ABSTRACT

CARLIN, KEVIN BRIAN. Engineering Multivalent Protein Affinity Ligands using the Sso7d Scaffold. (Under the direction of Balaji M. Rao).

Protein affinity ligands have applications in diagnostics, therapeutics and as research tools for probing biological systems. By specifically binding a given target a ligand is useful for detecting that target, blocking a given interaction or useful as non-perturbing biosensors which can monitor the state of a protein in a given system. There are established methods for generating affinity ligands which are continually being refined and improved through various means. However, systematically generating bivalency or multivalency for generating high affinity binding proteins is a potentially very powerful but underutilized approach.

We show that a combinatorial library constructed by random pairwise assembly of low affinity binders can efficiently generate binders with increased affinity. Such a library, from a pool of low affinity binders based on the Sso7d scaffold, contained putative high affinity clones for a model target (lysozyme) at higher frequency than a library of monovalent mutants with equivalent diversity (~ 10^7). Increased binding affinity was due to intramolecular avidity generated by linking binders targeting non-overlapping epitopes; individual binders of K_D ~ 1.3 μM and 250 nM produced a bivalent binder with K_D ~ 1.7 nM. Furthermore, the bivalent protein retained thermal stability (T_M= 84.5 °C) and high recombinant expression yields in *E. coli*. Finally, when binders comprising the bivalent protein are fused to two of the three fragments of tripartite split-green fluorescent protein (GFP), target-dependent reconstitution of fluorescence occurs, thereby enabling a "mix-and-read" assay for target quantification.

A separate study evaluated the ability of monovalent and multivalent ligands to bind the inherently disordered C-terminus of β-catenin. β-catenin is a human protein with orthologues in all vertebrates. Its most well-known roles are to functions as a component of cell to cell adhesion and as

key to transcription initiated by WNT signaling. A number of findings relevant to the Sso7d scaffold, yeast surface display and targeting inherently disordered domains are made. A self-assembling, disulfide bonded, homodimeric version of Sso7d, referred to as M1, is identified. M1 was originally selected to bind a peptide representing the C-terminal epitope of β-catenin and subsequently affinity matured for binding to the β-catenin protein. Multiple modes of affinity maturation were identified. The dimeric ligand gained affinity to both the originally targeted C-terminal epitope as well as to a second epitope on β-catenin. Additionally, both epitopes appear to be engaged simultaneously as the apparent K_D to either epitope is an order of magnitude higher than for the ligand to β-catenin. Multimeric Sso7d constructs were generated by linking the M1 dimerizing ligand to a library of Sso7d mutants. The library was selected for binding to β-catenin and a high affinity tetravalent ligand was identified. Affinity maturation of the selected Sso7d tetrameric ligand to β-catenin resulted in unique mutations on M1 which also increased the affinity of the tetramer to β-catenin through epitope addition.

Engineering Multivalent Protein Affinity Ligands using the Sso7d Scaffold

by
Kevin Brian Carlin

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

2016

APPROVED BY:

_____	_____
Dr. Balaji M. Rao Committee Chair	Dr. Jason M. Haugh
_____	_____
Dr. Gavin Williams	Dr. Gregory T. Reeves

BIOGRAPHY

I went to the University of Washington (UW) for a degree in Chemical Engineering. This motivation was driven by a chemistry teacher I had during my freshman year at the University of Portland (UP). I enjoyed her class and she took time to show talk about career options one could have with a chemistry or chemical engineering background. As UP did not offer a degree in Chemical Engineering course I transferred to (UW). I enjoyed my time there immensely until getting into the Chemical Engineering Department. Sure that graduate school was not a good option for me at that time I worked for Dow Chemical for five years. I found I did not enjoy following global work practices. The next three years were spent at a biofuel startup. I enjoyed being closer to the business factors influencing development timelines. I also enjoyed making major contributions to process development. I decided to pursue a PhD at NCSU to continue in the technology startup space afterwards. A protein engineering based PhD was a choice made so that I would be able to go into the therapeutics field.

ACKNOWLEDGMENTS

I would like to acknowledge Bala's efforts in securing the funding to make this research possible. Also, for all of his continued collaboration in publishing the key pieces of this thesis. I would also acknowledge Karthik Tiruthani and Carlos Cruz for their thoughtful contributions to my research throughout my time here. Their love of problem solving and open discussion were invaluable for the research and made my time at NCSU more enjoyable. Thanks to Prasenjit Sarkar for giving his input in group meetings and ad hoc. Thanks to Adam Mischler for bringing a high level of discipline and a strong voice to lab practices and cleanliness, also for his comments on making this work more accessible to a general audience. I am very grateful to the undergraduates I worked with; they were a great help in completing these projects. Special thanks to Apoorva Thavarty, Cathy Gomes, Kim Yejin, Ryan Vest, Kyle Goodman and Mark Dyson for their contributions. I hope their time in the lab and working with me was helpful in their decisions on where to take their careers. Lastly, thanks to Nimish Gera and Mahmud Hussain for teaching me the ropes of protein engineering.

TABLE OF CONTENTS

LIST OF FIGURES **8**

LIST OF TABLES **11**

BACKGROUND AND MOTIVATION FOR ENGINEERING AFFINITY LIGANDS **1**

1.1 Introduction 2

1.1.1 Protein Engineering 5

1.1.2 Engineering High Affinity ligands 8

1.1.3 Engineering Multivalent Affinity Ligands 10

1.1.4 Bivalent ligands as tools for manipulating biological phenomenona 15

1.2 Thesis Overview 16

1.3 Figures 18

COMBINATORIAL PAIRWISE ASSEMBLY EFFICIENTLY GENERATES HIGH AFFINITY BINDERS AND ENABLES A "MIX-AND-READ" DETECTION SCHEME. **23**

2.1 Introduction 24

2.2 Results 26

2.2.1 A bivalent library yields a higher frequency of putative high affinity clones. 26

2.2.2 Magnetic sorting and FACS identifies a pool of bivalent lysozyme binders with highest affinity. 28

2.2.3 Determination of intermolecular vs intramolecular binding of yeast displayed BVL-1; recombinant expression, stability and binding of lysozyme. 29

2.2.4 Binding of NTL-1 and CTL-1 to non-overlapping epitopes on the target can be exploited to design a “mix-and-read” assay for target quantification. 31

2.3 Materials and Methods 32

2.3.1 Library construction. 32

2.3.2 Comparison of monovalent and bivalent libraries. 33

2.3.3 Selection of the highest affinity bivalent ligands 34

2.3.4 Simultaneous Yeast Surface Display of CTL-1 and NTL-1 (Dual Display) 35

2.3.5 Recombinant expression of BVL-1. 35

2.3.6 Estimation of KD 36

2.3.7 Size Exclusion chromatography 37

2.3.8 Measurement of the Melting Temperature T_M. 37

2.3.9 Expression and purification of GFP1-9, NTL-GFP11 and GFP10-CTL-1. 37

2.3.10 *In vitro* “mix and read” fluorescence reconstitution assays. 38

2.4 Figures 39

FUNCTIONAL DIMERIC SSO7D LIGANDS AND EPITOPE ADDITION AS A MODE OF AFFINITY MATURATION IN BINDING INTRINSICALLY DISORDERED DOMAINS 48

3.1 Introduction 49

3.2 Results 53

3.2.1 L54W is a conserved mutation in the highest affinity ligands selected from the affinity matured monovalent library pre-targeted to the β-catenin C-terminal IDR. 53

3.2.2 The first generation tandem library assembled well and yielded high affinity ligands 54

3.2.3 A diversified polypeptide chain was enriched during selections for β-catenin binding. 55

3.2.4 Affinity maturation of T2 yields a diversity of advantageous mutations. 57

3.2.5 Selection of functional dimeric ligands 58

3.2.6 Measurement of soluble phase K_{DS} 62

3.2.7 Differential binding observed for ligands to βCP peptide sequence alone and in the context of the C-terminal IDR 64

3.2.8 ARM Domain Binding 67

3.2.9 No binding to β-catenin N-terminal IDR 68

3.2.10 Dimerization of T2 and T3 is required for 23CTL functionality. 69

3.3 Discussion 70

3.4 Conclusions 78

3.5 Materials and Methods 79

3.5.1 β-catenin Target Expression 79

3.5.2 β-catenin constructs cloning 80

3.5.3 GST-TCF Expression 80

3.5.4 Generation of the peptide pre-targeted library and selection of the M2 ligand. 80

3.5.5 Generation of the Tandem Library 82

3.5.6 Selection of Best Binders in Tandem Library 83

3.5.7 Assembly of the second generation tandem library 83

3.5.8 Ligand cloning into bacterial expression vectors 85

3.5.9 Blitz Experiments 86

3.5.10 K_D Measurements 87

3.6 Figures and Tables 88

CONCLUSIONS AND FUTURE WORK 118

4.1 Figures and Tables 127

Appendices 132

2. Appendix A1 132

3. Appendix A2 135

4. Appendix A3 144

References 146

LIST OF FIGURES

Figure 1.1: Flowchart of protein engineering by yeast surface display. 18

Figure 1.2: Chelating recombinant ligands generate bivalency regardless of target context 21

Figure 1.3: Estimated enhancement for linking ligands 21

Figure 2.1: A bivalent library yields a higher frequency of putative high affinity clones. 39

Figure 2.2: Analysis of nine clones from the highest affinity lysozyme ligands selected. 40

Figure 2.3: High affinity of bivalent binder BVL-1 is due to synergistic binding of low affinity NTL-1 and CTL-1 subunits. 41

Figure 2.4: Size exclusion chromatograms indicating BVL-1 is monomeric. 42

Figure 2.5: SDS page of elution samples of the BVL-1 and lysozyme complex from SEC 44

Figure 2.6: Differential scanning fluorimetry for estimation of T_M. 44

Figure 2.7: Binding of NTL-1 and CTL-1 to non-overlapping epitopes on the target can be exploited to design a “mix-and-read” assay for target quantification. 45

Figure 2.8: Comparison of monovalent and bivalent libraries by cytometry. 47

Figure 3.1: M1 and M2 differ by an L54W mutation. 88

Figure 3.2: Generation of a tandem library with M1.2 as the N-terminal ligand. 90

Figure 3.3: Generation of the T2 tandem library 91

Figure 3.4: PDB diagram of mutations observed in T3-NTL 93

Figure 3.5: SDS page gels indicating M2 and M1 exist predominantly as disulfide bonded homodimers 94

Figure 3.6: SDS page gels demonstrating T2, T3 and T3-NTL also exist as disulfide bonded dimers. 95

Figure 3.7: Densitometry analysis to determine the percentage of dimeric protein in each stock. 96

Figure 3.8: Biolayer interferometry data shows the dimeric form of the ligand is functional. 98

Figure 3.9: Calculated dissociation rate k_{off} and K_D for reduced (1 mM TCEP) and non-reduced M1 based on data recorded on the Blitz™. 99

Figure 3.10: Titration and competition based K_D curves for yeast displayed ligands to full length soluble β-catenin 101

Figure 3.11: Multimodal binding observed in yeast based K_D measuremnts. 102

Figure 3.12: K_D curves for yeast displayed ligands to soluble βCP. 105

Figure 3.13: Cytometry histograms for ligands binding the C-terminal IDR. 107

Figure 3.14: K_D curves for soluble M2 and T3 to yeast displayed C-terminal IDR (666-781). 108

Figure 3.15: Yeast surface display of the ARM domain, expression controls and binding by ligands. 109

Figure 3.16: Titration based K_D's curves for soluble ligands binding the yeast displayed ARM domain. 112

Figure 3.17: Expression level of N-terminal IDR's on yeast and binding by soluble ligands. 112

Figure 3.18: 23CTL requires avidity 113

Figure 3.19: Diagrams depicting how monomeric or dimeric M2 may interact with β-catenin immobilized on a Blitz sensor. 114

Figure 3.20: Proposed binding mechanisms of the ligands engineered in this work to β-catenin 116

Figure 4.1: A sensor for ligand dimerization in the yeast surface display system. 127

Figure 4.2: An alternative co-localization sensor using renilla GFP and PLCγ1-SH2. 129

Figure 4.3: A FRET based sort for heterodimeric self-associating ligands. 131

Figure A2.1 Biolayer interferometry data recorded for non-reduced M1. 136

Figure A2.2 Analysis of biolayer interferometry data recorded for non-reduced M1 137

Figure A2.3 Biolayer interferometry data for reduced M1..... 138

Figure A2.4 Analysis of biolayer interferometry data for reduced M1. 140

Figure A2.5 Biolayer interferometry data and analysis for non-reduced M2..... 141

Figure A2.6 biolayer interferometry data and analysis for reduced M2..... 142

Figure A2.7 Biolayer interferometry data and analysis for reduced and non-reduced T2. 143

LIST OF TABLES

Table 2.1: Primers used for library construction. 47

Table 3.1: Poly-peptide chain enriched in linker region of tandem library 89

Table 3.2 The highest affinity bivalent ligands selected for tend to have histidine's incorporated in the linker 90

Table 3.3: A diversity of mutations are seen in the highest affinity ligands selected from the T2 mutational library. 92

Table 3.4: Nomenclature of β-catenin affinity ligands 103

Table 3.5: Ligand K_D's and properties. 104

CHAPTER 1

BACKGROUND AND MOTIVATION FOR ENGINEERING AFFINITY LIGANDS

1.1 Introduction

Engineering protein affinity ligands is a discipline of generating affinity to a target protein (P) by another protein, referred to as a ligand (L), forming a complex (C). This is written as a reversible chemical reaction

$$P + L \leftrightharpoons C$$

First order kinetics are observed for both the forward reaction, which is referred to as the association or on rate and the reverse reaction, referred to as the dissociation or off rate. By convention the constant K_D defines this interaction:

$$\frac{[P]_{eq}[L]_{eq}}{[C]_{eq}} = \frac{k_{off}}{k_{on}} \equiv K_D$$

K_D is a useful term because it is the concentration at which binding is relevant for a given system. For instance, many diagnostic applications immobilize a ligand on a surface and then measure the binding of a target protein to that ligand. There are a great variety of methods through which this binding can be measured, e.g. surface plasma resonance (SPR), interferometry, sandwich enzyme-linked immunosorbent assay (ELISA), as well as many non-commercialized techniques which have been developed in published studies[1–3]. For each of these methods the fractional binding of the target (P_o) onto the surface-loaded ligand can be estimated in cases where the amount bound is small relative to the amount in solution, by the following equation:

$$Fraction\ bound = \frac{P_o}{(K_D + P_o)}$$

If K_D is much greater then P_o the fraction of the total possible amount of target that can be bound by the ligand is low. One method of increasing the signal is to engineer a K_D to be in the range of or lower than the concentration of the protein target of interest in the sample mixture. Decreasing detection limits is therefore one motivation for desiring a low K_D. An alternative motivation to engineer high affinity is in the context of binding a target of interest in a specific manner such that it is not readily displaced. Such ligands can act as inhibitors of protein-protein interactions and used as tools to study biological processes[4].

The specific binding of a ligand to a target is referred to as molecular recognition; it is characterized by low values of K_D for the target-specific interaction, relative to other interactions with non-target species. In case of protein-protein interactions, molecular recognition is mediated by interactions between amino acid residues of the ligand and target at the binding interface, including electrostatic, hydrophobic or Van der Waals interactions. In a survey of 15 different protease inhibitors and antibody complexes it was found that the binding interface consisted of 34 ±7 amino acids on the ligand[5]. Typically specific, high affinity interactions require many residues on the ligand to be placed in a very exact manner relative to the target in order to generate recognition.

Molecular recognition is often generated through mutagenesis of scaffold proteins. A scaffold is the protein from which the ligand is generated; it is the structure upon which affinity is engineered. Scaffolds are typically stable proteins which can accommodate a high degree of surface mutation and or insertion or deletion of amino acids. Therapeutically and diagnostically the dominant protein engineering scaffold has been the antibody. However, production

challenges continue to be a factor for the antibody because it is a large (150 kDa) multi-domain protein, requiring disulfide bond formation[6]. This results in the necessity of a eukaryotic expression system which remains a laborious and costly process[7]. Therapeutically, antibody size can also be a factor limiting its potency as there are tissue types which the antibody does not access well[7]. Outside of the brain and central nervous system this can be overcome with higher antibody dosages but this comes with increased cost and toxicity[8]. Therefore, smaller scaffolds based therapeutics may offer an advantage in these areas. The reduced in-vivo half-life can be an issue for alternative scaffolds, however despite this an alternative scaffold therapeutic with a 2 hour half-life has been approved for therapeutic use[8]. Also, for molecular imaging of cancers short half-life is a desired property as it lowers the background level of reagent detected[9]. For ex-vivo diagnostic applications the IgG format is unnecessary therefore it would be favored to have a reagent that can more easily engineered and produced. For these reasons and others the generation of affinity ligands from non-antibody scaffolds has been a major research area of interest[6,8–11]. Commonly used examples of non-antibody scaffolds are the 10th domain of fibronectin, designed ankyrin repeat proteins and affibodies[12]. For each of these scaffolds established libraries and mutational strategies have been developed which enable reliable generation of binding proteins to targets of interest. Ligands generated from scaffolds are characterized by having favorable properties such as low molecular mass, high stability, a lack of necessary disulfide bonds or a lack of disulfide bonds that do not form readily after recombinant expression in *Escherichia coli* and ease of recombinant expression in *Escherichia coli*.

Recent publications have been directed at engineering affinity ligands using the Sso7d protein as a scaffold[13–15]. Sso7d is a DNA binding protein from Sulfolobus solfataricus. It consists of an incomplete beta-barrel with five beta strands and a C-terminal alpha helix[16]. The protein has 63 residues (~7 kDa), has a very high melting temperature (100°C), does not require disulfide bond formation and expresses well in *E. coli*[16]. Sso7d is a highly stable form of an SH3 domain, a natural peptide binding domain found in nature[15–17]. Other homologous of Sso7d bind metal ions, oligosaccharides, nucleic acids and other proteins[15]. Therefore, efforts have been made to establish techniques and methods for selecting high affinity ligands using Sso7d or the highly homologous Sac7d domain. Initial studies focused on selecting affinity ligands from a library comprised of mutation of surface exposed residues on the 3 β-sheet face of the scaffold[13,15]. Moderate to high affinity ligands were selected from the naïve library with little to no loss of thermostability or stability losses due to pH environment or denaturants[13]. This was observed for each of six ligands studied in detail[13]. Alternative libraries have been designed for Sac7d and were successfully used to generate affinity ligands[18]. This thesis builds on these past studies with Sso7d. Techniques to preserve scaffold stability and generate affinity more expediently using combinatorial methods to generate bivalency are explored. A study is also done to bind an intrinsically disordered domain. Observations are made as to how these regions can be pre-targeted and to study mechanisms of affinity maturation for these targets.

1.1.1 Protein Engineering

Protein libraries for generation of affinity ligands are created at the DNA level. Targeted positions on the scaffold are selected based on generating a paratope, or binding face

in the region of those mutations. This DNA is inserted into some type of display cassette which strongly links the DNA to the translated protein. Five display methods are predominately used in the literature for selection of affinity ligands: phage, yeast, bacterial, ribosome or the very similar mRNA display system[19]. The first three methods are cell based, they incorporate the DNA for one library member into one cell, and that one library mutant protein is displayed on the cell surface. The library is comprised of a large collection of mutant expressing cells 10^8 to 10^{11} library members are possible depending on the display system[19]. Ribosome display and RNA display are cell-free systems; for each method RNA is transcribed and directly linked to the translated protein. To accomplish this translation machinery from prokaryotic or eukaryotic cells is extracted and used with DNA coding for the library to create the translated protein library[20]. Diversities of 10^{12} are possible for cell-free systems[20].

Each display system has advantages and disadvantages which have been reviewed [20–22]. The work described in this dissertation was carried out entirely using the yeast display system. Significant reasons for this choice are: protein expression level and binding can be directly measured and accounted for quantitatively throughout the selection and affinity maturation process; affinity measurements can be directly made for clones of interest using the yeast displayed ligand; yeast does not suffer from the amplification bias that phage does, this results in collapse of the library because some mutants can be parasitic[23]. This toxicity is a result of the phage system's inability to separate the expansion of the library from expression of the library mutant. Whereas, the yeast display system enables distinct modes of expansion of the library and induction of the ligand. Finally, we chose the yeast display method because

the library generation steps for RNA display are more cumbersome. This is an important factor as for the RNA system the library has to be re-created after each selection; whereas, for the yeast based system selected clones are simply amplified in selective media at yeasts natural growth rate.

There are variations of the yeast surface display system which could be considered and have been reviewed[24]. The originally developed Aga2p fusion system is used throughout this work[25]. This system displays mutant proteins by fusing them to the Aga2p subunit of the matting protein a-agglutinin in S. cervisae. Aga2p is disulfide bonded to the native Aga1p subunit which anchors itself to the cell well with β-glucan covalent linkeages. The displayed library protein is fused protein to Aga2p through a flexible linker and is therefore free to interact with species in solution. Aga1p is chromosomally expressed and the Aga2p and the library protein are encoded by a plasmid which is transformed into the yeast cell. Expression of both Aga1p and Aga2p are controlled by a galactose promoter. The library protein is labeled by an N-terminal HA tag followed by a 12 residue glycine serine linker and a C-terminal cMyc tag, which appears after the library protein. The HA and cMyc tags are used in sorting to normalize binding for cell surface expression levels, these vary between 10^4 and 10^5 copies[26]. The cMyc tag is used to determine full length expression of the library protein, i.e. to detect the cMyc tag the displayed protein has to have transcriptionally made it through the DNA strand without reaching a stop codon. For both the HA and cMyc tags to be detectable the translated protein has to clear the yeast surface display quality control system. There is data that indicates that misfolded proteins or molten globulins are not able to do this[27]. The yeast

surface display plasmid, referred herein as pCTCON, has a tryptophan marker and is maintained in yeast using casamino acids. Dextrose can be used to expand the yeast population and a galactose buffer is used to induce protein expression.

Display technologies provide methods through which extremely large pool of mutants can be generated; selections for productive binding mutants can be carried out, amplified and further diversified. Typically, initial selections from a naïve library are carried out to capture low and high affinity ligands[20,28]. Generally after enrichment the selected library is re-diversified through an error prone PCR step. For yeast surface display and other cell based display systems the DNA is extracted and purified from the cell and an error prone PCR is performed (EPCR). EPCR generally incorporates about 1-5 mutations/mutant and thereby re-diversifies the selected binders. Multiple rounds, often 4-8, consisting of enrichment followed by re-diversification through ePCR are generally done[20,28]. During these early enrichment steps negative selections against proteins or materials used in the selection process are done as well. Generally for the yeast display system for instance, streptavidin coated magnetic beads are used in the positive selections to label the target of interest. Therefore streptavidin would be used alone in negative selections to remove unwanted streptavidin binders. **Figure 1.1** shows a typical flow chart of how affinity ligands to targets of interest are generated using the yeast display system [24].

1.1.2 Engineering High Affinity ligands

A focus of the protein engineering literature is to create high quality libraries that decrease the time spent generating suitable affinity ligands, increase the number of possible

targets that are accessible by a given scaffold and increase the probability of the selection process yielding ligands that can be easily produced and are not prone to aggregation. An array of computational methods is being explored toward these ends. One method uses a scoring algorithm to find a suitable scaffold for a suitable targeted epitope based on structural data in the Protein Database (PDB) [29]. After a scaffold is selected further analysis is done in silico to select which residues will be mutated to create a library with an optimal probability for creating successful ligands[29]. Alternatively, in silico design can be used to determine a set of residues that will bind to a site of interest. These residues are then inserted into a scaffold and screened for binding[4]. A third approach uses in silico analysis to generate a focused library based on target docking studies by a specific ligand[30]. The in silico analysis enabled authors to select which residues to mutate furthermore, at each residue the diversity was reduced by further by choosing only a subset of amino acids to be evaluated at each positon. The authors demonstrated that a focused library yielded 30 times greater enrichment over an alternative naïve library based on saturated mutagenesis at a selected binding interface.

Another alternative method is to create what amounts to a super scaffold, such as the LoopDARPin[31]. The DARPIN is a well-established non-antibody scaffold that has been used to generate high affinity ligands to a number of targets[32–35]. To enhance the library design the authors inserted an antibody like CDR3 loop in place of a smaller loop segment which had formerly been part of the surface used to generate a binding interface. The authors found that a design in which loops were inserted such that they would not destabilize the scaffold was

highly successful at generating high affinity ligands by panning the naïve library without affinity maturation[31].

Techniques for improving how the affinity maturation process is carried out have also been developed for the Fibronectin scaffold. A 0.3 pM lysozyme binding mutant was selected after nine rounds of mutagenesis[36]. Affinity in Fibronectin is generated typically through three loop regions, in this work the authors showed that varying loop lengths and interchanging loops during affinity maturation cycles was effective. During rounds of mutagenesis error prone PCRs were either restricted to just the loop regions or to framework regions. During each round two libraries were made and then mixed after transformation in yeast. Interestingly, the loop sequences stopped evolving after the fifth round of affinity maturation but framework mutations continued to improve the affinity of the ligand. Lastly, in a final 9th round non-conserved mutations in the best binders were again mutated through saturated mutagenesis. This step led to an additional threefold improvement in K_D.

1.1.3 Engineering Multivalent Affinity Ligands

Bi- or multivalency are often used mechanism by which affinity is increased. Intermolecular bivalency or mutlivalency are commonly used in the human immune system to increase affinity through avidity[37]. A canonical example of this occurs when both complement determining regions of an antibody bind two distinct surface bound epitopes simultaneously. This dramatically changes the off-rate of the antibody, for instance a 10 nM antibody with a 12 minute half-life will be enhanced such that the half-life easily reaches 200 plus hours when both CDRs are engaged[38]. In the human immune system multivalence is typically generated in

a similar manner: a particular pair of identical binding domains and targets are present such that multiple contacts can be made. This is termed intermolecular avidity as two or more copies of either the target or ligand are required for the generation of affinity.

Chelating recombinant antibodies were developed to harness the power of avidity for intramolecular binding. This case involves one target being bound by one ligand through multi-epitope engagement. The key to this approach is that neither target, nor ligand, need be present on a cell surface for avidity to be observed. **Figure 1.2** shows how a chelating affinity ligand would compare in different circumstances to an antibody. Chelating is a term used by Neri et al. in one of the first papers linking two ligands which bind the same protein at different sites[39]. More often than not however the term chelating is dropped for bivalent or multivalent which is unfortunate as "bivalent" does not specify that the bivalency is generated by two epitopes on the ligand binding two epitopes on the target. The term chelating was applied to strictly identify the bivalent interaction as consisting of two proteins interacting through two binding sites. In the original work on chelating recombinant antibodies a low affinity lysozyme binding scFv (single chain fragment variable domain of an antibody), K_D of 1 μM, and a second scFv targeting a unique epitope, K_D of 10 nM were linked. Crystal structures with both ligands and lysozyme were used to determine an appropriate linker length between the scFv's (an 18 residue hydrophilic linker was used, GSSSGSAGKASGGSGSGG). Using competition ELISA the authors estimate a 20-100 fold improvement in the K_D relative to the 10 nM scFv.

Eight years later a 2003 modelling study endeavored to quantify the magnitude of the affinity enhancement that could be observed in a chelating system[40]. Thermodynamic analysis reduced the K_D observed in a bivalent system to this equation:

$$K_D^{AB} = \frac{K_D^A K_D^B}{A_{eff}}$$

The equation shows that the K_D for a bivalent system is equal to the product of the individual K_D's for each ligand divided by an affective concentration, A_{eff}. The effective concentration is defined by the distance between the epitopes of either ligand and the length of the linker used to fuse the two ligands together. A linker that is too short will of course not produce productive binding and a linker overly long will not be optimal. A model was developed to estimate how the effective concentration varies with the linker length and the distance between the ligand binding sites[41]. The model estimates that the effective concentration will vary between 0.01 mM and 100 mM for synergistic ligands. **Figure 1.3** shows the bivalent K_D expected based on effective concentrations in this range and two ligands with equal individual K_Ds. The chart shows that even linking the least synergistic pairs, i.e. even those with low micromolar K_Ds and relatively low effective concentrations that a 10 to 100 fold improvement in affinity should be observed. Substantially compounded affects are observed as K_D decreases and if the effective concentration increases. The first methods for systematically harnessing avidity were published shortly following this linker study.

Two separate publications validated the phage display method for the ability to pan multivalent combinatorial libraries against targets of interest and generate chelating affinity ligands. The first assembled unique, selected A-domains into trimers, these were termed

avimers. Avimers, were selected for binding to five different cellular receptors; each ligand evaluated had K_D's well below the nanomolar level, some below the picomolar level[42]. Multivalency in this case was generated by ligand addition. To generate functional trimeric ligands the authors first designed a monovalent A-domain library, selected the highest affinity clones to targets of interest and then evaluated soluble clones for binding and inhibition. The authors took a single clone that behaved favorably and then generated a bivalent library by fusing the selected clone back to the original library. They then repeated the selection and screening process with the bivalent library. Finally, after choosing a promising bivalent pair they fused the dimeric library back to the monovalent library to create a library of trimers. The avimer represented both a novel scaffold and a novel selection methodology. The scaffold has only 12 conserved residues, is highly mutateable and naturally contains linkers of various lengths that were also used for affinity generation. Given the plasticity of the scaffold the hypothetical library diversity generated is also extremely large, 10^{23}. Likely, the extreme affinity of these ligands is due to their ability to generate low nanomolar monovalent clones from a naïve library, the short linker between each A-domain and the compounding effect of tri-valency. Importantly, the authors validated the high affinity of their ligands in competition assays. This enables authors to look at soluble phase binding. Many affinity assays use surface bound ligand to capture a target protein whose surface bound presence will be quantified. The relationship between soluble phase concentration of the target protein and the surface loading is in part defined by the K_D. However, this approach can enable the formation of intermolecular affinity. A competition assay is identical except the surface bound ligand is also in solution.

The solution phase ligand binds the target and thus prevents the target from binding the ligand on the surface. The extent of inhibition is a function of the solution phase ligand concentration and the solution phase K_D[43]. Measuring low K_D's using competition experiments demonstrated that both chelating ligands were made and that it is possible for the phage display system to be used to select chelating affinity ligands. This use of phage display for this purpose may not necessarily have been possible. There are five copies of the library expressed on the surface of phage in the M13 phage display system, therefore it could have been possible for the target to only bind the phage displaying the multivalent affinity ligands when binding between separate phage surface fusions. While the authors did not rigorously show that this was not possible in the phage display system they do show that their ligands do not require intermolecular avidity to generate high affinity. Despite the promising results of the avimer scaffold and selection system few library based multivalent selections have been carried out with avimers or other scaffolds.

A second study published two years after the avimer work again used phage display as selection format for selecting functional high affinity bivalent ligands[44]. In this case the authors used the same scFv's as were used in the original Neri study and created a library to determine the optimal linker length. Eleven linkers were generated with sizes ranging between eleven and 54 residues. Additional diversity was added by the presence of a randomized residue in one or more positons in the linker. This study found that 16-21 residues was optimal and more importantly again this time took special effort to validate the phage display system as a selection tool for selection of bivalent scFv's. The authors showed that both scFv's were

functional in the phage system and that the selected bivalent construct gave high affinity in a competitive binding experiment.

A unique approach to multi-epitope engagement is the affinity clamp, for this construct two binding domains interface to create a peptide binding site. The original affinity clamp study demonstrated that a PDZ peptide binding domain could be linked to a fibronectin library and very specific high affinity peptide ligands could be generated[45]. The two affinity ligands were linked such that the fibronectin binding domain faces and can clamp down on the peptide and PDZ domain simultaneously. To enable this arrangement the PDZ domain had to be circularly permutated such that its C-terminus would line up appropriately with the N-terminus of fibronectin. A later publication adapted this technique to for use with an SH2 domain used in place of the PDZ domain[46]. This demonstrated that the clamp can likely be leveraged to all SH2 domains and affinity and specificity can be enhanced as desired.

1.1.4 Bivalent ligands as tools for manipulating biological phenomenona

Bivalent ligands have also been generated to probe biological interactions. The motivation for generating bivalency in these cases is independent of increasing the affinity of the ligand. Two such examples are to demonstrate a method for inhibiting an oncogenic kinase and to increase receptor down regulation[47,48]. The Grebien et al. study also highlights another mechanism by which tandem affinity ligands are selected. A first affinity ligand was selected and then a second ligand was selected in the presence of the first ligand. The first ligand in the work inhibited phosphotyrosine binding and the second ligand inhibited kinase activity, both ligands acted on the kinases SH2 binding domain. Linking the two ligands created a potent

inhibitor of the kinases activity. The study by Hackel et al. demonstrates that linking ligands which bind non-overlapping epitopes on epidermal growth factor receptor (EGFR) can cause receptor clustering, increase the rate of receptor recycle, resulting in downregulation of the receptor. In this case a series of high affinity monobodies were selected after several rounds of affinity maturation and analyzed for the ability to downregulate EGFR without activating downstream ERK signaling. The selected monobodies did not have the desired effect of decreasing the presence of EGFR on the cell surface. Ligands were therefore linked together in pairs as homobivalent or heterobivalent ligands. Pairs which were constituted by non-competitive binding heterobivalent ligands were found to be most effective at downregulating the receptor without agonizing signaling.

1.2 Thesis Overview

The focus of this thesis is based around generating intramolecular avidity using combinatorial approaches. Very few literature examples have pursued this type of approach despite the demonstration that it is possible and can be very effective. Our motivation for doing this is to more readily generate high affinity, stable ligands and to establish yeast surface display as an effective means of doing this.

Chapter 2 details an approach for linking a pool of low affinity ligands together in a combinatorial manner to generate synergistic bivalent ligands to the model target of lysozyme. Lysozyme represents a small (14.3 kDa) folded protein that has been commonly used as a target for protein engineering studies in the past. This study shows definitively that selected ligands bind with intramolecular affinity and that the yeast displayed bivalent ligands are not

able to bind lysozyme with intermolecular avidity. That is two different surface displayed ligands are not orientated such that they can synergistically bind lysozyme. We go on to show that the individual ligands in the bivalent construct can be adapted to the tripartite split GFP system and used for quantitative detection of lysozyme in an impure mixture.

Chapter 3 describes targeting the intrinsically disordered C-terminus of β-catenin using both non-tandem and multi-valent affinity ligands. For both cases a sequence within the intrinsically disordered region was pre-targeted using a twelve residue peptide sequence. Pre-targeting was carried out by first selecting Sso7d ligands from a naïve library which have affinity for the peptide. The selected pre-targeted ligands were diversified through mutagenic PCR, and used to create a second generation library. This second generation library was selected for binding to β-catenin. Pre-targeting was found to be effective as the highest affinity ligands selected from the second generation library bound specifically to the peptide epitope of interest. Also, it was found that for both non-tandem and multi-valent ligands affinity maturation resulted in gain of an epitope on β-catenin. Interestingly, this happened as affinity to the desired epitope was also improved. Despite the increased size of β-catenin (85 kDa) relative to lysozyme we were still able to generate bivalent affinity ligands that bind in a one to one manner, i.e. they do not require intermolecular binding to generate avidity. In addition the ligands selected were found to all be functional disulfide bonded dimers.

1.3 Figures

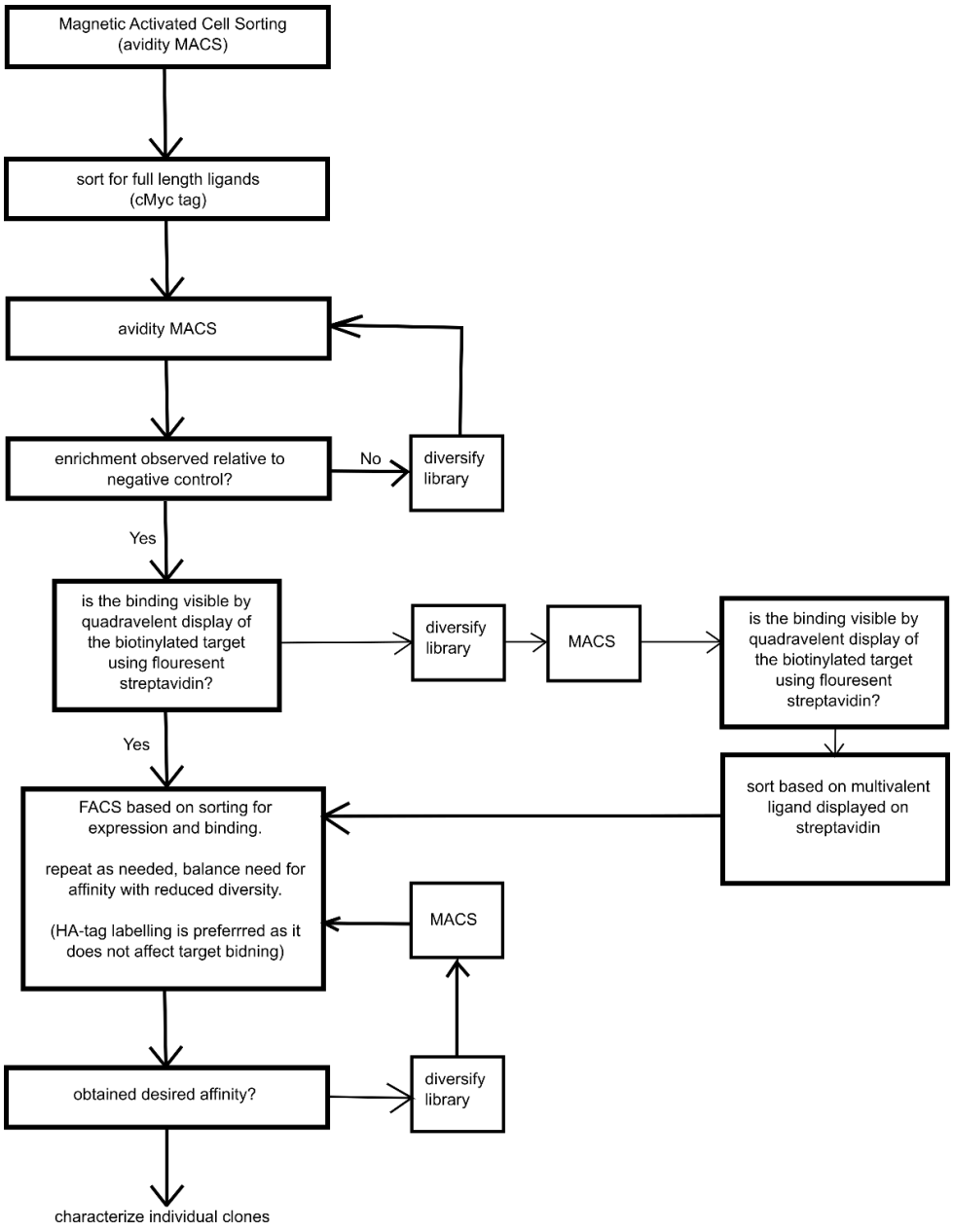


Figure 1.1: Flowchart of protein engineering by yeast surface display.

A flowchart which shows how the generation of high affinity ligands is carried out. This figure is adapted from[24]. Avidity MACS is a term used to describe a magnetic activated cell sorting system

where the target of interest is immobilized on magnetic yeast and then panned against the yeast display library. Cells with productive binding cells are selected by using a magnet to immobilize the magnetic particle. Low affinity ligands can be selected in this way as multi-epitope engagement is generated as many copies of the target can binding many yeast displayed ligands. This type of interaction results in a very low off-rate. Library diversification is achieved through mutagenic PCR.

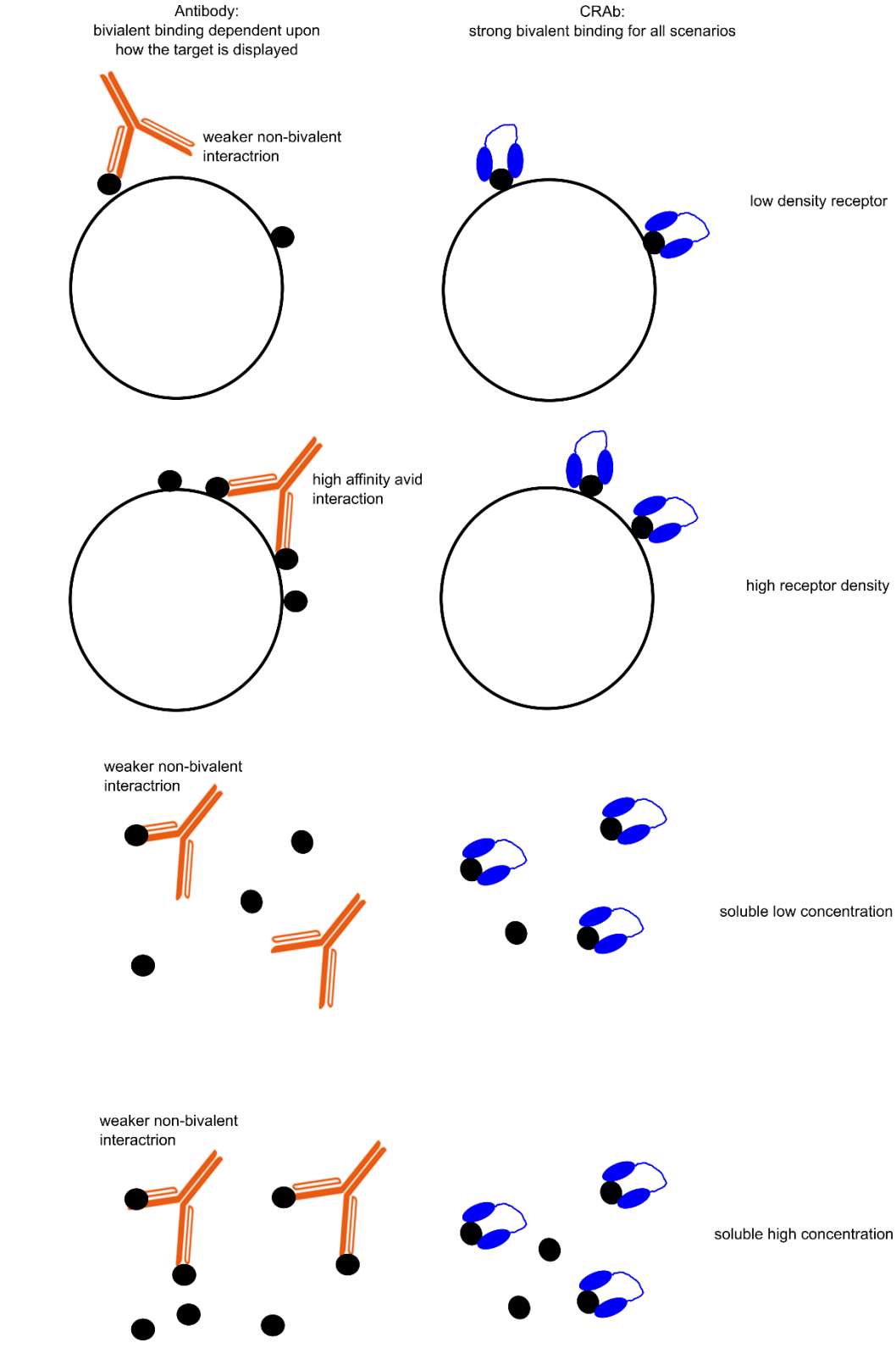

Antibody:
bivialent binding dependent upon
how the target is displayed
CRAb:
strong bivalent binding for all scenarios
weaker non-bivalent
interactrion
low density receptor
high affinity avid
interaction
high receptor density
weaker non-bivalent
interactrion
soluble low concentration
weaker non-bivalent
interactrion
soluble high concentration

Figure 1.2: Chelating recombinant ligands generate bivalency regardless of target context

The left panels depicts how an antibody's ability to generate a high affinity bivalent interaction is dependent upon how the target is displayed. The right panel depicts how the chelating recombinant antibody is designed to generate high affinity regardless of the context of the target. This figure is adapted from[44].

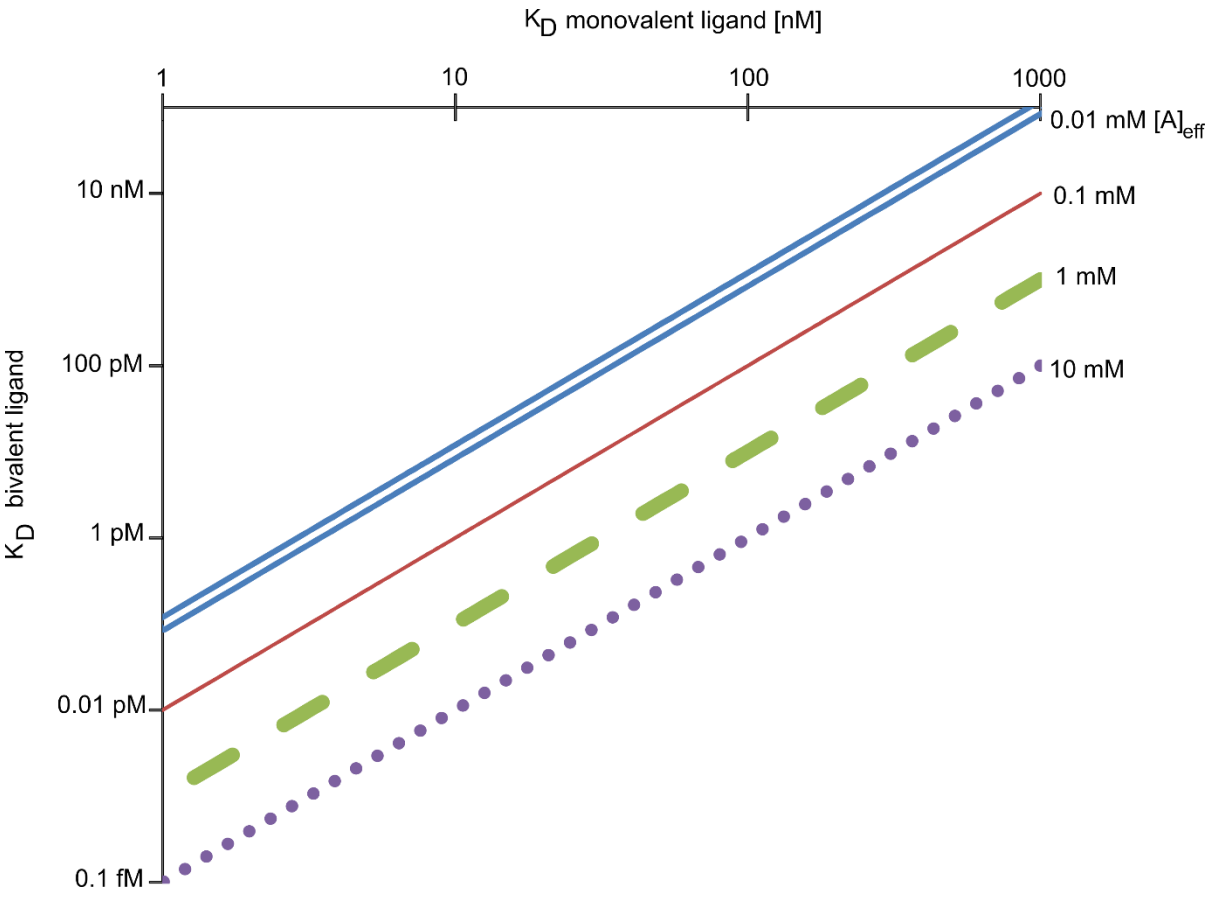


Figure 1.3: Estimated enhancement for linking ligands

This figure shows the estimated K_D of a bivalent ligand which is comprised of two linked monovalent ligands each having identical K_D's. The four different curves are shown for four different effective concentrations. Effective concentration is defined by the concentration an unbound ligand in a

bivalent construct has for its target once the alternate ligand in the bivalent construct binds to its specific epitope on the target. The effective concentration is determined by how far part the individual epitopes for each ligand are from each other and the linker length used. This figure is based on analysis in [41].

CHAPTER 2

COMBINATORIAL PAIRWISE ASSEMBLY EFFICIENTLY GENERATES HIGH AFFINITY BINDERS AND ENABLES A "MIX-AND-READ" DETECTION SCHEME.

(Adapted from manuscript in preparation by Carlin, Cruz et al.)

2.1 Introduction

Binding proteins derived from small non-immunoglobulin scaffolds have gained increasing prominence as affinity reagents and in therapeutic applications. Generation of high affinity binders often involves multiple rounds of random mutagenesis followed by combinatorial screening of the resultant libraries. However, high rates of mutation may result in loss of protein stability, particularly in the context of small protein scaffolds[36]. The use multivalency or intramolecular avidity is a powerful alternative for generating high affinity binding. In nature, proteins binding their cognate target with high affinity are commonly generated by combining multiple protein subunits that bind distinct, non-overlapping epitopes on the target. For instance, the very high affinity of binding of the transcription factor Oct-1 for its target DNA sequence (K_D = 71 pM) is a result of combining two subunits that bind with low to moderate affinities (K_D = 150 nM and 1.7 mM)[49]. Bivalency or multivalency has been exploited to engineer high affinity binding proteins. In one approach, individual ligands that have typically gone through several rounds of mutagenesis and selection, and are identified as binding non-overlapping epitopes, are linked to obtain high affinity binding[39,47,50]. A trial-and-error approach is used to identify binders where such linking results in increased affinity. In a second approach, an engineered affinity ligand is linked to a randomized scaffold; higher affinity binders are then selected from the resulting combinatorial library. Examples of this strategy include avimer binders [42]and affinity clamps[45,46]. Here we present an alternative strategy for efficiently generating high affinity binders using multivalency. In our approach, high affinity binders are isolated from a combinatorial library that is constructed by random pairwise assembly of low affinity binding proteins using a flexible linker.

We demonstrate our approach in the context of generating high affinity binders based on the highly stable Sso7d scaffold, for a model target, lysozyme. We have previously shown that the Sso7d

protein from the hyperthermophilic archaeon *Sulfolobus solfataricus* is a versatile scaffold that can be used to generate highly stable binding proteins for a wide spectrum of targets[13,14]. A pool of low affinity binders to lysozyme was subjected to random mutagenesis and individual mutants were randomly combined pairwise using a flexible $PG(PT)_8$ linker to construct a combinatorial library. We show that such a library has a greater frequency of high affinity clones than a library of equivalent diversity generated using random mutagenesis; notably, the latter is the most commonly used strategy for affinity maturation. High affinity of the selected clones using our approach is a consequence of combining low affinity binders targeting non-overlapping epitopes.

Finally, it is important to note that efficient identification of binders with non-overlapping epitopes on the target is implicit to our strategy; this in turn can be exploited in biosensing applications. In particular, we show that the two subunit proteins that compose the high affinity bivalent binder can be used to design a simple "mix-and-read" format assay for target quantification. In our scheme, binders that compose the bivalent protein are fused to two out of the three fragments of tripartite split GFP[51]. In the presence of the target, assembly of GFP and fluorescence reconstitution occurs in concentration-dependent manner. Mix-and-read assays, wherein a signal can be obtained by simply mixing reagents with the target analyte, have significant advantages over conventional assay formats such as an enzyme-linked immunosorbent assay (ELISA), wherein time-consuming wash steps and protein immobilization are required.

2.2 Results

2.2.1 A bivalent library yields a higher frequency of putative high affinity clones.

Affinity maturation of binding proteins is most commonly carried out by multiple rounds of mutagenesis and screening. We hypothesized that the pool of low affinity binders obtained after an initial round of library screening will contain binders to multiple epitopes, and that linking specific pairs of proteins with a flexible linker may generate high affinity binders. Indeed, examples in the literature demonstrate that linking binders with even weak affinities (micromolar to hundreds of nanomolar K_d) can generate high affinity binding[41]. To evaluate this hypothesis, we used as our starting point a pool of low affinity binders to lysozyme, obtained after magnetic sorting and a single round of screening by fluorescence activated cell sorting (FACS). The selected ligands originated from a yeast surface library of ~ 10^8 Sso7d mutants; the library was generated by randomization of ten surface residues on the Sso7d scaffold[13]. DNA from this pool of mutants was subjected to random mutagenesis and pairs of binders were randomly linked using a flexible $PG(PT)_8P$ linker[52] using homologous recombination in yeast[25], to construct a yeast display library of ~ 10^7 mutants (**Figure 2.1a)**. Notably, the PT linker used does not have any homology with the $(G_4S)_3$ linker in the yeast surface display vector. Also, this linker length represents a compromise between generating the highest affinity – shorter linkers connecting binders to epitopes that are closer in space result in greatest increases in affinity – and generating high affinity binding using pairs of ligands that may bind farther apart[41]. In parallel, a yeast display library of monovalent binders with similar diversity was also generated using the same pool of DNA (**Figure 2.1a**). Flow cytometry analysis was used to obtain a quantitative comparison between the bivalent and monovalent libraries as follows. Yeast cells, of identical sample size from each library (10^5 cells), were simultaneously labeled with varying concentrations of lysozyme and an antibody against the HA epitope tag; an HA tag is present on each

monovalent and bivalent protein fusion on the yeast surface (**Figure 2.5**). **Figure 2.1b** shows that the number of clones that fall in a conservatively drawn sort window is significantly higher for the bivalent library at lower concentrations of lysozyme (1 nM or 10 nM); this suggests a greater frequency of high affinity clones in the bivalent library. Mutants with higher affinity are characterized by a higher ratio of target binding fluorescence to the fluorescence corresponding to binding of the HA antibody. Indeed, this is the rationale for "sorting along the diagonal" during flow cytometry screening in yeast surface display[22]. Therefore, this ratio was used as a quantitative metric to further compare the number of putative high affinity clones in each library. Cells were binned into one of four bins based on this ratio. **Figure 2.1c** and **2.1d** show that at lysozyme concentrations of 1 nM and 10 nM, the bivalent library shows a greater frequency of yeast cells with higher values of the ratio across all bins. This result strongly suggests that the bivalent library has a greater frequency of putative high affinity than the monovalent library of similar diversity.

The combinatorial bivalent library was generated with the intent of creating pairs of ligands that bind synergistically when flexibly linked. The comparison of the monovalent and bivalent libraries shown in **Figure 2.1B** indicates that this mechanism is systemic within the bivalent library. The bivalent library was generated using the identical genetic library as the monovalent library. Therefore, the increase in frequency of selectable ligands as the labeling concentration decreases in **Figure 2.1B** is directly attributable to the assembly of the monovalent ligands into synergistic pairs. The 1 nM labeling concentration indicates there is approximately six times the number of bivalent selectable ligands then monovalent ligands. This would indicate that a random clone selected at 1 nM from the bivalent library has an approximately 83% (5/6) probability of having its affinity arise from synergistic binding of two ligands rather than being derived from a single high affinity ligand in the construct.

2.2.2 Magnetic sorting and FACS identifies a pool of bivalent lysozyme binders with highest affinity.

The bivalent library was screened using magnetic sorting and FACS to obtain a pool of binders with highest affinity as described[22]. DNA sequencing revealed that this pool predominantly contained bivalent binders comprising two distinct domains. 9 out of 14 clones analyzed contained two distinct Sso7d binder subunits. As shown in **Figure 2.2a**, 16 individual binders from 10 distinct families were identified constituted this pool of bivalent proteins; pairs of binders that form the nine distinct bivalent proteins are shown in **Figure 2.2b**. In addition to the nine bivalent proteins, four of the clones analyzed corresponded to a set of monovalent proteins derived from the lysozyme binder with highest affinity previously isolated from the naïve Sso7d library of 10^8 mutants[13]. The monovalent clones all contained a conserved G43D mutation; three had only this mutation and the fourth also had T2I and C20A mutations. Interestingly, however, none of the bivalent proteins contained this monovalent protein. Six out of the 16 individual binders and six out of the nine bivalent proteins contained mutations outside the ten residues mutagenized to generate the original Sso7d library[13]. Positions where these mutations occur are shown in **Figure 2.2c**. Residues 8, 27, 36, 38 and 60 are in spatial proximity to the residues on the three beta sheets randomized to generate the original Sso7d library, while residues 5 and 11 are a part of the hydrophobic core. In addition to the monovalent and bivalent proteins, DNA sequencing also identified a trivalent protein. Monovalent and trivalent proteins likely arise in the library due to undesired homologous recombination events during library construction. Further analysis was restricted to a single, randomly chosen bivalent protein, denoted as BVL-1. The N-terminal and C-terminal components of BVL-1 are denoted as NTL-1 and CTL-1 respectively.

The presence of monovalent and trivalent ligands in the highest affinity pool of ligands selected from the bivalent library motivates analysis on the quality of the bivalent library. This is necessary to determine if constructs with alternative numbers of ligands were enriched for. Sequencing 18 members of the bivalent library reveals that eleven code for bivalent genes, four code for ligands with a nearly completely truncated N-terminal ligand, one has a significantly truncated C-terminal ligand and two code for monovalent ligands; no trivalent ligands were observed. N-terminally truncated ligands would not be expressed as they are observed to be terminated by a stop codon prior to expression of the second ligand. This sequencing data indicates that over 10% of the bivalent library is comprised of monovalent ligands which do not contain a linker region. This population is likely highly diversified and therefore the selection of a single monovalent ligand, and mutants of this ligand, indicates that monovalent ligands were not enriched for. The presence of this single monovalent ligand, from this diversified pool, would on the contrary indicate that this ligand is an outlier of the monovalent library.

2.2.3 Determination of intermolecular vs intramolecular binding of yeast displayed BVL-1; recombinant expression, stability and binding of lysozyme.

We tested the hypothesis that high affinity binders isolated from the combinatorial bivalent library is largely due to synergistic binding of two low affinity binders separated by a flexible polypeptide linker by evaluating one bivalent ligand at random. We measured the binding affinities of the bivalent BVL-1 as well as the monovalent NTL-1 and CTL-1 subunits for lysozyme, using yeast surface titrations (**Figure 2.3a**). The K_D of BVL-1 for lysozyme was estimated as 1.7 nM, whereas the corresponding values for NTL-1 and CTL-1 were estimated as 1.3 μM and 250 nM, respectively. It has been previously demonstrated that K_D values estimated from yeast surface titrations are consistent with those estimated using soluble proteins[53]; notably, lysozyme was one of the target

proteins in this study. Nevertheless, K_D estimates from yeast surface titrations may be erroneous if the target molecule can simultaneously bind more than one cell surface fusion; in such a case, the strength of binding affinity is overestimated. In the context of BVL-1, binding of lysozyme to NTL-1 and CTL-1 subunits on two distinct cell surface fusions will result in an increase in apparent affinity. We therefore sought to rule out this possibility of intermolecular binding to lysozyme by distinct yeast surface display fusions. We therefore conducted titrations using a yeast dual-display system where CTL-1 and NTL-1 are expressed as distinct, unlinked cell surface fusions (**Figure 2.3b**). The apparent K_D for the dual-display configuration (160 nM) is ~ 100X higher than the K_D for BVL-1, indicating that K_D measurements for BVL-1 are not affected by the avidity effect arising from association of lysozyme with CTL-1 and NTL-1 on distinct cell surface fusions. We further expressed BVL-1 recombinantly and measured the K_D of binding with lysozyme using a competitive binding assay[53]. The K_D estimated in this assay (2.7 nM) is consistent with results from yeast surface titrations (**Figure 2.3d**). These results strongly indicate that the high affinity of BVL-1 for lysozyme arises from synergistic binding of the low affinity NTL-1 and CTL-1 subunits.

The residues selected in BVL-1 tend to be hydrophobic, therefore size exclusion chromatography (SEC) was used to validate that one BVL-1 protein binds to one lysozyme protein and that BVL-1 is a well behaved monomer. BVL-1 is observed to run through an SEC column as a nearly single peak protein with a residence time corresponding to a 38.1 kDa protein, 1.9 times larger than the expected 20.4 kDa. However, examples of elongated proteins, which exist as monomers, have been shown to run at five times their molecular weight[54]. The SEC data for the BVL-1, lysozyme complex indicates that this is likely the case for the BVL-1 protein as well. The complex residence time corresponded to a 38.4 kDa, which is just 10% larger than the expected 34.7 kDa. The flexibly linked binding domains of BVL-1 appear to take on a somewhat elongated shape when free

in solution but when bound to lysozyme the complex adopts a more spherical conformation. This evidence supports the conclusion that BVL-1 is a monomeric protein which binds lysozyme in a one to one manner. SEC data can be seen in **Figure 2.4**. **Figure 2.5** is an SDS page gel which confirms that both lysozyme and BVL-1 are observed in the complex elution peak shown in **Figure 2.4f**. The second peak in **Figure 2.4f** is shown to correspond to lysozyme alone.

The K_D of a clone with highest binding affinity for lysozyme, previously isolated from an Sso7d library of ~ 10^8 clones, was estimated as 349 nM[13]. Thus, BVL-1 represents a ~ 200X improvement in binding affinity in a single round of random mutagenesis, library construction and screening. Finally, it is important to note that BVL-1 retains thermal stability. The melting temperature of BVL-1 was estimated as 84.5 °C in a differential scanning fluorimetry assay (**Figure 2.6**). Furthermore, BVL-1 could be expressed recombinantly at high yields in *E. coli*. Yields in not yet optimized shake flask cultures were estimated as 40 mg/L.

2.2.4 Binding of NTL-1 and CTL-1 to non-overlapping epitopes on the target can be exploited to design a "mix-and-read" assay for target quantification.

NTL-1 and CTL-1 independently bind lysozyme, and when combined through a polypeptide linker yield the high affinity bivalent binder BVL-1; this is consistent with NTL-1 and CTL-1 binding non-overlapping epitopes on lysozyme. Binding proteins that bind non-overlapping epitopes on a target are extensively used in biomolecular detection schemes, e. g. sandwich ELISA. Of particular interest is a mix-and-read assay scheme that eliminates the need for wash steps or protein immobilization, which are necessary in case of ELISA. We hypothesized that the NTL-1 and CTL-1 can be used to construct a biosensor for a mix-and-read assay based on a modification of the tripartite split-GFP system[51]. In our scheme (**Figure 2.7a**), NTL-1 and CTL-1 are fused to the N-terminus and C-terminus of the GFP fragments GFP11 and GFP10, respectively. Binding of NTL-1 and CTL-1 to

lysozyme brings GFP10 and GFP11 in close proximity, and allows assembly with the GFP1-9 fragment that is present in solution. The reconstituted complex comprising GFP10, GFP11 and GFP1-9 is fluorescent. As seen in **Figure 2.7b**, the fluorescence readout is dependent on the concentration of lysozyme over a large concentration range (0-15 μM). This result provides direct evidence that NTL-1 and CTL-1 bind non-overlapping epitopes on lysozyme. Interestingly, the fluorescence readout increases as a function of incubation time. This increase may be explained by the irreversible nature of the assembly of the GFP fragments[51]. Consequently, fluorescence will persist even upon dissociation of lysozyme from the reconstituted complex (**Figure 2.7b**). The dissociated lysozyme may then further form another GFP complex. Thus lysozyme may act as a catalyst for reconstitution of GFP fluorescence. An alternative explanation for time-dependent increase in the fluorescent readout is the slow kinetics of GFP fluorescence reconstitution in our modified tripartite split-GFP system.

2.3 Materials and Methods

2.3.1 Library construction.

Lysozyme (Thermo Fisher) was biotinylated using a 3:1 molar excess of EZ-Link™ Sulfo-NHS-LC-Biotin (Thermo Fisher). Excess reagent was inactivated using 1mM Tris and removed by dialysis. The pool of low affinity binders to lysozyme in yeast display format, previously obtained after magnetic sorting of an Sso7d library of ~ 10^8 mutants[13], was labeled with 1 μM biotinylated lysozyme and a rabbit anti-HA antibody (Life Technologies; 1:100 dilution), in PBS containing 1 g/L bovine serum albumin (PBSA). Secondary labeling was carried out using streptavidin-R-phycoerythrin (SA-PE; Life Technologies; 1:500 dilution) and donkey anti-rabbit IgG DyLight 633 (DαR633; ImmunoReagents, Raleigh, NC). Labeled cells

were subjected to fluorescence-activated cell sorting (FACS), and all cells exhibiting binding to both lysozyme and the HA antibody above background were sorted. Plasmid DNA was extracted from these cells using the ZymoPrep kit (Zymo Research); this was used as template DNA for library construction.

Linear DNA used for generating both libraries were obtained by error-prone PCR using the aforementioned template DNA, as described[55]. Briefly, 35 cycles of PCR were conducted with an estimated error rate of ~ 3.8 nucleotides per template, using primers Vf and Vr, to obtain DNA product P1. All primers were purchased from Integrated DNA Technologies (IDT); sequences of all primers used can be found in **Table 2.1.** To obtain DNA products B1 and B2, coding for the N-terminal and C-terminal portions of the clones in the bivalent library respectively, P1 was amplified by PCR with two different primer sets. Primers NTLf and NTLr were used to generate B1; primers CTLf and CTLr were used to generate B2. To obtain DNA product M1 coding for the clones in the monovalent library, P1 was amplified with primers NTLf and CTLr. Finally, the bivalent and monovalent libraries were generated by homologous recombination in yeast (**see Figure 2.1**). Three transformations were carried out for each library. 4 μg of linearized pCTCON vector and 4 μg of M1 were used for each transformation in case of the monovalent library. For the bivalent library, 4 μg of linearized pCTCON vector and 4 μg each of B1 and B2 were used per transformation.

2.3.2 Comparison of monovalent and bivalent libraries.

At each concentration evaluated, $2x10^6$ yeast cells were labeled with biotinylated lysozyme and anti-HA antibody (Life Technologies; 1:100 dilution) in PBS containing 1 g/L bovine serum

albumin (PBSA), followed by secondary labeling with SA-PE and donkey anti-rabbit IgG DyLight 633 (ImmunoReagents, Raleigh, NC). Labeling volume was chosen to ensure ~ 10-fold excess of lysozyme molecules in solution relative to the yeast-displayed fusions as described[22]. A BD Accuri C6 cytometer was used to analyze 100,000 cells in all replicates and controls. The number of cells that fall within a conservatively drawn gate (see **Figure 2.8**) was assessed. For each cell found in this gate, a ratio of the fluorescence signal corresponding to lysozyme binding (FL2 in **Figure 2.8**) to the signal corresponding to anti-HA antibody binding (FL4 in **Figure 2.8)** was computed. Gated cells were placed in one of 4 bins (1-2, 2-3, 3-4 and 4+) based on the FL2/FL4 ratio. The fraction of cells in each bin relative to total number of cells analyzed (10^5) was used to generate histograms (**Figure. 2.1c, d**).

2.3.3 Selection of the highest affinity bivalent ligands

Library screening was conducted using a combination of magnetic sorting and FACS, as previously described[22]. Briefly, $2x10^8$ cells expressing cell surface fusions from the bivalent library were subjected to negative selection by incubation with 50 μL of Dynabeads Biotin Binder beads (Life Technologies) in 2 ml PBSA, at room temperature for 1 hr. Cells that did not bind the beads were recovered and incubated with 100 nM biotinylated lysozyme in PBSA (5 ml) for 1 hour at room temperature. Cells were then washed 3 times with PBSA and incubated with 25 mL of Dynabeads Biotin Binder in 2 mL PBSA, at 4 °C for 1 hr. Subsequently, bead-bound cells were isolated using a magnet and expanded in culture.

The pool of cells after magnetic sorting was further incubated with 10 nM biotinylated lysozyme in PBSA (1 ml) for four hours, washed three times with PBSA, and incubated with 10 mM non-biotinylated lysozyme in PBSA (100 μls) for one hour. Subsequently, cells were

labeled with an anti-HA antibody, followed by secondary labeling with SA-PE and GαR633, and subjected to FACS using a MoFlo cell sorter (Beckman Coulter) to isolate binders with the highest affinity for lysozyme. Cytometry analysis of cells expanded after FACS showed a largely homogenous population of cells, as assessed by labeling with 1 nM lysozyme. Therefore, plasmid DNA was extracted from this pool and sequenced.

2.3.4 Simultaneous Yeast Surface Display of CTL-1 and NTL-1 (Dual Display)

$PG(PT)_8P$-CTL-1 was cloned into the pCT302 yeast display vector containing the Trp selectable marker[56]. NTL-1 was cloned into a variant of the pCT302 vector containing a Leu selectable marker; this vector was a kind gift from Prof. Eric Boder (University of Tennessee, Knoxville). Additionally, the C-terminal c-myc tag flanking NTL-1 was replaced with a 6xHis tag. Both plasmids were transformed into the EBY100 yeast strain and transformants were selected, and maintained in cell culture, using plates/media lacking Trp and Leu. The presence of both NTL-1 and CTL-1 as distinct cell surface fusions was confirmed by flow cytometry; cells were simultaneously labeled with a chicken-anti-c-myc antibody (Thermo Fisher; 1:250 dilution) and an anti-6xHis antibody (Thermo Fisher; 1:250 dilution), followed by secondary labeling with goat-anti-chicken DyLight 488 (ImmunoReagents; 1:200 dilution) and donkey anti-mouse DyLight 633 (Immunoreagents; 1:200 dilution).

2.3.5 Recombinant expression of BVL-1.

BVL-1 was sub-cloned into the pET28a vector using NheI and XhoI restriction sites. BVL-1 was recombinantly expressed in Rosetta™, The protein was expressed in 2 XYT media, a 5 ml overnight culture was used to inoculate one liter culture was grown to O.D. 0.7-0.8 then induced for 4 hours with 250 mM IPTG at 37°C with 250 RPM shaking. The culture was pelleted and then lysed in 40 mls of half strength PBS (5 mM NaPO4 and 68 mM NaCl) at pH=7 using a sonicator. The

protein was then run on a Biorad FPLC using a cation exchange colum (Biorad High S par # 732-4132). A linear gradient over 25 mls was used to elute the protein using PBS with 1 M NaCl as the elution buffer. To inactivate proteases and further purify BVL-1, product from the chromatography step was heated to 70 °C for 20 minutes, centrifuged to remove precipitates and filtered using a 0.22 μm filter. Note that without this heat treatment step, cleavage of the bivalent protein was observed during storage. The protein stock was then dialyzed to PBS.

2.3.6 Estimation of KD

K_D values were estimated using yeast surface titrations or using a competition assay with soluble binding protein, as described [24,53]. Briefly, for yeast surface titrations, $2.5x10^6$ yeast cells were labeled with varying concentrations of biotinylated lysozyme and anti-HA antibody, followed by secondary labeling and analysis by flow cytometry. Volume of labeling reaction was adjusted to ensure that lysozyme molecules are in ~ 10-fold excess over cell surface fusions. To maintain this ratio at low concentrations fewer cells need to be used therefore, for BVL-1, titration yeast displaying BVL-1 fusions were combined with un-induced yeast cells to facilitate ease of obtaining cell pellets by centrifugation Samples were incubated with nutation at 4 °C for time corresponding to estimates to reach 98% of equilibrium labeling. For estimation of K_D by soluble competition yeast cells displaying BVL-1were labeled with lysozyme at a concentration corresponding to ~ 80% of saturating concentration, from the yeast titration experiment (6.25 nM) and varying concentrations of soluble BVL-l, followed by secondary labeling and analysis by flow cytometry. Primary labeling was carried out in 500 uls using 6.25 nM biotinylated lysozyme, $4x10^6$ yeast cells and varying amounts of BVL-1. Overnight incubations were done with at least 16 hours of incubation time. Secondary labeling was done with SAPE (streptavidin phycoerythrin). A global least squares fit for the solution phase K_D was carried out using a mol balance to determine the concentration of the complex in solution. To do

this the free unbound lysozyme fraction was calculated using yeast titration based K_D and the fluorescence readout from cytometry.

2.3.7 Size Exclusion chromatography

A GE Superdex 75 10/300 GL column was equilibrated with 50mM HEPES, pH 7.4, 150mM NaCl, 6% Glycerol and 2mM DTT. The protein sample (200ul) was loaded into the column and eluted with 50mM HEPES, pH7.4, 150mM NaCl, 6% Glycerol and 2mM DTT at a flow rate of 0.5 ml/min. UV absorbance at 280 nm was monitored. Eluted 0.5 ml samples were collected using the fraction collector. The purified protein was identified by the resulting absorbance peak and the respective fractions were chosen and pooled. Bio-rad gel filtration standard (catalog# 151-1901) was used to calibrate the column for molecular weight determination.

2.3.8 Measurement of the Melting Temperature T_M.

Differential scanning fluorimetry, as described[57], was used to determine T_M for BVL-1. Samples containing 300 mg/mL BVL-1 and SYPRO Orange according to the manufacturers protocol (Life Technologies) were heated in a real-time PCR machine (Bio-Rad CFX96), at 1°C/min from 25°C to 99°C.

2.3.9 Expression and purification of GFP1-9, NTL-GFP11 and GFP10-CTL-1.

Construction of plasmid vectors containing GFP1-9, NTL-GFP11 and GFP10-CTL-1 was carried out as follows: a synthetic gene block encoding the sequence of GFP1-9 (Genewiz), and oligonucleotides encoding NTL-linker-GFP11, and GFP10-linker-K1 (IDT) were obtained; sequences are shown in **Appendix 1.** The GFP1-9 gene was amplified by PCR using oligonucleotide primers PG9f and PG9r, and cloned into vector pET-22b(+) between NdeI and BamHI restriction sites. The oligonucleotide encoding NTL1-linker-GFP11 was amplified by PCR using primers PN1f and PN1r, and cloned into pET-28b(+) between NcoI and XhoI

restriction sites. To generate the plasmid vector containing GFP10-CTL1, a synthetic oligonucleotide containing the sequence of GFP10-linker-K1 was amplified by PCR using primers PK1f and PK1r, and cloned into pET-28b(+) to generate pET-GFP10; K1 is a coiled peptide[51]. Subsequently, the CTL-1 sequence was amplified by PCR using primers PC1f and PC1r, and cloned into pET-GFP10 between NdeI and XhoI restriction sites; this step removes the K1 sequence from pET-GFP10.

GFP1-9 was produced in E. coli Rosetta cells and purified from inclusion bodies as described[51]. NTL-1-GFP11 and GFP10-CTL-1 were produced in E. coli Rosetta cells with N- and C-terminal 6xHis-tags, respectively, and purified by immobilized metal affinity chromatography (IMAC) using a Biorad column (catalog # 7324614). Purified proteins were dialyzed into TNG buffer (100 mM Tris-HCl, 150 mM NaCl, 10% glycerol), pH 7.4. Protein concentrations were quantified by BCA assay.

2.3.10 *In vitro* "mix and read" fluorescence reconstitution assays.

Samples containing varying concentrations of lysozyme, along with1 μM NTL-1-GFP11 and 1 μM GFP10-CTL-1, were prepared in TNG buffer containing 0.5% BSA. After equilibration for 20 minutes at room temperature, samples were transferred to wells of a 96-well plate pre-coated with 0.5% BSA. GFP1-9 was added to each well to a final concentration of 8 μM, and emission of the samples at 505 nm and 530, upon excitation at 483 nm, was recorded every 15 minutes for 16 hours.

2.4 Figures

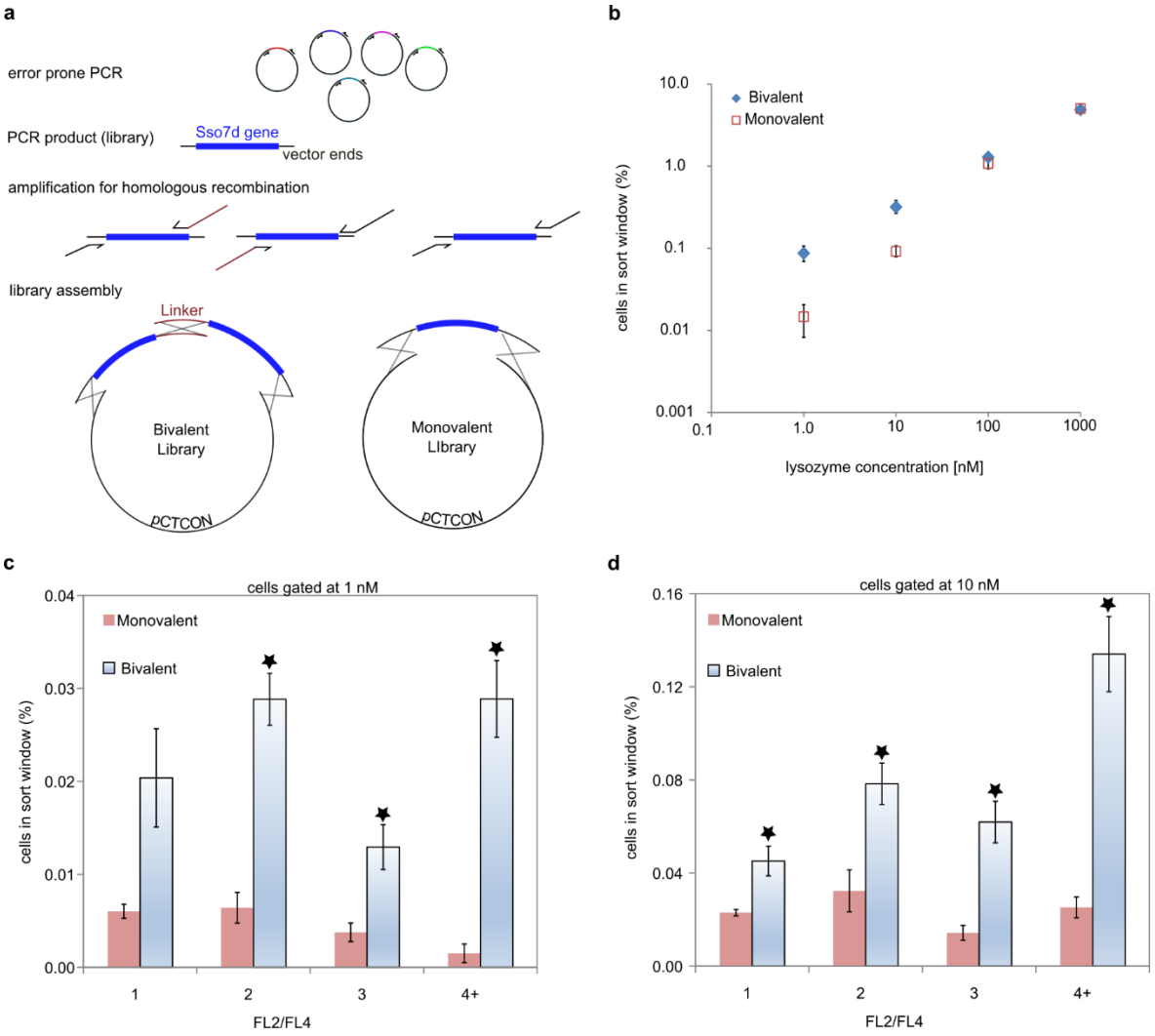


Figure 2.1: A bivalent library yields a higher frequency of putative high affinity clones.

(a) Scheme for construction of monovalent and bivalent libraries. **(b)** 10^5 cells from each library, labeled with varying concentrations of lysozyme and an anti-HA antibody, were analyzed using flow cytometry. The number of cells found in a conservatively drawn gate was assessed (see **Figure 9**) and is plotted as a fraction of the total number of cells analyzed. **(c), (d)** For gated cells, ratio of the fluorescence signal corresponding to lysozyme binding (FL2) to signal from anti-HA antibody binding

(FL4**)** was computed, and cells were placed in one of 5 bins (1-2, 2-3, 3-4, and 4+) based on the FL2/FL4 ratio. The largest fraction of either library was in the 0-1 bin and is not shown, this bin essentially accounts for those members of the library that do not bind lysozyme. The fraction of cells in each bin relative to total cells analyzed is shown for **(c)** 1 nM and **(d)** 10 nM lysozyme concentration. All error bars indicate standard error of the mean from triplicate experiments. * indicates $p<0.05$ based on a two-tailed t-test comparing results from monovalent and bivalent libraries.

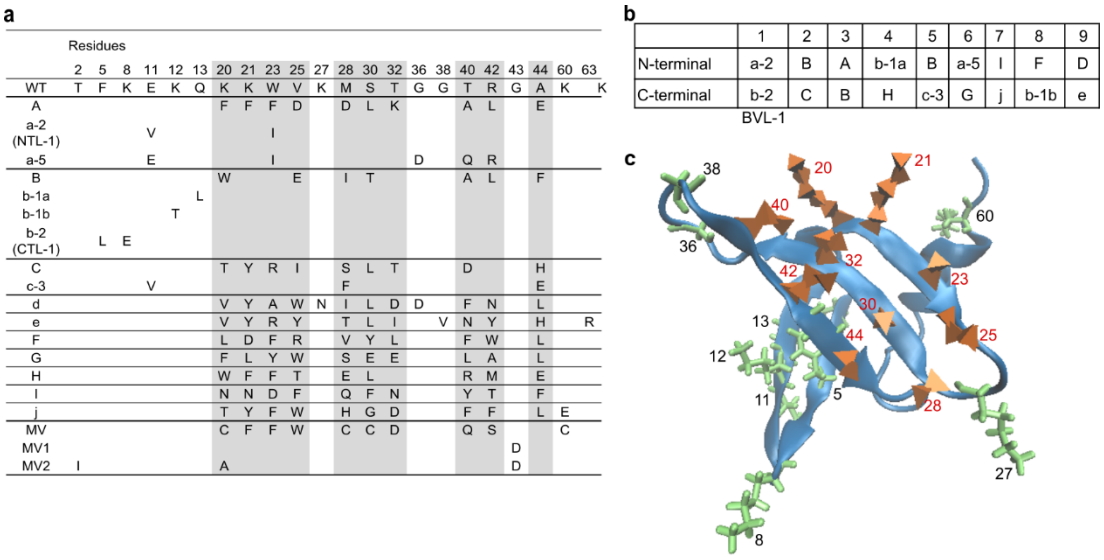
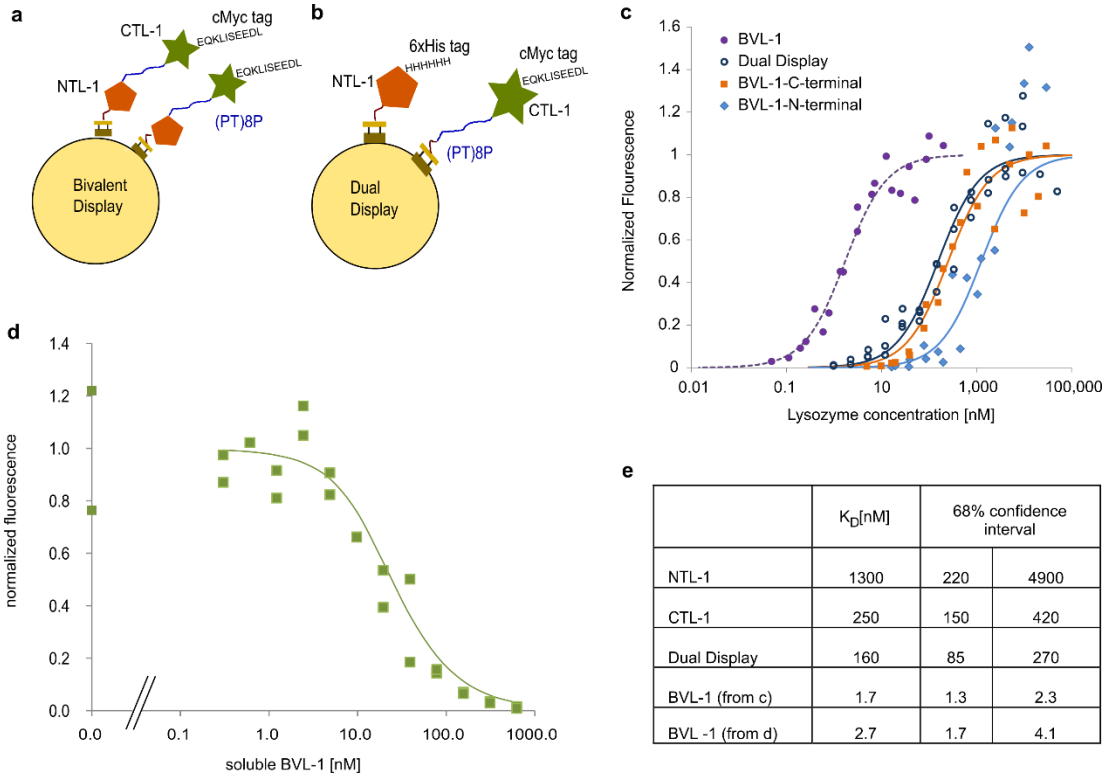


Figure 2.2: Analysis of nine clones from the highest affinity lysozyme ligands selected.

(a) Sequences of individual Sso7d binders that appear in nine bivalent ligands are shown (clones A-j). Residues randomized in the original Sso7d library are shaded gray. Clones with mutations only in these residues are denoted with upper case (e.g. A, B). Clones with similar mutations are designated by the same leading letter (e.g. A and a-1a); in such cases, only additional mutations relative the parent sequence are shown. MV is the monovalent binder with highest affinity previously obtained. MV1 and MV2 are monovalent clones obtained while screening the bivalent library. **(b)** Assembly of individual binders found in bivalent clones. **(c)** Location of mutations in the naïve library (orange)

and outside the 10 residues randomized in the original Sso7d library (green). Residues 2 and 43 cannot be seen in the figure as (2) is surface exposed and faces into the page and (43) faces towards the hydrophobic core.

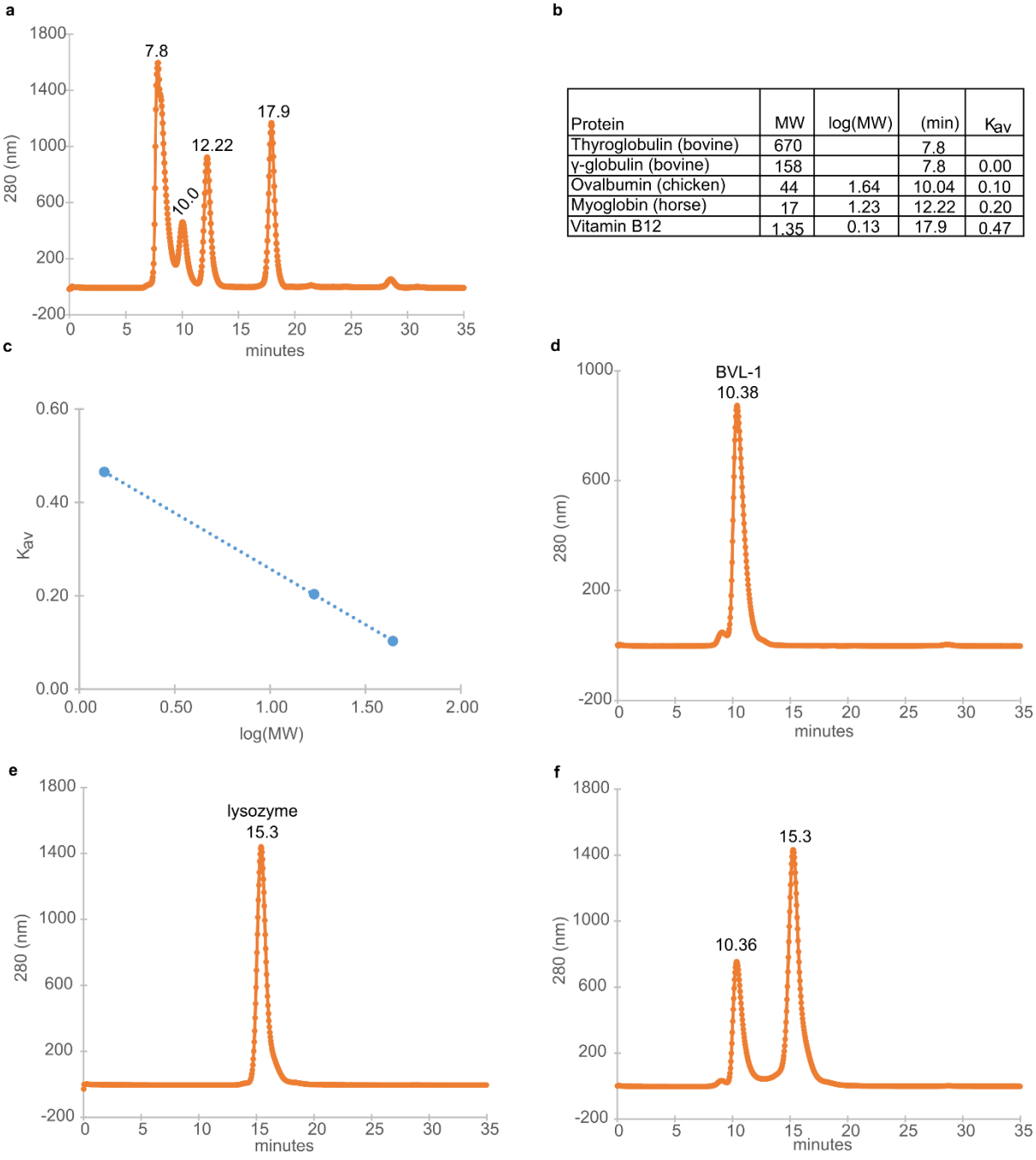


	K_D[nM]	68% confidence interval	
NTL-1	1300	220	4900
CTL-1	250	150	420
Dual Display	160	85	270
BVL-1 (from c)	1.7	1.3	2.3
BVL -1 (from d)	2.7	1.7	4.1

Figure 2.3: High affinity of bivalent binder BVL-1 is due to synergistic binding of low affinity NTL-1 and CTL-1 subunits.

(a) Schematic for yeast surface display of BVL-1. **(b)** Schematic for simultaneous yeast surface display of NTL-1 and CTL-1. **(c)** Yeast surface titrations for estimating K_D for BVL-1, NTL-1, and CTL-1, and the apparent K_D for cells simultaneously displaying NTL-1 and CTL-1 as distinct fusions. Data from two independent experiments is presented for each construct. Fluorescence data is normalized by the maximum fluorescence value for each construct, obtained from a global least squares fit. **(d)** Competition assay to estimate K_D of BVL-1 using soluble protein. Data from two independent

experiments is presented. **(e)** Estimates of K_D from data in (c) and (d) using a global least squares fit, assuming a 1:1 binding model.



Protein	MW	log(MW)	(min)	K_{av}
Thyroglobulin (bovine)	670		7.8	
γ-globulin (bovine)	158		7.8	0.00
Ovalbumin (chicken)	44	1.64	10.04	0.10
Myoglobin (horse)	17	1.23	12.22	0.20
Vitamin B12	1.35	0.13	17.9	0.47

Figure 2.4: Size exclusion chromatograms indicating BVL-1 is monomeric.

(a) The size exclusion column was calibrated using the Biorad Gel Filtration Standard, a homogenous mixture of the proteins shown (b). The UV absorbance reading at 280 (nm) was recorded over the course of the standard elution. (b) Table showing the proteins in the calibration curve, their molecular weights (MW) the residence time in minutes corresponding to the peak absorbance for each protein, the log of the MW and K_{av}. K_{av} is calculated as described in the GE Size Excluson Handbook. K_{av} and the log of the MW should be linearly correlated and this relationship can be used to determine the MW of unknown proteins. (c) The log of the molecular weight and the calculated K_{av}. (d) BVL-1 run by itself through the SEC (e) lysozyme run by itself (f) the BVL-1 and lysozyme complex were run through the column simultaneously after equilibration, excess lysozyme was used.

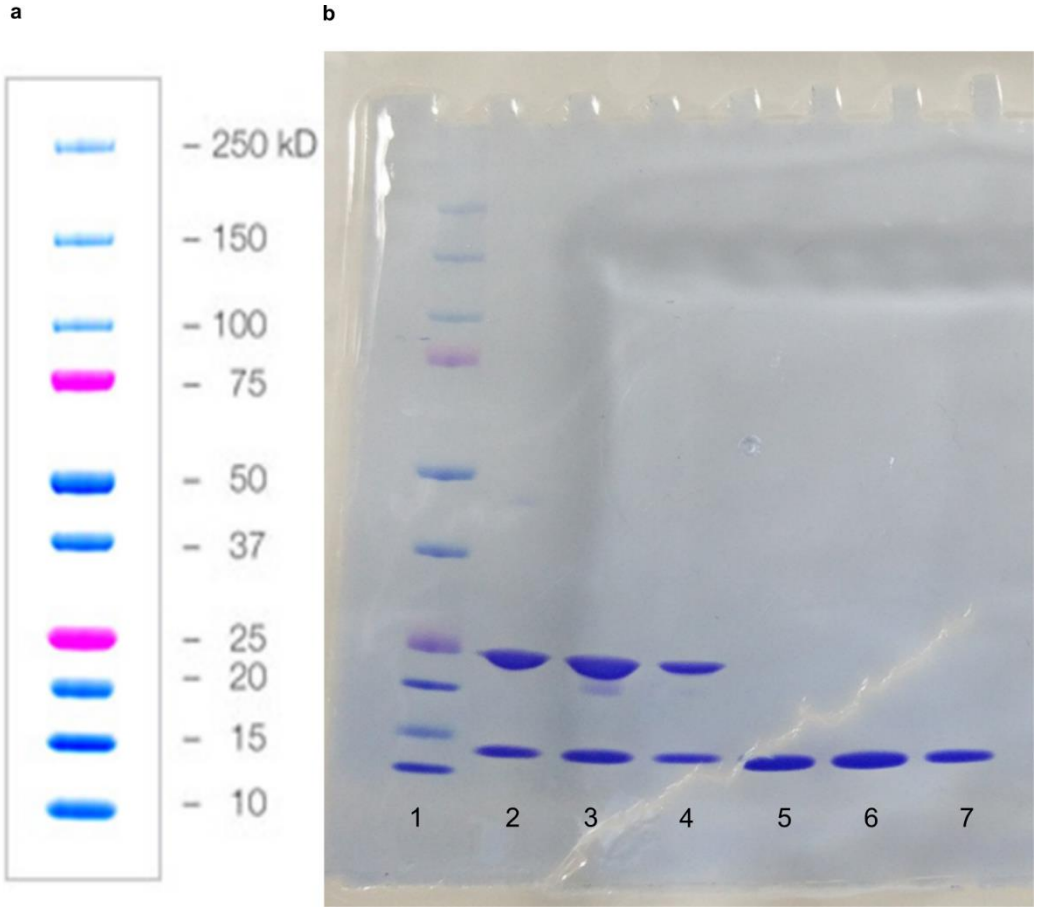


Figure 2.5: SDS page of elution samples of the BVL-1 and lysozyme complex from SEC

Non-reducing NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gel stained with Coomassie Brilliant Blue R-250 for elution samples taken from the lysozyme and BVL-1 SEC run shown in **Figure 2.4f**. (a) Is the Bio-Rad Precision Plus Protein Dual Color Standards molecular weight marker. (b) Lane 1 is the molecular weight standard. Lanes 2-4 were taken during the elution of the first peak and protein bands for both lysozyme and BVL-1 are observed. Lanes 5-7 were taken during the second peak elution peak and only lysozyme is observed.

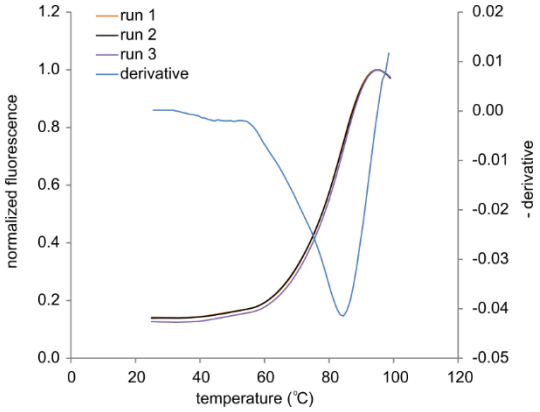


Figure 2.6: Differential scanning fluorimetry for estimation of T_M.

Baseline subtracted, normalized fluorescence signal is shown for three independent runs is shown. Also shown is the negative of the first derivative of the normalized fluorescence signal for one of the runs. The melting temperature, determined as the inflection point of the fluorescent signal, is 84.5°C for all three runs.

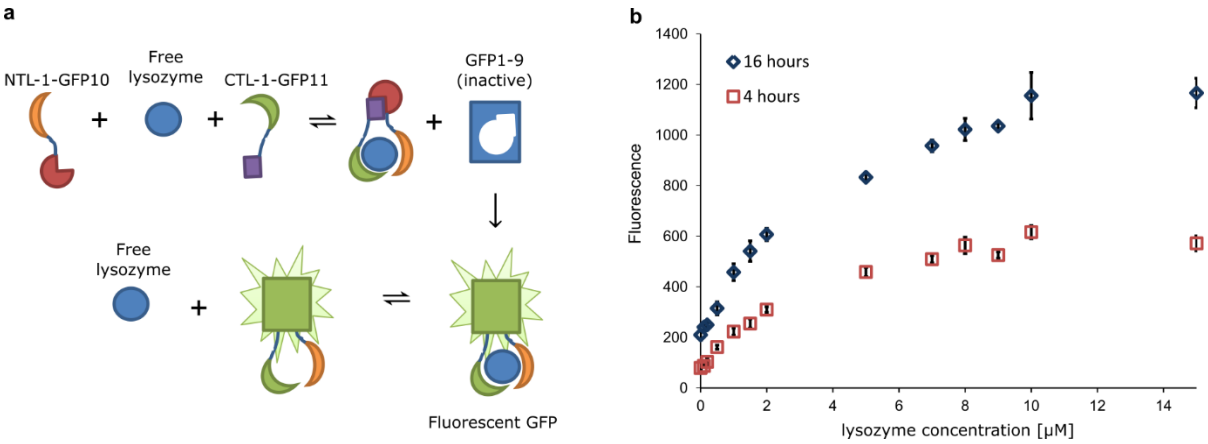


Figure 2.7: Binding of NTL-1 and CTL-1 to non-overlapping epitopes on the target can be exploited to design a "mix-and-read" assay for target quantification.

(a) Schematic for proposed "mix-and-read" assay. **(b)** Fluorescence readings, at two different times, for samples containing varying concentrations of lysozyme, along with1 μM NTL-1-GFP11 and 1 μM GFP10-CTL-1 are shown. Error bars indicate standard error of the mean for triplicate measurements.

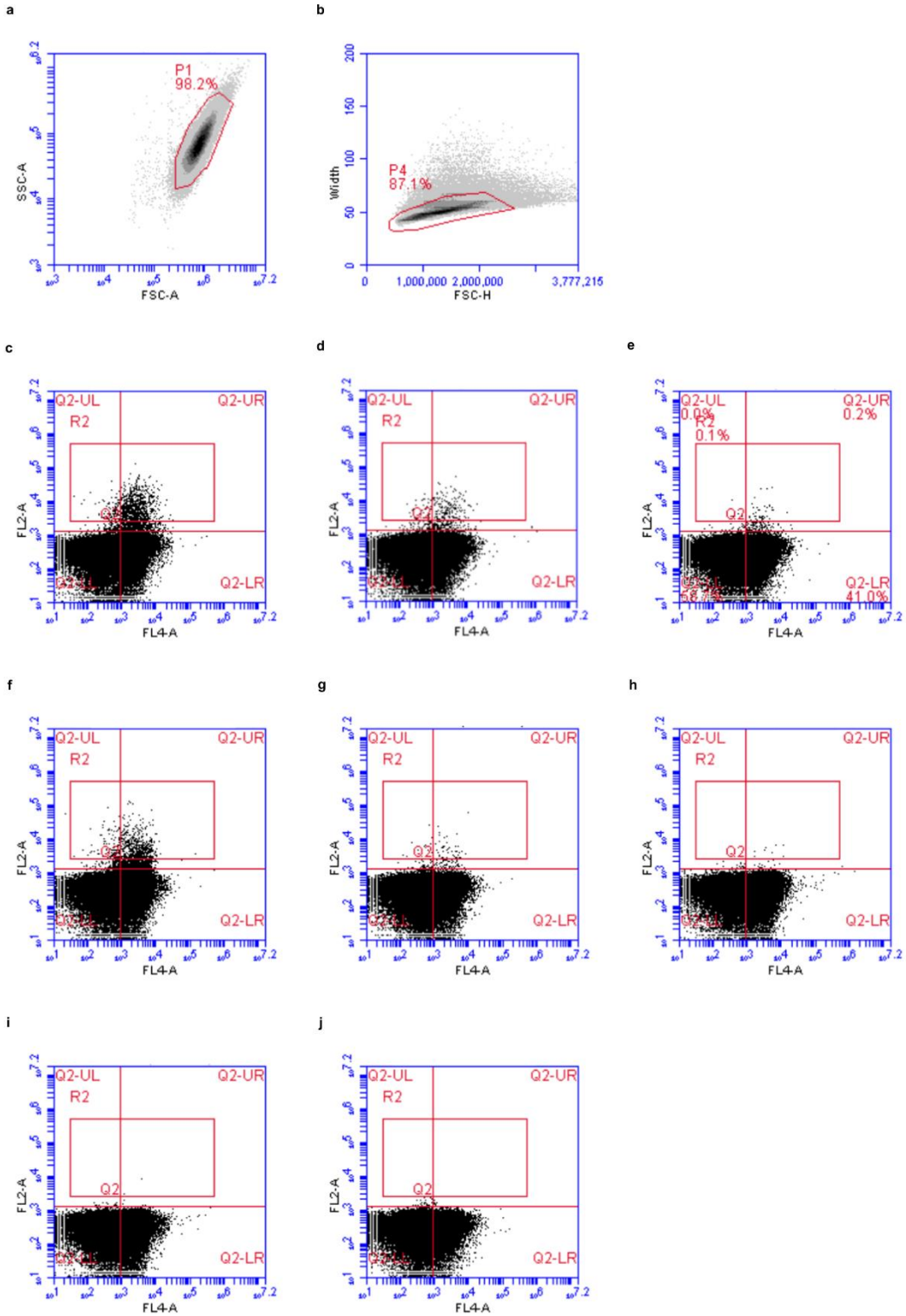

a
P1
98.2%
SSC-A
FSC-A
b
P4
87.1%
Width
FSC-H
0
1,000,000
2,000,000
3,777,215
c
d
e
f
g
h
i
j
Q2-UL
Q2-UR
R2
Q2
Q2-LR
FL2-A
FL4-A
0.0%
0.2%
0.1%
58.7%
41.0%

Figure 2.8: Comparison of monovalent and bivalent libraries by cytometry.

Yeast cells were labeled with biotinylated lysozyme and an anti-HA antibody, followed by secondary labeling with SA-PE (FL2) and DyLight 633 conjugated secondary antibody to detect the anti-HA antibody (FL4). **(a)** forward scatter and side scatter gating **(b)** gating to select singlet cells **(c)-(e)** bivalent library labeled at 100 nM, 10 nM and 1 nM lysozyme **(f)-(h)** monovalent library labeled at 100 nM, 10 nM and 1 nM lysozyme **(i)** no lysozyme control for bivalent library **(j)** no lysozyme control for monovalent library.

Vf	GTGGTGGTTCT GCTAGCATG
Vr	AGAAATAAGCTTTTGTTCGGATCC
NTLf	CTGCAGGCTAGTGGTGGTGGTGGTTCTGGTGGTGGTGGTTCTGGTGGTGGTGGTT CTGCTAGCATG
NTLr	CAGATGCTGGAAAAACAGAAAAAACCCGGGCCGACTCCTACGCCTACACCGACG CCCACACCAACTCCCACGCCGACACCTAGG
CTLf	CCCGGGCCGACTCCTACGCCTACACCGACGCCCACACCAACTCCCACGCCGACA CCTAGGAGCATGGCGACCGTG
CTLr	CTCGAGCTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGGATCC

Table 2.1: Primers used for library construction.

Primers are written 5' to 3'. Vf and Vr were forward and reverse primers used to generate the error prone library. NTLf and NTLr were used to amplify the library for insertion into an N-terminal positon and CTLf, CTLr were used to amplify the library for the C-terminal positon.

CHAPTER 3

FUNCTIONAL DIMERIC SSO7D LIGANDS AND EPITOPE ADDITION AS A MODE OF AFFINITY MATURATION IN BINDING INTRINSICALLY DISORDERED DOMAINS

3.1 Introduction

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions of proteins (IDRs) that compose over 40% of the human proteome. These proteins play important roles in cell signaling and transcription and their importance is highlighted by the fact they make up 80% of the proteins which have been identified as being responsible for neurodegenerative diseases and cancers[58,59]. They are defined by regions of 30 or more residues that do not form a stable tertiary structure. Their importance has led to the continued improvement in methods for identifying their presence in the proteome, modelling and measuring their dynamic behavior, as well as in searching for drugs to alter their behavior [58–60]. Targeting oncogenic transcription factors containing IDPs is an active area for small molecule drug discovery. One candidate that shows promise is a c-Myc inhibitor that prevents it from binding Max, a partner which makes c-Myc active[61]. The mechanism by which the drug works is to prevent the IDR of c-Myc from undergoing a change in conformation which is necessary to bind Max. Small molecules readily cross the cell membrane and therefore, in a therapeutic context, intracellular delivery does not represent the challenge that it does for larger engineered antibodies or protein affinity ligands. However, in the context of research tools for obtaining important biological information engineered protein affinity ligands have been successful[47,62,63]. The work herein expands the literature base on targeting a peptide sequence within an IDR using the non-antibody Sso7d scaffold. We make observations on the ability to pre-target a specific epitope within an IDR; and report findings made when affinity maturing pre-targeted ligands to a protein containing an IDR.

The model target we chose for the basis of our work is β-catenin an 85 kDa, 781 residue protein. This target was chosen as it can be recombinantly expressed, it is well-studied, and it contains N-terminal and C-terminal tails that are long, intrinsically disordered and as expected for an IDR, biologically important. B-catenin is a human protein involved in both cell to cell adhesion and in the

canonical Wnt signaling pathway[64]. Wnt signaling is critical to many cellular processes including proliferation, migration and stem cell differentiation[65,66]. In the absence of Wnt signaling, β-catenin binds E-cadherin and α-catenin near the cellular membrane, on the cytosolic side, and is involved in making cell to cell connections[67]. Cytoplasmic levels of β-catenin are maintained at very low levels because of a β-catenin destruction system[68]. However this changes when extracellular Wnt ligand is present. Secreted Wnt ligands bind to and co-localize the Frizzled and LRP5 or LPRP6 extracellular receptors. This co-localization results in the sequestration of the β-catenin destruction complex[69]. B-catenin then builds up in the cytoplasm and through a not well-understood mechanism translocates to the nucleus where it binds to DNA binding proteins and other transactivators to transcribe Wnt responsive genes[69]. B-catenin is composed of a 151 residue, intrinsically disordered N-terminus, a folded ARM domain composed of twelve alpha-helical repeat units (residues 151-666) and an intrinsically disordered C-terminus (666-781) [70]. The disordered C-terminus is a transactivation domain that binds to both transcriptional elements and chromatin modifying elements[66].

Two methods of targeting a peptide sequence within an IDR readily propose themselves. One approach would be to make a point mutation in the region of interest and use this as a target for negative selection. The resulting ligands are selected to bind at the site of the mutation in the non-mutated target. This is an often used approach in engineering affinity ligands to bind folded protein targets[71]. A mutation can typically easily be made to a surface exposed residue without substantially changing the target. However, such a mutation to an IDR could fundamentally change the behavior and dynamics of the IDR preventing the off-target IDR from directing the library towards a specific epitope as desired. An example of this is seen in the context of the target of interest for this work, the C-terminal domain of β-catenin. L774A and W776A mutations were shown to change the secondary structure of the C-terminus and because of this binding to the β-catenin co-factor TRRAP was lost[72].

A second approach would be to use a peptide itself as a targeted region in an IDR. This is a strategy used to make vaccines for targeted antibody generation by animals[73]. However, the resulting antibodies then need to be screened for functional binding to the target in the context applicable for the antibody. Affinity ligands engineered ex vivo have also been generated towards peptides which represent sequences of interest that appear in an IDR or IDP[46]. Importantly, in this example the peptide is the target, the ligand is never selected to bind the actual IDR of interest. Additionally, binding of the ligand to the actual IDR is typically not measured quantitatively, in its place selectivity measurements are typically done[46]. However, as the targeted peptide epitope may take on different behavior in the context of the entire IDR sequence, this could be an important issue in certain cases and could be a source of unexpected results.

Two methods were evaluated for generating ligands to intrinsically disordered regions, both of which relied upon peptide pre-targeted ligands generated in previous work[13]. In this previous study ligands which bound a twelve residue peptide from the C-terminal region of β-catenin (769-781) were selected from a naïve Sso7d library. This target will henceforth be referred to as βCP, β-catenin C-terminal Peptide. The highest affinity ligand selected from the naïve library towards βCP was termed Sso7d-β-catenin peptide and has a K_D of 2.8 μM. The ligand will be referred to in this work as M1. The first approach took as its starting point all ligands selected from this previous study capable of binding βCP through an avidity based interaction. Ligands that bound through an avid interaction were selected by incubation of the protein library with magnetic beads coated with βCP. As yeast displays 10^4-10^5 yeast display fusions, it is possible for multiple ligands on yeast to bind multiple targets on the biotin beads. This technique, which we refer to as an avidity MACS (magnetic activated cell sorting), selects for both very high and low affinity ligands simultaneously[28]. This population was then selected for ligands capable of binding full length β-catenin, recombinantly

expressed from E. coli through an avid interaction. The selection was carried out using biotin binder magnetic beads immobilized with biotinylated β-catenin protein. Subsequently, this pool of ligands was subject to a mutagenic PCR which covered the entire Sso7d gene and from this a second generation yeast surface display library was generated. The highest affinity members of this library to β-catenin were then selected using the avidity MACS technique and subsequent rounds of FACS (fluorescence activated cell sorting). One ligand from the selected highest affinity population, termed M2, was studied in detail. We refer to M2 as being pre-targeted to an epitope within an IDR and affinity matured, through mutagenic PCR, and subsequently selected for binding to the target protein containing the IDR, β-catenin.

The second approach taken to generate high affinity epitope specific, IDR binding ligands was to tether M1 via a flexible linker to a naïve Sso7d library. Each member of the library in this case will be composed of the N-terminal ligand (M1), a flexible proline threonine linker $(PT)_{15}P$ and the second ligand position would code for a library Sso7d library mutant. The linking region length was for this work is 31 residues in length. It was chosen based on previous linker modelling work that such a linker can provide synergistic linking at up to 70 angstroms apart.[41] As the folded region of β-catenin is 100 Å long (as determined by protein database data) this was felt to be a suitable length to bind nearly any site on β-catenin. The C-terminus because of its flexible nature adds additional length as well[74]. The motivation for this approach was to expand on the work described in Chapter 2 and evaluate the effect of selecting for a bivalent ligand in the context of an IDR target. This approach should generate ligands which bind to βCP and (1) potentially block cofactors that bind the βCP sequence (2) block a second epitope on β-catenin (3) in addition to blocking a second epitope it could alter the dynamic behavior of the C-terminus and affect interactions with other β-catenin binding partners. How these ligands affect the function of β-catenin's transcriptionally important C-terminus

are studies left for future work. Here, many important observations about the ligands generated through these methodologies have been made through in vitro assays and these are the focus of this chapter.

The highest affinity ligand chosen for further analysis using the tandem approach is termed T2, this is shorthand for tandem affinity ligand from a second generation library. As an affinity matured bivalent ligand to an IDR has not yet been generated to this author's knowledge, a T3 ligand was also generated. The name indicates that is a tandem ligand and generated after three different libraries have been generated. While profoundly high affinity was expected from T3 and observed to some extent, we were able to make other important observations into the nature of how a ligand with an IDR can bind a target with an IDR.

3.2 Results

3.2.1 L54W is a conserved mutation in the highest affinity ligands selected from the affinity matured monovalent library pre-targeted to the β-catenin C-terminal IDR.

The affinity matured monovalent library was generated using a nucleotide analogue based mutagenic PCR[75]. Primers coding for the N-terminal and C-terminal DNA sequences were used so that the entire Sso7d gene of the selected population would be subject to mutagenesis. The library diversity was estimated as $2x10^8$ and 1-3 residue mutations outside of the naïve library positions were observed when 10 clones form the library were sequenced. Library diversity is estimated by counting the number of plasmid transformed cells. This is accomplished by serial dilution and plating of transformed yeast on (-trytophan) plates prior to expansion of the library. Sequencing 10 members of the library revealed 9 of 10 clones were identical, neither of the clones were mutants of M1. The first selection from the library was carried out using the avidity MACS technique. β-catenin was immobilized on magnetic

beads and the library was incubated with the β-catenin displaying beads. After this selection four of five sequenced clones were found to be mutants of the M1 ligand. The highest affinity ligands were selected after several rounds of FACS, each selection was carried out at higher dilutions of biotinylated β-catenin. Six ligands were sequenced from the selected pool of ligands and they were all found to be mutants of M1. All six had a conserved L54W mutation; 2 had only this mutation; 2 others also contained K8R and Y33F mutations; and two others had the same 8R and 33F mutations as well as S17N and K48R. For this work we chose to study the single L54W mutant, referred to from here on as M2. **Figure 3.1** shows (**A**) the M1 sequence; (**B**) the mutations present in the pool of highest affinity ligands sequenced which differ from M1; (**C**) a PDB diagram display the location of the 54th residue position in Sso7d. The tryptophan in wild type Sso7d faces internally and is on the side opposite the original engineered binding interface on Sso7d. This interface is the location of the ten residues mutated in the naïve library[13]. These mutations are all on the 3 β-sheet surface which is also identified in **Figure 3.1C**.

3.2.2 The first generation tandem library assembled well and yielded high affinity ligands

The first generation tandem library was constructed by first creating a yeast surface display vector consisting of M1.2, followed by a (PT)15P linker and a stuffer fragment flanked by AvrII and BamhI restriction sites. The vector was digested by AvrII and BamhI, transformed with a naïve Sso7d library and assembled in transformed yeast cells through homologous recombination. The oligomer coding for the naïve Sso7d library had randomized nucleotides at the identical ten codon position as previously used[13]. However, NNK as opposed to NNN codons were used. Where N can be any nucleotide and K is restricted to being either the G or T nucleotide. The NNK codes for all 20 natural amino acids more efficiently, using only one stop codon and with 32 possible codons. Whereas the NNN library covers the same functional diversity with an additional two stop codons and uses 64

possible codons. Also for this library, the 5' end of the PCR product has homology to the linker region and not the yeast surface display vector. To enable precise recombination the codons used to code for the (PT)15P were highly diversified. **Figure 3.2** shows how the tandem library was assembled.

The transformed tandem library was estimated to have a diversity of $2x10^8$. Nine random library mutants sequenced contained the (PT)15P linker in frame and without mutations. Four of those nine mutants contained mutations only at the library positions of the C-terminal ligand; the remaining five mutants had stop codons in library positions or just following a library position. Sequence analysis also identified inadvertent mutations in each of the N-terminal M1 ligands: a K63 deletion and a K62Q mutation. These mutations are at the C-terminus of M1, K63 is the terminal residue in Sso7d. This mutant will be referred to as M1.2. It is shown that these mutations did not affect the ability of a tandem ligand selected in this library to bind the βCP sequence.

3.2.3 A diversified polypeptide chain was enriched during selections for β-catenin binding.

This first generation tandem library was initially sorted at 100 nM using biotinylated β-catenin. Cells which bound β-catenin at this concentration were collected after washing using magnetic biotin binding beads. This method is based on a previously published protocol[76]. The advantage of this magnetic bead capture step relative to FACS is the extreme high throughput and low cost. $2x10^9$ cells would take an unreasonable amount of time to sort whereas MACS can be readily done after a few hours or overnight incubation for a fraction of the cost and time. Typically an initial selection would be carried out using an avidity-based MACS wherein β-catenin is immobilized on magnetic beads prior to the selection. However, in this case each ligand already bears a β-catenin binding domain so an avidity based selection would be ineffective.

This 100 nM bead selected population was expanded, induced by incubation in galactose media for surface expression of the library ligand. Subsequently, FACS based on double labeling with 10 nM

biotinylated β-catenin and HA-tag expression was carried out. The HA tag was labelled with a primary rabbit antibody and then a goat anti rabbit conjugated with Dylight 633 secondary antibody; biotinylated β-catenin was labelled with streptavidin Phycoerythrin (PE). Sequencing nine members of the library at this point showed that only four were the expected M1.2 ligand, linker and a second Sso7d protein. The other five consisted of M1.2 but were truncated prior to the second ligand and contained a conserved peptide sequence. The unexpected peptide sequences seen in the linker region are shown in **Table 3.1**. The randomized polypeptide sequences likely arose from errors in the manner in which the yeast homologous recombination system combined the yeast display vector and library. The presence of the enriched sequence of ligand plus polypeptide linker would suggest that this would likely be a successful alternative approach for creating bivalent ligands. This alternative approach would use a short linker and a randomized peptide sequence, consisting of 8-10 sequential degenerate DNA codons. This randomized peptide would take the place of the Sso7d library that is used here.

To ensure that selection of a tandem Sso7d ligand a single label cMyc based FACS was subsequently carried out. The cMyc tag appears after the C-terminal ligand, therefore sorting for this selects for clones which are not truncated by a stop codon earlier in the sequence. This sort was carried out using a chicken anti-cMyc primary antibody and goat anti-chicken Dylight633 labelled secondary antibody. The best binders were then selected by FACS using a labeling concentration of 0.5 nM β-catenin. This final sort was gated based on double labeling of the HA tag, and β-catenin binding. The HA tag was again labelled using a rabbit anti-HA primary antibody and biotinylated β-catenin in a primary labeling incubation and a goat anti-rabbit Dylight 633 antibody and Streptavidin PE in a secondary incubation. Six of the clones from this sort were sequenced successfully, each was composed of the M1.2 ligand a linker region and a second Sso7d ligand. All but one clone contained four to five histidine mutations in the linker region. One clone contained a shorter (PT)8P linker, this mutant has

been termed T2 and was chosen for further characterization. **Table 3.2** shows the linker mutations observed as well as the library mutations selected in the C-terminal ligand for the six tandem ligands sequenced.

3.2.4 Affinity maturation of T2 yields a diversity of advantageous mutations.

T2 was affinity matured by randomly mutating the DNA for the N-terminal ligand (NTL) and linker sequence in a first PCR and in a second PCR the linker and C-terminal ligand (CTL) DNA sequence was mutated. The two libraries were generated in this case using a mutagenic PCR protocol based on Taq polymerase and an optimized buffer[55]. A T2 tandem library was assembled by the yeast homologous recombination system when the two libraries were transformed into electrocompetent yeast with the yeast display pCTCON vector. **Figure 3.3** depicts how this second generation tandem library was generated. The library diversity was estimated to be to be $1x10^8$. Library sequencing showed that 7 of 10 sequences contained two Sso7d mutants on either side of the linker with 0-3 amino acid mutations in either ligand; two library sequences had an additional linker segment in-between the two ligands and the other library member sequenced was a single Sso7d clone which was a mutant of the C-terminal ligand in the T2 construct.

The highest affinity clones from the mutated T2 library were selected after multiple rounds of selection and seven clones were sequenced. **Table 3.3** shows the residues that were mutated. The location where most mutations took place was in the region of residues 49-61 in the N-terminal ligand (NTL). All mutants had mutations in the NTL, two also contained linker mutations and four also contained mutations in the CTL. No clones in this pool contained the L54W previously observed in the M2 ligand. The T3 mutant (D49E, M57E and Q61R) depicted in **Figure 3.4** was used in this study for further characterization. For this mutant only the Q61R is surface exposed, the other two mutations are more inwardly directed.

The C-terminal ligand in both T2 and T3 are identical and therefore this ligand is referred to as 23CTL throughout the remainder of this work. **Table 3.4** shows the nomenclature for each of the mutants selected and characterized in this work, summarizes how they were engineered and the individual ligands which make up the T2 and T3 tandem ligands.

3.2.5 Selection of functional dimeric ligands

The oligomeric state of M1 was evaluated in previous work using a size exclusion column (SEC). The chromatogram indicated the ligand is predominantly observed to be in a multimeric state, likely a dimer[13]. Also, in the previous work the protein was found to be functional in soluble form in a binding inhibition experiment. Further studies were not done to characterize the behavior of the M1 ligand previously. During the course of this work it was observed that reducing agents had an effect on the binding of all the ligands selected in this work. Each of these ligands contain a version of the M1 protein, which has a cysteine residue at position 25. This mutation was selected from the naïve library and throughout the affinity maturation process for each ligand as well. Also, C25 is the only cysteine appearing in any of the Sso7d ligands presented in this chapter. This led us to hypothesize that dimerization was occurring by disulfide bond formation. To assess this dimerization reducing and non-reducing SDS page gels were run. **Figure 3.5** shows reducing and non-reducing SDS page gels for M2 and M1. **Figure 3.6** shows similar SDS page gels for T2, T3 and the T3-NTL clones. Both gels show that in the non-reduced state the majority of each of the ligands exists as disulfide bonded dimers. Each ligand was run at known dilutions so that densitometry could be used to quantify the fraction of each protein that exists as a dimer. Image J software was used to measure the band intensity of the dimeric and monomeric forms of each ligand in the non-reduced lanes. It was found that the Hill Equation could be used to define the relationship between band intensity and protein concentration. A global least squares approach was used to minimize the error between a calculated

band density using the Hill Equation and the band density measured by Image J software. The parameters used to minimize the least squares difference consisted of D_{max}, a maximum possible band density, a curve fitting parameter and the fraction of the ligand that appears as a dimer. More information can be found in Appendix A3 as to how this analysis was carried out.

The T3 ligand has additional minor bands which likely correspond to ligand which was cleaved by proteases. This can be concluded as the larger degradation product is also observed to be a disulfide induced dimer as it decreases in size in the reduced samples. The other portion of the protease product corresponds to the band which can be seen at 10 kDa, this likely represents a monomeric Sso7d protein. The fraction of these degradation bands for T3 were also quantified during the SDS page purity analysis. **Figure 3.7** shows the correlations between normalized band intensity and ligand concentration for each protein: (**A**) M1 (**B**) M2 (**C**) T2 (**D**) T3 (**E**) T3-NTL, this is the N-terminal ligand of T3 and (**F**) the concentration of dimer that was determined by the global least squares regression analysis using the Hill Equation. Normalized band density was determined by dividing the measured band density by the D_{max}. Throughout this work protein concentrations are determined based on the purity measured of the monomer or dimer as shown in the table in **Figure 3.7F**. Further purification was not carried out to separate the monomeric form of each protein from the dimeric form prior to ligand characterization. Based on the strong differential performance of the two forms of the ligand we concluded this was unnecessary.

Disulfide bond formation for all of the ligands proved difficult to prevent in the absence of a reducing agent. The standard protocol for removing undesired disulfide bonded proteins is to chemically reduce them and add 5 mM EDTA to chelate metal ions which will catalyze disulfide bond formation. Following a reduction treatment consisting of incubation with 1 mM TCEP for 30 minutes (reduction was validated by an SDS page gel) each ligand was dialyzed into PBS buffer

containing 5 mM EDTA. Dialysis took place over 24 hours and was expected to reduce the TCEP concentration to below 0.2 nM. Immediately following dialysis reducing and non-reducing protein gels were run and the proteins were found to be largely in the dimeric state, as seen in **Figures 3.5** and **3.6**. Furthermore, the extent of dimerization did not change over the course of months after the reducing agent was dialyzed from the solution.

To assess whether the reduced or disulfide bonded form is functional the Blitz™ instrument made by Forte Bio was used. Blitz™ uses biolayer interferometry (BLI) to measure the association and dissociation rates of one protein to another protein of interest that is first immobilized onto a Blitz sensor. For these experiments full length biotinylated β-catenin was immobilized onto a streptavidin coated Blitz™ sensor, subsequently the association rate was measured in the presence of each ligand at values at least 5 times greater than the K_D's measured for each ligand in this work. The ligands were reduced in two different assays using 5 mM DTT and 1 mM TCEP each individually. Using two different reducing agents provides strong evidence that observed differences in affinity are not due to inhibition caused by the reducing agent. The results of this assay for each of the ligands are shown in **Figure 3.8.** The M2, T2 andT3 ligands clearly show that the non-reduced ligand gives higher binding signal and a substantially lower off-rate in the non-reduced state, indicating that the dimeric form is in fact the functional form of the ligand and not an undesired artifact. Negligible binding signal was observed for each of the ligands binding the blank sensor, which does not have β-catenin immobilized onto its surface.

The small difference in apparent affinity of reduced and non-reduced M1 motivated us to carry out a more rigorous assessment of the binding behavior of this ligand to determine which ligand has higher affinity for β-catenin. This data and analysis is shown in Appendix A2, **Figures A2.3** through **A2.6**. **Figure 3.9** summarizes the differences between the (A) dissociation rate k_{off} (1/s) and

(B) K_Ds (μM) observed when comparing the two forms of M1 based on only a local, non-global analysis. Local analysis indicates that kinetic and equilibrium binding constants are calculated based on data taken for each BLI run individually. Global analysis by contrast would result from simultaneous regression analysis to fit one value of k_a and k_{off} that best fits all the data for each form of the ligand. This local analysis appears to be especially necessary for the non-reduced form of M1 as both the apparent k_{off} and apparent K_D for this ligand appear to be functions of the concentration at which the ligand is loaded at**.** We hypothesize that this behavior is observed as dimeric M1 is able to bind two different epitopes on β-catenin. However this two epitope binding mode is non-dominant and only becomes visible at lower loading concentrations. At higher concentrations the two-epitope binding appears to be masked by a dominant single epitope which is bound. **Figure 3.9A** and (**B**) show that even at higher loading concentrations non-reduced M1 has both a lower dissociation rate (k_{off}) and K_D then the reduced form. The highest off rate- measured for non-reduced M1 is 0.4 ± 0.0062 (1/s) (the standard error (SE) is shown. Whereas the lowest off-rate for reduced M1 measured is 0.87 ± 0.007 (1/s). A similar difference is measured between the lowest K_D for the reduced form (3.6 μM ± 0.043 μM SE) versus the highest K_D of the non-reduced form measured is (1.5 μM ± 0.48 μM SE).

The BLI data shown in **Figure 3.8A** for the 1mM TCEP treated M2 ligand shows some degree of association; the DTT condition shows this trend as well however the magnitude of the interferometry signal is reduced. BLI data was recorded at higher reduced concentrations to determine if the K_D of the reduced form of M2 could be determined. **Figure A2.6** shows that depending on whether the data is analyzed based on local or global fits the reduced form of the ligand has a K_D between 100-250 nM. Whereas the K_D of the dimeric form of M2 determined by BLI analysis is about two orders of magnitude lower. Appendix 2, **Figure A2.5** shows BLI raw data and analysis based on

loading dimeric M2 at 5.6 nM and 11.6 nM. Based on local fits of one to one binding models the non-reduced form of the ligand has K_D of 1.7 nM (standard error of 7.8e-3 nM) or 2.2 nM (standard error of 6.7e-3 nM).

Experiments done to determine the binding kinetics and K_D of protein interactions using the BLI are limited by the fact that the interaction measured is to a surface bound species. As the ligands generated in this work bind multiple epitopes (shown later) this enables two to one binding to occur. This would cause the K_D measurement to define the relationship between one soluble phase protein and possibly two distinct surface bound β-catenin proteins. Whereas, competition based affinity measurements enable the inference of solution phase binding to be made. In solution β-catenin proteins are not placed in close proximity as they are when surface immobilized. This allows the interaction to be interpreted as one ligand binding one β-catenin protein.

3.2.6 Measurement of soluble phase K_Ds

A two-step approach was used to evaluate the K_D of each ligand to full length β-catenin. Yeast titration curves were first carried out to determine a titration based K_D. These are shown in **Figure 3.10**, panes, (**A)**, (**C**), (**E**) for ligands M2, T2 and T3 respectively. These curves are generated by labeling yeast displaying each respective ligand with a range of β-catenin concentrations spanning three logs and quantifying the relative amount of β-catenin bound at each concentration. This is a cytometry based protocol and the presence of the biotinylated β-catenin is quantified using SAPE (streptavidin phycoerythrin). A K_D for each ligand was determined using a one to one isothermal binding model and a global least squares approach as previously described[22]. The least squares fit is based on data taken for two independent series of yeast titration measurements for each ligand. For each dataset the maximum fluorescent signal is fit along with a single K_D value.

Data recorded to determine a competition based K_D's are shown in **Figure 3.10B**, **D** and **F**, again for M2, T2 and T3 respectively. Competitive binding experiments were carried out by choosing a β-catenin concentration within a sensitive range on the titration based K_D curve. Each sample is allowed to equilibrate with a given amount of the same yeast display ligand also added to the solution. The ligand binds the β-catenin target in a manner defined by a solution phase K_D. This solution phase binding prevents β-catenin from reporting on the surface of the yeast cell and being reported as fluorescence in the cytometry measurement. The fluorescent signal recorded on the yeast cell is correlated to a solution phase β-catenin concentration. The relationship between soluble concentration and the yeast surface display signal is governed by the titration based yeast K_D. The solution phase β-catenin concentration is used as part of a mol balance to determine equilibrium concentration of the ligand and the solution phase K_D. The concentration of β-catenin at which each competition experiment was done is shown in **Figure 3.10**. Again, for each ligand two separate and independent competition experiments were run. The K_D values and 68% confidence intervals for the titration, competition and domain specific K_D's calculated for each ligand are shown in **Table 3.5**.

An important point for the calculation of the titration based K_D is that data points at high concentration were excluded from the global least squares fit. The entire data set included the fit excluded points are shown in **Figure 3.11**. High concentration being over about 90% of the fractional binding was determined by optimizing the K_D fit for accuracy on the low end of the titration curve. There are two reasons why the high concentration data was omitted from the K_D calculation (1) the titration curve serves as the calibration curve for the competition based K_D, which is crucial to validating the behavior of the ligand in solution. Accuracy of the calibration curve is only required below β-catenin concentration at which the competition is carried out. (2) As will be shown later in the results each ligand binds two or more epitopes, therefore the full K_D curve could contain multiple

inflection points. Synergistic binding of both ligands to one β-catenin will have a given K_D however each β-catenin binding site on each ligand has its own K_D. Therefore, it is possible that at higher concentration each yeast displayed ligand will bind 2 β-catenin proteins. Also, the selected ligands have higher affinity when in a disulfide bonded dimeric form and it is likely that the dimeric form predominantly exists on the surface of yeast. However, monomeric versions may also be present. At higher β-catenin loading concentrations, the lower affinity monomeric versions of the ligands, could also add to the total amount of β-catenin each yeast cell can bind.

The difference between the titration based K_D and the competition K_D for each ligand (M2, T2 and T3) are within a factor of at most 2.6 (values are shown in **Table 3.5**). Protein concentrations for the titration K_D are based on the concentration of β-catenin alone. While for the competition experiment the K_D is calculated based on the dimer fraction of the recombinanty produced and purified ligand as well as the β-catenin concentration used.

3.2.7 Differential binding observed for ligands to βCP peptide sequence alone and in the context of the C-terminal IDR

Table 3.5 shows the K_D values determined by titration and soluble competition of M2 (7.6 nM and 5.4 nM) respectively. The K_D to βCP for M1 is 2.8 μM, almost 1000 fold higher. The L54W mutation in M2 would therefore appear to have had a dramatic effect on how M2 binds βCP. The K_D of M2 to the βCP sequence was measured to determine to what extent the affinity of M2 is attributable to the binding of βCP. Yeast displaying M2 was again used and a titration curve was carried out using various concentrations of biotinylated βCP. SAPE was used as the secondary reagent. The amount of peptide binding to the yeast was quantified using cytometry. The K_D's for T2 and T3 were also determined for the βCP target in the same manner. The data collected is shown in **Figure 3.12** along with the global least squares curves for fitting the K_D using a one to one binding

model. The K_D values determined for each ligand to βCP along with 68% confidence intervals are shown in **Table 3.5**. Based on previously published data the K_D for M1 is in exact agreement with T2 and T3. Remarkably, for M2 the K_D to βCP decreased by approximately an order of magnitude to 26 μM with a 68% confidence interval of 18-46 μM. It should be noted that the fit to M2 is based on concentrations reaching only about 50% of what the maximum signal is expected to be. Labeling would have to done at ~250 μM to reach 90% of saturation. This is an exceedingly high concentration for this assay and background binding of βCP to yeast was expected and therefore binding at these high concentrations was not measured. However, given the range of the 68% confidence interval measured it is apparent that the affinity of M2 has dropped relative to M1.

To further probe the mechanism by which the affinity maturation occurred two versions of the C-terminal IDR were cloned into yeast display constructs. These constructs were the entire C-terminus which follows directly after the 12th ARM domain (residues 666-781); and (2) a truncated C-terminal IDR (residues 666-769), for which the last 12 residues representing the originally targeted βCP sequence are not present. **Figure 3.13A** shows the expression level of each construct in the yeast display format. Both constructs expressed on the surface of yeast extremely well despite each being over 100 residues in length and without a well-defined tertiary structure. Each of the recombinantly expressed and purified ligands, M1, M2, T2, T3 as well as T3's N-terminal ligand (T3-NTL), were N-terminally biotinylated and screened for binding to both C-terminal IDRs each at concentrations between 100 nM to 2 μM. Concentrations are based on the molecular weight and purity of the dimeric form of the ligand in each of the purified samples. SAPE was used as the secondary reagent. **Figure 3.13B** shows that M2 (900 nM) binds strongly and specifically to the full length C-terminus and not at all to the truncated C-terminus. No signal was observed for binding of the SAPE reagent to either IDR, also shown. Interestingly, almost no cytometry signal is observed by either M1 (at 220nM to

1700 nM) or T2 (140 nM-1,100 nM) to the C-terminal IDR. **Figure 3.13C** shows the cytometry signal measured for M1 and T2 as normalized by the signal observed for M2 at 900 nM. **Figure 3.13C** is used to indicate the trend in binding signal as concentration is increased and to show the dramatically decreased signal relative to ligands that do bind the C-terminal IDR strongly. T2 does show a trend of differential binding to the full length C-terminus relative to the truncated C-terminus (also shown in **3.11C**). Less signal was observed for M1 and only the binding to full length C-terminus is shown. **Figure 3.13D** shows cytometry histograms of the binding of M1 and T2 at the highest concentrations tested (1,700 and 1,110 nM) to the full C-terminal IDR. **Figure 3.13E** shows that the T3 ligand (650 nM) does again show strong specific binding to the C-terminus with almost no signal observed for binding to the truncated C-terminal IDR. As T3 differs from T2 only by mutations in the N-terminal ligand (NTL) it was determined if the T3-NTL expressed and purified by itself, without the linker or C-terminal ligand behaves similarly. **Figure 3.13E** also shows that T3-NTL (at 700 nM) is observed to give very little binding to the C-terminal construct. This indicates that the mutations in the T3-NTL enable the ligand to bind the C-terminus with substantially higher affinity when fused to the linker and the T3 C-terminal ligand.

To fully characterize the interaction of M2 and T3 to β-catenin yeast surface display based K_D curves were again generated. Yeast displaying the full length C-terminal construct (666-781) was titrated with either of the biotinylated ligands and SAPE signal from secondary labeling was measured using cytometry. **Figure 3.14A** and **3.14B** show the data collected and K_D fit. Again a one to one binding model is used and a global least squares fit is used. The K_D's of M2 and T3 to the C-terminus are 150 nM and 110 nM, confidence intervals for both K_D's are shown in **Table 3.5**. Hook effects were observed for both ligands above a concentration of about 1 μM, the data points above 800 nM were excluded from the global least squares fit because of this strong hook affect.

3.2.8 ARM Domain Binding

The measured K_D for M2 and T3 to the C-terminal IDR are 30 and over 100 times greater than the respective K_D's measured to full length β-catenin. Therefore, we hypothesized that the ligand gained an additional epitope to another domain on β-catenin. To evaluate this, the N-terminal domain (residues 1-150) and ARM domain (residues 151-666) of β-catenin were cloned into the yeast display construct individually and binding by soluble ligand was again measured.

Quality controls were done to ensure that the ARM domain was well displayed and folded on the yeast surface. **Figure 3.15A** shows the yeast displayed expression level of the ARM domain, which is seen to be adequate albeit lower than that observed for the C-terminal IDRs. T-cell Factor Four (TCF) is a canonical binding partner of β-catenin and the TCF 1-65 amino acid sequence is an intrinsically disordered domain that binds across the majority of the ARM domain[70]. Therefore, it was used as a screen to determine that the ARMS had indeed expressed well and were properly folded. TCF was expressed and purified as a GST fusion. The GST portion of the fusion has two cysteines that are accessible for biotinylation, therefore a maleamide based coupling to biotin was carried out to conjugate these residues. This enabled cytometric determination of the GST fusion by SAPE. **Figure 3.15B** shows labeling of ARM displaying yeast by the biotinylated TCF construct. To ensure TCF+GST fusion was not spuriously binding yeast it was also screened for binding to a β-catenin N-terminal construct expressing residues 50-100. **Figure 3.15C** shows that TCF+GST specifically labels the ARM domain and not NT-50-100. These measurements validated our use of the ARM domain. **Figure 3.15D** shows the signal generated by the N-terminally biotinylated ligands M1, M2, T2 and T3-NTL at (900 nM, 900 nM, 550 nM and 900 nM). All of the ligands including M1 give an above background signal to the ARM domain.

K_Ds for all of the ligands to the ARM domain were again determined by titration of the yeast displayed ARM domain with each of the ligands respectively. The K_D's were fit to a one to one binding model using a global least squares fit. Cytometry data recording SAPE signal from secondary labeling after primary labeling with the ligands are shown in the panels of **Figure 3.16**for M2 (A), T3 and T2 (B) and T3-NTL (C) M1 (D). **Figure 3.16E** is used to illustrate that the maximum fluorescent signals recorded for M1 is significantly lower than for T2 or M2. The M1 fluorescent signal cytometry readings in **Figure 3.16E** were normalized by the average of the F_{max} normalization constants determined for T2 and CFL.2. **Figure 3.16E** serves to illustrate that the M1 interaction with the ARM domain maybe non-specific. K_D's and 68% confidence intervals for all of the ARM binding titration curves are shown in Table 3.5.

3.2.9 No binding to β-catenin N-terminal IDR

Binding to the N-terminal IDR was also assessed. **Figure 3.17A** shows the yeast displayed expression levels of two N-terminal constructs: (1) the first 50 residues of β-catenin NT-1-50 and (2) the entire N-terminus which is 150 residues NT-1-150. N-terminal 1-50, 50-100 and 100-150 were all cloned into the yeast display system as it was unclear how well the entire 150 residue N-terminus would express on the surface of yeast. However, sufficient expression is observed for the NT-1-150 construct to be used to determine any binding interaction between it and the other ligands. **Figure 3.17B** shows that biotinylated M2 at 900 nM is not observed to bind the N-terminus. Binding of M1 at 1600 nM, T2 at 1100 nM and T3-NT at 1800 nM were also screened for binding against the NT-1-150 domain. No binding of any of the ligands to the N-terminal construct was observed.

3.2.10 Dimerization of T2 and T3 is required for 23CTL functionality.

The same C-terminal ligand (CTL) is present in the T2 and T3 constructs, we refer to this ligand as 23CTL. A consequence of both T2 and T3 being tandem disulfide bonded homodimers is that 23CTL appears twice in the dimeric (non-reduced) version of each construct. Two plausible binding mechanisms for 23CTL binding β-catenin are possible (1) each 23CTL present in T2 and T3 has one common epitope either 23CTL can bind alternatively or perhaps are oriented such that only one of the two ligands can bind the epitope and the second 23CTL serves no purpose. This mechanism would imply that 23CTL is fully functional when expressed individually, not as part of either T2 or T3 construct. (2) Each 23CTL binds distinct epitopes on β-catenin. This mechanism would imply that because of the avidity affect the apparent affinity of 23CTL within the tandem homodimer is much higher than when the ligand is expressed by itself. The **BLI** data in **Figure 3.18A** shows the reduced form of T2 shows negligible binding at 500 nM. This is inconsistent with the second proposed binding model. The T2 affinity for the ARM domain is 85 nM and this interaction is driven by the binding of 23CTL. Given this fact, a BLI signal corresponding to a ligand loaded at 5x its K_D was expected if 23CTL was fully functional in the reduced form. This is not observed here. The equivalent concentration of non-reduced T2 (138 nM) is also shown in **Figure 3.18A**. The equivalent concentration is determined by using the same dilution of the protein stock of T2 that was used to make the 500 nM reduced concentration. The 138 nM dimer condition is shown to give high association signal and slow dissociation of the ligand.

Figure 3.18B shows a second experiment carried out to measure the extent of binding of soluble 23CTL to yeast displayed ARM domain. 23CTL was cloned into an expression vector, purified recombinantly and then assessed for binding to the yeast displaying the ARM domain by cytometry. In this case the 6x Histidine tag was labelled with anti-his antibody and donkey anti-mouse Dylight633 conjugated secondary antibody was used to measure the extent of binding. **Figure 3.18B** shows that at 1 µM of biotinylated 23CTL very little binding is observed, only 18% of that seen for T2. Roughly equivalent levels of ARM domain binding between T2 and 23CTL would be expected if 23CTL behaved in the monomeric form as it does as part of the T2 tandem dimer.

BLI based kinetic analysis of the non-reduced T2 ligand is shown in Appendix 2 **Figure A2.7**. Local analysis of non-reduced T2 at 138 and 28 nM result in, 2.7 nM and 1.5 nM , K_D estimates when fit to a one to one binding model. The figure also indicates that the binding model does not exactly fit the association and dissociation data. This is in accord with the actual binding mechanism corresponding to a two or three epitope interaction.

3.3 Discussion

The intent of this study was to generate high affinity ligands to an inherently disordered domain using a peptide to pre-target ligands to a sequence within an IDR. However, we found we must first evaluate the unexpected result of having selected for Sso7d ligands which when recombinantly expressed and purified appear as disulfide bonded homodimers. The M1, M2 and T3-NTL ligands are seen to be ~90% dimeric and the tandem ligands T2 and T3 55% and 65% dimeric by mass. Given that C25 was selected for from the naïve library and throughout the affinity maturation process for ligands

in this work, we hypothesized that the self-affinity of M1, M2 and T3-NTL was driving association and disulfide bond formation inside the yeast cell. This would result in the expression of the disulfide bonded homodimer on the yeast surface. There are literature examples of the yeast surface display system displaying disulfide bonded hetero and homodimeric proteins[26] This self-association on the surface of yeast is however difficult to measure directly, usually a proxy which indicates dimerization or multimerization has taken place is used[26,77]. For example, in the streptavidin study referenced, it was inferred that yeast can display functional tetrameric streptavidin as the extremely high affinity interaction of the tetramer for biotin was observed[77]. We took a similar approach to determining whether the recombinantly expressed and purified ligands are self-associating and dimerizing into their functional forms; or alternatively, if the disulfide bond formation in the recombinant state is an unwanted artifact which results in non-productive ligands. As the yeast display fusions are themselves tethered to the yeast surface through disulfide bonds it is not possible to assess how the reduced ligand on the surface of yeast functions directly. We therefore measured how reduced and non-reduced forms of each ligand behaved in the BLI format for measuring the association, dissociation and K_D of protein affinity ligands. We then compared BLI data to the K_D of the yeast displayed ligand by competitive binding and titration based assays.

Two binding modes are possible in the BLI system for ligands which bind two epitopes on the same protein. **Figure 3.19A** shows a mode whereby one dimeric ligand binds to two different surface immobilized β-catenin proteins. Alternatively, **Figure 3.19B** illustrates an example whereby the dimeric ligand binds two epitopes on a single β-catenin protein. **Figures 3.19C** and **D** illustrate identical modes of interaction for reduced, monomeric forms of a ligand. The BLI data recorded indicates that monomeric forms of M2, T2 and T3 are not capable of binding surface immobilized β-catenin through an interaction characterized by a low K_D (<10 nM). This is evidenced by the near negligible association

observed in the BLI experiment at concentrations of 50 nM or higher (**Figure 3.8**). This indicates the binding modes depicted in **Figure 3.19C** and **3.19D** are not viable for the reduced ligands.

The BLI data does show that the non-reduced version of M2, T2 and T3 all do bind with high affinity to surface immobilized β-catenin. This indicates that either or both modes of binding as depicted in **Figure 3.19A** or **B** are possible and yield high affinity interactions. The competitive binding yeast surface display experiment also shows that all three ligands can bind in the soluble phase in a high affinity manner (**Figure 3.10**). In solution β-catenin proteins are not artificially brought into proximity of each other as they are when surface immobilized. Therefore, in solution, one ligand protein can readily interact with only one β-catenin protein.

The results of the competitive binding and BLI experiments taken together are indicative in the dimeric form of M2, T2 and T3 being the functional form of the ligand. However, one explanation of the data presented is that the reduced form of the ligands may be unable to bind β-catenin in a high affinity manner in the BLI format due to the surface immobilization of β-catenin. Immobilization may result in a restriction in β-catenin's conformational freedom. This could prevent simultaneously engagement of ligands which target multiple epitopes from binding β-catenin when in the monomeric form. By this reasoning the high affinity of the dimer observed in the non-reduced BLI experiment could also be attributed to binding of two different surface immobilized β-catenin proteins. Calculating the competition based K_D for the monomeric form of M2 indicates this is an unlikely scenario. The least squares fit to the competitive binding data considering only the monomeric form of M2 in the non-reduced sample results in a K_D of 0.76 nM (the 68% confidence interval is between 0.3 nM and 1.5 nM). This K_D is a value which is 10 fold lower than the yeast titration based K_D observed (the 68% confidence interval for the titration based K_D is 5.7 to 10 nM). The K_D considering only the dimeric fraction of M2 in the competitive binding experiment is 5.4 nM (the 68% confidence interval is between

3.9 nM to 7.3 nM) and corresponds much more closely to the yeast titration based K_D. Also, the least squares error calculated for the competitive binding K_D is 2.7 times lower when calculating the K_D of the dimer relative to that of the monomer. The closer agreement between the competition based K_D calculated for the dimeric form of the ligand and the titration based K_D indicates the dimer is the functional form of the ligand. Additionally, the hypothesis that β-catenin has limited dynamic motion on the BLI sensor is unlikely. β-catenin is not non-specifically absorbed onto the sensor surface. Rather immobilization occurs through a flexible chemical linker between biotin and β-catenin. This tethered form of immobilization likely enables β-catenin to retain sufficient conformational freedom.

BLI data indicates that reduced, monomeric M2 binds to β-catenin with a K_D between 100 and 250 nM (Appendix **Figure A2.6**). This is approximately equal to the K_D measured for binding of the dimeric M2 ligand to the ARM domain or C-terminal domain (150 nM and 110 nM respectively shown in **Table 3.5**). Again, however it is important to note that the K_D's of the ligands for the domains of β-catenin in **Table 3.5** are based on concentrations calculated for the dimeric fraction of each protein in the non-reduced protein stock. The ARM and C-terminal IDR K_D calculations based on the monomeric fraction of the non-reduced M2 sample are nine times lower then what is measured for the dimer (15 nM to the ARM domain and 11 nM to the C-terminal IDR). However, these values do not correspond to the K_D for the M2 monomer measured using BLI. Therefore, the data indicates monomeric M2 binds one or both of the ARM or C-terminal domains, but not both simultaneously, with the same affinity as the dimeric form of M2.

To appreciate the changes that were selected for in M2 it is useful to understand the parent ligand from which it was generated. The BLI data for M1 also indicates that M1 gives a lower K_D to β-catenin as a non-reduced dimer (**Figure 3.9**). Also, the BLI data collected for the non-reduced form of M1 (**Figure A2.2C**) may correspond to an interaction where both two epitope and one epitope binding

are observed. M1 has affinity to the βCP sequence and the ARM domain (**Figure 3.16D,E**). However it would appear that not every M1 ligand which binds to surface bound β-catenin is able to access both epitopes. We hypothesize that at higher BLI loading concentrations the two epitope binding M1 ligands are masked by the more dominant one epitope ligands. At lower loading concentrations we could be observing that only ligands capable of binding two different epitope's generate a BLI signal. The reason for non-uniform two epitope binding may be attributable to M1 not being able to simultaneously bind two epitopes on the same β-catenin protein. Rather M1 may be binding two epitope on two different surface immobilized β-catenin proteins. These two β-catenin would have to be orientated appropriately to enable two epitope engagement by M1.

The BLI data for the T2 ligand is even strongly indicative of the dimeric ligand being the functional form of the protein. This is because T2 is a flexible tandem ligand and should be able to accommodate potential restrictions placed on β-catenin as a result of being immobilized on the surface. However, the reduced form of T2, even when loaded at very high concentration, gives little to no binding signal (**Figure 3.18A**). Similar arguments could be made for T3 which also does not show binding of the reduced ligand even when loaded at 50 nM (**Figure 3.8**). This is a concentration over 50 times higher than the soluble K_D determined in the competitive binding experiment.

The tandem ligands T2 and T3 do not form disulfide bonded homodimers to the same extent as their N-terminal ligands do. The most direct hypothesis for this observation would be that the 23CTL and or linker domains interfere with the optimal dimerization for disulfide bond formation. As dimerization and disulfide bond formation occurred quickly but did not change over time, T2 and T3 likely do have self-affinity for each other. However, perhaps alternate and stable conformations of the dimers are possible such that disulfide bond formation is inhibited. A similar phenomenon is observed but to a lesser extent for M1, M2 and T3-NTL. 10% of each of these protein stocks is seen to remain in

the monomeric form. An alternative hypothesis is that the monomeric fraction could represent misfolded protein which cannot dimerize and form a disulfide bond.

T2 being functional in the disulfide bonded dimeric state adds further evidence to indicate that the M1 protein is functional in the dimeric disulfide bonded state. Additionally, this would indicate that the T2 surface display library effectively consisted of two disulfide bonded M1 proteins, each linked to a C-terminal naïve Sso7d library. The result was the selection of a tetravalent Sso7d ligand, which consists of two 23CTL proteins, each linked to the M1 dimerizing domain. The data collected indicates that 23CTL is bispecific for two epitopes on β-catenin. Considering the strength of the avidity affect this is not entirely unexpected. Two low affinity ligands, tethered together could easily give rise to higher affinity then a single protein which targets one domain. Additionally, β-catenin is comprised of twelve repetitive three alpha helix bundles, therefore homology on the target perhaps enabled a single ligand to have affinity for two different locations.

Evidence for low affinity multi-epitope targeting of 23CTL comes from a lack of ability to measure 23CTL's affinity in two different contexts. BLI data collected at the relatively high loading concentration of 500 nM reduced, monomeric T2 shows that little to no binding is observed. This concentration is almost 6 times the K_D to the ARM region measured for T2. A much stronger signal would be expected if a single 23CTL were solely responsible for the ARM binding of T2. Similarly, **Figure 3.18B** shows that 23CTL, when expressed by itself, shows very little binding signal to the yeast surface displayed ARM domain when labelled at 1 μM. Again in this case weak binding signal at high concentration relative to the K_D of T2 to the ARM domain indicates 23CTL is not a high affinity ARM binding ligand. However, these two data points would fit the hypothesis that 23CTL has weak affinity for two epitopes on β-catenin. The relatively high affinity of T2 for the ARM domain would then be

attributable to simultaneous binding of both of its 23CTL's to these two epitopes on β-catenin. This is an important consideration for any future use of 23CTL.

Important observations can also be made in regard to how these ligands bind the C-terminal IDR of β-catenin. Two techniques were employed in this study to increase the affinity of a peptide specific ligand. The first was to link the peptide binding ligand to a naïve Sso7d library and select for high affinity ligands. The T2 high affinity tandem ligand selected for evaluation in this study shows some degree of specificity for the C-terminal IDR relative to the C-terminus truncated by the βCP sequence (**Figure 3.13C**). However, definite association with the C-terminal IDR in β-catenin can be inferred as the K_D for T2 to β-catenin is over an order of magnitude lower than the K_D observed to the ARM domain (5.5 nM vs 85 nM). The difference in affinities is attributed to the contribution M1.2 has in binding the βCP sequence. This indicates that affinity maturation through epitope addition affectively increased the affinity of the βCP targeted domain.

A second approach evaluated was to affinity mature non-tandem ligands pre-targeted to the βCP sequence using the β-catenin protein. This resulted in an M2 ligand with dramatically improved affinity to the βCP sequence in the context of the IDR. This was seen without a similar increase for M2 to the βCP epitope when not fused to the IDR. Given that M1, the non-affinity matured ligand does not share this same affinity for the C-terminal IDR, indicates that K_D's ascribed for ligands or antibodies to peptides may only very inaccurately represent the K_D to the peptide in the context of the IDR. This leads to the non-surprising result that it is possible to use a peptide segment to pre-target a library towards a specific sequence, but that affinity maturation should be done using the IDR or protein of interest. This also then indicates that had affinity maturation been carried out using βCP as a target instead of full length β-catenin it is possible that the best binders would have had lower affinity towards the intended IDR as well as to β-catenin as a whole.

A similar result is observed when affinity maturing T3 for binding to β-catenin. Again, the resulting ligand strongly and specifically binds the C-terminal IDR and this increase in affinity is not associated with a change in the affinity for the ligand to βCP. Another important point can be made about the T3 ligand in this regard. T3 was affinity matured by mutating both the N-terminal and C-terminal ligands and linker region simultaneously as one construct. This resulted in the T3-NTL not being the only source of the increased affinity of T3 for the C-terminal IDR. T3-NTL expressed by itself gives negligible binding signal to the C-terminal IDR even at relatively high concentrations. Therefore, we infer that T3-NTL must act synergistically with another domain of the ligand in order to bind the C-terminal IDR.

T2 was designed to increase the affinity of M1.2 by addition of a second ligand. We hypothesized that this second ligand would target a second location on β-catenin and this is what is observed. The T2 ligand gained substantial affinity to the ARM domain relative to M1. Surprisingly, the M2 ligand is also observed to gain an epitope, in this case by evolution rather than by design. The single L54W point mutation in M2, resulting in a homodimer with two identical point mutations, is observed to have a three-fold affect: (1) the K_D to the pre-targeted βCP sequence is dramatically improved for βCP in the context of the entire C-terminal IDR; (2) affinity to the ARM domain increases and appears to have been made much more specific; and (3) synergistic and simultaneous binding of both epitopes on β-catenin by M2 is observed. The K_D of M2 to the C-terminus and ARM domains (150 and 110 nM) are an order of magnitude higher than the K_D measured for M2 to β-catenin (4.8 nM) indicating both epitopes are engaged simultaneously. We attribute this threefold change to the intrinsically disordered domain acting as a linker and enabling the ligand to engage two distinct epitopes.

The observations made for M2 would appear to be a one-off result and unlikely to be repeated. However, this conclusion is strongly challenged by the results observed for T3. T3 was mutated such that the K_D to the ARM domain decreased (85 nM to 5 nM); and the very minor affinity to the C-terminal IDR increases dramatically from a value we were not able to measure to 130 nM. T3-NTL is also seen to have significant affinity to the ARM domain (85 nM). Both changes are attributed to the three mutations in T3-NTL which are in a similar region of the Sso7d scaffold as the L45W present in the M2 ligand. Surprisingly, the K_D of T3 to β-catenin in the competition based K_D measurement increased by less than an order of magnitude from 5.5 nM for T2 to 0.74 nM. A possible reason for this relatively small increase in affinity could be attributable to a lack of synergy in how the individual changes in epitope specific binding contributed to the affinity of T3 for β-catenin.

Based on the data presented models of how the ligands interact with β-catenin are shown in **Figure 3.20**.

3.4 Conclusions

The selection of a functional, self-assembling, disulfide bonded, dimeric ligands is an unexpected but important result to have identified. The original Sso7d library was designed to bind a given target on the 3-β-sheet surface of the Ssso7d Scaffold. The M1, M2 and T3-NTL ligands however selected for self-affinity, disulfide bond formation and target binding. Given these results it is not clear which residues were selected to generate self-affinity and which are involved in target binding. This can be considered as a new and unique scaffold which has a K_D that is dependent on the reduction state of the buffer. For T2 the difference in K_D could be a factor of 1,000 or greater and 25 for M2. Additionally M2 appears to go from being a two epitope to one epitope ligand when reduced. The current ligand properties were selected in the context of being a homodimer, however alternative affinity maturation schemes could be considered whereby heterodimeric proteins could be selected.

This would greatly expand the functional diversity as homodimers have to be functional proteins which contain symmetric mutations. Ligands which have dramatic changes in affinity based when reduced could have applications in affinity chromatography for targets which do not contain disulfide bonds.

It is shown that directing a naïve library to bind a peptide which represents an epitope within an intrinsically disordered domain is observed to be highly effective. Affinity maturation carried out on pre-targeted ligands resulted in substantial gains of affinity to the targeted epitope within the IDR. However, this change did not correlate to any change in the affinity towards the peptide itself indicating that maturation towards the IDR containing protein is beneficial. However, it is also observed that gain of an epitope is a possible mode through which affinity maturation may occur when targeting an IDR. Interestingly, two epitope binding is also observed for many natural β-catenin C-terminal ligands. These natural ligands bind through a mechanism that requires the C-terminus and at least a portion of the ARM domain to be present[64].

3.5 Materials and Methods

3.5.1 β-catenin Target Expression

Full length human β-catenin in pet28a plasmid was purchased from Addgene (Plasmid #17198). The plasmid was transformed into BL21 Star cells and expressed as follows. A 5 ml overnight culture was used to inoculate a 1 liter culture of 2XYT media. Cells were incubated at 37°C until the O.D. reached 1.0. The culture was cooled to 20°C for 1 hour and then induced with 250 mM IPTG and incubated to 20-24 hours. The protein was purified at pH=7.8 using a Biorad column (catalog # 7324614). Elution product from the column was dialyzed into HEPES buffer at pH=8.0 (20 mM HEPES, 150 mM NaCl). Lower pH's were seen to precipitate the β- catenin. The protein was largely observed at this point to be of single band purity. β-catenin was biotinylated using ES-Link™

Sulfo-NHS-Biotin at a ratio of 8:1 (Thermo product #21217). This was carried out overnight at 4°C. The protein stock was dialyzed serveral times to remove excess biotin containing moiety.

3.5.2 β-catenin constructs cloning

Using the pet28a expression plasmid as a template the ARM domain was cloned into the pCTCON yeast display expression plasmid using Nhei and Bamhi cut stes. The ARM domain (residues 151-666) was cloned using the forward primer 5' GTTCTGCTAGCATG CGTGCAATCCCTGAACTG 3' and the reverse primer 5' GCTTTTGTTCGGATCC CTTGTCCTCAGACATTCGG 3'. The N-terminal IDR (residues 1-150) was clones using the forward primer 5' GCTATA GCTAGC ATG GCTACTCAAGCTGATTTGAT 3' and the reverse primer 5' GCTATAGTCGACTTATGTGGCAAGTTCTGCATCATCT 3'. The full length C-terminal construct (667-781) 5' CCACGT GCTAGC AAGCCACAAGATTACAAGAAA CG 3' and reverse primer 5'GCTATAGTCGACTTACAGGTCAGTATCAAACCAGG 3'. The C-terminal IDR minus the last twelve residues (667-769) used the same forward primer with the following reverse primer 5' CAGTATT GG ATCC ACCTGGAGGCAGCC 3'.

3.5.3 GST-TCF Expression

TCF(1-65) in pGEX-6P3 was a kind gift from the Liu Lab at the University of Kentucky. The protein was expresses in BL21 Star cells and purified on a GST column. The product was dialyzed to HEPES buffer and biotinylated using a maleamide biotinylation kit at a ratio of 8:1 (ThermoFisher product # 21901BID). Unreacted biotin was removed by dialysis.

3.5.4 Generation of the peptide pre-targeted library and selection of the M2 ligand.

A full description of yeast surface display methods is described here[22]. The referenced methods document gives a complete listing of buffers, media and incubation times.

For this work all inductions to express yeast surface display ligand were done at 20°C in SGCAA media. Also, all selections were carried out in PBS + 1 g/L BSA (PBSA).

The initial selection of peptide binding ligands for the biotinylated β-catenin C-terminal peptide (GDSNQLAWFDTD) purchased from Genescript is described previously[13]. Briefly a selection using naïve Sso7d library was carried out by pre-loading the peptide on magnetic biotin binder beads. This was done after a negative selection against the biotin binder beads was carried out. This population was expanded, subject to a second negative selection against biotin binder beads and then subject to a positive selection using full length biotinylated β-catenin pre-loaded onto biotin binder beads. The magnetic bead loading was carried out following the manufacturer's protocol (Thermofisher product #11047). This selected population was expanded and then zymoprepped (Zymo Research yeast plasmid miniprep kit product #D2004). The resulting DNA was then subject to an error prone PCR using primers starting and ending at the N and C-terminus of Sso7d: forward primer 5' ATG GCG ACC GTG AAA TTT AAA TAT 3' and reverse primer 5' TTT TTT CTG TTT TTC CAG CAT CTG 3'. The error prone PCR was carried out using nucleotide analogues 8-oxodG and dPTP the PCR was carried out using standard Taq polymerase (NEB #M0273L) with the standard buffer and 2 mM MgCl2. Three PCRs were carried out and the products were mixed evenly prior to being amplified with yeast surface display homolgous ends. The conditions for the three error prone PCRS were as follows: (1) 2 μM of each nucleotide analogue and 15 cycles (2) 10 μM of each analogue and 15 cycles and (3) 2 μM of each analogue and 20 cycles. The PCR cycler conditions were as follows 95°C for 2 minutes then repeat cycles of 95°C (30 seconds), 56°C (30 seconds) 72°C (20 seconds) with a 2 minute extension at 72°C. The PCR for the addition of overhangs with homologous ends to the yeast display vector has been described previously[13]. A library was made following the LiAc electroporation protocol[78]. 4 transformations were in 400 μls in 2 mm cuvettes (Biorad) each using 12 μg of insert and 12 μg of

vector. Serial dilution of the library before expansion showed the diversity to be $2x10^8$. The transformation of the vector control yielded 10^5 fewer transformants.

The library was expanded, induced for yeast surface display expression and subject to a negative selection with Dynabeads His-tag isolation beads (ThermoFisher part# 10103D). 10 times the library diversity was put in 6 mls of PBS with1 g/L BSA (PBSA) after multiple washes with PBSA. The library was incubated with 50 µls of His-tag beads overnight. The following day a 4 hour positive selection was done with β-catenin pre-loaded on His-tag beads, this loading was done following the manufacturers protocol. The beads were washed 3 times with PBSA and then allowed to grow up in SDCAA media. FACS based selections were then done successively at 100 nM, 10 nM and 0.5 nM. Each time double labeling with HA antibody (1:100 dilution) was used. HA antibody was purchased form Life technologies and goat anti-rabbit Dylight 633 (1:200) from ImmunoReagents Raleigh, NC. Biotinylated β-catenin was used for labeling in each case as well as Streptavidin PE (SAPE) (1:500) as a secondary reagent (Thermo Fisher part# S866). The highest affinity ligands from the final sort were zymoprepped and sequenced.

3.5.5 Generation of the Tandem Library

To generate the tandem library the M1.2 ligand was inserted following the HA tag and (G4S)3 linker, a (TP)15 linker then was inserted between Smai and Avrii cut sites followed by a stuffer fragment and the Bamhi site that precedes the cMyc tag. The library was generated by cutting the stuffer fragment between Avrii and Bamhi and using the (TP)15 and cMyc DNA coding regions for homologous recombination.

The codons used for the linker region are as follows: ccg acc cca aca cct acg ccc act ccc acg cct aca cca acc ccg act cct acc cct aca ccg acg ccc aca cca act ccc acc ccg aca. The forward primer used to generate the error prone library is as follows: 5’ CCG ACT CCT ACC CCT ACA CCG ACG CCC

ACA CCA ACT CCC ACC CCG ACA CCT AGG AGC ATG GCG ACC GTG AAA TTT AA 3'. The standard yeast surface display was used[13]. The library was transformed according to the LiAc protocol[78] using 4 transformations each with 4 μg vector and 4μg of insert DNA. $4x10^8$ transformants were observed with 10^4 fewer vector control transformants.

3.5.6 Selection of Best Binders in Tandem Library

The initial library selection was carried out using 10 times the diversity of the library ($4x10^9$ cells) after washing with PBSA with 100 nM biotinylated β-catenin in 40 mls PBSA. This volume and concentration of β-catenin represents 10 times excess over the number of library yeast surface display fusions present. The cells were incubated overnight at 4°C and then spun down and washed 3 times with PBSA. Cells which bound the β-catenin were then collected using biotin binder beads this was done following the manufacturers protocol.. The selected ligands were expanded, then re-induced and a FACS was carried out by double labeling for HA tag expression and binding to β-catenin at 10 nM. This population was expanded and re-induced and then subject to a cMyc based selection for full length clones. This was a single label sort using chicken anti-cMyc (Thermofisher A-21281) (1:250) and goat anti-chicken Dylight 633 from ImmunoReagents (1:200) The selected population was expanded and reinduced and finally a double labelled sort for HA expression and binding at 0.5 nM β-catenin was done to select the highest affinity clones. The same secondary reagents were used as detailed in the 3.5.4 section.

3.5.7 Assembly of the second generation tandem library

T2 ligand was affinity matured by mutagenesis across its entire gene. The tandem ligand does not PCR well in that DNA for a monovalent Sso7d ligand is consistently generated. As μgs of pure DNA are required for the transformation gel purification is not a viable option. To generate sufficiently

pure DNA the library was broken into two segments. The first insert consisted of the N-terminal ligand DNA and the linker region; the second consisted of the linker region and the C-terminal ligand DNA. The yeast homologous recombination system was used to reassemble a library of bivalent ligands. PCRs targeting an error rate of 2-3 residue mutations per insert were carried out for both inserts, in this case a purely Taq based method was used[55]. 31 total PCR cycles were carried out with 1:10 dilution of the PCR product after the first 8 cycles and then after every 4 cycles until 27 total cycles was reached; the final round of PCRs was allowed to run without further dilution for 5 cycles. Primers used for this erorr prone PCR were on the vector just before or after the Sso7d insert and at either end of the linker. 2 μM of each primer was used. The sequences used are as follows: for the N-terminal ligand and linker the forward used is 5'GTG GTG GTT CTG CTA GCA TG 3' and reverse primer 5' TGT CGG GGT GGG AGT T 3'. For the C-terminal ligand the forward primer 5' CCG ACT CCT ACC CCT ACA 3' and reverse primer 5'GGA TCC GAA CAA AAG CTT ATT TCT 3'. Cycler conditions were 95°C (3 minutes) then cycles of 95°C (30 seconds) 55°C (30 seconds) 72°C (25 seconds) and after the final cycle an extension of 5 minutes. After the introduction of mutations a second PCR was done to: 5'AGT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT GCT AGC ATG 3'. The same reverse primer was used. The DNA for the C-terminal ligand was amplified using the same forward primer as for the error prone library and this reverse primer 5' CTC GAG CTA TTA CAA GTC CTC TTC AGA AAT AAG CTT TTG TTC GGA TCC 3'. This PCR was carried out using phusion polymerase using 500 nM of each primer, 20 ngs of template DNA and the HF buffer supplied by the manufacturer (ThermoFisher). For transformations were carried out following the LiAc protocol again (4 μg of each insert and 4 μg of digested pCTCON). The transformed library contained a diversity of $1x10^8$ with 1000 fold fewer transformants in the vector control. Eight library members were

sequenced: two had no errors, four contained stop codons and one of these four contained an additional linker region prior to the second ligand. The remaining mutants had one to four mutations per Sso7d.

The initial library selection was carried out using 10 times the diversity of the library ($3x10^8$ cells) after washing with PBSA with 5 nM biotinylated β-catenin in 100 mls PBSA. The mixture was allowed to incubate overnight. Cells were then washed 3 times with PBSA. 50 μls of biotin binder beads were then used to collect cells which had been re-suspended in 2 mls of PBSA during a 20 minute incubation. The beads were then washed 3 times with 2 mls PBSA and then grown in SDCAA. Selected cells were expanded, re-induced and sorted by double labeling with 5 nM biotinylated β-catenin and HA antibody. This was done in two steps $5x10^6$ cells were labelled in 1 ml with 5 nM β-catenin for 2 hours while rotating at 4°C. Cells were spun down and then labelled in 100 uls with 1 ul HA antibody for 10 minutes. Secondary labels using SAPE and The highest affinity clones were selected by sorting along the diagonal. $2x10^4$ cells were selected after running 6 million through the cytometer. Selected cells were expanded, re-induced and sorted a final time using an off-rate based selection. $2x10^6$ cells were labelled with 10 nM biotinylated β-catenin for 2 hours in 200 uls, washed with 1 ml PBSA then labelled with 10 nM non-biotinylated β-catenin in 50 mls PBSA. This was incubated on a rotator for 24 hours at 4°C. The cells were washed 1 time with 1 ml PBSA and then labelled with SAPE and HA antibody for 10 minutes in 100 uls PBSA. A final labeling was done with goat anti rabbit Dylight 633 antibody, 10 minutes 100 uls PBSA.

3.5.8 Ligand cloning into bacterial expression vectors

M2 was cloned into pet22b, expressed and purified as described previously [13]. T2 was cloned into pet22b similarly except in this case the HA and cMyc tags were retained. T3 was cloned into pet28a using subcloning. The ligand was cut out of pCTCON using Nhei and Xhoi and then directly ligated into pet28a which was cut using the same primers. This results in an N-terminal His-tag and the C-

terminal cMyc tag is retained. The 23CTL was cloned by subcloning as well. In this case the T2 pet22b construct was cut using Avrii and Nhei the resulting plasmid can be re-ligated directly as the two cut sites have compatible overhangs. The T3-NTL ligand was subcloned from the T3 pet28a construct by cutting with Smai and Bamhi and inserting a (PT)3 stuffer fragment prior to the cMyc tag.

Protein purifications were carried out as described in[13] with the exception of using shorter induction times and less IPTG (4 hours with 250 mM IPTG). Also, for the bivalent ligands a heat treatment step at 65°C for 15 minutes was done as it is necessary to remove enzymes which would otherwise cleave the linker. Protein concentrations were determined using a BCA assays. Ligands were biotinylated using a lysine biotinylation (Thermo Fisher product# 20217). A manufacturers protocol for N-terminal biotinylation was followed, this was carried out at pH = 6.5 in PBS with a 5:1 ratio of biotinylation reagent at 4°C for 24 hours. Unreacted biotinylation reactant was removed by dialysis. Protein concentrations are based on BCA (ThermoFisher 23225) and protein purity by running NuPAGE Bis-Tris SDS page gels with reduced and non-reduced samples were run. Samples were denatured in SDS solution with or without the manufacturers reducing solution. Samples were denatured at 95°C for 5 minutes. We used GelQuant.NET software provided by biochemlabsolutions.com for densitometry analysis.

3.5.9 Blitz Experiments

For all experiments PBS with 1 g/L BSA and 0.04% Tween 20 was used as the buffer. Blitz™ experiments were carried out by loading 8:1 lysine biotinylated β-catenin at 1 μM for 2 minutes onto the Blitz streptavidin sensor. The sensor was then re-equilibrated in buffer containing 1 mM TCEP or 5 mM DTT for reduced conditions. Protein stocks at 30-70 μM protein based on a monomeric molecular weight were reduced for 30 minutes at room temperature with TCEP or DTT. Samples were then diluted as required for the assay into the appropriate reduced buffer.

3.5.10 K_D Measurements

KD values were estimated using yeast surface titrations or using a competition assay with soluble binding protein, as described [24,53]. Briefly, for yeast surface titrations, ~$1x10^4$ yeast cells expressing the soluble ligand of interest were diluted in $2x10^6$ un-induced yeast cells. This mixture was vortexed, centrifuged and re-suspended in 750 µls of PBS + 1g/L BSA buffer (PBSA). 50 µL aliquots of cells were then added to 12 micro-centrifuge tubes, the mixture was centrifuged at 13,000g and the supernatant aspirated. The highest concentration sample β-catenin sample was subsequently serially diluted eleven times, such that a 500 µL sample was left and could be directly to the pelleted cells. Samples were incubated with nutation at 4 °C for time corresponding to estimates to reach 98% of equilibrium labeling. Labeling as such requires double labeling for expression such that induced cells can be differentiated from un-induced cells. The HA-tag is best suited for this purpose as cMytc tag labeling has been seen to interfere with target binding. After incubation with β-catenin cells were washed and re-suspended in 50 µL of a secondary labeling mixture comprised of rabbit anti-HA antibody (Life Technologies) at 1:100 of rabbit anti-HA-tag antibody and a 1:500 diluton streptavidin (PE) (Life technologies). This mixture was incubated for six minutes, washed and then incubated with donkey anti-rabbit antibody conjugated to Dylight633 (ImmunoReagents). Cells were then analyzed on an Accuri C6 cytometer, for each tube the average PE fluorescence (FL2) was captured for 1,000 FL4 positive cells.

The K_D determination by soluble competition was carried out by labeling $2x10^6$ yeast cells displaying the ligands of interest with biotinylated β-catenin at a fixed concentration and the same yeast displayed ligand was also added to the solution in soluble form. 1250 µls of the highest concentration ligand stock was made and then serially diluted eleven times by adding 625 µl of the ligand stock to 625 µls of buffer. 400 µls of each serial dilution was then added to 50 µls of 10x β-

catenin stock and 50 µls of yeast cells containing $2x10^6$ yeast cells. One tube was also run which did not contain β-catenin, 50 µls of buffer was added in its place. PBSA buffer was used for all dilutions. This mixture was allowed to equilibrate (16 hours). After washing secondary labeling with SAPE was carried out. Average FL2 signal from 10,000 cells was recorded on an Accuri C6 cytometer. A global least squares fit for the solution phase K_D was carried out using a mol balance to determine the concentration of each of the components in solution. To do this the free unbound β-catenin fraction was calculated using yeast titration based K_D and the fluorescence readout from cytometry.

3.6 Figures and Tables

A

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
CT12.1	M	A	T	V	K	F	K	Y	K	G	E	E	K	Q	V	D	I	S	K	I	G	V	V	G	R	C	G	K	V	I	V	F
	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
	P	Y	D	L	G	G	G	K	L	G	L	G	T	V	S	E	K	D	A	P	K	E	L	L	Q	M	L	E	K	Q	K	K

B

	8	17	20	21	23	25	28	30	32	33	40	42	44	48	54
M1	K	S	G	V	G	C	V	V	P	Y	L	L	T	K	L
M2 (2 clones)															W
M2-2 (3)	R									F					W
M2-3 (1)	R	N								F				R	W

C

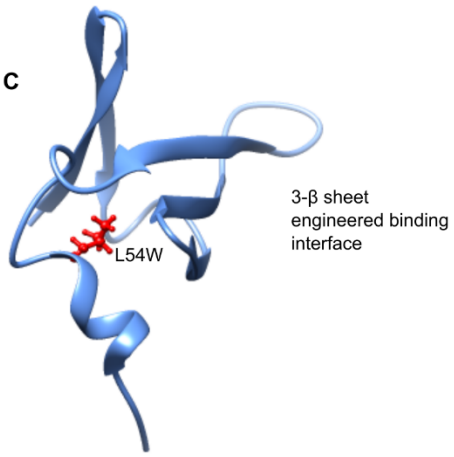


Figure 3.1: M1 and M2 differ by an L54W mutation.

(A) The residues which make up M1 are shown, the highlighted cells indicate the positions mutated in the naïve library. (B) A lower affinity pool of ligands from which M1 was selected from were affinity matured by mutagenic PCR and then inserted into the yeast surface display system. Six

ligands from the pool of ligands selected to bind β-catenin are shown here. L54W is a conserved mutation which appears to be the only position necessary for the increased affinity of the ligands. (C) The L54 position is shown in the PDB generated diagram. The mutation faces internally, towards the location of the 3-β sheets which is the location of the 10 library positions mutated in the naïve library.

	SmaI		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	AvrII		Sso7d						
EX	P	G	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	R	S	M	A	T	V	K	F
1	P	G	P	T	P	T	P	H	P	H	P	H	P	H	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	T	P	R	S	M	A	T	V	K	F	K	Y
2			P	G	P	T	P	T	P	T	P	T	P	T	P	T	P	H	P	L	L	P	L	H	R	R	P	H	Q	L	P	P	R	H	L	G	A	C	D	R	E	I
3				P	G	P	T	P	T	P	T	P	T	P	T	H	H	T	R	L	L	P	L	H	R	R	P	H	Q	L	P	P	R	H	L	G	A	W	R	P		
4	P	G	P	T	P	T	P	T	P	T	H	H	T	H	T	H	T	R	L	L	P	L	H	R	R	P	H	Q	L	P	P	R	H	L	G	A	W	R	P		
5			P	T	P	T	P	T	P	H	P	H	H	T	H	T	T	P	T	L	L	P	L	H	R	R	P	H	Q	L	P	P	R	H	L	G	A	W	R	P		
6	P	G	P	T	P	T	P	T	P	T	H	T	H	P	H	P	H	R	L	L	P	L	H	R	R	P	H	Q	L	P	P	R	H	L	G	A	W	R	P		

Table 3.1: Poly-peptide chain enriched in linker region of tandem library

Sequencing was done following a MACS and 10 nM FACS double labelled for β-catenin and HA-tag expression. Three of nine clones had the expected linker (EX) sequence as well as a C-terminal Sso7d library ligand. Clone 1 is almost identical to the expected sequence, with just three histidine's present. Clones with a conserved peptide sequence in the linker region were selected in five of nine clones, shown as clones 2-6 here. Empty boxes are placed where those sequences had to be right shifted to line up the homologues peptide region. The enrichment of this peptide sequence indicates that although not initially observed in library sequencing there are mutations in the linker region and this sequence were strongly selected for. This suggests that the ligand linked to a randomized peptide sequence may be another promising alternative method to generate high affinity tandem ligands.

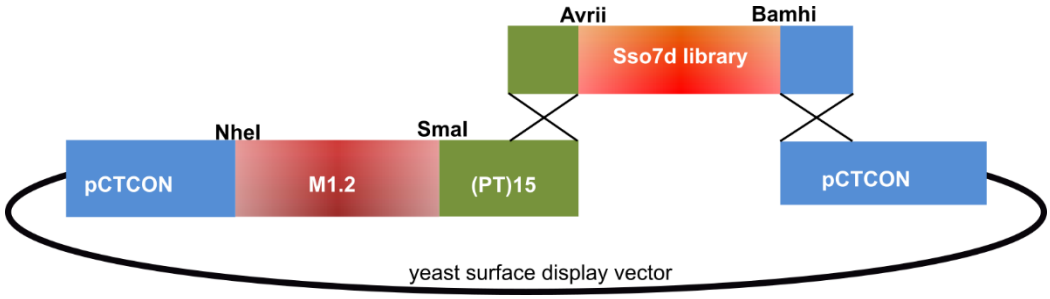


Figure 3.2: Generation of a tandem library with M1.2 as the N-terminal ligand.

The homologous recombination step used to generate the first generation tandem library is shown. A naïve Sso7d library is inserted into a modified yeast surface display vector. The modified vector was generated by inserting the following into the pCTCON yeast surface display vector between the Nhei and Bamhi cut sites: M1.2, a (PT)15 linker and a stuffer fragment (not shown) flanked by Avrii and Bamhi. The newly generated plasmid was double digest at the Avrii and Bamhi sites prior to transformation with the Sso7d library into electro competent yeast. The naïve Sso7d library consisted of the same ten residues surface exposed residues which were mutated in a prior publication[13]. The library was amplified with ends homologous to the yeast display vector and the (PT)15P linker such that homologous recombinbation could take place.

Clone	Linker	CTL Mutations									
		20	21	23	25	28	30	32	40	42	44
T2	(PT)8 P	D	R	W	L	S	A	S	N	F	H
T2-2	P T P T P T H T P H P H P H T (PT)7P	Q	F	R	L	H	Y	R	D	V	A
T2-3	P T P T P H P H T H T P H P T (PT)7P	C	D	M	A	D	L	I	R	I	G
T2-4 (x3)	P T P T P T H H T H T P H P H (PT)7P	T	A	S	I	G	F	L	W	T	T

Table 3.2 The highest affinity bivalent ligands selected for tend to have histidine's incorporated in the linker.

Mutations observed in clones sequenced from linking the M1.2 clone to a naïve Sso7d library and selecting the highest affinity lines to β-catenin. Not one of the clones sequenced contained the intended (PT)15P linker. Five of six clones sequenced contained multiple histidines in the linker. T2

was chosen for future work as it was suspected that the linker (PT)8P linker does not play an active role in binding other than to tether the ligands together. The mutations in selected for in the library positions of the C-terminal ligand (CTL) are shown on the right side of the table.

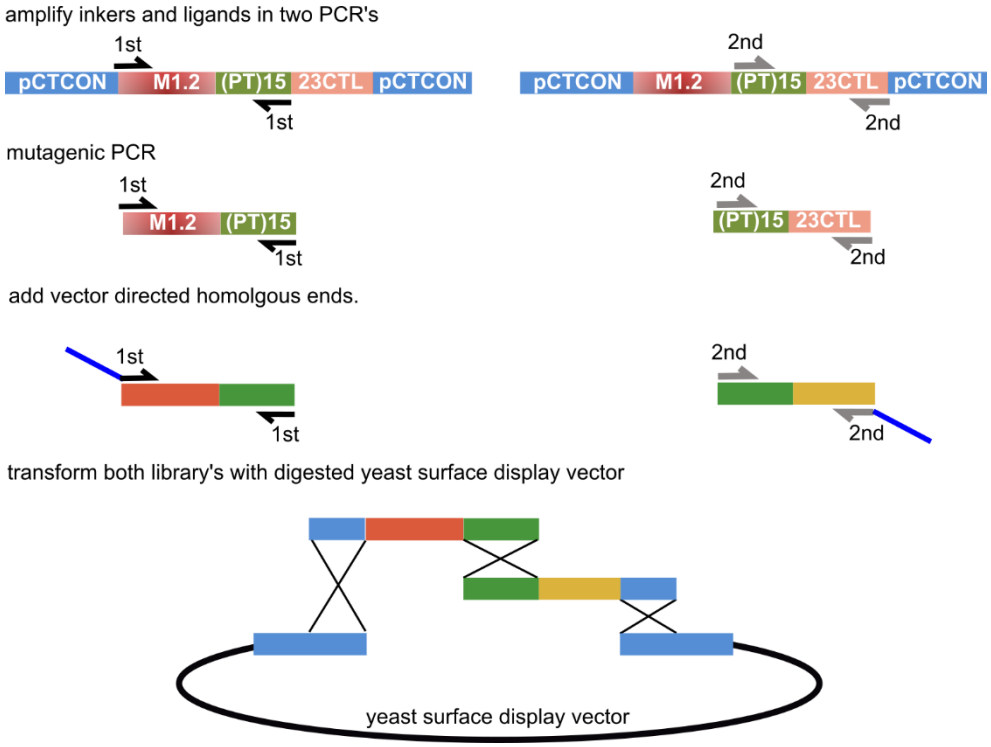


Figure 3.3: Generation of the T2 tandem library

The T2 library was generated by a two part amplification. The M1.2 ligand and linker region were amplified in a 1st PCR, in a second PCR the linker and 23CTL ligand were amplified. The two resulting PCR products were subject to identical but distinct error prone PCRs using the same primers as used for amplification. In a third PCR, vector directed homolgous ends were added to both libraries. The resulting library and NheI and BamhI digested pCTCON vector were transformed into electrocompetent yeast.

	Mutiations			NTL								Linker		CTL		
Clone	NTL	Linker	CTL	2	27	49	55	56	57	60	61	L17	L18	0	10	18
T2				T	K	D	L	Q	M	K	Q	P	T	M	E	K
T3	3	0	0			E			V		R					
T3-2	1	0	1						R					T		
T3-3	2	2	1	A							R	S	S	V		
T3-4	2	1	0		R				R			S				
T3-5	2	17*	0				P			E						
T3-6	1	0	2						V							E
T3-7	1	0	1						V						D	

*(PT)8P Linker insertion

Table 3.3: A diversity of mutations are seen in the highest affinity ligands selected from the T2 mutational library.

Shown are the mutants that were sequenced in the best binder pool from the affinity maturation of the T2 ligand. The table shows the total number of mutations observed in each portion of the T2 protein, the N-terminal ligand, linker region and the C-terminal region as well as what those mutations were. The M2, L54W, mutation is not observed in this case despite the T3 N-terminal ligand being preferentially mutated. T3 was chosen for further evaluation, its mutations are conserved to the N-terminal M1.2 ligand and it has a somewhat conserved M57V mutation.

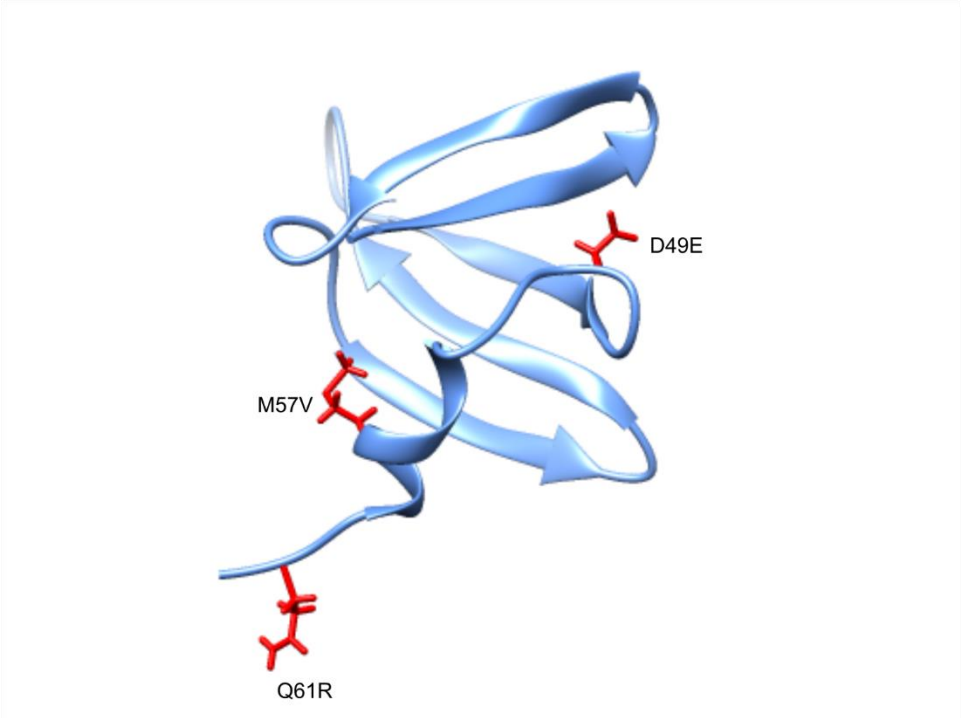


Figure 3.4: PDB diagram of mutations observed in T3-NTL.

The mutations for T3-NTL (D49E, M57V and Q61R). As was observed for M2 mutations occur on the Sso7d surface opposite the naïve library mutations and again the mutation(s) are predominantly inward facing. The Q61R residue is the only mutation that is surface exposed. The L54W mutation is not observed in the best binder pool that T3 was taken from; the M57V mutation is the dominant mutation seen in 3 of 7 clones in the pool that T3 was sequenced from.

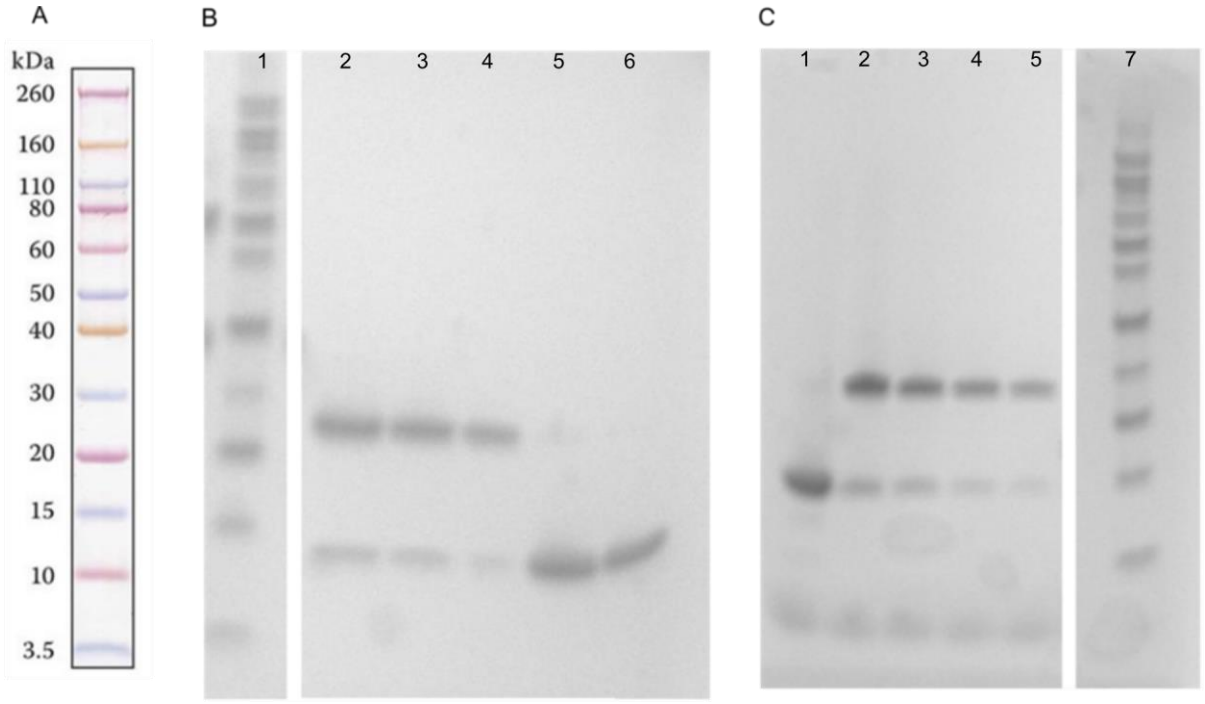


Figure 3.5: SDS page gels indicating M2 and M1 exist predominantly as disulfide bonded homodimers

SDS page gels demonstrating that all the ligands in this work exist predominantly as disulfide bonded dimers. A) The Novex® prestained protein ladder used in (B) and (C) lane 1. (B) Lanes 2-4 are dilutions from a recombinately made, purified and dialyzed (PBS) protein stock of the M2 protein. The fifth and six lanes are TCEP reduced stocks of M2 loaded at the same dilutions of the protein stock as in lanes 1 and 2. The weight of an M2 monomer is 8.1 kDa and 16.2 kDa for the dimer. (C) A similar gel for the M1 ligand. The first lane is a reduced concentrated sample and the following lanes are dilutions of the non-reduced protein stock. M1 has a nearly identical MW to M2.

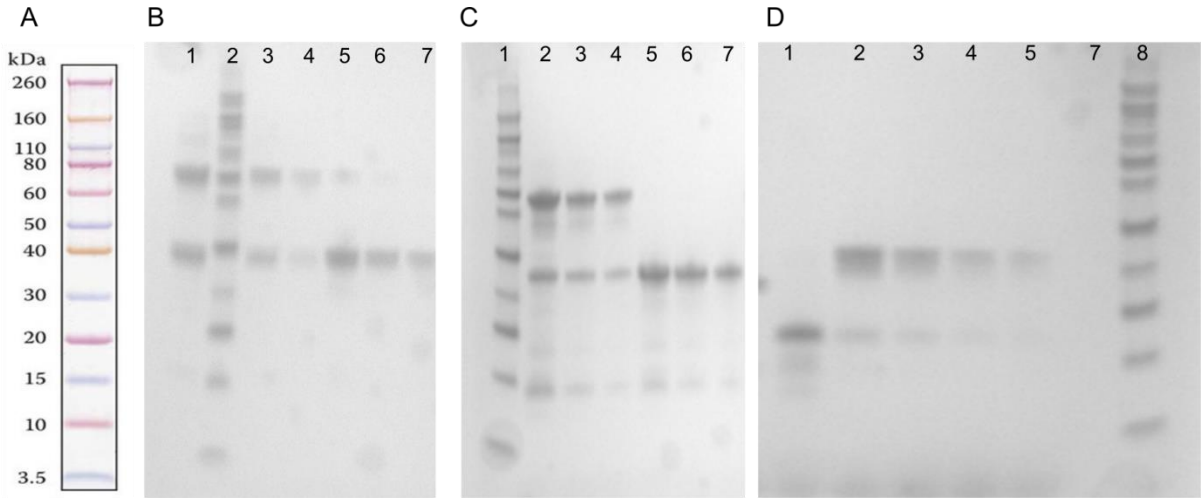
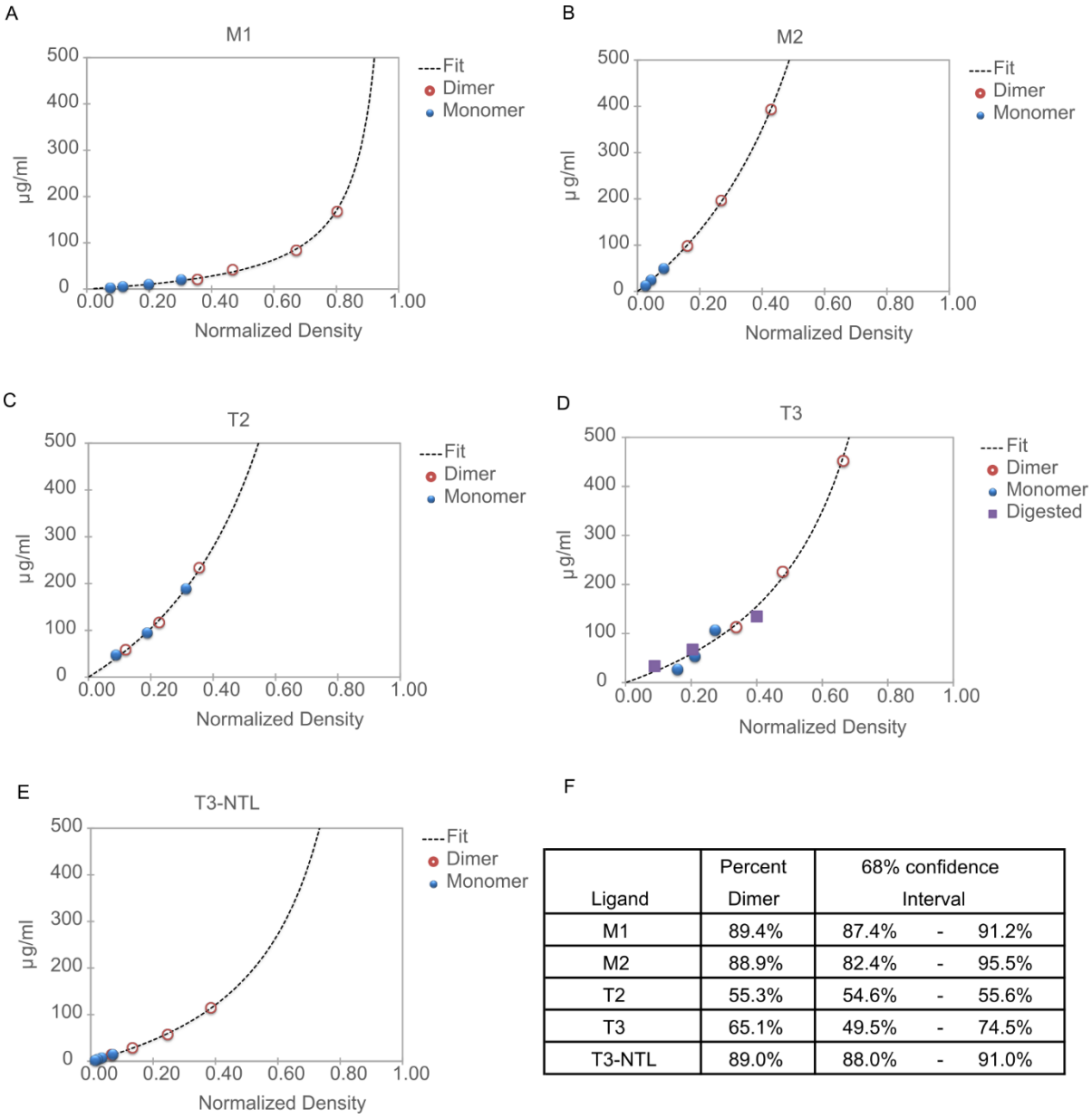


Figure 3.6: SDS page gels demonstrating T2, T3 and T3-NTL also exist as disulfide bonded dimers.

(A) The Novex Prestained Protein Ladder which was run in (B) lane 2, and in lane 2 for (C) and (D). (B) The T2 ligand: non-reduced samples can be seen at increasing dilutions in lanes 1, 3, 4 and reduced by TCEP in lanes 5-7 at the same dilutions as 1, 3, 4. The MW of T2 is 21.5 kDa for the monomer. The protein is observed to run somewhat larger then this in the SDS page gel. (C) T3 ligand: non-reduced in lanes 2-4 at increasing dilutions and reduced 5-7. The T3 monomer has a molecular weight of 20.1 kDa. (D) The T3-NTL ligand was also expressed and purified and is shown in the reduced state in lane 1 and at various dilutions in the non-reduced state, lanes 2-5. The MW for the T.3-NTL monomer is (10.4 kDa) this is as this protein was expressed with an additional N-terminal 6xHis tag.



F

Ligand	Percent Dimer	68% confidence Interval
M1	89.4%	87.4% - 91.2%
M2	88.9%	82.4% - 95.5%
T2	55.3%	54.6% - 55.6%
T3	65.1%	49.5% - 74.5%
T3-NTL	89.0%	88.0% - 91.0%

Figure 3.7: Densitometry analysis to determine the percentage of dimeric protein in each stock. Image J software was used to quantify the density of the SDS page bands in Figures 3.5 and 3.6. For all of the ligands except T3 only two significant bands are observed. T3 SDS page gel shows evidence protease degradation products. The concentration of protein run in each lane was quantified

based on BCA measurement of the undiluted protein stock and the dilution factor used for each lane. The SDS page band densities from each lane of the protein of interest were correlated to protein concentrations using a least squares fit to the Hill Equation. For each ligand two curve fitting constants were fit as well as the fraction of the protein stock that is dimeric. For T3 a fourth variable, the amount of proteolized T3 was also fit. More information can be found in Appendix A.2 regarding the least squares fit used to quantify the dimer purity.

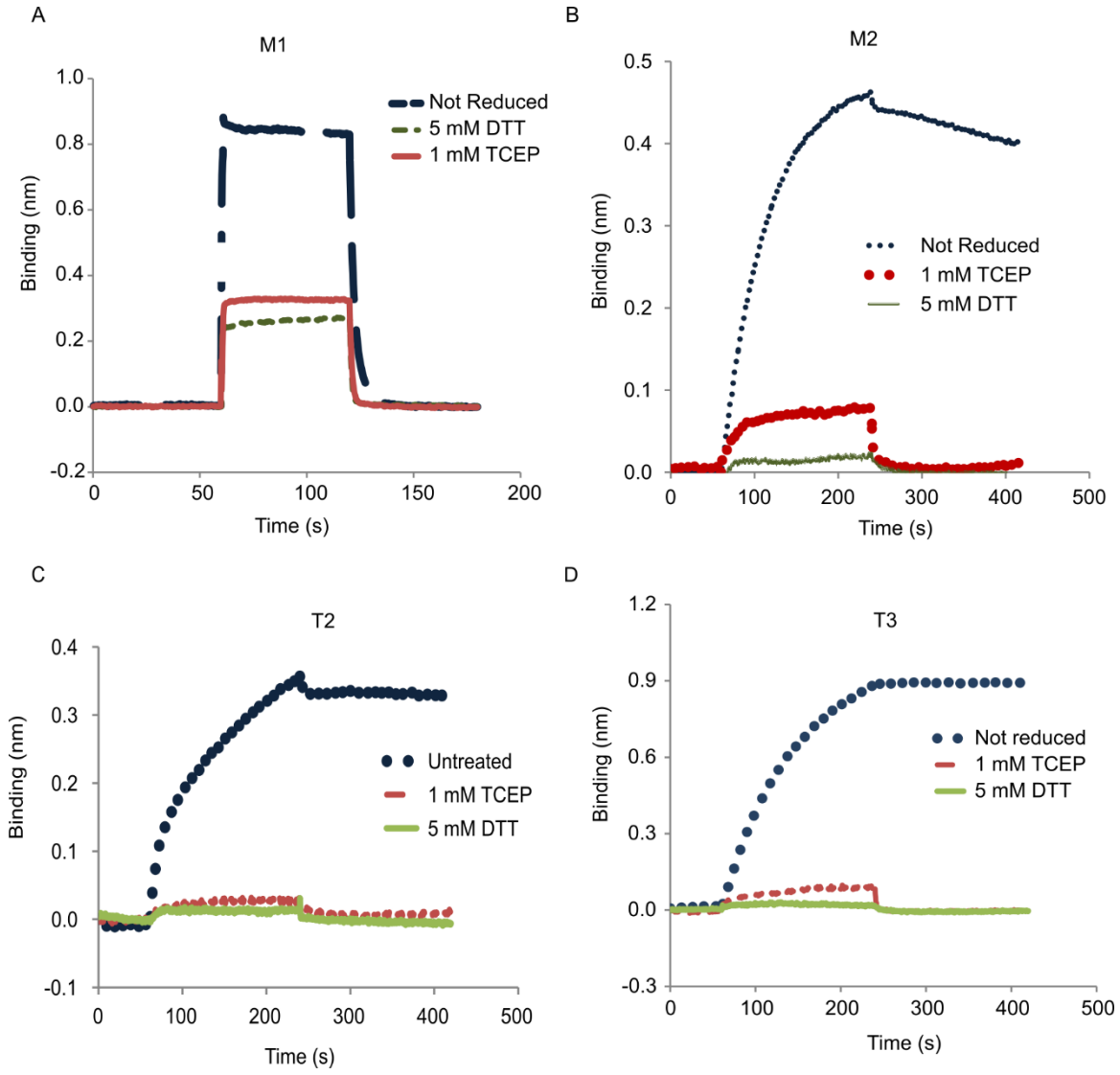


Figure 3.8: Biolayer interferometry data shows the dimeric form of the ligand is functional. For each experiment biotinylated β-catenin was immobilized in exactly the same manner, 1 μM biotinylated β-catenin was loaded for 120 seconds (data for loading not shown) onto a streptavidin sensor and the association and dissociation of the ligand was carried out at the same dilution of protein stock for each of the ligands at each of the conditions. The result of using constant dilution is that the non-reduced form is at a lower concentration, the concentration is halved based on the molecular weight of the dimer and lowered further based on residual monomer in the non-reduced state protein stock. The data represents an equilibration phase with the sensor in buffer, first 60 seconds, an association phase where the ligand is absorbing onto the ligand loaded sensor, 60-120 seconds, and an off-rate phase in which the ligand is de-complexing from the target into a solution containing just buffer. Samples were reduced using 5 mM DTT or 1 mM TCEP. All curves have been baseline subtracted for buffer effects which were accounted for using reduced and non-reduced ligand binding the sensor which has not had β-catenin immobilized on its surface. (A) The M1 ligand loaded at 10 μM reduced and 4.5 μM non-reduced protein. (B) M2 ligand at 50 nM reduced protein and 22 nM non-reduced protein. (C) T2 at 100 nM reduced ligand and 27 nM non-reduced. (D) T3 at 50 nM reduced and 16 nM non-reduced ligand.

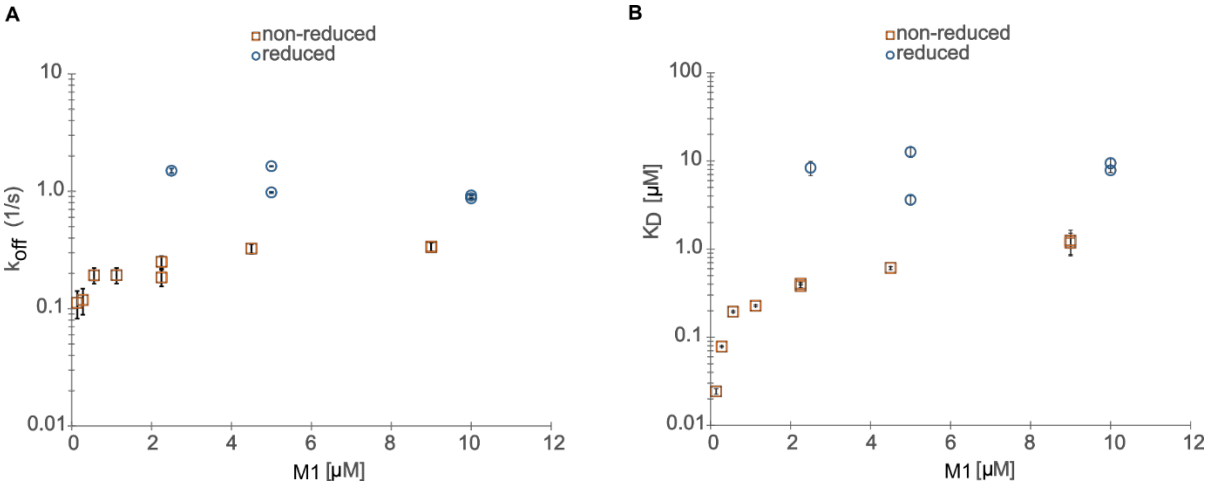


Figure 3.9: Calculated dissociation rate k_{off} and K_D for reduced (1 mM TCEP) and non-reduced M1 based on data recorded on the Blitz™.

The data on the y-axis is given on a log scale for both figures. The x-axis shows the concentration of M1 used during the association step. (A) The dissociation rate constant (k_{off}) as determined by linear regression based one a one to one binding model. The standard error is shown for each data point. (B) The K_D measured by the dissociation rate constant (k_{off}) divided by the association rate constant (k_a). The K_D value is measured based on a local analysis of association and dissociation data at each concentration. The Blitz data which is the basis for this analysis is shown in Appendix 2 **Figure A2.2**.

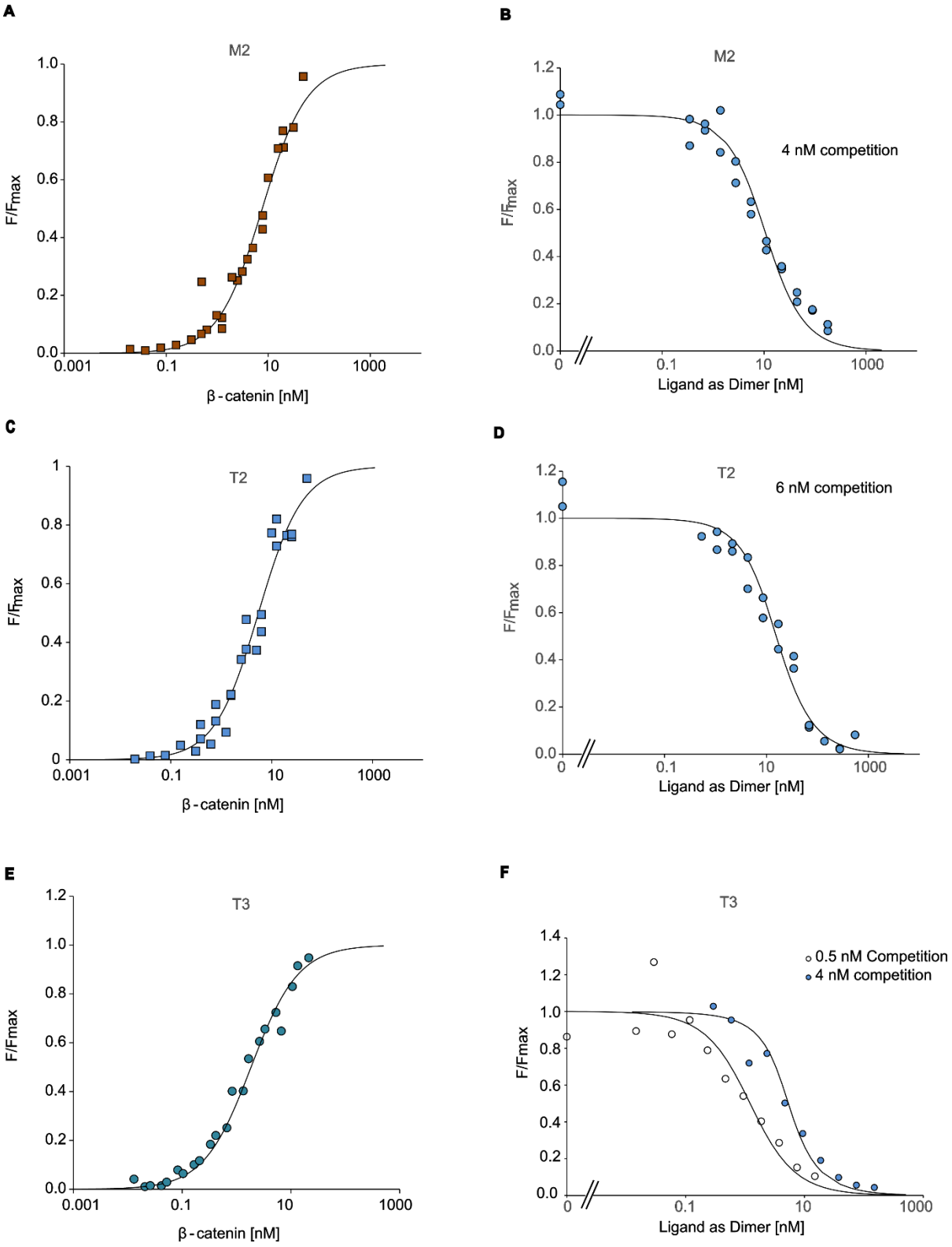

A
M2
F/Fmax
β-catenin [nM]
B
M2
4 nM competition
F/Fmax
Ligand as Dimer [nM]
C
T2
F/Fmax
β-catenin [nM]
D
T2
6 nM competition
F/Fmax
Ligand as Dimer [nM]
E
T3
F/Fmax
β-catenin [nM]
F
T3
0.5 nM Competition
4 nM competition
F/Fmax
Ligand as Dimer [nM]

Figure 3.10: Titration and competition based K_D curves for yeast displayed ligands to full length soluble β-catenin.

Titration based measurements are shown in the left panels and competition assays are in the right panels. Data and K_D based fits (dashed lines) using a one to one binding model are shown for M2 (A,B), T2 (C,D) and T3 (E,F). Titration based measurements were done with each respective yeast displayed ligand and soluble biotinylated β-catenin. The loading of β-catenin onto each respective yeast cell was quantified after secondary labeling with SAPE using cytometry. Competition based measurements were done similarly except in this case a fixed β-catenin concentration is chosen which lies within a sensitive region of the titration based K_D curve and β-catenin is competed from binding the yeast displayed ligand by soluble ligand. This mixture was allowed to equilibrate prior to washes and secondary labeling. Loading of β-catenin onto each yeast cell was quantified after equilibrium was reached. The concentration of β-catenin is shown for each ligand. The competition based method allows you to determine if the ligand behaves similarly in solution as it does on the yeast surface. K_D values determined using both methods are shown Table 3.5.

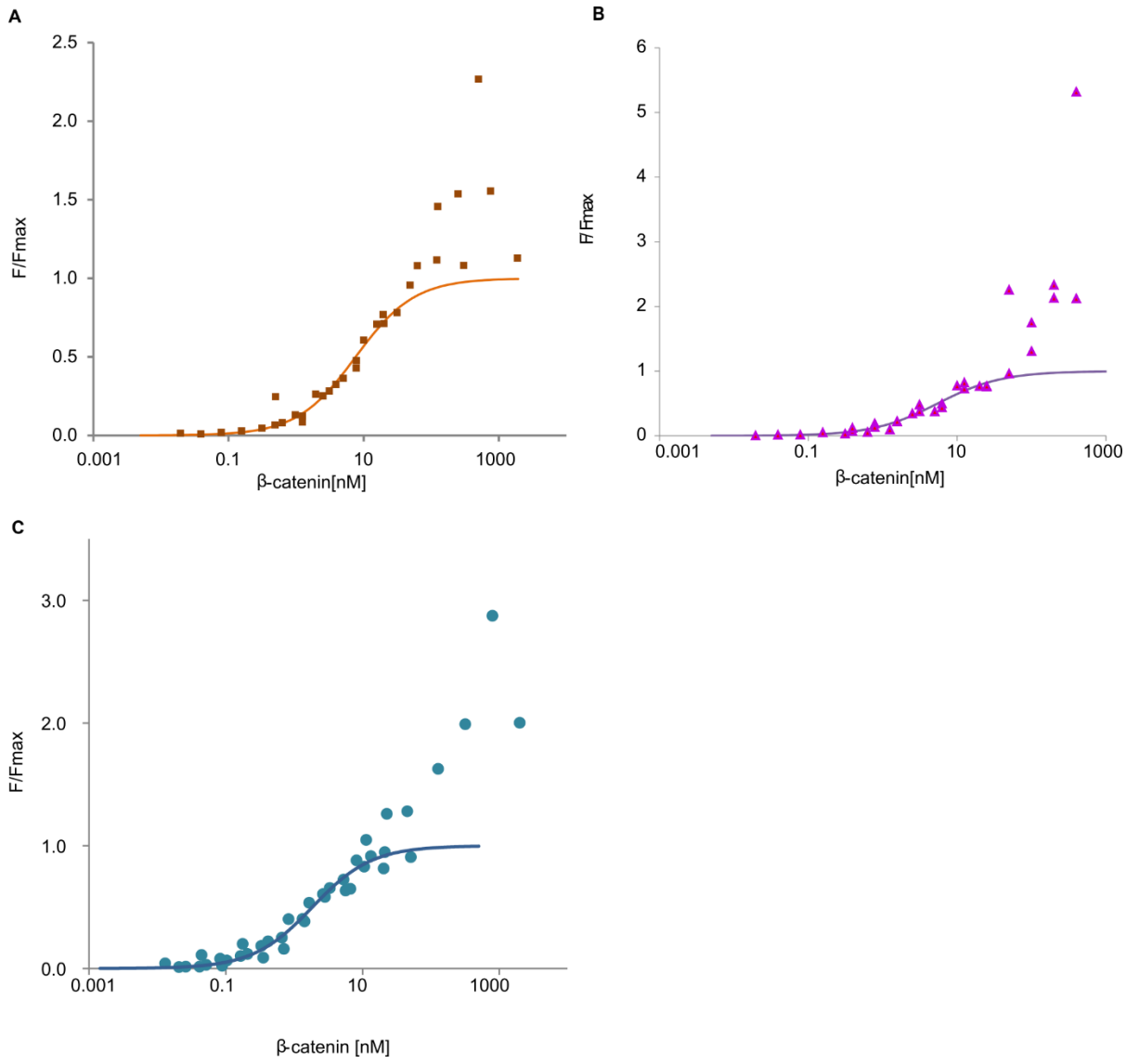


Figure 3.11: Multimodal binding observed in yeast based K_D measuremnts.

The figures show the complete data sets collected for determination of the titration based K_D's for each ligand again M2 is shown in (A), T2 (B) and T3 (C). The K_D fit was limited to data collected below about 0.9 fractional binding as seen in these Figures. It will be shown that each of the ligands in this work is capable of binding two sites on β-catenin therefore at higher concentrations two β-catenin proteins may bind to each site on the ligand. In addition, the yeast displayed ligand may show

up on the surface of yeast in both high affinity dimeric form and as a monomer. The monomeric versions will have lower apparent affinity and accordingly bind β-catenin only at higher concentrations.

Nomenclature	Method of generation	Ligand(s)
M1	Ligand selected to bind BCP sequence: SNQLAWFDTD	M1
M2	Ligand pre-targetted to BCP then affinity matured through mutagenic PCR to bind full length B-catenin.	M1 L54W
T2	One of the highest affinity ligands selected from the library consisting of M1.2 flexibly linked to a naive Sso7d library.	M1.2 linked to 23CTL
T3	One of the highest affinity ligands selected after creating a mutagenic T2 library.	T3-NTL linked to 23CTL

Table 3.4: Nomenclature of β-catenin affinity ligands.

This table shows the names given to each of the ligands characterized, how they were made and for the tandem ligands the names of the individual ligands that they are comprised of.

Protein Name	Oligomeric state	Number of Sso7ds per Ligand	Percent dimer in purrified protein stock	K_D's [nM] (68% Confidence Interval)				
				To full length β-catenin titration (1-781)	To full length β-catenin competition (1-781)	To C-terminal peptide βCP (769-781)	To C-terminal IDR (666-781)	To ARM Domain (151-666)
M1	Dimer	2	89.4 (87.4 - 91.2)			2800 (1800 - 4500) PP*	ND*	470 nM (450-870) *NS?
M2	Dimer	2	88.9 (82.4 - 95.5)	7.6 (5.7-10)	5.4 (3.9 - 7.3)	26000 (18,000-46,000)	150 (100-210)	110 (40-300)
T2	Tandem Dimer	4	55.3 (54.6 - 55.6)	5.5 (4.0-7.7)	5.5 (4.1-7.3)	3100 (1600-6300)	ND*	85 (47-150)
T3	Tandem Dimer	4	65.1 (49.5 - 74.5)	1.9 (1.6-2.1)	0.74 (0.45 -1.2)	2900 (1500-6200)	130 (70-250)	5.0 (3.5-7.0)
T3-NTL	Dimer	2	89 (88.0 - 91.0)				ND*	93 (67-130)

ND* not detected
PP* previously published
*NS? Potentially a non-specific interaction

Table 3.5: Ligand K_D's and properties.

Shown are all the K_D measurements made for the ligands described in this chapter. The table shows the number of Sso7s in each construct, the percentage of dimeric protein observed in the soluble protein samples as well as the K_D along with 68% percent confidence interval for each interaction measurement made to each domain of β-catenin. All K_Ds were determined using a one to one binding model using a global least squares fit method.

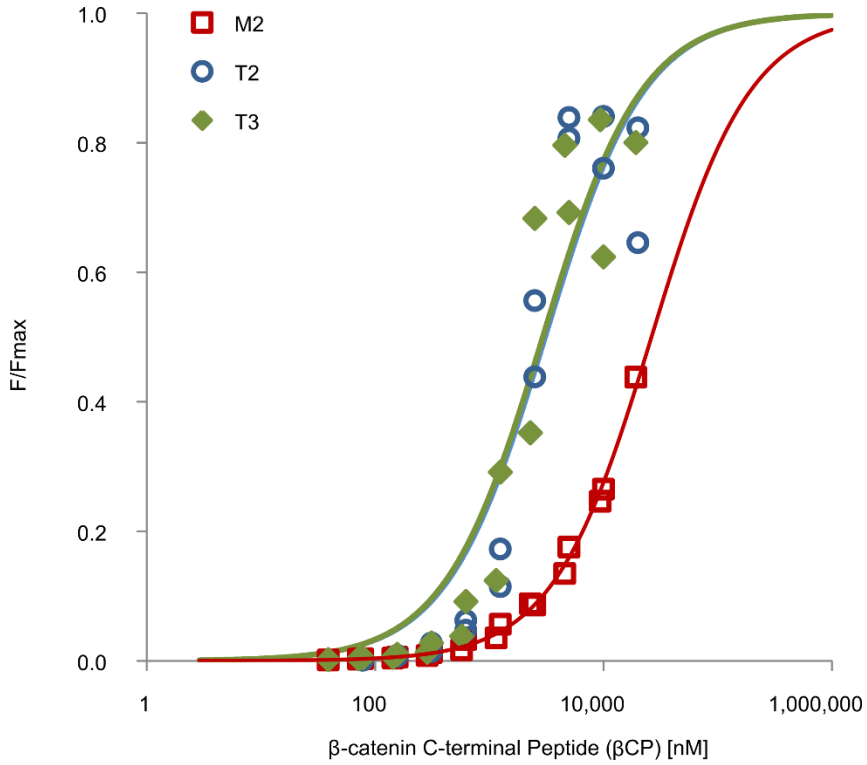


Figure 3.12: K_D curves for yeast displayed ligands to soluble βCP.

The figure shows cytometry based K_D curves which characterize the binding of yeast displayed M2, T2 and T3 to the biotinylated C-terminal β-catenin peptide sequence, βCP, (residues 770-781 of β-catenin). SAPE was used to quantify the presence of βCP loaded on the yeast surface at each concentration for each ligand. T2 and T3 both have as their N-terminal Ligand (NTL) the M1.2 ligand which differs from M1 by a Q62K mutation and deletion of 63K. These mutations are not observed to alter the affinity of either tandem ligand to βCP relative to M1. Furthermore, this data suggests that the tandem ligands did not gain affinity to β-catenin by gaining affinity to βCP. The M2 ligand despite having dramatically higher affinity to β-catenin loses affinity to the βCP sequence. The calculated K_D's are shown in Table 3.5.

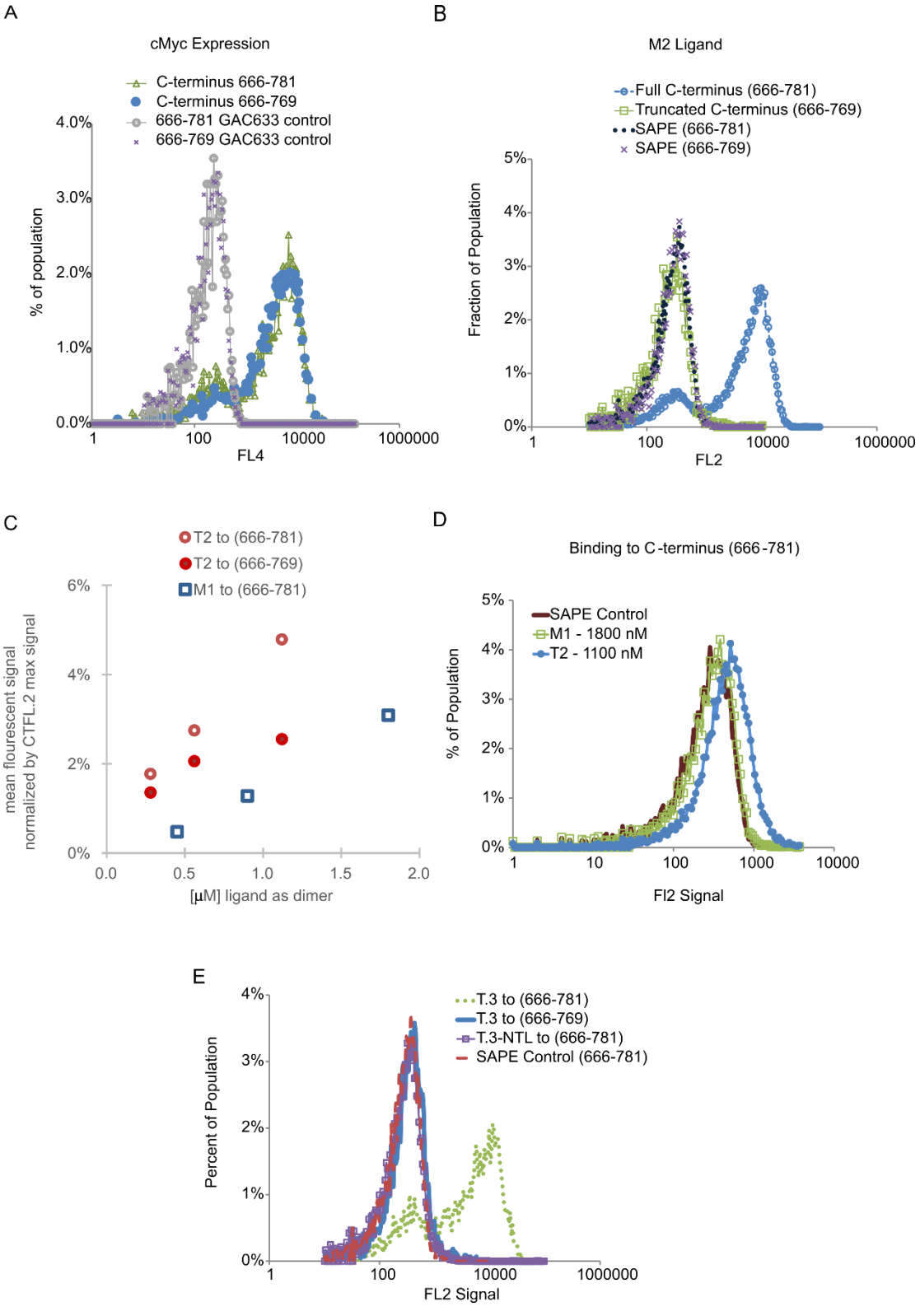

A
cMyc Expression
C-terminus 666-781
C-terminus 666-769
666-781 GAC633 control
666-769 GAC633 control
% of population
FL4
B
M2 Ligand
Full C-terminus (666-781)
Truncated C-terminus (666-769)
SAPE (666-781)
SAPE (666-769)
Fraction of Population
FL2
C
T2 to (666-781)
T2 to (666-769)
M1 to (666-781)
mean flourescent signal normalized by CTFL.2 max signal
[µM] ligand as dimer
D
Binding to C-terminus (666-781)
SAPE Control
M1 - 1800 nM
T2 - 1100 nM
% of Population
FI2 Signal
E
T.3 to (666-781)
T.3 to (666-769)
T.3-NTL to (666-781)
SAPE Control (666-781)
Percent of Population
FL2 Signal

Figure 3.13: Cytometry histograms for ligands binding the C-terminal IDR.

To determine the specificity of the ligands generated and to determine the affinity of the ligands to the C-terminal IDR a full length C-terminal construct (residues 666-781) of β-catenin was displayed on yeast as well as a βCP truncated construct (residues 666-769). Biotinylated ligand and Streptavidin, R-Phycoerythrin (SAPE) was used as the secondary reagent used to quantify the amount of ligand bound to the yeast displayed β-catenin construct. (A) Yeast surface display expression level of the cMy- tag which appears following the IDR construct. A chicken anti-cMyc primary antibody is used and a Dylight 633 conjugated goat anti-chicken secondary were used to detect the presence of the cMyc tag; the secondary control shows very little binding of the fluorescent antibody to both C-terminal IDR's. (B) Shows specific binding of soluble M2 protein (450 nM) to CTFT and no signal to the truncated construct (666-769). The controls for SAPE binding the C-terminal IDR's are also shown. (C) Shows that the T2 and M1 ligand (100 - 1,800 nM) gives relatively little binding signal to the full length C-terminal construct (666-781). The signal from M1 is normalized to the maximum signal observed for M2; the T2 signal is normalized relative to T3. The binding signal scales non-linearly with the concentration of either ligand the magnitude of the background subtracted signal is at most 5% of the signal from the M2 ligand. (D) Cytometry based histograms for the highest concentration points shown in Figure C. T2 (1.1 μM) and M1 (1.8 μM) give very low binding signal to either C-terminal IDR. (E) T3 (500 nM) again shows strong specific labeling to the full length C-terminal IDR (666-781) and very little signal to the truncated construct (666-769). The T3 N-terminal ligand (T3-NTL) was also, expressed, purified, biotinylated and its binding to the C-terminal IDRs was measured at (670 nM), negligible binding is observed to the C-terminal IDR (666-781).

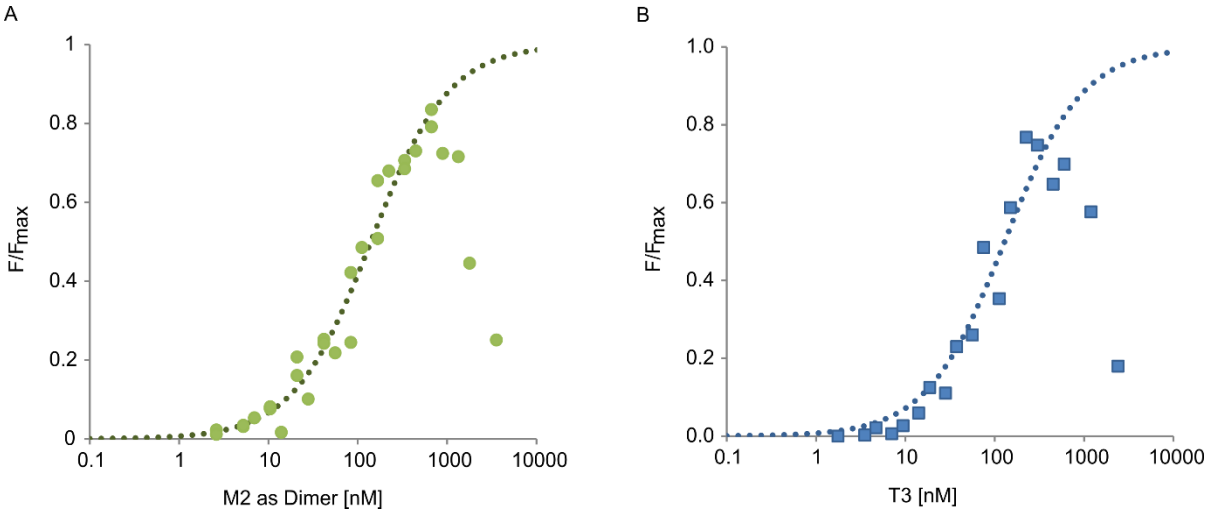


Figure 3.14: K_D curves for soluble M2 and T3 to yeast displayed C-terminal IDR (666-781).

Ligands were N-terminally biotinylated and ligand binding to yeast displaying the C-terminal IDR of β-catenin (666-781) was quantified using SAPE as the secondary reagent. (A) Titration curve for for M2. based on the concentration of the ligand as a dimer B) Titration curve for T3, again the concentration is as a dimer. The dotted line in both panels represents the response generated for a ligand with a K_D based on a least squares fit to the data. Both ligands show a strong hook affect at higher concentrations. The three highest concentration points for M2 and the two highest concentration points for T3 were not included in the dataset used for the K_D calculation. Calculated K_D's and 68% confidence intervals are shown in Table 3.5.

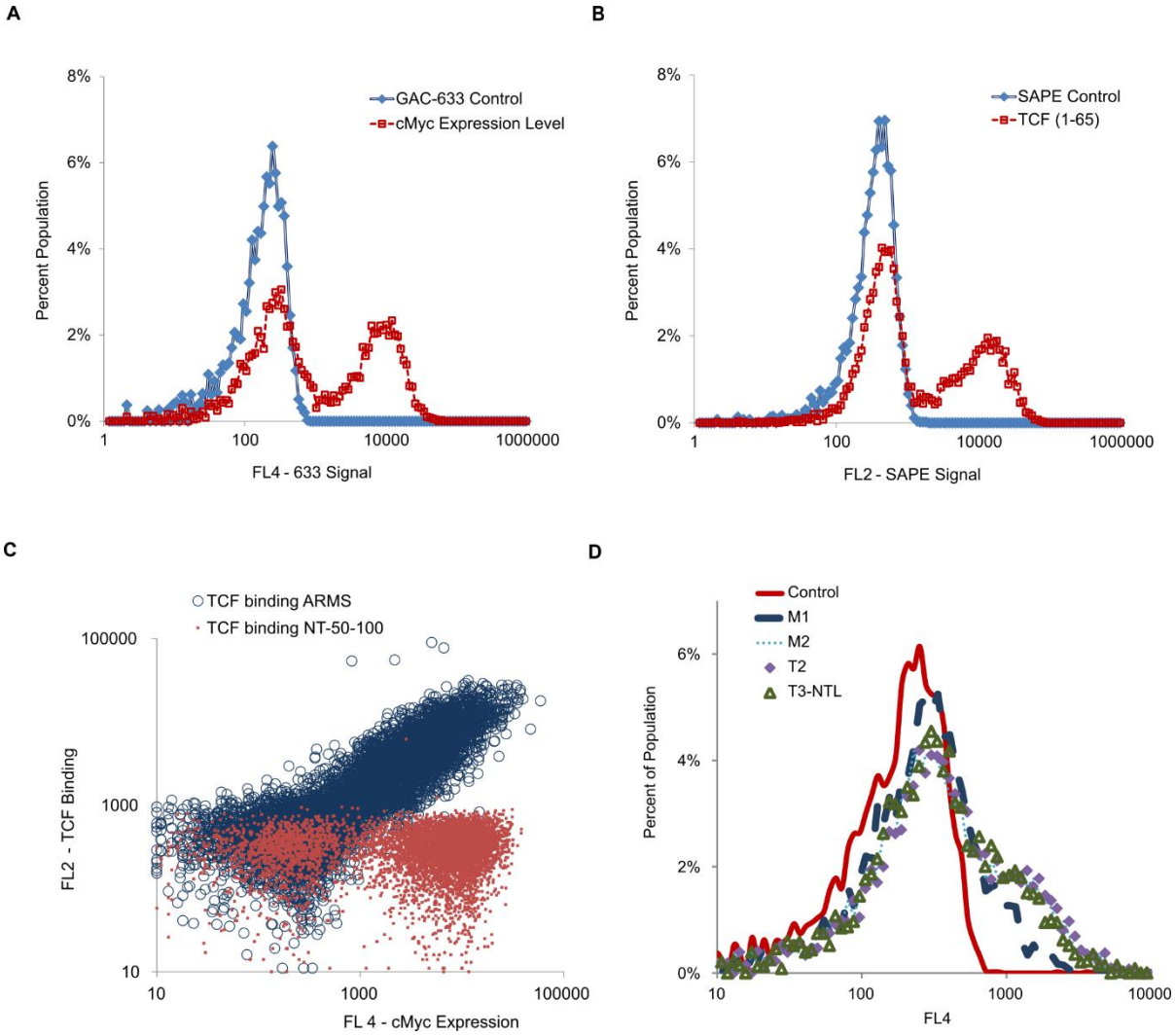


Figure 3.15: Yeast surface display of the ARM domain, expression controls and binding by ligands.

(A) Significant cMyc expression level is observed for the ARM domain (B) The TCF (1-65) domain is a canonical binding partner to the ARM domain. It was recombinately expressed and purified as a GST fusion, biotinylated (using cysteine labeling, maleamide chemistry) and is seen here to bind the ARM domain using SAPE as the secondary reagent. The control is shown for SAPE binding the ARM domain. (C) The TCF+GST construct is observed specifically to bind the yeast displayed ARM

domain and not to yeast displayed N-terminal construct (Residues 50-100). The figure shows cMyc labeling by an anti-cMyc antibody and Dylight 633 conjugated secondary antibody and binding to biotinylated GST+TCF (1-65) using SAPE. This validates the use of the ARM construct to determine domain specific binding of the engineered ligands in this work. (D) Histogram showing the binding of the ligands to the yeast displayed ARM domain. The SAPE signal is shown for a primary labeling done with N-terminally biotinylated ligands: M1 (900 nM) and M2 (900 nM), T2 (550 nM) and T3-NTL (900 nM).

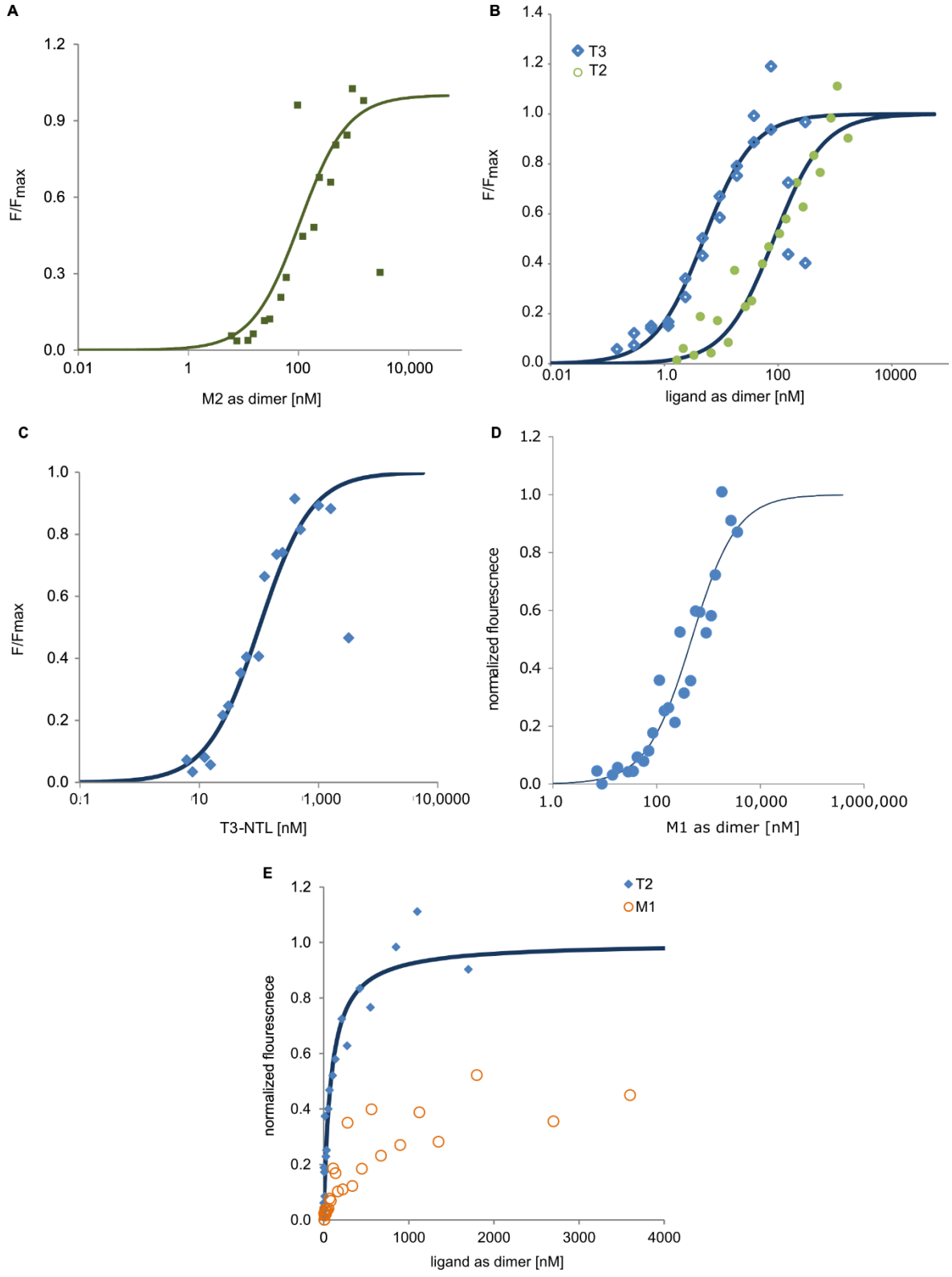
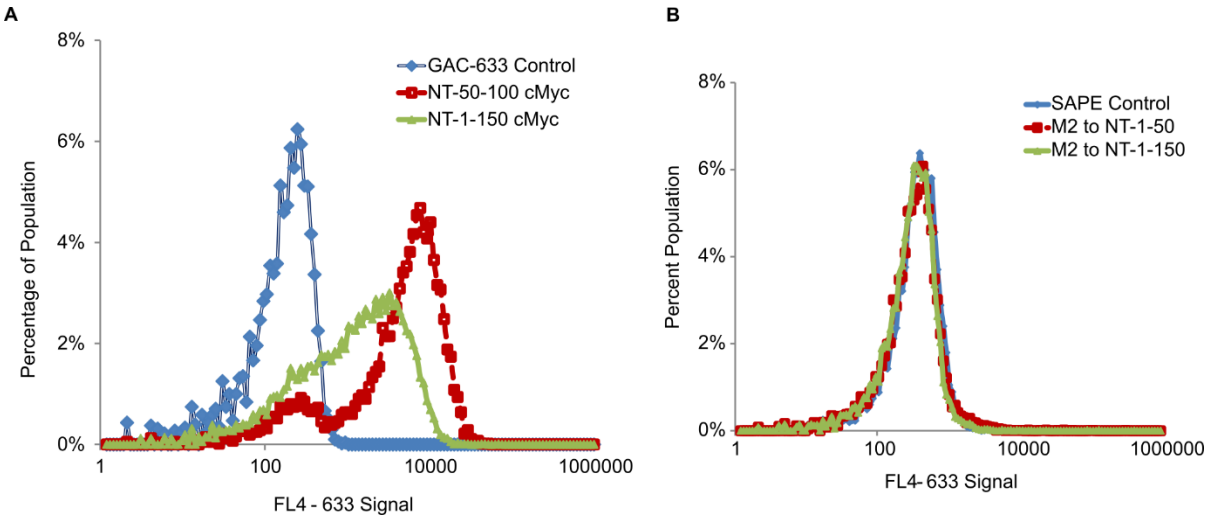

A
F/Fmax
M2 as dimer [nM]
B
T3
T2
F/Fmax
ligand as dimer [nM]
C
F/Fmax
T3-NTL [nM]
D
normalized flourescnece
M1 as dimer [nM]
E
T2
M1
normalized flourescnece
ligand as dimer [nM]

Figure 3.16: Titration based K_D's curves for soluble ligands binding the yeast displayed ARM domain.

N-terminally biotinylated ligands (A) M2 (B) T3 and T2 (C) T3-NTL, were labelled onto the yeast displayed ARM. The labeling concentration of the biotinylated dimeric form of each ligand with the corresponding SAPE secondary signal, quantified by cytometry is shown. The K_D values determined by a global least square fit to a one to one binding model are shown in Table 4. (D) (E)



Show a bar chart in figure c with binding data from other ligands?

Figure 3.17: Expression level of N-terminal IDR's on yeast and binding by soluble ligands.

The figure indicates that the ligands generated in this work do not bind the N-terminal IDR of β-catenin. (A) Cytometry histogram showing expression levels of the two β-catenin N-terminal IDRs, residues 50-100 and 1-150. Primary labeling was done with chicken anti-cMyc antibody and secondary labeling with goat anti-chicken 633 antibody. The figure shows that the sequence representing β-catenin residues 50-100 expressed very well and that the full length N-terminal

intrinsically disordered region (IDR) expressed to a lesser extent but still to a significant enough extent to determine if the ligands made target this domain. (B) Shows the signal from SAPE secondary labeling to a primary labeling done with N-terminally biotinylated M2 (800 nM) to both of the N-terminal IDRs shown in (A). Cytometry data was also collected for labeling the N-terminal 1-150 IDR with T2 and the T3-NTL (550 nM and 800 nM). This data is not shown but the same result is observed, no binding to the N-terminal IDR of β-catenin.

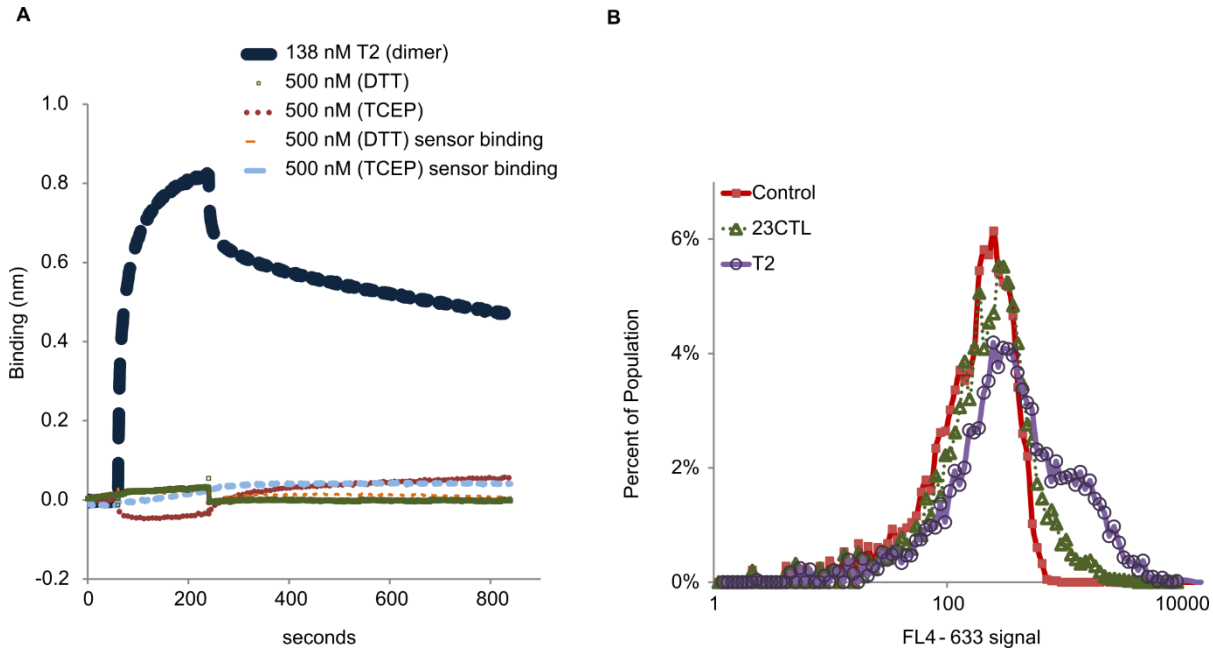


Figure 3.18: 23CTL requires avidity

(A) Blitz data for the reduced and non-reduced T2 at 500 nM reduced ligand and 138 nM non-reduced ligand. The T2 binding interaction with the ARM domain is attributed to all parts of the construct after the N-terminal M1.2 ligand. As T2 has a KD to the ARM domain of 85 nM if the ARM binding domain of T2 were functional in the monomeric form a stronger association signal should be seen at 500 nM labeling. (B) Cytometry data for T2 (250 nM as non-reduced dimer) and 23CTL (1 μM) binding the ARM domain. The interaction is measured using mouse anti-his antibody,

targeting the his-tag on each ligand and donkey anti-mouse 633 secondary. The low binding signal observed for 23CTL when soluble expressed on its own indicates it is more functional when a part of the tandem dimer T2.

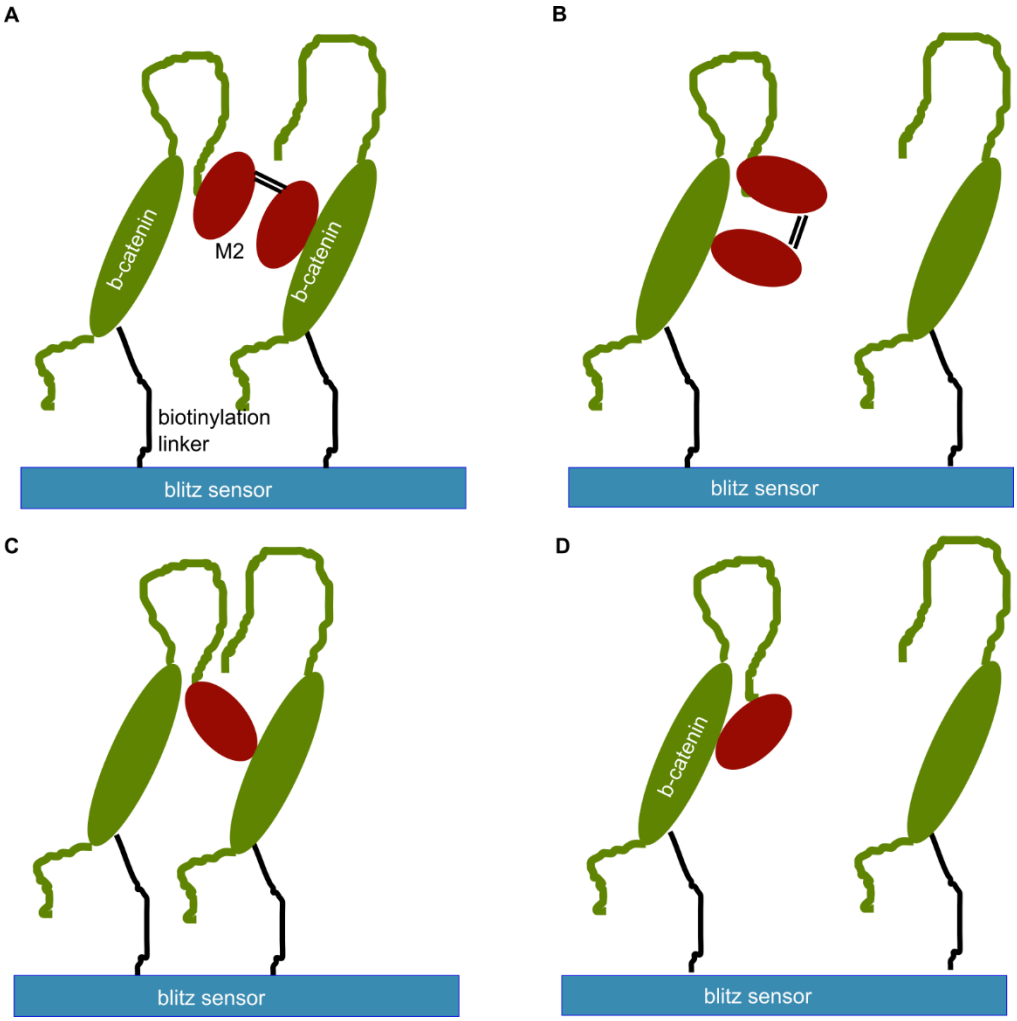


Figure 3.19: Diagrams depicting how monomeric or dimeric M2 may interact with β-catenin immobilized on a Blitz sensor.

The disordered N-terminal and C-terminal regions of β-catenin are shown. (A) Intermolecular avidity generated by one dimeric ligand binding two surface bound β-catenin proteins. (B) One dimeric

ligand binding one β-catenin protein. (C) a monomeric ligand binding two β-catenin proteins. (D) a monomeric protein binding one β-catenin protein by simultaneously engaging two epitopes.

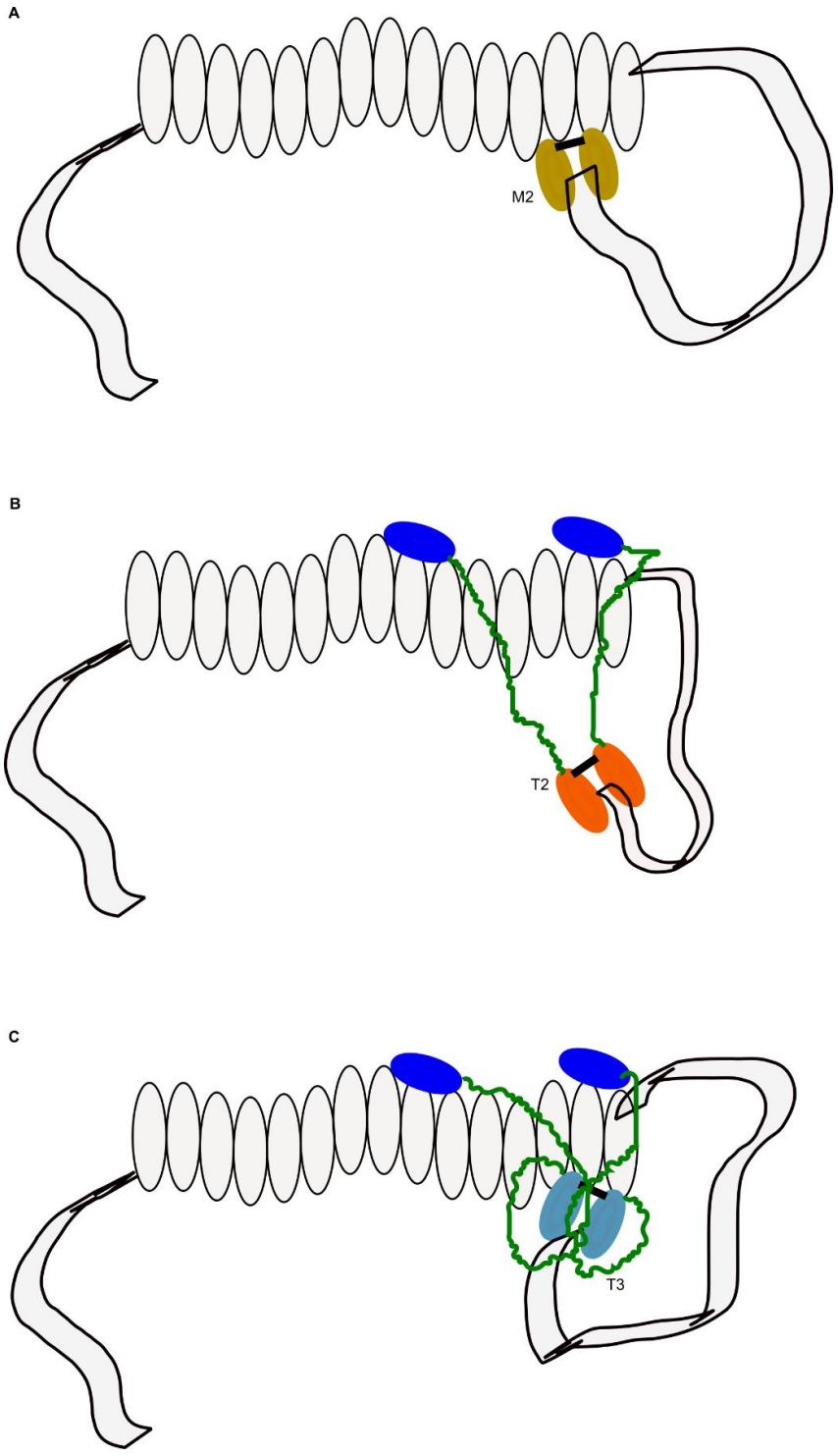


Figure 3.20: Proposed binding mechanisms of the ligands engineered in this work to β-catenin.

(A) The data collected indicates that M2 binds simultaneously to the ARM domain and the βCP epitope. The domain based K_D's of M2 to the ARM domain and C-terminal IDR are both over an order of magnitude lower than the affinity observed to the entire β-catenin protein. This indicates both binding sites need to be engaged in order to achieve the 4.8 nM measured in the competition experiment. (B) The T2 model involves simultaneous binding of both of the T2 C-terminal ligands (23CTL) to unique epitopes on the ARM domain and M1.2 engaging the βCP epitope (last 12 residues of β-catenin). Like M2, T2's K_D to the ARM domain is over an order of magnitude higher than its K_D to the entire β-catenin protein. This would indicate that M1.2 also does have affinity for the C-terminus despite T2 having a minimal interaction with the C-terminal IDR. The data also indicates that 23CTL requires the avidity provided to it by the dimeric nature of T2 in order to achieve the observed ARM K_D. (C) Mutations in the T3-N-terminal ligand cause both increased affinity to the ARM domain and to the C-terminal IDR. The net effect of both changes is for T3's affinity to improve 7 fold relative to T2. The binding mechanism is unique from M2 in that binding to the C-terminal IDR is not observed unless the T3-NTL is expressed as a fusion with the linker region and 23CTL. This leads to the conclusion that the linker or 23CTL acts with the T3-NTL to enable higher affinity C-terminal binding.

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

The motivation of this thesis was to develop methods through which avidity could be more readily harnessed to create high affinity ligands. Chapter 1 looked at successful methods for engineering high affinity ligands and establishes that multivalent ligands have been extremely successful in this regard. Despite this precedent, few publications exist which have used combinatorial selections to generate these ligands. Using different strategies in Chapters 2 and 3, each targeting dramatically different targets, we show that the yeast display platform can be harnessed to generate combinatorial bivalent libraries. Chapter 2 targeted the small, folded, lysozyme protein (14.3 kDa). Chapter 3 targeted the much larger β-catenin protein (85 kDa) which has extended inherently disordered regions. For both cases the resulting high affinity ligands selected were found to bind with intramolecular affinity, yielding the chelating ligands which were sought. The work in Chapter 3 also identified a functional dimeric disulfide bonded Sso7d ligand. This was an unexpected result as dimerization of engineered ligands is typically considered an unwanted artifact. Here we show that dimerization was selected for as it resulted in the highest affinity ligands present in the respective libraries from which they were drawn.

Chapter 2 details a study to select bivalent ligands in a combinatorial manner using yeast surface display. We do this by randomly pairing low affinity ligands. We find that there is systematic synergy of the linked ligands which results in a substantial enrichment in the number of high affinity ligands generated relative to a monovalent version of the same library. This result was shown to not be a result of intermolecular avidity caused by two different yeast surface display fusions binding the same lysozyme protein. This was demonstrated through both the dual display yeast display construct and through a competitive binding experiment. The former shows that for the two affinity ligands in BVL-1 when both are displayed simultaneously on distinct display fusions that they are not capable of simultaneously binding lysozyme. This indicates that intermolecular affinity was not selected for

as it does not appear to be possible for ligands on two different display fusions to bind one lysozyme protein. Rather intramolecular affinity of the two ligands within each bivalent ligand were both selected to bind lysozyme simultaneously. The later, competitive binding experiment confirms the desired high affinity chelating interaction in the soluble phase. We go on to show that a selected bivalent ligand can be broken into its subcomponent and used in a label free quantitative bio-sensing assay. Adaptation of the GFP tripartite system for this purposes enabled mix and read detection in an impure system. The biosensor was successful as the ligands were selected to bind distinct epitopes synergistically. Importantly, the bivalent ligand selected retained high thermostability.

The promising results demonstrated in Chapter 2 show that the bivalent library systematically generates synergy by linking low affinity ligands. This result, while encouraging, is likely not yet optimized. This conclusions is based on the observed 0.2 mM effective concentration measured for BVL-1. This concentration is defined by the product of the K_Ds of the individual ligands divided by the K_D of the resulting bivalent ligand. Previous work suggests that this could easily be a factor of 100 or possibly 1000 fold greater[40]. Increases in the effective concentration would result in the bivalent ligand having a K_D which is reduced by that factor. A study directed at determining how to improve the affinity of BVL-1, and or other ligands identified in the best binder pool of the bivalent library identified, is therefore worthwhile. Potentially a method of systematically generating improved synergy could be identified. For instance, perhaps the linker used in this study is not ideal. An alternative SH3 linking domain is found in nature which has a flexible but structured alpha-helical form[79]. Perhaps using this linker would generate higher effective concentrations. Alternatively, favorable high affinity mutations could be sought in an affinity maturation scheme as was done in Chapter 3, in which the linker was included in the affinity maturation of T3. This would allow us to see where mutations are favored, in the linker, in the engineered binding face or at other locations on

the Sso7d ligand. Toward this end a crystal structure of the individual ligands of BVL-1 and lysozyme could be beneficial in determining determine how to improve the synergy of BVL-1 ligands in binding lysozyme. For this crystal structure we believe it would be best to pursue each of the individual ligands with lysozyme individually rather than in the linked state. A crystallization effort for BVL-1 and lysozyme was undertaken by a collaborator but, after extensive efforts, was unsuccessful. Once an optimized strategy is developed for generating bivalent ligands perhaps studies with useful biological outcomes could be targeted.

Chapter 3 details several interesting findings for binding intrinsically disordered domains with ligands generated using the Sso7d scaffold. These are reviewed in detail in **Figure 3.20**. We see three main observations from this study: (1) gain of an epitope as a mode of affinity maturation when peptide targeted Sso7d ligands are mutated and selected to bind the protein of interest; (2) peptide pre-targeting to an IDR is effective but not perfect as strong differential binding of the peptide target and IDR are observed for the ligands characterized; (3) affinity maturation of a flexibly linked multivalent ligand to an IDR can generate ligands with extended paratopes (binding surface on the ligand), which cannot be readily broken into components. This third unexpected results is likely a consequence of the conformational freedom in the system, which exist as both the bivalent ligand and the IDR can likely adopt an array of conformations. It would be interesting to determine if any of these same observations are observed when targeting other proteins with an IDR or an IDP. It is not clear if the dimeric nature of the Sso7d ligands selected in Chapter 3 or the repeat regions within β-catenin contributed to the results observed.

Regarding the ligands generated in Chapter 3, it would be valuable to determine if when expressed in human cell lines what the degree of dimerization observed is. This information is necessary to complete a second study, initiated by a collaborator, which is directed at determining the intracellular

effect of the ligands generated in this work. Preliminary results show that in several human model cell lines not one of the ligands generated affects β-catenin's TCF dependent transcriptional activity. T-cell Factor (TCF) is a cofactor of β-catenin which binds both β-catenin and targeted DNA sites. This would indicate that the ligands generated may serve as useful β-catenin tags that do not disrupt β-catenin's function. We have considered using this as tool by which the effect of other β-catenin protein affinity ligands could be studied. Specifically, we have questioned whether we could generate a FRET based readout to determine if we could find an inhibitor that blocks β-catenin from entering the nucleus. The mechanism by which β-catenin enters the nucleus is currently only speculated at[69]. Current data suggests that β-catenin interacts with the nuclear pore complex directly and not through one of its many cofactors[80]. Finding inhibitors of this nuclear import would be a powerful tool for drug discovery as increased nuclear concentrations of β-catenin are associated with higher grade tumors[80]. We have considered that one of the ligands generated in this work could be fused to TagGFP and expressed solubly, a FRET based sort could then be carried out with a yeast displayed library of ligands which bind β-catenin and are fused to TagRFP. The productive FRET pairs could be screened after transfection into a human cell line, for diminished presence (FRET signal) of β-catenin in the nucleus. A bicistronic vector could be used which contains both FRET pairs, in this way equal amounts of both ligands will be delivered to each transfected cell. Both ligands could be expressed without nuclear localization sequences and the location and concentration of β-catenin could then be tracked after WNT stimulation of the cells, real time. Differences in the localization of the FRET signal could result in identifying RFP tagged ligands that modulate β-catenin's behavior. This approach could potentially identify the desired β-catenin nuclear inhibitor and perhaps a host of other ligands which affect β-catenin's localization.

Another important aspect of the Chapter 3 work is the selection of functional dimeric ligands which form on the surface of yeast. A hypothesis around this observation is that prior to being exported to the yeast surface M2 ligands self-associate and form an intermolecular-disulfide bond. This results in co-localization of the two dimerized M2 ligands on the surface of yeast. The same co-localization of ligands likely does not occur without the driving force of self-affinity. Accordingly, determining if yeast displayed ligands are co-localized should enable the selection of monomeric or dimeric ligands. It would be useful to screen for dimerization as this would remove the trial and error approach currently used to determine if a ligand is monomeric or dimeric. The current technique is to clone the yeast displayed ligand into an expression vector, express and purify the mutant and then evaluate the oligomeric state in a size exclusion column. One method to screen for co-localization of displayed ligands would be to develop a molecular ruler which would act as a biosensor. The biosensor would be comprised of two identical and linked scaffold derived ligands targeted to the HA-tag. The HA-tag ligands would need to be selected to have low affinity when only monovalent binding occurs but bind with high affinity when the two linked ligands bind distinct HA tags. This bivalent protein could be fused to a monomeric fluorescent protein such as TagGFP. FACS could then be used to select for GFP positive or negative yeast cells of interest. Additional examples of dimeric and monomeric ligands selected by yeast surface display exist in the literature and could serve as additional positive and negative controls to validate the tool[13,15]. **Figure 4.1** illustrates the labeling technique described. Linker length would have to be optimized so as to provide enough length such that bivalent binding of the two co-localized HA tags is possible.

Alternatively, constructs already present in the Rao Lab could be used to create the molecular ruler based biosensor. For instance, the PLCγ1 SH2 domain could be fused to Renilla GFP. Renilla GFP is an obligate dimer, as such the SH2 domain would appear twice in the dimeric construct[81]. This

would give rise to avidity such that if two EGFR epitopes (ADEYLIPQQGFF) are bound simultaneously a positive signal should be generated at a low labeling concentration. The low affinity of the wild type PLCγ1 SH2 domain should result in no signal if the ligands epitopes are not co-localized. The PLCγ1 SH2 tag would have to be inserted in the yeast surface display construct for this to work. Another alternative would be to evaluate the use of pegylated biotin linkers. In this case PLCγ1 SH2 could be biotinylated and pre-incubated with streptavidin R-Phycoerythrin to create a multivalent construct. Biotin chemically fused to pegylated linkers 20 angstrom in length with amine conjugating moieties are commercially available and could be evaluated. Many ligands and formats are possible and should be considered so as to reduce the time required to generate the biosensor.

The discovery that the M1 ligand is a disulfide bonded homodimeric ligand is interesting and suggests further study. Both M1, M2 and T3-NTL were selected in the context of having to be functional homodimers. The diversity and synergy could be greatly expanded by generating heterodimers. A simple and quick way to test this would be to generate the library from which M2 came from using two yeast display plasmids. Leucine and tryptophan yeast display plasmids were used in Chapter 2 and are readily available. This would enable each yeast cell to bear two different M1 derivative plasmids enabling the formation of a heterodimer. Expressing each construct with a different fluorescent protein that could be used as FRET pairs could allow for the selection of high levels of FRET. TagGFP and TagRFP are such pairs as has already been mentioned. High FRET signal could potentially correlate with a high ratio of heterodimer to homodimer. **Figure 4.3** illustrates how this FRET based sort could be carried out. Such a selection should enable self-assembly of the heterodimeric pairs of ligands in the soluble state. The selected pairs could then be purified separately, reduced and then allowed to self-assemble during dialysis. Also, incorporating a FRET selection prior to target binding could elucidate which residues are important for dimerization as ligands with both high and

low fret could be sorted and compared. Similarly, sequencing ligands which gain and lose affinity, while maintaining FRET, could help to determine which sequences are important for target binding. If the site of the paratope within the scaffold can be identified and distinguished from residues important to disulfide bond formation can be identified a new naïve library could be generated. This would be a disulfide induced heterodimeric library with potentially 5-10 library positions. This library could be evaluated for its ability to bind other peptide or IDR sequences. A potentially exploitable characteristic of this system would be that binding could be controlled by the presence of a reducing agent. Alternatively, lessons from this study could be applied such that a non-natural amino acid could be incorporated instead of the cysteine allowing dimerization to be controlled through an alternative, reversible chemistry.

A possible next step from the work in Chapter 3 would be to direct a naïve library to an epitope within an IDR or IDP and then carry out an intracellular selection for function. This could potentially be done in a yeast two hybrid screen between a known IDP or protein containing an IDR and a known co-factor. Pre-targeted ligands could be screened for an inhibitory affect such that interactions which are successfully blocked do not express GFP. Alternatively, if a known interaction partner does not exist perhaps a selection could be set up in a model human or mammalian cell line which naturally expresses the target of interest. An alphavirus transfection system may be adaptable for this purpose. A previous study shows that it is possible to infect a10^6 mammalian cells, each with DNA coding for one member of a library of proteins[82]. As the pre-targeted ligands will have a reduced diversity this is likely sufficient to identify a ligand which has a desired effect. A read-out would be required such that ligands with desired properties could be selected. There would be challenges with an intracellular selection in this manner, but it would advance the field of engineering affinity ligands as typically screens for intracellular functionality are done using just a few select

ligands. In these cases, affinity is generated using purified product target protein and then screens for functionality are carried out using a small group of ligands[47,48]. Alternatively, an extracellular target could be targeted. It is not uncommon for an extracellular receptor to have a disordered region, the BMP receptor is such an example. A peptide, corresponding to the disordered region involved in BMP binding could be used to pre-target a library to that location[83]. An affinity matured library could then be put into the phage display system and panned on BMP displaying cells directly. Phage display libraries have already been panned over mammalian cells displaying targets of interest previously[84].

Lastly, an effort to crystalize β-catenin with M2 would perhaps be successful. The ligand may be able to stabilize the C-terminus such that it is possible to crystalize the intrinsically disordered, and non-crystallizable, domain.

4.1 Figures and Tables

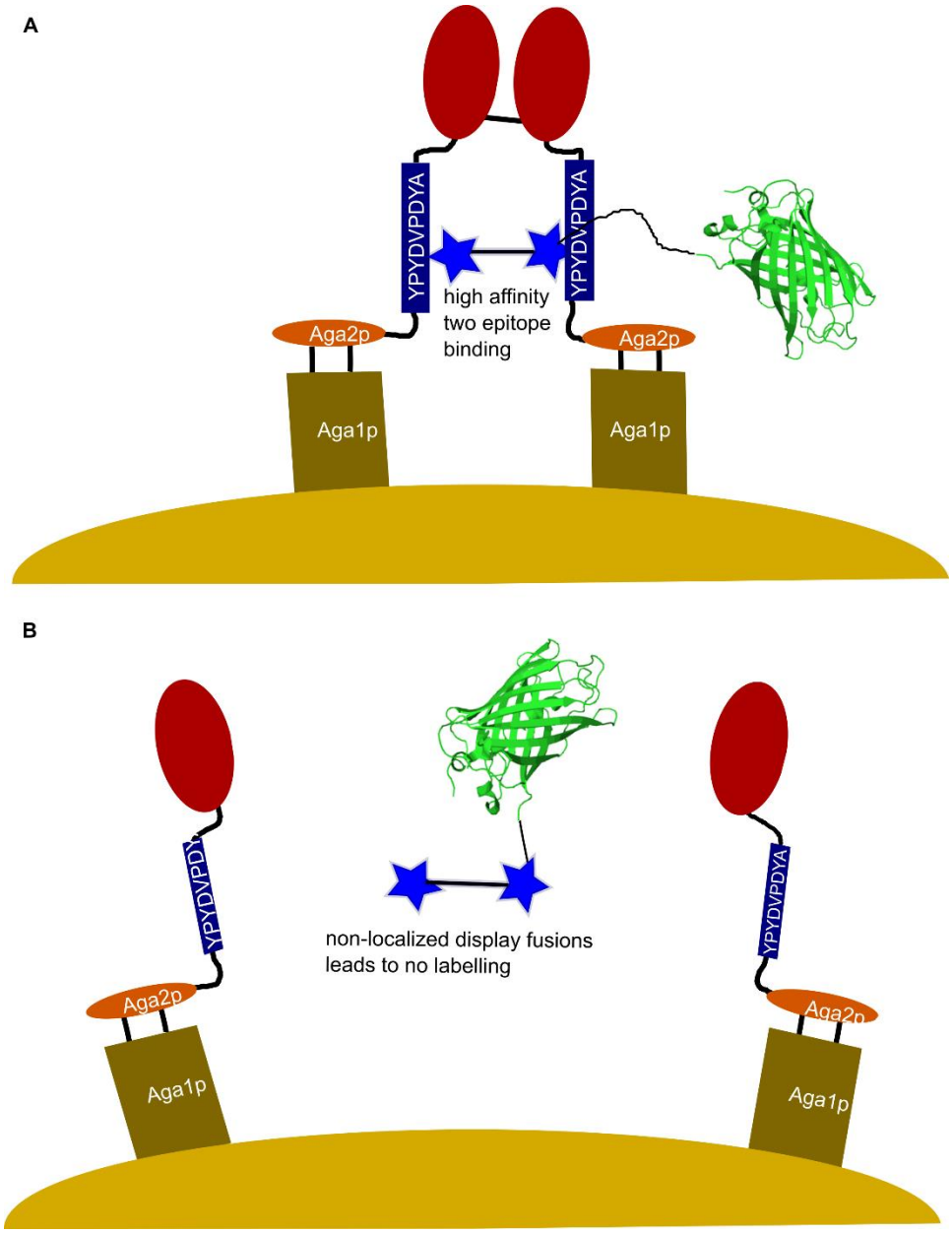


Figure 4.1: A sensor for ligand dimerization in the yeast surface display system.

Ligand dimerization results in co-localization of yeast surface display fusions, this in turn results in the HA tags from two separate fusions being in proximity of each other. The TagGFP biosensor

shown would only give signal when this HA-tag co-localization has taken place. (A) Positive labeling would be measured by a cytometer and is made possible because the two ligands which bind the HA tag are able to do so simultaneously. The resulting avidity would generate a binding signal despite the use of a labeling concentration well below the K_D of the monovalent HA-tag ligand for the HA-tag. (B) Yeast surface display fusions which are not co-localized result in HA-tags which cannot be labelled by the sensor at the same labeling concentration used in (A).

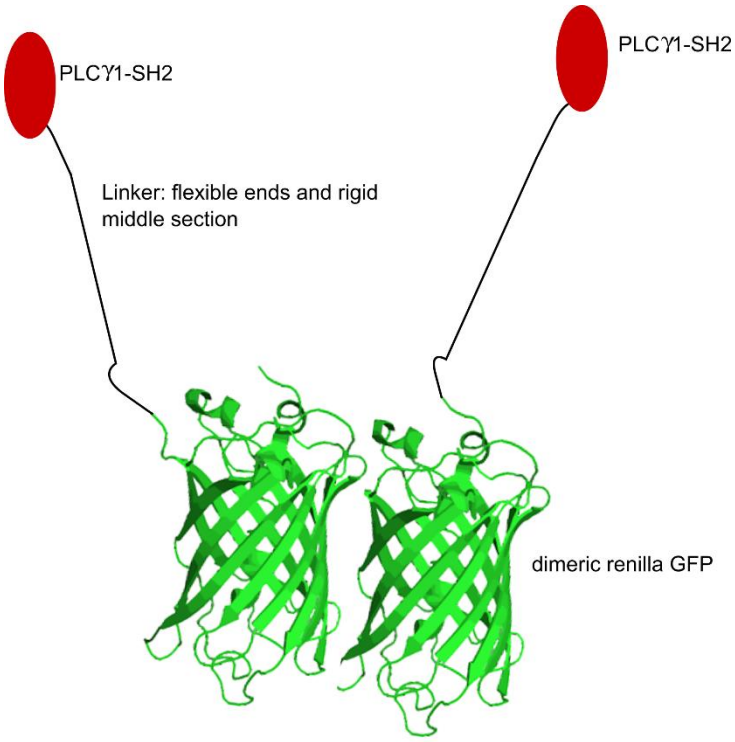


Figure 4.2: An alternative co-localization sensor using renilla GFP and PLCγ1-SH2.

PLCγ1-SH2 has low affinity for its peptide target sequence but may potentially generate the high affinity desired when present in a bivalent construct when two target sequences are engaged simultaneously. The PLCγ1-SH2 may therefore be a readily available ligand which could be used in place of the HA-tag biosensor described in Figure 4.1. In this case the biosensor is shown using Renilla GFP, which is a very strong dimerizing protein.

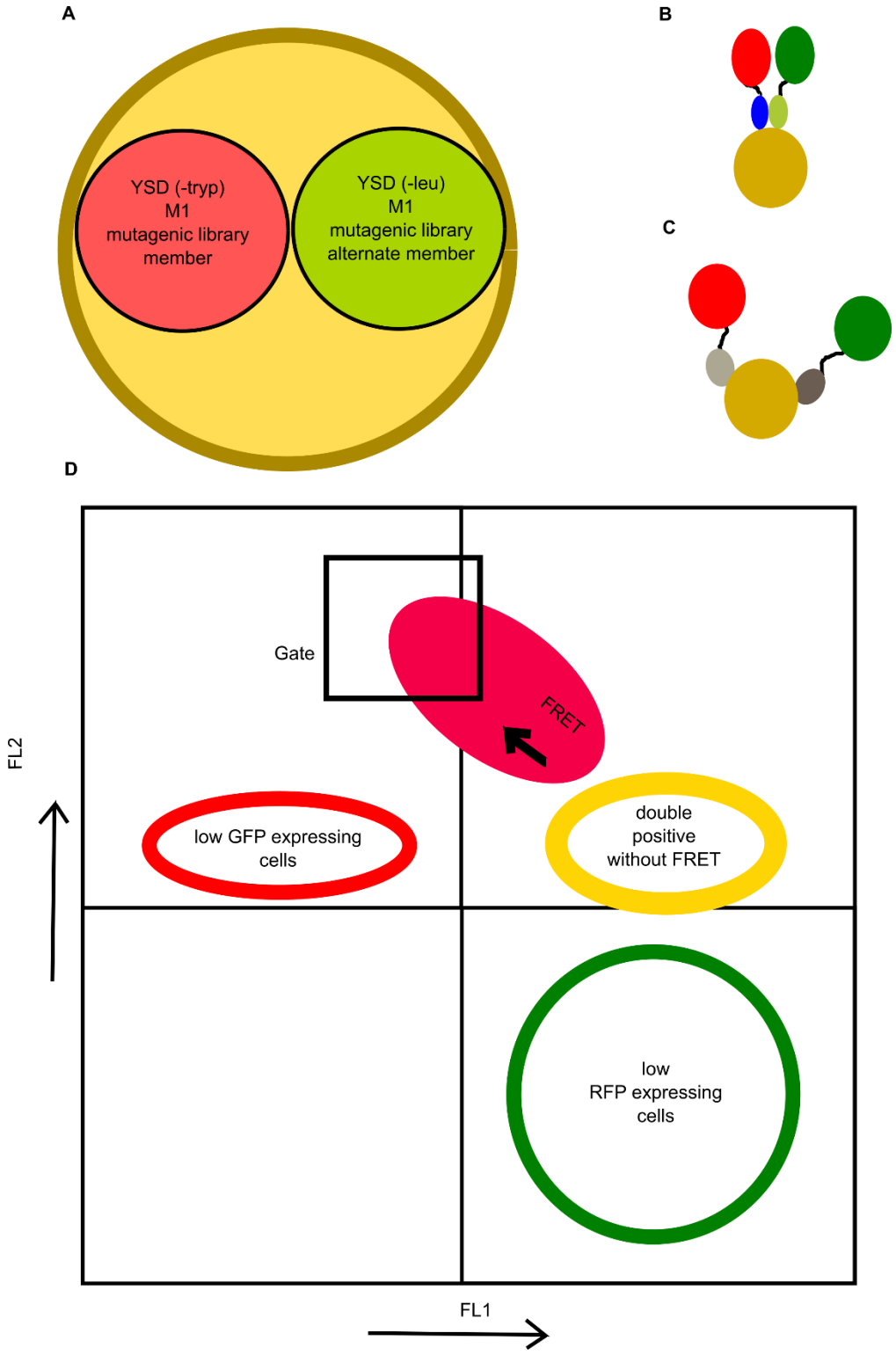

A
YSD (-tryp)
M1
mutagenic library
member
YSD (-leu)
M1
mutagenic library
alternate member
B
C
D
Gate
FRET
low GFP expressing
cells
double
positive
without FRET
low
RFP expressing
cells
FL2
FL1

Figure 4.3: A FRET based sort for heterodimeric self-associating ligands.

A library consisting of yeast cells bearing two plasmids is shown in (A) The M1 mutagenic library which was used to generate M2 is transformed along with (-leucine) and (-tryptophan) yeast surface display vectors. The (-tryp) plasmid would code for a tagRFP fusion protein which would be C-terminal to the library protein. The (-leu) vector would code for a tagGFP C-terminal fusion. The library is then grown up in double knock out media. (B) Dimeric ligands which self-associate should co-localize the tagGFP and tagRFP and should give rise to a FRET signal. (C) Ligands which do not co-localize should give both RFP and GFP signals but without FRET. (D) The expected cytometer sort plot for the library. Most cells should be double positive and be seen in the 1^{st} quadrant. Do to random and variable expression of the constructs in either plasmids some cells will have higher GFP then RFP fluorescence and vice versa, these will appear in the 2^{nd} and 4^{th} quadrants. RFP positive cells should have substantially less fluorescence relative to GFP cells as a 488 laser will result in very little RFP excitation. Some cells should give rise to FRET and these will be characterized by a decrease in GFP signal but an elevated RFP signal. The cells with the highest ratio of RFP to GFP fluorescence should have the highest ratio of heterodimeric ligands. These can be gated and sorted as shown. A third label for expression of the GFP tagged protein could also be used to ensure that the gated cell shown above to does not simply select for ligands with decreased GFP presence. The GFP protein could be expressed with a c-Myc tag such that the relative amount of GFP could be quantified independently from GFP fluorescence. A secondary antibody with a fluorescent tag in the 647 nm range could be used so as not interfere with the FRET system.

Appendices

2. Appendix A1

Novablue™ E. coli cells (EMD Biosciences, San Diego, CA) were used for all cloning steps. All primers and synthetic gene blocks were purchased from Integrated DNA Technologies (Coralville, IA), unless otherwise stated. Positive clones were confirmed by DNA sequencing (Genewiz, La Jolla, CA). A synthetic gene encoding the sequence of GFP1-9 was purchased from Genewiz. GFP1-9 sequence is[51]:

5'-ATG AGG AAA GGA GAA GAA CTT TTC ACT GGA GTC GTC CCA ATT CTT ATT GAA TTA GAT GGT GAT GTT AAT GGG CAC AAA TTT TTT GTC CGT GGA GAG GGT GAA GGT GAT GCT ACA ATC GGA AAA CTC AGC CTT AAA TTT ATT TGC ACT ACT GGA AAA CTA CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT CTG ACC TAT GGT GTT CAA TGC TTT TCC CGT TAT CCG GAT CAC ATG AAA CGG CAT GAC TTT TTC AAG AGT GCC ATG CCC GAA GGT TAT GTA CAG GAA CGC ACT ATA TAT TTC AAA GAT GAC GGG ACC TAC AAG ACG CGT GCT GAA GTC AAG TTT GAA GGT GAT ACC CTT GTT AAT CGT ATC GAG TTA AAG GGT ATT GAT TTT AAA GAA GAT GGA AAC ATT CTT GGA CAC AAA CTC GAG TAC AAC TTT AAC TCA CAC AAA GTA TAC ATC ACG GCA GAC AAA CAA AAT AAT GGA ATC AAA GCT AAC TTC ACA ATT CGC CAC AAC GTT GAA GAT GGT TCC GTT CAA CTA GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT GTC CTT TTA CCA -3'

The GFP1-9 gene was amplified by PCR using oligonucleotide primers PG9-F and PG9-R, and cloned into vector pET-22b(+) between NdeI and BamHI restriction sites. Primer sequences are:

PG9-F: 5'-GAG TCC *CAT ATG* ATG AGG AAA GGA GAA GAA CTT T -3'

PG9-R: 5'- GTT TCG *GGA TCC* TTA TGG TAA AAG GAC AGG GCC AT-3'

where italicized characters represent restriction sites.

NTL1-linker-GFP11 gene was amplified by PCR from a synthetic oligonucleotide using primers PN1-F and PN1-R, and cloned into pET-28b(+) between NcoI and XhoI restriction sites. The sequence of the synthetic gene block is:

5’-GCA GCA GTc acc atc acc acc atc at*G GAT CC*A TGG CGA CCG TGA AAT TTA AAT ATA AAG GCG AAG TAA AAC AGG TGG ATA TTA GCA AAA TTT TCT TTG TGA TCC GCG ACG GCA AAG ACA TTT TGT TTA AGT ATG ATC TGG GCG GCG GCA AAG CAG GCT TAG GTG AAG TGA GCG AAA AAG ATG CGC CGA AAG AAC TGC TGC AGA TGC TGG AAA AAC AGA AAA AA*G TCG AC*g gat ccg gcg ctg gcg gaa gcc ctg ggg gcg gga gcg gtg gct ctg gtt ctt ctg cta gtg gcg gct caa cat ct*g cta GC*G AAA AAC GGG ATC ATA TGG TTT TGC TGG AGT ACG TGA CGG CGG CTG GCA TTA CGG ATG CTT CG-3’

where the linker between NTL-1 and GFP11 is given in italicized lower-case letters, restriction sites are italicized, and a histidine tag is in lower-case letters. Primer sequences are:

PN1-F: 5’- GAG TCC *CCA TGG* GCA GCA GTC ACC ATC ACC A-3’

PN1-R: 5’-GAG TCC *CTC GAG* TTA CGA AGC ATC CGT AAT GCC-3’

where restriction sites are given in italics.

To create the GFP10-linker-CTL1, a synthetic oligonucleotide containing the sequence of GFP10-linker-K1 was amplified by PCR using primers PK1-F and PK1-R, and cloned into pET-28b(+) to generate pET-GFP10, where K1 is a coiled peptide[51]. The sequence of the oligonucleotide is:

5’-ATG GAC CTT CCA GAT GAT CAC TAT CTG AGT ACG CAG ACC ATT CTG AGC AAA GAC TTG AAC *GTC GAC gtt ggt ggt ggc gga tca gaa gga ggc ggt agc ggg ggc cct ggt tcg gga ggg gaa ggt tct gct ggg gga ggg agc gct ggc ggg ggg tct CAT ATG* AAG GTT AGT GCC CTG AAA GAA AAT GTC AGC GCA CTT AAA GAG AAA GTT AGT GCC CTT

ACA GAG AAA GTA TCT GCA TTA AAA GAA AAA GTA AGT GCC CTG AAA GAG-3’

Where the linker between GFP10 and K1 is given in italicized lower-case letters, and restriction sites are italicized. Primer sequences are:

PK1-F: 5’-GAG TCC *CCA TGG* ACC TTC CAG ATG ATC ACT ATC

PK1-R: 5’-GAC TCT *CTC GAG* CTC TTT CAG GGC ACT TAC TT

CTL-1 sequence was amplified by PCR using primers PC1-F and PC1-R, and cloned into pET-GFP10 between NdeI and XhoI restriction sites, which allow the removal of the K1 sequence. Primer sequences are:

PC1-F: 5’-GAG TCA *CAT ATG* GCG ACC GTG AAA TTA AAA TAT-3’

PC1-R: 5’-GAG TCA *CTC GAG* TTT TTT CTG TTT TTC CAG CAT CTG-3’

where restriction sites are italicized.

3. Appendix A2

All Blitz™ data was fit to a one to one binding model according to Forte Bio's Application Note 14. Association data is fit according to the following equation.

$$Binding\ signal\ (B) = B_{max}\left(1 - e^{-k_{obs}t}\right) \quad (1)$$

Where B_{max} is the maximum signal observed for each run. The dissociation rate follows first order decay kinetics and was fit as follows

$$Binding\ signal = B_D e^{-k_d t} \quad (2)$$

Where k_d is the dissociation constant. For data analysis both equations were linearized using the natural log function and calculated values and standard errors (SE) was carried out using statistical methods for linear regression. The local calculation of K_D is based on this equation, where k_a is calculated from k_{obs} and k_d

$$k_{obs} = [Ligand]k_a + k_d \quad (3)$$

And K_D is calculated as

$$K_D = \frac{k_d}{k_a} \quad (4)$$

Equation 4 can also be used to calculated k_a and k_d based on a linear regression for a series of k_{obs} calculations based on Blitz™ data collected for a range of ligand concentrations.

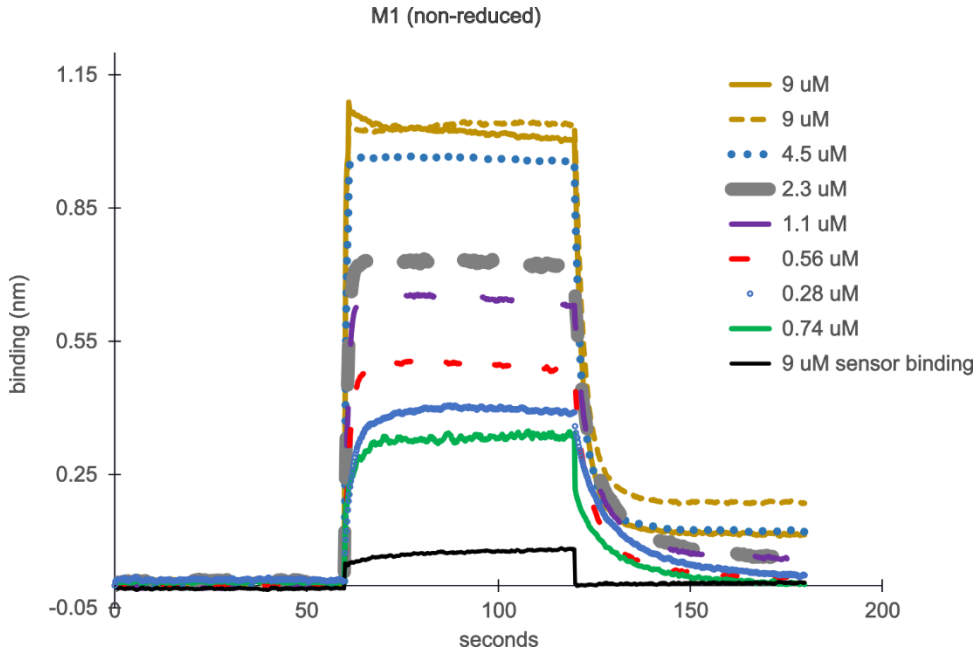
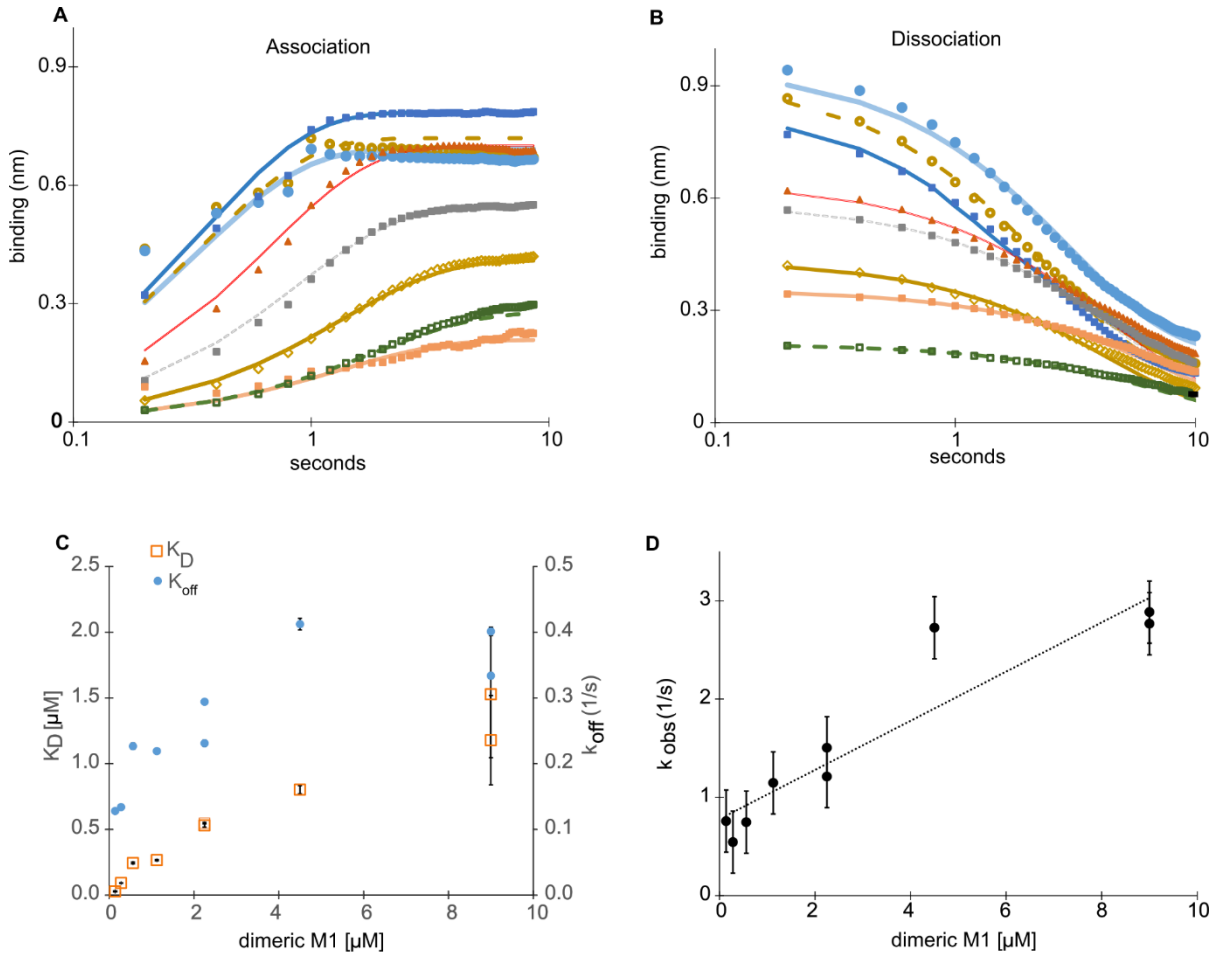


Figure A2.1 Biolayer interferometry data recorded for non-reduced M1.

β-catenin was immobilized in a previous step at 1 μM for 120 seconds prior to the addition of the ligand. The data represents a 60 second equilibration with buffer, the ligand was loaded for an additional 60 seconds and dissociation was recorded for a final 60 seconds. The sensor binding condition represents the control for which β-catenin was not immobilized on the sensor.



E

uM	kobs 1/(s)	k_{obs} SE 1/ s	koff (1/s)	koff error (1/s)	ka 1/ [uM] s	ka SE 1/ [uM] s	KD [uM]	KD SE [uM]
9.0	2.9	0.73	0.33	4.5E-03	0.28	8.2E-02	1.2	0.34
9.0	2.8	0.75	0.40	6.2E-03	0.26	8.3E-02	1.5	0.48
4.5	2.7	7.4E-02	0.41	8.8E-03	0.51	1.7E-02	0.80	3.1E-02
2.3	1.5	0.0E+00	0.29	3.3E-03	0.54	1.5E-03	0.55	6.2E-03
1.1	1.1	1.7E-02	0.22	2.1E-03	0.82	1.5E-02	0.27	5.5E-03
0.56	0.75	7.0E-03	0.23	4.6E-03	0.93	1.5E-02	0.24	6.4E-03
2.3	1.2	1.6E-02	0.23	4.7E-03	0.44	7.4E-03	0.53	1.4E-02
0.28	0.54	3.4E-03	0.13	2.0E-03	1.5	1.4E-02	0.092	1.6E-03
0.14	0.76	4.5E-02	0.13	3.8E-03	4.5	3.2E-01	0.029	2.2E-03
*LR			0.78	1.7E-01	0.25	0.04	3.1	0.8

Figure A2.2 Analysis of biolayer interferometry data recorded for non-reduced M1

(A) Association (B) dissociation data fits for non-reduced M1 at each of the concentration run from Figure A2.1. Each of the curves was fit using linear regression analysis and a one to one binding model after transformation with a natural log. (C) The equilibrium constant K_D (left y-axis) and dissociation rate k_{off} (right y-axis) as a function of the concentration of M1 loaded. (D) The observed association rate as a function of concentration. (E) At each concentration run local determination of K_D was carried out by fitting k_{obs} and k_{off}, k_a and K_D were calculated based on equations (3) and (4). observed data fit parameters for each run were used to calculate the association rate (k_a) and K_D. The row labelled "*LR" shows the result of k_a, k_d and K_D determination by linear regression carried out on the data in shown in figure E. "SE" stands for standard error.

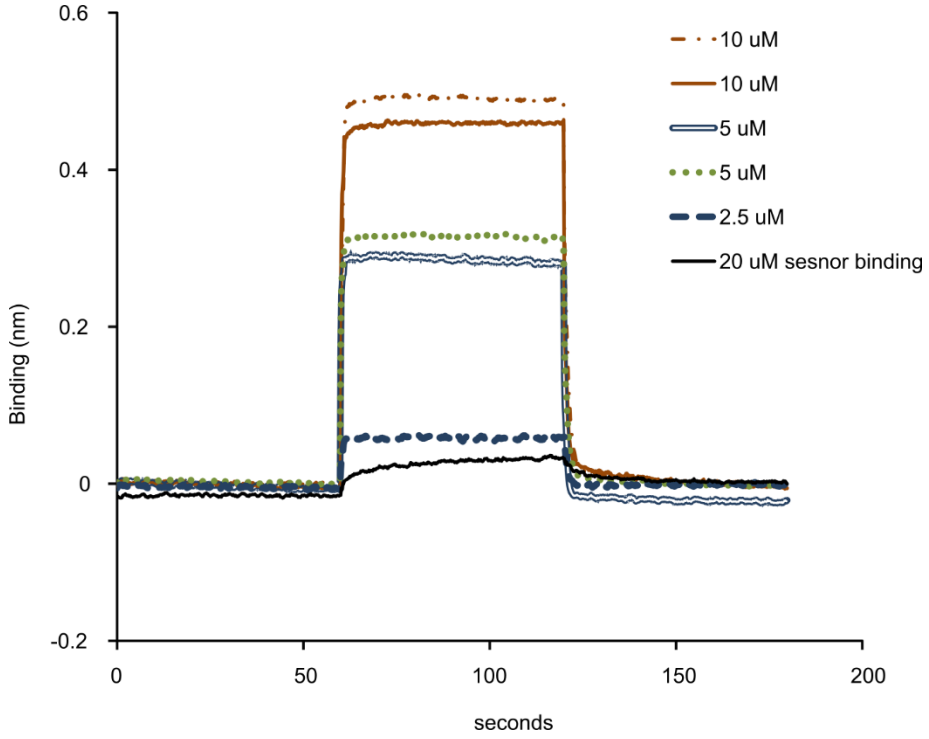
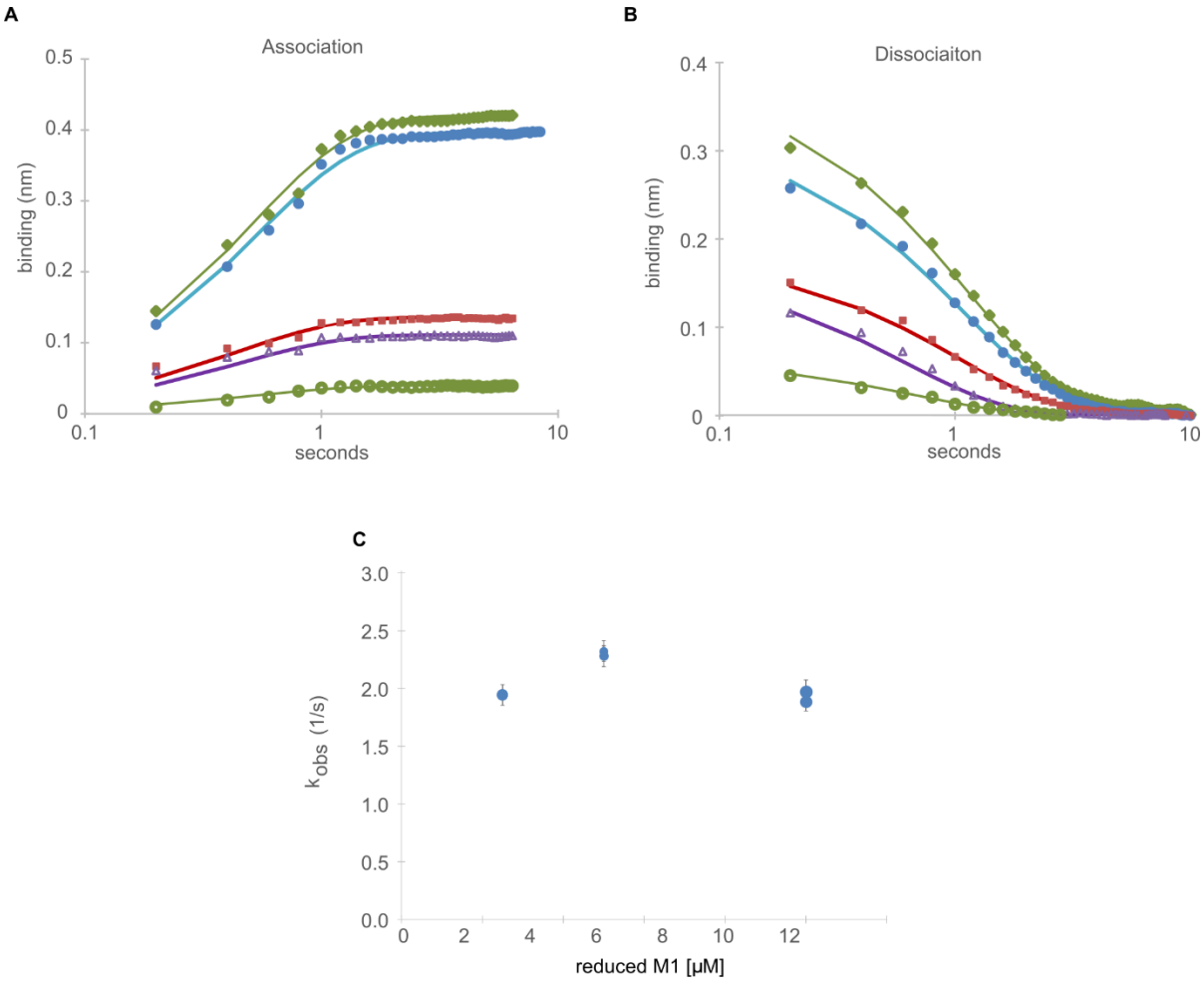


Figure A2.3 Biolayer interferometry data for reduced M1

1 mM TCEP was used as a reducing agent for each curve. β-catenin was immobilized in a previous step at 1 µM for 120 seconds prior to the addition of the ligand. The data represents a 60 second

equilibration with buffer, the ligand was loaded for an additional 60 seconds and dissociation was recorded for a final 60 seconds. The sensor binding condition represents the control for which β-catenin was not immobilized on the sensor.



D

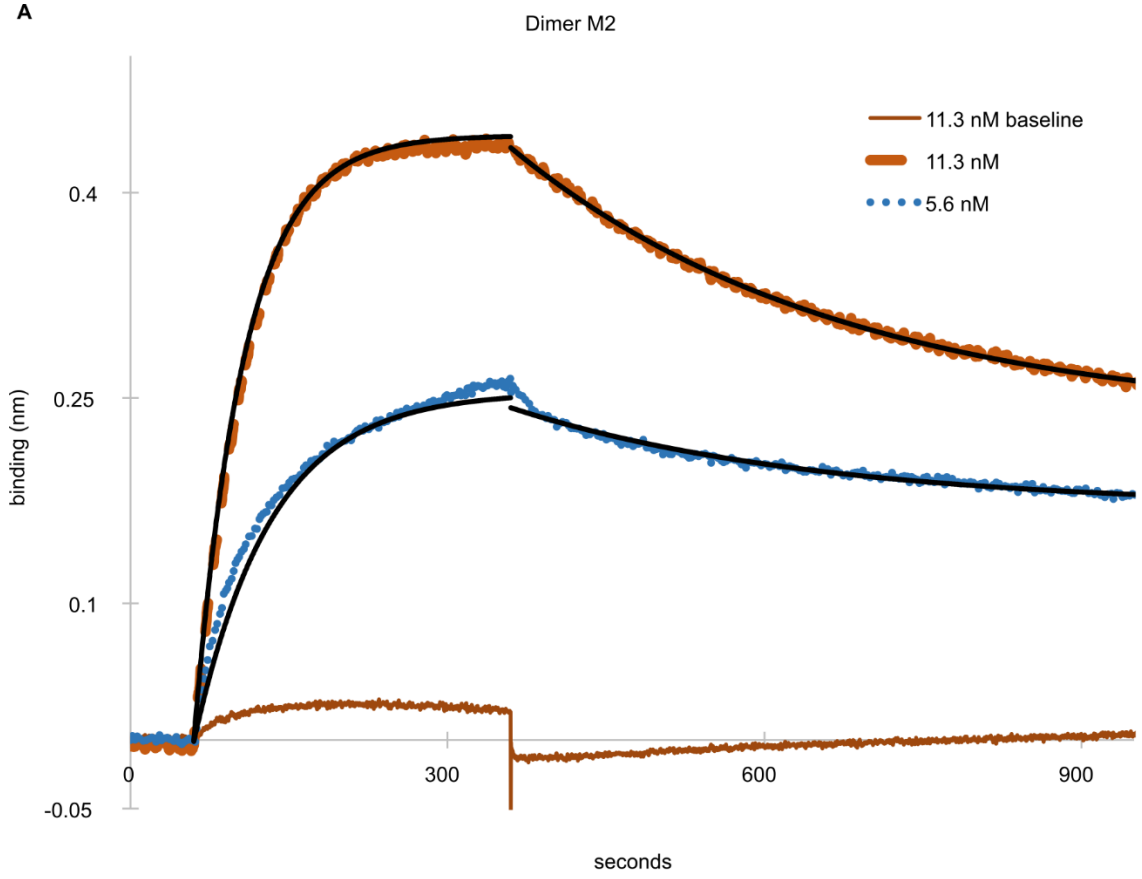
uM	k_{obs} (1/s)	k_{obs}SE (1/s)	k_{off} (1/s)	k_{off}error (1/s)	ka (1/[M] s)	ka SE (1/[M] s)	KD [uM]	KD SE [uM]
10	1.98	0.05	0.87	7.2E-03	1.1E+05	5.7E+03	7.8	0.40
10	1.90	0.13	0.9	6.6E-03	9.8E+04	1.2E+04	9.4	1.20
5.0	2.33	0.12	1.0	1.5E-02	2.7E+05	3.2E+04	3.6	0.43
5.0	2.28	0.14	1.63	7.2E-03	1.3E+05	1.7E+04	12.7	1.71
2.5	1.94	0.16	1.50	7.0E-02	1.8E+05	3.2E+04	8.4	1.52

Figure A2.4 Analysis of biolayer interferometry data for reduced M1.

(A) The association raw data and curve fit (B) Dissociation raw data and curve fitting. (C) k_{obs} as a function of the reduced M1 concentration. There is not a linear trend in this case so determination of

K_D by linear regression here was not possible. (D) For each concentration run of reduced M1 all of the kinetic parameters are shown along with their respective standard errors.

A



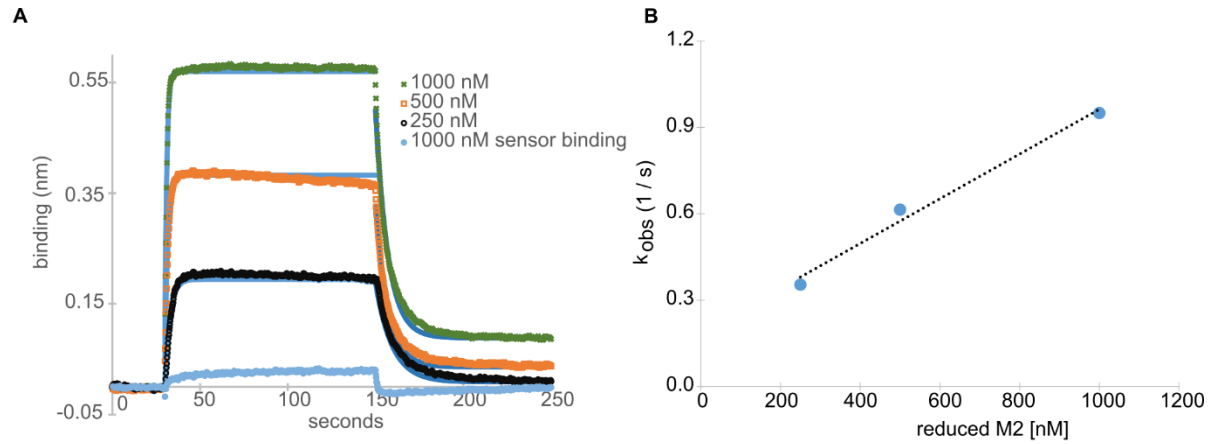
B

nM	kobs (1/s)	k_{obs} SE (1/ s)	koff (1/s)	koff error (1/s)	ka 1/ [nM] s	ka SE 1/ [nM] s	KD [nM]	KD SE [nM]
11.3	0.021	5.1E-05	3.5E-03	3.0E-06	1.6E-03	4.5E-06	2.2	6.7E-3
5.6	0.014	3.0E-05	3.2E-03	1.1E-05	1.9E-03	5.6E-06	1.7	7.8E-3

Figure A2.5 Biolayer interferometry data and analysis for non-reduced M2.

Blitz data and curve fits for non-reduced M2. Concentrations are based on the dimeric concentration of M2. β-catenin was immobilized in a previous step at 1 µM for 120 seconds prior to the addition of the ligand. The data represents a 60 second equilibration with buffer, the ligand

was loaded for an additional 60 seconds and dissociation was recorded for a final 60 seconds. The sensor binding condition represents the control for which β-catenin was not immobilized on the sensor. (B) Local determination of kinetic parameter and standard error for each measurement.



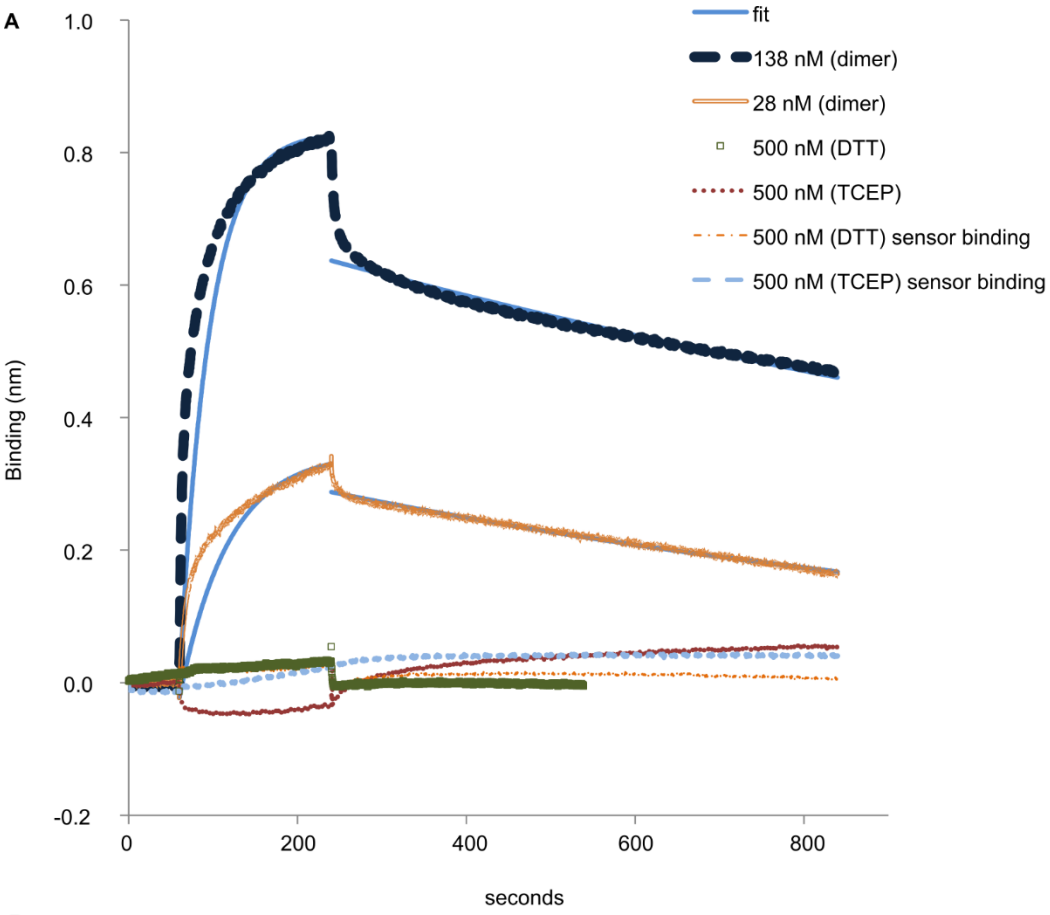
C

nM	k_{obs} (1/s)	k_{obs} SE (1/s)	koff (1/s)	k_{off} SE (1/s)	ka 1/ [nM] s	ka SE 1/ [nM] s	KD [nM]	KD SE [nM]
1000	0.95	0.021	0.14	2.0E-03	8.1E-04	2.1E-05	177	5.2
500	0.61	0.0041	0.14	1.8E-03	9.4E-04	9.1E-06	150	2.4
250	0.35	0.0076	0.11	7.8E-04	9.9E-04	3.0E-05	107	3.4
*LR			0.19	6.1E-02	7.8E-04	9.2E-05	239	83

Figure A2.6 biolayer interferometry data and analysis for reduced M2.

(A) Raw data representing a 30 second equilibration of the sensor with buffer, the association portion of the measurement was carried out for 120 seconds and dissociation was measured for 90 seconds. The sensor binding condition represents the control for which β-catenin was not immobilized on the sensor prior to the Blitz Run. The curve fit for each run is shown in the figure as well. (B) K_{obs} as a function of the concentration of M1 used during each run is shown to give a linear. (C) Local

determination of kinetic constants and standard error for each run as well as by linear regression analysis based on (B).



B

nM	k_{obs} (1/s)	k_{obs} SE (1/ s)	koff (1/s)	k_{off} SE (1/s)	ka (1/ [nM] s)	ka SE (1/ [nM] s)	KD [nM]	KD SE [nM]
138	2.8E-02	1.1E-04	5.4E-04	2.2E-06	2.0E-04	8.3E-07	2.7	1.5E-02
28	1.7E-02	8.1E-05	9.0E-04	1.4E-06	5.9E-04	2.9E-06	1.5	7.9E-03

Figure A2.7 Biolayer interferometry data and analysis for reduced and non-reduced T2.

Blitz data taken at higher concentrations to attempt to determine the KD of reduced T2 as well as to try and determine the binding mechanism of 23CTL. (A) Blitz data recorded for reduced and non-

reduced T2 (B) measurement of kinetic parameters and standard errors for the non-reduced state of the protein.

4. Appendix A3

Image J software™ was used to make a densitometry reading of each protein band in each lane of each gel. For each ligand a calculated density value was determined by using the total concentration of each protein (known value) which was added to each lane and the fraction of each protein that appears as a dimer (fit parameter, x). The Hill Equation (with n=1) was able to correlate monomer and dimer protein concentrations to the calculated density readings extremely well.

$$D_{calc\ dimer} = \frac{D_{max}^{n}}{(b + x(c_{lane})^{n})} \tag{A3.1}$$

$$D_{calc\ monomer} = \frac{D_{max}^{n}}{(b + (1 - x)(c_{lane})^{n})} \tag{A3.2}$$

D_{max} is the maximum density possible, c_{lane} is the total concentration of protein stock that was run in each lane and b is a curve fitting parameter. The parameters were determined by minimizing a least squares sum of the calculated density and the actual measured density. The data below shows the density measurements made for the M1 ligand, as well as the lane dilutions for each lane and the calculated concentration of the monomer and dimer within each lane. For M1 a D_{max} of 1.45e4 was measured and b = 42.3.

CT12.1 (250 ug/ml stock used)

	Lane	density measured	fraction dimer	Dilution	calculated protein concentration (ug/ml)	density calculated	squared difference	normalized density
Dimer	2	11630	89.4%	1.3	167.7	11565	4.12E+03	0.80
	3	9730	89.4%	2.7	83.8	9626	1.09E+04	0.67
	4	6773	89.4%	5.3	41.9	7208	1.90E+05	0.47
	5	5138	89.4%	10.7	21.0	4798	1.15E+05	0.35
monomer	2	4385	10.6%	1.3	19.8	4621	5.57E+04	0.30
	3	2883	10.6%	2.7	9.9	2749	1.79E+04	0.20
	4	1680	10.6%	5.3	5.0	1519	2.61E+04	0.12
	5	1099	10.6%	10.7	2.5	801	8.85E+04	0.08
						minimize	5.08E+05	

The 68% confidence interval was calculated as described in[22]. In this case however the fraction, x, was set to a given value above or below the fraction calculated for the least squares minimum. Dmax and b were subsequently calculated corresponding to a new least squares minimum for the given x. This was repeated until the least squares error calculated corresponded to the true least squares minimum multiplied by the F-statistic corresponding to a 68% confidence interval.

References

1. Wang, W., Chen, C., Qian, M. & Zhao, X. S. Aptamer biosensor for protein detection using gold nanoparticles. *Anal. Biochem.* **373,** 213–9 (2008).
2. Rodriguez, M. C., Kawde, A.-N. & Wang, J. Aptamer biosensor for label-free impedance spectroscopy detection of proteins based on recognition-induced switching of the surface charge. *Chem. Commun. (Camb).* 4267–9 (2005). doi:10.1039/b506571b
3. Wang, Y., Bao, L., Liu, Z. & Pang, D.-W. Aptamer biosensor based on fluorescence resonance energy transfer from upconverting phosphors to carbon nanoparticles for thrombin detection in human plasma. *Anal. Chem.* **83,** 8130–7 (2011).
4. Tinberg, C. E. *et al.* Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501,** 212–6 (2013).
5. Stites, W. E. Protein−Protein Interactions: Interface Structure, Binding Thermodynamics, and Mutational Analysis. *ACS Chemical Reviews* 1233−1250 (1997). at <http://pubs.acs.org/doi/pdf/10.1021/cr960387h>
6. Binz, H. K., Amstutz, P. & Plückthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* **23,** 1257–1268 (2005).
7. Skerra, A. Alternative non-antibody scaffolds for molecular recognition. *Curr. Opin. Biotechnol.* **18,** 295–304 (2007).
8. Vazquez-Lombardi, R. *et al.* Challenges and opportunities for non-antibody scaffold drugs. *Drug Discov. Today* **20,** 1271–1283 (2015).
9. Stern, L. A., Case, B. A. & Hackel, B. J. Alternative Non-Antibody Protein Scaffolds for Molecular Imaging of Cancer. *Curr. Opin. Chem. Eng.* **2,** 425–432 (2013).
10. Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V. & Koide, S. High-affinity single-domain binding proteins with a binary-code interface. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 6632–7 (2007).
11. Banta, S., Dooley, K. & Shur, O. Replacing antibodies: engineering new binding proteins. *Annu. Rev. Biomed. Eng.* **15,** 93–113 (2013).
12. Hosse, R. J., Rothe, A. & Power, B. E. A new generation of protein display scaffolds for molecular recognition. 14–27 (2006). doi:10.1110/ps.051817606.library
13. Gera, N., Hussain, M., Wright, R. C. & Rao, B. M. Highly stable binding proteins derived from the hyperthermophilic Sso7d scaffold. *J. Mol. Biol.* **409,** 601–16 (2011).
14. Gera, N., Hill, A. B., White, D. P., Carbonell, R. G. & Rao, B. M. Design of pH sensitive binding proteins from the hyperthermophilic Sso7d scaffold. *PLoS One* **7,** e48928 (2012).
15. Mouratou, B. *et al.* Remodeling a DNA-binding protein as a specific in vivo inhibitor of bacterial secretin PulD. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 17983–8 (2007).

16. Baumann, H., Knapp, S., Lundbäck, T., Ladenstein, R. & Härd, T. Solution structure and DNA-binding properties of a thermostable protein from the archaeon Sulfolobus solfataricus. *Nat. Struct. Biol.* **1,** 808–819 (1994).

17. Feng, S., Chen, J., Yu, H., Simon, J. & Schreiber, S. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science (80-.).* **266,** 1241–1247 (1994).

18. Béhar, G. *et al.* Tolerance of the archaeal Sac7d scaffold protein to alternative library designs: characterization of anti-immunoglobulin G Affitins. *Protein Eng. Des. Sel.* **26,** 267–75 (2013).

19. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16,** 379–394 (2015).

20. Dreier, B. & Plückthun, A. Ribosome Display and Related Technologies. **805,** 261–286 (2012).

21. Hoogenboom, H. R. Overview of antibody phage-display technology and its applications. *Methods Mol. Biol.* **178,** 1–37 (2002).

22. Gera, N., Hussain, M. & Rao, B. M. Protein selection using yeast surface display. *Methods* **60,** 15–26 (2013).

23. Matochko, W. L., Cory Li, S., Tang, S. K. Y. & Derda, R. Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Res.* **42,** 1784–98 (2014).

24. Desmet, J., Vanhoorelbeke, K. & Deckmyn, H. Antibody Engineering. **1,** 341–353 (2010).

25. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15,** 553–7 (1997).

26. Pepper, L. R., Cho, Y. K., Boder, E. T. & Shusta, E. V. A decade of yeast surface display technology: where are we now? *Comb. Chem. High Throughput Screen.* **11,** 127–34 (2008).

27. Shusta, E. V, Kieke, M. C., Parke, E., Kranz, D. M. & Wittrup, K. D. Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *J. Mol. Biol.* **292,** 949–56 (1999).

28. Ackerman, M. *et al.* Highly Avid Magnetic Bead Capture : An Efficient Selection Method for de novo Protein Engineering Utilizing Yeast Surface Display. (2009). doi:10.1021/bp.174

29. Procko, E. *et al.* Computational design of a protein-based enzyme inhibitor. *J. Mol. Biol.* **425,** 3563–75 (2013).

30. Guntas, G., Purbeck, C. & Kuhlman, B. Engineering a protein-protein interface using a computationally designed library. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 19296–301 (2010).

31. Schilling, J., Schöppe, J. & Plückthun, A. From DARPins to LoopDARPins: novel LoopDARPin design allows the selection of low picomolar binders in a single round of ribosome display. *J. Mol. Biol.* **426,** 691–721 (2014).

32. Boersma, Y. L., Chao, G., Steiner, D., Wittrup, K. D. & Plückthun, A. Bispecific designed ankyrin repeat proteins (DARPins) targeting epidermal growth factor receptor inhibit A431 cell proliferation and receptor recycling. *J. Biol. Chem.* **286,** 41273–85 (2011).

33. Stahl, A. *et al.* Highly potent VEGF-A-antagonistic DARPins as anti-angiogenic agents for topical and intravitreal applications. *Angiogenesis* (2012). doi:10.1007/s10456-012-9302-0

34. Stefan, N. *et al.* DARPins Recognizing the Tumor-Associated Antigen EpCAM Selected by Phage and Ribosome Display and Engineered for Multivalency. *J. Mol. Biol.* **413,** 826–843 (2011).

35. Dreier, B. *et al.* Development of a generic adenovirus delivery system based on structure-guided design of bispecific trimeric DARPin adapters. *Proc. Natl. Acad. Sci. U. S. A.* (2013). doi:10.1073/pnas.1213653110

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

37. Mammen, M., Choi, S.-K. & Whitesides, G. M. Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew. Chemie Int. Ed.* **37,** 2754–2794 (1998).

38. Kaufman, E. N. & Jain, R. K. Effect of Bivalent Interaction upon Apparent Antibody Affinity: Experimental Confirmation of Theory Using Fluorescence Photobleaching and Implications for Antibody Binding Assays. *Cancer Res.* **52,** 4157–4167 (1992).

39. Neri, D., Momo, M., Prospero, T. & Winter, G. High-affinity antigen binding by chelating recombinant antibodies (CRAbs). *J. Mol. Biol.* **246,** 367–73 (1995).

40. Zhou, H.-X. Quantitative Account of the Enhanced Affinity of Two Linked scFvs Specific for Different Epitopes on the Same Antigen. *J. Mol. Biol.* **329,** 1–8 (2003).

41. Zhou, H.-X. Quantitative Account of the Enhanced Affinity of Two Linked scFvs Specific for Different Epitopes on the Same Antigen. *J. Mol. Biol.* **329,** 1–8 (2003).

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

43. Martineau, P. Antibody Engineering. **1,** 657–665 (2010).

44. Wright, M. J. & Deonarain, M. P. Phage display of chelating recombinant antibody libraries. *Mol. Immunol.* **44,** 2860–9 (2007).

45. Huang, J., Koide, A., Makabe, K. & Koide, S. Design of protein function leaps by

directed domain interface evolution. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 6578–83 (2008).

46. Yasui, N. *et al.* Directed network wiring identifies a key protein interaction in embryonic stem cell differentiation. *Mol. Cell* **54,** 1034–41 (2014).

47. Grebien, F. *et al.* Targeting the SH2-kinase interface in Bcr-Abl inhibits leukemogenesis. *Cell* **147,** 306–19 (2011).

48. Hackel, B. J., Neil, J. R., White, F. M. & Wittrup, K. D. Epidermal growth factor receptor downregulation by small heterodimeric binding proteins. *Protein Eng. Des. Sel.* **25,** 47–57 (2012).

49. Klemm, J. D., Rould, M. a, Aurora, R., Herr, W. & Pabo, C. O. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* **77,** 21–32 (1994).

50. Järviluoma, A. *et al.* High-affinity target binding engineered via fusion of a single-domain antibody fragment with a ligand-tailored SH3 domain. *PLoS One* **7,** e40331 (2012).

51. Cabantous, S. *et al.* A new protein-protein interaction sensor based on tripartite split-GFP association. *Sci. Rep.* **3,** 2854 (2013).

52. Poon, D. K. Y., Withers, S. G. & McIntosh, L. P. Direct demonstration of the flexibility of the glycosylated proline-threonine linker in the Cellulomonas fimi Xylanase Cex through NMR spectroscopic analysis. *J. Biol. Chem.* **282,** 2091–100 (2007).

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

54. Klose, R. J. & Bird, A. P. MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. *J. Biol. Chem.* **279,** 46490–6 (2004).

55. Wilson, D. S. & Keefe, a D. Random mutagenesis by PCR. *Curr. Protoc. Mol. Biol.* **Chapter 8,** Unit8.3 (2001).

56. Orr, B. A., Carr, L. M., Wittrup, K. D., Roy, E. J. & Kranz, D. M. Rapid method for measuring ScFv thermal stability by yeast surface display. *Biotechnol. Prog.* **19,** 631–8

57. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2,** 2212–21 (2007).

58. Jensen, M. R. *et al.* Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings.

Structure **17,** 1169–1185 (2009).

59. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337,** 635–45 (2004).

60. Tsutakawa, S. E., Hura, G. L., Frankel, K. A., Cooper, P. K. & Tainer, J. A. Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. *J. Struct. Biol.* **158,** 214–23 (2007).

61. Dunker, A. K. & Uversky, V. N. Drugs for 'protein clouds': targeting intrinsically disordered transcription factors. *Curr. Opin. Pharmacol.* **10,** 782–788 (2010).

62. Wojcik, J. *et al.* A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nat. Struct. Mol. Biol.* **17,** 519–27 (2010).

63. Yeh, J. T.-H., Binari, R., Gocha, T., Dasgupta, R. & Perrimon, N. PAPTi: a peptide aptamer interference toolkit for perturbation of protein-protein interaction networks. *Sci. Rep.* **3,** 1156 (2013).

64. Mosimann, C., Hausmann, G. & Basler, K. Beta-catenin hits chromatin: regulation of Wnt target gene activation. *Nat. Rev. Mol. Cell Biol.* **10,** 276–86 (2009).

65. Castellone, M. D. *et al.* The beta-catenin axis integrates multiple signals downstream from RET/papillary thyroid carcinoma leading to cell proliferation. *Cancer Res.* **69,** 1867–76 (2009).

66. Davidson, K. C. *et al.* Wnt/β-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 4485–90 (2012).

67. Valenta, T., Hausmann, G. & Basler, K. The many faces and functions of β-catenin. *EMBO J.* **31,** 2714–36 (2012).

68. MacDonald, B. T., Tamai, K. & He, X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev. Cell* **17,** 9–26 (2009).

69. Clevers, H. & Nusse, R. Wnt/β-Catenin Signaling and Disease. *Cell* **149,** 1192–205 (2012).

70. Xing, Y. *et al.* Crystal structure of a full-length beta-catenin. *Structure* **16,** 478–87 (2008).

71. Levin, A. M. & Weiss, G. A. Optimizing the affinity and specificity of proteins with molecular display. *Mol. Biosyst.* **2,** 49–57 (2006).

72. Sustmann, C., Flach, H., Ebert, H., Eastman, Q. & Grosschedl, R. Cell-type-specific function of BCL9 involves a transcriptional activation domain that synergizes with beta-catenin. *Mol. Cell. Biol.* **28,** 3526–37 (2008).

73. Trier, N. H., Hansen, P. R. & Houen, G. Production and characterization of peptide antibodies. *Methods* **56,** 136–44 (2012).

74. Mo, R. *et al.* The Terminal Region of ␣ -Catenin Promotes Stability by Shielding the Armadillo Repeats from the Axin-scaffold Destruction Complex * □. (2009). doi:10.1074/jbc.M109.045039

75. Zaccolo, M., Williams, D. M., Brown, D. M. & Gherardi, E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* **255,** 589–603 (1996).

76. Yeung, Y. a & Wittrup, K. D. Quantitative screening of yeast surface-displayed polypeptide libraries by magnetic bead capture. *Biotechnol. Prog.* **18,** 212–20 (2002).

77. Lim, K. H., Hwang, I. & Park, S. Biotin-assisted folding of streptavidin on the yeast surface. *Biotechnol. Prog.* **28,** 276–83

78. Gietz, R. D. & Schiestl, R. H. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2,** 38–41 (2007).

79. Grabulovski, D., Kaspar, M. & Neri, D. A novel, non-immunogenic Fyn SH3-derived binding protein with tumor vascular targeting properties. *J. Biol. Chem.* **282,** 3196–204 (2007).

80. Jamieson, C., Sharma, M. & Henderson, B. R. Targeting the β-catenin nuclear transport pathway in cancer. *Semin. Cancer Biol.* **27,** 20–9 (2014).

81. Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67,** 509–44 (1998).

82. Koller, D. *et al.* A high-throughput alphavirus-based expression cloning system for mammalian cells. *Nat. Biotechnol.* **19,** 851–5 (2001).

83. Klages, J. *et al.* The solution structure of BMPR-IA reveals a local disorder-to-order transition upon BMP-2 binding. *Biochemistry* **47,** 11930–9 (2008).

84. Yoon, H. *et al.* An efficient strategy for cell-based antibody library selection using an integrated vector system. *BMC Biotechnol.* **12,** 62 (2012).