ABSTRACT

HUSSAIN, MD MAHMUD. Stable Affinity Ligands from Hyperthermophilic Scaffolds: Applications to Macromolecular Purification. (Under the direction of Dr. Balaji M. Rao).

We have designed highly stable affinity ligands derived from hyperthermophilic archea and bacteria for engineering molecular recognition using yeast surface display and mRNA display in conjunction with Fluorescence Activated Cell Sorting (FACS). Our model targets include a small molecule fluorescein, C-terminus peptide of β catenin, a cyclic peptide Brain Natriuretic Peptide (BNP32), hen egg lysozyme (HEL), rabbit IgG and a plant virus called *Red Clover Necrotic Mosaic Virus* (RCNMV). We demonstrate the use of these binding proteins in the context of macromolecular and virus purification. We also show that the binders obtained from what we call 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 a single, previously validated scaffold (Sso7d). Our results underscore the advantages of using an ensemble of hyperthermophilic scaffolds for library construction - the use of multiple scaffolds increases the topological diversity in the library. Taken together, 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*. In addition, the Sso7d-based RCNMV-binding protein (RBP) was used to purify RCNMV from plant extract, using a variation of affinity chromatography called "avidity chromatography". In this scheme, RCNMV from plant extract is captured on a nickel column that is pre-loaded

with hexahistidine tagged RBP. The highly avid interaction between RCNMV and immobilized RBP ensures efficient capture of RCNMV despite modest binding affinity (K_D ~ 100 nM) of the RBP-RCNMV interaction. This purification scheme eliminates the need for harsh elution conditions that are typically required for affinity chromatography of viruses and also eliminates the need for chemical conjugation of the affinity ligand to a resin. Further, the 7 kDa RBP has high thermal stability (T_m ~ 83 °C) and can be recombinantly expressed at high yield in *E. coli.*

Stable Affinity Ligands from Hyperthermophilic Scaffolds: Applications to Macromolecular Purification

by
Md Mahmud Hussain

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 Chair of Advisory Committee	Robert M. Kelly
_____	_____
Jason M. Haugh	Chris Ashwell

DEDICATION

Dedicated to my loving parents Shahana and Manir Hussain

BIOGRAPHY

I was born in Dhaka, the capital city of Bangladesh. I finished high school from the port city of Chittagong. I have earned my bachelors degree in Chemical Engineering in 2004 from Bangladesh University of Engineering and Technology (BUET), Dhaka and Masters degree in Chemical and Biomolecular Engineering in 2008 with a minor in Biotechnology from North Carolina State University. I have been pursuing graduate studies under the supervision of Dr. Balaji M. Rao since 2006 at NC State.

ACKNOWLEDGMENTS

I would like to express my heartiest gratitude to my PhD supervisor Dr. Balaji M. Rao for his constant support, encouragement and above all, patience while working with me. His expertise and valuable suggestions, scientific or otherwise, have been guiding me properly in this long stroll of PhD. I owe special thanks to Nimish Gera for help with numerous experiments. I would also like to thank other Rao group members: Prasenjit Sarkar, Kevin Carlin and Karthik Tiruthani for helpful discussions. I also thank Steve Cotten from UNC Chapel Hill, Sam Jenkins, Dustin Lockney and Ruqi Wang of Chemistry department and Nathaniel Hentz from BTEC for help with many experiments and expert comments. Finally, I would like to thank my beloved wife Sarah for bearing with me this whole time, supporting and encouraging me relentlessly.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Engineering Molecular Recognition

Molecular Recognition

Molecular recognition refers to highly specific and non-covalent interaction of one biomolecule with another (Protein-protein, protein-nucleic acid). Essentially, proteins (large multi-domain antibodies of 150 kDa or as small as less than 10 kDa) serve as the molecular recognition reagent i.e. the affinity reagent. Examples include, antibodies recognizing (binding) peptides derived from pathogens in the humoral immune system and numerous cell-surface receptors binding their cognate ligands. Molecular recognition can be mediated by all possible combinations of van der Waals force, hydrogen bond, hydrophobic and electrostatic interactions. Such biomolecular interactions are key steps to cellular processes, biomolecular detection, quantitative measurements and purification of targets of interest. Depending on the application, targets of interest can be as wide-ranging as small molecules through peptides, proteins and even virus. Traditionally, antibodies are used as the affinity reagents in purification or immunoassays like quantitative Western blot. Nevertheless, there are certain attributes of antibodies like their large multi-domain structure, presence of di-sulfide bonds, tedious and expensive generation process limit their use. Other than historic reason, lack of stable and robust alternatives to antibodies is the leading cause for widespread use of antibodies as affinity reagents.

Antibodies and antibody-based molecules as natural binders

Antibodies or Immunoglobulins are large (~150 kDa), multi-domain proteins (Fig 1.1). Evolution has equipped mammalian immune system with a highly diverse repertoire of immunoglobulins [1]. Immunoglobulins recognize their binding partners through the

mutations in the six hypervariable loop regions, called the Complementarity Determining Regions (CDRs). Undoubtedly, antibodies have been most widely used binding molecules in fundamental research, biotechnology and medicine [2]. Up until 20 years ago, antibodies generated in the immune system was the sole source of obtaining affinity reagents for a given target species [2]. The oldest approach to generating affinity reagents involves immunizing vertebrate animals such as mice or goats with the target (antigen) of interest. The immune system of the animal then generates antibodies that bind specifically to the target. The antibodies generated in this fashion are typically heterogeneous in nature and bind to different regions (epitopes) on the antigen, and are referred to as polyclonal antibodies (PAbs). While the use of PAbs as affinity reagents is very common, variability of the immune response in animals and the heterogeneity of PAbs can cause significant lot-to-lot variability in PAb reagents. Also, generation of PAbs is an expensive process. The problem of variability in PAbs can be overcome by the use of hybridoma technology to create mammalian cell lines that produce monoclonal antibodies (MAbs), i.e. defined antibodies binding to a single epitope on the antigen. However, generation of cell lines is tedious, time consuming and expensive. The recombinant production of antibodies in bacterial systems has been reported. However, this has proved to be difficult in general due to the large multi subunit disulfide-bonded structure of antibodies, which also leads to low stability. Thus both MAbs and PAbs suffer from low stability and high cost of production. On the other hand, advances in the field of protein engineering have enabled scientists to create molecular diversity *in vitro* from antibody, antibody-based molecules or protein of non-immunoglobulin origin [2]. For instance, single chain variable fragments (scFv) have been

successfully engineered against various targets with subnanomolar or even picomolar affinity and high specificity [3-6]. However, conflict with intellectual property over antibodies also present a challenge to engineering antibodies in many applications [1]. Besides, for non-therapeutic applications, immunoglobulin framework is not necessary [2]. In order to address these issues associated with immunoglobulins, non-immunoglobulin or the so called 'alternate scaffolds' have emerged.

The alternate scaffold paradigm

In order to circumvent the limitations of antibodies, over 50 different alternate scaffolds have been reported so far [1, 2]. Analogous to the mutations in the CDR region of the antibodies, mutagenizing certain residues on non-immunoglobulin small proteins (~10 kDa) can present enormous diversity in the topological framework of the given protein. An alternate scaffold can be thought of a 'template' protein that can tolerate mutations to certain amino acid residues and give rise to a library of mutant proteins. Desirables properties of a scaffold proteins include small size (~ 10 kDa), lack of di-sulfide bond(s), high thermodynamic stability, facile expression in *E. coli* in large titers and above all, their ability to give as high specificity and affinity towards the 'target' of interest as an immunoglobulin binder. In recent times, alternate protein scaffolds have been successfully engineered for molecular recognition in therapeutics, *in vivo* diagnostics, intracellular targeting, co-crystallization, purification of MAbs, inhibition of proteases, antagonists to receptors as well as antidotes for neutralizing intoxicating agents [1, 2]. 10th domain of the human fibronectin type III [7], DARPins [8, 9], Affibodies [10, 11], Avimer [12], AdNectins [13], Lipocalins [14], Affilins

[15], Knottins [16] are examples of some of the non-immunoglobulin scaffolds that have been engineered for molecular recognition for a wide range of targets and hence applications. The most recent addition to these alternate scaffolds is the hyperthermophilic protein Sso7d from *Sulfolobus solfataricus*, which has been engineered for molecular recognition against a wide range of targets [17]. Because of its hyperthermophilic origin, binding molecules obtained from mutagenizing ten solvent-exposed residues on Sso7d are highly stable in terms of melting temperature (as high as 103^{o} C), resistance to Guanidine hydrochloride denaturation and maintaining secondary structure under extended incubation in extreme pH condition [17]. Tolerance to extensive mutations (~16% residues) of Sso7d is not surprising owing to its hyperthermophilic origin. Indeed, it has been shown previously by the Arnold group that highly stable proteins are more tolerant to a wide range of mutations [18]. Yeast surface display in conjunction with Fluorescence Activated Cell Sorting (FACS) was used to obtain binding molecules from a combinatorial Sso7d library. A few other state-of-the art technologies for protein engineering are discussed below.

Display technologies for engineering molecular recognition

Mainly, there are two approaches to protein engineering: one is rational design and the other is directed evolution. In the rational design approach, specific amino acid residue(s) of a protein is replaced by another residue and thus a mutant protein is created. The product of such an approach is a *point mutant* because only one amino acid is substituted in the sequence of the wild type protein. Point mutation can be introduced to significantly alter the biophysical properties of a protein [19]. Contrary to rational design, directed evolution

employs random mutagenesis and/or gene recombination to create mutant libraries [20]. Simplistically put, directed protein evolution can be thought of evolution in test tubes. The idea is to isolate the best pool of mutants that have the desired property which the other billions of mutants don't have (Figure 1.2). Just the way natural selection favors certain phenotypes and rejects others, directed protein evolution also applies a selection pressure on the library of mutant proteins. For engineering molecular recognition, the selection pressure can be successive lower concentration of the target, slower dissociation rate and/or stringent wash condition of the protein-target complex. In random mutagenesis, any amino acid residue can be replaced by 19 other amino acids either at given positions or completely randomly at any position. Gene recombination by DNA shuffling is also an elegant way of creating molecular diversity [21]. A combinatorial library can also be created by mutagenizing only selected residues on a protein template instead of mutagenizing randomly. Point mutation is achieved by site-directed mutagenesis and with the current techniques in molecular biology, the protocols for performing site-directed mutagenesis is relatively straightforward. Screening for the best mutant from a combinatorial library, however, is challenging. The challenge is to isolate the mutants with improved properties at each stage of selection. This challenge is overcome by 'Display Technology'. Display technology links the protein (phenotype) to its nucleic acid moiety (genotype), either using a cellular platform or in a completely cell-free environment. The first of its kind was phage display invented in the mid 80's by George P. Smith [22]. This method uses bacteriophage to express the foreign gene library as fusion protein on phage surface. Bratkovic discussed the progress in phage display techniques and its applications in his recent review [23]. Phage display has been

extensively used in antibody engineering. Yeast Surface Display (YSD), developed by KD Wittrup came into picture in the late 90's [24]. YSD has been particularly useful because it uses eukaryotic protein processing machineries and thus enables expression of mammalian proteins with post-translational modifications better than phage display [24]. In YSD, the gene library is expressed on the surface of the yeast as a fusion protein to Aga2p mating protein of the yeast strain *Saccharomyces cerevisiae* (Figure 1.3). Another display technique, called messenger RNA display is different from the former two in that mRNA display is completely a cell-free, *in vitro* platform. Ribosome display is also a cell-free display technique. Lipovsek et al presented a review on the latter two display technologies [25]. Binz et al [2] enumerated a list of protein scaffolds that were engineered for molecular recognition alongside the technology that was used for their selection.

References

[1] A. Skerra, "Alternative non-antibody scaffolds for molecular recognition," *Curr Opin Biotechnol*, vol. 18, pp. 295-304, Aug 2007.

[2] H. K. Binz, P. Amstutz, and A. Pluckthun, "Engineering novel binding proteins from nonimmunoglobulin domains," *Nat Biotechnol*, vol. 23, pp. 1257-68, Oct 2005.

[3] E. T. Boder, 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*, vol. 97, pp. 10701-5, Sep 26 2000.

[4] M. J. Feldhaus, R. W. Siegel, L. K. Opresko, J. R. Coleman, J. M. Feldhaus, Y. A. Yeung, J. R. Cochran, P. Heinzelman, D. Colby, J. Swers, C. Graff, H. S. Wiley, and K. D. Wittrup, "Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library," *Nat Biotechnol*, vol. 21, pp. 163-70, Feb 2003.

[5] C. P. Graff, K. Chester, R. Begent, and K. D. Wittrup, "Directed evolution of an anti-carcinoembryonic antigen scFv with a 4-day monovalent dissociation half-time at 37 degrees C," *Protein Eng Des Sel*, vol. 17, pp. 293-304, Apr 2004.

[6] M. C. Kieke, B. K. Cho, E. T. Boder, D. M. Kranz, and K. D. Wittrup, "Isolation of anti-T cell receptor scFv mutants by yeast surface display," *Protein Eng*, vol. 10, pp. 1303-10, Nov 1997.

[7] H. I. Liao, C. A. Olson, S. Hwang, H. Deng, E. Wong, R. S. Baric, R. W. Roberts, and R. Sun, "mRNA display design of fibronectin-based intrabodies that detect and inhibit severe acute respiratory syndrome coronavirus nucleocapsid protein," *J Biol Chem*, vol. 284, pp. 17512-20, Jun 26 2009.

[8] M. T. Stumpp, H. K. Binz, and P. Amstutz, "DARPins: a new generation of protein therapeutics," *Drug Discov Today*, vol. 13, pp. 695-701, Aug 2008.

[9] D. Steiner, P. Forrer, and A. Pluckthun, "Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display," *J Mol Biol,* vol. 382, pp. 1211-27, Oct 24 2008.

[10] M. Lindborg, E. Cortez, I. Hoiden-Guthenberg, E. Gunneriusson, E. von Hage, F. Syud, M. Morrison, L. Abrahmsen, N. Herne, K. Pietras, and F. Y. Frejd, "Engineered high-affinity affibody molecules targeting platelet-derived growth factor receptor beta in vivo," *J Mol Biol,* vol. 407, pp. 298-315, Mar 25 2011.

[11] J. Lofblom, J. Feldwisch, V. Tolmachev, J. Carlsson, S. Stahl, and F. Y. Frejd, "Affibody molecules: engineered proteins for therapeutic, diagnostic and biotechnological applications," *FEBS Lett,* vol. 584, pp. 2670-80, Jun 18 2010.

[12] J. Silverman, Q. Liu, A. Bakker, W. To, A. Duguay, B. M. Alba, R. Smith, A. Rivas, P. Li, H. Le, E. Whitehorn, K. W. Moore, C. Swimmer, V. Perlroth, M. Vogt, J. Kolkman, and W. P. Stemmer, "Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains," *Nat Biotechnol,* vol. 23, pp. 1556-61, Dec 2005.

[13] D. Lipovsek, "Adnectins: engineered target-binding protein therapeutics," *Protein Eng Des Sel,* vol. 24, pp. 3-9, Jan 2010.

[14] S. Schlehuber and A. Skerra, "Lipocalins in drug discovery: from natural ligand-binding proteins to "anticalins"," *Drug Discov Today,* vol. 10, pp. 23-33, Jan 1 2005.

[15] T. Hey, E. Fiedler, R. Rudolph, and M. Fiedler, "Artificial, non-antibody binding proteins for pharmaceutical and industrial applications," *Trends Biotechnol,* vol. 23, pp. 514-22, Oct 2005.

[16] S. Krause, H. U. Schmoldt, A. Wentzel, M. Ballmaier, K. Friedrich, and H. Kolmar, "Grafting of thrombopoietin-mimetic peptides into cystine knot miniproteins yields high-affinity thrombopoietin antagonists and agonists," *FEBS J,* vol. 274, pp. 86-95, Jan 2007.

[17] N. Gera, M. Hussain, R. C. Wright, and B. M. Rao, "Highly stable binding proteins derived from the hyperthermophilic sso7d scaffold," *J Mol Biol,* vol. 409, pp. 601-16, Jun 17 2011.

[18] J. D. Bloom, S. T. Labthavikul, C. R. Otey, and F. H. Arnold, "Protein stability promotes evolvability," *Proc Natl Acad Sci U S A,* vol. 103, pp. 5869-74, Apr 11 2006.

[19] B. Nilsson, T. Moks, B. Jansson, L. Abrahmsen, A. Elmblad, E. Holmgren, C. Henrichson, T. A. Jones, and M. Uhlen, "A synthetic IgG-binding domain based on staphylococcal protein A," *Protein Eng,* vol. 1, pp. 107-13, Feb-Mar 1987.

[20] F. H. Arnold, "Combinatorial and computational challenges for biocatalyst design," *Nature,* vol. 409, pp. 253-7, Jan 11 2001.

[21] W. P. Stemmer, "DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution," *Proc Natl Acad Sci U S A,* vol. 91, pp. 10747-51, Oct 25 1994.

[22] G. P. Smith, "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface," *Science,* vol. 228, pp. 1315-7, Jun 14 1985.

[23] T. Bratkovic, "Progress in phage display: evolution of the technique and its application," *Cell Mol Life Sci,* vol. 67, pp. 749-67, Mar 2010.

[24] E. T. Boder and K. D. Wittrup, "Yeast surface display for screening combinatorial polypeptide libraries," *Nat Biotechnol,* vol. 15, pp. 553-7, Jun 1997.

[25] D. Lipovsek and A. Pluckthun, "In-vitro protein evolution by ribosome display and mRNA display," *J Immunol Methods,* vol. 290, pp. 51-67, Jul 2004.

Figures

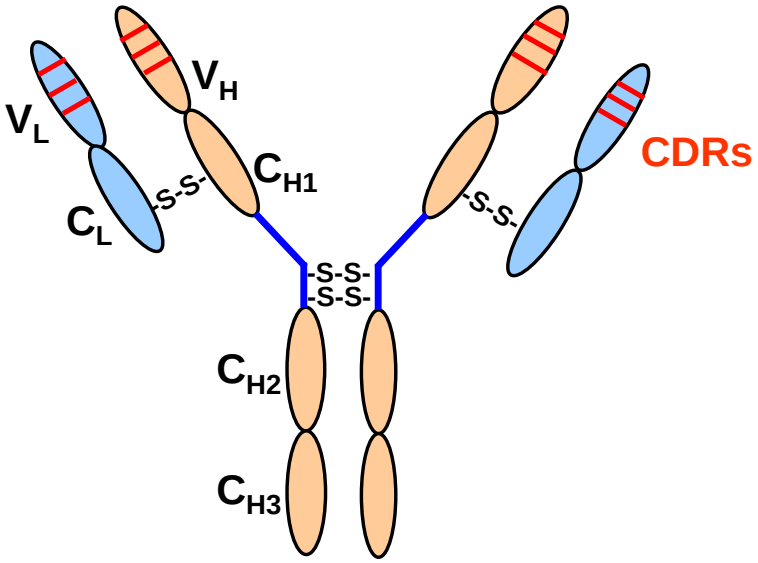


Figure 1.1: Schematics of an antibody. V and C refer to variable and constant regions, whereas subscripts L and H refer to light and heavy chains, respectively. Disulfide bonds are shown as –S-S-, CDR denotes Complementarity Determining Region.

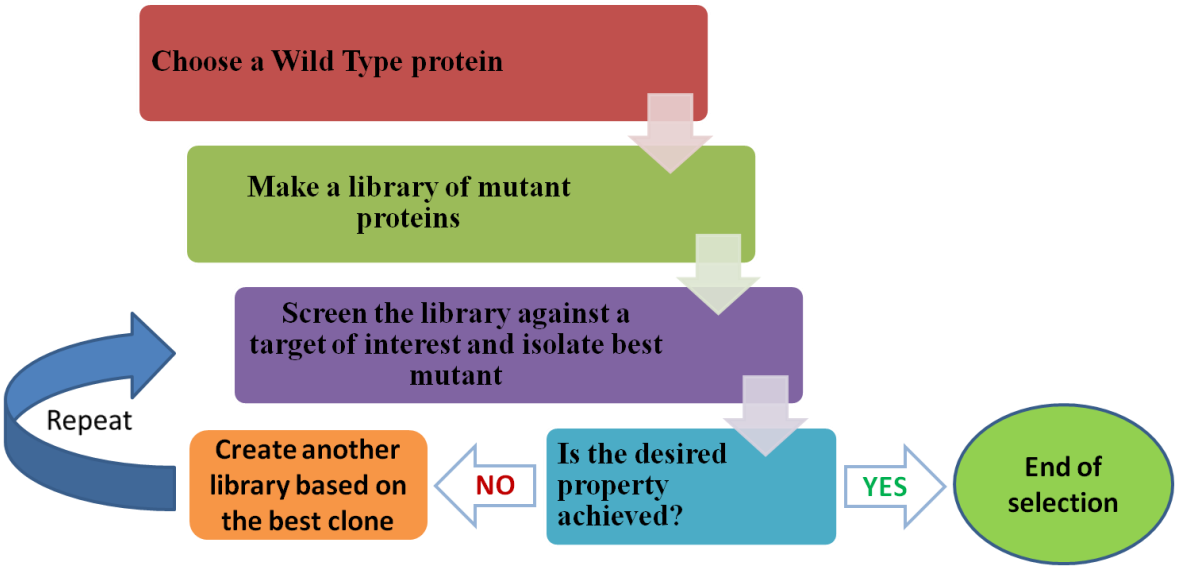


Figure 1.2: Outline of a directed protein evolution strategy

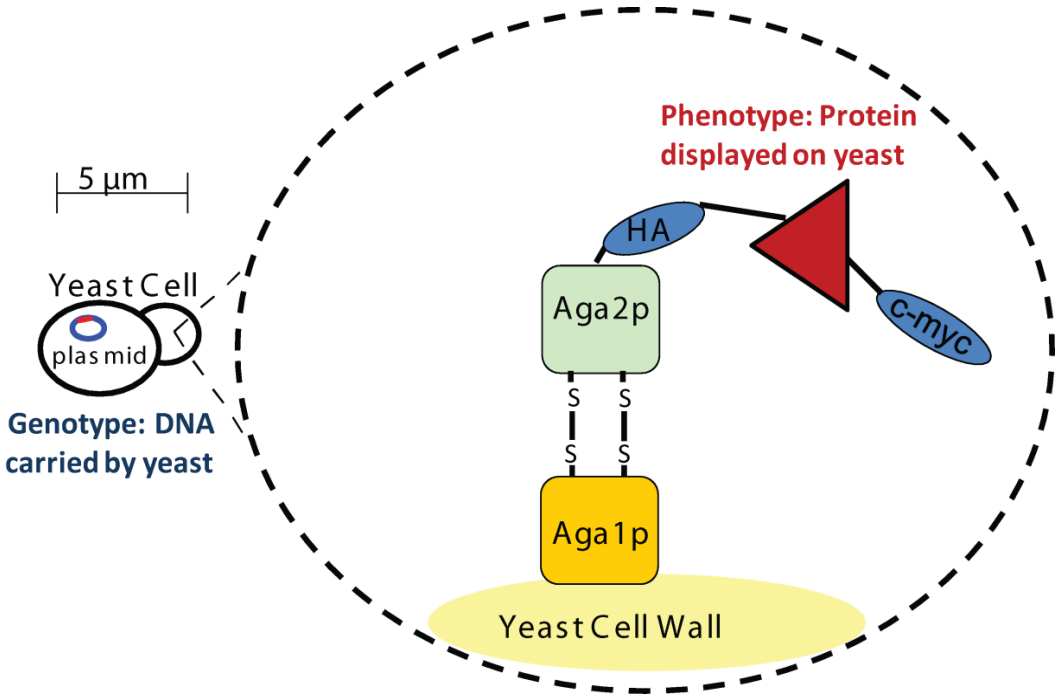


Figure 1.3: Yeast surface display as a platform to link genotype to phenotype. The mutant protein is expressed on yeast cell surface as a fusion to Aga2p subunit of yeast *Saccahromyces cerevisiae*. Aga2p is in turn linked to the Aga1p subunit of yeast mating protein a-agglutinin. Aga1p anchors the entire assembly on yeast such that the fusion protein can interact with other molecules in solution [24].

Chapter 2

A Super-library of Hyperthermophilic Protein Scaffolds for Engineering Molecular Recognition

Mahmud Hussain[1], Nimish Gera[1], Andrew B. Hill and Balaji M. Rao

Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 2769-7905

[1]These authors contributed equally

Abstract

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 a single, previously validated scaffold. Our results underscore the advantages of using an ensemble of hyperthermophilic scaffolds for library construction - the use of multiple scaffolds increases the topological diversity in the library, and due to their higher tolerance to mutation, hyperthermophilic proteins are likely to yield libraries containing fewer misfolded mutant proteins. Taken together, 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*. Therefore, we expect that proteins based on these scaffolds will be useful for applications in research, biotechnology and medicine.

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 ~ 5×10^{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.

Results

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 2.1** and **Table 2.1**. 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 $\alpha 2$ and $\alpha 3$; 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 2.1**. 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 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 2.2**. 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.

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 2.3**. 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 2.3**, 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].

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 2.4**. Briefly, we generated an Sso7d library with an estimated diversity of ~ 5 x 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 2.2**. 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 2.3**. 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.

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**

2.5, 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 2.6**, 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 2.4 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 2.7**. 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).

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 2.2**). 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 2.1**). 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 2.2**). 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 ~ 5×10^{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 2.3**). 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 2.4**). 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.

Materials and Methods

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 2.5**. 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 2.1**. Individual libraries were pooled together to obtain the super-library, with overall estimated diversity of 4.4×10^8 mutants.

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™ 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.

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.

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].

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^{TM} HF DNA polymerase (New England Biolabs, Ipswich, MA; 1U/50µl) in 1X HF Phusion^{TM} 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 µl T7 RNA polymerase (20 U/µl) 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 μl reaction for 4 hours at 37 ^{o}C with RNase-free Turbo DNase (1 U/μl, 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-dA$_{15}$-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 μl 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 μl 75% ethanol, centrifuged for 15 min and washed with 100% ethanol. The mRNA pellet was air-dried and resuspended in 50 μl 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 μg crosslinked mRNA, 50 μM methionine, 1X buffer without leucine, 340 μl retic lysate, to a final 500 μl 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 µl 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 µl PBS-BSA-DNA. Subsequently, beads were incubated with 100 µl 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 µl SuperSignal^{TM} 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 µM 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 µl 5X first strand buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM $MgCl_2$), 50 µl 10 mM dNTPs and 100 µl 0.1 M DTT were added to the reverse transcription reaction, the total volume was adjusted to 995 µl using 0.1% DEPC water, and the mixture was incubated at 42 ^{o}C

for 2 min. Finally, 5 μl Superscript™ 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 ~ $5x10^{11}$.

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^{o}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^{o}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^{o}C, the supernatant was separated from the beads and subsequently incubated with beads coated with BNP-32 for 1 hr at 4^{o}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 µM 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.

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.

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.

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.

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].

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 θ_{min} and θ_{max} are the minimum and maximum values of baseline-corrected molar ellipticity.

Acknowledgments

We thank Prof. Rihe Liu and Steve Cotten (University of North Carolina, Chapel Hill) for help with mRNA display protocols. We also thank Paige Luck from the Department of Food Science, North Carolina State University (NCSU) and Dr. John van Zanten from the Biomanufacturing Training and Education Center (BTEC), NCSU for help with Differential Scanning Calorimetry and CD experiments, respectively. This work was supported by the Defense Threat Reduction Agency (grant HDTRA1-10-1-0024).

References

[1] H. K. Binz, P. Amstutz, and A. Pluckthun, "Engineering novel binding proteins from nonimmunoglobulin domains," *Nat Biotechnol,* vol. 23, pp. 1257-68, Oct 2005.

[2] A. Skerra, "Alternative non-antibody scaffolds for molecular recognition," *Curr Opin Biotechnol,* vol. 18, pp. 295-304, Aug 2007.

[3] H. I. Liao, C. A. Olson, S. Hwang, H. Deng, E. Wong, R. S. Baric, R. W. Roberts, and R. Sun, "mRNA display design of fibronectin-based intrabodies that detect and inhibit severe acute respiratory syndrome coronavirus nucleocapsid protein," *J Biol Chem,* vol. 284, pp. 17512-20, Jun 26 2009.

[4] D. Lipovsek, "Adnectins: engineered target-binding protein therapeutics," *Protein Eng Des Sel,* vol. 24, pp. 3-9, Jan 2011.

[5] D. Steiner, P. Forrer, and A. Pluckthun, "Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display," *J Mol Biol,* vol. 382, pp. 1211-27, Oct 24 2008.

[6] M. T. Stumpp, H. K. Binz, and P. Amstutz, "DARPins: a new generation of protein therapeutics," *Drug Discov Today,* vol. 13, pp. 695-701, Aug 2008.

[7] J. Lofblom, J. Feldwisch, V. Tolmachev, J. Carlsson, S. Stahl, and F. Y. Frejd, "Affibody molecules: engineered proteins for therapeutic, diagnostic and biotechnological applications," *FEBS Lett,* vol. 584, pp. 2670-80, Jun 18 2010.

[8] M. Lindborg, E. Cortez, I. Hoiden-Guthenberg, E. Gunneriusson, E. von Hage, F. Syud, M. Morrison, L. Abrahmsen, N. Herne, K. Pietras, and F. Y. Frejd, "Engineered high-affinity affibody molecules targeting platelet-derived growth factor receptor beta in vivo," *J Mol Biol,* vol. 407, pp. 298-315, Mar 25 2011.

[9] J. D. Bloom, S. T. Labthavikul, C. R. Otey, and F. H. Arnold, "Protein stability promotes evolvability," *Proc Natl Acad Sci U S A,* vol. 103, pp. 5869-74, Apr 11 2006.

[10] N. Gera, M. Hussain, R. C. Wright, and B. M. Rao, "Highly stable binding proteins derived from the hyperthermophilic sso7d scaffold," *J Mol Biol,* vol. 409, pp. 601-16, Jun 17 2011.

[11] Y. Feng, H. Yao, and J. Wang, "Crystal structure of the crenarchaeal conserved chromatin protein Cren7 and double-stranded DNA complex," *Protein Sci,* vol. 19, pp. 1253-7, Jun 2010.

[12] L. Guo, Y. Feng, Z. Zhang, H. Yao, Y. Luo, J. Wang, and L. Huang, "Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among Crenarchaea," *Nucleic Acids Res,* vol. 36, pp. 1129-37, Mar 2008.

[13] Z. Zhang, Y. Gong, L. Guo, T. Jiang, and L. Huang, "Structural insights into the interaction of the crenarchaeal chromatin protein Cren7 with DNA," *Mol Microbiol,* vol. 76, pp. 749-59, May 2010.

[14] S. Mine, T. Nakamura, Y. Hagihara, K. Ishikawa, T. Ikegami, and K. Uegaki, "NMR assignment of the chitin-binding domain of a hyperthermophilic chitinase from Pyrococcus furiosus," *J Biomol NMR,* vol. 36 Suppl 1, p. 70, 2006.

[15] T. Nakamura, K. Ishikawa, Y. Hagihara, T. Oku, A. Nakagawa, T. Inoue, M. Ataka, and K. Uegaki, "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,* vol. 61, pp. 476-8, May 1 2005.

[16] T. Nakamura, S. Mine, Y. Hagihara, K. Ishikawa, T. Ikegami, and K. Uegaki, "Tertiary structure and carbohydrate recognition by the chitin-binding domain of a hyperthermophilic chitinase from Pyrococcus furiosus," *J Mol Biol,* vol. 381, pp. 670-80, Sep 5 2008.

[17] T. Obita, S. Saksena, S. Ghazi-Tabatabai, D. J. Gill, O. Perisic, S. D. Emr, and R. L. Williams, "Structural basis for selective recognition of ESCRT-III by the AAA ATPase Vps4," *Nature,* vol. 449, pp. 735-9, Oct 11 2007.

[18] D. McMullan, R. Schwarzenbacher, L. Jaroszewski, F. von Delft, H. E. Klock, J. Vincent, K. Quijano, P. Abdubek, E. Ambing, T. Biorac, L. S. Brinen, J. M. Canaves, X. Dai, A. M. Deacon, M. DiDonato, M. A. Elsliger, S. Eshaghi, R. Floyd, A. Godzik, C. Grittini, S. K. Grzechnik, E. Hampton, C. Karlak, E. Koesema, A. Kreusch, P. Kuhn, I. Levin, T. M. McPhillips, M. D. Miller, A. Morse, K. Moy, J. Ouyang, R. Page, R. Reyes, F. Rezezadeh, A. Robb, E. Sims, G. Spraggon, R. C. Stevens, H. van den Bedem, J. Velasquez, X. Wang, B. West, G. Wolf, Q. Xu, K. O. Hodgson, J. Wooley, S. A. Lesley, and I. A. Wilson, "Crystal structure of a novel Thermotoga maritima enzyme (TM1112) from the cupin family at 1.83 A resolution," *Proteins,* vol. 56, pp. 615-8, Aug 15 2004.

[19] M. S. Almeida, T. Herrmann, W. Peti, I. A. Wilson, and K. Wuthrich, "NMR structure of the conserved hypothetical protein TM0487 from Thermotoga maritima: implications for 216 homologous DUF59 proteins," *Protein Sci,* vol. 14, pp. 2880-6, Nov 2005.

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

[21] D. Lipovsek, S. M. Lippow, B. J. Hackel, M. W. Gregson, P. Cheng, A. Kapila, and K. D. Wittrup, "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,* vol. 368, pp. 1024-41, May 11 2007.

[22] B. J. Hackel 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,* vol. 23, pp. 211-9, Apr 2010.

[23] M. Ackerman, D. Levary, G. Tobon, B. Hackel, K. D. Orcutt, and K. D. Wittrup, "Highly avid magnetic bead capture: an efficient selection method for de novo protein engineering utilizing yeast surface display," *Biotechnol Prog,* vol. 25, pp. 774-83, May-Jun 2009.

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

[25] D. S. Wilson, 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,* vol. 98, pp. 3750-5, Mar 27 2001.

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

[27] J. Hanes and A. Pluckthun, "In vitro selection and evolution of functional proteins by using ribosome display," *Proc Natl Acad Sci U S A,* vol. 94, pp. 4937-42, May 13 1997.

[28] E. T. Boder and K. D. Wittrup, "Yeast surface display for screening combinatorial polypeptide libraries," *Nat Biotechnol,* vol. 15, pp. 553-7, Jun 1997.

[29] E. T. Boder and K. D. Wittrup, "Optimal screening of surface-displayed polypeptide libraries," *Biotechnol Prog,* vol. 14, pp. 55-62, Jan-Feb 1998.

[30] J. J. VanAntwerp and K. D. Wittrup, "Fine affinity discrimination by yeast surface display and flow cytometry," *Biotechnol Prog,* vol. 16, pp. 31-7, Jan-Feb 2000.

[31] B. A. Orr, L. M. Carr, K. D. Wittrup, E. J. Roy, and D. M. Kranz, "Rapid method for measuring ScFv thermal stability by yeast surface display," *Biotechnol Prog,* vol. 19, pp. 631-8, Mar-Apr 2003.

[32] E. V. Shusta, M. C. Kieke, E. Parke, D. M. Kranz, and K. D. Wittrup, "Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency," *J Mol Biol,* vol. 292, pp. 949-56, Oct 8 1999.

[33] A. Chakrabartty, T. Kortemme, S. Padmanabhan, and R. L. Baldwin, "Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities," *Biochemistry,* vol. 32, pp. 5560-5, Jun 1 1993.

[34] P. O. Freskgard, L. G. Martensson, P. Jonasson, B. H. Jonsson, and U. Carlsson, "Assignment of the contribution of the tryptophan residues to the circular dichroism spectrum of human carbonic anhydrase II," *Biochemistry,* vol. 33, pp. 14281-8, Nov 29 1994.

[35] S. Vuilleumier, J. Sancho, R. Loewenthal, and A. R. Fersht, "Circular dichroism studies of barnase and its mutants: characterization of the contribution of aromatic side chains," *Biochemistry,* vol. 32, pp. 10303-13, Oct 5 1993.

[36] B. J. Hackel, A. Kapila, and K. D. Wittrup, "Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling," *J Mol Biol,* vol. 381, pp. 1238-52, Sep 19 2008.

[37] Y. Barrios, P. Jirholt, and M. Ohlin, "Length of the antibody heavy chain complementarity determining region 3 as a specificity-determining factor," *J Mol Recognit,* vol. 17, pp. 332-8, Jul-Aug 2004.

[38] U. Lamminmaki, S. Pauperio, A. Westerlund-Karlsson, J. Karvinen, P. L. Virtanen, T. Lovgren, and P. Saviranta, "Expanding the conformational diversity by random insertions to CDRH2 results in improved anti-estradiol antibodies," *J Mol Biol,* vol. 291, pp. 589-602, Aug 20 1999.

[39] C. V. Lee, W. C. Liang, M. S. Dennis, C. Eigenbrot, S. S. Sidhu, and G. Fuh, "High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold," *J Mol Biol,* vol. 340, pp. 1073-93, Jul 23 2004.

[40] E. P. Rock, P. R. Sibbald, M. M. Davis, and Y. H. Chien, "CDR3 length in antigen-specific immune receptors," *J Exp Med,* vol. 179, pp. 323-8, Jan 1 1994.

[41] A. Koide, R. N. Gilbreth, K. Esaki, V. Tereshko, and S. Koide, "High-affinity single-domain binding proteins with a binary-code interface," *Proc Natl Acad Sci U S A,* vol. 104, pp. 6632-7, Apr 17 2007.

[42] G. Guntas, C. Purbeck, and B. Kuhlman, "Engineering a protein-protein interface using a computationally designed library," *Proc Natl Acad Sci U S A,* vol. 107, pp. 19296-301, Nov 9 2010.

[43] S. Park, Y. Xu, X. F. Stowell, F. Gai, J. G. Saven, and E. T. Boder, "Limitations of yeast surface display in engineering proteins of high thermostability," *Protein Eng Des Sel,* vol. 19, pp. 211-7, May 2006.

[44] E. Shehi, V. Granata, P. Del Vecchio, G. Barone, P. Fusi, P. Tortora, and G. Graziano, "Thermal stability and DNA binding activity of a variant form of the Sso7d protein from the archeon Sulfolobus solfataricus truncated at leucine 54," *Biochemistry,* vol. 42, pp. 8362-8, Jul 15 2003.

[45] G. Chao, W. L. Lau, B. J. Hackel, S. L. Sazinsky, S. M. Lippow, and K. D. Wittrup, "Isolating and engineering human antibodies using yeast surface display," *Nat Protoc,* vol. 1, pp. 755-68, 2006.

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

[47] B. C. Huang and R. Liu, "Comparison of mRNA-display-based selections using synthetic peptide and natural protein libraries," *Biochemistry,* vol. 46, pp. 10102-12, Sep 4 2007.

[48] C. Gaillard and F. Strauss, "Ethanol precipitation of DNA with linear polyacrylamide as carrier," *Nucleic Acids Res,* vol. 18, p. 378, Jan 25 1990.

Tables

Table 2.1: 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 PMSKKVTKEDVLNALKNVIDFELGLDVVSLGLV**YD**I**Q**I**DDQ**N**N**V**K**V**L**MTM
5.3x10^7	.....60.....70.....80.....90.....100. TTPMCPLAGMILSDAEEAIKKIEGV**NN**V**E**V**EL**TFDPPWTPERM**S**PELREKFGV
TM1112[b] (1LKN)	1.....10.....20.....30.....40.....50 MEVKIEKPTPEKLKELSVEKWPIWEKEVSEFDWYYDTNET**S**YILEG**K**V**E**V
7x10^7	.....60.....70.....80..... **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 MSAQVMLEDMARKYAILAVKADKEGKV**E**DAI**T**YY**K**KAI**E**VL**SQ**IIVLYPE
6.7x10^7	.....60.....70.....80.. SVA**RT**AY**E**QMI**N**EY**KK**RI**S**YL**E**KVLPASSDGSG
Ph1500 (2JV2)	1.....10.....20.....30.....40.....50 HHHHHHMEGVIMS**E**L**K**L**K**PLPKVELPPDFVDVIRIKLQGKTVRTGD**V**I**G**I
6.2x10^7	.....60.....70.....80.. **S**ILGK**E**V**K**F**K**V**VQ**A**Y**PSPLRVEDRT**K**I**T**LVTHP

Table 2.1 Continued

Sso7d[c] (1SSO)	1.....10.....20.....30.....40.....50 ATVKFKYKGEEKQVDISKI**KK**V**W**R**V**GK**M**I**S**F**T**YD***L***GGGK**T**G**R**G**A**VSEKDA
$1x10^8$	.....60.. PKELLQMLEKQKK
Sso6901 (2JTM)	1.....10.....20.....30.....40.....50 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] (2CZN) $1.6x10^7$)	258.260.....270.....280.....290 .....300..305 TTPVPVSG**S**L**E**V**K**V**N**D**W**GSGA**E**Y**D**V**T**L**N**LDGQYDW**T**TVKVKLAPGATVG ....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[17]. The E35L mutation was introduced to abolish RNase activity[64].

[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 2.2: 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	**E**DAI**T**YY**K**KAI**E**VL**SQ**IIVLYPESVA**RT**AY**E**QMI**N**EY**KK**RI**S**YL**E**
MIT-rIgG[a]	**G**DAI**S**YY**W**KAI**V**VL**QR**IIVLYPESVA**TG**AY**L**QMI**G**EY**AR**RI**D**YL**N**
	......30.....40.....50..
Sso6901-WT	**K**V**W**A**L**AP**K**G**R**KG**V**K**I**G**L**F**K**DPETGK**Y**F**R**H**K**
Sso6901-β-catenin peptide[a]	**L**V**Q**A**R**AP**F**G**R**KG**V**K**R**G**L**F**R**DPETGK**A**F**F**H**L**
Sso6901-BNP-32[a]	**S**V**Y**A**F**AP**C**G**L**KG**S**K**W**G**S**F**L**DPETGK**Y**F**D**H**V**
	20.....30.....40...
Sso7d-WT	**KK**V**W**R**V**GK**M**I**S**F**T**YDLGGGK**T**G**R**G**A**
Sso7d-HEL	**CF**V**F**R**W**GK**C**I**C**F**D**YDLGGGK**Q**G**S**G**C**
Sso7d- β-catenin peptide	**NP**V**V**R**Y**GK**L**I**F**F**A**YDLGGGK**L**G**A**G**W**
Sso7d-Fluorescein[a]	**KF**V**L**R**P**GK**A**I**L**F**Y**YDLGGGK**Y**G**F**G**L**
Sso7d-Fluorescein-2	**LK**V**F**R**I**GK**V**I**F**F**R**YDLGGGK**F**G**Y**G**Y**
Sso7d-BNP-32-2	**IN**V**N**R**G**GK**F**I**R**F**T**YDLGGGK**F**G**S**G**R**
From mRNA display library	
Sso7d-rIgG[a]	**YR**V**F**R**S**GK**T**I**F**F**R**YDLGGGK**L**G**V**G**I**
Sso7d-rIgG-2	**YW**V**R**R**H**GK**S**I**T**F**Q**YDLGGGK**N**G**L**G**F**
Sso7d-rIgG-3	**QL**V**R**R**R**GK**R**I**T**F**R**YDLGGGK**K**G**V**G**Y**
Sso7d-BNP-32	**YC**V**K**R**S**GK**K**I**R**F**F**YDLGGGK**R**G**I**G**T**
Sso7d-BNP-32-3	**AR**V**W**R**V**GK**R**I**L**F**G**YDLGGGK**L**G**I**G**R**
Sso7d-BNP-32-4	**LY**V**Y**R**I**GK**R**I**I**F**A**YDLGGGK**V**G**V**G**W**

Table 2.2	**Continued**
Sso7d-BNP-32-6	**SW**V**F**R**R**GK**Y**I**I**F**A**YDLGGGK**G**G**H**G**K**
Sso7d-BNP-32-7	**QA**V**S**R**I**GK**H**I**N**F**K**YDLGGGK**Q**G**S**G**C**
	...50.....60.....70.....80
TM1112-WT	**K**V**E**V**T**T**E**D**G**K**K**Y**V**I**EK**GDLVTFPKGLRC**R**W**K**V**LE**
Tm1112-BNP-32[a, b]	**F**V**C**V**L**T**M**D**G**K**T**Y**G**I**D**----------
	.267.....277.....287...I....297.....307..
ChBD-WT[c]	**S**L**E**V**K**V**N**D**W**GSGA**E**Y**D**V**T**L**N**LDGQYDW***T***TVKVKLAPGATVGSF**W**SAN
	.313.....323.....333..
	KQEGNGYVIFTPVS**W**NKGP**T**A**T**F**G**F**I**
	.267.....277.....287...I....297.....307..
ChBD-rIgG[a, c]	**P**L**R**V**R**V**V**D**L**GSGA**W**Y**L**V**L**L**H**LDGQYDW***T***TVKVKLAPGATVGSF**Y**SAN
	.313.....323.....333..
	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 2.3: 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[a]	515-1700	-
Beta catenin C-terminus peptide	1.36	Sso6901	3297	2120-5300	-
BNP32	4	Sso6901	2100	1560-2850	-
		Tm1112	653	565-755	-
		Sso7d[b]	2696	2000-3650	-
Lysozyme	14.3	Sso7d	349[c]	225-540	-
Rabbit IgG[d]	150	MIT	183	100-380	105±8
		ChBD	271	150-530	163±9
		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. This corresponds to the concentration at which the fluorescence is half the value of the experimentally determined maximum fluorescence. 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

[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, 50]. 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 2.4: 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. 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 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.

Wild-Type or Target-binding Mutant	**T_m or $T_{1/2}$ (^oC)**
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]

[a] $T_{1/2}$ value

[b] The wild-type ChBD protein has been reported to retain its secondary structure at 85°C [14].

Table 2.5: Oligonucleotides and primers used for construction of the super-library. 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.**

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'

Table 2.5 Continued

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’

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’

Table 2.5 Continued

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’
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’

Table 2.5 Continued

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'

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'

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'

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'

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'

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'

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'

[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

Figures

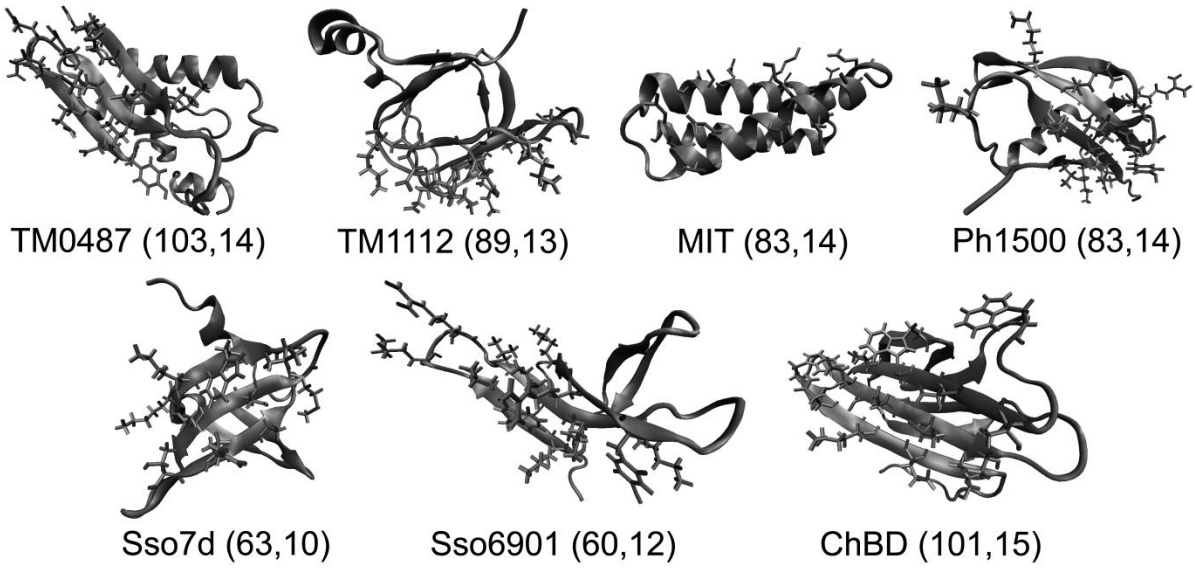


Figure 2.1: Selection of surface-accessible residues on scaffolds for randomization. Residues chosen for randomization are shown in licorice representation; this figure was generated using Visual Molecular Dynamics (VMD) software. For each scaffold, the total number of amino acid residues in the wild-type protein, followed by the number of residues randomized, is shown in parentheses.

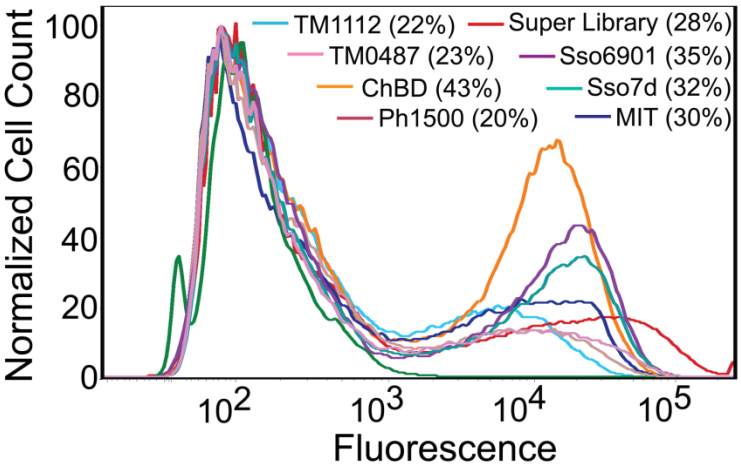


Figure 2.2: Analysis of cell surface protein expression for individual scaffold libraries and the combined super-library, through immunofluorescent detection of the c-myc epitope tag. Yeast cells displaying mutant proteins as cell surface fusions were labeled with an anti-c-myc antibody followed by a 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.

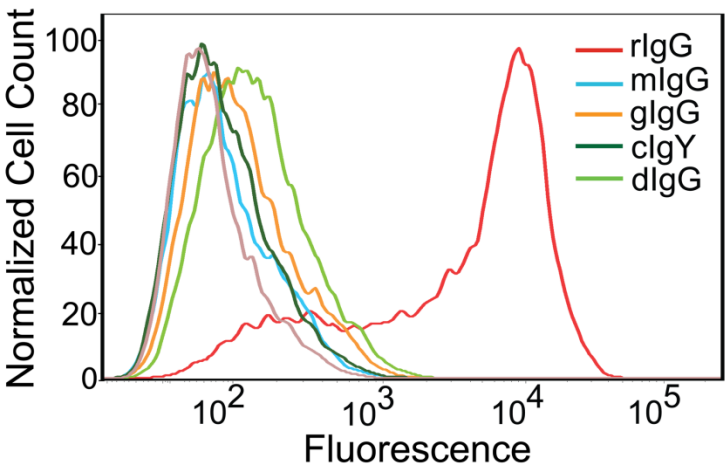


Figure 2.3: Analysis of binding specificity of MIT-rIgG. Yeast cells displaying MIT-rIgG as cell surface fusions were labeled with 1 ###M rIgG or an equivalent concentration of the non-target immunoglobulin species donkey IgG (dIgG), goat IgG (gIgG), chicken IgG (cIgY) and mouse IgG (mIgG). Fluorescence histograms confirm the high specificity of MIT-rIgG binding to rIgG; binding to other non-target immunoglobulins is insignificant.

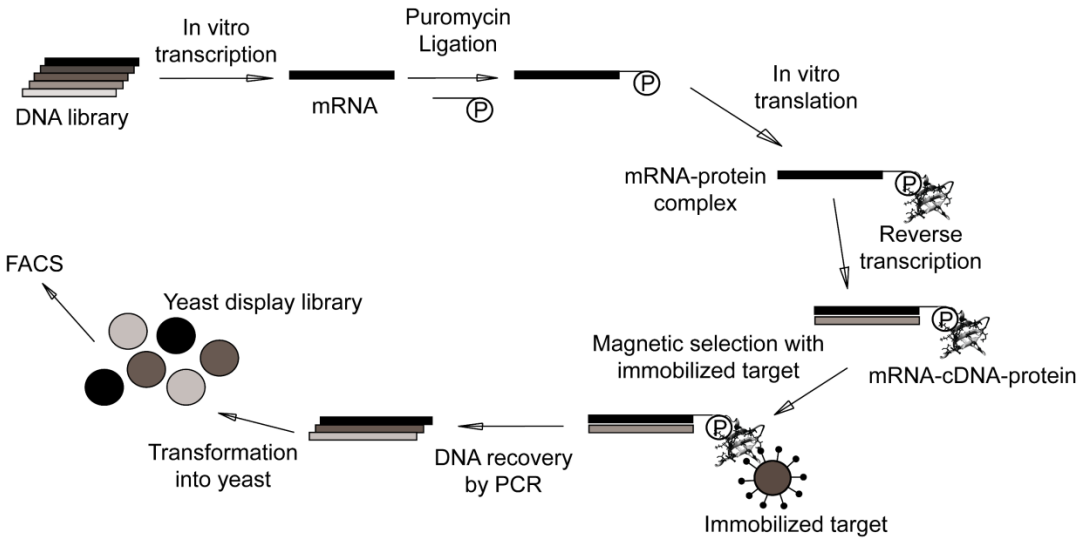


Figure 2.4: Schematic for screening of a high diversity Sso7d library using a combination of mRNA display and yeast surface display. The mRNA display was screened against magnetic beads coated with rIgG or BNP-32, and bead-bound mRNA-cDNA-protein fusions were isolated. DNA was recovered by PCR and transformed 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.

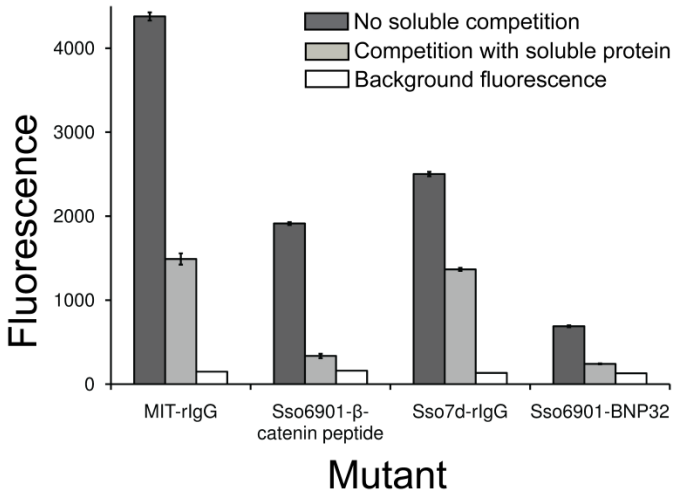


Figure 2.5: Mutant proteins recombinantly expressed in *E. coli* are functional. Yeast cells expressing mutant proteins as cell surface fusions were labeled with their corresponding targets in the absence (dark grey bars) or presence (light grey bars) of 100-200 fold excess soluble mutant protein, and 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.

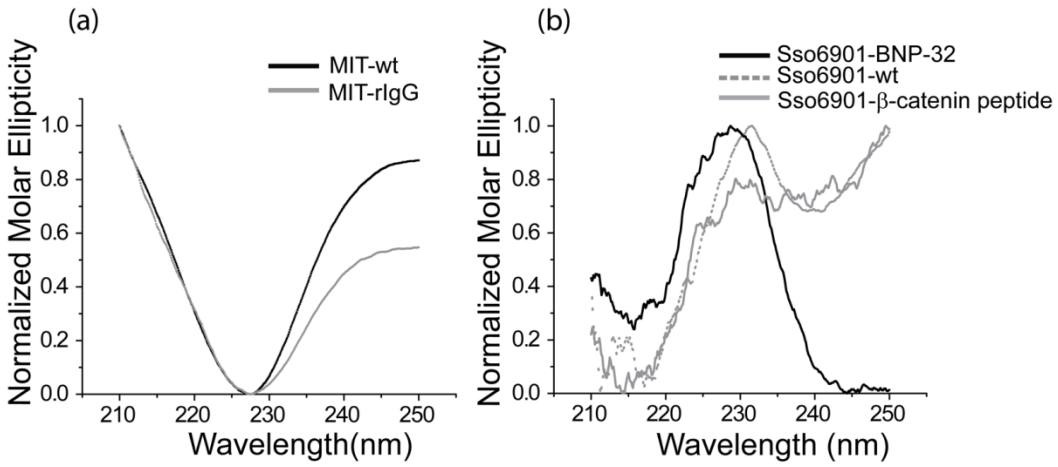


Figure 2.6: CD spectra of (a) wild-type MIT, (b) wild-type Sso6901 and mutants in 20 mM sodium phosphate buffer (pH 7.4). The baseline corrected molar ellipticity (θ) was normalized as $\bar{\theta} = \frac{\theta - \theta_{\min}}{\theta_{\max} - \theta_{\min}}$, where $\theta_{\min}$ and $\theta_{\max}$ are the minimum and maximum values of baseline-corrected molar ellipticity. Normalized ellipticity values are plotted here.

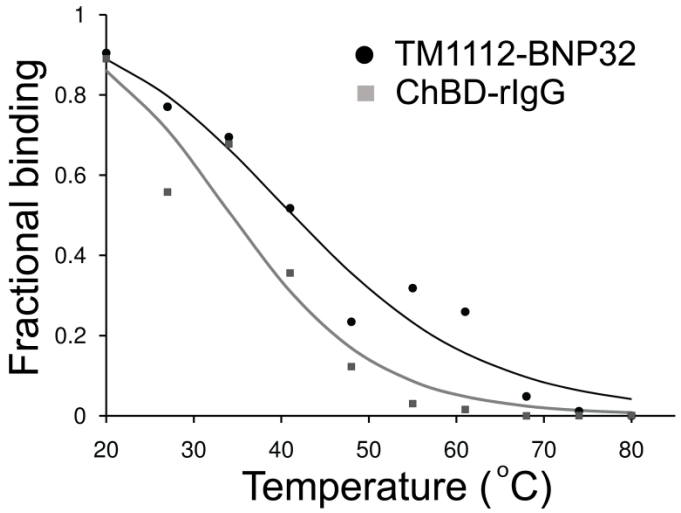
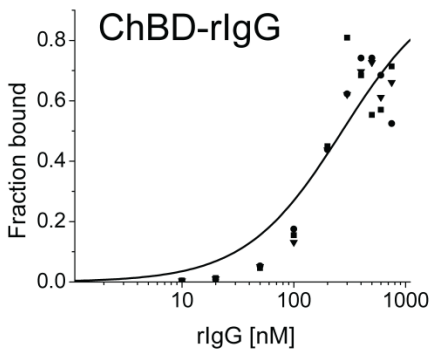


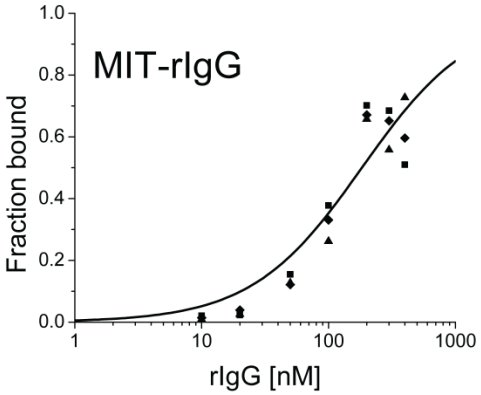
Figure 2.7: Analysis of thermal stability through determination of the mid-point of irreversible thermal denaturation ($T_{1/2}$) of yeast surface displayed protein. Yeast cells expressing Tm1112-BNP-32 and ChBD-rIgG as cell surface fusions were incubated at different temperatures in the range 20-80 ^{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.

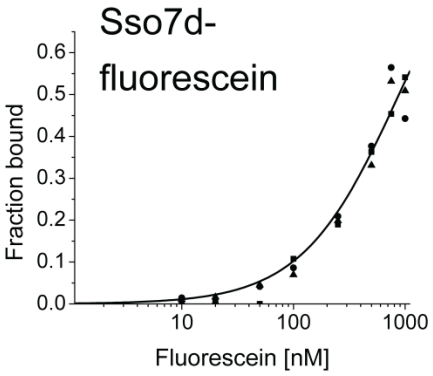
Appendices

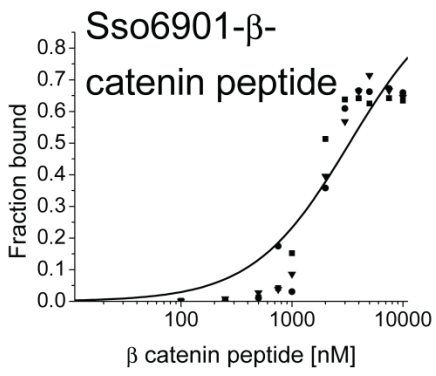
Appendix A: Supplementary Figures

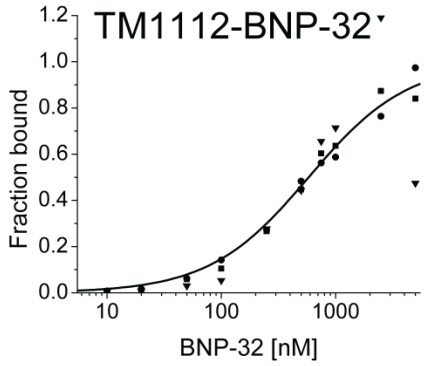
Supplementary Figure 2.1: Yeast cell surface titration plots for estimation of K_D. Yeast cells displaying mutant proteins as cell surface fusions were labeled with biotinylated target followed by streptavidin-phycoerythrin (strep-PE). Subsequently, cells were analyzed using flow cytometry. Different symbols denote data from different experiments. The solid line shows the fraction of cell surface bound fusions bound to the target, as estimated by a global fit of the experimental data to a monovalent binding isotherm. Sso7d (Y+M) refers to the mutants obtained from the mRNA display library; all other mutants were isolated from the super-library.

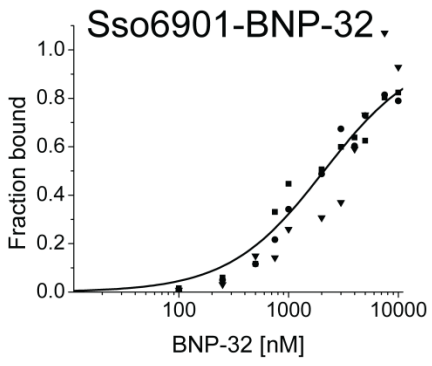

ChBD-rIgG
Fraction bound
rIgG [nM]

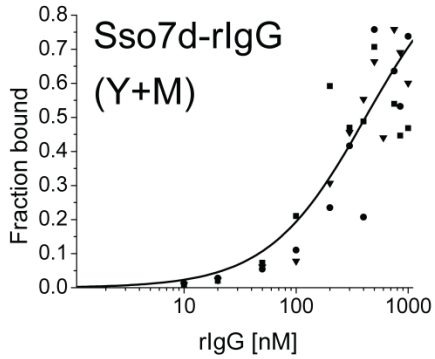

MIT-rIgG
Fraction bound
rIgG [nM]

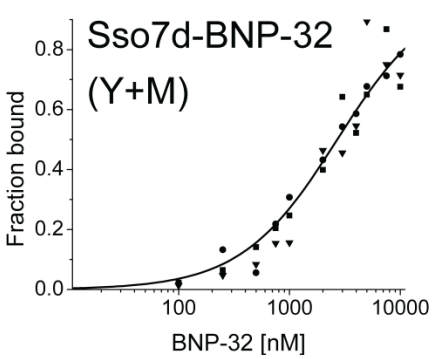

Sso7d-
fluorescein
Fraction bound
Fluorescein [nM]


Sso6901-β-
catenin peptide
Fraction bound
β catenin peptide [nM]


TM1112-BNP-32
Fraction bound
BNP-32 [nM]


Sso6901-BNP-32
Fraction bound
BNP-32 [nM]


Sso7d-rIgG
(Y+M)
Fraction bound
rIgG [nM]


Sso7d-BNP-32
(Y+M)
Fraction bound
BNP-32 [nM]

Chapter 3

A Hyperthermophilic Affinity Ligand for Virus Purification by Avidity Chromatography

Mahmud Hussain[1], Dustin Lockney[2], Ruqi Wang[2], Nimish Gera[1], Balaji M. Rao[1, 3]

[1]Department of Chemical and Biomolecular Engineering
[2]Department of Chemistry, North Carolina State University, Raleigh, NC

Abstract

We have generated a binding protein for Red clover necrotic mosaic virus (RCNMV), a plant virus, through mutagenesis of the highly stable Sso7d protein from the hyperthermophilic archaeon *Sulfolobus solfataricus*. The Sso7d-based RCNMV-binding protein (RBP) was then used to purify RCNMV from plant sap, using a variation of affinity chromatography. In this scheme – termed avidity chromatography – RCNMV from plant sap is captured on a nickel column that is pre-loaded with hexahistidine tagged RBP. The highly avid interaction between RCNMV and immobilized RBP ensures efficient capture of RCNMV despite modest binding affinity (K_D ~ 100 nM) of the RBP-RCNMV interaction. Subsequently, RBP is eluted from the column along with RCNMV; a single density gradient ultracentrifugation step is used to separate the RBP from RCNMV, as well as concentrate pure RCNMV. This purification scheme precludes the need for harsh elution conditions that are typically required for affinity chromatography of viruses and also eliminates the need for chemical conjugation of the affinity ligand to a resin. Further, the 7 kDa RBP has high thermal stability (T_m ~ 83 °C) and can be recombinantly expressed at high yield in *E. coli.* We have previously shown that highly-specific binding proteins with high thermal and pH stability can be generated for a wide spectrum of target species through mutagenesis of Sso7d [1]. Thus, taken together, Sso7d-based binding proteins are well suited for use as affinity ligands for viruses in general, and may reduce the cost and complexity of virus purification.

Introduction

Virus purification is central to applications such as production of viral vaccines and vectors for gene delivery. Typically, viral purification involves multiple centrifugation and density gradient ultracentrifugation steps at the laboratory scale, or membrane-based filtration and chromatography steps at larger scales [2]. The use of affinity chromatography can significantly reduce the cost and complexity of virus purification. In particular, immunoaffinity chromatography, which exploits the highly selective interaction between the virus particle and an immobilized antibody or antibody fragments, can separate viruses from complex mixtures in a single step. Indeed, this strategy has been previously used for virus purification in several different systems [3-6]. However, the use of antibody-based affinity chromatography for virus purification suffers from two key disadvantages. First, the use of antibodies for large-scale virus purification is impractical in general due to their relatively high cost of production and low stability. Second, immunoaffinity chromatography for viruses typically requires stringent elution conditions that may be unsuitable for labile viruses, as well as further accelerate the degradation of column performance due to antibody denaturation. The need for harsh elution conditions arises in large part due to the highly avid – and hence strong – interaction between the immobilized antibody molecules and multiple copies of their cognate targets on the viral surface. Here we describe the development of a stable non-antibody affinity ligand for a plant virus, Red clover necrotic mosaic virus (RCNMV), and subsequent purification of RCNMV from plant sap using a variation of affinity chromatography that eliminates the need for stringent elution conditions.

Plant viruses are of great interest due to their role in plant disease. More recently, the use of plant viruses in imaging and drug delivery applications has been explored. Fluorescently labeled cowpea mosaic virus (CPMV) particles have been used for vascular imaging *in vivo* [7]. Also, the hollow protein shell of plant viruses can be filled with therapeutic cargo to produce drug-filled plant virus nanoparticles (PVNs); subsequently, the outer surface of these PVNs can be modified to target specific cell types upon *in vivo* administration [8-10]. Of the several plant viruses investigated for drug delivery, RCNMV has shown significant promise[11]. RCNMV is a spherical RNA virus, 36 nm in diameter, belonging to the *Dianthovirus* genus, family *Tombusviridae* [12]. The T=3 virus capsid consists of 180 copies of a 37 kDa capsid protein; three chemically equivalent copies of the capsid protein form a trimer. Depletion of divalent cations such as Ca^{2+} and Mg^{2+} results in conformational changes in the capsid protein and formation of pores (11-13 A°) in the viral shell [13]. Small molecule drugs can be infused into the virus through these pores. Subsequently, re-addition of divalent cations can be used to close the pores and encapsulate the drug [14].

RCNMV is typically propagated by infecting a suitable host plant under standard greenhouse conditions for 7 to 10 days, and purified using multiple centrifugation and ultracentrifugation steps [15]. In this study, we generated an RCNMV binding protein (RBP) through mutagenesis of the highly stable Sso7d protein from the hyperthermophilic archaeon *Sulfolobus solfataricus*. RBP – with modest binding affinity for RCNMV (K_D ~ 100 nM) – was then used as an affinity ligand to isolate RCNMV from plant sap, using a variation of affinity chromatography termed "avidity chromatography". In our purification scheme, RBP containing a hexahistidine (6xHis) tag is immobilized on a nickel-nitriloacetic acid (Ni-NTA) column. RCNMV from plant sap is

captured, with high specificity, on the Ni-NTA column due to the highly avid interaction between RCNMV and RBP. Subsequently, RBP is eluted from the column along with RCNMV; this is different from conventional affinity chromatography schemes where the target species to be purified is eluted. Due to the large difference in molecular weight between RCNMV (~ 8.4 MDa) and RBP (~ 7 kDa), and the relatively low affinity of their binding interaction, a single ultracentrifugation step can be used to further separate RCNMV from RBP. This step also serves as a concentration and buffer-exchange step for RCNMV; RBP can be re-used for further cycles of virus purification. Notably, RBP has several advantages over antibodies such as low molecular weight, high thermal stability (T_m ~ 83°C) and ease of recombinant expression at high yield in *E. coli.* We have previously shown that highly specific binding proteins for a wide spectrum of targets can be obtained through mutagenesis of the Sso7d protein. Therefore, we expect that our approach can be extended to purification of other viruses. Taken together, our approach may lead to significant reduction in the cost and complexity of virus purification schemes.

Materials and Methods

RCNMV purification using conventional protocol. RCNMV was obtained as described by Lockney et al [15]. Briefly, leaves of *Nicotiana clevelandii* plants grown in a greenhouse at 20-24 °C for 4-6 weeks were rub-inoculated with the RNA1 and RNA2 transcripts of RCNMV. 7-10 days post-inoculation, plants were harvested by cutting the stem at the root and stored at -80 °C. Plant tissue (300 g) from -80 °C stock was combined with 900 ml of 200 mM Sodium acetate, pH 5.3 and 0.1% β-mercaptoethanol (v/v) and homogenized in a Waring Pro Stainless

Steel blender (Waring Products, Torrington, CT) set low for 30 sec on and 30 sec off, 3 times. The homogenized slurry was incubated on ice and slowly stirred for 10 min, and the slurry was strained through 4 layers of cheesecloth. The resulting sap was centrifuged at 6600xg for 25 min to remove particulate matter and the supernatant was strained through miracloth (EMD Biosciences, San Diego, CA) to remove pelleted plant debris from the sap. 450 ml of sap was mixed with 900 ml 40% PEG 8000/1 M NaCl and the virus was precipitated out of suspension by incubating on ice at 4 °C with slow stirring for 2 hours. Subsequently, the precipitant was pelleted out of suspension by centrifugation at 6600xg for 20 min at 4 °C and the virus pellets were suspended in 53 ml 50 mM HEPES pH 7.2 buffer at 4 °C. After 30 min, the suspension was centrifuged at 6600xg for 10 min. The supernatant was filtered through miracloth and further purified by ultracentrifugation through a 1-inch 20% sucrose pad at 167,000xg for 2 hours at 5 °C in a Beckman L8-70M ultracentrifuge (Beckman Coulter Inc., Brea, CA). Pellets were resuspended in 1 ml 20 mM HEPES pH 7.2 overnight at 4 °C and further centrifuged at 10,000xg for 10 min to remove insoluble material. The supernatant was subjected to ultracentrifugation again at 167,000xg for 2 hours at 5 °C through a 20% sucrose pad. Pellets were immediately suspended in 20 mM sodium phosphate buffer pH 7.2 or 20 mM sodium acetate buffer pH 5.3. After overnight suspension, the virus concentration was measured based on UV absorbance at 260 nm (A_{260}); molar absorptivity of RCNMV at 260 nm is 6.46 ml mg^{-1} cm^{-1}. An A_{260}/A_{280} ratio between 1.63-1.72 is indicative of pure virus.

Biotinylation of RCNMV. 2 mg of EZ-Link Maleimide PEG2-Biotin (Thermo Scientific, Rockford, IL) in 50 μL of DMSO was added to 5 mg of RCNMV (3 mg/ml) in 50 mM phosphate buffer, pH 7.1 and the mixture was incubated for 2 hours at room temperature.

Subsequently, the mixture was dialyzed against 4 L 1X DPBS (Dulbecco's PBS; Sigma-Aldrich, St. Louis, MO) at 4 °C for 16 hours. Biotinylated-RCNMV was then further purified by desalting on a NAP-25 column (GE Healthcare, UK) equilibrated with 1X DPBS. Fractions containing the highest concentration of virus were pooled and centrifuged for 5 min at 10,000xg to remove any aggregates. Integrity of the viral capsid was confirmed by Dynamic Light Scattering (DLS) studies at room temperature, using a Malvern 1000 ES Zetasizer (Malvern Instruments, UK). The biotin-to-RCNMV ratio was estimated as 58.5 ± 3.8:1 using a biotin quantitation kit as per the manufacturer's protocol (Thermo Scientific, Rockford, IL).

Isolation of RCNMV-binding Sso7d mutants. RCNMV-binding proteins were isolated from a combinatorial library of ~ 10^8 Sso7d mutants, using previously described protocols [1]. Briefly, biotinylated RCNMV was incubated overnight at 4 °C with 100 μl Dynal^{TM} biotin binder beads (Invitrogen, Carlsbad, CA) to generate magnetic beads coated with RCNMV. Yeast cells (~ 10^9) expressing Sso7d mutants as cell surface fusions were incubated with 100 μl Dynal^{TM} biotin binder beads pre-washed in PBS (8g/L NaCl, 0.2g/L KCl, 1.44 g/L Na_2HPO_4, 0.24 g/L KH_2PO_4, pH 7.4) containing 0.1% Bovine Serum Albumin (PBS-BSA) and bead-bound cells were discarded using a magnetic particle concentrator (Invitrogen); this step serves to eliminate mutants that bind streptavidin on the biotin-binder beads. The remaining yeast cells were incubated with ~ $2x10^7$ RCNMV-coated beads washed with PBS-BSA, for 1 hr with gentle rotation at 4 °C. Beads were subsequently isolated using a magnetic particle concentrator, washed in PBS-BSA and resuspended in 5 ml 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) with 1:100 penicillin-streptomycin solution (Invitrogen); bead-bound cells that express RCNMV-binding

mutants were expanded by culturing in a shaker (250 rpm) for 48 hrs at 30 °C.

The pool of Sso7d mutants isolated using the magnetic selection step described above was further sorted using Fluorescence Activated Cell Sorting (FACS) to isolate mutants with the highest binding affinity for RCNMV. Briefly, yeast cells expressing Sso7d mutants as cell surface fusions were simultaneously labeled with biotinylated RCNMV and a chicken antibody against the c-myc epitope tag (Invitrogen). Immunofluorescent detection was achieved using secondary labeling with streptavidin-phycoerythrin (strep-PE) and a goat ant-chicken antibody conjugated with Alexa Fluor 633 (Invitrogen). Two rounds of FACS sorting, at 350 nM and 100 nM RCNMV respectively, were carried out using a BD FacsAria^{TM} flow cytometer (BD, San Jose, CA). The pool of mutants isolated after the second sort were plated onto an SDCAA plate (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, 15 g/L agar). Plasmid DNA was isolated from six clones and sequenced. One of the clones was denoted as RCNMV-binding protein (RBP) and used for subsequent analysis.

Recombinant expression and purification of RBP. RBP was cloned into the pET22b (+) vector (EMD Biosciences, San Diego, CA) for recombinant expression with a C-terminal 6xHis tag. RBP was expressed in Rosetta^{TM} (EMD Biosciences) or BL21 (New England Biolabs (NEB), Ipswich, MA) *E. coli* cells and purified as previously described [1]. RBP was also expressed as a fusion protein with enhanced Green Fluorescent Protein (EGFP) in Rosetta^{TM} cells and purified using similar protocols. The RBP-EGFP fusion contains a $[(Gly)_4 Ser]_3$ linker between the N-terminal RBP and the C-terminal EGFP; a C-terminal 6xHis tag flanks the EGFP. Concentration of the RBP-EGFP fusion was measured using absorbance at 488 nm using an extinction co-efficient of 55,000 $M^{-1}cm^{-1}$ for EGFP [16].

Biophysical characterization of RBP. The change in apparent size of RCNMV upon incubation with RBP was analyzed using DLS. RCNMV in 50 mM Tris, 300 mM NaCl, pH 7 was equilibrated with RBP (15 μM) in a molar ratio of 1:300 (RCNMV: RBP) in a 1 ml volume. The RCNMV: RBP mixture or RCNMV alone was analyzed in a Malvern 1000 ES Zetasizer using a nonnegative least squares (NNLS) algorithm. Melting temperature of RBP dialyzed in PBS was determined by Differential Scanning Calorimetry (DSC) using a Nano DSC II differential scanning calorimeter (TA instruments, Newcastle, DE)

Estimation of K_D. $2x10^7$ Dynal™ biotin-binder beads were washed with PBS-BSA and incubated with 42.5 μg biotinylated RCNMV in 1 ml PBS-BSA overnight at 4 °C with gentle rotation. Beads were washed with PBS-BSA and incubated with varying concentrations of RBP-EGFP for 1 hr at 4 °C with gentle rotation. After the incubation step, beads were further washed with PBS-BSA and fluorescence due to bead-bound RBP-EGFP was analyzed using flow cytometry. The equilibrium dissociation constant (K_D) was estimated through a global fit of data from three separate experiments to the equation $F=\frac{c[L]_0}{K_D+[L]_0}$, where, F is the observed mean fluorescence intensity, $[L]_0$ is the concentration of RBP-EGFP, and c and K_D are fitted parameters. 68% confidence intervals were calculated as described [17]; these are similar to the commonly reported standard deviation of triplicate measurements.

Purification of RCNMV by avidity chromatography. Initial evaluation of RCNMV capture by immobilized RBP was carried out using pure RCNMV. 1 ml RCNMV (250 μg/ml) was loaded at 0.2 ml/min on a 1 ml Ni-NTA column (GE Healthcare, UK) connected to the AKTAprime™ FPLC system (GE Healthcare, UK). Alternately, 2 ml RBP (2.8 mg/ml) containing a C-terminus

6xHis tag was loaded at 0.2 ml/min on to the column, followed by washing with 5 ml Buffer A (50 mM Tris, 300 mM NaCl, pH 7.5) and loading of 1 ml RCNMV (250 μg/ml) at a flow rate of 0.2 ml/min. In each case, elution was carried out using Buffer B (50 mM Tris, 300 mM NaCl, 500 mM Imidazole, pH 7.5) at 1 ml/min, and chromatograms were recorded. Size Exclusion Chromatography (SEC) studies were carried out on an S200 16/60 column (GE Healthcare). RBP (0.5 mg/ml) was equilibrated overnight with RCNMV at 4 °C at a RBP:RCNMV molar ratio of 300:1. Subsequently, the equilibrated mixture, and RCNMV and RBP individually were flowed through the column at 1 ml/min.

To purify RCNMV from plant sap, in one experiment, 2 ml RBP (2.8 mg/ml) was loaded at 1 ml/min on a 1 ml Ni-NTA column followed by 100 ml plant sap at a flow rate of 1 ml/min. The column was washed with 16 ml 20 mM sodium-acetate, pH 5.5 and subsequently elution of protein in 1 ml fractions was obtained by flowing Buffer B at 1 ml/min. The elution fractions were pooled and subjected to ultracentrifugation at 167,000xg for 2 hrs at 5 °C through a 1-inch 20% sucrose pad. The pellet from the ultracentrifugation step was immediately resuspended in 1 ml of 20 mM sodium phosphate, pH 7.2 and incubated overnight. In parallel, supernatant from the ultracentrifugation step was dialyzed overnight in PBS to recover imidazole-free RBP for recycle. In other experiments, a 5 ml Ni-NTA column loaded with 40 mg of RBP was used for purifying RCNMV from plant sap.

Transmission Electron Microscopy (TEM). 10 μl of an elution fraction of RCNMV purified from plant extract was prepared on a copper grid (Ted Pella Inc., Redding, CA) followed by the addition of 10 μl saturated uranyl acetate (2% aqueous solution) for staining. The sample was

imaged at 80 kV; the image was captured using Kodak 4489 sheet film (Rochester, NY) in a JEOL JEM-1200EX (JEOL U.S.A., Peabody, MA) at the North Carolina State University Center for Electron Microscopy.

Results

Isolation of RCNMV-binding Sso7d mutants. The Sso7d protein from *Sulfolobus solfataricus* is a small (63 amino acids, ~ 7kDa) and highly stable protein (T_m ~ 98°C) with DNA binding activity [18-21]. We have previously shown that binding proteins for a wide spectrum of targets can be isolated from a yeast surface display library of ~ 10^8 Sso7d mutants, generated through mutagenesis of 10 amino acid residues in the DNA-binding surface of Sso7d [1]. We screened this Sso7d library for binders to RCNMV using a two-step procedure. First, yeast cells displaying Sso7d mutants as cell surface fusions were incubated with micron-sized streptavidin magnetic beads that were coated with biotinylated RCNMV; cells displaying mutants that bind RCNMV were isolated using a magnet. Subsequently, mutants isolated after magnetic selection was further screened using Fluorescence Activated Cell Sorting (FACS). Yeast cells were simultaneously labeled with RCNMV and an antibody against the c-myc epitope tag; each Sso7d mutant contains a C-terminal c-myc tag. After two rounds of FACS at 350 nM and 100 nM RCNMV respectively, a smaller pool of RCNMV-binding mutants was obtained, as shown in **Figure 3.1**. Plasmid DNA was isolated from six individual clones in this population and sequenced. Sequencing results are shown in **Table 3.1**; four distinct clones were identified. Clone Sso7d-RCNMV-4 appeared in 3 out of the six sequences analyzed and was denoted as RCNMV-binding protein (RBP). Flow cytometric analysis was used to further confirm that yeast

cells expressing RBP as cell surface fusions binds specifically to RCNMV, and not the secondary detection reagent (**Supplementary Figure 3.1**). RBP was used for all further analysis.

Biophysical characterization of RBP. RBP was recombinantly expressed in the *E. coli* cytoplasm with a C-terminal 6xHis tag and purified in a single step using Immobilized Metal Affinity Chromatography (IMAC). The unoptimized protein yield in shake flasks was estimated as ~ 25-30 mg of pure protein per liter of bacterial culture, using a bicinchoninic acid (BCA) assay with BSA as a standard. As is the case with several other Sso7d mutants [1], RBP has high thermal stability. The melting temperature (T_m) of RBP was estimated to be 83 °C using Differential Scanning Calorimetry. We used Dynamic Light Scattering (DLS) studies to confirm the RCNMV-binding activity of recombinantly expressed RBP. The diameter of RCNMV was estimated using DLS in the presence and absence of RBP. As shown in **Figure 3.2**, the average diameter of RCNMV observed increases in the presence of RBP; this suggests binding of RBP to RCNMV. We further estimated the equilibrium dissociation constant (K_D) of the binding interaction between RBP and RCNMV as follows. RBP was recombinantly expressed with enhanced Green Fluorescent Protein (EGFP) as a C-terminal fusion. Micron-sized beads coated with RCNMV were incubated with varying concentrations of RBP-EGFP and the fluorescence corresponding to bead-bound RBP was measured using flow cytometry. Subsequently the data from three separate experiments were fit to single step binding isotherm and the K_D was estimated as 93 nM using non-linear least squares regression; the associated 68% confidence interval – akin to the commonly reported standard deviations for triplicate measurements – is 61-138 nM. The binding isotherm is shown in **Figure 3.3**. This measurement corresponds to the K_D of interaction between RBP and its corresponding binding site on RCNMV. It is highly likely

that RCNMV contains multiple binding sites for RBP; indeed, there are 180 copies of the capsid protein on the viral surface. However, since RCNMV is immobilized, the measured K_D is not affected by the multivalency of RBP binding sites.

Avidity-driven capture of RCNMV by RBP. We hypothesized that a chromatography column with surface-bound RBP can efficiently capture RCNMV, due to the highly avid interaction between the multivalent RCNMV and RBP immobilized on a surface. To test this hypothesis, we immobilized RBP containing a 6xHis tag on a 1 ml Ni-NTA column. Next, we loaded 250 μg of pure RCNMV on this column. As shown in **Figure 3.4**, RCNMV is completely captured on the column. In contrast, RCNMV is not retained on the column in the absence of immobilized RBP. Subsequently, both RBP and RCNMV were eluted in the presence of imidazole. Interestingly, the elution chromatogram shows two distinct peaks. SDS-PAGE analysis confirmed that the eluted fractions corresponding to the peak with lower retention time predominantly contains RCNMV. This observation suggests that RCNMV dissociates from RBP as it traverses through the column upon elution. The very large difference in size between RCNMV and RBP – molecular weights ~ 8400 kDa and ~ 7 kDa respectively – may result in these species having different retention times in the column due to a size exclusion effect. Notably, a previous study has reported the inaccessibility of pores in an NI-NTA resin to virus particles greater than 4000 kDa [22].

These inferences were consistent with additional studies using size exclusion chromatography (SEC). Pure RCNMV, RBP or RBP equilibrated with RCNMV were loaded on to an SEC column. As seen in **Figure 3.5**, the chromatogram for the equilibrated mixture of RBP and

RCNMV shows two peaks that exactly overlay those corresponding to the pure RCNMV and

RBP samples; no peak corresponding to an RBP-RCNMV complex is observed. As the RCNMV-RBP complexes flow through the column, complex dissociation as well as separation of RCNMV and RBP due to differences in molecular size takes place simultaneously. Complexes are completely dissociated before they exit the column. Note that the time scale for complex dissociation is significantly shorter than the retention times of RBP or RCNMV in the column. Using a conservative estimate of the association rate constant of RBP binding to RCNMV (10^5 M^{-1} s^{-1}) and experimentally determined K_D, the dissociation half-life of RBP-RCNMV complexes is ~ 1.2 min; retention times for RCNMV and RBP are greater than 20 min.

Taken together, our results highlight two key points. First, RCNMV can be efficiently captured on an RBP column, due to the highly avid interaction between RCNMV and immobilized RBP. Second, due to the relatively low affinity of the RCNMV-RBP interaction in solution, RCNMV and RBP can be separated using a strategy that exploits the large difference in their molecular sizes. Examples of such a strategy include SEC or density gradient ultracentrifugation, as discussed in the following section.

Purification of RCNMV from plant sap. Current protocols for purification of RCNMV from plant sap are time-consuming and involve multiple centrifugation and ultracentrifugation steps. We hypothesized that RCNMV could be purified from plant sap using avidity chromatography, as described in the previous section, followed by a single ultracentrifugation step to simultaneously separate RBP from RCNMV and also concentrate the isolated RCNMV.

Accordingly, RBP with a 6xHis tag was immobilized on a 5 ml Ni-NTA column and plant sap was loaded on to the column. Following a wash step, RBP and RCNMV were eluted with buffer containing imidazole. The relevant chromatogram for these steps is shown in **Figure 3.6**. The elution step results in two peaks, similar to those seen in **Figure 3.4**. We analyzed the elution fractions using SDS-PAGE to confirm the presence of the virus through detection of the viral capsid protein. We further used Transmission Electron Microscopy (TEM) to verify that intact RCNMV was eluted from the column (**Figure 3.7**).

All elution fractions were combined and sucrose gradient ultracentrifugation was carried out. Subsequently, the pellet obtained was resuspended in phosphate buffer to recover RCNMV. The supernatant containing RBP was dialyzed to remove imidazole and enable reuse of RBP. In parallel, we also purified RCNMV from the same batch of plant sap using conventional protocols. RCNMV purified in this way, the corresponding product from avidity chromatography and supernatant from the ultracentrifugation step were analyzed on an SDS-PAGE gel, as shown in **Figure 3.8**. The purity of RCNMV obtained using avidity chromatography is comparable to that isolated from plant sap using conventional procedures; purity was estimated as over 90% based on densitometry analysis of the band corresponding to the 37 kDa protein. Also, the A_{260}/A_{280} ratio was determined as 1.69; A_{260}/A_{280} ratio between 1.63-1.72 is indicative of pure virus. Strikingly, the supernatant from the ultracentrifugation step contains almost pure RBP; this underscores the specificity of RCNMV capture from plant sap by RBP. Approximately 1 mg of virus was isolated with a 5 ml column in this experiment.

Discussion

We have shown that RBP, an affinity ligand obtained through mutagenesis of the highly stable Sso7d protein, can be used to isolate RCNMV from plant sap using avidity chromatography followed by a single ultracentrifugation step. Our approach reduces the overall complexity of RCNMV purification; a comparison with the currently used approach is shown in **Figure 3.9**. In addition, RBP provides significant advantages over antibodies that have been traditionally used in immunoaffinity chromatography. RBP has low molecular weight, can be produced recombinantly at high yields in *E. coli* and has high thermal stability (T_m = 83° C). In the course of our experiments, RBP (1 mg/ml) has been stored at 4 °C for several months without precipitation or loss of binding activity.

The highly avid interaction between immobilized RBP and multiple binding sites on RCNMV leads to extremely efficient capture of RCNMV, despite modest binding affinity of the RBP-RCNMV interaction (K_D ~ 100 nM). While the avidity effect enables efficient capture of polyvalent target species, the strong column-target interaction also necessitates harsh elution conditions. To put this in context, the K_D of interaction between Protein G and Immunoglobulin G (IgG) is ~ 100 nM; yet, elution of the divalent IgG requires a pH of ~ 3 [23]. Not surprisingly, the need for harsh elution conditions, primarily due to the avidity effect, is a distinct disadvantage of affinity chromatography for polyvalent virus purification. Notably, the need for harsh elution conditions is bypassed in our purification scheme by elution of RBP – as opposed to RCNMV – from the column. We used a buffer containing imidazole to elute the 6xHis tagged RBP. Alternately, a low pH buffer may be used.

The eluted fractions contain a mixture of RBP and RCNMV. Subsequently, RCNMV can be separated from RBP by exploiting the large difference in their molecular sizes. The relatively low binding affinity of RBP for RCNMV ensures that the time-scale of dissociation of the RCNMV-RBP complexes is much smaller than the time-scale of the process used to separate RBP from RCNMV. This in turn allows near complete separation of RBP and RCNMV using sucrose gradient ultracentrifugation or SEC. Alternately, an ultrafiltration step may be used to separate RBP from RCNMV. In our scheme, ultracentrifugation simultaneously serves to concentrate RCNMV and enables buffer exchange. Ultracentrifugation may also be beneficial as a final polishing step after chromatography in the broader context of virus purification. In our studies, supernatant from the ultracentrifugation step contained near pure RBP. This RBP can be re-used after a buffer exchange step; indeed the ability to recycle RBP is critical to the practical applicability of our approach. We have confirmed that recycled RBP retains RCNMV-binding activity (**Supplementary Figure 3.2**).

The yield of RCNMV was estimated as ~ 1 mg from a 5 ml Ni-NTA column loaded with ~ 40 mg of RBP; over 90% of the RBP was recovered after ultracentrifugation. In contrast, the conventional protocol yielded ~ 10 mg of RCNMV from the same quantity of plant sap. Nevertheless, as seen in **Figure 3.4**, complete capture of RCNMV can be achieved by an RBP column. Therefore it is likely that use of a bigger column will result in higher yield of RCNMV. Interestingly, chromatograms in **Figure 3.4** and **Figure 3.6** show the presence of two distinct elution peaks; upon elution, the retention time of RCNMV in the Ni column is lower than that of RBP. One possible explanation for this observation is that due to the large difference in their

molecular sizes, RBP follows a more tortuous path through the column than RCNMV. Pores in the Ni-NTA resin may be inaccessible to RCNMV due to its large size (MW > 8 MDa). Note that the inaccessibility of pores in an Ni-NTA resin to virus particles has been previously reported [22]. Also, these pores are also likely occupied by RBP, further reducing accessibility to RCNMV; the resin is pre-loaded with RBP in our purification scheme. Thus, a significant fraction of immobilized RBP may not be accessible to RCNMV, thereby reducing the yield per mg of immobilized RBP. The use of resins with larger pore sizes may increase the overall yield of RCNMV. Another possibility is to use an alternate mode of operation wherein, RCNMV in plant sap is equilibrated with RBP prior to loading onto the Ni-NTA column; in this case, RBP is not pre-loaded on the column. In a pilot experiment, we confirmed that this mode of operation does indeed result in complete capture of the virus with significantly lower amounts of RBP used per mg of virus captured (**Supplementary Figure 3.3**). We achieved complete capture of ~ 160 μg of RCNMV with ~ 0.5 mg of RBP, using a 1 ml Ni-NTA column.

In summary, we have shown that RBP – a stable non-immunoglobulin affinity ligand – can be used to design a convenient purification scheme for RCNMV. This approach can be easily extended for purification or selective capture of other viruses provided suitable affinity ligands can be identified. In this context, we have previously shown that binding proteins for a wide spectrum of targets can be obtained from a combinatorial library of Sso7d mutants. Most importantly, Sso7d mutants are highly stable, can be produced at high yields in *E. coli*, and can withstand pH extremes [1]. These properties make Sso7d-based binding proteins well suited for use as affinity ligands in virus purification. Due to their pH stability, Sso7d ligands may be used in conventional affinity chromatography schemes if the virus can withstand harsh elution

conditions. Alternately, an avidity chromatography scheme as described here may be used. In addition to precluding harsh elution conditions, a distinct advantage of our approach is elimination of the need to chemically conjugate the affinity ligand on a resin. Thus, Sso7d-based affinity ligands may reduce the cost and complexity of virus purification.

Acknowledgments

We thank Dr. Stefan Franzen, Department of Chemistry, North Carolina State University (NCSU) for helpful discussions, and Valerie Knowlton from the Center for Electron Microscopy, NCSU and Paige Luck from the Department of Food Science, NCSU for assistance with TEM and DSC, respectively. We gratefully acknowledge funding from the Defense Threat Reduction Agency (DTRA; grant HDTRA1-10-1-0024) and the National Science Foundation (NSF; grant CBET-0853771).

References

[1] N. Gera, M. Hussain, R. C. Wright, and B. M. Rao, "Highly stable binding proteins derived from the hyperthermophilic Sso7d scaffold," *J Mol Biol,* vol. 409, pp. 601-16, Jun 17 2011.

[2] M. M. Segura, A. A. Kamen, and A. Garnier, "Overview of current scalable methods for purification of viral vectors," *Methods Mol Biol,* vol. 737, pp. 89-116, 2011.

[3] R. Diaco, J. H. Hill, and D. P. Durand, "Purification of Soybean Mosaic Virus by Affinity Chromatography Using Monoclonal Antibodies," *Journal of General Virology,* vol. 67, pp. 345-351, 1986.

[4] D. Grimm, A. Kern, K. Rittner, and J. A. Kleinschmidt, "Novel tools for production and purification of recombinant adenoassociated virus vectors," *Hum Gene Ther,* vol. 9, pp. 2745-60, Dec 10 1998.

[5] M. Njayou and G. Quash, "Purification of measles virus by affinity chromatography and by ultracentrifugation: a comparative study," *J Virol Methods,* vol. 32, pp. 67-77, Apr 1991.

[6] R. H. Smith, J. R. Levy, and R. M. Kotin, "A simplified baculovirus-AAV expression vector system coupled with one-step affinity purification yields high-titer rAAV stocks from insect cells," *Mol Ther,* vol. 17, pp. 1888-96, Nov 2009.

[7] J. D. Lewis, G. Destito, A. Zijlstra, M. J. Gonzalez, J. P. Quigley, M. Manchester, and H. Stuhlmann, "Viral nanoparticles as tools for intravital vascular imaging," *Nat Med,* vol. 12, pp. 354-60, Mar 2006.

[8] G. Destito, R. Yeh, C. S. Rae, M. G. Finn, and M. Manchester, "Folic acid-mediated targeting of cowpea mosaic virus particles to tumor cells," *Chem Biol,* vol. 14, pp. 1152-62, Oct 2007.

[9] Y. Ren, S. M. Wong, and L. Y. Lim, "Folic acid-conjugated protein cages of a plant virus: a novel delivery platform for doxorubicin," *Bioconjug Chem,* vol. 18, pp. 836-43, May-Jun 2007.

[10] Y. Ren, S. M. Wong, and L. Y. Lim, "Application of plant viruses as nano drug delivery systems," *Pharm Res,* vol. 27, pp. 2509-13, Nov 2010.

[11] D. M. Lockney, R. N. Guenther, L. Loo, W. Overton, R. Antonelli, J. Clark, M. Hu, C. Luft, S. A. Lommel, and S. Franzen, "The Red clover necrotic mosaic virus capsid as a multifunctional cell targeting plant viral nanoparticle," *Bioconjug Chem,* vol. 22, pp. 67-73, Jan 19 2011.

[12] V. R. Basnayake, T. L. Sit, and S. A. Lommel, "The genomic RNA packaging scheme of Red clover necrotic mosaic virus," *Virology,* vol. 345, pp. 532-9, Feb 20 2006.

[13] M. B. Sherman, R. H. Guenther, F. Tama, T. L. Sit, C. L. Brooks, A. M. Mikhailov, E. V. Orlova, T. S. Baker, and S. A. Lommel, "Removal of divalent cations induces structural transitions in red clover necrotic mosaic virus, revealing a potential mechanism for RNA release," *J Virol,* vol. 80, pp. 10395-406, Nov 2006.

[14] L. Loo, R. H. Guenther, S. A. Lommel, and S. Franzen, "Infusion of dye molecules into Red clover necrotic mosaic virus," *Chem Commun (Camb),* pp. 88-90, Jan 7 2008.

[15] D. Lockney, S. Franzen, and S. Lommel, "Viruses as nanomaterials for drug delivery," *Methods Mol Biol,* vol. 726, pp. 207-21, 2011.

[16] S. R. McRae, C. L. Brown, and G. R. Bushell, "Rapid purification of EGFP, EYFP, and ECFP with high yield and purity," *Protein Expr Purif,* vol. 41, pp. 121-7, May 2005.

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

[18] H. Baumann, S. Knapp, A. Karshikoff, R. Ladenstein, and T. Hard, "DNA-binding surface of the Sso7d protein from Sulfolobus solfataricus," *J Mol Biol,* vol. 247, pp. 840-6, Apr 14 1995.

[19] H. Baumann, S. Knapp, T. Lundback, R. Ladenstein, and T. Hard, "Solution structure and DNA-binding properties of a thermostable protein from the archaeon Sulfolobus solfataricus," *Nat Struct Biol,* vol. 1, pp. 808-19, Nov 1994.

[20] S. P. Edmondson and J. W. Shriver, "DNA binding proteins Sac7d and Sso7d from Sulfolobus," *Methods Enzymol,* vol. 334, pp. 129-45, 2001.

[21] S. Knapp, A. Karshikoff, K. D. Berndt, P. Christova, B. Atanasov, and R. Ladenstein, "Thermal unfolding of the DNA-binding protein Sso7d from the hyperthermophile Sulfolobus solfataricus," *J Mol Biol,* vol. 264, pp. 1132-44, Dec 20 1996.

[22] V. Krasnykh, N. Belousova, N. Korokhov, G. Mikheeva, and D. T. Curiel, "Genetic targeting of an adenovirus vector via replacement of the fiber protein with the phage T4 fibritin," *J Virol,* vol. 75, pp. 4176-83, May 2001.

[23] H. Watanabe, H. Matsumaru, A. Ooishi, Y. Feng, T. Odahara, K. Suto, and S. Honda, "Optimizing pH response of affinity between protein G and IgG Fc: how electrostatic modulations affect protein-protein interactions," *J Biol Chem,* vol. 284, pp. 12373-83, May 1 2009.

Tables

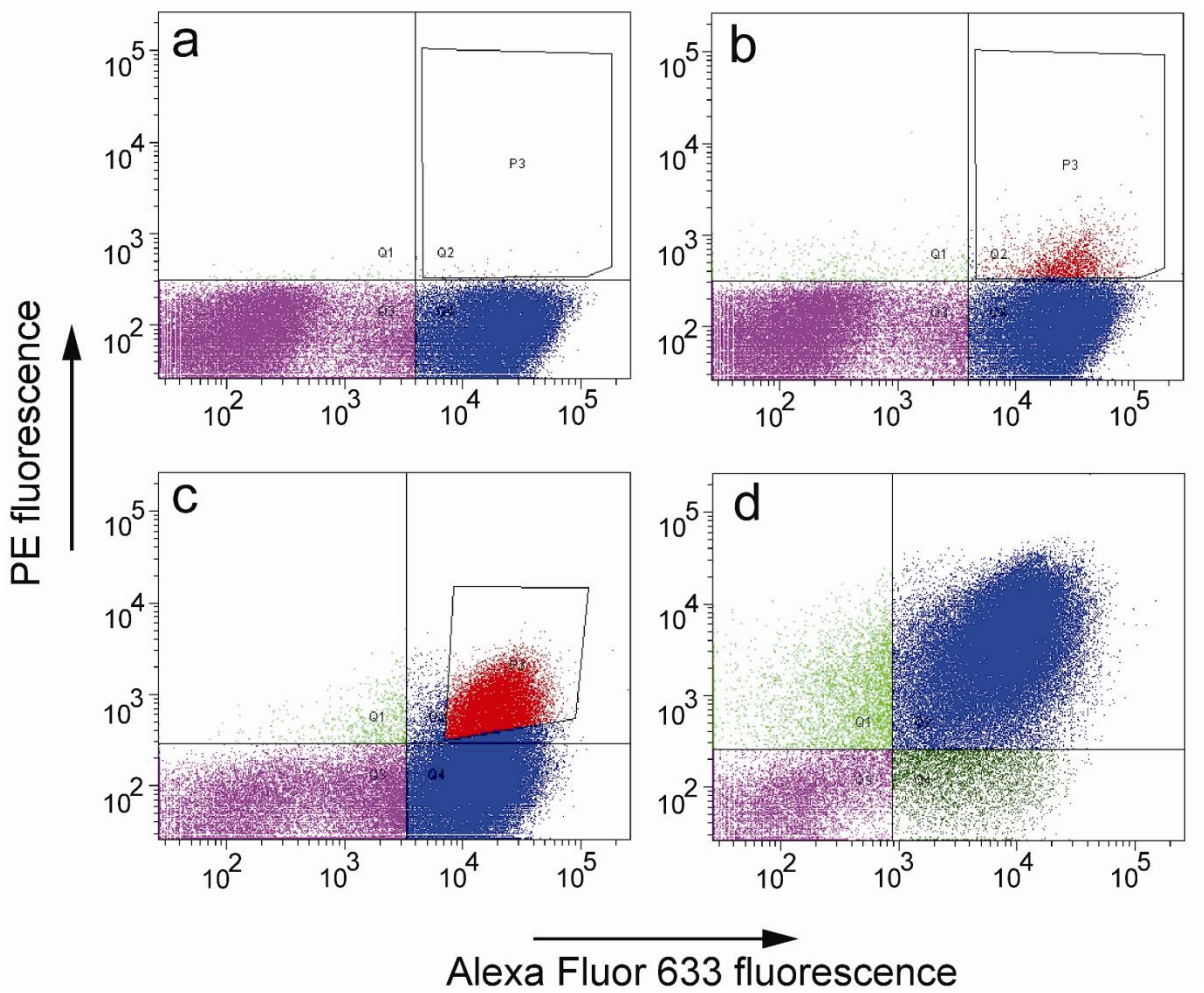
Table 3.1: Amino acid sequences of RCNMV-binding Sso7d mutants. Mutated residues are in bold font and underlined, only residues in the mutated region of Sso7d is shown.

Wild type/mutant	Sequence
	20.....30.....40...
Sso7d-WT	**KK**V**W**R**V**GK**M**I**S**F**T**YDLGGGK**T**G**R**G**A**
Sso7d-RCNMV-1	**LR**V**Y**R**T**GK**M**I**F**F**S**YDLGGGK**Y**G**S**G**L**
Sso7d-RCNMV-2	**IR**V**Y**R**T**GK**S**I**Y**F**G**YDLGGGK**F**G**Y**G**F**
Sso7d-RCNMV-3	**LR**V**Y**R**S**GK**Q**I**F**F**A**YDLGGGK**F**G**G**G**N**
Sso7d-RCNMV-4 (RBP)[a]	**KY**V**L**R**R**GK**G**I**W**F**I**YDLGGGK**T**G**F**G**D**

[a]Sequence corresponding to RBP was indentified in 3 out of 6 clones sequenced; RBP was used for all further analysis.

Figures

Figure 3.1: Isolation of RCNMV-binding Sso7d mutants by FACS. Yeast cells displaying Sso7d mutants as cell surface fusions were simultaneously labeled with biotinylated RCNMV and a saturating concentration of chicken anti-c-myc antibody, followed by secondary labeling with streptavidin-phycoerythrin (strep-PE) and a goat anti-chicken antibody conjugated with Alexa Fluor 633. **(a)** Yeast cells labeled only with the anti-c-myc antibody (no RCNMV control) **(b)** The pool of yeast cells obtained after magnetic selection was labeled at 350 nM RCNMV. Cells in the polygonal region were sorted by FACS.
(c) Yeast cells from the 350 nM sort were labeled with 100 nM RCNMV and cells in the polygonal region were sorted again by FACS. **(d)** Yeast cells from the 100 nM sort were labeled with 10 nM RCNMV. Individual clones from this population were sequenced.


a
b
c
d
P3
Q1
Q2
Q3
Q4
PE fluorescence
Alexa Fluor 633 fluorescence

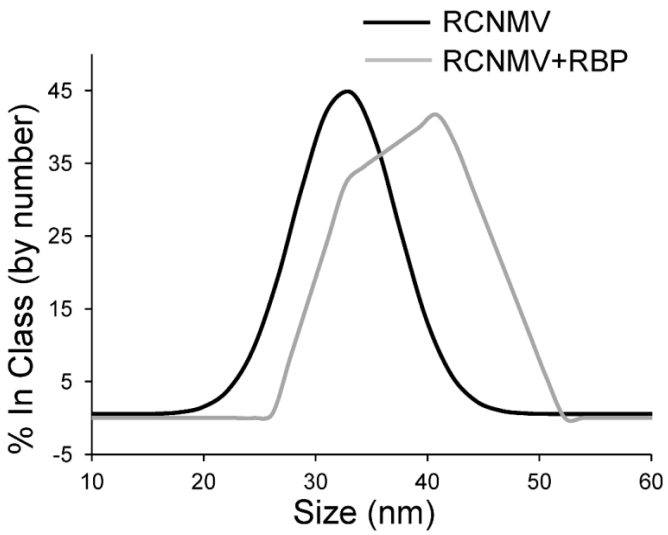


Figure 3.2: Dynamic Light Scattering (DLS) spectra of RCNMV and an equilibrated mixture of RCNMV and RBP. RCNMV was equilibrated with RBP (15 μM) at a molar ratio of 1:300 (RCNMV: RBP), in a 1 ml volume. DLS spectra for this mixture and RCNMV alone were recorded. The increase in particle size distribution of RCNMV in the presence of RBP (grey curve) relative to RCNMV alone (black curve) indicates binding of RBP to RCNMV.

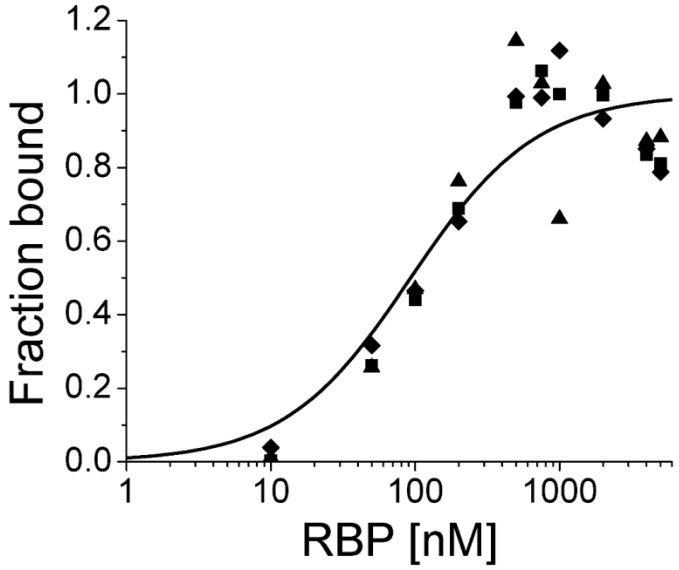


Figure 3.3: Titration plot for estimation of K_D. Biotinylated RCNMV immobilized on biotin-binder magnetic beads was incubated with varying concentrations of RBP-EGFP and beads were analyzed using flow cytometry. Different symbols indicate data from three separate experiments. The solid line shows the fraction of immobilized RCNMV bound to RBP-EGFP, as estimated by a global fit of the experimental data to a one-step binding isotherm.

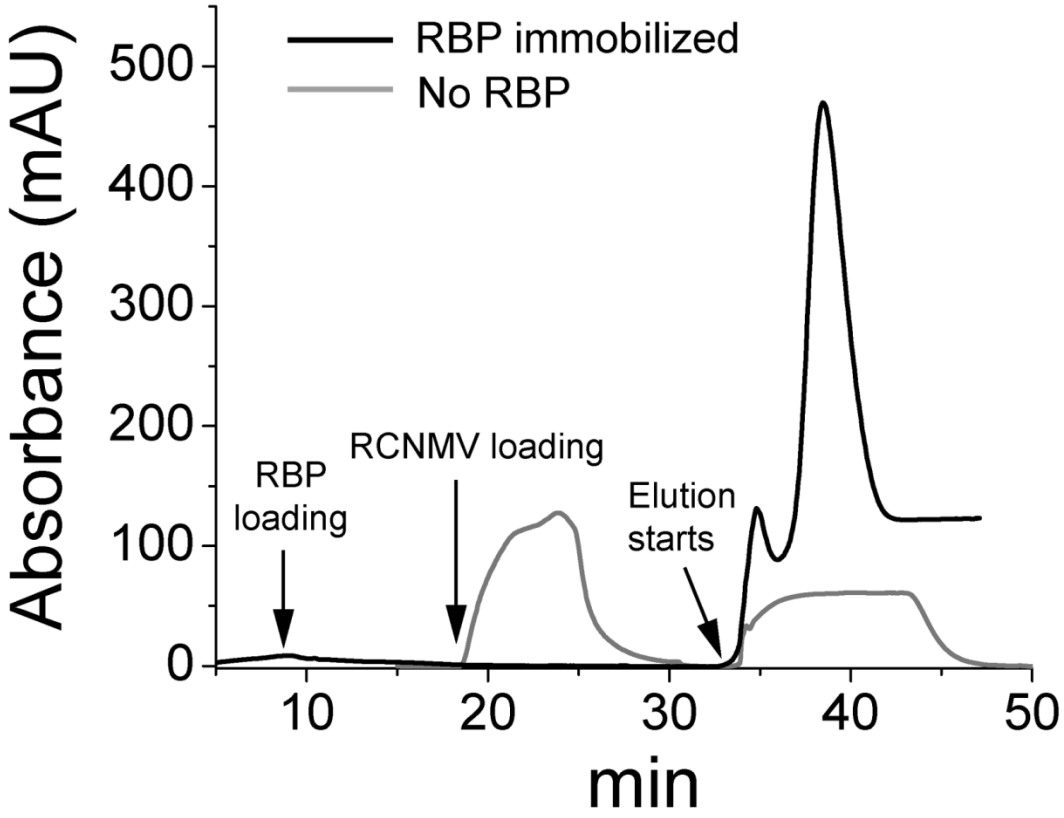


Figure 3.4: RCNMV is captured by immobilized RBP. Pure RCNMV (250 μg) was loaded on a 1 ml Ni-NTA column. No RCNMV is captured in the absence of immobilized RBP, as seen in the peak between 15 and 30 min (grey curve). In a separate experiment, (black curve), the Ni-NTA column was loaded with 6xHis-tagged RBP followed prior to introduction of RCNMV (250 μg). All RCNMV is captured on the column. Upon introduction of elution buffer containing imidazole (at ~ 32 min on the chromatogram), RCNMV is co-eluted with RBP.

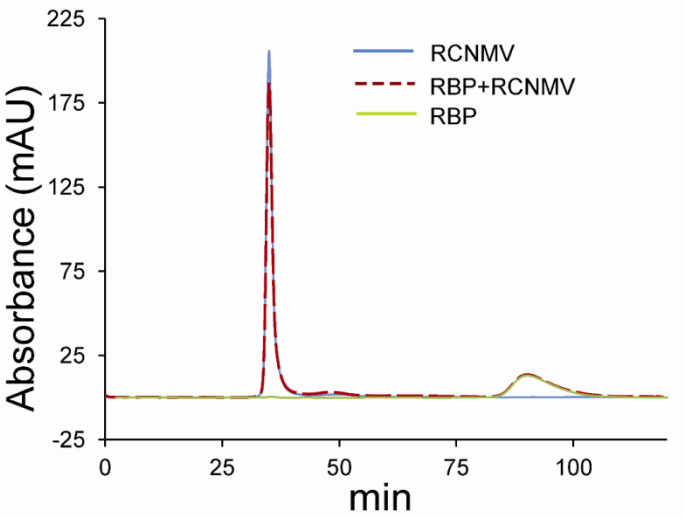


Figure 3. 5: Size exclusion chromatography analysis of RBP-RCNMV binding. Pure RCNMV (0.5 mg/ml) was incubated overnight with RBP in a 1:300 molar ratio of RCNMV:RBP. Pure RCNMV, RBP or the equilibrated mixture of RBP and RCNMV were loaded on to an SEC column. The chromatogram for the equilibrated mixture shows two peaks that exactly overlay those corresponding to the pure RCNMV and RBP samples; no peak corresponding to an RBP-RCNMV complex is observed.

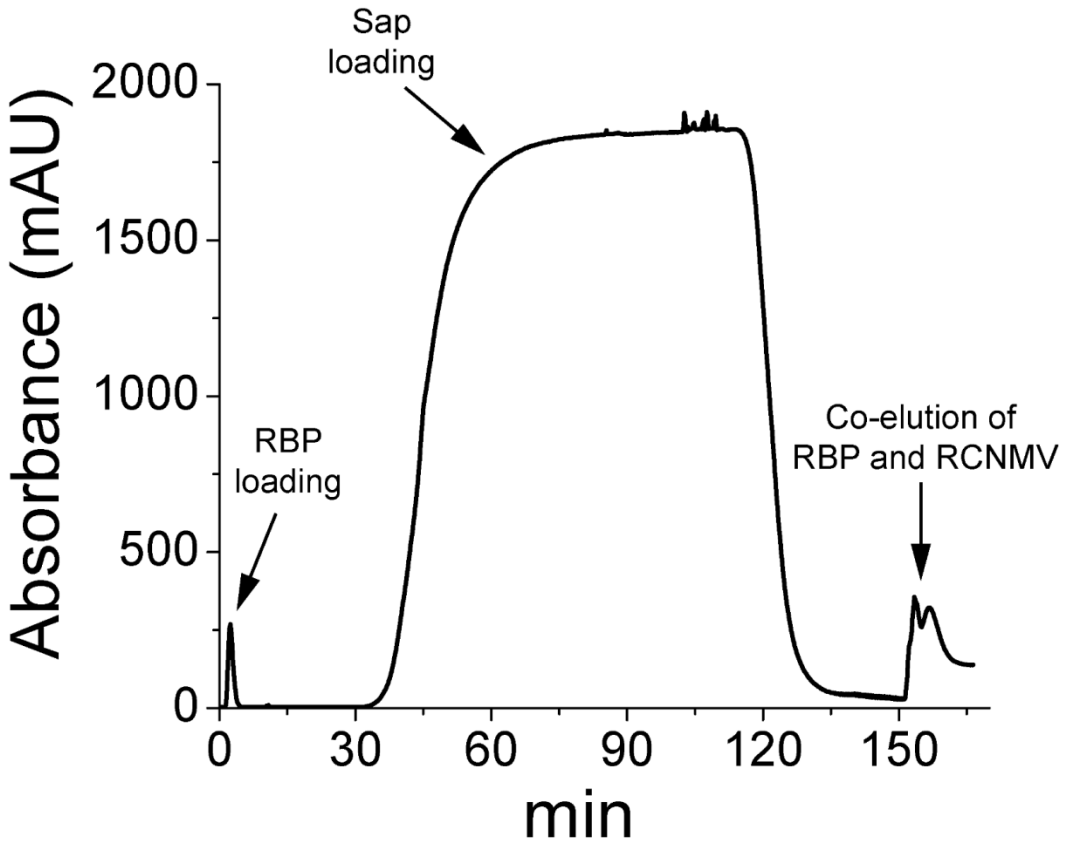


Figure 3.6: RCNMV purification from plant sap by avidity chromatography. A 1 ml Ni-NTA column was saturated with 6XHis tagged RBP, and loaded with plant sap. RCNMV was subsequently co-eluted with RBP using buffer containing imidazole.

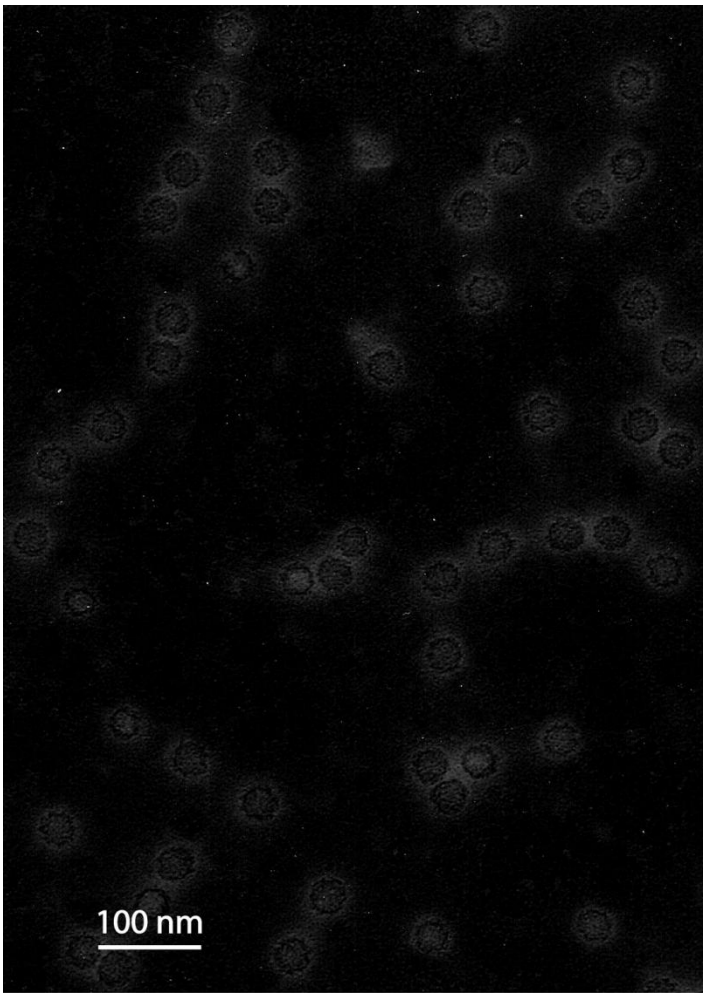


Figure 3.7: TEM image of RCNMV particles purified from plant sap by avidity chromatography. Plant sap was flowed through a Ni-NTA column pre-loaded with 6xHis tagged-RBP. TEM grids were prepared from eluted fractions containing RCNMV, as identified by analysis on an SDS-PAGE gel.

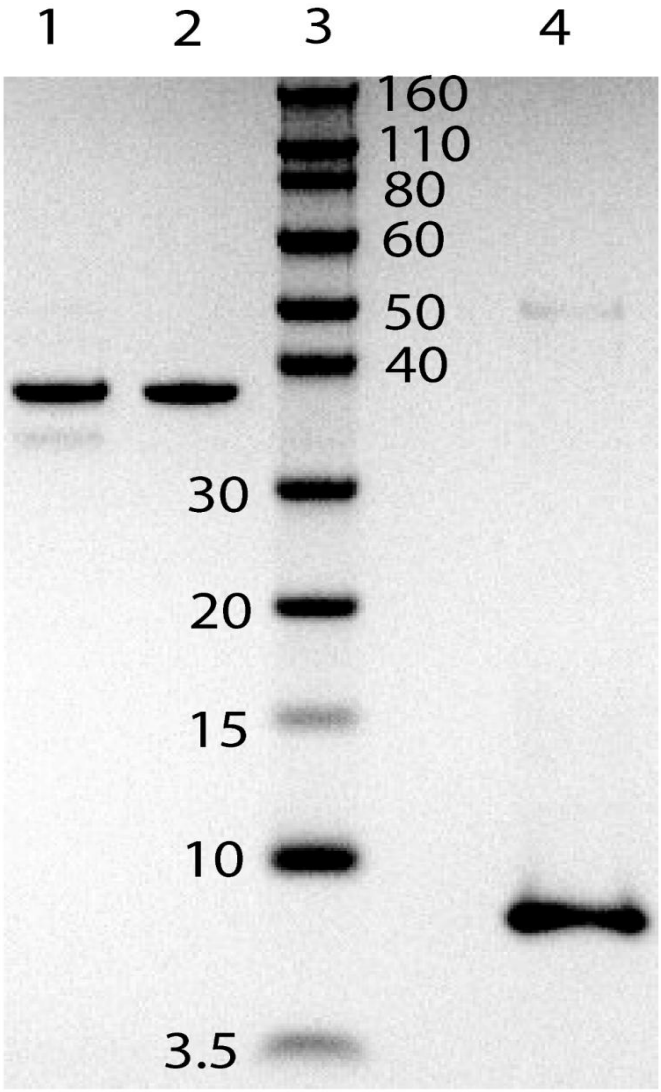
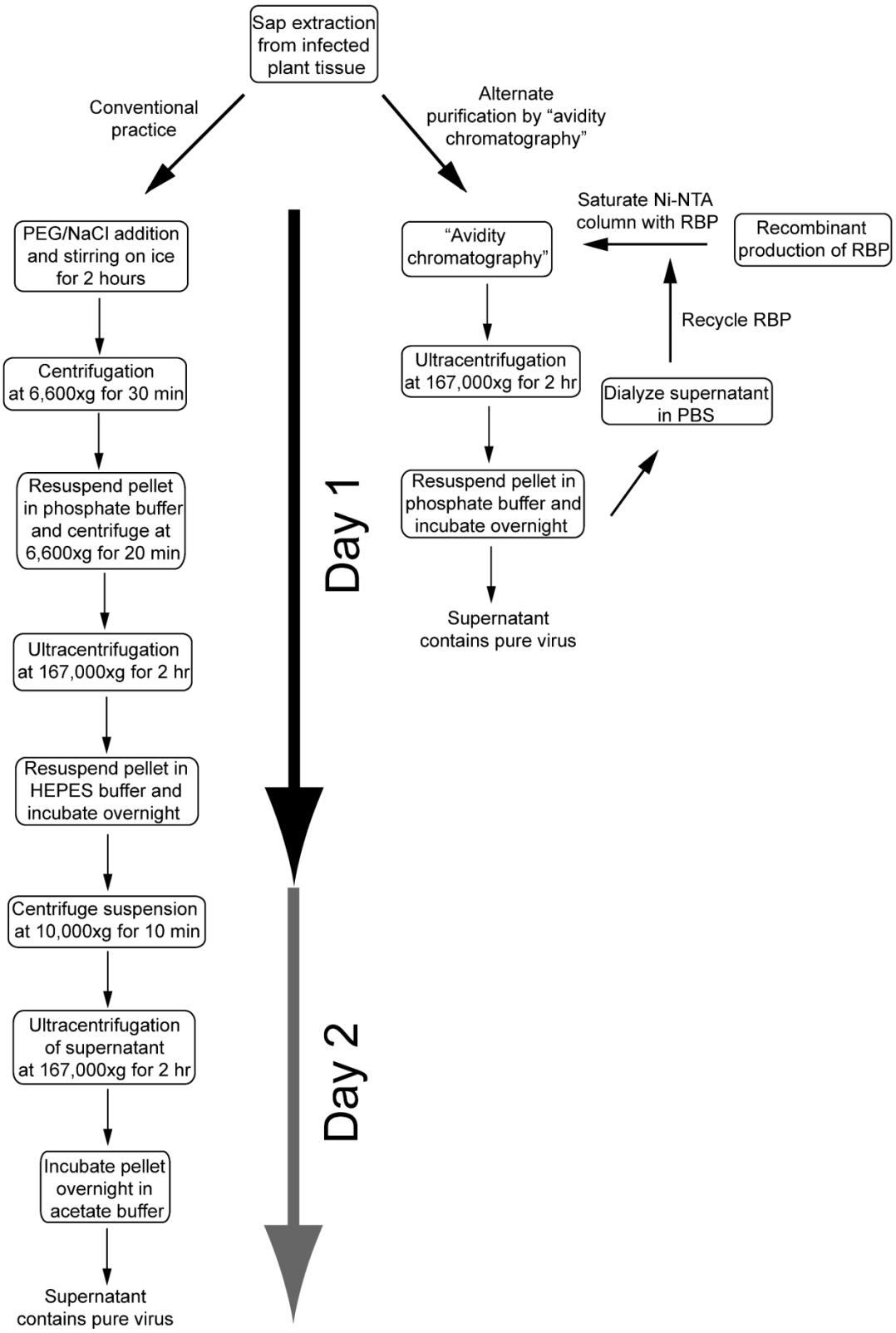


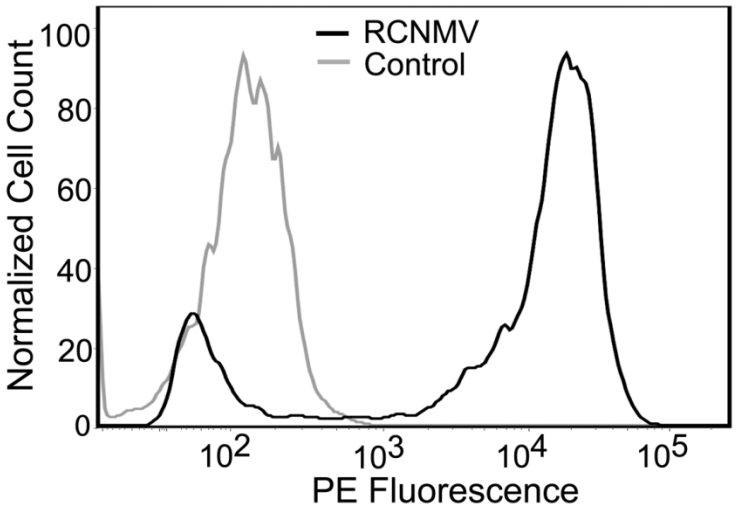
Figure 3.8: SDS-PAGE analysis of RCNMV purified by avidity chromatography, and RCNMV obtained using the conventional method. **Lane 1:** RCNMV obtained using avidity chromatography followed by a single ultracentrifugation step. A protein band corresponding to the capsid protein (37 kDa) is observed. **Lane 2:** RCNMV purified by conventional method **Lane 3:** Molecular weight marker. **Lane 4:** RBP recovered from supernatant in the ultracentrifugation step.

Figure 3.9: Comparison between RCNMV purification from plant sap by avidity chromatography and the conventional method.

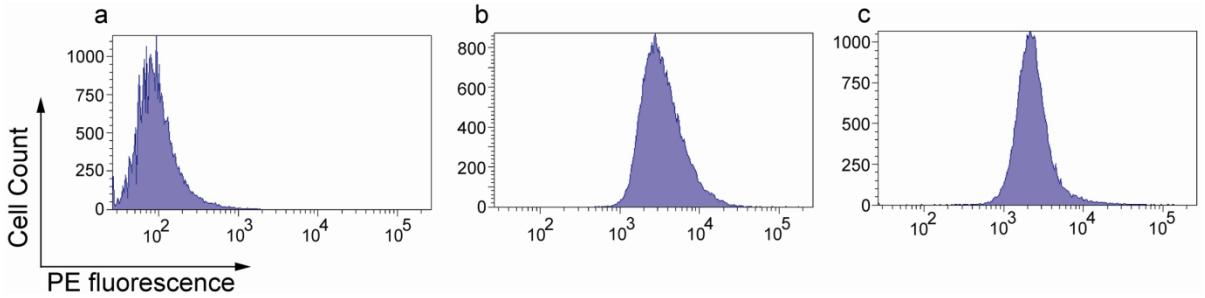

Sap extraction from infected plant tissue
Conventional practice
Alternate purification by "avidity chromatography"
PEG/NaCl addition and stirring on ice for 2 hours
Centrifugation at 6,600xg for 30 min
Resuspend pellet in phosphate buffer and centrifuge at 6,600xg for 20 min
Ultracentrifugation at 167,000xg for 2 hr
Resuspend pellet in HEPES buffer and incubate overnight
Centrifuge suspension at 10,000xg for 10 min
Ultracentrifugation of supernatant at 167,000xg for 2 hr
Incubate pellet overnight in acetate buffer
Supernatant contains pure virus
Day 1
Day 2
"Avidity chromatography"
Saturate Ni-NTA column with RBP
Recombinant production of RBP
Recycle RBP
Ultracentrifugation at 167,000xg for 2 hr
Dialyze supernatant in PBS
Resuspend pellet in phosphate buffer and incubate overnight
Supernatant contains pure virus

Appendices

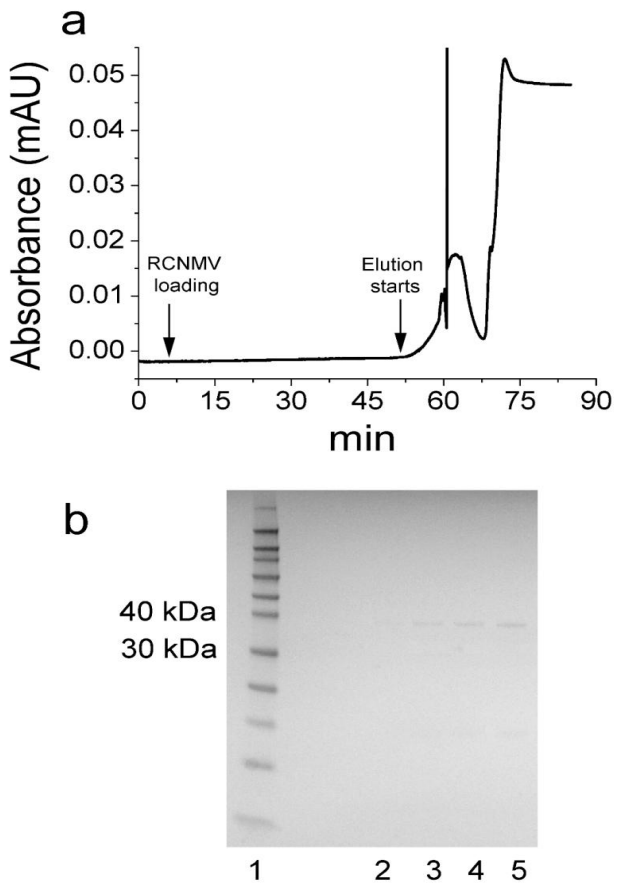
Appendix A: Supplementary Figures



Supplementary Figure 3.1: RBP binds specifically to RCNMV and not the secondary detection reagent. Yeast cells expressing RBP as cell surface fusions were labeled with 10 nM biotinylated RCNMV (black histogram) or PBS-BSA only (grey histogram), followed by secondary labeling with strep-PE, and analyzed using flow cytometry. The higher PE fluorescence in the presence of RCNMV (black histogram) shows specific binding of yeast cell surface displayed RBP to RCNMV.



Supplementary Figure 3.2: RBP from the supernatant in the ultracentrifugation step retains RCNMV-binding activity. 40 μg each of fresh RBP and RBP from the supernatant in the ultracentrifugation step were immobilized on Ni-NTA magnetic beads through overnight incubation at 4 °C. Beads were washed in PBS-BSA and labeled with 30 nM biotinylated RCNMV, followed by secondary labeling with strep-PE, and analyzed by flow cytometry. **(a)** Fluorescence histogram of control beads (without immobilized RBP) labeled with strep-PE only. **(b)** Fluorescence histogram of beads with fresh RBP immobilized, and labeled with biotinylated RCNMV followed by strep-PE. **(c)** Fluorescence histogram of beads coated with RBP obtained from the supernatant of the ultracentrifugation step.



Supplementary Figure 3.3: Capture of RCNMV on a Ni-NTA column when RBP incubated with RCNMV prior to loading on the column. **(a)** Pure RCNMV (0.162 mg) and RBP (0.5 mg) in 1 ml were incubated on ice for 40 min and loaded on a 1 ml Ni-NTA column at 0.2 ml/min; loading commences at ~ 5 min on the chromatogram. The column was washed with Buffer A (50 mM Tris, 300 mM NaCl, pH 7.5) for 5 min at 0.2 ml/min and eluted with buffer containing imidazole; elution commences at ~ 55 min on the chromatogram. RCNMV is completely captured on the column, and subsequently eluted in the presence of imidazole. **(b)** SDS-PAGE analysis of eluted fractions confirms the presence of a 37 kDa band corresponding to the coat protein of RCNMV. Lane 1: molecular weight marker. Lanes 2-5: 1 ml eluted fractions collected between 62-65 min, as shown in the chromatogram.

Chapter 4

Free Solution Capillary Electrophoresis (FSCE) for Biomolecular separation

Abstract

The key to understanding any cellular process is to measure the concentrations of relevant biomolecules such as proteins and messenger RNAs. The heart of any measurement process is a biomolecular probe that specifically recognizes and binds the target of interest with high affinity. When it comes to measuring the concentrations of proteins, antibodies are the first choice as probe. Antibodies however have large, multi-domain structure containing disulfide bridges and hence they cannot be easily produced in common bacterial expression systems. Also, antibodies against the ever increasing list of target-proteins are not commercially available. Apart from antibodies, another challenge facing quantitative biology is the detection of proteins and mRNAs present in low concentration (~ picogram/liter) as well as in low volume (~ tens of microliters). This is especially true in the context of quantitative cell biology where the number of cells available for analysis may be too few to gather sufficient amount of proteins or RNA for detection. We show here the proof-of-concept of a novel FSCE-based approach to isolate binders from library screening as well as mRNA quantification.

Free Solution Capillary Electrophoresis-based probe separation

Biomolecular probes that uniquely recognize targets of interest are generated by introducing mutations in the scaffold protein. The specific mutations that need to be introduced in the scaffold protein are most commonly identified through combinatorial methods. Simplistically, a large library of mutant proteins based on the scaffold is created and the proteins that exhibit the highest affinity of binding for the target are selected. The key challenge in this process is to identify and isolate the protein with desired properties and is achieved using tools such as mRNA display that links the genotype to the phenotype [1-10]. Other combinatorial screening techniques such as yeast surface display and phage display have also been reported elsewhere [11-18]. In mRNA display, a protein/peptide is attached to the nucleic acid sequence that codes for it via puromycin (**Figure 4.1**). Puromycin is an antibiotic that can form a covalent bond between the C-terminus of the protein and the mRNA that encodes it. The DNA encoding the protein is *in vitro* transcribed into an mRNA without a stop codon. The 3' end of the mRNA is then ligated with a nucleic acid linker modified to have puromycin at the 3' end. Since there is no stop codon, the growing peptide does not cleave off of the mRNA during the *in vitro* translation step and allows puromycin to form a covalent bond with the C-terminus of the protein. Puromycin acts as a translation inhibitor [19, 20]. A reverse transcription step is used to obtain an mRNA-cDNA-protein hybrid. In a typical selection scheme using mRNA display (**Figure 4.1**), a library of the hybrid molecules is incubated with an immobilized target. The hybrid-target complexes are isolated, amplified by error-prone PCR and the entire process is repeated. After a few rounds, a pool of protein variants (in the mRNA-cDNA-protein format) with high affinity for the target is obtained. The identity of the protein scaffold variants is established using PCR.

Thus, linking of genotype (mRNA) to the phenotype (protein) can be achieved by mRNA display which allows retrieval of the identity of a functional protein after affinity maturation. MRNA display has been employed in obtaining mutant proteins with nanomolar affinity for 10th domain of Fibronectin type III and ScFv [1, 21]. MRNA display has also been used for improving the binding affinity of various proteins for the elucidation of many different phenomena, such as protein-DNA interaction and identification of drug target [21-28]. MRNA display has also been used for improving the binding affinity of various proteins for the elucidation of many different phenomena, such as protein-DNA interaction and identification of drug target [1, 29]. The reason for obtaining such an enormous number of mutants in a library is that no host organism is involved, which obviates the limitation inherent in the *in vivo* methods of selection. However, following limitations exist in the selection of proteins using mRNA display:

1) *Labeling of the library members:* In the *in vitro* translation step, radioactive methionine is used to detect the translation product [2, 4, 8, 22, 23, 30]. Radioactivity is also required for carrying out the binding assay for the detection of potential mutants.

2) *Immobilization of the target molecule:* A common feature of library screening procedures for selecting functional mutants from a library is immobilization of the target molecule. This immobilization may compromise the probable binding site of the target molecule which leads to the problem of "False Negatives".

3) *Soluble target protein:* The selection process typically requires at least several tens of micrograms of target protein. This may be difficult, especially in the case of targets such as membrane receptors.

We have developed a novel method based on free solution capillary electrophoresis (FSCE) to isolate binders from an mRNA display library. FSCE refers to the migration of charged species under the influence of an electric field, in a narrow channel containing a conducting buffer. The differential rate of migration can be used to separate the charged species. A key step in directed evolution by mRNA display involves the separation of the mRNA/cDNA-protein hybrids that binds the target of interest. We hypothesize that binding to a target results in different rates of electrophoretic migration of the free (unbound) and target-bound mRNA/cDNA-protein hybrids.

Library screening using FSCE

Upon equilibration of an mRNA display library with the target of interest, some of the probes will exhibit affinity to the target and hence bind to the target. However, most of the probes in the library will not bind to the target. So, the critical step here is the separation of the target-bound probes from the unbound ones. Free Solution Capillary Electrophoresis (FSCE) – a process that separates species based on their charge to friction ratio – will be implemented to achieve this separation with high efficiency. FSCE refers to the migration of a charged species in free solution under the influence of an electric field in a capillary. In free solution, the migration of a species is based on its charge to friction ratio, rather than size only. Therefore, DNA or RNA of any size (in terms of base pairs) travels with the same electrophoretic mobility i.e. velocity inside the capillary *[31]*. However, once a label or a tag is attached to a nucleotide, the theory of End Labeled Free Solution Electrophoresis (ELFSE) predicts that the DNA which is longer in size will travel faster than a shorter one in the absence of the bulk flow or the so called EOF (Electro Osmotic Flow) *[31-34]*. Conversely, in the presence of EOF, labeled DNA/RNA which is longer

in size will travel slower than a shorter one. The labeling of ssDNA or dsDNA can be achieved by covalently attaching a protein/peptide *[34]*. The charge to friction ratio of nucleotide alters upon binding to a protein moiety termed as a "drag-tag". This drag causes an additional friction which slows down the mobility of the nucleotide moiety *[32]*. Addition of another protein to the nucleotide moiety that is already bound to a protein will further slow down the migration of the species. In the case of library screening using mRNA display, let us assume that the mRNA-cDNA-Sso7d hybrid molecule is the probe. The electrophoretic mobility of this probe relative to the mRNA-cDNA moiety can be related by to the following equation *[34]*:

$$\mu_{probe} = \mu_0 \frac{M - \beta_{probe}}{M + \alpha_{probe}} \qquad (4.1)$$

Where, μ_0 = electrophoretic mobility of the mRNA-cDNA portion; M = number of base pairs in the mRNA-cDNA portion; β_{probe} = effective charge of the Sso7d mutant of the probe expressed in terms of the charge of base pairs and α_{probe}= friction contributed by the Sso7d mutant and it is expressed in terms of the friction of one base pair of the mRNA-cDNA moiety. The charge of the sso7d mutant of the probe, β_{probe} is negligible compared to the highly negatively charged DNA moiety ($M >> \beta_{probe}$). This simplifies the numerator of the right hand side of equation 4.1 to M . However, once this probe is bound to the target molecule; the electrophoretic mobility of this newly formed 'complex' can be related to that of the probe according to the following simplification of equation 4.1,

$$\mu_{probe} = \mu_0 \frac{M}{M + \alpha_{probe}} \quad \text{and} \quad \mu_{complex} = \mu_0 \frac{M}{M + \alpha_{complex}} \text{ simplify to}$$

$$\mu_{complex} = \mu_{probe} \frac{M + \alpha_{probe}}{M + \alpha_{complex}} \qquad (4.2)$$

'Complex' here refers to the mRN-cDNA-Sso7d probe bound to any target molecule. The value of α_i can be determined using the following relationship as per reference *[34]*:

$$\alpha_i = 23\left(\frac{MW_i(kDa)}{52}\right)^{\frac{1}{3}} \qquad (4.3)$$

Here, i refers to the probe or the complex. For the probe, MW_i is the molecular weight of Sso7d while in the case of the complex, MW_i refers to the combined molecular weight of Sso7d and the target moelcule. 23 is the experimentally determined α value and 52 kDa is the molecular weight of streptavidin respectively. If we plug in the values of M and those of α_i in equation 4.2, the reduction in the electrophoretic mobility of the complex, compared to the probe is 4.3%. We hypothesize that this difference in the electrophoretic mobility is sufficient to separate the two species by FSCE.

As outlined in **Figure 4.2**, after putting the Sso7d library into mRNA display construct, the library will be incubated with a target and the equilibrium mix will be subjected to FSCE. In the presence of EOF, the complex will elute faster while the unbound ones will elute slower because of the difference in the electrophoretic mobility between the two species. The fraction that corresponds to the bound probes will be collected at the outlet of the capillary while the unbound

probes will be discarded. This fraction collected will then be subjected to error-prone PCR and a few more rounds of selection will be repeated. At the end of the final round, we will be left with the Sso7d mutants that have the highest affinity for the target molecule. These probes will be sequenced, cloned and expressed in *E. coli* for subsequent use as biomolecular probe. Indeed, FSCE has been employed in separating aptamers that bind a target from an aptamer-library *[35-41]*.

The selection process outlined here has several distinct advantages over common selection techniques. First of all, in developing the mRNA display library, no radio-active labeling is necessary which is the most common tag used for quantification *[2, 4, 8, 22, 23, 30]*. Secondly, non-specific binding is completely eliminated since the bound and unbound probes are separated on the basis of their difference in electrophoretic mobility. Third of all, no modification of the target (such as fluorescent labeling) is required. Also, CE is able to pull out even low affinity binders that are often lost in a wash step in common affinity maturation techniques based on immobilized targets.

Evidence of Successful *In Vitro* Translation

We have successfully implemented the mRNA display of a chitin binding domain (ChBD) from *Pyrococcus furiosus* (encoded in the ORF Pf1234). Briefly, the DNA coding for the ChBD flanked by a C-terminal 6x-His tag was *in vitro* transcribed to produce mRNA lacking a stop codon. The mRNA was then fused with a puromycin linker containing a poly-dA stretch using protocol described in *[42]*. The mRNA-linker species was subjected to *in vitro* translation in the presence of radioactive methionine and the reaction was purified using oligo-dT beads to obtain

mRNA-linker and mRNA-linker-ChBD. **Figure 4.3** shows radioactive counts from the wash steps and the elution fractions in the oligo-dT purification procedure. The eluate fractions are radioactive, indicating the presence of radio-labeled methionine and hence presence of the mRNA-linker-ChBD species.

Resolution of mRNA Display Products by ELFSE: Proof-of-concept

We have successfully demonstrated the proof of our theoretical analysis based on ELFSCE in the context of mRNA display. As described in the earlier section, the smaller ChBD was also put into mRNA display using non-radioactive material. An *in vitro* translation step was carried out, followed by oligo d(T) purification. The eluate from the oligo-dT purification was subjected to reverse transcription and further purified using a Ni-NTA beads. The flow through (which should contain mRNA-cDNA) and the eluate (mRNA-CDNA-ChBD) from the Ni-NTA bead purification were mixed and subjected to FSCE with UV detection. As predicted by the ELFSE theory, two distinct peaks were observed (**Figure 4.4**). This result validates our theoretical analysis and provides proof-of-concept that FSCE can be used in mRNA display to efficiently separate target-bound probes from the unbound probes.

Resolution of Bound and Unbound DNA Probes by ELFSE

DNA encoding the ribosomal protein L29 (207 bp) from *Thermotoga maritima* was amplified from the genomic DNA using biotinylated primer. DNA encoding Sso7d (from *Sulfolobus solfataricus*) was also PCR-amplified from a pET22b vector containing the coding sequence (216 bp long) but with ordinary oligos (not biotinylated). The two DNA samples were mixed to a

final concentration of 500 nM each in 1X Tris-Glycine (TG) buffer (Sample 1). Another sample containing the same DNA species (500 nM final concentration of each kind) was also prepared in 1X TG buffer. However, 5 uM streptavidin (SA) was added into the latter sample (Sample 2). The idea was to allow the formation of a complex between biotin and SA. These two species were chosen because of the very high affinity of SA for biotin ($K_d \approx 10^{-14}$ M) *[43]*.

All the CE experiments were performed in a ProteomeLab PA 800 CE instrument (Beckman Coulter). Each experiment was performed in triplicate. A bare fused-silica capillary that is 75 um in diameter and 60 cm long (50 cm to the detector) was used in all the experiments. The field strength was 500 V/cm, injection was at the positive electrode for 5 sec at 0.5 psi. Run buffer used was 1X TG. Since a fused-silica capillary was used, EOF (the bulk flow of ions) was present in all the experiments. Capillary temperature was maintained at 25^0 C. At first, Sample 1 (Sso7d and biotinylated L29) was injected showing the electropherogram depicted in **Figure 4.5** was observed. Next, sample 2 (DNA+SA) was injected in the capillary using the same CE conditions. The electropherogram observed in this case is shown in **Figure 4.6**.

As seen from **Figure 4.6**, introduction of SA to the DNA species causes two extra peaks to show up in comparison to **Figure 4.5**. There is still a peak at 300 sec which shows the presence of DNA that are still eluting at the same time as it did when no SA was present (**Figure 4.5** and **Figure 4.6** are overlaid in **Figure 4.7** to prove this point). However, right before this peak corresponding to free DNA, there is now another peak (**Figure 4.7**, DNA bound to SA) that is eluting earlier than free DNA and equal in height of the peak corresponding to free DNA. The

identity of this very peak (DNA bound to SA) is evident from two facts, one is that it is eluting earlier than free DNA, which is because of a greater electrophoretic mobility of this species (DNA bound to SA) arising from the presence of EOF. The second reason that confirms the identity of this species is the same height of this peak as that of free DNA. Because, the only DNA that could have been bound to streptavidin is the biotinylated L29 which was spiked at the same concentration (500 nM) as that of Sso7d. But since the heights of these two peaks are the same, it is evident that each peak corresponds to DNA of the same concentration.

References

[1] L. Xu, P. Aha, K. Gu, R. G. Kuimelis, M. Kurz, T. Lam, A. C. Lim, H. Liu, P. A. Lohse, L. Sun, S. Weng, R. W. Wagner, and D. Lipovsek, "Directed evolution of high-affinity antibody mimics using mRNA display," *Chem Biol,* vol. 9, pp. 933-42, Aug 2002.

[2] D. S. Wilson, 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,* vol. 98, pp. 3750-5, Mar 27 2001.

[3] C. A. Valencia, S. W. Cotten, B. Dong, and R. Liu, "mRNA-display-based selections for proteins with desired functions: a protease-substrate case study," *Biotechnol Prog,* vol. 24, pp. 561-9, May-Jun 2008.

[4] R. W. Roberts, "Totally in vitro protein selection using mRNA-protein fusions and ribosome display," *Curr Opin Chem Biol,* vol. 3, pp. 268-73, Jun 1999.

[5] S. E. Osborne and A. D. Ellington, "Nucleic Acid Selection and the Challenge of Combinatorial Chemistry," *Chem Rev,* vol. 97, pp. 349-370, Apr 1 1997.

[6] N. Nemoto, E. Miyamoto-Sato, Y. Husimi, and H. Yanagawa, "In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro," *FEBS Lett,* vol. 414, pp. 405-8, Sep 8 1997.

[7] R. Liu, J. E. Barrick, J. W. Szostak, and R. W. Roberts, "Optimized synthesis of RNA-protein fusions for in vitro protein selection," *Methods Enzymol,* vol. 318, pp. 268-93, 2000.

[8] A. D. Keefe and J. W. Szostak, "Functional proteins from a random-sequence library," *Nature,* vol. 410, pp. 715-8, Apr 5 2001.

[9] G. Cho, A. D. Keefe, R. Liu, D. S. Wilson, and J. W. Szostak, "Constructing high complexity synthetic libraries of long ORFs using in vitro selection," *J Mol Biol,* vol. 297, pp. 309-19, Mar 24 2000.

[10] J. E. Barrick, T. T. Takahashi, A. Balakin, and R. W. Roberts, "Selection of RNA-binding peptides using mRNA-peptide fusions," *Methods,* vol. 23, pp. 287-293, Mar 2001.

[11] E. T. Boder and K. D. Wittrup, "Yeast surface display for screening combinatorial polypeptide libraries," *Nat Biotechnol,* vol. 15, pp. 553-7, Jun 1997.

[12] J. S. Pedersen, D. E. Otzen, and P. Kristensen, "Directed evolution of barnase stability using proteolytic selection," *J Mol Biol,* vol. 323, pp. 115-23, Oct 11 2002.

[13] M. J. Feldhaus, R. W. Siegel, L. K. Opresko, J. R. Coleman, J. M. Feldhaus, Y. A. Yeung, J. R. Cochran, P. Heinzelman, D. Colby, J. Swers, C. Graff, H. S. Wiley, and K. D. Wittrup, "Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library," *Nat Biotechnol,* vol. 21, pp. 163-70, Feb 2003.

[14] D. W. Colby, B. A. Kellogg, C. P. Graff, Y. A. Yeung, J. S. Swers, and K. D. Wittrup, "Engineering antibody affinity by yeast surface display," *Methods Enzymol,* vol. 388, pp. 348-58, 2004.

[15] P. A. Alexander, D. A. Rozak, J. Orban, and P. N. Bryan, "Directed evolution of highly homologous proteins with different folds by phage display: implications for the protein folding code," *Biochemistry,* vol. 44, pp. 14045-54, Nov 1 2005.

[16] Y. Li, R. Moysey, P. E. Molloy, A. L. Vuidepot, T. Mahon, E. Baston, S. Dunn, N. Liddy, J. Jacob, B. K. Jakobsen, and J. M. Boulter, "Directed evolution of human T-cell receptors with picomolar affinities by phage display," *Nat Biotechnol,* vol. 23, pp. 349-54, Mar 2005.

[17] G. Chao, W. L. Lau, B. J. Hackel, S. L. Sazinsky, S. M. Lippow, and K. D. Wittrup, "Isolating and engineering human antibodies using yeast surface display," *Nat Protoc,* vol. 1, pp. 755-68, 2006.

[18] M. Friedman, A. Orlova, E. Johansson, T. L. Eriksson, I. Hoiden-Guthenberg, V. Tolmachev, F. Y. Nilsson, and S. Stahl, "Directed evolution to low nanomolar affinity of a tumor-targeting epidermal growth factor receptor-binding affibody molecule," *J Mol Biol,* vol. 376, pp. 1388-402, Mar 7 2008.

[19] R. W. Roberts and J. W. Szostak, "RNA-peptide fusions for the in vitro selection of peptides and proteins," *Proc Natl Acad Sci U S A,* vol. 94, pp. 12297-302, Nov 11 1997.

[20] S. R. Starck and R. W. Roberts, "Puromycin oligonucleotides reveal steric restrictions for ribosome entry and multiple modes of translation inhibition," *RNA,* vol. 8, pp. 890-903, Jul 2002.

[21] I. Fukuda, K. Kojoh, N. Tabata, N. Doi, H. Takashima, E. Miyamoto-Sato, and H. Yanagawa, "In vitro evolution of single-chain antibodies using mRNA display," *Nucleic Acids Res,* vol. 34, p. e127, 2006.

[22] P. W. Hammond, J. Alpin, C. E. Rise, M. Wright, and B. L. Kreider, "In vitro selection and characterization of Bcl-X(L)-binding proteins from a mix of tissue-specific mRNA display libraries," *J Biol Chem,* vol. 276, pp. 20898-906, Jun 15 2001.

[23] T. P. Cujec, P. F. Medeiros, P. Hammond, C. Rise, and B. L. Kreider, "Selection of v-abl tyrosine kinase substrate sequences from randomized peptide and cellular proteomic libraries using mRNA display," *Chem Biol,* vol. 9, pp. 253-64, Feb 2002.

[24] M. McPherson, Y. Yang, P. W. Hammond, and B. L. Kreider, "Drug receptor identification from multiple tissues using cellular-derived mRNA display libraries," *Chem Biol,* vol. 9, pp. 691-8, Jun 2002.

[25] K. Horisawa, S. Tateyama, M. Ishizaka, N. Matsumura, H. Takashima, E. Miyamoto-Sato, N. Doi, and H. Yanagawa, "In vitro selection of Jun-associated proteins using mRNA display," *Nucleic Acids Res,* vol. 32, p. e169, 2004.

[26] X. Shen, C. A. Valencia, J. W. Szostak, B. Dong, and R. Liu, "Scanning the human proteome for calmodulin-binding proteins," *Proc Natl Acad Sci U S A,* vol. 102, pp. 5969-74, Apr 26 2005.

[27] S. Tateyama, K. Horisawa, H. Takashima, E. Miyamoto-Sato, N. Doi, and H. Yanagawa, "Affinity selection of DNA-binding protein complexes using mRNA display," *Nucleic Acids Res,* vol. 34, p. e27, 2006.

[28] W. Ju, C. A. Valencia, H. Pang, Y. Ke, W. Gao, B. Dong, and R. Liu, "Proteome-wide identification of family member-specific natural substrate repertoire of caspases," *Proc Natl Acad Sci U S A,* vol. 104, pp. 14294-9, Sep 4 2007.

[29] A. Fischer, C. von Eiff, T. Kuczius, K. Omoe, G. Peters, and K. Becker, "A quantitative real-time immuno-PCR approach for detection of staphylococcal enterotoxins," *J Mol Med,* vol. 85, pp. 461-9, May 2007.

[30] B. C. Huang and R. Liu, "Comparison of mRNA-display-based selections using synthetic peptide and natural protein libraries," *Biochemistry,* vol. 46, pp. 10102-12, Sep 4 2007.

[31] C. Heller, G. W. Slater, P. Mayer, N. Dovichi, D. Pinto, J. L. Viovy, and G. Drouin, "Free-solution electrophoresis of DNA," *Journal of Chromatography A,* vol. 806, pp. 113-121, May 8 1998.

[32] R. J. Meagher, L. C. McCormick, R. D. Haynes, J. I. Won, J. S. Lin, G. W. Slater, and A. E. Barron, "Free-solution electrophoresis of DNA modified with drag-tags at both ends," *Electrophoresis,* vol. 27, pp. 1702-12, May 2006.

[33] R. J. Meagher, J. I. Won, J. A. Coyne, J. Lin, and A. E. Barron, "Sequencing of DNA by free-solution capillary electrophoresis using a genetically engineered protein polymer drag-tag," *Anal Chem,* vol. 80, pp. 2842-8, Apr 15 2008.

[34] R. J. Meagher, J. I. Won, L. C. McCormick, S. Nedelcu, M. M. Bertrand, J. L. Bertram, G. Drouin, A. E. Barron, and G. W. Slater, "End-labeled free-solution electrophoresis of DNA," *Electrophoresis,* vol. 26, pp. 331-350, Jan 2005.

[35] S. D. Mendonsa and M. T. Bowser, "In vitro selection of high-affinity DNA ligands for human IgE using capillary electrophoresis," *Anal Chem,* vol. 76, pp. 5387-92, Sep 15 2004.

[36] S. D. Mendonsa and M. T. Bowser, "In vitro evolution of functional DNA using capillary electrophoresis," *J Am Chem Soc,* vol. 126, pp. 20-1, Jan 14 2004.

[37] S. D. Mendonsa and M. T. Bowser, "In vitro selection of aptamers with affinity for neuropeptide Y using capillary electrophoresis," *J Am Chem Soc,* vol. 127, pp. 9382-3, Jul 6 2005.

[38] R. K. Mosing, S. D. Mendonsa, and M. T. Bowser, "Capillary electrophoresis-SELEX selection of aptamers with affinity for HIV-1 reverse transcriptase," *Anal Chem,* vol. 77, pp. 6107-12, Oct 1 2005.

[39] M. Berezovski and S. N. Krylov, "Nonequilibrium capillary electrophoresis of equilibrium mixtures--a single experiment reveals equilibrium and kinetic parameters of protein-DNA interactions," *J Am Chem Soc,* vol. 124, pp. 13674-5, Nov 20 2002.

[40] M. Berezovski, M. Musheev, A. Drabovich, and S. N. Krylov, "Non-SELEX selection of aptamers," *J Am Chem Soc,* vol. 128, pp. 1410-1, Feb 8 2006.

[41] M. Berezovski, R. Nutiu, Y. Li, and S. N. Krylov, "Affinity analysis of a protein-aptamer complex using nonequilibrium capillary electrophoresis of equilibrium mixtures," *Anal Chem,* vol. 75, pp. 1382-6, Mar 15 2003.

[42] M. Kurz, K. Gu, and P. A. Lohse, "Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA-protein fusions," *Nucleic Acids Res,* vol. 28, p. E83, Sep 15 2000.

[43] A. Holmberg, A. Blomstergren, O. Nord, M. Lukacs, J. Lundeberg, and M. Uhlen, "The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures," *Electrophoresis,* vol. 26, pp. 501-10, Feb 2005.

Figures

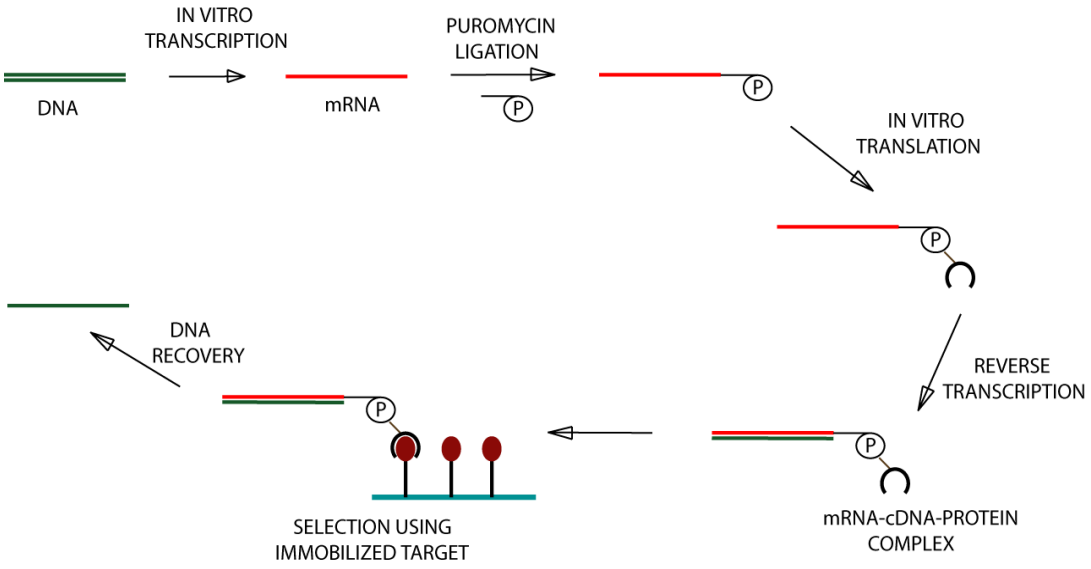


Figure 4.1: Selection of high affinity probes by mRNA display

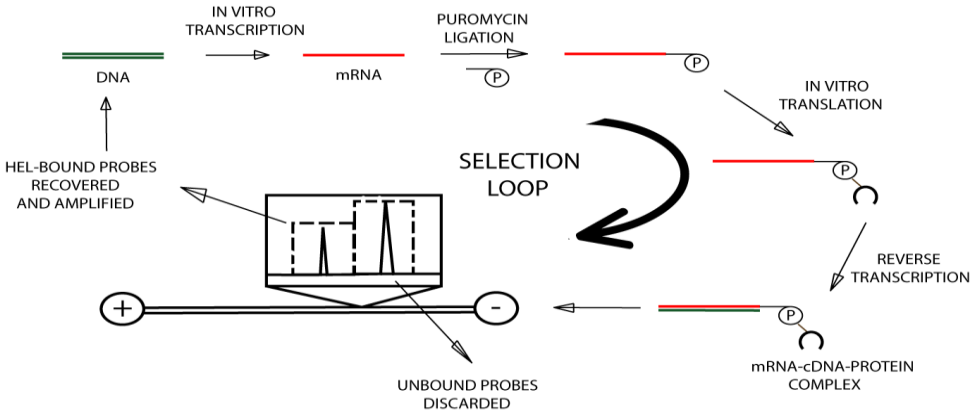


Figure4.2: Screening of mRNA display library by ELFSE (migration of species is from left to right)

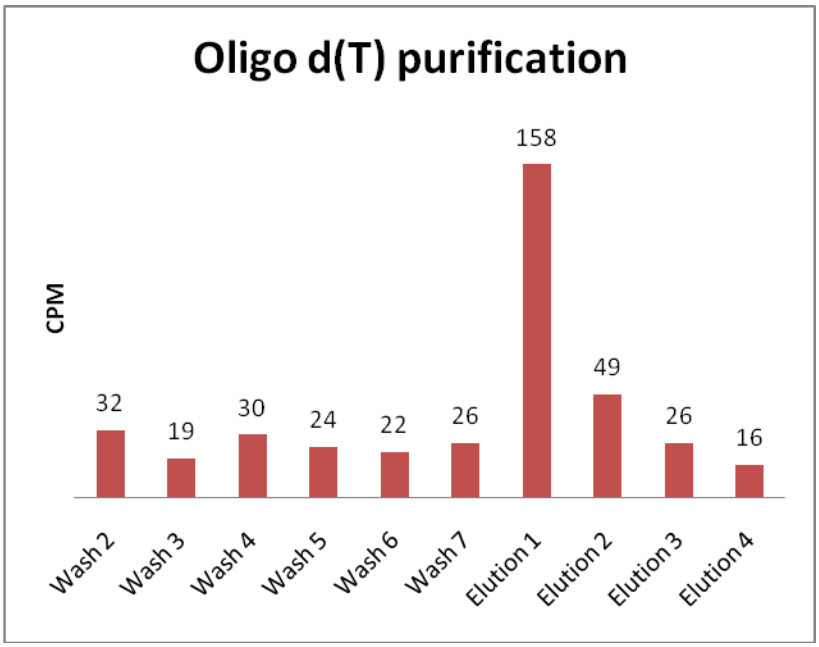


Figure 4.3: The Elution fraction from oligo (dT) purification showing radioactivity indicating the presence of mRNA-linker-ChBD fusion molecule.

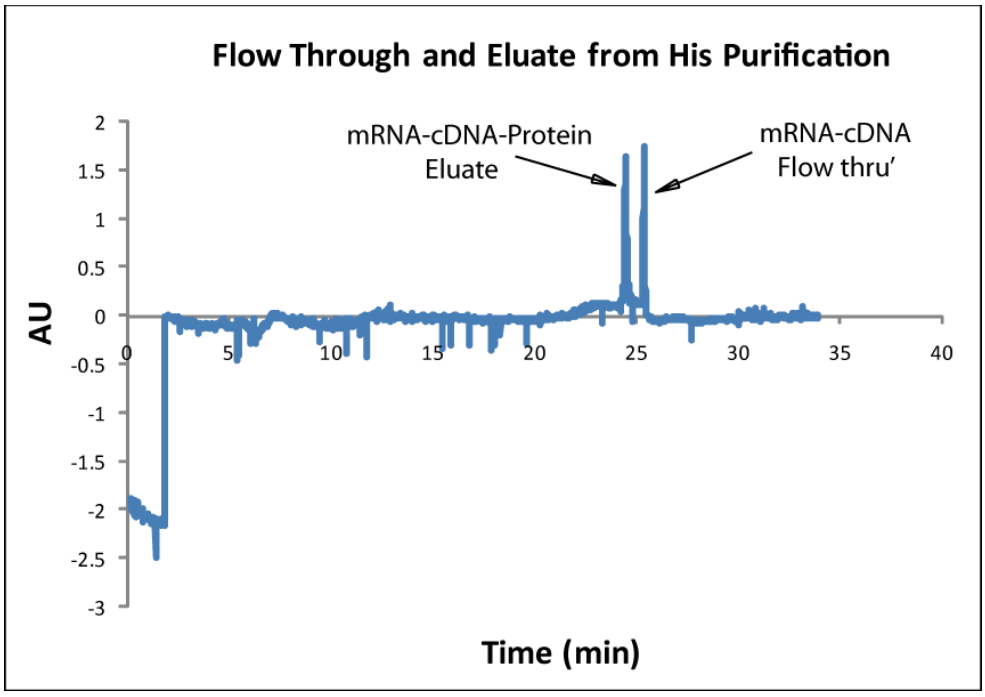


Figure 4.4: Separation of Flow through and Eluate of the ChBD mRNA display library

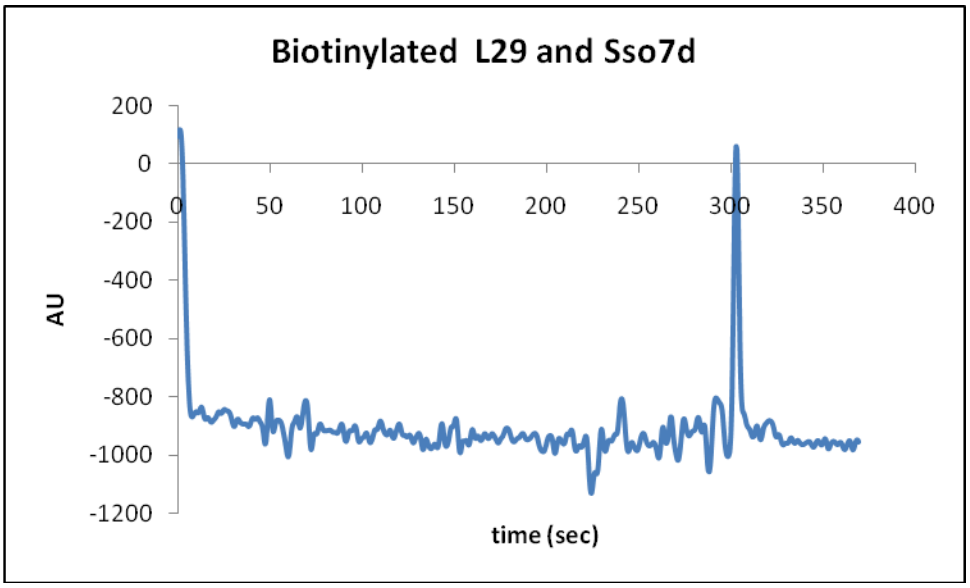


Figure 4.5: Electropherogram detecting DNA

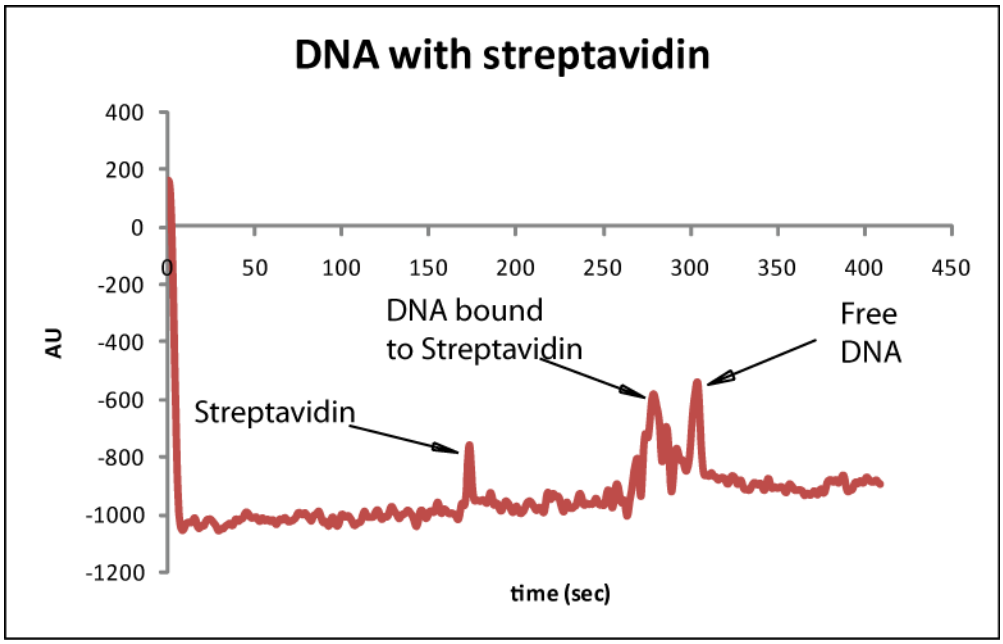


Figure 4.6: Electropherogram showing bound and free DNA

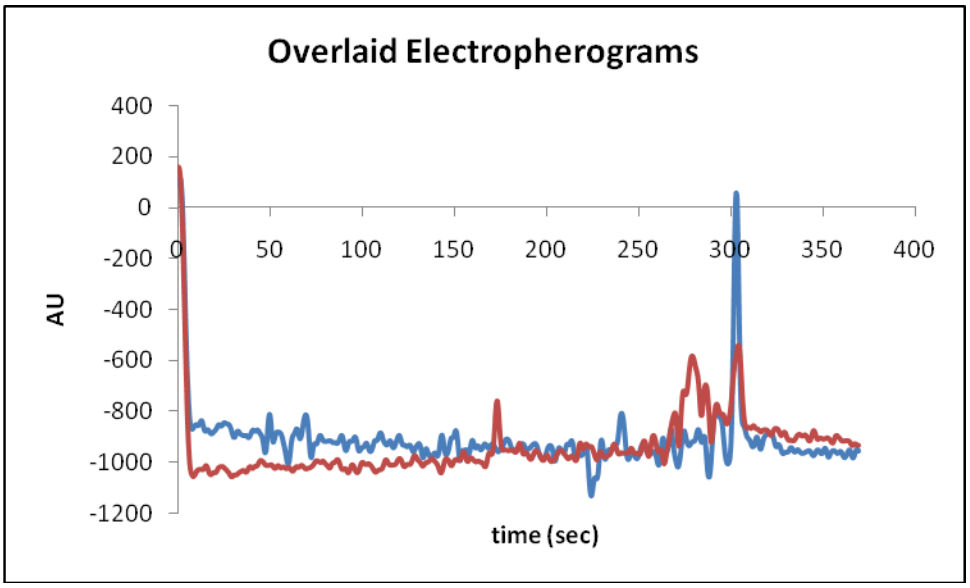


Figure 4.7: Fig. 4.8 and Fig. 4.9 are overlaid to demonstrate the same elution time of free DNA in either case.

Appendices

Appendix A: Theoretical analysis of ELFSE

Theoretical Analysis of ELFSE

We propose to use protein-DNA probes based on Sso7d for quantification of target species. As shown in equation B.2, the theory of ELFSE predicts the change in electrophoretic mobility of a protein-DNA probe upon binding a target as following:

$$\mu_{complex} = \mu_{probe} \frac{M + \alpha_{probe}}{M + \alpha_{complex}}$$

Where, μ = electrophoretic mobility; M = number of bases (or base pairs) in the nucleic acid portion; and α = friction contributed by the protein species expressed in terms of the friction of one base pair of DNA. Consider the binding of a probe *P* to its target *T* forming a complex, *C*. The equilibrium reaction is given by the following, where k_{on} and k_{off} are the rate constants of the association and dissociation reactions respectively.

$$T + P \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C$$

At equilibrium, the following relationship holds

$$K_d = \frac{C_{P,eq} C_{T,eq}}{C_{C,eq}} = \frac{k_{off}}{k_{on}}$$

Where, K_d is the equilibrium dissociation constant of the binding interaction and $X_{i,eq}$ (i = *P*, *T* or *C*) are the concentrations at equilibrium. The Krylov group introduced the concept of Non-Equilibrium Capillary Electrophoresis of Equilibrium Mixture (NECEEM) that can be used to theoretically determine the electrophoretic migration of a probe *P*, it's binding target *T* and the complex *C*, formed by *P* and *T [119, 128, 129]*. Briefly, the equilibrium mixture of *P* and *T* is prepared with *P* in large excess. The mixture contains three components, *P*, *T* and *C*. A plug of

the equilibrium mixture is introduced in the capillary by pressure and the electrophoretic separation of all three components take place inside the capillary. The assumption here is that the forward reaction (complex formation) is negligible compared to the reverse reaction, the dissociation of complex *C* will take place inside the capillary during NECEEM *[128]*. The following system of partial differential equations can describe the mass transfer of the three components with diffusion during NECEEM [128]:

$$\frac{\partial P(t,x)}{\partial t} + v_P \frac{\partial P(t,x)}{dx} - D_P \frac{\partial^2 P}{\partial x^2} = k_{off} C(t,x) \qquad \text{(A.1)}$$

$$\frac{\partial T(t,x)}{\partial t} + v_T \frac{\partial T(t,x)}{dx} - D_T \frac{\partial^2 T}{\partial x^2} = k_{off} C(t,x) \qquad \text{(A.2)}$$

$$\frac{\partial C(t,x)}{\partial t} + v_C \frac{\partial C(t,x)}{dx} - D_C \frac{\partial^2 C}{\partial x^2} = -k_{off} C(t,x) \qquad \text{(A.3)}$$

Solution to the above equations describe the concentrations as a function of time (*t*) and distance (*x*), where *x* is the distance beginning at the injection end and extends throughout the entire length of the capillary. Also, v_p, v_T and v_C are effective velocities of *P*, *T* and *C* respectively; D_P, D_T and D_C are diffusion coefficients of *P*, *T* and *C* respectively *[128]*. In developing the solution of equations B.5, B.6 and B.7, it has also been assumed that the effective velocities and diffusion coefficients are constant during NECEEM [128]. If the length of the plug injected into the capillary is *l*, the initial condition (at t=0) assumed for the three species here is as following:

$P(x)=P_0$, $T(x)= T_0$ $C(x)=C_0$ $x \in$ *[130]*

$P(x)=0$, $T(x)= 0$ $C(x)=0$ $x \notin$ *[130]*

Where, P_0, T_0 and C_0 are equilibrium concentrations of P,T and C respectively. Solution of equation B.7 is straightforward and can be given by the following eq. *[128]*:

$$C(t,x) = C_0 \frac{\exp(-tk_{off})}{2}\left(erf\left(\frac{1-x+tv_C}{\sqrt{4t\mu_C}}\right) - erf\left(\frac{tv_C - x}{\sqrt{4t\mu_C}}\right)\right) \qquad \text{(A.4)}$$

The solution to equations B.5 and B.6 are similar. However, in this case, there are two components of the solution namely,

$$P(t,x) = P_{eq}(t,x) + P_{dis} \qquad \text{(A.5)}$$

and

$$T(t,x) = T_{eq}(t,x) + T_{dis} \qquad \text{(A.6)}$$

The first terms in the right hand sides of equations B.9 and B.10 i.e. $P_{eq}(t,x)$ and $T_{eq}(t,x)$ describe the expression of the equilibrium fraction of the species whereas, the terms P_{dis} and T_{dis} arise from the dissociation of the complex, C into P and T during NECEEM. Complete solution of equations B.5 and B.6 are given below *[128]*:

$$P(t,x) = P_0\left(\frac{\exp(-tk_{off})}{2}\left(erf\left(\frac{1-x+tv_P}{\sqrt{4t\mu_P}}\right) - erf\left(\frac{tv_P - x}{\sqrt{4t\mu_P}}\right)\right) + \left(\frac{k_{off}\exp\left(\frac{(x-v_P t)k_{off}}{v_C - v_P}\right)}{v_C - v_P}\right)\right. \qquad \text{(A.7)}$$

$$\left.\varepsilon_P I\left(l/\varepsilon_P, \frac{(v_C t - x)}{\varepsilon_P}, \frac{(v_P t - x)}{\varepsilon_P}\right)\right)$$

$$T(t,x) = T_0\left(\frac{\exp(-tk_{off})}{2}\left(erf\left(\frac{1-x+tv_T}{\sqrt{4t\mu_T}}\right) - erf\left(\frac{tv_T - x}{\sqrt{4t\mu_T}}\right)\right) + \left(\frac{k_{off}\exp\left(\frac{(x-v_T t)k_{off}}{v_C - v_T}\right)}{v_C - v_T}\right)\right. \qquad \text{(A.8)}$$

$$\left.\varepsilon_T I\left(l/\varepsilon_T, \frac{(v_C t - x)}{\varepsilon_T}, \frac{(v_T t - x)}{\varepsilon_T}\right)\right)$$

Where, $\varepsilon_P = 2\sqrt{(D_C(x - v_P t) - D_P(x - v_C t)) / (v_C - v_P)}$ and

$$\varepsilon_T = 2\sqrt{(D_C(x - v_T t) - D_T(x - v_C t)) / (v_C - v_T)}$$

The above results can be summarized with the help of **Figure 4.A.1**. Let us consider the FSCE of probe, *P*; target, *T* and complex *C*. The equilibrated mixture is introduced as a plug at the inlet to the capillary and subjected to capillary electrophoresis. Upon application of an electric field, peaks corresponding to the unbound probe P and the target-probe complex C arise in the capillary, as shown in **Figure 4.A.1**. (The target protein *T* will typically have a significantly lower electrophoretic mobility relative to the probe *P* and complex *C*. We do not consider the target *T* for this analysis.) Let us assume that the detector is located right at the exit of the capillary. When probe, *P* is at the detector, *C* is yet to arrive at the detector. Between the peaks, *P* and *C*, there is a smear arising from the dissociation of *C* into *P* and *T*. If the contents of the capillary are collected, right after the exit of the 'P' peak, we will collect all the probes that are (i) bound to the target, and (ii) dissociated from the target while migrating through the capillary. **Figure 4.A.1 (b)** shows this collection window. Since the dissociated probes are collected and used for analysis, we can use low affinity probes for our analysis (at least up to K_D ~ 100 nM as per our preliminary analysis).

Figures

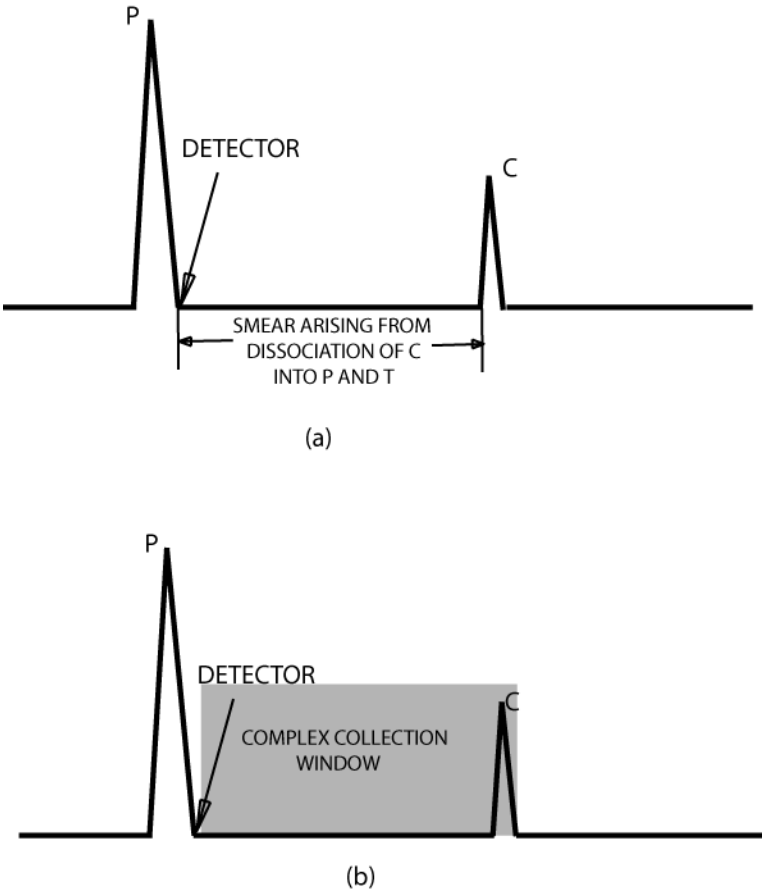


Figure 4.A.1: Schematic diagram of a simulation showing the separation of probe and complex (migration of species is from right to left)

Chapter 5

Conclusion

Molecular recognition is the cornerstone in eliciting numerous cellular responses. Antibodies have been the key molecule in molecular recognition. Highly specific antibodies for peptide, proteins, post-translational modification in proteins have been generated which underscores the flexibility of antibodies as binding molecules. Despite the plasticity of antibodies as binders, some of their attributes as discussed in chapter 1 limit their use in many applications. This has given rise to the alternate protein scaffolds for engineering molecular recognition. While as many as 50 non-immunoglobulin proteins have been reported for engineering molecular recognition [1, 2], there is still room for improvement in terms of melting temperature, resistance to chemical denaturation, maintaining structure under extreme pH conditions. Hyperthermophilic protein scaffolds are hence good candidates for engineering molecular recognition. We have explored the possibility of generating stable binding molecules from up to seven different scaffolds from four bacteria and archea. We have shown that binding molecules from five of these seven scaffolds can be engineered against fluorescein, C-terminus β-catenin peptide, a cyclic peptide BNP32, lysozyme, rabbit IgG and a plant virus nanoparticle without any affinity maturation. Yeast surface display and mRNA display in conjunction with Fluorescence Activated Cell Sorting (FACS) has been employed for generating these binding molecules. The apparent dissociation constant (K_D) ranges from hundreds of nanomolar to micromolar. In addition, the melting temperature of the binding proteins are > 70^o C. These binding proteins are also easy to produce in common bacterial expression system in high titer. In addition, we tested MIT, Sso7d and Sso6901 mutant for their maintenance of secondary structure by circular dichroism. We have also demonstrated the use of an Sso7d mutant for affinity purification of RCNMV, a plant virus nanoparticle that has been previously characterized for packaging cancer drug in the viral capsid

protein [3, 4]. In addition, we have shown proof-of-concept for biomolecular separation using Free Solution Capillary Electrophoresis (FSCE). Taken together, we have generated stable hyperthermophilic binding molecules for macromolecular purification.

High throughput screening process was used to isolate the binding molecules from yeast surface display. The theoretical diversity obtained by randomizing 12 residues on a scaffold is ~ 10^{15} whereas even the biggest combinatorial libraries typically have only 10^{12} scaffold mutants. Thus any combinatorial screening strategy samples a very small fraction of the theoretical library diversity. As a result, in many cases, it is possible that a high-affinity binder may not be present in the library. On the other hand, our results show that low affinity binders (at the minimum) can be reliably isolated from our "super-library" using magnetic sorting and yeast surface display. Hence, "*intramolecular avidity*" can be taken advantage of to construct high-affinity binders from the low affinity binding proteins isolated through magnetic sorting. To elaborate, Consider two proteins (A and B) that bind a target at non-overlapping regions separated by a distance d_o (**Figure 5.1**). Let $K_{D,A}$ and $K_{D,B}$ be the equilibrium dissociation constants of binding of A and B to the target respectively. If A and B are tethered to each other through a flexible linker, 'L' amino acids in length, then the resulting "tandem-binder" A-B has a much lower equilibrium dissociation constant ($K_{D,A,B}$) than $K_{D,A}$ or $K_{D,B}$ (i.e., much higher binding strength than A alone or B alone). The relationship between $K_{D,A,B}$, $K_{D,A}$ and $K_{D,B}$ can be obtained from thermodynamics as:

$$K_{D,A,B} = \frac{K_{D,A}K_{D,B}}{[A]_{eff}}$$

where $[A]_{eff}$ is the effective concentration of A close to its binding site when B is bound to the target. $[A]_{eff}$ can be calculated as [5],

$$[A]_{eff}(M) = 5.03L^{-3/2}e^{-6.58\times10^{-2}\,d_0^2/L}\left\{\begin{array}{l}1-\dfrac{0.987}{L}+0.139\dfrac{d_0^2}{L^2}-2.51\times10^{-3}\dfrac{d_0^4}{L^3}-\dfrac{0.308}{L^2}\\ -0.150\dfrac{d_0^2}{L^3}+0.0204\dfrac{d_0^4}{L^4}-5.17\times10^{-4}\dfrac{d_0^6}{L^5}+3.14\times10^{-6}\dfrac{d_0^8}{L^6}\end{array}\right\}$$

In this equation, d_0 is expressed in A° and 'L' is number of amino acids in the linker.

At the end of the magnetic selection step described in chapters 2 and 3, all proteins that bind a given target (even with low affinities) can be isolated. These proteins are likely derived from various starting scaffolds and potentially bind several different epitopes on the target. One an create a comprehensive yeast display library of "*tandem-binders*" by linking proteins pair-wise through flexible linkers of variable length (5, 10 or 15 amino acids). Thus if 10^3 proteins are isolated as was shown in chapter 2, the resultant tandem-binder library will be $3x10^6$ in size. Tandem-binders consisting of proteins that bind non-overlapping regions on the target will have a significantly higher binding affinity for the target and can easily be isolated using flow cytometry.

Recent advances in computational design of proteins for altered specificity, affinity, enzymatic activity, stability or even *de novo* protein design is a promising alternative or at least complementary to the combinatorial library screening [6]. An extensive review on the progress in computational design of proteins can be found here [6]. There have been tremendous improvements in recent times in developing algorithms and speeding up the computational

expenses required for protein design. A 4-fold speedup was attained by removing atom-atom calculations in a trie data structure designed by Lever-Fay, Kuhlman and Snoeyink [7]. Very recently, Baker and co-workers have computationally designed proteins against a conserved stem region of influenza hemagglutinin [8]. Crystal structure of HB36 in complex with the 1918/H1 HA showed that the binding interface is nearly identical to that in the computational model. Schief, Baker and co-workers have also shown computation-guided grafting of backbone and side chains in an unrelated protein scaffold to obtain as high affinity and specificity as the neutralizing antibody b12 against HIV gp120 epitope [9]. These findings emphasize the power and necessity of computational approaches to protein design. In near future, it is likely that all protein engineering problems will be steered by some extent of computational design.

References

[1] A. Skerra, "Alternative non-antibody scaffolds for molecular recognition," *Curr Opin Biotechnol,* vol. 18, pp. 295-304, Aug 2007.

[2] H. K. Binz, P. Amstutz, and A. Pluckthun, "Engineering novel binding proteins from nonimmunoglobulin domains," *Nat Biotechnol,* vol. 23, pp. 1257-68, Oct 2005.

[3] D. M. Lockney, R. N. Guenther, L. Loo, W. Overton, R. Antonelli, J. Clark, M. Hu, C. Luft, S. A. Lommel, and S. Franzen, "The Red clover necrotic mosaic virus capsid as a multifunctional cell targeting plant viral nanoparticle," *Bioconjug Chem,* vol. 22, pp. 67-73, Jan 19 2011.

[4] D. Lockney, S. Franzen, and S. Lommel, "Viruses as nanomaterials for drug delivery," *Methods Mol Biol,* vol. 726, pp. 207-21, 2011.

[5] H. X. Zhou, "Quantitative account of the enhanced affinity of two linked scFvs specific for different epitopes on the same antigen," *J Mol Biol,* vol. 329, pp. 1-8, May 23 2003.

[6] S. M. Lippow and B. Tidor, "Progress in computational protein design," *Curr Opin Biotechnol,* vol. 18, pp. 305-11, Aug 2007.

[7] A. Leaver-Fay, B. Kuhlman, and J. Snoeyink, "Rotamer-pair energy calculations using a trie data structure," *Lecture notes in computer science*, vol. 3692, pp. 389-400, 2005.

[8] S. J. Fleishman, T. A. Whitehead, D. C. Ekiert, C. Dreyfus, J. E. Corn, E. M. Strauch, I. A. Wilson, and D. Baker, "Computational design of proteins targeting the conserved stem region of influenza hemagglutinin," *Science,* vol. 332, pp. 816-21, May 13 2011.

[9] M. L. Azoitei, B. E. Correia, Y. E. Ban, C. Carrico, O. Kalyuzhniy, L. Chen, A. Schroeter, P. S. Huang, J. S. McLellan, P. D. Kwong, D. Baker, R. K. Strong, and W. R. Schief, "Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold," *Science,* vol. 334, pp. 373-6, Oct 21 2011.

Figures

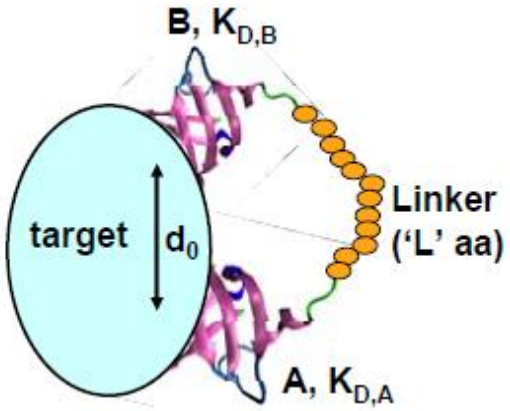


Figure 5.1: A tandem binder obtained by linking two proteins binding distinct regions on the target