## ABSTRACT

GERA, NIMISH. Hyperthermophilic Protein Scaffolds for Engineering Biomolecular Recognition. (Under the direction of Balaji M. Rao).

Specific and non-covalent interaction between species, referred to as molecular recognition, is central to several diverse biological processes. Antibodies have been extensively used as binding molecules for molecular recognition. However, their applications are limited by low thermodynamic stability, large size and expensive production. Therefore, there is a keen interest in alternative non-immunoglobulin scaffold frameworks. Hyperthermophilic organisms are a rich source of small proteins with no cysteines that are excellent candidates for molecular recognition scaffolds.

We have shown that highly stable binding proteins for a wide spectrum of targets can be generated through mutagenesis of the Sso7d protein from the hyperthermophilic archaeon *Sulfolobus solfataricus.* Sso7d is a small (~ 7 kDa, 63 amino acids) DNA binding protein that lacks cysteine residues and has a melting temperature of nearly 100°C. Sso7d-derived mutants have high thermal stability, resistant to chemical denaturation by guanidine hydrochloride and retain their secondary structure after extended incubation in extreme pH conditions.

The high thermal, chemical and pH stability of Sso7d-derived proteins is particularly useful for chromatographic separations. Therefore, to further explore the use of Sso7d-based ligands in the context of protein purification, we isolated Sso7d-based protein ligands that bind the

Fc portion of human IgG (hFc). These hFc binders could be used to isolate human IgG from complex media. We systematically characterized the binding of the Sso7d-based ligands to various hIgG isotypes as well the effect of glycosylation on binding. Further, using a combination of histidine scanning mutagenesis and directed evolution, we isolated Sso7d variants that can elute human IgG under milder conditions.

Lastly, to further explore the potential of hyperthermophilic proteins as scaffolds, we generated a library-of-libraries, or a combinatorial "super-library", of ~ $4x10^8$proteins by randomizing surface accessible residues on seven different non-immunoglobulin scaffolds of hyperthermophilic origin. Binding proteins with high specificity and to a diverse set of model targets could be isolated from this super-library. Interestingly, the pool of highest affinity binders for each target contained proteins derived from a distinct subset of scaffolds, suggesting that certain scaffolds may be more suited to generate binders to a specific target. Also strikingly, binders from the super library have higher or similar affinities than those from a library with three orders of magnitude greater overall diversity, but derived by randomizing the previously validated Sso7d scaffold. Therefore, this approach produces a combinatorial library of higher quality. Our results are particularly relevant since current screening technologies sample only a fraction of the theoretical diversity generated by randomizing 10-15 residues. Further, mutant proteins from multiple scaffolds described in this study have favorable properties such as low molecular weight (< ~ 10kDa), high thermal stability ($T_m$> 74 °C) and ease of recombinant expression in *E. coli*.

Hyperthermophilic Protein Scaffolds for Engineering Biomolecular Recognition

by
Nimish Gera

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Chemical Engineering

Raleigh, North Carolina

2012

APPROVED BY:

| | |
|---|---|
| _____ | _____ |
| Balaji M. Rao<br>(Chair of Advisory Committee) | Robert M. Kelly |
| _____ | _____ |
| Ruben M. Carbonell | Carla Mattos |

## DEDICATION

To my friends and family

*"There is no substitute for hard work"- Thomas Edison*

## BIOGRAPHY

Nimish Gera was born on January 28th 1984 in the city of Kota in the western Indian state of Rajasthan. He finished high school in Kota in 2002 and started his undergraduate studies in the Department of Chemical Engineering at the Indian Institute of Technology, Guwahati in the fall. During his freshman year, he became highly interested in biology and its applications. He went on to take a number of classes and laboratory modules in biotechnology to further his interest. In the summer of 2005, Nimish did a research internship at the Jacob's University Bremen, Germany in the Protein Engineering Laboratory of Dr. Ulrich Schwaneberg which further elevated his interest in directed evolution and protein engineering. After finishing his undergraduate studies in Chemical Engineering with a concentration in Biotechnology, he started his graduate studies in the Department of Chemical and Biomolecular Engineering at the North Carolina State University, Raleigh, USA. Nimish joined the Protein Engineering Laboratory of Dr. Balaji M. Rao in Spring 2007 as one of the first students in the lab. During the summer of 2011, Nimish did an industrial internship at the Department of Antibody Engineering at Genentech in South San Francisco, CA. Nimish wants to pursue a career as a protein engineering scientist in the biopharmaceutical industry.

## ACKNOWLEDGEMENTS

My graduate school experience has been extremely satisfying, rewarding and the most cherished moments of my career. Starting up a new lab, managing and training people have been immense learning experiences which not everyone gets to do in graduate school. I was fortunate to have an extremely helpful, enthusiastic and thoughtful advisor in Bala who made this journey worthwhile. I have seen myself learn and grow in protein engineering during the last five years under his guidance. I am proud to be the first PhD graduate from his lab.

My advisory committee has been critical in the completion of my PhD degree. I am very grateful to Dr. Robert M. Kelly for his help, guidance and suggestions right from the first year. Without his support, it would have been impossible to set up the lab from scratch. I am indebted to him for allowing me to use equipment in his lab for a long time. Dr. Ruben Carbonell has been instrumental in the work on affinity ligands. Additionally, I would also like to acknowledge Dr. Carla Mattos and members of the Mattos lab for providing the purified Ras proteins and the discussions on protein crystallography.

The most important people who have led to the completion of this thesis are my undergrads Clay, Dalon and Andrew. Clay was very cooperative in setting up the lab and with the Sso7d library work which eventually became the foundation of my PhD thesis. Dalon helped me with numerous small projects which would never have been possible without his dedication and support. Thank you Dalon for being around and believing in me for two years even when I didn’t know what I was doing ☺. Andrew has been critical in the completion of the super

library work and the affinity ligands project. Thank you Andrew for all those late nights and as you said ‘me stealing your youth’. I hope I made it up to you with those last few lunches.

I am grateful to Paige Luck and Dr. Allen Foegeding from the Department of Food Science at NCSU for allowing me to use their DSC machine, Dr. John van Zanten and the staff at BTEC for help with Circular dichroism spectroscopy, Dr. Munir Alam and Dr. Moses Sekaran at the Duke Human Vaccine Institute for help with Biacore experiments and analysis. I would like to make a special mention of Dr. Saad Khan, who has served as a second advisor. I thoroughly enjoyed all the four classes I TA’d with him and all our interactions and discussions. These classes helped me be sane outside the lab.

I owe a special thanks to Prasenjit for being a great mentor, friend and roommate for the last 10 years. I don’t know what I would have done without you during my undergrad, grad studies or otherwise. Mahmud has been a great support in tough times in the lab and has been always around. Thank you Amr, Kevin and Karthik for making the lab environment enjoyable, entertaining and fun. I would also like to mention Adam, Jose, Sindhu, and Megan who worked with me at several stages during my PhD.

Sincere thanks to people in the Kelly Group for being great friends, constant help with experiments and scientific or not–so-scientific discussions. Thank you Arpan, Inci, Jaspreet and Hong for listening to me, Sara for being a great lab manager and Amy for insightful discussions on protein purification. I always look up to you guys. The Carbonell group has

been extremely helpful with experimental details and suggestions in the affinity ligands work especially Amith, Stefano and Nafisa.

I owe a lot to my friends Abhijeet, Chhavi, Anurag and Shamit for being great listeners and support in the past five years online, on the phone and in person. Additionally, I would like to acknowledge Gargi, Rishi, Pratik, Nishit, Sharad, Chintan and Chinmay for those wonderful times together. Special thanks to my friends in India- Akash and Abhishek and their wonderful wives Twinkle and Swati for hearing me almost every weekend.

My acknowledgements to all the CBE staff who always welcomed me with open arms. Sandra, you are doing an awesome job! Thank you Saundra and Diane for all the ordering and invoice help, Sheila for shipping and facility help, Angela for the seminars and Rajani for marketplace. This work would not have been possible without the funding support from National Science Foundation and Defense Threat Reduction Agency.

Last but not the least; I would like to thank my parents for being patient and supportive during these five years.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1

# NON-IMMUNOGLOBULIN SCAFFOLDS FOR BIOMOLECULAR RECOGNITION

Nimish Gera and Balaji M. Rao

Department of Chemical and Biomolecular Engineering

North Carolina State University

Raleigh, NC 27695-7905

## 1.1. Introduction

Specific and non-covalent interaction between bio-molecular species is called molecular recognition. Molecular recognition is central to many biological processes such as signal transduction or antigen-antibody recognition. Antibodies serve as universal molecular recognition agents since they can generate binding molecules to virtually any target antigen. The complementarity determining regions (CDR's) of antibodies can adapt themselves to different topologies and mediate surface interactions. However, recent advances in synthetic library construction and library screening technologies have led to the use of several non-immunoglobulin scaffolds as templates for engineering molecular recognition. Display techniques such as phage display [1, 2], yeast surface display [3-6], mRNA display[7, 8] and bacterial display[9, 10] have been used to engineer novel recognition agents using 'alternative' protein scaffolds. A central theme in these display techniques is a physical link between the phenotype (protein) and its genotype (coding DNA), which allows to recover the identity of the improved variant proteins.

Historically, antibodies have been used as the preferred scaffold architecture; however, certain molecular and biophysical properties of antibodies limit their applications. The large multi-domain structure of antibodies is difficult to produce in a simple bacterial system such as *E.coli* because antibodies require post-translational modifications. The antibody heavy and light chains are held together by disulfide bonds that can fall apart in the reducing *E.coli* cytoplasm. The large size also limits tissue penetration of antibodies which impairs specific tumor targeting in therapeutic and imaging applications. Antibody fragments such as single chain variable fragments (scFv's)[11] or single domain antibodies( $V_H$ and $V_L$) [12, 13] have

been developed to address size constraints as the binding surface of antibodies is only limited to the variable region which is quite small as compared to the whole antibody itself. These fragments are also easier to engineer using standard protein engineering methods. The poor thermal and chemical stability of antibodies impose constraints on their storage as well as applicability. Additionally, Fc mediated effector function may be unnecessary for applications such as imaging with payloads or antagonist blocking[14] .Lastly, intellectual property concerns such as humanization or antibody production also limit the applications of antibodies. Hence, there is a growing interest in non-antibody or non-immunoglobulin or alternative scaffolds to provide adaptable binding surface topologies to engineer molecular recognition.

### 1.2. Non immunoglobulin scaffolds

The human immune system employs antibodies as universal recognition agents against target antigens. However, full length antibodies are not the only bio-molecular scaffold that can be utilized to generate binding repertoires. Immunoglobulins from camelids and shark contain single domains of either $V_H$ or $V_L$ serving as the antigen recognition agents [15-17]. Recently, jawless vertebrates such as sea lampreys were found to contain highly diverse leucine rich repeats as the adaptive immunity scaffold [18, 19]. Interestingly, designed ankyrin repeat proteins (DARPins) were shown to serve as a protein scaffold confirming the fact that repeat proteins indeed can serve as molecular recognition agents. Numerous other non-immunoglobulin protein scaffolds have gathered a lot of interest over the past fifteen years or so[20].

A protein scaffold can be defined as a 'template' protein which can tolerate amino acid substitutions or insertions and still maintain its structural integrity. Success of a scaffold depends on its ability to perform at least as good as antibodies if not better and therefore the attributes of high affinity and specificity are critical. Not only that, a scaffold should have potentially superior biophysical properties. Therefore, an ideal scaffold is one that has a small size, high stability and no disulfide bonds to facilitate easy expression in the reducing cytoplasm of *E.coli*. However, other properties can also be engineered by fine tuning the library screening conditions.

Simplistically, a DNA library of the protein scaffold is created by first identifying key residues to be mutated. The next step is to establish a link between the protein and its coding DNA (i.e. phenotype and genotype) which is achieved by either cell surface display (bacteria, yeast or mammalian display) or covalent linkage (ribosome or mRNA display). The protein library thus generated can be screened for properties such as binding affinity, specificity, thermal or chemical stability, enzyme activity and many more to obtain candidates that can serve as molecular recognition agents. Almost all the display screening technologies sparsely sample the theoretical diversity and generation of a high affinity reagent by a simple screening methodology is a rarity. Affinity maturation strategies are typically required to achieve the desired affinity of the binding molecule.

Non-immunoglobulin scaffolds have been used as affinity reagents in several diagnostic, imaging, research and therapeutic applications which may require varying molecular or biophysical properties. For example, in a diagnostic application- small size, high affinity and

specificity with rapid renal clearance would be preferred in contrast to a therapeutic application where along with high affinity and specificity a long in-vivo half-life would be critical. Diagnostic or imaging applications may also require higher thermal and chemical stability because conjugation to probes requires higher temperatures and harsh chemical conditions. Biotechnological applications such as protein purification ligands demand higher stability for tolerating harsh clean in place regeneration procedures. Generation of multi-specific binding molecules involving the combination of molecules with varying specificity is also a rapidly growing field due to advantages associated with cost and FDA approval. It is therefore apparent that biophysical properties are as important as high affinity in determining the success of a scaffold in various applications. Engineered affinity reagents should either have these properties or be amenable to engineering them. Multiple engineered proteins from the Adnectins, Affibodies and the DARPins are now in clinical trials and have shown promise as low cost alternatives to antibodies [20, 21].

Novel screening strategies are being employed by combining two different display methodologies to take advantage of the benefits. Recently, affibody candidates against ErbB3 (HER3) were identified using staphylococcal display and then fine affinity discrimination was performed using phage display[22]. Our group has integrated mRNA display and yeast surface display by first isolating low affinity binding proteins using mRNA display and then further library screening using yeast surface display (Chapter 4). Another study has highlighted the use of yeast surface displayed antigen as a target for screening phage display libraries[23].

Although several new scaffolds have been identified, this review highlights the most extensively used scaffolds that have made a mark during the past fifteen years such as the $10^{th}$ domain of human fibronectin (also known as Monobodies, Adnectins or Trinectins), a three helix bundle protein from staphylococcal protein A (Affibody), designed ankyrin repeat proteins (DARPins), Lipocalins (Anticalins) and more recently cystine knot miniproteins family (knottins). Figure 1.1 depicts these non-immunoglobulin scaffolds. Additionally, most common applications of these scaffolds have also been described. Some examples of targets for which these scaffolds have been engineered and the mutation strategy has been highlighted in Table 1.

### 1.2.1. Tenth domain of human fibronectin III or 'Adnectins'

Fibronectin domains are multifunctional proteins responsible for integrin mediated cell adhesion found at high concentrations in the extracellular matrix and human serum[24]. The tenth domain of human fibronectin (10FnIII) is a small (94 aa, 10 kDa) protein with a β-sandwich fold structurally resembling the variable heavy chain of immunoglobulin. The loops BC, DE and FG of fibronectin structurally resemble the complementarity determining regions of immunoglobulin heavy chain CDR1, CDR2 and CDR3.The solution structure as determined by NMR suggests two anti-parellel β-sheets consisting of seven β-strands that pack together to form a hydrophobic core(Figure 1.1 A)[25]. Solvent exposed loops mediate cell adhesion interactions through an Arg-Gly-Asp (RGD) sequence on the FG loop. 10FnIII is a highly stable protein domain ($T_m$> 80˚C) with no cysteines[26, 27]. Moreover, 10FnIII does not require any post translational modifications and therefore can be expressed at high yields in *E.coli.*

Extensive mutagenesis of the 10FnIII scaffold has been performed with most studies involving generation of loop libraries. The first use of 10FnIII as a scaffold was described by Koide et al[28] wherein five residues each in two solvent exposed loops, BC and FG were mutagenized to create a phage displayed library of $10^8$ protein variants. The FG loop was also shortened by four residues to provide a contiguous binding surface. Ubiquitin binding proteins were isolated and the best mutant had an $IC_{50}$ of 5 μM. This mutant however also showed specificity towards dextran and was lower in stability than the wild type protein. Subsequently, structural stability of 10FnIII was investigated involving removal of unfavorable electrostatic interactions and stability enhancement using fragment complementation [29, 30]. Loop elongation studies were performed to identify stable loops and the EF loop was found to be the most destabilizing of all[31]. An alternative display method, mRNA display was used to create a greater diversity ($10^{12}$ mutants) library by randomizing 21 residues on the BC,DE and FG loops and screened to isolate specific and higher affinity binding proteins ($K_D$ ~1-24 nM) for TNF-α[32]. Furthermore, it was validated that 10FnIII based mutants can also be affinity matured ($K_D$~20 pM). In another study, five residues around the RGD peptide on the FG loop were mutated to generate a mutant with $K_D$~800 pM against integrin αvβ3 [33]. This mutant could also stain αvβ3 positive cells in flow cytometry. These studies established the use of 10FnIII as a scaffold to generate specific, high affinity mutants that were amenable to further affinity maturation. Conformation specific antibody mimics were generated that recognize ligand bound estrogen receptor alpha inside cells. Two combinatorial libraries based on the FG and the AB loops were constructed and screened using yeast two hybrid system. Isolated mutants could detect

different conformations of estrogen receptor alpha inside living cells where short peptides would get degraded due to proteolysis[34]. Further studies on 10FnIII involved generation of binding proteins to biologically relevant proteins to show their therapeutic potential. A phage displayed library of $2x10^9$ members was generated by mutating the FG and BC loops and screened against the SH3 domain of oncogenic human c-Src protein. The obtained binding proteins had proline rich sequences mimicking the natural peptide ligands [35]. It was shown that both BC and FG loops contribute to binding. The binding protein was shown to pull down full length c-Src from a mouse fibroblast cell line. The protein could also be used in a western blotting assay as a primary detection reagent.

Biologically active 10FnIII derived mutants were engineered as antagonists for human and mouse VEGF-R2 using an mRNA display library of $10^{13}$ variants [36]. Some of these binding proteins (VR28) inhibited the action of the natural VEGF ligand binding to the VEGF-R2. The native VEGF ligand has an affinity of 75-125 pM. In comparison, VR28 showed a moderate binding affinity (11-13 nM) but not great efficacy and therefore needed affinity maturation. Deletion of eight N terminal residues in VR28 seemed to improve binding to the KDR (human VEGF-R2). Binders were further evolved to bind mouse-VEGFR2 thereby generating dual specificity. In a follow up study, biophysical properties of these fibronectin derived antibody mimics were extensively studied. Antibody mimics were characterized for their stability, solubility and oligomerization properties. Point mutations led to the isolation of 159(A56E) -a relatively stable (53˚C), high affinity (0.59 nM) and predominantly monomeric (75-100%) VEGF-R2 binding antibody mimic[37].

The binary code of amino acids tyrosine and serine had been tested for antibodies and this strategy was further applied to the fibronectin domain to test its applicability for a smaller scaffold. A phage display library of $10^{10}$ variants for screened for maltose binding protein, human small ubiquitin like modifier and yeast small ubiquitin like modifier[38] to create high affinity ($K_D$~ low nanomolar) binding proteins. Three rounds of phage display selection were performed and the library transformed into yeast surface display format for another round of sorting. This study validated that even though the two amino acids have a low chemical diversity, the conformational diversity achieved by the amino acids might be more important.

10FnIII is found in a multimeric format in nature and are therefore well suited for generating multi-functional binding molecules[14]. An engineered fibronectin domain specific to integrin αvβ3 was fused to a pentamerization domain to generate a fusion protein with $K_D$< 13 pM. This protein was also tested on K562 cell lines for biological efficacy[39]. Further, it was also shown that two loops (BC and FG) can provide enough surface area for generation of a low picomolar affinity ($K_D$~350 pM) binding fibronectin variant. A disulfide bonded antibody mimic was isolated similar to shark Immunoglobulin new antigen receptor and single domain camelid antibodies [40]. In a subsequent study, a much higher affinity ($K_D$~1.1pM) fibronectin mutant was generated using a combination of loop length diversity, loop shuffling and recursive mutagenesis[41].

### 1.2.2. Z domain scaffold or 'Affibody'

The Affibody scaffold (Z-domain) is an engineered version of the B-domain from *Staphylococcus aureus* protein A and is a small (58 aa), cysteine free, three α-helix bundle protein that binds the Fc region of IgG(Figure 1.1B). The B-domain also shows some binding to the Fab fragment of IgG. The Z-domain is an engineered variant with higher alkaline and proteolytic stability, solubility and no Fab binding. Affibody molecule has one of the fastest measured folding kinetics in a two state process with a rate constant of 120000 $s^{-1}$[42].

In the earliest studies for engineering the affibody scaffold, thirteen surface exposed residues including nine residues responsible for binding to IgG were mutated with a few additional residues and a combinatorial phage display library was created by solid phase assembly[43]. This library was further selected against three targets namely *Taq* DNA polymerase, human insulin and a human apolipoprotein A-1 variant [44]. Binding proteins obtained had micromolar affinities and retained the three helical bundle structure of the wild type Z-domain with no affinity to the Fc domain of IgG. These binding proteins could be easily produced in *E.coli* at yields of 3-10 mg per liter of culture. The *Taq* polymerase binding affibody was further affinity matured to an affinity between 30-50 nM by randomizing six alpha helical residues on the binding interface [45]. Additionally, it was also shown that dimeric form of affibodies can be constructed. The binding proteins developed for *Taq* DNA polymerase and human apolipoprotein A-1 variants were shown to be useful as ligands in affinity chromatography and regeneration of the columns was also demonstrated[46].

Affibody variants that recognize the G protein subtypes A and B of respiratory syncytial virus (RSV) were also isolated from a previously generated Affibody library [43, 44]. Affibody molecules for the extracellular domain of human epidermal growth factor receptor two (HER2) were selected and a 50 nM affinity variant was selected by phage display [47, 48]. Further affinity maturation led to an affinity of 5-10 nM [49].Subsequently, this variant was affinity matured to generate an affibody molecule with the highest affinity (22 pM) [50] and shown to be useful for tumor imaging applications in mice xenografts. With the recent understanding that HER3/Erb3 also plays a critical role in resistance to Her2 targeted therapies, nanomolar affinity ErbB3 targeted affibody molecules have also been developed[22].

Since the affibody scaffold is derived from protein A, affinity chromatography was an obvious application for using this scaffold. Binding proteins have been created not just for IgG, but also for other human Immunoglobulin classes such as IgA and IgE[51-53]. These affibody molecules were used as capturing ligands in protein microarrays. Ligands to a predefined fragment of human factor VIII were isolated by using competitive biopanning in phage display format. These ligands had an initial affinity of $K_D$~100-200 nM which could be further affinity matured to a $K_D$~5 nM[54]. Affibody molecules for amyloid beta peptides[55], human serum albumin[56], Ras proteins[57] and human Raf-1 proteins[58] have also been developed.

### 1.2.3. Designed ankyrin repeat proteins or DARPins.

Repeat proteins based on the leucine-rich repeat family have been found to be an integral part of the immune system of jawless vertebrates[19]. Hence, exploring their potential as an alternative to antibodies seems promising. Repeat proteins comprise of small repeat motifs of 20-50 residues stacked in an elongated fashion to provide a greater potential binding interface than globular proteins(Figure 1.1C)[59]. Multiple repeat proteins have been discussed such as the tetratricopeptide repeat protein [60-62], armadillo repeat protein[63], alpha helicoidal repeat proteins based on HEAT-like repeats[64] but the most widely used repeat protein scaffold is the designed ankyrin repeat protein (DARPin). DARPin is based on ankyrin repeat proteins that mediate protein-protein interactions in bacteria, fungi, plants and animals[65]. The individual repeat motifs in ankyrin repeat proteins comprise of anti-parellel α-helices connected by a β-turn and a single repeat protein can have from four to six of these individual repeat motifs[66]. Ankyrin repeat proteins do not contain any disulfide bonds. Stability of DARPins depends on the number of repeat motifs. DARPins with more than three repeats are highly stable to temperature and chemical denaturation[67]. Salt bridges on the DARPin surface are thought to be responsible for such exceptional thermal stability[68].

A consensus designed ankyrin repeat protein was generated by fixing the conserved residues in the framework by comparing several protein sequences/structures and creating a random surface for interaction with target proteins. Capping repeats with hydrophilic sequences are included to shield the hydrophobic surfaces. This architecture was further chosen as a scaffold to generate binding proteins for molecular recognition [69, 70]. Since DARPins fold

fast, surface display techniques like phage display could not be used and most combinatorial libraries of designed ankyrin repeat proteins have been constructed mostly in ribosome display[67]. Subsequently, phage display based on a different translocational pathway in *E.coli* called as signal recognition particle phage display was developed allowing the display of DARPin's on phages [71, 72].

In the earliest library selections, binders were generated for maltose binding protein and two eaukaryotic mitogen-activated protein kinases from a ribosome display library of $10^{10}$ members[73]. The crystal structure of the maltose binding protein DARPin was also solved. The proteins were found to express at 200 mg/liter in shake flasks. The use of DARPin's in the context of intracellular inhibition was further validated by generating intracellular kinase inhibitors for aminoglycoside phosphotransferase[74]. Aminoglycoside phosphotransferase is responsible for mediating resistance to aminoglycoside antibiotics in pathogenic bacteria. Subsequently, an allosteric inhibition mechanism of interaction was confirmed along with the crystal structure of the designed ankyrin repeat with aminoglycoside transferase[75].In another study, intracellular proteinase inhibitors were isolated against $Nia^{pro}$ proteinase of tobacco etch virus[76].A two-step selection based on ribosome display and protein fragment complementation assay (PCA) was used for isolating binders to MAP-kinase. Two rounds of ribosome display and and PCA led to an enrichment of $10^5$ per round[77].

DARPins for the human epidermal growth factor receptor two (HER2) were isolated using ribosome display. Specific and high affinity ($K_D$~7.3 nM) DARPins were isolated and tested against the extracellular domain of HER2 and HER2 positive breast cancer tissues[78].

Further, a picomolar affinity DARPin was developed for HER2 [79]. DARPins have also been generated as binding proteins for Caspase-2[80], neurotension receptor-1[81] and tumor-associated antigen epithelial cell adhesion molecule(EpCAM)[82]. Multimerization of DARPins specific to EpCAM has also been demonstrated[83].

### 1.2.4. Lipocalins or 'Anticalins'

Lipocalins are a family of proteins highly conserved in several species such as bacteria, plants, insects and even humans and are responsible for transporting small molecules like vitamins, steroids and other metabolites[84]. A lipocalin is a small protein (< 20 kDa) consisting of a conserved β- barrel structure generated by eight anti-parellel β- strands that give rise to four loops on one side forming a ligand binding pocket and an α-helix against the β-barrel on the other side(Figure 1.1D). As in other scaffolds, the loop regions are variable in sequence and length among different members of the family but the β-barrel structure is conserved. Lipocalins and their engineered variants have a high thermal stability ($T_m$> 70˚C) and can be produced as monomers in *E.coli*[84, 85]. Lipocalins are also present in human plasma at concentrations up to 1 mg/ml which makes them as ideal candidates for generation of therapeutic drugs[84].

In the first study for engineering the lipocalin scaffold, the bilin binding lipocalin from *Pieris brassicae* was engineered to bind fluorescein[86]. Sixteen residues in the bilin binding interface of the lipocalin protein were subjected to random mutagenesis and the library was screened using phage display. A lipocalin variant with a dissociation constant of 35.2 nM

was isolated from the screen. This parental anticalin was further rationally engineered by mutagenizing just two amino acid residues to enhance the affinity to 1 nM [87]. Anticalins against digoxigenin were obtained using phage display and a bacterial colony screening assay. Rational affinity maturation was also performed where the affinity was increased by an order of magnitude from $K_D$~300 nM to $K_D$~30 nM. These anticalins could be useful in bioanalytical assays and fusions with alkaline phosphatase were also generated [88]. Fusion proteins using lipocalins have been developed for use as bi-specific affinity reagents, called as 'duo-calins' where anticalins against fluorescein and digoxigenin were expressed in *E.coli*[89]. A step wise *in vitro* affinity maturation of the digoxigenin binding anticalin has also been reported where the affinity was increased to $K_D$~12 nM[90]. Crystal structure of the digoxigenin binding anticalin in complex with digoxigenin revealed structural changes in mutated loop regions[91]. In the structure of the fluorescein binding anticalin, the β-barrel framework is conserved but the loops show significant conformational changes from the wild type[92]. Multiple biophysical studies have been carried out on lipocalin variants since and other lipocalins from the family have also been explored as scaffolds [93-97].

### 1.2.5. Knottins

Cystine-knot miniproteins are a family of small (30 aa) triple disulfide bonded proteins (Figure 1.1 E). The three disulfide bonds give them a knot like structure and therefore they are also called as 'knottins'[98, 99]. Knottins with variable structural characteristics such as a triple-stranded β- sheet, a $3_{10}$ α-helix or a head to tail cyclized version called as cyclotides

have been found in nature. The high conformational rigidity imparts exquisite thermal, pH, chemical and proteolytic stability to knottins[98].

Loop regions of knottins have been shown to be tolerant to amino acid mutations, thereby suggesting their promise as molecular recognition scaffolds. Knottins have also been engineered using combinatorial approaches in several studies. A phage display library of the C-terminal cellulose binding domain of cellobiohydrolase I from the fungus *Trichoderma reesei* was generated and screened against cellulose, bovine alkaline phosphatase, α- amylase and *E.coli* β- glucuronidase[100]. Other sources have been reported for knottins and the two most commonly used topologies are: *Ecballium elaterium* trypsin inhibitor II (EETI-II) from plants and the Agouti-Related-Protein (AgRP) from humans. It was demonstrated that these knottins are amenable to loop grafting where biologically active epitopes could be substituted in their loops [101, 102]. A yeast surface display library of the EETI-II knottin was created and screened against αvβ3 integrin. Isolated peptides bound with low nanomolar affinities to αvβ3 and αvβ5 integrins[103]. Another yeast display library was created with the AgRP knottin and peptides specific to αvβ3 integrin were generated. These peptides had much higher affinities than the ones generated by using the EETI knottin. A second generation library was created using a high affinity αvβ3 binding knottin to engineer specificity for αvβ3 over αiibβ3 [104]. AgRP knottins were also specifically engineered for αiibβ3 with low nanomolar affinity and prevented platelet aggregation slightly better than the FDA approved cyclic peptide eptifibatide[105]. Engineered knottin peptides have also been extensively used for imaging applications (see section 1.3.1 for details).

## 1.3. Applications of non-immunoglobulin scaffolds

Non immunoglobulin scaffold generated affinity reagents can be used in multiple applications. Diagnostic applications are more popular because of the non-human nature of many of these scaffolds which raise potential concerns regarding their use as therapeutic reagents. Affinity and specificity are the primary criteria in most applications and depending on the application several other characteristics may be required. For example for a tumor targeting application the targeting agent or the affinity reagent must be small in size for rapid renal clearance (Size < 60 kDa) with high affinity to allow for high contrast imaging, however for an affinity chromatography ligand the specificity is more important than the affinity as a moderate affinity should suffice to allow for elution of the target protein. Additionally, high thermal and chemical stability may be important for resin regeneration. Similarly, large tumor residence time, non-immunogenicity and high affinity is imperative for a therapeutic application in which case a large size of the targeting molecule may be preferable. Specific recognition of the target molecule is critical for all the applications in biotechnology and medicine.

### 1.3.1. Imaging and diagnostics

Affinity reagents in diagnostic or imaging applications typically rely on cell surface markers that are overexpressed in malignant tumors and detection of these cell bound proteins would enable diagnosis of cancer or any particular disease. However, these reagents need certain favorable characteristics- high affinity and specificity for efficient detection and rapid

clearance from the body, hence small size. Smaller size also allows for better tissue penetration and higher stability is required for radionuclide labeling. Not only that, unbound targeting molecules should also be rapidly cleared to allow high contrast imaging. The rather large size of antibody prevents them from having high level of tissue penetration or faster kinetics.

In diseases such as cancer, early detection can lead to an increased chance of survival. Traditional imaging techniques such as magnetic resonance imaging (MRI) detect tumors greater than a centimeter in diameter which is already about $10^9$ cells[106]. Therefore, specific molecular targeting of tumors is imperative for early detection. Monoclonal antibodies are expensive and have a longer serum half-life, therefore high contrast imaging with antibodies may not be very efficient. Antibodies have poor tissue penetration[107] and long *in vivo* half-life which leads to poor tumor-to-blood contrasts[22]. Therefore, alternative scaffold based probes have recently gained impetus for cancer imaging[108].

Small peptides provide the advantage of a small size and rapid clearance but fail to provide the required affinity for specific tumor detection[109]. Smaller scaffolds such as affibodies and knottins have been extensively used for imaging applications. The small size of these scaffolds also allows chemical synthesis of the binding proteins. Non-immunoglobulin scaffold based cancer biomarkers have been engineered for receptor tyrosine kinases such as HER2 and EGFR, integrins and TNF-α. Affibodies have shown great potential in HER2 imaging as compared to antibodies and also DARPins[50, 110-112]. Imaging reagents for

EGFR and TNF-α have also been generated using affibodies[113-115]. Affibody molecules have been shown to have specific targeting within an hour of administration and have a rapid in vivo pharmacokinetics. Knottins such as AgRP have also been used to generate high affinity targeting reagents for integrins αvβ3, αvβ5 and α5β1 [116, 117].

### 1.3.2. Therapeutics

Non-immunoglobulin scaffolds are starting to make their mark in the therapeutic realm with multiple scaffolds reaching the clinic. As of 2008, as many as eight different non immunoglobulin based proteins had entered clinical trials and about thirty eight were in pre-clinical development[118]. Similar to antibodies and antibody fragments, affinity proteins from non-immunoglobulin domains are now being conjugated to payloads such as toxins or radionuclides for therapeutic applications. A longer half-life is an important attribute of any therapeutic and therefore there have been efforts in half-life extension [111]. Moreover, Fc mediated cellular toxicity functions might not be required for certain applications. If specific targeting agents have been developed for imaging applications, a payload can be conjugated to the protein moiety and half-life extension strategies can be employed. Such efforts have been employed for the HER-2 specific affibody molecule and fusions with abumin binding domain (ABD) have been generated. The fusion to ABD slows the kinetic profile and decreases kidney uptake. The dose delivery to the tumor was also enhanced fivefold as compared to the non-ABD fused affibody [111]. Radiotherapy applications for the EETI-II knottin have also been demonstrated [119]. Enzyme inhibitors for extracellular proteases were developed based on the Kunitz type scaffold such as DX-88[120] for hereditary

angioedema and DX-890 which is an inhibitor of human neutrophil elastase and is therefore used as an anti-inflammatory drug for cystic fibrosis[121]. An adnectin, CT-322 (angiocept) against the VEGFR2 has been shown to have efficacy in the mouse model of pancreatic cancer [122]. Avimers developed against IL-6 are also currently in clinical development[123]. A neutralizing anticalin against human and murine VEGF with picomolar affinity and another anticalin against CTLA4 are currently in preclinical development [97, 124, 125].

### 1.3.3. Other biotechnology applications

As true alternatives to antibodies, non-immunoglobulin scaffolds have been used in a number of biotechnological applications apart from imaging and therapeutics. Adnectins have been used in protein microarrays, western blotting and immuoprecipitation applications [32, 35]. Functional fusions of affibody molecules to enzymes such as horseradish peroxidase and β-galactocidases have been reported as well [126, 127]. Affibodies have also been validated for use as detection reagents in protein microarrays [52, 53, 128].

Affinity chromatography is another application where there is a need to stable and low cost alternatives. Stable versions of the affibody scaffold have been created for protein purification applications [46]. Additionally, an active form of estrogen receptor was also isolated from an *E.coli* lysate using an engineered fibronectin domain[34].

Non-immunoglobulin scaffolds are particularly useful for intracellular applications because most of these do not have any disulfide bonds and therefore are stable inside cells. Fibronectin domains can differentiate and inhibit two different states of estrogen receptor have been generated[34]. DARPins inhibiting proteases and kinases have been shown useful in such applications [76, 77]. Other antagonist and blocking applications have also been described. Affibodies have been used for inhibiting the CD28-CD80 interaction[129] and other cell surface receptor mediated interactions[130-132]

Non immunoglobulin proteins are also useful as crystallization chaperones. Chaperones reduce the conformational flexibility, shield the hydrophobic surfaces and thereby improve solubility. Infact, a number of crystal structures have been reported with various scaffolds [69, 92, 133]. DARPins and Adnectins have been extensively used as crystallization chaperones [134, 135]. Lastly, bispecific or other multivalent moieties are routinely constructed[136] using scaffolds such as affibodies[137] , DARPins[138], Adnectins[139].

## 1.4. Thesis overview

We proposed a novel class of non-immunoglobulin scaffolds- small and highly stable protein scaffolds from hyperthermophilic organisms. Hyperthermophilic organisms grow in extreme conditions of temperature, pH and pressure and therefore it is likely that proteins from these organisms have higher stability. This class of proteins has received scant attention for engineering molecular recognition.

We have described the engineering of as many as seven different protein scaffolds from hyperthermophilic organisms. We identified several protein domains and engineered them with novel binding specificities. Further, we explored use of these engineered affinity reagents for applications in two diverse areas- ligands protein purification and intracellular inhibitors in human embryonic stem cells.

### 1.4.1. Hypothesis

Engineering a protein for a novel function involves generating a library of mutant proteins and screening the library using high throughput screening techniques to identify the novel improved variants. Firstly, a DNA library is created by introducing mutations in the gene and the DNA library is translated into a protein library by either cell transformation or a covalent linking of the gene and the protein. The fraction of functional or folded mutants determines the functional library diversity from which the improved variants would be isolated. Some mutations may destabilize the protein scaffold thereby lowering the overall fraction of stable

and folded protein variants and hence the likelihood of isolating a functional protein variant from the library.

Our central hypothesis is that highly stable proteins can tolerate mutations without compromising on their stability. In other words, if a highly stable scaffold is mutagenized, the reduction is stability would be acceptable enough so that the final variant has a reasonable stability and maintains its functionality. Therefore, small protein domains from hyperthermophilic organisms are an ideal candidate for protein engineering since they typically have high thermal, chemical or proteolytic stability. High thermal and chemical stability is important for reagents in imaging and other biotechnological applications such as ligands for protein purification. Proteolytic and low pH stability would be beneficial for using reagents as intracellular inhibitors.

### 1.4.2. Thesis layout

Chapter 2 is a review on yeast surface display, a central technology used in the engineering of protein scaffolds, antibodies and antibody fragments. We have used this technology to engineer small proteins from hyperthermophilic organisms against a wide spectrum of targets. Yeast surface display allows quantification and fine affinity discrimination between protein variants. Certain biophysical properties such as binding affinity and thermal stability can also be determined on yeast cell surface without the need for separate sub-cloning and protein production.

Chapter 3 describes engineering of a novel protein scaffold called Sso7d from *Sulfolobus solfataricus*. Sso7d is a small and highly stable DNA binding protein whose DNA binding surface was engineered to bind a variety of targets with different molecular weights. Specific binding proteins were obtained from a library of $10^8$ members using yeast surface display. Sso7d-derived variants were shown have high thermal and chemical stability and can be easily expressed at high yields in *E.coli*.

Chapter 4 describes a fundamental protein engineering problem of comparing scaffold diversity against sequence diversity. A mixed library of seven different protein scaffolds from hyperthermophilic organisms was screened against several targets. Different scaffold topologies were found to be optimum for different targets confirming that a particular scaffold may be optimum for a given target. This study showed that scaffold diversity is superior to sequence diversity for isolation of target specific variants.

Chapter 5 illustrates the application of Sso7d-derived proteins in the context of protein purification. A human-Fc portion (hFc) binding Sso7d variant (Sso7d-hFc) has been described. This protein variant can be used for the purification of human or humanized-IgG's. pH sensitive affinity ligands for hFc were generated by using the approaches of rational design and directed evolution. A pH based elution strategy can be employed where the target protein (human or humanized IgG) can be eluted from the ligand (Sso7d-variant) by using pH 4.5.

Chapter 6, the final chapter; underlines the conclusions from these studies and describes the further applicability of these binding proteins. A broader perspective on how these scaffolds can be used to address other biological problems is also explained. Furthermore, future work that can help advance the research and provide clues about biophysical properties of small hyperthermophilic scaffolds is also outlined.

## 1.5. References


1. Finlay, W.J., L. Bloom, and O. Cunningham, *Phage display: a powerful technology for the generation of high specificity affinity reagents from alternative immune sources.* Methods Mol Biol, 2011. **681**: p. 87-101.

2. Pande, J., M.M. Szewczyk, and A.K. Grover, *Phage display: concept, innovations, applications and future.* Biotechnol Adv, 2010. **28**(6): p. 849-58.

3. Boder, E.T. and K.D. Wittrup, *Yeast surface display for screening combinatorial polypeptide libraries.* Nat Biotechnol, 1997. **15**(6): p. 553-7.

4. Chao, G., et al., *Isolating and engineering human antibodies using yeast surface display.* Nat Protoc, 2006. **1**(2): p. 755-68.

5. Kieke, M.C., et al., *Isolation of anti-T cell receptor scFv mutants by yeast surface display.* Protein Eng, 1997. **10**(11): p. 1303-10.

6. Gai, S.A. and K.D. Wittrup, *Yeast surface display for protein engineering and characterization.* Curr Opin Struct Biol, 2007. **17**(4): p. 467-73.

7. Lipovsek, D. and A. Pluckthun, *In-vitro protein evolution by ribosome display and mRNA display.* J Immunol Methods, 2004. **290**(1-2): p. 51-67.

8. Roberts, R.W., *Totally in vitro protein selection using mRNA-protein fusions and ribosome display.* Curr Opin Chem Biol, 1999. **3**(3): p. 268-73.

9. Georgiou, G., et al., *Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines.* Nat Biotechnol, 1997. **15**(1): p. 29-34.

10. Little, M., et al., *Human antibody libraries in Escherichia coli.* J Biotechnol, 1995. **41**(2-3): p. 187-95.

11. Huston, J.S., et al., *Protein engineering of single-chain Fv analogs and fusion proteins.* Methods Enzymol, 1991. **203**: p. 46-88.

12. Holt, L.J., et al., *Domain antibodies: proteins for therapy.* Trends Biotechnol, 2003. **21**(11): p. 484-90.

13. Vincke, C., et al., *General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold.* J Biol Chem, 2009. **284**(5): p. 3273-84.

14. Lipovsek, D., *Adnectins: engineered target-binding protein therapeutics.* Protein Eng Des Sel, 2011. **24**(1-2): p. 3-9.

15. Muyldermans, S., et al., *Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains.* Protein Eng, 1994. **7**(9): p. 1129-35.

16. Roux, K.H., et al., *Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins.* Proc Natl Acad Sci U S A, 1998. **95**(20): p. 11804-9.

17. Nuttall, S.D., et al., *Isolation of the new antigen receptor from wobbegong sharks, and use as a scaffold for the display of protein loop libraries.* Mol Immunol, 2001. **38**(4): p. 313-26.

18. Pancer, Z., et al., *Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey.* Nature, 2004. **430**(6996): p. 174-80.

19. Pancer, Z. and M.D. Cooper, *The evolution of adaptive immunity.* Annu Rev Immunol, 2006. **24**: p. 497-518.

20. Lofblom, J., F.Y. Frejd, and S. Stahl, *Non-immunoglobulin based protein scaffolds.* Curr Opin Biotechnol, 2011.

21. Gebauer, M. and A. Skerra, *Engineered protein scaffolds as next-generation antibody therapeutics.* Curr Opin Chem Biol, 2009. **13**(3): p. 245-55.

22. Kronqvist, N., et al., *Combining phage and staphylococcal surface display for generation of ErbB3-specific Affibody molecules.* Protein Eng Des Sel, 2011. **24**(4): p. 385-96.

23. Hu, X., et al., *Combinatorial libraries against libraries for selecting neoepitope activation-specific antibodies.* Proc Natl Acad Sci U S A, 2010. **107**(14): p. 6252-7.

24. Cembrowski, G.S. and D.F. Mosher, *Plasma fibronectin concentration in patients with acquired consumptive coagulopathies.* Thromb Res, 1984. **36**(5): p. 437-45.

25. Main, A.L., et al., *The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions.* Cell, 1992. **71**(4): p. 671-8.

26. Litvinovich, S.V. and K.C. Ingham, *Interactions between type III domains in the 110 kDa cell-binding fragment of fibronectin.* J Mol Biol, 1995. **248**(3): p. 611-26.

27. Cota, E., et al., *The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold.* J Mol Biol, 2001. **305**(5): p. 1185-94.

28. Koide, A., et al., *The fibronectin type III domain as a scaffold for novel binding proteins.* J Mol Biol, 1998. **284**(4): p. 1141-51.

29. Koide, A., et al., *Stabilization of a fibronectin type III domain by the removal of unfavorable electrostatic interactions on the protein surface.* Biochemistry, 2001. **40**(34): p. 10326-33.

30. Dutta, S., et al., *High-affinity fragment complementation of a fibronectin type III domain and its application to stability enhancement.* Protein Sci, 2005. **14**(11): p. 2838-48.

31. Batori, V., A. Koide, and S. Koide, *Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain.* Protein Eng, 2002. **15**(12): p. 1015-20.

32. Xu, L., et al., *Directed evolution of high-affinity antibody mimics using mRNA display.* Chem Biol, 2002. **9**(8): p. 933-42.

33. Richards, J., et al., *Engineered fibronectin type III domain with a RGDWXE sequence binds with enhanced affinity and specificity to human alphavbeta3 integrin.* J Mol Biol, 2003. **326**(5): p. 1475-88.

34. Koide, A., et al., *Probing protein conformational changes in living cells by using designer binding proteins: application to the estrogen receptor.* Proc Natl Acad Sci U S A, 2002. **99**(3): p. 1253-8.

35. Karatan, E., et al., *Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain.* Chem Biol, 2004. **11**(6): p. 835-44.

36. Getmanova, E.V., et al., *Antagonists to human and mouse vascular endothelial growth factor receptor 2 generated by directed protein evolution in vitro.* Chem Biol, 2006. **13**(5): p. 549-56.

37. Parker, M.H., et al., *Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two.* Protein Eng Des Sel, 2005. **18**(9): p. 435-44.

38. Koide, A., et al., *High-affinity single-domain binding proteins with a binary-code interface.* Proc Natl Acad Sci U S A, 2007. **104**(16): p. 6632-7.

39. Duan, J., et al., *Fibronectin type III domain based monobody with high avidity.* Biochemistry, 2007. **46**(44): p. 12656-64.

40. Lipovsek, D., et al., *Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: molecular convergence with single-domain camelid and shark antibodies.* J Mol Biol, 2007. **368**(4): p. 1024-41.

41. Hackel, B.J., A. Kapila, and K.D. Wittrup, *Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling.* J Mol Biol, 2008. **381**(5): p. 1238-52.

42. Myers, J.K. and T.G. Oas, *Preorganized secondary structure as an important determinant of fast protein folding.* Nat Struct Biol, 2001. **8**(6): p. 552-8.

43. Nord, K., et al., *A combinatorial library of an alpha-helical bacterial receptor domain.* Protein Eng, 1995. **8**(6): p. 601-8.

44. Nord, K., et al., *Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain.* Nat Biotechnol, 1997. **15**(8): p. 772-7.

45. Gunneriusson, E., et al., *Affinity maturation of a Taq DNA polymerase specific affibody by helix shuffling.* Protein Eng, 1999. **12**(10): p. 873-8.

46. Nord, K., et al., *Ligands selected from combinatorial libraries of protein A for use in affinity capture of apolipoprotein A-1M and taq DNA polymerase.* J Biotechnol, 2000. **80**(1): p. 45-54.

47. Wikman, M., et al., *Selection and characterization of HER2/neu-binding affibody ligands.* Protein Eng Des Sel, 2004. **17**(5): p. 455-62.

48. Friedman, M., et al., *Phage display selection of Affibody molecules with specific binding to the extracellular domain of the epidermal growth factor receptor.* Protein Eng Des Sel, 2007. **20**(4): p. 189-99.

49. Friedman, M., et al., *Directed evolution to low nanomolar affinity of a tumor-targeting epidermal growth factor receptor-binding affibody molecule.* J Mol Biol, 2008. **376**(5): p. 1388-402.

50. Orlova, A., et al., *Tumor imaging using a picomolar affinity HER2 binding affibody molecule.* Cancer Res, 2006. **66**(8): p. 4339-48.

51. Ronnmark, J., et al., *Human immunoglobulin A (IgA)-specific ligands from combinatorial engineering of protein A.* Eur J Biochem, 2002. **269**(11): p. 2647-55.

52. Renberg, B., et al., *Affibody molecules in protein capture microarrays: evaluation of multidomain ligands and different detection formats.* J Proteome Res, 2007. **6**(1): p. 171-9.

53. Renberg, B., et al., *Affibody protein capture microarrays: synthesis and evaluation of random and directed immobilization of affibody molecules.* Anal Biochem, 2005. **341**(2): p. 334-43.

54. Nord, K., et al., *Recombinant human factor VIII-specific affinity ligands selected from phage-displayed combinatorial libraries of protein A.* Eur J Biochem, 2001. **268**(15): p. 4269-77.

55. Hoyer, W. and T. Hard, *Interaction of Alzheimer's A beta peptide with an engineered binding protein--thermodynamics and kinetics of coupled folding-binding.* J Mol Biol, 2008. **378**(2): p. 398-411.

56. Jonsson, A., et al., *Engineering of a femtomolar affinity binding protein to human serum albumin.* Protein Eng Des Sel, 2008. **21**(8): p. 515-27.

57. Grimm, S., et al., *Selection and characterisation of affibody molecules inhibiting the interaction between Ras and Raf in vitro.* N Biotechnol, 2010. **27**(6): p. 766-73.

58. Grimm, S., S. Salahshour, and P.A. Nygren, *Monitored whole gene in vitro evolution of an anti-hRaf-1 affibody molecule towards increased binding affinity.* N Biotechnol, 2011.

59. Grove, T.Z., A.L. Cortajarena, and L. Regan, *Ligand binding by repeat proteins: natural and designed.* Curr Opin Struct Biol, 2008. **18**(4): p. 507-15.

60. Main, E.R., et al., *Local and long-range stability in tandemly arrayed tetratricopeptide repeats.* Proc Natl Acad Sci U S A, 2005. **102**(16): p. 5721-6.

61. Cortajarena, A.L., F. Yi, and L. Regan, *Designed TPR modules as novel anticancer agents.* ACS Chem Biol, 2008. **3**(3): p. 161-6.

62. Jackrel, M.E., et al., *Screening libraries to identify proteins with desired binding activities using a split-GFP reassembly assay.* ACS Chem Biol, 2010. **5**(6): p. 553-62.

63. Parmeggiani, F., et al., *Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core.* J Mol Biol, 2008. **376**(5): p. 1282-304.

64. Urvoas, A., et al., *Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (alphaRep) based on thermostable HEAT-like repeats.* J Mol Biol, 2010. **404**(2): p. 307-27.

65. Bork, P., *Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally?* Proteins, 1993. **17**(4): p. 363-74.

66. Binz, H.K., et al., *Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins.* J Mol Biol, 2003. **332**(2): p. 489-503.

67. Wetzel, S.K., et al., *Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins.* J Mol Biol, 2008. **376**(1): p. 241-57.

68. Merz, T., et al., *Stabilizing ionic interactions in a full-consensus ankyrin repeat protein.* J Mol Biol, 2008. **376**(1): p. 232-40.

69. Kohl, A., et al., *Designed to be stable: crystal structure of a consensus ankyrin repeat protein.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1700-5.

70. Forrer, P., et al., *Consensus design of repeat proteins.* Chembiochem, 2004. **5**(2): p. 183-9.

71. Steiner, D., P. Forrer, and A. Pluckthun, *Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display.* J Mol Biol, 2008. **382**(5): p. 1211-27.

72. Steiner, D., et al., *Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display.* Nat Biotechnol, 2006. **24**(7): p. 823-31.

73. Binz, H.K., et al., *High-affinity binders selected from designed ankyrin repeat protein libraries.* Nat Biotechnol, 2004. **22**(5): p. 575-82.

74. Amstutz, P., et al., *Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins.* J Biol Chem, 2005. **280**(26): p. 24715-22.

75. Kohl, A., et al., *Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein.* Structure, 2005. **13**(8): p. 1131-41.

76. Kawe, M., et al., *Isolation of intracellular proteinase inhibitors derived from designed ankyrin repeat proteins by genetic screening.* J Biol Chem, 2006. **281**(52): p. 40252-63.

77. Amstutz, P., et al., *Rapid selection of specific MAP kinase-binders from designed ankyrin repeat protein libraries.* Protein Eng Des Sel, 2006. **19**(5): p. 219-29.

78. Zahnd, C., et al., *Selection and characterization of Her2 binding-designed ankyrin repeat proteins.* J Biol Chem, 2006. **281**(46): p. 35167-75.

79. Zahnd, C., et al., *A designed ankyrin repeat protein evolved to picomolar affinity to Her2.* J Mol Biol, 2007. **369**(4): p. 1015-28.

80. Schweizer, A., et al., *Inhibition of caspase-2 by a designed ankyrin repeat protein: specificity, structure, and inhibition mechanism.* Structure, 2007. **15**(5): p. 625-36.

81. Milovnik, P., et al., *Selection and characterization of DARPins specific for the neurotensin receptor 1.* Protein Eng Des Sel, 2009. **22**(6): p. 357-66.

82. Winkler, J., et al., *EpCAM-targeted delivery of nanocomplexed siRNA to tumor cells with designed ankyrin repeat proteins.* Mol Cancer Ther, 2009. **8**(9): p. 2674-83.

83. Stefan, N., et al., *DARPins Recognizing the Tumor-Associated Antigen EpCAM Selected by Phage and Ribosome Display and Engineered for Multivalency.* J Mol Biol, 2011. **413**(4): p. 826-43.

84. Schlehuber, S. and A. Skerra, *Lipocalins in drug discovery: from natural ligand-binding proteins to "anticalins".* Drug Discov Today, 2005. **10**(1): p. 23-33.

85. Skerra, A., *Alternative binding proteins: anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities.* FEBS J, 2008. **275**(11): p. 2677-83.

86. Beste, G., et al., *Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold.* Proc Natl Acad Sci U S A, 1999. **96**(5): p. 1898-903.

87. Vopel, S., H. Muhlbach, and A. Skerra, *Rational engineering of a fluorescein-binding anticalin for improved ligand affinity.* Biol Chem, 2005. **386**(11): p. 1097-104.

88. Schlehuber, S., G. Beste, and A. Skerra, *A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin.* J Mol Biol, 2000. **297**(5): p. 1105-20.

89. Schlehuber, S. and A. Skerra, *Duocalins: engineered ligand-binding proteins with dual specificity derived from the lipocalin fold.* Biol Chem, 2001. **382**(9): p. 1335-42.

90. Schlehuber, S. and A. Skerra, *Tuning ligand affinity, specificity, and folding stability of an engineered lipocalin variant -- a so-called 'anticalin' -- using a molecular random approach.* Biophys Chem, 2002. **96**(2-3): p. 213-28.

91. Korndorfer, I.P., S. Schlehuber, and A. Skerra, *Structural mechanism of specific ligand recognition by a lipocalin tailored for the complexation of digoxigenin.* J Mol Biol, 2003. **330**(2): p. 385-96.

92. Korndorfer, I.P., G. Beste, and A. Skerra, *Crystallographic analysis of an "anticalin" with tailored specificity for fluorescein reveals high structural plasticity of the lipocalin loop region.* Proteins, 2003. **53**(1): p. 121-9.

93. Breustedt, D.A., et al., *The 1.8-A crystal structure of human tear lipocalin reveals an extended branched cavity with capacity for multiple ligands.* J Biol Chem, 2005. **280**(1): p. 484-93.

94. Breustedt, D.A., L. Chatwell, and A. Skerra, *A new crystal form of human tear lipocalin reveals high flexibility in the loop region and induced fit in the ligand cavity.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 10): p. 1118-25.

95. Mills, J.L., et al., *NMR structure and dynamics of the engineered fluorescein-binding lipocalin FluA reveal rigidification of beta-barrel and variable loops upon enthalpy-driven ligand binding.* Biochemistry, 2009. **48**(31): p. 7411-9.

96. Binder, U., et al., *High-throughput sorting of an Anticalin library via EspP-mediated functional display on the Escherichia coli cell surface.* J Mol Biol, 2010. **400**(4): p. 783-802.

97. Schonfeld, D., et al., *An engineered lipocalin specific for CTLA-4 reveals a combining site with structural and conformational features similar to antibodies.* Proc Natl Acad Sci U S A, 2009. **106**(20): p. 8198-203.

98. Kolmar, H., *Alternative binding proteins: biological activity and therapeutic potential of cystine-knot miniproteins.* FEBS J, 2008. **275**(11): p. 2684-90.

99. Kolmar, H., *Engineered cystine-knot miniproteins for diagnostic applications.* Expert Rev Mol Diagn, 2010. **10**(3): p. 361-8.

100. Smith, G.P., et al., *Small binding proteins selected from a combinatorial repertoire of knottins displayed on phage.* J Mol Biol, 1998. **277**(2): p. 317-32.

101. Christmann, A., et al., *The cystine knot of a squash-type protease inhibitor as a structural scaffold for Escherichia coli cell surface display of conformationally constrained peptides.* Protein Eng, 1999. **12**(9): p. 797-806.

102. Reiss, S., et al., *Inhibition of platelet aggregation by grafting RGD and KGD sequences on the structural scaffold of small disulfide-rich proteins.* Platelets, 2006. **17**(3): p. 153-7.

103. Kimura, R.H., et al., *Engineered cystine knot peptides that bind alphavbeta3, alphavbeta5, and alpha5beta1 integrins with low-nanomolar affinity.* Proteins, 2009. **77**(2): p. 359-69.

104. Silverman, A.P., et al., *Engineered cystine-knot peptides that bind alpha(v)beta(3) integrin with antibody-like affinities.* J Mol Biol, 2009. **385**(4): p. 1064-75.

105. Silverman, A.P., M.S. Kariolis, and J.R. Cochran, *Cystine-knot peptides engineered with specificities for alpha(IIb)beta(3) or alpha(IIb)beta(3) and alpha(v)beta(3) integrins are potent inhibitors of platelet aggregation.* J Mol Recognit, 2011. **24**(1): p. 127-35.

106. Weissleder, R., *Molecular imaging in cancer.* Science, 2006. **312**(5777): p. 1168-71.

107. Schmidt, M.M. and K.D. Wittrup, *A modeling analysis of the effects of molecular size and binding affinity on tumor targeting.* Mol Cancer Ther, 2009. **8**(10): p. 2861-71.

108. Miao, Z., J. Levi, and Z. Cheng, *Protein scaffold-based molecular probes for cancer molecular imaging.* Amino Acids, 2011. **41**(5): p. 1037-47.

109. Landon, L.A. and S.L. Deutscher, *Combinatorial discovery of tumor targeting peptides using phage display.* J Cell Biochem, 2003. **90**(3): p. 509-17.

110. Tolmachev, V., *Imaging of HER-2 overexpression in tumors for guiding therapy.* Curr Pharm Des, 2008. **14**(28): p. 2999-3019.

111. Tolmachev, V., et al., *Radionuclide therapy of HER2-positive microxenografts using a 177Lu-labeled HER2-specific Affibody molecule.* Cancer Res, 2007. **67**(6): p. 2773-82.

112. Zahnd, C., et al., *Efficient tumor targeting with high-affinity designed ankyrin repeat proteins: effects of affinity and molecular size.* Cancer Res, 2010. **70**(4): p. 1595-605.

113. Jonsson, A., et al., *Generation of tumour-necrosis-factor-alpha-specific affibody molecules capable of blocking receptor binding in vitro.* Biotechnol Appl Biochem, 2009. **54**(2): p. 93-103.

114. Malmberg, J., V. Tolmachev, and A. Orlova, *Imaging agents for in vivo molecular profiling of disseminated prostate cancer--targeting EGFR receptors in prostate cancer: comparison of cellular processing of [111In]-labeled affibody molecule Z(EGFR:2377) and cetuximab.* Int J Oncol, 2011. **38**(4): p. 1137-43.

115. Nordberg, E., et al., *In vivo and in vitro uptake of 111In, delivered with the affibody molecule (ZEGFR:955)2, in EGFR expressing tumour cells.* Oncol Rep, 2008. **19**(4): p. 853-7.

116. Miao, Z., et al., *An engineered knottin peptide labeled with 18F for PET imaging of integrin expression.* Bioconjug Chem, 2009. **20**(12): p. 2342-7.

117. Kimura, R.H., et al., *Engineered knottin peptides: a new class of agents for imaging integrin expression in living subjects.* Cancer Res, 2009. **69**(6): p. 2435-42.

118. Wurch, T., et al., *Development of novel protein scaffolds as alternatives to whole antibodies for imaging and therapy: status on discovery research and clinical validation.* Curr Pharm Biotechnol, 2008. **9**(6): p. 502-9.

119. Jiang, L., et al., *Preliminary evaluation of (177)Lu-labeled knottin peptides for integrin receptor-targeted radionuclide therapy.* Eur J Nucl Med Mol Imaging, 2011. **38**(4): p. 613-22.

120. Lehmann, A., *Ecallantide (DX-88), a plasma kallikrein inhibitor for the treatment of hereditary angioedema and the prevention of blood loss in on-pump cardiothoracic surgery.* Expert Opin Biol Ther, 2008. **8**(8): p. 1187-99.

121. Devy, L., et al., *PEGylated DX-1000: pharmacokinetics and antineoplastic activity of a specific plasmin inhibitor.* Neoplasia, 2007. **9**(11): p. 927-37.

122. Mamluk, R., et al., *Anti-tumor effect of CT-322 as an adnectin inhibitor of vascular endothelial growth factor receptor-2.* MAbs, 2010. **2**(2): p. 199-208.

123. Silverman, J., et al., *Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains.* Nat Biotechnol, 2005. **23**(12): p. 1556-61.

124. Skerra, A., *Anticalins as alternative binding proteins for therapeutic use.* Curr Opin Mol Ther, 2007. **9**(4): p. 336-44.

125. Hohlbaum, A.M. and A. Skerra, *Anticalins: the lipocalin family as a novel protein scaffold for the development of next-generation immunotherapies.* Expert Rev Clin Immunol, 2007. **3**(4): p. 491-501.

126. Ronnmark, J., et al., *Affibody-beta-galactosidase immunoconjugates produced as soluble fusion proteins in the Escherichia coli cytosol.* J Immunol Methods, 2003. **281**(1-2): p. 149-60.

127. Lundberg, E., et al., *Site-specifically conjugated anti-HER2 Affibody molecules as one-step reagents for target expression analyses on cells and xenograft samples.* J Immunol Methods, 2007. **319**(1-2): p. 53-63.

128. Ronnmark, J., et al., *Construction and characterization of affibody-Fc chimeras produced in Escherichia coli.* J Immunol Methods, 2002. **261**(1-2): p. 199-211.

129. Sandstrom, K., et al., *Inhibition of the CD28-CD80 co-stimulation signal by a CD28-binding affibody ligand developed by combinatorial protein engineering.* Protein Eng, 2003. **16**(9): p. 691-7.

130. Gronwall, C., et al., *Generation of Affibody ligands binding interleukin-2 receptor alpha/CD25.* Biotechnol Appl Biochem, 2008. **50**(Pt 2): p. 97-112.

131. Li, J., et al., *Selection of affibody molecules to the ligand-binding site of the insulin-like growth factor-1 receptor.* Biotechnol Appl Biochem, 2010. **55**(2): p. 99-109.

132. Nordberg, E., et al., *Cellular studies of binding, internalization and retention of a radiolabeled EGFR-binding affibody molecule.* Nucl Med Biol, 2007. **34**(6): p. 609-18.

133. Bloom, L. and V. Calabro, *FN3: a new protein scaffold reaches the clinic.* Drug Discov Today, 2009. **14**(19-20): p. 949-55.

134. Sennhauser, G. and M.G. Grutter, *Chaperone-assisted crystallography with DARPins.* Structure, 2008. **16**(10): p. 1443-53.

135. Koide, S., *Engineering of recombinant crystallization chaperones.* Curr Opin Struct Biol, 2009. **19**(4): p. 449-57.

136. Cochran, J.R., *Engineered proteins pull double duty.* Sci Transl Med, 2010. **2**(17): p. 17ps5.

137. Friedman, M., et al., *Engineering and characterization of a bispecific HER2 x EGFR-binding affibody molecule.* Biotechnol Appl Biochem, 2009. **54**(2): p. 121-31.

138. Eggel, A., et al., *DARPins as bispecific receptor antagonists analyzed for immunoglobulin E receptor blockage.* J Mol Biol, 2009. **393**(3): p. 598-607.

139. Emanuel, S.L., et al., *A fibronectin scaffold approach to bispecific inhibitors of epidermal growth factor receptor and insulin-like growth factor-I receptor.* MAbs, 2011. **3**(1): p. 38-48.

140. Koide, A., et al., *The fibronectin type III domain as a scaffold for novel binding proteins.* Journal of Molecular Biology, 1998. **284**(4): p. 1141-1151.

141. Lipovsek, D., et al., *Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: Molecular convergence with single-domain camelid and shark antibodies.* Journal of Molecular Biology, 2007. **368**(4): p. 1024-1041.

142. Xu, L.H., et al., *Directed evolution of high-affinity antibody mimics using mRNA display.* Chemistry & Biology, 2002. **9**(8): p. 933-942.

143. Parker, M.H., et al., *Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two.* Protein Engineering Design & Selection, 2005. **18**(9): p. 435-444.

144. Nord, K., et al., *Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain.* Nature Biotechnology, 1997. **15**(8): p. 772-777.

145. Nord, K., et al., *Recombinant human factor VIII-specific affinity ligands selected from phage-displayed combinatorial libraries of protein A.* European Journal of Biochemistry, 2001. **268**(15): p. 4269-4277.

146. Gronwall, C., et al., *Selection and characterization of affibody ligands binding to Alzheimer amyloid beta peptides.* Journal of Biotechnology, 2007. **128**(1): p. 162-183.

147. Orlova, A., et al., *Tumor Imaging using a picomolar affinity HER2 binding affibody molecule.* Cancer Research, 2006. **66**(8): p. 4339-4348.

148. Beste, G., et al., *Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(5): p. 1898-1903.

149. Schlehuber, S., G. Beste, and A. Skerra, *A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin.* Journal of Molecular Biology, 2000. **297**(5): p. 1105-1120.

150. Schlehuber, S. and A. Skerra, *Lipocalins in drug discovery: from natural ligand-binding proteins to 'anticalins'.* Drug Discovery Today, 2005. **10**(1): p. 23-33.

151. Skerra, A., *Anticalins as alternative binding proteins for therapeutic use.* Current Opinion in Molecular Therapeutics, 2007. **9**(4): p. 336-344.

152. Wyler, E., et al., *Inhibition of NF-kappaB activation with designed ankyrin-repeat proteins targeting the ubiquitin-binding/oligomerization domain of NEMO.* Protein Sci, 2007. **16**(9): p. 2013-22.

**Figures**

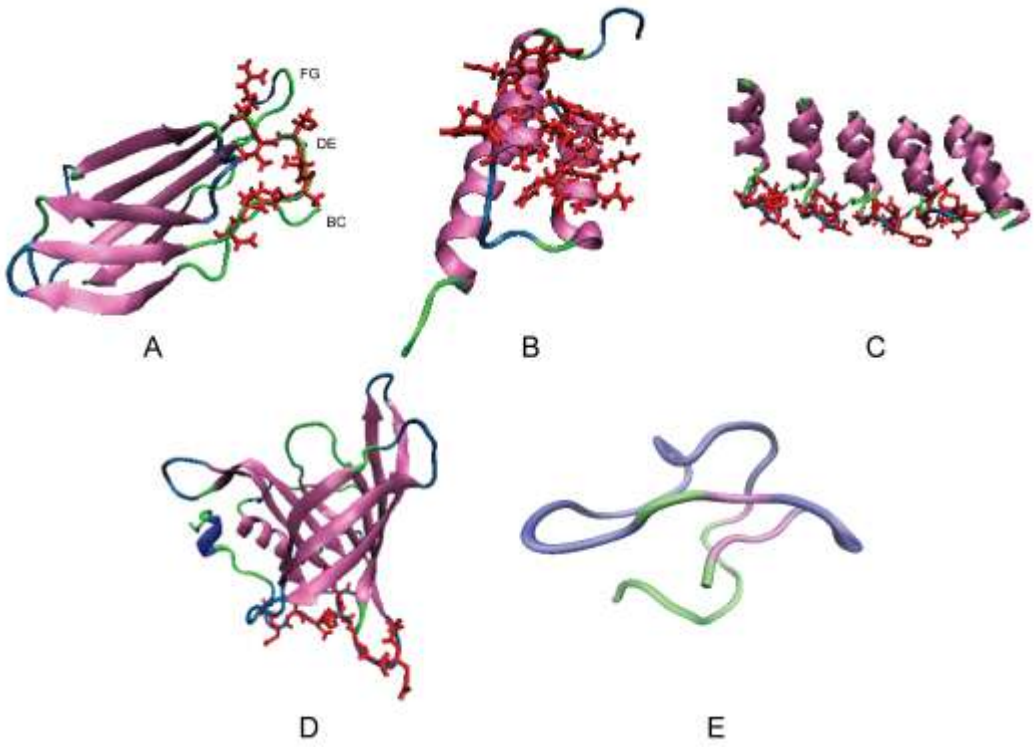



**Figure 1.1**: Non-immunoglobulin scaffolds for molecular recognition, (A) 10th domain of fibronectin , (B) a three helix bundle from staphylococcal protein A (Affibody) , (C) Designed ankyrin repeat protein , (D) Lipocalin (retinol binding protein) , (E) Cystine-knot miniprotein (Knottin from *Ecballium elaterium* Trypsin inhibitor-II).

**Table 1.1**: Common non-immunoglobulin scaffolds engineered for a variety of targets

| *Scaffold(a.a.)*[a] | *Mutation Strategy* | *Targets(partial list)* |
|---|---|---|
| $10^{th}$ domain of fibronectin (94 aa), Adnectins[b] | Loop mutagenesis | Ubiquitin [140],Hen Egg Lysozyme[141],TNF-α[142], Vascular endothelial growth factor receptor two[143] |
| Immunoglobulin binding domain of protein A(58aa), Affibody[b] | Flat surface mutagenesis | Taq DNA polymerase[144], Human factor VIII[145], Alzheimer amyloid β-peptides[146],Breast cancer marker HER2[147] |
| Lipocalins(160-180aa), Anticalins[b] (Retinol binding protein, 174 aa) | Loop mutagenesis | Small molecules: Fluorescein[148], Digoxigenin[149], Macromolecules: Cytotoxic T lymphocyte-associated antigen[150], Vascular endothelial growth factors[151] |
| Designed Ankyrin Repeat proteins, DARPins[b] (67+n.33)aa where n is number of repeat units, which can be varied | Flat surface mutagenesis | NF-kappaB[152], HER2[79], MAP-kinase[77] |
| Cystine-knot miniproteins, Knottins[b](20-60aa) | Loop grafting/mutagenesis | Alkaline phosphatase[100], integrins αvβ3, αiibβ3[103-105] |

**a**: size in terms of amino acid length,
**b**: common name of the scaffold

# CHAPTER 2

# PROTEIN SELECTION USING YEAST SURFACE DISPLAY

Nimish Gera and Balaji M. Rao

Department of Chemical and Biomolecular Engineering

North Carolina State University

Raleigh, NC 27695-7905

**(**Adapted from Gera et al. 2011, *Methods* (In Press))

## 2.1. Introduction

The ability to generate proteins that bind with high affinity and specificity to any given target is critical to many applications in medicine, biotechnology and research. Binding proteins are typically isolated from a combinatorial library of mutant proteins, generated through mutagenesis of certain regions on a stable protein framework. The key challenge in this task is the identification of mutants from the library that bind the desired target with the highest affinity, and can be accomplished using one of the many library screening tools available. Despite their differences, the common principle underlying the use of any screening tool is the linkage of each mutant protein to its genetic code, thereby "bar-coding" the mutant. The protein(s) that bind the target are isolated and subsequently identified through DNA sequencing. In phage display, the oldest and most widely cited tool for screening combinatorial libraries, the mutant proteins are fused to the bacteriophage coat protein[1]. Thus the phage particle carries both the mutant protein and the DNA encoding the mutant protein, linking phenotype to genotype. Similarly, other cell-based systems rely on the expression of mutant proteins as fusions to cell surface proteins in bacterial [2-4], yeast [5-8] or mammalian cells [9, 10]; the host cell carries a plasmid vector that encodes the mutant protein. In addition to these systems where a phage particle or a host cell links the protein to its genetic sequence, two cell-free systems have also been extensively used for screening combinatorial libraries of proteins-ribosome display and mRNA display [11]. The mutant protein is physically linked to its coding mRNA through the ribosome in case of ribosome display [12-14]. Along similar lines, in case of mRNA display, the protein is covalently linked to the mRNA through a puromycin adaptor [15-18].

In this review, we predominantly focus on the use of yeast surface display for the isolation of binding proteins from combinatorial libraries. We present a broad overview of the yeast surface display platform, discuss its application to a wide array of problems in protein engineering and compare it with other tools for screening libraries of mutant proteins. We also provide detailed protocols for generation and screening of combinatorial libraries and biophysical characterization of binding proteins using yeast surface display.

### 2.1.1. The yeast surface display platform

As discussed earlier, the central feature of any library screening tool is the linkage of protein to its corresponding genetic sequence. The yeast cell provides this linkage in yeast surface display; the mutant protein is expressed as a fusion to a cell wall protein. Different yeast strains and cell surface receptors have been used for cell surface display [7]. However, the *Saccharomyces cerevisiae*-Aga2p system [5, 19] remains the most commonly used. In this system, mutant proteins are expressed as cell surface fusions to the Aga2p subunit of the mating protein a-agglutinin in *Saccharomyces cerevisiae* (Figure 2.1). The Aga2p subunit is linked by two disulfide bonds to the Aga1p subunit, which anchors the entire assembly to the yeast cell wall while allowing the mutant protein to interact with species in solution. The gene encoding the mutant protein is cloned into a single-copy yeast shuttle plasmid vector as a fusion to the Aga2p gene; the yeast strain used has the Aga1p gene stably integrated in the chromosome. Expression of both Aga1p and Aga2p is controlled by the galactose-inducible GAL1 promoter.

In the presence of galactose, the mutant protein is displayed on yeast cell surface as a C-terminal fusion to the Aga2p subunit, as shown in Figure 2.1. Control of cell surface protein expression using the galactose promoter ensures that cells displaying proteins that are toxic to cell growth are not subject to negative selection; protein expression is induced after cell growth. Certain proteins show differential binding affinities when linked to the C or N-terminal of Aga2p subunit due to steric effects [20]. Nevertheless, surface display as C-terminal Aga2p fusions is most common.

In a yeast surface display library, each cell displays ~ 50,000 copies of the mutant protein; in some cases, the expression level may be lower. An N-terminal hemagglutinin (HA) tag and a C-terminal c-myc tag allow for immunofluorescent detection. As described in greater detail in Section 2.3, yeast libraries of up to $10^9$ mutants can be generated with reasonable effort using homologous recombination in yeast. Binding proteins are isolated through a protocol that combines magnetic sorting and flow cytometry. Simplistically, cells are incubated with target-coated magnetic beads or fluorescently labeled soluble target. Cells displaying mutants that bind the target are isolated through magnetic separation or Fluorescence Activated Cell Sorting (FACS).

### 2.1.2. Applications of the yeast surface display platform

Since its inception, the yeast surface display platform has been used in a wide array of applications including generation of *de novo* binders from combinatorial libraries, affinity maturation, engineering of binding specificity and stability, and biophysical characterization of binding proteins. Here we review the use of yeast surface display in the context of these applications.

#### 2.1.2.1. Isolation of *de novo* binders and affinity maturation

The earliest studies on yeast surface display reported the generation of high affinity binders through mutation of an existing binding protein. Indeed, Boder and Wittrup demonstrated the use of the yeast surface display system in the context of affinity maturation; a small library($10^5$ variants) of 4-4-20 single chain variable fragment (scFv) mutants was screened for fluorescein and a three-fold reduced off-rate was obtained [5]. In another study, an scFv(KJ16) specific to the Vβ8 region of a T-cell receptor was affinity matured to have a three-fold higher affinity[21]. Since then, yeast surface display has also been used to isolate *de novo* binding proteins from combinatorial libraries of scFvs as well as a variety of non-immunoglobulin scaffolds. A naive non immune human scFv library of $10^9$ clones was created using yeast surface display and screened using magnetic screening and flow cytometry to isolate high affinity scFvs to a wide spectrum of targets including proteins, peptides and haptens [22]. Needless to say, yeast display is a powerful tool to increase the binding affinity of the *de novo* binders isolated. Target-scFv interactions with very high

binding affinity have been engineered by directed evolution by yeast surface display. An scFv against fluorescein with an equilibrium dissociation constant ($K_D$) of 48 fM, one of the strongest protein-ligand interactions reported to date, was engineered using yeast surface display[23]. In another study, a scFv with a 4 day long dissociation half time at 37°C against the carcinoembryonic antigen was isolated [24].

In addition to scFvs, several other antibody-related proteins have been engineered using yeast surface display to enhance affinities of existing binding interactions or to introduce *de novo* binding function. Structurally similar to scFvs, a single chain T cell receptor (scTCR) was engineered to obtain scTCR mutants that bind the peptide/MHC ligand with high affinity ($K_D$=9 nM) [25]. This study shows that the intrinsic low affinity of the TCR-peptide/MHC interaction is not due to inherent structural or genetic limitations. In a more recent study, the Fc portion of human IgG1 was engineered to bind HER2/neu; mutant proteins with a $K_D$ of 69 nM were obtained after affinity maturation [26]. Interestingly, these Fc mutants retain binding activity to protein A and CD64. Thus a single 50 kDa homodimeric protein integrates antigen binding function as well as effector function.

Over the last several years, there has been increasing interest in the use of non-immunoglobulin scaffolds for generating binding proteins[27]. Their favorable properties such as low molecular weight and relatively easy recombinant expression makes them attractive candidates for a variety of clinical, diagnostic and biotechnology applications. In principle, yeast surface display can be used to isolate binders from combinatorial libraries of alternate scaffold proteins; indeed, several examples have been previously reported. The 10$^{th}$

domain of fibronectin type III (10Fn3) adopts a β-sandwich type fold similar to the immunoglobulins. The three loops in fibronectin are similar to the Complementarity Determining Regions (CDRs) in immunoglobulins. Yeast surface display has been used extensively for generating binding proteins based on the fibronectin scaffold; combinatorial libraries of 10Fn3 mutants have been generated by altering the loop length and composition [28, 29].

Green Fluorescent protein (GFP) serves as a useful scaffold because the isolated binders have intrinsic fluorescence, allowing for easy detection. Yeast surface display and directed evolution was used to isolate nanomolar affinity binding proteins to four different targets; variable loop regions were inserted in the GFP scaffold [30, 31]. Another scaffold where binding proteins were generated through mutagenesis of loop regions, using yeast surface display, is the cystine knot peptide (knottin). Binding proteins to the $\alpha_v\beta_3$ integrin were engineered from the Agouti-related protein (AgRP) scaffold, a knottin. These proteins bound with nanomolar affinities to cells expressing $\alpha_v\beta_3$ integrin and had no binding to the other closely related integrins [32, 33]. In another study, the *Ecballium elaterium* trypsin inhibitor-II (EETI-II), another knottin, was engineered using yeast surface display to bind $\alpha_v\beta_3$, $\alpha_v\beta_5$ and $\alpha_5\beta_1$ integrins with high affinities[34]. Further, an EETI-II mutant containing two separate integrin binding epitopes was also engineered using yeast surface display. This engineered mutant bound with high affinity to $\alpha_v\beta_3$, $\alpha_v\beta_5$ but relatively lower affinities to $\alpha_5\beta_1$ and $\alpha_{iib}\beta_3$[35]. Yeast surface display has also been used to engineer binding proteins through mutagenesis of surface accessible residues on β-strands of a scaffold protein; binding proteins based on the hyperthermophilic Sso7d scaffold were isolated against six different

target proteins. These proteins are highly stable against thermal and chemical denaturation and can tolerate a range of pH extremes [36]. Other examples of alternate scaffolds engineered using yeast surface display include the Kringle domain [37] and Leptins [38].

Table 2.1 provides an overview of the numerous examples of binding proteins isolated from combinatorial libraries of scFvs and other scaffolds. A striking observation from Table 2.1 is that binding proteins to a wide spectrum of targets have been obtained using a variety of different scaffolds. This is despite the use of modestly sized combinatorial libraries; yeast libraries tend to be one to two orders of magnitude smaller than those screened using other screening techniques (see section 2.1.3). Apart from these examples, yeast surface display has also been used for increasing the binding affinity of the interaction between cytokines and their receptors. Interleukin-2 (IL-2) mutants with picomolar binding affinity for the alpha subunit of the IL-2 receptor (IL-2R□) were generated by directed evolution; the wild-type protein binds IL-2Rα with nanomolar affinity. The resultant mutants could stimulate the proliferation of activated T-cells in *in vitro* assays to a significantly greater extent than wild-type IL-2 [39, 40]. Similarly, Epidermal Growth Factor (EGF) mutants that bind with higher affinity to the Epidermal Growth Factor receptor (EGFR) have also been generated using yeast surface display[41].

Finally, in addition to engineering binding affinity, yeast surface display has also been used to engineer the specificity of binding. Antibody specificity and cross reactivity of a botulinum neurotoxin type A (BoNT/A) binding scFv was tuned using yeast surface display. Specifically, the scFv affinity for BoNT/A2 subtype was enhanced by 1250 fold while

keeping its affinity for the BoNT/A1 subtype constant [42]. In another study, yeast surface display and directed evolution was used to generate conformation-specific binders against mammalian calmodulin (CaM) which changes conformation upon calcium binding. Binding proteins that preferentially bind either forms $Ca^{+2}$ bound/unbound were isolated from a human non immune scFv library displayed on yeast [43].

Thus, taken together, the examples described in this section illustrate the versatility and broad applicability of the yeast surface display system in the context of engineering affinity and specificity of molecular recognition.

#### 2.1.2.2. Characterization of binding interactions

The yeast surface display platform can be used to quantitatively characterize biophysical properties of cell surface displayed proteins such as affinity of binding and thermal stability. This provides a convenient method to characterize binding proteins isolated from combinatorial libraries, without additional sub-cloning and soluble protein expression. The $K_D$ of interaction between the binding protein and its target can be measured using yeast cell surface titrations; the binding protein is expressed as a cell surface fusion (see section 2.3.5). Briefly, yeast cells displaying the protein are labeled with varying concentrations of soluble target. Subsequently, the fraction of cell surface fusions bound to the target is determined using flow cytometry and the data are fit to a single step binding isotherm to estimate the $K_D$. Two points are important to note here. First, several studies have shown that $K_D$ of a binding interaction determined using yeast cell surface displayed protein is essentially equivalent to the $K_D$ measured using soluble protein[28]. Second, yeast cell surface titrations can

differentiate between mutants with similar affinity; fine affinity discrimination between mutants with a two-fold difference in $K_D$ has been achieved [44]. Fine affinity discrimination is crucial for understanding the importance of point mutations. Indeed, significant improvements in affinity are usually additive contributions from single point mutations that impart subtle changes. Yeast surface display can also be used to assess the thermal stability of binding proteins. As with determination of $K_D$, the cell surface fusions act as a substitute for soluble protein. The relative thermal stabilities of scFvs have been determined by assessing the extent of irreversible protein denaturation in yeast cell surface fusions, as a function of temperature [45].

Another application of yeast surface display is in the identification of binding epitopes on a target protein. Domain-level and fine epitope mapping has been achieved using yeast surface display. Small domains of the target protein or full-length protein can be displayed on the yeast cell surface and labeled with an antibody or binding protein to determine epitopes on the target. Stable fragments of EGFR were displayed on yeast cell surface and tested for binding to conformation-specific antibodies [46]. Linear and conformational antibody epitopes were easily distinguished by thermal denaturation of the cell surface displayed protein. In a follow up study, fine epitope mapping of EGFR-specific antibodies has also been achieved. The EGFR ectodomain was randomized to create a yeast surface displayed library; this library was tested for antibody binding to determine the antibody contact residues [47]. Similar studies have also been performed for epitope mapping of BoNT/A [48] and the West Nile virus NS1 glycoprotein [49].

### 2.1.2.3. Other applications

In this review, we have largely focused on the use of yeast surface display for the isolation and of binding proteins from combinatorial libraries and their subsequent characterization. Nevertheless, the yeast surface display platform has been used in the context of a broad range of other applications such as engineering protein stability, development of chaperones, identification of protein-protein interactions and engineering enzyme function.

Higher thermodynamic stability is linked to efficient intracellular processing of proteins in the yeast secretory pathway and consequently higher surface expression levels [50]. Therefore, higher expression levels in yeast surface display can be used as a surrogate for higher thermodynamic stability; a protein that is engineered for higher yeast cell surface expression may be expected to have higher stability. This strategy has been used to engineer stable scTCR mutants. A yeast surface display library of scTCR mutants, generated through random mutagenesis, was screened to identify mutations using that resulted in higher surface expression as well as higher stability[51, 52]. Interestingly, proteins with extremely high stability can form molten globules which can escape the yeast quality control mechanism and still be displayed on yeast cell surface[53]. Therefore, the correlation between thermodynamic stability and yeast surface expression should be interpreted with caution.

Generation of ordered crystals for protein X-ray crystallography is a challenge in structural biology. Yeast surface display has been used to identify crystallization chaperones, i.e., proteins assisting in enhancing crystal packing, from a library of camelid single-domain ($V_HH$) antibodies [54]. Multiple crystal forms could be found that can diffract up to atomic

resolution. It is conceivable that other non-immunoglobulin scaffolds can also be used to create such chaperones using yeast surface display. Note that designed ankyrin repeat proteins (DARPins) selected by ribosome display have been useful as crystallization chaperones for co-crystallization of multiple target proteins such as maltose binding protein[55], polo like kinase-1[56], caspase-2[57] and a membrane protein AcrB[58] to name a few[59].

Yeast surface display is also well-suited for proteome-wide analysis of protein-protein interactions. A yeast surface display human cDNA library has been screened to identify protein fragments that bind to tyrosine phosphorylated peptides from auto-phosphorylated EGFR and focal adhesion kinase [60] and phosphatidylinositides [61]. In another study, novel tumor markers that are recognized by antibodies in the serum of breast cancer patients were identified using cell surface display of a cDNA library derived tumor tissue [62]. Thus, yeast surface display of human cDNA libraries can enable the identification of disease-related targets.

Directed evolution of enzymes using cell surface display systems has been limited due to the challenge in maintaining the physical link between DNA and enzymatic activity; upon completion of the enzymatic reaction, the product is readily detached from the cell surface. Yeast surface display has been used to enhance the enantioselectivity of horseradish peroxidase (HRP). The critical step here involves the trapping of a fluorescently labeled substrate on yeast cell surface. Alternative positive/negative screening between L- or D-tyrosinol was used to isolate an HRP mutant with an 8-fold altered enantioselectivity than the

wild type, including a reversal of selectivity towards the D-tyrosinol over L-tyrosinol [63]. In another study, combinatorial yeast surface displayed libraries of the lid domain of *Rhizopus oryzae* lipase were screened to isolate mutants with higher specificity towards short chain substrates [64, 65].

### 2.1.3 Comparison with other screening tools

The central feature of all cell surface or cell-free display systems for screening combinatorial libraries is the linking of each mutant protein to its genetic code. However, each screening platform has its unique features and consequently certain advantages or disadvantages [66]. Table 2.2 summarizes key features of several different display technologies. Here we discuss the important differences between yeast surface display and other screening platforms.

As discussed earlier, in yeast surface display, the mutant proteins are expressed as cell surface fusions to a cell wall protein. The yeast surface displayed fusion protein is easily accessible to large macromolecular proteins like antibodies. In comparison, bacterial surface display is limited by steric hindrance from the lipo-polysaccharide layer [67]. Yeast surface display also has the unique advantage of eukaryotic post translational machinery such as disulfide isomerization and glycosylation. Despite the differential glycosylation patterns between yeast and mammalian cells, glycosylated proteins have been successfully displayed on yeast. Recent efforts to express proteins with human-like glycosylation structures in yeast suggests that this limitation can be overcome[68]. Therefore, it follows that yeast surface display is an excellent platform for the engineering eukaryotic proteins or proteins with

multiple disulfide bonds. Use of the GAL1 promoter to drive expression of cell surface proteins provides another important benefit in case of yeast surface display. Since expression of displayed proteins is not induced until after the growth phase, proteins that are toxic to cell growth are not subject to negative selection [5]. On the other hand, in some cases, heterologous protein display may cause toxicity in bacterial display; this results in elimination of cells expressing protein variants that affect cell growth and consequently library bias.

In a yeast display library, each yeast cell expresses ~ 50,000 copies of a particular mutant protein. Therefore, magnetic screening with micron-sized target-coated magnetic beads results in a high avidity target-protein interaction; multiple copies of a mutant protein on the yeast cell surface can interact with multiple copies of the target species immobilized on the bead. Consequently, even binding proteins with low affinity for the target can be isolated in the magnetic screening step [69]. By the same token, the highly avid interaction between yeast cells and magnetic beads can be exploited to achieve a very stringent negative selection step. In such a scheme, the yeast library may be incubated with magnetic beads, or that are coated with a non-relevant protein such as secondary antibodies to be used in further analysis or a closely related non-target protein. Subsequently, yeast cells binding the beads are discarded. Due to the high avidity yeast-bead interaction, mutants binding non-target species with even low affinity can be rejected.

The high copy number of cell surface displayed fusions and the size of the yeast cell make the yeast surface display platform compatible with flow cytometry. This enables quantitative

screening of combinatorial libraries. While the magnetic screening step isolates all mutants that bind the target with at least low affinity of interaction, FACS can be used to discriminate between binders with varying affinities for the target. Immunofluorescent labeling of the HA and c-myc epitope tags also allows for normalization of protein expression on yeast cell surface. Indeed, FACS can be used to achieve fine affinity discrimination and mutants with subtle affinity improvements (up to two fold) over the wild-type protein can be isolated [44]. Further, flow cytometry can be used for biophysical characterization of yeast displayed proteins; properties such as $K_D$ of binding interaction and thermal stability can be conveniently measured, without the need for sub-cloning and soluble protein expression [44, 45, 52].

A limitation of yeast surface display is that library sizes are at least 1-2 orders of magnitude smaller than those obtained using other display systems (Table 2.2). However, it can be argued that yeast display libraries have higher functional diversity. In one study comparing yeast display to phage display, an HIV-1 immune scFv library was screened by both yeast surface display and phage display. The yeast display screen yielded additional novel binders to the target antigen HIV-1 gp120 glycoprotein [70]. Further, as shown in Table 2.1, binding proteins for a wide spectrum of targets have been obtained from a variety of scaffold proteins using yeast surface display, suggesting that the lower library sizes do not preclude the isolation of binders. Interestingly, a highly efficient yeast transformation protocol developed recently allows construction of yeast displayed human antibody libraries of up to $10^{10}$ members[71]. By comparison, phage display and ribosome display allow generation of libraries of up to $10^{11}$ and $10^{13}$-$10^{14}$ members respectively[72].

Several recent studies have used a combination of display platforms. Such a strategy harnesses the advantages of multiple display technologies and enables previously unexplored applications. A method for combining yeast display and phage display systems for isolating antigen-antibody pairs has been described [73]. In this study, a phage library displaying fragments of the antigen (HIV-1 gp160) was screened against a yeast library of scFvs obtained from an HIV-1 infected individual displayed on yeast. In another study, yeast surface display library of Mac-1 inserted (I) domains was screened for active mutations against HeLa cells expressing Mac-1 ligands. Subsequently, a phage library was screened against a yeast displayed active I domain [74]. In these examples, the use of yeast display in conjunction with phage display enabled the isolation of antibody-antigen pairs and bypassed the need for soluble target protein. We have also used the yeast surface display platform in conjunction with the mRNA display platform to exploit the high diversity available with mRNA display libraries as well as the advantages of yeast surface display such as quantitative screening and biophysical characterization using flow cytometry. In our protocol, first a high diversity mRNA display library screened against the target. The pool of binders isolated in this screen is subsequently cloned into the yeast surface display platform. Further rounds of FACS-based screening and biophysical characterization are carried out using the yeast platform (unpublished results).

## 2.2. Materials

### 2.2.1. Strains and Plasmids

*Saccharomyces cerevisiae* yeast strain EBY 100[5] is used for yeast surface display. The EBY100 strain is *ura 3-52, trp1, leu 2, leu 2Δ1, his 3Δ200, pep4::HIS2, prb1Δ1.6R* and *can1*. This strain is $Leu^-$ and $Trp^-$ which allows for selection based on the presence of a suitable plasmid vector. In addition it has the Aga1 gene stably integrated in the chromosome with a URA3 selectable marker which is controlled by a GAL promoter. The pCTCON yeast surface display vector[75]can be used for library generation. This plasmid has a tryptophan marker for selective growth in yeast and an ampicillin marker for selective growth in *E. coli*.

### 2.2.2. Media, Agar Plates and Buffers

- YPD(Yeast Peptone Dextrose) Media-10g/L yeast extract, 20g/L peptone and 20g/L dextrose in deionized $H_2O$; filter sterilize with a 0.2-micron filter and store at 4°C. For making plates add 15g/L agar to this solution and autoclave.

- SDCAA Media-20g/L dextrose, 6.7g/L yeast nitrogen base(Becton Dickinson, catalog no.-291940), 5g/L casamino acids (Fisher Scientific, catalog no.-BP1424), 5.4g/L $Na_2HPO_4$, 8.6 g/L $NaH_2PO_4.H_2O$ in deionized $H_2O$, filer sterilize with a 0.2-micron filter and store at 4°C.

- SGCAA Media- can be prepared the same way as SDCAA media by replacing dextrose with galactose. 2g/L dextrose can be added to SGCAA media to promote cell growth during protein induction

- SDCAA plates-First dissolve 5.4g $Na_2HPO_4$, 8.6 g of $NaH_2PO_4.H_2O$, 182 g sorbitol and 15g agar in 900 ml deionized $H_2O$ and autoclave. Also dissolve 20g dextrose, 6.7 g yeast nitrogen base and 5g casamino acids in 100 ml deionized $H_2O$ separately and filter sterilize with a 0.2-micron filter. Cool the autoclaved solution to 50°C and add the filter sterilized solution. Store plates at 4°C.

- Luria-Bertani (LB) Media-10g/L tryptone, 5g/L yeast extract, 10g/L NaCl in deionized $H_2O$ and autoclave. For preparing plates add 15g/L agar to this solution and autoclave. 50 mg/L ampicillin can be added for selection separately after autoclaving and cooling the solution to 50°C.

- Tris-DTT buffer-0.39g/ml 1, 4-dithiothreitol and 1 M Tris pH 8.0, filter sterilize with a 0.2-micron filter. Fresh preparation is recommended. However, this buffer can be stored for up to 6 months at -20°C.

- E buffer-1.2g/L Tris base, 92.4 g/L sucrose, 0.2g/L $MgCl_2$ in deionized water. Adjust pH to 7.5 and filter sterilize with a 0.2-micron filter. Store at room temperature or 4°C.

- PBS-BSA buffer-8g/L NaCl, 0.2g/L KCl, 1.44 g/L $Na_2HPO4$,0.24 g/L $KH_2PO_4$, 1g/L bovine serum albumin (BSA), adjust pH to 7.4 and filter sterilize with a 0.2-micron filter. Store at 4°C.

- Low dextrose SDCAA-5g/L dextrose, 6.7 g/L yeast nitrogen base, 5g/L casamino acids, 5.4g/L $Na_2HPO_4$, 8.6 g/L $NaH_2PO_4.H_2O$ in deionized $H_2O$ and adjust the pH to 6.0, filter sterilize and store at 4°C.

- Yeast freezing solution-10% glycerol**,** 6.7 g/L yeast nitrogen base, autoclave and store at 4°C.

### 2.2.3. Selection Reagents (Magnetic selection and FACS)

- DynaMag$^{TM}$-2 Magnet(Invitrogen, catalog no.-123-21D)
- Dynabeads® Biotin binder(Invitrogen, catalog no.-110-47)
- Biotinylated target antigen (Biotinylation Kit- Thermo Scientific, catalog no.-21435)
- Chicken anti-c-myc IgY (Invitrogen, catalog no.-A21281)
- Mouse anti-HA clone 12CA5 (Roche, catalog no.- 11 583 816 001)
- Alexa Fluor 488 goat anti-chicken IgG (H+L) (Invitrogen, catalog no.-A11039)
- Alexa Fluor 633 goat anti-chicken IgG (H+L) (Invitrogen, catalog no.- A21103)
- Alexa Fluor 488 goat anti-mouse IgG (H+L) (Invitrogen, catalog no.- A11029)
- Alexa Fluor 633 goat anti-mouse IgG (H+L) (Invitrogen, catalog no.-A21052)
- Streptavidin R-phycoerythrin conjugate (Invitrogen, catalog no.- S-866)
- Neutravidin-fluorescein conjugate (Invitrogen, catalog no.-A2662)

### 2.2.4. Transformation and Characterization Reagents

- Zymoprep II yeast plasmid miniprep kit (Zymo Research, catalog no.- D2004)
- Qiaquick PCR purification kit (Qiagen, catalog no.-28104)
- Competent *E. coli* NovaBlue$^{TM}$ cells (EMD Biosciences, catalog no.- 69825)
- Phusion® High Fidelity DNA polymerase ( New England Biolabs, catalog no.-F-530)
- 0.2 cm electroporation cuvettes (Fisher Scientific, catalog no.-FB102)
- Pellet Paint co-precipitant(EMD Biosciences, catalog no.-69049)
- Restriction Enzymes *NheI*(New England Biolabs, catalog no.-R0131), *SalI*(New England Biolabs, catalog no.-R0138), *BamHI* (New England Biolabs, catalog no.-R0136)
- Gene Pulser X cell microbial system (BioRad, catalog no.-165-2662)
- Penicillin-streptomycin/pen-strep (Invitrogen catalog no.-15140-122)

## 2.3. Methods

In this section we describe detailed protocols for isolation of binding proteins from combinatorial libraries of scaffold proteins using yeast surface display; an overview of this process is shown in Ta 2.2**.** Protocols for isolation of scFvs from a yeast display library and subsequent affinity maturation have been previously described [19, 76]. Here we focus on the generation of combinatorial libraries of non-immunoglobulin scaffolds and isolation of *de novo* binders using yeast surface display. The protocols described here can be easily modified for affinity maturation of the isolated binders. Estimation of binding affinities ($K_D$) of individual clones using yeast cell surface titrations is also discussed.

### 2.3.1. Library Construction

The first step in constructing a combinatorial library of yeast surface displayed scaffold mutants is the generation of linear DNA that codes for the mutant proteins. A set of residues on the scaffold protein is identified for randomization. Prior studies on protein structure and stability can help in determining the residues most likely to tolerate mutations. In general, residues important for the structure and stability of a scaffold should not be mutated, because the resultant mutants will be unstable and reduce the quality of the library. The linear DNA library, with selected positions randomized, can be generated using synthetic oligonucleotides that contain degenerate codons. This is illustrated through two examples in Figure 2.3. For small scaffolds, the DNA library can be constructed using a single PCR step, as shown in Figure 2.3A. Here, the synthetic oligonucleotide U1 carries degenerate NNN,

NNK or NNS codons in randomized regions, where N represents an equimolar mixture of A, T, G and C nucleotides, K represents an equimolar mixture of T and G nucleotides and S represents an equimolar mixture of C and G nucleotides. The use of the NNS or NNK codons results in a lower frequency of stop codons as compared to the NNN codon. Alternately, randomization can be achieved by using a mixture of trimer phosphoramidites at selected positions, during oligonucleotide synthesis. For larger scaffolds, the linear DNA library can be constructed by overlap-extension PCR using synthetic oligonucleotides incorporating randomized regions. Figure 2.3B illustrates this procedure for the case where two synthetic oligonucleotides, U2 and U3, contain the randomized regions. A PCR reaction is carried out with U2 as the template and the forward and reverse primers P2f and P2r respectively, to obtain Cassette 1. Similarly, Cassette 2 is obtained by PCR with U3 as the template and primers P3f and P3r. Cassettes 1 and 2 are then combined using primerless PCR. Subsequently, the product of the primerless PCR is amplified with forward primer P4f and reverse primer P4r, to obtain the linear DNA library.

The yeast surface display library is generated using homologous recombination mediated plasmid gap repair [77]. The linear DNA library is transformed into yeast along with linearized pCTCON vector; a detailed protocol for transformation is described in section 2.3.2. The pCTCON vector is linearized by serial digestion with *NheI*, *BamHI* and *SalI* enzymes. The primers P1f and P2f are chosen such that the linear DNA library has a 30-50 bp homology at each end with the corresponding ends of the linearized pCTCON vector. Homologous recombination results in plasmid gap repair and generation of the yeast display library (Figure 2.3C). A very useful feature of this procedure is that one can bypass the

primerless PCR step, if so desired, and directly transform the yeast with the linearized pCTCON and two or more linear DNA fragments. As shown in Figure 2.3D, primers P2f, P2r, P3f, P3r, P4f and P4r can be designed to ensure homologous recombination. Using two linear DNA fragments as opposed to one fragment does not appreciably alter the yeast transformation efficiency or library diversity[77]. In our experience, libraries of $10^7$-$10^8$ mutants can be routinely created using two linear DNA fragments and linearized pCTCON plasmid.

The methods described here focus on the generation of combinatorial libraries of scaffold proteins using yeast surface display. However, these methods can be easily modified to generate yeast surface display libraries for affinity maturation of *de novo* binders previously isolated. DNA from a single binding protein or a pool of *de novo* binders can be used as a starting point for affinity maturation. The linear DNA library can be generated using multiple strategies such as error prone PCR with nucleotide analogs [78-80], DNA shuffling or recombination [81, 82].

### 2.3.2. Yeast transformation of naive scaffold libraries

Linearized pCTCON vector along with one or more linear DNA fragments is transformed into yeast. Typically, multiple yeast transformations are required to get the desired library diversity. We routinely achieve a library diversity of $5x10^7$ with 60 µg of 300 bp linear DNA library (insert) and 20 µg of cut pCTCON vector. The following steps describe the protocol for transformation of linear DNA library (single fragment as in Figure 2.3A) into yeast EBY100.

1. Pick a single EBY colony from a freshly streaked YPD plate and inoculate a 5 ml YPD culture. Grow the culture overnight for 14-16 hrs at 30°C.

2. Inoculate a 50 ml YPD culture at a starting $OD_{600}$ of 0.1 from the overnight culture. It is important to check for any bacterial contamination at this point.

3. Grow the cells to an $OD_{600}$ of 1.3-1.5 (~ 6-7 hrs).

4. Meanwhile, concentrate the insert and cut plasmid DNA using Pellet Paint™. We routinely obtain a library size of ~$5x10^7$ with 20 transformations through electroporation; each transformation uses 3 µg of insert DNA (300 bp) and 1 µg digested vector DNA.

*Note:* The amount of insert DNA should be adjusted depending on the size of the insert. For transformations with two 300 bp insert fragments, we use 3 µg of each fragment and 1 µg digested vector DNA.

5. At an $OD_{600}$ = 1.3-1.5, add 500 µl of Tris-DTT buffer to the 50 ml YPD culture and incubate the cells for another 15 min, at 30°C and 250 rpm in a shaker incubator.

6. Harvest cells at 2500g for 3 min. Discard the supernatant and resuspend in 25 ml ice-cold E buffer and centrifuge again at 2500g for 3 min. Wash cells again in 1 ml ice-cold E buffer.

7. Resuspend the final pellet to a volume of 300 µl in E buffer. Six aliquots of 50 µl each can be prepared from a 50 ml culture. If more transformations are required, multiple 50 ml YPD cultures can be used.

8. Mix 3 µg insert DNA and 1 µg cut vector DNA with one 50 µl aliquot and keep on ice. Add this mixture to a pre-chilled 2 mm electroporation cuvette. Prepare a negative control sample with just the cut vector.

9. Electroporate the cuvettes at 540 volts, 25µF and 1000Ω; we use the BioRad Gene Pulser X-cell microbial system for electroporation. Immediately add 1 ml of warm (30 °C) YPD media. Recover remaining cells from each cuvette by washing with further 1 ml of warm YPD media and collect all samples (except the negative control) in a 15-ml Falcon tube. For good transformation efficiency, time constants should range from 15-20 ms.

10. Incubate cells for 1 hr at 30°C in a shaker incubator.

11. Harvest cells, discard supernatant, and resuspend in 10 ml SDCAA. Take a small aliquot (100 □L) and plate serial dilutions on SDCAA plates to determine transformation efficiency and library diversity. Typically, colonies can be observed on the plates after 36-48 hrs growth in a 30°C incubator. The negative control plate should have less than 1% of the colonies on the corresponding plate for the library.

12. Grow cells for 24-48 hrs in 250 ml SDCAA media supplemented with 1:100 dilution of pen-strep.

The cells should be passaged in SDCAA at least once before inducing expression for screening. To passage cells, inoculate fresh SDCAA media with yeast cells corresponding to a 10-20 fold excess over the calculated library diversity. Aliquots of the yeast library should also be frozen for long term storage, as follows.

***Freezing of yeast libraries.*** Yeast libraries can be frozen for long term storage either at -80°C or in liquid nitrogen. The following steps describe the protocol for freezing aliquots of the yeast library. Each aliquot should sufficient oversample the library diversity; this is to ensure that upon thawing, the yeast library retains its original diversity.

1. Inoculate freshly grown yeast cells corresponding to a 10-20 fold excess over the estimated library diversity in low dextrose SDCAA media. Incubate the cells at 30°C, 250 rpm for approximately 3 days

*Note*: The volume of SDCAA media used here depends on the library diversity. Typically cells grow up to an $OD_{600}$ of 8.0. To ensure having sufficient cells for freezing, large volume cultures should be grown.

2. After the cells have grown make sure that there is no bacterial contamination and harvest the cells at 2500 g for 5 min at 4°C, discard supernatant.

3. Resuspend pellet in freezing solution to a cell concentration such that every vial has 20-100 fold excess of library diversity. For example- a $10^8$ diversity library should be stored as at least $2x10^9$-$10^{10}$ cells per cryogenic vial.

4. Yeast cell vials should be slowly frozen first in an isopropanol bath and can then be transferred to long term storage such as in a -80°C freezer or in liquid nitrogen.

For revival of the library, the vial can be thawed at either 30°C or room temperature. Thawed cells can be grown in SDCAA media and passaged at least once before further analysis.

### 2.3.3. Magnetic screening

Binding proteins are isolated from the yeast display library using a two-step procedure involving magnetic screening followed by FACS-based screening. In the first step, the library of yeast cells expressing mutant proteins is incubated with target-coated micron-sized magnetic beads. The yeast-bead complexes are subsequently isolated using a magnet. A schematic of this process is shown in Figure 2.4. A key feature of magnetic screening is that binding proteins with even low affinity for the target are isolated, due to the highly avid yeast-bead interaction. The following steps describe the protocol for magnetic screening.

1. Inoculate yeast cells from the library, freshly grown in SDCAA, into SGCAA media at an $OD_{600}$ of 1 (i.e. $10^7$cells/mL) to induce expression of cell surface fusions. The number of yeast cells used for inoculation should be at least 10-20 times the estimated library

diversity. The SGCAA culture is maintained at 20°C and 250 rpm in a shaker. Cells are ready for screening after 20-24 hours of induction in SGCAA.

*Note:* The yeast library may be induced at a different temperature. Usually 20-24 hours of induction in SGCAA result in satisfactory expression of cell surface fusions. However, the time of induction may need to be optimized in certain cases.

2. Wash the biotin binder beads twice with PBS-BSA using the Dyna-Mag2 magnet as per the manufacturer's protocol and incubate 2.5 x $10^6$ beads with 0.8-2.5 picomoles of biotinylated target in 100 µL of PBS-BSA, overnight at 4°C in a rotator.

3. Wash the target preloaded beads twice with PBS-BSA, using the Dyna-Mag2 magnet to remove excess antigen.

4. Pellet the SGCAA induced library by centrifugation at 2500 g for 3 min. Wash the cells twice with PBS-BSA and resuspend in 1 ml PBS-BSA.

5. For negative selection against beads, incubate the cells with the washed beads for 1 hr at 4°C in a micro-centrifuge tube, in a rotator. Apply the micro-centrifuge tube to the Dyna-Mag2 magnet and collect the unbound yeast cells.

6. Repeat the negative selection as in step 5 with the unbound yeast cells.

7. Apply the unbound cells again to the magnet and remove any residual cells that are bound to beads.

8. Incubate the unbound cells from step 7 with antigen preloaded beads from step 2 for 1 hr at 4°C in a rotor in a 2ml micro-centrifuge tube.

*Note:* This incubation step can be carried out at a different temperature, if desired.

9. Isolate the beads (and hence the yeast-bead complexes) using the magnet and discard the unbound yeast cells. Wash the beads at least 3 times with 1 ml PBS-BSA at room temperature for 5-10 min.

10. The beads (including yeast-bead complexes) are transferred to a culture tube containing 5 ml SDCAA with pen-strep (1:100 dilution) and incubated for 36-48 hrs in a shaker at 30°C, 250 rpm. The beads are then removed using the magnet and the yeast cells are used to further inoculate a 50 ml SDCAA culture.

***Additional comments.*** Additional negative selection steps can be carried out after step 7. Typically, we use a negative selection step to eliminate yeast cells that bind to the secondary antibodies. Steps 5-7 are carried out using beads preloaded with the secondary antibody.

### 2.3.4. Screening using FACS

The yeast cells obtained after magnetic screening are further screened using flow cytometry to isolate the pool of binders with the highest affinity for the target. In this step, cells are simultaneously labeled with biotinylated target and an antibody against the c-myc or HA epitope tags. Subsequently, immunofluorescent detection is achieved by fluorescent labeling with strep-PE and a secondary antibody conjugated to a suitable fluorophore such as

Alexa Fluor 488. The Alexa Fluor 488 fluorescence intensity corresponds to the total number of protein fusions on the cell surface while the PE fluorescence intensity corresponds to the number of fusions that are bound to the target.

Figure 2.5 shows a representative plot for such a labeling scheme. Each dot on the plot represents the PE and Alexa-Fluor 488 fluorescence values for a single cell. At a given target concentration, cells that have a higher fraction of cell surface fusions bound to the target are those that express mutants with higher binding affinity. The top 0.1-1% of cell is collected; this corresponds to the polygonal region in Figure 2.5. The polygonal region sort gate is drawn to capture cells along the diagonal (i.e. same ratio of binding to expression) allowing normalization of binding affinity to cell surface expression. Initial rounds of screening are usually less stringent and up to 1% of the population is captured to enrich all the binding proteins (low and high affinity). Subsequently, gates can be more stringent to capture fewer cells (~0.1-0.2%), hence selecting for higher affinity. The collected cells are expanded in culture and further rounds of FACS screening are carried out. The concentration of target labeling can be reduced in progressive rounds to increase the stringency of the screen. In general, the target concentration used should be high enough so as to obtain a fluorescent signal over background and yet not so high as to cause saturation. In this regard, Boder and Wittrup have discussed the optimal target concentrations for efficient discrimination between clones of different binding affinities [83].Typically, after 3-5 rounds of FACS screening, a pool of mutants with the highest binding affinity for the target can be isolated. Here we present a detailed protocol for FACS screening of yeast display libraries.

1. Inoculate SGCAA medium with yeast cells freshly grown in SDCAA at a starting $OD_{600}$ of ~0.5 to 1. The number of yeast cells used in the inoculum should be 10-20 times the calculated library diversity. Note that the number of cells collected during a round of FACS determines the maximum possible diversity for the following round. Incubate the SGCAA culture at 20°C, 250 rpm for 20-24 h in a shaker incubator.

2. Pellet $10^7$ cells or 10-fold oversampling of the library diversity, whichever is greater, in micro-centrifuge tubes in a tabletop centrifuge at 12000g for 30 sec. Wash the pellet with 1 ml of PBS-BSA. Typically, the pellets obtained after the wash step are used to prepare six control samples in addition to the test sample for sorting the library. The controls contain $10^7$cells; the sort sample contains $10^7$ cells or more, as required to oversample the library diversity. The following samples are prepared:

    a) *Unlabeled control:* yeast cells + PBS-BSA
    b) *Secondary reagent control-1:* Goat anti-chicken-IgG-Alexa Fluor 488/633 conjugate only
    c) *Secondary reagent control-2:* Strep-PE/Neutravidin-FITC only
    d) *Expression control:* Primary labeling with chicken anti-c-myc, secondary with goat anti-chicken-IgG-Alexa Fluor 488/633 conjugate
    e) *Binding control:* Target + Strep-PE/Neutravidin-FITC only

f) *No Target control:* Primary labeling with chicken anti-c-myc, secondary labeling with goat anti-chicken- IgG-Alexa Fluor488/ 633 conjugate and Strep-PE/Neutravidin-FITC

g) *Sort sample:* Primary labeling with chicken anti-c-myc and biotinylated antigen, secondary labeling with goat anti-chicken-IgG-Alexa-Fluor-488/633 conjugate and Strep-PE/Neutravidin-FITC

*Note:* The secondary reagent controls (a and b) test if the mutant proteins bind non-specifically to the secondary reagents. The expression control (d) is used to assess the fraction of the library that expresses cell surface fusions. Comparison of the Alexa Fluor 488/633 fluorescent signal corresponding to c-myc antibody binding in the no target control (f) and the sort sample (g) can be used to diagnose any steric effects that are associated with simultaneous detection of the target and the c-myc epitope tag. For instance, lower Alexa Fluor 488/633 fluorescence in the presence of the target may indicate a steric effect. A similar comparison can be made between the PE fluorescence in the binding control (e) and the sort sample (g). Lower PE fluorescence in the sort sample indicates that the binding of the c-myc antibody interferes with target binding. If such a steric effect is observed, an anti-HA antibody may be used instead of the anti-c-myc antibody.

3. *Primary labeling:* Incubate samples d), f) and g) with a 1:250 dilution of chicken anti-c-myc antibody in 100 µl PBS-BSA at room temperature. Alternatively, d), f) and g) can be labeled with a mouse anti-HA antibody (1:100 dilution). For samples e) and g) incubate

with an appropriate concentration of the biotinylated target; sample g) contains both target and the c-myc/HA antibody. Ensure that the target is in ten-fold molar excess over the number of yeast displayed mutant proteins in solution (see additional notes for explanation). The incubation time for samples e) and g) should be such that the binding reaction approaches equilibrium (see additional notes for calculation of incubation time). All other samples are incubated for the same time for convenience. During this primary labeling step, samples a), b) and c) may be incubated in PBS-BSA.

4. Pellet cells at 12,000g for 30 s at 4°C and wash with 1 mL ice-cold PBS-BSA. All subsequent steps for yeast labeling should use ice-cold PBS-BSA. Tubes should be placed on ice during all incubation steps and use of a refrigerated centrifuge is recommended to avoid isolation of binding proteins that bind the secondary reagent.

5. *Secondary labeling*: Incubate yeast cells with appropriate secondary reagents as described in step 1, goat anti-chicken IgG-Alexa Fluor 488/633 (1:250) and Strep-PE (1:250), in 100 µl for $10^7$ cells. Incubate cells on ice, protected from light for 15 min.

   *Note:* Alternate use of secondary reagents i.e. strep-PE and neutravidin-FITC in successive rounds of FACS screening is recommended. Note that goat anti-chicken IgG-Alexa Fluor 633 needs to be used in conjunction with neutravidin-FITC

6. Pellet cells for 12,000g for 30 s at 4°C and wash with 1 mL PBS-BSA. Keep cell pellets on ice.

7. Resuspend cells in 1 ml PBS-BSA just prior to flow cytometric analysis. For the sort sample, collect cells based on considerations discussed earlier.

***Additional notes.*** Consider the single-step binding interaction between the mutant protein and the target

$$P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C \qquad K_D = \frac{k_{off}}{k_{on}}$$

Here *P* is the cell surface displayed protein, *L* is the target and *C* is the protein-target complex.

$k_{on}$ and $k_{off}$ are the association and dissociation rate constants respectively. For this system, the fraction of cell surface fusions that are bound to the target (*y*), at equilibrium is given by the following relationship

$$y = \frac{[L]_0 - \frac{nyR_0}{N_{Av}}}{K_D + \left([L]_0 - \frac{nyR_0}{N_{Av}}\right)} \qquad 0 \le y \le 1 \quad \textbf{(Eq. 1)}$$

Here, $[L]_0$ is the initial target concentration (in mol/L), *n* is the yeast cell density in solution (#cells/L), $R_0$ is the number of cell surface fusions displayed per cell and $N_{av}$ is the Avogadro number ($6.023 \times 10^{23}$/mol).

**Eq. 1** simplifies to

$$y = \frac{[L]_0}{K_D + [L]_0} \quad \textbf{(Eq. 2)}$$

when $[L]_0 >> \frac{nyR_0}{N_{Av}}$ **(Eq. 3)**

Ensuring **Eq. 3** is satisfied allows the target concentration at equilibrium to be approximated as the known initial target concentration. This also enables one to easily estimate the time required for the system to approach equilibrium (see discussion below). The target concentration satisfies **Eq. 3** when the target is in a ten-fold or more molar excess over the total number of cell surface displayed proteins in solution. This happens when *N* yeast cells are incubated in a solution volume (*V*) such that

$$V \geq \frac{10NR_0}{N_{Av}[L]_0} \quad \textbf{(Eq. 4)}$$

Thus, for a fixed number of cells, the volume of incubation can be tailored to ensure molar excess of the target. For a conservative estimate, $R_0$ can be assumed to be $10^5$.

Assuming that **Eq. 3** is satisfied, the approach to equilibrium can be calculated using the following relationship

$$[C] = [C]_{eq}(1 - e^{-k_{obs}t}) \quad \text{where} \quad k_{obs} = k_{on}[L]_0 + k_{off} \quad \textbf{(Eq. 5)}$$

Here, *[C]* and $[C]_{eq}$ are the concentrations of the target-protein complex at time *t* and equilibrium respectively. The incubation time needed to reach 95% of equilibrium ($t_{95\%}$) can be estimated using the following relationship.

$$t_{95\%} = \frac{3}{k_{obs}} \textbf{(Eq. 6)}$$

Here $k_{obs}$ can be estimated using approximate values of $k_{on}$ and $k_{off}$ as follows. For a typical protein-protein interaction, $k_{on}$ is estimated as $10^5$ $M^{-1}$ $s^{-1}$. $k_{off}$ can be calculated from $k_{on}$ and an approximate estimate of the expected $K_D$ for the best binder.

### 2.3.5. Clone Analysis and biophysical characterization using yeast surface titrations

After 3-5 rounds of screening using flow cytometry, a pool of mutants with the highest affinity for the target is isolated. Yeast cells from this pool are plated on SDCAA plates and incubated at 30°C for 36-48 hrs. 10-25 randomly picked colonies are used to inoculate 5 mL SDCAA cultures; the cultures are incubated for 24-36 hrs at 30°C, 250 rpm in a shaker incubator. Subsequently, plasmid DNA is isolated from the SDCAA cultures using the Zymoprep kit, following the manufacturer's protocol. The isolated DNA is further

transformed into NovaBlue$^{TM}$ *E. coli* cells. Plasmid DNA is extracted from *E. coli* cells and used for DNA sequencing to identify the selected yeast clones.

Binding affinity of unique clones is determined using yeast cell surface titrations. Briefly, yeast cells expressing the mutant protein are labeled with varying concentrations of the target. The fraction of cell surface mutants bound to the target is determined using flow cytometry. Subsequently the $K_D$ is determined by fitting the data to a single-step binding isotherm. A detailed protocol for measuring the $K_D$ of the target-protein interaction and estimating the confidence intervals for this measurement is presented here.

1. Inoculate 5 mL of SGCAA medium with yeast cells from a freshly grown SDCAA culture of the yeast clone, at a starting $OD_{600}$ of ~0.5 to 1. Incubate the SGCAA culture at 20°C, 250 rpm for 20-24 h in a shaker incubator to induce expression of cell surface fusions.

2. Pellet $2x10^6$ cells per sample and wash the pellet with 1 ml PBS-BSA. We prepare at least 10-12 samples with varying concentrations of the target, ideally spanning two orders of magnitude both below and above the expected $K_D$. The titration experiments are usually performed in triplicate. Therefore, after the first titration the concentrations can be adjusted to span the range if required.

3. *Primary labeling*: Cells are incubated with varying concentrations of the target in PBS-BSA, typically in micro-centrifuge tubes. Ensure that the target is in at least a 10-fold molar excess over the total yeast displayed mutant proteins in solution. This can be

achieved by choosing a suitable incubation volume, as discussed in the previous section. The incubation time is chosen such that the binding reaction in lowest concentration sample approaches close to equilibrium; this is also described in the previous section. Samples are incubated at the desired temperature at which the $K_D$ is to be determined.

4. Wash the cells with 1 ml of ice cold PBS-BSA.

5. *Secondary labeling:* Incubate the samples with strep-PE (1:250 dilution) in 100 µl of PBS-BSA, on ice and shielded from light.

6. Pellet cells for 12,000g for 30 s at 4°C and wash with 1 mL PBS-BSA. Keep cell pellets on ice

7. Resuspend the pellet in 1 ml PBS-BSA and analyze the sample on a flow cytometer. The mean-PE fluorescence for all samples is recorded.

The background-subtracted mean PE fluorescence intensity of the population ($F$) is directly proportional to the average fraction of the yeast cell surface fusions bound to the target ($y$), at a particular initial target concentration ($[L]_0$). The relationship between these quantities is given by

$$F = \frac{F_{\max}[L]_0}{K_D + [L]_0} \quad \textbf{(Eq. 7)}$$

Here, $F_{max}$ is the background subtracted fluorescence intensity when $y$=1. Experimental measurements using flow cytometry provide values of $F$ for varying $[L]_0$; three different sets of data are obtained. Subsequently, the data are fit to **Eq. 7** using global non-linear least squares regression across all three data sets. A single $K_D$ value and $F_{max}$ corresponding to each titration are the fitted parameters.

***Evaluating confidence intervals for estimated $K_D$.*** The global non-linear least squares regression gives the most likely estimate of $K_D$. Here we describe a procedure to estimate the upper and lower bounds of the 68% confidence interval for the $K_D$ estimate [84]. The 68% confidence interval is analogous to the standard deviation of triplicate measurements that are commonly reported as error bars; in this case, the mean of triplicate measurements is the most likely estimate. The following steps can be followed to estimate the 68% confidence interval.

1. Identify the degrees of freedom (*f*) in the data. The degrees of freedom (*f*) can be calculated by subtracting the total number of parameters (*p*) from the number of data points. For example- if three thirteen point titrations are performed, there are a total of 13x3=39 data points and four parameters- $F_{max1}$, $F_{max2}$, $F_{max3}$ (one $F_{max}$ for each titration) and $K_D$.

2. Calculate the F-statistic, F (P, *p*, *f*), for *p* parameters, *f* degrees of freedom and probability P. When the value of P is 0.32 or less, there is a 68% probability that the $K_D$ estimate is consistent with the data. The F-statistic can be calculated using any statistical software program. In Microsoft Excel the F-statistic can be calculated using the FINV function – FINV(P, *p*, *f*)

3. For the most likely (global least squares) estimates of $K_D$, $F_{max1}$, $F_{max2}$ and $F_{max3}$, calculate the sum of squares of the differences between the observed background-subtracted mean PE-fluorescence intensity ($F_{obs}$) and the corresponding model-predicted value (from **Eq. 7**) ($F_{calc}$). This is normalized by $F_{obs}$ to give the quantity $E^2$.

$$E^2 = E^2_{\min} = \sum \frac{(F_{obs} - F_{calc})^2}{F_{obs}} \quad \textbf{(Eq. 8)}$$

4. Next, calculate $E^2$ for different values of $K_D$, greater than and lesser than the least squares estimate. In order to do this, first a particular value of $K_D$ is fixed as a parameter and the data sets are subjected to a global least squares minimization by varying $F_{max1}$, $F_{max2}$ *and* $F_{max3}$. Then $E^2$ is calculated as

$$E^2_{KD} = \sum \frac{(F_{obs} - F'_{calc})^2}{F_{obs}} \quad \textbf{(Eq. 9)}$$

Here, $F'_{calc}$ is the model-predicted fluorescence intensity (**Eq. 7**), where $K_D$ is the fixed value chosen and $F_{max1}$, $F_{max2}$ and $F_{max3}$ are the least squares estimates of these parameters, with $K_D$ held fixed. For values of $K_D$ greater than or lesser than the most likely estimate of $K_D$, $E^2_{KD}$ is greater than $E^2_{min}$. The upper and lower bounds of the 68% confidence interval for the $K_D$ estimate are the values of $K_D$ for which the following equation is satisfied.

$$\frac{E^2_{KD}(f-p)}{E^2_{\min}(f-p-1)} = 1 + \frac{p}{f}\mathrm{F}(0.32, p, f) \quad \textbf{(Eq. 10)}.$$

## 2.4. Conclusions

Over the years, yeast surface display has emerged as a powerful display platform for protein engineering. The distinct advantages of yeast surface display include: eukaryotic post-translational machinery that makes it well suited for display of a wide spectrum of complex proteins, the ability to use quantitative flow cytometric sorting as well screening schemes that exploit the avidity effect to isolate low affinity binders and carry out stringent negative selections, and the ability to efficiently characterize binding proteins as cell surface fusions, while bypassing sub-cloning steps and recombinant protein expression. Yeast surface display has been used in a variety of applications such as the isolation of *de novo* binders from combinatorial libraries of a wide spectrum of scaffold proteins, affinity maturation, engineering binding specificity and protein stability, epitope mapping and protein-protein interaction studies. Combining yeast surface display with other combinatorial

screening platforms can harness the unique advantages of each technique and enable previously unexplored applications.

## 2.5. References


1. Finlay, W.J., L. Bloom, and O. Cunningham, *Phage display: a powerful technology for the generation of high specificity affinity reagents from alternative immune sources.* Methods Mol Biol, 2011. **681**: p. 87-101.

2. Daugherty, P.S., et al., *Antibody affinity maturation using bacterial surface display.* Protein Eng, 1998. **11**(9): p. 825-32.

3. Daugherty, P.S., et al., *Development of an optimized expression system for the screening of antibody libraries displayed on the Escherichia coli surface.* Protein Eng, 1999. **12**(7): p. 613-21.

4. Daugherty, P.S., *Protein engineering with bacterial display.* Curr Opin Struct Biol, 2007. **17**(4): p. 474-80.

5. Boder, E.T. and K.D. Wittrup, *Yeast surface display for screening combinatorial polypeptide libraries.* Nat Biotechnol, 1997. **15**(6): p. 553-7.

6. Gai, S.A. and K.D. Wittrup, *Yeast surface display for protein engineering and characterization.* Curr Opin Struct Biol, 2007. **17**(4): p. 467-73.

7. Kondo, A. and M. Ueda, *Yeast cell-surface display--applications of molecular display.* Appl Microbiol Biotechnol, 2004. **64**(1): p. 28-40.

8. Pepper, L.R., et al., *A decade of yeast surface display technology: where are we now?* Comb Chem High Throughput Screen, 2008. **11**(2): p. 127-34.

9. Zhou, C., et al., *Development of a novel mammalian cell surface antibody display platform.* MAbs, 2010. **2**(5): p. 508-18.

10. Akamatsu, Y., et al., *Whole IgG surface display on mammalian cells: Application to isolation of neutralizing chicken monoclonal anti-IL-12 antibodies.* J Immunol Methods, 2007. **327**(1-2): p. 40-52.

11. Lipovsek, D. and A. Pluckthun, *In-vitro protein evolution by ribosome display and mRNA display.* J Immunol Methods, 2004. **290**(1-2): p. 51-67.

12. Hanes, J., L. Jermutus, and A. Pluckthun, *Selecting and evolving functional proteins in vitro by ribosome display.* Methods Enzymol, 2000. **328**: p. 404-30.

13. Hanes, J., et al., *Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries.* Proc Natl Acad Sci U S A, 1998. **95**(24): p. 14130-5.

14. Hanes, J. and A. Pluckthun, *In vitro selection and evolution of functional proteins by using ribosome display.* Proc Natl Acad Sci U S A, 1997. **94**(10): p. 4937-42.

15. Liu, R., et al., *Optimized synthesis of RNA-protein fusions for in vitro protein selection.* Methods Enzymol, 2000. **318**: p. 268-93.

16. Roberts, R.W. and J.W. Szostak, *RNA-peptide fusions for the in vitro selection of peptides and proteins.* Proc Natl Acad Sci U S A, 1997. **94**(23): p. 12297-302.

17. Shen, X., et al., *Scanning the human proteome for calmodulin-binding proteins.* Proc Natl Acad Sci U S A, 2005. **102**(17): p. 5969-74.

18. Wilson, D.S., A.D. Keefe, and J.W. Szostak, *The use of mRNA display to select high-affinity protein-binding peptides.* Proc Natl Acad Sci U S A, 2001. **98**(7): p. 3750-5.

19. Chao, G., et al., *Isolating and engineering human antibodies using yeast surface display.* Nat Protoc, 2006. **1**(2): p. 755-68.

20. Wang, Z., et al., *A new yeast display vector permitting free scFv amino termini can augment ligand binding affinities.* Protein Eng Des Sel, 2005. **18**(7): p. 337-43.

21. Kieke, M.C., et al., *Isolation of anti-T cell receptor scFv mutants by yeast surface display.* Protein Eng, 1997. **10**(11): p. 1303-10.

22. Feldhaus, M.J., et al., *Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library.* Nat Biotechnol, 2003. **21**(2): p. 163-70.

23. Boder, E.T., K.S. Midelfort, and K.D. Wittrup, *Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity.* Proc Natl Acad Sci U S A, 2000. **97**(20): p. 10701-5.

24. Graff, C.P., et al., *Directed evolution of an anti-carcinoembryonic antigen scFv with a 4-day monovalent dissociation half-time at 37 degrees C.* Protein Eng Des Sel, 2004. **17**(4): p. 293-304.

25. Holler, P.D., et al., *In vitro evolution of a T cell receptor with high affinity for peptide/MHC.* Proc Natl Acad Sci U S A, 2000. **97**(10): p. 5387-92.

26. Wozniak-Knopp, G., et al., *Introducing antigen-binding sites in structural loops of immunoglobulin constant domains: Fc fragments with engineered HER2/neu-binding sites and antibody properties.* Protein Eng Des Sel, 2010. **23**(4): p. 289-97.

27. Binz, H.K., P. Amstutz, and A. Pluckthun, *Engineering novel binding proteins from nonimmunoglobulin domains.* Nat Biotechnol, 2005. **23**(10): p. 1257-68.

28. Lipovsek, D., et al., *Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: molecular convergence with single-domain camelid and shark antibodies.* J Mol Biol, 2007. **368**(4): p. 1024-41.

29. Hackel, B.J., A. Kapila, and K.D. Wittrup, *Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling.* J Mol Biol, 2008. **381**(5): p. 1238-52.

30. Pavoor, T.V., Y.K. Cho, and E.V. Shusta, *Development of GFP-based biosensors possessing the binding properties of antibodies.* Proc Natl Acad Sci U S A, 2009. **106**(29): p. 11895-900.

31. Huang, D. and E.V. Shusta, *Secretion and surface display of green fluorescent protein using the yeast Saccharomyces cerevisiae.* Biotechnol Prog, 2005. **21**(2): p. 349-57.

32. Silverman, A.P., M.S. Kariolis, and J.R. Cochran, *Cystine-knot peptides engineered with specificities for alpha(IIb)beta(3) or alpha(IIb)beta(3) and alpha(v)beta(3) integrins are potent inhibitors of platelet aggregation.* J Mol Recognit, 2011. **24**(1): p. 127-35.

33. Silverman, A.P., et al., *Engineered cystine-knot peptides that bind alpha(v)beta(3) integrin with antibody-like affinities.* J Mol Biol, 2009. **385**(4): p. 1064-75.

34. Kimura, R.H., et al., *Engineered cystine knot peptides that bind alphavbeta3, alphavbeta5, and alpha5beta1 integrins with low-nanomolar affinity.* Proteins, 2009. **77**(2): p. 359-69.

35. Kimura, R.H., et al., *Functional mutation of multiple solvent-exposed loops in the Ecballium elaterium trypsin inhibitor-II cystine knot miniprotein.* PLoS One, 2011. **6**(2): p. e16112.

36. Gera, N., et al., *Highly stable binding proteins derived from the hyperthermophilic Sso7d scaffold.* J Mol Biol, 2011. **409**(4): p. 601-16.

37. Lee, C.H., et al., *Engineering of a human kringle domain into agonistic and antagonistic binding proteins functioning in vitro and in vivo.* Proc Natl Acad Sci U S A, 2010. **107**(21): p. 9567-71.

38. Shpilman, M., et al., *Development and characterization of high affinity leptins and leptin antagonists.* J Biol Chem, 2011. **286**(6): p. 4429-42.

39. Rao, B.M., et al., *Interleukin 2 (IL-2) variants engineered for increased IL-2 receptor alpha-subunit affinity exhibit increased potency arising from a cell surface ligand reservoir effect.* Mol Pharmacol, 2004. **66**(4): p. 864-9.

40. Rao, B.M., et al., *High-affinity CD25-binding IL-2 mutants potently stimulate persistent T cell growth.* Biochemistry, 2005. **44**(31): p. 10696-701.

41. Cochran, J.R., et al., *Improved mutants from directed evolution are biased to orthologous substitutions.* Protein Engineering Design & Selection, 2006. **19**(6): p. 245-53.

42. Garcia-Rodriguez, C., et al., *Molecular evolution of antibody cross-reactivity for two subtypes of type A botulinum neurotoxin.* Nat Biotechnol, 2007. **25**(1): p. 107-16.

43. Weaver-Feldhaus, J.M., et al., *Directed evolution for the development of conformation-specific affinity reagents using yeast display.* Protein Eng Des Sel, 2005. **18**(11): p. 527-36.

44. VanAntwerp, J.J. and K.D. Wittrup, *Fine affinity discrimination by yeast surface display and flow cytometry.* Biotechnol Prog, 2000. **16**(1): p. 31-7.

45. Orr, B.A., et al., *Rapid method for measuring ScFv thermal stability by yeast surface display.* Biotechnol Prog, 2003. **19**(2): p. 631-8.

46. Cochran, J.R., et al., *Domain-level antibody epitope mapping through yeast surface display of epidermal growth factor receptor fragments.* J Immunol Methods, 2004. **287**(1-2): p. 147-58.

47. Chao, G., J.R. Cochran, and K.D. Wittrup, *Fine epitope mapping of anti-epidermal growth factor receptor antibodies through random mutagenesis and yeast surface display.* J Mol Biol, 2004. **342**(2): p. 539-50.

48. Levy, R., et al., *Fine and domain-level epitope mapping of botulinum neurotoxin type A neutralizing antibodies by yeast surface display.* J Mol Biol, 2007. **365**(1): p. 196-210.

49. Chung, K.M., et al., *Antibodies against West Nile Virus nonstructural protein NS1 prevent lethal infection through Fc gamma receptor-dependent and -independent mechanisms.* J Virol, 2006. **80**(3): p. 1340-51.

50. Shusta, E.V., et al., *Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency.* J Mol Biol, 1999. **292**(5): p. 949-56.

51. Kieke, M.C., et al., *Selection of functional T cell receptor mutants from a yeast surface-display library.* Proc Natl Acad Sci U S A, 1999. **96**(10): p. 5651-6.

52. Shusta, E.V., et al., *Directed evolution of a stable scaffold for T-cell receptor engineering.* Nat Biotechnol, 2000. **18**(7): p. 754-9.

53. Park, S., et al., *Limitations of yeast surface display in engineering proteins of high thermostability.* Protein Eng Des Sel, 2006. **19**(5): p. 211-7.

54. Tereshko, V., et al., *Toward chaperone-assisted crystallography: protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold.* Protein Sci, 2008. **17**(7): p. 1175-87.

55. Binz, H.K., et al., *High-affinity binders selected from designed ankyrin repeat protein libraries.* Nat Biotechnol, 2004. **22**(5): p. 575-82.

56. Bandeiras, T.M., et al., *Structure of wild-type Plk-1 kinase domain in complex with a selective DARPin.* Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 4): p. 339-53.

57. Schweizer, A., et al., *Inhibition of caspase-2 by a designed ankyrin repeat protein: specificity, structure, and inhibition mechanism.* Structure, 2007. **15**(5): p. 625-36.

58. Sennhauser, G., et al., *Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors.* PLoS Biol, 2007. **5**(1): p. e7.

59. Sennhauser, G. and M.G. Grutter, *Chaperone-assisted crystallography with DARPins.* Structure, 2008. **16**(10): p. 1443-53.

60. Bidlingmaier, S. and B. Liu, *Construction and application of a yeast surface-displayed human cDNA library to identify post-translational modification-dependent protein-protein interactions.* Mol Cell Proteomics, 2006. **5**(3): p. 533-40.

61. Bidlingmaier, S. and B. Liu, *Interrogating yeast surface-displayed human proteome to identify small molecule-binding proteins.* Mol Cell Proteomics, 2007. **6**(11): p. 2012-20.

62. Wadle, A., et al., *Serological identification of breast cancer-related antigens from a Saccharomyces cerevisiae surface display library.* Int J Cancer, 2005. **117**(1): p. 104-13.

63. Lipovsek, D., et al., *Selection of horseradish peroxidase variants with enhanced enantioselectivity by yeast surface display.* Chem Biol, 2007. **14**(10): p. 1176-85.

64. Shiraga, S., et al., *Creation of Rhizopus oryzae lipase having a unique oxyanion hole by combinatorial mutagenesis in the lid domain.* Appl Microbiol Biotechnol, 2005. **68**(6): p. 779-85.

65. Shiraga, S., et al., *Enhanced reactivity of Rhizopus oryzae lipase displayed on yeast cell surfaces in organic solvents: potential as a whole-cell biocatalyst in organic solvents.* Appl Environ Microbiol, 2005. **71**(8): p. 4335-8.

66. S.J. Park , J.R.C., ed. *Protein Engineering and Design*. 2010, CRC Press, Florida.

67. Roberts, I.S., *The biochemistry and genetics of capsular polysaccharide production in bacteria.* Annu Rev Microbiol, 1996. **50**: p. 285-315.

68. Wildt, S. and T.U. Gerngross, *The humanization of N-glycosylation pathways in yeast.* Nat Rev Microbiol, 2005. **3**(2): p. 119-28.

69. Ackerman, M., et al., *Highly avid magnetic bead capture: an efficient selection method for de novo protein engineering utilizing yeast surface display.* Biotechnol Prog, 2009. **25**(3): p. 774-83.

70. Bowley, D.R., et al., *Antigen selection from an HIV-1 immune antibody library displayed on yeast yields many novel antibodies compared to selection from the same library displayed on phage.* Protein Eng Des Sel, 2007. **20**(2): p. 81-90.

71. Benatuil, L., et al., *An improved yeast transformation method for the generation of very large human antibody libraries.* Protein Eng Des Sel, 2010. **23**(4): p. 155-9.

72. Dufner, P., L. Jermutus, and R.R. Minter, *Harnessing phage and ribosome display for antibody optimisation.* Trends Biotechnol, 2006. **24**(11): p. 523-9.

73. Bowley, D.R., et al., *Libraries against libraries for combinatorial selection of replicating antigen-antibody pairs.* Proc Natl Acad Sci U S A, 2009. **106**(5): p. 1380-5.

74. Hu, X., et al., *Combinatorial libraries against libraries for selecting neoepitope activation-specific antibodies.* Proc Natl Acad Sci U S A, 2010. **107**(14): p. 6252-7.

75. Colby, D.W., et al., *Engineering antibody affinity by yeast surface display.* Methods in Enzymology, 2004. **388**: p. 348-58.

76. Orcutt, K.D. and K.D. Wittrup, *Yeast Display and Selections*, in *Antibody Engineering Vol. 1*, R. Kontermann and S. Dubel, Editors. 2010, Springer-Verlag: Berlin Heidelberg. p. 207-233.

77. Swers, J.S., B.A. Kellogg, and K.D. Wittrup, *Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display.* Nucleic Acids Res, 2004. **32**(3): p. e36.

78. Zaccolo, M. and E. Gherardi, *The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase.* J Mol Biol, 1999. **285**(2): p. 775-83.

79. Zaccolo, M., et al., *An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues.* J Mol Biol, 1996. **255**(4): p. 589-603.

80. Cadwell, R.C. and G.F. Joyce, *Mutagenic PCR.* PCR Methods Appl, 1994. **3**(6): p. S136-40.

81. Stemmer, W.P., *DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution.* Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10747-51.

82. Stemmer, W.P., *Rapid evolution of a protein in vitro by DNA shuffling.* Nature, 1994. **370**(6488): p. 389-91.

83. Boder, E.T. and K.D. Wittrup, *Optimal screening of surface-displayed polypeptide libraries.* Biotechnol Prog, 1998. **14**(1): p. 55-62.

84. Lakowicz, J.R., ed. *Principles of Fluorescence Spectroscopy*. Third Edition ed. 2006. 129-137.

**Figures**

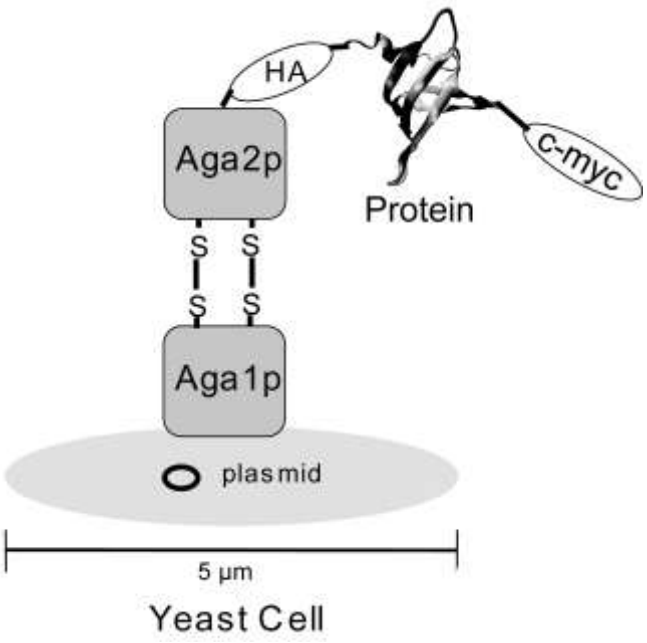



**Figure 2.1:** The yeast surface display platform. The mutant protein is displayed on the yeast cell surface as a fusion to the Aga2p cell wall protein. The HA and c-myc epitope tags can be used for immunofluorescent labeling and detection.

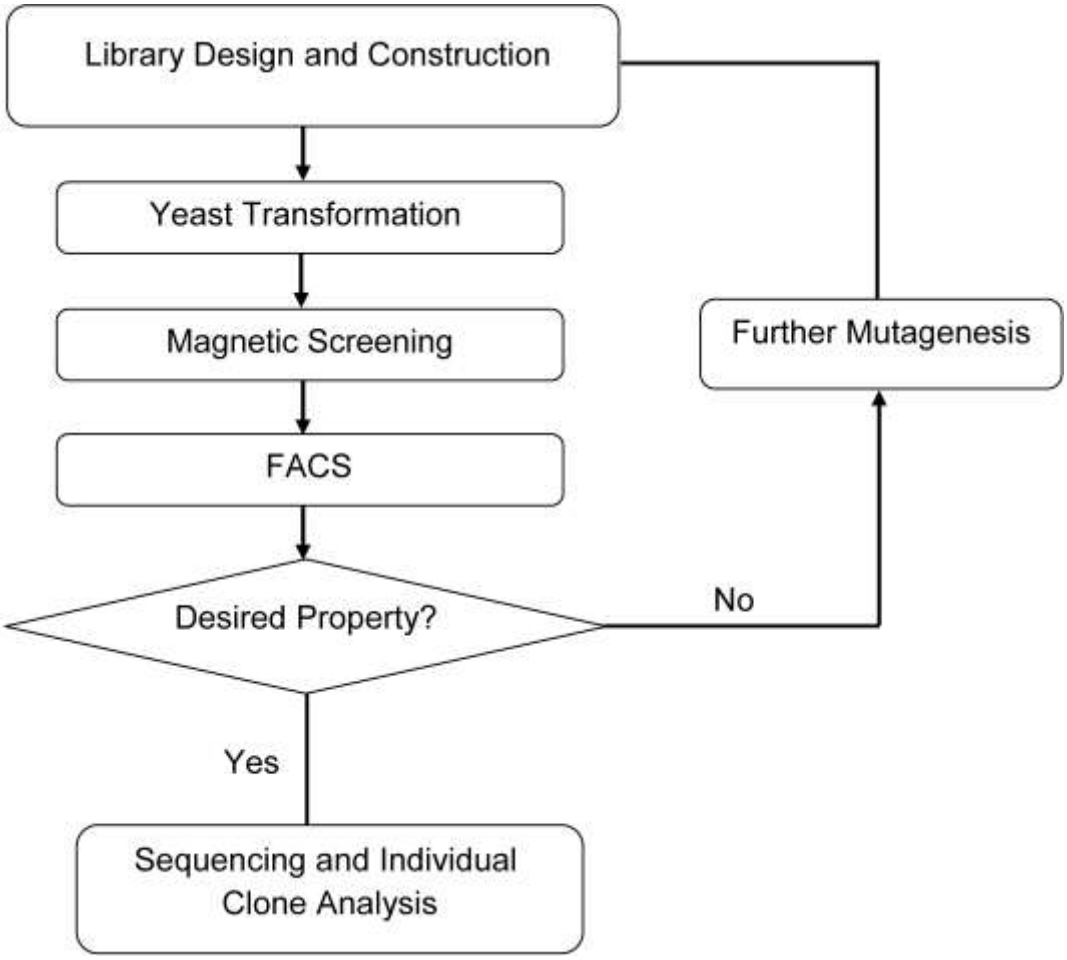



**Figure 2.2:** Isolation of binding proteins using yeast surface display. A flow chart for isolation of binding proteins from a combinatorial library of a scaffold protein is shown.

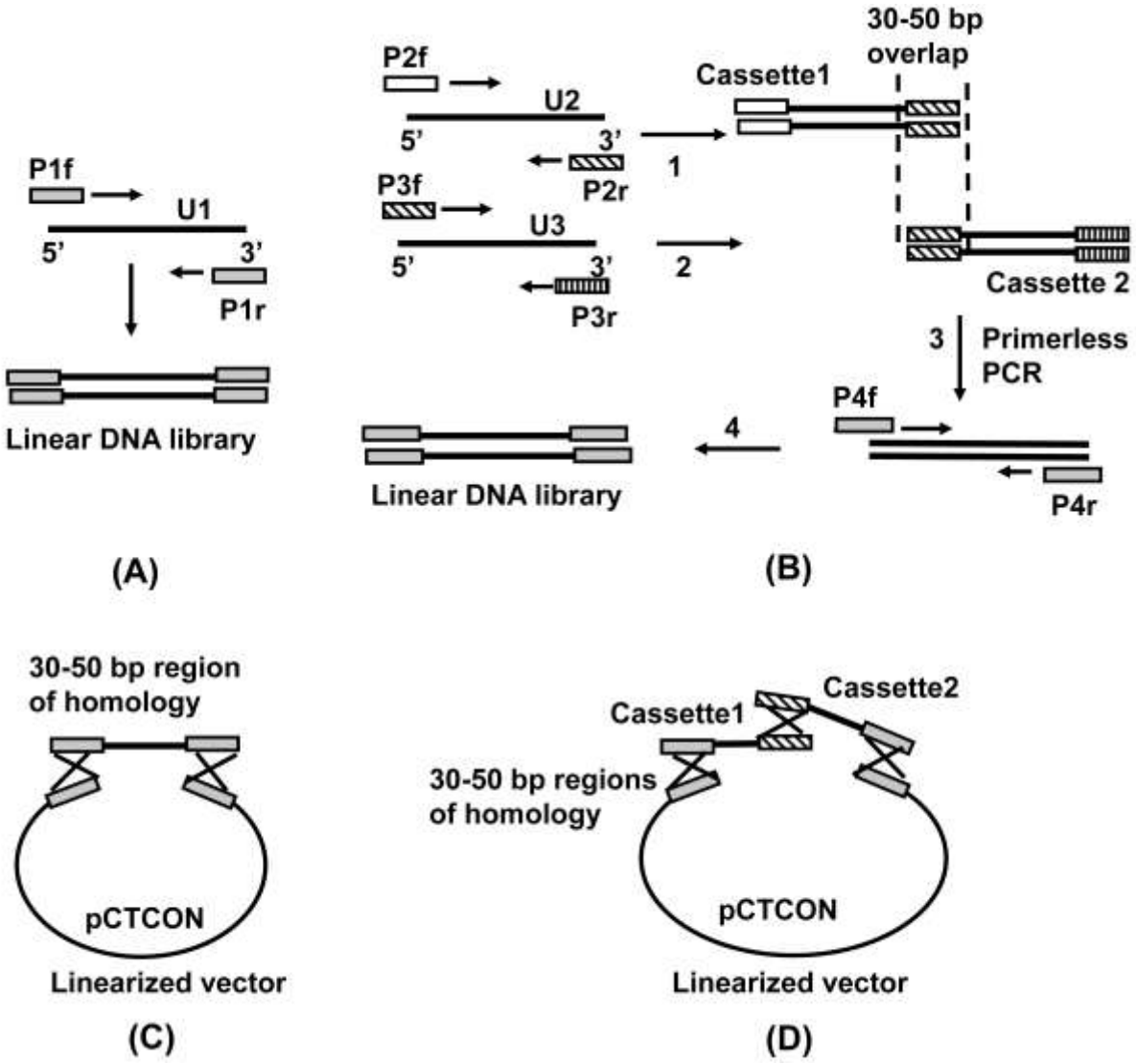



**Figure 2.3**: Generation of a yeast surface display library. (A) A linear DNA library can be generated in a single PCR step, using a synthetic oligonucleotide with randomized regions. (B) A linear DNA library can be generated by overlap-extension PCR, using two synthetic oligonucleotides with randomized regions. (C) The yeast surface library can be generated by homologous recombination mediated plasmid gap repair. Cells are transformed with

linearized plasmid vector and a single DNA fragment encoding the library. (D) The yeast surface library can also be generated by transforming cells with the linearized plasmid vector and two DNA fragments.

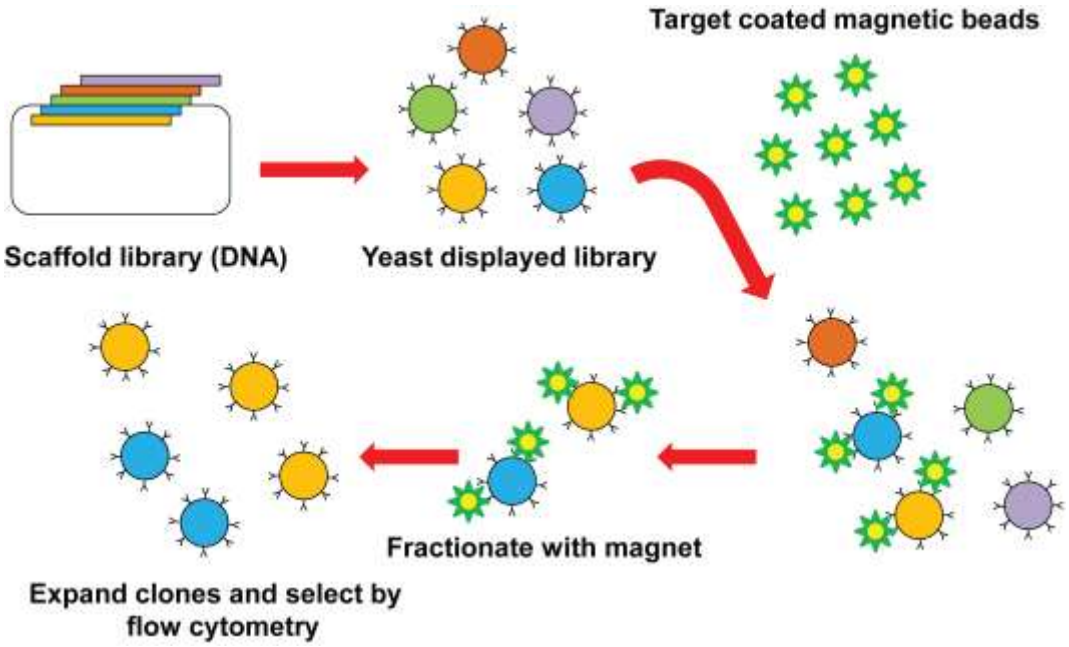



**Figure 2.4**: Magnetic screening of a yeast surface displayed library. Yeast cells in the library are incubated with target-coated micron-sized magnetic beads. Yeast cells expressing binding proteins that bind the target form yeast-bead complexes that are isolated using a magnet.

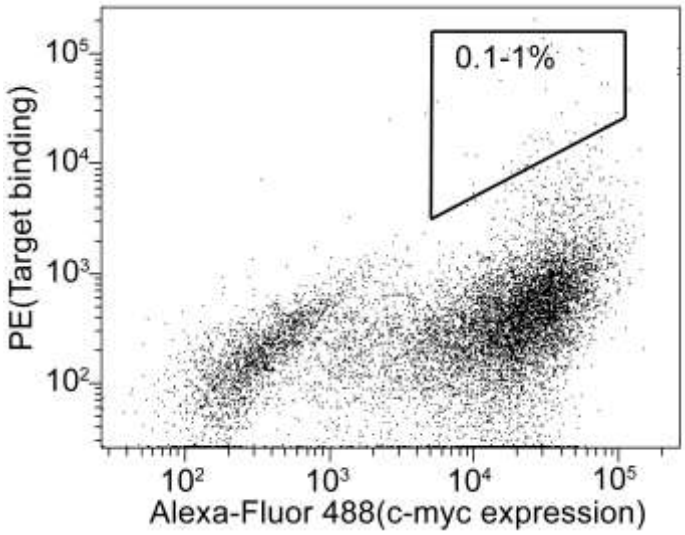



**Figure 2.5**: Screening using Fluorescence Activated Cell Sorting (FACS). Yeast cells expressing the mutant protein are labeled with the biotinylated target and an anti-c-myc antibody. Subsequently, immunofluorescent detection is achieved by secondary labeling with strep-PE and a secondary antibody conjugated with a fluorophore, say Alexa Fluor 488. The Alexa Fluor 488 fluorescence intensity (X-axis) corresponds to the total number of protein fusions on the cell surface while the PE fluorescence intensity (Y-axis) corresponds to the number of fusions that are bound to the target. Each dot on the plot represents the PE and Alexa-Fluor 488 fluorescence values for a single cell. At a given target concentration, cells that have a higher fraction of cell surface fusions bound to the target are those that express mutants with higher binding affinity. The top 0.1-1% of cell is collected; this corresponds to the polygonal region.

**Tables**

**Table 2.1**: Protein scaffolds engineered using yeast surface display

| Scaffold | | Mutation Strategy | Library Size | Targets(partial list) | Affinity Maturation |
|---|---|---|---|---|---|
| Single Chain Variable Fragment (scFv) | | - | $10^9$ | Lysozyme, fluorescein, epidermal growth factor[22] | No |
| $10^{th}$ domain of fibronectin | | Loop Mutagenesis and Loop Length diversification | 0.2-6 $x10^8$ | Lysozyme[28, 29] | Yes |
| Green Fluorescent Protein | | Randomized loop Insertions | [a]$6x10^6$ | Neutrophin receptor (TrkB), GAPDH, Strep-PE[30] | No |
| Knottins | Agouti related protein (AgRP) | Loop grafting | $0.5\text{-}2X10^7$ | $\alpha_v\beta_3$, $\alpha_{iib}\beta_3$[32, 33] | Yes |
| | *Ecballium elaterium* trypsin inhibitor(EETI-II) | Loop grafting | $1\text{-}2x10^7$ | $\alpha_v\beta_3$,$\alpha_v\beta_5$, $\alpha_5\beta_1$, $\alpha_{iib}\beta_3$[34, 35] | Yes |
| Sso7d | | β-sheet mutagenesis | $10^8$ | mIgG, Lysozyme, β-catenin peptide, fluorescein[36] | No |
| Kringle domain | | Loop mutagenesis | $2x10^6$ | Human death receptors 4 and 5, tumor necrosis factor-α[37] | No |

[a]expressed library diversity

**Table 2.2:** Comparison of display methods

| Property | Display Systems | | | |
|---|---|---|---|---|
| | **Yeast** | **Bacteria** | **Phage** | **mRNA** |
| Post translational processing | Yes | No | No | No |
| Cellular Toxicity | No | Yes | Yes | No |
| Compatibility with FACS | Yes | Yes | No | No |
| Typical Library Size | $10^8$-$10^9$ | $10^9$-$10^{11}$ | $10^9$-$10^{11}$ | $10^{13}$-$10^{14}$ |
| Avidity Effect | Yes | Yes | Yes | No |

# CHAPTER 3

# HIGHLY STABLE BINDING PROTEINS DERIVED FROM THE HYPERTHERMOPHILIC Sso7d SCAFFOLD

Nimish Gera, Dalon P. White, Robert C.Wright and Balaji M. Rao

Department of Chemical and Biomolecular Engineering

North Carolina State University

Raleigh, NC 27695-7905

**(**Adapted from Gera et al. 2011, *Journal of Molecular Biology* 409, 601-616**)**

### 3.1. Introduction

The generation of binding molecules that specifically recognize target species is of great importance in clinical applications, biotechnology and fundamental research in biology. Historically, antibodies or immunoglobulins have been the most commonly used reagents for molecular recognition. Antibodies that bind specifically to a wide variety of targets can be obtained through alteration of the amino acid composition of six hypervariable loop regions, called Complementarity Determining Regions (CDRs), on the immunoglobulin framework. However, antibodies are limited by low stability resulting from their complex multi-subunit structure, and relatively high cost of production.

To overcome the limitations of antibodies, several alternate protein frameworks have been used as scaffolds or "templates" for engineering molecular recognition [1-3]. A protein variant that binds a particular target with high affinity and specificity is generated through a combination of amino acid substitutions and sequence insertions or deletions in certain regions of the scaffold protein. Typically, binding proteins are isolated from a combinatorial library of scaffold protein mutants. Common examples include the $10^{th}$ domain of fibronectin[4-6], Designed Ankyrin Repeat Proteins (DARPins)[7, 8] and affibodies[9, 10]. Proteins derived from alternate non-immunoglobulin scaffolds can provide several advantages over antibodies such as low molecular weight, higher stability and ease of recombinant expression in bacterial systems.

Since binding proteins are obtained through mutagenesis of the scaffold protein, tolerance to a diverse set of amino acid changes while retaining its native structure is the hallmark of an

ideal scaffold. Highly stable proteins are more tolerant to a wide range of mutations and therefore have enhanced evolvability[11]. Consequently, highly stable scaffold proteins are more likely to yield binding proteins to a wide variety of target species. In nature, hyperthermophilic archaea and bacteria have evolved to withstand extreme conditions of temperature and pH. Not surprisingly, several proteins from these organisms have very high thermal stability and resistance to proteolysis[12, 13]. Further, many of these proteins also have low molecular weight and lack disulfide bonds, enabling their facile expression in the reducing cytoplasm of *E. coli*. As a result, hyperthermophilic proteins are excellent candidates for use as protein scaffolds to engineer molecular recognition, in addition to being attractive model proteins for obtaining insight into protein stability. Nevertheless, the generation of binding proteins from proteins of hyperthermophilic origin using combinatorial library screening methods has received scant attention. Here we show that the ultrastable Sso7d protein from the hyperthermophilic archaeon *Sulfolobus solfataricus* can be used as a versatile scaffold to generate highly stable binding proteins for a wide range of target species.

The Sso7d protein is a small (~ 7 kDa, 63 amino acids) DNA binding protein containing a SH3-like fold consisting of an incomplete beta-barrel with five beta strands and a C-terminal alpha helix[14-17]. Sso7d is highly stable with a melting temperature of nearly 100 °C and lacks cysteines[18]. We investigated the possibility of generating binding proteins to a wide spectrum of target species through mutagenesis of DNA binding surface of Sso7d. Towards this end, we generated a modestly sized combinatorial library of ~ $10^8$ Sso7d mutants using yeast surface display[19, 20], through randomization of 10 amino acid residues on the DNA

binding surface of Sso7d. Subsequently, we screened this library and successfully isolated binding proteins for a diverse set of model targets, with equilibrium dissociation constants ($K_D$) in the micromolar to nanomolar range. Our set of chosen targets comprised of a small organic molecule, a peptide and model proteins of varying size, including immunoglobulin proteins from different species. Isolation of binding proteins that can discriminate between closely related immunoglobulins demonstrates the ability to engineer highly specific molecular recognition through mutagenesis of the Sso7d scaffold. Further, we recombinantly expressed several Sso7d-derived binding proteins in *E. coli* and evaluated their stability under denaturing conditions. Despite mutation of 10 out of 63 residues (15%) in Sso7d, the mutant binding proteins show remarkable stability against thermal and chemical denaturation as well as extended exposure to a wide range of pH (0.33 -12.5). Previously, binding proteins to the membrane protein PulD have been isolated using ribosome display, from a combinatorial library of $3x10^{12}$ mutants obtained through randomization of 14 residues on the nearly identical Sac7d from *Sulfolobus acidocaldarius* [21]. Taken together, our results show that stable binding proteins for diverse target species can be generated using the Sso7d scaffold. Due to their favorable properties such as low molecular weight, ease of bacterial expression and high stability, we expect that the Sso7d-derived binding proteins can be used in several clinical and biotechnology applications.

## 3.2. Results

### 3.2.1. Construction of Sso7d mutant library using yeast surface display

We constructed a library of Sso7d variants through mutation of 10 surface exposed residues in the DNA binding region of Sso7d, as shown in Figure 3.1. The wild-type Sso7d protein is reported to have RNase activity which can limit the application of Sso7d-derived mutants in the context of specific intracellular inhibition of target proteins or domains thereof. Interestingly, single point mutations K12L or E35L completely abolish the RNase activity of Sso7d without affecting its thermal stability [22]. Therefore, an E35L mutant of Sso7d was used as the starting point for library construction. The DNA sequences encoding the Sso7d mutant library were obtained through synthesis of a single oligonucleotide with NNN degenerate codons at positions encoding the residues K20, K21, W23, V25, M28, S30, T32, T40, R42 and A44. Subsequently, a double stranded DNA library was generated through a single PCR step, and homologous recombination mediated plasmid gap repair[23] during yeast transformation was used to obtain a yeast surface display library of Sso7d mutants.

The library diversity was estimated as $1x10^8$, based on the number of yeast transformants obtained. We sequenced 57 individual clones from this library; 16 out of 57 clones sequenced, i.e. 28%, were full length Sso7d clones without any frameshift mutations or deletions. Each Sso7d mutant in the yeast surface display library is expressed as a C-terminal fusion to the yeast cell wall protein Aga2 and is flanked by an N-terminal HA epitope tag and a C-terminal c-myc epitope tag. Flow cytometric analysis following immunofluorescent labeling of the c-myc epitope tag showed that approximately 30% of the mutants in the yeast

library express full-length Sso7d variants (Figure 3.2). It is thought that the quality control mechanism of the yeast endoplasmic reticulum ensures that only correctly folded proteins are exported to the yeast cell surface[24]. Therefore, it is likely that proteins expressed as full-length fusions on the yeast cell surface are correctly folded. It must be noted that based on sequencing data only about 30% of the yeast cells contain DNA encoding full-length Sso7d mutants. Further, a fraction of the yeast cells in the library may not express cell surface protein fusions due to plasmid loss[25]. Taken together, our data on c-myc epitope expression suggests that a very large fraction of the Sso7d mutants expressed as cell surface fusions are correctly folded. However, these results must be interpreted with caution, since incorrectly folded proteins or molten globules may be expressed as cell surface fusions, especially in the case of small proteins with high thermal stability[26].

### 3.2.2. Library screening to isolate binding proteins for model targets

To assess the likelihood of generating highly specific binding proteins based on the Sso7d scaffold, for a wide spectrum of target species, we screened the Sso7d mutant library against a diverse set of model targets. Our representative targets included a small organic molecule (fluorescein), a 12 amino acid containing peptide from the C-terminal end of the beta-catenin protein, the model proteins hen egg lysozyme and streptavidin, and two closely related proteins - immunoglobulin G from mouse (mIgG) and chicken (cIgY).

The screening process consisted of two steps. In the first step, the Sso7d library was incubated with micron-sized magnetic beads that were coated with the target. Subsequently,

yeast cells that bound the target-coated beads were isolated using a magnet. This magnetic selection procedure allows capture of even low affinity binders, due to the avidity effect arising from the interaction between multiple copies of the Sso7d mutant protein on the yeast cell surface and multiple target molecules immobilized on the magnetic beads[27]. Prior to isolation of binders, negative selection steps using magnetic beads or beads coated with non-target proteins were employed. This step enables the rejection of yeast cells displaying Sso7d mutants that bind the beads and non-target proteins from further analysis. In the second step of the screening process, the pool of mutants obtained after magnetic selection was subjected to Fluorescence Activated Cell Sorting (FACS). Multiple rounds of FACS resulted in a pool of mutants with the highest binding affinity for a given target. For each target, we sequenced plasmid DNA isolated from five individual yeast clones in the final pool of mutants after FACS. Our results are shown in Table 3.1

Five distinct Sso7d mutants each were obtained from the pool of binders for lysozyme and the peptide fragment from beta-catenin, two distinct clones were obtained in case of binders to mIgG and a single dominant clone was found for binders to cIgY, streptavidin and fluorescein. Based on analysis of the fifteen distinct Sso7d mutants obtained for the six targets using their GRAVY score[28], it is seen that the mutant Sso7d proteins are in general more hydrophobic than the wild-type Sso7d protein. An overall increase in aromaticity of the binding interface, relative to the wild-type protein, is observed in case of binders to mIgG, streptavidin, lysozyme and fluorescein; three or more aromatic residues are present in these proteins.

### 3.2.3. Analysis of binding affinity and specificity

A single clone from the pool of mutants with the highest affinity for each target was selected for further analysis of binding affinity and specificity; selected clones are indicated in Table 3.1 The equilibrium dissociation constants ($K_D$) for the binding interaction between the selected Sso7d mutants and their respective targets were determined through titrations using proteins displayed on yeast cells. Briefly, yeast cells expressing the Sso7d mutants as cell surface fusions were incubated with varying concentrations of the target species. At each target concentration, the fraction of cell surface fusions bound by the target was determined by flow cytometric analysis following indirect immunofluorescent labeling of the target. Subsequently, for each mutant, data from at least three different experiments were fit to a simple one-step binding isotherm and the $K_D$ was estimated as previously described[29]. It is important to note that multiple prior studies have demonstrated the equivalence of $K_D$ values determined using yeast cell surface titrations to those obtained when soluble protein is used for analysis[6]. Estimated $K_D$ values and their associated 68% confidence intervals are shown in Table 3.2. The Sso7d mutants analyzed have a range of binding affinities for their respective targets. Not surprisingly, the lowest binding affinities were observed in case of Sso7d mutants that bind the small molecule fluorescein and the peptide fragment from beta-catenin. The estimates of $K_D$ values for Sso7d mutants binding immunoglobulins and streptavidin may be influenced by the avidity effect; the immunoglobulins and streptavidin have two and four identical subunits respectively and a single target molecule may interact with more than one cell surface displayed Sso7d protein. Nevertheless, assuming that steric considerations allow the immunoglobulins and streptavidin to bind two cell surface fusions

each, the $K_D$ can still be determined from the half-maximal binding concentration[30]. The concentration at which half the yeast cell surface displayed proteins are bound by the target species is the $K_D$ of the interaction; this concentration can be easily determined from the cell surface titration plots. As seen in Table 3.2, the $K_D$ values estimated from the half-maximal binding concentrations are in agreement with values obtained by fitting data from cell surface titrations to a monovalent binding isotherm.

The Sso7d mutants analyzed did not exhibit any significant binding to secondary reagents used for flow cytometry analysis. The specificity of the mIgG-binding Sso7d mutant was further evaluated through binding studies with closely related non-target proteins. As seen in Figure 3.3, yeast cells expressing Sso7d-mIgG do not exhibit significant binding to immunoglobulin G from rabbit (rIgG), chicken and goat (gIgG), and the $F_c$ portion of human IgG (hFc), even when these proteins are present at a high concentration (1 μM). This result further underlines the role of a highly stringent negative selection included in our procedure to isolate binding proteins from the yeast library. The negative selection step using micron sized magnetic beads ensures the effective removal of yeast cells expressing Sso7d mutants that bind non-target species; the avidity effect due to the multivalent interaction between the yeast cell and the magnetic bead ensures removal of cells expressing proteins with even low binding affinity for the non-target species. Binders to streptavidin were isolated using streptavidin-phycoerythrin (Strep-PE) as the target during FACS. To confirm that the isolated Sso7d mutant indeed binds streptavidin and not phycoerythrin, a competition experiment was carried out as follows. Yeast cells displaying Sso7d-streptavidin were incubated with Strep-

PE in the presence or absence of 100-fold excess streptavidin and subsequently, the fluorescent signal from cell surface bound Strep-PE was measured using flow cytometry. In the presence of soluble streptavidin, binding of Strep-PE is completely eliminated; this result confirms that Sso7d-Streptavidin does not bind to phycoerythrin.

### 3.2.4. Soluble expression of Sso7d mutants

Six different Sso7d mutants, each binding to a different target, were recombinantly expressed in the cytoplasm of *E. coli*, with a C-terminal 6xHistidine (6xHis) tag. The recombinant proteins were purified in a single step by Immobilized Metal Affinity Chromatography (IMAC) using a Nickel column. Estimates of purified protein yield ranged from 5-60 mg per liter of bacterial culture. It should be noted that this is a conservative estimate based on elution fractions that contained pure protein; SDS-PAGE analysis of these fractions showed a single band of the Sso7d mutant protein. Eluate fractions with other contaminating proteins were not considered, even though they contained a substantial amount of the Sso7d mutant. Protein concentrations were measured through densitometry on Coomassie blue stained SDS-PAGE gels, using bovine serum albumin (BSA) as a standard. An interesting point to note here is that the estimates of protein concentrations for the Sso7d mutants obtained using absorbance measurements at 280 nm and theoretically calculated molar extinction coefficients, and a Bradford assay using BSA as a standard, differed significantly with each other and also from those obtained by densitometry. The theoretically calculated molar extinction coefficients are likely to be inaccurate since multiple Sso7d mutants analyzed do not contain any tryptophan residues. Similarly, multiple Sso7d mutants lack surface

accessible arginine residues that the Bradford reagent predominantly interacts with; it is likely that BSA is not a good reference standard for estimating concentrations of Sso7d protein variants using a Bradford assay.

Analytical Size Exclusion Chromatography was used to determine the oligomeric state of a subset of the Sso7d mutant binding proteins. Our results show that at the high concentrations used in our analysis (~ 4 mg/mL or ~ 482 μM) the mutant Sso7d proteins analyzed may exist as oligomers. Based on molecular weights estimated from a calibration curve, it is likely that Sso7d-Lysozyme exists predominantly as a monomer, while Sso7d-cIgY and Sso7d-beta-catenin peptide are present in the dimeric state. Nevertheless, all soluble Sso7d mutants are functional. To confirm that the solubly expressed recombinant Sso7d proteins retain binding to their targets, competition experiments were carried out. Briefly, yeast cells displaying a particular Sso7d variant were incubated with the target species in the presence or absence of the soluble Sso7d protein. Subsequently, the fraction of cell surface Sso7d fusions bound to the target was determined through indirect immunofluorescent labeling of the target and flow cytometry. For each Sso7d mutant, the presence of soluble protein leads to a decrease in the fraction of cell surface fusions bound to the target; this result confirms that the recombinant proteins retain their ability to bind their respective target species. All Sso7d mutants analyzed in this study were stored for several weeks at 4 °C and 1-10 mg/mL; all proteins remained soluble under these conditions.

### 3.2.5. Biophysical characterization of Sso7d mutants

The protein sequences of Sso7d-derived binding proteins differ from the wild-type protein at eleven residues, including the E35L substitution introduced to abolish RNase activity. Circular Dichroism (CD) spectroscopy was used to investigate if the large number of mutations introduced (over 17% of the original amino acid sequence is changed) cause a change in secondary structure relative to the wild-type protein. As seen in Figure 3.4, CD spectra for the wild-type Sso7d protein and the six different mutants are similar over the range of wavelengths from 210 nm to 240 nm. The CD spectra suggest that the Sso7d mutants are likely to have the same beta-stranded structure as the wild-type protein. To further investigate the effect of extensive mutagenesis on the Sso7d scaffold, we carried out multiple experiments that assess the stability of the Sso7d-derived binding proteins in different contexts. Specifically, we evaluated the stability of the Sso7d mutants under conditions of thermal and chemical denaturation, and exposure to a wide range of pH.

Differential Scanning Calorimetry (DSC) was used to measure the melting temperature ($T_m$) of the Sso7d mutants. As seen in Table 3.3, the Sso7d mutants analyzed retain high melting temperatures. Sso7d-Streptavdin has the lowest $T_m$ (72 °C) and all other mutants have $T_m$ values greater than 89°C. In this regard, it is interesting to note that Sso7d-Streptavidin carries the W23A mutation, which by itself has been reported to cause a 6°C reduction in $T_m$ of the wild-type Sso7d protein[31]. A remarkable feature of the Sso7d mutants binding mIgG and cIgY is that despite 11 amino acid substitutions, the $T_m$ values for these proteins are essentially identical to the wild-type Sso7d protein. Further, in case of Sso7d-mIgG, Sso7d-

Fluorescein and Sso7d-Streptavidin, melting curves obtained in two successive heating and cooling cycles were nearly identical. These data suggests that a significant fraction of these three mutant proteins undergo a reversible folded-to-unfolded state transition under thermal denaturation conditions.

CD spectroscopy was used to monitor the unfolding of the Sso7d mutants in the presence of Guanidine Hydrochloride (Gdn-HCl). The Sso7d mutants were incubated with varying concentrations of Gdn-HCl at 4 °C and pH 7.3 for 12-18 hours. Subsequently, molar ellipticity measurements at 222 nm were used to estimate the fraction of unfolded protein at a given concentration of Gdn-HCl. Figure 3.5 shows the Gdn-HCl denaturation curves for the six Sso7d mutants analyzed. Gdn-HCl concentrations corresponding to the mid-point of the folded-to-unfolded state transition for the Sso7d mutants ($[\text{Gdn-HCl}]_{1/2}$) are listed in Table 3.3**.** Most Sso7d mutants are less stable to Gdn-HCl denaturation than the wild-type protein ($[\text{Gdn-HCl}]_{1/2}$ = 4 M). Of the mutants analyzed, Sso7d-beta-catenin peptide has the lowest $[\text{Gdn-HCl}]_{1/2}$ (2.5 M) while Sso7d-Lysozyme has a higher value of $[\text{Gdn-HCl}]_{1/2}$ (5M) than the wild-type protein. To put this data in context, the $[\text{Gdn-HCl}]_{1/2}$ for a typical single chain antibody has been reported as 1.5 M[32].

To evaluate pH stability of the Sso7d mutants, protein denaturation as a result of extended exposure to varying pH conditions was measured using CD spectroscopy. The Sso7d mutants and the wild-type protein were incubated at 20°C for 12-16 hours, under varying pH conditions. Figure 3.6 shows the CD spectra over a range of pH conditions for the wild-type Sso7d protein and Sso7d-Lysozyme. As with the wild-type protein, the CD spectrum of

Sso7d-Lysozyme over the range of wavelengths 210-240 nm remains largely unchanged under a wide range of pH conditions. CD spectra for the other mutants show similar behavior. The pH range over which the CD spectrum remains unchanged is listed for each Sso7d mutant in Table 3.3. The CD spectra at pH 13.5 were very noisy, likely due to the high ionic strength, and were not considered in this analysis.

These results indicate that despite extensive mutagenesis, binding proteins derived from the Sso7d scaffold are highly stable to thermal and chemical denaturation and can withstand extended exposure to pH extremes.

### 3.2.6. Mutational Analysis of Sso7d-Lysozyme

The Sso7d library was generated through mutagenesis of 10 residues on the DNA-binding surface of Sso7d. To assess the number and location of mutations that contribute towards the generation of binding affinity for a specific target, we conducted a detailed mutational analysis on Sso7d-Lysozyme. Specifically, the mutations in Sso7d-Lysozyme were reverted back to the wild type Sso7d residues, one by one. Subsequently, we quantified the lysozyme-binding activity of these single mutants. Briefly, yeast-surface displayed mutants were labeled with varying concentrations of soluble lysozyme, ranging from 25 nM to 1 μM. The fraction of yeast cell surface fusions that bind lysozyme at each concentration was used to classify the mutants in the following four categories: binding similar to Sso7d-Lysozyme (+++), binding somewhat reduced (++), binding greatly reduced (+) and binding completely abolished in the given concentration range (-). This analysis allows us to assess the

contribution of a particular mutation to the overall binding to determine the energetic hotspot in the binding site. The results of our analysis are summarized in Figure 3.7.

Out of the ten mutations analyzed, seven (F21K, F23W, W25V, C30S, D32T, S42R, C44A) have a profound effect on binding affinity, as shown by the complete loss of binding to lysozyme upon reversion to the wild-type residue. The reversion mutants corresponding to Q40 and M28 retain some binding affinity for lysozyme, albeit reduced, suggesting that these positions contribute somewhat to the binding interaction with lysozyme. On the other hand, the C20K mutant retains binding affinity to lysozyme similar to Sso7d-Lysozyme suggesting that C20 is not important. Taken together, our results show that nine of the ten mutations in Sso7d-Lysozyme have a role in binding to lysozyme.

### 3.3. Discussion

We have shown that the hyperthermophilic Sso7d protein from *Sulfolobus solfataricus* can be mutagenized to generate highly stable binding proteins for a variety of different targets. Sso7d is a DNA binding protein with a beta-barrel structure that topologically resembles the Src Homology 3 (SH3) domain fold and the oligonucleotide/oligosaccharide (OB) fold; the Sso7d protein is classified as an SH3-fold protein[33]. Together, the SH3- and OB-fold proteins bind a wide spectrum of target species including DNA/RNA, oligosaccharides, peptides and small molecules[34, 35]. Sso7d ($T_m$ = 98 °C) is the most thermostable protein possessing the SH3-fold[36]. Mutagenesis of multiple residues in a protein, as in the case of generating binding proteins through mutation of a scaffold protein, often results

destabilization of the protein and a concomitant decrease in $T_m$ of the mutant relative to the wild-type protein. It can therefore be naively reasoned that a scaffold protein with high thermal stability is more likely to yield mutant binding proteins that are stable. Indeed, it has been shown that highly thermostable proteins are more tolerant to mutation and consequently have enhanced evolvability[11]. Due to the seemingly inherent adaptability of the SH3- and OB-fold and the highly thermostable nature of Sso7d, we hypothesized that the Sso7d protein is likely to be versatile scaffold for generating binding proteins. In this study, we investigated the ability of the Sso7d scaffold to yield binding proteins to a diverse set of targets.

We mutagenized 10 surface accessible residues on beta-strands in the DNA-binding interface of Sso7d (Figure 3.1) to generate a library of $10^8$ Sso7d mutants using yeast surface display. DNA sequencing suggests that this yeast library contains about $3x10^7$ full-length Sso7d mutants. We were successfully able to isolate binding proteins to a set of six diverse model targets from this modestly-sized library; our targets included a small organic molecule (fluorescein), a peptide (a 12 amino acid peptide from the C-terminus of the beta-catenin protein), hen egg lysozyme, streptavidin and immunoglobulins. Further, the selected Sso7d mutants could discriminate between the closely related immunoglobulin targets (Figure 3.3). Previous studies have reported the isolation of *de novo* binding proteins from small yeast surface display libraries[5, 37-39]. In such cases, the number of correctly folded mutants in the library can be maximized through careful choice of residues to be mutagenized. Due to the high thermal stability of the Sso7d protein, and choice of surface accessible residues for mutagenesis based on data from previous studies with Sso7d, it is likely that a very large

fraction of the full length Sso7d mutants are correctly folded. Indeed, the high thermal stability of the six different mutants characterized in this study strongly suggests that the Sso7d scaffold is highly tolerant to mutation. In this regard, it is also interesting to note that indirect immunofluorescent detection of the c-myc epitope tag shows 30% of the total library, i.e. ~ $3x10^7$ mutants, to be comprised of full-length Sso7d mutants displayed as yeast cell surface fusions (Figure 3.2). To put this number in context, sequencing data shows that there are only approximately $3x10^7$ (30%) full-length Sso7d mutants in the library. It has been previously proposed that the quality control mechanism in the yeast endoplasmic reticulum ensures cell surface display of only correctly folded proteins[24], although there is some evidence to suggest that incorrectly folded proteins or molten globules may be displayed as cell surface fusions[26].

The Sso7d mutants obtained have binding affinities in the nanomolar to micromolar range for their respective targets (Table 3.2). It is important to note that no affinity maturation strategies were applied in this study. Additional rounds of mutagenesis and screening may be used to obtain binders with higher affinities. Alternatively, it might be possible to obtain binders with higher affinity through screening of an Sso7d library with higher diversity. In this study, we used a yeast surface display library of ~ $10^8$ Sso7d mutants, with ~ $3x10^7$ full length clones. In yeast surface display, library diversity is limited by the transformation efficiency of yeast; the largest reported diversity of a yeast display library is ~$10^9$ mutants[40]. A lysozyme-binding protein with a $K_D$ of 5 nM has been isolated from a library of ~ $10^{12}$ mutants derived from the nearly identical Sac7d scaffold, using ribosome

display[41]. By contrast, the lysozyme-binding Sso7d mutant isolated in this study using yeast surface display has a $K_D$ of ~ 350 nM. Nevertheless, the yeast display system provides certain key advantages such as the use of flow cytometric sorting, the ability to discriminate between clones with small differences in binding affinity and ease of quantitative determination of binding affinities ($K_D$) through cell surface titrations. Therefore, a screening strategy that harnesses the large library diversity accessible using ribosome display or mRNA display in a preliminary screening step, followed by subsequent steps using yeast surface display may be optimal.

An interesting question that arises in the context of generating *de novo* binders from scaffold proteins is: how many mutations play a role in the binding interaction with the target? To address this question, we conducted a mutational analysis on Sso7d-Lysozyme wherein each mutation was reverted to the wild-type residue and the binding affinity of the reversion mutant was compared to Sso7d-Lysozyme. Strikingly, our analysis (Figure 3.7) showed that nine of the ten mutations in Sso7d-Lysozyme contribute to the binding interaction with lysozyme. While this result cannot be extrapolated to any target, our data shows that the DNA-binding surface of Sso7d can be mutagenized in a manner such that nearly all the mutated residues contribute to target-binding.

The mutant Sso7d proteins analyzed in this study could be easily expressed at reasonable yields in the reducing cytoplasm of *E. coli* and purified using conventional IMAC protocols. Despite extensive mutagenesis relative to the wild-type protein (11 mutations out of 63 residues including E35L), the Sso7d mutants exhibited remarkably high melting

temperatures. Five of the six mutants analyzed had melting temperatures in excess of 89 °C. To the best of our knowledge, Sso7d-mIgG ($T_m$ =102.8 °C) has the highest reported melting temperature amongst mutant proteins derived from non-immunoglobulin scaffolds. Table 3.4 lists the $T_m$ ranges for mutants derived from commonly used non-antibody scaffolds. Simplistically, mutants derived from a protein scaffold will be thermally stable if the non-mutagenized regions of the scaffold are key determinants of scaffold protein stability. For instance, mutation of 11 amino acid residues in the non-binding surface results in an optimized affibody scaffold with greater thermal stability[42]. This optimized scaffold, in turn, yields binding proteins with improved thermal stability. In case of Sso7d, the hydrophobic core comprised of eleven residues (V3, F5, V14, I19, I29, F31, Y33, G43, V45, A50 and L54) as well as the C-terminal alpha helix play a major role in protein stability[31, 43]. Mutagenesis of residues in the hydrophobic core results in a significant decrease in the melting temperature of Sso7d[44]. In particular, a single point mutation F31A was shown to decrease the melting temperature by 24 °C[45]. Similarly, deletion of the C-terminal alpha helix causes a 46 °C decrease in melting temperature[43]. The Sso7d mutants analyzed were also resistant to chemical denaturation mediated by Gdn-HCl and extreme conditions of pH. Of the mutants analyzed, Sso7d-beta-catenin-peptide has the lowest $[\text{Gdn-HCl}]_{1/2}$ of 2.5 M. By comparison, the corresponding $[\text{Gdn-HCl}]_{1/2}$ for a typical single chain antibody was reported as 1.5 M[32]. All Sso7d mutants analyzed could also withstand extended incubation in a wide range of pH conditions (0.33-12.5), while retaining their secondary structure; data obtained at pH 13.5 was inconclusive. Taken together, our results show that binding proteins

derived from the Sso7d scaffold exhibit several desirable properties such as easy recombinant expression, high thermal stability and resistance to chemical denaturation and pH extremes.

To date, antibodies are the most commonly used scaffolds for generating binding proteins. Apart from historical reasons, this is largely because the topological diversity generated through altering the length and amino acid compositions of six hypervariable loops in the CDRs of antibodies can yield high affinity binding proteins to a wide spectrum of targets. In our work, the lowest binding affinities were obtained in case of binders to fluorescein ($K_D$ ~ 7 μM) and the beta-catenin peptide fragment ($K_D$ ~ 2.5 μM). It is possible that the flat surface topology of the DNA-binding surface of Sso7d is not optimal for generating high affinity binding proteins to these targets. Nevertheless, the Sso7d protein yields binding proteins with several desirable properties such as small size, easy expression in *E. coli* and high thermal stability. The tolerance to mutation that makes Sso7d a versatile scaffold is arguably a result of its high thermal stability and hyperthermophilic origin. Hyperthermophilic archaea and bacteria may provide a rich source of small stable proteins that can be used as scaffolds for generating binding proteins. An ensemble of hyperthermophilic scaffolds where multiple different topologies are mutagenized may be well suited for generating binding proteins for a wide spectrum of targets. Further, a different topology on the Sso7d protein may be mutagenized to obtain binding proteins. Previously, binding proteins have been isolated from a library generated through mutagenesis of the RT- and src-loops in an SH3-fold containing protein[46]. Therefore, it is conceivable that the loop regions in Sso7d near the ATP-binding pocket[47] (residues Y7 and K39) may be targeted

for mutagenesis. We are currently investigating the use of an ensemble of hyperthermophilic protein scaffolds as well as mutagenesis of loop regions of Sso7d to generate binding proteins.

In summary, we have shown that the Sso7d protein scaffold can be engineered to generate binding proteins with high specificity for a wide range of targets. Sso7d variants are highly stable to thermal and chemical denaturation as well as pH extremes. Consequently, Sso7d-derived binding proteins are well-suited for use in several applications where antibodies cannot or need not be used. The lack of disulfide bonds and low molecular weight make them ideal for use as intracellular inhibitors for "protein interference". Unlike RNA interference, mutant proteins based on the Sso7d scaffold can be used to bind and inhibit the function of a specific domain or a post-translationally modified form of a protein[21]. Sso7d proteins are also well-suited for use as affinity ligands in protein purification applications. Due to their resistance to pH extremes, the Sso7d mutants are likely to be compatible with clean-in-place sterilization cycles involving wash steps with sodium hydroxide solution. It is worthwhile to note that a very high binding affinity of interaction is not required in both these aforementioned applications. Finally, due to their high thermal stability and ease of recombinant expression, Sso7d mutant proteins can also be used as low-cost affinity reagents in imaging and diagnostics. Thus the Sso7d protein is a versatile scaffold that is relevant to multiple applications in biotechnology and medicine.

## 3.4. Materials and Methods

### 3.4.1. Sso7d library construction

Surface exposed residues in the DNA binding surface of the wild-type Sso7d protein were identified using Visual Molecular Dynamics Software (VMD, University of Illinois, Urbana-Champaign). Ten residues in the flat DNA binding surface of Sso7d were chosen for randomization. Subsequently, the Sso7d library was synthesized as a single oligonucleotide U1 containing degenerate NNN codons (Integrated DNA Technologies, Coralville, IA). The DNA sequence of U1 is 5'- ATG GCG ACC GTG AAA TTT AAA TAT AAA GGC GAA GAA AAA CAG GTG GAT ATT AGC AAA ATT **NNN NNN** GTG **NNN** CGC **NNN** GGC AAA **NNN** ATT **NNN** TTT **NNN** TAT GAT CTG GGC GGC GGC AAA **NNN** GGC **NNN** GGC **NNN** GTG AGC GAA AAA GAT GCG CCG AAA GAA CTG CTG CAG ATG CTG GAA AAA CAG AAA AAA - 3'. The non-randomized regions of U1 contain codons that are optimized for expression in *E. coli.*

A double stranded DNA library of Sso7d mutants was generated by amplification of U1 in a single PCR step using oligonucleotide primers Pf1 and Pr1 that contain yeast surface display consensus sequences as well as the restriction sites for the enzymes *NheI* (Pf1) and *BamHI* (Pr1). The primer sequences are: Pf1 - 5'AGT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT *GCT AGC* ATG GCG ACC GTG AAA TTT AAA TAT AAA G 3' and Pr1 - 5'CTC GAG CTA TTA CAA GTC CTC TTC AGA AAT AAG CTT TTG TTC *GGA TCC* TTT TTT CTG TTT TTC CAG CAT CTG 3'; restriction sites are italicized. Each 50µl PCR reaction had the following components: Phusion$^{TM}$ high fidelity

DNA polymerase (New England Biolabs, Ipswich, MA; 1U/50µl) in 1X high fidelity Phusion™ buffer, 0.2mM deoxynucleotide triphosphate (dNTPs), 0.5µM primers, 1M betaine and 3% dimethyl sulfoxide (DMSO). The mixture was denatured at 98°C for 2min followed by 25 cycles of 98°C for 1 min, 68°C for 1 min, 72°C for 15 sec and a final extension of 72°C for 10 min. PCR products were concentrated using PelletPaint™ (Novagen, San Diego, CA).

A yeast surface display library of Sso7d mutants was generated by homologous recombination mediated plasmid gap repair, using previously described protocols as a guideline[23, 29]. The *Saccharomyces cerevisiae* strain EBY100 and the yeast surface display plasmid vector (pCTCON) were kind gifts from Prof. K. Dane Wittrup (Massachusetts Institute of Technology, Cambridge, MA). Briefly, the pCTCON vector was linearized by digestion with the restriction enzymes *NheI*, *BamHI* and *SalI*. Linearized vector was PCR purified using Qiagen PCR purification kit (Qiagen, Valencia, CA) and then concentrated using PelletPaint™. Subsequently, the linearized vector was transformed into yeast along with the linear DNA library of Sso7d mutants. Electroporation was carried out at 0.54 kV, 25µF and 1000Ω in a Bio-Rad Gene Pulser system (Bio-Rad, Hercules, CA) with 1 µg of cut pCTCON vector and 3µg of concentrated Sso7d library PCR product. Electrocompetent EBY100 cells were prepared as described[29]. Transformed cells were grown in YPD medium (10g/L yeast extract, 20g/L peptone, 20g/L dextrose) for 1 hour at 30°C and 250 rpm. At this point, serial dilutions of a small fraction of the transformed cells were plated on SDCAA plates (20g/L dextrose, 5g/L casamino acids, 6.7g/L yeast nitrogen

base, 182 g/L sorbitol, 5.40 g/L $Na_2HPO_4$, 7.45 g/L $NaH_2PO_4$, 15g/L agar) to determine library diversity. The remainder of the transformed cells were grown in 250 ml of selective medium SDCAA (20g/L dextrose, 5g/L casamino acids, 6.7g/L yeast nitrogen base, 5.40 g/L $Na_2HPO_4$, 7.45 g/L $NaH_2PO_4$) with 1:100 pen-strep solution (Invitrogen, Carlsbad, CA) at 30°C and 250 rpm for 24-48 hours. Subsequently, the yeast library was passaged 3 times in SDCAA medium and frozen in aliquots of $2x10^9$ cells per vial. The library diversity was determined to be $1x10^8$ transformants.

### 3.4.2. Magnetic Selection

Magnetic selection was carried out using previously described protocols[27]. Yeast cells grown in SDCAA were pelleted and re-suspended to 1x $10^7$ cells/mL ($OD_{600}$=1) in SGCAA(20g/L galactose, 5g/L casamino acids, 6.7g/L yeast nitrogen base, 5.40 g/L $Na_2HPO_4$, 7.45 g/L $NaH_2PO_4$). Cells were incubated at 20°C and 250 rpm for 20-24 hours to induce expression of yeast cell surface protein fusions. 100 µl of Dynal™ biotin binder beads ($4 x10^8$ beads/ml; Invitrogen, Carlsbad, CA) were washed in Phosphate Buffered Saline (8g/L NaCl, 0.2g/L KCl, 1.44 g/L $Na_2HPO_4$ , 0.24g/L $KH_2PO_4$, pH 7.4) containing 0.1 % BSA (PBS-BSA) and incubated with biotinylated target (mIgG, cIgY, hen egg lysozyme, beta-catenin C-terminal peptide, fluorescein) overnight at 4°C with rotation, to obtain the target-coated magnetic beads. Biotinylated versions of these targets were commercially obtained from the following sources: mIgG and cIgY from Jackson Immunoresearch (Westgrove, PA), hen egg lysozyme from Biomeda Corporation (Foster City, CA) and Sigma-Aldrich (St Louis, MO), beta-catenin C-terminal peptide from Genscript (Piscataway,

NJ) and fluorescein from Thermo Scientific (Rockford, IL). Biotin binder beads were used as a target for isolating binders to streptavidin.

Negative selection against the biotin binder beads was performed for all targets (except streptavidin) by incubating ~2 x $10^9$ induced cells (20 times the library diversity) with the washed beads for 1-2 hours at 4°C with rotation in a 1.5 mL tube. For mIgG and cIgY, additional negative selections were performed against cIgY, rIgG and hFc, and mIgG, rIgG and hFc respectively. The tube was then placed on a magnetic particle concentrator, and bead-bound cells were discarded. Unbound cells were added to target pre-coated beads washed with PBS-BSA and incubated for 1-2 hours at 4°C. Finally, bead-bound cells were collected using a magnetic particle concentrator, washed four times with 1 ml of PBS-BSA and grown in 5 ml of SDCAA for 24-48 hours at 30°C and 250 rpm.

### 3.4.3. Fluorescence Activated Cell Sorting (FACS)

Magnetically sorted cells were expanded and induced in SGCAA at $10^7$ cells/mL at 20°C and 250 rpm. FACS was used to isolate binders with highest affinity for the targets using protocols that have been previously described[19, 29, 48, 49]. Briefly, 2x$10^7$ cells were labeled simultaneously with biotinylated target (mIgG, cIgY, hen egg lysozyme, beta-catenin C-terminal peptide, and fluorescein) and an anti-c-myc chicken antibody (Invitrogen, Carlsbad, CA) or an anti-HA mouse antibody (Roche, Indianapolis, IN). Indirect immunofluorescent detection of the target was achieved using secondary labeling with Strep-PE or neutravidin-FITC (Invitrogen, Carlsbad, CA). A goat anti-chicken antibody conjugated

with Alexa Fluor-488 or Alexa Fluor-633, and a goat anti-mouse antibody conjugated with Alexa Fluor-488 or Alexa Fluor-633 (Invitrogen, Carlsbad, CA) were used for secondary detection of the anti-c-myc and the anti-HA antibodies respectively. Strep-PE was used as a target for isolation of streptavidin binders. Cell samples were analyzed and sorted on a FACS Aria (Becton Dickinson, San Jose, CA) flow cytometer. The magnetically sorted population was sorted four times by FACS for cIgY and mIgG, three times for fluorescein and streptavidin and two times for hen egg lysozyme and beta catenin peptide, with successively lower target concentrations. Cells from the final sort were plated on an SDCAA plate. For each target, DNA sequences were obtained for five individual clones and the equilibrium dissociation constant ($K_D$) was determined for a single clone.

### 3.4.4. Measurement of $K_D$

The equilibrium dissociation constant ($K_D$) was determined using cell surface titrations as previously described[29]. Briefly, yeast cells displaying cell surface Sso7d mutant were incubated with varying concentrations of the biotinylated target at room temperature and pH 7.4, followed by Strep-PE; in case of Sso7d-Streptavidin, cells were directly labeled with Strep-PE. Subsequently, flow cytometric analysis was used to measure the mean fluorescence intensity of the cells. The $K_D$ of the binding interaction between the Sso7d mutant and its target was estimated using the following relationship:

$$F = \frac{c[L]_0}{K_D + [L]_0}$$

where, F is the observed mean fluorescence intensity, $[L]_0$ is the corresponding concentration of target used for labeling the cells, and c is a constant. The constant c and $K_D$ were determined through a global fit, using data from at least three different experiments. 68% confidence intervals were calculated as described[50]. The 68% confidence intervals correspond to the commonly reported standard deviations associated with triplicate experiments.

### 3.4.5. Sequence determination of Sso7d variants

Plasmid DNA from the Sso7d library and five Sso7d variants for each target was isolated using the Zymoprep Kit II (Zymoresearch Corporation, Orange, CA). The isolated DNA was subsequently transformed into Novablue$^{TM}$ (*E. coli)* cells (EMD Biosciences, San Diego, CA). Plasmids were sequenced by Genewiz (La Jolla, CA). Oligonucleotide P3 (5' ACT ACG CTC TGC AGG CTA GT 3') was used as the primer for sequencing.

### 3.4.6. Recombinant Expression and Purification of Sso7d mutants

Sso7d variants were cloned into the pET22b (+) vector using *NdeI* and *XhoI* restriction sites. Oligonucleotide primers Pf2 and Pr2 (Integrated DNA Technologies, Coralville, IA) were used to introduce the restriction sites. The primer sequences are as follows: Pf2- 5' GAA TCC *CAT ATG* ATG GCG ACC GTG AAA TTT AAA 3' and Pr2-5' CCG CCG *CTC GAG* TTT TTT CTG TTT TTC CAG CAT C 3'; restriction sites have been italicized. Plasmid constructs were subsequently transformed into Rosetta$^{TM}$ (*E. coli*) cells (EMD Biosciences, San Diego, CA). A 5 ml overnight culture in LB medium (10g/L Bacto-tryptone, 5g/L Yeast

Extract, 10g/L NaCl) containing 1% glucose and 50µg/ml ampicillin at 37°C and 250 rpm was used to inoculate a 1L culture in 2XYT medium(16g/L Bacto-tryptone, 10g/L Yeast Extract, 5g/L NaCl). The cells were induced with 0.5 mM IPTG at an $OD_{600}$ of 1.0 and grown for 19-20 hrs at 37°C and 250 rpm. Subsequently, cells were harvested by centrifugation at 5000 rpm for 20 min, re-suspended in Buffer A (50 mM Tris pH 7.6 , 300 mM NaCl) and lysed by sonication for 10 min in a sonicator. The cell lysate was centrifuged and the final supernatant filtered with a 0.22 µm filter. The filtered sample was loaded on a Hi-Trap 5 ml Ni-column (GE Healthcare Bio-Sciences, Piscataway, NJ) on a Bio-Rad BioLogic Duoflow FPLC (Hercules, CA). 6xHis-tagged Sso7d mutants were eluted with a linear gradient using Buffer A and Buffer B (50 mM Tris pH 7.6, 300 mM NaCl, 500 mM Imidazole). The eluate fractions containing pure protein were pooled, dialyzed with PBS (pH 7-8) (Sso7d-beta-catenin peptide and cIgY binding proteins) or 50 mM sodium acetate buffer, pH 5.0 (Sso7d-mIgG, Sso7d-Lysozyme, Sso7d-Fluorescein and Sso7d-Streptavidin) and concentrated. Protein concentrations were determined by densitometry on Coomassie blue stained SDS-PAGE gels using ImageJ[51] software.

### 3.4.7. Soluble competition experiments

Yeast cells expressing an Sso7d variant as cell surface fusions were incubated with biotinylated target in the presence or absence of 100-200 fold excess of the corresponding soluble Sso7d mutant protein, recombinantly expressed in *E. coli.* Subsequently, yeast cells were labeled with Strep-PE (Invitrogen, Carlsbad, CA) to detect target-bound cell surface fusions.

### 3.4.8. Differential Scanning Calorimetry

Differential Scanning Calorimetry (DSC) experiments were carried out on a Nano DSC II differential scanning calorimeter (TA Instruments, Newcastle, DE). PBS was used as a buffer for Sso7d-beta-catenin peptide and Sso7d-cIgY. 50 mM sodium acetate buffer was used for Sso7d-Fluorescein, Sso7d-Streptavidin, Sso7d-mIgG and Sso7d-Lysozyme. Protein samples and the corresponding buffers were degassed for 30 min-1 hr before analysis. Four scans (1°C/min) consisting two heating and two cooling cycles were carried out for each sample. In each case, a buffer baseline run consisting of the same four scans, in the same temperature range used for the sample, was also carried out. Data was recorded in the DSC Run software and exported to CpCalc software for analysis. Each buffer and sample run was performed twice.

### 3.4.9. Circular dichroism spectroscopy

***Gdn-HCl stability experiments***

Protein samples were incubated overnight (12-16 hrs) in 20 mM sodium phosphate buffer (pH 7.3) and 4°C with varying concentrations of Gdn-HCl (0-8 M). Samples were analyzed on a JASCO-815 spectropolarimeter and CD spectra over the range of wavelengths from 210 to 240 nm were recorded. A 50 nm/min scan rate, 0.1 nm pitch, 1 nm bandwidth and 2 sec. response time were used. Three accumulation scans were performed for every sample. Buffer runs were performed for baseline correction. The molar ellipticity at 222 nm was used as a metric to track protein unfolding upon chemical denaturation. For each protein, the baseline-

corrected molar ellipticity values (θ) were normalized to estimate the fraction of folded protein as:

$$f = \frac{\theta - \theta_{max}}{\theta_{min} - \theta_{max}}$$

where $\theta_{min}$ and $\theta_{max}$ are minimum and maximum values of baseline-corrected molar ellipticity. Normalized ellipticity values are plotted against Gdn-HCl concentration in Figure 3.5.

***pH stability experiments***

Protein samples were incubated in 20 mM sodium phosphate buffer with pH ranging from pH 0.33 to pH 13.5 at 20°C for 12-18 hours. Subsequently, samples were analyzed on a JASCO-815 spectropolarimeter to record CD spectra over the range of wavelengths 210-240 nm, at 50 nm /min, 0.1 nm pitch, 1 nm bandwidth and 2 sec. response time. Three accumulation scans were performed for every sample. Each spectrum was normalized by the molar ellipticity value at 222 nm.

### 3.4.10. Analytical Size Exclusion Chromatography

Size Exclusion Chromatography/gel filtration was performed to determine the oligomeric state of wild type Sso7d and Sso7d-derived binding proteins (Sso7d-Lysozyme, Sso7d-cIgY and Sso7d-beta-catenin peptide) using a Superdex 75 HR 10/300 column (GE Healthcare Biosciences, Piscataway, NJ). The calibration standards used for molecular weight determination were BSA (66 kDa), Carbonic anhydrase (29 kDa), Cytochrome C (12.4 kDa),

Aprotinin (6.5kDa). Blue dextran (2000 kDa) was used for the determination of void volume. All proteins were purified and extensively dialyzed in 1x PBS prior to analysis. The column was equilibrated with 1X PBS for 5 column volumes and 500µl of each sample (~4mg/ml) was loaded on the column. Absorbance at 280 nm was monitored and the time of elution of the sample was compared to the elution time of the calibration standards to estimate molecular weights.

### 3.4.11. Mutational Analysis for Sso7d-Lysozyme

Overlap extension PCR and homologous recombination were used to generate Sso7d-Lysozyme reversion mutants, wherein each mutation was reverted to the wild-type sequence, one at a time. Two different fragments (F1 and F2) were created from the Sso7d-Lysozyme mutant using internal primers, and primers Pf1 and Pr1 as described in the library construction protocol. Fragment F1 was obtained by using primer Pf1 and a reverse internal primer. Fragment F2 was generated using a forward internal primer and the Pr1 primer. The mutation sites were introduced in the internal primers and were incorporated during the PCR steps. The two fragments F1 and F2 had 30-40 bp homology with each other and 50bp homology with the pCTCON vector to facilitate homologous recombination. Subsequently, yeast transformations were performed as described previously and the identity of each reversion mutant was confirmed by DNA sequencing.

Yeast-displayed proteins were used to compare the binding affinities of the reversion mutants to Sso7d-Lysozyme. Briefly, yeast cells expressing the mutants or Sso7d-Lysozyme were

labeled with 5 different concentrations of biotinlyated lysozyme at pH 7.4, followed by strep-PE. Subsequently, flow cytometric analysis was conducted to measure the mean fluorescence intensity of the cells. Simultaneously, in a parallel experiment, mean fluorescence intensity of cells due to immunofluorescent labeling of the c-myc tag was measured. The ratio of mean fluorescence intensity due to lysozyme binding to that due to c-myc labeling was used as a quantitative metric for the fraction of cell surface fusions that bind lysozyme at a given concentration (*f*). Plots of *f* vs. lysozyme concentration for each mutant were compared with the corresponding plot for Sso7d-Lysozyme and each mutant was classified into one of the following four categories: binding similar to Sso7d-Lysozyme (+++), binding somewhat reduced (++), binding greatly reduced (+) and binding completely abolished in the given concentration range (-).

### 3.5. References


1. Binz, H.K., P. Amstutz, and A. Pluckthun, *Engineering novel binding proteins from nonimmunoglobulin domains.* Nat Biotechnol, 2005. **23**(10): p. 1257-68.

2. Skerra, A., *Alternative non-antibody scaffolds for molecular recognition.* Curr Opin Biotechnol, 2007. **18**(4): p. 295-304.

3. Gebauer, M. and A. Skerra, *Engineered protein scaffolds as next-generation antibody therapeutics.* Curr Opin Chem Biol, 2009. **13**(3): p. 245-55.

4. Koide, A. and S. Koide, *Monobodies: antibody mimics based on the scaffold of the fibronectin type III domain.* Methods Mol Biol, 2007. **352**: p. 95-109.

5. Hackel, B.J., A. Kapila, and K.D. Wittrup, *Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling.* J Mol Biol, 2008. **381**(5): p. 1238-52.

6. Lipovsek, D., et al., *Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: molecular convergence with single-domain camelid and shark antibodies.* J Mol Biol, 2007. **368**(4): p. 1024-41.

7. Stumpp, M.T., H.K. Binz, and P. Amstutz, *DARPins: a new generation of protein therapeutics.* Drug Discov Today, 2008. **13**(15-16): p. 695-701.

8. Zahnd, C., et al., *A designed ankyrin repeat protein evolved to picomolar affinity to Her2.* J Mol Biol, 2007. **369**(4): p. 1015-28.

9. Nygren, P.A., *Alternative binding proteins: affibody binding proteins developed from a small three-helix bundle scaffold.* FEBS J, 2008. **275**(11): p. 2668-76.

10. Lofblom, J., et al., *Affibody molecules: engineered proteins for therapeutic, diagnostic and biotechnological applications.* FEBS Lett. **584**(12): p. 2670-80.

11. Bloom, J.D., et al., *Protein stability promotes evolvability.* Proc Natl Acad Sci U S A, 2006. **103**(15): p. 5869-74.

12. Karshikoff, A. and R. Ladenstein, *Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads.* Trends Biochem Sci, 2001. **26**(9): p. 550-6.

13. Ladenstein, R. and G. Antranikian, *Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water.* Adv Biochem Eng Biotechnol, 1998. **61**: p. 37-85.

14. Baumann, H., et al., *Solution structure and DNA-binding properties of a thermostable protein from the archaeon Sulfolobus solfataricus.* Nat Struct Biol, 1994. **1**(11): p. 808-19.

15. Edmondson, S.P. and J.W. Shriver, *DNA binding proteins Sac7d and Sso7d from Sulfolobus.* Methods Enzymol, 2001. **334**: p. 129-45.

16. Knapp, S., et al., *Thermal unfolding of small proteins with SH3 domain folding pattern.* Proteins, 1998. **31**(3): p. 309-19.

17. Baumann, H., et al., *DNA-binding surface of the Sso7d protein from Sulfolobus solfataricus.* J Mol Biol, 1995. **247**(5): p. 840-6.

18. Knapp, S., et al., *Thermal unfolding of the DNA-binding protein Sso7d from the hyperthermophile Sulfolobus solfataricus.* J Mol Biol, 1996. **264**(5): p. 1132-44.

19. Boder, E.T. and K.D. Wittrup, *Yeast surface display for screening combinatorial polypeptide libraries.* Nat Biotechnol, 1997. **15**(6): p. 553-7.

20. Gai, S.A. and K.D. Wittrup, *Yeast surface display for protein engineering and characterization.* Curr Opin Struct Biol, 2007. **17**(4): p. 467-73.

21. Mouratou, B., et al., *Remodeling a DNA-binding protein as a specific in vivo inhibitor of bacterial secretin PulD.* Proc Natl Acad Sci U S A, 2007. **104**(46): p. 17983-8.

22. Shehi, E., et al., *The Sso7d DNA-binding protein from Sulfolobus solfataricus has ribonuclease activity.* FEBS Lett, 2001. **497**(2-3): p. 131-6.

23. Swers, J.S., B.A. Kellogg, and K.D. Wittrup, *Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display.* Nucleic Acids Res, 2004. **32**(3): p. e36.

24. Shusta, E.V., et al., *Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency.* J Mol Biol, 1999. **292**(5): p. 949-56.

25. Kieke, M.C., et al., *Isolation of anti-T cell receptor scFv mutants by yeast surface display.* Protein Eng, 1997. **10**(11): p. 1303-10.

26. Park, S., et al., *Limitations of yeast surface display in engineering proteins of high thermostability.* Protein Eng Des Sel, 2006. **19**(5): p. 211-7.

27. Ackerman, M., et al., *Highly avid magnetic bead capture: an efficient selection method for de novo protein engineering utilizing yeast surface display.* Biotechnol Prog, 2009. **25**(3): p. 774-83.

28. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* J Mol Biol, 1982. **157**(1): p. 105-32.

29. Chao, G., et al., *Isolating and engineering human antibodies using yeast surface display.* Nat Protoc, 2006. **1**(2): p. 755-68.

30. Hackel, B.J. and K.D. Wittrup, *The full amino acid repertoire is superior to serine/tyrosine for selection of high affinity immunoglobulin G binders from the fibronectin scaffold.* Protein Eng Des Sel. **23**(4): p. 211-9.

31. Catanzano, F., et al., *Differential scanning calorimetry study of the thermodynamic stability of some mutants of Sso7d from Sulfolobus solfataricus.* Biochemistry, 1998. **37**(29): p. 10493-8.

32. Worn, A. and A. Pluckthun, *Different equilibrium stability behavior of ScFv fragments: identification, classification, and improvement by protein engineering.* Biochemistry, 1999. **38**(27): p. 8739-50.

33. Agrawal, V. and R.K. Kishan, *Functional evolution of two subtly different (similar) folds.* BMC Struct Biol, 2001. **1**: p. 5.

34. Kishan, K.V. and V. Agrawal, *SH3-like fold proteins are structurally conserved and functionally divergent.* Curr Protein Pept Sci, 2005. **6**(2): p. 143-50.

35. Arcus, V., *OB-fold domains: a snapshot of the evolution of sequence, structure and function.* Curr Opin Struct Biol, 2002. **12**(6): p. 794-801.

36. Granata, V., et al., *Guanidine-induced unfolding of the Sso7d protein from the hyperthermophilic archaeon Sulfolobus solfataricus.* Int J Biol Macromol, 2004. **34**(3): p. 195-201.

37. Pavoor, T.V., Y.K. Cho, and E.V. Shusta, *Development of GFP-based biosensors possessing the binding properties of antibodies.* Proc Natl Acad Sci U S A, 2009. **106**(29): p. 11895-900.

38. Hackel, B.J., et al., *Stability and CDR Composition Biases Enrich Binder Functionality Landscapes.* J Mol Biol.

39. Silverman, A.P., et al., *Engineered cystine-knot peptides that bind alpha(v)beta(3) integrin with antibody-like affinities.* J Mol Biol, 2009. **385**(4): p. 1064-75.

40. Feldhaus, M.J., et al., *Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library.* Nat Biotechnol, 2003. **21**(2): p. 163-70.

41. Cinier, M., et al., *Bisphosphonate adaptors for specific protein binding on zirconium phosphonate-based microarrays.* Bioconjug Chem, 2009. **20**(12): p. 2270-7.

42. Feldwisch, J., et al., *Design of an optimized scaffold for affibody molecules.* J Mol Biol. **398**(2): p. 232-47.

43. Shehi, E., et al., *Thermal stability and DNA binding activity of a variant form of the Sso7d protein from the archeon Sulfolobus solfataricus truncated at leucine 54.* Biochemistry, 2003. **42**(27): p. 8362-8.

44. Consonni, R., et al., *Structural determinants responsible for the thermostability of Sso7d and its single point mutants.* Proteins, 2007. **67**(3): p. 766-75.

45. Consonni, R., et al., *A single-point mutation in the extreme heat- and pressure-resistant sso7d protein from sulfolobus solfataricus leads to a major rearrangement of the hydrophobic core.* Biochemistry, 1999. **38**(39): p. 12709-17.

46. Bertschinger, J., D. Grabulovski, and D. Neri, *Selection of single domain binding proteins by covalent DNA display.* Protein Eng Des Sel, 2007. **20**(2): p. 57-68.

47. Renzone, G., et al., *Structural characterization of the functional regions in the archaeal protein Sso7d.* Proteins, 2007. **67**(1): p. 189-97.

48. VanAntwerp, J.J. and K.D. Wittrup, *Fine affinity discrimination by yeast surface display and flow cytometry.* Biotechnol Prog, 2000. **16**(1): p. 31-7.

49. Boder, E.T. and K.D. Wittrup, *Optimal screening of surface-displayed polypeptide libraries.* Biotechnol Prog, 1998. **14**(1): p. 55-62.

50. Rao, B.M., et al., *Interleukin-2 mutants with enhanced alpha-receptor subunit binding affinity.* Protein Eng, 2003. **16**(12): p. 1081-7.

51. Rasband, W.S., *ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/.* 1997-2009.

52. Bloom, L. and V. Calabro, *FN3: a new protein scaffold reaches the clinic.* Drug Discov Today, 2009. **14**(19-20): p. 949-55.

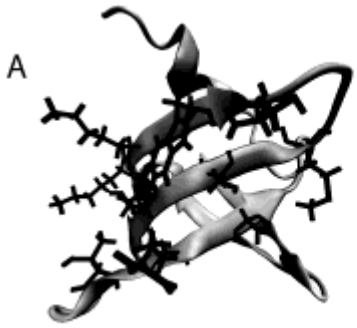
53. Parker, M.H., et al., *Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two.* Protein Eng Des Sel, 2005. **18**(9): p. 435-44.

54. Wetzel, S.K., et al., *Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins.* J Mol Biol, 2008. **376**(1): p. 241-57.

55. Skerra, A., *Alternative binding proteins: anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities.* FEBS J, 2008. **275**(11): p. 2677-83.

56. Schlehuber, S. and A. Skerra, *Tuning ligand affinity, specificity, and folding stability of an engineered lipocalin variant -- a so-called 'anticalin' -- using a molecular random approach.* Biophys Chem, 2002. **96**(2-3): p. 213-28.

**Figures**




B

1.....10.....20.....30.....40
ATVKFKYKGEEKQVDISKI**KK**V**W**R**V**GK**M**I**S**F**T**YD***L***GGGK**T**
.....50.....60..
G**R**G**A**VSEKDAPKELLQMLEKQKK

**Figure 3.1:** Selection of amino acid positions for mutagenesis in the Sso7d protein from *Sulfolobus solfataricus.* **(A)** Residues on the DNA binding surface selected for randomization are shown in licorice representation. (PDB accession code 1SSO; this image was generated using Visual Molecular Dynamics (VMD) software) **(B)** The amino acid sequence for Sso7d is shown. The residues at positions to be randomized are bold-faced and underlined. The specific mutation E35L (italicized and bold-faced) was introduced to abolish RNase activity present in wild-type Sso7d.

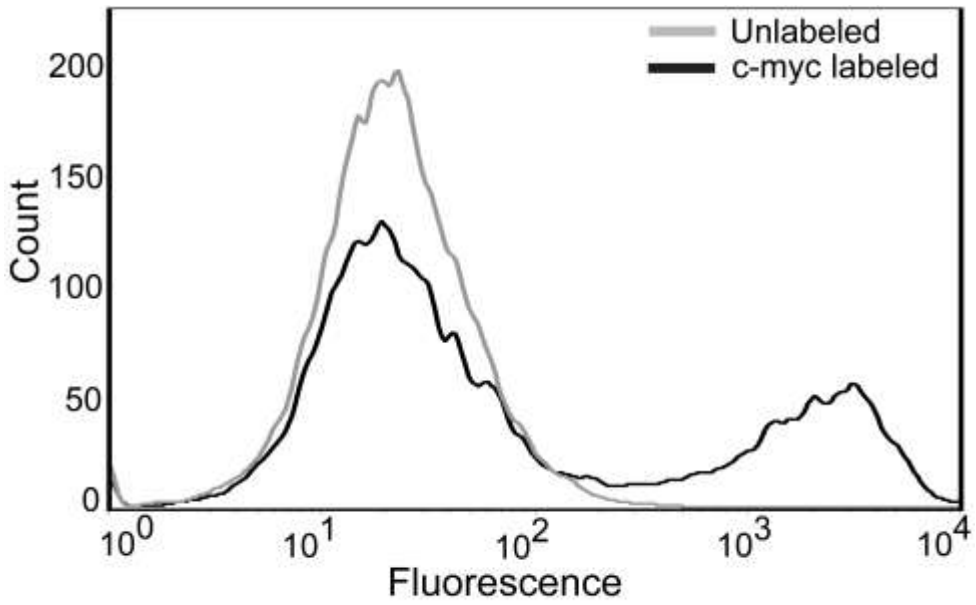



**Figure 3.2:** Expression analysis of Sso7d yeast library through immunofluorescent detection of the c-myc epitope tag. Yeast cells displaying Sso7d mutants as cell surface fusions were labeled with an anti-c-myc antibody followed by a secondary antibody conjugated with Alexa Fluor-488 (black histogram) or just with the fluorescent secondary antibody (grey histogram). Subsequently cells were analyzed by flow cytometry. Approximately 30% of the cells express the c-myc epitope tag, and therefore by inference, full-length Sso7d proteins as cell surface fusions.

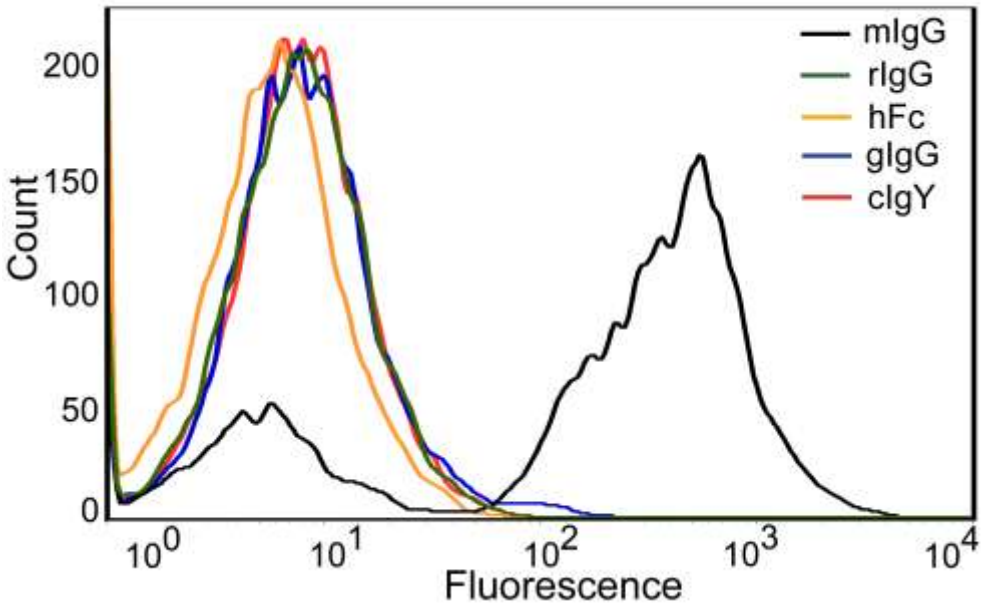



**Figure 3.3:** Specificity analysis for Sso7d-mIgG. Yeast cells displaying Sso7d-mIgG as cell surface fusions were labeled with 1 μM biotinylated mIgG or an equivalent concentration of the non-target immunoglobulin species (biotinylated) chicken IgG (cIgY), goat IgG (gIgG), rabbit IgG (rIgG) and the Fc portion of human IgG (hFc). Subsequently, cells were labeled with streptavidin-phycoerythrin (Strep-PE) and analyzed using flow cytometry. The fluorescence histograms show that Sso7d-mIgG binds to mIgG while binding to other non-target proteins is insignificant.

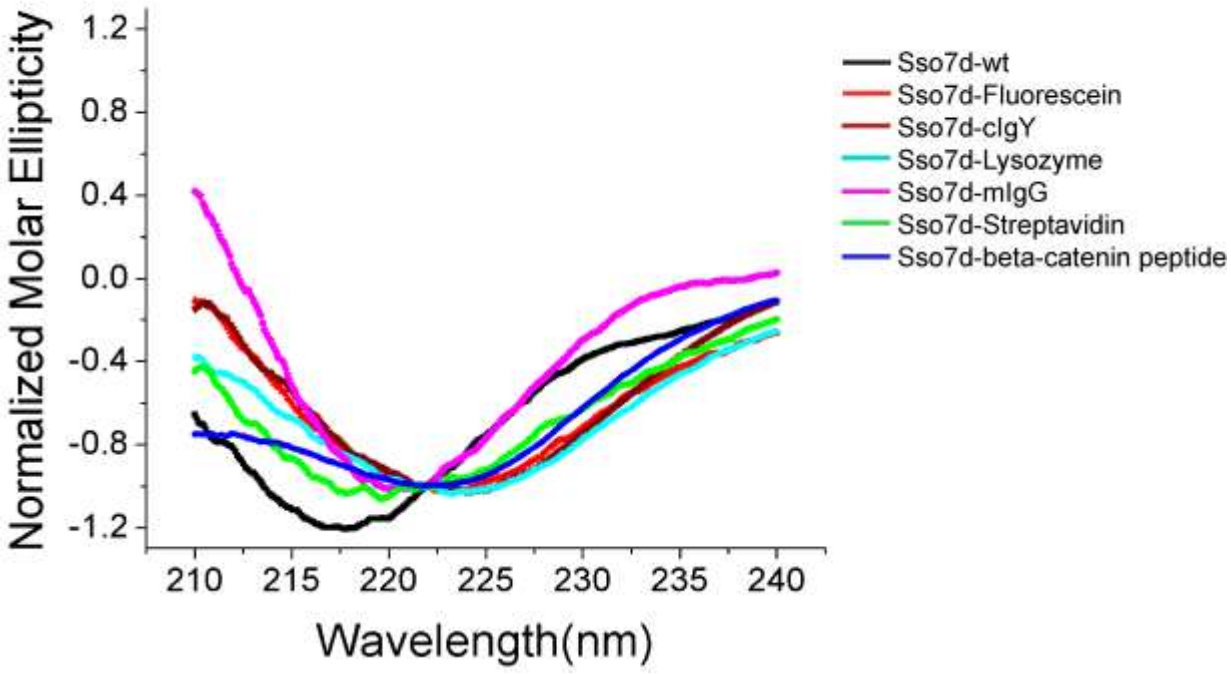
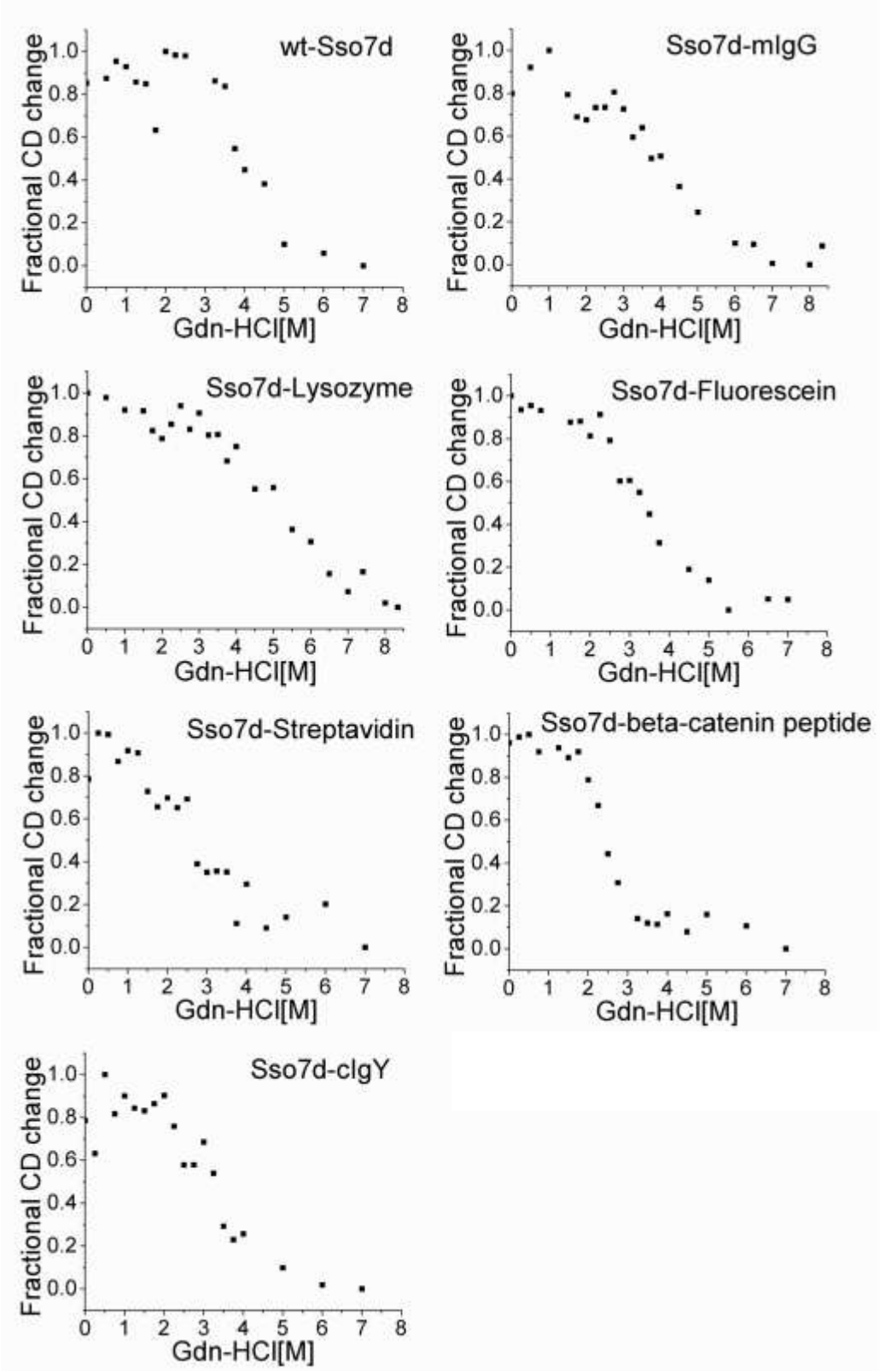



**Figure 3.4**: CD spectra of wild-type Sso7d and six different Sso7d mutants at pH 6.0. Wild-type Sso7d and Sso7d mutants were incubated for 12-18 hours at 20°C in 20 mM sodium phosphate buffer (pH 6.0) and subsequently CD spectra were recorded. Each spectrum was normalized by the molar ellipticity value at 222 nm. The CD spectra for the wild-type Sso7d and the Sso7d mutants are similar and indicate the presence of secondary structure.

**Figure 3.5:** Protein denaturation in the presence of Guanidine-HCl. Wild-type Sso7d and Sso7d mutants were incubated in 20 mM sodium phosphate buffer with varying concentrations of Gdn-HCl for 12-16 hours at 4°C, following which CD spectra were recorded. The fraction of folded protein was estimated as $f = \frac{\theta - \theta_{max}}{\theta_{min} - \theta_{max}}$, where θ is the baseline corrected observed molar ellipticity at 222 nm, and $\theta_{max}$ and $\theta_{min}$ are the maximum and minimum values of molar ellipticity for a given protein respectively. The fraction of folded protein is plotted as a function of Gdn-HCl concentration.


wt-Sso7d
Sso7d-mIgG
Sso7d-Lysozyme
Sso7d-Fluorescein
Sso7d-Streptavidin
Sso7d-beta-catenin peptide
Sso7d-clgY
Fractional CD change
Gdn-HCl[M]

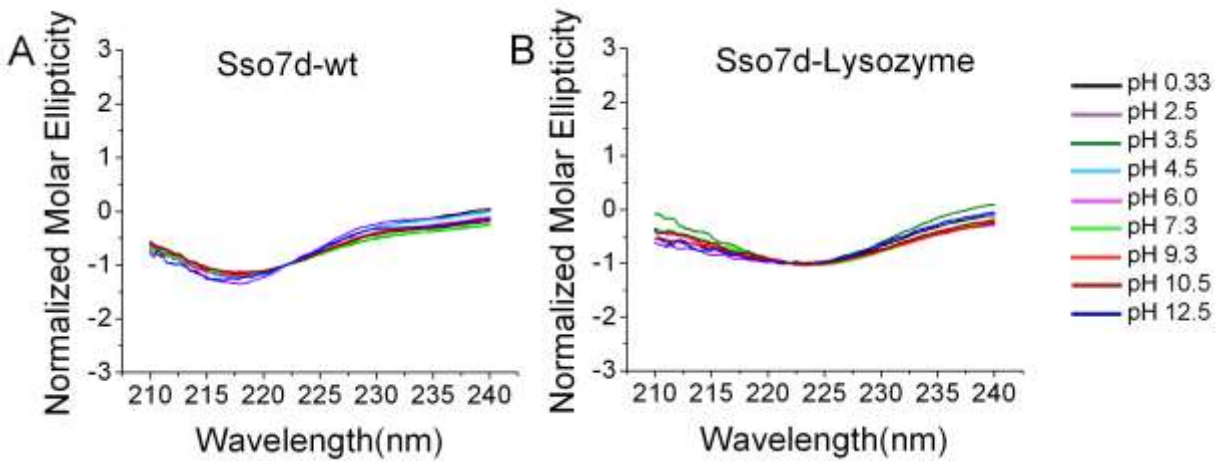
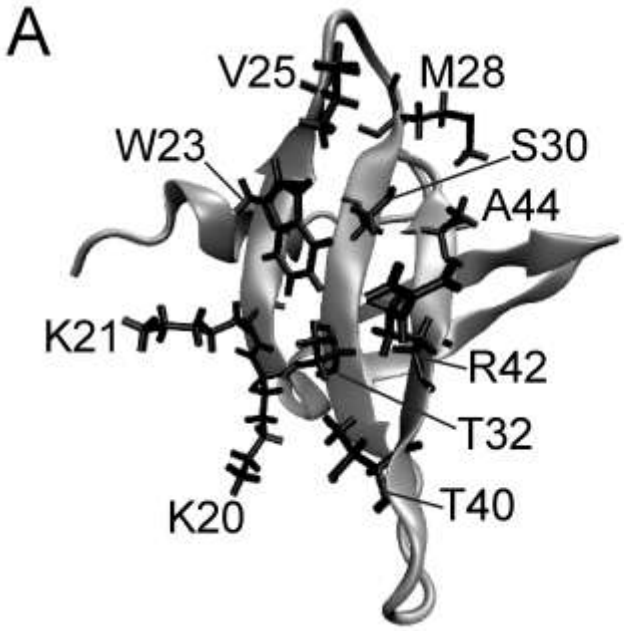



**Figure 3.6:** pH stability analysis of **(A)** wild type Sso7d and **(B)** Sso7d-Lysozyme. Protein samples were incubated in 20 mM sodium phosphate buffer under varying pH conditions for 12-18 hours at 20°C and subsequently CD spectra in the 210-240 nm wavelength range was measured. Each CD spectrum was normalized by the molar ellipticity value at 222 nm. As with the wild-type Sso7d, normalized CD spectra for Sso7d-Lysozyme are essentially identical, suggesting no appreciable change in secondary structure over a wide range of pH.




B

| | | |
|---|---|---|
| W25V<br>- | | C28M<br>+ |
| F23W<br>- | C30S<br>- | C44A<br>- |
| F21K<br>- | D32T<br>- | S42R<br>- |
| C20K<br>+++ | | Q40T<br>++ |

**Figure 3.7:** Mutational analysis for Sso7d-Lysozyme. Each mutated residue in Sso7d-Lysozyme was reverted back to the corresponding residue in wild type Sso7d. **(A)** Sso7d scaffold with randomized residues shown in licorice representation. (PDB accession code 1SSO; this image was generated using Visual Molecular Dynamics (VMD) software) **(B)** Mutational Analysis grid showing the effect of reversion mutations; lysozyme binding of each reversion mutant was compared to Sso7d-Lysozyme and each mutant was classified into one of four categories as follows: binding similar to Sso7d-Lysozyme (+++), binding somewhat reduced (++), binding greatly reduced (+) and binding completely abolished in the given concentration range (-). Shades of gray are used to indicate importance of residue for maintaining binding to lysozyme, a darker shade of gray indicates greater contribution to binding.

**Table 3.1:** Protein sequences of the Sso7d mutants. Corresponding sequence for the wild-type protein is shown as a reference. Mutated positions are bold-faced and underlined.

| | 20.....30.....40.... |
|---|---|
| **wt-Sso7d** | **KK**V**W**R**V**GK**M**I**S**F**T**YDLGGGK**T**G**R**G**A**V |
| **Target** | **Sso7d Mutant Sequence** |
| **mIgG** | |
| Sso7d-mIgG[a] (4)[b] | **YL**V**L**R**L**GK**F**I**Y**F**Y**YDLGGGK**L**G**L**G**H**V |
| Sso7d-mIgG-21 | **LY**V**K**R**A**GK**Y**I**T**F**W**YDLGGGK**L**G**S**G**F**V |
| **Streptavidin** | |
| Sso7d-Streptavidin[a] (5)[b] | **WI**V**A**R**D**GK**W**I**D**F**S**YDLGGGK**S**G**I**G**Y**V |
| **cIgY** | |
| Sso7d-cIgY[a] (5)[b] | **DG**V**S**R**M**GK**L**I**V**F**T**YDLGGGK**R**G**I**G**I**V |
| **Fluorescein** | |
| Sso7d-Fluorescein[a] (5)[b] | **FR**V**I**R**S**GK**A**I**R**F**L**YDLGGGK**F**G**Y**G**V**V |
| **Beta Catenin peptide** | |
| Sso7d-beta-catenin-peptide[a] | **GV**V**G**R**C**GK**V**I**V**F**P**YDLGGGK**L**G**L**G**T**V |
| Sso7d-beta-catpep-1 | **KS**V**L**R**T**GK**I**I**I**F**P**YDLGGGK**L**G**I**G**I**V |
| Sso7d-beta-catpep-6 | **AI**V**A**R**F**GK**Q**I**L**F**P**YDLGGGK**L**G**I**G**I**V |
| Sso7d-beta-catpep-9 | **SW**V**S**R**Y**GK**A**I**L**F**L**YDLGGGK**M**G**R**G**V**V |
| Sso7d-beta-catpep-10 | **KC**V**H**R**C**GK**A**I**F**F**L**YDLGGGK**F**G**A**G**L**V |
| **Lysozyme** | |
| Sso7d-Lysozyme[a] | **CF**V**F**R**W**GK**C**I**C**F**D**YDLGGGK**Q**G**S**G**C**V |
| Sso7d-Lysozyme-12 | **FL**V**N**R**G**GK**W**I**F**F**L**YDLGGGK**E**G**E**G**N**V |
| Sso7d-Lysozyme-13 | **FL**V**Y**R**W**GK**S**I**E**F**E**YDLGGGK**L**G**A**G**L**V |
| Sso7d-Lysozyme-18 | **SL**V**L**R**H**GK**L**I**L**F**I**YDLGGGK**L**G**G**G**E**V |
| Sso7d-Lysozyme-20 | **AY**V**N**R**S**GK**F**I**L**F**S**YDLGGGK**F**G**T**G**E**V |

[a]These proteins were chosen for further characterization

[b]For each target, five individual clones after the final cytometric sort were sequenced. The number in parentheses indicates the number of identical DNA sequences obtained.

**Table 3.2:** Estimates of $K_D$ for the Sso7d mutants. Yeast cells displaying the Sso7d mutants were labeled with biotinylated target followed by streptavidin-phycoerythrin (Strep-PE), or Strep-PE alone in case of Sso7d-Streptavidin, and analyzed using flow cytometry. For each mutant, data from three different experiments was fit to a monovalent binding isotherm. $K_D$ values were obtained from a global fit and 68% confidence intervals were calculated [50].

| **Sso7d mutant** | **Target Size (kDa)** | **$K_D$ (nM)** | **68% confidence interval(nM)** | **[a]Half maximal $K_D$ (nM)** |
|---|---|---|---|---|
| Sso7d-Fluorescein | 0.390 | 7569 | 3700-20500 | - |
| Sso7d-beta-catenin peptide | 1.36 | 2853 | 1800-4500 | - |
| Sso7d-Lysozyme | 14.3 | 349 | 225-540 | - |
| Sso7d-Streptavidin | 52.8 | 12 | 7-20 | 17.7±1.5 |
| Sso7d-mIgG | 150 | 26[b] | 16-41 | 22.5±2.5 |
| Sso7d-cIgY | 150 | 30[b] | 18-49 | 26.7 ±4.4 |

[a] The half-maximal $K_D$ is the target concentration at which the fluorescence is half of the maximum value of experimentally observed fluorescence. The mean value determined from three different experiments is reported. Error bars indicate the standard error of the mean.

**[b]** In case of labeling with mIgG and cIgY, fluorescence signal decreased at high target concentrations (hook effect). Such a decrease has been previously reported in case of yeast cell surface titrations[37]. In these cases, fluorescence values at higher concentrations were not included in calculations to estimate $K_D$.

**Table 3.3:** Biophysical characterization of Sso7d mutants and comparison with wild-type Sso7d

| **Sso7d Variant** | **Tm(°C)[a]** | **[Gdn-HCl]$_{1/2}$(M)[c]** | **pH stability range[d]** |
|---|---|---|---|
| Sso7d-wt | 98[b] | 4 | 0.33-12.5 |
| Sso7d-Fluorescein | 89.6 (89.5,89.6) | 3.4 | 0.33-12.5 |
| Sso7d-beta-catenin peptide | 92.7 (92.5,92.8) | 2.5 | 0.33-12.5 |
| Sso7d-Lysozyme | 92.7 (92.6,92.7) | 5 | 0.33-12.5 |
| Sso7d-Streptavidin | 72.5 (72.4,72.5) | 2.8 | 0.33-12.5 |
| Sso7d-mIgG | 102.8 (102.8,102.8) | 4 | 0.33-12.5 |
| Sso7d-cIgY | 100 (100.3,99.7) | 3.2 | 0.33-12.5 |

[a]Average value from duplicate experiments is reported. Numbers in parentheses show experimentally determined $T_m$ in each experiment.

[b]$T_m$ for wild-type Sso7d reported in literature[18]

[c]$[GdnHCl]_{1/2}$ calculated from Figure 5

[d]This is the range of pH over which the CD spectrum remains largely unchanged and indicates presence of secondary structure.

**Table 3.4:** Comparison of melting temperatures of Sso7d mutants with those of mutants obtained from other non-immunoglobulin scaffolds.

| **Scaffold** | **Size(aa)** | **$T_m$- Wild type(°C)** | **$T_m$- Mutants (°C)** |
|---|---|---|---|
| Affibody[9] | 58 | 78 | 40 -57 |
| $10^{th}$ domain of Fibronectin[5, 52, 53] | 94 | 84 | 56 -74 |
| DARPins[7, 54] | 130-190 | -[a] | 66-100 |
| Lipocalin (bilin-binding protein)[55, 56] | <180 | 61 | 62-73 |
| Sso7d(this study) | 60 | 98 | 72-103 |

[a]Melting temperature depends on the number of ankyrin repeats

# CHAPTER 4

# A SUPER LIBRARY OF HYPERTHERMOPHILIC PROTEIN SCAFFOLDS FOR ENGINEERING MOLECULAR RECOGNITION

Mahmud Hussain*, Nimish Gera*, Andrew B. Hill and Balaji M. Rao

Department of Chemical and Biomolecular Engineering

North Carolina State University

Raleigh, NC 27695-7905

***These authors contributed equally**

## 4.1. Introduction

Highly specific molecular recognition is of great importance for a broad range of applications. In particular, antibodies are commonly used as binding proteins in research, biotechnology and medicine. The immunoglobulin framework of antibodies can give rise to highly specific binding proteins for a wide spectrum of targets through mutagenesis of hypervariable loop regions called the Complementarity Determining Regions (CDRs). However, due to their large multi-subunit structure with multiple disulfide bonds, antibodies have relatively low thermal and chemical stability. They cannot be easily produced in common bacterial expression systems, resulting in relatively high cost of production. Further, the generation of antibodies for new targets through the frequently used procedure of immunizing animals is a lengthy and tedious process, with no guarantees of success.

Analogous to generation of antibodies through mutagenesis of CDRs, binding proteins can also be generated by altering the amino acid composition of certain regions in non-immunoglobulin scaffold proteins. Indeed, the generation of binding proteins from over 50 different protein scaffolds has been previously reported [1, 2]; the more widely cited among these alternate scaffolds include the $10^{th}$ domain of fibronectin[3, 4], DARPins [5, 6]and affibodies[7, 8]. Binding proteins based on alternate scaffolds can offer several potential advantages over antibodies; these include low molecular weight, high stability and easy recombinant expression at high yields in bacterial systems. In spite of these significant advantages, antibodies remain the most widely used proteins for molecular recognition applications. Indeed, despite being poorly characterized and potential lot-to-lot variability, polyclonal antibodies generated by immunizing animals are arguably the most common

affinity reagents for research applications. This is, in large part, due to historical reasons; highly specific antibodies for a diverse set of targets including proteins, peptides, post-translational modifications in proteins such as phosphorylation and haptens have been generated, highlighting the remarkable plasticity of the immunoglobulin framework. On the other hand, fewer data exist on binding proteins from specific alternate scaffolds. It is plausible that randomization of a small scaffold protein may not provide the topological diversity required to reliably obtain binders to all targets. Further, screening strategies used to isolate binders from a combinatorial library of scaffold mutants sample a small fraction of the theoretical diversity arising from randomization of the scaffold. For instance, randomization of 12 residues on a scaffold results in a theoretical diversity of $10^{15}$, while the largest combinatorial libraries have a diversity of $10^{12\text{-}13}$ mutants.

The likelihood of isolating binders through randomization of an alternate scaffold protein is greatest when the resulting library has few misfolded proteins due to deleterious mutations, and includes large topological diversity as obtained by altering the amino acid composition of CDRs on antibodies. Highly stable proteins are more likely to be tolerant to mutation and retain their native structure. Consequently, they are highly evolvable and are excellent candidates for use as scaffolds[9]. Indeed, we have previously shown that highly stable binding proteins for a wide spectrum of targets can be isolated from a modestly-sized library (~ $10^8$ mutants) generated through mutagenesis of the Sso7d protein from the hyperthermophilic archaeon *Sulfolobus solfataricus*[10]. Notably, Sso7d-based proteins have high thermal stability, are resistant to chemical denaturation and can withstand pH extremes. Here, we hypothesize that a library generated through randomization of multiple stable

scaffolds will result in greater topological diversity and is, therefore, more likely to yield binders to a given target than a library derived from a single scaffold. To assess this hypothesis, we generated a combinatorial "super-library" of scaffold proteins or a "library-of-libraries", by randomizing secondary structure elements of seven different proteins of hyperthermophilic origin. Based on structural data, 10-15 surface accessible residues on each scaffold were randomized to generate a super-library of ~ $4x10^8$ mutants. Subsequently, we isolated binding proteins for five different model targets – including a small organic molecule, linear and cyclic peptides, and proteins – from this library. Binding proteins from five out of the seven scaffolds in the super-library were isolated in our screen. Interestingly, the pool of highest affinity binders for each target contained proteins derived from a distinct subset of scaffolds, suggesting that specific scaffolds may be optimal for generating binders to a given target. Thus, scaffold diversity may be advantageous from the perspective of designing a combinatorial library suitable for isolating binders to a wide spectrum of targets. To further explore the role of scaffold diversity, we compared the affinities of binding proteins isolated from the super-library with scaffold diversity, and a library with significantly higher sequence diversity, but derived from a single scaffold – the Sso7d protein that has been shown to be a highly versatile scaffold. Strikingly, our results show that binders to two different targets isolated from the super-library of $4x10^8$ mutants have higher target-binding affinities than those isolated from a library of ~ $5 \times 10^{11}$ Sso7d mutants. Taken together, our results show the importance of scaffold diversity in the design of combinatorial libraries from alternate scaffolds. Our results also highlight the effectiveness of hyperthermophilic proteins as alternate scaffolds for generating highly specific binding

proteins. Mutant proteins derived from multiple scaffolds described in this paper have several desirable properties such as low molecular weight and high stability. We expect that binding proteins from these scaffolds will be useful for applications in biotechnology and medicine.

## 4.2. Results

### 4.2.1. Construction of a super-library of alternate scaffolds using yeast surface display

We hypothesized that randomization of secondary structure elements on multiple, scaffold proteins of hyperthermophilic origin will result in a combinatorial library with increased topological diversity. Further, due to their high thermal stability, the scaffold proteins are more likely to be tolerant to mutation; consequently, mutant proteins in the library are less likely to be misfolded. To test this hypothesis, we constructed a combinatorial super-library that consists of component libraries obtained by randomizing specific regions on seven different proteins – TM0487 and TM1112 from *Thermotoga maritima*, Sso7d, Sso6901and the Microtubule Interacting and Transport (MIT) domain of Sso0909, from *Sulfolobus solfataricus*, the chitin binding domain (ChBD) of chitinase A from *Pyrococcus furiosus* and the N-terminal domain of the Ph1500 protein from *Pyrococcus horikoshii*. Candidate proteins from hyperthermophilic organisms were chosen based on two criteria – a known NMR or crystal structure that allows careful selection of surface-accessible residues for mutagenesis, and overall length of ~ 100 amino acids or less. The seemingly arbitrary restriction on the overall length of scaffold proteins was necessary to ensure that all individual scaffold libraries could be generated using one or two oligonucleotides containing

degenerate codons. Six of the seven proteins chosen – all proteins except Sso7d – have not been used as scaffolds for generating binding proteins prior to this study; we have previously isolated binding proteins for a wide spectrum of target species from a library of $10^8$ Sso7d mutants [10].

Combinatorial libraries of scaffold proteins were generated by randomizing 10-15 residues on each scaffold, as shown in Figure 4.1 and Table 4.1. In general, surface-accessible residues were chosen for randomization; where available, additional data in the literature were used to guide the choice of residues. Construction of the Sso7d library has been previously described [10]. Like Sso7d, the Sso6901 protein contains an SH3-like fold and has intrinsic DNA binding activity[11-13]. 12 amino acids in the putative DNA binding interface of Sso6901 were chosen for randomization. Along similar lines, 15 residues on the chitin binding interface of ChBD – including the three tryptophan residues W274, W308 and W326 (numbering as in chitinase A) – were randomized[14-16]. The MIT domain from Sso0909 contains a putative binding region in the helices α2 and α3; 12 residues in this region were selected for mutagenesis[17]. Thirteen residues on the MIT domain from Sso0909 – A18, D22, A30, Y34, A37, I38, L41, I45, Y57, I61, Y64, R67 and L71 – are highly conserved across MIT domains from multiple species. Interestingly, these include D22 and R67 that are though to stabilize the structure of the MIT domain through formation of a salt bridge[17]. These conserved residues were excluded from randomization. Finally, information on specific surface accessible residues that might be involved in binding or important for protein stability was not available in case of TM1112, TM0487 and Ph1500. The TM1112 protein containing seven β-strands has significant sequence similarity to the

cupin family[18], while the TM0487 protein belongs to the Domain of Unknown Function (DUF59) family of proteins[19]. Cysteine residues in both these proteins are not involved in disulfide bond formation and were mutated to serines.

Yeast display libraries for individual scaffolds were generated using homologous recombination mediated plasmid gap repair during yeast transformation[20]; a single DNA fragment or two overlapping DNA fragments encoding the mutant libraries were used in the transformation step. Based on the number of yeast transformants obtained, library diversity was estimated as $10^7$-$10^8$ mutants for each scaffold, as shown in Table 4.1. Subsequently, individual libraries were combined to obtain a super-library with an overall diversity of ~ 4.4 x $10^8$ mutants. All mutant proteins are expressed as a C-terminal fusion to the yeast cell wall protein Aga2, and are flanked by an N-terminal HA and C-terminal c-myc epitope tags respectively. Expression of the c-myc epitope tag confirms expression of the full-length mutant protein as a yeast cell surface fusion. Indirect immunofluorescent labeling of the c-myc epitope tag shows that 20-43% of yeast cells in the individual libraries express full-length fusions; overall, 28% of yeast cells in the super-library express full-length scaffold variants, as shown in Figure 4.2.

### 4.2.2. Isolation of binders to model targets

We sought to investigate if the use of multiple scaffolds in a combinatorial library was indeed advantageous by screening the super-library for binders to multiple targets – if the pool of binders with the highest affinity was derived from distinct scaffolds for different targets, then such a result would indicate the importance of topological diversity arising from

the use of multiple scaffolds. In this context, it is worthwhile to note that Sso7d variants constitute the largest fraction of the super-library (~ 25%). The Sso7d protein has been previously shown to be a versatile scaffold for generating binders to a wide spectrum of targets[10], and serves as a rigorous internal control to assess the advantage of using multiple scaffolds while generating a combinatorial library.

We screened the super-library to isolate binders for five different model targets; our selected targets included a small organic molecule (fluorescein), a 12 amino acid peptide from the C-terminus of the β-catenin protein (β-catenin peptide), Brain Natriuretic Peptide-32 (BNP-32) – a 32 amino acid cyclic peptide containing a disulfide bridge, hen egg lysozyme (HEL), and immunoglobulin G from rabbit (rIgG). The chosen targets represent a wide spectrum of target species with vastly different molecular weights. Further, the specificity of rIgG binders could be rigorously assessed by evaluating their binding to other closely related immunoglobulins. Previously, we have obtained highly stable binding proteins for fluorescein, β-catenin peptide and HEL from a library of Sso7d mutants[10].

A two step procedure – using magnetic selection followed by flow cytometry, as previously described[10] – was used to isolate binding proteins from the super-library. Briefly, yeast cells were incubated with micron-sized magnetic beads that had been coated with the target species and bead-bound cells were isolated using a magnet. Subsequently, a pool of mutants with the highest affinity for the targets was isolated using multiple rounds of Fluorescence Activated Cell Sorting (FACS); for each target, we isolated and sequenced plasmid DNA from ten individual yeast clones. Our results are shown in Table 4.2. The pool of mutants with highest binding affinity for the five targets included proteins derived from

five of the seven scaffolds used to construct the super-library; binding proteins based on the Sso7d, Sso6901, MIT, TM1112 and ChBD scaffolds were isolated. No mutants derived from the Tm0487 and Ph1500 scaffolds were found among mutants sequenced. Binding proteins from multiple scaffolds were obtained in the pool of highest affinity binders for all targets, with the exception of HEL.

A single distinct clone was obtained in the pool of HEL binders. Interestingly, the Sso7d mutant isolated from the super-library is identical to the mutant that was previously isolated by screening the Sso7d library[10]. The pool of binders for fluorescein contained three distinct clones – two from Sso7d and one from Sso6901 respectively. Two distinct mutants – one each from Sso7d and Sso6901 – were obtained in case of binders to the β-catenin peptide fragment; the pool of binders to rIgG comprised two distinct clones – one each from MIT and ChBD. During FACS, yeast cells were simultaneously labeled with the target, and an antibody against the c-myc epitope tag or the HA epitope tag. Interestingly, the pool of binders for BNP-32 from sorts using the anti-HA antibody contained a truncated TM1112 mutant. Subsequently, we sorted the super-library to isolate binders for BNP-32 using the anti-c-myc antibody; in this case, we obtained one distinct clone from each Sso7d and Sso6901.

Our results show that the pool of binders with highest affinity for each target is derived from distinct subsets of scaffolds. For instance, binders to rIgG are based on the MIT and ChBD scaffolds, while fluorescein binders are derived from Sso7d and Sso6901. This is particularly noteworthy since the Sso7d mutants constitute the largest fraction of the super-

library. Taken together, our results strongly underscore the advantage of using multiple scaffolds in generating a combinatorial library.

### 4.2.3. Characterization of binding affinity and specificity

We measured the equilibrium dissociation constants ($K_D$) of the binding interaction between mutant proteins and their targets for a selected subset of mutant proteins, using yeast cell surface titrations[10]. Note that multiple previous reports have shown the consistency between $K_D$ values estimated using yeast surface titrations and those obtained using soluble proteins[21]. Briefly, yeast cells displaying mutant proteins were incubated with varying concentrations of target protein, and the fraction of cell surface bound fusions was measured using flow cytometry. Data from at least three different experiments for each mutant was combined and a global fit to a one-step binding isotherm was used to estimate the $K_D$. The estimated binding affinities and their associated 68% confidence intervals – analogous to the commonly reported standard deviations from triplicate measurements – are shown in Table 4.3. Since rIgG is a dimeric molecule and our estimates for $K_D$ are based on the assumption of a monovalent binding isotherm, $K_D$ estimates from yeast surface titrations may be influenced by the avidity effect in case of rIgG; a single rIgG molecule may bind to two cell surface fusions. Nevertheless, it can be shown that the concentration at which half the cell surface fusions are bound to a divalent target – the half-maximal binding concentration – is equal to the $K_D$[22]. We estimated the $K_D$ for rIgG-binding mutants using the half-maximal binding concentration determined from cell surface titration plots. The $K_D$ values obtained in this manner were found to be consistent with those estimated using the monovalent binding

isotherm. Our analysis shows that the binding affinities of mutants obtained from the super-library are in the 100nM-micromolar range; these values are consistent with those typically obtained from naïve combinatorial libraries, without an affinity maturation step.

Binding of mutant proteins to the secondary reagents used for flow cytometric analysis was insignificant. We further analyzed the specificity of MIT-rIgG by assessing its binding to immunoglobulins from species other than rabbit. Briefly, flow cytometry was used to detect the binding of yeast cells displaying MIT-rIgG to its cognate target, rIgG, or immunoglobulins from mouse, goat, chicken and donkey. As shown in Figure 4.3, MIT-rIgG shows negligible binding to non-target immunoglobulins. We attribute the high specificity of MIT-rIgG to the stringent negative selection steps employed during magnetic selection; yeast cells binding to streptavidin magnetic beads and beads coated with non-target immunoglobulins were rejected. The highly avid interaction between yeast cells displaying multiple cell surface fusions and the beads ensures that cells expressing mutants that bind – even with low affinity – to non-target species are eliminated from consideration[23].

### 4.2.4. Sequence diversity vs. "scaffold diversity" in combinatorial libraries

Our results suggested that the use of multiple different scaffolds while constructing a combinatorial library is advantageous. In particular, our data suggests that a particular scaffold may be more suited than others for generating binding proteins to a given target. Indeed, the pool of highest affinity binders for rIgG contained a ChBD mutant but no Sso7d or Sso6901 mutants, even though the ChBD library has significantly lower diversity than the Sso7d and Sso6901 libraries. However, the overall diversity of a yeast surface display library

is restricted by the transformation efficiency in yeast; the highest reported diversities are in the order of ~ $10^9$ mutants, whereas the theoretical diversity generated by randomizing 10 residues is ~ $10^{13}$. More pertinently, the component of the super-library with the highest diversity – the Sso7d library – has a diversity of only ~ $10^8$ mutants. The limited diversity of our individual scaffold libraries raises the following question: is the ostensible optimality of a particular scaffold or subset of scaffolds for generating binders to a given target simply an artifact of library size limitation? In other words, can a library with very high sequence diversity but generated from a single versatile scaffold such as Sso7d, yield binders to multiple targets with higher affinities than those obtained from the super-library with lower diversity but derived from multiple scaffolds?

To address this question, we sought to generate, and subsequently screen, an Sso7d library with high sequence diversity. *In vitro* combinatorial screening tools such as mRNA display[24, 25] and ribosome display[26, 27] can be used to screen libraries consisting of ~ $10^{13}$ mutants. Yet, the yeast display system provides the distinct advantages of quantitative screening and subsequent characterization of binding affinities [28-30] or thermal stability[31, 32] using flow cytometry. Therefore, we developed a protocol that combined mRNA display and yeast surface display to exploit the advantages of each method. Our overall strategy is schematically illustrated in Figure 4.4. Briefly, we generated an Sso7d library with an estimated diversity of ~ $5 \times 10^{11}$ mutants, in the mRNA display format. Mutants that bound to target-coated magnetic beads were isolated, amplified by PCR and subsequently transformed into yeast. This pool of mutants, in yeast display format, was further screened using FACS.

Using this method, we screened a library of ~ 5 x $10^{11}$ Sso7d mutants to isolate binding proteins for two targets: rIgG and BNP-32; the pool of binders with highest affinity for these targets, isolated from the super-library, did not contain Sso7d mutants. For each target, plasmid DNA was sequenced from 6 clones in the pool of mutants obtained after multiple rounds of FACS. Our results are shown in Table 4.2. Six and three distinct clones were obtained in case of binders to BNP-32 and rIgG respectively. Subsequently, we estimated binding affinities ($K_D$) for one mutant binding each rIgG and BNP-32, using yeast cell surface titrations as described earlier. $K_D$ for Sso7d-rIgG was also estimated using the half-maximal binding concentration. These values are reported in Table 4.3. Strikingly, binding affinities of mutants isolated from a 5 x $10^{11}$ library of Sso7d mutants are lower, or at best comparable, to those obtained from the super-library with overall diversity lower by three orders of magnitude (~ 4 x $10^8$). These results further underscore the advantage of "scaffold-diversity" in a combinatorial library.

### 4.2.5. Biophysical characterization of mutant proteins

Binding proteins for five different targets, derived from five different scaffolds were obtained from the super-library and the high diversity Sso7d library. For a subset of mutants, we carried out further experiments to investigate the following properties: recombinant expression in *E. coli*, functionality of soluble protein with respect to target binding, retention of secondary structure, and thermal stability.

Wild-type proteins and mutants were recombinantly expressed in the *E. coli* cytoplasm with C- or N-terminal 6xHistidine (6xHis) tag. Subsequently, all proteins were purified in a

single step using a nickel column. Where data was available, the choice of using an N- or C-terminal 6xHis tag was based on previous reports of expression of the wild-type protein[10, 17, 18]. Wild-type proteins and mutants from MIT, Sso6901 and TM1112 were expressed with an N-terminal 6xHis tag. All Sso7d mutants were expressed with a C-terminal 6xHis tag, as previously described[10]. Based on bicinchoninic acid (BCA) assay measurements using Bovine Serum Albumin (BSA) as a standard, purified protein yields were estimated in the range of 5-25 mg/L of unoptimized, shake flask culture. The one notable exception was, not surprisingly, the truncated TM1112 mutant that binds BNP-32. Yields for TM1112-BNP-32 were low and the protein was prone to aggregation. The highest protein yields were obtained in case of Sso7d and Sso6901 mutants. Note that these yields are conservative estimates since only those elution fractions containing pure protein, as determined by SDS-PAGE analysis, were considered; for reasons of expediency, fractions containing contaminating proteins were rejected.

To validate that the recombinantly expressed proteins retain binding to their targets, we carried out competition assays wherein yeast cells expressing mutant proteins were labeled with their cognate target in the presence or absence of a large excess of soluble mutant protein. Flow cytometry was used to detect binding of cell surface fusions to the target. As shown in Figure 4.5, there is a significant decrease in fluorescent signal due to cell surface bound target, in the presence of soluble mutant proteins. These data confirm that the recombinantly expressed mutant proteins are functional. We also used Circular Dichroism (CD) spectroscopy to compare the secondary structure of Sso6901 and MIT wild-type and mutant proteins. As shown in Figure 4.6, the CD spectra for wild-type and mutant proteins

are reasonably similar over the range of wavelengths from 210 to 240 nm. All Sso6901 mutants show a characteristic peak at ~ 230 nm in their CD spectra that can be attributed to the presence of aromatic residues in wild-type and mutant Sso6901 proteins[33-35]. In previous studies, the CD spectra of several Sso7d mutants have been shown to be similar to that of the wild-type protein[10]. The TM1112 mutant was not analyzed using CD spectroscopy due to poor recombinant expression and its aggregation-prone nature. Along similar lines, the ChBD mutant was not expressed solubly due to its low thermal stability (see below).

To assess the thermal stability of proteins derived from different scaffolds in the super-library, we used a combination of Differential Scanning Calorimetry (DSC) and thermal denaturation studies on yeast cell surface displayed proteins. DSC was used to measure the melting temperatures ($T_m$) for the wild-type proteins TM1112, Sso6901 and MIT, as well as selected mutants derived from Sso7d, Sso6901 and MIT. For ChBD-rIgG and TM1112-BNP-32, we used yeast cell surface displayed protein to determine the temperature of half-maximal irreversible thermal denaturation ($T_{1/2}$), as previously described[36]. Briefly, yeast cells expressing cell surface protein are incubated at different temperatures. Subsequently, the fraction of cell surface fusions that retain binding to the target is determined using flow cytometry. Progressive loss of binding to the target at higher temperatures can be attributed to the irreversible thermal denaturation of cell surface fusions. The yeast display system enables the evaluation of thermal stability without recombinant protein expression. Pertinently, DSC experiments on TM1112-BNP-32 were not feasible due to the low protein yield and aggregation-prone nature of the protein under the buffer

conditions tested. Also, we chose to use yeast displayed protein to assess the thermal stability of ChBD-rIgG due to amino acid insertion in a β-strand; we anticipated – as indeed confirmed by experiment – that this may lead to loss of protein stability.

Table 4.4 shows the $T_m$ and $T_{1/2}$ values for the wild-type and mutant proteins; the thermal denaturation curves for ChBD-rIgG and TM1112-BNP-32 are shown in Figure 4.7. Mutant proteins derived from the Sso7d, Sso6901 and MIT scaffolds show high thermal stability ($T_m$ = 74-93 °C). On the other hand, the TM1112 and ChBD mutants analyzed have significantly lower thermal stability (50 °C and 38 °C respectively). The low thermal stability of these mutants is not surprising. The deletion of two β-strands in TM1112-BNP-32 and the insertion of an amino acid in a β-strand are the likely reasons for loss of thermal stability for these mutants.

Taken together, our data shows that mutant proteins derived from the super-library are functional when recombinantly expressed and retain their secondary structure. Further, several mutants retain high thermal stability ($T_m$ > 74 °C).

### 4.3. Discussion

We have shown that a combinatorial super-library generated by randomizing secondary structure elements on multiple stable scaffolds of hyperthermophilic origin can yield binders to a broad spectrum of targets. Strikingly, even though a significant fraction of the super-library is composed of a single scaffold – Sso7d mutants make up $10^8$ out of the overall 4.4 x $10^8$ mutants in the super-library – the pool of binders with the highest affinity for each target is derived from a distinct subset of scaffolds. For example, the "best" binders for rIgG were

derived from MIT and ChBD, while binders to BNP-32 were derived from TM1112, Sso7d and Sso6901 (Table 4.2). Our results suggest that specific scaffolds may be more suited than others to generate binders for a given target. Therefore, by inference, inclusion of multiple scaffolds in a combinatorial super-library increases the likelihood of generating binders to a given target.

In generating our super-library, we chose to randomize 10-15 amino acid residues on β-strands or α-helices of various scaffold proteins; the basic flat surface topology of the randomized region on the scaffold was maintained (Figure 4.1). Arguably, inclusion of additional scaffolds where loop regions are mutagenized might enhance the binding capabilities of the super-library. Nevertheless, it is interesting to note that the possibly subtle variations in topology and/or amino acid composition of the scaffold framework may favor a particular scaffold in the context of generating binding proteins to a given target. In some ways, this is analogous to antibodies for two different targets arising from two different germ lines. Parallels can also be drawn to the introduction of loop length diversity in CDRs of antibodies[37-40] or the tenth domain of fibronectin to generate high affinity binders[36, 41].

It is reasonable to assume that the likelihood of finding a binder in a combinatorial library increases if the number of correctly folded mutants in the library is high. Indeed, a computationally designed library, wherein mutations that destabilize the scaffold were eliminated, significantly outperforms a library obtained through random mutagenesis in its ability to generate high affinity binders[42]. Proteins with high thermal stability are more tolerant to a wider range of mutations[9]. Therefore, we hypothesized that a library generated through mutagenesis of an ensemble of hyperthermophilic scaffolds will have higher

topological diversity, as well as a high fraction of correctly folded mutants. In this context, it is instructive to examine the fraction of full-length yeast display fusions in the individual libraries (Figure 4.2). The ChBD library was constructed using oligonucleotides with degenerate trimer phosphoramidites coding for twenty amino acids; unlike the NNN codons, all stop codons are eliminated. Despite this, only 43% of this ChBD library is comprised of full-length ChBD variants, as detected by immunofluorescent labeling of the c-myc epitope tag. In contrast, 30% of the Sso7d library expresses full-length fusions; note that sequencing data showed that only 30% of the $10^8$ library consists of full-length Sso7d variants due to stop codons introduced by the degenerate NNN codons use for oligonucleotide synthesis[10]. Thus, a larger fraction of Sso7d mutants are expressed as yeast cell surface fusions. It has been previously proposed that the quality control mechanism in the yeast endoplasmic reticulum allows cell surface expression of only correctly folded proteins, although there is some evidence of expression of misfolded proteins or molten globules[43]. Therefore, it is likely that a larger fraction of mutants in the Sso7d library are correctly folded relative to ChBD library. Interestingly, an amino acid insertion was inadvertently introduced in a β-sheet in the ChBD scaffold; this might destabilize the ChBD scaffold. Indeed, a ChBD mutant isolated from the super-library had the lowest thermal stability among all proteins analyzed in this study. Taken together, these data support the idea that highly stable scaffold proteins can yield combinatorial libraries containing a larger fraction of correctly folded proteins.

A key limitation of combinatorial approaches is that any library screening method samples only a very small fraction of the sequence space generated by randomizing a

scaffold protein. This is particularly true in case of yeast surface display, where the library sizes range from $10^7 – 10^9$ mutants; individual scaffold libraries in this study have a diversity of $10^8$ or less. Therefore, two different yeast display libraries from the same scaffold may yield different results, due to differences in the mutant sequence space sampled in each library. As discussed earlier, our studies on screening the super-library with multiple targets suggest that a particular scaffold may be more suited than others for generating binding proteins to given target. To confirm that this result was not a consequence of a sampling artifact, and to rigorously assess the benefit of scaffold diversity in a combinatorial library, we sought to directly compare the use of single scaffold versus an ensemble of scaffolds. We screened a library of ~ 5 x $10^{11}$ Sso7d mutants for binders to two targets, BNP-32 and rIgG, in addition to screening the super-library. Interestingly, binding affinities of mutants isolated from the super-library are comparable or higher than those obtained from an Sso7d library with significantly higher diversity (Table 4.3). These results confirm that the incorporation of scaffold diversity in a combinatorial library is advantageous. Clearly, sequence diversity is important; it is intuitive that a scaffold library with very low diversity will be limited in its capability to generate binders. However, on the evidence of our data, it appears that the benefits of a modest increase in overall library size through scaffold diversification outweigh the gains from increasing the sequence diversity of a single scaffold by three orders of magnitude.

Binding proteins from five different scaffolds in the super-library were obtained in our screening experiments. To the best of our knowledge, this is the first demonstration of the use of Sso6901, MIT, TM1112 and ChBD as scaffolds for engineering molecular

recognition, adding to the palate of alternate scaffolds available for generating binding proteins. Mutant proteins derived from scaffolds in the super-library have several desirable properties such as low molecular weight, lack of disulfide bonds and ease of recombinant expression in *E. coli*. Also, several proteins retain high thermal stability (Table 4.4). The high thermal stability of mutant proteins, despite extensive mutagenesis of the wild-type proteins, provides additional support to the idea that hyperthermophilic proteins are likely to be more tolerant to a wide range of mutations. The two exceptions, however, are mutant proteins derived from TM1112 and ChBD. The low thermal stability of TM1112-BNP-32 can be attributed to the deletion of multiple β-strands that likely results in significant disruption of the wild-type structure. In case of ChBD-rIgG, loss of thermal stability is likely due to an inadvertent amino acid insertion in a β-strand during oligonucleotide synthesis; all mutants in the ChBD library carry this insertion. These mutants with low thermal stability underscore an important point: the choice of residues to be randomized on a scaffold is extremely important, even in case of hyperthermophilic scaffolds; mutations that destabilize the scaffold must be avoided. Put differently, stability of mutant proteins derived from a thermostable scaffold will be retained if the mutated regions are not critical contributors to protein stability. Given their small size, deletion of a portion of the scaffold is also likely to cause a loss of stability, as in the case of TM1112-BNP-32. Similar results are also seen in case of Sso7d, where deletion of the C-terminal α helix leads to a decrease in melting temperature by 46 °C[44].

In summary, we have shown that an ensemble of hyperthermophilic scaffolds can be used to generate binders for a wide spectrum of targets. A combinatorial library-of-

libraries derived from hyperthermophilic proteins, such as our super-library, is likely to have two key advantages – increased topological diversity and a high fraction of correctly folded mutants – that result in an increased likelihood of isolating stable binders to any target. Despite extensive research on alternate scaffolds, antibodies remain the most widely used molecules for molecular recognition. This is particularly true in case of affinity reagents for research applications, where the use of poorly characterized polyclonal antibodies is largely the norm. The ability to reliably generate stable and low-cost binding proteins for any target will facilitate the adoption of alternate scaffolds as a matter of routine; the use of a combinatorial library sourced from multiple highly stable scaffolds is a step in this direction. Here, we have explored the use of a super-library derived from seven different scaffolds. However, several other proteins of hyperthermophilic origin with known structures may be attractive candidates for use as alternate scaffolds.

Mutant proteins isolated in our study have binding affinities in the 100 nM-micromolar range, typical of binders isolated from a naïve library of alternate scaffolds. Further increases in affinity may be achieved by additional rounds of mutagenesis and screening. Nevertheless, it is important to note that several applications such as the design of affinity ligands for chromatographic separations do not require binding proteins with high affinity; on the other hand, stable binding proteins are highly desirable. Other potential applications of proteins based on hyperthermophilic scaffolds include their use as affinity reagents in low-cost diagnostics and imaging, and intracellular inhibitors to selectively bind and block the function of proteins, specific domains of proteins, or their post-translational modifications. Hyperthermophilic scaffolds lacking disulfide bonds are

particularly well-suited for use in the context of such intracellular "protein interference". Thus proteins based on hyperthermophilic scaffolds can find wide applicability in research, biotechnology and medicine.

## 4.4. Materials and Methods

### 4.4.1. Construction of yeast surface display libraries

The plasmid vector for yeast surface display (pCTCON) and the yeast strain EBY100 were kind gifts from Prof. K. Dane Wittrup (Massachusetts Institute of Technology, Cambridge, MA). Construction of the yeast surface display library of Sso7d mutants has been previously described[10]. All other libraries were obtained using similar protocols, through homologous recombination mediated plasmid gap repair. Briefly, libraries for Sso6901, TM1112 and MIT scaffolds were synthesized as a single oligonucleotide containing degenerate NNN codons. The oligonucleotides were amplified by PCR with scaffold-specific forward and reverse primers to produce linear DNA fragments containing the yeast surface display consensus sequences. Subsequently, DNA fragments were transformed into yeast, along with linearized pCTCON.

Libraries for ChBD, TM0487 and Ph1500 were synthesized as a pair of oligonucleotides containing degenerate codons. Oligonucleotides for ChBD contained a mixture of trimer phosphoramidites coding for all twenty amino acids at randomized positions; degenerate NNN codons were used in oligonucleotides for TM0487 and Ph1500. Each oligonucleotide was amplified by PCR using suitable forward and reverse primers so as

to generate overlapping DNA fragments. For instance, in case of ChBD, oligonucleotides U1_ChBD and U2_ChBD were amplified with the primer sets P1f_YSD_ChBD/P1r_YSD_ChBD and P2f_YSD_ChBD/P2r_YSD_ChBD respectively, to yield DNA fragments D1 and D2. The 5' end of D1 and 3' end of D2 contain yeast display consensus sequences; 3' end of D1 and 5' end of D2 have sequence homology. Subsequently, the two DNA fragments were transformed into yeast along with the linearized pCTCON vector to construct the yeast display library.

The sequences for all oligonucleotides and primers used are given in Table 4.5. Oligonucleotides contained codons that were optimized for expression in *E. coli.* Primers and oligonucleotides for construction of the ChBD library were purchased from Trilink Biotechnologies (San Diego, CA). All other oligonucleotides and primers were obtained from Integrated DNA Technologies (Coralville, IA). The diversity of each individual scaffold library was estimated based on the number of yeast transformants, as determined by plating serial dilutions on SDCAA plates (20 g/L dextrose, 5 g/L casamino acids, 6.7 g/L yeast nitrogen base, 182 g/L sorbitol, 5.40 g/L $Na_2HPO_4$, 7.45 g/L $NaH_2PO_4$, 15 g/L agar); library diversities are reported in Table 4.1. Individual libraries were pooled together to obtain the super-library, with overall estimated diversity of $4.4 \times 10^8$ mutants.

### 4.4.2. Isolation of binders by magnetic selection

Yeast cells expressing mutant proteins that bind a specific target were isolated using a magnetic selection step, as previously described[10]. Briefly, 100 µl Dynal$^{TM}$ biotin binder magnetic beads (Invitrogen, Carlsbad, CA) were washed with Phosphate Buffered Saline

(PBS) containing 0.1% BSA (PBS-BSA; 8g/L NaCl, 0.2g/L KCl, 1.44 g/L $Na_2HPO_4$ , 0.24 g/L $KH_2PO_4$, pH 7.4 with 0.1% BSA). Subsequently, beads were incubated overnight at 4°C with rotation, with the biotinylated targets – rIgG, BNP-32, fluorescein, β-catenin peptide and HEL – to obtain the target-coated magnetic beads. In parallel, beads were also incubated with biotinylated mouse IgG (mIgG), chicken IgY (cIgY), goat IgG (gIgG) and the Fc portion of human IgG (hFc) for negative selection experiments. All biotinylated immunoglobulin species were obtained from Jackson Immunoresearch (Westgrove, PA). Biotinylated BNP-32 (Bachem Inc., Torrance, CA) was a kind gift from Prof. David C. Muddiman (North Carolina State University, Raleigh, NC). Biotinylated HEL was purchased from Biomeda Corporation (Foster City, CA) and Sigma-Aldrich (St. Louis, MO), fluorescein-biotin from Thermo Scientific (Rockford, IL), and biotinylated β-catenin peptide was obtained from Genscript (Piscataway, NJ).

Negative selections against the biotin binder beads as well as beads coated with mIgG, cIgY, gIgG and hFc were performed as follows. Yeast cells expressing mutant proteins (~5 x $10^9$ cells) were incubated with washed beads in PBS-BSA, in a 2 ml tube for 1 hour at 4°C with rotation. Subsequently, the tube was placed on a magnetic particle concentrator and bead-bound cells were discarded. Unbound cells from this step were screened sequentially by incubation with target-coated magnetic beads in the following order: rIgG, BNP-32, fluorescein, β-catenin peptide and HEL; unbound cells from incubation with one target was incubated with the next target and so on. Bead-bound cells were collected, washed 3-4 times with PBS-BSA and grown in 5 ml SDCAA (20g/L dextrose, 5g/L casamino acids, 6.7g/L yeast nitrogen base, 5.40 g/L $Na_2HPO_4$, 7.45 g/L $NaH_2PO_4$)

with 1:100 pen-strep solution (Invitrogen, Carlsbad, CA) for 48 hours, in a shaker at 30°C and 250 rpm.

### 4.4.3. Isolation of binders by Fluorescence Activated Cell Sorting (FACS)

The pool of binders for each target obtained after magnetic selection was expanded in SDCAA and cell surface protein expression was induced by culturing in SGCAA (20g/L galactose, 5g/L casamino acids, 6.7g/L yeast nitrogen base, 5.40 g/L $Na_2HPO_4$, 7.45 g/L $NaH_2PO_4$) for 24 hours, at a starting cell density of $10^7$ cells/mL, in a 20°C shaker at 250 rpm. FACS was used to isolate a pool of mutants with the highest affinity for each target, using previously described protocols[10]. Briefly, ~ 2 x $10^7$ were labeled simultaneously with an antibody against the HA or the c-myc epitope tags (Roche, Indianapolis, IN, and Invitrogen, Carlsbad, CA respectively), and the biotinylated target species. Subsequently, secondary labeling was carried out with a goat anti-mouse antibody conjugated with Alexa Fluor-488 or -633 for the anti-HA antibody, a goat anti-chicken antibody conjugated with Alexa Fluor-488 or -633 for the anti-c-myc antibody, and streptavidin-phycoerythrin (strep-PE) or neutravidin-FITC (Invitrogen, Carlsbad, CA) for all biotinylated targets. Samples were analyzed and sorted on a BD FACS Aria (Beckton Dickinson Biosciences, San Jose, CA) flow cytometer. For each target, multiple sorts at successively lower concentrations were carried out.

### 4.4.4. Clone sequencing and measurement of $K_D$

Cells from the pool of binders after the final sort were plated on SDCAA plates. For each target, 6-10 clones were randomly picked for sequencing. Plasmid DNA from these clones was isolated using the Zymoprep Kit II (Zymoresearch, Irvine, CA). The isolated DNA was transformed into Novablue$^{TM}$ (*E.coli*) cells (Novagen, San Diego, CA) and plasmid DNA was obtained using the QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA). Subsequently, plasmid DNA was sequenced by Genewiz (La Jolla, CA).

The equilibrium dissociation constant ($K_D$) was determined for a subset of mutants through yeast cell surface titrations, as previously described[10, 45]. Briefly, yeast cells expressing mutant proteins as cell surface fusions were labeled with varying concentrations of biotinylated target species, followed by strep-PE; labeling was carried out on ice and pH 7.4. The mean fluorescence intensity of cells, corresponding to cell surface bound target, was measured using flow cytometry. Subsequently, $K_D$ of the binding interaction was determined by fitting fluorescence data, from at least three separate experiments for each mutant, to the following equation:

$$F = \frac{c[L]_0}{K_D + [L]_0}$$

where, F is the observed mean fluorescence intensity, $[L]_0$ is the concentration of target used for labeling the cells, and c is a fitted parameter along with $K_D$. The 68% confidence intervals, corresponding to the commonly reported standard deviation from triplicate measurements, were calculated as described[46].

### 4.4.5. Construction of mRNA display library

An mRNA display library of Sso7d mutants was constructed using a previously described protocol as a guideline[47]. Oligonucleotide U1_Sso7d was amplified using two rounds of PCR to attach the 5' and 3' mRNA display consensus sequences. The 5' consensus sequence contains a T7 RNA polymerase promoter, a TMV translation enhancer and a sequence coding the FLAG epitope tag; the 3' consensus sequence includes a 6xHis tag coding region followed by a sequence required for conjugation of the puromycin linker [47]. The primers used for the first round of PCR were: *Sso7d_mRNA_Rd1_fwd* – 5' CAA TTA CTA TTT ACA ATT ACA ATG GCG ACC GTG AAA TTT AAA TAT AAA 3' and *Sso7d_mRNA_Rd1_Rev* – 5'GTG ATG GTG GTG ATG GCT GCC GCC TTT TTT CTG TTT TTC CAG CAT CTG 3'. Primers used for the second PCR were: *Sso7d_mRNA_Rd2_fwd*- 5'GCA AAT TTC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG G 3' and *Sso7d_mRNA_Rd2_rev* -5'ATA GCC GGT GCC AGA TCC AGA CAT TCC CAT ATG GTG ATG GTG GTG ATG GCT GC-3'. Primers were obtained from Integrated DNA Technologies (Coralville, IA).

The reaction mixture for the first round of PCR had the following components: Phusion™ HF DNA polymerase (New England Biolabs, Ipswich, MA; 1U/50µl) in 1X HF Phusion™ buffer, 0.2 mM deoxynucleotide triphosphate (dNTPs), 0.1 M of the forward and reverse primer each, 1M betaine, 3% dimethyl sulfoxide (DMSO) and 500 ng U1_Sso7d as the template. PCR conditions were as follows: Initial denaturation at 98°C for 2 min, followed by 30 cycles of denaturation at 98°C for 1 min, annealing at 67°C for 1 min, extension at 72°C for 20 sec, and a final extension at 72°C for 10 min. The product from the

first round of PCR was purified using the Qiagen PCR-purification kit (Qiagen, Valencia, CA) and was used as the template (100 ng per PCR) for the second round of PCR along with corresponding primers; other components of the reaction mixture were the same as those used in the first round of PCR. Also, PCR conditions used in round 2 were identical to round 1, except that an annealing temperature of 66$^{o}$C was used. DNA obtained after the second round of PCR was purified using phenol: chloroform: isoamyl alchohol (Fisher, Waltham, MA) precipitation and DNA pellets were resuspended in water containing 0.1% diethylpyrocarbonate (DEPC).

DNA from the round 2 PCR was used in an *in vitro* transcription reaction. A 300 μL reaction containing 200 nM DNA, 5 mM ribonucleotide triphosphate (rNTPs), 19 mM $MgCl_2$, 45 l T7 RNA polymerase (20 U/ l) and 1X transcription buffer (Ambion, Austin, TX) was incubated for 8 hours at 37 $^{o}$C. The mRNA obtained was purified using acidic phenol chloroform (Ambion, Austin, TX) and a Nap 5 column (GE Healthcare, UK), followed by DNA digestion in a 850 l reaction for 4 hours at 37 $^{o}$C with RNase-free Turbo DNase (1 U/ l, 5% of the reaction volume) and 1X DNase buffer (Promega, Madison, WI).The purified mRNA was conjugated with a puromycin linker ([psoralen-(ATAGCCGGTG) $_2$-OMe-d$A_{15}$-C9C9-ACC-puromycin]; Keck Oligo Synthesis Lab, Yale University). The 250 μL conjugation reaction mix was comprised of 20 mM HEPES, 100 mM KCl and the puromycin linker at 2.5 times the total molar concentration of mRNA; 200 μg of mRNA was used. The reaction mixture was incubated in a thermocycler with the following program: 85 $^{o}$C for 8 min, followed by 60 cycles for 20 sec each, with 1 $^{o}$C decrease in each cycle from 85 $^{o}$C to 25 $^{o}$C, and 25 $^{o}$C for 25 min. Subsequently, the

puromycin linker was annealed to mRNA by crosslinking with ultraviolet (UV) light at 360 nm for 20 min. Subsequently, 125 l LiCl was added to the crosslink reaction and was incubated at -20 $^{o}$C overnight. Following incubation, the sample was centrifuged at maximum speed on a table-top centrifuge and the supernatant was removed. The resulting mRNA pellet was washed with 500 l 75% ethanol, centrifuged for 15 min and washed with 100% ethanol. The mRNA pellet was air-dried and resuspended in 50 l 0.1% DEPC water.

*In vitro* translation on the mRNA-puromycin fusion was carried out in rabbit reticulocyte lysate (Ambion, Austin, TX) without the use of any radioactive amino acid. The translation reaction was composed of 282 g crosslinked mRNA, 50 M methionine, 1X buffer without leucine, 340 l retic lysate, to a final 500 l volume adjusted by nuclease-free water. The reaction was incubated at 30 $^{o}$C for 1.5 hr. Subsequently, $MgCl_2$ (76 mM) and KCl (880 mM) was added and the mixture was further incubated at room temperature for 1 hr.

A bead-based assay was used to detect the presence of the translated protein. 50 µl of the translation product or a control sample containing crosslinked mRNA was incubated with His-binding magnetic beads (Invitrogen, Carlsbad, CA) pre-washed in binding buffer (50 mM sodium phosphate, 300 mM NaCl and 0.01% Tween-20), for 1 hr. The volume was then adjusted to 100 l by adding PBS-BSA containing 0.1% salmon sperm DNA (PBS-BSA-DNA) and the beads were separated from the solution using a magnetic particle concentrator, washed with PBS-BSA-DNA and resuspended in 100 l PBS-BSA-DNA. Subsequently, beads were incubated with 100 l PBS-BSA-DNA containing an anti-FLAG antibody conjugated with Horse Radish Peroxidase (HRP) (Sigma-Aldrich, St. Louis, MO) at

1:20000 dilution, for 1 hr at room temperature. After the incubation, beads were washed twice with PBS-BSA-DNA, resuspended in 100 l SuperSignal ELISA femto maximum sensitivity substrate (Thermo Scientific, Rockford, IL) and transferred to a 96 well plate; chemiluminescence was recorded using a microplate reader (Perkin Elmer, Waltham, MA) .

mRNA-puromycin and mRNA-protein fusions were isolated from the *in vitro* translation mix using oligo(dT) cellulose beads (Ambion, Austin, TX); the puromycin linker has a poly(dA) stretch. Subsequently, the entire oligo(dT) purified product was subjected to reverse transcription. 25 μL of 100 M reverse transcription primer (5'-TTT TTT TTT TNN CCA GAT CCA GAC ATT CCC AT-3', Integrated DNA Technologies, Coralville, IA) was added to the entire oligo (dT) purified product (600 μL) and incubated for 15 min at room temperature. Next, 200 l 5X first strand buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM $MgCl_2$), 50 l 10 mM dNTPs and 100 l 0.1 M DTT were added to the reverse transcription reaction, the total volume was adjusted to 995 l using 0.1% DEPC water, and the mixture was incubated at 42 $^{o}$C for 2 min. Finally, 5 l Superscript$^{TM}$ reverse transcriptase (200 U/μL; Invitrogen, Carlsbad, CA) was added and the reaction was incubated further for 50 min at 42 $^{o}$C.

The 6xHis tag on the translated protein was used to isolate the mRNA-cDNA-proteins, using Ni-NTA agarose suspension (Qiagen, Valencia, CA). Finally, a Nap-5 column was used to purify the mRNA-cDNA-protein fusions. Based on $A_{260}$ measurement, the library diversity was estimated as ~ $5x10^{11}$.

### 4.4.6. Screening the mRNA display library

Negative selection with Dnyal™ biotin binder beads, as well as beads coated with mIgG, cIgY, gIgG and hFc, was carried out on the mRNA display library. 100 µl Dynal™ biotin binder beads were washed twice in PBS-BSA-DNA. The entire mRNA display library was added to the washed beads and incubated at 4°C with gentle rotation for 1 hr. The supernatant was separated from the beads using a magnetic particle concentrator and applied sequentially to beads coated with mIgG, gIgG and hFc, in that order, with incubation at 4°C with rotation for 1 hr in each step.

Supernatant from the negative selection step was incubated with beads coated with rIgG After a 1 hr incubation at 4°C, the supernatant was separated from the beads and subsequently incubated with beads coated with BNP-32 for 1 hr at 4°C. The beads from incubation with rIgG and BNP-32 were washed in their respective tubes with 200 µl PBS-BSA-DNA. Next, the Sso7d mutants bound to rabbit IgG and BNP-32 cDNA were eluted using 200 µl 0.15 M KOH at room temperature with shaking for 1 hr, and the eluate was neutralized by 5N HCl. cDNA was precipitated using linearized acrylamide made from poly acrylamide (Thermo Fisher Scientific, NJ) as described[48]. The precipitate was amplified by PCR with the primers *Sso7d_mRNA_Rd2_fwd* and *Sso7d_mRNA_Rd2_rev*. The PCR mix was composed of Phusion™ HF DNA polymerase (1U/50µl) in 1X HF Phusion™ buffer, 0.2mM dNTPs, 0.2 M of each primer, 1M betaine, 3% dimethyl sulfoxide and the linear acrylamide precipitated DNA as template. PCR conditions used were as follows: Initial denaturation at 98°C for 2 min, followed by 30 cycles of 98°C for 1 min, 66°C for 1 min, 72°C for 15 sec, and a final extension at 72°C for 10 min. The PCR product was precipitated

by Pellet Paint™ (Novagen, San Diego, CA) using the manufacturer's protocol, and used as the template in a second PCR with identical conditions. The product from the second PCR was subsequently amplified with the primers P1f-Sso7d and P1r-Sso7d to attach the 5' and 3' yeast surface display consensus sequences. This PCR product was then transformed into yeast, along with linearized pCTCON. The resultant yeast surface display library was screened by FACS using protocols described in a previous section.

### 4.4.7. Recombinant expression and purification of mutant proteins

Wild-type scaffold proteins and mutants were cloned into the pET22b(+) and pET28b(+) vectors for expression in *E. coli*, with C-terminal and N-terminal 6xHis tags respectively. NdeI and XhoI restriction sites were used for cloning, except in case of TM1112-BNP-32, where NheI and XhoI sites were used. All Sso7d mutants were cloned in pET22b(+).Wild type MIT, Sso6901, Tm1112 and their corresponding mutants were all cloned in pET28b(+). Restriction sites were introduced by PCR with primers (Integrated DNA Technologies, Coralville, IA) as shown; restriction sites have been italicized:

*wild type MIT*- Pf1-5' GAA TCC *CAT ATG* ATG AGT GCA CAA GTA ATG TTA 3'and Pr1-5' CCG CCG *CTC GAG TTA* TCC ACT ACC ATC ACT AGA TG 3',

*wild type Sso6901*-Pf2-5' GAA TCC *CAT ATG* ATG AGT TCG GGT AAA AAA CCA 3'and Pr2- 5' CCG CCG *CTC GAG* TTA TAT TGG ATA ATC ATC TGG TA 3',

*wild type Tm1112*-Pf3-5' GCT ATC *CAT ATG* ATG GAA GTG AAG ATA GAA AAG 3' and Pr3-5' ACT CTG *CTC GAG* GAA GAG GTT GTA GTG CTT TCT 3'

*BNP-32 Sso6901 and β-catenin peptide-Sso6901*-Pf4-5'- G ATA *CAT ATG* AGC AGC GGC AAA AAA CCG GTG -3' and Pr4-5'- C TTA *CTC GAG* TTA AAT CGG ATA ATC ATC CGG CAG -3'

*rIgG-MIT*-Pf5-5' T CGA *CAT ATG* ATG AGC GCG CAG GTG ATG CTG -3' and Pr5- 5' G AAT *CTC GAG* TTA GCG CTG CCA TCG CTG CTC GC -3'

*rIgG-Sso7d and Lysozyme-Sso7d*-Pf6-5' GAA TCC *CAT ATG* ATG GCG ACC GTG AAA TTT AAA 3' and Pr6-5' CCG CCG *CTC GAG* TTT TTT CTG TTT TTC CAG CAT C 3'

*BNP-32-Tm1112*-Pf7-5' GAG TCT *GCT AGC* ATG GAA GTG AAA ATT GAA AAA CCG 3' and Pr7-5'GTG TAC *CTC GAG* TTA AAA CAG GTT ATA ATG TTT GCG CAC 3'

For all mutants, plasmid DNA from yeast surface display selections was used as template DNA in PCRs; genomic DNA from *Sulfolobus solfataricus* and *Thermotoga maritima* (a kind gift from Prof. Robert M. Kelly, North Carolina State University, Raleigh, NC) was used as the template in case of wild-type MIT, Sso6901 and TM1112.

Plasmid constructs were transformed into Rosetta$^{TM}$ (*E.coli*) cells (EMD Biosciences, San Diego, CA) for protein expression. 1L of 2XYT medium (16 g/L Bacto-tryptone, 10 g/L Yeast Extract, 5 g/L NaCl) was inoculated using a 5 ml overnight culture in LB medium( 10 g/L Bacto-tryptone, 5 g/L Yeast Extract, 10 g/L NaCl) with 1% glucose. Protein expression was induced by addition of 0.5 mM IPTG at an $OD_{600}$ of 1.0, and cells were cultured in a 37 °C shaker at 250 rpm for 19-20 hours in case of all proteins except MIT-rIgG; for MIT-rIgG, 1mM IPTG was used to induce protein expression and cells were cultured for 4 hours only. Subsequently, cell extracts were prepared and purified using the Bio-Rad Biologic LP system (Hercules, CA), as previously described[10]. Eluate fractions containing pure protein were

pooled and dialyzed with PBS (wild type Sso6901, wild type MIT, Sso6901- BNP-32 and Sso6901- β catenin peptide) or 50 mM sodium acetate buffer (wild type TM1112, MIT-rIgG, Sso7d- rIgG) and concentrated for analysis. BCA assay was used for the determination of protein concentrations using BSA as a standard.

### 4.4.8. Soluble competition experiments

Yeast cells expressing mutant proteins as cell surface fusions were incubated with biotinylated target, with or without 100-200 fold excess of the corresponding soluble mutant protein. Subsequently, yeast cells were labeled with strep-PE (Invitrogen, Carlsbad, CA) and mean fluorescent intensity of cells due to target-bound cell surface fusions was measured using flow cytometry.

### 4.4.9. Differential Scanning Calorimetry (DSC)

Proteins dialyzed in PBS (wild-type Sso6901, wild-type MIT, Sso6901- BNP-32 and Sso6901-β-catenin peptide), or 50 mM sodium acetate buffer (wild type TM1112, MIT-rIgG, Sso7d- rIgG) were used for DSC experiments, as previously described[10]. A Nano DSC II differential scanning calorimeter (TA instruments, Newcastle, DE) was used for analysis. DSC run software was used for recording the data and exported to Nano Analyze software for analysis.

### 4.4.10. Yeast surface thermal denaturation experiments

The thermal denaturation of ChBD-rIgG and TM1112-BNP-32 was studied using yeast cell surface experiments, previously described [32]. Briefly, yeast cells expressing mutant proteins as cell surface fusions protein were incubated at different temperatures from 20-80 °C, for 10 min. Then, cells were washed with ice-cold PBS-BSA and incubated with their respective targets (2 µM BNP-32 and 500 nM rIgG) for 1 hour on ice. Subsequently, cells were washed with 1 ml PBS-BSA and labeled with strep-PE for 15 min on ice. After this incubation step, cells were washed again with 1 ml PBS-BSA and mean fluorescent intensity of cells was measured using flow cytometry. Experimental data was fit to the theoretical relationship between temperature and the unfolded protein fraction; non-linear least squares regression was used to estimate the mid-point of thermal denaturation ($T_{1/2}$), as previously described[36].

### 4.4.11. Circular Dichroism (CD) Spectroscopy

Protein samples in 20 mM sodium phosphate buffer (pH 7.4) were used for CD spectrocopy on a JASCO-815 spectropolarimeter (Jasco Inc., Easton, MD). CD spectra was recorded over 210-250 nm using a 50 nm/min scan rate, 0.1 nm pitch, 1nm bandwidth and 2 sec response time. Triplicate runs were performed for each sample; 20 mM sodium phosphate buffer (pH 7.4) was used to generate the baseline. The baseline corrected molar ellipticity (θ) was normalized as follows:

$$\bar{\theta} = \frac{\theta - \theta_{\min}}{\theta_{\max} - \theta_{\min}}$$

where $\theta_{min}$ and $\theta_{max}$ are the minimum and maximum values of baseline-corrected molar ellipticity.

## 4.5. References


1. Binz, H.K., P. Amstutz, and A. Pluckthun, *Engineering novel binding proteins from nonimmunoglobulin domains.* Nat Biotechnol, 2005. **23**(10): p. 1257-68.

2. Skerra, A., *Alternative non-antibody scaffolds for molecular recognition.* Curr Opin Biotechnol, 2007. **18**(4): p. 295-304.

3. Liao, H.I., et al., *mRNA display design of fibronectin-based intrabodies that detect and inhibit severe acute respiratory syndrome coronavirus nucleocapsid protein.* J Biol Chem, 2009. **284**(26): p. 17512-20.

4. Lipovsek, D., *Adnectins: engineered target-binding protein therapeutics.* Protein Eng Des Sel, 2011. **24**(1-2): p. 3-9.

5. Steiner, D., P. Forrer, and A. Pluckthun, *Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display.* J Mol Biol, 2008. **382**(5): p. 1211-27.

6. Stumpp, M.T., H.K. Binz, and P. Amstutz, *DARPins: a new generation of protein therapeutics.* Drug Discov Today, 2008. **13**(15-16): p. 695-701.

7. Lofblom, J., et al., *Affibody molecules: engineered proteins for therapeutic, diagnostic and biotechnological applications.* FEBS Lett, 2010. **584**(12): p. 2670-80.

8. Lindborg, M., et al., *Engineered high-affinity affibody molecules targeting platelet-derived growth factor receptor beta in vivo.* J Mol Biol, 2011. **407**(2): p. 298-315.

9. Bloom, J.D., et al., *Protein stability promotes evolvability.* Proc Natl Acad Sci U S A, 2006. **103**(15): p. 5869-74.

10. Gera, N., et al., *Highly stable binding proteins derived from the hyperthermophilic sso7d scaffold.* J Mol Biol, 2011. **409**(4): p. 601-16.

11. Feng, Y., H. Yao, and J. Wang, *Crystal structure of the crenarchaeal conserved chromatin protein Cren7 and double-stranded DNA complex.* Protein Sci, 2010. **19**(6): p. 1253-7.

12. Guo, L., et al., *Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among Crenarchaea.* Nucleic Acids Res, 2008. **36**(4): p. 1129-37.

13. Zhang, Z., et al., *Structural insights into the interaction of the crenarchaeal chromatin protein Cren7 with DNA.* Mol Microbiol, 2010. **76**(3): p. 749-59.

14. Mine, S., et al., *NMR assignment of the chitin-binding domain of a hyperthermophilic chitinase from Pyrococcus furiosus.* J Biomol NMR, 2006. **36 Suppl 1**: p. 70.

15. Nakamura, T., et al., *Crystallization and preliminary X-ray diffraction analysis of a chitin-binding domain of hyperthermophilic chitinase from Pyrococcus furiosus.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2005. **61**(Pt 5): p. 476-8.

16. Nakamura, T., et al., *Tertiary structure and carbohydrate recognition by the chitin-binding domain of a hyperthermophilic chitinase from Pyrococcus furiosus.* J Mol Biol, 2008. **381**(3): p. 670-80.

17. Obita, T., et al., *Structural basis for selective recognition of ESCRT-III by the AAA ATPase Vps4.* Nature, 2007. **449**(7163): p. 735-9.

18. McMullan, D., et al., *Crystal structure of a novel Thermotoga maritima enzyme (TM1112) from the cupin family at 1.83 A resolution.* Proteins, 2004. **56**(3): p. 615-8.

19. Almeida, M.S., et al., *NMR structure of the conserved hypothetical protein TM0487 from Thermotoga maritima: implications for 216 homologous DUF59 proteins.* Protein Sci, 2005. **14**(11): p. 2880-6.

20. Swers, J.S., B.A. Kellogg, and K.D. Wittrup, *Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display.* Nucleic Acids Res, 2004. **32**(3): p. e36.

21. Lipovsek, D., et al., *Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: molecular convergence with single-domain camelid and shark antibodies.* J Mol Biol, 2007. **368**(4): p. 1024-41.

22. Hackel, B.J. and K.D. Wittrup, *The full amino acid repertoire is superior to serine/tyrosine for selection of high affinity immunoglobulin G binders from the fibronectin scaffold.* Protein Eng Des Sel, 2010. **23**(4): p. 211-9.

23. Ackerman, M., et al., *Highly avid magnetic bead capture: an efficient selection method for de novo protein engineering utilizing yeast surface display.* Biotechnol Prog, 2009. **25**(3): p. 774-83.

24. Takahashi, T.T. and R.W. Roberts, *In vitro selection of protein and peptide libraries using mRNA display.* Methods Mol Biol, 2009. **535**: p. 293-314.

25. Wilson, D.S., A.D. Keefe, and J.W. Szostak, *The use of mRNA display to select high-affinity protein-binding peptides.* Proc Natl Acad Sci U S A, 2001. **98**(7): p. 3750-5.

26. Dreier, B. and A. Pluckthun, *Ribosome display: a technology for selecting and evolving proteins from large libraries.* Methods Mol Biol, 2011. **687**: p. 283-306.

27. Hanes, J. and A. Pluckthun, *In vitro selection and evolution of functional proteins by using ribosome display.* Proc Natl Acad Sci U S A, 1997. **94**(10): p. 4937-42.

28. Boder, E.T. and K.D. Wittrup, *Yeast surface display for screening combinatorial polypeptide libraries.* Nat Biotechnol, 1997. **15**(6): p. 553-7.

29. Boder, E.T. and K.D. Wittrup, *Optimal screening of surface-displayed polypeptide libraries.* Biotechnol Prog, 1998. **14**(1): p. 55-62.

30. VanAntwerp, J.J. and K.D. Wittrup, *Fine affinity discrimination by yeast surface display and flow cytometry.* Biotechnol Prog, 2000. **16**(1): p. 31-7.

31. Orr, B.A., et al., *Rapid method for measuring ScFv thermal stability by yeast surface display.* Biotechnol Prog, 2003. **19**(2): p. 631-8.

32. Shusta, E.V., et al., *Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency.* J Mol Biol, 1999. **292**(5): p. 949-56.

33. Chakrabartty, A., et al., *Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities.* Biochemistry, 1993. **32**(21): p. 5560-5.

34. Freskgard, P.O., et al., *Assignment of the contribution of the tryptophan residues to the circular dichroism spectrum of human carbonic anhydrase II.* Biochemistry, 1994. **33**(47): p. 14281-8.

35. Vuilleumier, S., et al., *Circular dichroism studies of barnase and its mutants: characterization of the contribution of aromatic side chains.* Biochemistry, 1993. **32**(39): p. 10303-13.

36. Hackel, B.J., A. Kapila, and K.D. Wittrup, *Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling.* J Mol Biol, 2008. **381**(5): p. 1238-52.

37. Barrios, Y., P. Jirholt, and M. Ohlin, *Length of the antibody heavy chain complementarity determining region 3 as a specificity-determining factor.* J Mol Recognit, 2004. **17**(4): p. 332-8.

38. Lamminmaki, U., et al., *Expanding the conformational diversity by random insertions to CDRH2 results in improved anti-estradiol antibodies.* J Mol Biol, 1999. **291**(3): p. 589-602.

39. Lee, C.V., et al., *High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold.* J Mol Biol, 2004. **340**(5): p. 1073-93.

40. Rock, E.P., et al., *CDR3 length in antigen-specific immune receptors.* J Exp Med, 1994. **179**(1): p. 323-8.

41. Koide, A., et al., *High-affinity single-domain binding proteins with a binary-code interface.* Proc Natl Acad Sci U S A, 2007. **104**(16): p. 6632-7.

42. Guntas, G., C. Purbeck, and B. Kuhlman, *Engineering a protein-protein interface using a computationally designed library.* Proc Natl Acad Sci U S A, 2010. **107**(45): p. 19296-301.

43. Park, S., et al., *Limitations of yeast surface display in engineering proteins of high thermostability.* Protein Eng Des Sel, 2006. **19**(5): p. 211-7.

44. Shehi, E., et al., *Thermal stability and DNA binding activity of a variant form of the Sso7d protein from the archeon Sulfolobus solfataricus truncated at leucine 54.* Biochemistry, 2003. **42**(27): p. 8362-8.

45. Chao, G., et al., *Isolating and engineering human antibodies using yeast surface display.* Nat Protoc, 2006. **1**(2): p. 755-68.

46. Lakowicz, J.R., *Principles of Fluorescence Spectroscopy.* 2nd edn Kluwer Academic/Plenum Publishers, New York., 1999.

47. Huang, B.C. and R. Liu, *Comparison of mRNA-display-based selections using synthetic peptide and natural protein libraries.* Biochemistry, 2007. **46**(35): p. 10102-12.

48. Gaillard, C. and F. Strauss, *Ethanol precipitation of DNA with linear polyacrylamide as carrier.* Nucleic Acids Res, 1990. **18**(2): p. 378.

49. Pavoor, T.V., Y.K. Cho, and E.V. Shusta, *Development of GFP-based biosensors possessing the binding properties of antibodies.* Proc Natl Acad Sci U S A, 2009. **106**(29): p. 11895-900.

**Figures**

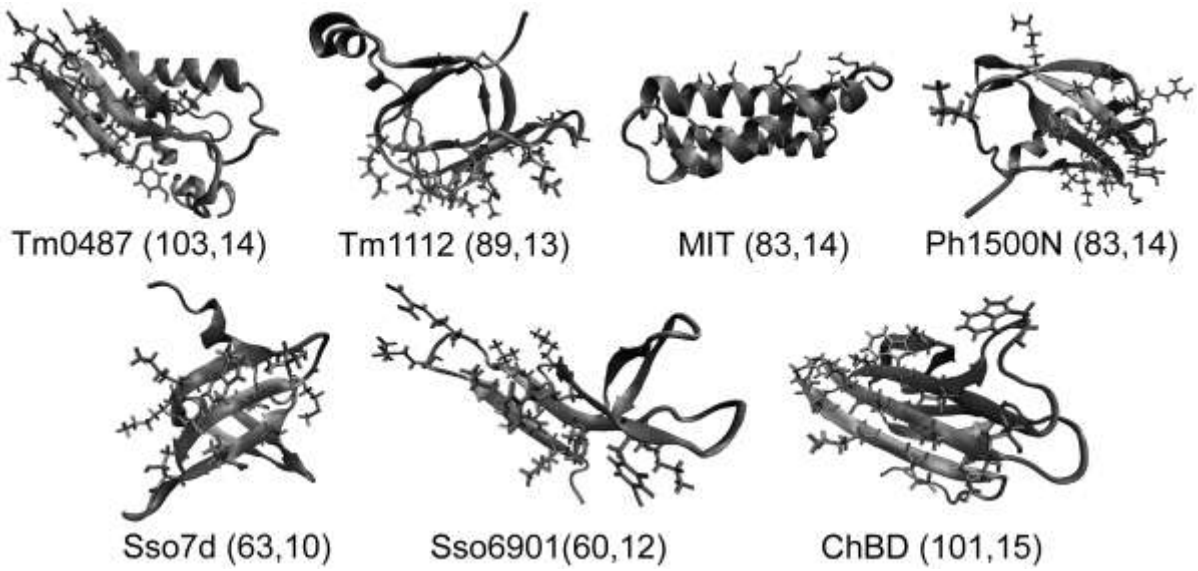



**Figure 4.1:** Selection of surface-accessible residues on scaffolds for randomization. Residues chosen for randomization are shown in licorice representation; this figure was generated using Visual Molecular Dynamics (VMD) software. For each scaffold, the total number of amino acid residues in the wild-type protein, followed by the number of residues randomized, is shown in parentheses.

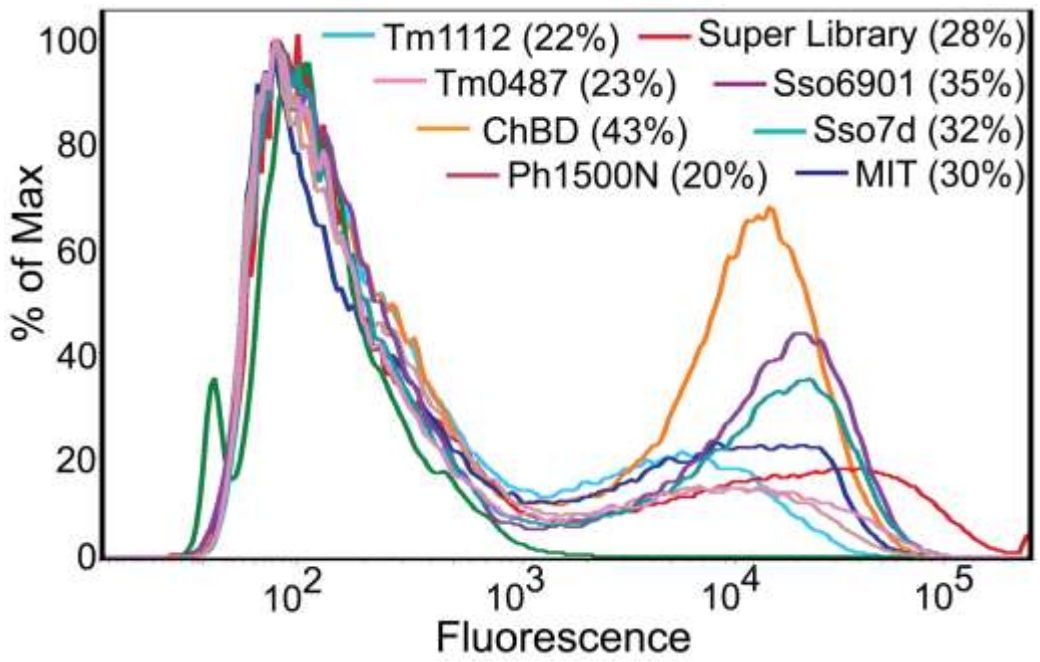



**Figure 4.2:** Analysis of cell surface protein expression for individual scaffold libraries and the combined super-library, through immunofluorescent detection of the c-myc epitope tag. Yeast cells displaying mutant proteins as cell surface fusions were labeled with an anti-c-myc antibody followed by a secondary antibody conjugated with Alexa Fluor 488, or the fluorescent secondary antibody alone; cells were subsequently analyzed using flow cytometry. Fluorescence histograms for each scaffold library, as well as the combined super-library are shown. The green histogram corresponds to the control sample labeled with just the secondary antibody; controls for individual scaffold libraries are not shown in the interest of clarity. Numbers in parentheses in the legend show the approximate percentage of yeast cells in each library expressing the c-myc epitope tag, and therefore by inference, full-length proteins as cell surface fusions.

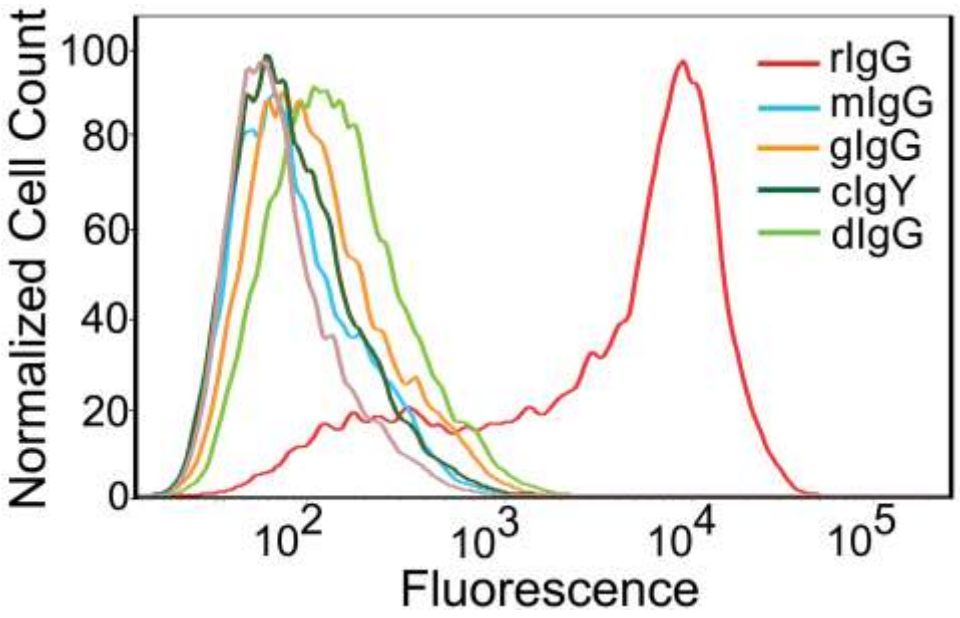



**Figure 4.3:** Analysis of binding specificity of MIT-rIgG. Yeast cells displaying MIT-rIgG as cell surface fusions were labeled with 1 M rIgG or an equivalent concentration of the non-target immunoglobulin species donkey IgG (dIgG), goat IgG (gIgG), chicken IgG (cIgY) and mouse IgG (mIgG). Fluorescence histograms confirm the high specificity of MIT-rIgG binding to rIgG; binding to other non-target immunoglobulins is insignificant.

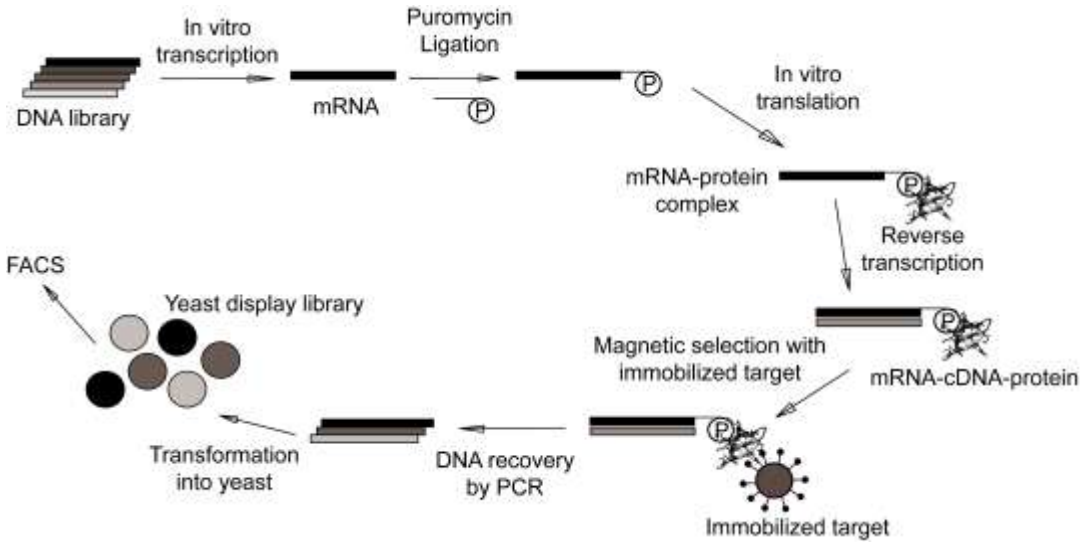



**Figure 4.4:** Schematic for screening of a high diversity Sso7d library using a combination of mRNA display and yeast surface display. The mRNA display was screened against magnetic beads coated with rIgG or BNP-32, and bead-bound mRNA-cDNA-protein fusions were isolated. DNA was recovered by PCR and transformed into yeast to construct a yeast surface display library. Subsequently FACS was used to isolate the pool of binders with highest affinity for the target from the yeast display library.

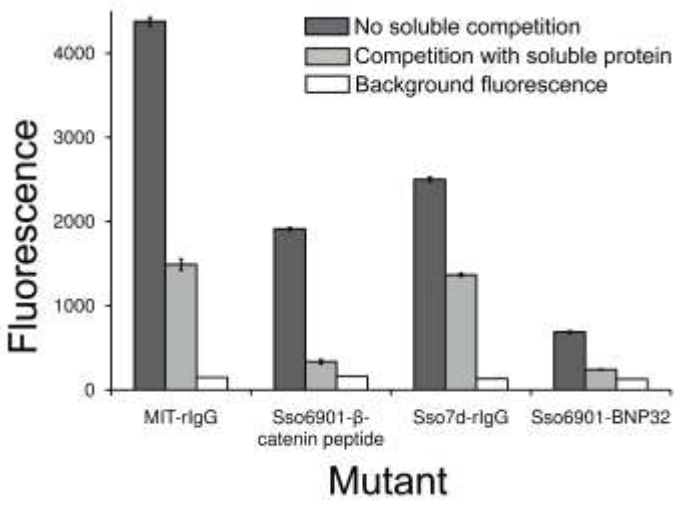



**Figure 4.5:** Mutant proteins recombinantly expressed in *E. coli* are functional. Yeast cells expressing mutant proteins as cell surface fusions were labeled with their corresponding targets in the absence (dark grey bars) or presence (light grey bars) of 100-200 fold excess soluble mutant protein, and fluorescence intensity of cells due to cell surface bound target was recorded using flow cytometry. Fluorescence intensity decreases in the presence of the soluble protein; this shows that the recombinant protein is functional and binds the target. Error bars indicate the standard error of the mean for all cells assayed.

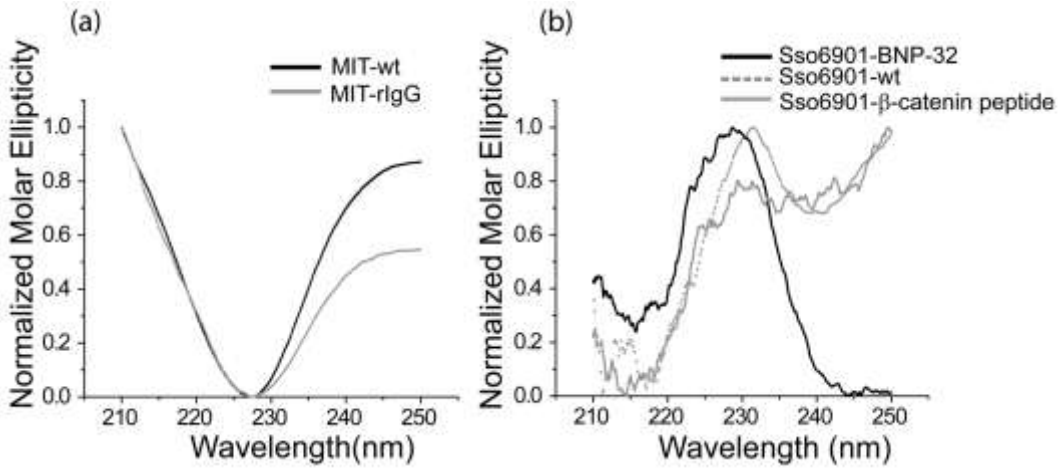



**Figure 4.6:** CD spectra of (a) wild-type MIT, (b) wild-type Sso6901 and mutants in 20 mM sodium phosphate buffer (pH 7.4). The baseline corrected molar ellipticity (θ) was normalized as $\bar{\theta} = \frac{\theta - \theta_{\min}}{\theta_{\max} - \theta_{\min}}$, where $\theta_{\min}$ and $\theta_{\max}$ are the minimum and maximum values of baseline-corrected molar ellipticity. Normalized ellipticity values are plotted here.

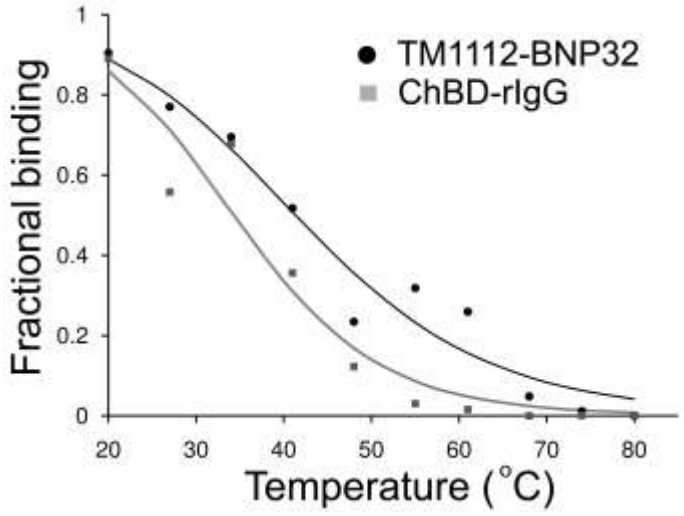



**Figure 4.7:** Analysis of thermal stability through determination of the mid-point of irreversible thermal denaturation ($T_{1/2}$) of yeast surface displayed protein. Yeast cells expressing Tm1112-BNP-32 and ChBD-rIgG as cell surface fusions were incubated at different temperatures in the range 20-80 $^{o}$C. Subsequently, cells were labeled with the corresponding targets (2 M BNP32 and 500 nM rIgG) and fraction of cell surface fusions that retain binding to the target was assessed using flow cytometry. The solid curves were obtained by fitting experimental data to the theoretical relationship between the fraction of protein fusions retaining binding and the incubation temperature. Representative data from one of three separate experiments is shown.

## Tables

**Table 4.1:** Amino acid sequences of scaffold proteins used in the construction of the super-library is shown. Randomized positions are shown in bold font and underlined.

| Scaffold (PDB ID) Estimated Library Diversity | Wild type sequence |
|---|---|
| **TM0487**[a] (1WCJ) | 1.....10.....20.....30.....40.....50<br>PMSKKVTKEDVLNALKNVIDFELGLDVVSLGLV**YD**I**Q**I**DD**N**Q**NN**V**K**V**L**MTM |
| 5.3x10$^7$ | .....60.....70.....80.....90.....100.<br>TTPMCPLAGMILSDAEEAIKKIEGV**NN**V**E**V**EL**TFDPPWTPERM**S**PELREKFGV |
| **TM1112**[b] (1LKN) | 1.....10.....20.....30.....40.....50<br>MEVKIEKPTPEKLKELSVEKWPIWEKEVSEFDWYYDTNET**S**YILEG**K**V**E**V |
| 7x10$^7$ | .....60.....70.....80.....<br>**T**T**E**D**G**K**K**Y**V**I**EK**GDLVTFPKGLR**SR**W**K**V**LE**PVRKHYNLF |
| **MIT** (2V6Y) | 1.....10.....20.....30.....40.....50<br>MSAQVMLEDMARKYAILAVKADKEGKV**E**DAI**T**YY**K**KAI**E**VL**SQ**IIVLYPE |
| 6.7x10$^7$ | .....60.....70.....80..<br>SVA**RT**AY**E**QMI**N**EY**KK**RI**S**YL**E**KVLPASSDGSG |
| **Ph1500** (2JV2) | 1.....10.....20.....30.....40.....50<br>HHHHHHMEGVIMS**E**L**K**L**K**PLPKVELPPDFVDVIRIKLQGKTVRTGD**V**I**G**I |
| 6.2x10$^7$ | .....60.....70.....80..<br>**S**ILGK**E**V**K**F**K**V**VQ**A**Y**PSPLRVEDRT**K**I**T**LVTHP |
| **Sso7d**[c] (1SSO) | 1.....10.....20.....30.....40.....50<br>ATVKFKYKGEEKQVDISKI**KK**V**W**R**V**GK**M**I**S**F**T**YD**L**GGGK**T**G**R**G**A**VSEKDA |
| 1x10$^8$ | .....60..<br>PKELLQMLEKQKK |

**Table 4.1** continued

| | |
|---|---|
| **Sso6901** | 1.....10.....20.....30.....40.....50 |
| (2JTM) | MSSGKKPVKVKTPAGKEAELVPE**K**V**W**A**L**AP**K**G**R**KG**V**K**I**G**L**F**K**DPETGK**Y**F |
| $6.2x10^7$ | .....60 |
| | **R**H**K**LPDDYPI |
| **ChBD**[d] | 258.260.....270.....280.....290 .....300..305 |
| (2CZN) | TTPVPVSG**S**L**E**V**K**V**N**D**W**GSGA**E**Y**D**V**T**L**N**LDGQYDW**T**TVKVKLAPGATVG |
| $1.6x10^7$) | |
| | ....310.....320.....330.....340.....350.....358 |
| | SF**W**SANKQEGNGYVIFTPVS**W**NKGP**T**A**T**F**G**F**I**VNGPQGDKVEEITLEINGQVI |

[a] C94S mutation was introduced; the serine residue introduced is shown in bold font. Two additional positions, D40 and L82 were inadvertently randomized during oligonucleotide synthesis; these residues are shown in bold font, italicized and underlined.

[b] C41S and C74S mutations were introduced; the serine residues introduced are shown in bold font.

[c] Construction of the Sso7d library was previously reported[10]. The E35L mutation was introduced to abolish RNase activity[49].

[d] Numbering of residues is based on chitinase A from *Pyrococcus furiosus* (ORF PF1233). A threonine insertion was inadvertently introduced between W292 and T293 during oligonucleotide synthesis; this insertion is shown in bold font and highlighted in grey.

**Table 4.2:** Protein sequences of mutants obtained from the super-library, or the high diversity Sso7d library screened using a combination of mRNA display and yeast surface display. Corresponding wild-type sequences are shown as a reference. Mutated residues are in bold font and underlined.

| Scaffold/mutant | Sequence |
|---|---|
| | .30.....40.....50.....60.....70.. |
| **MIT-WT** | EDAITYYKKAIEVLSQIIVLYPESVARTAYEQMINEYKKRISYLE |
| *MIT-rIgG*[a] | GDAISYYWKAIVVLQRIIVLYPESVATGAYLQMIGEYARRIDYLN |
| | ......30.....40.....50.. |
| **Sso6901-WT** | KVWALAPKGRKGVKIGLFKDPETGKYFRHK |
| *Sso6901-β-catenin peptide*[a] | LVQARAPFGRKGVKRGLFRDPETGKAFFHL |
| *Sso6901-BNP-32*[a] | SVYAFAPCGLKGSKWGSFLDPETGKYFDHV |
| | 20.....30.....40... |
| **Sso7d-WT** | KKVWRVGKMISFTYDLGGGKTGRGA |
| Sso7d-HEL | CFVFRWGKCICFDYDLGGGKQGSGC |
| Sso7d- β-catenin peptide | NPVVRYGKLIFFAYDLGGGKLGAGW |
| *Sso7d-Fluorescein*[a] | KFVLRPGKAILFYYDLGGGKYGFGL |
| Sso7d-Fluorescein-2 | LKVFRIGKVIFFRYDLGGGKFGYGY |
| Sso7d-BNP-32-2 | INVNRGGKFIRFTYDLGGGKFGSGR |
| ***From mRNA display library*** | |
| *Sso7d-rIgG*[a] | YRVFRSGKTIFFRYDLGGGKLGVGI |
| Sso7d-rIgG-2 | YWVRRHGKSITFQYDLGGGKNGLGF |
| Sso7d-rIgG-3 | QLVRRRGKRITFRYDLGGGKKGVGY |
| *Sso7d-BNP-32* | YCVKRSGKKIRFFYDLGGGKRGIGT |
| Sso7d-BNP-32-3 | ARVWRVGKRILFGYDLGGGKLGIGR |
| Sso7d-BNP-32-4 | LYVYRIGKRIIFAYDLGGGKVGVGW |
| Sso7d-BNP-32-5 | NVVSRIGKTITFLYDLGGGKMGSGV |
| Sso7d-BNP-32-6 | SWVFRRGKYIIFAYDLGGGKGGHGK |
| Sso7d-BNP-32-7 | QAVSRIGKHINFKYDLGGGKQGSGC |
| | ...50.....60.....70.....80 |
| **TM1112-WT** | KVEVTTEDGKKYVIEKGDLVTFPKGLRCRWKVLE |
| *Tm1112-BNP-32*[a,b] | FVCVLTMDGKTYGID---------- |

**Table 4.2** continued

| | |
|---|---|
| **ChBD-WT**[c] | .267.....277.....287...I....297.....307..<br>**SLEVKVNDWGSGAEYDVTLNLDGQYDW*T*TVKVKLAPGATVGSF*W*SAN**<br>.313.....323.....333..<br>KQEGNGYVIFTPVS**W**NKGP**TA*T*FG*F*I** |
| *ChBD-rIgG*[a, c] | .267.....277.....287...I....297.....307..<br>**PLRVRV*V*DLGSGA*W*Y*L*VLL*H*LDGQYDW*T*TVKVKLAPGATVGSF*Y*SAN**<br>.313.....323.....333..<br>KQEGNGYVIFTPVS**Q**NKGP**Q**A*W*F*E*F*E** |

[a] These proteins were chosen for further analysis

[b] A truncated TM1112 mutant was obtained due to the introduction of a stop codon at K62

[c] A threonine insertion was inadvertently introduced between W292 and T293 (indicated by I) during oligonucleotide synthesis; this insertion is shown in bold font and italicized.

**Table 4.3:** $K_D$ estimates for selected mutants. Yeast cells displaying mutant proteins as cell surface fusions were labeled with varying concentrations of soluble target, and the fraction of cell surface fusions bound to the target was measured using flow cytometry. Data from at least three separate experiments for each mutant was fit to a monovalent binding isotherm and the $K_D$ values were estimated from a global fit. 68% confidence intervals were calculated as previously described [46]

| Target molecule | Target size (kDa) | Mutant-scaffold | $K_D$ (nM) | 68% confidence interval (nM) | Half maximal[a] $K_D$ (nM) |
|---|---|---|---|---|---|
| Fluorescein | 0.39 | Sso7d | 879 | 515-1700 | - |
| β catenin peptide | 1.36 | Sso6901 | 3297 | 2120-5300 | - |
| BNP-32 | 4 | Sso6901 | 2100 | 1560-2850 | - |
| | | Tm1112 | 653 | 565-755 | - |
| | | Sso7d[b] | 2696 | 2000-3650 | - |
| HEL | 14.3 | Sso7d | 349[c] | 225-540 | - |
| rIgG | 150 | MIT | 183 | 100-380 | 105±8 |
| | | ChBD | 271 | 160-480 | 170±15 |
| | | Sso7d[b] | 416 | 215-900 | 332±54 |

[a] The half-maximal $K_D$ is the target concentration at which fraction of cell surface fusions bound to the target is 0.5. Values are reported as mean of triplicate measurements ± standard error of the mean.

[b] These binders were isolated from the mRNA display library of Sso7d mutants

**Table 4.3** continued

[c] The HEL binder isolated from the super-library was identical to a mutant previously isolated from the Sso7d yeast surface display library[10].

[d] In case of titrations with rIgG, fluorescence signal decreased at high target concentrations (hook effect). Such a decrease has been previously reported in yeast cell surface titrations[10, 49]. Fluorescence values at higher concentrations were not included in calculations to estimate $K_D$; this leads to a more conservative estimate of $K_D$, i.e. a higher $K_D$ value.

**Table 4.4:** Analysis of thermal stability for wild-type scaffold proteins and mutants. The melting temperature ($T_m$) was determined by Differential Scanning Calorimetry. The temperature corresponding to the mid-point of irreversible thermal denaturation was determined from experiments with yeast cell surface displayed proteins.

| Wild-Type or Target-binding Mutant | $T_m$ or $T_{1/2}$ (°C) |
|---|---|
| **TM1112-WT** | 99.4 |
| Tm1112-BNP-32 | 38±1.7[a] |
| **MIT-WT** | 116.9 (116.7,117.2) |
| MIT-rIgG | 74.2 (74.1,74.3) |
| **Sso7d-WT** | 98[51] |
| Sso7d-HEL | 92.7[10] |
| Sso7d-rIgG | 85.5 (85.5,85.6) |
| **Sso690-WT** | 97.2 (97.1,97.3) |
| Sso6901-β-catenin peptide | 84.5 (84.5,84.6) |
| Sso6901-BNP-32 | 76.2 (76.0,76.4) |
| **ChBD-WT** | >85[b] |
| ChBD-rIgG | 50±4.3[a] |

Unless stated otherwise, all data reported are $T_m$ values. The average value from duplicate measurements is reported for all proteins, with the exception of TM1112-wild-type, where a

**Table 4.4** continued

single measurement was performed. Numbers in parentheses show $T_m$ values obtained in each experiment. The $T_{1/2}$ values reported are the mean from three separate experiments ± standard error of the mean.

[a] $T_{1/2}$ value

[b] The wild-type ChBD protein has been reported to retain its secondary structure at 85°C[16].

**Table 4.5:** Oligonucleotides and primers used for construction of the super-library.

| Scaffold | Sequence |
|---|---|
| TM1112 | ***Oligonucleotide U1_TM1112:*** 5'- GAA TTT GAT TGG TAT TAT GAT ACC AAC GAA ACC AGC TAT ATT CTG GAA GGC **NNN** GTG **NNN** GTG **NNN** ACC **NNN** GAT **NNN** AAA **NNN** TAT **NNN** ATT **NNN NNN** GGC GAT CTG GTG ACC TTT CCG AAA GGC CTG CGC AGC **NNN** TGG **NNN** GTG **NNN** **NNN** CCG GTG CGC AAA CAT TAT AAC CTG TTT -3' |
| | ***Forward Primer (P1f_YSD_TM1112):***5'- AGT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GCT AGC ATG GAA GTG AAA ATT GAA AAA CCG ACC CCG GAA AAA CTG AAA GAA CTG AGC GTG GAA AAA TGG CCG ATT TGG GAA AAA GAA GTG AGC GAA TTT GAT TGG TAT TAT GAT ACC -3' |
| | ***Reverse Primer (P1r_YSD_TM1112):***5'-CTC GAG CTA TTA CAA GTC CTC TTC AGA AAT AAG CTT TTG TTC GGA TCC AAA CAG GTT ATA ATG TTT GCG CA -3' |
| TM0487[a] | ***Oligonucleotide U1_TM0487:*** 5'- ATG CCG ATG AGC AAA AAA GTG ACC AAA GAA GAT GTG CTG AAC GCG CTG AAA AAC GTG ATT GAT TTT GAG CTG GGC CTG GAT GTG GTG AGC CTG GGC CTG GTG **NNN NNN** ATT **NNN** ATT **NNN** ***NNN*** **NNN** AAC **NNN** GTG **NNN** GTG **NNN** ATG ACC ATG ACC ACC CCG ATG AGC CCG -3' |
| | ***Forward Primer (P1f_YSD_TM0487):*** 5' - AGT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GCT AGC ATG CCG ATG AGC AAA AAA GTG A -3' |
| | ***Reverse Primer (P1r_YSD_TM0487):*** 5'- AAT CGC TTC TTC CGC ATC GCT CAG AAT CAT GCC CGC CAG CGG GCT CAT CGG GGT GGT - 3' |
| | ***Oligonucleotide U2_TM0487:*** 5'- GCG ATT AAA AAA ATT GAA GGC GTG **NNN NNN** GTG **NNN** GTG **NNN** ***NNN*** ACC TTT GAT CCG CCG TGG ACC CCG GAA CGC ATG AGC CCG GAA CTG CGC GAA AAA TTT GGC GTG -3' |
| | ***Forward Primer (P2f_YSD_TM0487):*** 5'- CCG CTG GCG GGC ATG ATT CTG AGC GAT GCG GAA GAA GCG ATT AAA AAA ATT GAA GGC GTG- 3' |
| | ***Reverse Primer (P2r_YSD_TM0487):*** 5'- CTC GAG CTA TTA CAA GTC CTC TTC AGA AAT AAG CTT TTG TTC *GGA TCC* CAC GCC AAA TTT TTC GCG CAG - 3' |

**Table 4.5** continued

| | |
|---|---|
| **MIT** | ***Oligonucleotide U1_MIT:*** 5'-AAA GCG GAT AAA GAA GGC AAA GTG **NNN** GAT GCG ATT **NNN** TAT TAT **NNN** AAA GCG ATT **NNN** GTG CTG **NNN** **NNN** ATT ATT GTG CTG TAT CCG GAA AGC GTG GCG **NNN** **NNN** GCG TAT **NNN** CAG ATG ATT **NNN** GAA TAT **NNN** **NNN** CGC ATT **NNN** TAT CTG **NNN** AAA GTG CTG CCG GCG AGC- 3' |
| | ***Forward Primer (P1f_YSD_MIT):*** 5'- AGT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GCT AGC ATG AGC GCG CAG GTG ATG CTG GAA GAT ATG GCG CGC AAA TAT GCG ATT CTG GCG GTG AAA GCG GAT AAA GAA GGC AAA- 3' |
| | ***Reverse Primer (P2r_YSD_MIT):*** 5'- CTC GAG CTA TTA CAA GTC CTC TTC AGA AAT AAG CTT TTG TTC *GGA TCC* GCC GCT GCC ATC GCT GCT CGC CGG CAG CAC TTT -3' |
| **Ph1500** | ***Oligonucleotide U1_Ph1500:*** 5'-GGT GGT TCT GCT AGC CAT CAT CAT CAT CAT CAT ATG GAA GGC GTG ATT ATG AGC **NNN** CTG **NNN** CTG **NNN** CCG CTG CCG AAA GTG GAA CTG CCG CCG GAT TTT GTG GAT G-3' |
| | ***Forward Primer (P1f_YSD_Ph1500):*** 5'-AGT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GCT AGC CAT C-3' |
| | ***Reverse Primer (P1r_YSD_Ph1500):*** 5'-GCC CTG CAG TTT AAT GCG AAT CAC ATC CAC AAA ATC CGG CGG-3' |
| | ***Oligonucleotide U2_Ph1500:*** 5'- ATT AAA CTG CAG GGC AAA ACC GTG CGC ACC GGC GAT **NNN** ATT **NNN** ATT **NNN** ATT CTG GGC AAA **NNN** GTG **NNN** TTT **NNN** GTG **NNN** **NNN** GCG **NNN** CCG AGC CCG CTG CGC GTG GAA GAT CGC ACC **NNN** ATT **NNN** CTG GTG ACC CAT CCG GGA TCC G-3' |
| | ***Forward Primer (P2f_YSD_Ph1500):*** 5'- CCG CCG GAT TTT GTG GAT GTG ATT CGC ATT AAA CTG CAG GGC AAA ACC-3' |
| | ***Reverse Primer (P2r_YSD_Ph1500):*** 5'- CTC GAG CTA TTA CAA GTC CTC TTC AGA AAT AAG CTT TTG TTC GGA TCC CGG ATG GGT CA-3' |

**Table 4.5** continued

| | |
|---|---|
| **Sso7d**[10] | ***Oligonucleotide U1_Sso7d:*** 5'- ATG GCG ACC GTG AAA TTT AAA TAT AAA GGC GAA GAA AAA CAG GTG GAT ATT AGC AAA ATT **NNN NNN** GTG **NNN** CGC **NNN** GGC AAA **NNN** ATT **NNN** TTT **NNN** TAT GAT CTG GGC GGC GGC AAA **NNN** GGC **NNN** GGC **NNN** GTG AGC GAA AAA GAT GCG CCG AAA GAA CTG CTG CAG ATG CTG GAA AAA CAG AAA AAA - 3' |
| | ***Forward Primer (P1f_YSD_Sso7d):*** 5'-AGT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GCT AGC ATG GCG ACC GTG AAA TTT AAA TAT AAA G - 3' |
| | ***Reverse Primer (P1r_YSD_Sso7d):*** 5'- CTC GAG CTA TTA CAA GTC CTC TTC AGA AAT AAG CTT TTG TTC *GGA TCC* TTT TTT CTG TTT TTC CAG CAT CTG -3' |
| **Sso6901** | ***Oligonucleotide U1_Sso6901:*** 5'- *AGC* AGC AGC GGC AAA AAA CCG GTG AAA GTG AAA ACC CCG GCG GGC AAA GAA GCG GAA CTG GTG CCG GAA *NNN* GTG *NNN* GCG *NNN* GCG CCG *NNN* GGC *NNN* AAA GGC *NNN* AAA *NNN* GGC *NNN* TTT *NNN* GAT CCG GAA ACC GGC AAA *NNN* TTT *NNN* CAT *NNN* CTG CCG GAT GAT TAT CCG ATT-3' |
| | ***Forward Primer (P1f_YSD_Sso6901):*** 5'- AGT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT *GCT AGC* AGC AGC GGC AAA AAA CC-3' |
| | ***Reverse Primer (P1r_YSD_Sso6901):*** 5'- CTC GAG CTA TTA CAA GTC CTC TTC AGA AAT AAG CTT TTG TTC GGA TCC AAT CGG ATA ATC ATC CGG CAG-3' |

**Table 4.5** continued

| | |
|---|---|
| **ChBD**[b] | ***Oligonucleotide U1_ChBD:*** 5'-ACC CCT GTC CCA GTC TCA GGA ***X*** CTA ***X*** GTA ***X*** GTA ***X*** GAT ***X*** GGT AGT GGT GCT ***X*** TAT ***X*** GTG ***X*** CTT ***X*** TTG GAT GGA CAG TAT GAC TGG- 3'<br><br>***Forward Primer (P1f_YSD_ChBD):*** 5'- AGT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GCT AGC ACC ACC CCT GTC CCA GTC TCA G - 3'<br><br>***Reverse Primer (P1r_YSD_ChBD):*** 5'- GAA GCT TCC TAC AGT GGC TCC CGG CGC CAG TTT CAC TTT CAC GGT ***AGT*** CCA GTC ATA CTG TCC ATC CAA - 3'<br><br>***Oligonucleotide U2_ChBD:*** 5'- GGG AAT GGC TAT GTC ATC TTC ACT CCA GTA AGC **X** AAT AAA GGG CCG **X** GCA **X** TTT **X** TTC **X** GTA AAC GGA CCA CAA GGA GAC AAA - 3'<br><br>***Forward Primer (P2f_YSD_ChBD):*** 5'- GGA GCC ACT GTA GGA AGC TTC **X** AGC GCT AAC AAA CAA GAG GGG AAT GGC TAT GTC ATC TTC -3'<br><br>***Reverse Primer (P2r_YSD_ChBD):*** 5'- CTC GAG CTA TTA CAA GTC CTC TTC AGA AAT AAG CTT TTG TTC *GGA TCC* AAT CAC CTG GCC GTT AAT TTC CAG GGT AAT TTC TTC CAC TTT GTC TCC TTG TGG TCC GTT - 3' |

Randomized positions are shown in bold font. For all libraries, the degenerate NNN codon was used, with the exception of ChBD where an equimolar mixture of trimer phosphoramidites coding for all twenty amino acids was used at randomized positions; these are indicated by **X.**

[a] Two additional positions were inadvertently mutagenized; these are indicated in bold font and italicized

[b] An amino acid insertion was inadvertently introduced; this is indicated in bold font and italicized

# CHAPTER 5

# AFFINITY LIGANDS DERIVED FROM THE Sso7d SCAFFOLD FOR HUMAN IMMUNOGLOBULIN PURIFICATION

Nimish Gera, Andrew B. Hill, Dalon P. White, Ruben M. Carbonell and Balaji M. Rao

Department of Chemical and Biomolecular Engineering

North Carolina State University

Raleigh, NC 27695-7905

## 5.1. Introduction

Downstream processing of protein biopharmaceuticals accounts for more than 50% of the total cost of production [1, 2]. Increasing number of purification steps lead to higher purity but substantially reduced yields. Therefore, single step purification methods such as affinity chromatography are commonly used. Affinity chromatography relies on a highly specific interaction between a protein and an affinity ligand and can be readily used to purify the protein of interest from a complex mixture. However, the purity of the product is governed by the specificity of the protein-ligand interaction. Immunoglobulins (IgGs) are by far the largest class of protein biopharmaceuticals and conventional ligands such as Protein A and Protein G are traditionally used for their purification. A complex cell lysate containing IgG is loaded on a column containing Protein A or Protein G conjugated resin and the bound IgG is eluted using a low pH buffer solution. Such harsh pH elution conditions can lead to loss of activity of the IgG and may necessitate an additional refolding step [3]. Therefore, there is a need for generating new ligands that are low cost, stable, and specific and provide milder elution conditions.

Generation and discovery of affinity ligands for protein purifications is an extremely active area of research [2, 4]. Affinity ligands such as peptides and small proteins are promising because of higher stability and lower cost relative to large multi-domain protein ligands such as antibodies. Specific and reversible association of the affinity ligand with the target protein is the mainstay of an affinity chromatography separation. Higher affinity for the target protein may guarantee specific association and separation from complex mixtures, but

reversibility of the association ensures recovery of the target protein from an affinity chromatography resin. Therefore, moderate affinity whereby specificity is ensured and recovery can be maximized would be the desired attribute of any affinity ligand. Affinity should be high enough to guarantee selectivity and also maintain high binding capacity of the target protein. On the contrary, too high an affinity may require multiple elution steps to completely recover the IgG product.

Protein A, the most widely used affinity ligand for immunoglobulin purifications; specifically binds the Fc portion of human immunoglobulin G (hFc) [5, 6]. Therefore, it has been used extensively in the biopharmaceutical industry for human or humanized antibody purifications. However, only three (IgG1, IgG2 and IgG4) of the four human IgG(hIgG) subclasses are recognized[7]. Protein G, on the other hand has broader specificity and binds intact immunoglobulin as well as the Fab2 and Fc regions[8]. Interestingly, Protein G also recognizes three (IgG1, IgG2 and IgG3) of the four hIgG subclasses. More recently, Protein A/G fusion has been created that identifies all four human subclasses and is therefore useful for purification processes where the subclass is unknown [9, 10]. Protein L, though less commonly used; binds to the framework regions of the IgG variable light chain[11].

A low pH for elution of IgGs may cause denaturation or aggregation of the IgG product and necessitate further refolding steps making the overall process expensive and complicated [3]. Therefore, ligands that can be used to elute IgGs at a somewhat higher pH than currently used are of great value. Engineered versions of Protein A or Protein G have been created with greater stability and mild elution capabilities [12-14]. Additionally, a great deal of

research is in progress to discover novel protein/peptide ligands in addition to other synthetic ligands for protein purification. Human Fc binding peptides have been engineered using phage display [15, 16]. N-terminal histidine based hexapeptides HWRGWV, HYFKFD and HFRRHL that selectively bind hFc have also been demonstrated as ligands for purification of IgGs[17]. HWRGWV binds all hIgG subclasses and also IgA and IgM[18]. It has broad species specificity towards immunoglobulins from goat, mouse, rabbit and cow. Chimeric and humanized monoclonal antibodies from commercial cell culture media were purified using HWRGWV to a purity of 94%[19].

An hFc binding protein called as protein Z has been engineered to allow elution at mild pH conditions such as pH 4.5[20]. Protein Z is derived from the B-domain of Staphylococcal protein A and is also known as the 'Affibody'. Apart from IgG binding, affibody was also engineered using phage display to broaden its specificity and bind IgA[21]. Engineered single domain camelid antibodies based on a single variable heavy chain have also been used as ligands for IgG purifications [22, 23]. These ligands were isolated by screening an immune phage display library. These scaffolds are small in size (13-15 kDa) and can be produced at high yields using microbial systems [24]. Capture select affinity ligands based on camelid single domain antibodies are commercially available. Carbohydrate binding proteins such as lectins can also be used for purifications of IgGs which possess an N-linked glycosylation[25]. Binding of IgG to lectins takes place at neutral pH and elution can be carried out by using high sugar concentrations (0.1-0.5 M). There is also an increasing interest in additives to elution that can help elute as well as stabilize IgGs. Arginine is very

commonly used to prevent aggregation and allow mild elution conditions [26-29]. Salts such as NaCl and $Na_2SO_4$ can also help in reducing aggregation.

Most protein or peptide ligands suffer from limitations such as low stability, higher manufacturing costs and inability to withstand harsh elution conditions[25]. High thermal and chemical stability are desirable for the harsh clean in place sterilization conditions. Recently, we have shown that Sso7d-derived affinity reagents have high thermal and chemical stability and can withstand pH extremes[30]. Therefore, we reasoned that affinity ligands based on the Sso7d scaffold would be well suited for affinity chromatography.

Three different approaches were taken to obtain ligands with different biophysical properties. Firstly, an affinity based ligand (Sso7d-hFc) where selection was solely based on higher affinity was developed by screening a yeast surface displayed library of the Sso7d protein scaffold. Higher affinity can guarantee specificity and higher binding capacity but elution of the IgG product is difficult. Further, this affinity based ligand was engineered for pH sensitivity using a rational approach of histidine scanning to generate a pH sensitive version (Sso7d-his-hFc). In addition, pools of mutants derived from different rounds of selection of the affinity based ligand were recombined using error prone PCR to create a random yeast surface displayed library. This library was screened for higher binding affinity (positive selection) at physiological pH (pH 7.4) and lower binding affinity (negative selection) at pH 4.5 to generate another pH sensitive ligand (Sso7d-ev-hFc).

## 5.2. Results

### 5.2.1. Construction and screening of Sso7d library using yeast surface display

Binding proteins for hFc were isolated from the Sso7d yeast surface display library described previously[30]. The screening process comprised of magnetic selection followed by four rounds of flow cytometry selection. Magnetic screening enables negative selection which allows generation of highly specific binders for hFc. Negative selection was done with biotin binder beads followed by negative selection against chicken IgY(cIgY), mouse IgG( mIgG) and rabbit IgG(rIgG) proteins. The fractionated cells were further subjected to flow cytometry selections to isolate a pool of mutants with the best binding affinities (Table 5.1). Three distinct hFc binding Sso7d clones were obtained. All the selections were done in the presence of PBS-BSA at pH 7.4. We selected Sso7d-hFc since it had the highest affinity for hFc compared to other mutants in the pool.

### 5.2.2. Column assay for human-IgG purification using Sso7d-hFc

As a preliminary test to evaluate the ability of Sso7d-hFc as an affinity ligand, we used the 6X-his-tag on the purified protein as a means to immobilize the protein on a Ni-NTA column used for traditional IMAC purifications. Briefly, purified Sso7d-hFc was immobilized by sample loading on a Ni-NTA column. After washing with five column volumes of Buffer A (50 mM Tris pH 7.4, 300 mM NaCl), 2 ml of complete minimal essential medium (cMEM) with 3 mg/ml of hIgG was loaded on the prepared column. The column was again washed with five column volumes of Buffer A. Sso7d-hFc bound to hIgG was eluted by using a

linear gradient from 0-100% Buffer B(50 mM Tris pH 7.4, 300 mM NaCl, 500 mM imidazole). As shown in Figure 5.1 the elution fractions show that Sso7d-hFc binds specifically to hIgG in complex media.

In large scale chromatographic separations, the above described method is infeasible particularly because of the histidine tag based immobilization which would not withstand harsh regeneration and elution conditions. Additionally, this necessitates an extra separation step for separating hIgG from the Sso7d-hFc ligand thereby making the overall process time-consuming and expensive. A process where the binding of hIgG in complex medium takes place at physiological pH (pH 7.4) and can be eluted by at a lower pH(pH > 4) would be useful. The mild elution condition (pH>4) used will assure intact and stable hIgG product. Therefore, an ideal ligand would be one that binds at pH 7.4 and has a greater off-rate in a lower pH buffer.

### 5.2.3. pH sensitivity of Sso7d-hFc

We wanted to assess the pH sensitivity of Sso7d-hFc to validate use in large scale chromatographic purifications. Yeast surface displayed Sso7d-hFc was incubated with 100 nM biotinylated hFc in P1-BSA (100 mM sodium phosphate buffer pH 7.4) for 15 min at 4˚C. Further, labeled cells were washed with P1-BSA and incubated with 1 ml of P2-BSA (100 mM sodium phosphate buffer pH 4.5) with shaking at room temperature. A control sample, where the cells were incubated in 1 ml of P1-BSA was also prepared. After 30 min of incubation the cells were labeled with Strep-PE and analyzed on the flow cytometer. A higher off rate ($k_{off}$) will deplete the fraction of bound hFc from the complex and reduce the

fluorescence signal. The results are shown in Figure 5.2. The binding of hFc to the yeast surface displayed Sso7d-hFc protein stays essentially constant in buffers of pH 7.4 and 4.5, confirming that the off rate is unchanged and the binding is not pH sensitive. Therefore, the Sso7d-hFc protein might not be useful in a large scale purification process unless harsh elution conditions are used. However, we further investigated if we could generate pH sensitive affinity ligands against hIgG that would elute at milder conditions.

### 5.2.4. Engineering pH sensitive binding against hFc

Two different approaches were taken for engineering pH sensitive affinity ligands for hIgG. The first method was a more rational approach of histidine scanning mutagenesis, where each amino acid in the binding interface of Sso7d-hFc was mutated to a histidine and the pH sensitivity was assessed at pH 7.4 and 4.5. The second was a directed evolution approach where a random yeast surface display library was created with DNA isolated from different flow cytometry sorts for Sso7d-hFc and pH sensitive binding proteins were isolated.

#### 5.2.4.1. Histidine Scanning

We wanted to engineer a pH sensitive hFc binding protein so that the hIgG could be eluted at mild pH. To this end, we performed a histidine scanning mutagenesis of the Sso7d-hFc mutant. Simplistically, the ten mutations in the binding interface of Sso7d were replaced with histidine one at a time and the pH sensitivity of the binding was tested. The rationale behind this study was that the histidine side chain has a pKa of about 6.8 and therefore at a lower pH it would be positively charged and repel any positively charged interactions in the binding

interface. This approach has been previously described in generating a pH sensitive protein G mutant and also in antibody recycling in endosomes [31, 32]

The histidine scanning mutants were generated using a combination of overlap extension PCR and homologous recombination in yeast. Sso7d-hFc histidine scanning variants were first labeled with 100 nM of hFc-biotin for 15 min in P1-BSA at pH 7.4, washed with P1-BSA and then further incubated in P2-BSA at pH 4.5 for 30 min with shaking. The secondary incubation was performed as described in the previous section. Three out of the ten mutations (L20H, M32H and I42H) led to loss of binding. Interestingly, only one mutation Y40H showed significant sensitivity to pH (Figure 5.3). This mutant was selected for further analysis and denoted as Sso7d-his-hFc.

#### 5.2.4.2. Directed Evolution

We used a directed evolution approach to isolate binding proteins with pH sensitivity. DNA from the magnetically sorted (low affinity) population in the Sso7d-hFc sorts and also a high affinity population from the later sorts was isolated to create a new randomized library. Briefly, error-prone PCRs with varying random oligonucleotides (8-oxo-dGTP and dPTP) concentrations and different cycles were performed on plasmid DNA isolated from yeast cells. An error prone PCR library with a diversity of $1x\ 10^7$ was created in yeast surface display.

A two-step library screening strategy consisting of magnetic screening followed by flow cytometry was employed. In the magnetic screening step, after performing a negative

selection with biotin binder beads, a positive selection for binding to hFc was performed in P1-BSA at pH 7.4. A magnetic field was applied to separate the cell bound beads from unbound cells. After washing twice with P1-BSA, the cell bound beads were incubated with shaking in P2-BSA pH 4.5. The eluted cells were collected and grown in SDCAA media for two days.

Further sorting was performed using flow cytometry by alternating between a sort at pH 7.4 for binding to hFc followed by a sort at pH 4.5(where the cells were allowed to dissociate in P2-BSA as described earlier) where the non-binding but surface displaying cells were collected. The schematic in Figure 5.4 describes this process. Three rounds of flow cytometry were performed and the population was plated. The sequences of clones from the final population are shown in Table 5.2. The pH sensitivity of these evolution mutants was tested by labeling yeast surface displayed Sso7d-ev-hFc with 2000 nM of hFc-biotin in P1-BSA at pH 7.4 and then allowing the bound hFc to dissociate in P2-BSA pH 4.5. The results are shown in Figure 5.5. Sso7d-ev-hFc was selected as the best binding clone based on its ability to have the maximum change in binding between pH 7.4 and pH 4.5.

### 5.2.5. Specificity Analysis of Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc

The specificity of binding of Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc to other closely related IgGs and Fab fragments was evaluated through a flow cytrometric assay. Yeast cell surface displayed Sso7d-hFc was labeled at a high concentration (1 μM) of hFc, mIgG, cIgY, rIgG, goat-IgG(gIgG), donkey-IgG(dIgG), Fab and Fab2. As shown in Figure 5.6a, Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc bind specifically to hFc with high affinity. However,

Sso7d-ev-hFc shows some binding to rIgG as well. All three mutants do not bind the Fab and Fab2 fragments (Figure 5.6b).

### 5.2.6. Soluble Expression of Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc

Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc were all recombinantly expressed in the *E.coli* cytoplasm with a C- terminal 6xHistidine tag[30]. The recombinant proteins were purified in a single step by Immobilized Metal Affinity Chromatography (IMAC) using a Nickel column. Estimates of purified protein yield ranged from 40-50 mg per liter of bacterial culture. The purified proteins were dialyzed in 50 mM phosphate buffer with 300 mM NaCl at pH 7.4 and pH 4.5. Protein concentrations were measured using bicinchoninic acid (BCA) assay with Bovine Serum Albumin (BSA) as a standard. Analytical size exclusion chromatography experiments were performed to confirm the oligomeric state of the binding proteins at pH 7.4 and 4.5. It was found that all three mutant proteins exist as monomers at both pH values

### 5.2.7. Binding to human IgG subclasses

A western blot was performed to evaluate the binding of Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc to the hIgG subclasses hIgG1, hIgG2, hIgG3 and hIgG4. Soluble biotinylated proteins Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc were used as primary reagents in the western blot experiments and an anti-biotin antibody conjugated to HRP was used as the secondary reagent. As shown in Figure 5.7, all mutants recognize all four hIgG subclasses.

Finally, a PNGase digestion was performed on the hIgG and binding was again probed on a western blot with the variants and anti-biotin-HRP. All three mutants recognize the digested hIgG confirming that the binding to hIgG is not carbohydrate mediated (data not shown).

## 5.3. Discussion

Affinity chromatography is particularly useful as a rapid purification step in industrial antibody manufacturing. However, stable affinity ligands with low handling cost are in increasing demand. Hyperthermophilic organisms grow in extreme conditions of temperature and pH and therefore hyperthermophilic proteins are promising candidates as affinity ligands due to their inherent stability under these conditions. Recently, we validated the Sso7d protein from *Sulfolobus solfataricus* as a scaffold for molecular recognition where binding proteins for six different target proteins of varying size and topology were generated from a yeast surface display library of $10^8$ mutants[30]. Further, we wanted to investigate whether this scaffold can be used to generate binding proteins for the Fc portion of hIgG; which can serve as a universal ligand for purification of human, humanized or chimeric antibodies.

We engineered pH sensitive mutants against hIgG which would be ideal in a purification process where a low pH is used for elution. Previous studies have shown that a low pH of elution can cause aggregation or denaturation of IgG which may increase the overall cost and complexity of the process [3].

pH sensitive binding is usually governed by ionizable amino acid residues such as histidine, aspartate and glutamine. Natural proteins such as hemoglobin show pH dependent binding to their cognate antigens [33, 34].pH induced structural or allosteric changes and signaling interactions have also been observed or engineered.[35-40].The human prolactin signaling protein shows a 500-fold decrease in binding affinity to its receptor when the pH is changed from 8 to 6[41]. This pH sensitivity is also regulated by two histidine residues[42]. In a

recent study, pH sensitive protein G variants were engineered by introduction of histidines in the binding interface of Protein G and hIgG[31]. hIgG could be eluted from this engineered protein G at a pH of 4.3. Additionally a histidine scanning library was created using histidine and other wild type residues using a camelid single domain VHH antibody to generate a pH sensitive binding interaction with RNase A[43] . Glutathione-S-transferase was engineered to create a pH dependent allosteric mutant to be used as a pH dependent ligand in affinity chromatography [44].

A high affinity binding protein against the human Fc portion of IgG could be isolated from the yeast surface display library of $10^8$ mutants. However, binding of this protein to human IgG was not found to be pH sensitive. Ideally, an affinity ligand with a low off rate ($k_{off}$) at pH 7.4, and a relatively higher off rate at pH 4.5 is desirable. Higher $k_{off}$ at pH 4.5 will allow dissociation of the IgG from the immobilized ligand and hence elution. We further investigated if we could engineer this pH dependent off-rate phenomenon in two scenarios a) by utilizing a high affinity binding protein already isolated from a yeast surface displayed library and rationally introducing pH sensitivity using histidine scanning or b) by screening a random yeast surface display library for pH sensitive mutant proteins generated from a combination of low and high affinity binding Sso7d mutants.

Our histidine scanning studies revealed that three mutations (L20H, M32H and I42H) lead to loss of binding to hIgG with M32H and I42H causing complete loss of binding. Therefore, these mutations play a role in binding to hIgG. Interestingly, one out of the ten mutations

Y40H leads to pH dependent binding where the binding is reduced when the hFc bound yeast surface displayed Sso7d complex is allowed to dissociate in a buffer of pH 4.5.

A directed evolution approach was further undertaken to generate pH dependent binding. A library of low and high affinity mutants was subjected to random mutagenesis using nucleotide analogues to generate a second generation library. This library was screened by sorting for mutants with higher off rates after isolating a pure binding population. The final pool of mutants were sequenced and found to have one-two mutations in addition to the ten mutations in previously selected positions. These additional mutants were charged residues such as glutamic acid or arginine, which implies that the corresponding epitope on the human IgG could also be a charged residue that might mediate the pH sensitive binding.

All the three mutants (Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc) specifically recognize human Fc portion, while only Sso7d-ev-hFc recognizes rIgG to some extent. The initial screening for Sso7d-hFc involved a stringent negative selection step against other closely related immunoglobulins such as cIgY, mIgG and rIgG which finally leads to its specificity. However, introduction of random mutations at a later step to generate a second generation library might not yield a specific mutant as the negative selection step was not performed. Not surprisingly, Sso7d-ev-hFc also recognizes rIgG to some extent when the specificity was examined. Yeast displays multiple protein copies on its surface and therefore the negative selection step leads to depletion of non-specific binding proteins owing to the avid nature of fractionation. All three mutant proteins also specifically bound the hIgG in comparison to the Fab and Fab2 regions.

Binding to all human IgG subclasses (hIgG1-4) was investigated using western blotting analysis. Protein A and Protein G both do not recognize all four human IgG isotypes. The hIgG isotype 3 has been shown to be a marker of active systemic lupus erythematosus, particularly lupus nephritis [45]. In treatment of certain autoimmune diseases, removal of circulating autoantibodies is essential and therefore affinity ligands against IgG3 would be useful [46]. The Sso7d based affinity ligands (Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc) recognize all four hIgG isotypes and therefore would be useful for immune-adsorption applications in the above mentioned diseases.

Elucidating the binding epitope of these ligands might suggest ways to improve the elution conditions. As mentioned before, lectins bind the glycosylated regions in IgG's and therefore can be eluted using high sugar concentrations. We further investigated the binding of these ligands to glycosylation free hIgG. PNGase digestion was carried out to remove the N-linked glycosylation of the human IgG and the digested glycoprotein was then run on a western blot to assess binding to Sso7d mutants. The binding was retained even after digestion, therefore confirming that this binding is not carbohydrate mediated.

Using a combination of the specificity data, western blotting and sequence alignments, we proposed a method to identify potential binding epitopes of these Sso7d based ligands on human IgG. Sequence alignments for hIgG subclasses, mouse IgG and rabbit IgG were performed using UniProt and the results are shown in Figure 5.8. Based on the specificity data for Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc we identified several peptide stretches which might be a part of putative binding sites. Furthermore, our western blot experiments

suggest that the hIgG epitope is at least partially (if not completely) linear, which enables binding of Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc to the denatured IgG subclasses in a western blot.

We also expect that the binding epitope for Sso7d-his-hFc would be the same as Sso7d-hFc. It is also reasonable to assume that there should be a charged residue on the epitope because the binding is pH sensitive. Since histidine should be positively charged at pH 4.5, there should be another positive charge causing repulsion and therefore LDSD is not the likely epitope for Sso7d-hFc and Sso7d-his-hFc. However, the pKa of aspartic acid side chain is 4.0 and since protein microenvironments can have substantial effects on the intrinsic pKa of any side chain, this assumption may not hold true. The binding site for Sso7d-ev-hFc may be different from Sso7d-hFc and Sso7d-his-hFc because of no rIgG binding in these cases.

The Sso7d-based affinity ligands isolated in this study can be used to elute hIgG at mild elution conditions in affinity chromatography. Further information on the binding site of these ligands on hIgG would also be useful in designing elution conditions. Effect of additives such as high salt and arginine can also be examined to improve elution. Kinetic and thermodynamic data providing information about the effects of mutations on binding and stability would also be useful.

## 5.4. Materials and Methods

### 5.4.1. Sso7d library screening against human IgG

The Sso7d library was generated as described previously[30]. The screening process involved magnetic screening followed by fluorescence activated cell sorting. Magnetic screening was performed as described[47]. Briefly, $10^7$/ml of yeast cells grown in SDCAA(20g/L dextrose, 5g/L casamino acids, 6.7g/L yeast nitrogen base, 5.40 g/L $Na_2HPO_4$, 7.45 g/L $NaH_2PO_4$) were induced in SGCAA(20g/L galactose, 5g/L casamino acids, 6.7g/L yeast nitrogen base, 5.40 g/L $Na_2HPO_4$, 7.45 g/L $NaH_2PO_4$). Cells were grown at 20˚C and 250 rpm for 20-24 hours for protein induction on yeast cell surface. 100 µl of Dynal™ biotin binder beads (4 x $10^8$ beads/ml, Invitrogen, Carlsbad, CA) were precoated with biotinylated human Fc protein (Jackson Immunoresearch, Westgrove, PA) overnight at 4˚C. 2 x$10^9$ cells (20X library diversity) were incubated with biotin binder beads for negative selection for 1 hour at 4˚C. Bead bound cells were discarded and further negative selection was performed against mIgG, cIgY and rIgG. Finally, the unbound cells were used for a positive selection against hFc coated beads for 1 hour at 4˚C. The bead bound cells were washed four times with PBS-BSA(8g/L NaCl, 0.2g/L KCl, 1.44 g/L $Na_2HPO_4$ , 0.24g/L $KH_2PO_4$, pH 7.4 containing 0.1 % BSA) and resuspended in SDCAA. The cells were grown for 24-48 hours at 30˚C and 250 rpm.

Following magnetic selection the cells were expanded and induced again at $10^7$ cells/ml in SGCAA for flow cytometric sorting. Sorting was performed using previous protocols to select for the highest affinity binders for Sso7d-hFc[48]. For Sso7d-ev-hFc, a pure population

of cells binding at pH 7.4 in P1-BSA(100 mM sodium phosphate buffer pH 7.4) was obtained and in further sorts an off rate based selection was carried out where yeast displayed protein-biotinylated hFc complexes were allowed to dissociate in P2-BSA(100 mM sodium phosphate buffer pH 4.5) at pH 4.5. Subsequently, the cells that lost binding to hFc but were still expressing the c-myc tag were sorted. hFc binding was indirectly detected using Strep-PE (Invitrogen, Carlsbad, CA). The expression was monitored by c-myc tag labeling using an anti-c-myc chicken antibody followed by a goat anti-chicken antibody conjugated to Alexa Fluor-633 (Invitrogen, Carlsbad, CA). Cell sorting was performed on a FACS Aria (Becton Dickinson, San Jose, CA) flow cytometer. Four rounds of sorts were performed for Sso7d-hFc. In case of Sso7d-ev-hFc, two rounds of FACS sorts were performed to get a pure population and two further off-rate sorts were performed. Final sorts were plated on SDCAA plates and clones picked for sequencing.

### 5.4.2. Sequence determination of Sso7d based affinity ligands

Plasmid DNA was isolated from the Sso7d mutants using the Zymoprep Kit II (Zymoresearch Corporation, Orange, CA). The DNA was further transformed into Novablue$^{TM}$(*E.coli*) cells (EMD Biosciences, San Diego, CA). A Qiagen miniprep kit was used to isolate plasmids from *E.coli* (Qiagen, Valencia, CA) and sequenced by Genewiz (South Plainfield, NJ). The following primer sequence was used for sequencing- 5’ ACT ACG CTC TGC AGG CTA GT 3’

### 5.4.3. Histidine Scanning Analysis

The pCTCON plasmid carrying the Sso7d-hFc gene was used as a template for creating the histidine scanning mutants. Briefly, two fragments were created from the Sso7d-hFc gene using internal primers that introduced the histidine mutation one by one at ten previously mutated positions. The complete gene for the histidine scanning mutants was obtained by homologous recombination where the two fragments share 30-40 bp homology with each other and 50 bp homology with the pCTCON vector. Linearized pCTCON vector was prepared by digestion with *NheI* and *Bam HI* restriction enzymes (New England Biolabs, Ipswich, MA). Homologous recombination mediated plasmid gap repair facilitated the ligation of two fragments with linearized pCTCON vector. The identity of the mutants was confirmed by DNA sequencing.

### 5.4.4. Random Mutagenesis library construction

Plasmid DNA was isolated using Zymoprep Kit II (Zymoresearch Corporation, Orange, CA) from two different stages in the sorting of the Sso7d library for Sso7d-hFc. A low affinity magnetically screened pool and a higher affinity pool after two round of FACS were used as a template for random mutagenesis. Error prone PCR's were performed for each pool using the 8-oxo-dGTP and dPTP nucleotide analogues (Trilink Biotechnologies, San Diego, CA) as described [49, 50]. Four different combinations of PCR cycles and nucleotide analogue concentrations were used: 10 cycles with 10 μM each of 8-oxo-dGTP and dPTP, 20 cycles with 10 μM each of 8-oxo-dGTP and dPTP, 20 cycles with 2 μM each of 8-oxo-dGTP and dPTP, 30 cycles with 2 μM each of 8-oxo-dGTP and dPTP. A 50 μl PCR reaction mixture

consisted of 1X Taq Buffer without $MgCl_2$, 2 mM $MgCl_2$, primers Pf1 and Pr1 0.1 μM each, 200 μM dNTPs, 20 ng template DNA and 0.05 U/μl of Taq DNA polymerase(Invitrogen, Carlsbad, CA).The primer sequences were: Pf1 - 5'AGT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT *GCT AGC* ATG GCG ACC GTG AAA TTT AAA TAT AAA G 3' and Pr1 - 5'CTC GAG CTA TTA CAA GTC CTC TTC AGA AAT AAG CTT TTG TTC *GGA TCC* TTT TTT CTG TTT TTC CAG CAT CTG 3'; restriction sites are italicized. The mixture was denatured at 94˚C for 3 min followed by 10, 20 or 30 cycles of 94˚C for 45 s, 58˚C for 30 s, 72˚C for 90 s and a final extension step at 72˚C for 10 min. The PCR products were PCR purified using Qiagen PCR purification kit (Qiagen, Valencia, CA) and then 200 ng of each product was mixed to create a combined DNA pool from all different reactions. This mixed product was further used as a template for amplification PCR with following components: 1X high fidelity Phusion buffer, 0.1 μM primers, 200 μM dNTPs and Phusion high fidelity DNA polymerase (1U/ 50 μl, New England Biolabs, Ipswich, MA). PCR reaction was performed with an initial denaturation at 98˚C for 2 min and 30 cycles of 98˚C for 1 min, 61˚C for 30 s, 72˚C for 90 s and a final extension step at 72˚C for 10 min. PCR products were concentrated using Pellet Paint™ (Novagen, San Diego, CA).

Homologous recombination mediated gap repair was used to generate a yeast surface display library. 1 μg of linearized pCTCON plasmid with 3 μg of insert was transformed in *Sacchromyces cerevisiae* strain EBY100 as described[30]. The electroporation was performed using a Biorad Gene Pulser X cell system (Bio-Rad, Hercules,CA) at 0.54 kV, 25 μF and 1000 Ω. Transformed EBY 100 cells were grown in YPD medium (10g/L yeast

extract, 20g/L peptone, 20g/L dextrose) for 1 hour at 30°C and 250 rpm. Serial dilutions were plated on SDCAA plates (20g/L dextrose, 5g/L casamino acids, 6.7g/L yeast nitrogen base, 182 g/L sorbitol, 5.40 g/L $Na_2HPO_4$, 7.45 g/L $NaH_2PO_4$, 15g/L agar) to estimate the library diversity of about $10^7$ mutants. The rest of the library was grown in SDCAA medium (20g/L dextrose, 5g/L casamino acids, 6.7g/L yeast nitrogen base, 5.40 g/L $Na_2HPO_4$, 7.45 g/L $NaH_2PO_4$) for 24-48 hours with 1:100 pen-strep solution (Invitrogen. Carlsbad, CA). The library was passaged twice in SDCAA medium before proceeding to negative selection using magnetic screening (as described above).

### 5.4.5. Specificity Analysis

$10^7$ cells/ml yeast cells carrying Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc plasmids were induced in SGCAA. After 20-24 hours of induction at 20˚C and 250 rpm the yeast surface displayed proteins were labeled with high concentrations (1 μM) of hFc, mIgG, cIgY, rIgG, goat-IgG(gIgG), donkey-IgG(dIgG), Fab and Fab2( Jackson Immunoresearch, Westgrove, PA) . The cells were analyzed on a BD FACS Aria (Becton Dickinson, San Jose, CA) flow cytometer.

### 5.4.6. Recombinant expression and purification of affinity ligands

Sso7d mutants were cloned into pET22b(+) vector as previously described[30]. One liter of 2XYT medium was inoculated with a 5 ml overnight culture of Rosetta Cells harboring the corresponding plasmid for the Sso7d mutant. At an $OD_{600}$ of 0.7 the cells were induced with 0.5 mM of IPTG and grown for 19-20 hours at 37˚C and 250 rpm. The cells were harvested

by centrifugation at 4700 rpm for 20 min and supernatant was discarded. Cell pellet was resuspended in Buffer A (50 mM Tris pH 7.4, 300 mM NaCl) with 2mM PMSF. Cell lysis was performed by sonication for 10 min. The lysed cells were centrifuged at 15000 rpm and the supernatant was filtered with a 0.22 µm filter. The filtered sample was loaded onto a 5 ml Bio-Scale™Mini Profinity™ IMAC Cartridge( Biorad, Hercules, CA) and purified on a Bio-rad Biologic LP system (Bio-Rad, Hercules, CA). Sso7d variants with a 6X-his tag were eluted using a linear gradient of Buffer A and B (50 mM Tris pH 7.6, 300 mM NaCl, 500 mM Imidazole). The collected fractions were analyzed by SDS-PAGE analysis using Novex 10% Bis-Tris Gels (Invitrogen, Carlsbad, CA) and the pure protein fractions were pooled together. Dialysis was carried out at two different pH values (pH 7.4 and pH 4.5) with 50 mM sodium phosphate buffer and 300 mM NaCl. The protein concentrations were determined using a BCA assay with BSA as the standard.

### 5.4.7. Western Blotting Analysis

Three identical 8% SDS-PAGE gels were run with the four human IgG isotypes-hIgG1,hIgG2, hIgG3 and hIgG4 (Sigma Aldrich, St. Louis, MO) the composite hIgG (Equitech bio, Kerrville, TX), PNGase F (New England Biolabs, Ipswich, MA) digested hIgG and an undigested control using standard procedures as described. The proteins were transferred onto a PVDF membrane (Whatman, a part of GE Healthcare Bio-Sciences, Piscataway, NJ) using a semi-dry blotting unit (Fisher Scientific, Pittsburgh, PA) with 1X Transfer buffer(25 mM Tris, 192 mM glycine, 10% methanol, 0.5% SDS). The membrane was blocked with 5% non-fat dry milk (Lab scientific, Livingston, NJ) in 1X TBS-T(0.01M

Tris HCl, 0.15M NaCl, 0.05% Tween-20). Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc were biotinylated using EZ-Link Sulfo-NHS-LC biotinylation kit (Thermo Scientific, Rockford, IL) and the number of biotins per protein was between 3 and 4. These biotinylated proteins were used as primary reagents in the western blot and incubated overnight at 4˚C. Secondary labeling was performed using an anti-biotin-HRP conjugated antibody (Cell Signaling Technology Inc, Danvers, MA) for 1 hour. The blot was developed using a SuperSignal West Femto Chemiluminescent Substrate (Thermo Scientific, Rockford, IL) and imaged on a Bio-Rad chemiluminescence Imager.

### 5.4.8. Column Assay using Sso7d-hFc

Purified and his-tagged Sso7d-hFc protein was loaded on a 5ml Bio-Scale™Mini Profinity™ IMAC Cartridge( Biorad, Hercules, CA). The column was then washed with Buffer A. A complex mixture was prepared using 3 mg/ml of hIgG in 2ml cMEM medium (Invitrogen, Carlsbad, CA) which was loaded on the column and washed again with Buffer A. Finally, a linear gradient with buffer B was used to elute the complex of hIgG and Sso7d-hFc. Fractions were collected throughout the process on a Biologic LP system fraction collector and analyzed on an SDS-PAGE Gel.

## 5.5. References


1. Roque, A.C., C.R. Lowe, and M.A. Taipa, *Antibodies and genetically engineered related molecules: production and purification.* Biotechnol Prog, 2004. **20**(3): p. 639-54.

2. Roque, A.C., C.S. Silva, and M.A. Taipa, *Affinity-based methodologies and ligands for antibody purification: advances and perspectives.* J Chromatogr A, 2007. **1160**(1-2): p. 44-55.

3. Shukla, A.A., P. Gupta, and X. Han, *Protein aggregation kinetics during Protein A chromatography. Case study for an Fc fusion protein.* J Chromatogr A, 2007. **1171**(1-2): p. 22-8.

4. Linhult, M., S. Gulich, and S. Hober, *Affinity ligands for industrial protein purification.* Protein Pept Lett, 2005. **12**(4): p. 305-10.

5. Sjoquist, J., *Structure and immunology of protein A.* Contrib Microbiol Immunol, 1973. **1**: p. 83-92.

6. Hjelm, H., K. Hjelm, and J. Sjoquist, *Protein A from Staphylococcus aureus. Its isolation by affinity chromatography and its use as an immunosorbent for isolation of immunoglobulins.* FEBS Lett, 1972. **28**(1): p. 73-6.

7. Kronvall, G. and R.C. Williams, Jr., *Differences in anti-protein A activity among IgG subgroups.* J Immunol, 1969. **103**(4): p. 828-33.

8. Bjorck, L. and G. Kronvall, *Purification and some properties of streptococcal protein G, a novel IgG-binding reagent.* J Immunol, 1984. **133**(2): p. 969-74.

9. Eliasson, M., et al., *Chimeric IgG-binding receptors engineered from staphylococcal protein A and streptococcal protein G.* J Biol Chem, 1988. **263**(9): p. 4323-7.

10. Eliasson, M., et al., *Differential IgG-binding characteristics of staphylococcal protein A, streptococcal protein G, and a chimeric protein AG.* J Immunol, 1989. **142**(2): p. 575-81.

11. Akerstrom, B. and L. Bjorck, *Protein L: an immunoglobulin light chain-binding bacterial protein. Characterization of binding and physicochemical properties.* J Biol Chem, 1989. **264**(33): p. 19740-6.

12. Gulich, S., et al., *Engineering streptococcal protein G for increased alkaline stability.* Protein Eng, 2002. **15**(10): p. 835-42.

13. Bottomley, S.P., B.J. Sutton, and M.G. Gore, *Elution of human IgG from affinity columns containing immobilised variants of protein A.* J Immunol Methods, 1995. **182**(2): p. 185-92.

14. Palmer, B., et al., *Design of stability at extreme alkaline pH in streptococcal protein G.* J Biotechnol, 2008. **134**(3-4): p. 222-30.

15. DeLano, W.L., et al., *Convergent solutions to binding at a protein-protein interface.* Science, 2000. **287**(5456): p. 1279-83.

16. Krook, M., K. Mosbach, and O. Ramstrom, *Novel peptides binding to the Fc-portion of immunoglobulins obtained from a combinatorial phage display peptide library.* J Immunol Methods, 1998. **221**(1-2): p. 151-7.

17. Yang, H., P.V. Gurgel, and R.G. Carbonell, *Hexamer peptide affinity resins that bind the Fc region of human immunoglobulin G.* J Pept Res, 2005. **66 Suppl 1**: p. 120-37.

18. Liu, Z., P.V. Gurgel, and R.G. Carbonell, *Effects of peptide density and elution pH on affinity chromatographic purification of human immunoglobulins A and M.* J Chromatogr A, 2011. **1218**(46): p. 8344-52.

19. Naik, A.D., et al., *Performance of hexamer peptide ligands for affinity purification of immunoglobulin G from commercial cell culture media.* J Chromatogr A, 2011. **1218**(13): p. 1691-700.

20. Gulich, S., M. Uhlen, and S. Hober, *Protein engineering of an IgG-binding domain allows milder elution conditions during affinity chromatography.* J Biotechnol, 2000. **76**(2-3): p. 233-44.

21. Ronnmark, J., et al., *Human immunoglobulin A (IgA)-specific ligands from combinatorial engineering of protein A.* Eur J Biochem, 2002. **269**(11): p. 2647-55.

22. Klooster, R., et al., *Improved anti-IgG and HSA affinity ligands: clinical application of VHH antibody technology.* J Immunol Methods, 2007. **324**(1-2): p. 1-12.

23. Zandian, M. and A. Jungbauer, *Engineering properties of a camelid antibody affinity sorbent for Immunoglobulin G purification.* J Chromatogr A, 2009. **1216**(29): p. 5548-56.

24. Frenken, L.G., et al., *Isolation of antigen specific llama VHH antibody fragments and their high level secretion by Saccharomyces cerevisiae.* J Biotechnol, 2000. **78**(1): p. 11-21.

25. Ayyar, B.V., et al., *Affinity chromatography as a tool for antibody purification.* Methods, 2011.

26. Tsumoto, K., et al., *Role of arginine in protein refolding, solubilization, and purification.* Biotechnol Prog, 2004. **20**(5): p. 1301-8.

27. Ejima, D., et al., *Effective elution of antibodies by arginine and arginine derivatives in affinity column chromatography.* Anal Biochem, 2005. **345**(2): p. 250-7.

28. Tsumoto, K., et al., *Review: Why is arginine effective in suppressing aggregation?* Protein Pept Lett, 2005. **12**(7): p. 613-9.

29. Baynes, B.M., D.I. Wang, and B.L. Trout, *Role of arginine in the stabilization of proteins against aggregation.* Biochemistry, 2005. **44**(12): p. 4919-25.

30. Gera, N., et al., *Highly stable binding proteins derived from the hyperthermophilic sso7d scaffold.* J Mol Biol, 2011. **409**(4): p. 601-16.

31. Watanabe, H., et al., *Optimizing pH response of affinity between protein G and IgG Fc: how electrostatic modulations affect protein-protein interactions.* J Biol Chem, 2009. **284**(18): p. 12373-83.

32. Igawa, T., et al., *Antibody recycling by engineered pH-dependent antigen binding improves the duration of antigen neutralization.* Nat Biotechnol, 2010. **28**(11): p. 1203-7.

33. Kilmartin, J.V., J.H. Fogg, and M.F. Perutz, *Role of C-terminal histidine in the alkaline Bohr effect of human hemoglobin.* Biochemistry, 1980. **19**(14): p. 3189-83.

34. Perutz, M.F., et al., *Identification of residues contributing to the Bohr effect of human haemoglobin.* J Mol Biol, 1980. **138**(3): p. 649-68.

35. Ambroggio, X.I. and B. Kuhlman, *Computational design of a single amino acid sequence that can switch between two distinct protein folds.* J Am Chem Soc, 2006. **128**(4): p. 1154-61.

36. Ambroggio, X.I. and B. Kuhlman, *Design of protein conformational switches.* Curr Opin Struct Biol, 2006. **16**(4): p. 525-30.

37. Nordlund, H.R., et al., *Introduction of histidine residues into avidin subunit interfaces allows pH-dependent regulation of quaternary structure and biotin binding.* FEBS Lett, 2003. **555**(3): p. 449-54.

38. Kampmann, T., et al., *The Role of histidine residues in low-pH-mediated viral membrane fusion.* Structure, 2006. **14**(10): p. 1481-7.

39. Rotzschke, O., et al., *A pH-sensitive histidine residue as control element for ligand release from HLA-DR molecules.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16946-50.

40. Nakajima, K., et al., *Identification and modulation of the key amino acid residue responsible for the pH sensitivity of neoculin, a taste-modifying protein.* PLoS One, 2011. **6**(4): p. e19448.

41. Keeler, C., et al., *The kinetics of binding human prolactin, but not growth hormone, to the prolactin receptor vary over a physiologic pH range.* Biochemistry, 2007. **46**(9): p. 2398-410.

42. Kulkarni, M.V., et al., *Two independent histidines, one in human prolactin and one in its receptor, are critical for pH-dependent receptor recognition and activation.* J Biol Chem, 2010. **285**(49): p. 38524-33.

43. Murtaugh, M.L., et al., *A combinatorial histidine scanning library approach to engineer highly pH-dependent protein switches.* Protein Sci, 2011. **20**(9): p. 1619-31.

44. Sagermann, M., et al., *Using affinity chromatography to engineer and characterize pH-dependent protein switches.* Protein Sci, 2009. **18**(1): p. 217-28.

45. Amoura, Z., et al., *Presence of antinucleosome autoantibodies in a restricted set of connective tissue diseases: antinucleosome antibodies of the IgG3 subclass are markers of renal pathogenicity in systemic lupus erythematosus.* Arthritis Rheum, 2000. **43**(1): p. 76-84.

46. Staudt, A., et al., *Potential role of autoantibodies belonging to the immunoglobulin G-3 subclass in cardiac dysfunction among patients with dilated cardiomyopathy.* Circulation, 2002. **106**(19): p. 2448-53.

47. Ackerman, M., et al., *Highly avid magnetic bead capture: an efficient selection method for de novo protein engineering utilizing yeast surface display.* Biotechnol Prog, 2009. **25**(3): p. 774-83.

48. Chao, G., et al., *Isolating and engineering human antibodies using yeast surface display.* Nat Protoc, 2006. **1**(2): p. 755-68.

49. Zaccolo, M., et al., *An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues.* J Mol Biol, 1996. **255**(4): p. 589-603.

50. Zaccolo, M. and E. Gherardi, *The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase.* J Mol Biol, 1999. **285**(2): p. 775-83.

**Figures**

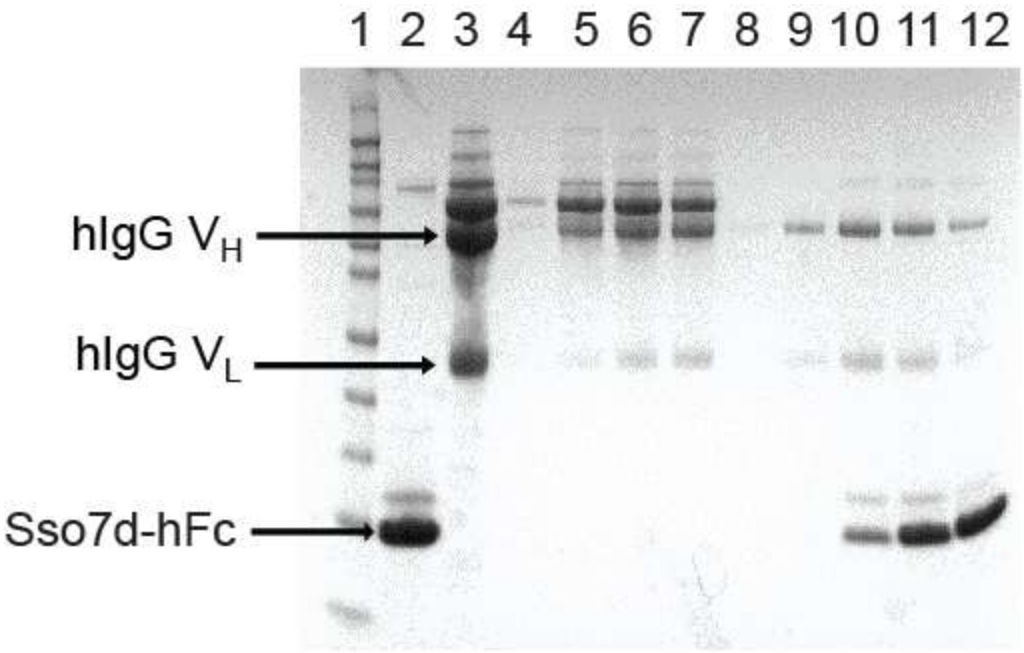



**Figure 5.1**: Sso7d-hFc used as an affinity ligand on a Ni-column. Sso7d-hFc was immobilized on a Ni-column and complex cMEM media containing hIgG was loaded on the column. An imidazole linear gradient was used to elute the hIgG bound Sso7d-hFc. Lane 1 shows Sso7d-hFc loaded on the Ni-column, Lane 2- hIgG in cMEM media before loading, Lane 4-7 describe the flow through when hIgG in cMEM was loaded on the column and Lanes 8-12 are the elution fractions when a linear gradient of imidazole was flown through the column.

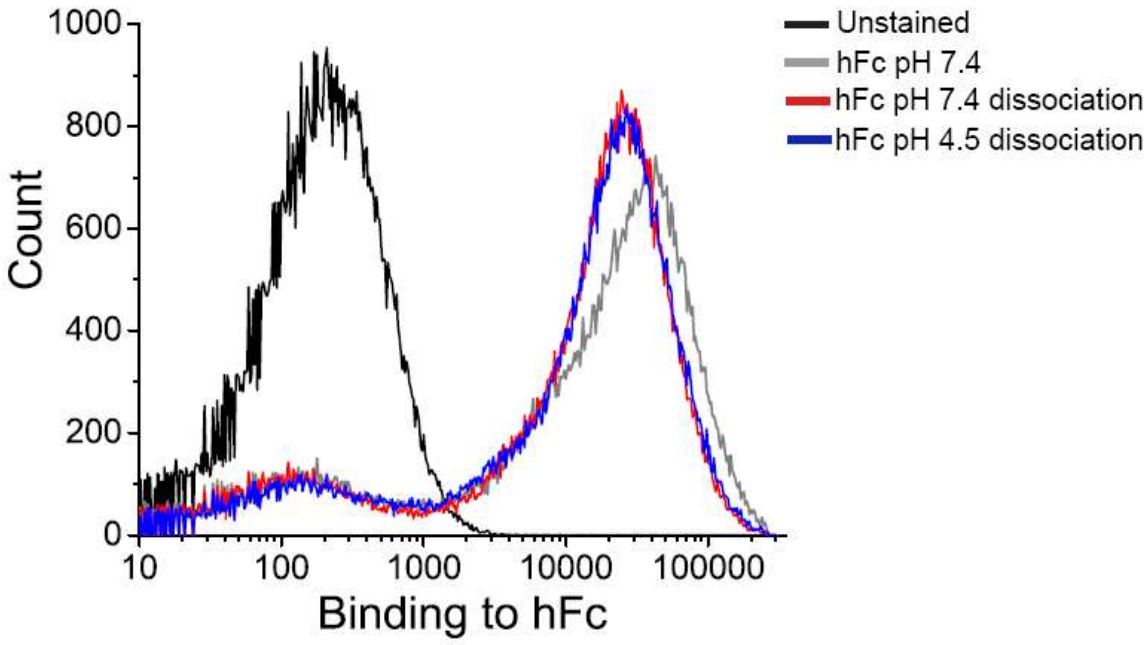



**Figure 5.2:** pH sensitivity of Sso7d-hFc. Yeast surface displayed Sso7d-hFc was incubated with 100 nM of hFc-biotin for 15 min. Three different conditions were studied, gray curve shows fluorescence from a sample labeled with Strep-PE right after incubation, red curve shows sample allowed to dissociate in a buffer of pH 7.4 for 30 min and then labeled with Strep-PE, blue curve shows sample allowed to dissociate in buffer of pH 4.5 for 30 min and then labeled with Strep-PE. Binding of hFc-biotin to Sso7d-hFc is insensitive to pH.

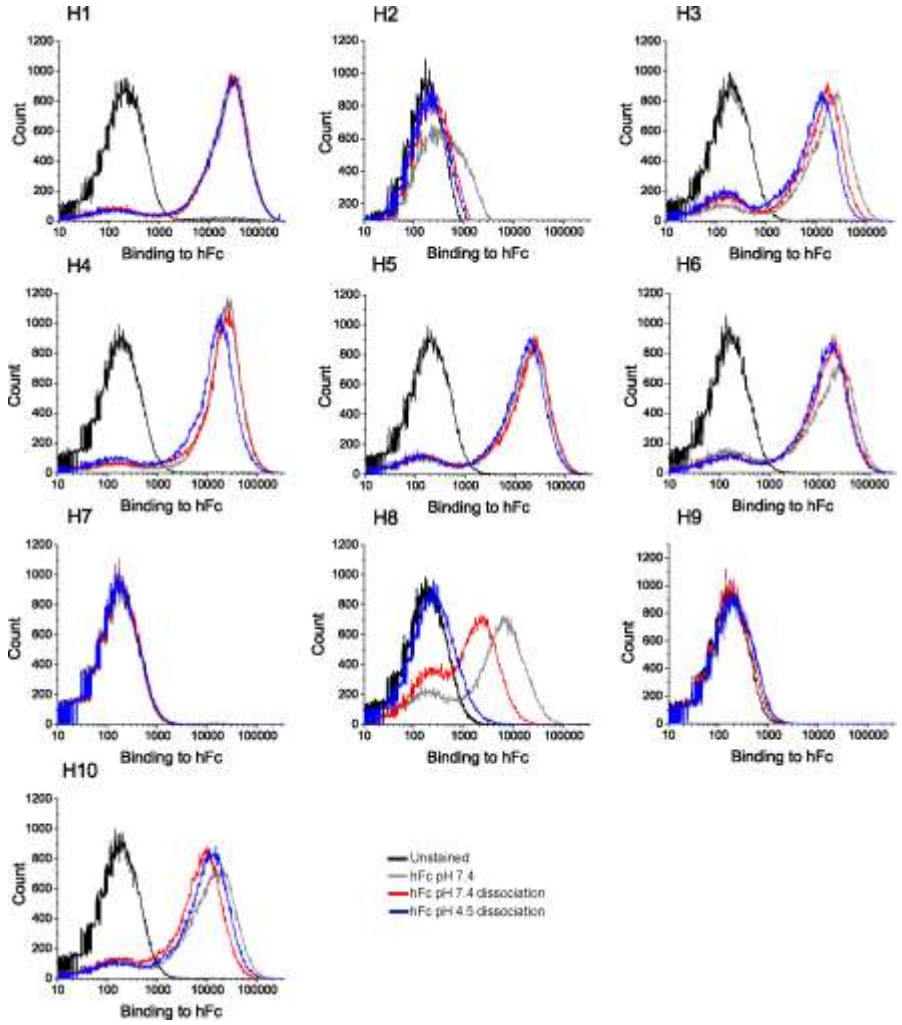



**Figure 5.3:** pH sensitivity of histidine scanning mutants. Yeast surface displayed mutants were incubated with 100 nM of hFc-biotin for 15 min. Three different conditions were studied, gray curve shows fluorescence from a sample labeled with Strep-PE right after incubation, red curve shows sample allowed to dissociate in a buffer of pH 7.4 for 30 min and then labeled with Strep-PE, blue curve shows sample allowed to dissociate in buffer of pH 4.5 for 30 min and then labeled with Strep-PE.

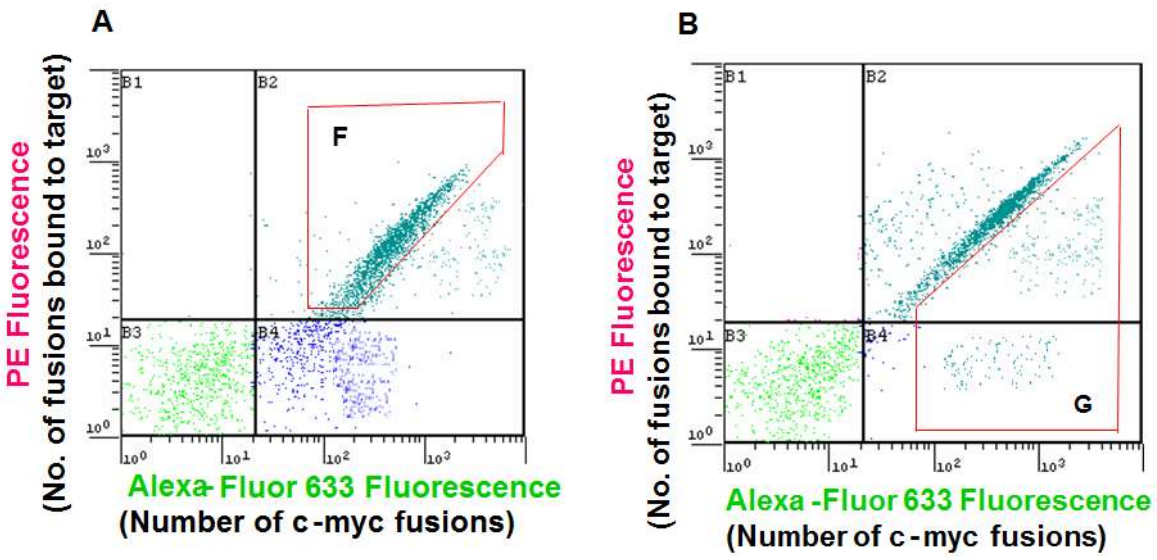



**Figure 5.4**: Representative flow cytometry data to illustrate the flow cytometric sorting for Sso7d-ev-hFc. A) A pure population of cells binding at pH 7.4 is first obtained and all cells in region F are collected. B) Yeast displayed protein-hFc complexes are allowed to dissociate in a buffer of pH 4.5 for 30 min and then run on the cytometer. All cells in region G, which have a higher off rate at pH 4.5 are collected.

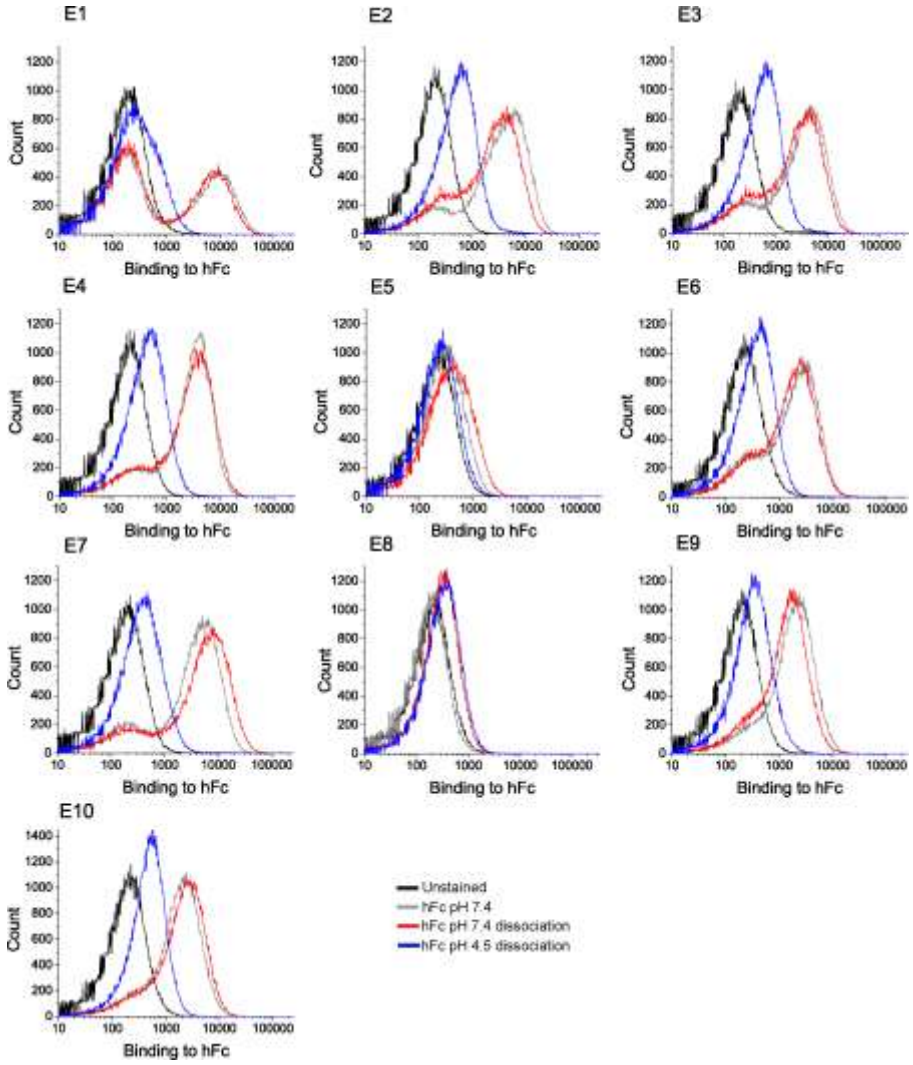
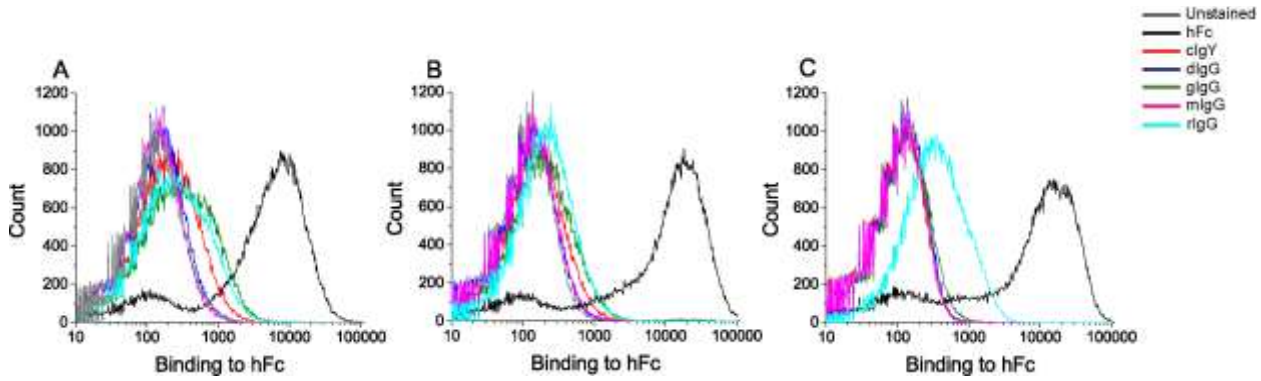



**Figure 5.5:** pH sensitivity of directed evolution mutants. Yeast surface displayed mutants were incubated with 2000 nM of hFc-biotin for 15 min. Three different conditions were studied, gray curve shows fluorescence from a sample labeled with Strep-PE right after incubation, red curve shows sample allowed to dissociate in a buffer of pH 7.4 for 30 min and then labeled with Strep-PE, blue curve shows sample allowed to dissociate in buffer of pH 4.5 for 30 min and then labeled with Strep-PE.

**a)**




**b)**

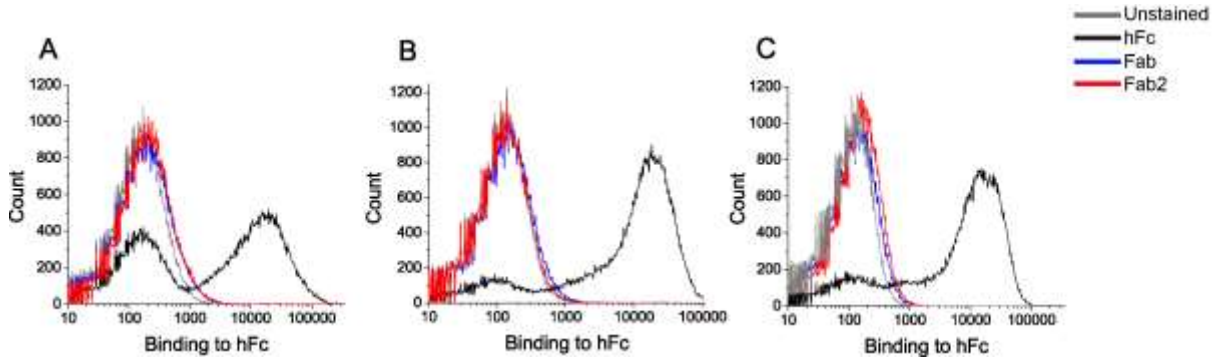



**Figure 5.6**: Specificity analysis of hFc-binding Sso7d mutants. Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc were labeled with 1 µM of closely related immunoglobulin species, a) chicken-IgY (cIgY), donkey-IgG(dIgG), goat-IgG(gIgG), mouse-IgG(mIgG) and rabbit-IgG(rIgG) and b) Fab and Fab2 fragments. Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc show specific binding to hFc, Sso7d-ev-hFc also shows some binding to rIgG.

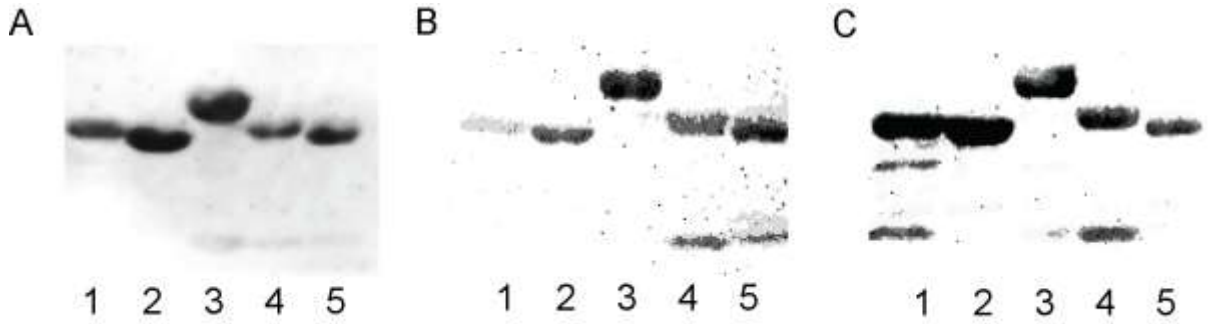



**Figure 5.7**: Western blotting analysis of Sso7d-hFc, Sso7d-his-hFc and Sso7d-ev-hFc. The mutant binding proteins were used as primary reagents in western blotting and secondary incubation was carried out with a HRP conjugated anti-biotin antibody. The following samples were loaded on the gel for western blot: lane 1- human IgG isotype1, lane 2- human IgG isotype 2, lane 3-human IgG isotype 3, lane 4- human IgG isotype4, lane 5- human IgG composite.

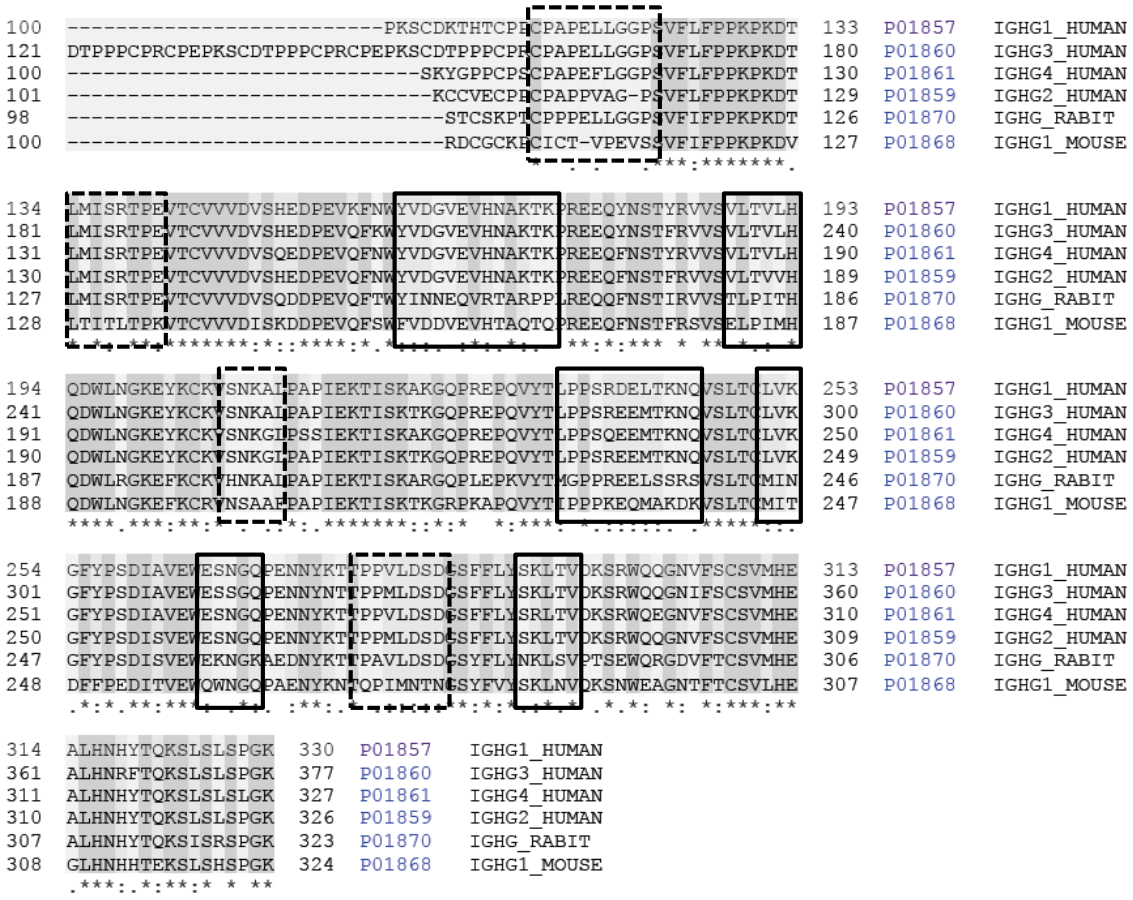



**Figure 5.8:** Sequence alignments for epitope mapping. The sequences of hIgG subclasses 1, 2, 3, 4 rIgG and mIgG are aligned in Uniprot. Sso7d-hFc and Sso7d-his-hFc are specific to all four hIgG subclasses and not to rIgG and mIgG. The solid boxes represent putative peptide stretches which might be epitopes for binding. Dashed boxes represent peptide stretches as potential epitopes for Sso7d-ev-hFc which is specific to the hIgG subclasses but also shows some binding to rIgG

## Tables

**Table 5.1:** Pool of mutants binding to hFc with the highest affinity. Sso7d-hFc had the highest binding affinity among the pool. Sso7d-his-hFc is a histidine scanning mutant (Y40H) from the parent clone Sso7d-hFc

| | 20.....30.....40... |
|---|---|
| wt Sso7d | KKVWRVGKMISFTYDLGGGKTGRGA |
| hFc-1(4) | SIVPRSGKYIHFIYDLGGGKTGRGN |
| hFc-2(4) | CVVRRFGKVISFDYDLGGGKSGRGC |
| Sso7d-hFc | YLVSRIGKRILFMYDLGGGKYGIGR |
| Sso7d-his-hFc | YLVSRIGKRILFMYDLGGGKHGIGR |

**Table 5.2:** Pool of mutants with pH sensitive binding from the random yeast display library. Sso7d-ev-hFc was selected as the best mutant from this pool.

| | .....20.....30.....40.....50.... |
|---|---|
| Sso7d-hFc | DISKI**YL**V**S**R**I**GK**R**I**L**F**M**YDLGGGK**Y**G**I**G**R**VSEKDAPKELL |
| hFcev1 | DIS**E**I**YR**V**S**R**R**GK**S**I**A**F**M**YDLGGGK**Y**G**I**G**Y**VSEKDAPKELL |
| hFcev2(2) | DISKI**FF**V**R**R**ID**K**L**I**A**F**S**YDLGGGK**Y**G**L**G**F**VSEKDAPKELL |
| hFcev4 | DISKI**RL**V**S**R**N**G**RR**I**S**F**M**YDLGGGK**H**G**I**G**Q**VSE**R**DAPKELL |
| hFcev6 | DISKI**RL**V**S**R**Q**GK**I**I**K**F**T**YDLGGG**EL**G**M**G**R**VSEKDAPKELL |
| Sso7d-ev-hFc | DISKI**YR**V**F**R**S**GK**T**I**F**F**R**YDLGGG**EL**G**V**G**I**VSEKDAPKELL |
| hFcev9 | DISKI**RL**V**A**R**T**GK**I**I**R**F**Q**YDLGGGK**Y**G**L**G**R**VSEKDAPKELL |
| hFcev10 | DISKI**YR**V**S**R**R**G**ES**I**A**F**M**YDLGGGK**Y**G**I**G**Y**VSE**E**DAP**E**ELL |

# CHAPTER 6

# CONCLUSIONS AND FUTURE WORK

Nimish Gera

Department of Chemical and Biomolecular Engineering

North Carolina State University

Raleigh, NC 27695-7905

## 6.1. Conclusions

Non-immunoglobulin scaffolds is a rapidly growing area of research and hyperthermophilic protein scaffolds are an addition to the protein engineering toolbox for engineering bio-molecular recognition. The yeast surface display system as described in Chapter 2 can be routinely used to generate large libraries ($10^8$-$10^9$ members) of these scaffolds and engineer binding proteins for a variety of targets. As shown in Chapter 3, we have used the Sso7d scaffold to generate a library of $10^8$ members and isolate binding proteins to six different targets of varying topology. These proteins were found to be highly stable to thermal($T_m$> 70˚C), chemical and pH (pH 0.33~12.5) denaturation. These properties make hyperthermophilic scaffolds as excellent candidates for applications such as affinity chromatography.

In Chapter 5, we have validated that human IgG (hIgG) specific affinity ligands can be isolated from the same Sso7d library. The generated ligands can be engineered to have pH sensitivity using rational design and/or directed evolution. These ligands recognize all four hIgG isotypes (hIgG1, 2, 3 and 4), are specific to hIgG and do not cross-react with other species such as mouse, rabbit, goat or chicken-IgY. Additionally, the binding is not carbohydrate mediated.

We have also used other hyperthermophilic proteins as scaffolds such as Sso6901 and MIT domain from *Sulfolobus solfataricus*, Tm1112 and Tm0487 from *Thermatoga maritima*, Ph1500 from *Pyrococcus horikoshii* and the chitin binding domain from *Pyrococcus*

*furiosus*. Binding proteins from these scaffolds were also demonstrated to have high thermal stability. In Chapter 4, we also validated that scaffold diversity is at least as good as or better than using sequence diversity.

## 6.2. Future Work

Non-immunoglobulin scaffolds have been used in myriad applications such as imaging, diagnostics, therapeutics, affinity chromatography and crystallization chaperones to name a few. On the contrary, the potential of hyperthermophilic scaffolds is largely un-utilized. Here, we describe some potential applications where the use of hyperthermophilic scaffolds would be useful.

Firstly, after having generated binding proteins from the Sso7d scaffold and characterizing the biophysical properties of the Sso7d-derived proteins, a thorough biochemistry analysis of the scaffold would be useful. To this end, other regions in the Sso7d scaffold can be mutated and binding proteins can be generated. Having a thorough biochemical analysis of the mutants along with engineering molecular recognition would provide a deeper understanding of the stability of the scaffold. Targeted library generation can then be accomplished by focusing on residues that are evolvable and do not lead to instability of the Sso7d molecule.

Using a different region on the Sso7d scaffold such as the double stranded β-sheet and ATP binding pocket might provide higher affinity binding proteins which were previously unachievable by the triple-stranded β-sheet library.

Further, affinity maturation of the Sso7d scaffold has not been tested yet. Therefore, the ability of hyperthermophilic scaffolds to evolve into extremely high affinity binders ($K_D$~pM) still needs to be examined.

The Sso7d-derived ligands generated in Chapter 5 can be characterized further. The binding capacity and recovery achieved with these ligands is an important parameter. Stability of these ligands to harsh regeneration procedures and several cycles of clean in place sterilization also need to be investigated. Elucidating the crystal structure of these ligands with hIgG would be useful. This would provide information about interactions involved in mediating the binding and allow further rational engineering to improve the pH sensitive interactions.

The Sso7d scaffold has excellent thermal, chemical and pH stability as demonstrated in Chapter 3. High thermal and chemical stability ensures that hyperthermophilic protein scaffolds can serve as excellent candidates for molecular imaging in cancer related applications. Affinity reagents can be generated for the Ras oncogenic mutants Q61L and G12V. The Ras protein family is also found in human embryonic stem cells and binding proteins for the Ras proteins would also be useful.

A growing area of research is the integration of display systems for protein engineering. Obtaining crystal structures for membrane proteins is an ardent task because of the difficulty in producing these proteins in soluble form. An interesting possibility would be to explore

screening yeast surface displayed membrane protein as a target with a phage/mRNA displayed library of hyperthermophilic scaffold such as Sso7d. This would not only validate the compatibility of all three display systems with Sso7d but also allow us to isolate binding proteins for targets which are difficult to produce in soluble form otherwise. Once the binding partner has been obtained, it could be used for efficient purification of the target protein from the membrane fraction as validated in Chapter 5 using a his-tagged Sso7d variant and the complex can be used for crystallography. Here, the Sso7d protein provides a means to purify the membrane protein as well as stabilize it for complex formation. This strategy would also be useful for other proteins which might be difficult to crystallize or those that form a tertiary structure only in the presence of a binding partner.

With the above studies, the potential of hyperthermophilic variants in multiple applications can be explored. Success of a scaffold depends on its ability to generate specificity for a spectrum of targets, its practical usefulness such as robustness, smaller size and ease of expression; and benefits over most conventional antibody fragments and existing scaffolds. Most of these properties have been validated in this study on hyperthermophilic scaffolds. Validation of hyperthermophilic protein scaffolds will also involve affinity generation/ maturation *in vitro* followed by commercial therapeutic or diagnostic applications *in vivo* which needs to be explored further.