Ubiquitin Engineering: Understanding Ubiquitin Recognition and Generating Affinity Reagents

by

Isabel Leung

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Molecular Genetics
University of Toronto

© Copyright by Isabel Leung 2017
Abstract

Ubiquitin Engineering: Understanding Ubiquitin Recognition and Generating Affinity Reagents

Isabel Leung

Doctor of Philosophy

Department of Molecular Genetics
University of Toronto

2016

Protein-protein interactions are necessary for virtually all biological processes. There have been tremendous efforts to document the diversity of molecular recognition, and to understand how molecular recognition occurs. The understanding of molecular interaction has also served as the foundation for designing novel protein interactions for use in therapeutics, diagnostics and basic sciences. An attractive system for studying protein-protein interactions is the ubiquitin (Ub) system. Ub is a protein modifier that is combinatorially ligated onto substrate proteins to influence substrate turnover and function. Ub uses a common surface to interact with more than 1000 proteins and plays pivotal roles in cell physiology. Despite the substantial structural information on Ub mediated interactions, there is no clear understanding of how individual Ub residues contribute to Ub’s broad scope of interactions. To address this question, I used affinity enhanced Ub variants (Ubvs) as proxies of native Ub in saturation scanning. Using saturation scanning, I studied the interactions between Ubvs and two Ub specific proteases (USP), USP2 and USP21, and elucidated a common functional epitope that is critical for USP recognition. The functional epitope recognizes USP residues that are conserved among the human USP family, suggesting it may make functional contributions in many other USP interactions. While
the functional epitopes for Ub binding to USP2 and USP21 share many identical residues, there exist distinct clusters of residues that are responsible for gaining high affinity and specificity to either USP2 or USP21. In this thesis, I also describe the usage of Ub as an alternative scaffold for generating affinity reagents. I developed a novel Ub phage library for the generation of Ubvs to targets of interest. I present validation data on Ubvs selected to bind mammalian cell surface receptor, Her3, and an intracellular protein, Grb2, showing that Ubvs can be used like conventional antibodies in cell biological experiments and can be utilized to antagonize cell signaling within a cell. My work demonstrates that protein-engineering approaches can reveal novel aspects of Ub biology and generate novel affinity reagents to probe intracellular signaling pathways.
Acknowledgments

First and foremost, I thank my parents for leaving behind their established lives in Hong Kong and immigrating to Canada so that my brother and I can have a better future. Without their sacrifices, this PhD degree would likely have been impossible.

I sincerely thank my supervisor, Sachdev Sidhu, for taking a chance on me. Not only have I learnt the art of doing science from him, but I also learnt from him the meaning of hard work, determination and most of all what it means to have a vision. I thank my committee members Jason Moffat and James Ellis for their scientific insights and support as well.

This degree would have been far less enjoyable without the help from the fun loving, past and present colleagues of the Sidhu lab: Andreas Ernst, Rachel Hanna, Amandeep Gakhal, Nish Patel, Linda Beatty, Esther Lau, Rashida Williams, Joan Teyra, Maryna Gorelik, Lia Cardarelli, Jaspal Singh, Max London, Kristin Kantautus, Natasha Pascoe, Valencio Salema, Moshe Ben-David, Johan Nilvebrant, Jan Tykvart, Gianluca Veggiani, Wei Ye, Alia Pavlenco, Lori Moffat and Danielle Carranza.

Finally, I owe a big part of this degree to my fiancé and soon to be husband, Mike Eastwood, for pushing me to the finish line. I will forever be grateful for his, sometimes feeble, attempts to put a smile on my face when times were tough. Thanks for putting up with me. Finally, special shout outs to my friends - Leena Baker, Praan Misir, Tina Shum, and Paul Lee.

Thank you.
# Table of Contents

Abstract ............................................................................................................................................. ii  
Acknowledgments .......................................................................................................................... iv  
Table of Contents ........................................................................................................................... v  
List of Tables ....................................................................................................................................... viii  
List of Figures ...................................................................................................................................... ix  
List of Abbreviations ......................................................................................................................... xi  

1 Introduction ....................................................................................................................................... 1  
1.1 Protein-Protein Interactions ........................................................................................................... 1  
   1.1.1 Affinity and Specificity ............................................................................................................. 1  
   1.1.2 Binding mechanisms and their effects on affinity and specificity ........................................... 3  
1.2 Characteristics of protein binding surfaces .................................................................................. 6  
   1.2.1 Amino Acid Composition ........................................................................................................ 6  
   1.2.2 Hot spots ................................................................................................................................ 7  
   1.2.2 Studying PPIs .......................................................................................................................... 9  
   1.2.2.1 Phage Display ...................................................................................................................... 9  
   1.2.2.2 Applications of phage display in PPI studies ................................................................... 11  
1.3 Protein Engineering of Affinity Reagents ..................................................................................... 13  
   1.3.1 Antibodies ............................................................................................................................. 14  
      1.3.1.1 Antibody Limitations ....................................................................................................... 15  
   1.3.2 Alternative Scaffolds .............................................................................................................. 16  
   1.3.3 In vitro display systems for affinity reagent engineering ....................................................... 17  
   1.3.4 Examples of Successful Alternative Scaffolds ..................................................................... 19  
      1.3.4.1 Affibodies ....................................................................................................................... 19  
      1.3.4.2 DARPins ......................................................................................................................... 20  
      1.3.4.3 Monobodies .................................................................................................................... 20  
      1.3.4.4 Anticalins ....................................................................................................................... 21  
   1.3.5 Practical considerations for scaffold engineering ................................................................... 22  
1.4 Ubiquitin ......................................................................................................................................... 23
2 Saturation Scanning of Ubiquitin Variants reveals a common hot spot for binding
to USP2 and USP21 ................................................................. 31
  2.1 Statement of Contribution .................................................. 31
  2.2 Introduction ........................................................................ 31
  2.3 Methods and Materials .......................................................... 33
    2.3.1 Library construction and analysis ...................................... 33
    2.3.2 Statistical Analysis .......................................................... 35
    2.3.3 Enzyme activity assays ..................................................... 35
    2.3.4 Computational saturation mutagenesis ............................. 36
  2.4 Results ................................................................................. 36
    2.4.1 Saturation Scanning of Ubv.2.1 and Ubv.21.4 ................... 36
    2.4.2 The Ub-binding site of USPs ............................................ 43
    2.4.3 Interactions at the USP:Ubv interface .............................. 48
  2.5 Discussion ............................................................................ 52

3 Ubiquitin as an alternative scaffold for the development of affinity reagents .......... 55
  3.1 Statement of contribution ...................................................... 55
  3.2 Introduction ........................................................................ 55
  3.3 Methods and Materials .......................................................... 58
    3.3.1 Library construction .......................................................... 58
    3.3.2 Mutation tolerance scan .................................................... 60
    3.3.3 Ag binding selections ......................................................... 60
    3.3.4 Protein expression and purification .................................... 61
    3.3.5 Phage and protein ELISAs ............................................... 61
    3.3.6 Immunoprecipitations ....................................................... 62
    3.3.7 Cell fractionation .............................................................. 63
    3.3.8 Western blotting ............................................................... 63
    3.3.9 EGFR signaling assays ..................................................... 64
    3.3.10 SPR kinetic binding analysis ......................................... 64
  3.4 Results ................................................................................ 64
List of Tables

Table 2.1 Sequences of mutagenic oligonucleotides used to construct saturation scanning libraries .................................................. 34

Table 3.1 Degenerate codons used in the mutation tolerance scan............................................. 59

Table 3.2 Oligonucleotides used to construct naïve Ub libraries. ............................................ 60

Table 3.3 Kinetic analysis of Ubv binding to corresponding targets........................................... 76
List of Figures

Figure 1.1 Models for protein-binding mechanisms .......................................................... 6

Figure 1.2 Phage Display workflow .................................................................................. 11

Figure 1.3 Antibody structure .......................................................................................... 15

Figure 1.4 Structures of binding domains ........................................................................ 21

Figure 1.5 The β-grasp fold .............................................................................................. 26

Figure 1.6 The structure of human Ub .............................................................................. 28

Figure 1.7 Interactions of Ub with partner proteins ......................................................... 28

Figure 2.1 Ubv saturation scanning library designs ......................................................... 37

Figure 2.2 The saturation scan database ......................................................................... 38

Figure 2.3 Saturation scans of Ubvs ................................................................................ 39

Figure 2.4 Dose response curves for inhibition of USP2 by Ubvs ..................................... 40

Figure 2.5 Computational saturation scanning of Ubvs ................................................... 42

Figure 2.6 Sequence alignment of the Ub-binding sites of the human USP family .......... 44

Figure 2.7 Sequence conservation across the Ub-binding sites of human USPs ............. 45

Figure 2.8 Alanine-scanning of USP2 site-2 ..................................................................... 47

Figure 2.9 Superposition of USPs in complex with Ub.wt or Ubvs .................................... 49

Figure 2.10 Interactions between the Ubv core functional epitopes and USPs ................. 50

Figure 3.1 Phage ELISA of Ub library ............................................................................. 65
Figure 3.2 Library designs for mutational tolerance scan. ........................................... 66

Figure 3.3 Ub positions that exhibit low tolerance to mutations. ................................ 68

Figure 3.4 Amino acid frequencies in libraries KHT and NNK. ...................................... 69

Figure 3.5 Binding selection outcomes for a panel of protein antigens. ......................... 71

Figure 3.6 Sequences of binding Ubvs ........................................................................ 73

Figure 3.7 Competitive phage ELISAs for Her3 binding Ubvs. ....................................... 74

Figure 3.8 Analysis of purified Her-binding Ubvs. ......................................................... 75

Figure 3.9 Kinetic analysis of Ubv-Her3.2 binding to Her3 .............................................. 75

Figure 3.10 Ubv-Her3.2 as a tool to immunoprecipitate Her3 and co-immunoprecipitate Her3 interacting Her members from cell lysates ................................................................. 77

Figure 3.11 Competition phage ELISA for Grb2 selected Ubvs. ...................................... 78

Figure 3.12 Analysis of purified Grb2-binding Ubvs..................................................... 79

Figure 3.13 Grb2 immunoprecipitation by Ubvs............................................................ 80

Figure 3.14 Kinetic analysis of Ubvs binding to Grb2-SH2 .............................................. 81

Figure 3.15 Inhibition of the interaction of Grb2-SH2 and p-Tyr peptide ligand by Ubvs ...... 82

Figure 3.16 Effects of Ubv-L12 on Grb2-mediated EGFR signaling. ............................... 83
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ab</td>
<td>antibody</td>
</tr>
<tr>
<td>Ag</td>
<td>antigen</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA binding domain</td>
</tr>
<tr>
<td>DUB</td>
<td>deubiquitinase</td>
</tr>
<tr>
<td>ECD</td>
<td>extracellular domain</td>
</tr>
<tr>
<td>EGFR</td>
<td>epidermal growth factor receptor</td>
</tr>
<tr>
<td>ErbB</td>
<td>erythroblastosis oncogene B</td>
</tr>
<tr>
<td>Fab</td>
<td>Fragment antigen-binding</td>
</tr>
<tr>
<td>Fc</td>
<td>Fragment crystalizable</td>
</tr>
<tr>
<td>Fv</td>
<td>Fragment variable</td>
</tr>
<tr>
<td>G</td>
<td>Gibbs free energy</td>
</tr>
<tr>
<td>Grb2</td>
<td>Growth factor receptor-bound protein 2</td>
</tr>
<tr>
<td>Her</td>
<td>human epidermal growth factor receptor</td>
</tr>
<tr>
<td>hGH</td>
<td>human growth hormone</td>
</tr>
<tr>
<td>hGHR</td>
<td>human growth hormone receptor</td>
</tr>
<tr>
<td>$K_D$</td>
<td>dissociation constant</td>
</tr>
<tr>
<td>$K_{off}$</td>
<td>dissociation rate</td>
</tr>
<tr>
<td>$K_{on}$</td>
<td>association rate</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>nM</td>
<td>Nanomolar</td>
</tr>
<tr>
<td>PDZ</td>
<td>postsynaptic density protein 95; discs large; and zonula occludens-1</td>
</tr>
<tr>
<td>pM</td>
<td>picomolar</td>
</tr>
<tr>
<td>PPI</td>
<td>protein-protein interaction</td>
</tr>
<tr>
<td>PPIC</td>
<td>protease phosphatase inhibitor cocktail</td>
</tr>
<tr>
<td>PRM</td>
<td>peptide recognition module</td>
</tr>
<tr>
<td>PTI</td>
<td>pancreatic trypsin inhibitor</td>
</tr>
<tr>
<td>s</td>
<td>second</td>
</tr>
<tr>
<td>scFV</td>
<td>single chain fragment variable</td>
</tr>
<tr>
<td>SH</td>
<td>Src Homology</td>
</tr>
<tr>
<td>TF</td>
<td>transcription factor</td>
</tr>
<tr>
<td>Ub</td>
<td>ubiquitin</td>
</tr>
<tr>
<td>UBA</td>
<td>ubiquitin association domain</td>
</tr>
<tr>
<td>UBD</td>
<td>ubiquitin binding domain</td>
</tr>
<tr>
<td>UIM</td>
<td>ubiquitin interacting motif</td>
</tr>
<tr>
<td>wt</td>
<td>wildtype</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>alpha</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Beta</td>
</tr>
<tr>
<td>$\beta$-GF</td>
<td>beta grasp fold</td>
</tr>
<tr>
<td>$\mu$M</td>
<td>micromolar</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Protein-Protein Interactions

Proteins are the workhorses of the cell. They maintain homeostasis and cell survival by performing functions such as catalysis, activation or inhibition of other proteins, translocation of biological molecules between compartments, immune recognition to destroy foreign entities, and assembly of macromolecular architectures. All of these functions are rooted in protein-protein interactions (PPIs), the act of one protein contacting another. Not only are PPIs critical for cell survival, they are also fascinating in their own right. Far from the textbook view, PPIs are not blocks that merely complement each other in shape and exist only in either bound or unbound form. Rather, PPIs are extremely dynamic and vary in affinity, specificity and binding mechanisms.

1.1.1 Affinity and Specificity

At the molecular level, protein interactions occur because there exists complementarity between surfaces of two proteins. Such complementarity include the shape of the two molecules, charge, electrostatics, dipole-dipole forces (e.g. hydrogen bonds), van der Waals forces and hydrophobic packing to shield non-polar groups from solvent or other polar environments. PPIs can be characterized by two parameters: affinity and specificity. Affinity is the strength of the interaction between the two proteins, and is quantified by the equilibrium dissociation constant, $K_d$, (units in mol/L or M), which also equates to the concentration required to reach 50% binding. As such, the amount of protein complexes formed within a cell is influenced by local concentrations. $K_d$ is also equivalent to the ratio of association and dissociation rates ($K_d=k_{off}/k_{on}$) of an interaction. In physical terms, affinity is related to the Gibbs free energy of association ($\Delta G=RT\ln K_d$), which must be negative for a complex to form. Gibbs free energy can also be represented in terms of changes in enthalpy ($\Delta H$) and entropy ($\Delta S$), as follows: $\Delta G= \Delta H-T\Delta S$. The concept of entropy is relevant to protein binding mechanisms discussed in the next section.

Affinity is an important parameter in PPIs, and contributes to the proper execution of biological processes. An example that highlights the importance of affinity is the barnase-barstar
complex. Barnase is a secreted ribonuclease first discovered in the bacterium, *Lactobacillus amyloliquefaciens*. Barnase is lethal to bacteria when expressed in the absence of its inhibitor Barstar, as the enzyme readily cleaves RNA within the cell. Due to the inhibitors critical role in preventing cell death, the affinity of the interaction is extremely high with an association rate ($k_{on}$) of $> 5 \times 10^9 \text{M}^{-1} \text{s}^{-1}$ and a $K_d$ of $1.3 \times 10^{-14} \text{M}$ (Schreiber and Fersht 1993). Thus, binding is practically irreversible at the concentrations present in the cell, and this ensures that barnase activity is potently inhibited.

Yet in other scenarios, low affinity interactions are optimal. PPIs in signaling are generally much more transient. In signaling pathways, interactions mediated by many peptide recognition modules (PRMs), such as Src homology-2 (SH2) and SH3 domains, often recognize their binding partners with affinities in the $10^{-6} \text{M}$ to $10^{-3} \text{M}$ range, and this is thought to enable rapid response to external stimuli (Pawson and Nash 2000). Therefore, in the context of biological efficacy, differing PPI affinities are necessary to accurately execute and regulate cellular processes.

The other parameter of PPIs is specificity. A specific interaction is defined as a functional biologically relevant recognition event, while non-specific interactions are viewed as off target binding, which are usually several orders of magnitude lower in $K_d$ and do not contribute to purposeful biological output (Kastritis and Bonvin 2013). Specificity can also be used to describe the number of functional binding partners a protein has, a protein can be monospecific or multispecific in its interactions. For example, some proteins recognize few partners that are usually similar in structure, while others mediate many interactions with diverse structural partners. In an extreme case of binding promiscuity, the transcription factor p53 contains numerous binding sites and is known to interact with at least 84 proteins (Schreiber and Keating 2011), which enable responses myriad of cellular insults to initiate processes such as apoptosis, DNA damage response, cell cycle arrest and autophagy modulation (Zilfou and Lowe 2009). As such, protein-binding sites can be plastic to accommodate different binding partners, and it is hypothesized that extreme multispecificity is a hallmark of “hub” proteins for mediating cross talk between different pathways (Peleg, Choi et al. 2014).

PPI specificity is especially important when viewed within the context of the cell. When proteins reach high enough concentrations, they will inevitably interact. Thus, in a crowded cellular environment, where total protein concentrations have been estimated to reach mM range
(Milo 2013), protein-binding specificity plays an important role in avoiding miscommunication. Furthermore, the chances of binding within a cell are compounded by other cellular processes such as expression levels, compartmentalization, post-translational modification and pH. Thus, affinity and specificity measurements in vitro might not necessarily translate to the cellular context. However, by comparing specificity and affinity of proteins measured under identical in vitro conditions, one can gain invaluable insights into molecular determinants that dictate binding and biology.

From the above examples, it is clear that affinity and specificity of PPIs are integral to the proper execution of biological processes. Due the importance of affinity and specificity in shaping biological processes, it is important to understand the molecular underpinnings by which the two parameters are derived.

1.1.2 Binding mechanisms and their effects on affinity and specificity

Understanding how proteins establish binding is pivotal to understanding how affinity and specificity are derived. The processes by which two proteins come together in the thermodynamic sense can be very different, and can explain why two proteins have high affinity for each other or why a protein is specific or promiscuous in its interactions.

Three dominant binding mechanisms have been proposed overtime, each offering a different proposition on how proteins associate. The first mechanism, the lock and key model (Fig 1.1A), was proposed by Emil Fischer more than 100 years ago (Fischer 1894). According to this model, proteins possess shape complementarity like pieces of a puzzle, and binding depends on the degree of complementarity in the pre-bound state. This description was founded on observations of enzyme-substrate complexes (Ruhlmann, Kukla et al. 1973), which tend to have rigid binding pockets. Interactions that follow the lock and key model have excellent superposition of unbound structures to the complex. According to this theory, the inflexibility of the binding interface disallows the protein from binding others that are very different from the shape of preferred partner, and thus binding is usually limited to a few partners that are structurally similar. Lock and key interactions are usually high in affinity because they tend to incur a low binding entropic penalty compared to interactions that follow other binding mechanisms, because the starting pre-bound proteins have already restricted motion. Furthermore, a protein that has a preformed binding configuration usually has a faster
association rate, $k_{on}$, to its cognate partner because it does not require time to reorient into a binding competent molecule. This can be seen in the comparison between trypsinogen and trypsin. At the sequence level, the only difference between trysin and trysinogen is that trypsinogen has a 15 residue N-terminal extension compared to trypsin. However, at the structure level, trypsinogen possess a disordered binding site for the pancreatic trypsin inhibitor (PTI), which becomes ordered when bound to PTI, while trypsin has an ordered PTI binding site prior to PTI binding. (Kossiakoff, Chambers et al. 1977, Bode, Schwager et al. 1978). Comparison of binding kinetics revealed that trypsinogen has a $k_{on}$ for PTI that is several orders of magnitude lower than that of trypsin and PTI, likely reflecting the transition from the disordered to the ordered state.

The second theory, the induced-fit model, was proposed more than 50 years after the Lock and Key model, by Daniel Koshland (Koshland 1959). He proposed that rather than rigid bodies, proteins structures are dynamic and the binding shapes are only induced or formed when two proteins come together (Fig 1.1B). In other words, proteins adjust to each other to adopt complementing shapes. At the temporal scale, induced fit binding requires reorientation of the proteins and thus retards the association rate, resulting in interactions that can be lower in affinity.

A third model, conformation selection, was developed in 1965 and describes binding as a process that selects one of many pre-existing conformations in equilibrium (Fig 1.1C) (Monod, Wyman et al. 1965, Rubin and Changeux 1966). Upon binding, the ligand stabilizes one conformation of the partner and thus shifts the equilibrium to favour that particular conformation. Evidence for this mechanism was seen in NMR studies of unbound ubiquitin (Ub), which showed an ensemble of solution conformations that covered a span of Ub conformations observed in previously solved crystal structures of Ub complexes (Lange, Lakomek et al. 2008). This theory has been used to explain binding promiscuity, as one protein can bind to different partners using different conformations within its naturally occurring ensemble. With regards to affinity, such a theory would result in interactions that have a lower association rate than the lock and key model, because time is required for a protein to search for its binding conducive conformation and accompanies a general lost in entropy as proteins will become restricted in their bound conformations. Evidence can be seen in antibodies (Abs), which are known to bind through the conformation selection model (Berger, Weber-Bornhauser...
et al. 1999, James, Roversi et al. 2003). However, affinity matured Abs that exhibit tighter affinity and specificity towards their cognate antigens, often have much reduced flexibility compared to germline Abs, and it is believed that the rigidification of the complementarity determining region (CDR) loops changes the binding process to more closely resemble the lock and key model (Manivel, Sahoo et al. 2000, Jimenez, Salazar et al. 2003, Sagawa, Oda et al. 2003).

More recently, a combination of the induced fit model and conformation selection model have been observed to apply simultaneously to a binding interaction, albeit at differing segments of a protein. Many examples have been observed where the binding conformation at the global scale already exists in solution, but the local binding site rearranges only in the presence of ligand (Wlodarski and Zagrovic 2009, Csermely, Palotai et al. 2010, Kurisaki, Takayanagi et al. 2014).
Figure 1.1 Models for protein-binding mechanisms

(A) Lock and key. (B) Induced-fit. (C) Conformational selection. (D) Conformation selection and induced-fit.

1.2 Characteristics of protein binding surfaces

A holy grail in protein science has been to accurately predict where PPIs will occur and in turn decipher the connectivity of PPIs within a cell. To narrow down the search of potential binding events, much research has been dedicated to answering the question: what makes a protein binding site? The question is particularly interesting to researchers within the field of synthetic biology, where we hope to harness some of the basic biochemical principles of PPIs to create customized or de novo binding proteins. The characteristics of ligand binding sites on enzymes are easy to describe because they are usually structurally obvious, forming deep clefts, pockets or crevices. However, protein binding sites are much more difficult to differentiate from the rest of the exposed surface, in terms of structure and character.

1.2.1.1 Amino Acid Composition

It is widely accepted that PPI are mainly driven by packing of hydrophobic residues at the binding interface (Jones and Thornton 1996, Keskin, Gursoy et al. 2008), with aromatic residues (His, Tyr, Trp, and Phe) more prevalent than aliphatic residues (Met, Leu, Ile, and Val) when compared to the average protein surface (Lo Conte, Chothia et al. 1999). However, hydrogen bonding and ion pair interactions mediated by hydrophilic and charged residues also play important roles in binding (Keskin, Gursoy et al. 2008). In addition, binding sites do not have to be the largest exposed hydrophobic surface on a protein (Keskin, Gursoy et al. 2008). Therefore, while hydrophobicity is an important parameter for binding, it is only a loose criterion for defining PPIs. In addition, amino acid propensities also vary depending on the type of PPI. Obligatory protein complexes (proteins that must oligomerize to function) have an even higher preference for hydrophobic residues.

Many insights in molecular recognition emerged from studies of Ab-antigen (Ag) interactions. The six CDRs or hypervariable loops of Abs form the antigen-binding site, which is also known as the paratope (Wu and Kabat 1970). In the immune system, the diverse sequence variation in CDRs can create large Ab repertoires that can theoretically recognize any
foreign molecule (discussed further in section 1.2.1). Interestingly, sequence analysis of natural paratopes revealed that Tyr residues are prevalent in CDRs (Janin and Chothia 1990, Mian, Bradwell et al. 1991), and if considering only functional contacts to Ag, Tyr abundance is even higher at \~25\% (Mian, Bradwell et al. 1991). It was hypothesized that the aromatic nature of Tyr in conjunction with a hydroxyl group on the side chain makes it a versatile amino acid for molecular recognition. The aromatic ring can take part in \(\pi\)-stacking and large van der Waals interactions, while the hydroxyl group can form a hydrogen bond with partner proteins.

Furthermore, the functionality of Tyr in Ab-Ag interactions was highlighted in synthetic Ab studies, where one can precisely control the sequence composition of CDRs and in turn test hypotheses regarding the functionality of particular amino acids. Minimalist antibody libraries with only Tyr and Ser residues in the CDRs have generated binding Abs to diverse Ags with high affinity and specificity (Fellouse, Wiesmann et al. 2004, Fellouse, Li et al. 2005, Fellouse, Barthelemy et al. 2006, Fellouse, Esaki et al. 2007). Ser is a small polar amino acid that can play an auxiliary role by providing flexibility to position Tyr for optimal antigen contacts. However, Tyr dominated CDRs are not optimal for all Ags, and additional chemical diversity has been shown to improve binding to some antigens (Birtalan, Fisher et al. 2010). Another notable observation was that while Arg is favourable for affinity, it is deleterious for specificity (Birtalan, Zhang et al. 2008). Many of these insights were applied to protein engineering efforts and will be discussed in Chapter 3.

1.2.1.2 Hot spots

Another method to describe the binding interface surface relies on the architectural organization of the surface rather than the amino acid composition.

When proteins interact, previously exposed protein surfaces become buried. Analyses of many protein complexes have shown that about 1000-2000 Å\(^2\) of surface area on each protein is buried (Hwang, Vreven et al. 2010). Interestingly, despite the large area that participates in binding, the binding energy released upon association of a pair of proteins is not equally distributed amongst the binding surface, and this discovery has had profound implications on drug design for perturbing PPIs (Arkin, Tang et al. 2014).

Wells and colleagues pioneered alanine-scanning mutagenesis as a methodology to assess energetic contributions by individual side chains within the binding surface area.
Substitution with Ala removes all side chain atoms past the β-carbon of the residue without altering main chain conformations or introducing electrostatics, thus allowing one to infer consequential binding effects of individual side chains (Clackson and Wells 1995, Wells 1996, Wells and de Vos 1996). By individually replacing all residues within the binding surface area, one can scan the entire buried surface and determine the residues that functionally contribute to binding. When Wells and colleagues applied alanine-scanning to human growth hormone receptor (hGHR), and measured resulting binding to human growth hormone (hGH), they discovered that of the 33 residues that became buried upon complexation, only nine residues had substantial negative effects on binding to HGH (ΔΔG ≥ 2 kcal/mol). Hence within the buried surface area, also termed the ‘structural epitope’, most residues were actually energetically neutral for binding, and only a subset of residues within the structural epitope are responsible for the affinity of the complex. The residues that are responsible for the affinity of a PPI constitute the “hot spot” or functional epitope within the structural epitope, and together often form a contiguous patch.

Much work has been done to characterize the anatomy of hot spots. In a study where 2325 Ala mutants of protein complexes were analyzed, Tyr, Trp and Arg dominated hot spots (Bogan and Thorn 1998), which further reinforced the findings derived from Ab studies that showed that Tyr is important for molecular recognition. Another intriguing aspect of hot spots is that hot spot residues are often more dynamic than the remaining binding surface, and the hot spot residues accommodate diverse rotameric conformations (DeLano, Ultsch et al. 2000, Sundberg and Mariuzza 2000). Moreover, peptide phage display studies have shown that protein binding peptides that exhibit no homology to cognate ligands also bind to hot spots of protein, and thus suggest that hot spots are better adapted for binding, and possibly possess a unique molecular anatomy (Fairbrother, Christinger et al. 1998, DeLano, Ultsch et al. 2000, Schaffer, Deshayes et al. 2003).

From a therapeutic standpoint, PPIs were traditionally assumed to be intractable to inhibition by small molecules. Unlike enzymes and receptors that have deep binding pockets for small ligands, PPI surfaces cover expansive flat areas that were thought to be too shallow to accommodate small molecules. However, the understanding of hot spots and their anatomy led to the realization that targeting PPI for therapeutic intervention is feasible, as one only needs to target the hot spot when designing drugs. The hot spot area covers a much smaller area relative
to the structural epitope, and has been observed to possess dynamic behavior that can accommodate small molecules. In fact, while unbound structures might seem to be devoid of binding clefts or pockets, the few available solved structures of small molecules bound to hot spots of protein-protein interfaces have revealed unexpected cavities that were only visible when bound to the small molecules (Arkin, Randal et al. 2003, Grasberger, Lu et al. 2005, Bruncko, Oost et al. 2007). As of 2014, 26 PPI inhibitors were reported (Rognan 2015), which further reinforces the notion that PPIs are tractable for small molecular targeting.

1.2.2 Studying PPIs

While structural elucidation by x-ray crystallography, NMR, and more recently high-resolution electron microscopy enable the visualization of protein complexes in three dimensions, these techniques only provide limited insights into relationships between structure and function. Structures of static protein complexes do not give information on the plasticity of the binding site nor on the functional epitope that is energetically responsible for interaction. Site-directed mutagenesis is an invaluable technique for elucidating structure-function relationships. It endowed biochemists with the ability to alter protein structure by making specific changes at the primary sequence level, and subsequently testing for consequential functional changes (Flavell, Sabo et al. 1975, Hutchison, Phillips et al. 1978). By identifying and mapping functional epitopes onto three-dimensional structures, one can understand the organization and composition of functional epitopes. This approach is not only important for answering the fundamental question of molecular recognition, but in more practical terms, it provides clues for protein and drug design.

As described above, alanine-scanning mutagenesis is a powerful method for probing the functional contributions of individual side chains in PPIs. However, the technique is labourious as it requires cloning, purification and binding analysis of individual Ala point mutants that together cover the entire binding surface. To expedite the process, a combinatorial approach called shotgun alanine-scanning was developed to produce combinatorial phage-displayed libraries of mutants and collectively assess binding in a pooled format.

1.2.2.1 Phage Display

Phage display is a powerful technology that is now routinely used to study PPIs and for protein engineering (see also section 1.2.3). In 1985, George Smith demonstrated that peptides can be
genetically fused to the phage outer coat protein, which resulted in the external display of a heterologous peptide on phage (Fig. 1.2)(Smith 1985). In 1990, the Wells and Winter groups demonstrated that large proteins can also be functionally displayed on phage (Bass, Greene et al. 1990, McCafferty, Griffiths et al. 1990), and this accelerated the field of protein engineering. The power of phage display and other in vitro display technologies (Lipovsek and Pluckthun 2004, Pepper, Cho et al. 2008, Lofblom 2011) stems from the fact that there exists a physical linkage between genotype (gene coding the displayed protein) and phenotype (the displayed protein), permitting for in vitro evolution to occur. By exploiting the life cycle of phage, one can amplify and enrich for ‘fitter’ molecules on phages that exhibit desirable qualities in parameters such as affinity, specificity and function.

The most common species exploited for phage display is the non-lytic filamentous bacteriophage M13, which infects a variety of gram-negative bacteria including the laboratory workhorse Escherichia coli (Smith and Scott 1993). The workflow of phage display is simple and cost effective, and thus can be easily adopted in a common laboratory setting. The concept relies on the production of a massive library of diversified protein variants on phage (routinely >10^10)(Fellouse 2007), and subjecting the library to a panning process to select for binding variants, which can then be amplified in the E.coli host (Fig 1.2). By iteratively performing the selection, one can enrich for binding molecules. Since the selection process is performed independently of a living cell, one can employ different in vitro conditions to select for displayed molecules exhibiting desired properties such as stability (Jung, Honegger et al. 1999) or slow off rates (Levin and Weiss 2006). Another benefit is that one can easily identify the binding proteins by sequencing the encoding gene encapsulated in the phage, which allows for binder characterization independent of phage or further engineering efforts.
A Phage pool representing a displayed combinatorial library of protein variants is added to immobilized antigen to allow binding phages to interact with antigen. Non-binding phages are washed away and bound phages are subsequently amplified in a bacterial host. The selection process can be repeated with the amplified phage pool until desirable binding phages are identified.

Other display systems have emerged over the years, such as bacterial (Lofblom 2011), yeast (Pepper, Cho et al. 2008) and ribosomal display (Lipovsek and Pluckthun 2004), each aimed to add a different functionality to the display technology at the expense of other limitations.

1.2.2.2 Applications of phage display in PPI studies

The advent of phage display provided a combinatorial mutagenesis approach to tackle the structure-function problem in a high-throughput (HTP) manner. It enabled the facile production of libraries of mutant proteins that subsequently can be rapidly tested for binding. In shotgun alanine-scanning, one can produce a combinatorial library of Ala mutants and enrich for binding mutants by panning the library against the cognate binding partner. Subsequent sequence analysis of binding variants can reveal positions that are depleted in Ala, and thus one can deduce (wild-type) wt residues that are critical for binding. Under the assumption that the wt occurrence to Ala occurrence ratio (wt/Ala) is proportional to equilibrium binding constants of the wt or Ala mutants, one can perform statistical analyses of the wt/Ala ratios (after proper
correction for display biases) to calculate $\Delta \Delta G_{\text{mut-wt}}$ at each tested position. Combinatorial Ala scanning by phage display was validated by Sidhu and colleagues when they benchmarked alanine-scanning data produced from the combinatorial method to that produced by conventional alanine-scanning (Weiss, Watanabe et al. 2000). Thus, shotgun alanine-scanning circumvented the need to make kinetic measurements on individual mutants, and expedited the process of mapping binding hotspots.

In addition to alanine-scanning, other combinatorial scanning strategies have also emerged to unveil other relationships between protein structure and function. Recognizing that substitution by an apolar Ala side chains may overstate the binding contributions of hydrophilic side chains, serine-scanning was devised to assess if binding contributions will differ when scanned with a small polar side chain (Pal, Fong et al. 2005). Results from serine-scanning of hGH were in good accord with results from alanine-scanning, indicating that both apolar and polar residues were good substitutes for assessing binding contributions in this case. Homolog-scanning is an alternative method in which side chains are substituted with similar side chains, and this method was designed to assess binding effects introduced by subtle structural changes at the binding surface (Pal, Fong et al. 2005). Interestingly, the homolog scan of hGH showed little effect on binding to the hGHR, indicating that affinity cannot be improved substantially with only subtle changes to the binding surface.

An even more powerful method is the combinatorial saturation scan by phage display, which involves making libraries where the entire binding surface is substituted by all 20 amino acids (Pál, Kouadio et al. 2006). Such a scan provides a comprehensive assessment of the relationship between function and all the possible structural diversity afforded by the 20 amino acids. Results generated from such a scan offer tremendous insight into the adaptability of binding surfaces. Such a scan is practically inconceivable if not for display technologies, and will be discussed further in chapter 2.

Others have applied phage display to uncover relationships between protein structure and dynamics. Rather than strictly exploring binding surface structure-function relationships, Zhang and colleagues designed libraries to study protein interior core substitutions and consequential functional effects. Interestingly, protein core substitutions altered protein structural dynamics and stabilized some conformations over others, which resulted in biasing certain interactions at the expense of others in cells (Zhang, Zhou et al. 2013).
Yet another application of phage display in exploring PPIs involved tuning the affinity or specificity of naturally low affinity binding sites. These applications are highlighted in studies of SH3 and PDZ domains (Ernst, Gfeller et al. 2010, Teyra, Sidhu et al. 2012), which in nature often recognize many linear peptide ligands with low affinity. These studies proved to be powerful in elucidating the molecular determinants of affinity and specificity. Furthermore, by expressing engineered PRM variants in cells, one can study the consequential changes in PPI networks and signaling effects of a monospecific PRM (Findlay, Smith et al. 2013).

1.3 Protein Engineering of Affinity Reagents

The discipline of protein engineering emerged in 1980s, initially out of the necessity to improve industrial enzyme properties such as efficiency, robustness or specificity (Ulmer 1983). The idea of improving enzymes was posited upon the observation that nature had similar enzymes that exhibited wide variations in traits such as thermo stability and catalytic rate constants. In addition, natural enzymes harbouring genetic variations showed distinct properties (Scaccioch and Sealy-Lewis 1978). Thus, it was logical to propose that one could use site-directed mutagenesis to change structural features and endow an enzyme with desirable traits.

In addition to enzyme enhancements, it became apparent that the idea of changing structure and thus function of a molecule could be applied to another branch of biotechnology, the development of tailored affinity reagents. Affinity reagents are valuable tools in detecting analytes in basic research and in medicine. Not only can affinity reagents be used as diagnostics, but they can also be used as therapeutics that function to neutralize infectious agents or harmful overexpressed soluble proinflammatory molecules such as TNF-α (Feldmann and Maini 2001), to activate beneficial surface receptors such as CD40 (Beatty, Chiorean et al. 2011) or to antagonize unwanted immunosuppressive signals or disease causing receptors such as EGFR, (Van Cutsem, Kohne et al. 2009). In addition, small molecules or other toxins can be ligated to an affinity binder to deliver payloads to diseased cells. In biotechnology, affinity proteins are conjugated to resin for use in affinity purification of recombinant proteins. Traditionally, affinity reagents are antibody based. However, over the years, new classes of affinity reagents based on alternative frameworks have been engineered to bind targets of interest.
1.3.1 Antibodies

Abs are naturally occurring proteins in the immune system, evolved as part of the humoral immune response to recognize Ags, which are usually foreign immunogenic molecular entities that can induce an immune response. The ability of Abs to recognize diverse molecular patterns lies in the molecule’s unique structure. Each Ab is a “Y” shaped assembly of two heavy and two light chains (Fig 1.3). The stem of the Y structure is a constant “fragment crystallizable” (Fc) domain that is responsible for extending in vivo half-life and for mediating immune effector functions. Each arm of the Y structure is a ‘fragment antigen binding’ (Fab) domain that is responsible for antigen recognition.

The Fab contains a fragment variable (Fv) domain that is composed of two juxtaposed immunoglobulin (Ig) folds. The Ig fold is a beta sandwich made of two $\beta$-sheets held together by a disulfide bond, and the loops at one end of the fold are the CDR or hypervariable loops. In total, each Fv has six CDR loops that are responsible for engaging antigens. The surface area on the CDRs that becomes buried upon binding to the antigen is known as the paratope. The B cells that produce Abs undergo various genetic shuffling events known as VDJ recombination to produce sequence diversity within the CDR loops. Usually the naïve Ab repertoire will only produce low affinity antibodies. However, under antigen stimulation, B cells producing low affinity Abs will undergo an affinity maturation process to produce higher affinity Abs. In addition to natural repertoires (diversity encoded by B cells), synthetic repertoires, where CDR diversities are man-made, have also proven to be useful for generating binding molecules to diverse antigens (see Section 1.2.3).
Figure 1.3 Antibody structure

(A) Antibody schematic. (B) Fab structure. Light and heavy chain frameworks are colored tan and blue, respectively. Light and heavy chain CDRs are colored magenta and green, respectively (PDB entry 1ZTI).

1.3.1.1 Antibody Limitations

The relationship between Ab CDR sequence variation and the ability to form novel binding interfaces to any molecular entity with high affinity and specificity have long been exploited for the development of binding reagents for basic research and medicine. However, the continuous development of novel affinity reagent applications has demanded more than what is possible with natural Abs.

As previously mentioned, Abs are large (150 kDa) tetrameric proteins. They are heavily disulfide bonded and glycosylated assemblies that were evolved to exist in the oxidizing extracellular environment. Not only does the complex structure result in cost-intensive tissue culture production methods, the complexity limitations are not ideal for certain applications. From an in vivo diagnostic standpoint, Abs face two challenges. First, due to their large size, Abs cannot penetrate tissues easily. Second, Abs have a slow circulation clearance rate, because Abs are above the renal clearance size of ~70kDa and they possess an Fc domain that mediates recycling by binding to the neonatal Fc receptor (Chaudhury, Mehnaz et al. 2003, Montoyo, Vaccaro et al. 2009). As such, Abs are not ideal for diagnostic applications where clearance rate
needs to be rapid so that imaging tools such as radioisotope conjugated Abs are excreted as soon as possible post imaging to reduce toxic effects. While smaller Ab fragments such as Fabs or scFv domains, which lack the Fc region and are smaller in size (refer to Fig 1.3) can theoretically be substituted for full size Ab for diagnostic purposes, they remain disulfide bonded, which require production methods that are usually more expensive and/or result in lower production yields (Frenzel, Hust et al. 2013, Schlegel, Rujas et al. 2013). In addition, since these fragments lack the Fc portion, they are often less structurally stable and hence more prone to aggregation (Wörn and Plückthun 2001, Demarest, Chen et al. 2006, Perchiacca and Tessier 2012, Gil and Schrum 2013).

Another application limitation is that Ab tools are not suitable for intracellular applications. In basic research, Abs are used to tease out surface receptor function. Since natural ligands of receptors are often multispecific and can stimulate different receptors simultaneously (Yarden and Sliwkowski 2001, MacEwan 2002, Lisabeth, Falivelli et al. 2013), Abs are especially useful in deciphering receptor biology as they can be designed to specifically activate or inhibit a single receptor with high affinity. In addition, Abs are valuable in teasing out receptor structure and function relationships by trapping receptors in particular conformations or linking targeted epitopes on receptors to downstream functional consequences. However, Abs are less effective tools for probing intracellular protein function, because Abs are not cell membrane permeable. Moreover, even with gene delivery methods coupled with intracellular expression, Abs are structurally reliant on disulfide bonds, and thus do not fold and function well in the reducing intracellular environment. While small molecules exist to dissect some protein functions, a large fraction of these only target enzymes, which have deep binding pockets. Since many protein functions are a consequence of PPIs and PPIs generally lack deep pockets, it has been especially challenging to develop small molecule based tools to probe intracellular PPIs. Thus, there remains a severe paucity of tools for investigating intracellular protein functions.

1.3.2 Alternative Scaffolds

Alternative scaffolds were first conceptualized in the early 1990s. The goal was to create antibody mimetics to overcome some of the limitations of Ab reagents. The idea revolved around constructing synthetic antigen-binding sites on other protein scaffolds. The invention of alternative scaffolds was spurred by the success of synthetic antibody libraries and display
technologies, which enabled HTP protein engineering. Like Abs, alternative scaffolds have a stable three-dimensional structure and are structurally robust so that they can tolerate extensive sequence variation in the designed binding site without compromising structural integrity. Additionally, these alternative scaffolds must be able to form high affinity and specific interactions to antigens of interest. However, to overcome Ab limitations, these alternative scaffolds should ideally be structurally simple, single domain, small in size (<15kDa) and lack disulfides so they can be produced in high yields using recombinant bacterial methods, have broad applicability, and can be easily linked in modular fashion for novel effector functions.

Over the last 25 years, more than 50 alternative scaffolds have been reported (Gebauer and Skerra 2009). The proteins chosen to be alternative scaffolds originate from diverse species and perform a wide array of native functions. Few groups have taken advantage of extremophile bacteria, which naturally express proteins evolved to withstand extreme conditions, and therefore are naturally structurally robust (Gera, Hussain et al. 2011, Behar, Bellinzoni et al. 2013). Other scaffolds are protease inhibitors (Dennis and Lazarus 1994, Smith, Patel et al. 1998, Williams and Baird 2003) or toxins (Pierret, Virelizier et al. 1995, Li, Dowd et al. 2001), which have small compact robust folds, most probably evolved to be exceptionally stable to escape proteolytic cleavage while inhibiting protease active sites. While others chose proteins that resemble the immunoglobulin fold (Fig 1.4A), such as fibronectins (Fig 1.4B), which also has a beta sandwich fold with loops connecting two beta sheets. Based on the concept of Ab CDRs, scaffolds such as Anticalins (Beste, Schmidt et al. 1999), Knottins (Moore and Cochran 2012), and Kunitz domains (Hosse, Rothe et al. 2006) also utilize combinatorial loops to create novel binding sites. Yet others have co-opted other secondary structures that are entirely comprised of helices or a beta sheet to support a universal binding site. Thus, while nature has utilized flexible loops for Ab paratopes, the ability to form a complementarity region is not restricted to loop based architectures and can be achieved with other structures.

1.3.3 In vitro display systems for affinity reagent engineering

The discipline of structural biology has provided much insight into molecular recognition, and the knowledge governing PPIs have been in turn used to develop algorithms for interface design. Although computational protein designs have made remarkable achievements in designing de novo proteins or re-designing enzymes (Koga, Tatsumi-Koga et al. 2012, Siegel, Smith et al. 2015, Boyken, Chen et al. 2016), current methods are only applicable for designing
binding molecules to certain proteins that have existing high-resolution structural information and predicted candidate sequences are only partially accurate. Furthermore, binding sequences still require additional experimental optimization to achieve high affinity and specificity. As such, *in vitro* evolution systems still remain the dominant choice for affinity reagent engineering, especially for scaffold-based binders.

Molecular biologists have long taken advantage of natural immune systems to produce high affinity and specificity Abs. However, many practical drawbacks exist in the immune system when it comes to antibody production. First, antigens used to raise antibodies cannot be toxic or non-immunogenic. Second, it is difficult to produce antibodies against antigens that are either identical or have high similarities to host molecules because the immune system goes through a negative selection process to rid cells that produce antibodies that recognize self-epitopes. Third, the conditions in which an immune system develops antibodies cannot be controlled, and as such one cannot tailor Abs against a particular epitope of a domain or tweak antibody biochemical properties such as solubility. Fourth, the sequence of antibodies raised in an animal cannot be easily identified, which makes it difficult to streamline antibody production in recombinant systems, add functional modifications, or humanize for therapeutic development. In fact, the lack of sequence information on research grade Abs has been criticized as a major contributor to published data irreproducibility issues (Baker 2015, Bradbury and Plückthun 2015). Since Ab sequences are unknown, Ab production is tied to the immunized animals, which can produce different Ab clones when immunized with the same Ag. Even when Abs are produced using hybridoma technology, which in theory should produce monoclonal Abs, variability can be introduced in the produced Abs because biological changes can occur in hybridoma cell lines over long-term propagation (Couture and Heath 1995, Barnes, Bentley et al. 2003). Thus, it is difficult to produce the same Abs consistently using animals or hybridoma methods, and the batch-to-batch Ab variations have led to inconsistent experimental outcomes. The inability to define Ab based on its sequence poses another problem for users, as there is no way to confirm which Abs were actually used in published literature, so it can be difficult to match experimental setups to reproduce published data.

*In vitro* evolution using display methods has served as an attractive alternative method to the natural immune system for Ab engineering, and has been a powerful HTP platform for scaffold engineering. Combinatorial scaffold libraries, synthetic Ab libraries, or natural immune
repertoires can be cloned and displayed on various vectors such as phage, RNA or yeast. Subsequently, the libraries are subjected to selection pressure in \textit{in vitro} settings to select for affinity molecules to targets of interest. As discussed above in section 1.1.4.1, in display systems such as phage display, selection settings are tunable to develop binders with desirable biochemical properties and competing target binding molecules can be included in the selection procedure to enrich for binders to other epitopes of the target. The ability to bias binder development against particular protein epitopes is important since binder engagement of different epitopes can result in different functional outcomes. For example, affinity reagent binding to the ligand-binding site on a receptor may mimic ligand activity and stimulate the receptor, where as an affinity reagent binding to the dimerization interface of a receptor may inhibit receptor activity. Another benefit of display systems is that the DNA coding sequence of the binder can be easily identified, which enables the binding clone to be indefinitely produced by recombinant methods, and may alleviate the data irreproducibility issues surrounding the use of affinity reagents lacking sequence information (Bradbury and Plückthun 2015, Bradbury and Plückthun 2015). The known binder sequence will also allow for further engineering, if necessary. In addition, in display systems one can precisely define the diversity of the binding sites so that residues that are known to be advantageous or deleterious for molecular recognition can be biased or depleted, respectively.

1.3.4 Examples of Successful Alternative Scaffolds

To date, numerous small biotechnology companies have been built based on alternative scaffolds. The following four scaffolds stand out as achieving much success in broad applications.

1.3.4.1 Affibodies

Affibodies, which are based on the Z domain of \textit{Staphylococcus} Protein A (Fig. 4C), is a 58 residue scaffold comprised of three helices (Nord, Nilsson et al. 1995, Lofblom, Feldwisch et al. 2010), and to date over 100 papers has been published regarding this scaffold. Thirteen residues along the face formed by two helices were chosen as the binding site. Phage-displayed Affibody libraries have been successful for selecting μM to pM affinity binders to diverse antigens including protein A, CD-28 (a T-cell activating co-receptor) (Sandstrom, Xu et al. 2003), and Her2 (a cell-surface receptor) (Orlova, Magnusson et al. 2006). Currently, over 20 Affibody-
based research reagents have been marketed, and at least two affibodies are in clinical trials (Affibody 2016). In addition, Gallium conjugated anti-Her2 affibodies have been tested in clinical studies and proved to provide better imaging contrasts than Abs (Sorensen, Velikyan et al. 2016).

1.3.4.2 DARPins

Designed ankyrin repeat proteins (DARPins) are based on the natural ankyrin repeat proteins (Fig 4D), which typically consist of four to six consecutively linked 33-residue motifs containing two alpha helices connected by a loop (Plückthun 2015). The idea of adopting repeat proteins for building synthetic binding proteins is based on the observation that in nature, repeat proteins are abundantly present in PPIs and can vary in length to create larger or smaller binding surfaces. Prior to library construction, a consensus strategy based on sequence analysis of hundreds of ankyrin repeats was adopted to design a framework with superb biochemical properties (Binz, Stumpp et al. 2003). On a single motif, six residues located on the loop and helix were randomized and a larger complementarity region was formed by assembling multiple units together. Phage-displayed and ribosome-displayed ankyrin libraries have yielded binders to hundreds of molecules and pM affinities have been reported in some cases (Plückthun 2015). Currently more than 200 publications are available on DARPins, and four DARPin molecules are in clinical trials (Molecular_Partners 2016).

1.3.4.3 Monobodies

Monobodies are based on the fibronectin type III domain (FN3) (Fig. 4B), which has been estimated to account for 2% of animal proteins. It is 96 residues long and has a fold similar to that of an Ig domain. The binding site was developed by randomizing 21 residues on the loops situated between the two beta sheets (Koide, Bailey et al. 1998). Recombinant monobody libraries based on phage, yeast and mRNA display have been used to select for binders to diverse protein folds including post-translational modification, PRMs, surface receptors and kinases, with nM to pM affinities (Lipovsek 2011, Koide, Koide et al. 2012). Currently, there are more than 30 papers published on the monobody scaffold with another 20 papers published by the company Adnexus, which has commercialized a FN3 library, and developed binders called Adnectins. An Adnectin raised against VEGFR2 was advanced through a Phase 2 clinical trial to treat recurrent glioblastoma but did not meet efficacy requirements (Schiff, Kesari et al. 2016).
Many monobodies have demonstrated success in intervening with intracellular protein function when expressed within the cell (Wojcik, Hantschel et al. 2010, Grebien, Hantschel et al. 2011, Sha, Gencer et al. 2013). Of particular note, a monobody targeting specifically an epitope at the interface between an SH2 and kinase domain within the oncogenic fusion protein BCR-Abl has validated this region as important for oncogenesis and could be a target for therapeutic intervention (Grebien, Hantschel et al. 2011).

### 1.3.4.4 Anticalins

Anticalins are based on members of the lipocalin protein family (Fig. 4E), which naturally exist in animal blood as carriers to store or transport small ligands (Beste, Schmidt et al. 1999). Lipocalins possess a beta-barrel fold with a cup-like structure that supports four loop extensions, and it is within the cup where natural ligands are known to bind. Exploiting the natural function of lipocalins, synthetic binding proteins have been isolated from phage or bacterial display libraries in which the loops and inner beta barrel residues were randomized (Gebauer and Skerra 2012). Nanomolar to pM affinity binders have been selected to bind small molecules such as steroids and digoxigenin (Schlehuber, Beste et al. 2000), and disease relevant proteins such as VEGF and CTLA4 (Richter, Eggenstein et al. 2014). Currently, there are more than 20 papers published regarding the Anticalin scaffold and two molecules are in clinical trials (Pieris 2016).

**Figure 1.4 Structures of binding domains**
(A) Antibody light chain variable domain (PDB entry 1ZTI). (B) 10th human fibronectin type III domain (PDB entry 1FNF). (C) Z domain of Staphylococci Protein A (PDB entry 1Q2N) (D) Ankryin repeats (PDB entry 1MJO). (E) lipocalin domain of Bilin binding protein (PDB entry 1BBP). β-sheets and α-helices are colored blue, loops are colored gray, and diversified positions are represented as red spheres.

1.3.5 Practical considerations for scaffold engineering

As mentioned before, computational approaches have made great strides in binding interface design. However, it continues to face many challenges, and thus, there is still a great need for carefully designed scaffold libraries for affinity reagent development.

As mentioned before, it is of paramount importance for binders to have excellent biochemical properties as this correlates with longer shelf life, higher yields, lower off target binding, and also proposed to have lower antigenicity for therapeutic applications (Pluckthun 2009). Thus, a scaffold should be naturally stable, so that it can afford to incur some instability attributed by sequence variation. As a result, some current scaffolds are natural proteins that have gone through framework engineering to further improve structural stability (Binz, Amstutz et al. 2004, Lorey, Fiedler et al. 2014). Typically, scaffold choice also takes into consideration the type of surface for binding site construction. The surface to be diversified can be loops or secondary structures, and each has advantages and disadvantages. Loop flexibility adds another dimension of diversity, as it can adopt different conformations, and can increase the chances of arriving at a binding configuration. However, once bound, flexible loops are fixed spatially, which can result in unfavorable entropic costs, and this is contrary to complementarity regions based on fixed beta sheets or helices. Regardless of surface chosen, diversified positions should be surface exposed to prevent disturbing the hydrophobic core and the overall fold of the protein.

Since every display method has a practical limit in the achievable library size, one must carefully consider the diversification strategy to maximize coverage of the theoretical diversity and to minimize the number of non-functional members due to instability issues. The amino acids used in scaffold diversification should be devoid of Cys, to prevent formation of disulfides, which can complicate folding. Gly and Pro should also be avoided on secondary structure based complementarity regions, as they can disrupt helices and beta sheets. Arg has been shown to be beneficial for affinity but deleterious for specificity, thus Arg should also be
excluded or used sparingly (Birtalan, Zhang et al. 2008). As mentioned in section 1.1.3.1, Ab studies have revealed Tyr and Ser to be important in Ab-Ag interactions and this also holds true for monobody engineering (Koide and Sidhu 2009). Thus, to maximize contacting residues one can bias diversification to favour Tyr and Ser residues. For scaffold libraries produced with oligonucleotide-directed mutagenesis strategies, precise amino compositions can be achieved with oligonucleotides synthesized with trinucleotide phosphoramidite codons (Sondek and Shortle 1992, Persson, Ye et al. 2013).

Finally, it is important to note that library design is an iterative process. Binder characterization will often provide valuable lessons for improved library designs. For example, observations have been made that Abs rich in negative charges are more soluble and less prone to aggregation (Lee, Perchiacca et al. 2013), thus future Ab library design might benefit from incorporation of additional negative charges. Structural determination of binders can also give valuable insight into randomized positions that may not be suitable for randomization or vice versa. For example, second generation monobody libraries included randomization of the “framework” because it was observed in a structure that the framework residues can also make contact (Koide, Wojcik et al. 2012).

1.4 Ubiquitin

1.4.1 Ub in biology and disease

Ub is a 76-residue post-translational modifier that is covalently attached to substrate lysines or N termini, either singly or in chains of differing topologies. These modifications serve to target substrate for degradation or to alter PPIs for orchestrating a variety of cellular processes such as gene silencing, receptor internalization (Terrell, Shih et al. 1998, Haglund, Sigismund et al. 2003), DNA repair (Ulrich and Walden 2010) and cell cycle progression (Spence, Gali et al. 2000, Jin, Williamson et al. 2008). Ub attachment to substrates is catalyzed by the sequential action of three classes of enzymes: E1, E2 and E3. (Schulman and Harper 2009, Ye and Rape 2009, Berndsen and Wolberger 2014). Ubiquitination can also be reversed by deubiquitinases (DUBs), which are proteases that cleave Ub moieties from substrates (Komander, Clague et al. 2009). In addition to the plethora of enzymes that recognize Ub, there are at least 20 structural families of Ub-binding domains (UBDs), which recognize Ub non-covalently and serve to
translate Ub modifications into signaling cascades (Hurley, Lee et al. 2006, Husnjak and Dikic 2012, Harrison, Jacobs et al. 2015). Thus, Ub function is the product of finely balanced interactions between Ub and diverse enzymes and binding domains.

Given that Ub’s pivotal role in maintaining cell physiology is dependent on the proper integration of diverse Ub-mediated interactions, deregulation of Ub interactions can have broad consequences on the cell. In fact, mutations or alterations affecting expression in many Ub associated genes are strongly implicated in diseases such as cancer, immune pathologies, and neurodegeneration (reviewed in (Popovic, Vucic et al. 2014).

Due to the role of ubiquitination in targeting proteins for proteasomal degradation, it directly controls the abundance of oncogenic proteins or tumor suppressors, as such genetic alterations of E2s, E3s and DUBs are strongly implicated in malignancies. Typical examples of E3 ligases in cancer include MDM2 (Wade, Li et al. 2013) and the SCF ligase complex (Frescas and Pagano 2008), which are responsible for regulating levels of tumor suppressor p53 or cyclin-dependent cell cycle kinases, respectively. As for DUBs, a number of them are tumor suppressors, including A20, CYLD and BAP1, which all have been found to be genetically mutated in many immune cancers such as B cell lymphomas and lymphoblastic leukemia (Popovic, Vucic et al. 2014). Furthermore, ubiquitination has been shown to directly effect oncogenes or suppressors activity, and thus offers another mechanism in which misregulation of ubiquitination can lead to malignancy (Baker, Lewis et al. 2013).

In immune pathologies, misregulation of ubiquitination is often described in the context of the NFκB pathway, which has critical roles in coordinating cellular response to stress and immune response to infection (Kanarek, London et al. 2010, Chen and Chen 2013). The canonical NFκB pathway can be activated through many cytokine or pathogen sensing receptors, and once activated, many downstream components are ubiquitinated in different topologies, which serve as a signal for degradation or as a scaffold for complex formation. Ubiquitination is critical for proper NFκB signaling, and this is reflected by the many genetic alternations in ubiquitination along the NFκB pathway in immune disorders such as immunodeficiencies, chronic auto-inflammation, rheumatoid arthritis, psoriasis, systemic lupus erythematosus and celiac disease(Popovic, Vucic et al. 2014).

Neurodegenerative diseases such as Parkinson’s or Alzheimer’s are characterized by
toxic accumulation of protein aggregates. Not surprisingly, both diseases are linked to defects in Ub mediated degradation or autophagic pathways that are normally responsible for removing aggregates (Lam, Pickart et al. 2000, Hara, Nakamura et al. 2006). Interestingly, ubiquitination at specific sites of α-synuclein, the main aggregate forming protein in Parkinson’s, seems to promote toxic fibril formation, while ubiquitination at other sites strongly inhibits fibril formation (Meier, Abeywardana et al. 2012). This suggests that differential Ub interactions are responsible for pathogenesis.

All of the above diseases highlight the importance of regulating ubiquitination and speak to the importance of Ub-mediated interactions.
1.4.2 The Ub fold

Ub is the prototype of the β-grasp fold (β-GF). The β-GF superfamily is widespread amongst the animal kingdom, and in eukaryotes there are at least 70 subfamilies identified (Burroughs, Iyer et al. 2012). According to Burroughs et al., the common β-GF topology is a four or five-stranded beta-sheet and a helical segment. The first and last β-strands are adjacent to each other and located in parallel in the center of the domain. The two remaining strands are found flanking the middle two, in anti-parallel fashion. The single helix is packed against one face of the sheet, and it appears as if the β-sheet is grasping the helical segment (Fig. 1.5A).

![Diagram of the Ub fold](image)

Figure 1.5 The β-grasp fold

Schematics (upper panel) and structures (lower panel) are shown for (A) the typical five-stranded β-GF fold, for which Ub (PDB entry: 1UBQ) is an example, and (B) the β-GF fold of the Fasciclin subfamily, for which the Fasciclin domain from *Drosophila* fasciclin I (PDB entry 1O70) is an example. Arrows and cylinders indicate β-strands and α-helices, respectively. The orange strand is absent in the four-stranded β-GF fold. Insertions in the basic β-GF fold are indicated by green arrows (β-strands) and magenta cylinders (α-helices).

The β-GF is thought to be already present in the last common organismal ancestor. In prokaryotes, the fold went through much structural diversification. Many extensions and modifications are found in numerous present prokaryotic β-GF domains, some with elaborate
modular insertions of β-sheets and helices such as in the Fasiciclin superfamily (Fig. 1.5B) (Burroughs, Iyer et al. 2012). The structural diversification enabled the prokaryotic β-GF to adopt a wide array of biological and biochemical niches such as iron-sulphur transfer (e.g. ThiS/MoaD) (Lake, Wuebbens et al. 2001), RNA recognition (e.g. threonyl-tRNA synthetase) (Sankaranarayanan, Dock-Bregeon et al. 1999), toxin function (e.g. Staphylococci enterotoxin A) (Schad, Zaitseva et al. 1995), and catalysis (e.g. fibrinolytic enzymes) (Rabijns, De Bondt et al. 1997). However, evolution of the β-GF in eukaryotes did not mirror that in prokaryotes, and instead of evolving new biochemical functions, expanded the β-GF members to impart function by mainly altering PPIs (Burroughs, Balaji et al. 2007). It is clear that the rich evolutionary history of the β-GF structure and hence function, demonstrates that this fold is a multifunctional scaffold that is highly versatile and robust to structural perturbations. As such, the Ub fold is an attractive candidate for an alternative scaffold, as discussed further in Chapter 3.

1.4.3 Ub Interactions

To date, two E1 enzymes (Schulman and Harper 2009), ~45 E2 conjugation enzymes (Valimberti, Tiberti et al. 2015), ~600 E3 ligases (Li, Bengtson et al. 2008), ~80 DUBs (Komander, Clague et al. 2009) and more than 20 structural families of UBDs have been identified in humans. Analysis of the ~250 solved Ub complexes revealed that Ub contacts are predominately found on the flat beta sheet surface (Fig. 1.5 and 1.6), with very rare contacts on the Ub backside encompassing segments α1 and β2 (Lange, Lakomek et al. 2008, Harrison, Jacobs et al. 2015). This observation corroborates with a functional study that showed Ub substitution on the beta sheet was deleterious for yeast growth (Roscoe, Thayer et al. 2013).

Remarkably, evolution has resulted in the dedication of one side of Ub for interactions to accommodate distinct structural binding features. For example, Ub interacting motifs (UIMs) and Ub-associated domains (UBAs) recognize Ub with single or double helices, respectively, while zinc fingers utilize loops for interaction. E2s, which all contain a UBC domain, recognize Ub with a mix of helices and loops, but in HECT E3s, a mix of helices and a distinct β-hairpin is used for Ub binding (Maspero, Valentini et al. 2013). In USPs, the Ub-binding site consists of three subdomains that form a hand-like structure (Fig. 1.7F).
Figure 1.6 The structure of human Ub

The left panel shows the β-sheet on the helix unobstructed side, which is the surface used predominately in Ub non-covalent interactions. The right panel shows a 180° rotation, and this “back” side rarely participates in interactions.

Figure 1.7 Interactions of Ub with partner proteins

Ub (blue) bound to partner proteins (gray): (A) UIM of RAP80 (PDB entry 3A1Q), (B) UBA of Dsk2p (PDB entry 1WR1), (C) Zinc finger of Npl4 (PDB entry 1Q5W), (D) UBC domain of UbcH5 (PDB entry 5FER), (E) HECT domain of NEDD4 (PDB entry 4BBN), (F) USP domain of USP2 (PDB entry 3V6C).
While the binding partners have been categorized according to their structural domains, many of them remain functionally elusive and only a fraction of these complexes have been solved up to date. To aid the elucidation of functions for Ub-associated proteins, our group developed a phage display strategy to rapidly develop Ub variants (Ubvs) that can act as tight binding reagents for specific Ub system proteins (Ernst, Avvakumov et al. 2013). This strategy emerged from the observations that Ub interactions predominately took place on the β-sheet and that the binding interfaces are not complementarily optimized. With a combinatorial Ub library with low levels of β-sheet substitutions, Ub variants (Ubvs) were selected with more optimal contacts to Ub system proteins, and hence the Ubvs had improved affinity and specificity over Ub.wt. Most Ubvs occluded natural Ub-binding sites and thus acted as inhibitors of protein function. Most interestingly, some Ubvs occupied natural Ub-binding regulatory sites on enzymes, and thus acted as modulators or activators inside cells, and both of these functions are not achievable with knockdown techniques (Zhang, Wu et al. 2016). With these tools in hand, Ub-binding protein function can be probed inside cells.

Although many structures have unveiled insights into Ub recognition at the molecular level, and this information has been successfully translated into engineering recombinant Ubvs for tools, many questions regarding Ub recognition remain unanswered. For example, within the large buried surface areas, what is the functional epitope for Ub interactions, and are there distinct functional epitopes for different proteins that utilize the same fold to bind Ub? How do the functional epitopes differ between different Ub binding features? How are the binding determinants spatially organized within the large binding surface? All these questions have important implications for drug design targeting specific enzymes and future development of computational protein design. Some of these questions will be addressed in Chapter 2.

1.5 Thesis Overview

A detailed understanding of the molecular underpinnings of how Ub recognizes enzymes in the Ub proteasome system is crucial for understanding Ub biological function. Many structures of Ub complexes have been solved and, in most cases, reveal a large structural epitope on a common face of the Ub molecule. However, due to the generally weak nature of these interactions, it has been difficult to map in detail the functional contributions of individual
Ub side chains to affinity and specificity. In chapter 2, I describe how I took advantage of Ubvs that bind tightly to two particular Ub specific proteases (USPs), USP2 and USP21, and employed saturation scanning mutagenesis by phage display to comprehensively map functional epitopes within the structural epitopes. I found that Ubvs that bind to USP2 or USP21 contain a remarkably similar core functional epitope, or hot spot, consisting mainly of positions that are conserved as the wt sequence but also some positions that preferred mutant sequences. The Ubv core functional epitope contacted residues that are conserved in the human USP family, and thus it is likely to be important for the interactions across many family members. In addition, despite the high identity between USP2 and USP21, we discovered distinct molecular preferences on each Ubv that enhanced binding to its cognate USP, and these molecular differences could be exploited for specific drug design.

Given the high adaptability of the β-GF in evolutionary history, the diverse interactions Ub can make and its unusual biochemical characteristics, we speculated that Ub could serve as a robust alternative scaffold for the development of affinity reagents. In chapter 3, I present my work on a novel combinatorial phage-displayed Ub library that is based on randomization of the β-sheet. In an attempt to reduce structural destabilization, prior to library construction, I performed a mutation tolerance scan to identify positions along the β-sheet that were important for stability, and these were fixed as wt in the final library design. In the end, I co-opted an expansive surface on the β-sheet for the construction of a binding site by sequence diversification. The resultant library was successful for generating binders to 30 of 43 diverse proteins. Validation data are presented for Ubvs selected to bind a cell surface receptor, human epidermal growth factor receptor (Her3), and Ubvs selected to bind the SH2 domain of the intracellular adaptor, growth factor receptor bound-2 (Grb2). My results show that Ubvs are specific to the proteins that they were raised against and have affinities that are comparable to antibodies. In particular, I show that Her3 binding Ubv can be used similarly to antibodies in immunoprecipitation experiments and that expression of Grb2 binding Ubv in mammalian cells has inhibitory effects on Grb2 mediated signaling.
2 Saturation Scanning of Ubiquitin Variants reveals a common hot spot for binding to USP2 and USP21

2.1 Statement of Contribution

This project was a collaboration with Dr. Julia Shifman from the Hebrew University of Jerusalem, and it began when she was a visiting scientist at the Sidhu lab. Dr. Shifman and I designed and constructed the phage libraries and performed selections. Computational scanning (Fig. 2.5) was done by Dr. Shifman and her student, Ayelet Dekel. Sequence analysis was done by me and Dr. Shifman. All other experiments were performed by me.


2.2 Introduction

Ub binds to thousands of proteins in the human proteome, but typically with low affinities in the $10^{-5}$-$10^{-3}$ M range. Structural studies have delineated that most interactions occur on the beta-sheet face (Lange, Lakomek et al. 2008, Harrison, Jacobs et al. 2015), which is thus highly versatile in recognizing a wide variety of domain folds and surface topologies. Despite low affinity, many Ub interactions bury substantial surface areas that can exceed 2000 Å$^2$ (Ernst, Avvakumov et al. 2013), defying the conventional view that buried surface area positively correlates with binding affinity of protein-protein interactions (Chen, Sawyer et al. 2013).

We have shown that the affinity of particular Ub-protein interactions can be enhanced by several orders of magnitude by mutating the surface to produce Ubvs that function as potent and specific inhibitors of the enzymes they target (Ernst, Avvakumov et al. 2013). Although this
work demonstrated that Ub interactions are not optimal and can be improved by mutations, we still do not understand which residues contribute the most to binding energetics and which residues are sub-optimal for binding. Many questions remain to be resolved. For example: what is the functional epitope, how are the binding determinants spatially organized within the large interface, and how do binding energetics differ depending on the interacting partner?

To understand molecular recognition by Ub, we need detailed molecular binding landscapes of Ub-protein interactions. Such binding landscapes can be obtained by introducing all possible single mutations in the binding site and assessing the consequent effects on affinities (Wells 1996, Wells and de Vos 1996, Pál, Kouadio et al. 2006, Moreira, Fernandes et al. 2007, Sharabi, Erijman et al. 2013, Aizner, Sharabi et al. 2014). However, the low affinities of typical Ub-protein interactions (Hurley, Lee et al. 2006) make it difficult to accurately measure the effects of mutations, and thus, detailed empirical binding landscapes of Ub-protein interactions have not been reported.

Here I used twoUbvs, which were evolved as high affinity inhibitors of different Ub specific proteases (USPs), as proxies to explore the functional details of Ub-protein interactions (Ernst, Avvakumov et al. 2013). USPs constitute the largest class of DUBs and several have been implicated as potential therapeutic targets (Yang, Kitagaki et al. 2009, Sippl, Collura et al. 2011). USPs share a structurally similar catalytic domain, which binds Ub and positions its C-terminal tail into the conserved catalytic cleft. I analyzed Ubv.2.1 and Ubv.21.4, distinct triple mutants with high affinity for USP2 or USP21, respectively (Ernst, Avvakumov et al. 2013). Comparisons of crystal structures of each Ubv or Ub.wt in complex with USP2 or USP21 have revealed that each Ubv binds its cognate enzyme in a manner that is virtually identical to the binding mode of Ub.wt (Ernst, Avvakumov et al. 2013). Thus, the fact that only three mutations yield Ubvs that bind in a similar manner but with dramatically enhanced affinities relative to Ub.wt makes these Ubvs ideal candidates for mutational studies to shed light on both the molecular bases for enhanced affinity and on the molecular details of natural Ub-USP interactions.

To obtain a comprehensive assessment of the mutational tolerance of the binding surface of each Ubv, without having to make and test thousands of point mutants, I employed combinatorial saturation scanning by phage display. This method uses combinatorial libraries of protein variants designed for the rapid and facile assessment of the effects of all possible point
mutations across a large protein binding surface and has been used previously to survey the binding energy landscape of human growth hormone (hGH) for its receptor (hGHR) (Pál, Kouadio et al. 2006). Saturation scanning analysis of the USP-binding sites of Ubv.2.1 and Ubv.21.4 reveals many common elements and some differences between the two interactions. Moreover, interpretation of the results in the context of the human USP family shows that many conserved functional interactions are likely generalizable to most family members. Our group and collaborators have reported structures of additional Ubvs in complex with E3 ligases of the Homologous-to-E6AP-C-terminus (HECT), Skp1-Cul1-F-box (SCF) and Really-Interesting-New-Gene (RING) families (Gorelik, Orlicky et al. 2016, Zhang, Wu et al. 2016), and the extension of the methods described here to these and other types of Ub-protein interactions may identify commonalities and differences in how Ub recognizes diverse structural folds to mediate biological effects.

2.3 Methods and Materials

2.3.1 Library construction and analysis

Each Ubv was displayed on M13 bacteriophage, as described (Ernst, Avvakumov et al. 2013). Phage-displayed Ubv libraries were constructed, sorted and analyzed, as described (Pál, Kouadio et al. 2006). Briefly, a Ubv-displaying phagemid vector containing TAA stop codons in the regions to be mutagenized was used as a template for oligonucleotide-directed mutagenesis (Kunkel 1985). Mutagenic oligonucleotides were designed to replace positions to be scanned with degenerate NNK codons (N = A/G/C/T, K = G/C) that collectively code for all 20 natural amino acids. Five libraries were designed for each Ubv, with each combining three mutagenic oligonucleotides to introduce the mutations into the Ubv sequence (Table 1). Each mutagenesis reaction was electroporated separately into E.coli SS320 and each yielded a library of >10⁹ unique members. The libraries were handled separately and each was sorted for four rounds to select for binding to immobilized USP. Individual clones from the fourth round of selection were grown in 96-well format, and the culture supernatants were used directly in phage ELISAs (Sidhu, Lowman et al. 2000) to detect clonal Ubv-phage that bound to USP. 50-100 positive binding clones from each library were subjected to DNA sequencing and statistical analysis.
<table>
<thead>
<tr>
<th>Ubv</th>
<th>Library</th>
<th>Region</th>
<th>Oligonucleotide Sequencesa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC GAT GAC AAA ATG (NNK) ATT TTC GTG AAC ACC (NNK) ACG GGG ACC (NNK) ATC ACC CTC GAG GTT GAA CCC</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>GGA ATT CCT CCT GAT CAG AGA CTG (NNK) TTG GCT GCC AAG CAG CTG GAA GAT GGA CGT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC TAC AAT ATT CAA AAG GAG TCT ACT CCT CAT CTG TTG (NNK) CTG GCT CAC GTG ACT GTG GGC</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>GAC GAT GAC AAA ATG CAG ATT (NNK) GTG AAC ACC CT T (NNK) GGG ACC CAC ATCC CTC GAG GTT GAA CCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGA ATT CCT CCT GAT CAG CAG (NNK) CTG ATC TT (NNK) GCC AAG CAG CTG GAA GAT GGA CGT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC TAC AAT ATT CAA AAG GAG TCT ACT CCT CAT CTG TTG AGA CT T (NNK) CAC GTG ACT GTG GGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC GAT GAC AAA ATG CAG ATT TT C GTG (NNK) ACC CT C TCG GGG ACC CAC ATC (NNK) CTC GAG GTT GAA CCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGA ATT CCT CCT GAT CAG AGA CTG ATC TT (NNK) GCT GCC AAG CAG CTG GAA GAT GGA CGT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC TAC AAT ATT CAA AAG (NNK) GTG TTG AGA CT T (NNK) GCT CAC GTG ACT GTG GGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC GAT GAC AAA ATG (NNK) ATT TTC GTG AAA ACC (NNK) ACG GGG AAG NNK ATC ACC CTC GAG GTT GAA CCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>same as Ubv2.1 Library 1 Region2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC TAC AAT ATT CAA AAG TTG TCT ACT CCT CAT CTG TTG (NNK) CTG GCT CAC GTG ACT GTG GGC</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>GAC GAT GAC AAA ATG CAG ATT (NNK) GTG AAA ACC CTT (NNK) GGG AAG ACC ATC ACC CTC GAG GTT GAA CCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>same as Ubv2.1 Library 2 Region2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC TAC AAT ATT CAA AAG TTG TCT ACT CCT CAT CTG TTG AGA CT T (NNK) CAC GTG ACT GTG GGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC GAT GAC AAA ATG CAG ATT TT C GTG (NNK) ACC CT C TCG GGG ACC CAC ATC (NNK) CTC GAG GTT GAA CCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>same as Ubv2.1 Library 3 Region2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC TAC AAT ATT CAA (NNK) TTG TCT ACT C TT CAT CTG TTG AGA (NNK) CTG CAC GTG ACT GTG GGC</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>GAC GAT GAC AAA ATG CAG ATT TT C GTG AAA ACC CTT AG (NNK) AAG ACC ATC ACC CTC GAG GTT GAA CCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>same as Ubv2.1 Library 4 Region2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC TAC AAT ATT CAA AAG TGG TCT (NNK) CT T TT CCT (NNK) TTG AGA CT T (NNK) CAC GTG ACT GTG GGC</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>GAC GAT GAC AAA ATG CAG ATT TT C GTG AAA ACC CTT AG (NNK) AAG ACC ATC ACC CTC GAG GTT GAA CCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>same as Ubv2.1 Library 5 Region2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAC TAC AAT ATT CAA AAG (NNK) TCT ACT CTT (NNK) CTT CTG (NNK) AGA CT T (NNK) CAC GTG ACT GTG GGC</td>
</tr>
</tbody>
</table>

*aCodons enclosed in brackets indicate randomized positions. “N” indicate equimolar mix of all four nucleotides. “K” indicate equimolar mix of G and T nucleotides.
2.3.2 Statistical Analysis.

The binding sequences from each library were aligned. The occurrence of each amino acid at each scanned position was corrected for bias by dividing the counts by the number of codons for that amino acid contained within the NNK degenerate codon (Bond, Wiesmann et al. 2005). Normalized sequences were used to produce an alignment in LOGOS consensus format (http://weblogo.berkeley.edu/logo.cgi) (Schneider and Stephens 1990). Bits scores at position \( i \) were calculated according to the formula: 

\[
R_i = \log_2 20 - (H_i - e_n),
\]

where \( H_i \), Shannon Entropy at position \( i \), \( = -\sum f_{a,i} \log_2 f_{a,i} \), and \( f_{a,i} \) is the relative frequency of amino acid, \( a \), at position \( i \). Small sample correction, \( e_n \), was assigned as 0, so that bits scores ranged from 0 to 4.32 (\( \log_2 20 \)).

The statistical analysis of the Ub-binding site of USPs was performed similarly, using an alignment of the catalytic domains of 48 human USPs (Fig. 2.6).

2.3.3 Enzyme activity assays

Enzymatic assays were performed in assay buffer (50 mM HEPES, pH 7.5, 0.01% Tween 20, 10 mM DTT). Serial dilutions were performed in 96-well plates and then transferred to 384-well plates for measurements. USP2 inhibition assays were performed by measuring the release of fluorogenic amido-4-methylcoumarin (AMC) from the substrate, Ub-AMC (Boston Biochem, Boston, MA), as described (Ernst, Avvakumov et al. 2013), with 1 mM Ub-AMC substrate, 7.5 nM USP2 and serial dilutions of Ubv. USP2 and Ubv were mixed in assay buffer and incubated at room temperature for 2 min prior to the addition of Ub-AMC. Proteolytic activity was measured by monitoring the increase of fluorescence emission at 460 nm (excitation at 360 nm) for 30 min using a Synergy2 plate reader (BioTek Instruments, Winooski, VT). IC\(_{50}\) was defined as the concentration of Ubv that inhibits 50% of USP2 activity, and was fitted using the sigmoidal 4PL equation in GraphPad Prism software.

For determination of kinetic parameters, a fixed concentration of USP2 was incubated with serial dilutions of Ub-AMC. Duplicates of initial reaction velocity (nM/s) were determined at each substrate concentration by determining the linear slope from plotting fluorescence signal versus time and converting to molarity by interpolating from a standard curve of known AMC concentrations. Velocity versus substrate concentration was plotted to determine the \( K_M \) and \( V_{max} \) using Graphpad Prism with the Michaelis-Menten equation 

\[
V_0 = V_{max} \times [S] / (K_M + [S]).
\]

\( k_{cat} \) was
obtained from the equation $k_{\text{cat}} = \frac{V_{\text{max}}}{[E]_0}$, where $[E]_0$ is the total enzyme concentration. For velocity versus substrate plots that could not be fitted using the Michaelis Menten equation but could be fitted using a linear line, I determined an estimate of the $k_{\text{cat}}/K_M$ value by dividing the linear slope (nMs$^{-1}$M$^{-1}$) by $[E]_0$.

### 2.3.4 Computational saturation mutagenesis

Computational scanning of Ubv.2.1 or Ubv.21.4 in complex with USP2 or USP21 (PDB file 3V6E or 3MTN), respectively, was performed as described (21). Each substitution was introduced into the Ubv, and the intermolecular energy $D_{\text{E}_{\text{inter}}}$ was calculated by subtracting the energies of the unbound chains from that of the complex. $DDG_{\text{bind}}$ was calculated as the difference between $D_{\text{E}_{\text{inter}}}$ of the Ubv and Ub.wt complexes. Backrub algorithm was used to model backbone flexibility including movements of -15, -10, -5, +5, +10, +15 degrees relative to the original conformation. The lowest $DDG_{\text{bind}}$ from all backrub moves was assigned to each substitution, and a substitution was considered allowed if it exhibited a negative $DDG_{\text{bind}}$ value. The position was categorized as high, medium or low conservation if <4, 4-7, or >7 substitutions were allowed, respectively.

### 2.4 Results

#### 2.4.1 Saturation Scanning of Ubv.2.1 and Ubv.21.4

In saturation scanning, precisely designed combinatorial libraries of phage-displayed protein variants are used to assess the functional importance of individual residues for binding to a partner protein (Clackson and Wells 1995). The basic principle of the method is that the frequency of a particular amino acid at a particular position within a protein interaction site reflects how favourable it is for the function of the site. By allowing all 20 genetically encoded amino acids to compete against each other, the relative frequencies can be used to simultaneously compare the effects of all possible substitutions across the interface.

I defined the USP-binding site on Ub as 25 residues across the β-sheet face, and for each Ubv, I divided these residues into five groups of five for randomization in phage-displayed libraries (Fig. 2.1). To minimize cooperative effects between mutated positions, each library was designed to randomize positions that are not in direct contact with each other (Pál, Kouadio et
al. 2006). Ubv.2.1 or Ubv.21.4 libraries, displayed on M13 bacteriophage as fusions to the N-terminus of the gene-3 minor coat protein, were subjected to selections for binding to the catalytic domain of USP2 or USP21, respectively. Individual clones were tested for cognate USP binding by phage enzyme-linked immunosorbent assays (ELISAs) and 50-100 clones with specific binding signals were sequenced from each library. Binding sequences were used to create a position weight matrix (PWM) consisting of frequencies of each amino acid type at each position (Fig. 2.2). Each matrix column depicts the amino acid preference at each position as a probability distribution. From this PWM, the distribution at each position was visualized as a sequence logo (Fig. 2.3). The importance of each Ubv position to USP binding was quantified as a bits score plotted along the y-axis of the sequence logo (Fig. 2.3). The bits score measures the amount of information that is required to describe the amino acid frequency pattern at a particular position and is related to Shannon Entropy, which measures the degree of randomness within a population (Schneider and Stephens 1990) and is routinely used to identify and predict DNA (D’Haeseleer 2006, Badis, Berger et al. 2009), RNA (Ray, Kazan et al. 2013) and protein binding motifs (Tonikian, Xin et al. 2009, Gfeller, Butty et al. 2011, Andreatta and Nielsen 2012). For a frequency distribution of 20 amino acids, the bits score varies between zero and 4.32 for positions that are completely random or completely conserved, respectively.

**Figure 2.1 Ubv saturation scanning library designs**

(A) The USP-binding site mapped onto the structure of Ub (PDB entry 1UBQ). The site contains 25 residues distributed across three regions: region 1 (residues 2-14), region 2 (residues 40-49) and region 3 (residues 63-74). Five libraries were designed to scan five residues each, in a manner that maximized the distances between residues in any one library. Scanned residues are shown as spheres and are color-coded according to the library they share. (B) U bv sequences subjected to saturation scanning. Dashes (-) indicate residues that are identical to Ub.wt and positions that differ are indicated by the amino acid substitution. Five libraries were designed for each U bv, and the randomized positions in each library are denoted by (X).
Figure 2.2 The saturation scan database

Data shown for (A) Ubv.2.1 and (B) Ubv.21.4. Following selection for binding to cognate USP, the percent occurrence of each amino acid was calculated after normalization for codon bias at each scanned position. The wt occurrences are boxed and values are colored as follows: yellow ≥10%, blue ≤2%. Bits score were calculated as described in “Materials and Methods.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Bits</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2</td>
<td>F4</td>
<td>K6</td>
<td>L8</td>
</tr>
<tr>
<td>T9</td>
<td>G10</td>
<td>T11</td>
<td>T12</td>
</tr>
<tr>
<td>T14</td>
<td>Q40</td>
<td>R42</td>
<td>Q49</td>
</tr>
<tr>
<td>K46</td>
<td>G47</td>
<td>I44</td>
<td>G48</td>
</tr>
<tr>
<td>I45</td>
<td>K48</td>
<td>Q49</td>
<td>L51</td>
</tr>
<tr>
<td>P</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>L</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Y</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>W</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>N</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Q</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>K</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>R</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Figure 2.3 Saturation scans of Ubvs

(A and B) The results of saturation scanning mapped on the structure of (A) Ubv.2.1 (PDB entry 3V6C) or (B) Ubv.21.4 (PDB entry 3MTN). The Ubv is shown as a surface colored according to bit values derived from amino acid frequencies (see Fig. S2), as follows: green $\leq 2$, $2 < yellow < 2.7$, red $\geq 2.7$, grey = unscanned. The white dashed lines demarcate the core functional epitope. (C and D) The results of saturation scanning presented as sequence logos for (C) Ubv.2.1 or (D) Ubv.21.4. The sequence of each Ubv is shown at the bottom and boxes signify sequences that differ from Ub.wt. Residues marked by blue or red boxes indicate sequence changes in Ubv.2.1 predicted to enhance inhibition of USP2 for which enzyme inhibition assays did or did not validate predictions, respectively (see Fig. 2.4).

To assess the accuracy of the saturation scanning data, I focused on inhibitors of USP2 and measured the effects of Ub.wt and a panel of Ubvs on the hydrolysis of the model substrate Ub-
AMC (Fig. 2.4). Ub.wt exhibited an IC\(_{50}\) value of ~60 μM and Ubv.2.1 (K6N/K11T/T12H) was ~1000-fold more potent (IC\(_{50}\) = 56 nM). Amongst the three mutated positions in Ubv.2.1, we noted that position 11 exhibited a low bits score and that Thr was relatively uncommon (Fig 1D). Thus, we predicted that Thr\(^{11}\) does not contribute significantly to the enhanced affinity of Ubv.2.1 for USP2 and confirmed this by showing that a variant containing two substitutions (K6N/T12H) was virtually equipotent (IC\(_{50}\) = 56 nM) to Ubv.2.1. I assessed the effects of individual substitutions at positions 6 and 12 and found that, relative to Ub.wt, the substitution Asn\(^6\) or His\(^{12}\) enhanced potency by ~300-fold (IC\(_{50}\) = 200 nM) or ~2-fold (IC\(_{50}\) = 24 μM), respectively. Taken together, these results show that the two mutations K6N/T12H are necessary and sufficient to confer high potency inhibition of USP2.

**Figure 2.4 Dose response curves for inhibition of USP2 by Ubvs**

The IC\(_{50}\) value was determined as the concentration of UBV that reduced USP2 activity by 50% (n = 2).

Next, in the background of the double-mutant K6N/T12H, I tested the effects of substitutions that were abundant in the selected sequences (Fig. 2.4). Substitutions at seven positions (Q2D, N6M, K11R, T14L, E64F, H68F, V70N) either enhanced or did not affect inhibitor potency, confirming the predictions of the saturation scan analysis. However, substitutions at two positions (R42G, L73W) reduced inhibitor potency, and we speculate that these substitutions may have been selected due to favorable effects on protein stability or levels of protein display on phage, factors that were not corrected for in the analysis. Taken together, these results generally confirm the accuracy of predictions derived from the saturation scan data, but the
method was inaccurate for two of nine positions at which substitutions to the Ubv.2.1 sequence were predicted to enhance binding to USP2.

To visualize the saturation scanning data in a structural context, I assigned the positions on the basis of bits scores into high, moderate or low conservation groups and mapped the data onto the structure of Ubv.2.1 (Fig. 2.3A). The high conservation positions form a large, contiguous functional epitope including the C-terminal tail region and a band of residues across the center of the interface. Strikingly, the functional epitope of Ubv.21.4 (Fig. 2.3B) is very similar to that of Ubv.2.1, as both exhibit high sequence conservation at nine positions across the central band (4, 6, 10, 12, 42, 44, 49, 66, 68), which I henceforth refer to as the core functional epitope.

To further validate and understand the saturation scanning data, we used an in silico saturation mutagenesis protocol (Sharabi, Erijman et al. 2013, Aizner, Sharabi et al. 2014) to predict tolerance of each Ub position to mutation. In this protocol, we substituted all Ub residues randomized in this study with the nineteen other amino acids and calculated changes in free energy of Ub binding ($\Delta G_{\text{bind}}$) to USP2 or USP21 (Fig. 2.5 C and D). The positions were then classified according to the number of substitutions that produced negative $\Delta G_{\text{bind}}$ values. Our computational results largely confirm the conservation patterns observed by saturation scanning, producing a similar Ub functional epitope for binding to USP2 or USP21 (Fig. 2.5A and B). A few Ub positions exhibited slightly higher variability in the computational analysis relative to the empirical analysis, and this may be due to the inability to model certain experimental phage selection pressures such as higher stability and expression.
Figure 2.5 Computational saturation scanning of Ubvs

Computational saturation scanning data mapped onto the structures of (A) Ubv.2.1 and (B) Ubv.21.4. Residues shaded in red indicate positions that tolerate less than or equal to 3 substitutions, yellow indicates positions that tolerate between 4-7 substitutions and green indicate positions that tolerate 8 or more substitutions. (C and D) Database of \( \Delta \Delta G_{bind} \) values calculated from in silico substitution of the 19 non-wt amino acids at each indicated position for
(C) Ubv.2.1 binding to USP2 and (D) Ubv21.4 binding to USP21. The wt $\Delta \Delta G_{\text{bind}}$ values are boxed, and negative $\Delta \Delta G_{\text{bind}}$ values are highlighted in red. Dash indicates calculations that were not completed most probably due to largely unfavorable energy upon introduction of the particular mutation.

### 2.4.2 The Ub-binding site of USPs

I next examined the interactions of Ubv.2.1 and Ubv.21.4 with their cognate USPs. I examined the structure of the USP2:Ubv.2.1 complex and defined the Ub-binding site on USP2 as 52 residues that are within 4.5 Å of Ubv.2.1 (Fig. 2.6). In a manner analogous to the analysis of the saturation scan data for the Ubvs, an alignment of these 52 positions across 48 human USPs (Fig. 2.6), and the data were used to calculate bits scores to quantify sequence conservation within the human USP family. Based on the same cut-offs used for the saturation scan analysis of the Ubvs, the Ub-binding site contains 18 residues that exhibit high conservation and 12 that exhibit moderate conservation, and are shown as a sequence logo (Fig. 2.7B). The sequence conservation data were visualized by mapping the bits scores onto the USP2 structure (Fig. 2.7A).
### Figure 2.6 Sequence alignment of the Ub-binding sites of the human USP family.

The Ub-binding site was defined as the 52 residues in USP2 that are within 4.5 Å of Ubp2.1 in the structure of the USP2:Ubp2.1 complex (PDB entry 3V6C), and the alignment for these residues is shown for 48 human USPs. Residues are shaded based on bits score: green ≤2 (non-conserved), 2< yellow <2.7 (conserved), red ≥ 2.7 (highly conserved).
Figure 2.7 Sequence conservation across the Ub-binding sites of human USPs.

(A) The structure of USP2 in complex with Ubv.2.1 (PDB entry 3V6C). USP2 is shown as a grey surface with residues in the Ub-binding site (residues within 4.5 Å of any residue on Ubv.2.1) colored according to bits scores derived from the alignment of 48 human USPs, as follows: green \( \leq 2 \), yellow \( 2 < \text{bits} < 2.7 \), red \( \geq 2.7 \). The main-chain of Ubv.2.1 is shown as a wheat tube and nine residues that show high conservation in the saturation scan of both Ubv.2.1 and Ubv.21.4 (core functional epitope) are shown as ruby spheres. White or black dashed lines demarcate site-1 or site-2, respectively. The structure on the right is Ubv.2.1 rotated 180° relative to Ubv.2.1 in complex. (B) The sequence logo for highly conserved (bits \( \geq 2.7 \)) and moderately conserved positions (2 < bits < 2.7) within the Ub-binding sites of human USPs. Positions are numbered according to the sequence of USP2, and sequences of USP2 and USP21 are shown. Asterisks (*) or arrowheads (^) indicate residues that interact with the Ub C-terminal tail or the core functional epitope, respectively.

Viewed in a structural context, the conserved residues form two sites that interact with Ubv.2.1. Eight highly conserved residues and one moderately conserved residue form the active site cleft (site-1) that interacts with the C-terminal tail of Ubv.2.1 (residues 70-74). Ten of the remaining highly conserved residues and the remaining 11 moderately conserved residues form
a second site (site-2) adjacent to site-1 (Fig. 2.7). Site-2 interacts with the flat beta-sheet face of Ubv.2.1, and notably, 14 of the residues in site-2 interact with the nine residues that form the core functional epitope shared by Ubv.2.1 and Ubv.21.4.

I speculated that the USP2 site-2 revealed to be important for binding to Ubv.2.1 may also be involved in the recognition of native Ub during catalysis. I explored this possibility using alanine-scanning mutagenesis (Morrison and Weiss 2001) by constructing a panel of USP2 variants containing single-site alanine substitutions at each of the 14 conserved site-2 residues that make contact with the core functional epitope of Ubv.2.1 and measuring their effects on the catalytic efficiency for the hydrolysis of Ub-AMC substrate. These assays showed that 10 of the 14 alanine substitutions reduced catalytic efficiency by at least two-fold relative to wt USP2 (Fig. 2.8), thus verifying that USP2 uses site-2 for the recognition of Ub substrates.
### Table 2.8

<table>
<thead>
<tr>
<th>USP2</th>
<th>$k_{\text{cat}}$ (s$^{-1}$)</th>
<th>$K_M$ (M)</th>
<th>$k_{\text{cat}}/K_M$ (M$^{-1}$ s$^{-1}$)</th>
<th>Fold Effect**</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>(1.4 ± 0.1) x 10$^{-1}$</td>
<td>(1.9 ± 0.3) x 10$^{-6}$</td>
<td>(8.0 ± 1) x 10$^4$</td>
<td>1.0</td>
</tr>
<tr>
<td>F462A</td>
<td>(2.9 ± 0.2) x 10$^{-1}$</td>
<td>(7.0 ± 0.7) x 10$^{-6}$</td>
<td>(4.1 ± 0.4) x 10$^4$</td>
<td>2.0</td>
</tr>
<tr>
<td>D466A</td>
<td>(2.0 ± 0.1) x 10$^{-1}$</td>
<td>(6.0 ± 0.5) x 10$^{-6}$</td>
<td>(3.3 ± 0.3) x 10$^4$</td>
<td>2.4</td>
</tr>
<tr>
<td>P474A</td>
<td>(6.0 ± 0.2) x 10$^{-2}$</td>
<td>(2.5 ± 0.3) x 10$^{-6}$</td>
<td>(2.4 ± 0.3) x 10$^4$</td>
<td>3.3</td>
</tr>
<tr>
<td>K503A</td>
<td>(5.7 ±0.2) x 10$^{-1}$</td>
<td>(5.1 ± 0.5) x 10$^{-6}$</td>
<td>(1.2 ± 0.1) x 10$^5$</td>
<td>0.7</td>
</tr>
<tr>
<td>Y592A</td>
<td>(4.8 ± 0.4) x 10$^{-1}$</td>
<td>(1.3 ± 0.2) x 10$^{-5}$</td>
<td>(3.9 ± 0.7) x 10$^4$</td>
<td>2.0</td>
</tr>
<tr>
<td>D357A*</td>
<td></td>
<td>(5.9 ± 0.2) x 10$^2$</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>Q359A*</td>
<td></td>
<td>(3.7 ± 0.1) x 10$^4$</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>E360A*</td>
<td></td>
<td>(1.3 ± 0.1) x 10$^3$</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>L419A*</td>
<td></td>
<td>(7.3 ± 0.2) x 10$^4$</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>S421A*</td>
<td></td>
<td>(1.2 ± 0.1) x 10$^4$</td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td>D436A*</td>
<td></td>
<td>(9.7 ± 0.5) x 10$^3$</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td>L468A*</td>
<td></td>
<td>(3 ± 0.1) x 10$^3$</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>K487A*</td>
<td></td>
<td>(2.7 ± 0.1) x 10$^4$</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>H501A*</td>
<td></td>
<td>(2.9 ± 0.1) x 10$^4$</td>
<td>2.8</td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 2.8 Alanine-scanning of USP2 site-2**

(A) Kinetic values for Ub-AMC cleavage by USP2 variants. Asterisk (*) Only $k_{\text{cat}}/K_M$ values could be determined. See “Materials and Methods” for details. Double asterisks (**) Fold effect was calculated by dividing $k_{\text{cat}}/K_M$ of wt USP2 by the $k_{\text{cat}}/K_M$ of the variant. (B) Alanine-scanning data mapped on to the structure of USP2 bound to Ubv.2.1 (PDB entry: 3V6C). USP2 is shown as a surface colored according to fold reduction in catalytic efficiency relative to wt caused by alanine substitutions at the labeled positions, as follows: *green* < 2, *yellow* < 10, *red* > 10, *grey* = unscanned. The Ubv.2.1 main chain is shown as a tube and residues in the core functional epitope are shown as numbered spheres.
2.4.3 Interactions at the USP:Ubv interface

Saturation scanning enabled elucidation of the core functional epitope of the two Ubvs for binding to their cognate USPs and examination of the USP:Ubv complex structures identified the USP residues interacting with this epitope. Notably, the core functional epitopes of Ubv.2.1 and Ubv.21.4 differ from the sequence of Ub.wt at only two or one positions, respectively, suggesting that each Ubv likely interacts with its cognate USP in a manner very similar to that of Ub.wt. Indeed, comparison of the USP2:Ubv.2.1 and USP2:Ub.wt (Fig. 2.9A and B, respectively) or the USP21:Ubv.21.4 and USP21:Ub.wt complex structures (Fig. 2.9 C and D, respectively) revealed excellent superposition of the C-α atoms, with a root-mean-square deviation (rmsd) less than 0.28 Å or 0.32 Å, respectively. Moreover, there was excellent superposition of the six or seven side-chains of the core functional epitope that are identical between Ub.wt and Ubv.2.1 or Ubv.21.4, respectively (rmsd less than 0.13 Å or 0.20 Å, respectively) and for the 19 USP2 side-chains or 18 USP21 side-chains that make contact with any of the eight side-chains of the Ubv.2.1 or Ubv.21.4 core functional epitope, respectively (rmsd less than 0.18 Å or 0.36 Å, respectively).
Figure 2.9 Superposition of USPs in complex with Ub.wt or Ubvs.

Ribbon representations of superposed (A) USP2 from USP2:Ub.wt (PDB entry 2HD5) and USP2 from USP2:Ubv2.1 (PDB entry 3V6C), (B) Ub.wt from USP2:Ub.wt (PDB entry 2HD5) and Ubv2.1 from USP2:Ubv2.1 (PDB entry 3V6C), (C) USP21 from USP21:Ub.wt (PDB entry 3I3T) and USP21 from USP21:Ubv21.4 (PDB entry 3MTN), and (D) Ub.wt from USP21:Ub.wt (PDB entry 3I3T) and Ubv21.4 from USP21:Ubv21.4 (PDB entry 3MTN). USP and Ub.wt in complex are colored green or light green, respectively. USP and Ubv in complex are colored purple or pink, respectively. Side-chains are shown as sticks for the Ubv core functional epitope and for USP side-chains that contact the core functional epitope (USP side-chains that are within 4.5 Å of side-chains in the Ubv core functional epitope). For residues that differ between Ub.wt and Ubv, the Ubv sequence is shown after the residue number. Both USP2 and USP21 are numbered according to the sequence of USP2, based on a sequence alignment (Fig. 2.6).

I next compared the USP2:Ubv.2.1 and USP21:Ubv.21.4 complex structures to understand how the core functional epitopes on the Ubvs interact with USPs (Fig. 2.10A). Superposition of the two complexes showed the overall structures to be highly similar (rmsd less than 0.55 Å for C-α atoms). Notably, six of nine residues in the core functional epitopes of the Ubvs are identical and 12 of 19 USP residues that contact this epitope are identical. Thus, both in terms of the overall fold and in terms of side-chain interactions, the two Ubvs bind to their cognate USPs in a very similar manner.
Figure 2.10 Interactions between the Ubv core functional epitopes and USPs.

(A) Superposition of USP2 (purple, pink and magenta) and USP21 (lime, sand and yellow) in complex with Ubv.2.1 (green) or Ubv.21.4 (blue), respectively. The USP main-chains are shown as ribbons, the catalytic Cys is shown as a purple sphere and side-chains are shown as sticks for residues that contact the Ubv core functional epitope (within 4.5 Å). For clarity, the Ubv main-chains are not shown and only the side-chains of the core functional epitopes are shown. (B, C, and D) Interactions between USP side-chains and Ubv side-chains of (B) patch A residues 42, 44 and 49, (C) non-wt patch B residues 6, 12 and 68, and (D) wt patch B residues 4 and 66. USP residues are labeled in italics and numbered according to the sequence of USP2. For residues that differ between USP2 and USP21, the USP21 sequence is shown after the residue number. Ubv residues are labeled in bold, and for residues that differ between Ubv.2.1 and Ubv.21.4, the Ubv.21.4 sequence is shown after the residue number.

The USP fold is comprised of three sub-domains, likened to a palm, thumb and fingers, with the catalytic site located between the palm and the thumb (Hu, Li et al. 2002). The eight side-chains in the Ubv core functional epitope form two patches, and one patch (patch A) docks on to the palm while the other (patch B) docks on to the fingers (Fig. 2.10A). Patch A contains three residues (Arg42, Ile44, Gln49) and these are all conserved as the Ub.wt sequence in both Ubvs (Fig. 2.10B). Patch B contains five residues, which can in turn be grouped into two clusters, with one cluster containing two residues (Phe4, Thr66) that are conserved as the Ub.wt
sequence in both Ubvs (Fig. 2.10D) and the second cluster containing three positions (6, 12, 68) that differ between the Ubvs and show preference for particular non-wt sequences in the saturation scans (Fig. 2.10C).

Each patch is anchored by a highly conserved hydrophobic residue: Ile44 or Phe4 in patch A or B, respectively. In patch A of Ubv.2.1, the hydrophobic side-chain of Ile44^{2.1} resides in a predominantly hydrophilic environment formed by the side-chains of Gln359^{USP2} and Arg363^{USP2}, which are either identical (position 359) or conservatively substituted (position 363) in USP21 (Fig. 2.10B). The other two residues in patch A that are conserved as the Ub.wt sequence (Arg42 and Gln49) also reside in a hydrophilic environment that is conserved in the two USPs. In particular, the Arg42^{2.1} side-chain forms hydrogen bonds with the side-chains of Gln359^{USP2} and Gln49^{2.1} and makes favorable electrostatic interactions with the side-chain of Asp357^{USP2}, and the Gln49^{2.1} side-chain also forms a hydrogen bond with the Glu360^{USP2} side-chain. Notably, all of these side-chains are identical in USP2 and USP21 and are highly conserved within the USP family (Fig. 2.6 and 2.7B), and consequently, the interactions between Ubv.21.4 and USP21 are virtually identical in this region of the interface. In patch B of Ubv.2.1, the Phe4^{2.1} side-chain interacts with Pro474^{USP2}, the side-chains of Leu423^{USP2} and Leu468^{USP2}, and the aliphatic portion of the Lys487^{USP2} side-chain. The amino group of the Lys487^{USP2} side-chain makes a hydrogen bond with the side-chain of Thr66^{2.1}, the second residue in patch B that is conserved as the Ub.wt sequence (Fig. 2.10D). In addition, the aliphatic portion of the Thr66^{2.1} side-chain packs against the side-chains of Asp436^{USP2} and Phe489^{USP2}. Notably, aside from positions 436 and 489 that have conservative substitutions, these residues are identical and make very similar interactions in the USP21:Ubv.21.4 complex structure.

Between the conserved hydrophobic anchor residues (Ile44 and Phe4) and the associated conserved hydrophilic residues (Arg42, Gln49, Thr66), resides a cluster of three residues (6, 12, 68) that differ between Ubv.2.1 and Ubv.21.4 and tend to vary from the sequence of Ub.wt (Fig. 2.10C). At position 68, Ubv.2.1 contains a wt His residue while Ubv.21.4 contains a Phe substitution, but the saturation scan shows that Ubv.2.1 also prefers a Phe at this position (Fig. 2.3C). Consistent with a preference for a hydrophobic Phe, the side-chains at position 68 of both Ubvs interact with three hydrophobic residues in their cognate USPs that are either identical (positions 419 and 462) or conserved (position 489). At position 6 of Ubv.2.1, the wt Lys
residue is replaced by an Asn and Met is also highly prevalent in the saturation scan dataset (Fig. 2.3C). The Asn6^{2.1} side-chain forms a hydrogen bond with the side-chain of Lys503^{USP2}. The side-chain at position 6 is close to the hydrophobic side-chains of Phe462^{USP2} and Phe489^{USP2}, and this likely explains why a hydrophobic Met can effectively substitute for Asn at this position (Fig. 2.4). Notably, although Ubv.21.4 contains a wt Lys at position 6, the saturation scan also shows a preference for Asn, which could be explained by a putative hydrogen bond with Gln503^{USP21} (Fig. 2.10B). At position 12 of Ubv.2.1, His dominates over the wt Thr and all other substitutions in the saturation scan (Fig. 2.10C), suggesting that this side-chain contributes to recognition of USP2. This preference can be explained by a hydrogen bond between the His12^{2.1} side-chain and the side-chain of Asp466^{USP2}. In contrast, in the saturation scan data for Ubv.21.4, His is completely absent at position 12, which instead prefers the wt Thr and also tolerates Ser and Tyr. In this case, the structure shows that a longer Glu466^{USP21} side-chain makes a hydrogen bond with Thr12^{21.4}, and it is likely that a larger His residue at position 12 could not make the same favorable interactions.

2.5 Discussion

By applying phage display and saturation scanning analyses, we have obtained detailed views of how Ubv.2.1 and Ubv.21.4 bind with high affinity to their respective targets USP2 and USP21. Moreover, the availability of structures of each USP in complex with either its cognate Ubv or Ub.wt enabled us to extend our insights to the interactions between USPs and native Ub. Our structural comparison shows remarkable similarity at both the main-chain and side-chain level between key residues on both sides of the interface (Fig. 3), suggesting that our conclusions regarding the Ubvs are likely also applicable to native Ub.

Overall, the saturation scans reveal that a large number of residues in each Ubv exhibit significant sequence conservation, and thus, a large surface area at the binding interface is recruited for productive binding contacts (Fig. 1). Notably, nine positions exhibit high conservation in both Ubvs, and this core functional epitope is involved in interactions with many other Ub-binding proteins (Roscoe, Thayer et al. 2013, Harrison, Jacobs et al. 2015). Moreover, six of the nine positions (4, 6, 10, 42, 44, 68) have been shown to be highly intolerant
to substitutions in a study that probed the effects of point mutations on the biological function of Ub in yeast (Roscoe, Thayer et al. 2013).

Interestingly, the core functional epitope docks on a region of the USP fold that is conserved within the family and is distinct from the active site (site-2, Fig. 2.7). Six of the nine residues within the core functional epitope are conserved as the wt sequence and interact with USP residues that are also highly conserved within the USP family, suggesting that these interactions may make similar energetic contributions to the recognition of Ub substrates by many USPs (Fig. 2.6 and 2.7). The importance of site-2 for the recognition of Ub substrates was confirmed by alanine-scanning mutagenesis of USP2, which showed that alanine substitutions at many positions in site-2 reduced catalytic efficiency for hydrolysis of Ub-AMC (Fig. 2.8). The remaining three residues are also highly conserved but favor non-wt sequences. Notably, these three residues cluster together (Fig. 2.10C) and include residues that were mutated in the Ubvs to gain high affinity and specificity for their cognate USPs, and thus, they represent a critical cluster that can be exploited for specific inhibitor design.

Taken together, our results show that site-2 is important for the recognition of Ubvs and Ub substrates, and thus, this region may be an attractive alternative target for the design of small molecules that inhibit USP function with higher specificity than has been possible with inhibitors targeting the highly conserved active site (Colland, Formstecher et al. 2009, Colland 2010, Pal, Young et al. 2014). Indeed, site-2 on the USPs and the core functional epitope on the Ubvs conform to the hot spot model of protein-protein interactions, in which a subset of side-chains on each side of an interface form contiguous patches that interact with each other and contribute a large fraction of the binding energy (Clackson and Wells 1995, Wells 1996, Moreira, Fernandes et al. 2007). While inhibition of protein-protein interactions with small molecules remains challenging, significant progress has been made in targeting sites other than classical active site clefts (Arkin and Wells 2004, Arkin, Tang et al. 2014, Guo, Wisniewski et al. 2014).

Our phage-displayed technology for developing high affinity Ubvs has proven successful against not only USPs, but also against other DUB families, E3 ligases, E2 conjugating enzymes and Ub-binding domains (Ernst, Avvakumov et al. 2013). Indeed, we have recently reported structures of Ubvs that target E3 ligases of the HECT (Zhang, Wu et al. 2016), SCF (Gorelik, Orlicky et al. 2016) and RING families (Brown, VanderLinden et al. 2016).
Thus, the methods described here can be applied to dissect the functional epitopes for Ubvs targeting many other components of the Ub proteasome system. These studies should prove useful for guiding new inhibitor design strategies, and also, for understanding the molecular basis for the myriad of protein interactions that mediate Ub biology.
3 Ubiquitin as an alternative scaffold for the development of affinity reagents

3.1 Statement of contribution

I constructed all libraries described in this chapter. I performed all selections, except for Fab library selections on the following targets: Grb2, Her3, ROR-alpha, MBP, and BATF, and selections on targets: ASH1L, BRD1, BRD2, BRD3, and BRD4.

HEK293T cell lines stably overexpressing ErbB family member(s) were engineered by Dr. Nish Patel, a former post-doc in the Sidhu/Moffat lab.

All surface plasmon resonance (SPR) experiments were designed by Nick Jarvik and myself and performed by Nick Jarvik.

3.2 Introduction

Affinity reagents, such as Abs, are powerful binding tools that are used to identify and detect the presence of virtually any analyte in a sample. As such, they are the cornerstone of biological research and biomedicine. The demand for affinity reagents has accelerated over the past decade due to the expansion of proteomic approaches aimed at uncovering the diverse functions of gene products in biology, health and disease.

Although, the global affinity reagent market in 2014 has been estimated at a worth of $2.2 billion, with a staggering 2 million products available (Fan 2015), there still exists an immense need for additional specific affinity reagents, because tools useful for identifying many proteins, sequence variants, and post-translational modifications still do not exist (Taussig, Stoevesandt et al. 2007, Colwill, Renewable Protein Binder Working et al. 2011, Edwards, Isserlin et al. 2011). The direness of the situation has been highlighted in a number of recent opinion pieces (Edwards, Isserlin et al. 2011, Anonymous 2015, Bradbury and Plückthun 2015,

For decades, antibodies have been the gold standard for affinity tools, due to their unique structure that permits a high tolerance for sequence variation to make diverse binding surfaces. However, a number of structural complexities inherent to the antibody fold have limited their utility. First, the complex, heterotetrameric structure of conventional IgGs makes production technically challenging, which has been linked to batch-to-batch variations and problems with data reproducibility (Bradbury and Plückthun 2015). Secondly, the large number of disulfides in the native structure prevents antibody usage in live, intracellular protein studies because they do not readily fold within the reducing environment of the cytoplasm. Another class of affinity reagents known as alternative scaffolds has been emerging for the past decade. These reagents, which lack disulfides and are typically less than 10 KDa, have provided relief from antibody limitations, and have been successfully used in diverse applications in research, diagnostics, and therapeutics (Lofblom, Feldwisch et al. 2010, Wurch, Pierre et al. 2012, Hassanzadeh-Ghassabeh, Devoogdt et al. 2013, Tanaka, Takahashi et al. 2015). The lack of disulfides in these affinity reagents opens up an immense potential for intracellular applications, including protein localization studies, the discovery of unknown gene functions, and the validation of potential drug targets.

Alternative scaffolds are typically much less structurally complex when compared to antibodies. Like other successful scaffolds such as Anticalins, Affibodies, Monobodies, and DARPs (Fig. 1.4), Ub also possesses favourable biochemical properties that make it suitable as a scaffold. Ub is remarkably stable at high temperatures in a wide spectrum of pH and denaturation agents (Wintrode, Makhatadze et al. 1994, Ibarra-Molero, Loladze et al. 1999, Kony, Hunenberger et al. 2007), which enables it to tolerate extensive sequence variation for forming new binding surfaces. It is highly soluble with no cysteines, thus allowing for less complex bacterial production methods. The lack of disulfide bonds also permits Ub based reagents for usage in the reducing intracellular milieu, which is currently intractable for traditional antibody formats. Furthermore, the small size of the 76-residue Ub protein also supports modular design for novel capabilities.
Previously, our group and a company, Scil Proteins, have already validated Ub as a viable scaffold option. Our lab has demonstrated Ub as a scaffold to generate activators or inhibitors of enzymes, such as E3s, DUBs, and UBDs in the ubiquitin proteasome system (UPS) (Ernst, Avvakumov et al. 2013, Gorelik, Orlicky et al. 2016, Zhang, Wu et al. 2016). Ub phage libraries were produced by randomizing the natural protein-binding site situated on the flat β-sheet of Ub, in a manner that biased the occurrence of wt sequence. Through panning, Ubvs exhibiting enhanced affinity and specificity for targets were selected. The Ubvs that inhibited or activated ubiquitination were active inside cells. Thus, our previous work validated Ub as a scaffold for making extracellular and intracellular functional reagents.

At Scil Proteins, while the Ub β-sheet was also chosen as the designed binding site, the randomization approach differed substantially, with the effort focused on producing binders to targets that do not naturally bind Ub. Two Ub phage libraries were produced, one of a mono-Ub and another of a di-Ub joined by a linker. The former library had six positions “hard randomized”, which allowed for all 20 amino acids (Hoffmann, Kovermann et al. 2012). The resulting library thus had a restricted complementarity region and a limited library diversity of $6.4 \times 10^7$, which likely explained why only low affinity binders ($\mu$M range) were isolated from this library. The di-Ub library expanded the engineered binding site to 15 residues spread over the two Ub molecules (Lorey, Fiedler et al. 2014), and was successful at isolating nM affinity binders to targets of interest.

Here, I present a phage-displayed Ubv library that differs substantially from previously reported libraries, and applied it to select for tight and selective Ubvs binding to diverse proteins of interest. Like our group’s previous libraries and the libraries reported by Scil Proteins, my naïve Ubv library uses the β-sheet of Ub as the binding surface. However, unlike the previous libraries, I diversified a surface comprising an expansive 24 positions, and thus increased the likelihood for making contacts with antigens. I also biased the binding surface for Tyr and Ser residues, which have proven to contribute functionally to binding (Fellouse, Wiesmann et al. 2004, Fellouse, Li et al. 2005, Fellouse, Barthelemy et al. 2006). In addition, I incorporated numerous quality control measures into the library design and production processes to maximize Ubv stability and display, which are critical parameters that contribute to the success of a library. Finally, I demonstrated the feasibility of making affinity reagents from the Ub library by validating affinity reagents for two proteins of interest: the cell-surface receptor Her3 and the
intracellular adaptor Grb2. Finally, I applied a novel anti-Grb2 Ubv as a genetically encoded intracellular inhibitor of receptor signaling.

3.3 Methods and Materials

3.3.1 Library construction

For mutation tolerance scan libraries, FLAG-Ub was cloned into a previously described phagemid for gene-3 minor coat protein display (Ernst, Avvakumov et al. 2013). Three stop templates were constructed, where each had a different region replaced with a stop codon, and were subsequently used for library randomization. For library design, each position was randomized using degenerate codons shown in Table 3.1, and in total three libraries were produced following procedures described previously (Fellouse 2007). For naïve Ub libraries, a dimerization sequence was added between the FLAG-Ub and the phage protein coding sequence for bivalent display, as described previously for Fab display (Lee, Sidhu et al. 2004). A stop template in which region 1 and 2 of the wt Ub sequence were replaced by stop codons was constructed and used to first construct a library where only region 1 and 2 were randomized using the KHT (K= G/C, H=A/C/T) degenerate codon. An equal mix of two mutagenic oligonucleotides was used to randomize region 2 (Table 3.2), and the phage library was produced as described previously (Fellouse 2007). The resultant phage library was subjected to two rounds of selection for binding anti-FLAG Ab (Sigma F3165) immobilized on 48 wells of a 96-well NUNC Maxisorp plate (Thermo Scientific 12565135) to enrich for variants that were stably displayed. Subsequently, the phage output was amplified in 200 ml of E.coli CJ236 culture, and the isolated ssDNA were used as template for a mutagenesis reaction that randomized region 3. Two Ub naïve libraries were produced that differed in how region three was randomized. In library KHT, KHT degenerate codons were used, while in library NNK, NNK (N=A/C/G/T) degenerate codons were used. The libraries were made by electroporating separately into E. coli SS320 cells. For amino acid frequency determination, 201 and 58 sequences were analyzed from the naïve library KHT or NNK, respectively. Phage pools from the libraries were concentrated to $10^{13}$ CFU/ml, aliquoted and flash frozen in 50% glycerol.
Table 3.1 Degenerate codons used in the mutation tolerance scan

<table>
<thead>
<tr>
<th>Position</th>
<th>wt Residue</th>
<th>Degenerate Codon^</th>
<th>Encoded Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Q</td>
<td>HMK</td>
<td>K N T T Q H P P * Y S S</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>KHT</td>
<td>D V A Y F S</td>
</tr>
<tr>
<td>6</td>
<td>K</td>
<td>HMK</td>
<td>K N T T Q H P P * Y S S</td>
</tr>
<tr>
<td>8</td>
<td>L</td>
<td>HHT</td>
<td>N I T Y F S H L P</td>
</tr>
<tr>
<td>9</td>
<td>T</td>
<td>HHT</td>
<td>N I T Y F S H L P</td>
</tr>
<tr>
<td>10</td>
<td>G</td>
<td>DVT</td>
<td>G D A C Y S S N T</td>
</tr>
<tr>
<td>11</td>
<td>K</td>
<td>HMK</td>
<td>K N T T Q H P P * Y S S</td>
</tr>
<tr>
<td>12</td>
<td>T</td>
<td>HHT</td>
<td>N I T Y F S H L P</td>
</tr>
<tr>
<td>14</td>
<td>T</td>
<td>HHT</td>
<td>N I T Y F S H L P</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>R</td>
<td>YNT</td>
<td>H L P Y F S R C</td>
</tr>
<tr>
<td>44</td>
<td>I</td>
<td>HHT</td>
<td>N I T Y F S H L P</td>
</tr>
<tr>
<td>46</td>
<td>A</td>
<td>KHT</td>
<td>D V A Y F S</td>
</tr>
<tr>
<td>47</td>
<td>G</td>
<td>DVT</td>
<td>G D A C Y S S N T</td>
</tr>
<tr>
<td>48</td>
<td>K</td>
<td>HMK</td>
<td>K N T T Q H P P Y S S</td>
</tr>
<tr>
<td>49</td>
<td>Q</td>
<td>HMK</td>
<td>K N T T Q H P P Y S S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>Q</td>
<td>HMK</td>
<td>K N T T Q H P P * Y S S</td>
</tr>
<tr>
<td>63</td>
<td>K</td>
<td>HMK</td>
<td>K N T T Q H P P * Y S S</td>
</tr>
<tr>
<td>64</td>
<td>E</td>
<td>KV K</td>
<td>G G A A E D W C S S * Y</td>
</tr>
<tr>
<td>66</td>
<td>T</td>
<td>YHT</td>
<td>H L P Y F S</td>
</tr>
<tr>
<td>68</td>
<td>H</td>
<td>YHT</td>
<td>H L P Y F S</td>
</tr>
<tr>
<td>70</td>
<td>V</td>
<td>KHT</td>
<td>D V A Y F S</td>
</tr>
<tr>
<td>71</td>
<td>L</td>
<td>YHT</td>
<td>H L P Y F S</td>
</tr>
<tr>
<td>72</td>
<td>R</td>
<td>YNT</td>
<td>H L P Y F S R C</td>
</tr>
<tr>
<td>73</td>
<td>L</td>
<td>YHT</td>
<td>H L P Y F S</td>
</tr>
<tr>
<td>74</td>
<td>R</td>
<td>YNT</td>
<td>H L P Y F S R C</td>
</tr>
<tr>
<td>75</td>
<td>G</td>
<td>KVT</td>
<td>G D A C Y S V</td>
</tr>
<tr>
<td>76</td>
<td>G</td>
<td>KVT</td>
<td>G D A C Y S V</td>
</tr>
</tbody>
</table>

* indicates a stop
^ all letters follow the International Union of Biochemistry nucleotide code.
Table 3.2 Oligonucleotides used to construct naïve Ub libraries.

<table>
<thead>
<tr>
<th>ID</th>
<th>Region randomized</th>
<th>Oligonucleotide Sequence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL_Reg1</td>
<td>1</td>
<td>AAA GCA GGC TCC ATG KHT ATT KHT GTG KHT ACC KHT KHT GGG KHT KHT ATC KHT CTC GAG GTT GAA CCC</td>
</tr>
<tr>
<td>IL_Reg2 1a</td>
<td>2</td>
<td>CCT CCT GAT CAG CAG KHT CTG KHT TTT GCT GCC KHT YMW CTG GAA GAT GGA CGT</td>
</tr>
<tr>
<td>IL_Reg22b</td>
<td>2</td>
<td>CCT CCT GAT CAG CAG KHT CTG KHT TTT GCT GCC AAG CAG CTC GAA GAT GGA CGT</td>
</tr>
<tr>
<td>IL_Reg3 1c</td>
<td>3</td>
<td>TCT GAC TAC ATT KHT KHT TCT KHT TCT KHT KHT KHT KHT TAC CCA GCT TCC TTT</td>
</tr>
<tr>
<td>IL_Reg3 2d</td>
<td>3</td>
<td>TCT GAC TAC ATT NNK NNK NNK TCT NNK CTT NNK NNK NNK NNK NNK NNK NNK NNK TAC CCA GCT TCC TTT</td>
</tr>
</tbody>
</table>

* all letters follow the International Union of Biochemistry nucleotide code.

a Oligonucleotide randomizing positions 42, 44, 48 and 49.
b Oligonucleotide randomizing positions 42, 44, and 48.
c Oligonucleotide used to construct KHT Ub library.
d Oligonucleotide used to construct NNK Ub library.

3.3.2 Mutation tolerance scan

The three scanning libraries were handled separately, and each was subjected to two rounds of panning on anti-FLAG antibody immobilized on eight wells of a 96-well NUNC Maxisorp plate. Individual clones that bound to the anti-FLAG Ab were identified by phage ELISA (Cunningham, Lowe et al. 1994), and 100-200 binding clones were sequenced from the round 2-selection pool of each library. The frequency for each amino acid occurrence was normalized by adjusting for codon bias, which was done by dividing the occurrence of a particular amino acid by the number of times that amino acid was encoded in the degenerate codon. The mutant/wt ratio was calculated by dividing the normalized frequency of an amino acid substitution by the frequency of the wt.

3.3.3 Ag binding selections

Glutathione S transferase (GST)-Ag fusion proteins were immobilized onto 96-well NUNC Maxisorp plates by incubating 100 μ of protein (2-10 ng/ml) in each well overnight at 4 °C, and blocked with PBS/0.5% BSA for 1 hour at room temperature on the next day. Four rounds of selections were performed as described (Fellouse 2007). All Ags except for Her3, Grb2, MBP, BAFT, ROR-α, and FOXP1 were subjected to HTP selection. The protocol for HTP binding selections differed at the phage output amplification step between selection rounds. Instead of amplifying binding phages in 30 ml volumes of culture in baffled shaker flasks overnight at 37 °C, for the HTP method phage were amplified in 1.4 ml of 2YT broth supplemented with 50 μg/ml carbenicilllin, 25 μg/ml kanamycin, and 10^10 K07 helper phage in 96-deep-well plates.
Subsequently, the amplified phage supernatant was adjusted to neutral pH by adding 155 μl of 10x PBS and used directly in the next round of selection. For round one, a mix of phage from libraries KHT and NNK were incubated with immobilized Ags. To eliminate non-specific binding phages in rounds two through four, phage pools were preincubated in wells coated with GST for 30 minutes at room temperature and then transferred to selection plates.

3.3.4 Protein expression and purification

MBP, and Grb2-SH2 used for SPR were cloned into the pET15 vector (Invitrogen) using Gateway methods as described in (Liang, Peng et al. 2013) for fusion to the C-terminus of a 6xHIS tag. For FLAG-Grb2-SH2 protein used for peptide-binding ELISAs, a FLAG tag was cloned N-terminally to Grb2-SH2 in the pET15 Grb2-SH2 vector using transfer PCR method (Erijman, Dantes et al. 2011). All TFs were cloned into a custom-made isopropyl β-D-1-thiogalactopyranoside (IPTG) expression vector as described in (Huang, Economopoulos et al. 2015) and Ubvs were cloned into pDEST15 vector (Invitrogen) using Gateway methods, which all produced fusions to the C-terminus of GST. Her3-Fc fusion protein was purchased from R&D Systems (348-RB-050). Plasmids designed for protein expression were transformed into E.coli BL21 DE3 pLysS cells. Cultures were grown from single colonies and protein expression was induced with 500μM of IPTG when bacterial growth reached mid-log phase. Induced cultures were grown overnight at 18 °C, collected by centrifugation and lysed using standard methods (Gasteiger 2005). 6xHIS or GST tagged proteins were purified using Ni-NTA affinity resin (Qiagen 30230) or glutathione resin (GE healthcare 17-0756-05), respectively, according to manufacturer’s instructions. The purity of eluted proteins was determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Protein concentrations were determined by measuring the absorption at 280 nm with Nanodrop 1000 (Thermo Scientific) and converted to molarity using the extinction-coefficient determined by Protparam (http://web.expasy.org/protparam/).

3.3.5 Phage and protein ELISAs

All ELISAs were performed in 384-well Maxisorp plates (Thermoscientific 12665347). Assay volumes were 25 μl except for blocking, which was done with 50 μl volumes. For clonal phage ELISAs, 5 ng/ml of Ags or 5ng/ml of BSA (Bioshop ALB001.500) were immobilized as in section 3.3.3. Binding to immobilized anti-FLAG Ab (2 ng/ml) was used to assess display
levels. Clonal phage cultures were grown by picking single colonies into 400 μl of 2YT broth supplemented with 50 μg/ml carbenicillin and 10^10 CFU/ml of K07 helper phage and incubated overnight at 37 °C. Supernatants containing phage were used directly for ELISAs, as described (Tonikian, Zhang et al. 2007) except for the following modifications: phages were incubated with immobilized Ag for 15 min and four washes were conducted to remove non-binding phages. The procedure for specificity ELISAs conducted with purified GST-Ubv proteins was the same as that for clonal phage ELISAs, except that binding to Ubv was detected with anti-GST-HRP Ab (1:8000, Sigma A7340).

For EC_{50} determination by ELISA, phages or Ubvs were serially diluted in 96-well plates, transferred to plates that were coated with immobilized Ag, and incubated for 15 min. EC_{50} values were calculated by fitting binding curves with a four parameter logistic nonlinear regression model using Graphpad Prism software and were defined as the concentration required to achieve 50% of the maximum binding signal.

For competitive phage ELISAs, phage supernatants were first diluted 3 fold in PBS and then incubated with serial dilutions of soluble Ag in 96 well plates for 1 hour at room temperature. Subsequently, the solutions were transferred to 384 well ELISA plates, and the assay was completed using same protocol as described for clonal phage ELISAs.

For IC_{50} assays, ELISA plates were prepared by first immobilizing 25 μl of 5ng/ml of streptavidin overnight. Next day, plates were blocked with 50 μl of PBS/0.5% BSA and then 25 μl of 5 μM biotinylated p-Tyr SHC peptide (sequence: DHQpYYNDFP, GenScript) or SHC peptide (sequence: DHQYYNDFP, GenScript) were added. In a 96 well plate, serial dilutions of Ubv protein were incubated with 25 nM His-FLAG-Grb2-SH2 for 1 hour at room temperature before transfer to the ELISA plate. After 15 min of binding, the plate was washed four times with PBS and Grb2-SH2 binding was detected by anti-FLAG-HRP Ab (Sigma A8592).

### 3.3.6 Immunoprecipitations

For immunoprecipitation (IP) of Her3, all HEK293T cell lines expressing Her receptors were stable expression lines generated by Lentiviral transduction, except for HEK293T expressing EGFR. HEK293T cells overexpressing EGFR were made by transfecting a 15-cm dish of HEK293T cells at ~70 confluency with 5 μg of pCDNA3.1 EGFR plasmid using X-
tremeGENE9 transfection reagent (Roche 06365809001) and according to manufacturer’s instructions. Unless stated otherwise, all DNA transfections into mammalian cells were performed with X-tremeGENE 9 transfection reagent and used according to manufacturer’s instructions. Stable cell lines or transfected cells were grown to ~95% confluency on 15 cm plates and were lysed directly on plates by adding lysis buffer (Sigma S8830) supplemented with protease and phosphatase inhibitor cocktail (PPIC) (ThermoScientific 78440). Collected lysates were incubated on ice for 20 minutes and centrifuged at 21,130 g for 10 minutes. Protein concentrations were determined by bicinchoninic acid assay (ThermoScientific 23225). For each sample, 500 μg of lysate protein in 1 ml of lysis buffer were precleared by incubating with 50 ml glutathione sepharose slurry (GE #17-0756-05) by nutating at 4°C for 1 hour. Subsequently, 3 mg of Ubv or Ub.wt (negative control) were added to each sample and incubated at 4 °C from 4 hours or overnight. IPs were performed using 40 μl glutathione sepharose slurry according to manufacturer’s instructions. The resin was washed three times with 500 ml of lysis buffer and proteins were eluted by adding 50 μl SDS-PAGE loading buffer (Invitrogen NP0007).

For IP of Grb2, HEK293T cells were grown to ~70% confluency in 6-well plates and were transfected with 1.5 μg of pCAGGS GST-Ubv plasmid alone or in combination with 0.3 μg pCDNA3 3xFLAG-Grb2 plasmid. Two days after transfection, cells were lysed, IPs were performed using 20 μl glutathione sepharose slurry and proteins were eluted by adding 20 μl SDS-PAGE loading buffer, as described above.

3.3.7 Cell fractionation

HEK293T cells were grown to ~70% confluency in 6-well plates and were transfected with 1.5 μg of pCAGGS GST-Ubv plasmid. Two days post transfection, cells were washed with ice-cold PBS. Cytosolic and membrane fractions were isolated with a Subcellular Protein Fractionation Kit (ThermoScientific 78840) used according to manufacturer’s instructions, with the exception that an additional wash with 500 μl PBS supplemented with 1xPPIC was performed after isolation of cytosolic fraction and prior to membrane fraction isolation.

3.3.8 Western blotting

The following sample volumes were loaded for SDS-PAGE: 10 μl for Her3 IP experiment and EGFR signaling experiment, 15 μl for Grb2 IP experiment, and 25 μl for fractionation experiment. Proteins were transferred to PVDF member using standard methods. Instead of
stripping and reprobing blots, multiple fresh blots were used to probe for different proteins of the same size in each experiment to avoid possible residual signals. Blots were probed with antibodies against EGFR (Invitrogen 44798G), p-EGFR (Invitrogen 44794ZG), Her2 (LS Biosciences N12), Her3 (ThermoScientific 2F12), GST (Sigma A7340), HSP90 (SantaCruz sc-7947), Na^+K^+ATPase (Abcam ab7671), and Abs targeting the following proteins were purchased from Cell Signaling Technology: Her4 (4795), GAPDH (2118), Grb2 (3972), ERK (9101), p-ERK (4370), MEK (4694), and p-MEK (9154).

3.3.9 EGFR signaling assay

HEK293T cells in 6-well plates were grown to ~70% confluency and transfected with 1.5 μg of pCAGGS GST-Ubv plasmid. After 6-8 hours, cells were transfected with siRNA designed to knock down Grb2 (ThermoScientific S6121) or mock transfected with Lipofectamine RNAiMAX (Invitrogen 13778-075). After 12-16 hours, cells were washed with PBS and starved by switching to media without FBS supplement. The next day, cells were stimulated with EGF (1 ng/ml), and after 10 minutes, cells were harvested by washing with ice-cold PBS. Cells were scraped off and each sample was split into two Eppendorf tubes. Cells in one tube were lysed for total protein analysis and those in the other tube were processed by fractionation for Grb2 localization analysis (see Section 3.3.7).

3.3.10 SPR kinetic binding analysis

SPR measurements were performed at 25 °C using a ProteOn XPR36 instrument (Bio-Rad). Proteins were immobilized by amine coupling to GLC sensor chip surfaces. GST-Ubv proteins were diluted with PBS buffer and injected for 360 s at a flow rate of 50 μl/min. For binding kinetics, sensorgrams were fitted to 1:1 Langmuir model using ProteOn Manager Software (Bio-Rad).

3.4 Results

3.4.1 Mutation Tolerance Scan

To produce a combinatorial phage-displayed Ubv library to generate high affinity and selective binders to targets of interest, I aimed to use a large surface on Ub as the binding site. Previously, we identified a β-sheet surface on Ub consisting of 27 residues that was amenable for
engineering by phage display (Ernst, Avvakumov et al. 2013). By mutating a few residues on this surface, we were able to develop Ubv inhibitors and activators of enzymes in the Ub proteasome system. To extend the utility of the Ub scaffold and to develop Ubvs capable of binding to proteins that do not naturally interact with Ub, I decided to diversify all 27 residues with a subset of amino acids (Tyr, Ala, Asp, Ser, Phe, and Val) that have been proven to function in the binding sites of synthetic antibody and alternative scaffolds (Fellouse, Wiesmann et al. 2004, Gilbreth, Esaki et al. 2008, Mandal, Uppalapati et al. 2012). However, this resulted in a non-optimal library in which >95% of phage clones did not display Ubvs at detectable levels in phage ELISAs (Fig. 3.1), thus severely jeopardizing the probabilities of deriving functional Ubvs. We attributed low display levels to the high mutation frequency resulting in unstable variants incorporating destabilizing mutations, which would then be proteolytically cleaved in the *E.coli* periplasm during phage assembly (Betton, Sassoon et al. 1998, Ernst, Sazinsky et al. 2009). As a first step towards designing a library of more stable Ubvs, I designed an experiment to investigate which of the 27-targeted positions were intolerant to mutation.

![Figure 3.1 Phage ELISA of Ub library](image)

**Figure 3.1 Phage ELISA of Ub library**

Twenty-four randomly selected clones from the library constructed by randomizing 27 residues Ub were subjected to anti-FLAG ELISA to assess display. Ub.wt denotes Ub.wt displayed on phage.

Since detection of an epitope tag, such as FLAG, fused to the N-terminus of a displayed protein can be used as a proxy for stability (Betton, Sassoon et al. 1998, Ernst, Sazinsky et al. 2009), I designed a phage display approach called mutation tolerance scan to systematically assess and compare FLAG tag detection when wt residues were substituted by other residues.
First, I designed three phage libraries that together covered the 27 residues (Fig. 3.1). Each position was randomized with a degenerate codon that only encoded six to nine residues including the wt, Tyr and Ser (Table S1), and this allowed me to completely sample the theoretical diversities and enabled the quantitative assessment of the relative frequency of each substitution compared with the wt. The wt residue was important for comparing the occurrence of mutations relative to wt. Tolerance to substitution by Tyr or Ser was of interest because of their importance in mediating interaction in antibodies and scaffold based complementarity regions (Fellouse, Wiesmann et al. 2004, Fellouse, Barthelemy et al. 2006).

Figure 3.2 Library designs for mutational tolerance scan.

Residues randomized in each of the three scanning libraries are colored differently and mapped onto (A) the Ub primary sequence and (B) the Ub structure (PDB entry: IUBQ). Secondary structure elements, β-strands (blue arrows) and α-helices (red zigzags), are indicated above sequence.

Following selections for binding to anti-FLAG Ab, many clones that were positive for anti-FLAG Ab binding by phage ELISA were sequenced. The sequences were used to calculate a mutant/wt ratio at each scanned position, and based on this ratio, positions were categorized as exhibiting, low, medium or high tolerance to mutations (Fig. 3.2). The analysis showed that of the 27 scanned positions, 23 positions exhibited either medium tolerance, defined as 0.5 > mutant/wt ratio < 1.5, or high tolerance, defined as mutant/wt ratio ≥ 1.5, for at least half of the encoded mutations. Only four positions exhibited low tolerance, defined as mutant/wt ratio < 0.5 for all or all but one tested substitutions (Fig. 3.2). Surprisingly, many positions tolerated side-chains that were substantially different than wt and in some cases multiple mutations with different chemical properties were preferred over wt. For example, position 62 preferred Pro or Tyr over the wt Gln (Fig. 3.2). Our data showed that Ub is structurally robust, and this is in
agreement with literature showing that Ub is very stable protein (Ibarra-Molero, Loladze et al. 1999, Lange, Lakomek et al. 2008).

![Mutation Tolerance Scan](image)

**Figure 3.2 Mutant/wt ratios derived from mutation tolerance scanning data.**

The scan sampled 6-9 amino acids at each position, and amino acids not sampled are omitted. Ratios are shaded as follows: red ≤ 0.5, 0.5 > grey < 1.5, 1.5 ≥ green, and blue indicates the wt. Ub sequences shaded in yellow represent residues that were intolerant to mutation and were retained as wt in the final library design.

### 3.4.2 Library Design

The mutation tolerance scan identified four positions, Gly10, Ala46, Gly47 and Gln49 that exhibited low tolerance to mutation (Fig. 3.2). Within the Ub structure, G10, A46, and G47 are found in β-turns (Fig. 3.3), hence it is likely that the small residues were required to maintain the loop conformation. Thus, we fixed positions 10, 46, and 47 as wt in the library design. Gln49 forms a a two-residue β strand with Lys48 (Fig. 3.3), which is highly tolerant to mutation. Since Gln49 and Lys48 together form a distinct surface, but one is highly tolerant of mutation and one is intolerant to mutation, we decided to use a partial randomization strategy at these two positions, where wt was allowed at least 50% of the time. At position 48, wt Lys was allowed at 50%, while 50% of the library was diversified with the KHT degenerate codon, which codes for Tyr, Ala, Asp, Ser, Phe, and Val. At position 49, while wt Gln was allowed at 50%, the other 50% of the library was diversified with the degenerate codon YMW. This codon was chosen to
further increase the variants stability, as it coded for the wt Gln in addition to Tyr, Ser. Due to
codon degeneracy, His and Pro were also included.

**Figure 3.3 Ub positions that exhibit low tolerance to mutations.**
The main-chain of the Ub structure (PDB entry: IUBQ) is shown as a ribbon, and spheres indicate positions that exhibit low tolerance to mutations in the mutation tolerance scan, except for position 48, which is tolerant to mutations. Scanned positions are colored as in Fig. 3.1.

Phage-displayed protein libraries that contain a high proportion of non-displaying phages can severely jeopardize selection outcomes (Kramer, Cox et al. 2003), and because display is linked to stability (Betton, Sassoon et al. 1998, Ernst, Sazinsky et al. 2009), we sought to further increase the structural stability of displayed Ubvs by integrating a stability selection step in the library construction process. The mutation tolerance scan revealed regions one and two to be more structurally sensitive to mutation than region 3 (Fig. 3.2). Thus, we first constructed a library in which regions 1 and 2 were randomized with the degenerate codon KHT (encoding for Tyr, Ala, Asp, Ser, Val, and Phe) and subjected this library to a selection for binding to an anti-FLAG Ab to enrich for Ubvs that were well-displayed. The resulting pool of binding phage were used make a template in which region 3 was randomized. For region 3, in addition to using the KHT degenerate codon for randomization (referred to as Library KHT), a second library was constructed where positions were randomized with the NNK degenerate codon, which encodes for all 20 amino acids (referred to as Library NNK), because I wanted to study whether additional diversity may provide better binding Ubvs.

In total, 24 of the 76 positions in the Ub sequence were diversified, and we constructed libraries that contained $6.8 \times 10^{10}$ or $4 \times 10^{9}$ unique Ubvs in Libraries KHT or NNK, respectively. Sequencing of the naïve libraries showed that region one had an enrichment for Phe at the
expense of Ala, and regions two and three showed unbiased distributions of each allowed residue (Fig. 3.4). Libraries KHT and NNK had some wt sequence in region 3 due to incomplete oligonucleotide incorporation, at a frequency of 8% or 14%, respectively. Notably, 15% of the library NNK sequences had at least one stop codon incorporated in region 3.

Figure 3.4 Amino acid frequencies in libraries KHT and NNK.
Bar graphs show the amino acid frequencies for (A) library KHT and (B) library NNK. Asterisks (*) indicate the wt residues.
3.4.3 Binding selections

I pooled phage from libraries KHT and NNK and assessed the ability to derive binders to a panel of 40 diverse proteins (Fig. 3.5). For comparison, 37 of the 40 proteins were also subjected to binding selections with the highly validated Fab library F (Persson, Ye et al. 2013). I successfully isolated specific binding Ubv-phage clones for 30 of the 40 Ags (Fig. 3.5), and I determined the sequences of a total of 99 unique binding clones for 14 Ags (Fig. 3.6). Sequences of many clones that bound to the same target showed conservation at particular positions, suggesting that the conserved positions may be critical for binding to a common epitope. Of the 37 proteins that were subjected to selection with both Ubv and Fab libraries, the Ubv libraries yielded binders to 28 antigens, the Fab library yielded binders to 26 antigens, and 21 targets yielded binders with both libraries. Neither libraries were successful against four targets (ZFP691, EOMES, SFP1, and ORS2), suggesting that these targets may be unstable or misfolded and thus poorly antigenic. Taken together, these results show that Ubv-phage library KHT is robust and as functional as the highly validated Fab library F in terms of generating binders against a large and diverse panel of proteins.
Figure 3.5 Binding selection outcomes for a panel of protein antigens.

For each antigen, the indicated domain was used as a target for binding selections. A plus sign (+) indicates that library F yielded binding Fabs but percent positives were not determined. Results with library F were taken from previous studies for Grb2 (Colwill, Renewable Protein Binder Working et al. 2011) and MBP (Persson, Ye et al. 2013). “NA” indicates that the selection was not attempted. Green and red shading denote successful or failed selections, respectively. A selection was considered to be successful if it yielded at least some clones exhibiting at least five-fold higher binding signal by phage ELISA against antigen compared with an irrelevant protein, and numbers in cells show the number of positive clones followed by the number of clones screened.
<p>| Antigen | Ubr | 2 | 4 | 6 | 8 | 9 | 11 | 12 | 14 | 22 | 44 | 48 | 49 | 62 | 63 | 64 | 66 | 68 | 70 | 72 | 73 | 74 | 75 | 76 |
|---------|-----|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Her3.1  | Y   | Y | Y | Y | Y | D | S | F | F | F | F | K | K | A | A | D | D | D | Y | A | V | D | A | A | V | 37 | 1  |
| Her3.2  | F   | Y | Y | V | A | F | F | D | Y | S | K | K | Q | V | D | V | Y | A | D | D | D | V | Y | D | V | 23 | 1  |
| Her3.3  | Y   | F | Y | A | D | F | Y | Y | V | A | K | Q | Y | D | D | F | Y | D | V | D | S | V | S | 28 | 1  |
| Her3.4  | A   | F | F | F | D | Y | F | D | S | Y | K | Q | D | D | V | Y | D | V | D | V | V | D | V | D | A | A | 15 | 1  |
|         | Y   | F | Y | D | V | F | F | F | F | V | S | K | Q | V | D | D | D | V | D | V | V | A | S | D | D | 12 | 1  |
|         | V   | F | F | S | F | F | F | F | K | K | Q | V | D | D | S | A | V | F | V | F | D | A | Y | 24 | 2  |
|         | F   | F | Y | V | V | V | F | F | D | S | Y | K | Q | D | F | S | I | L | N | I | G | A | A | 14 | 1  |
|         | F   | F | S | Y | F | F | D | S | F | F | Q | T | E | L | O | T | E | Y | Q | P | D | R | G | 13 | 1  |
|         | V   | F | Y | A | D | F | F | Y | Y | K | Q | S | D | Y | D | Y | D | D | D | V | V | D | 23 | 1  |
|         | Y   | F | D | F | Y | F | D | Y | S | K | Q | D | A | F | F | V | V | A | D | D | D | F | 15 | 1  |
|         | Y   | F | D | F | F | F | D | Y | V | K | Q | A | D | V | D | F | A | A | D | D | D | 26 | 1  |
|         | V   | Y | D | S | Y | F | D | Y | Y | Y | P | V | D | V | D | Y | D | V | A | D | D | 20 | 1  |
| Her3    | Y   | Y | V | D | F | F | D | S | F | K | Q | A | D | V | D | F | S | S | D | V | D | V | V | 32 | 1  |
|         | Y   | Y | A | D | F | F | D | Y | F | A | P | D | D | Y | A | D | F | D | A | A | V | S | 11 | 1  |
|         | Y   | Y | V | D | F | F | S | F | K | Q | V | D | D | V | D | A | D | V | Y | A | A | D | 30 | 1  |
|         | Y   | Y | V | D | F | F | D | S | F | K | Q | V | D | D | V | D | A | D | V | A | D | Y | A | A | D | 14 | 1  |
|         | A   | Y | F | D | F | F | D | S | F | K | Q | V | D | D | V | D | A | D | V | A | D | V | A | 14 | 1  |
|         | V   | Y | D | F | F | Y | D | S | F | K | Q | V | D | D | V | D | A | D | V | A | D | V | A | 14 | 1  |
|         | Y   | Y | V | D | F | F | D | S | F | K | Q | V | D | D | V | D | A | D | V | A | D | V | A | 14 | 1  |
|         | Y   | Y | V | D | F | F | D | S | F | K | Q | V | D | D | V | D | A | D | V | A | D | V | A | 14 | 1  |
| Grb2.1  | D   | S | V | S | D | A | V | D | A | D | Y | P | S | I | N | V | W | M | H | P | V | N | L | G | 23 | 8  |
| Grb2.2  | D   | S | V | S | D | A | V | D | A | D | Y | P | S | I | N | V | W | M | H | P | V | N | L | G | 23 | 8  |
| Grb2    | D   | S | V | S | D | A | V | D | A | D | Y | P | S | I | N | V | W | M | H | P | V | N | L | G | 23 | 8  |
| BATF    | F   | F | D | D | D | D | Y | D | D | S | D | S | S | S | F | S | F | D | Y | A | V | Y | S | S | S | 6  | 1  |
| FOXO3   | Y   | Y | D | S | F | F | Y | F | P | S | A | Y | Y | A | D | D | D | D | F | F | V | D | A | F | Y | 13 | 2  |
| Goseoid2 | S   | D | A | V | F | D | V | Y | Y | K | Q | S | S | S | F | D | Y | D | D | Y | V | A | F | 10 | 9  |
| MBP     | A   | V | Y | A | A | A | S | A | Y | D | F | V | S | A | F | F | S | A | F | V | D | F | V | A | D | 8  | 41 |
| Sna1    | D   | V | V | D | D | D | D | D | Y | K | K | D | D | F | D | D | D | T | D | D | D | 14 | 1  |
|         | V   | F | F | F | F | D | D | D | D | D | D | D | D | D | D | D | D | D | D | D | D | D | D | D | D | 11 | 2  |
|         | Y   | V | D | V | A | Y | S | D | F | Y | Y | P | D | F | D | Y | S | Y | D | D | V | Y | Y | F | 29 | 2  |
|         | F   | V | Y | D | V | F | D | Y | P | D | S | Y | D | Y | Y | Y | Y | Y | D | A | V | Y | Y | Y | 11 | 1  |
|         | Y   | V | D | V | D | D | Y | F | Y | V | F | S | D | Y | Y | Y | Y | D | Y | S | S | S | S | S | S | 9  | 1  |
|         | V   | Y | D | V | D | D | Y | F | Y | V | F | S | D | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | 15 | 1  |
|         | A   | Y | D | D | Y | D | F | V | Y | K | Q | D | Y | D | F | F | F | Y | D | D | D | Y | D | D | D | 22 | 9  |
|         | F   | V | D | D | Y | D | F | V | Y | K | Q | D | Y | V | F | Y | V | F | Y | D | Y | D | D | D | D | 8  | 2  |
|         | Y   | F | D | Y | D | F | D | Y | V | D | S | F | D | Y | V | D | D | D | D | D | D | D | D | D | D | 15 | 1  |
| ROR-alpha | D | D | V | V | A | S | F | D | D | A | K | Q | Y | F | F | F | D | V | D | F | D | S | F | S | 9 | 2  |
|         | V   | V | D | S | Y | A | A | V | V | F | F | F | D | D | D | D | D | D | D | D | D | D | D | D | D | 10 | 8  |
|         | Y   | Y | V | F | F | D | D | D | V | F | F | F | D | D | D | D | D | D | D | D | D | D | D | D | D | 10 | 2  |</p>
<table>
<thead>
<tr>
<th>Antigen</th>
<th>Uvb</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Signal/Noise Ratio</th>
<th>Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>76</td>
</tr>
</tbody>
</table>

Figure 3.6 Sequences of binding Uvbs
Sequences are shown for positions diversified in the library design and grey shading indicates at least 50% identity amongst clones binding to a given antigen. “Signal:noise” indicates the ratio of the phase ELISA signals for binding to the antigen and an irrelevant protein. “Count” indicates the number of times the sequence occurred amongst sequenced binding clones.

3.4.4 Characterization of Uvbs targeting Her3

Her3 is a cell-surface receptor tyrosine kinase and is of great biological significance both in development (Britsch, Li et al. 1998, Schmucker, Ader et al. 2003, Stern 2008, Buac, Xu et al. 2009) and cancer biology (Stern 2008, Amin, Campbell et al. 2010, Campbell, Amin et al. 2010). Due to its importance in biology, I sought to generate binding Uvbs to the extracellular domain (ECD) of Her3. I identified 29 unique Her3 binding Uvbs that showed conservation at
many positions (Fig. 3.6), suggesting that they bind to a common epitope. Based on relative affinities determined by competitive phage ELISAs (Fig. 3.7), the four tightest binders (Her3.1-Her3.4) were chosen for further characterization as purified proteins.

**Figure 3.7 Competitive phage ELISAs for Her3 binding Ubvs.**
Thirty unique phage clones were preincubated with 250 nM, 100 nM, 25 nM or 0 nM of recombinant Her3 in solution, and then transferred to ELISA plates with immobilized Her3. Binding to immobilized Her3 was detected and binding signals were normalized to signal obtained with 0 nM Her3 preincubation.

First, I performed dose response binding ELISA against Her3 to obtain relative affinities for the four purified Her3 selected Ubvs (Ubv-Her3.1 – 3.4). The half-maximum effective concentration (EC$_{50}$) was calculated to be the lowest, and hence the highest affinity, for Ubv-Her3.2 (EC$_{50}$ = 90 nM) (Fig. 3.8A). I also conducted binding ELISA against all four ErbB members and confirmed that the four Her3 selected Ubvs bound only to Her3 (Fig. 3.8B), despite homology between EGFR members (Her3 to EGFR identity: 44%, Her3 to Her2 identity: 42%, Her3 to Her4 identity: 57%). Next, I conducted a competition ELISA with the Her3 ligand Neregulin (NRG) (Carraway, Sliwkowski et al. 1994). In the presence of NRG, Ubvs bound Her3 with similar intensities as in the absence of NRG, indicating Ubvs do not compete with NRG binding (Fig. 3.8C). Only Ubv-Her3.2 was carried forward, as it bound Her3 with the highest affinity according to dose response ELISA. To measure the affinity of Ubv-Her3.2 for Her3, we studied binding kinetics by surface plasmon resonance (SPR). While an exact affinity value could not be determined because the Langmiur binding model cannot
accurately fit the SPR binding data, it nonetheless provided us with an approximate affinity measurement in the low nM range (Fig. 3.9).

**Figure 3.8 Analysis of purified Her-binding Ubvs.**

(A) ELISAs for Ubvs binding to immobilized Her3. EC\textsubscript{50} was determined as the Ubv concentration at which 50% of the saturation signal is reached for binding to immobilized Her3. (B) Specificity ELISAs for Ubvs binding to the human ErbB family members. (C) Competitive ELISA for Ubvs binding to immobilized Her3 in the presence or absence of the Her3 ligand NRG. Fab D5 is an antibody that competes with NRG and served as a positive control.

**Figure 3.9 Kinetic analysis of Ubv-Her3.2 binding to Her3**

Sensorogram showing binding to immobilized Her3 by 2-fold serial dilutions of Ubv-Her3.2 starting at 50 nM. Colored lines represent actual binding data and black lines represent fits by Langmiur binding model.
Table 3.3 Kinetic analysis of Ubv binding to corresponding targets.

<table>
<thead>
<tr>
<th>Ubv</th>
<th>Target</th>
<th>$k_a$(M$^{-1}$ sec$^{-1}$)</th>
<th>$k_d$(sec$^{-1}$)</th>
<th>Apparent $K_D$(M)</th>
<th>IC50 (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Her3.2</td>
<td>Her3 ECD</td>
<td>1.7E+06</td>
<td>1.5E-03</td>
<td>4.7E-09</td>
<td></td>
</tr>
<tr>
<td>Grb2.1</td>
<td>Grb2-SH2</td>
<td>1.6E+06</td>
<td>2.5E-03</td>
<td>1.6E-09</td>
<td>NA</td>
</tr>
<tr>
<td>Grb2.2</td>
<td>Grb2-SH2</td>
<td>7.4E+05</td>
<td>7.5E-04</td>
<td>1.0E-09</td>
<td>1.5E-08</td>
</tr>
<tr>
<td>Grb2.1-2.2 L12</td>
<td>Grb2-SH2</td>
<td>1.0E+06</td>
<td>8.3E-04</td>
<td>8.3E-10</td>
<td>7E-09</td>
</tr>
<tr>
<td>Grb2.1-2.2 L16</td>
<td>Grb2-SH2</td>
<td>1.0E+06</td>
<td>1.0E-03</td>
<td>1.0E-09</td>
<td></td>
</tr>
<tr>
<td>Grb2.1-2.2 L20</td>
<td>Grb2-SH2</td>
<td>7.8E+05</td>
<td>1.1E-03</td>
<td>1.4E-09</td>
<td></td>
</tr>
<tr>
<td>Grb2.2-2.1 L12</td>
<td>Grb2-SH2</td>
<td>1.3E+06</td>
<td>3.6E-04</td>
<td>2.8E-10</td>
<td></td>
</tr>
<tr>
<td>Grb2.2-2.1 L16</td>
<td>Grb2-SH2</td>
<td>9.8E+05</td>
<td>4.6E-04</td>
<td>4.7E-10</td>
<td></td>
</tr>
<tr>
<td>Grb2.2-2.1 L20</td>
<td>Grb2-SH2</td>
<td>9.5E+05</td>
<td>5.1E-04</td>
<td>5.3E-10</td>
<td></td>
</tr>
</tbody>
</table>

To determine whether the anti-Her3 Ubvs can recognize Her3 specifically in the cellular context, I performed immunoprecipitation (IP) with Ubv-Her3.2. IP was conducted with lysates from four HEK293T cell lines that overexpressed individual ErbB family members, and four cell lines that expressed multiple ErbB family members (Fig. 3.10A). Western blotting analysis of IPs from HEK293T overexpression cells showed that UbV-Her3.2 effectively pulled down Her3, but not other ErbB family members (Fig. 3.10B), confirming specificity. For HEK293T cells that overexpressed Her2 and Her3, which result in constitutive heterodimerization (Holbro, Beerli et al. 2003, Amin, Campbell et al. 2010), IP of Her3 resulted in co-IP of Her2. Similarly, for the three cancer cell lines (A431, MCF7 and SkBr3), which expressed Her3 in combination with other ErbB members, IP of Her3 resulted in co-IP of other ErbB family members (Fig. 3.10B). Taken together, these results show that Ubv-Her3.2 is Her3 specific and can be used to IP Her3 and to co-IP associated proteins to map Her3 receptor interactions.
**Figure 3.10 UBV-Her3.2 as a tool to immunoprecipitate Her3 and co-immunoprecipitate Her3 interacting Her members from cell lysates**

(A) Profiling of ErbB family members from nine cell lines by western blot. Superscripts “hi” and “lo” denote high and low exposure for blot development, respectively. GAPDH serves as loading control. (B) Western blot analysis of IPs with GST UBV-Her3.2. Ub.wt serves as a negative control. WCL indicates whole cell lysate that had the expression of the ErbB member being probed for that particular blot.

3.4.5 Characterization of Ubvs targeting the Grb2 SH2 domain.

Grb2 is an intracellular adaptor molecule with one SH2 domain flanked by two SH3 domains (Lowenstein, Daly et al. 1992), and it is central to transducing signals resulting from activation of receptor tyrosine kinases (RTKs) (e.g. EGFR, c-MET and FGFR) into intracellular signaling such as the MAPK pathway (Tari and Lopez-Berestein 2001). Selections against the Grb2 SH2 domain (Grb2-SH2) yielded on binder (UBV-Grb2.1) that originated from library NNK and four binders (UBV-Grb2.2-5) that originated from library KHT (Fig. 3.6). The binders derived from library KHT exhibited sequence similarities that suggested that they likely bind the same epitope, while the binder derived from library NNK exhibited a unique sequence (Fig. 3.6).
Ubv-Grb2.1 and U bv-Grb2.2 had the highest relative affinities as determined by competitive phage ELISA (Fig. 3.11) and thus were selected for purification and further characterization.

**Figure 3.11 Competition phage ELISA for Grb2 selected Ubvs.**

Ubv-phage were preincubated with 250 nM, 100 nM, 25 nM or 0 nM of Grb2-SH2 in solution, and then transferred to ELISA plates with immobilized Grb2-SH2. Binding to immobilized Grb2-SH2 was measured and binding signals were normalized to signal obtained with 0 nM Grb2-SH2 preincubation.

I first assessed binding of each U bv to Grb2-SH2 and the two SH2 domains from GRAP and GADS, which are most closely related to Grb2-SH2, exhibiting 63% and 56% identity, respectively (Liu, Jablonowski et al. 2006). ELISAs revealed that U bv-Grb2.1 is highly specific for Grb2-SH2 while U bv-Grb2.2 exhibited low but detectable binding to the other two SH2 domains (Fig. 3.12A). This suggested that Grb2.1 bound to a non-conserved epitope on Grb2-SH2 while U bv-Grb2.2 bound to an epitope that was somewhat conserved amongst the three SH2 domains. Mapping of sequence conservation onto the structure of Grb2-SH2 revealed that many of the residues that are identical across all three SH2 domains are located in the ligand-binding site (Fig. 3.12B), suggesting that U bv-Grb2.2 likely binds to this site and may thus act as an antagonist of ligand binding. Indeed, a competition ELISA confirmed that U bv-Grb2.2 inhibited binding of Grb2-SH2 to an immobilized phospho-Tyrosine (p-Tyr) peptide ligand with an IC$_{50}$ value of 16 nM, while U bv.Grb2.1 had no effect (Fig. 3.12C).
Figure 3.12 Analysis of purified Grb2-binding Ubvs

(A) ELISAs for Ubvs binding to immobilized SH2 domains of Grb2, GADS and GRAP. (B) Surface representation of Grb2-SH2 structure (PDB entry: 1JYR). Identical residues between GRAP, GADS and Grb2 are shaded in blue. p-Tyr peptide ligand is shown in yellow. (C) Dose response curve for Grb2-SH2 binding to immobilized p-Tyr SHC peptide. Ubv-Grb2.1 and non-biotinylated P-Tyr SHC peptide inhibit Grb2-SH2 binding to immobilized peptide with IC<sub>50</sub> of 16 nM and 30 μM, respectively. (N = 3).

To assess whether the Ubvs could bind to full-length Grb2 in cells, I immunoprecipitated intracellularly expressed Ubvs from HEK293T cells and detected Grb2 by western blotting. In addition, I performed the same IP experiment with HEK293T cells overexpressing 3xFLAG-Grb2 (referred to as 293T-3FGrb2). These experiments showed that IP of Ubv-Grb2.1 and Ubv-Grb2.2 co-immunoprecipitated overexpressed 3xFLAG-Grb2 but no detectable levels of endogenous Grb2 (Fig. 3.13), which was expressed at a much lower level (see expression comparison in anti-Grb2 western blots in Fig. 3.13). Importantly, a negative control Ubv that bound to maltose binding protein, Ubv-MBP.1, did not bind Grb2 (Fig. 3.13). Hence, Grb2-SH2 binding Ubvs recognize full length Grb2 intracellularly but affinity was apparently not sufficient to pull down the less abundant endogenous Grb2. To boost affinity by avidity, I took advantage
of the small size of Ubvs and linked them in a modular fashion, which has been shown as a viable strategy to improve target binding in cells (Grebien, Hantschel et al. 2011).

**Figure 3.13 Grb2 immunoprecipitation by Ubvs**

Co-IP of transiently expressed Grb2 (3xFLAG Grb2) or endogenous Grb2 (Endo Grb2) with Ubvs. Indicated Ubvs were expressed in HEK293T cells or in HEK293T cells over-expressing Grb2 with an N-terminal triple-FLAG epitope fusion. Grb2.1 – Grb2.2 and G2.2 – Grb2.1 indicate tandem fusions of Grb2.1 and Grb2.2 with Grb2.1 at the N- or C-terminus, respectively, and the numbers “20”, “16” and “12” indicate intervening linkers of 20 (GGSGGSGGSGGSGS), 16 (GGSGGSGS)GSGSGGSGS) or 12 (GGSGGSGGSGGGS) residues, respectively. “MBP.1” indicates an Ubv that binds to MBP, which was used as a negative control. Precipitated endogenous Grb2, transiently expressed Grb2 and GST-Ubv fusions were detected by western blotting with anti-Grb2, anti-FLAG or anti-GST antibodies, respectively. “WCL” indicates western blotting of the whole cell lysate with an anti-Grb2 antibody to compare relative levels of transiently expressed and endogenous Grb2.

Plasmids designed to express Ubvs linked in different orientations (Grb2.1-Grb2.2 or Grb2.2-Grb2.1) with Gly-Gly-Ser linkers at different lengths (12, 16, or 20 residues) were transfected into HEK293T or 293T-3FGrb2 cells, and IP experiments were conducted. Compared to unlinked Ubvs, linked Ubvs immunoprecipitated 3XFLAG-Grb2 much more effectively, but linker length had no effect on binding (Fig. 3.13). Notably, the Ubvs in the Grb2.1-Grb2.2 orientation pulled down endogenous Grb2 much more efficiently than those in the Grb2.2-Grb2.1 orientation. Interestingly, surface plasmon resonance (SPR) did not show enhanced affinities for the Ubv linked in the Grb2.1-Grb2.2 orientation (Fig. 3.14 D-F and Table 3.3), and only showed two- to three-fold enhancements for Ubvs in the Grb2.2-Grb2 orientation (Fig. 3.14 G-I and Table 3.3), suggesting that other cellular factors might be contributing to the vastly improved interaction observed by IP. Since, the Ubvs in the Grb2.2-Grb2.1 orientation exhibited the best IP efficiency and had improved affinity by SPR, I chose to conduct further intracellular
studies with the Ubv in the Grb2.1-Grb2.2 with the shortest linkage of 12 residues (referred to as Ubv-L12).

Figure 3.14 Kinetic analysis of Ubvs binding to Grb2-SH2
Sensorgrams showing binding to immobilized Grb2-SH2 by serial diluted (A) Ubv-Grb2.1 (B) Ubv-Grb2.2 (C) Ub.wt (D - F) Ubv-Grb2.1-Grb2.2 fusion with a 12, 16, 20 residue linker, respectively (G-I) Ubv-Grb2.2-Grb2.1 fusion with a 12, 16, 20 residue linker, respectively. Ubv concentrations assayed were 50 nM (magenta), 25 nM (cyan), 12.5 nM (navy), 6.3 nM (green), and 3.2 nM (purple), and the black line represents the fit by 1:1 Langmiur binding model.

Since Ubv-L12 had improved Grb2 binding capacity inside cells, and retained the ability to interfere with the interaction between Grb2-SH2 and p-Tyr peptide ligand (Fig. 3.15A), I rationalized that it may have functional effects inside cells. First, I assessed whether Ubv-L12 could block Grb2 from localizing to the cell membrane by virtue of Grb2-SH2 and RTK cytoplasmic tail interactions (Lowenstein, Daly et al. 1992, Rozakis-Adcock, Fernley et al. 1993, Margolis and Skolnik 1994). HEK293T fractionation experiment showed that cells expressing Ubv-L12 had reduced endogenous Grb2 levels in the membrane fraction compared...
to mock transfected cells or cells expressing Ubv.MBP.1 (FIG. 3.15B), indicating that Ubv-L12 binds to Grb2-SH2 in cells and competes with endogenous p-Tyr ligands.

**Figure 3.15 Inhibition of the interaction of Grb2-SH2 and p-Tyr peptide ligand by Ubvs.**

(A) Competition ELISAs to assess inhibition of Grb2-SH2 binding to immobilized p-Tyr peptide by the indicated Ubvs. (B) Effects of the intracellular expression of Ubv-L12 on the levels of Grb2 in the membrane fraction (MEB) of HEK293T cells. WCL indicate whole cell lysate. Na^+K^+ATPase and HSP90 serve as loading control for MEB and WCL, respectively.

Next, I investigated whether Ubv-L12 can antagonize Grb2 mediated signaling in the EGFR pathway. The SH3 domain of Grb2 is known to complex with Sons of Sevenless1, which is a guanine nucleotide exchange factor of the membrane anchored RAS. Upon EGFR stimulation, the EGFR C-terminal tail is phosphorylated and recruits the Grb2:SOS1 complex to the membrane through the Grb2-SH2, and this in turn activates RAS and downstream MAPK pathway (Lowenstein, Daly et al. 1992, Margolis and Skolnik 1994). I reconstituted the EGFR system in HEK293T cells with transfection of a plasmid that expressed EGFR (referred to as 293T-EGFR). 293T-EGFR cells responded to EGF stimulation leading to phosphorylation of EGFR and the downstream kinases MEK and ERK (Fig. 3.16). Grb2 depletion by siRNA led to a reduction in pERK and pMEK (Fig. 3.16), which has been observed in other cell lines (Huang and Sorkin 2005, Modi, Li et al. 2011, Qu, Chen et al. 2014). Grb2 depletion also resulted in a decrease in pEGFR levels. Although, to our knowledge this has never been reported before, it has been reported that knockdown of Grb2 reduces phosphorylation of the RTK, FGFR2 (Lin, Melo et al. 2012, Ahmed, Lin et al. 2013). Interestingly, Grb2 knockdown also reduced total EGFR levels in 293T cells (Fig. 3.16), and this may be a feature of an artificial EGFR system.
introduced in HEK293T cells, as this has not been reported in other cell lines. Under EGF stimulation, cells that expressed Ubv-L12 exhibited a reduction in levels of pEGFR, total EGFR, pMEK and pERK (Fig. 3.16). Reductions in pMEK and pERK levels by Ubv-L12 were not as robust as siGrb2 treatment, but nonetheless, the effects of Ubv-L12 phenocopied those of an siRNA targeting Grb2. Under unstimulated conditions, Ubv-L12 expression reduced levels of pEGFR, total EGFR and pERK similarly to siGrb2 (Fig. 3.16). Taken together, these results show that in cells, Ubv-L12 binds to Grb2, competes with endogenous p-Tyr ligands, and this in turn reduces Grb2 membrane association and antagonizes Grb2 mediated signaling.

**Figure 3.16 Effects of Ubv-L12 on Grb2-mediated EGFR signaling.**

Effects of Ubv-L12 on Grb2-mediated EGFR signaling. Both membrane fraction (MEB) and whole cell lysate (WCL) were isolated from the same samples. Cells were either mock transfected (Mock), transfected with siRNA designed to knock down Grb2 (siGrb2), or transfected with plasmid encoding for GST fusion of Ubv-MBP.1 (MBP.1), or GST fusion of Ubv-L12 (L12). Cells treated with or without EGF treatment are denoted by “+EGF” or “-
EGF”, respectively. Blotting of Na\textsuperscript{+}K\textsuperscript{+}ATPase and HSP90 were used as loading controls for the membrane fraction or whole cell lysate samples, respectively. Superscripts “hi” and “lo” denote high or low exposure for blot development, respectively. Numbers are quantification of signal intensities of bands above and were normalized to mock transfection condition.

3.5 Discussion

I have described the design, construction and validation of a highly diverse combinatorial library that co-opted an expansive surface consisting of 24 residues within the 76-residue Ub protein. Our mutation tolerance scan and anti-FLAG stability selection methods proved to increase display levels substantially. By reducing the proportion of non-displayed variants, we effectively improved the functional diversity of the library, a feature that has been positively correlated with selection outcomes (Kramer, Cox et al. 2003). Furthermore, we have demonstrated that Ubvs derived from my naïve Ub libraries can be used as affinity reagents. Notably, Ubv-Her3.2 was very effective recognizing Her3 in IP experiments and could be used to co-IP binding partners, which could be useful in mapping protein-protein interactions. In addition, I have demonstrated that Ubvs can be expressed in cells to perturb signaling, which is an invaluable tool for uncovering protein function inside cells.

Unlike previously reported Ub libraries, I have made a library that utilizes a restricted diversity biased in favor of Tyr and Ser. Minimalistic Fab and FN3 libraries with loops biased towards Tyr and Ser are well known to form high affinity and specific interactions targeting diverse protein folds (Fellouse, Wiesmann et al. 2004, Fellouse, Li et al. 2005, Fellouse, Barthelemy et al. 2006, Gilbreth, Esaki et al. 2008, Koide and Sidhu 2009). It was thought that the binary diversity was sufficient because flexible loops have additional conformational diversity, which aids in forming complementarity shapes to binding antigen. However, my current study shows that even a β-sheet, which has much reduced conformational flexibility, can still form high affinity interactions with restricted chemical diversity. Whether Tyr residues contribute the bulk of the binding energy in Ubvs, like in the antigen-binding sites of antibodies (Fellouse, Barthelemy et al. 2006), remains to be determined.

Engineered binding proteins lacking disulfides are conducive for intracellular applications, and hold great potential for basic research. While there have been some successes with engineered intracellular Abs (intrabodies), it still remains very challenging due to folding
issues of Abs in the cytoplasmic environment. Attempts to improve intrabody folding have been very time-consuming and relied on trial and error approaches with no standardized procedures (Marschall, Dubel et al. 2015). Others (Mirecka, Hey et al. 2009, Grebien, Hantschel et al. 2011, Parizek, Kummer et al. 2012, Ernst, Avvakumov et al. 2013, Shibasaki, Karasaki et al. 2014) and I have demonstrated the ease of producing functional intracellular binders from alternative scaffolds. Once a high quality scaffold library is obtained, one can have a functional binder in as little as four weeks, given all downstream procedures and assays have been optimized. Due to the specificity of protein binders, there should be a much-reduced chance of off target effects compared to small molecule tools or siRNA for probing protein function. As such, binders derived from alternative scaffolds are ideal for uncovering novel protein functions, as they can be rapidly generated and should have lower off target effects when compared to other tools such as small molecule or siRNA. In addition, in the context of drug target validation, direct protein inhibition or activation by designed binders maybe a superior strategy than siRNA or CRISPR/Cas methods, which rely on abolishing expression of the target protein, and do not accurately mimic drug mediated mechanisms. With constant innovations in technology, it is hoped that the direct delivery of proteins into cells will become a reality in the near future.

Another advantage of alternative scaffolds is that due to their small size, they can be linked in a modular fashion to target different epitopes on the same target to boost affinity, which I demonstrated by the linkage of two Grb2-binding Ubvs. It will also be feasible to fuse Ubvs with differing functions or different targets to produce bi-functional molecules. Notably, heterologous fusion Ubvs can be linked in one genetic fusion owing to their small size, which alleviates the problem of producing mis-paired molecules, like those observed in the production of bi-specific antibodies (De Lau, Heije et al. 1991, Klein, Sustmann et al. 2012).

Finally, as we move further into the post-genomic era, elucidation of protein function will become increasingly more important in the context of disease and drug development. In basic science, intracellular perturbation tools derived from alternative scaffolds can fill the long-standing gap of selectively inhibiting not only enzymes but also protein-protein interactions. In addition to intracellular application, affinity reagents based on alternative scaffolds exceed antibodies in the areas of production and modification for multifunctional purposes. Given that protein folds are highly diverse, and different scaffolds offer different geometries for binding,
the scientific community may benefit from expanding the repertoire of scaffolds to tackle the problem of identifying the tightest and most specific affinity reagent to targets of interest.
4 Conclusions and Future Perspectives

A long-standing goal in protein science has been to understand how the observable characteristics of an interaction (i.e. affinity and specificity) are explained by the molecular features of the proteins involved. Understanding these molecular underpinnings is the foundation for PPI predictions and molecular design, which encompasses both proteins and small molecules. A deeper understanding will continue to help refine current methodologies.

In Chapter 2 of this thesis, I first explored the molecular landscape of Ubv interactions with USP2 and USP21, and also deduced insights into how wt Ub is recognized. In Chapter 3, I utilized Ub as an alternative scaffold and used previously identified molecular insights to design a phage-displayed Ubv library for the isolation of binding molecules to targets of interest. This library was successful in deriving functional binding reagents against numerous protein targets, and two of these targets, Her3 and Grb2, were investigated in detail.

4.1 Ubvs in saturation scanning

Although many structural studies have elucidated the residues that participate in Ub interactions, it still remains unclear how Ub can recognize diverse protein folds using a single surface composed mainly of a β-sheet. In particular, what are the Ub residues responsible for energetically driving interaction and how are they distributed within the binding surface? How do the functional residues differ among interactions to different family folds and how do they differ between interactions to members within the same family? In Chapter 2, I begin to tackle these questions by using phage display saturation scanning to map in detail the binding landscape of enhanced affinity Ubvs to two DUBs, USP2 and USP21. Since Ub.wt binds partner proteins with very low affinity, interactions mediated by Ub.wt are intractable for saturation scanning, as it is difficult to accurately measure substitution associated affinity changes of a weak interaction. As such, engineered Ubvs with few substitutions that recognize binding partners with enhanced affinity and preserve Ub.wt binding mode (Fig. 2.9) serve as ideal substitutes for Ub.wt in saturation scanning.
Using saturation scanning, I assessed the functional importance of each amino acid substitution on the entire Ubv surface that becomes buried when bound to USP2 or USP21. This comprehensive assessment revealed three different categories of residues across the binding surface (Fig. 2.3): residues that were intolerant to mutation and thus were already optimized for binding, residues that exhibited no preferences, suggesting they do not energetically contribute to the interaction, and residues that were highly conserved but favoured residues other than wt, representing positions that are functionally important for binding but not optimized.

From the scanning data, I identified a common functional epitope on Ubvs that is important for binding to their cognate USPs (Fig. 2.3), and this docks onto a USP region (site-2) that is distinct from the active side (site-1). I also proved that site-2 on USP2 is critical for recognition and hydrolysis of native Ub substrates, thus supporting the supposition that Ubvs are viable proxies for Ub.wt and that insights derived from Ubv saturation scanning are applicable to Ub.wt. Additionally, from USP sequence alignments, I determined that the USP site-2 residues are strongly conserved amongst the USP family indicating the Ubv functional core likely makes similar energetic contributions in many other USP interactions. Importantly, I identified a cluster of Ubv residues within the functional core that are different from wt and also different between the two scans (Fig. 2.10, Patch B non-wt cluster). This cluster reflects the different binding preferences of the two investigated USPs and we hypothesize that it can have significant implications in designing specific drugs that target site-2 rather than the highly conserved catalytic site.

This study using Ubvs as proxies for Ub.wt in saturation scanning paves the way for more Ub PPIs analysis. By identifying Ub functional residues for binding to different family folds, we will gain a clearer picture of how Ub uses the same surface to bind diverse protein folds. It is likely that different Ub-binding protein folds (e.g. UBC, UIM and HECT domains), which all recognize Ub, will utilize different functional residues within the overlapping buried surface on Ub for recognition. Identification of hot spots for each Ub interaction will undoubtedly provide valuable information on the epitope that should be targeted for potential therapy. With a complete hot spot database for all Ub binding partners, one can focus drug design on epitopes that are more likely to compete with Ub binding and use structural information to rationally add elements to small molecules to gain affinity and specificity towards a particular Ub-binding protein. In addition, the vast molecular knowledge on Ub
interactions will undoubtedly uncover trends regarding Ub biology that are not observable with few samples. In the context of broader protein science, the vast molecular knowledge will aid the development of computational algorithms for PPI prediction and protein design.

4.1.1 Technical improvements to saturation scanning

Prior to expanding the saturation scan methodologies to other Ub interactions, some technical adjustments should be made to improve the accuracy of the scanning results. As shown in Fig. 2.3 and Fig. 2.4 not all highly conserved substitutions revealed in the scan actually translated to a gain in affinity (e.g. R72W and R42G), and I attributed this to expression or display biases that are intrinsic to phage display technique. To reduce biases, it would be useful to conduct another selection in parallel against anti-FLAG Ab (Ubvs are FLAG tagged within the phage libraries) or an Ab that recognizes the other side of Ub. Sequences enriched in this selection will reflect sequences that are more highly displayed or expressed, and can be subtracted from the saturation scan frequencies to correct for biases. In addition, deep sequencing the phage libraries prior and post selections will ensure the libraries indeed cover the entire sequence space, and give the scanning results more statistical confidence.

4.2 Naïve Ub library

In Chapter 3, I utilized Ub as an alternative scaffold and designed a phage-displayed Ubv library for the isolation of binding molecules to targets of interest. I showed that a minimalistic randomization strategy biased towards Tyr and Ser is feasible for binding site formation on a β-sheet secondary structure. Previous minimalistic libraries (Fab and Monobody) utilized loops for binding surfaces, and it was suggested that the conformational flexibility of loops enables diverse binding interactions to form with limited chemical diversities (Koide, Gilbreth et al. 2007). However, the success of the Ub library presented in this thesis suggests that loop based binding surfaces are not essential for forming interactions to diverse protein topologies, because small scaffolds with restricted conformationally flexible binding surfaces are still capable of using minimalistic chemical diversities to form diverse interactions, suggesting that conformational diversity is not necessary to compensate for restricted chemical diversities in forming binding surfaces to diverse Ags.
Using this naive Ubv combinatorial library, I generated binding proteins to 30 of the 40 proteins tested. Of the 37 proteins that were also subjected to selections with Fab library F, the Ubv libraries performed comparably to Library F as I was able to generate binders to 28 and 26 of targets with Ubv and Fab libraries, respectively. With the characterization of Ubvs that bound to Her3, I proved that Ubvs can be used in place of antibodies in biochemical experiments such as IPs and could potentially be used to map PPI network. In addition, I proved that Grb2-binding Ubvs could be used as intracellular tools to modulate signaling, which is a function that is not easily achieved with antibodies.

### 4.2.1 Ubvs targeting transcription factors

Many of the proteins subjected to Ub library selection within this thesis were DNA binding domains (DBDs). DBDs are found in transcription factors (TFs), and the human genome is predicted to encode ~1900 DNA binding TFs (Vaquerizas, Kummerfeld et al. 2009). TFs are critical to all cellular functions as they control transcriptional regulation of genes. Although Ubvs targeting DBDs were not extensively characterized in this thesis, I would like to dedicate a section to explain how such tools could be applied to uncover TF biology.

Within the field of TF biology, it has been difficult to accurately describe DNA binding specificities of DBDs. Many HTP techniques, such as chromatin immunoprecipitation sequencing (ChIP-Seq), have been developed to identify binding motifs (Geertz and Maerkl 2010, Ogawa and Biggin 2012). However, a TF can exhibit different DNA binding modes as binding is often complicated by combinatorial factors such as having multiple DBDs linked within a TF, and consequently DNA binding can be mediated by a single DBD or combinations of DBDs within a single TF (Siggers and Gordân 2014). This is exemplified by the Oct1 TF, which has two DBDs, a Pit-Oct-Unc (POU) homeodomain (POU_{HD}) and a POU-specific domain (POU_{S}), and can singularly or combinatorially bind to a total of three different DNA motifs (Verrijzer, Alkema et al. 1992, Klemm, Rould et al. 1994). Additionally, many TFs are capable of homodimerization or heterodimerization with other TFs, which can alter TF binding specificity (Siggers and Gordân 2014). While it is possible to identify the existence of different DNA recognition modes from distinct binding motifs generated from HTP methods, it is possible that differences can go undetected due to technical limitations, and even if detected, the data do not provide insights into what contributes to the different binding modes.
One method to dissect DBD contributions to TF binding is to intracellularly express a designed Ubv that inhibits a DBD from binding DNA. By comparing ChIP-Seq generated binding motifs of cells expressing protein inhibitors of different DBDs of a TF, one can determine the DNA binding contributions from individual DBDs. In addition, expression of a protein inhibitor that recognizes homodimerization or heterodimerization surfaces that can disrupt TF complex formation can shift the DNA binding specificity to that of the monomeric state. Application of affinity reagents that target and inhibit DBDs may also shed light on the functional roles of each DBD in cellular processes by enabling observation of phenotypic outcomes mediated by selectively manipulating an individual DBD. While ChIP-seq of genetically engineered TFs with missing DBDs or mutations that disrupt homodimerization or heterodimerization can in principle be used to answer the same questions, protein inhibitors have the advantage of modulating endogenous TFs. Engineered TFs may have compromised folds or may not express at endogenous levels and thus might not provide an accurate depiction of binding specificities of native TFs.

### 4.2.2 Validation of TFs as drug targets

Many TFs, including JUN, MYC and p53, have long been recognized as critical and central players in cancer and immune diseases, as TFs are often converging points of pathogenic signaling (Bhagwat and Vakoc 2015, Fontaine, Overman et al. 2015, Lazo and Sharlow 2016). However, just like proteins that function by PPIs, TF has been considered to be intractable for drug intervention due the large binding surfaces that lack obvious small molecule binding pockets (Fontaine, Overman et al. 2015, Lazo and Sharlow 2016). Although possible, it remains difficult to develop small molecules to disrupt PPI or protein-DNA interactions (Scott, Bayly et al. 2016), and thus it may be beneficial to first apply TF-binding Ubvs to determine if disruption of a particular TF interaction will truly lead to a desired outcome in *in vitro* disease models before dedicating resources and time to the development of small molecules. TF modulation by Ubvs would be particularly useful for cases where overexpression of a TF has generated conflicting results in both support and repression of malignancy (Sporn and Liby 2012). In addition, if expression of a Ubv that inhibits a particular DBD does indeed result in therapeutic benefit, then it may be useful to solve the structure of the Ubv-DBD
complex. The structure can offer clues as to what epitope on a DBD should be targeted by small molecules and the Ubv residues responsible for engaging DBD maybe used as a template for drug design.

While most TFs are difficult to drug, one class of TFs that have been successfully targeted for therapy is the nuclear hormone receptor (NHR) family. In addition to DBDs, NHRs also have ligand binding domains that are conducive to small molecule binding (Bhagwat and Vakoc 2015). However, resistance to drugs targeting NHR frequently develop in patients (Lewis and Jordan 2005, Joseph, Lu et al. 2013), and thus much effort has been devoted to developing combinatorial therapies that can target multiple domains of a NHR. In this scenario, inhibitors of the DBDs of NHR DBDs may validate the DBD as a viable targeting domain and provide structural insights for drug design.

4.3 Future of affinity reagents based on alternative scaffolds

Affinity reagents based on alternative scaffolds have many benefits over conventional Ab based reagents. First, current Ab reagents are typically derived from animal sources and do not have a known DNA coding sequence. This is problematic because production of animal sourced Abs rely on hybridoma technology, which involves indefinitely maintaining the Ab producing B cells by fusion to myeloma immortal cell lines (Kohler and Milstein 1975). Not only is hybridoma production cost intensive but it also suffers from issues inherent to cell culture propagation such as loss of expression over time, mutations and spurious expression of other Ab genes (Zack, Wong et al. 1995, Bradbury and Plückthun 2015). As a result, the identity of Abs produced from hybridomas are difficult to confirm and this often results in Ab batch-to-batch variations, which contributes to data irreproducibility. Since alternative scaffolds are synthetically engineered usually by display technologies, these binders are recombinant and by default have a known sequence. Therefore, production does not require indefinitely maintaining the cells producing these binders, and identification of a recombinant binder can be easily confirmed by comparing to its known sequence, which helps to keep production
consistent. Secondly, alternative scaffolds are carefully chosen or more recently, can be created *de novo* to be small modular proteins with superior biophysical properties (Khoury, Smadbeck et al. 2014, Woolfson, Bartlett et al. 2015). As such, recombinant binders based on alternative scaffolds can be produced much more cost effectively than Abs. While synthetic Abs developed using display technologies are becoming more prominent and can bypass the requirement for hybridoma production, downstream production of synthetic Abs still necessitates expensive and resource intensive tissue culture methods. With the increased emphasis on reducing data irreproducibility due to Ab reagent inconsistencies, and the declining funding climate, the use of alternative scaffolds will likely rise.

Currently, the intracellular usage of engineered binding proteins is underappreciated, with only a handful of studies published. Most binding protein engineering projects have been directed towards targeting surface receptors for therapeutic purposes, and this is intuitively justified as investments into such ventures provide a larger financial incentive. Many published papers regarding the intracellular use of affinity reagents have been proof of principle studies that demonstrated the possibility to achieve expected functional outcomes by benchmarking to other established molecular tools, such as sh/siRNA (Amstutz, Binz et al. 2005, Mirecka, Hey et al. 2009, Ernst, Avvakumov et al. 2013). However, as we move further into the post-genomic era, the elucidation of protein function rather than gene function will become more important for understanding cell physiology, and the need for protein modulation tools will rise.

As more scientists dedicated to developing affinity reagents continue to publish work dealing with creative usage of alternative scaffolds in uncovering new protein function, the adoption of these binders will most likely expand. To date, binders derived from alternative scaffolds have been used to selectively inhibit protein domain function (Sha, Gencer et al. 2013) or activate proteins (Zhang, Wu et al. 2016), trace proteins in live cells without the need of fusing fluorescent proteins to target proteins (Riedl, Flynn et al. 2010, Burgess, Lorca et al. 2012, Zolghadr, Gregor et al. 2012), and identify novel epitopes for drug targeting (Grebien, Hantschel et al. 2011, Gorelik, Orlicky et al. 2016). Currently, no other tools can parallel the intracellular utility of engineered protein binders. I envision that the usage of intracellular
protein binders will be a routine methodology in cell biology in the future, just as genetic knock down techniques have evolved to become over the years.

I support the development of more scaffolds that utilize secondary structural elements for binding surface formation. This is because loop based binding surfaces are more conformationally flexible and thus are likely to pay a higher entropic cost when bound to target proteins, which equates to lower affinity. In addition, flexibility can result in non-specific interactions and thus affinity maturation procedures are usually needed to enhance affinity and specificity. It is possible that a single alternative scaffold library using a fixed structural binding site may not perform well against all protein folds or may not be able to target specific epitopes owing to geometric incompatibilities. Thus, it would be wise to develop a collection of scaffolds with varying binding topologies (i.e. concave, convex, or flat) based on different structural elements (i.e. helices, beta-sheets or combinations of), and the entire collection may be pooled together to select for binders against a protein of interest. This will enable competition between different scaffolds for the selection of the best binders. However, precautions must be taken so that certain scaffold libraries will not dominate selection outcomes due to expression or display biases, leading to wrongly identifying enriched binders as the ones possessing the tightest affinity or other desirable qualities. First, such a library collection must be made within the same plasmid vector so that expression and identification of binding variants are standardized for all libraries. Second, the abundance of each scaffold library may need to be adjusted for the proportions of members that are actually displaying a protein variant in each library. A library with all of its members displaying a protein is more likely to dominate a selection outcome compared to a library in which only a fraction of its members are displayed, regardless of the quality of the derived binders. As selection successes accumulate, one might be able to predict which libraries will perform better for particular target proteins.
5 References


