, W389-W394.

Duan, Z. and Siegumfeldt, H. (2010). An efficient method for isolating antibody fragments against small peptides by antibody phage display. Combinatorial Chemistry and High Throughput Screening, 13, 818-828.

Eklund, M., Sandstrom, K., Teeri, T. T. and Nygren, P. (2004). Site-specific and reversible anchoring of active proteins onto cellulose using a cellulosome-like complex. Journal of Biotechnology, 109, 277-286.

Esposito, E. and Cuzzocrea, S. (2009). TNF-alpha as a therapeutic target in inflammatory diseases, ischemia-reperfusion injury and trauma. Current Medicinal Chemistry, 16(24), 3152-3167.

Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins, Nature Biotechnology, 22, 1302-1306.

Ferrigno, P. K. (2016). Non-antibody protein-based biosensors. Essays in Biochemistry, 60, 19-25.

Finch, A. J. and Kim, J. R. (2018). Thermophilic proteins as versatile scaffolds for protein engineering. Microorganisms, 6, 97.

Frei, J. C. and Lai, J. R. (2016). Protein and antibody engineering by phage display. Methods in Enzymology, 580, 45-87.

Frigault, M. J., Lee, J., Basil, M. C., Carpenito, C., Motohashi, S., Scholler, J., Kawalekar, O. U., Guedan, S., McGettigan, S. E., Posey Jr., A. D., Ang, S., Cooper, L. J. N., Platt, J. M., Johnson, F. B., Paulos, C. M., Zhao, Y., Kalos, M., Milone, M. C. and June, C. H. (2015). Identification of chimeric antigen receptors that mediate constitutive or inducible ploriferation of T cells. Cancer Immunology Research, 3(4), 356-367.

Gad, S. and Ayakar, S. (2021). Protein scaffolds: a tool for multi-enzyme assembly. Biotechnology Reports, 32, e00670.

Ganesan, A., Siekierska, A., Beerten, J., Brams, M., Van Durme, J., De Baets, G., Van Der Kant, R., Gallardo, R., Ramakers, M., Langenberg, T., Wilkinson, H., De Smet, F., Ulens, C., Rousseau, F. and Schymkowitz, J. (2016). Structural hot spots for the solubility of globular proteins. Nature Communications, 7, 10816.

Gebauer, M. and Skerra, A. (2009). Engineered protein scaffolds as next-generation antibody therapeutics. Current Opinion in Chemical Biology, 13(3), 245-255.

Gebauer, M. and Skerra, A. (2020). Engineered protein scaffolds as next-generation therapeutics. Annual Reviews of Pharmacology and Toxicology, 60, 391-415.

Gil, A. A., Carrasco-Lopez, C., Zhu, L., Zhao, E. M., Ravindran, P. T., Wilson, M. Z., Goglia, A. G., Avalos, J. L. and Toettcher, J. E. (2020). Optogenetic control of protein binding using light-switchable nanobodies. Nature Communications, 11, 4044.

Gilbreth, R. N. and Koide, S. (2012). Structural insights for engineering binding proteins based on non-antibody scaffolds. Current Opinion in Structural Biology, 22(4), 413-420.

Glatzova, D., Mavila, H., Saija, M. C., ChuM, t., Cwiklik, L., Brdicka, T. and Cebecauer, M. (2021). The role of prolines and glycine in the transmembrane domain of LAT. FEBS Journal, 288, 4039-4052.

Godde, C., Sahm, K., Brouns, S. J. J., Kluskens, L. D., van der Oost, J., de Vos, W. M. and Garabed A. (2005). Cloning and expression of islandisin, a new thermostable subtilisin from *Fervidobacterium islandicum*, in *Escherichia coli*. Applied and Environmental Microbiology, 71(7), 3951-3958.

Greenfield, N.J. (2007). Using circular dichroism spectra to estimate protein secondary structure. Nature Protocols, 1, 2876–2890.

Gronwall, C. and Stahl, S. (2009). Engineered affinity proteins – Generation and applications. Journal of Biotechnology, 140, 254-269.

Guex, N. and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis, 18, 2714-2723.

Hannula, L., Kuivanen, S., Lasham, J., Kant, R., Kareinen, L., Bogacheva, M., Strandin, T., Sironen, T., Hepojoki, J., Sharma, V., Saviranta, P., Kipar, A., Vapalahti, O., Huiskonen, J. T. and Rissanen, I. (2024). Nanobody engineering for SARS-CoV-2 neutralization and detection. Microbiology Spectrum, 12(4), 1-18.

Hantschel, O., Biancalana, M. and Koide, S. (2021). Monobodies are enabling tools for structural and mechanistic biology. Current Opinion in Structural Biology, 60, 167-174.

Harmansa, S. and Affolter, M. (2018). Protein binders and their applications in developmental biology. Development, 145, dev148874. Doi:10.1242/dev.148874.

Hata, Y. and Lida, J. (2009). Scaffold protein. In: Binder, M. D., Hirokawa, N. and Windhorst, U. (Eds). Encyclopedia of Neuroscience. Springer, Berlin, Heidelberg. pp. 3313-3316. https://doi.org/10.1007/978-3-540-29678-2_5231

Hirao, A., Nagatoishi, S., Ikeuchi, E., Yamawaki, T., Mori, C., Nakakido, M. and Tsumoto, K. (2023). Design of single-domain VHH antibodies to increase the binding activity in SPR amine coupling. Biochemical and Biophysical Research Communication, 663, 54-60.

Hober, S., Lindbo, S. and Nilvebrant, J (2019). Bispecific applications of non-immunoglobulin scaffold binders. Methods, 154, 143-152.

Hoffmann, A., Kovermann, M., Lilie, H., Fiedler, M., Balbach, J., Rudolph, R. and Pfeifer, S. (2012). New binding mode of TNF-alpha revealed by ubiquitin-based artificial binding protein. PLoS ONE, 7(2), e31298.

Holm, L. and Sander, C. (1995). Dali: a network tool for protein structure comparison. Trends in Biochemical Sciences, 20 (11), 478-480.

Hosse, R. J., Rothe, A. and Power, B. E. (2006). A new generation of protein display scaffolds for molecular recognition. Protein Science, 15, 14-27.

Hou, Q., Bourgeas, R., Pucci, F. and Rooman, M. (2018). Computational analysis of the amino acid interactions that promote or decrease protein solubility, Scientific Reports, 8, 1–13.

Huang, R., Fang, P. and Kay, B. K. (2013). Improvements to the Kunkel mutagenesis protocol for constructing primary and secondary phage-display libraries. Methods, 58(1), 10-17.

Huang, W. C., Ellis, J., Moody, P. C. E., Raven, E. L. and Roberts, G. C. K. (2013). Redox-linked domain movements in the catalytic cycle of cytochrome P450 reductase. Structure, 21, 1581-1589.

Huber, R., Woese, C. R., Langworthy, T. A., Kristjansson, J. K. and Stetter, K. O. (1990). *Fervidobacterium islandicum* sp. nov., a new extremely thermophilic eubacterium belonging to the “Thermotogales”. Archives of Microbiology, 154, 105-111.

Hutchinson, E. G. and Thornton, J. M. (1994). A revised set of potentials for beta-turn formation in proteins. Protein Science, 3, 2207-2216.

Isom, D. G., Castañed, C. A., Cannon, B. R. and García-Moreno, B. E. (2011). Large shifts in pKa values of lysine residues buried inside a protein. Proc Natl Acad Sci U S A, 108, 5260–5265.

Isom, D. G., Castañeda, C. A., Cannon, B. R., Velu, P. D., and García-Moreno, E. B. (2010). Charges in the hydrophobic interior of proteins, Proc Natl Acad Sci U S A, 107, 16096–16100.

Jang, D., Lee, A., Shin, H., Song, H., Park, J., Kang, T., Lee, S. and Yang, S. (2021). The role of tumor necrosis factor alpha (TNF-α) in autoimmune disease and current TNF-α inhibitors in therapeutics. International Journal of Molecular Sciences, 22, 2719.

Jarrot, P. and Kaplanski, G. (2014). Anti-TNF-Alpha therapy and systemic vasculitis. Mediators of Inflammation, volume 2014, article ID 493593, 1-9. http://dx.doi.org/10.1155/2014/493593.

Jefferis, R. (2016). Posttranslational modifications and the immunogenicity of biotherapeutics. Journal of Immunology Research, 2016:5358272. doi: 10.1155/2016/5358272.

Jenkins, P. T., Fryer, T., Dehli, R. I., Jurgensen, J. A., Fuglsang-Madsen, A., Fons, S. and Lausten, A. H. (2019). Toxin neutralization using alternative binding proteins. Toxins, 11, 53; doi.10.3390/toxins11010053.

Josephs, S. F., Ichim, T. E., Prince, S. M., Kesari, S., Marincola, F. M., Escobedo, A. R. and Jafri, A. (2018). Unleashing endogenous TNF-alpha as a cancer immunotherapeutic. Journal of Translational Medicine, 16, 242.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A. et al. (2021). Highly accurate protein structure prediction with Alphafold. Nature, 596, 583-589.

Kato, A., Maki, K., Ebina, T., Kuwajima, K., Soda, K. and Kuroda, Y. (2007). Mutational analysis of protein solubility enhancement using short peptide tags. Biopolymers, 85, 12-18.

Keane, J. T. and Posey Jr, A. D. (2021). Chimeric antigen receptors expand the repertoire of antigenic macromaolcules for cellular immunity. Cells, 10 (12), 3356.

Kelm, J. M., Pandey, D. S., Malin, E., Kansou, H., Arora, S., Kumar, R. and Gavande, N. S. (2023). PROTAC'ing oncoproteins: targeted protein degradation for cancer therapy. Molecular Cancer, 22, 62. https://doi.org/10.1186/s12943-022-01707-5

Khan, M. A., Islam, M. M. and Kuroda, Y. (2013). Analysis of protein aggregation kinetics using short amino acid peptide tags. Biochemica et Biophysica Acta, 1834, 2107-2115.

Kille, S., Acevedo-Rocha, C. G., Parra, L. P., Zhang, Z., Opperman, D. J., Reetz, M. T. and Acevedo, J. P. (2012). Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. ACS Synthetic Biology, 2, 83-92.

Kim, J. S., Kluskens, L. D., De Vos, W. M., Huber, R. and Van Der Oost, J. (2004). Crystal structure of fervidolysin from *Fervidobacterium pennivorans*, a keratinolytic enzyme related to subtilisin. Journal of Molecular Biology, 335, 787-797.

Kim, S. G., Min, W. K., Rho, Y. T. and Seo, J. H. (2012). Electrostatic interaction-induced inclusion body formation of glucagon-like peptide-1 fused with ubiquitin and cationic tag, Protein Expression and Purification, 84, 38–46.

Kitching, A. R. and Jaw, J. (2018). Chimeric antigen receptor T (CAR T) cells: another cancer therapy with potential applications in kidney disease and transplantation. Kidney International, 94, 4-6.

Kneuttinger, A. C. (2022). A guide to designing photocontrol in proteins: methods, strategies and applications. Biological Chemistry, 403(5-6), 573-613.

Kohara, S., Matsuzawa, Y. and Kuroda, Y. (2022). Lattice-model analysis of the effect of protein surface charge distribution on amorphous aggregation and condensation. Chemical Physics Letters, 802, 139767.

Koide, A., Bailey, C. W., Huang, X. and Koide, S. (1998). The fibronectin type III domain as a scaffold for novel binding proteins. Journal of Molecular Biology, 284, 1141-1151.

Koide, A., Wojcik, J., Gilbreth, R. N., Reichel, A., Peihler, J. and Koide, S. (2009). Accelerating phage-display library selection by reversible and site-specific biotinylation. Protein Engineering, Design and Selection, 22(11), 685-690.

Kougentakis, C. M., Majumdar, A. and Bertrand Garcia-Moreno, E. (2020). Electrostatic effects in proteins are governed by pH-redistribution of the conformational ensemble. Biophysics, **doi:** https://doi.org/10.1101/2020.02.02.931253

Kramer, R. M., Shende, V. R., Motl, N., Pace, C. N., and Scholtz, J. M. (2012). Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility, Biophysical Journal, 102, 1907.

Kronqvist, N., Lofblom, J., Jonsson, A., Wernerus, H. and Stefan S. (2008). A novel affinity protein selection system based on the staphylococcal cell surface display and flow cytometry. Protein Engineering, Design and Selection, 21, 247-255.

Kwasnikowski, P., Kristensen, P. and Markiewicz, W. (2005). Multivalent display systems on filamentous bacteriophage pVII minor coat protein. Journal of Immunological Methods, 307, 135-143.

Lai, J. R., Chen, G. and Sidhu, S. S. (2018). Synthetic antibody engineering: Concepts and Applications; in Putterman, C., Cowburn, D. and Almo, S. (ed), Structural Biology in Immunology, Academic Press, pp 81-100. ISBN 9780128033692, https://doi.org/10.1016/B978-0-12-803369-2.00003-6.

Land, H. and Humble, M. S. (2018). YASARA: A tool to obtain structural guidance in biocatalytic investigations. Methods in Molecular Biology, 1685, 43-67.

Lawrence, M. S., Philips, K. J., Liu, D. R. (2007). Supercharging proteins can impart unusual resilience, Journal of American Chemical Society, 129, 10110-10112.

Lee, W. S., Wheatley, A. K., Kent, S. J. and DeKosky, B. J. (2020). Antibody-dependent enhancement and SARS-CoV-2 vaccines and therapies. Nature Immunology, 5, 1185-1191.

Liljas, A., Liljas, L., Lindblom, G., Nissen, P., Kjeldgaard, M., Ash, M. R. (2016). Testbook of structural biology, World scientific.

Limsakul, P., Peng, Q., Wu, Y., Allen, M. E., Liang, J., Remacle, A. G., Lopez, T., Ge, X., Kay, B. K., Zhao, H., Strongin, A. Y., Yang, X., Lu, S. and Wang, Y. (2018). Directed evolution to engineer monobody for FRET biosensor assembly and imaging in live-cell surface. Cell Chemical Biology, 25, 370-379.

Linse, S. S. (2018). Directed evolution of enzymes and binding proteins in “scientific background on the noble price in chemistry 2018”. The Royal Swedish Academy of Sciences. pp1-24.

Liu, B., Long, S. and Liu, J. (2020). Improving the mutagenesis efficiency of the Kunkel method by codon optimization and annealing temperature adjustment. New Biotechnology, 56, 46-53.

Lofblom, J., Frejd, F. Y. and Stahl, S. (2011). Non-immunoglobulin based protein scaffolds. Current Opinion in Biotechnology, 22, 843-848.

Long, A. H., Haso, W. M., Shern, J. F., Wanhainen, K. M., Murgai, M., Ingaramo, M., Smith, J. P., Walker, A. J., Kohler, M. E., Venkateshwara, V. R., Kaplan, R. N., Patterson, G. H., Fry, T. J., Orentas, R. J. and Mackall, C. L (2015). 4-1BB costimulation ameliorates T cell exhaustion induced by tonic signaling of chimeric antigen receptors. Nature Medicine, 21(6), 581-590.

Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R. and Shi, J. (2020). Recent advances in the development of protein-protein interactions modulators: mechanisms and clinical trials. Signal Transduction and Targeted Therapy, 5, 213.

Luo, R., Liu, H. and Cheng, Z. (2022). Protein scaffolds: antibody alternatives for cancer diagnosis and therapy. RSC Chemical Biology, 3, 830 – 847.

Mabonga, L. and Kappo, A. P. (2019). Protein-protein interaction modulators: advances, successes and remaining challenges. Biophysical Reviews, 1194), 559-581.

Maghraby, Y. R., El-Shabasy, R. M., Ibrahim, A. H. and Azzazy, H. M. E. 2023). Enzyme immobilization technologies and industrial applications. ACS Omega, 8, 5184-5196.

Martin, M. L., Bedford, R., Heseltine, S. J., Tang., A.A., Haza, K. Z., Rao, A., McPherson, M. J. and Tomlinson, D. C. (2018). Non-immunoglobulin scaffold proteins: Precision tools for studying protein-protein interactions in cancer: New Biotechnology, 45, 28-35.

McCue, A. C. and Kuhlman, B. (2022). Design and engineering of light-sensitive protein switches. Current Opinion in Structural Biology, 74, 102377.

Miersch, S. and Sidhu, S. S. (2016). Intracellular targeting with engineered proteins. F1000Research, doi:10.12688/f1000research.8915.1

Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovichinnikov, S. and Steinegger, M. (2022). ColaFold: making protein folding accessible to all. Nature Methods, 19, 679-682.

Mitra, A., Barua, A., Huang, L., Ganguly, S., Feng, Qin, and He, B. (2023). From bench to bedside: the history and progress of CAR T cell therapy. Frontiers in Immunology, 14, 1188049.

Mullard, A. (2016). 2015 FDA drug approvals. Nature Reviews Drug Discovery, 15, 73-76.

Munoz-Lopez, P., Ribas-Aparicio, R. M., Becerra-Baez, E. I., Fraga-Perez, K., Flores-Martinez, L. F., Mateos-Chavez, A. A. and Luria-Perez, R. (2022). Single-chain fragment variable:

Recent progress in cancer diagnosis and therapy. Cancers, 14(17), 4206. doi: 10.3390/cancers14174206

Nie, J., Ma, X., Hu, F., Miao, H., Feng, X., Zhang, P., Han, M. H., You, F., Yang, Y., Zhang, W. and Zheng, W. (2021). Designing and constructing a phage display synthesized single domain antibodies library based on camel VHHs frame for screening and identifying humanized TNF-α specific nanobody. Biomedicine and Pharmacotherapy, 137, 111328

Norn, C., Wicky, B. M., Juergens, D., Liu, S., Kim, S., Tischer, D., Koepnick, B., Anishchenko, I., Players, F., Baker, D. and Ovchinnikov, S. (2021). Protein sequence design by conformational landscape optimization. PNAS, 118(11), e2017228118.

Nuttall, S. D., Irving, R. A and Hudson, P. J. (2000). Immunoglobulin VH domains and beyond: design and selection of single-domain binding and targeting reagents. Current Pharmaceutical Biotechnology, 1(3), 253-263.

Nygren, P. and Skerra, A. (2004). Binding proteins from alternative scaffolds. Journal of Immunological Methods, 290, 3-28.

Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., German Grehov, O. G., Kandrov, D., Rasputin, K., Syabro, M. and Tleukenov, T. (2012). Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics, 28, 1166-1167.

Olaleye, O., Govorukhina, N., van de Merbel, N. and Bischoff, R. (2021). Non-antibody-based binders for the enrichment of proteins for analysis by mass spectrometry. Biomolecules, 11(12), 1791.

Oliver, H. (2017). Monobodies as possible next-generation protein therapeutics – a perspective. Swiss Medical Weekly, 147, w14545.

Ononugbo, C. M., Shimura, Y., Yamano-Adachi, N., Omasa, T. and Koga, Y. (2024). Rational design approach to improve the solubility of the β-sandwich domain 1 of a thermophilic protein. Journal of Bioscience and Bioengineering, 138(4), 271-282.

Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Schell, D., Thurlkill, R. L., Imura, S., Scholtz, J. M., Gajiwala, K., Sevcik, J., Urbanikova, L., Myers, J. K., Takano, K., Hebert, E. J., Shirley, B. A. and Grimsley, G. R. (2014). Contribution of hydrogen bonds to protein stability. Protein Science, 23, 652.

Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Shirley, B. A., Hendricks, M. M. N., Iimura, S., Gajiwala, K., Scholtz, J. M. and Grimsley, G.R. (2011). Contribution of hydrophobic interactions to protein stability. Journal of Molecular Biology, 408, 514–528.

Pandey, K. R. and Mehrotra, S. (2024). Engineering high affinity antigen-binders: Beyond conventional antibodies. Advances in Protein Chemistry and Structural Biology, 140, 37-57.

Paraskevopoulou, V. and Falcone, F. H. (2018). Polyionic tags as enhancers of protein solubility in recombinant protein expression, Microorganisms, 6, 47.

Petrovskaya, L. E., Shingarova, L. N., Kryukova, E. A., Boldyreva, E. F., Yakimov, S. A., Guryanova, S. V., Novoseletsky, V. N., Dolgikh, D. A. and Kirpichnikov, M. P. (2012). Construction of TNF-binding proteins by grafting hypervariable regions of F10 antibody on human fibronectin domain scaffold. Biochemistry (Moscow), 77, 62-70.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. and Ferrin, T. E. (2021). USCF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Science, 30, 70-82.

Porebski, B. T., Nickson, A. A., Hoke, D. E., Hunter, M. R., Zhu, L., McGowan, S., Webb, G. I., and Buckle, A. M. (2015). Structural and dynamic properties that govern the stability of an engineered fibronectin type III domain, Protein Engineering Design and Selection, 28, 67–78.

Prajapati, R. S., Das, M., Sreeramulu, S., Sirajuddin, M., Srinivasan, S., Krishnamurthy, V., Ranjani, R., Ramakrishnan, C. and Varadarajan, R. (2007). Thermodynamic effects of proline introduction on protein stability. Proteins, 66, 480-491.

Qiao, H., Li, L., Wang, L., Yu, H., Hu, F., Zhou, X., Yang, H., Xu, J., Meng, X., Geng, Y. and Dai, Y. (2023). Preparation and characterization of nanobodies targeting SARS-CoV-2 papain-like protease. Protein Expression and Purification, 207, 106267.

Qing, R., Hao, S., Smorodina, E., Jin, D., Zalevsky, A. and Zhang, S. (2022). Protein design: From the aspect of water solubility and stability. Chemical Reviews, 122, 14085-14179.

Ramamurthy, V., Krystek, S. R., Bush, A., Wei, A., Emanuel, S. L., Gupta, R. D., Janjua, A., Cheng, L., Murdock, M., Abramczyk, B., Cohen, D., Lin, Z., Morin, P., Davis, J. H., Dabritz, M., McLaughlin, D. C., Russo, K. A., Chao, G., Wright, M. C., Jenny, V. A., Engle, L. J., Furfine, E. and Sheriff, S. (2012). Structures of adnectin/protein complexes reveal an expanded binding footprint. Structure, 20, 259 – 269.

Reiter, Y., Shuck, P., Boyd, L. F. and Plaksin, D. (1999). An antibody single-domain phage display library of a native heavy chain variable region: isolation of functional single-domain VH molecules with unique interface. Journal of Molecular Biology, 290, 685-698.

Rodriguez-Nava, C., Ortuno-Pineda, C., Illades-Aguiar, B., Flores-Alfaro, E., Leyva-Vazquez, M. A., Parra-Rojas, I., del Moral-Hernandez, O., Vences-Velazquez, A., Cortes-Sarabia, K. and Alarcon-Romero, L. C. (2023). Mechanisms of action and limitations of monoclonal antibodies and single chain fragment variable (scFv) in the treatment of cancer. Biomedicines, 11(6), 1610.

Rosano, G. L. and Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. Frontiers in Microbiology, 5(172), 1- 17.

Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., Zehfus, M. H. (1985). Hydrophobicity of amino acid residues in globular proteins. Science, 229, 834-838.

Sachsenhauser, V., and Bardwell, J. C. (2018). Directed evolution to improve protein folding in vivo. Current Opinion in Structural Biology, 48, 117-123.

Samaranayake, H., Wirth, T., Scenkwein, D., Raty, J. K. and Yla-Herttuala, S. (2009). Challenges in monoclonal antibody-based therapies. Annals of Medicine, 41, 322-331.

Sant'Anna, R., Braga, C., Varejao, N., Pimenta, K. M., Grana-Montes, R., Alves, A., Cortines, J., Cordeiro, Y., Ventura, S. and Foguel, D. (2014). The importance of a gatekeeper residue on the aggregation of transthyretin: Implications for transthyretin-related amyloidoses. Journal of Biological Chemistry, 289, 28324.

Sato., Y., Okano, K., Kimura, H. and Honda, K. (2020). TEMPURA: Database of growth TEMPeratures of Unsual and RAre prokaryotes. Microbes and Environments, 35, ME20074.

Schmid, M., Ernst, P., Honegger, A., Suomalainen, M., Zimmermann, M., Braun, L., Stauffer, S., Thom, C., Dreier, B., Eibauer, M., Kipar, A., Vogel, V., Greber, U. F., Medalia, O. and Pluckthun, A. (2018). Adenoviral vector with shield and adapter increases tumor specificity and escapes liver and immune control. Nature Communications, 9, 450.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005). The FoldX web server: an online force field. Nucleic Acids Research, 33, W382-W388.

Serna, N., Pallares, V., Unzueta, U., Garcia-Leon, A., Volta-Duran, E., Sanchez-Chardi, A., Parlade, E., Rueda, A., Casanova, I., Falgas, A., Alba-Castellon, L., Sierra, J., Villaverde, A., Vazquez, E. and Mangues, R. (2022). Engineering non-antibody human proteins as efficient scaffolds for selective, receptor-targeted drug delivery. Journal of Controlled Release, 343, 277-287.

Sethi, G., Sung, B., Kunnumakkara, A. B. and Aggarwal, B. B. (2009). Targeting TNF for treatment of cancer and autoimmunity. Advances in Experimental Medicine and Biology, 647, 37-51.

Sethi, J. K. and Hotamisligil, G. S. (2021). Metabolic Messengers: tumor necrosis factor. Nature Metabolism, 3, 1302-1312.

Sha, F., Salzman, G., Gupta, A. and Koide, S. (2017). Monobodies and other synthetic binding proteins for expanding protein science. Protein Science, 26, 910-924.

Shaytan, A. K., Shaitan, K. V., Khokhlov, A. R. (2009). Solvent accessible surface area of amino acid residues in globular proteins: correlation of apparent transfer free energies with experimental hydrophobicity scales. Biomacromolecules, 10, 1224-1237.

Shen, F. and Dassama, L. M. K. (2023). Opportunities and challenges of protein-based targeted protein degradation. Chemical Science, 14, 8433-8447.

Shepard, H. M., Phillips, G. L., Thanos, C. D. and Feldmann, M. (2017). Developments in therapy with monoclonal antibodies and related proteins. Clinical Medicine, 17(3), 220-232.

Shingarova, L. N., Petrovskaya, L. E., Zlobinov, A. V., Gapizov, S. S., Kryukova, E. A., Birikh, K. R., Boldyreva, E. F., Yakimov, S. A., Dolgikh, D. A. and Kirpichnikov, M. P. (2018). Construction of artificial TNF-binding proteins based on the 10th human fibronectin type III domain using bacterial display. Biochemistry (Moscow), 83, 708-716

Shipunova, V. O. and Deyev, S. M. (2022). Artificial scaffold polypeptides as an efficient tool for the targeted delivery of nanostructures in vitro and in vivo. Acta Naturae, 14(1), 54-72.

Sidhu, S. S., Lowman, H. B., Cunningham, B. C. and Wells, J. A. (2000). Phage display for selection of novel binding peptides. Methods in Enzymology, 328, 333-363.

Silva, D-A., Yu, S., Ulge, U. Y., Spangler, J. B., Jude, K. M., Labao-Almeida, C., Ali, L. R., Quijano-Rubio, A., Ruterbusch, M., Leung, I., Biary, T., Crowlwy, S. J., Marcos, E., Walkey, C. D., Weitzner, B. D., Pardo-Avila, F., Castellanos, J., Carter, L., Stewart, L., Riddell, S. R., Pepper, M., Bernardes, G. J. L., Dougan, M., Garcia, K. C. and Baker, D. (2019). De novo design of potent and selectiove mimics of IL-2 and IL-15. Nature, 565, 186 – 214.

Simeon, R. and Chen, Z. (2018). In vitro-engineered non-antibody protein therapeutics. Protein Cell, 9(1), 3-14.

Skerra, A. (2001). 'Anticalins': a new class of engineered ligand-binding proteins with antibody-like proteins. Reviews in Molecular Biotechnology, 74, 257-275.

Skerra, A. (2007). Alternative non-antibody scaffolds for molecular recognition. Current Opinion in Biotechnology, 18, 295-304.

Smolen, J. S. and Aletaha, D. (2015). Rheumatoid arthritis therapy reappraisal: strategies, opportunities and challenges. Nature Reviews Rhematology, 11, 276-289.

Song, B. P. C., Ch'ng, A. C. W. and Lim, T. S. (2024). Review of phage display: A jack-of-all-trades and master of most biomolecule display. International Journal of Biological Macromolecules, 256, 128455.

Sterner, R. C. and Sterner, R. M. (2021). CAR-T cell therapy: current limitations and potential strategies. Blood Cancer Journal, 11, 69.

Stoffer, M. A., Schoels, M. M., Smolen, J. S., Aletaha, D., Breedveld, F. C., Burmester, G., Bykerk, V., Dougados, M., Emery, P., Haraoui, B., Gomez-Reino, J., Kvien, T. K., Nash, P., Navarro-Compan, V., Scholte-Voshaar, M., van Vollenhoven, R., van der Heijde, D. and Stamm, T. A. (2016). Evidence for treating rheumatoid arthritis to target: results of a systematic literature search update. Annals of the Rheumatic Diseases, 75(1), 16-22.

Storz, U. (2011). International intellectual property strategies for therapeutic antibodies. mAbs, 3(6), 596-606.

Storz, U. (2015). Antibody-drug conjugates: Intellectual property considerations. mAbs, 7(6), 989-1009.

Stumpp, M. T., Binz, H. K. and Amstutz, P. (2008). DARPins: A new generation of protein therapeutics. Drug Discovery Today, 13, 695-701.

Tang, L., Gao, H., Zhu, X., Wang, X., Zhou, M. and Jiang, R. (2012). Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. BioTechniques, 52, 149-158.

Thomas, S., Smatti, M. K., Ouhtit, A., Cyprian, F. S., Almaslamani, M. A., Thani, A. A. and Yassine, H. M. (2022). Antibody-dependent enhancement (ADE) and the role of complement system in disease pathogenesis. Molecular Immunology, 152, 172-182.

Trevino, S. R., Scholtz, J. M. and Pace, C. N. (2007). Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in Rnase Sa. Journal of Molecular Biology, 366, 449-460.

Tungekar, A. A. and Ruddock, L. W. (2023). Design of an alternative antibody fragment format that can be produced in the cytoplasm of *Escherichia coli*. Scientific Reports, 13, 14188.

Van Durme, J., De Baets, G., Van Der Kant, R., Ramakers, M., Ganesan, A., Wilkinson, H., Gallardo, R., Rousseau, F. and Schymkowitz, J. (2016). Solubis: a webserver to reduce protein aggregation through mutation. Protein Engineering Design and Selection, 29, 285–289.

Vazquez-Lombardi, R., Phan, T. G., Zimmermann, C., Lowe, D., Jermutus, L. and Christ, D. (2015). Challenges and opportunities for non-antibody scaffold drugs. Drug Discovery today, 20(10), 1271-1283.

Walker, J. M., Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server, The proteomics protocols handbook, pp 571-607.

Wang, J., Lisanza, S., Juergens, D., Tischer, D., Watson, J. L., Castro, K. M., Ragotte, R., Saragovi, A., Milles, L. F., Baek, M., Anishchenko, I., Yang, W., Hicks, D. R., Exposit, M., Schlichthaerle, T., Chun, J., Daupara, J., Bennett, N., Wicky, B. I. M., Muenks, A., DiMaio, F., Correia, B., Ovchinnikov, S. and Baker, D. (2022). Scaffolding protein functional sites using deep learning. Science, 377, 387-394.

Wang, X., Li, F., Qui, W., Xu, B., Li, Y., Lian, X., Yu, H., Zhang, Z., Wang, J., Li, Z., Xue, W. and Zhu, F. (2022). SYNBIP: synthetic binding proteins for research, diagnosis and therapy. Nucleic Acids Research, 50, D560 – D570.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B./ L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Hanikel, N., Pellock, S. J., Courbet, A., Sheffler, W., Wang, J., Vankatesh, P., Sappington, I., Torres, S. V., Lauko, A., De Bortoli, V., Mathieu, E., Ovchinnikov, S., Barzilay, R., Jaakkola, T. S., DiMaio, F., Baek, M. and Baker, D. (2023). De novo design of protein structure and function with Rfdiffusion. Nature, 620, 1089 – 1126.

Weidle, U. H., Auer, J., Brinkmann, U., Georges, G. and Tiefenthaler, G. (2013). The emerging role of new protein scaffold-based agents for treatment of cancer. Cancer Genomics and Proteomics, 10, 155-168.

Woloschuk, R. M., Reed, P. M. M., Jaikaran, A., S., I., Demmans, K. Z., Youn, J., Kanelis, V., Uppalapati, M. and Woolley, G. A. (2021). Structure-based design of a photoswitchable affibody scaffold. Protein Science, 30, 2359-2372.

You, K., Gu, H., Yuan, Z. and Xu, X. (2021). Tumor necrosis factor alpha signaling and organogenesis. Frontiers in Cell and Developmental Biology, 9, 727075.

Zajc, C. U., Salzer, B., Taft, J. M., Reddy, S. T., Lehner, M. and Traxlmayr, M. W. (2021). Driving CARs with alternative navigation tools – the potential of engineered binding scaffolds. The FEBS Journal, 288, 2103-2118.

Zhao, A., Xue, Y., Zhang, J., Gao, B., Feng, J., Mao, C., Zheng, L., Liu, N., Wang, F. and Wang, H. (2004). A conformation-constrained peptide library based on insect defensin A. Peptides, 25, 629-635.

Zhao, N., Schmitt, M. A. and Fisk, J. D. (2016). Phage display selection of tight specific binding variants from a hyperthermostable Sso7d scaffold protein library. The FEBS Journal, 283, 1351-1367.

Zhao, Y., Yang, J., Niu, Q., Wang, J., Jing, M., Guan, G., Liu, M., Luo, J., Yin, H. and Liu, Z. (2023). Identification and characterization of nanobodies from a phage display library and their application in an immunoassay for the sensitive detection of African swine fever virus. Journal of Clinical Microbiology, 61(6), e0119722.

Zheng, K., Bantog, C. and Bayer, R. (2011). The impact of glycosylation on monoclonal antibody conformation and stability. mAbs, 3(6), 568-576.

Zimmermann, I., Egloff, P., Hutter, C. A. J., Kuhn, B. T., Brauer, P., Newstead, S., Dawson, R. J. P., Geertsma, E. R. and Seeger, M. A. (2020). Generation of synthetic nanobodies against delicate proteins. Nature Protocols, 15, 1707-1741.

Publication List

Ononugbo, C. M., Shimura, Y., Yamano-Adachi, N., Omasa, T., and Koga, Y. Rational design approach to improve the solubility of the beta-sandwich domain 1 of a thermophilic protein. *Journal of Bioscience and Bioengineering*, 138 (4), 271-282.

Acknowledgment

I would like to express my sincere gratitude to Professor Takeshi Omasa and Professor Yuichi Koga for their superb supervision. This journey would not have been easy without your constant support and advice. I appreciate the opportunity you gave me five years ago, Professor Takeshi Omasa, to join your lab under the MEXT scholarship. Thank you for believing in me and supporting my application to join this program. Since joining your laboratory, you have provided me with the opportunity and the enabling environment to work independently. Dear Professor Yuichi Koga, I appreciate the constant support you have given me since joining your research team and the continuous encouragement and motivation. Thank you very much.

I would also like to thank the Department of Biotechnology at Osaka University for opening its doors to students from Sub-Saharan Africa. Throughout this academic journey, I have learned not just research skills but also many transferrable skills. I hope the doors remain open to future students from Sub-Saharan Africa who want to join the program.

I would like to thank Associate Professor Noriko Yamano-Adachi for giving me one or two pieces of advice and for always being supportive. I would also like to extend my gratitude to all the members of the Omasa laboratory for always being willing to assist, especially in learning the Japanese language. I appreciate you all.

I would like to thank my wonderful friends, Kulachatr Panyawechamontri AKA (UE) and Sereirath Soth, with whom I joined this program together with. You have been very supportive and I appreciate everything, you are more like a family away from home. Thank you so much!

To my ever-supportive parents and siblings who have been on this journey together with me all these years, even while being far away, I appreciate you. Thank you very much for your moral support.

Finally, I want to thank God almighty for the wisdom, knowledge, and perseverance, as well as for bringing to me all the great and wonderful people whom I have worked with in the course of this program. I appreciate you, my heavenly father.