Computational design of protein-protein interactions to engineer novel inhibitors of the p53 pathway and the polycomb repressive complex

James D. Moody

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

David Baker, Chair

Ning Zheng

Dustin Maly

Program Authorized to Offer Degree:

Molecular and Cellular Biology

©Copyright 2014
James D. Moody

University of Washington

Abstract

Computational design of protein-protein interactions to engineer novel inhibitors of the p53 pathway and the polycomb repressive complex

James D. Moody

Chair of the Supervisory Committee:
Professor David Baker
Biochemistry

Here I present 2 examples of the application of the Rosetta macromolecular modeling software suite to the successful development of novel protein-based inhibitors with the aim of creating useful tools for the greater cancer research community. Rosetta was used to engineer the first examples of high affinity, highly specific inhibitors of Mdmx that are useful in mammalian cells. These inhibitors bind either Mdmx or Mdm2 with sub-nanomolar affinities and specificities for one or the other estimated at 1000 to 10,000-fold. In mammalian cancer cell lines, these inhibitors retain their affinities and specificities, halt the cell cycle, and trigger apoptosis. The x-ray crystal structures of 3 of these Mdmx inhibitors were determined and validate the design models and methodology. Rosetta was also employed to engineer high affinity inhibitors of the EED-Ezh2 interaction. These inhibitors bind a large, charged surface of EED with sub-nanomolar affinities.

Table of Contents

Acknowledgements……………………………………………………………………………… 7
Introduction…………………………………………………………………………………..... 9

Section 1: Computational design of proteins to inhibit the interaction of Mdmx-p53

1.1 Background and Motivation……………………………………………………………... 12
1.2 Differences between Mdmx and Mdm2 and design strategy………………………………. 13
1.3 Experimental validation and characterization of MB17……………………………………. 15
1.4 Binding mode validation and affinity and specificity maturation of MB17………………... 17
1.5 Crystal structures of MB17 and variants and comparisons to models……………………… 23
1.6 MB17 and variants retain affinity and specificity in mammalian cells…………………….. 28
1.7 MB17 variants halt the cell cycle and trigger apoptosis in colon cancer cell lines………… 30
1.8 Balancing the charges on the surface of top MB17 variants……………………………….. 34
1.9 Conclusions and future directions………………………………………………………….. 38

Section 2: Computational design of proteins to inhibit the interaction of EED-Ezh2

2.1 Background and Motivation……………………………………………………………... 42
2.2 Design strategy……………………………………………………………………………... 43
2.3 Experimental validation and characterization of EB15 and EB22…………………………. 46
2.4 Binding mode validation and affinity maturation of EB15 and EB22……………………... 48
2.5 Conclusions and future directions………………………………………………………….. 54

Section 3: (Appendix) Supplemental data for the computational design of proteins to inhibit the interaction of Mdmx-p53

3.1 MB17 single site mutation enrichment heatmaps…………………………………………... 55
3.2 Yeast surface display and ITC affinities for Mdmx binders and variants………………….. 58
3.3 Scaffolds and amino acid sequences of Mdmx binders and variants………………………. 63
3.4 MB17 combinatorial library positions and mutations……………………………………… 71

Section 4: (Appendix) Supplemental data for the computational design of proteins to inhibit the interaction of EED-Ezh2

4.1 EB15 and EB22 single site mutation enrichment heatmaps……………………………….... 73
4.2 Yeast surface display and BLI affinities for EED binders and variants……………………. 74
4.3 Scaffolds and amino acid sequences of EED binders and variants……………………….... 76
4.4 EB15 and EB22 combinatorial library positions and mutations…………………………… 79

Section 5: (Appendix) General computational methods

5.1 Fastdesign with atompair constraints……………………………………………………….. 81
5.2 Automatically identifying positions of key interacting residues…………………………… 83

Section 6: (Appendix) Computational methods specific to the design of proteins to inhibit the interaction of Mdmx-p53

6.1 Side-chain grafting with backbone flexibility…………………………………………… 86
6.2 Filtering for clashes with Mdm2 and multi-state design…………………………………… 88
6.3 Homology modeling…………………………………………………………………... 90
6.4 Automated design of Kunkel oligonucleotides…………………………………………... 90
6.5 Illumina sequence data processing of MB17 SSM pools………………………………... 92
6.6 Balancing the surface charges of the top MB17 variants………………………………... 98

Section 7: (Appendix) Computational methods specific to the design of proteins to inhibit the interaction of EED-Ezh2

7.1 Computational methods: Design of EED binding proteins……………………………... 101
7.2 Redesign of EB22 loop………………………………………………………………. 115
7.3 Illumina sequence data processing of EB15 and EB22 SSM pools……………………….. 116

Section 8: (Appendix) General experimental methods

8.1 Yeast Colony PCR……………………………………………………………………… 122
8.2 E. coli colony PCR……………………………………………………………………… 126
8.3 Chemically competent yeast transformation……………………………………………. 129

Section 9: (Appendix) Experimental methods specific to the design of proteins to inhibit the interaction of Mdmx-p53

9.1 Biotinylation and size-exclusion chromatography of Mdmx and Mdm2…………………. 132
9.2 Small volume yeast surface titration of Mdmx binders…………………………………… 134
9.3 Kunkel mutagenesis………………………………………………………………….. 140
9.4 Labeling MB17 SSM library for FACS selection………………………………………. 149
9.5 Extraction and high-throughput sequencing of MB17 gene pools from yeast……………. 156
9.6 MB17 Combinatorial library construction via ultramer assembly………………………... 166
9.7 Overlap PCR mutagenesis and assembly………………………………………………….. 172

Section 10: (Appendix) Experimental methods specific to the design of proteins to inhibit the interaction of EED-Ezh2

10.1 Biotinylation and size-exclusion chromatography of EED3…………………………….. 177
10.2 Small volume yeast surface titration of EED binders……………………………………. 179
10.3 Overlap PCR mutagenesis and assembly………………………………………………… 185
10.4 Labeling EB15 and EB22 libraries for FACS…………………………………………. 189
10.5 Extraction and high-throughput sequencing of EB15 and EB22 gene pools from yeast… 193
10.6 EB15 and EB22 combinatorial library construction via ultramer assembly……………... 202
10.7 Sub-cloning, expressing, and purifying top EB15 and EB22 combinatorial variants…... 207

Section 11: References

11.1 References for the introduction………………………………………………………….. 213
11.2 References for the design of proteins to inhibit the interaction of Mdmx-p53…………... 215
11.3 References for the design of proteins to inhibit the interaction of EED-Ezh2…………... 220
11.4 References for computational methods…………………………………………………... 221
11.5 References for Experimental methods…………………………………………………… 222

Acknowledgements

I first met David Baker as a graduate school recruit the first time I applied to graduate school. I wasn't accepted by any of the 3 schools at which I interviewed, so I resolved to try again the following year. I contacted David to ask about what I might improve on in the intervening year. To my surprise, he needed a laboratory technician position in his laboratory and offered me the position, which I gratefully accepted. The following year, I was accepted at all 8 of the schools at which I interviewed, a feat I surely owe to my experiences as a lab tech in the Baker laboratory. I thank David for providing this singular opportunity, because without his help, I might not have attended graduate school, or at least not at a school so prestigious as the University of Washington. I thank David for providing a challenging and enriching research environment where I could learn and apply both the computational and experimental sides of protein design. I also thank David for excellent mentorship and enlightening discussions.

I thank the members of my Doctoral Advisory Committee, Barry Stoddard, Roland Strong, Ning Zheng, and Dustin Maly for taking time out of their busy schedules to provide insightful advice and helpful discussions about various aspects of my research projects.

I thank a large number of former and current members of the Baker lab for advice and tutelage in the art of protein design. Rocco Moretti, Javier Castellanos, Tom Linsky, Possu Wang, David La, Yifan Song, Sarel Fleishman, TJ Brunette, Shawn Yu, David Kim, Darwin Alonso, Will Sheffler, Alex Ford, Yu-Ru Lin, Chris King, Gustav Oberdorfer, Fabio Parmeggiani, Lei Shi, Srivatsan Raman, Florian Richter, Frank DiMaio, Sagar Khare, Firas Khatib, Nobuyasu Koga, and Lucas Nivon are among those I thank for guidance in computational matters. Michelle Scalley-Kim, Aaron Chevalier, Erik Procko, Eva Strauch, Christy Tinberg, Luke Dang, Cassie Bryan, Chris Bahl, Matt Bick, Yakov Kipnis, Jeremy Mills, Rijhu Das, Erik Althoff, Justin

Siegel, Matt Smith, Jacob Corn, Niel King, Ling Wang, Wenxin Hu, and Tim Whitehead, are among those I thank for guidance in experimental matters.

Lastly, and most of all, I thank Amelia, my best friend and sweetheart, who has put up being married to a graduate student for the past 5 years, all while raising 2 rambunctious redheaded little girls and managing an apartment complex. May her many sacrifices be rewarded.

Introduction

The first time that I saw a ribbon diagram of a protein in my molecular biology textbook, I naively thought that proteins looked like ribbon diagrams and marveled that something so simple and elegant could be the foundation of all life of Earth. Later I realized that a ribbon diagram was only a representation of a protein backbone and that in fact protein atoms also fill all of the space between the ribbons. Since that time, I have imagined designing proteins as a civil engineer might design a bridge. Over time I have come to know that designing proteins is far more difficult than designing a bridge, due in part to the fact that as yet we cannot perfectly simulate all of the forces responsible for protein shape and function (1, 2). Proteins consist of long chains, each unit of which normally can be 1 of 20 unique building blocks and as such, proteins can access a vast sequence space. If we consider a relatively small protein of 100 amino acids, then the total number of unique amino acid sequences available is 20^100 = 1.27*10^130, a staggering number. As large as this number is, computational methods can often, given a backbone conformation, identify the best combination of amino acid identities to allow the protein to stably adopt that backbone conformation (3). Proteins are also highly flexible molecules rich in rotatable bonds and as such, can access to an even more vast conformational space. Let us assume that if we correctly sample the backbone conformation of a given protein of known sequence, we can then correctly sample the conformations of all its side chains. If we once again consider a protein of 100 residues of known sequence and sample the backbone Φ and ψ dihedral angles at 1° increments, then we must generate and evaluate 360^200 = 1.82*10^511 different backbone conformations in order to identify conformations close to that adopted by such a protein in reality. Computational methods can only sample a small fraction of that vast conformational space. One way around this is to predict the secondary structure of a

protein from its amino acid sequence and bias the backbone conformational sampling to protein fragments of known structure that have similar amino acid sequences and secondary structures (4).

Natural selection has evolved a large and diverse universe of proteins to carry out a wide range of useful biological functions, necessary for all life on this Earth. More and more, however, we encounter new problems for which no natural protein is currently known, encouraging us to engineer novel proteins to fill the need. The traditional route to engineer new proteins has been directed evolution. Directed evolution involves creating a large number of variants of a starting point protein and screening the variants for those having the desired properties. Directed evolution requires that a screening method be available with sufficient throughput to handle the desired number of variants. This works well for some target properties, such as a protein that binds to another protein, but less well for other target properties, such as an enzymatic activity that cannot be measured in a high-throughput manner. Even in the most optimal of cases, directed evolution techniques can only sample libraries comprised of no more that 10^15 different variants, such as in mRNA display techniques (5). If we remember the relatively small protein of 100 amino acids, the total number of unique amino acid sequences available to it, 1.27*10^130, is over 1*10^115 times greater than our ability to sample. Current directed evolutions techniques side step this problem by creating libraries were only a subset of the residues in a protein are varied (6). There are impressive examples of the evolution of complicated enzymatic activities from starting point proteins that have no such activity, however, the starting point proteins were relatively small (7, 8). If we hope to engineer larger protein tools, we must rely on computational modeling techniques to trim down the number of different amino acid sequences that must be tested. Although computational protein design in in its infancy,

numerous impressive examples of computationally designed proteins have been published, including: novel protein cages (9, 10), digoxygenin binding proteins (11), BHRF1 binding proteins (12), IgG constant region binding proteins (13), metal-binding proteins (14), lysozyme inhibitors (15), proteins capable of binding and neutralizing Influenza hemagglutinin (16), proteins capable of eliciting anti-HIV antibodies (17), Kemp elimimases (18), Diels-Alderases (19), retro-aldolases (20), enzymes capable of catalyzing the Morita-Baylis-Hillman reaction (21), ester hydrolases (22), organophosphate hydrolases (23, 24, 25), and numerous idealized protein folds (2).

The Rosetta macromolecular modeling software suite can be alternately viewed as a challenging tool to be developed or as a powerful tool to be applied to the solution of difficult problems. I set out to use Rosetta to solve such problems, aiming to engineer proteins to solve real-world problems and hence be of use to the greater scientific community. I present 2 examples of difficult problems facing the cancer community that I will attempt to solve using computationally designed proteins. The first is the creation of highly specific, high affinity inhibitors of the p53 regulator Mdmx and the second is the creation of high-affinity inhibitors of the EED-Ezh2 interaction. As one reviews my development process, it may occur to the reader that experiments of interest have been ignored. One should remember that I approached these challenging problems as a protein engineer and that my primary goal here was to generate protein-based tools that other scientists could use toward the understanding and eventual eradication of cancer.

Section 1

Computational design of proteins to inhibit the interaction of Mdmx-p53

1.1 Background and Motivation

The p53 pathway is disrupted in virtually every human tumor, allowing cancer cells to evade apoptosis (26). In about 50% of tumors, p53 is mutated or deleted. If p53 is intact, often expression of p53 inhibitors Murine Double Minute 2 (Mdm2) or Murine Double Minute 4 (Mdmx) is up-regulated (27, 28). Small-molecule drugs such as Nutlin3a selectively inhibit Mdm2, but are not effective against Mdmx (29, 30). The last several years have seen considerable effort by the cancer pharmacology community to develop small molecule inhibitors of Mdmx and Mdm2, but thus far, highly potent, specific inihibitors of Mdmx have not been reported (31, 32). There has been some success in the development of small molecules that bind Mdmx less tightly, but such also bind Mdm2 (33). There are also stapled p53 peptides that bind Mdmx in the low nanomolar range while also binding Mdm2 (34).

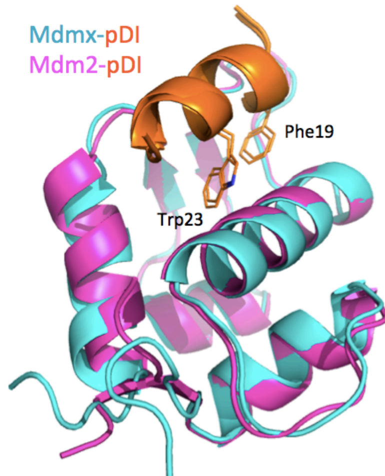
The individual contribution of Mdmx and Mdm2 to p53 regulation in healthy and cancer cells are not clearly understood (34). A selective Mdmx inhibitor may be able to provide such clarification. The only highly specific Mdmx inhibitor reported to date is the VH9 antibody (reported to be 140-fold specific for Mdmx over Mdm2) (35). Unfortunately, VH9 is not useful within the reducing environment of mammalian cells. We surmised that a computationally designed small protein could achieve increased specificity for Mdmx by making additional contacts with Mdmx around the rim of the p53-binding cleft. We also surmised that such a protein could additionally achieve increased affinity for Mdmx through such additional contacts and by stabilizing the p53 transactivation helix (the region of p53 bound by Mdmx and Mdm2) in a rigid, preformed, helical state. To that end, we sought to computationally design a small

protein analog of the p53 trans-activation helix (hereafter termed p53) that would both demonstrate increased affinity and specificity for Mdmx over Mdm2 and be useful inside mammalian cells. We hoped such a binding protein would better elucidate importance of the Mdmx-p53 interaction in normal and cancer cells and possibly inform structure-based drug design efforts toward cancer therapeutics.

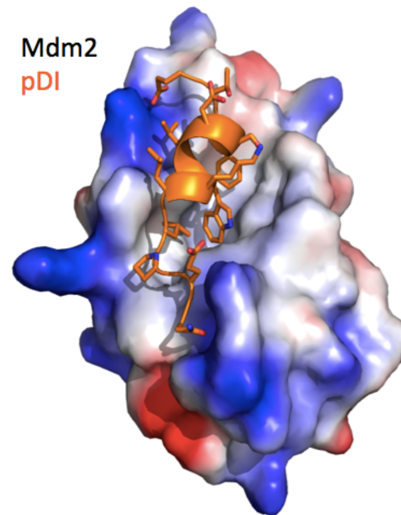
1.2 Differences between Mdmx and Mdm2 and design strategy

The p53-binding domains of Mdm2 and Mdmx (hereafter termed Mdm2 and Mdmx, respectively) share a 54% sequence identity and a protein backbone RMSD of 0.7 Å (**Figure 1.2.A**). In general Mdm2 is more positively charged than Mdmx, with net charges of +5 and +1, respectively. In particular, the region of Mdm2 that makes up half of the rim of the p53 binding pocket (residues 65-104) features the amino acid sequence $Rx_3EKQx_{25}RxKx_2T$ with a net charge of +3 (**Figure 1.2.B**). The structurally homologous region of Mdmx (residues 64-103) features the amino acid sequence $Qx_3QQEx_{25}SxPx_2D$ with a net charge of -2 in Mdmx (**Figure 1.2.C**). Phenylalanine 55, located on the rim of the p53-binding cleft in Mdm2, is substituted with Histidine in Mdmx (**Figure 1.2.D**). Superposition of available Mdmx and Mdm2 crystal structures revealed that the C-terminal helix of the Mdm2 p53-binding domain protrudes approximately 3.5 Å further into solvent than does the corresponding helix of Mdmx due to a Proline to Histidine substitution at Mdm2 residue 96 and a Proline to Lysine substitution at Mdm2 residue 98 (**Figure 1.2.E**)(30, 33-75). While this C-terminal helix of Mdm2 is frequently observed at crystal contacts due to the presence of two surface-exposed Tyrosine residues thereon, 4 available NMR structures of Mdm2 confirm that the position of the helix in Mdm2 is not a crystallographic artifact (64, 76-78).

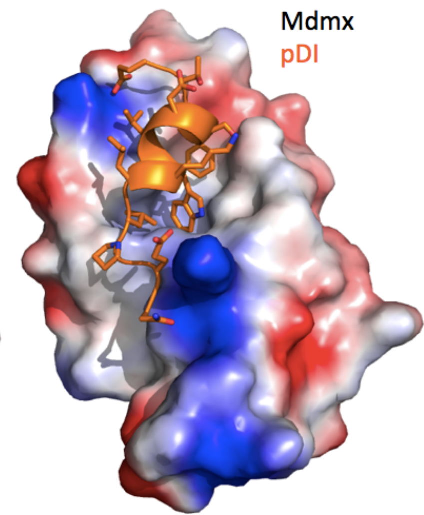
1.2.A.



1.2.B.



1.2.C.



1.2.D.

1.2.E.

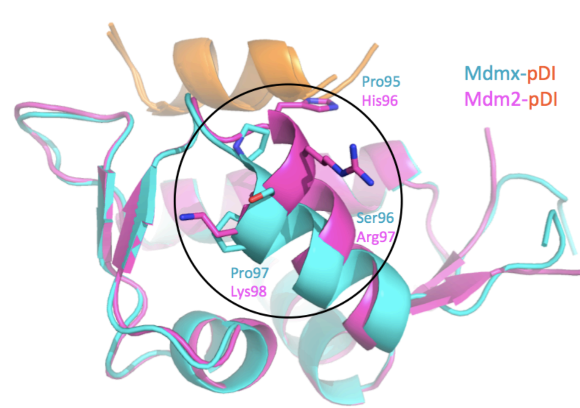


Figure 1.2: Mdmx and Mdm2 are nearly identical with subtle differences.
1.2.A. Superposition of the crystal structures of Mdmx (cyan) and Mdm2 (magenta) bound to pDI (orange) showing the protein backbone in cartoon representation and the key p53 binding residues as sticks.
1.2.B. pDI (shown as an orange cartoon with side chains as sticks) bound to Mdm2. Mdm2 is shown as an electrostatics map with regions of positive charge shown in blue, regions of negative charge shown in red, and regions of neutral charge shown in white.
1.2.C. pDI (shown as an orange cartoon with side chains as sticks) bound to Mdmx. Mdmx is shown as an electrostatics map with regions of positive charge shown in blue, regions of negative charge shown in red, and regions of neutral charge shown in white.
1.2.D. Superposition of the crystal structures of Mdmx (cyan) and Mdm2 (magenta) bound to p53 (orange) showing the protein backbone in cartoon representation and side chains as sticks. The Phenylalanine 55 to Histidine substitution is highlighted with a black circle.
1.2.E. Superposition of the crystal structures of Mdmx (cyan) and Mdm2 (magenta) bound to p53 (orange) showing the protein backbone in cartoon representation. The protruding helix of Mdm2 is highlighted with a black circle.

To identify scaffold proteins capable of hosting Phenylalanine 19 and Tryptophan 23 from p53 close to their natively observed positions relative to Mdm2 and Mdmx, we used the PlaceStub protocol within the RosettaScripts application of Rosetta as previously described (79-81). We filtered models output from RosettaScripts for those that placed a beta-carbon within 6.5 Å of the Proline beta-carbon at residue 95 of Mdmx and visually inspected passing models for those likely have significant clashes with Mdm2 but not Mdmx. We selected 14 candidate designs and, using Foldit (82), carried out simultaneous manual multi-state design against Mdmx (PDB ID 3jzp)(38) and Mdm2 (PDB ID 1ycr)(62) to identify mutations predicted to have a positive or neutral effect on binding to Mdmx but a neutral or negative effect on Mdm2 binding. Specifically, mutations were chosen which introduced charge repulsion and/or steric clashes with Mdm2.

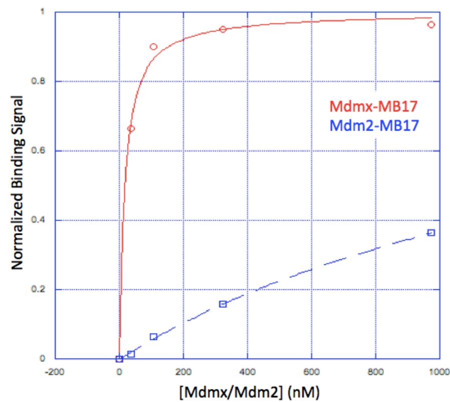
1.3 Experimental validation and characterization of MB17

We synthesized the genes for each of the candidate designs and controls using a commercial vendor (IDT, San Jose, California, USA, Gen9, Cambridge, Mass, USA, or Genescript, Piscataway, New Jersey, USA) and cloned these into pETCON, a modified version of the yeast display vector pCTCON2 (83). We carried out yeast display as previously described and titrated soluble biotinylated Mdmx and Mdm2 against the designed proteins displayed on the yeast surface under non-avid conditions (83). We initially identified a single design, termed Mdmx Binder 17 (MB17) that appeared to be 75-fold specific for Mdmx over Mdm2, binding Mdmx with an apparent dissociation constant (Kd) of 22 ± 6.7 nM and Mdm2 with a Kd of 1500 ± 340 nM by yeast surface titration (YST)(**Figure 1.3.A**)(83). YST data was processed using Flowjo (Tree Star Inc, Ashland, OR, USA) and fit using either Kaleidagraph (Synergy

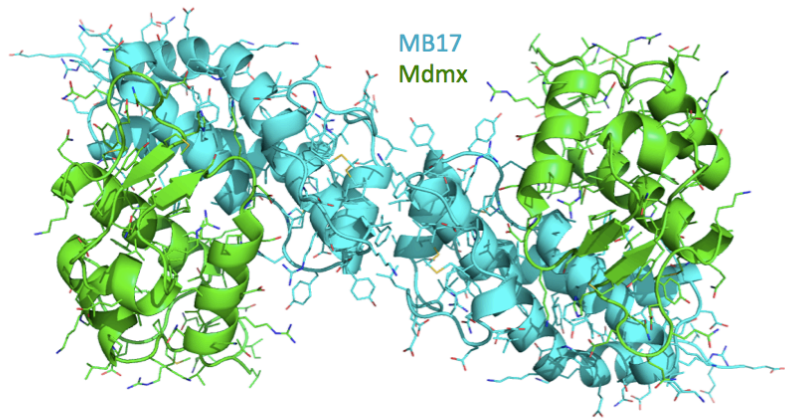
Software, Reading, PA, USA) or a Java nonlinear least squares algorithm (http://statpages.org/nonlin.html). Isothermal titration calorimetry (ITC) revealed that purified MB17 is 97-fold specific for Mdmx over Mdm2, binding Mdmx with a Kd of 39 nM and Mdm2 with a Kd of 3800 nM. The fact that they do not bind many of the other designed binders when added to yeast cells at a concentration of 1 uM suggests that our soluble biotinylated Mdmx and Mdm2 preparations are neither nonspecific binders nor misfolded. The antibody VH9 was previously shown to be 140-fold specific for Mdmx over Mdm2, we further validated our biotinylated Mdmx and Mdm2 preparations by confirming that VH9 displayed in the surface of yeast binds Mdmx more tightly than Mdm2 (35).

MB17 is based on the dimeric Enterobacteria T7 phage protein OCR, a type I nuclease inhibitor (PDB ID 1s7z)(84). As such, MB17 is presumably also a dimer capable of simultaneously binding 2 molecules of Mdmx (**Figure 1.3.B**). This hypothesis supported by size-exclusion chromatography (SEC) that gives an estimated molecular weight of 22.1 kDa for MB17 (**Figure 1.3.C**). The calculated molecular weight for the MB17 monomer is roughly half that, at 13.2 kDa. MB17 was designed to discriminate between Mdmx and Mdm2 by disfavoring binding to Mdm2 in 3 different ways: 1. MB17 places a helical region so as to introduce steric clashes with the protruding helix of Mdm2. 2. MB17 positions positively charged amino acids near the more positively charged rim of the Mdm2 p53-binding pocket that should repel the positive charges thereon. 3. MB17 surrounds Mdm2 Phenylalanine 55 with charged groups that should disfavor binding to the Phenylalanine in favor of the Histidine of Mdmx (**Figure 1.3.D**).

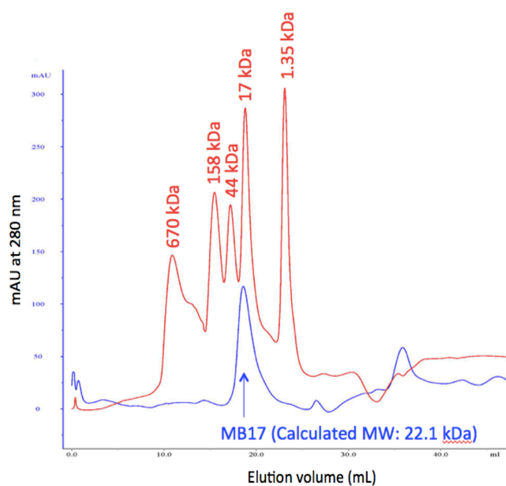
1.3.A.



1.3.B.



1.3.C.



1.3.D.

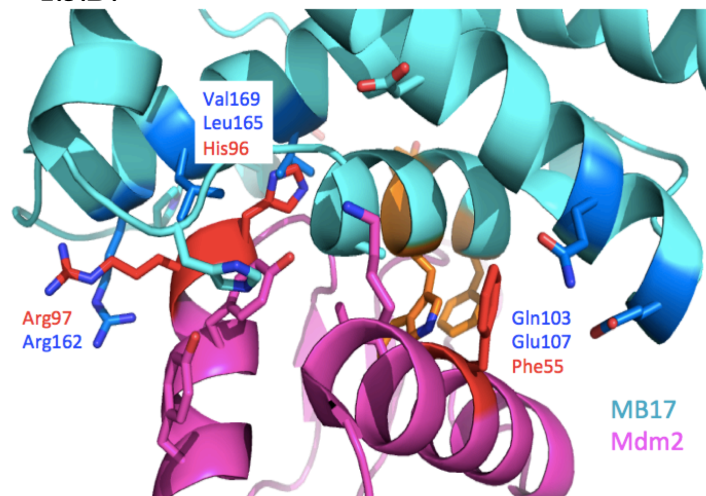


Figure 1.3: Initial Characterization of MB17.

1.3.A. Representative normalized yeast surface titration binding curves of displayed MB17 labeled with soluble Mdmx (red) or Mdm2 (blue).

1.3.B. Initial design model of MB17 dimer (cyan) binding 2 molecules of Mdmx (green) shown with the protein backbone as a cartoon and side chains as sticks.

1.3.C. SEC traces of molecular weight standards (red) or purified MB17 (blue).

1.3.D. Close-up of model of MB17 (cyan) bound to Mdm2 (magenta). MB17 residues predicted to disfavor binding are colored blue while Mdm2 residues predicted to disfavor binding are colored red.

1.4 Binding mode validation and affinity and specificity maturation of MB17

We sought to validate the folded state and binding mode of MB17 and improve the affinity and specificity of MB17 for Mdmx through deep mutational scanning as previously described (85, 86). We also sought to determine whether the specificity of MB17 for Mdmx was partially encoded in the backbone conformation of MB17 or was merely a product of the residue

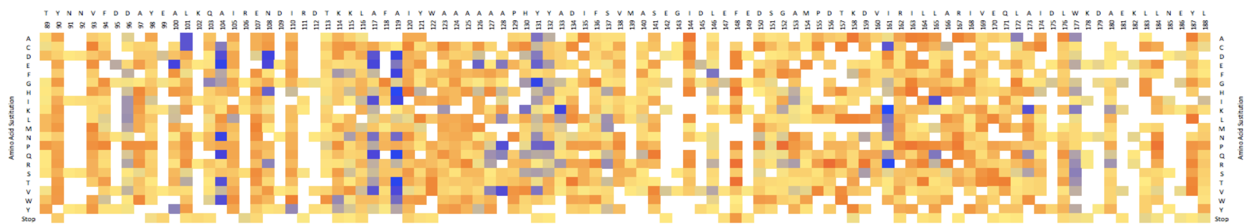
identities at the interface of MB17 and Mdmx. To that end, we investigated whether mutations to the surface of MB17 could reverse its specificity (becoming specific for Mdm2 instead of Mdmx). Using Kunkel mutagenesis (87), we individually mutated to all possible amino acids each of the 80 MB17 residues nearest the interface with Mdmx (1680 unique variants). We displayed the resulting library on the surface of yeast and screened it using fluorescent-activated cell sorting (FACS)(83, 88) under 4 different selection conditions for 4 rounds each: 1. Three-color labeling to simultaneously sort for maximal Mdmx affinity and maximal Mdmx selectivity over Mdm2, 2. Three-color labeling to simultaneously sort for maximal Mdm2 affinity and maximal Mdmx selectivity over Mdmx, 3. Two-color labeling to sort for maximal Mdmx affinity, or 4. Two-color labeling to sort for maximal Mdm2 affinity.

We extracted the plasmid DNA from cells saved from the unselected library and from successive rounds of sorting against each of the 4 selection conditions and subjected the pools of MB17 genes to high-throughput next-generation sequencing on an Illumina Miseq (Illumina, San Diego, CA, USA). The fitness (enrichment ratio) of a given mutation under a given selection condition was calculated as the ratio of the frequency of that mutation in the selected pool relative to the frequency of that mutation in the pool selected only for display on the yeast surface. Under all selection conditions involving Mdmx, core MB17 residues and MB17 residues modeled at the binding interface with Mdmx were strictly conserved, validating that MB17 is both properly folded and binds in the modeled binding mode (**Figure 1.4.A**). While selection conditions involving Mdm2 also strictly conserved MB17 residues modeled at the binding interface with Mdm2, core positions frequently saw enrichment of mutations not consistent with the modeled conformation of MB17, suggesting that if unfolded, MB17 becomes more specific for Mdm2. This is not surprising since the key binding residues of MB17 comprise a linear

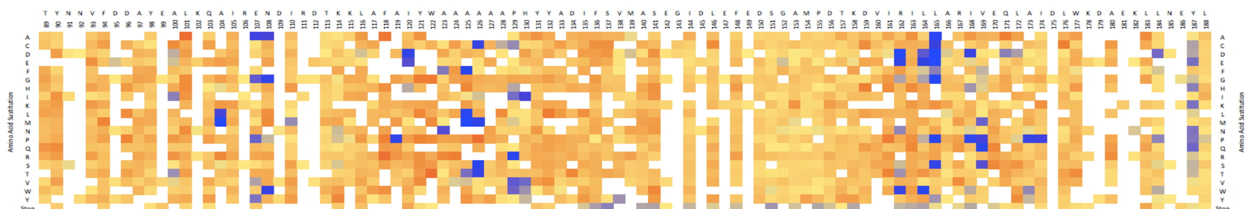
epitope that should retain binding even if the pseudo-three-helix bundle of MB17 adopted instead an extended partially helical peptide conformation as observed for the native p53 peptide. As an extended partially helical peptide would also bind Mdmx, these results suggest that a folded conformation of MB17 is required to attain the observed specificity for Mdmx over Mdm2 (**Figure 1.4.B**). Full Log2 enrichment heatmaps for all selection conditions are given in the Appendix (**Section 3.1**).

Comparing the enrichment ratios for each substitution across 2 different selection conditions reveals new insights. As expected, comparing enrichments from selections for Mdmx and Mdm2 specificity shows no variants that confer increased specificity for both targets (**Figure 1.4.C**). On the other hand, comparing enrichments from selections for Mdmx and Mdm2 affinity shows that there are variants that simultaneously confer increased affinity for both targets (**Figure 1.4.D**). Comparing enrichments from selections for affinity and specificity for Mdmx shows that there are variants that confer both increased specificity and affinity for Mdmx (**Figure 1.4.E**). Likewise, Comparing enrichments from selections for affinity and specificity for Mdm2 shows that there are variants that confer both increased specificity and affinity for Mdm2 and these are more enriched than those observed for Mdmx (**Figure 1.4.F**).

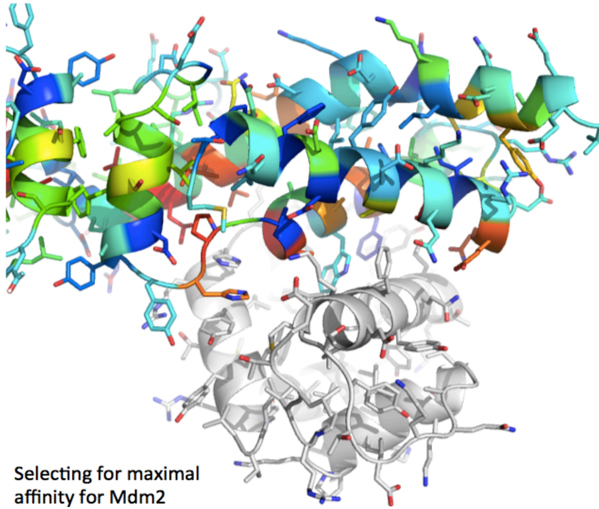
1.4.A.



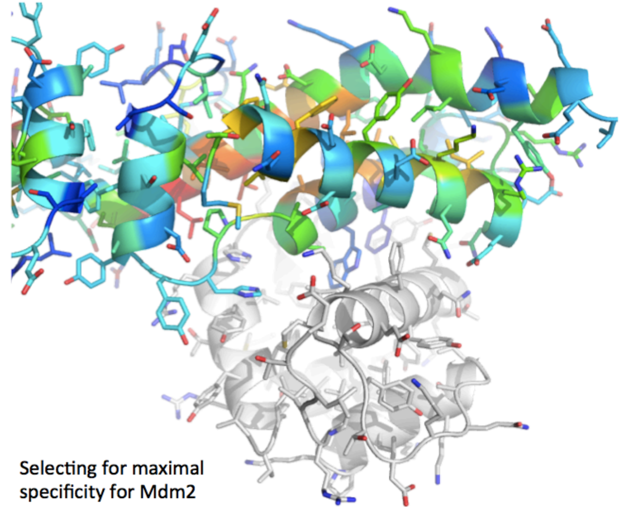
1.4.B.



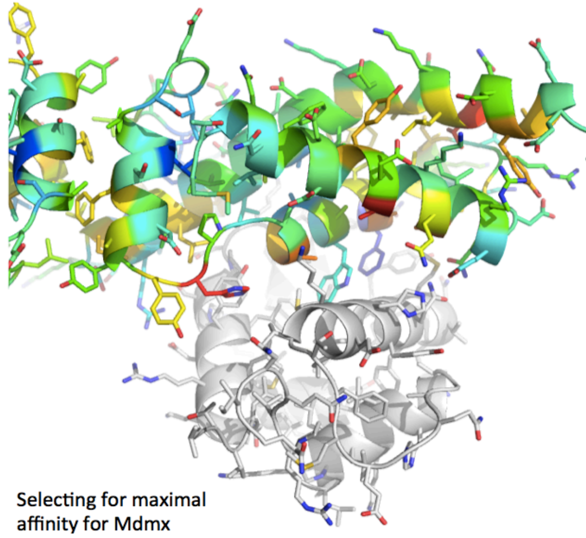
1.4.C.



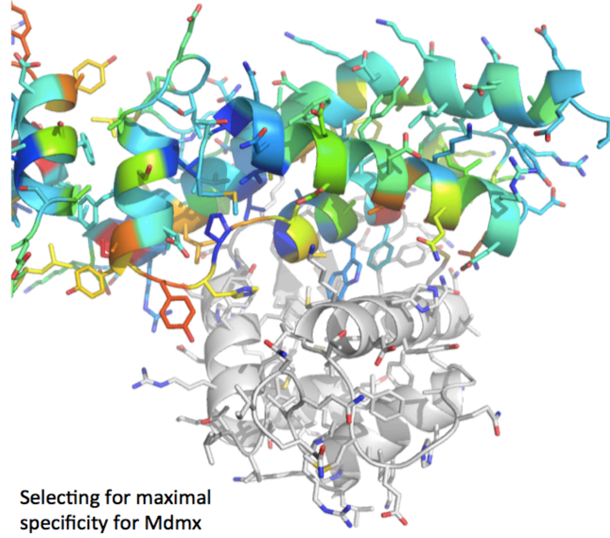
1.4.D.



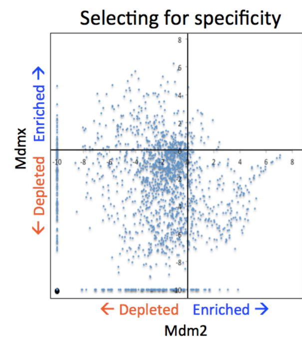
1.4.E.



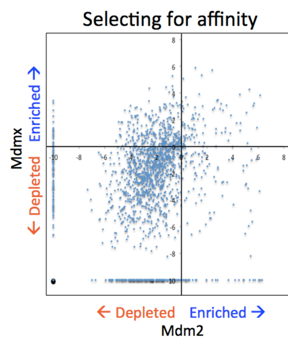
1.4.F.



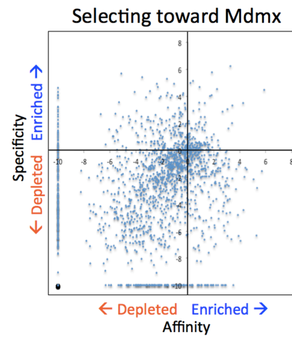
1.4.G.



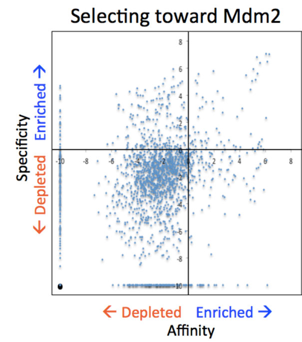
1.4.H.



1.4.I.



1.4.J.



1.4.K.

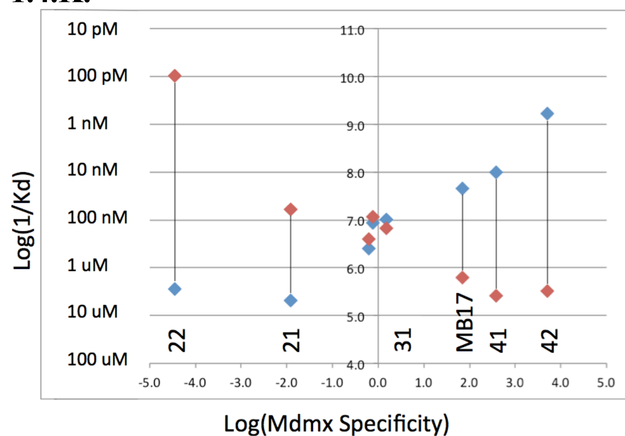


Figure 1.4: Validation of MB17's folded state and binding mode.

1.4.A. Heatmap of Log2 enrichments of single-site variants when selected for maximal specificity for Mdmx over Mdm2, 3rd round. The X-axis gives sequential positions in the protein while the Y-axis gives all possible amino acid substitutions. Orange-colored boxes denote a depleted substitution, yellow boxes denote a neutral substitution, blue boxes denote an enriched substitution, and white boxes denote no data. The scale goes from -9.1 to +6.2.

1.4.B. Heatmap of Log2 enrichments of single-site variants when selected for maximal affinity for Mdm2, 2nd round. The X-axis gives sequential positions in the protein while the Y-axis gives all possible amino acid substitutions. Orange-colored boxes denote a depleted substitution, yellow boxes denote a neutral substitution, blue boxes denote an enriched substitution, and white boxes denote no data. The scale goes from -7.3 to +6.4.

1.4.C. Cartoon and stick representation of model of Mdm2 (white) bound to MB17, colored according to the enrichment ratio of the most-enriched substitution at each position when sorted for affinity, 2nd round. The scale goes from blue for a Log2 enrichment of -2.0 to red for an enrichment of +6.4.

1.4.D. Cartoon and stick representation of model of Mdm2 (white) bound to MB17, colored according to the enrichment ratio of the most-enriched substitution at each position when sorted for specificity, 3rd round. The scale goes from blue for a Log2 enrichment of -1.5 to red for an enrichment of +7.2.

1.4.E. Cartoon and stick representation of model of Mdmx (white) bound to MB17, colored according to the enrichment ratio of the most-enriched substitution at each position when sorted for affinity, 2nd round. The scale goes from blue for a Log2 enrichment of -2.4 to red for an enrichment of +5.8.

1.4.F. Cartoon and stick representation of model of Mdm2 (white) bound to MB17 colored according to the enrichment ratio of the most-enriched substitution at each position when sorted for affinity, 2nd round. The scale goes from blue for a Log2 enrichment of -1.5 to red for an enrichment of +6.2.

1.4.G. Scatterplot of Mdmx specificity 3rd round Log2 enrichments plotted as a function of Mdm2 specificity 3rd round enrichments. The scale on both axes goes from -10 to +8.

1.4.H. Scatterplot of Mdmx affinity 2nd round Log2 enrichments plotted as a function of Mdm2 affinity 2nd enrichments. The scale on both axes goes from -10 to +8.

1.4.I. Scatterplot of Mdmx specificity 3rd round Log2 enrichments plotted as a function of Mdmx affinity 2nd round enrichments. The scale on both axes goes from -10 to +8.

1.4.J. Scatterplot of Mdm2 specificity 3rd round Log2 enrichments plotted as a function of Mdm2 affinity 2nd round enrichments. The scale on both axes goes from -10 to +8.

1.4.K. Scatterplot of Mdmx and Mdm2 Kds for MB17 variants plotted as a function of their Log10 specificity for Mdmx. Mdmx Kds are denoted with blue points while Mdm2 Kds are denoted with red points. A black line connects Kds for each respective variant, identified below the points. Points for MB17.31, MB17.32, and MB17.33 are simply labeled as "31". The X-axis goes from -5 to +5 while the Y-axis goes from 10 pM to 100 uM.

We sought to rapidly evolve MB17 for maximal specificity or affinity for Mdmx or Mdm2 by combining the most enriched mutations from each of the previous selection conditions

and screening for improved properties. Nine to eleven of the most enriched positions from each of the 4 selection conditions were simultaneously mutated to a small number of enriched residue types in 4 combinatorial libraries and screened against their previous selection conditions, respectively. A 5th condition was also added, 3-color labeling to simultaneously select for maximal but equal affinity for Mdmx and Mdm2. The location on the Mdmx-binding versus Mdm2-binding FACS plot of cells displaying the nonspecific p53 peptide was used to set the specificity gate for each sort using the 5th selection condition. As the library selected for maximal specificity for Mdm2 had the best chance of containing a dual-specificity MB17 variant, only that library was subjected to the 5th selection condition. Libraries made from selections involving Mdm2 intentionally excluded positions and mutations predicted to destabilize the folded state of MB17. The most represented clones after 6 to 8 rounds of sorting were further characterized by YST and ITC (**Table 1.4**). The top clone selected for maximal affinity and specificity for Mdmx was termed MB17.41, while the top clone selected for only maximal affinity was termed MB17.42. The top clone selected for maximal affinity and specificity for Mdm2 was termed MB17.21, while the top clone selected for only maximal affinity was termed MB17.22. The top clones selected for maximal but equal affinities for Mdmx and Mdm2 were termed MB17.31, MB17.32, and MB17.33. MB17 and its variants were expressed and their ITC affinities obtained by Elih Velasquez, Zhengding Su, and David Duda in the laboratories of Brenda Shulman and Richard Kriwacki at St. Jude Children's Research Hospital. YST and ITC affinities for all variants so tested are given in the Appendix (**Section 3.2**). Sequences for all variants tested are given in the Appendix (**Section 3.3**). Mutations made in the combination libraries and observed in the top clones are given in the Appendix (**Section 3.4**). If we assume that MB17.21 and MB17.22 have the same backbone conformation as MB17, then the fact that 7-9 mutations could

cause a roughly million-fold switch in specificity supports the hypothesis that MB17's specificity for Mdmx or Mdm2 is due entirely to the identities of its side chains. Size Exclusion chromatography of MB17, MB17.41, MB17.42, MB17.21, and MB17.22 shows that all are dimers, in spite of the fact that MB17.21 and MB17.22 bear mutations at the homodimer interface.

Table 1.4: ITC and average YST dissociation constants for MB17 and variants. All values are given in nanomolar.

Variant	Mdmx Kd (YST)	Mdmx Kd (ITC)	Mdm2 Kd (YST)	Mdm2 Kd (ITC)
MB17	22	39	1600	3800
MB17.21	4900	9200	59	21
MB17.31	115	Not determined	88	Not determined
MB17.32	100	Not determined	151	Not determined
MB17.33	400	Not determined	250	Not determined
MB17.41	10	Not determined	3900	Not determined
MB17.22	2800	Not determined	0.099	Not determined
MB17.42	0.61	Not determined	3100	Not determined

1.5 Crystal structures of MB17 and variants and comparisons to models

MB17 and its variants were expressed and crystallized by Elih Velasquez, Zhengding Su, and David Duda in the laboratories of Brenda Shulman and Richard Kriwacki at St. Jude Children's Research Hospital. The crystal structure of MB17 bound to Mdmx, determined by molecular replacement to 2.80 Å resolution, closely matches the design model, with a Cα RMSD of 0.808 Å, and reveals that MB17 is indeed a dimer that simultaneously binds 2 molecules of Mdmx (**Figure 1.5.A**). The crystal structure of MB17 more closely matches the crystal structure of OCR (PDB ID: 1s7z)(84) with a Cα RMSD of 0.576 Å (**Figure 1.5.B**). OCR methionine residue 6 is the first residue visible in the OCR crystal structure and the corresponding residue from MB17 and variants will be arbitrarily and uniformly referred to throughout this dissertation as residue 88. In the crystal structure, the C-terminal helix of the Mdmx p53-binding domain

(corresponding to the aforementioned protruding helix of Mdm2) adopts a position midway between that previously observed for Mdmx and that previously observed for Mdm2 (**Figure 1.5.C**)(30, 33-75). This observation suggests the selectivity of MB17 for Mdmx may be less dependent on this structural polymorphism than previously suspected. The backbone of residues 114-117 of MB17 (corresponding to p53 residues 14-18) was observed in two crystallographic forms, one similar to the design model, having a marked break in the helix, and one very different from the model, having a closed canonical helix (**Figure 1.5.D**). The p53 peptide is not helical in this region, allowing Glutamine 71 from Mdmx to hydrogen bond the backbone amide proton of Phenylalanine 19 (Phenylalanine 118 in MB17). The modeled break in the corresponding helix of MB17 presumably allows Glutamine 71 from Mdmx to make the corresponding hydrogen bond and provides a backbone carbonyl from Lysine 114 to additionally hydrogen bond the side-chain amide group of Glutamine 71. Conversely, a closed, helical form of this region appears to block access to these hydrogen bonding partners and force Glutamine 71 into an unfavorable side-chain conformation. This hypothesis supported by the fact that Proline substitutions are enriched at appropriate positions in this region of MB17, presumably to enforce the break in the helix. RMSD measurements were made using the Python script match.py (http://boscoh.com/protein/matchpy.html).

The crystal structure of MB17.41 bound to Mdmx was determined by molecular replacement to 2.79 Å resolution and closely matches the crystal structure of MB17 bound to Mdmx, with a Cα RMSD of 0.754 Å. The crystal structure of the MB17.41 molecule less closely matches the crystal structure of OCR, with a Cα RMSD of 1.11 Å. The C-terminal helix of the Mdmx p53-binding domain adopts the same position as that observed in the crystal structure of MB17 bound to Mdmx. The crystal structure differs, however, from the crystal structure of OCR

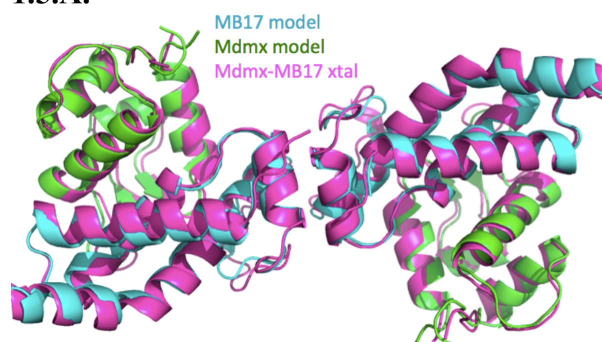
in that the backbone of MB17.41 residues 112-115 differs significantly from the corresponding region of OCR, with a Cα RMSD of 2.31 Å and with corresponding shifts in the N-terminal and C-terminal helices of MB17.41 (**Figure 1.5.E**). This change in the backbone still allows the side-chain of Glutamine 71 from Mdmx to hydrogen bond both the amide nitrogen from Phenylalanine 118 and the backbone carbonyl of Lysine 114. The change, however, reorients MB17.41 Lysines 114 and 115 toward Mdmx Glutamine 42 instead of Mdmx Glutamine 46. The change in the backbone of this region of MB17.41 relative to MB17 and OCR appears to be caused by an Alanine to Glutamine substitution at MB17.41 residue 119. The polar functional group of Glutamine 119 faces the C-terminus of MB17.41, packs on Mdmx Methionine 39, and forces MB17.41 loop residues 109-112 outward into solvent with a maximum Cα displacement of 3.3 Å. The side-chain carbonyl of Glutamine 119 is accessible to bulk solvent via a small channel while the side-chain amide hydrogen bonds to the hydroxyl of Serine 104 from MB17.41. Interestingly, A119Q was the most enriched mutation when selecting the aforementioned MB17 single site mutant library for maximal selectivity for Mdmx over Mdm2, supporting the hypothesis that adjustment of the backbone of MB17.41 residues 112-115 enforces the break in the corresponding helix of MB17.41, allowing Mdmx Glutamine 71 to hydrogen bond the backbone of Phenylalanine 118 and Lysine 114.

The crystal structure of MB17.42 bound to Mdmx was determined by molecular replacement to 2.59 Å resolution and closely matches the crystal structures of both OCR (84), with a Cα RMSD of 0.615 Å, and MB17 bound to Mdmx, with a Cα RMSD of 0.718 Å. Aligning just the MB17.42 molecule with the MB17 molecule gives a closer Cα RMSD of 0.419 Å and reveals that the MB17.42 does not adjust its fold relative to MB17 but rather rotates relative to the binding pocket on Mdmx (**Figure 1.5.F**). This rotation appears to be caused by a

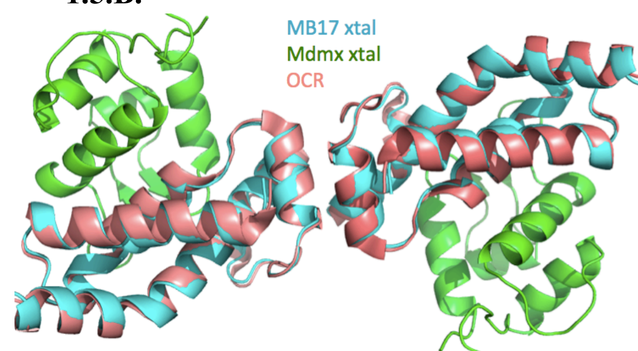
Histidine to Tryptophan substitution at MB17.42 residue 130 such that binding of Tryptophan 130 to Mdmx pulls the N-terminus of MB17.42 slightly closer to Mdmx and causes an approximately 6° rotation of MB17.42 around the Mdmx-MB17.42 axis relative to MB17 and MB17.41. The C-terminal helix of the Mdmx p53-binding domain adopts a position midway between that previously observed for Mdmx and that observed in the crystal structure of MB17 bound to Mdmx, suggesting that the position of this helix is largely dependent on the identity of molecules bound in the adjacent portion of the p53-binding pocket (**Figure 1.5.G**). Additionally, the backbone of residues 114-117 of MB17.42 was observed in the open conformation in both molecules in the asymmetric unit, likely due to an Alanine to Proline substitution at MB17.42 position 119 (**Figure 1.5.H**).

Phenylalanine 118 from MB17, MB17.41, and MB17.42 sits approximately 1.0 Å less deeply into the p53-binding pocket of Mdmx than does the corresponding Phenylalanine 19 from p53. Likewise, Tryptophan 122 from MB17, MB17.41, and MB17.42 sits approximately 0.4 Å less deeply into the p53-binding pocket of Mdmx than does the corresponding Tryptophan 23 from p53. Adjustment of the N-terminus of the MB17 helix supporting Phenylalanine 118 and Tryptophan 122 by MB17.41 and MB17.42 appears to make no difference in the position of these key binding residues (**Figure 1.5.I**).

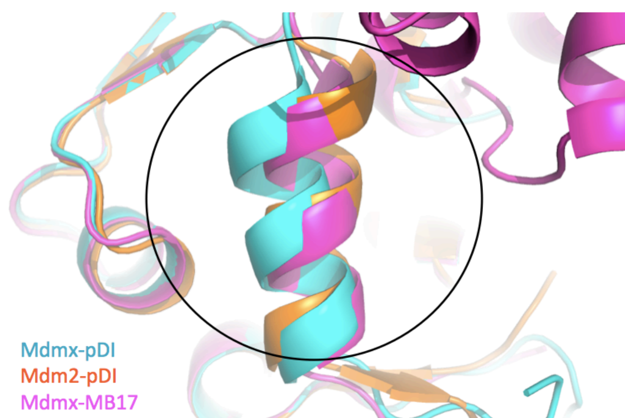
1.5.A.



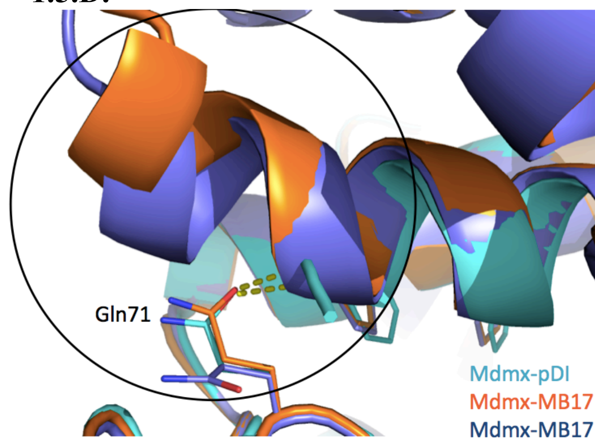
1.5.B.



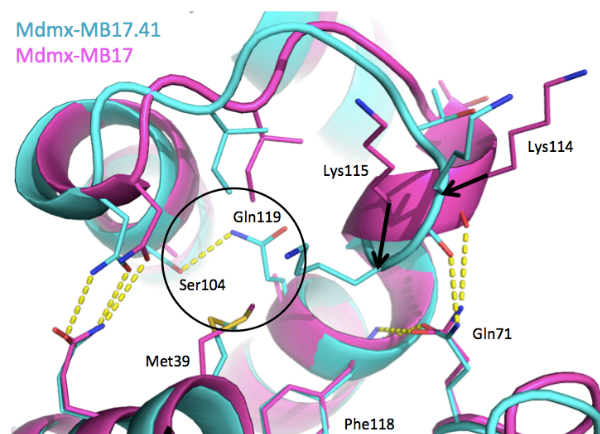
1.5.C.



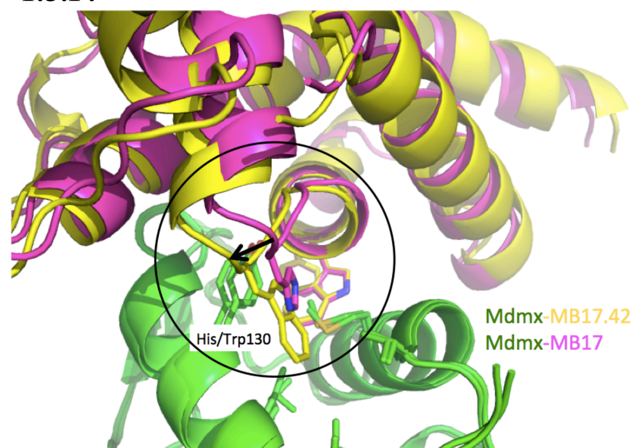
1.5.D.



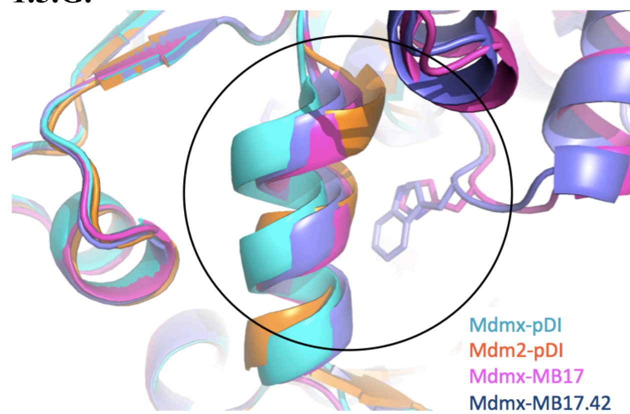
1.5.E.



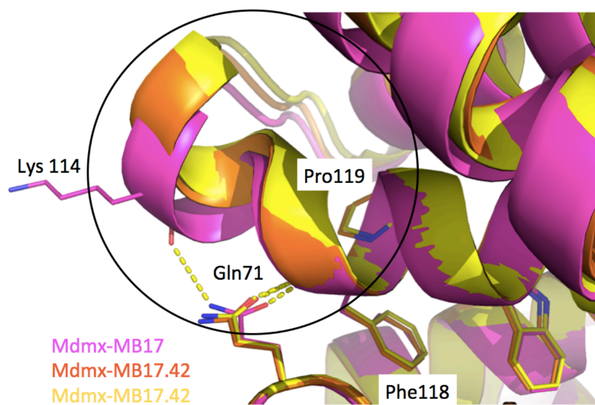
1.5.F.



1.5.G.



1.5.H.



1.5.I.

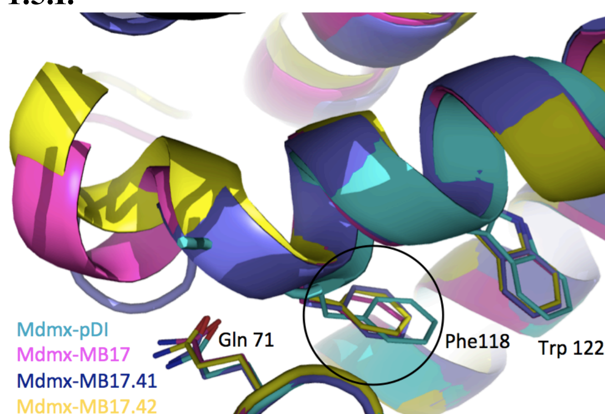


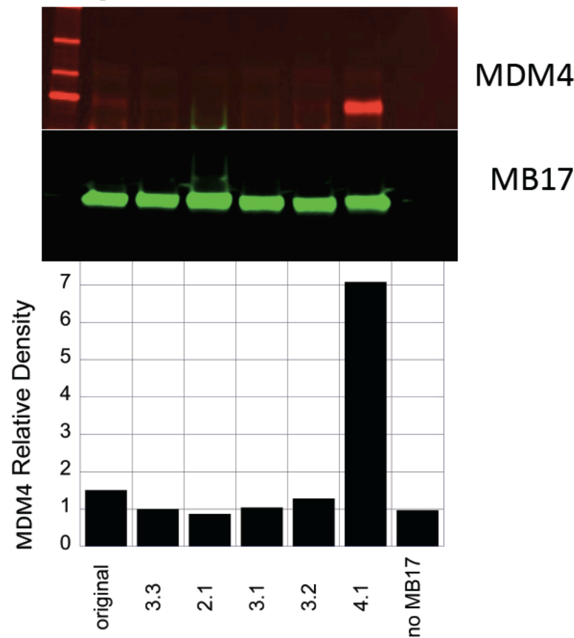
Figure 1.5: Crystal structures of MB17 and its variants are similar to design models.
1.5.A. The design model of MB17 (cyan) bound to Mdmx (green) superimposed onto the crystal structure (magenta) of MB17 bound to Mdmx, depicted in cartoon representation.
1.5.B. The design model of MB17 (cyan) bound to Mdmx (green) superimposed onto the crystal structure of OCR (salmon, PDB ID 1s7z), depicted in cartoon representation.
1.5.C. The crystal structure of MB17-Mdmx (magenta) superimposed onto the crystal structure of pDI-Mdmx (cyan, PDB ID 3fdo) and pDI-Mdm2 (orange, PDB ID 3g03), each depicted in cartoon representation.
1.5.D. Both molecules from the asymmetric unit of the crystal structure of MB17-Mdmx (blue and orange) superimposed onto the crystal structure of pDI-Mdmx (cyan), depicted in cartoon representation. Hydrogen bonds are denoted with yellow dashes and the hydrogen-bonding Glutamine 71 from Mdmx and the region of MB17 observed in 2 crystallographic forms are highlighted with a black circle.
1.5.E. The crystal structure of MB17.41-Mdmx (cyan) superimposed onto the crystal structure of MB17-Mdmx (magenta), each represented as a cartoon and sticks. The movement of the Cα atoms supporting MB17 Lysines 114 and 115 is denoted with black arrows, the Alanine 119 to Glutamine mutation is highlighted with a black circle, and hydrogen bonds are denoted with yellow dashes.
1.5.F. The crystal structure of MB17.42 (yellow) bound to Mdmx (green) superimposed onto the crystal structure of MB17 (magenta) bound to Mdmx (green), each represented as a cartoon and sticks. The rotation of MB17.42 relative to Mdmx is denoted with a black arrow and highlighted with a black circle.
1.5.G. Superposition of the crystal structures of MB17.42-Mdmx (blue), MB17-Mdmx (magenta), pDI-Mdmx (cyan), and pDI-Mdm2 (orange), each represented as a cartoon and sticks. The C-terminal helix of Mdmx/Mdm2 is highlighted with a black circle.
1.5.H. Both molecules from the asymmetric unit of the crystal structure of MB17.42-Mdmx (yellow and orange) superimposed onto the crystal structure of MB17-Mdmx (magenta) each represented as a cartoon and sticks. The open helix of MB17.42 relative to MB17 is highlighted with a black circle and hydrogen bonds are shown as yellow dashes.
1.5.I. Superposition of the crystal structures of Mdmx-pDI (cyan), Mdmx-MB17 (magenta), Mdmx-MB17.41 (blue), and Mdmx-MB17.42 (yellow), each represented as a cartoon and sticks. The position from each structure of the key Phenylalanine residue relative to Mdmx is highlighted with a black circle.

1.6 MB17 and variants retain affinity and specificity in mammalian cells

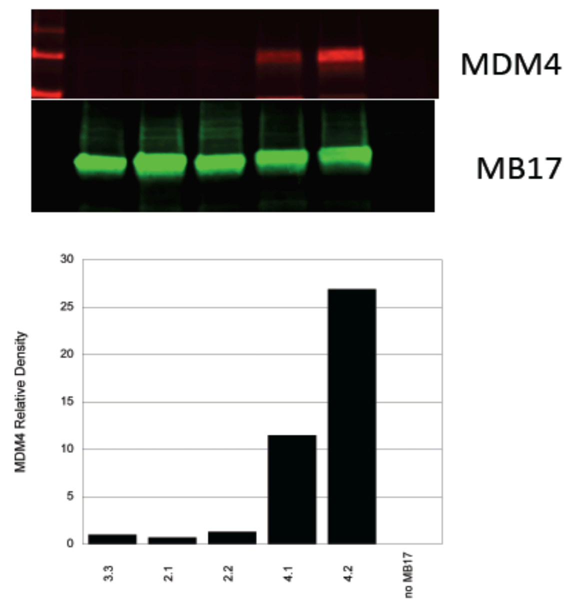
Analysis of the binding of MB17 and its variants to Mdmx and Mdm2 in mammalian cells was carried out by Lyra Griffiths and Jongrye Jeon in the laboratory of Dr. Michael Dyer at St. Jude Children's Research Hospital. Genes encoding MB17, MB17.21, MB17.22, MB17.31,

MB17.32, MB17.33, MB17.41, and MB17.42, each fused to enhanced green fluorescent protein (EGFP), having an N terminal flag tag and a C-terminal avi tag (bearing a biotin moiety) were transfected into mammalian 293T cells. The proteins were expressed in the mammalian cells, which were then lysed. An anti-flag antibody was used to extract the MB17 variants from the lysate and the bound protein was subjected to polyacrylamide gel electrophoresis (PAGE). Western blot of the PAGE gels reveals that MB17.41 and to a lesser degree, MB17, pull down more endogenous Mdmx from mammalian cells than any other variant (**Figure 1.6.A**). MB17.42, in turn, pulls down even more Mdmx than MB17.41 (**Figure 1.6.B**). MB17.21 pulls down more Mdm2 from mammalian cells than any other variant (**Figure 1.6.C**). MB17.22, in turn, pulls down even more Mdm2 than MB17.21 (**Figure 1.6.D**).

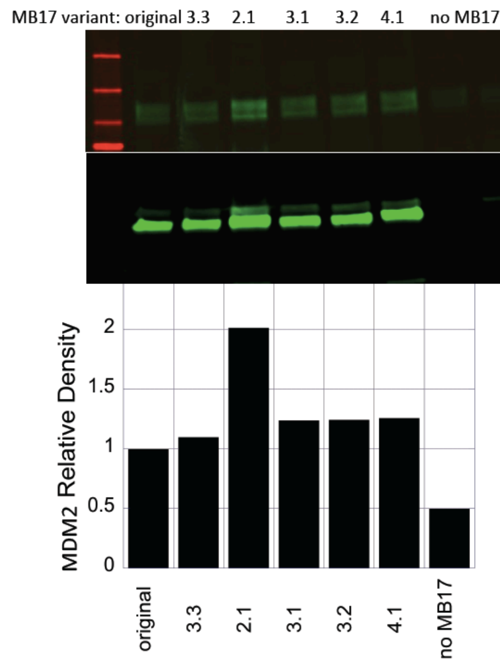
1.6.A.



1.6.B.



1.6.C.



1.6.D.

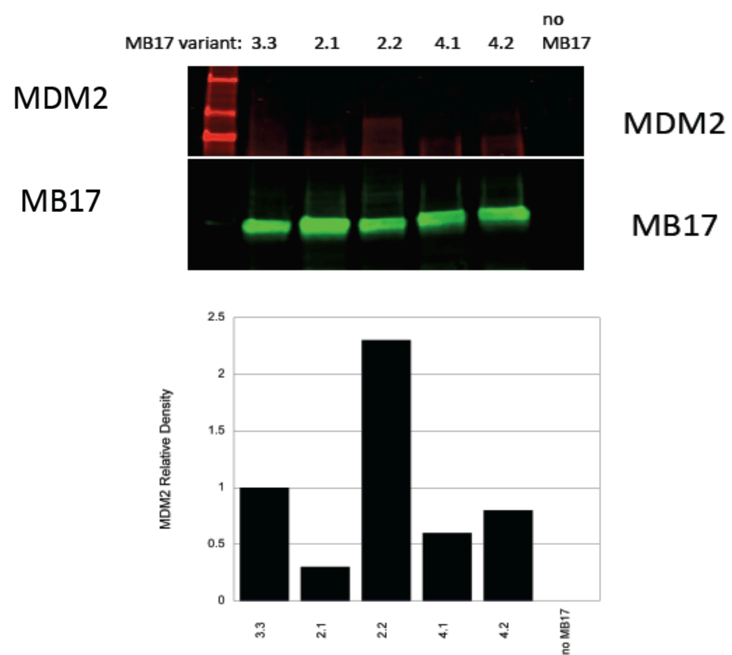


Figure 1.6: MB17 and variants retain affinity and specificity in mammalian cells.
1.6.A. Western blot and densitometry of MB17 variants extracted from 293T cells and blotted for Mdm4 (Mdmx, red).
1.6.B. Western blot and densitometry of MB17 variants, including MB17.42 and MB17.22, extracted from 293T cells and blotted for Mdm4 (Mdmx, red).
1.6.C. Western blot and densitometry of MB17 variants extracted from 293T cells and blotted for Mdm2 (green).
1.6.D. Western blot and densitometry of MB17 variants, including MB17.42 and MB17.22, extracted from 293T cells and blotted for Mdm2 (red).

1.7 MB17 variants halt the cell cycle and trigger apoptosis in colon cancer cell lines

Analysis of the effect of MB17 and its variants on the cell cycle in mammalian cells was carried out by Lyra Griffiths and Jongrye Jeon in the laboratory of Dr. Michael Dyer at St. Jude Children's Research Hospital. Mammalian 293T cells transfected with either Flag-MB17-EGFP-Avi or Flag-MB17-EGFP-6XHis were treated with 5-Ethynyl-2'-deoxyuridine (EdU). The cells were either induced to express MB17 or not and the fraction of transfected cells that also incorporated EdU was determined by flow cytometry. Induction of MB17 expression resulted in a 52% drop in the number of transfected cells that also incorporated EdU (**Figure 1.7.A**). As EdU incorporation is a reporter for DNA synthesis, this result suggests that MB17 expression in

mammalian cells halts DNA synthesis in 293T cells. Microscopic analysis of the transfected cells revealed that MB17 localized to the nucleus of mammalian cells in the absence of a nuclear localization signal and that EdU uptake only appears to occur in cells not expressing MB17 (**Figure 1.7.B**).

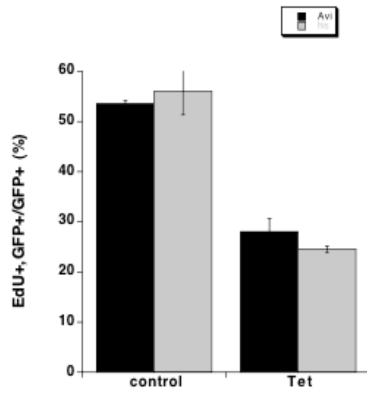
We sought to determine whether MB17 and its variants would have this same effect on cellular behavior in cancer cell lines and whether p53 was involved. To that end, we transfected MB17.21, MB17.41, MB17.31, MB17.32, and MB17.33 into HCT116 cells, a colon cancer cell line. We also transfected the variants into an HCT116 cell line that lacked p53 (HCT116 p53 -/-). Microscopic analysis of HCT116 cells transfected with MB17.41 revealed that caspase 3 becomes activated only in cells expressing MB17.41 (**Figure 1.7.C**). The fraction of transfected cells also positive for active caspase 3 was determined using flow cytometry (89). MB17.31, MB17.32, and MB17.33 had little effect on caspase 3 activation above that of negative controls in HCT116 cells. MB17.21 and MB17.41, however, showed a 3.4-fold and 3.1-fold increase in the number of cells with active caspase 3, respectively. As caspase 3 activation is a sign of the execution phase of apoptosis, this result suggests that MB17 variants have the ability to trigger apoptosis in cancer cells. The increase in caspase 3 activation was not seen in HCT116 p53 -/- cells, indicating that p53 is required for MB17 variants to trigger apoptosis in cancer cells (**Figure 1.7.D**).

We sought to determine whether coexpression of MB17.21 with MB17.41 or MB17.22 with MB17.42 might have synergistic activity in cancer cells and whether any additional gains in cell-cycle arrest and apoptosis activation could be achieved with the addition of Nutlin3a, a small molecule Mdm2 inhibitor. To that end, HCT116 and HCT116 p53 -/- cells were co-transfected with MB17.21 and MB17.41, MB17.22 and MB17.42, or GFP alone. The cells were also

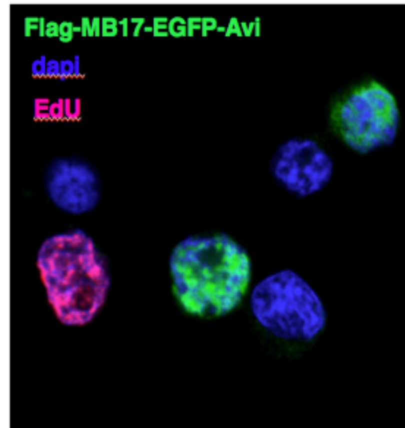
untreated or treated with Nutlin3a and the fraction of transfected cells that also incorporated EdU or exhibited caspase 3 activation was determined by flow cytometry. Nutlin3a treatment halted all EdU incorporation in the GFP-transfected cells in a p53-dependent manner. MB17.21/MB17.41 co-expression and MB17.22/MB17.42 coexpression also halted all EdU incorporation in a p53-dependent manner, as did concurrent MB17.21/MB17.41 or MB17.22/MB17.42 coexpression and Nutlin3a treatment (**Figure 1.7.E**). Nutlin3a treatment resulted in a subtle 1.1-fold increase in the number of GFP-transfected cells positive for caspase 3 activation in a p53-dependent manner. MB17.21/MB17.41 or MB17.22/MB17.42 co-expression resulted in a 1.8-fold increase in the number of GFP-transfected cells positive for caspase 3 activation in a p53-dependent manner, while concurrent MB17.21/MB17.41 or MB17.22/MB17.442 coexpression and Nutlin3a treatment resulted in a 1.6-fold increase in the number of GFP-transfected cells positive for caspase 3 activation in a p53-dependent manner (**Figure 1.7.F**).

These results, taken together with the fact that MB17 variants have been shown to bind Mdmx and Mdm2 at the same site occupied by the p53 transactivation helix, suggest that MB17 variants do in fact block Mdmx and Mdm2-mediated repression of p53 activity by competitively occupying the sites on Mdmx and Mdm2 that normally bind p53. These results also further validate Mdmx as a target for anti-cancer therapy.

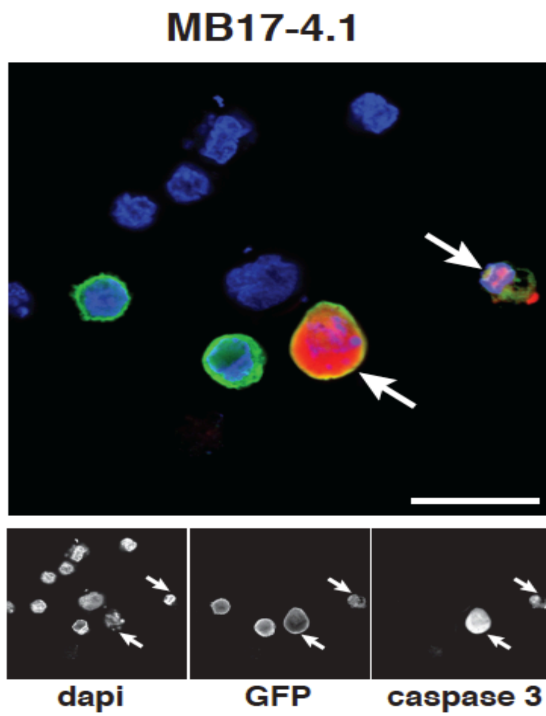
1.7.A.



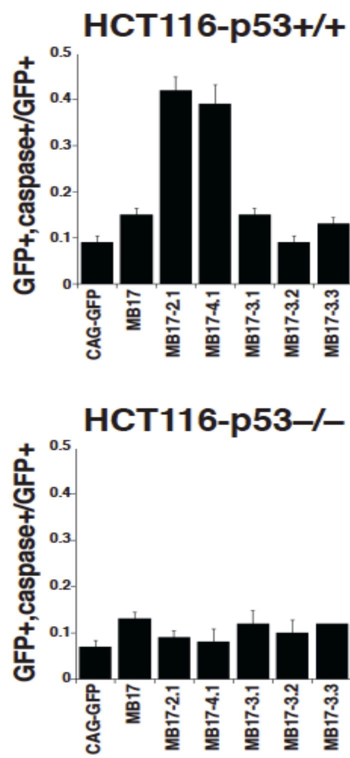
1.7.B.



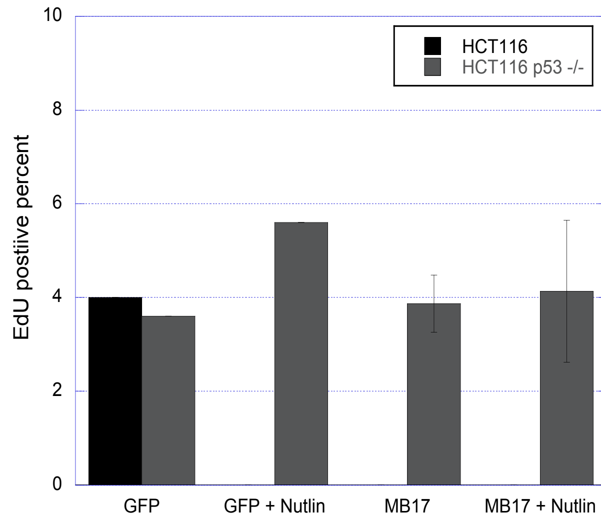
1.7.C.



1.7.D.



1.7.E.



1.7.F.

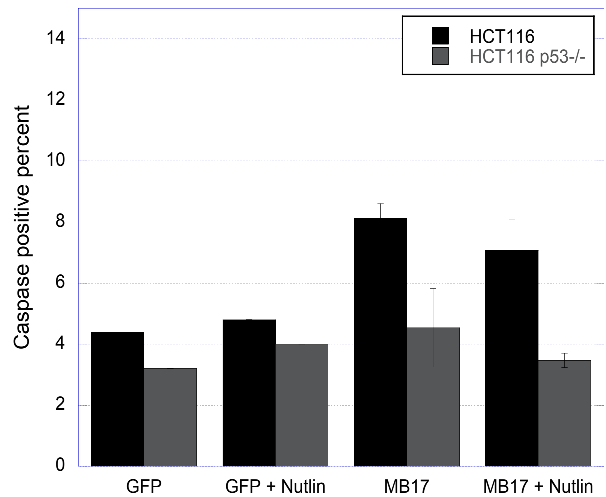


Figure 1.7: MB17 variants halt the cell cycle and trigger apoptosis in colon cancer cell lines in a p53-dependent manner.

1.7.A. Percent of transfected 293T cells undergoing cell division (as detected by EdU uptake) when MB17 is not expressed (Control) or expressed (Tet). Black bars denote Flag-MB17-EGFP-Avi while gray bars denote Flag-MB17-EGFP-6XHis

1.7.B. Image of 293T cells expressing Flag-MB17-EGFP-Avi, stained with Dapi, and treated with 5-Ethynyl-2'-deoxyuridine (EdU).

1.7.C. Image of HCT116 cells expressing Flag-MB17-EGFP-Avi, stained with Dapi, and treated with EdU.

1.7.D. Percent of transfected HCT116 p53 +/+ and HCT116 p53 -/- cells that undergo caspase activation when untreated (CAG-GFP) or when induced to express various MB17 variants.

1.7.E. Percent of transfected HCT116 p53 +/+ and HCT116 p53 -/- cells that are EdU positive when untreated (GFP), treated with Nutlin3a, when MB17.21/MB17.41 or MB17.22/MB17.42 are expressed, or when MB17.21/MB17.41 or MB17.22/MB17.42 are coexpressed in addition to treatment with Nutlin3a. Black bars denote HCT116 p53 +/+ cells and gray bars denote HCT116 p53 -/- cells. MB17.21/MB17.41 and MB17.22/MB17.42 co-transfections gave similar results and the results reported are the average of both co-transfections.

1.7.F. Percent of transfected HCT116 p53 +/+ and HCT116 p53 -/- cells that undergo caspase 3 activation when untreated (GFP), treated with Nutlin3a, when MB17.21/MB17.41 or MB17.22/MB17.42 are expressed, or when MB17.21/MB17.41 or MB17.22/MB17.42 are coexpressed in addition to treatment with Nutlin3a. Black bars denote HCT116 p53 +/+ cells and gray bars denote HCT116 p53 -/- cells. MB17.21/MB17.41 and MB17.22/MB17.42 co-transfections gave similar results and the results reported are the average of both co-transfections.

1.8 Balancing the charges on the surface of top MB17 variants

Mass spectrometry analysis of MB17 variants was carried out by Vishwajeeth Pagala in the laboratory of Junmin Peng at St. Jude Children's Research Hospital. Biotinylated MB17.41 was expressed in 293T mammalian cells, extracted therefrom, and proteins bound to it were subjected to identification and quantification by mass spectrometry (MS)(**Figures 1.8.A, 1.8.B,**

and **1.8.C**). MS revealed that MB17.41 bound 251 proteins in the cell other than Mdmx and Mdm2, at least 37% of which were proteins known to bind carboxylates, phosphates, nucleotides, RNA, or DNA. Binding to other intracellular proteins might render MB17 variants unavailable to inhibit Mdmx or Mdm2, a hypothesis supported by the fact that MB17 variants affect p53 activity only when expressed at high levels. As they are based on OCR, a known DNA mimic (84), MB17 variants are enriched for surface Aspartates and Glutamates. The OCR dimer has a net charge of -56, while dimers of MB17, MB17.41, MB17.42, MB17.21, and MB17.22 have net charges of -28, -34, -28, -38, and -36, respectively (**Figure 1.8.D**).

In an effort to reduce off-target binding of the top MB17 variants and improve their potency in mammalian cells, the surfaces of MB17.41, MB17.42, MB17.21, and MB17.22 were manually redesigned to remove large areas of negative charge and superfluous hydrophobic residues. To illustrate the degree of charge balance attained, dimers of some representative redesigned variants, MB17.28, MB17.47, and MB17.48, have net charges of -10, -2, and 0, respectively (**Figure 1.8.E**). Two to four redesigned variants, each having similar mutations to charged residues but with subtly different combinations of mutations that reverted designed positions to the identity observed in OCR (84), were made for each of the input parent variants. Additionally, a Proline mutation was tested in place of Tyrosine 132 (Histidine 132 in MB17.41) in an attempt to further stabilize the observed backbone conformation of the MB17 variants. Genes for the resulting re-designed variants were either synthesized by a commercial vendor (IDT, San Jose, California, USA) or synthesized in house using an overlap PCR and assembly technique, cloned into pETCON, and their affinities and specificities determined using YST (**Table 1.8**). Charge-balanced variants were obtained for MB17.22, MB17.41, and MB17.42 that had comparable affinities and specificities to their patent variants. Further ongoing work is

required to determine whether balancing the charges on the surface of the MB17 variants will reduce nonspecific binding and increase potency in mammalian cells. All of the varaints that reverted designed or evolved mutations to the wild type identitity showed reduced affinities for their targets, supporting the need for the designed or evolved mutations to achieve tight binding.

Interestingly, one of the additional mutations to MB17.21, Glutamate 129 to Arginine, necessary to eliminate a large area of negative charge, resulted in a re-designed variant with 7.5-fold less affinity for Mdm2. This unexpected result suggests that Glutamate 129 is key for MB17.21 to bind Mdm2 tightly. Glutamate 129 replaces Proline 129 from MB17 and is modeled as having the same backbone conformation in the associated region as Proline. If Glutamate does indeed have the same backbone conformation as Proline 129, one would expect Arginine 129 to serve equally well. In fact, one would expect the Arginine to even more readily adopt the backbone conformation observed with the Proline since the Arginine's charged functional group and associated hydration sphere would be further away from the core of MB17.21 than would the charged functional group and associated hydration sphere of the Glutamate (**Figure 1.8.F**). Other mutations made to MB17.21 were identical to those made to MB17.22, suggesting that the Glutamate to Arginine mutation is indeed responsible for the loss in affinity. Taken together, these results and observations suggest that the region surrounding MB17.21 Glutamate 129 has a different backbone conformation than what is observed in the crystal structures of MB17, MB17.41, and MB17.42, and the model of MB17.21. That MB17.21 may have a backbone conformation different than MB17 suggests that simply altering the side chain identities of MB17 is not enough to change it's specificity and that MB17's specificity for Mdmx must also be encoded in its backbone conformation. This poses the question as to whether MB17.22 also alters its backbone conformation in order to achieve the observed affinity and specificity for

Mdm2. MB17.22 has Proline at position 129, which suggests that it may have the same conformation in this region as MB17. These questions will only be definitively answered once the crystal structures of MB17.21 and MB17.22 bound to Mdm2 become available.

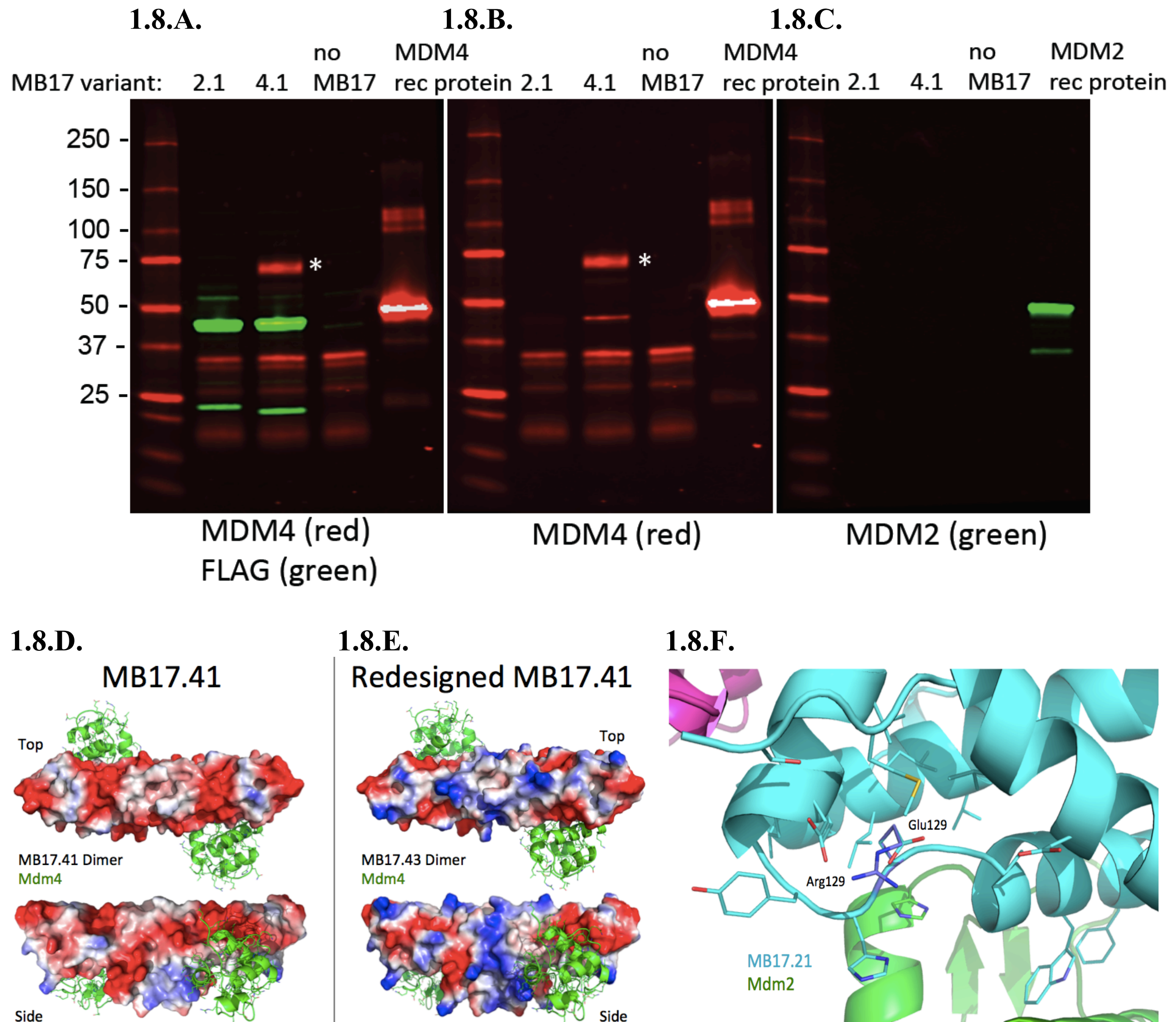


Figure 1.8: Western blots of biotinylated MB17.21 and MB17.41 extracted from 293T cells.
1.8.A. MB17.21 and MB17.41 blotted with anti-Mdm4 (Mdmx) antibody (red) and anti-flag antibody (green). White asterisks denote endogenous Mdmx.
1.8.B. MB17.21 and MB17.41 blotted with anti-Mdm4 (Mdmx) antibody (red). White asterisks denote endogenous Mdmx.
1.8.C. MB17.21 and MB17.41 blotted with anti-Mdm2 antibody (green).
1.8.D. Top and side view of a model of MB17.41-Mdmx. MB17.41 is shown as an electrostatics map and Mdmx is shown as a green cartoon and sticks.
1.8.E. Top and side view of a model of MB17.43-Mdmx, a charge-balanced variant of MB17.41. MB17.43 is shown as an electrostatics map and Mdmx is shown as a green cartoon and sticks.
1.8.F. Model of MB17.21 (cyan) bound to Mdm2 (green) shown as cartoon and sticks. The Arginine 129 substitution is shown as blue sticks.

Table 1.8: Average YST Affinities of re-designed MB17 variants. All values are given in nanomolar.

Parent	Charge-balance Variant	Mdmx Kd (YST)(nM)	Mdm2 Kd (YST)(nM)	Mutations to non-charged residues
MB17.21		4900	59	none
	MB17.23	>10000	8200	E129P, Y132P, Q164L
	MB17.25	>10000	4500	E129P, Y132A, Q164L
	MB17.26	>10000	440	E129R, Y132P
	MB17.27	>10000	7200	E129R, Y132A
MB17.22		2800	.099	none
	MB17.24	1400	0.20	Y132P
	MB17.28	738	0.14	Y132A
MB17.41		10	3900	none
	MB17.43	260	>10000	S104N, Q119A, H132P, L165Q
	MB17.45	370	>10000	S104N, Q119A, H132A, L165Q
	MB17.46	10	>10000	H132P
	MB17.47	10	>10000	H132A
MB17.42		0.61	3100	none
	MB17.44	220	>10000	Y132P
	MB17.48	0.35	>10000	Y132A

1.9 Conclusions and future directions

We have used Rosetta to stabilize the bound state of the p53 Mdmx-binding motif by incorporating it into a number of larger host proteins and have designed the resulting binding proteins to be specific for Mdmx over Mdm2. Using ITC and YST, we have shown that 1 of the resulting binding proteins, MB17, is highly specific for Mdmx over Mdm2. We have validated that our preparations of Mdmx and Mdm2 do not bind nonspecifically. We evolved MB17 for greater affinity or specificity for Mdmx, greater affinity or specificity for Mdm2, or equal affinity for Mdmx and Mdm2. MB17 and its variants can be expressed in E. coli and specifically and tightly bind to either Mdmx or Mdm2 as purified proteins by ITC. Using deep mutational scanning and x-ray crystallography, we have validated that MB17, MB17.41, and MB17.42 are properly folded and bind to human Mdmx at the correct site using the designed binding surface.

We have shown that MB17 and its variants retain their observed affinities and specificities in mammalian cells and that they halt the cell cycle and trigger apoptosis in cancer cell lines in a p53-dependent manner. MB17 and its variants represent the first highly specific, high affinity inhibitors of Mdmx that can be used in mammalian cells. Ongoing work will determine what effect inhibiting just Mdmx will have on normal and cancer cells.

One question immediately arises: Why is MB17 so specific for Mdmx over Mdm2 while 4 other designed binders displayed on the yeast surface fail to be? A number of possible, non-mutually exclusive explanations present themselves. First, a lesson learned from other protein design studies is that many designed proteins fail to fold into the modeled conformation (90). Second, comparison of MB17 to models of MB14, MB19, MB25, and MB26 reveal that MB17 completely desolvates Histidine 96 from Mdm2 (Proline 95 in Mdmx) while MB14, MB19, and MB25 allow the Mdm2 Histidine access to solvent. MB26 also desolates the Histidine in a manner similar to MB17, and in spite of binding Mdmx only weakly, is about 3-fold specific for Mdmx over Mdm2. The hypothesis that Histidine 96 desolvation contributes to MB17 selectivity is also supported by the fact that selecting for increased Mdm2 affinity or specificity enriches polar substitutions at the MB17 hydrophobic residues responsible for desolvating the Mdm2 Histidine 96. The hypothesis is additionally supported by the observation that MB17.22 incorporates a Glycine in place of MB17 Asparagine 108. This Glycine is on the end of a helix and thus is not reasonably expected to be destabilizing. An alternate explanation for the role of the Glycine mutation is to allow the MB17.22 C-terminus to tilt closer to Mdm2 while allowing the N-terminus to tilt away from the Mdm2 protruding helix and Histidine 96, as observed in a homology model of MB17.22 bound to Mdm2. Third, MB17 surrounds Mdm2 Phenylalanine 55 with polar residues that should favor binding the Mdmx Histidine 54 over the Mdm2

Phenylalanine. MB14, MB25, and MB26 leave the Phenylalanine exposed to bulk solvent and are therefore unable to distinguish it from the Histidine. MB19 places the Phenylalanine in an environment predicted to favor the Phenylalanine and Histidine equally well and actually appears to be a slight 1.3-fold specific for Mdmx over Mdm2.

Notably, simply aligning all structures in the PDB to the p53 transactivation helix in the prescence of Mdmx fails to return OCR as a match. To properly mimic p53 as bound to Mdmx without causing the remainder of OCR to clash with the rims of the binding pocket, the corresponding OCR helix must move as much as 0.81 Å relative to the remainder of OCR. If the structure of OCR is aligned to the MB17 molecule in the crystal structure of MB17-Mdmx, the MB17 helix mimicking p53 has an RMSD to its original position in OCR of 0.747 Å. Neither the Epigraft/MotifGraft nor DockDesign strategies are thus capable of finding OCR as a scaffold for an Mdmx binder. Only PlaceStub is capable of identifying OCR as a potential scaffold. FoldFromLoops is capable of designing binders that exactly mimic the position of the p53 helix, but has no inherent mechanism for identifying OCR as a potential scaffold. One cannot rule out the possibility that if supplied with a large enough number of starting topologies, FoldFromLoops could design other binding proteins capable of high specificity for Mdmx over Mdm2. If PlaceStub is to be matched to productive starting-point topologies, a method must be developed that can match known PDB structure and/or topologies to a given target surface or binding motif that includes backbone flexibility.

With regards to MB17.21 and MB17.22, other questions yet remain to be answered: Do the mutations necessary to confer Mdm2 selectivity to MB17 truly destabilize or alter its backbone conformation, and if so, in what way and to what degree? Do MB17.2.1 and MB17.2.2 even adopt a folded structure similar to MB17? Is the core three-helix bundle motif conserved at

the very least? These questions will only be definitively answered once the crystal structures of MB17.21 and MB17.22 bound to Mdm2 become available.

Section 2

Computational design of proteins to inhibit the interaction of EED-Ezh2

2.1 Background and Motivation

Ezh2, the catalytic subunit of the polycomb repressive complex 2 (PRC2), is overexpressed or up-regulated in many cancer types and mounting evidence implicates Ezh2 as critical in all steps of cancer progression, including initiation, development, and metastasis, as well as in anti-cancer drug resistance. Additionally, factors important in stem cell maintenance, cell proliferation, and tumorigenesis also regulate Ezh2 expression and increased expression of Ezh2 correlates with increased cancer aggressiveness (91). Ezh2 is a histone lysine methyl-transferase and numerous drugs that block its active site are undergoing clinical development as cancer therapeutics (91). The interaction of Ezh2 with another PRC2 component, EED, has been shown to be critical for Ezh2 catalytic activity, presumably by localizing Ezh2 to its substrates. The N-terminal domain of Ezh2 consists of a long helix that interacts with a long cleft on the lower face of EED, a 7-bladed beta-propeller adaptor protein (92, 93). Another route for anti-cancer therapeutics targeting Ezh2 activity involves blocking this interaction between Ezh2 and EED. To that end, our collaborators in the Stuart Orkin lab at Harvard Medical School developed stapled peptide mimetics of the N-terminal EZH2 helix to serve as competitive inhibitors of the Ezh2-EED interaction (94). Stapled Ezh2 N-terminal peptide mimetics have affinities for EED ranging from 260 nM to around 10 uM. These stapled peptides would be far more effective if their affinity for EED were tighter. A tightly binding inhibitor of the EED-Ezh2 interaction could further elucidate importance of the EED-Ezh2 interaction in normal and cancer cells and possibly inform structure-based drug design efforts.

We sought to apply computational protein design to engineer a synthetic protein that incorporated the Ezh2 N-terminal peptide in a preformed helical state so as to achieve tighter binding to EED. A synthetic protein could also make additional interactions with EED to further increase affinity. Additionally, targeting the Ezh2 binding cleft of EED represents a challenge in computational protein design, as the Ezh2-binding cleft of EED is very large and quite charged (**Figure 2.2.A**). Protein modeling software has been shown to models charged interactions poorly (95). We also wanted to see how tightly a computationally designed protein could be made to bind such a difficult surface.

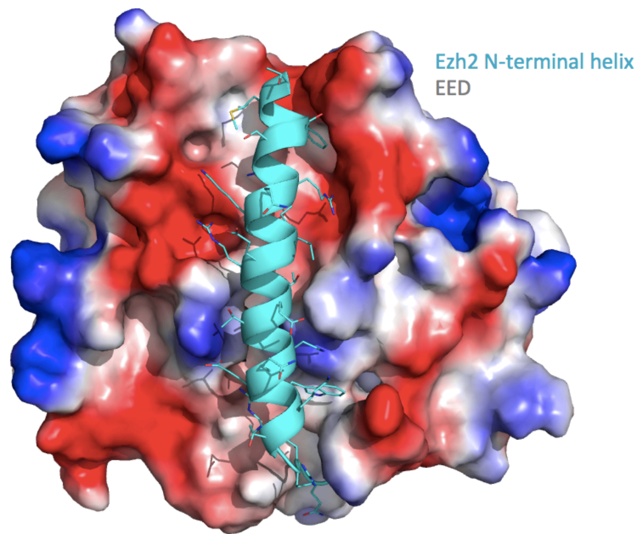
2.2 Design strategy

Using the Epigraft application in Rosetta (97-99), we searched through all monomeric proteins in the Protein Data Bank (PDB) for proteins that could host the 19-residue region of the Ezh2 N-terminal helix from Phenylalanine 42 to Tryptophan 60, inclusive. Specifically, we searched for proteins with an exterior helix that closely matched the backbone conformation of the Ezh2 helix and that failed to clash with EED when that exterior helix was superimposed onto the Ezh2 helix as it was bound to EED. Because the Ezh2 N-terminal helix binds its cognate cleft on EED with a distinctive curved backbone conformation (93), a potential host protein (scaffold) must have an exterior helix with a concave shape in order to match. Concave exterior helices of sufficient length are rare in monomeric proteins, based on our search of the PDB. The closest-matching protein in the PDB was the core domain of apolipoprotein E2 (ApoE2), a component of blood-borne lipid particles, with a backbone RMSD of 0.569 Å over the matched region (100). Unfortunately, because the backbone of ApoE2 and the other close matches from the PDB do not have exactly the same backbone conformation as that of the Ezh2 N-terminal helix, the scaffold

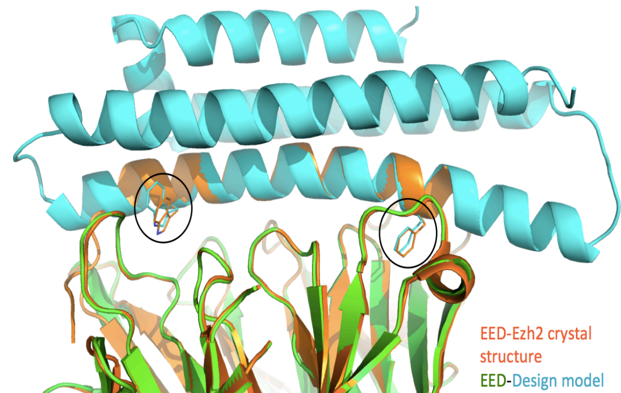
protein cannot position the aromatic binding residues at either end of the helix exactly as in Ezh2 (**Figure 2.2.B**). This is a problem because improperly positioned Phenylalanine and Tryptophan residues may clash with EED, abrogating binding. Superposition of the crystal structure of EED-Ezh2 onto 11 crystal structures of EED not bound to Ezh2 revealed that the loops comprising the Ezh2-binding cleft are flexible and appear to move in the bound state to accommodate the aromatic binding residues at either end of the Ezh2 N-terminal helix (**Figure 2.2.C**)(93, 101, 102). We surmised that the flexible loops comprising the Ezh2-binding pocket on EED might adjust their conformation to accommodate a binding protein with a subtly different backbone than Ezh2.

We required candidate scaffolds to match the 19-residue region of the Ezh2 N-terminal helix with a backbone RMSD of 1.5 Å and visually screened the resulting matches to remove unsuitable starting points for design (See **Section 7.1** in the Appendix for details). The side chains from the key binding residues from the Ezh2 N-terminal helix were grafted onto the resulting 8 candidate scaffolds (binders) and any binder residues that clashed with EED were mutated (**Figure 2.2.D**). The identities of surface and interfacial side chains of the binders were optimized while allowing the backbone of both the binder and EED to move, in an attempt to simulate the subtle adjustments that the EED loops might need to make to accommodate the binders. Backbone distance constraints were placed on the binder and EED to prevent clashes between the two from forcing the backbone into unrealistic conformations. No constraints were placed on the orientation of the binder relative to EED to allow dissociation of binders that couldn't be accommodated realistically.

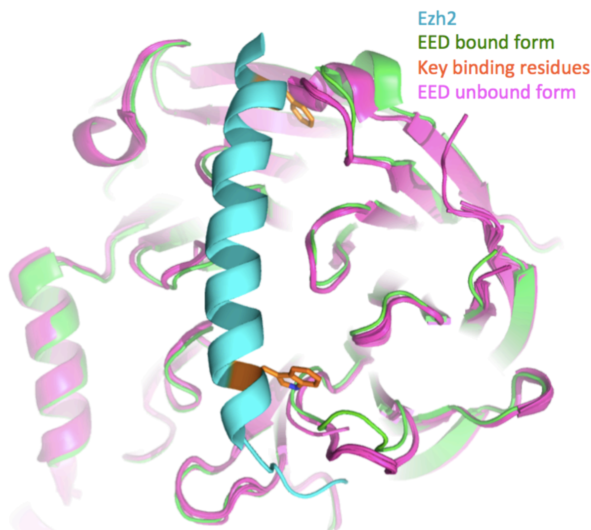
2.2.A.



2.2.B.



2.2.C.



2.2.D.

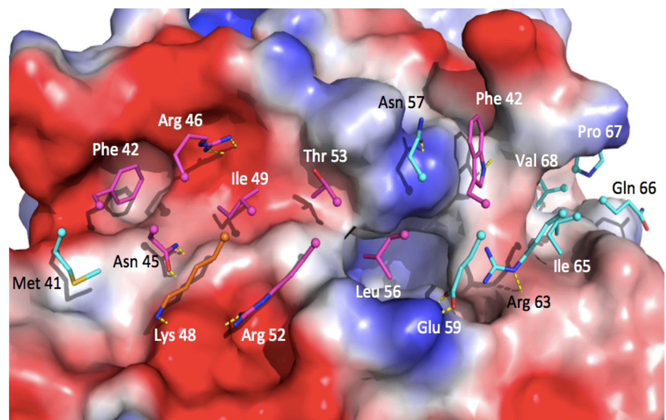


Figure 2.2: Observed flexibility of EED loops and design strategy

2.2.A. The crystal structure of the Ezh2 N-terminal helix (shown as a cyan cartoon and sticks) bound to EED (shown as an electrostatics map) from PDB ID 2qxv.

2.2.B. A preliminary Epigraft match using ApoE2 (PDB ID 1le2) as a scaffold (cyan) bound to EED (green) superimposed onto the crystal structure of EED-Ezh2 (orange) represented as cartoons. The key aromatic binding residues are shown as sticks and are highlighted with black circles.

2.2.C. The crystal structure of Ezh2 (cyan) bound to EED (green) superimposed onto 11 unbound structures of EED (PDB IDs 3iiw, 3iiy, 3ij0, 3ij1, 3ijc, 3jpx, 3jzg, 3jzh, 3jzn, 3k26, and 3k27)(magenta), each represented as cartoons. The aromatic binding residues are shown as orange sticks.

2.2.D. Ezh2 side chains (shown as sticks) making contact with EED (shown as an electrostatics map). The Cα atoms are shown as small spheres at the base of each side chain. Magenta side chains were grafted into all 8 binders. Orange side chains were grafted onto a subset of the binders. Cyan side chains were not grafted onto the binders.

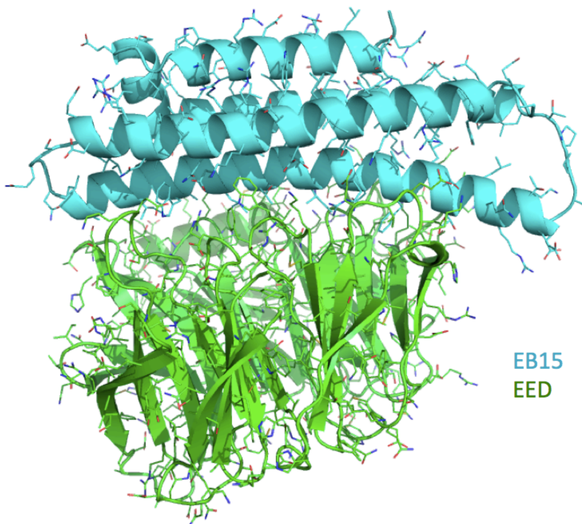
2.3 Experimental validation and characterization of EB15 and EB22

We synthesized the genes for each of the controls and candidate binders using a commercial vendor, (IDT, San Jose, California, USA, or Genescript, Piscataway, New Jersey, USA) and cloned these into pETCON, a modification of the yeast display vector pCTCON2 (103). We carried out yeast display as previously described and titrated soluble biotinylated human EED3 against the binders and controls displayed on the yeast surface under non-avid conditions (103). Soluble EED3 from 3 different biotinylation preparations showed no binding to 6 un-designed starting point proteins displayed on the surface of yeast (PDB IDs: 1le2, 3lf9, 1oz9, 1u84, 1wpa, and 1ya9) when labeled at 1 uM. This behavior of EED3 was unchanged regardless of the concentration of reducing agent or the amount of time the EED3 had been thawed and stored at 4°C, up to 2 weeks.

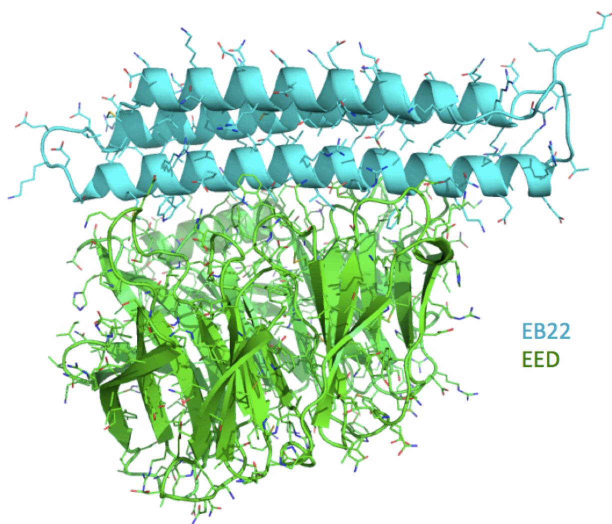
EB15, EB16, EB17, EB18, EB20, and EB22 displayed on the yeast showed 1.9 to 47-fold binding to 10 nM EED3 over background. Pre-incubating the EED with 32 uM soluble un-biotinylated Ezh2 dropped the binding signal to 0.92 to 2.1-fold over background, indicating that the designed binders must compete with Ezh2 for binding to EED3. EB15, EB16, EB17, EB18, and EB22 surface showed no binding to 1 uM soluble Mdm2, Mdm4, or RhoA, with the following exceptions: EB17 and EB18, both based on human occludin, showed 1.5-fold and 1.3-fold binding to RhoA over background, respectively, and EB20 showed 1.3-fold binding to Mdm2 over background. Yeast surface titration (YST) of EED3 against EB15, EB16, EB17, EB18, EB20, and EB22 showed that they bound EED3 with dissociation constants (Kds) ranging from 0.56 ± 0.049 nM to 13 ± 2.8 nM (**Table 2.3**). YST data was processed using Flowjo (Tree Star Inc, Ashland, OR, USA) and fit using either Kaleidagraph (Synergy Software, Reading, PA, USA) or a Java nonlinear least squares algorithm (http://statpages.org/nonlin.html). EB15 and

EB22 were selected for further characterization and affinity maturation due to their tight binding, high degree of display on the yeast surface, and central location of the grafted Ezh2 region on the binder (**Figures 2.3.A, 2.3.B, 2.3.C,** and **2.3.D**).

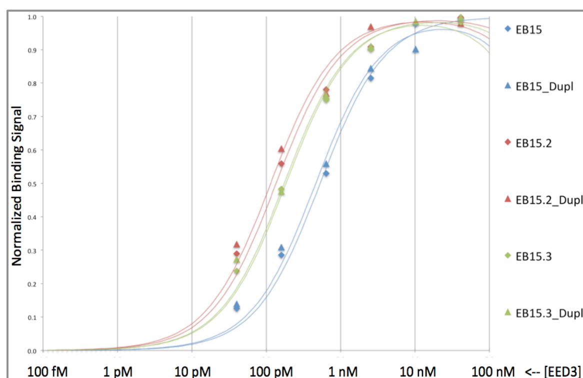
2.3.A.



2.3.B.



2.3.C.



2.3.D.

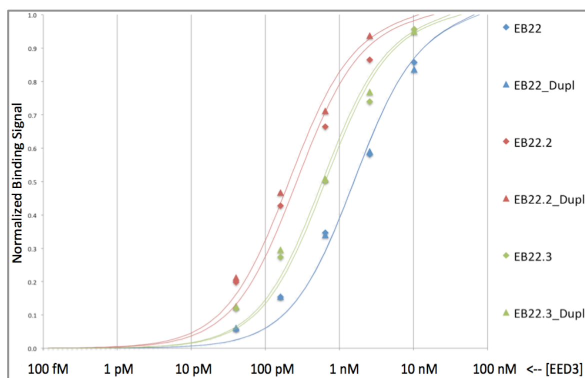


Figure 2.3: Experimental validation and characterization of EB15 and EB22.
2.3.A. Model of EB15 (cyan) bound to EED (green) shown in cartoon and stick representation.
2.3.B. Model of EB22 (cyan) bound to EED (green) shown in cartoon and stick representation.
2.3.C. YST titrations of EB15 (blue curve) and selected point variants (red and green curves).
2.3.D. YST titrations of EB22 (blue curve) and selected point variants (red and green curves).

Table 2.3: Average YST dissociation constants for initial EED binding proteins. All values are given in nanomolar.

Binder	YST EED3 Kds (nM)	Fold signal loss at 10 nM EED3 if pre-incubated with Ezh2
EB15	0.56	33
EB16	2.1	5.5
EB17	12	7.6
EB18	13	19
EB19	Not determined	Not determined
EB20	Not determined	2.1
EB21	Not determined	Not determined
EB22	1.9	27

2.4 Binding mode validation and affinity maturation of EB15 and EB22

We sought to validate the folded state and binding mode of EB15 and EB22 and improve their affinity for EED3 through deep mutational scanning as previously described (104, 105). Using overlap PCR, we individually mutated to all possible amino acids every residue of EB15 and EB22 (3024 unique variants for EB15 and 2352 unique variants for EB22). We displayed the resulting library on the surface of yeast and screened it using fluorescent-activated cell sorting (FACS)(103, 106) for 3 rounds, stringently selecting for clones having maximal affinity, maximal thermal stability, maximal on-rate, and minimal off-rate. We extracted the plasmid DNA from cells saved from the unselected library and from each round of sorting and subjected the pools of EB15 and EB22 variants to high-throughput next-generation sequencing. The fitness (enrichment ratio) of a given mutation under a given selection condition was calculated as the ratio of the frequency of that mutation in the selected pool relative to the frequency of that mutation in the pool selected only for display on the yeast surface. Throughout this chapter, those residues of EED3 visible in the crystal structure 2qxv (93) are arbitrarily numbered 1-352, with designed binding proteins beginning at residue 353.

Under all selection conditions residues modeled at the N-terminal half of the EB15 binding interface with EED3 were strictly conserved. Many residues modeled at the C-terminal half of the EB15 binding interface with EED3 showed enrichment of polar and positively charged residues under less-stringent selection conditions, suggesting that EED3 bears negative character in this region not sufficiently countered by the EB15 design and further that EB15 buries hydrogen-bonding groups on EED3. For example, EB15 Threonine 440 (Ezh2 Threonine 53) shows enrichment for positively charged residues, Methionine 442 (Ezh2 Threonine 55) shows enrichment for Asparagine, Leucine 443 (Ezh2 Leucine 56) shows enrichment for Alanine, Methionine, positively charged residues, and uncharged polar residues, Glutamate 444 (Ezh2 Asparagine 57) shows enrichment for Glutamine and Threonine, Leucine 446 (Ezh2 Glutamate 59) shows enrichment for Valine, Tyrosine, negatively charged residues, or uncharged polar residues, and Isoleucine 450 (Ezh2 Arginine 63) shows enrichment for Methionine and all polar residues. In most cases, the enriched mutations allowed formation of a new polar interaction with a complementary charge or partial charge on either EB15 or EED3 (**Figure 2.4.A**).

Under all selection conditions, EB15 Tryptophan 447 (Ezh2 Tryptophan 60) shows enrichment for every other amino acid except Cysteine, Phenylalanine, Isoleucine, Leucine, Methionine, and Valine, suggesting that if the Tryptophan is not positioned exactly as in Ezh2, it hinders binding rather than stabilizing it. The most enriched amino acids at this position are Arginine and Histidine, likely to balance the negative charge of the disfavored EB15 Glutamate 444 nearby. All of the enriched mutations were consistent with the design model if side-chain conformational sampling of the mutation and surrounding residues is allowed. Mutations L446VY, W447PY, I450P, and A454R do not fit well with the model, suggesting that these mutations require subtle modification of the backbone of either EB15 or EED3 or adjustment of

the rigid-body transformation between them. EB15 core residues were strictly conserved even at less stringent selection conditions, validating that EB15 is properly folded and suggesting that it binds EED3 using the modeled binding mode (**Figures 2.4.C** and **2.4.D**). Full Log2 enrichment heatmaps for all selection conditions are given in the Appendix (**Section 4.1**).

Under all selection conditions the key residues of the modeled EB22 binding interface with EED3 were strictly conserved. EB22 residues Alanine 394, Leucine 405, Threonine 410, Alanine 412, Leucine 413, Alanine 416, Tryptophan 417, Leucine 420, and Methionine 421, however, showed enrichment of polar residues, suggesting that EB22 either buries hydrogen-bonding groups on EED3 and/or does not complement nearby EED3 charged residues well. In most cases, the enriched mutations allowed formation of a new polar interaction with a complementary charge or partial charge on either EB22 or EED3. Nearly all of the enriched mutations are consistent with the design model if side-chain conformational sampling of the mutation and surrounding residues is allowed. Similar to the corresponding position of EB15, EB22 Tryptophan 417 shows enrichment of nearly all other amino acids, except Cysteine, Phenylalanine, Glycine, Isoleucine, Leucine, Methionine, Threonine, and Valine, once again suggesting that if the Tryptophan is not positioned exactly as in Ezh2, it precludes binding rather than aiding it. Notably, some EB22 enrichments recapitulate interactions made by Ezh2 with EED3 that were not included in the design model, such as A416DE (which would interact with EED3 Arginine 200), L420KR (which would interact with the backbone carbonyls of EED3 Arginine 200 and Proline 199) and L408KR (which would interact with Ezh2 Aspartate 309). EB22 mutations A394R, T410H, A412NKQR, M421H, and Proline substitutions at EB22 residues 371, 375, 394, 411, 412, 416, 417, 419, 420, 421, 422, 431, 432, 435, 437, 444, and 445 do not fit well with the model, suggesting that these mutations require modification of the

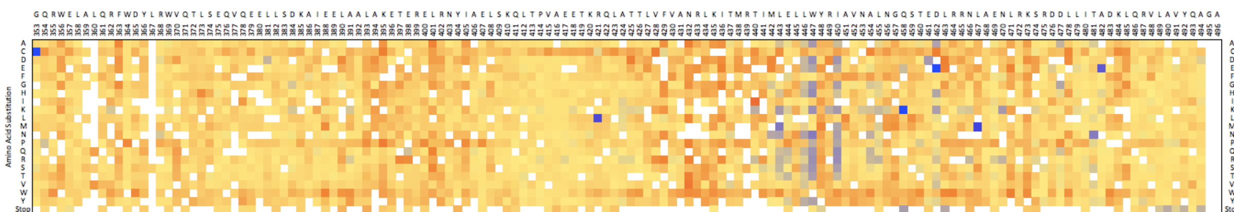
backbone of either EB22 or EED3 or adjustment of the rigid-body transformation between them. Strict conservation of key binding residues on EB22 suggests that it binds with the correct binding mode, but enrichment of substitutions not consistent with the modeled backbone conformation of EB22 suggests that it may not be well folded (**Figure 2.4.B**).

Single-site mutant enrichment data from less-stringent selection conditions showed that 9 EB22 core hydrophobic residues (Alanine 361, Methionine 365, Leucine 372, Leucine 376, Valine 397, Alanine 400, Alanine 436, Isoleucine 440, Isoleucine 451, and Leucine 462) became enriched for polar substitutions, all of which would be buried in the core of EB22 in the modeled 3-helix bundle conformation. Interestingly, if the N-terminal helix of EB22 detached from the other 2 helices, all of these substitutions would become solvent exposed. Additionally, each of the residues in the 4-Glycine linker that replaced the previously removed beta sheet domain showed enrichments for larger aliphatic and aromatic hydrophobic residues. This enrichment data suggested to us that the N-terminal helix might less stably interact with the other 2 helices and that the region surrounding the 4-Glycine linker lacked sufficient hydrophobic packing in its core. The data don't suggest that EB22 is inherently unstable as much as that removal of the N-terminal helix can aid in binding to EED3 relative to other mutations (**Figures 2.4.E** and **2.4.F**).

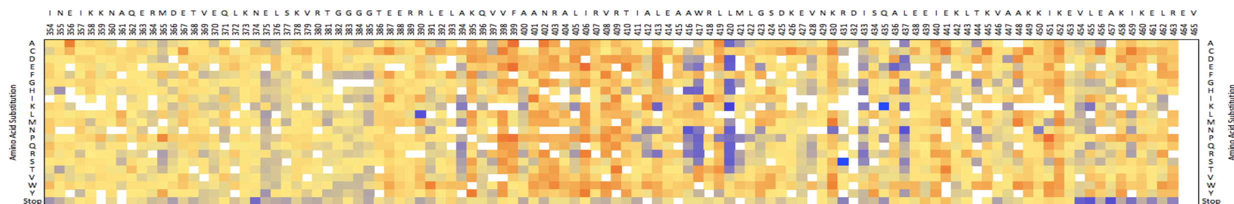
As EB22 is based on the PDB structure 3lf9 and both 3lf9 and PDB structure 3lhp are based on the PDB structure 1ise (Escherichia coli ribosome recycling factor), EB22 and 3lhp share nearly identical fold topologies and core residue sequences (107). EB22 differs from 3lhp in that EB22 has 7 residues of loop secondary structure from residues 379-385 while 3lhp has 6 residues of helical secondary structure followed by 4 residues of loop secondary structure in the corresponding region. 3lhp also has more hydrophobic packing in the underlying core region, including a Phenylalanine and Tryptophan at positions where 3lf9 carries Valine and Leucine,

respectively. We sought to improve the stability of EB22 by replacing the 7 residues of loop secondary structure with the corresponding region from 3lhp and modifying the underlying core region to have the sequence observed in 3lhp. These changes were incorporated into EB22 during construction of the combinatorial library (**Figures 2.4.G** and **2.4.H**).

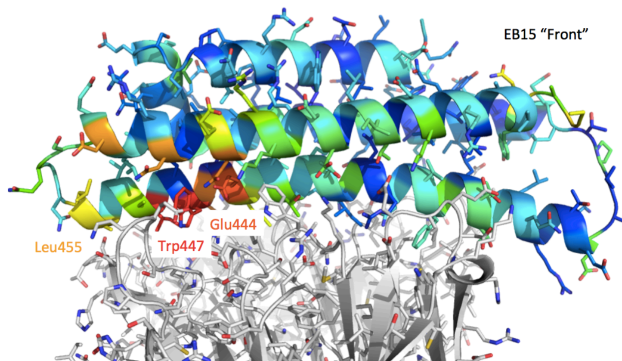
2.4.A.



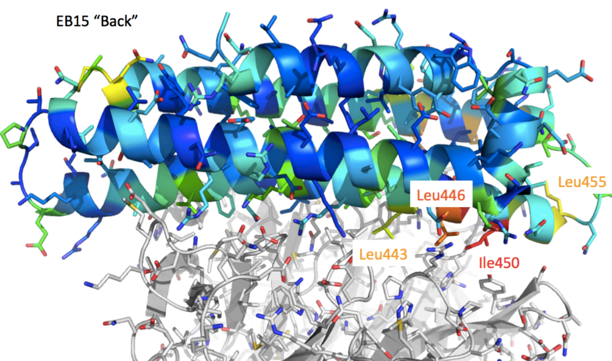
2.4.B.



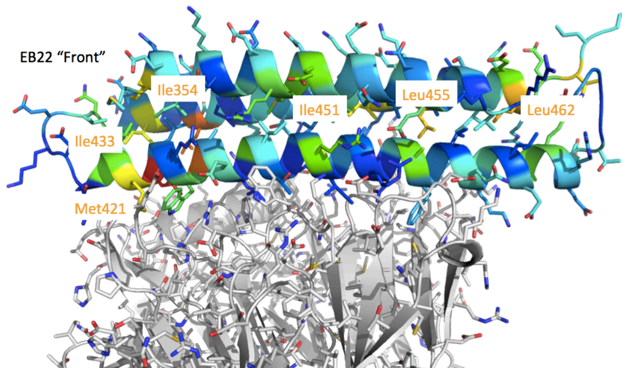
2.4.C.



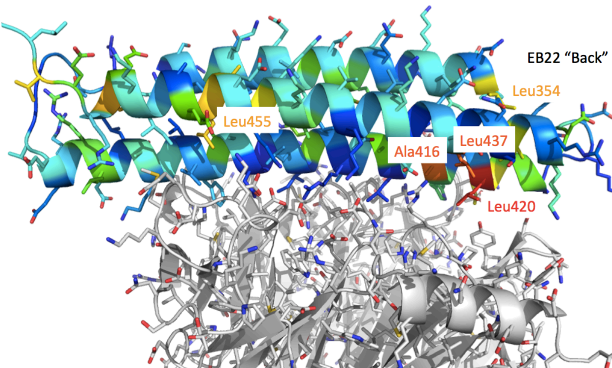
2.4.D.



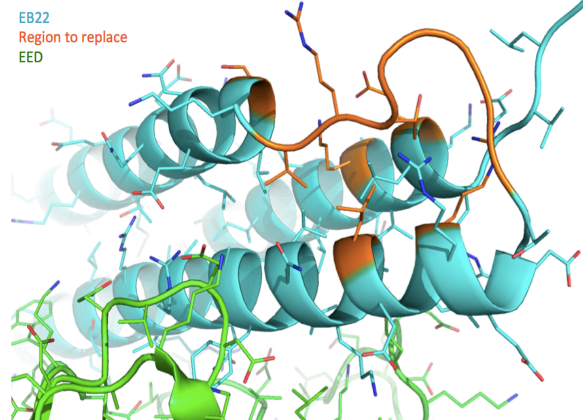
2.4.E.



2.4.F.



2.4.G.



2.4.H.

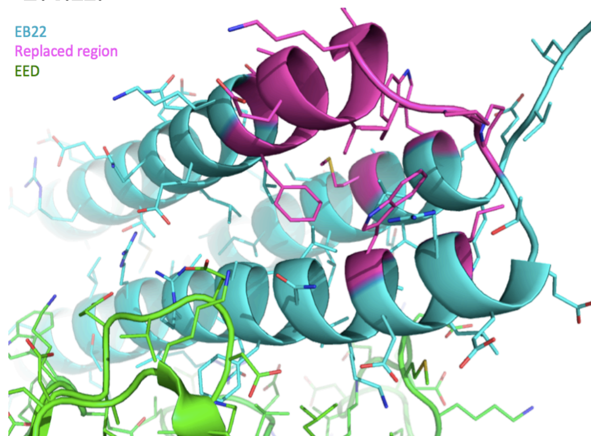


Figure 2.4: Binding mode validation and affinity maturation of EB15 and EB22
2.4.A. Heatmap of Log2 enrichments of EB15 single-site variants when selected for maximal affinity for EED3, 1st round. The X-axis gives sequential positions in the protein while the Y-axis gives all possible amino acid substitutions. Orange-colored boxes denote a depleted substitution, yellow boxes denote a neutral substitution, blue boxes denote an enriched substitution, and white boxes denote no data. The scale goes from -8.6 to +6.3.
2.4.B. Heatmap of Log2 enrichments of EB22 single-site variants when selected for maximal affinity for EED3, 1st round. The X-axis gives sequential positions in the protein while the Y-axis gives all possible amino acid substitutions. Orange-colored boxes denote a depleted substitution, yellow boxes denote a neutral substitution, blue boxes denote an enriched substitution, and white boxes denote no data. The scale goes from -7.8 to +6.5.
2.4.C. Cartoon and stick representation of model of EED (white) bound to EB15, "front" view, colored according to the enrichment ratio of the most-enriched substitution at each position when sorted for affinity, 2nd round. The scale goes from blue for a Log2 enrichment of -3.8 to red for an enrichment of +7.6. Highly enriched residues are labeled.
2.4.D. Cartoon and stick representation of model of EED (white) bound to EB15, "back" view, colored according to the enrichment ratio of the most-enriched substitution at each position when sorted for affinity, 2nd round. The scale goes from blue for a Log2 enrichment of -3.8 to red for an enrichment of +7.6. Highly enriched residues are labeled.
2.4.E. Cartoon and stick representation of model of EED (white) bound to EB22, "front" view, colored according to the enrichment ratio of the most-enriched substitution at each position when sorted for affinity, 2nd round. The scale goes from blue for a Log2 enrichment of -2.7 to red for an enrichment of +8.8. Highly enriched residues are labeled.
2.4.F. Cartoon and stick representation of model of EED (white) bound to EB22, "back" view, colored according to the enrichment ratio of the most-enriched substitution at each position when sorted for affinity, 2nd round. The scale goes from blue for a Log2 enrichment of -2.7 to red for an enrichment of +8.8. Highly enriched residues are labeled.
2.4.G. Image of model of EB22 (cyan) bound to EED (green) shown in cartoon and stick representation. The region of EB22 that was replaced with the analogous region of 3lhp is colored orange, shown before the replacement.
2.4.H. Image of model of EB22 (cyan) bound to EED (green) shown in cartoon and stick representation. The region of EB22 that was replaced with the analogous region of 3lhp is colored orange, shown after the replacement.

Seven to eight of the most enriched positions from EB15 and EB22 were simultaneously mutated to a small number of enriched residue types in 2 combinatorial libraries and screened against their previous selection conditions, respectively. The most represented 6 clones each for EB15 and EB22 after 5 rounds of sorting were sub-cloned into pET29b (Novagen (EMD) Biosciences, Inc, Madison, WI, USA), expressed in BL21star E. coli cells (ThermoFischer Scientific, Waltham, MA, USA), and further characterized by biolayer interferometry (BLI)

using an OctetRED96 system (Forte Bio, Menlo Park, California, USA). Preliminary results show that the EB15, EB22, and their variants all bind EED3 and suggest that their Kds lie below 1 nM. YST and BLI affinities for all variants so tested are given in the Appendix (**Section 4.2**). Amino acid sequences for all variants tested are given in the Appendix (**Section 4.3**). Mutations made in the combination libraries and observed in the top clones are given in the Appendix (**Section 4.4**). Further ongoing work is required to determine what effect the top EB15 and EB22 variants will have in human cancer and stem cells.

2.5 Conclusions and future directions

We have used Rosetta to stabilize the bound state of the Ezh2 EED-binding motif by incorporating it into a number of larger host proteins. Using deep mutational scanning and YST, we have validated that 2 of the resulting binding proteins, EB15 and EB22, bind to human EED3 at the correct site and using the designed binding surface. We have validated that our preparations of EED3 and our binders do not bind nonspecifically and that EB15 is properly folded. Using YST, we evolved EB15 and EB22 for greater affinity EED3. EB15, EB22, and their improved variants can be expressed in E. coli and tightly bind EED3 as purified proteins. As these binding proteins present backbone conformations to EED that subtly differ from Ezh2, our results suggest that the Ezh2-binding cleft on EED exhibits conformational flexibility. We have shown that the Rosetta models are sufficiently accurate to successfully target the large, polar Ezh2-binding pocket of EED. We also further validate the idea that stabilization of a disordered peptide results in massive affinity gains. Further work will determine how well variants of EB15 and EB22 behave in mammalian cells, whether they bind EED3 in vitro, and what effect this binding has on PRC2 activity in cancer and stem cells.

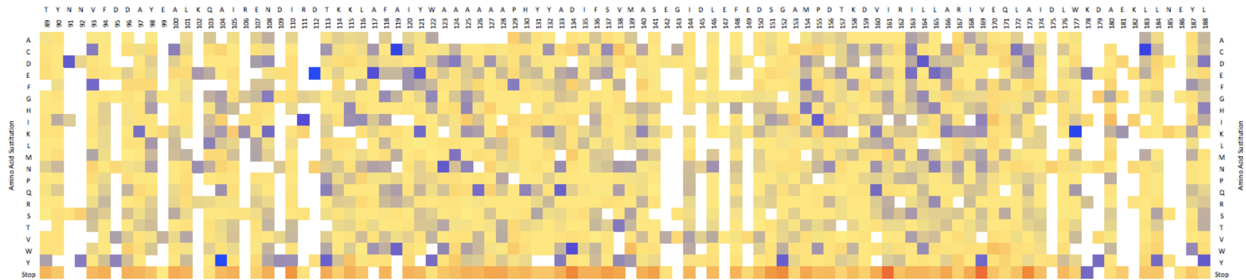
Section 3 (Appendix)

Supplemental data for the design of proteins to inhibit the interaction of Mdmx-p53

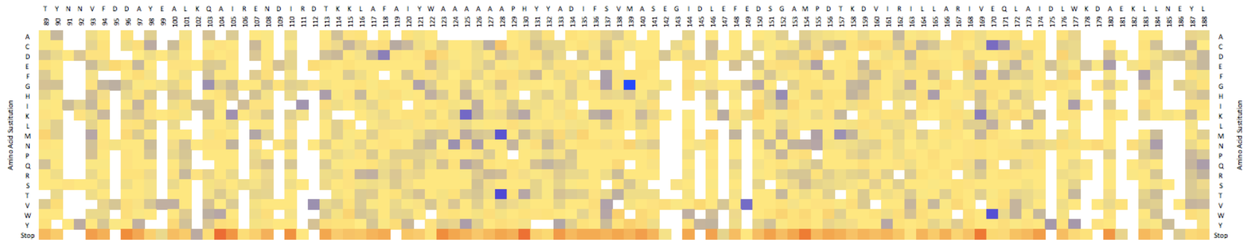
3.1 (Appendix) MB17 single site mutation enrichment heatmaps

Figure 3.1: Heatmaps showing the Log2 enrichment ratio for each unique sequence from each round of selection are shown below. Orange color denotes sequences that were depleted relative to the parent pool. Blue color denotes sequences that were enriched relative to the parent pool. Yellow color denotes sequences that were neither depleted nor enriched relative to the parent pool. Enrichment ratios for display-only selections were calculated relative to the unselected pool, while enrichment ratios for binding selections were calculated relative to the display-only selected pool.

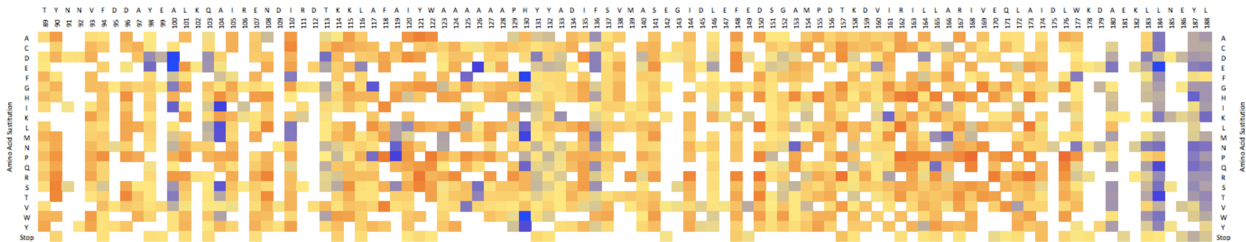
MB17 SSM sorted for display after induction at 30°C in CG-UT media:



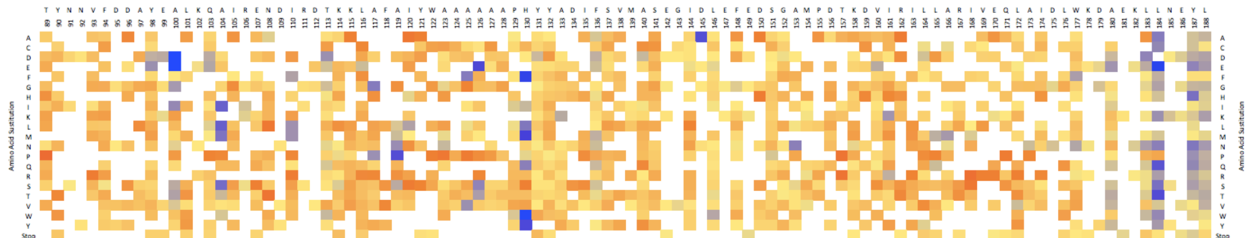
MB17 SSM sorted for display after induction at 22°C in SGCAA media:



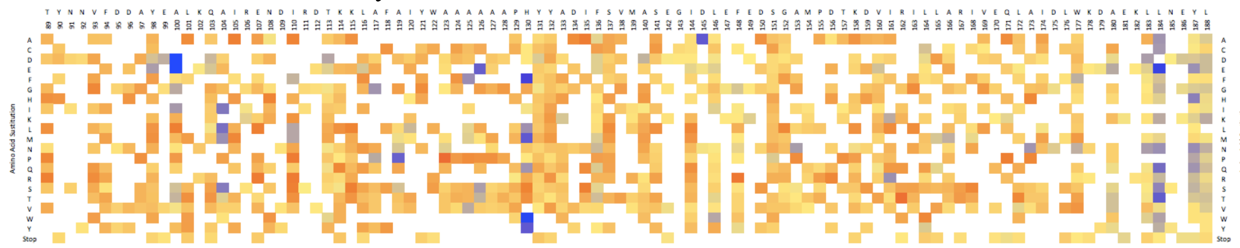
MB17 SSM sorted for affinity toward Mdmx, 2nd round:



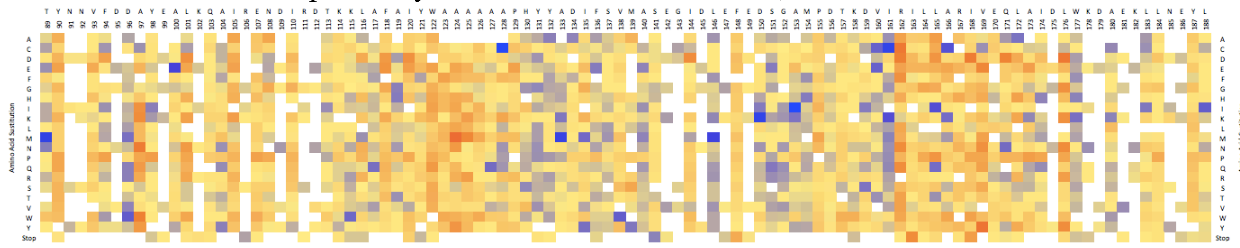
MB17 SSM sorted for affinity toward Mdmx, 3rd round:



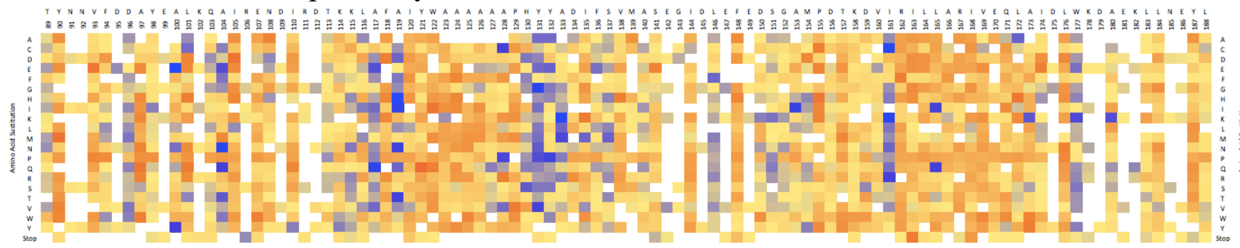
MB17 SSM sorted for affinity toward Mdmx, 4th round:



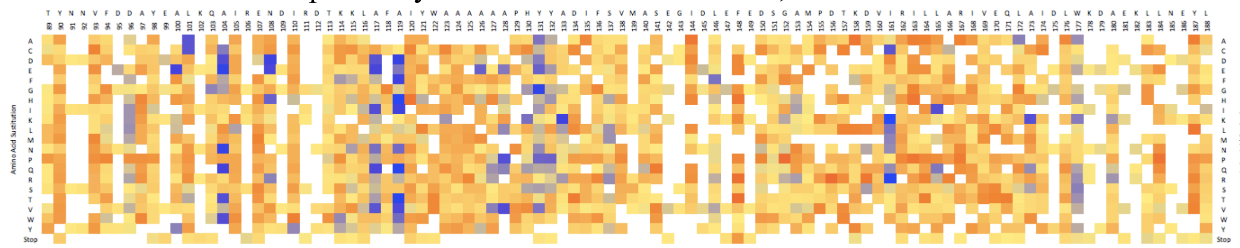
MB17 SSM sorted for specificity toward Mdmx over Mdm2, 1st round:



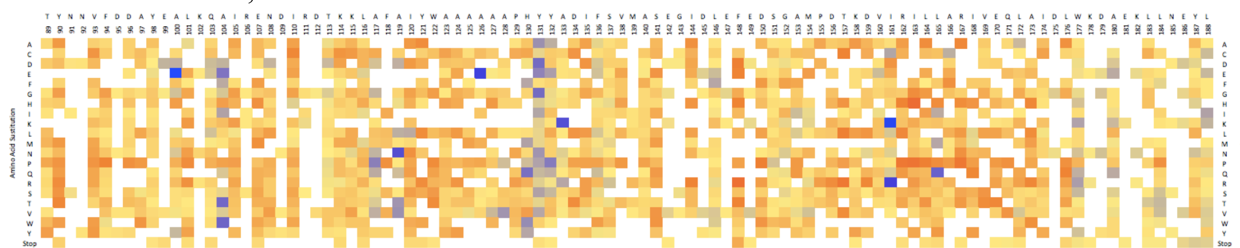
MB17 SSM sorted for specificity toward Mdmx over Mdm2, 2nd round:



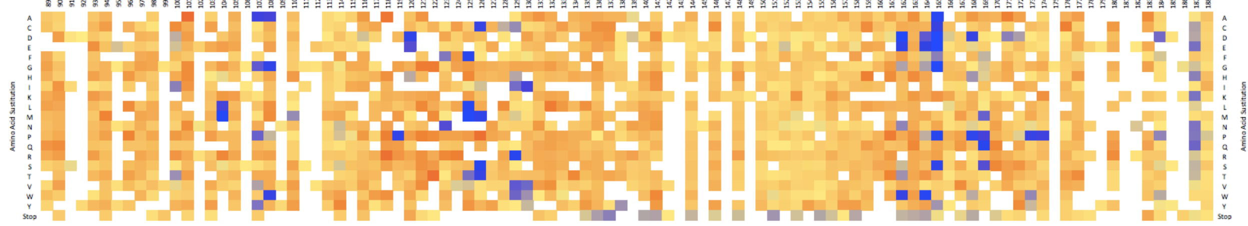
MB17 SSM sorted for specificity toward Mdmx over Mdm2, 3rd round:



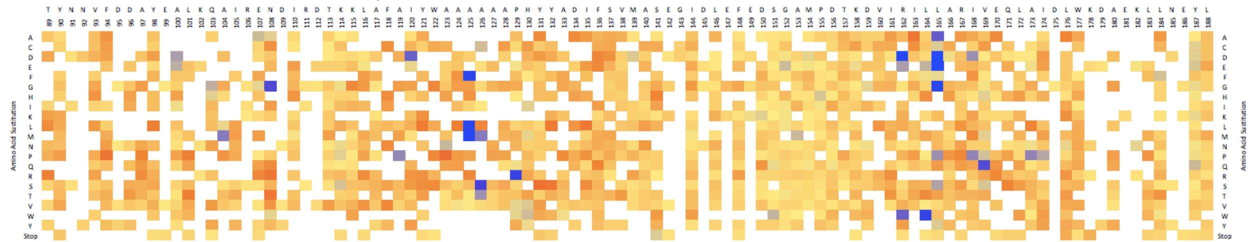
MB17 SSM sorted for affinity, 2nd round added to MB17 SSM sorted for specificity toward Mdmx over Mdm2, 3rd round:



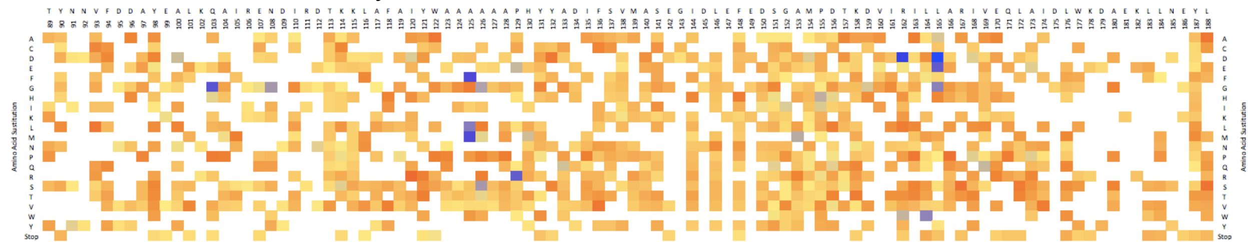
MB17 SSM sorted for affinity toward Mdm2, 2nd round:



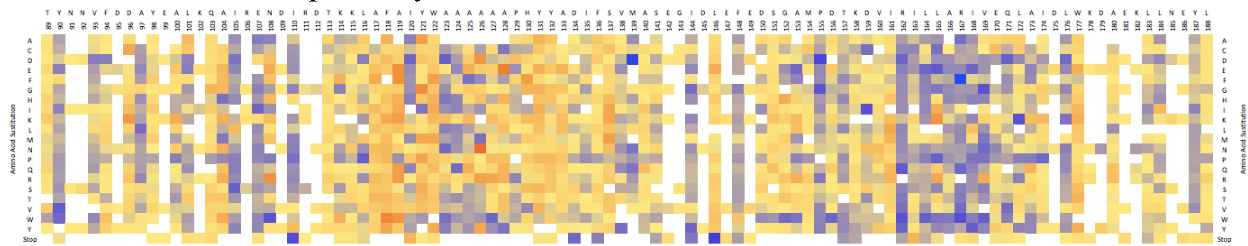
MB17 SSM sorted for affinity toward Mdm2, 3rd round:



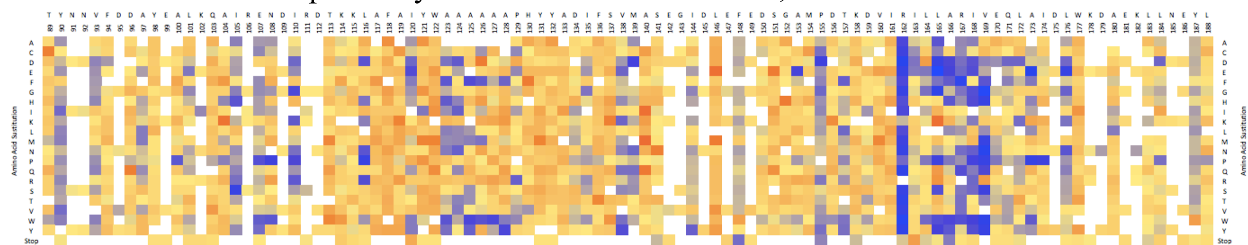
MB17 SSM sorted for affinity toward Mdm2, 4th round:



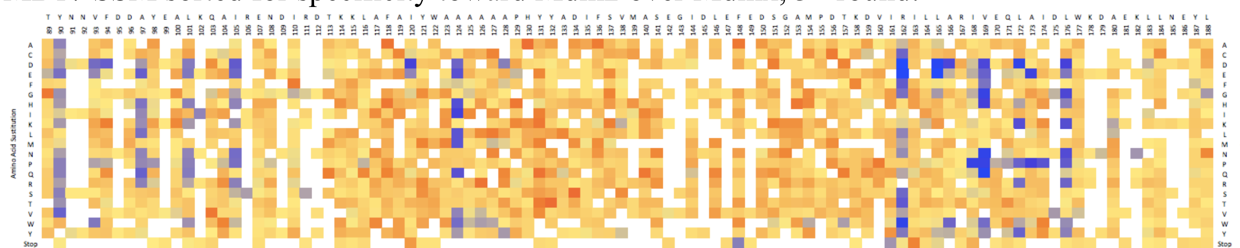
MB17 SSM sorted for specificity toward Mdm2 over Mdmx, 1st round:



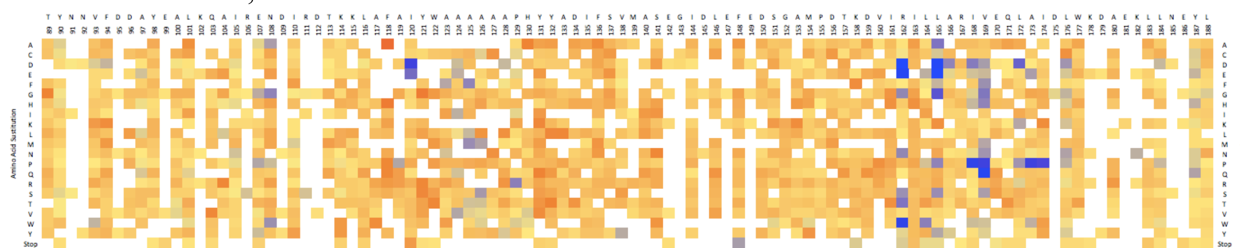
MB17 SSM sorted for specificity toward Mdm2 over Mdmx, 2nd round:



MB17 SSM sorted for specificity toward Mdm2 over Mdmx, 3rd round:



MB17 SSM sorted for affinity, 2nd round added to MB17 SSM sorted for specificity toward Mdmx over Mdm2, 3rd round:



3.2 (Appendix) Yeast surface display and ITC affinities for Mdmx and Mdm2 binders and variants

Table 3.2: YST and ITC affinities of designed binders and evolved variants are given in the table below. The degree of display of each protein on the surface of yeast is denoted with "+++" for 30-fold or greater display signal above that of the non-displaying population, "++" for 7 to 30-fold, "+" for 1 to 7-fold, or "-" for no display signal above that of the non-displaying population. Only non-avid YST KD values are reported. Bolded YST and ITC values are of higher confidence than non-bolded values. Lower confidence YST values were due to insufficient labeling time to reach equilibrium at low Mdmx and Mdm2 concentrations, YST fitted Kds above 2 uM, disagreement between replicates, lack of replicates, or overly long incubation with the secondary label. Lower confidence ITC values were due to inaccurate quantification of the Mdmx and Mdm2 proteins used for the titration.

Design or variant name	**Degree of Display**	**YST Mdmx Kds (nM)**	**ITC Mdmx Kds (nM)**	**YST Mdm2 Kds (nM)**	**ITC Mdm2 Kds (nM)**
Mdmx controls					
p53 (peptide)	+++	1200, 4300		**1900, 2000,** 3900	
VH9_SCFV	+++	22		180	
Round 1 designs					
MB1	-	Not determined		Not determined	
MB2	+++	2200		1400	
MB3	++	No binding		No fit	
MB4	++	850		69, 250	
MB5	+	No fit		No binding	
MB6	-	Not determined		Not determined	

MB7	+	No fit		No binding	
MB8	+++	No binding		No fit	
MB9	+	No binding		No binding	
MB10	-	Not determined		Not determined	
MB11	++	No fit		540	
MB12	++	No fit		No binding	
Round 2 designs					
MB13	-	Not determined		Not determined	
MB14	+++	16000		16000	
MB15	++	No binding		No binding	
MB16	+	Not determined		Not determined	
MB17	+++	**20, 17, 30**, 46	22, **39**	**1100, 1700, 1600, 1900**	**3800**
MB18	-	Not determined		Not determined	
MB19	+	3500		4600	
MB20	-	Not determined		Not determined	
MB21	-	Not determined		Not determined	
MB22	-	Not determined		Not determined	
MB23	-	Not determined		Not determined	
MB24	-	Not determined		Not determined	
MB25	++	3000		110	
MB26	++	32000, **480, 330**		**1500**, 480, **930**	
MB19CFX	+	2700, 4100		130, 880	
MB25CFX	+	**490, 490**		210, 600	
MB26CFX	+	2500, 380		570, 1000	
MB17 Mdmx specificity single-point variants					
MB17_A119D	+++	**22, 39**		7600	
MB17_A119Q	+++	**9.4, 24**		6400, 8100	
MB17_A119T	+++	**12, 26**		5700	
MB17_N108D	+++	73		1200	
MB17_N108E	+++	86		870	
MB17_I161K	+++	12		2000	
MB17_I161R	+++	16		2800	
MB17_L165R	+++	34		5900	
MB17_A128P	+++	16		2200	
MB17_Y131G	+++	16		3200	
MB17_Y131S	+++	13		1800	

MB17 specificity combinatorial library averages					
MB17_Speci_CL_Rd_2.5	+++	1000		4200	
MB17_Speci_CL_Rd_2.6	+++	2100		750	
MB17_Speci_CL_Rd_2.7	+++	4000		190	
MB17_Speci_CL_Rd_3.4	+++	72		190	
MB17_Speci_CL_Rd_3.5	+++	41		120	
MB17_Speci_CL_Rd_3.6	+++	53		150	
MB17_Speci_CL_Rd_4.5	+++	6.9		1000	
MB17_Speci_CL_Rd_4.6	+++	6.1		1000	
MB17_Speci_CL_Rd_4.7	+++	6.1		870	
MB17_Speci_CL_Rd_4.8	+++	5.6		950	
MB17 specificity combinatorial library clones					
MB17_clone_4.2	+++	14		1800	
MB17_clone_4.4	+++	20		1500	
MB17_clone_4.5	+++	15		370	
MB17_clone_4.6	+++	13		490	
MB17_clone_4.19	+++	15		1000	
MB17_clone_2.27	+++	190		420	
MB17_clone_2.30	+++	8800		200	
MB17_clone_2.36	+++	6500		94	

MB17_clone_ 2.33	+++	13000		1500	
MB17_clone_ 2.39	+++	2200		230	
MB17_clone_ 5.48	+++	400		13	
MB17.38 (MB17_clone_ 5.49)	+++	380		460	
MB17_clone_ 5.53	+++	2900		980	
MB17.35 (MB17_clone_ 5.55)	+++	31		120	
MB17_clone_ 5.60	+++	160		31	
MB17_clone_ 5.61	+++	740		24	
MB17_clone_ 5.62	+++	4200		83	
MB17.36 (MB17_clone_ 5.63)	+++	180		57	
MB17_clone_ 5.64	+++	4800		330	
MB17.38 (MB17_clone_ 5.67)	+	190, 1100		180, 47	
MB17_clone_ 5.68	+++	160		22	
MB17_clone_ 6.50	+++	11000		100	
MB17_clone_ 6.53	+++	6500		11000	
MB17_clone_ 6.55	+++	2400		17000	
MB17.37 (MB17_clone_ 6.56)	+	**120, 86**		150, 2300	
MB17_clone_ 6.58	+++	340		28	
MB17_clone_ 6.59	+++	1600		170	
MB17_clone_ 6.60	+++	440		50	

MB17_clone_ 6.61	+++	4100		No fit	
MB17_clone_ 6.62	+++	110		21	
MB17_clone_ 6.63	+++	1000		43	
MB17_clone_ 6.64	+++	170		33	
MB17_clone_ 6.65	+++	3100		490	
MB17_clone_ 6.66	+++	93		5.3	
MB17_clone_ 6.69	+++	5100		39	
MB17 specificity combinatorial library top variants					
MB17.21	+++	5000, 27000, 4700	**9200**	**42, 46, 56**, 120, **98**, 430, **55**	57, **21**
MB17.31	+++	**120, 110**		55, 120	
MB17.32	+++	140, 59		72, 230	
MB17.33	+++	400		250	
MB17.41	+++	**9.9**, 23, **7.5**, **15**, **13**, 27, **4.5**	**3**	3700, 1600, 53000, 6300	
MB17 affinity combinatorial library variants					
MB17.22 (FMW)	+++	1400, 2600, 4300		1.7, 1.6, 0.019, **0.096, 0.101**	
MB17.MAL	+++	**17, 24**		**2.2, 1.3**	
MB17.AML	+++	440, 180		**2.9, 2.9**	
MB17.VAW	+++	**150, 120**		**4.4, 3.4**	
MB17.42 (TAR)	+++	9.5, 17, **0.83**, 1.3, **0.38**		2300, 1000, 4600, 4500	
MB17.TVR	+++	9.2, 14, **0.56**		1600, 7900	
MB17.TAK	+++	7.9, 11, **0.32**		**310, 300**, 4030	
MB17.TVK	+++	10, 16, **0.56**		**210, 270**, 4300	
MB17.VVR	+++	10, 22		**850, 920**	
MB17 charge-balance redesigns					
MB17.23	+++	9900, 26000		12000, 7900, 4700	

MB17.24	+++	**1300, 1400**		**0.22, 0.25, 0.13**	
MB17.25	+++	51000		6800, 3300, 3400	
MB17.26	+++	63000		1900, **430, 440**	
MB17.27	+++	No fit		7200	
MB17.28	+++	738		**0.17, 0.13, 0.13**	
MB17.43	+++	**160**, 660, **355**		84000	
MB17.44	+++	**170**, 710, **270**		No fit	
MB17.45	+++	1100, **370, 360**		No fit	
MB17.46	+++	110, **11, 9.0**		No fit	
MB17.47	+++	58, **9.9, 11**		No fit	
MB17.48	+++	1.3, **0.33, 0.37**		No fit	

3.3 (Appendix) Scaffolds and amino acid sequences of Mdmx binders and variants

Table 3.3: Amino Acid sequences of designed binders and their evolved variants are given in the table below. The NdeI site on the N-terminus and the XhoI site on the C-terminus are given for reference, but other residues pertaining to the yeast display vector are omitted.

Name	Scaffold PDB ID	Amino Acid Sequence
Mdmx controls		
p53 (peptide)	1ycr	HMETFSDLWKLLPELE
VH9_SCFV	2vyr	HMEVQLLESGGGLVQPGGSLRLSCAASGFTFEEYAMLWVRQAPGKGLEWVSGINARGYTTYYADSVKGRFTISRDNSKNTLYLQMNSLRTEDTAVYYCAKPWYPFMASKGSEFDYWGQGTLVTVSSAAALEIKRASQPELAPEDPEDVELE
Round 1 designs		
MB1	1pgv	HMTDVESAINRLREDDTDLKEVNINNMKRVSKERIRSLIEAAVNSKHIEKHMFSLANTAISDSEARGLIELIETSPSLRVLNVESNFLTPELLARLLRSTLVTQSIVEFKADNQRQSVLGNQVEMDMMMAIEQNESLLRVGIEFKSFLAIWLVMVALYFNYEEVQLRRLGKLE
MB2	2byg	HMFQSMTVVEIKLFKGPKGLGFSIAGGVGNQHIPGDNSIYVTKIIDGGAAQKDGRLQVGDRLLMVNNYALEEVTHFWAAWILMATDEVVYLKVGKPTTIY
MB3	2fi9	HMHFPGRAPIDAYGNGGFRFADMSHRGSIICIPSGIYGIDMTGPVPTQEDISRVLQESRQIEVLLIGTGVELLQLPFLLIWLLMMEEISSDTMSTGAAVRVFNVLLAEDRAVAALLFAVELE
MB4	2olm	HMNFWQQWQEAFHLWMLREMTGLPHNRKCFDCDQRGPTYVNMTVGSFVCTSCSGSLRGLNPPHRVKSISMTTFTQQEIEFLEKHGNEVAKQIWLGLFDDRSSAIPDFRDPQKVKEFLQEKYEKKRWYVPPEQAKVLE

MB5	2pww	HMAENWPHTGLEQKEVAFSIVNHAAKSLGFIHVDQWDYERVM FDYKIVHHQGTFYLRVPAYAVKGEIPRPSTIVQIMTPILGKY YYPHGVEYEGRTFPFLVLWEAMATLWFLALTIDLE
MB6	2vac	HMSMGIHASFLLIWFFMWALWGSVRLIKVVIEDEQLVLGASQ EPVGRWDQDYDRAVLPLLDAQQPCYLLYRLDAQNAQGFEWLF LAWSPDNSPVRLKMLYAATRATVKKEFGGGHIKDELFGTVKD DLSFAGYQKHLE
MB7	2vim	HMQVLATAFTLYWLIMSNKGQLIVVDFFAQWCGPCRNIAPKV EALAKEIPEVEFAKVDVDQNEQAAANYSVTAMPTFVFIKDGE EVDRFSGANEAKLRETITRHKLE
MB8	3eye	HMNILLTRIDNQLVDGRVGVTWTSTIGANLLVVVDQVVADEP IQQWIMAFVAAAYGFGIQFFTIEQTINVIGKAAPHQKIFLIC RTPQTVRKLVEGGIDLKDVNVGNMHFSEGKKQISSKVYVDDQ DLTDLRFIKQRGVNVFIQDVPGDQKEQIPLE
MB9	3foj	HMESITVFLLLWFFLLADPVNIVDVRTDQETAMGIIPGAETI PMNSIPDNLNYFNDNEVYYIICKAGGRSAQVVQYLEQNGVNA VNVQGGMDQFGDQGLEHLE
MB10	3fyq	HMSRGTQACINAAHTVSGIIGDLNTTIMFATAGTLHSDGDGS FADHREHILQTAKALVEDTQVLETGASGTQDQLANAAQNAVS TITQLAEAVKRGASSLGSTQPQSQVMVISAVQQVAQALGDLI NSTKLASGKAINDPSMFLLAWSAMFMVWMVSQLLETVKAVED QHTRGTRAMEATVEAISQEIRALEHLE
MB11	2o37	HMVKETKLYDLLGVSPSAIEFLLIWGYMFAAYEYHPDEPTGD TEKFKEISEAFEILDDPQKREIYDQYGLEAARSGGPSFGPLE
MB12	1zma	HMSNAMEQFLDNIKDLEVTTVVRALEAALKKETATFFIGRKT CPYCRKFAGTLSGVVAETKAHIYFINSEEPSQFFALFWFMMA YGIPTVPGFVHITDGDANVRSDSSMSAQEIKDFAGLLE
Round 2 designs		
MB13	1ggz	HMSGGSGGLTELLVTIFKEAFSKFDKDGDGAITTRELGTVMR SLGKDPTEAELRDKMSEIDRDGNGAVDFPEFLAMKARAAEKF AKDWAIAADFLAFDKDGNGRVSAAELRHVMTELGEKLSDEEV DEMIRAADTDGDGQVNFQEFVRVLVSGGSGGSLE
MB14	1k6k	HMSGGSGGMLNQELELSLNMAFARAREHRHEFMTVEHLLLAL LSNPSAREALEAASVDLVALRQELEAFIEQTTPVLPASEEER DTQPTLSFQRVLQAAVEFVQSKGKNEVTGANVLMAIFSEKES QAAYLLQKHKVSMFAVIWFILAGTGGSGGSLE
MB15	1mij	HMSGGSGGSSTLTPMHLRKAKLMFFWVRYPSSKTLQKYFPDI KFNKQNTAQLRQWFANFADFYYIQMEKYARQAVTEGIKTPDD LLIAGDSELYRVLNLHYNRNNHIEVPQNFRFVVESTLREFFR AIQGGNDTKEAWKFAIQWAIAAMDDPVPEYFKSPGGSGGSLE
MB16	1s69	HMSGGSGGSTLYENLGGTTAVDLAVDKFYEAVLQDDRIKHFF ADVDMAKQRAHQKAFLTYAFGGTDKYDGRYMREAHKELVENH GLNGKHFVAVMANMVGTLLEMGVPFALVWAVMAVMGEPAHKD DVLNQGGSGGSLE
MB17	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFAIY WAAAAAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE

MB18	1tu9	SGGSGGDAAERVMQSYGRAAASTGFFDDFYRHFLASSPQIRA KFATTDMTAQKHLLRAGIMNLVMYARGMSDSKLRALGASHSR AALDIQPKLYALWLAALLGAVAEHDRDADFATLWAWMAVMGL GIEVIKSYYGSGGSGGSLE
MB19	1v2z	HMSGGSGGSTAFFFRRMSPANKKKLALELMDIYRTIVLEYFN TDAKVNERIDEFVSKAFFADISRFAVYWIAMALWYLFAVELK LEGKSKAILRDYQLTAEDVAAHLAEMYRRSGGSGGSLE
MB20	1y6i	HMSGGSGGMSDQLTELSQQLHDASEKKQLTAIAALAEMGKGG QGILLDYLAKNVPLEKPVLAVGNVYQTLLNLAQFAIVWALMA DFPTGIFPLQSAQGIDYRPLQYALGSGDFEKADEITRDKLAE LAGPGASQRQWLYFTEVEKFPALDLHTINALWWLHSNGNFGF SVQRRLWLASGKEFTKLWPKIGWKSGNVWTRKPKGFTWDLSA PQGHLPLLNQLRGVRVAESLYRHPVWSQYGWGGSGGSLE
MB21	1z6g	HMSGGSGGNIYPLVICGPSGVGKGTLIKKLLNEFPNYFYFSV SCTTRKKREKEKEGVDYYFIDKTIFEDKLKNEDFLEYDNYAN NFYGTLKSEYDKAKEQNKICLFEMNINGVKQLKKSTHIKNAL YIFIKPPSTSVLVKRLLERSTENFAQLWARMAQLLLELQEAN NLNFNLSIINDDLTLTYQQLKNYLLNSYIHLGGSGGSLE
MB22	2hje	HMSGGSGGSKQQTSALIHNIFDKHFEAIAFLHAWMSASQVIR DFYTDRDTNVLNEFFLAWDQGQPSHTPEFRFLTDHKGIIWDD GNAHFYGVNDLILDSLANRVSFSDRWYYINVMTSIGSRHMLV IRVPIVDPSTGKVLGFSFNAVVLDNNFALMEQLKSNSNVDNV VLVANSVPLANSLIGDEPYNVADVLQRKSSDKRLDKLLVIET PIVVNAVTTELALLTVQDGGSGGSLE
MB23	2nr7	HMSGGSGGNAMANVKLLLPYILKWEGGFVHDPADAGGATNKG VTIATWKRVGYDKDGDGDIDVEDLKLLTDDDVLNRVLKPFYW DRWKADLIESQKVANILVDWVWGSGKYGIVIPQQILGVQADG IVGQKTLQAVNSADPDELFESIFRARSNFLFAITWASAADYA NDIGRSATQREAERHTNAAFAAGWAKRLQDIRNLGGSGGSLE
MB24	2oc5	HMSGGSGGQALPDFTSDRYKDAYSRINAIVIEGEQEAHDNYI AIGTLLPDHVEELKRLAKMEMRHKKGFTAAGKNLGVEADMDF AREFFAPLRDNFQTALGQGKTPTALLIQALLIEAFAISAYLT YIPVSDPFARKITEGVVKDEYTHLNYGEAWLKANLESAREEL LEANRENLPLIARMLEQVAGDAAVLRMYAFLLIWHFAAAYSA SLAEIGFSEREIARMLAAAGGGSGGSLE
MB25	3bci	HMSGGSGGGKPLVVVYGDYKAPYAKELDEKVMPKLRQNYIDN HKVEYQFVNLAFLGKDSIVGSRASHAVLMYAPKSFLDFQEAL FASQQDENKQWLTAFLLAWHIAALNLNAKTAQKIAADYETKD SKSWKAAEKDKKIAKDNHIKTTPTAFINGEKVEDPYDYESYE KLLKDGGSGGSLE
MB26	3bw6	HMSGGSGGMAMRIYYIGVFRSGGVMANNLANVSDLSQFGRRN SFAVLWFMLAFAMVVASDTGAGQRQSIEDGNYIGHVYARSEG IAGVLITDKEYPVRPAYTLLNKILDEYLVAHPKEQWADVDSS NEALRMNQLDTYISKYQDPGGSGGSLE
MB19CFX	1v2z	HMSGGSGGSTAFFFRRMSPANKKKLALELMDIYRTIVLEYFN TDAKVNERIDEFVSKAFFADASRFMAYWIAMALWYLFAVELK LEGKSKAILRDYQLTAEDVAAHLAEMYRRSGGSGGSLE

MB25CFX	3bci	HMSGGSGGGKPLVVVYGDYKAPYAKELDEKVMPKLRQNYIDN HKVEYQFVNLAFLGKDSIVGSRASHAVLMYAPKSFLDFQEAL FASQQDENKQSLTAFLLAWHVAALNLNAKTAQKIAADYETKD SKSWKAAEKDKKIAKDNHIKTTPTAFINGEKVEDPYDYESYE KLLKDGGSGGSLE
MB26CFX	3bw6	HMSGGSGGMAMRIYYIGVFRSGGVMANNLANVSDTSQFGRRN SFLVLWFLAAFAAVVASDTGAGQRQSIEDGNYIGHVYARSEG IAGVLITDKEYPVRPAYTLLNKILDEYLVAHPKEQWADVDSS NEALRMNQLDTYISKYQDPGGSGGSLE
MB17 specificity single-point variants		
MB17_A119D	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFDIY WAAAAAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_A119Q	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFQIY WAAAAAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_A119T	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFTIY WAAAAAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_N108D	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIREDDIRDTKKLAFAIY WAAAAAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_N108E	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIREEDIRDTKKLAFAIY WAAAAAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_I161K	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFAIY WAAAAAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVKRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_I161R	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFAIY WAAAAAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVRRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_L165R	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFAIY WAAAAAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIRI LRARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_A128P	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFAIY WAAAAAPPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_Y131G	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFAIY WAAAAAAPHGYADIFSVMASEGIDLEFEDSGAMPDTKDVIRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_Y131S	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFAIY WAAAAAAPHSYADIFSVMASEGIDLEFEDSGAMPDTKDVIRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE

MB17 specificity combinatorial library variants		
MB17_clone_4.2	1s7z	HMSGGSGGMTYNNVFDDAYEELKQGIRENDIRDTKKLAFTIY WAAAAAAPHDYADIFSVMASEGIDLEFEDSGAMPDTKDVKRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 4.4	1s7z	HMSGGSGGMTYNNVFDDAYEELKQEIRENDIRDTKKLEFAIY WAAAAAAPHDPEDIFSVMASEGIDLEFEDSGAMPDTKDVKRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 4.5	1s7z	HMSGGSGGMTYNNVFDDAYEEVKQWIRENDIRDTKKLDFAIY WAAAAAVPHYSADIFSVMASEGIDLEFEDSGAMPDTKDVKRI LLARIVEQLKIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 4.6	1s7z	HMSGGSGGMTYNNVFDDAYEELKQEIRENDIRDTKKLEFPIY WAAAAAAPHDYEDIFSVMASEGIDLEFEDSGAMPDTKDVKRI LLARIVEQLKIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 4.19	1s7z	HMSGGSGGMTYNNVFDDAYEELKQEIRENDIRDTKKLEFAIY WAAAAAAPHDHADIFSVMASEGIDLEFEDSGAMPDTKDVKRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 2.27	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAAGAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIEI LLDRIVEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 2.30	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAAWAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIEI KLARIEEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 2.36	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAAGAAEHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIEI LLDRIEEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 2.33	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAGWAAAHYYEDIFSVMASEGIDLEFEDSGAMPDTKDVIDI IVDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 2.39	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAASGAAAHYYEDIFSVMASEGIDLEFEDSGAMPDTKDVIDI LVDRIVEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 5.48	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFAIY WAAWVAAPHYYQDIFSVMASEGIDLEFEDSGAMPDTKDVIEI QVDRIQEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 5.49	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAWGAAEHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIDI LQDRIGEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 5.53	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAWAAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIDI QHDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 5.55	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAWGAAPHYYQDIFSVMASEGIDLEFEDSGAMPDTKDVIGI QHDRIGEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE

MB17_clone_ 5.60	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAWVAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIDI QEDRIQEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 5.61	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIRENDIRDTKKLAFAIY WAAASAAAHYYPDIFSVMASEGIDLEFEDSGAMPDTKDVIEI KHDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_5.62	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAWVAAAHYYEDIFSVMASEGIDLEFEDSGAMPDTKDVIQI QLDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 5.63	1s7z	HMSGGSGGMTYNNVFDDAYEALKRAIRENDIRDTKKLAFAIY WAAWSAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIEI QDARIQEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 5.64	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAWVAAEHYYPDIFSVMASEGIDLEFEDSGAMPDTKDVIEI KVDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 5.67	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAWSAAAHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIDI KDDRIEEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 5.68	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAWAAAPHYYEDIFSVMASEGIDLEFEDSGAMPDTKDVIQI QHDRIGEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.50	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAAVAAPHYYQDIFSVMASEGIDLEFEDSGAMPDTKDVIRI KHDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.53	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAWAAAPHYYEDIFSVMASEGIDLEFEDSGAMPDTKDVIEI KLDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.55	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAWVAAEHYYPDIFSVMASEGIDLEFEDSGAMPDTKDVIDI QHDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.56	1s7z	HMSGGSGGMTYNNVFDDAYEALKRAIRENDIRDTKKLAFAIY WAAALAAEHYYPDIFSVMASEGIDLEFEDSGAMPDTKDVIDI QLDRILEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.58	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAWVAAAHYYPDIFSVMASEGIDLEFEDSGAMPDTKDVIDI QLDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.59	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAWAAAQHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIEI QQARIGEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.60	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAWAAAPHYYEDIFSVMASEGIDLEFEDSGAMPDTKDVIEI QDDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.61	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAWVAAQHYYPDIFSVMASEGIDLEFEDSGAMPDTKDVIHI QEDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE

MB17_clone_ 6.62	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAWVAAPHYYEDIFSVMASEGIDLEFEDSGAMPDTKDVIQI IDDRIQEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.63	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAWVAAPHYYEDIFSVMASEGIDLEFEDSGAMPDTKDVIDI QHDRIQEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.64	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAWVAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIEI QQDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.65	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAWAAAAHYYEDIFSVMASEGIDLEFEDSGAMPDTKDVIQI QEDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.66	1s7z	HMSGGSGGMTYNNVFDDAYEALKRAIRENDIRDTKKLAFAIY WAAWAAAPHYYPDIFSVMASEGIDLEFEDSGAMPDTKDVIEI QHARIGEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17_clone_ 6.69	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIRENDIRDTKKLAFAIY WAAWVAAPHYYEDIFSVMASEGIDLEFEDSGAMPDTKDVIGI IDARIVEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17 top specificity combinatorial library variants		
MB17.21 (MB17_clone_2.26)	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAAGAAEHYYQDIFSVMASEGIDLEFEDSGAMPDTKDVIDI QLDRIVEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17.31 (MB17_clone_6.49)	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAWVAAPHYYQDIFSVMASEGIDLEFEDSGAMPDTKDVIGI IDARIQEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17.32 (MB17_clone_6.67)	1s7z	HMSGGSGGMTYNNVFDDAYEALKGAIRENDIRDTKKLAFAIY WAAWVAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIGI QLDRIQEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17.33 (MB17_clone_5.54)	1s7z	HMSGGSGGMTYNNVFNDAYEALKEAIRENDIRDTKKLAFAIY WAAWVAAPHYYEDIFSVMASEGIDLEFEDSGAMPDTKDVIEI QLARIEEQLEIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17.41 (MB17_clone_4.1)	1s7z	HMSGGSGGMTYNNVFDDAYEELKQSIRENDIRDTKKLEFQIY WAAAAAAPHDHEDIFSVMASEGIDLEFEDSGAMPDTKDVRRI LLARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17 affinity combinatorial library variants		
MB17.22 (MB17.FMW)	1s7z	HMSGGSGGMTYNNVFDDAYEALKQAIREGDIRDTKKLAFAIY WAAFSAAPHYYADIFSVMASEGIDLEFEDSGMMPDTKDVIDI WEDRIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17.MAL	1s7z	SGGSGGMTYNNVFDDAYEALKEAIREGDIRDTKKLAFAIYWA AMSAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIDILD ARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE

MB17.AML	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIREGDIRDTKKLAFAIY WAAASAAPHYYADIFSVMASEGIDLEFEDSGMMPDTKDVIDI LDARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17.VAW	1s7z	HMSGGSGGMTYNNVFDDAYEALKEAIREGDIRDTKKLAFAIY WAAVSAAPHYYADIFSVMASEGIDLEFEDSGAMPDTKDVIDI WDARIVEQLAIDLWKDAEKLLNEYLEEVEEGGSGGSLE
MB17.42 (MB17.TAR)	1s7z	HMSGGSGGMTYNNVFDDAYEDLKQTIRENDIRDTKKLAFPIY WAAAAAAPWYYADIFSVMASEGIDLEFEDSGAMPDTKDVRRI LLARIVEQLAIDLWKDAEKLVNEHLEEVEEGGSGGSLE
MB17.TVR	1s7z	HMSGGSGGMTYNNVFDDAYEDLKQTIRENDIRDTKKLAFPIY WAAVAAAPWYYADIFSVMASEGIDLEFEDSGAMPDTKDVRRI LLARIVEQLAIDLWKDAEKLVNEHLEEVEEGGSGGSLE
MB17.TAK	1s7z	HMSGGSGGMTYNNVFDDAYEDLKQTIRENDIRDTKKLAFPIY WAAAAAAPWYYADIFSVMASEGIDLEFEDSGAMPDTKDVKRI LLARIVEQLAIDLWKDAEKLVNEHLEEVEEGGSGGSLE
MB17.TVK	1s7z	HMSGGSGGMTYNNVFDDAYEDLKQTIRENDIRDTKKLAFPIY WAAVAAAPWYYADIFSVMASEGIDLEFEDSGAMPDTKDVKRI LLARIVEQLAIDLWKDAEKLVNEHLEEVEEGGSGGSLE
MB17.VVR	1s7z	HMSGGSGGMTYNNVFDDAYEDLKQVIRENDIRDTKKLAFPIY WAAVAAAPWYYADIFSVMASEGIDLEFEDSGAMPDTKDVRRI LLARIVEQLAIDLWKDAEKLVNEHLEEVEEGGSGGSLE
MB17 charge-balance redesigns		
MB17.23	1s7z	HMSGGSHMTYNNVFKDAYEALKGAIRANDIRDTKKLAFAIYW AAAGAAPHNPKDIFSVMASDGIDLEFKDSGAMPDTKDVIDIL LNRIVEQLAIDLAKDAEKLLNEYLKEVKKLE
MB17.24	1s7z	HMSGGSHMTYNNVFKDAYEALKQAIRAGDIRDTKKLAFAIYW AAFSAAPHNPKDIFSVMASDGIDLEFKDSGMMPDTKDVIDIW ENRIVEQLAIDLAKDAEKLLNEYLKEVKKLE
MB17.25	1s7z	HMSGGSHMTYNNVFKDAYEALKGAIRANDIRDTKKLAFAIYW AAAGAAPHNAKDIFSVMASDGIDLEFKDSGAMPDTKDVIDIL LNRIVEQLAIDLAKDAEKLLNEYLKEVKKLE
MB17.26	1s7z	HMSGGSHMTYNNVFKDAYEALKGAIRANDIRDTKKLAFAIYW AAAGAARHNPKDIFSVMASDGIDLEFKDSGAMPDTKDVIDIQ LNRIVEQLAIDLAKDAEKLLNEYLKEVKKLE
MB17.27	1s7z	HMSGGSHMTYNNVFKDAYEALKGAIRANDIRDTKKLAFAIYW AAAGAARHNAKDIFSVMASDGIDLEFKDSGAMPDTKDVIDIQ LNRIVEQLAIDLAKDAEKLLNEYLKEVKKLE
MB17.28	1s7z	HMSGGSHMTYNNVFKDAYEALKQAIRAGDIRDTKKLAFAIYW AAFSAAPHNAKDIFSVMASDGIDLEFKDSGMMPDTKDVIDIW ENRIVEQLAIDLAKDAEKLLNEYLKEVKKLE
MB17.43	1s7z	HMSGGSHMTYNNVFKAAYEELKQNIRENDIRDTKKLEFAIYW AAAAAAPHNPKDIFSVMASDGIDLKFEDSKAMPDTKDVRRIL QARIVEQLAIDLRKDAEKLLNEYLKEVKKLE
MB17.44	1s7z	HMSGGSHMTYNNVFKAAYEDLKQTIRENDIRDTKKLAFPIYW AAAAAAPWNPKDIFSVMASDGIDLKFEDSKAMPDTKDVRRIL LARIVEQLAIDLKKDAEKLVNEHLKEVKKLE

MB17.45	1s7z	HMSGGSHMTYNNVFKAAYEELKQNIRENDIRDTKKLEFAIYW AAAAAAPHNAKDIFSVMASDGIDLKFEDSKAMPDTKDVRRIL QARIVEQLAIDLRKDAEKLLNEYLKEVKKLE
MB17.46	1s7z	HMSGGSHMTYNNVFKAAYEELKQSIRENDIRDTKKLEFQIYW AAAAAAPHNPKDIFSVMASDGIDLKFEDSKAMPDTKDVRRIL LARIVEQLAIDLRKDAEKLLNEYLKEVKKLE
MB17.47	1s7z	HMSGGSHMTYNNVFKAAYEELKQSIRENDIRDTKKLEFQIYW AAAAAAPHNAKDIFSVMASDGIDLKFEDSKAMPDTKDVRRIL LARIVEQLAIDLRKDAEKLLNEYLKEVKKLE
MB17.48	1s7z	HMSGGSHMTYNNVFKAAYEDLKQTIRENDIRDTKKLAFPIYW AAAAAAPWNAKDIFSVMASDGIDLKFEDSKAMPDTKDVRRIL LARIVEQLAIDLKKDAEKLVNEHLKEVKKLE

3.4 (Appendix) MB17 combinatorial library positions and mutations

Table 3.4: MB17 combinatorial library positions, degenerate codons used, and mutations made. Aminao acid mutations are denoted with 1-letter codes. Undesired mutations that contributed to the top clones are bolded and colored orange.

Residue position	Wild type residue	Degenerate codon used	Desired mutations	Undesired mutations	Mutations seen in top clone
MB17 Affinity for Mdmx					
100	A	GMW	ADE	none	D
104	A	VYT	IL	APTV	**T**
119	A	SCT	AP	none	P
125	A	KYT	AFV	S	A
126	A	GMA	AE	none	A
130	H	HDK	FHMWY	CIKLNQRSstop	W
161	I	ADA	IKR	none	R
184	L	SWG	ELQ	V	**V**
187	Y	HAK	HNQY	Kstop	H
MB17 Affinity for Mdm2					
103	Q	SRA	GQ	ER	Q
108	N	RRT	GN	DS	G
125	A	DYK	AFLM	IPSTV	F
126	A	RST	AGS	T	S
129	P	CST	PR	none	P
153	A	RYG	AM	TV	M
162	R	SRT	DR	GH	D
164	L	TKG	LW	none	W

165	L	SDK	DEGL	HQRV	E
166	A	Spontaneous mutation	A	D	**D**
MB17 Specificity for Mdmx					
100	A	GMG	AE	none	A
101	L	SYG	AL	PV	L
104	A	KVG	AEW	GSstop	**S**
117	A	GMS	ADE	none	E
119	A	VMK	AHQT	DEKNP	Q
128	A	SYG	APV	L	A
131	Y	KRT	GY	CD	**D**
132	Y	YMC	PY	SH	**H**
133	A	RMA	AK	ET	**E**
161	I	ADA	IKR	none	R
173	A	RMA	AK	ET	A
MB17 Specificity for Mdm2					
103	Q	SRG	GQ	ER	G
125	A	KSG	AGW	S	A
126	A	KBG	AVW	GLS	**G**
129	P	SMG	EP	AQ	E
133	A	SMG	AQ	EP	Q
162	R	SRW	DER	GHQ	D
164	L	MWA	KL	IQ	**Q**
165	L	SWS	DEL	HQV	L
166	A	GMC	AD	none	D
169	V	SDG	GQV	ELR	V
173	A	GMG	AE	none	E

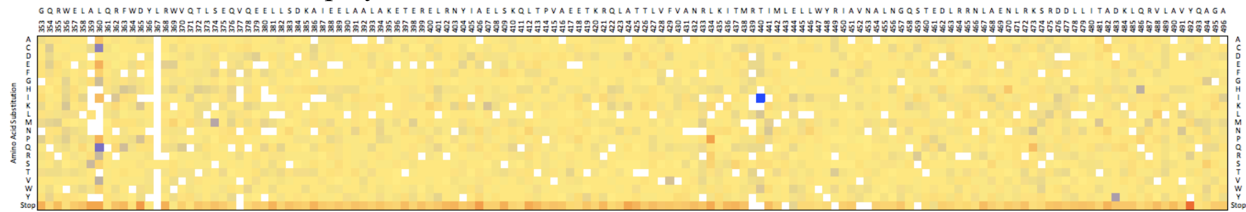
Section 4 (Appendix)

Supplemental data for the design of proteins to inhibit the interaction of EED-Ezh2

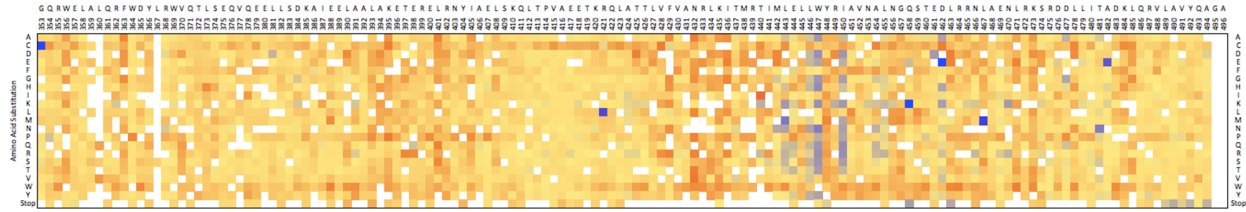
4.1 EB15 and EB22 single site mutation enrichment heatmaps

Figure 4.1: Heatmaps showing the Log2 enrichment ratio for each unique sequence from each round of selection are shown below. Orange color denotes sequences that were depleted relative to the parent pool. Blue color denotes sequences that were enriched relative to the parent pool. Yellow color denotes sequences that were neither depleted nor enriched relative to the parent pool. Enrichment ratios for display-only selections were calculated relative to the unselected pool, while enrichment ratios for binding selections were calculated relative to the display-only selected pool.

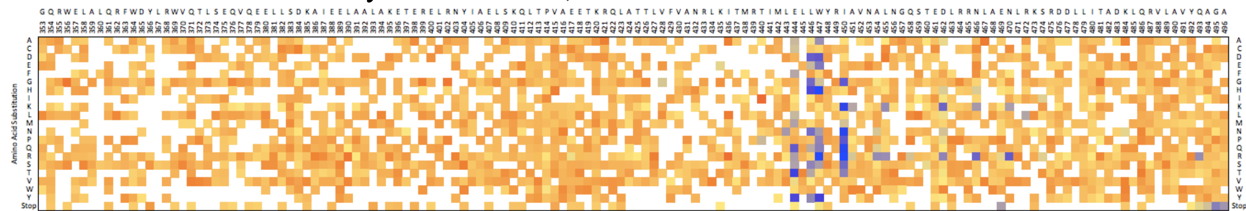
EB15 SSM sorted for display after induction at 30°C in SGCAA media:



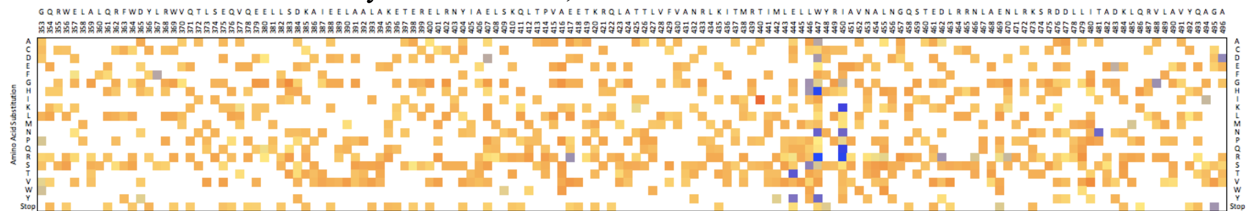
EB15 SSM sorted for affinity toward EED3, 1st round:



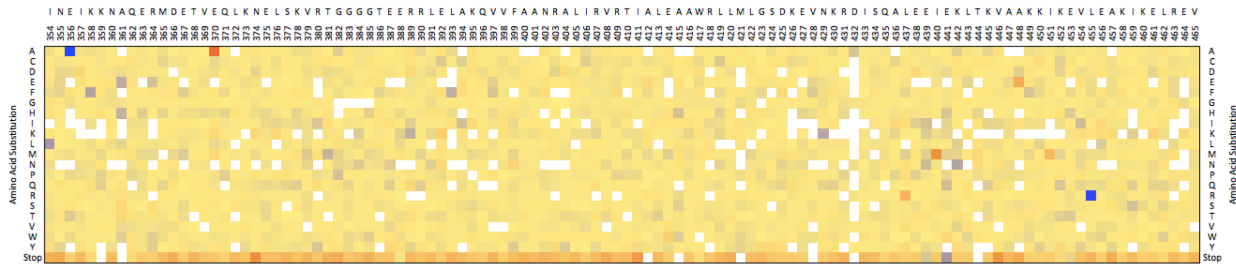
EB15 SSM sorted for affinity toward EED3, 2nd round:



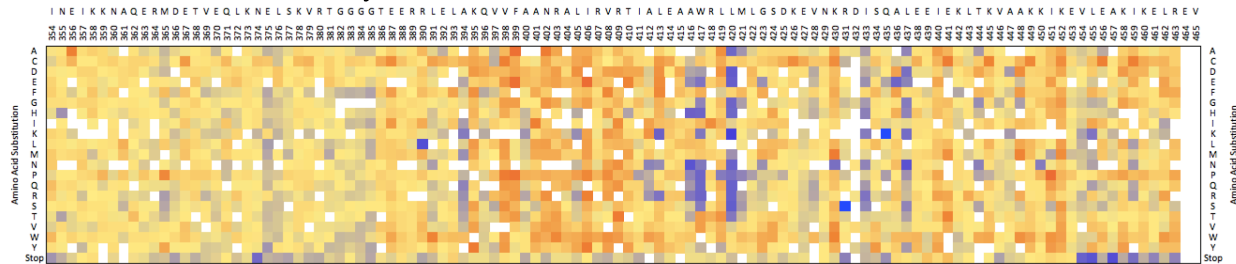
EB15 SSM sorted for affinity toward EED3, 3rd round:



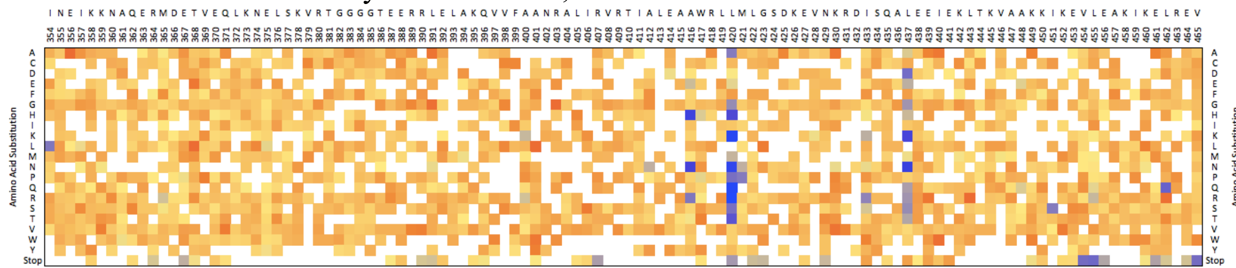
EB22 SSM sorted for display after induction at 30°C in SGCAA media:



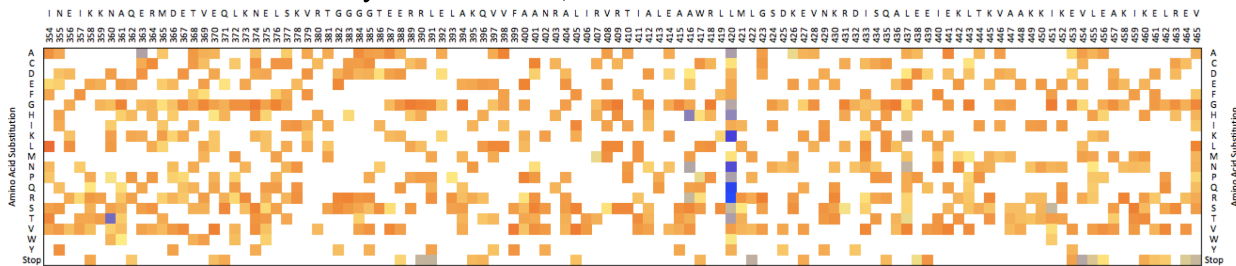
EB22 SSM sorted for affinity toward EED3, 1st round:



EB22 SSM sorted for affinity toward EED3, 2nd round:



EB22 SSM sorted for affinity toward EED3, 3rd round:



4.2 (Appendix) Yeast surface display and ITC affinities for EED binders and variants

Table 4.2: YST and BLI affinities of designed binders and evolved variants are given in the table below. The degree of display of each protein on the surface of yeast is denoted with "+++" for 30-fold or greater display signal above that of the non-displaying population, "++" for 7 to 30-fold, "+" for 1 to 7-fold, or "-" for no display signal above that of the non-displaying population. Only non-avid YST Kd values are reported. Bolded YST and BLI values are of higher confidence than non-bolded values. Lower confidence YST values were due to insufficient labeling time to reach equilibrium at low EED3 concentrations, YST fitted Kds above 2 uM, disagreement between

replicates, lack of replicates, or overly long incubation with the secondary label. Lower confidence BLI values are rough estimated based on visual inspection of the binding curves.

Name	**Degree of Display**	**YST EED3 Kds (nM)**	**BLI EED3 Kds (nM)**	**Fold signal loss at 10 nM EED3 if pre-incubated with Ezh2**
EED Controls				
Ezh2 (peptide)	++	**100, 44**		Not determined
1le2	+++	No binding		Not determined
3lf9	+++	No binding		Not determined
Round 1 designs				
EB15	+++	5.8, 5.3, **0.59, 0.52**		33
EB16	++	2.5, 1.7		5.5
EB17	++	14, 10		7.6
EB18	+++	15, 11		19
EB19	++	310		Not determined
EB20	+	1.8		2.1
EB21	++	16		Not determined
EB22	+++	8.67, 9.92, **1.8, 1.9**		27
EB15 top single-point variants				
EB15.2	+++	**0.14, 0.12**		Not determined
EB15.3	+++	**0.19, 0.19**		Not determined
EB15 top combinatorial library variants				
EB15.A		≤ 0.067		Not determined
EB15.B		≤ 0.067		Not determined
EB15.C		≤ 0.067		Not determined
EB15.D		≤ 0.067		Not determined
EB15.E		≤ 0.067		Not determined
EB15.F		≤ 0.067		Not determined
EB22 top single-point variants				
EB22.2	+++	**0.30, 0.22**		Not determined
EB22.3	+++	**0.77, 0.72**		Not determined
EB22 loop graft redesign				
EB22.4		Not determined		Not determined

EB22 loop graft redesign L423R				
EB22.5		Not determined		Not determined
EB22 top combinatorial library variants				
EB22.G		≤ 0.071	≤ 1	Not determined
EB22.H		≤ 0.071		Not determined
EB22.I		≤ 0.071		Not determined
EB22.J		≤ 0.071		Not determined
EB22.K		≤ 0.071		Not determined
EB22.L		≤ 0.071		Not determined

4.3 (Appendix) Scaffolds and amino acid sequences of EED binders and variants

Table 4.3: Amino Acid sequences of designed binders and their evolved variants are given in the table below. The NdeI site on the N-terminus and the XhoI site on the C-terminus are given for reference, but other residues pertaining to the yeast display or bacterial expression vectors are omitted.

Name	**Scaffold PDB ID**	**Amino Acid Sequence**
Controls		
Ezh2 (peptide)	2qxv	KTMFSSNRQKILERTETLNQEWKQRRIQPV
1le2_scaffold	1le2	MGQRWELALGRFWDYLRWVQTLSEQVQEELLSSQVTQELRALMD ETMKELKAYKSELEEQLTPVAEETRARLSKELQAAQARLGADME DVCGRLVQYRGEVQAMLGQSTEELRVRLASHLRKLRKRLLRDAD DLQKCLAVYQAGA
3lf9_scaffold	3lf9	MINEIKKDAQERMDKSVEALKNNLSKVRTGGGGTEERRKDLVKI VRGEAEGGRVAVRNIARDAANDLAALGKDKEVNWFDISQALWEI QKLTDVAVKKIDEVLAAKEKELMEV
Round 1 designs		
EB15	1le2	HMGQRWELALQRFWDYLRWVQTLSEQVQEELLSDKAIEELAALA KETERELRNYIAELSKQLTPVAEETKRQLATTLVFVANRLKITM RTIMLELLWYRIAVNALNGQSTEDLRRNLAENLRKSRDDLLITA DKLQRVLAVYQAGALE
EB16	1le2	HMGQRWELALQRFWDYLRWVQTLSEQVQEELLTKQVTRELSELR SNTLRELAAYKSELEEQLTPVAEETRARLSKELATTAKALLFVM NRILIALRTYILAVLWMDGTSTEKLRVQLASDLRQLRDKLLRAA DELQKVLAVYQAGALE
EB17	1wpa	HMGGWRREYPPITSDQQRQEYKRNFDTGLREAARLVFILNRIRI QLRTLILELIWADEESRRYKQAADEYNRLKQVKGSADYKSKRDI VLELAKKLEHIAKMVKDYDRQKTLE

EB18	1wqg	HMIREALKDAQEKMKKAVQVAEDDLSTIRTGGGGTQERRKELVD QAIHKGKEAEQSVKKIMEEAQKELRRIRKEGEAGEDEVGKASAM LTFITNRYKITIRTLVLEKMWRLLAVLE
EB19	1xaw	HMGGWRREYPPITSDQQRQRYVEDSKRGAFIYNRLRIVLRTIEL ELIWLDIILRSLREESEDYMRAAERYNRLKQVKGSAEYKSAKNH AEQLKKKLDHLHKMVEDYLRQKTLE
EB20	2wy7	HMTSKQRQVFIANRRKISARTAILELMWQDSERNRRLAQREVNK APQESKEKLQKTLDQLVADKDAEKLE
EB21	3dai	HMSMQEEDTFRELRIFLRQVTHRLAIREALRVFTKPVDPDEVPD YVTVIEQPMDLSSVISKIDLHKYLTVKDYLRDIDLIMRNALKYN PRASFKNNRIAIAARTLALEAYWIIEMELDRKFEQLAEEIQKSR LE
EB22	3lf9	HMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAK QVVFAANRALIRVRTIALEAAWRLLMLGSDKEVNKRDISQALEE IEKLTKVAAKKIKEVLEAKIKELREVLE
EB15 top single point variants		
EB15.2	1le2	HMGQRWELALQRFWDYLRWVQTLSEQVQEELLSDKAIEELAALA KETERELRNYIAELSKQLTPVAEETKRQLATTLVFVANRLKITM RTIMLELLRYRIAVNALNGQSTEDLRRNLAENLRKSRDDLLITA DKLQRVLAVYQAGALE
EB15.3	1le2	HMPQRWELALQRFWDYLRWVQTLSEQVQEELLSDKAIEELAALA KETERELRNYIAELSKQLTPVAEETKRQLATTLVFVANRLKITM RTIMLELLNYRIAVNALNGQSTEDLRRNLAENLRKSRDDLLITA DKLQRVLAVYQAGALE
EB15 top combinatorial library variants		
EB15.A	1le2	HMGQRWELALQRFWDYFRWVQTLSEQVQEELLSDKAIEELAALA KETERELRNYIAELSKQLTPVARETKRQLATTLVFVANRLKITM RTIMLNLHYYRNAVNALNGQSTENLRRNLAENLRKSRDDLLITA DKLQRVLAVYQAGALE
EB15.B	1le2	HMGQRWELALQRFWDYFRWVQTLSEQVQEELLSDKAIEELAALA KETERELRNYFAELSKQLTPVAEETKRQLATTLVFVANRLKITM RTIMLNLHYYRRAVNALNGQSTENLRRNLAENLRKSRDDLLINA DKLQRVLAVYQAGALE
EB15.C	1le2	HMGQRWELALQRFWDYLRWVQTLSEQVQEELLSDKAIEELAALA KETERELRNYIAELSKQLTPVAEETKRQLATTLVFVANRLKITM RTIMLHLYRYRLAVNALNGQSTEGLRRNLAENLRKSRDDLLITA DKLQRVLAVYQAGALE
EB15.D	1le2	HMGQRWELALQRFWDYLRWVQTLSEQVQEELLSDKAIEELAALA KETERELRNYIAELSKQLTPVAEETKRQLATTLVFVANRLKITM RTIMLTLYRYRNAVNALNGQSTERLRRNLAENLRKSRDDLLINA DKLQRVLAVYQAGALE

EB15.E	1le2	HMGQRWELALQRFWDYLRWVQTLSEQVQEELLSDKAIEELAALA KETERELRNYIAELSKQLTPVAKETKRQLATTLVFVANRLKITM RTIMLTLYRYRNAVNALNGQSTEELRRNLAENLRKSRDDLLITA DKLQRVLAVYQAGALE
EB15.F	1le2	HMGQRWELALQRFWDYLRWVQTLSEQVQEELLSDKAIEELAALA KETERELRNYIAELSKQLTPVARETKRQLATTLVFVANRLKITM RTIMLTLHRYRIAVNALNGQSTERLRRNLAENLRKSRDDLLITA DKLQRVLAVYQAGALE
EB22 top single point variants		
EB22.2	3lf9	HMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAK QVVFAANRALIRVRTIALEAAWRLRMLGSDKEVNKRDISQALEE IEKLTKVAAKKIKEVLEAKIKELREVLE
EB22.3	3lf9	HMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAK QVVFAANRALIRVRTIALEAAWRLLMLGSDKEVNKRDISQALEE IEKLTKVAAKKIKEVREAKIKELREVLE
EB22 loop graft redesign		
EB22.4	3lf9/3lhp	HMINEIKKNAQERMDETVEQLKNELDKFKAEVRKVPPTEERILE WAKQVVFAANRALIRVRTIALEAAWRLLMLGSDKEVNKRDISQA LEEIEKLTKVAAKKIKEVLEAKIKELREVLE
EB22 loop graft redesign L423R		
EB22.5	3lf9/3lhp	HMINEIKKNAQERMDETVEQLKNELDKFKAEVRKVPPTEERILE WAKQVVFAANRALIRVRTIALEAAWRLRMLGSDKEVNKRDISQA LEEIEKLTKVAAKKIKEVLEAKIKELREVLE
EB22 top combinatorial library variants		
EB22.G	3lf9/3lhp	HMINEIKKNAQERMDETVEQLKNELDKFKAEVRKVPPTEERILE WAKQVVFAANRALIRVRTITLEANWRLRMLGSDKEVNKRDKSQA LEEIEKLTKVAAKKIKEVREAMIKWLREVLE
EB22.H	3lf9/3lhp	HMLNEIKKNAQERMDETVEQLKNELDKFKAEVRKVPPTEERILE WAKQVVFAANRALIRVRTINLEATWRLRMLGSDKEVNKRDKSQA NEEIEKLTKVAAKKIKEVKEAMIKWLREVLE
EB22.I	3lf9/3lhp	HMINEIKKNAQERMDETVEQLKNELDKFKAEVRKVPPTEERILE WAKQVVFAANRALIRVRTISLEANWRLRMLGSDKEVNKRDKSQA NEEIEKLTKVAAKKIKEVREAMIKWLREVLE
EB22.J	3lf9/3lhp	HMINEIKKNAQERMDETVEQLKNELDKFKAEVRKVPPTEERILE WAKQVVFAANRALIRVRTIGLEAAWRLRMLGSDKEVNKRDRSQA NEEIEKLTKVAAKKIKEVREAMIKWLREVLE

EB22.K	3lf9/3lhp	HMINEIKKNAQERMDETVEQLKNELDKFKAEVRKVPPTEERILE WAKQVVFAANRALIRVRTIPLEAGWRLHMLGSDKEVNKRDISQA KEEIEKLTKVAAKKIKEVREAMIKWLREVLE
EB22.L	3lf9/3lhp	HMLNEIKKNAQERMDETVEQLKNELDKFKAEVRKVPPTEERILE WAKQVVFAANRALIRVRTINLEANWRLQMLGSDKEVNKRDKSQA QEEIEKLTKVAAKKIKEVQEAMIKWLREVLE

4.4 (Appendix) EB15 and EB22 combinatorial library positions and mutations

Table 4.4: EB15 and EB22 combinatorial library positions, degenerate codons used, and mutations made. Undesired mutations that contributed to the top clones are bolded and colored orange. Mutations made to EB22 as part of the 3lhp grafting procedure are colored blue.

Residue position	**Wild type residue**	**Degenerate codon used**	**Desired mutations**	**Undesired mutations**	**Mutations seen in top clones**
EB15 Affinity for EED3					
367	L	TTK	FL	none	FL
417	E	RRA	ER	GK	E**K**R
444	E	NMK	EQTY	ADHKNPS	**HN**T
446	L	YHT	LHS	FPY	H**Y**
447	W	HRK	HNRYW	CKQS	R**Y**
450	I	MDM	IKNQR	HLS	I**L**NR
462	D	RRK	DKR	EGNS	**EGN**R
481	T	AMT	NT	none	NT
EB22 Affinity for EED3					
354	I	MTT	IL	none	IL
377	S	GAT	D	none	D
378	K	AAA	K	none	K
379	V	TTT	F	none	F
380	R	AAG	K	none	K
381	T	GCG	A	none	A
382	G	GAA	E	none	E
383	G	GTT	V	none	V
384	G	CGT	R	none	R
385	G	AAA	K	none	K
Insertion		GTG	V	none	V
Insertion		CCG	P	none	P
Insertion		CCG	P	none	P

386 (389 after insertion)	T	ACC	T	none	T
387 (390)	E	GAA	E	none	E
388 (391)	E	GAG	E	none	E
389 (392)	R	CGC	R	none	R
390 (393)	R	ATT	I	none	I
391(394)	L	CTG	L	none	L
392 (395)	E	GAA	E	none	E
393 (396)	L	TGG	W	none	W
412 (415)	A	VVT	ANR	DGHPST	**G**N**PST**
416 (419)	A	VVT	AHNR	DGPST	A**G**N**T**
420 (423)	L	MDM	HKNLQR	IS	HQR
433 (436)	I	ADA	IKR	none	IKR
437 (440)	L	MWM	KNL	HIQ	KLN**Q**
455 (458)	L	MDA	KLQR	I	KQR
458 (461)	K	ATG	M	none	M
461 (464)	E	TTG	W	none	W

Section 5 (Appendix)

General computational methods

5.1 Fastdesign with atompair constraints

All protein structure modeling was carried out in Rosetta (108) or Foldit (109). All protein structure visualization was carried out in PyMOL (Schrödinger LLC, New York City, New York, USA). The input crystal structure or model is subjected to 4 cycles of side-chain conformational sampling followed by minimization of all side-chain and backbone bond dihedral angles and the rigid body transform between each of the protein chains in the model. The atom-to-atom repulsive term in the Talaris2013 score function was reweighted to 8%, 20%, 50%, and 100% of its standard value during cycles 1, 2, 3, and 4, respectively. Applying iterative cycles of minimization of dihedral angles with a reduced weight on the atom-atom repulsion term tends to overly distort the protein backbone and so atom-atom distance constraints were applied to every pair of Cα atoms within each chain that were within 12 Å of each other in the input crystal structure or model and greater than 7 residues apart in primary sequence. No constraints were applied between atoms residing on different chains in the model to allow free sampling of the rigid body transform between each of the protein chains in the model. Residue identities can either be held constant or allowed to vary. This protocol allows the backbone of each partner in a protein-protein complex to move to better accommodate binding to the other partner while not forcing unrealistic binding interfaces to remain in the bound state.

A sample command line is given below:

```
/path_to_rosetta_scripts_binary/rosetta_scripts.default.linuxgccrelease -
database /path_to_rosetta_database/database -ignore_zero_occupancy false -
ignore_unrecognized_res -overwrite -out:file:renumber_pdb false -ex1 -ex2 -
nstruct 1 -s input_structure.pdb -parser:protocol input_protocol.xml -
score:weights talaris2013
```

A sample xml protocol is given below:

```
<ROSETTASCRIPTS>
      <SCOREFXNS>
            <sfxn_hard weights=talaris2013>
                  <Reweight scoretype=atom_pair_constraint weight=1.0/>
            </sfxn_hard>
      </SCOREFXNS>
      <TASKOPERATIONS>
            <OperateOnCertainResidues name=chainA>
                  <PreventRepackingRLT/>
                  <ChainIs chain=A/>
            </OperateOnCertainResidues>
            <OperateOnCertainResidues name=chainB>
                  <PreventRepackingRLT/>
                  <ChainIs chain=B/>
            </OperateOnCertainResidues>
            <LimitAromaChi2 name=arochi2/>
            <IncludeCurrent name=inclcur/>
            <RestrictChainToRepacking name=rctr1 chain=1/>
            <RestrictChainToRepacking name=rctr2 chain=2/>
      </TASKOPERATIONS>
      <MOVERS>
            <AtomTree name=ftree simple_ft=1/>
            <MinMover name=min_bb_sc bb=1 chi=1 jump=1 scorefxn=sfxn_hard>
                  <MoveMap name=minmvrmap>
                        <Chain number=2 chi=1 bb=1/>
                        <Chain number=1 chi=1 bb=1/>
                        <Jump number=1 setting=1/>
                  </MoveMap>
            </MinMover>
            <FastRelax name=fstrlx_no_dsn scorefxn=sfxn_hard repeats=1 
task_operations=arochi2,rctr1,rctr2,inclcur>
                  <MoveMap name=fstrlxmap>
                        <Chain number=2 chi=1 bb=1/>
                        <Chain number=1 chi=1 bb=1/>
                        <Jump number=1 setting=1/>
                  </MoveMap>
            </FastRelax>
            <AddConstraintsToCurrentConformationMover 
name=add_pair_chainA_cst use_distance_cst=1 coord_dev=0.5 bound_width=0.1 
min_seq_sep=8 max_distance=12.0 cst_weight=1.0 task_operations=chainA/>
            <AddConstraintsToCurrentConformationMover 
name=add_pair_chainB_cst use_distance_cst=1 coord_dev=0.5 bound_width=0.1 
min_seq_sep=8 max_distance=12.0 cst_weight=1.0 task_operations=chainB/>
            <ClearConstraintsMover name=clear_cst/>
      </MOVERS>
      <FILTERS>
            <Ddg name=binding_energy_AB threshold=0 scorefxn=sfxn_hard 
confidence=0 jump=1 repack=1 relax_mover=min_bb_sc repeats=3/>
            <Sasa name=dsasa threshold=500 confidence=0/>
            <ShapeComplementarity name=shape_comp_AB jump=1 verbose=0 
min_sc=0.60 confidence=0/>
            <SymUnsatHbonds name=unsat_AB jump=1 cutoff=1000 confidence=0/>
            <ScoreType name=total_score_complex scorefxn=sfxn_hard 
score_type=total_score confidence=0 threshold=0/>
```

```
	</FILTERS>
	<PROTOCOLS>
		<Add mover_name=ftree/>
		<Add mover=add_pair_chainA_cst/>
		<Add mover=add_pair_chainB_cst/>
		<Add mover_name=fstrlx_no_dsn/>
		<Add mover=clear_cst/>
		<Add filter_name=binding_energy_AB/>
		<Add filter_name=dsasa/>
		<Add filter_name=shape_comp_AB/>
		<Add filter_name=unsat_AB/>
		<Add filter_name=total_score_complex/>
	</PROTOCOLS>
</ROSETTASCRIPTS>
```

5.2 Automatically identifying positions of key interacting residues

When optimizing the residue identities of designed binding proteins, its necessary to hold fixed the identities of those residues corresponding to the key binding residues from the native binder, p53 in this case. RosettaScripts includes functions to prohibit sequence design at specified residue positions, but identifying the positions of key binding residues becomes nontrivial when using a wide variety of host proteins where the key residues may occupy different numerical positions from host to host. To solve this problem, a bash script, get_stubs.sh, was written that identifies the residue position of that residue on the design molecule closest to a specified residue on the target molecule. In the case of designs against Mdmx, the script was instructed to look for all Cζ on Phenylalanine residues of the design chain B and measure the distance between each and the Cγ on Isoleucine 39 of the target chain A (Mdmx). The Phenylalanine Cζ closest to the Isoleucine 39 Cγ can only correspond to the key interacting Phenylalanine taken from p53. The residue positions of the key Tyrosine and Tryptophan residues are then found by adding 3 or 4 to the value reported for the key Phenylalanine residue. In the case of designs made against EED, the script was instructed to look for all Cζ on Phenylalanine residues of the design chain B and measure the distance between each and the Cζ

on Phenylalanine 268 of the target chain A (EED). The Phenylalanine Cζ closest to the Phenylalanine 268 Cζ can only correspond to the key interacting Phenylalanine taken from Ezh2. The residue positions of the other key binding residues are then found by adding the appropriate number of residues to the value reported for the key Phenylalanine residue. This script was automatically run immediately before running any Rosetta application that carried out sequence optimization.

The script is reproduced here:

```
#!/bin/bash

### This script takes a PDB file and given a specific residue type, residue
number, atom name, and chain ID, will find the closest instance of another
given residue type, printing out to standard output the found residue's atom
name, residue type, chain ID, residue number, and the distance between the
anchor atom and the stub atom.
### Usage: ./get_chn_B_stubs.sh <pdbfilename> <anchor_restype>
<anchor_resnum> <anchor_atomname> <anchor_chnid> <stub_restype>
<stub_atomname> <stub_chnid>
basename=`echo $1 | sed 's|\.pdb||'`
anchor_restype=$2
anchor_resnum=$3
anchor_atomname=$4
anchor_chnid=$5
stub_restype=$6
stub_atomname=$7
stub_chnid=$8
grep 'ATOM' ${basename}.pdb | grep $stub_restype | grep $stub_atomname | grep
$stub_chnid > ${basename}_stub_list.tmp
lines=`wc ${basename}_stub_list.tmp | awk '{print $1}'`
for i in `seq 1 $lines`; do
      anchor_xcoord=`grep 'ATOM' ${basename}.pdb | grep " $anchor_restype " |
grep " $anchor_resnum " | grep " $anchor_atomname " | grep " $anchor_chnid "
| awk '{print $7}'`
      anchor_ycoord=`grep 'ATOM' ${basename}.pdb | grep " $anchor_restype " |
grep " $anchor_resnum " | grep " $anchor_atomname " | grep " $anchor_chnid "
| awk '{print $8}'`
      anchor_zcoord=`grep 'ATOM' ${basename}.pdb | grep " $anchor_restype " |
grep " $anchor_resnum " | grep " $anchor_atomname " | grep " $anchor_chnid "
| awk '{print $9}'`
      awk 'NR=='$i'{print $0}' ${basename}_stub_list.tmp >
${basename}_stub_list_${i}.tmp
      awk '{stub_xcoord += $7} {stub_ycoord += $8} {stub_zcoord += $9}
{anchor_xcoord += '$anchor_xcoord'} {anchor_ycoord += '$anchor_ycoord'}
{anchor_zcoord += '$anchor_zcoord'} END {print $3,$4,$5,$6,((stub_xcoord-
anchor_xcoord)^2+(stub_ycoord-anchor_ycoord)^2+(stub_zcoord-
anchor_zcoord)^2)^0.5}' ${basename}_stub_list_${i}.tmp  >>
${basename}_stub_list_dist.tmp
done
```

```
sort -nk5 ${basename}_stub_list_dist.tmp | head -n1
rm ${basename}_stub_list*tmp
```

Section 6 (Appendix)

Computational methods specific to the design of proteins to inhibit the interaction of Mdmx-p53

6.1 Side-chain grafting with backbone flexibility

A set of 866 unrelated protein crystal structures from the PDB (hereafter termed scaffolds) underwent low-resolution docking against the crystal structure of Mdmx taken from PDB ID 3jzp using the Patchdock algorithm. We refined the docked orientations of each of the scaffolds using RosettaDock and iteratively superimposed the key Phenylalanine 19 and Tryptophan 23 residues isolated from an evolved p53 peptide, pDI6W, of PDB ID 3jzp, onto the top 500 scoring Patchdock solutions for each of the 866 scaffolds using the PlaceStub protocol in the RosettaScripts application of Rosetta (110, 111). We included the Mdmx domain during the placement process to prune solutions that introduced clashes between the scaffold and Mdmx. We identified scaffolds and docked orientations that allowed the key Phenylalanine 19 and Tryptophan 23 residues to remain close to their original positions and used flexible-backbone protein design to subtly adjust the backbone of the scaffold to optimally host these key residues. We then optimized the identities of scaffold residues near Mdmx to promote maximal binding affinity.

The PlaceStub application requires the following inputs: 1. A PDB file containing both the target protein (Mdmx) and a potential scaffold protein as separate chains, 2. A patchdock file containing a list of 6-dimensional transforms to be randomly applied to the scaffold chain in the PDB file relative to the target chain in the pdb file, 3. A Stub file for each residue to be placed, containing a set of coordinates for each side chain conformation to be tested of the disembodied amino acid, 4. A Sequence Profile file containing a line for every residue in the target-scaffold

complex, and 5. A Constraint file containing a line for every residue in the target-scaffold complex.

An example commandline to run the PlaceStub protocol is given below:

```
/path_to_rosetta_scripts_binary/rosetta_scripts.default.linuxgccrelease -
database /path_to_rosetta_database/database  -write_failures false -
boinc:max_nstruct 200000 -s input_complex_structure.pdb.gz -run:protocol
jd2_scripting -jd2:dd_parser -parser:protocol input_protocol.xml -out:nstruct
10000 -ex1 -ex2aro -extrachi_cutoff 5 -out:file:silent default.out -silent_gz
-docking:no_filters -use_input_sc -parser:patchdock -
parser:patchdock_random_entry 1 500 -no_his_his_pairE -lj_hbond_hdis 1.75 -
lj_hbond_OH_donor_dis 2.6 -dun08 false -correct -linmem_ig 10
```

An example Patchdock file to use in the PlaceStub step is given below: ("~~~" denotes portions omitted for brevity).

```
*********************************************************
Program parameters

~~~

   # | score | pen.  | Area    | as1   | as2   | as12  | ACE     | hydroph |
Energy  |cluster| dist. || Ligand Transformation
   1 | 13604 | -3.41 | 2197.30 |  4099 |     0 |     0 |  305.76 | 1360.63 |
0.00 |     0 | 0.00 || 2.38983 -1.05293 -0.88813 56.55492 13.01158 74.02017
   2 | 12864 | -3.38 | 1884.70 |  4946 |     0 |     0 |  223.94 | 1251.31 |
0.00 |     0 | 0.00 || 0.07525 0.19143 2.98156 67.88561 2.71938 37.81353
   3 | 12512 | -3.20 | 1705.00 |  6061 |     0 |     0 |  -54.08 | 1093.44 |
0.00 |     0 | 0.00 || -0.50266 -0.54353 -1.65889 47.87671 11.68168 50.90615

~~~

2131 |  5442 | -3.33 |  846.00 |   859 |     0 |     0 |  240.55 |  812.89 |
0.00 |     0 | 0.00 || 3.10953 -0.47926 -1.36943 68.00933 27.96250 77.11030
2132 |  5442 | -3.91 |  921.30 |   623 |     0 |     0 |  248.17 |  786.35 |
0.00 |     0 | 0.00 || 2.05089 0.48245 1.64009 37.78255 -21.45059 22.98504
2133 |  5442 | -3.14 |  751.50 |   724 |     0 |     0 |  246.22 |  660.91 |
0.00 |     0 | 0.00 || -0.87704 -0.84688 -0.72093 28.15167 12.35644 85.10198
Best Rmsd Result: 100000000.00000 rank -1
Best Rank Result: 100000000.00000 rank 100000
```

An example Sequence Profile file to use in the PlaceStub step is given below: ("~~~" denotes portions omitted for brevity).

```
made for pdb 3jzp_1s7z.pdb from matrix /path_to_matrix/BLOSUM62 with factor -
1. (note: SequenceProfile::read_from_checkpoint() skips this line)
F 2 2 3 3 -6 3 1 -0 3 -0 -0 3 4 3 3 2 2 1 -1 -3
K 1 3 1 -1 3 2 1 3 -5 2 1 -0 1 -1 -2 -0 1 2 3 2
C -0 -9 3 4 2 3 3 1 3 1 1 3 3 3 3 1 1 1 2 2

~~~
```

```
V -0 1 3 2 1 3 3 -3 2 -1 -1 3 2 2 3 2 -0 -4 3 1
E 1 4 -2 -5 3 2 -0 3 -1 3 2 -0 1 -2 -0 -0 1 2 3 2
E 1 4 -2 -5 3 2 -0 3 -1 3 2 -0 1 -2 -0 -0 1 2 3 2
```

An example Constraint file to use in the PlaceStub step is given below: ("~~~" denotes portions omitted for brevity).

```
SequenceProfile 1 /work/moody/PatchDock/Pairs/3jzp_1s7z.sequence_profile
SequenceProfile 2 /work/moody/PatchDock/Pairs/3jzp_1s7z.sequence_profile
SequenceProfile 3 /work/moody/PatchDock/Pairs/3jzp _1s7z.sequence_profile

~~~

SequenceProfile 191 /work/moody/PatchDock/Pairs/3jzp _1s7z.sequence_profile
SequenceProfile 192 /work/moody/PatchDock/Pairs/3jzp _1s7z.sequence_profile
SequenceProfile 193 /work/moody/PatchDock/Pairs/3jzp _1s7z.sequence_profile
```

6.2 Filtering for clashes with Mdm2 and multi-state design

Models output from the PlaceStub and the subsequent design algorithm were filtered on the basis of shape complementarity, predicted binding energy, and surface area buried in the interface with Mdmx, yielding ~15,000 output models. We filtered these output models for those that placed a beta-carbon within 6.5 Å of the beta-carbons of either residue 96, 98, or 99 or two beta-carbons within 6.5 Å of the beta-carbon of residue 95 of Mdmx and visually inspected passing models for those likely have significant clashes with Mdm2 but not Mdmx, yielding 28 models. We selected 14 candidate designs and reverted unnecessary mutations back to their wild-type identities using the RevertDesignToNative application in Rosetta (108). We used Foldit (109) to carry out manual multi-state design simultaneously against Mdmx (taken from PDB ID 3jzp)(112) and Mdm2 (taken from PDB ID 1ycr)(113) to identify mutations predicted to have a positive or neutral effect on binding to Mdmx but a neutral or negative effect on Mdm2 binding. Specifically, mutations were chosen predicted to introduced charge repulsion and/or steric clashes with Mdm2 but have negligible effect on Mdmx binding. The electrostatics calculator available in PyMOL (Schrödinger LLC, New York City, New York, USA) was used to estimate

the distribution of charge across the surface of Mdmx, Mdm2, and the designed binder during the multistate design process. At this point, N-linked glycosylation sites on the binder were identified by the consensus sequence NXS/T and removed by mutating either the Asparagine or the Serine/Threonine in Foldit. Short Glycine-Serine linkers were added to all binder N and C-termini so as to avoid the yeast-display machinery from interfering with proper binder folding or target binding. All binder Cysteine residues were removed, usually by mutation to Alanine. DnaWorks was used to generate a nucleotide sequence for each designed amino acid sequence (114).

A sample command line to filter the candidate designs is given below:

```
/path_to_rosetta_scripts_binary/rosetta_scripts.default.linuxgccrelease -
database /path_to_rosetta_database/database  -
parser:protocol ./input_protocol.xml -ex1 -ex2 -dun08 –overwrite -nstruct 1 -
correct -s input_structure.pdb
```

A sample xml protocol is given below:

```
<dock_design>
      <FILTERS>
            <ResidueBurial name=burial_73 res_num=73 distance=6.5
neighbors=2/>
            <ResidueBurial name=burial_74 res_num=74 distance=6.5
neighbors=1/>
            <ResidueBurial name=burial_76 res_num=76 distance=6.5
neighbors=1/>
            <ResidueBurial name=burial_77 res_num=77 distance=6.5
neighbors=1/>
            <CompoundStatement name=burial_all>
                  <OR filter_name=burial_73/>
                  <OR filter_name=burial_74/>
                  <OR filter_name=burial_76/>
                  <OR filter_name=burial_77/>
            </CompoundStatement>
      </FILTERS>
      <PROTOCOLS>
            <Add filter_name=burial_all/>
      </PROTOCOLS>
</dock_design>
```

6.3 Homology modeling

The sequences of MB17.4.1, MB17.4.2, MB17.2.1, MB17.2.2, MB17.3.1, and MB17.3.2 were manually threaded onto each molecule of MB17 of the Mdmx-MB17 crystal structure. In the case of MB17.2.1 and MB17.2.2, each molecule of Mdmx was also replaced with a superimposed Mdm2 molecule taken from PDB ID 1ycr. The threaded models were then subjected to the fastdesign with atompair constraints protocol while holding all residue identities fixed. Chains A, B, C, and D corresponded to Mdmx or Mdm2, MB17 variant model, MB17 variant model, and Mdmx or Mdm2, respectively, to allow rigid body sampling and interface quality metrics to be applied to all three protein-protein interfaces in the models.

The relaxed homology models were subjected to fragment-based ab initio structure prediction using only their primary amino acid sequences and the top scoring models were compared to the design models (115). OCR (116), the protein used as the starting point of the MB17 design, was also subjected to ab initio structure prediction as a control. Generally, ab initio structure prediction correctly predicted the orientation of the 4 helices of OCR but could not correctly orient the long loop connecting the 3rd and 4th helices, presumably because this loop participates extensively in the interface between 2 molecules of OCR, is likely only ordered in the dimeric state, and ab initio structure prediction was carried out using only 1 molecule of the OCR dimer. Similar results were obtained for all of the designed proteins based on OCR.

6.4 Automated design of Kunkel oligonucleotides

A simple shell script was used to design NNK Kunkel primers for every amino acid position in the MB17 for the purposes of synthesizing an SSM library. The script takes in a DNA

sequence of the gene and outputs a list of 45-mer antisense oligonucleotides with an MNN codon (NNK in the sense strand) centered in the middle. The script is reproduced below:

```
#!/bin/bash

### This script takes a DNA sequence (in capital letters) and, given the 3rd
base numbers of the first and last residue to be covered by the primers,
generates a list of antisense NNK Kunkel primers.

### Usage: ./dsn_NNK_kunkel_45_mer_primers.sh <DNA sequence filename> <base
number of 3rd base in codon of residue to be covered by first primer,
relative to start of DNA sequence file> <base number of 3rd base in codon of
residue to be covered by last primer, relative to start of DNA sequence file>
<residue number to be covered by first primer relative to another source,
such as a .pdb file>

name=$1 ###<DNA sequence filename>
dna_first=$2 ###<base number of 3rd base in codon of residue to be covered by
first primer, relative to start of DNA sequence file>
dna_last=$3 ###<base number of 3rd base in codon of residue to be covered by
last primer, relative to start of DNA sequence file>
prot_first=`echo "$dna_first" | awk '{print $1/3}'`
prot_last=`echo "$dna_last" | awk '{print $1/3}'`
range=`echo "$prot_first" "$prot_last" | awk '{print $2-$1+1}'`
pdb_first=$4 ###<residue number to be covered by first primer relative to
another source, such as a .pdb file>
### Make a list of the residue positions to be covered by each primer
for i in `seq $prot_first $prot_last`; do echo $i | awk '{print $1*3-23}';
done > ${name}.pos_lst
### From the DNA sequence, extract 45-base-pair blocks of sense-strand
sequence corresponding to each forward primer to be designed, substituting
the middle 3 base pairs with "NNK"
for j in `cat ${name}.pos_lst`; do awk '{print
substr($_,'$j',21)"NNK"substr($_,'$j'+24,21)}' $name; done >
${name}fwd_prim_lst
### Generate a name for each forward primer, corresponding to the residue
number it covers
for k in `seq 1 ${range}`; do echo $k $pdb_first | awk '{print ">pos_"$1+$2-
1"_fwd"}'; awk 'NR=='$k'{print $0}' ${name}.fwd_prim_lst; done >
${name}.fwd_prim_lst_w_label
### From the forward primer list, generate a list of reverse compliment
sequences to serve as reverse primers
for l in `cat "$name".fwd_prim_lst`; do echo "$l" | awk '{print
substr($_,45,1)substr($_,44,1)substr($_,43,1)substr($_,42,1)substr($_,41,1)su
bstr($_,40,1)substr($_,39,1)substr($_,38,1)substr($_,37,1)substr($_,36,1)subs
tr($_,35,1)substr($_,34,1)substr($_,33,1)substr($_,32,1)substr($_,31,1)substr
($_,30,1)substr($_,29,1)substr($_,28,1)substr($_,27,1)substr($_,26,1)substr($
_,25,1)substr($_,24,1)substr($_,23,1)substr($_,22,1)substr($_,21,1)substr($_,
20,1)substr($_,19,1)substr($_,18,1)substr($_,17,1)substr($_,16,1)substr($_,15
,1)substr($_,14,1)substr($_,13,1)substr($_,12,1)substr($_,11,1)substr($_,10,1
)substr($_,9,1)substr($_,8,1)substr($_,7,1)substr($_,6,1)substr($_,5,1)substr
($_,4,1)substr($_,3,1)substr($_,2,1)substr($_,1,1)}' | sed 's|A|t|g' | sed
's|T|a|g' | sed 's|C|g|g' | sed 's|G|c|g' | sed 's|K|m|g' | sed 's|a|A|g' |
sed 's|t|T|g' | sed 's|c|C|g' | sed 's|g|G|g' | sed 's|m|M|g'; done >
${name}.rev_prim_lst
```

```
### Generate a name for each reverse primer, corresponding to the residue
number it covers
for m in `seq 1 ${range}`; do echo $m $pdb_first | awk '{print ">pos_"$1+$2-
1"_rev"}'; awk 'NR=='$m'{print $0}' ${name}.rev_prim_lst; done >
${name}.rev_prim_lst_w_label
### Combine the lists of forward and reverse primers
#cat ${name}.fwd_prim_lst_w_label ${name}.rev_prim_lst_w_label >
${name}_prim_lst.txt
### Delete the intermediate files
rm ${name}.pos_lst
rm ${name}.fwd_prim_lst
rm ${name}.fwd_prim_lst_w_label
rm ${name}.rev_prim_lst
rm ${name}.rev_prim_lst_w_label
```

An example input sequence file is given below:

```
AGTGGTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTCGGCTAGCCATATGTCTGGTGGTAGTGGAGG
TATGACTTACAATAATGTTTTTGATGATGCTTATGAAGCTTTGAAACAAGCTATTAGAGAGAATGATATAAGAGATA
CTAAAAAGTTAGCTTTTGCTATATACTGGGCAGCCGCTGCGGCAGCTCCACATTACTATGCTGATATTTTCTCTGTT
ATGGCTTCTGAAGGTATTGATCTTGAATTTGAAGATTCTGGCGCTATGCCAGATACAAAAGATGTTATTAGAATTTT
GTTGGCTAGAATTGTCGAGCAATTGGCTATTGATTTGTGGAAAGATGCTGAAAAATTGTTGAATGAATATTTGGAAG
AAGTTGAAGAAGGAGGTTCTGGGGGGTCTCTCGAGGGGGGCGGATCCGAACAAAAGCTTATTTCTGAAGAGGACTTG
TAATAG
```

An example output list of Kunkel primers is given below: ("~~~" denotes portions omitted for brevity).

```
MB17_NNK_88_rev,TAAGCATCATCAAAAACATTATTGTAAGTMNNACCTCCACTACCACCAGACATATGGCTA
MB17_NNK_89_rev,TCATAAGCATCATCAAAAACATTATTGTAMNNCATACCTCCACTACCACCAGACATATGG
MB17_NNK_90_rev,GCTTCATAAGCATCATCAAAAACATTATTMNNAGTCATACCTCCACTACCACCAGACATA

~~~

MB17_NNK_187_rev,CCAGAACCTCCTTCTTCAACTTCTTCCAAMNNTTCATTCAACAATTTTTCAGCATCTTTC
MB17_NNK_188_rev,CCCCAGAACCTCCTTCTTCAACTTCTTCMNNATATTCATTCAACAATTTTTCAGCATCTT
```

6.5 Illumina sequence data processing of MB17 SSM pools

High-throughput sequencing was carried out on an Illumina MiSeq (Illumina, San Diego, CA, USA). Each run generated 7-15 million paired-end sequences which were output along with their associated quality scores in compressed .fastq files. The resulting data was analyzed using a modified version of the Enrich package (117). In brief, the reads were first spit up into their respective pools according to their barcodes and then the forward and reverse reads for each individual molecule of DNA were fused together to form a single continuous read. Next the

fused sequences present in each pool were aligned to both the wild-type nucleotide and amino acid sequences and both nucleotide and amino acid mutations relative to wild type were identified. Then the number of instances of each unique amino acid sequence were tallied and their frequencies calculated relative to the number or reads in their pool. Finally, the tallied sequence counts were imported into Microsoft Excel for Mac 2011 (Microsoft Corporation, Redmond, WA, USA), sequences occurring less than 30 times in at least 1 pool were removed, and the enrichment or depletion of each unique sequence relative to its frequency in a baseline pool was calculated. Pools selected for both display and binding were compared to the pool selected only for display, while pools selected only for display were compared to the unselected pool. Commandlines to run each of the data processing steps and example inputs and outputs are given below:

1. Split the reads into separate pools according to their barcodes:

```
python /path_to_scripts/Index_parser_PE.py --inpath ./ --readi <name
of .fastq file containing forward read> --read1
<forward_read_.fastq_filename> --read2 <reverse_read_.fastq_filename> --index
<barcode_sequence> --iname <barcode_name> --outpath <barcode_name> --
avg_quality 10 --mutations 3
```

An example .fastq file is given below: (“~~~” denotes portions omitted for brevity)

```
@M00777:19:000000000-A7L7K:1:1107:15354:4614 1:N:0:1
GGTCGCGGACCAGGGTCGGCTAGCCATATGATCAATGAGATCAAAAAGAATGCGCAGGAGCGTATGGATGAAACCGT
GGAACAACTGAAAAATGAACTGAGCAAAGTTCGTACCGGTGGTGGCGGCACCGAAGAGCGCCGTCTGGAACTGGCGA
AACAAGTGGTTTTTGCGGCGAATCGCGCCTTAATTCGTGTGCGTACCATCGCCCTGGAGGCGGCGTGGCGTTTACTG
ATGCTGGGCAGCGATAAGGAAGTGAATA
+
CCABCCCCB@BBGGGGGGGGGGHHHHHHHHHHHHHHHHHHHHHHHHHHGGHHHGGGGGGGHGGGGHHHHHHHHHHGG
HGGHHHHHHHHHHHHHHHHHGHHHHHHHHHHHHGGHHGGCGGGGGGHGGGGGGGGGGGHHGGGGCGHGHGGGGGGGG
GGGGGGFFGGGGGGFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFDFFFFFFFFFFFFBDFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFFFFF
@M00777:19:000000000-A7L7K:1:1107:24218:4614 1:N:0:1
GGTCGCGGACCAGGGTCGGCTAGCCATATGATCAATGAGATCAAAAAGAATGCGCAGGAGCGTATGGATGAAACCGT
GGAACAACTGAAAAATGAACTGAGCAAAGTTCGTACCGGTGGTGGCGGCACCGAAGAGCGCCGTCTGGAACTGGCGA
AACAAGTGGTTTTTGCGGCGAATCGCGCCTTAATTCGTGTGCGTACCATCGCCCTGGAGGCGGCGTGGCGTTTACTG
ATGCTGGGCAGCGATAAGGAAGTGAATA
+
BBAABBDBCCCBGGGGGGGGGGHHHHHHHHHHHHGHHHHHHHHHHGHHGGHHHHGGGGGGHGGGGHHHHHHHHHHGG
HGEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHGGHHGGGDGGGGGHGGGGGGGGGGGGHGGGGDGHGHGGGGGGGG
```

```
GGGGGGFGGGGGGGFFFFFFFFFFFFFFFFFFFFFFFF.AFFFFFFFFFFFFFFFFFEFDFFFFFCFFFDFFFFFFF
FFFFFFFFFFFFFFFEFEFFBFFFFB/BF
@M00777:19:000000000-A7L7K:1:1107:24501:4615 1:N:0:1

~~~
```

2. Fuse the paired end reads into a single read:

```
python /path_to_scripts/quickunforgivingfuser1.3.py --read1
<forward_read_.fastq_filename> --read2 <reverse_read_.fastq_filename> --
paired_mismatch_threshold 200 --read1_start 1 --read2_start 1 --
length_overlap
<size_of_the_overlap_region_between_the_forward_and_reverse_read> --
include_nonoverlap_region 1 --wtseq <non-mutated_gene_nucleotide_sequence> --
mode B --trim_5p <number_of_unneeded_bases_at_5'_end_of_fused_gene> --trim_3p
<number_of_unneeded_bases_at_3'_end_of_fused_gene>
```

An example fused read file is given below: ("~~~" denotes portions omitted for brevity)

```
read1.ID    fused_sequence    total_bases paired_mismatch_count
      paired_match_count      gap_count   paired_unresolvable_count
      mutations_to_wt   read1_avgquality  read1_chastity    read1_Ncount
      read2_avgquality  read2_chastity    read2_Ncount
@M00777:19:000000000-A7L7K:1:1107:15354:4614 1:N:0:1
      GGGTCGGCTAGCCATATGATCAATGAGATCAAAAAGAATGCGCAGGAGCGTATGGATGAAACCGTGGAACA
ACTGAAAAATGAACTGAGCAAAGTTCGTACCGGTGGTGGCGGCACCGAAGAGCGCCGTCTGGAACTGGCGAAACAAG
TGGTTTTTGCGGCGAATCGCGCCTTAATTCGTGTGCGTACCATCGCCCTGGAGGCGGCGTGGCGTTTACTGATGCTG
GGCAGCGATAAGGAAGTGAATAAACGCGATATTAGCCAGGCGCTGGAGGAGATTGAAAAACTGACCAAGGTTGCCGC
GAAGAAAATCAAAGAGGTGCTGGAAGCCAAGATCAAGGAACTGCGTGAAGTTCTCGAGGGTGGCGGA     395
      0     123   272   0     0     37.6486486486     1     0
      36.8918918919     2     0
@M00777:19:000000000-A7L7K:1:1107:24218:4614 1:N:0:1
      GGGTCGGCTAGCCATATGATCAATGAGATCAAAAAGAATGCGCAGGAGCGTATGGATGAAACCGTGGAACA
ACTGAAAAATGAACTGAGCAAAGTTCGTACCGGTGGTGGCGGCACCGAAGAGCGCCGTCTGGAACTGGCGAAACAAG
TGGTTTTTGCGGCGAATCGCGCCTTAATTCGTGTGCGTACCATCGCCCTGGAGGCGGCGTGGCGTTTACTGATGCTG
GGCAGCGATAAGGAAGTGAATAAACGCGATTGTAGCCAGGCGCTGGAGGAGATTGAAAAACTGACCAAGGTTGCCGC
GAAGAAAATCAAAGAGGTGCTGGAAGCCAAGATCAAGGAACTGCGTGAAGTTCTCGAGGGTGGCGGA     395
      0     123   272   0     0     37.3938223938     1     0
      35.8841698842     2     0
@M00777:19:000000000-A7L7K:1:1107:24501:4615 1:N:0:1
      GGGTCGGCTAGCCATATGATCAATGAGATCAAAAAGAATGCGCAGGAGCGTATGGATGAAACCGTGGAACA
ACTGAAAAATGAACTGAGCAAAGTTCGTACCGGTGGTGGCGGCACCGAAGAGCGCCGTCTTGAACTGGCGGAACAAG
GGGTTTTTGCGGCGGATCGCGCCTTAATTCGTGGTCGTACCATTCGTCTGGGGGGCGGCGTGCGTTTAATTGATGCG
GGCAGCCGATAAGAAGTTGATAAAACGCGATATTAGCCAGGCGCTGGAGGAGATTGAAAAACTGACCAAGGTTGCCG
CGAAGAAAATCAAAGAGGTGCTGGAAGCCAAGATCAAGGAACTGCGTGAATGCTCGAGGGTGGCGGA     395
      91    32    272   0     0     36.6486486486     1     0
      34.9691119691     2     0

~~~
```

3. Align fused reads to wild-type sequence and identify mutations:

```
python /path_to_scripts/Fused_read_aligner.py --path ./ --infile
<fused_reads_filename> --referenceAA <wild_type_amino_acid_sequence> --
referenceDNA <wild_type_nucleotide_sequence> --gap_max 1000 --
```

```
unresolvable_max 1000 --maxmutrun 1000 --avg_quality 1 --chaste 0 --
Ncount_max 1000 --use_N 1 --mode B > /dev/null
```

An example aligned reads file is given below: ("~~~" denotes portions omitted for brevity)

```
readID      sequence    match_count mutation_count    mutation_location
      mutation_identity max_mutation_run
@M00777:19:000000000-A7L7K:1:1107:15354:4614 1:N:0:1
      GSASHMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAKQVVFAANRALIRVRTIALEAAWR
LLMLGSDKEVNKRDISQALEEIEKLTKVAAKKIKEVLEAKIKELREVLEGGG  123   0     NA    NA
      0
@M00777:19:000000000-A7L7K:1:1107:24218:4614 1:N:0:1
      GSASHMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAKQVVFAANRALIRVRTIALEAAWR
LLMLGSDKEVNKRDCSQALEEIEKLTKVAAKKIKEVLEAKIKELREVLEGGG  122   1     85    C
      0
@M00777:19:000000000-A7L7K:1:1107:24501:4615 1:N:0:1
      GSASHMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAEQGVFAADRALIRGRTIRLGGGVR
LIDAGSR*EVDKTRY*PGAGGD*KTDQGCREENQRGAGSQDQGTA*MLEGGG  74    49
      47,49,54,60,64,66,67,68,69,72,73,74,77,78,81,83,84,85,86,87,88,89,90,91
,92,93,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113
,114,115,116,117
      E,G,D,G,R,G,G,G,V,I,D,A,R,*,D,T,R,Y,*,P,G,A,G,G,D,*,T,D,Q,G,C,R,E,E,N,Q
,R,G,A,G,S,Q,D,Q,G,T,A,*,M    0

~~~
```

4. Tally each unique sequence:

```
python /path_to_scripts/mapCounts.py --path ./ --infile
<aligned_reads_filename>
```

An example tallied reads file is given below: ("~~~" denotes portions omitted for brevity)

```
seqID sequence    match_count mutation_count    mutation_location
      mutation_identity max_mutation_run  sequence_frequency
      sequence_count
NA-NA
      GSASHMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAKQVVFAANRALIRVRTIALEAAWR
LLMLGSDKEVNKRDISQALEEIEKLTKVAAKKIKEVLEAKIKELREVLEGGG  123   0     NA    NA
      0     0.164877234271    196303
95-R
      GSASHMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAKQVVFAANRALIRVRTIALEAAWR
LLMLGSDKEVNKRDISQALEEIEKRTKVAAKKIKEVLEAKIKELREVLEGGG  122   1     95    R
      0     0.0151444522556   18031
22-A
      GSASHMINEIKKNAQERMDETVAQLKNELSKVRTGGGGTEERRLELAKQVVFAANRALIRVRTIALEAAWR
LLMLGSDKEVNKRDISQALEEIEKLTKVAAKKIKEVLEAKIKELREVLEGGG  122   1     22    A
      0     0.0121308482019   14443
```

5. Reformat tallied sequence counts for Excel:

```
#!/bin/bash
### Please note that the stop codons are represented as a "Z" in this script
and in the subsequent output files, since "*" has special meanings in bash
and causes problems if used to search patterns.
num_residues=<number_of_amino_acids_in_sequenced_region>
start_range=`echo 0`
end_range=`echo $num_residues | awk '{print $1-1}'`
filename=<tallied_reads_filename>
sed 's|*|Z|g' ${filename} > ${filename}_2
awk '/NA-NA\t[A-Z].*[A-Z]\t'$num_residues'\t0\t/ {print "num_wt_seqs:\t"$9}'
${filename}_2 > ${filename}_3
echo `seq $start_range $end_range` | tr ' ' '\t' | awk '{print
"Residues\t"$0}' >> ${filename}_3
for i in A C D E F G H I K L M N P Q R S T V W Y Z; do
      echo $i | awk '{print $1"\t"}' >> ${filename}_3
      for j in `seq $start_range $end_range`; do
            count=`awk '/.*-[A-Z]\t[A-Z].*[A-
Z]\t'$end_range'\t1\t'$j'\t'$i'\t/ {print $9}' ${filename}_2`
            value=`awk '/.*-[A-Z]\t[A-Z].*[A-
Z]\t'$end_range'\t1\t'$j'\t'$i'\t/ {print $4}' ${filename}_2`
            if [ "$value" == "1" ]; then
                  echo $count | awk '{print $1"\t"}' >> ${filename}_3
                  else echo NA | awk '{print $1"\t"}' >> ${filename}_3
            fi
      done
      echo end_of_line >> ${filename}_3
done
sed ':a;N;$!ba;s|\([A-Z0-9]\t\)\n|\1|g' ${filename}_3 > ${filename}_4
sed 's|end_of_line||g' ${filename}_4 > ${filename}_ready_for_excel.txt
rm ${filename}_[234]
```

An example reformatted tallied reads file is given below: ("~~~" denotes portions omitted for brevity)

```
num_wt_seqs:    196303
Residues  0     1     2     3     4     5     6     7     8     9     10
      11    12    13    14    15    16    17    18    19    20    21    22
      23    24    25    26    27    28    29    30    31    32    33    34
      35    36    37    38    39    40    41    42    43    44    45    46
      47    48    49    50    51    52    53    54    55    56    57    58
      59    60    61    62    63    64    65    66    67    68    69    70
      71    72    73    74    75    76    77    78    79    80    81    82
      83    84    85    86    87    88    89    90    91    92    93    94
      95    96    97    98    99    100   101   102   103   104   105   106
      107   108   109   110   111   112   113   114   115   116   117   118
      119   120   121   122
A     3     37    NA    NA    NA    NA    166   166   198   143   173   179
      135   NA    93    105   146   155   210   188   498   566   14443 127
      5     176   170   128   157   89    95    141   103   142   85    73
      117   86    90    133   221   138   103   247   76    25    NA    65
      138   145   219   245   NA    NA    84    124   NA    132   141   186
      214   131   178   139   NA    47    73    NA    NA    144   136   221
      220   47    215   158   132   115   107   186   186   196   191   96
```

```
        15      600     144     130     NA      149     106     118     507     186     95      123
        100     84      120     NA      NA      107     220     150     159     115     143     162
        121     NA      83      152     164     119     177     94      90      284     NA      2
        1       NA      5
C       1       NA      NA      14      NA      NA      79      91      72      98      83      85
        72      33      109     124     137     95      95      81      195     189     81      86
        100     131     81      65      66      99      82      95      107     78      88      123
        103     140     69      75      70      144     121     73      48      44      47      84
        90      91      89      125     115     282     98      78      47      185     71      100
        86      307     66      78      46      67      79      41      97      117     112     78
        98      43      105     154     67      65      45      54      66      117     82      145
        10      322     69      84      107     126     72      80      259     65      91      53
        52      62      107     98      55      93      161     93      54      73      80      101
        71      68      49      57      69      89      77      120     37      196     NA      NA
        45      73      NA
D       NA      NA      39      NA      1       NA      38      72      58      53      93      40
        95      34      60      77      50      90      NA      82      166     128     61      54
        54      70      81      91      30      42      59      72      32      56      73      79
        105     138     51      84      72      51      49      48      65      29      54      81
        61      52      75      86      166     183     99      52      72      140     95      65
        66      58      68      87      65      51      89      39      83      77      25      67
        52      19      90      116     60      NA      53      83      40      114     73      18
        NA      206     59      95      61      92      194     119     159     67      61      45
        33      42      80      122     21      62      66      44      48      102     45      87
        58      50      47      38      57      92      57      38      53      109     NA      50
        24      19      NA

~~~
```

6. Process the counts in Excel:

As the MB17 gene was split into front and back halves prior to Illumina sequencing, it was necessary to normalize the number of reads in each front pool to the number of reads in the corresponding back pool. The total number of reads in each front pool was normalized to the total number of reads in the corresponding back pool according to the following formulas:

```
Normalized number of front reads = (total number of front
reads)/(normalization factor)

Normalization factor = ((number of front wild type reads + number of front
mutant reads)/(number of back wild type reads + number of back mutant reads))
```

Splitting the gene into 2 parts also meant that for every mutant read in the normalized front pool, there was a wild type read in the back pool and vice versa. The true number of wild type reads was calculated according to the following formula:

```
True number of wild type reads = ((normalized number of front wild type reads
- number of back mutant reads) + (number of back wild type reads - normalized
number of front mutant reads))/2
```

Enrichment ratios were calculated as follows:

```
Enrichment ratio for unique sequence X = (count of sequence X in selected
pool/total count of all sequences in selected pool)/(count of sequence X in
unselected pool/total count of all sequences in unselected pool)
```

The following formula was used in Excel to filter the front sequence counts for those seen more than a threshold number of times, to normalize each individual front sequence count to the ratio of total front sequence counts to total back sequence counts, and to calculate Log 2 enrichment ratios:

```
=IF(AND(OR(count_of_sequence_X_in_selected_pool>counts_threshold,count_of_seq
uence_X_in_unselected_pool>counts_threshold),((count_of_sequence_X_in_selecte
d_pool*count_of_sequence_X_in_unselected_pool)>0)),(LOG(((count_of_sequence_X
_in_selected_pool/(total_count_of_all_sequences_in_front_pool/total_count_of_
all_sequences_in_back_pool)/total_count_of_all_sequences_in_selected_pool(cou
nt_of_sequence_X_in_unselected_pool/total_count_of_all_sequences_in_unselecte
d_pool)),2)),ND)
```

As the number of front reads was normalized to the number of back reads, no normalization of the number of back reads was needed. The following formula was used in Excel to filter the back sequence counts for those seen more than a threshold number of times and to calculate Log 2 enrichment ratios:

```
=IF(AND(OR(count_of_sequence_X_in_selected_pool>counts_threshold,count_of_seq
uence_X_in_unselected_pool>counts_threshold),((count_of_sequence_X_in_selecte
d_pool*count_of_sequence_X_in_unselected_pool)>0)),(LOG(((count_of_sequence_X
_in_selected_pool/total_count_of_all_sequences_in_selected_pool(count_of_sequ
ence_X_in_unselected_pool/total_count_of_all_sequences_in_unselected_pool)),2
)),ND)
```

6.6 Balancing the surface charges of the top MB17 variants

In an attempt to improve the behavior of the MB17.4.1, MB17.4.2, MB17.2.1, and MB17.2.2 in mammalian cells, we sought to reduce the number of negative charges and

hydrophobic residues on their surfaces. First we identified the following 6 subsets of residues on each of the MB17 variants: 1. Core hydrophobic residues, 2. Surface polar residues believed to contribute to stability, 3. Surface residues believed to contribute to binding or specificity for Mdmx or Mdm2, 4. Surface hydrophobic residues not essential for binding to Mdmx or Mdm2, 5. Surface polar residues not essential for protein stability, and 6. Surface residues that could be reverted to the residue identity found in OCR (116). Next we mutated residues in subset 4 to polar residues that were enriched during deep mutational scanning and scored well in Rosetta (108). Then we mutated acidic residues in subset 5 to basic residues if the mutation was not depleted during deep mutational scanning, the mutation scored well in Rosetta, and the mutation balanced the surface charge distribution better than the original residue identity. Finally we reverted residues in subset 6 above to the identity found in OCR if the native residue was enriched during deep mutational scanning and if the native residue scored well in Rosetta. Mutations to residues in subset 6 were not critical to balancing the surface charge distribution or reducing the surface hydrophobic load of the MB17 variants and could potentially negatively affect binding to the desired target. We thus made a number of sub-variants for each starting MB17 variant that incorporated all of the described changed to residues in subsets 4 and 5, but incorporated mutations to none, some, or all of the residues in subset 6. The electrostatics calculator available in PyMOL (Schrödinger LLC, New York City, New York, USA) was used to estimate the distribution of charge across the surface of Mdmx, Mdm2, and the designed binder during the redesign process.

The Mdmx specificity-only enrichments were used during the redesign of MB17.4.1 and MB17.4.2 because the Mdmx specificity-only selection tended to enrich mutations that appeared to stabilize the structure of MB17 while the Mdmx affinity-only selection tended to enrich

mutations that introduced more hydrophobic residues to the surface of MB17 (presumably to increase the number of hydrophobic interactions across the MB17-Mdmx interface). The Mdm2 affinity-only enrichments were used during the redesign of MB17.2.1 and MB17.2.2 because the Mdm2 affinity-only selection tended to enrich mutations that appeared to stabilize the structure of MB17 while the Mdm2 specificity-only selection tended to enrich mutations that appeared to destabilize the structure of MB17 presumably because an unfolded MB17 peptide would bind Mdmx and Mdm2 equally well and this be more specific for Mdm2 than was the parent MB17). Between 2 and 4 redesigned proteins were produced for each of the 4 starting point variants. A representative redesigned protein for each starting point variant was subjected to ab initio structure prediction using only its primary amino acid sequence and the top scoring models were compared to the design models (115). DnaWorks was used to generate a nucleotide sequence for each designed amino acid sequence (114).

Section 7 (Appendix)

Computational methods specific to the design of proteins to inhibit the interaction of EED-Ezh2

7.1 Computational methods: Design of EED binding proteins

1. Identifying candidate scaffolds:

Residues 42-60 from the crystal structure of Ezh2 (taken from PDB ID 2qxv)(118) were iteratively aligned with every possible 19-residue stretch within a set of 34467 monomeric proteins (scaffolds) taken from the Protein Data Bank (PDB) using the Epigraft application in Rosetta (108, 119, 120).

A sample commandline to run the Epigraft matching step is given below:

```
/path_to_rosetta/rosetta.intel -paths ../paths.txt -epi_graft -match -
rough_match -input_file <scaffold_list_file> -output_file out_rough -nres_Ab
<number_of_residues_in_target_protein_chain(in this case 352)> -E_align -
S_align -fluidize_takeoff -max_closure_rms 1.5 -fluidize_landing -
native_complex <pdb_structure_of_native_complex> -loop_ranges <ranges file> -
compute_cbeta_neighbors -rough_match_closure_rms 1.5 -max_intra_clash 3000 -
max_inter_clash 10000 -termini_residue_skip 4
```

An example "paths.txt" file is given below:

```
Rosetta Input/Output Paths (order essential)
path is first '/', './',or  '../' to next whitespace, must end with '/'
INPUT PATHS:
pdb1                          ./
pdb2                            /net/pdb/
alternate data files            /net/shared/rosetta_database/
fragments                       ./
structure dssp,ssa (dat,jones)  ./
sequence fasta,dat,jones        ./
constraints                     ./
starting structure            ./
data files                      /scratch/ROSETTA/rosetta_database/
OUTPUT PATHS:
movie                           ./
pdb path                        ./output/
score                           ./
status                          ./
user                            ./
FRAGMENTS: (use '*****' in place of pdb name and chain)
2                                      number of valid fragment files
3                                      frag file 1 size
aa*****03_05.200_v1_3                              name
9                                      frag file 2 size
aa*****09_05.200_v1_3                              name
-------------------------------------------------------------------------
CVS information:
$Revision: 1.10 $
$Date: 2002/07/16 16:47:31 $
```

```
$Author: rohl $
--------------------------------------------------------------------------
```

An example “scaffold list file” is given below: (“~~~” denotes portions omitted for brevity).

```
path_to_scaffolds/01/101ma.pdb
path_to_scaffolds/01/102la.pdb
path_to_scaffolds/01/102ma.pdb

~~~

path_to_scaffolds/ra/9rata.pdb
path_to_scaffolds/ra/9rnta.pdb
path_to_scaffolds/ra/9rnta.pdb
```

An example ranges file is given below:

```
loop: 1
full_range: 353 381
nranges: 1
range: 355 373
```

An example of the output from the Epigraft matching step is given below: (“~~~” denotes portions omitted for brevity).

```
a1/2a1jB.pdb           1             E             6           23
355                 373          5:0:-64.951065         5:1:-44.615498
6:0:-72.98307          6:1:-49.676495         23:0:-85.84198        23:1:-
5.6372075        24:0:-69.804115        24:1:-39.077366       0.44630823
-0.87222779      -0.20006934       0.89104259       0.41246533
0.18951385      -0.08277757      -0.26285189       0.96127874      -
18.22481346       51.68355942      -30.45912743           1.109
1.220              1.073                 inf          907.089        3054.459
61.607        30.799                 13                  -

~~~

k5/1k53A.pdb           1             S            22          40
355                 373                        -                       -
-                     -                      -                       -
-                     -      -0.79238760      -0.34577596       0.50255436
0.55912358      -0.08224291       0.82499516      -0.24393195
0.93470591       0.25849974      115.39505005        8.01579285      -
40.18924713          1.095             1.606              2.193
0.058         372.293        8511.128        10.285         22.899
9                     -
```

2. Grafting key side chains onto candidate scaffolds:

To be considered for grafting, scaffolds had to match the backbone of the Ezh2 19-residue stretch with a backbone RMSD ≤ 1.5 Å and had to allow the matched region to bind the Ezh2-binding pocket of EED (taken from PDB ID 2qxv) in the same position as Ezh2 without

introducing significant clashes between the backbone of the scaffold and the backbone or side chains of EED.

An example formatted list of matches is given below: ("~~~" denotes portions omitted for brevity).

```
1le2_1_rlx.pdb 1 S 73 91 355 373 $ keep_bb graft_sc
1le2_1_rlx.pdb 1 S 77 95 355 373 $ keep_bb graft_sc
1le2_1_rlx.pdb 1 S 84 102 355 373 $ keep_bb graft_sc

~~~

3ef8_1_rlx.pdb 1 S 7 25 355 373 $ keep_bb graft_sc
3f4m_1_rlx.pdb 1 S 24 42 355 373 $ keep_bb graft_sc
3lf9_1_rlx.pdb 1 S 47 65 355 373 $ keep_bb graft_sc
```

An example commandline to run the Epigraft grafting step is given below:

```
/path_to_rosetta/rosetta.intel -epi_graft -multigraft -nres_Ab 352 -
native_complex <pdb structure of the native complex> -loop_ranges <ranges
filename> -input_file <formatted_list_of_matches> -
use_non_monotone_line_search -atom_vdw_set highres -ex1 -ex1aro -ex2 -
extrachi_cutoff 0 -try_both_his_tautomers -dump_predesign
<output_design_filename> -output_file <output_data_filename> -
Ab_epitope_optimize -repack_Ab -design_attempts 10 -store_n_best_designs 1 -
paths ./paths.txt -keep_natro <list_of_residues_to_graft> -
design_after_closure
```

An example list of residues to graft is given below:

```
355
358
362
365
366
369
373
```

3. Screening Epigraft match output for candidate designs:

A total of 306 grafted models in 291 scaffolds were visually screened in PyMOL to remove scaffolds that were not truly monomeric, that required metals to fold, that were over 150 residues in length, that introduced unresolvable clashes with EED, whose endpoints differed to much from the aligned region of Ezh2, or that had large cavities, yielding 8 candidates (in 6 scaffolds) for design.

4. Remodeling selected scaffolds for improved experimental properties:

Scaffold 1wqg (Mycobacterium tuberculosis ribosome recycling factor)(122) consists of a 3 helix bundle with an inserted beta sheet domain at one end of the bundle. Homologous scaffold 3lf9 (HIV 4E10 epitode scaffold), based on PDB ID 1ise (Escherichia coli ribosome recycling factor), replaces the beta sheet domain with a 4-Glycine loop (121). In order to reduce

the size of 1wqg and increase it's chances for expression in yeast and bacteria, this beta sheet domain was removed in PyMOL and replaced with 4 Glycine residues using the Rosetta Remodel application (123).

An example commandline to run Remodel is given below:

```
/path_to_rosetta/remodel.static.linuxiccrelease -database
/path_to_rosetta_database/rosetta_database/ -s <input_structure> -
remodel:blueprint <blueprint file> -ex1 -ex2 -correct -num_trajectory 1 -
remodel:use_blueprint_sequence -use_input_sc -save_top 1 -chain A -
remodel:quick_and_dirty -remodel:use_pose_relax -run_confirmation true -
preserve_header true -use_clusters false -overwrite -score:weights
score12prime -hbond_params sp2_params -corrections:score:hb_sp2_chipen -
lj_hbond_hdis 1.75 -lj_hbond_OH_donor_dis 2.6 -ignore_zero_occupancy false
```

An example blueprint file is given below: ("~~~" denotes portions omitted for brevity).

```
1 I .
2 D .
3 E .

~~~

25 T .
26 I .
27 R .
28 T L PIKAA T
0 x L PIKAA G
0 x L PIKAA G
0 x L PIKAA G
0 x L PIKAA G
29 T L PIKAA T
30 E .
31 E .
32 R .

~~~

106 L .
107 E .
108 V .
```

5. Preparing Epigraft output models for surface design:

Since only select side chains and not the backbone of the Ezh2 N-terminnus are grafted onto the scaffold protein, the scaffold cannot orient the grafted side chains exactly as does the N-terminus of Ezh2. As the Epigraft grafting step automatically adjusts side chain orientations during the grafting step, these residues were sometimes modeled in positions incompatible with binding to EED. This was particularly a problem with Tryptophan 60 from Ezh2, but was also observed with Phenylalanine 42 and Asparagine 45 from Ezh2 and Glutamate 26, Phenylalanine 284, Tryptophan 285, and Arginine 332 from EED. These incompatible side chain positions were

manually adjusted to the orientations observed in the crystal structure of EED-Ezh2 complex using Foldit (Ref for Foldit).

Further, as the Epigraft grafting step alters the identities of scaffold residues using an older score function, a fixed-backbone design step was included to reverse unwanted mutations and allow redesign of other mutations. For each binder-EED complex, a resfile was manually generated that 1. prevented movement or design of manually adjusted residues, 2. forced reversal of mutations to core residues, 3. allowed other mutated residues to design to any residue except Cysteine, Glycine, Histidine, Phenylalanine, Tryptophan, or Tyrosine, 4. removed unwanted Glycine or Cysteine residues from the scaffold, and 5. added additional binding residues taken from Ezh2 in a scaffold-dependent manner. This resfile was used by RosettaScripts to guide side-chain conformational sampling and mutation of selected binder residues in the presence of EED, without backbone flexibility. The score function was additionally biased using a position-specific scoring matrix (pssm) based on alignments of each scaffold with homologous proteins.

A sample commandline to run the fixed-backbone design step is given below:

```
/path_to_rosetta/rosetta_scripts.static.linuxiccrelease -database
/path_to_rosetta_database/database -ignore_zero_occupancy false -
ignore_unrecognized_res -overwrite -out:file:renumber_pdb false -ex1 -ex2 -
nstruct 1 -parser:script_vars resfile=<resfile_filename> pssm=<pssm_filename>
-s <input_structure> -parser:protocol ./fixbb_dsn.xml -score:weights
talaris2013 \
```

A sample xml protocol to run the fixed-backbone design step is given below:

```
<ROSETTASCRIPTS>
	<SCOREFXNS>
		<sfxn_hard weights=talaris2013/>
	</SCOREFXNS>
	<TASKOPERATIONS>
		<LimitAromaChi2 name=arochi2/>
		<DisallowIfNonnative name=nocys disallow_aas=CGH/>
		<RestrictChainToRepacking name=rctr chain=1/>
		<ReadResfile name=resfile filename="%%resfile%%"/>
	</TASKOPERATIONS>
	<MOVERS>
		<FavorSequenceProfile name=fsp_pssm scaling=global chain=2
pssm="%%pssm%%" weight=1.0 scorefxns=sfxn_hard/>
		<AtomTree name=ftree simple_ft=1/>
		<PackRotamersMover name=packdsn scorefxn=sfxn_hard
task_operations=rctr,arochi2,nocys,resfile/>
		<MinMover name=min_bb_sc bb=1 chi=1 jump=0 scorefxn=sfxn_hard>
			<MoveMap name=pin_down_hs>
				<Chain number=2 chi=1 bb=1/>
				<Chain number=1 chi=1 bb=1/>
				<Jump number=1 setting=0/>
			</MoveMap>
		</MinMover>
	</MOVERS>
	<FILTERS>
		<Ddg name=binding_energy threshold=0 scorefxn=sfxn_hard
confidence=0 jump=1 repack=1 relax_mover=min_bb_sc repeats=3/>
```

```
            <Sasa name=dsasa threshold=500 confidence=0/>
            <ShapeComplementarity name=shape_comp jump=1 verbose=0
min_sc=0.60 confidence=0/>
            <SymUnsatHbonds name=unsat jump=1 cutoff=1000 confidence=0/>
      </FILTERS>
      <PROTOCOLS>
            <Add mover_name=ftree/>
            <Add mover_name=fsp_pssm/>
            <Add mover_name=packdsn/>
            <Add filter_name=binding_energy/>
            <Add filter_name=dsasa/>
            <Add filter_name=shape_comp/>
            <Add filter_name=unsat/>
      </PROTOCOLS>
</ROSETTASCRIPTS>
```

A sample resfile is used during the fixed-backbone design step given below: ("~~~" denotes portions omitted for brevity).

```
NATAA
START
332 A NATRO
284 A NATRO
285 A NATRO
364 B NOTAA CGHWYF
368 B PIKAA T
394 B PIKAA V
395 B NOTAA CGHWYF
396 B NOTAA CGHWYF
398 B NOTAA CGHWYF
399 B NATRO
400 B NOTAA CGHWYF
403 B PIKAA R
404 B NOTAA CGHWYF
405 B NOTAA CGHWYF
406 B NATRO
407 B NOTAA CGHWYF
409 B NATRO
410 B NATRO
413 B NATRO
414 B NOTAA CGHWYF
417 B NATRO
418 B NOTAA CGHWYF
424 B NOTAA CGHWYF
437 B PIKAA L
441 B NOTAA CGHWYF
444 B PIKAA T
448 B NOTAA CGHWYF
452 B NOTAA CGHWYF
455 B PIKAA L
459 B NOTAA CGHWYF
```

A sample pssm file is given below: ("~~~" denotes portions omitted for brevity).

```
Last position-specific scoring matrix computed, weighted observed percentages
rounded down, information per position, and relative weight of gapless real
matches to pseudocounts
           A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V   A   R
N   D   C   Q   E   G   H   I   L   K   M   F   P   S   T   W   Y   V
    1 M   -2 -3 -3 -4 -2 -2 -3 -4 -3  2  2 -2  8 -1 -4 -3 -2 -2 -2  1    0
0   0   0   0   0   0   0   0  12  10   0  75   0   0   0   0   0   0   3
0.78 0.15
    2 I   -2 -4 -4 -4 -2 -3 -4 -5 -4  6  1 -3  1 -1 -4 -3 -1 -3 -2  3    0
0   0   0   0   0   0   0   0  85   3   0   0   0   0   0   2   0   0  10
0.69 0.13
    3 N   -2 -1  7  0 -3 -1 -1 -1  0 -4 -4  0 -3 -4 -3  1  1 -4 -3 -3    0
0  80   0   0   0   1   0   0   0   0   3   0   0   0   6  10   0   0   0
0.90 0.15

~~~

  118 H   -2  0  1 -1 -3  0  0 -2  8 -3 -3 -1 -2 -1 -2 -1 -2 -2  2 -3    0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0.00    0.00
  119 H   -2  0  1 -1 -3  0  0 -2  8 -3 -3 -1 -2 -1 -2 -1 -2 -2  2 -3    0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0.00    0.00
  120 H   -2  0  1 -1 -3  0  0 -2  8 -3 -3 -1 -2 -1 -2 -1 -2 -2  2 -3    0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0.00    0.00

                      K         Lambda
Standard Ungapped    0.1301     0.3128
Standard Gapped      0.0410     0.2670
PSI Ungapped         0.1883     0.3179
PSI Gapped           0.0576     0.2670
```

6. Optimizing the identities of surface residues of the designed binders:

The models output from fixed_backbone design were then redesigned 20 times in parallel using the fastdesign with atompair constraints protocol. A resfile was manually generated that allowed surface residues on the binder to change to any amino acid except Cysteine, Glycine, Histidine, Phenylalanine, Proline, Tryptophan, or Tyrosine, allowed binder core residues and all EED residues to sample different side chain conformations without changing their identities, and prevented movement of selected binding residues. The score function was additionally biased using a position-specific scoring matrix (pssm) based on alignments of each scaffold with homologous proteins.

A sample commandline used for the surface design step is given below:

```
/path_to_rosetta/rosetta_scripts.default.linuxgccrelease -database
/path_to_rosetta_databse/database -ignore_zero_occupancy false -
ignore_unrecognized_res -overwrite -out:file:renumber_pdb false -ex1 -ex2 -
nstruct 1 -parser:script_vars resfile=<resfile_filename> pssm=<pssm_filename>
```

```
-s <input_pdb_structure> -
parser:protocol ./flxbb_fstdsn_resfile_atompaircst.xml -score:weights
talaris2013
```

A sample xml protocol used for the surface design step is given below:

```
<ROSETTASCRIPTS>
      <SCOREFXNS>
            <sfxn_hard weights=talaris2013>
                  <Reweight scoretype=atom_pair_constraint weight=1.0/>
            </sfxn_hard>
      </SCOREFXNS>
      <TASKOPERATIONS>
            <OperateOnCertainResidues name=chainA>
                  <PreventRepackingRLT/>
                  <ChainIs chain=A/>
            </OperateOnCertainResidues>
            <OperateOnCertainResidues name=chainB>
                  <PreventRepackingRLT/>
                  <ChainIs chain=B/>
            </OperateOnCertainResidues>
            <LimitAromaChi2 name=arochi2/>
            <DisallowIfNonnative name=nocys disallow_aas=CGHWYF/>
            <ProteinInterfaceDesign name=pido repack_chain1=1 repack_chain2=1
design_chain1=0 design_chain2=1 jump=1 interface_distance_cutoff=10
allow_all_aas=1/>
            <RestrictChainToRepacking name=rctr chain=1/>
            <ReadResfile name=resfile filename="%%resfile%%"/>
      </TASKOPERATIONS>
      <MOVERS>
            <FavorSequenceProfile name=fsp_pssm scaling=global chain=2
pssm="%%pssm%%" weight=1.0 scorefxns=sfxn_hard/>
            <AtomTree name=ftree simple_ft=1/>
            <MinMover name=min_bb_sc bb=1 chi=1 jump=0 scorefxn=sfxn_hard>
                  <MoveMap name=minmvrmap>
                        <Chain number=2 chi=1 bb=1/>
                        <Chain number=1 chi=1 bb=1/>
                        <Jump number=1 setting=1/>
                  </MoveMap>
            </MinMover>
            <FastRelax name=fstrlx_dsn scorefxn=sfxn_hard repeats=1
task_operations=pido,arochi2,rctr,nocys,resfile>
                  <MoveMap name=fstrlxmap>
                        <Chain number=2 chi=1 bb=1/>
                        <Chain number=1 chi=1 bb=1/>
                        <Jump number=1 setting=1/>
                  </MoveMap>
            </FastRelax>
            <AddConstraintsToCurrentConformationMover
name=add_pair_chainA_cst use_distance_cst=1 coord_dev=0.5 bound_width=0.1
min_seq_sep=8 max_distance=12.0 cst_weight=1.0 task_operations=chainA/>
            <AddConstraintsToCurrentConformationMover
name=add_pair_chainB_cst use_distance_cst=1 coord_dev=0.5 bound_width=0.1
min_seq_sep=8 max_distance=12.0 cst_weight=1.0 task_operations=chainB/>
            <ClearConstraintsMover name=clear_cst/>
      </MOVERS>
```

```
	<FILTERS>
		<Ddg name=binding_energy threshold=0 scorefxn=sfxn_hard
confidence=0 jump=1 repack=1 relax_mover=min_bb_sc repeats=3/>
		<Sasa name=dsasa threshold=500 confidence=0/>
		<ShapeComplementarity name=shape_comp jump=1 verbose=0
min_sc=0.60 confidence=0/>
		<SymUnsatHbonds name=unsat jump=1 cutoff=1000 confidence=0/>
		<ScoreType name=total_score_complex scorefxn=sfxn_hard
score_type=total_score confidence=0 threshold=0/>
	</FILTERS>
	<PROTOCOLS>
		<Add mover_name=ftree/>
		<Add mover_name=fsp_pssm/>
		<Add mover=add_pair_chainA_cst/>
		<Add mover=add_pair_chainB_cst/>
		<Add mover_name=fstrlx_dsn/>
		<Add mover=clear_cst/>
		<Add filter_name=binding_energy/>
		<Add filter_name=dsasa/>
		<Add filter_name=shape_comp/>
		<Add filter_name=unsat/>
		<Add filter_name=total_score_complex/>
	 </PROTOCOLS>
</ROSETTASCRIPTS>
```

A sample resfile used for the surface design step is given below: ("~~~" denotes portions omitted for brevity).

```
NATAA
START
332 A NATRO
284 A NATRO
285 A NATRO
355 B NOTAA CGHWYFP
356 B NOTAA CGHWYFP
358 B NOTAA CGHWYFP

~~~

464 B NOTAA CGHWYFP
466 B NOTAA CGHWYFP
467 B NOTAA CGHWYFP
399 B NATRO
406 B NATRO
409 B NATRO
410 B NATRO
413 B NATRO
417 B NATRO
```

7. Refining the designed binders using greedy optimization:

For each binder, the best-scoring model according to total score, shape complementarity, binding energy, and buried unsatisfied hydrogen bonds was then subjected to greedy optimization 20 times in parallel. The greedy algorithm individually scores all allowed mutations according to predefined combination of filters and then attempts to incorporate all of the

mutations into the design in order of best scoring to worst scoring. A mutation is only incorporated if it improves the combined score of all filter criteria. Otherwise that mutation is skipped (124). Running the greedy algorithm 20 X in parallel allows the time-consuming step of scoring all allowed mutations to be split up among 20 different processors, which then share the resulting data during the incorporation step. The filter criteria used here were total score, shape complementarity, binding energy, and buried unsatisfied hydrogen bonds, where the value of each filter criteria was divided by a factor equal to that filter's value in the input model. The filters total score, shape complementarity, and binding energy were additionally multiplied by a factor of negative 1. These factors had the effect of giving each of the four filters equal weight in determining whether a mutation should be incorporated or skipped during the design step. The residue positions and identities to be scored during the greedy run were defined using the same resfile used for the surface design run and the score function was additionally biased using the aforementioned pssm.

A sample script used to automatically set the factors for the combined filters and launch the greedy MPI run is given below:

```
#!/bin/bash

pssm=<pssm_filename>
resfile=<resfile_filename>
pdb=<input_structure>
total_score=`grep ^total_score $pdb | awk '{print 1/$2*(0-1)}'`
shape=`grep ^shape $pdb | awk '{print 1/$2*(0-1)}'`
unsat=`grep ^unsat $pdb | awk '{print 1/$2}'`
binding=`grep ^binding $pdb | awk '{print 1/$2*(0-1)}'`

mpirun -np 20 /path_to_rosetta/rosetta_scripts.mpi.linuxgccrelease -database
/path_to_rosetta_database/database -ignore_zero_occupancy false -
ignore_unrecognized_res -overwrite -out:file:renumber_pdb false -ex1 -ex2 -
nstruct 19 -parser:script_vars resfile=${resfile} pssm=${pssm}
total_score=${total_score} shape=${shape} binding=${binding} unsat=${unsat} -
s $pdb -parser:protocol ./greedy_no_min.xml -score:weights talaris2013
```

A sample xml protocol used for the MPI greedy run is given below:

```
<ROSETTASCRIPTS>
      <SCOREFXNS>
            <sfxn_hard weights=talaris2013/>
      </SCOREFXNS>
      <TASKOPERATIONS>
            <LimitAromaChi2 name=arochi2/>
            <DisallowIfNonnative name=nocys disallow_aas=CGHWYFP/>
            <ProteinInterfaceDesign name=pido repack_chain1=1 repack_chain2=1
design_chain1=0 design_chain2=1 jump=1 interface_distance_cutoff=10
allow_all_aas=1/>
            <RestrictChainToRepacking name=rctr chain=1/>
            <ReadResfile name=resfile filename="%%resfile%%"/>
      </TASKOPERATIONS>
      <MOVERS>
            <FavorSequenceProfile name=fsp_pssm scaling=global chain=2
pssm="%%pssm%%" weight=1.0 scorefxns=sfxn_hard/>
            <AtomTree name=ftree simple_ft=1/>
```

```
			<MinMover name=min_bb_sc bb=0 chi=1 jump=1 scorefxn=sfxn_hard>
				<MoveMap name=minmvrmap>
					<Chain number=2 chi=1 bb=0/>
					<Chain number=1 chi=1 bb=0/>
					<Jump number=1 setting=1/>
				</MoveMap>
			</MinMover>
		</MOVERS>
		<FILTERS>
			<Ddg name=binding threshold=0 scorefxn=sfxn_hard confidence=0
jump=1 repack=1 relax_mover=min_bb_sc repeats=3/>
			<Sasa name=dsasa threshold=500 confidence=0/>
			<ShapeComplementarity name=shape jump=1 verbose=0 min_sc=0.60
confidence=0/>
			<SymUnsatHbonds name=unsat jump=1 cutoff=1000 confidence=0/>
			<ScoreType name=total_score scorefxn=sfxn_hard
score_type=total_score confidence=0 threshold=0/>
			<CombinedValue name=combo>
				<Add filter_name=total_score factor="%%total_score%%"/>
				<Add filter_name=unsat factor="%%unsat%%"/>
				<Add filter_name=binding factor="%%binding%%"/>
				<Add filter_name=shape factor="%%shape%%"/>
			</CombinedValue>
		</FILTERS>
		<MOVERS>
			<GreedyOptMutationMover name=greedy
task_operations=pido,arochi2,rctr,nocys,resfile filter=combo
scorefxn=sfxn_hard relax_mover=min_bb_sc sample_type=low rtmin=0
design_shell=-1 repack_shell=8.0 parallel=1/>
		</MOVERS>
		<PROTOCOLS>
			<Add mover_name=ftree/>
			<Add mover_name=fsp_pssm/>
			<Add mover_name=greedy/>
			<Add filter_name=binding/>
			<Add filter_name=dsasa/>
			<Add filter_name=shape/>
			<Add filter_name=unsat/>
			<Add filter_name=total_score/>
		</PROTOCOLS>
</ROSETTASCRIPTS>
```

For each candidate binder, the amino acid sequences of the models output from each of the 20 greedy runs were aligned in Geneious (Biomatters Inc. San Francisco, CA, USA) and a consensus sequence was determined. The model with the sequence closest to the consensus was modified to have the consensus sequence using Foldit. An electrostatics map was created using PyMOL for each modified model and non-complementary charge interactions across the binder-EED interface were flagged. These flagged interactions were manually adjusted in Foldit by

mutating the residue on the binder side of the interface to the most appropriate identity and the resulting model was re-evaluated in PyMOL. At this point, N-linked glycosylation sites on the binder were identified by the consensus sequence NXS/T and removed by mutating either the Asparagine or the Serine/Threonine in Foldit.

Once all non-complementary charged interactions and N-linked glycosylation sites were corrected, the modeled interface was validated by subjecting the model to side chain conformational sampling and side chain and backbone minimization without any constraints using the fast design with atompair constraints protocol in both the bound and unbound states. None of the 8 candidate designs saw dissociation of the modeled complex when relaxed in the bound state with no constraints. The validated models were subjected to fragment-based ab initio structure prediction using only their primary amino acid sequences and the top scoring models were compared to the design models.

A sample script to generate an unbound binder model, identify the grafted region, and relax the binder in the unbound state is given below:

```
#!/bin/bash

hs_finder=`/path_to_residue_finder/residue_finder.sh <input_model> PHE 268 CZ
A PHE CZ B`
hs_list=`echo $hs_finder | awk '{print
$4","$4+3","$4+7","$4+10","$4+11","$4+14","$4+18}'`
hs_range_start=`echo $hs_finder | awk '{print $4-352}'`
hs_range_end=`echo $hs_finder | awk '{print $4+18-352}'`
mnmr_flank_start=`echo $hs_finder | awk '{print $4-7-352}'`
mnmr_flank_end=`echo $hs_finder | awk '{print $4+25-352}'`
pdb=`echo <input_model> | sed 's|\.pdb||'`
grep ^ATOM <input_model> | grep ' B ' > ${pdb}_chain_B.pdb
/path_to_scripts/convpdb.pl -renumberAcrossChains ${pdb}_chain_B.pdb >
${pdb}_chain_B_rlx.pdb
rm -f ${pdb}_chain_B.pdb

/path_to_rosetta/rosetta_scripts.default.linuxgccrelease -database
/path_to_rosetta_database/database -ignore_zero_occupancy false -
ignore_unrecognized_res -overwrite -out:file:renumber_pdb false -ex1 -ex2 -
nstruct 1 -parser:script_vars hs_list="$hs_list"
mnmr_flank_start=${mnmr_flank_start} mnmr_flank_end=${mnmr_flank_end}
hs_range_start=${hs_range_start} hs_range_end=${hs_range_end} -s
```

```
${pdb}_chain_B_rlx.pdb -in:file:native ${pdb}_chain_B_rlx.pdb -
parser:protocol ./validate_monomer.xml -score:weights talaris2013
```

A sample xml protocol to relax and score the binder in the unbound state is given below:

```
<ROSETTASCRIPTS>
	<SCOREFXNS>
		<sfxn_hard weights=talaris2013/>
	</SCOREFXNS>
	<TASKOPERATIONS>
		<LimitAromaChi2 name=arochi2/>
		<RestrictChainToRepacking name=rctr2 chain=1/>
	</TASKOPERATIONS>
	<MOVERS>
		<FastRelax name=fstrlx_monomer scorefxn=sfxn_hard repeats=8
task_operations=rctr2,arochi2/>
		<Superimpose name=super ref_start=%%mnmr_flank_start%%
ref_end=%%mnmr_flank_end%% target_start=%%mnmr_flank_start%%
target_end=%%mnmr_flank_end%%/>
	</MOVERS>
	<FILTERS>
		<ScoreType name=total_score_monomer_after scorefxn=sfxn_hard
score_type=total_score confidence=0 threshold=0/>
		<Rmsd name=rmsd_all_rlx chains=B threshold=5 confidence=0
superimpose=1/>
		<Rmsd name=rmsd_graft_rlx threshold=5 confidence=0 superimpose=1>
			<span begin_res_num=%%hs_range_start%%
end_res_num=%%hs_range_end%%/>
		</Rmsd>
	</FILTERS>
	<PROTOCOLS>
		<Add mover_name=fstrlx_monomer/>
		<Add mover_name=super/>
		<Add filter_name=total_score_monomer_after/>
		<Add filter_name=rmsd_all_rlx/>
		<Add filter_name=rmsd_graft_rlx/>
	</PROTOCOLS>
</ROSETTASCRIPTS>
```

A sample script to identify the grafted region and relax the binder and EED in the bound state is given below:

```
#!/bin/bash

hs_finder=`/ path_to_residue_finder/residue_finder.sh <input_model> PHE 268
CZ A PHE CZ B`
hs_list=`echo $hs_finder | awk '{print
$4","$4+3","$4+7","$4+10","$4+11","$4+14","$4+18}'`
hs_range_start=`echo $hs_finder | awk '{print $4}'`
hs_range_end=`echo $hs_finder | awk '{print $4+18}'`

/path_to_rosetta/rosetta_scripts.default.linuxgccrelease -database
/path_to_rosetta_database/database -ignore_zero_occupancy false -
ignore_unrecognized_res -overwrite -out:file:renumber_pdb false -ex1 -ex2 -
nstruct 1 -s <input_model>  -parser:script_vars hs_list=${hs_list}
```

```
hs_range_start=${hs_range_start} hs_range_end=${hs_range_end} -
parser:protocol ./validate_complex.xml -score:weights talaris2013
```

A sample xml protocol to relax and score the binder and EED in the bound state is given below:

```
<ROSETTASCRIPTS>
	<SCOREFXNS>
		<sfxn_hard weights=talaris2013/>
	</SCOREFXNS>
	<TASKOPERATIONS>
		<LimitAromaChi2 name=arochi2/>
		<RestrictChainToRepacking name=rctr chain=1/>
		<RestrictChainToRepacking name=rctr2 chain=2/>
	</TASKOPERATIONS>
	<MOVERS>
		<AtomTree name=ftree simple_ft=1/>
		<MinMover name=min_bb_sc bb=0 chi=0 jump=0 scorefxn=sfxn_hard>
			<MoveMap name=minmvrmap>
				<Chain number=2 chi=0 bb=0/>
				<Chain number=1 chi=0 bb=0/>
				<Jump number=1 setting=0/>
			</MoveMap>
		</MinMover>
		<FastRelax name=fstrlx_dsn scorefxn=sfxn_hard repeats=1
task_operations=arochi2,rctr,rctr2>
			<MoveMap name=fstrlxmap>
				<Chain number=2 chi=1 bb=0/>
				<Chain number=1 chi=1 bb=0/>
				<Jump number=1 setting=1/>
			</MoveMap>
		</FastRelax>
	</MOVERS>
	<FILTERS>
		<Ddg name=binding_energy threshold=0 scorefxn=sfxn_hard
confidence=0 jump=1 repack=1 relax_mover=min_bb_sc repeats=3/>
		<Sasa name=dsasa threshold=500 confidence=0/>
		<ShapeComplementarity name=shape_comp jump=1 verbose=0
min_sc=0.60 confidence=0/>
		<SymUnsatHbonds name=unsat jump=1 cutoff=1000 confidence=0/>

		<ScoreType name=total_score_complex scorefxn=sfxn_hard
score_type=total_score confidence=0 threshold=0/>
		<Rmsd name=rmsd_all chains=B threshold=5 confidence=0
superimpose=1/>
		<Rmsd name=rmsd_graft threshold=5 confidence=0 superimpose=1>
			 <span begin_res_num=%%hs_range_start%%
end_res_num=%%hs_range_end%%/>
		</Rmsd>
	</FILTERS>
	<PROTOCOLS>
		<Add mover_name=ftree/>
		<Add mover_name=fstrlx_dsn/>
		<Add filter_name=binding_energy/>
		<Add filter_name=dsasa/>
		<Add filter_name=shape_comp/>
		<Add filter_name=unsat/>
```

```
            <Add filter_name=rmsd_all/>
            <Add filter_name=rmsd_graft/>
            <Add filter_name=total_score_complex/>
      </PROTOCOLS>
</ROSETTASCRIPTS>
```

7.2 Redesign of EB22 loop

To identify the region and residues from EB22 to replace with the corresponding region and residues from 3lhp, the model of EB22 bound to EED was aligned in PyMOL to the PDB structure 3lhp (4E10 HIV Epitope scaffold)(121). Next a blueprint file was generated that directed the Remodel application (123) to perform the replacement and Remodel was run 20 X. The 20 output models were aligned to the model of EB22 bound to EED and the model that most closely matched the input EB22 model in regions outside the grafted region was used for further work.

An example commandline used to run Remodel is given below:

```
/path_to_rosetta/remodel.linuxiccrelease -database
/path_to_rosetta_databse/rosetta_database/ -s <input_model> -
remodel:blueprint <blueprint_file> -ex1 -ex2 -correct -num_trajectory 10 -
save_top 1 -remodel:quick_and_dirty -use_clusters false  -overwrite -
jd2:no_output
```

An example blueprint file is given below: ("~~~" denotes portions omitted for brevity)

```
1 F .
2 K .
3 C .

~~~

372 L .
373 K .
374 N .
375 E H PIKAA E
376 L H PIKAA L
377 S H PIKAA D
378 K H PIKAA K
379 V H PIKAA F
380 R H PIKAA K
381 T H NOTAA C
382 G H NOTAA C
```

```
383 G H PIKAA V
384 G L PIKAA R
0 X L PIKAA K
0 X L PIKAA V
0 X L NOTAA C
385 G L PIKAA P
386 T L PIKAA T
387 E H PIKAA E
388 E H PIKAA E
389 R H PIKAA R
390 R H PIKAA I
391 L H PIKAA L
392 E H PIKAA E
393 L H PIKAA W
394 A .
395 K .
396 Q .

~~~

465 V .
466 L .
467 E .
```

Next, a python script developed by Alex Ford was used to test how well the backbone conformation of the region grafted into EB22 was represented in a library of protein fragments harvested from the PDB. A second python script developed by Alex Ford used the amino acid sequence of the region grafted into EB22 to pick fragments from the fragment library and test how well the picked fragments matched the modeled backbone conformation of the grafted region. These steps gave an idea of whether the modeled backbone was physically reasonable and how well the designed sequence might favor the modeled conformation.

7.3 Illumina sequence data processing of EB15 and EB22 SSM pools

High-throughput sequencing was carried out on an Illumina MiSeq (Illumina, San Diego, CA, USA). Each run generated roughly 4 million paired-end sequences which were output along with their associated quality scores in compressed .fastq files. The resulting data was analyzed using a modified version of the Enrich package (117). In brief, the reads were first spit up into their respective pools according to their barcodes and then the forward and reverse reads for each

individual molecule of DNA were fused together to form a single continuous read. Next the fused sequences present in each pool were aligned to both the wild-type nucleotide and amino acid sequences and both nucleotide and amino acid mutations relative to wild type were identified. Then the number of instances of each unique amino acid sequence were tallied and their frequencies calculated relative to the number or reads in their pool. Finally, the tallied sequence counts were imported into Microsoft Excel for Mac 2011 (Microsoft Corporation, Redmond, WA, USA), sequences occurring less than 30 times in at least 1 pool were removed, and the enrichment or depletion of each unique sequence relative to its frequency in a baseline pool was calculated. Pools selected for both display and binding were compared to the pool selected only for display, while pools selected only for display were compared to the unselected pool.

Commandlines to run each of the data processing steps and example intputs and outputs are given below:

1. Split the reads into separate pools according to their barcodes:

```
python /path_to_scripts/Index_parser_PE.py --inpath ./ --readi <name
of .fastq file containing forward read> --read1
<forward_read_.fastq_filename> --read2 <reverse_read_.fastq_filename> --index
<barcode_sequence> --iname <barcode_name> --outpath <barcode_name> --
avg_quality 10 --mutations 3
```

An example .fastq file is given below: (“~~~” denotes portions omitted for brevity)

```
@M00777:19:000000000-A7L7K:1:1107:15354:4614 1:N:0:1
GGTCGCGGACCAGGGTCGGCTAGCCATATGATCAATGAGATCAAAAAGAATGCGCAGGAGCGTATGGATGAAACCGT
GGAACAACTGAAAAATGAACTGAGCAAAGTTCGTACCGGTGGTGGCGGCACCGAAGAGCGCCGTCTGGAACTGGCGA
AACAAGTGGTTTTTGCGGCGAATCGCGCCTTAATTCGTGTGCGTACCATCGCCCTGGAGGCGGCGTGGCGTTTACTG
ATGCTGGGCAGCGATAAGGAAGTGAATA
+
CCABCCCCB@BBGGGGGGGGGGHHHHHHHHHHHHHHHHHHHHHHHHHGGHHHGGGGGGGGHGGGGHHHHHHHHHHGG
HGGHHHHHHHHHHHHHHHHGHHHHHHHHHHHGGHHGGCGGGGGGHGGGGGGGGGGGGHHGGGGCGHGHGGGGGGGGG
GGGGGGFFGGGGGGFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFDFFFFFFFFFFFFBDFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFFFFF
@M00777:19:000000000-A7L7K:1:1107:24218:4614 1:N:0:1
GGTCGCGGACCAGGGTCGGCTAGCCATATGATCAATGAGATCAAAAAGAATGCGCAGGAGCGTATGGATGAAACCGT
GGAACAACTGAAAAATGAACTGAGCAAAGTTCGTACCGGTGGTGGCGGCACCGAAGAGCGCCGTCTGGAACTGGCGA
```

```
AACAAGTGGTTTTTGCGGCGAATCGCGCCTTAATTCGTGTGCGTACCATCGCCCTGGAGGCGGCGTGGCGTTTACTG
ATGCTGGGCAGCGATAAGGAAGTGAATA
+
BBAABBDBCCCBGGGGGGGGGGHHHHHHHHHHHGHHHHHHHHHGHHGGHHHHGGGGGGGHGGGGHHHHHHHHHHHGG
HGEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHGGHHGGGDGGGGGHGGGGGGGGGGGGHGGGGDGHGHGGGGGGGGG
GGGGGGFGGGGGGGFFFFFFFFFFFFFFFFFFFFFFF.AFFFFFFFFFFFFFFFFEFDFFFFFCFFFDFFFFFFF
FFFFFFFFFFFFFFEFEFFBFFFFB/BF
@M00777:19:000000000-A7L7K:1:1107:24501:4615 1:N:0:1

~~~
```

2. Fuse the paired end reads into a single read:

```
python /path_to_scripts/quickunforgivingfuser1.3.py --read1
<forward_read_.fastq_filename> --read2 <reverse_read_.fastq_filename> --
paired_mismatch_threshold 200 --read1_start 1 --read2_start 1 --
length_overlap
<size_of_the_overlap_region_between_the_forward_and_reverse_read> --
include_nonoverlap_region 1 --wtseq <non-mutated_gene_nucleotide_sequence> --
mode B --trim_5p <number_of_unneeded_bases_at_5'_end_of_fused_gene> --trim_3p
<number_of_unneeded_bases_at_3'_end_of_fused_gene>
```

An example fused read file is given below: ("~~~" denotes portions omitted for brevity)

```
read1.ID	fused_sequence	total_bases paired_mismatch_count
	paired_match_count	gap_count	paired_unresolvable_count
	mutations_to_wt	read1_avgquality	read1_chastity	read1_Ncount
	read2_avgquality	read2_chastity	read2_Ncount
@M00777:19:000000000-A7L7K:1:1107:15354:4614 1:N:0:1
	GGGTCGGCTAGCCATATGATCAATGAGATCAAAAAGAATGCGCAGGAGCGTATGGATGAAACCGTGGAACA
ACTGAAAAATGAACTGAGCAAAGTTCGTACCGGTGGTGGCGGCACCGAAGAGCGCCGTCTGGAACTGGCGAAACAAG
TGGTTTTTGCGGCGAATCGCGCCTTAATTCGTGTGCGTACCATCGCCCTGGAGGCGGCGTGGCGTTTACTGATGCTG
GGCAGCGATAAGGAAGTGAATAAACGCGATATTAGCCAGGCGCTGGAGGAGATTGAAAAACTGACCAAGGTTGCCGC
GAAGAAAATCAAAGAGGTGCTGGAAGCCAAGATCAAGGAACTGCGTGAAGTTCTCGAGGGTGGCGGA	395
	0	123	272	0	0	37.6486486486	1	0
	36.8918918919	2	0
@M00777:19:000000000-A7L7K:1:1107:24218:4614 1:N:0:1
	GGGTCGGCTAGCCATATGATCAATGAGATCAAAAAGAATGCGCAGGAGCGTATGGATGAAACCGTGGAACA
ACTGAAAAATGAACTGAGCAAAGTTCGTACCGGTGGTGGCGGCACCGAAGAGCGCCGTCTGGAACTGGCGAAACAAG
TGGTTTTTGCGGCGAATCGCGCCTTAATTCGTGTGCGTACCATCGCCCTGGAGGCGGCGTGGCGTTTACTGATGCTG
GGCAGCGATAAGGAAGTGAATAAACGCGATTGTAGCCAGGCGCTGGAGGAGATTGAAAAACTGACCAAGGTTGCCGC
GAAGAAAATCAAAGAGGTGCTGGAAGCCAAGATCAAGGAACTGCGTGAAGTTCTCGAGGGTGGCGGA	395
	0	123	272	0	0	37.3938223938	1	0
	35.8841698842	2	0
@M00777:19:000000000-A7L7K:1:1107:24501:4615 1:N:0:1
	GGGTCGGCTAGCCATATGATCAATGAGATCAAAAAGAATGCGCAGGAGCGTATGGATGAAACCGTGGAACA
ACTGAAAAATGAACTGAGCAAAGTTCGTACCGGTGGTGGCGGCACCGAAGAGCGCCGTCTTGAACTGGCGGAACAAG
GGGTTTTTGCGGCGGATCGCGCCTTAATTCGTGGTCGTACCATTCGTCTGGGGGGCGGCGTGCGTTTAATTGATGCG
GGCAGCCGATAAGAAGTTGATAAAACGCGATATTAGCCAGGCGCTGGAGGAGATTGAAAAACTGACCAAGGTTGCCG
CGAAGAAAATCAAAGAGGTGCTGGAAGCCAAGATCAAGGAACTGCGTGAATGCTCGAGGGTGGCGGA	395
	91	32	272	0	0	36.6486486486	1	0
	34.9691119691	2	0

~~~
```

3. Align fused reads to wild-type sequence and identify mutations:

```
python /path_to_scripts/Fused_read_aligner.py --path ./ --infile
<fused_reads_filename> --referenceAA <wild_type_amino_acid_sequence> --
referenceDNA <wild_type_nucleotide_sequence> --gap_max 1000 --
unresolvable_max 1000 --maxmutrun 1000 --avg_quality 1 --chaste 0 --
Ncount_max 1000 --use_N 1 --mode B > /dev/null
```

An example aligned reads file is given below: ("~~~" denotes portions omitted for brevity)

```
readID      sequence    match_count mutation_count    mutation_location
      mutation_identity max_mutation_run
@M00777:19:000000000-A7L7K:1:1107:15354:4614 1:N:0:1
      GSASHMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAKQVVFAANRALIRVRTIALEAAWR
LLMLGSDKEVNKRDISQALEEIEKLTKVAAKKIKEVLEAKIKELREVLEGGG  123   0     NA    NA
      0
@M00777:19:000000000-A7L7K:1:1107:24218:4614 1:N:0:1
      GSASHMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAKQVVFAANRALIRVRTIALEAAWR
LLMLGSDKEVNKRDCSQALEEIEKLTKVAAKKIKEVLEAKIKELREVLEGGG  122   1     85    C
      0
@M00777:19:000000000-A7L7K:1:1107:24501:4615 1:N:0:1
      GSASHMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAEQGVFAADRALIRGRTIRLGGGVR
LIDAGSR*EVDKTRY*PGAGGD*KTDQGCREENQRGAGSQDQGTA*MLEGGG  74    49
      47,49,54,60,64,66,67,68,69,72,73,74,77,78,81,83,84,85,86,87,88,89,90,91
,92,93,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113
,114,115,116,117
      E,G,D,G,R,G,G,G,V,I,D,A,R,*,D,T,R,Y,*,P,G,A,G,G,D,*,T,D,Q,G,C,R,E,E,N,Q
,R,G,A,G,S,Q,D,Q,G,T,A,*,M    0

~~~
```

4. Tally each unique sequence:

```
python /path_to_scripts/mapCounts.py --path ./ --infile
<aligned_reads_filename>
```

An example tallied reads file is given below: ("~~~" denotes portions omitted for brevity)

```
seqID sequence    match_count mutation_count    mutation_location
      mutation_identity max_mutation_run  sequence_frequency
      sequence_count
NA-NA
      GSASHMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAKQVVFAANRALIRVRTIALEAAWR
LLMLGSDKEVNKRDISQALEEIEKLTKVAAKKIKEVLEAKIKELREVLEGGG  123   0     NA    NA
      0     0.164877234271    196303
95-R
      GSASHMINEIKKNAQERMDETVEQLKNELSKVRTGGGGTEERRLELAKQVVFAANRALIRVRTIALEAAWR
LLMLGSDKEVNKRDISQALEEIEKRTKVAAKKIKEVLEAKIKELREVLEGGG  122   1     95    R
      0     0.0151444522556   18031
22-A
      GSASHMINEIKKNAQERMDETVAQLKNELSKVRTGGGGTEERRLELAKQVVFAANRALIRVRTIALEAAWR
LLMLGSDKEVNKRDISQALEEIEKLTKVAAKKIKEVLEAKIKELREVLEGGG  122   1     22    A
      0     0.0121308482019   14443
```

5. Reformat tallied sequence counts for Excel:

```
#!/bin/bash
### Please note that the stop codons are represented as a "Z" in this script
and in the subsequent output files, since "*" has special meanings in bash
and causes problems if used to search patterns.
num_residues=<number_of_amino_acids_in_sequenced_region>
start_range=`echo 0`
end_range=`echo $num_residues | awk '{print $1-1}'`
filename=<tallied_reads_filename>
sed 's|*|Z|g' ${filename} > ${filename}_2
awk '/NA-NA\t[A-Z].*[A-Z]\t'$num_residues'\t0\t/ {print "num_wt_seqs:\t"$9}'
${filename}_2 > ${filename}_3
echo `seq $start_range $end_range` | tr ' ' '\t' | awk '{print
"Residues\t"$0}' >> ${filename}_3
for i in A C D E F G H I K L M N P Q R S T V W Y Z; do
      echo $i | awk '{print $1"\t"}' >> ${filename}_3
      for j in `seq $start_range $end_range`; do
            count=`awk '/.*-[A-Z]\t[A-Z].*[A-
Z]\t'$end_range'\t1\t'$j'\t'$i'\t/ {print $9}' ${filename}_2`
            value=`awk '/.*-[A-Z]\t[A-Z].*[A-
Z]\t'$end_range'\t1\t'$j'\t'$i'\t/ {print $4}' ${filename}_2`
            if [ "$value" == "1" ]; then
                  echo $count | awk '{print $1"\t"}' >> ${filename}_3
                  else echo NA | awk '{print $1"\t"}' >> ${filename}_3
            fi
      done
      echo end_of_line >> ${filename}_3
done
sed ':a;N;$!ba;s|\([A-Z0-9]\t\)\n|\1|g' ${filename}_3 > ${filename}_4
sed 's|end_of_line||g' ${filename}_4 > ${filename}_ready_for_excel.txt
rm ${filename}_[234]
```

An example reformatted tallied reads file is given below: ("~~~" denotes portions omitted for brevity)

```
num_wt_seqs:      196303
Residues    0     1     2     3     4     5     6     7     8     9     10
      11    12    13    14    15    16    17    18    19    20    21    22
      23    24    25    26    27    28    29    30    31    32    33    34
      35    36    37    38    39    40    41    42    43    44    45    46
      47    48    49    50    51    52    53    54    55    56    57    58
      59    60    61    62    63    64    65    66    67    68    69    70
      71    72    73    74    75    76    77    78    79    80    81    82
      83    84    85    86    87    88    89    90    91    92    93    94
      95    96    97    98    99    100   101   102   103   104   105   106
      107   108   109   110   111   112   113   114   115   116   117   118
      119   120   121   122
A     3     37    NA    NA    NA    NA    166   166   198   143   173   179
      135   NA    93    105   146   155   210   188   498   566   14443 127
      5     176   170   128   157   89    95    141   103   142   85    73
      117   86    90    133   221   138   103   247   76    25    NA    65
      138   145   219   245   NA    NA    84    124   NA    132   141   186
      214   131   178   139   NA    47    73    NA    NA    144   136   221
      220   47    215   158   132   115   107   186   186   196   191   96
```

```
        15      600     144     130     NA      149     106     118     507     186     95      123
        100     84      120     NA      NA      107     220     150     159     115     143     162
        121     NA      83      152     164     119     177     94      90      284     NA      2
        1       NA      5
C       1       NA      NA      14      NA      NA      79      91      72      98      83      85
        72      33      109     124     137     95      95      81      195     189     81      86
        100     131     81      65      66      99      82      95      107     78      88      123
        103     140     69      75      70      144     121     73      48      44      47      84
        90      91      89      125     115     282     98      78      47      185     71      100
        86      307     66      78      46      67      79      41      97      117     112     78
        98      43      105     154     67      65      45      54      66      117     82      145
        10      322     69      84      107     126     72      80      259     65      91      53
        52      62      107     98      55      93      161     93      54      73      80      101
        71      68      49      57      69      89      77      120     37      196     NA      NA
        45      73      NA
D       NA      NA      39      NA      1       NA      38      72      58      53      93      40
        95      34      60      77      50      90      NA      82      166     128     61      54
        54      70      81      91      30      42      59      72      32      56      73      79
        105     138     51      84      72      51      49      48      65      29      54      81
        61      52      75      86      166     183     99      52      72      140     95      65
        66      58      68      87      65      51      89      39      83      77      25      67
        52      19      90      116     60      NA      53      83      40      114     73      18
        NA      206     59      95      61      92      194     119     159     67      61      45
        33      42      80      122     21      62      66      44      48      102     45      87
        58      50      47      38      57      92      57      38      53      109     NA      50
        24      19      NA

~~~
```

6. Calculate Enrichment ratios in Excel:

```
Enrichment ratio for unique sequence X = (count of sequence X in selected
pool/total count of all sequences in selected pool)/(count of sequence X in
unselected pool/total count of all sequences in unselected pool)
```

The following formula was used in excel to filter the sequence counts for those seen more than a threshold number of times and to calculate Log 2 enrichment ratios:

```
=IF(AND(OR(count_of_sequence_X_in_selected_pool>counts_threshold,count_of_seq
uence_X_in_unselected_pool>counts_threshold),((count_of_sequence_X_in_selecte
d_pool*count_of_sequence_X_in_unselected_pool)>0)),(LOG(((count_of_sequence_X
_in_selected_pool/total_count_of_all_sequences_in_selected_pool(count_of_sequ
ence_X_in_unselected_pool/total_count_of_all_sequences_in_unselected_pool)),2
)),ND)
```

Section 8 (Appendix)

General experimental methods

8.1 Yeast Colony PCR

PCR directly from yeast colonies allows more rapid identification of the gene sequences contained therein. The protocol is a modification of a protocol developed by Dr. Eva Strauch. In short, a yeast colony is treated with Zymolase (Zymo Research, Irvine, California, USA) for 15 minutes at 37°C, vortexed hard for 10 seconds, boiled at 97°C for 15 minutes, and vortexed again. Then the gene contained therein is amplified by PCR and the PCR product is treated with Exonµease I (Epicentre, Madison, Wisconsin, USA) and Shrimp Alkaline Phosphatase (Affymetrix, Santa Clara, California, USA) prior to Sanger sequencing. The protocol can easily be run in 96-well format. A detailed protocol is given below:

1. Make up Zymolase solution according to the following recipe:

Component:	For 1 sample:	For 96 samples:
PBS	25 uL	2463 uL
Zymolase (2U/uL)	0.4 uL	37.5 uL
Total	25 uL	2500 uL

2. Prepare a new yeast plate with a numbered grid with a square for each colony you want to pick, if you desire to save samples of each clone.
3. Add 25 uL of the Zymolase solution to the desired number of wells of a PCR strip. (I use a repeater pipette to do this.)
4. Using pipette tips, pick colonies from a yeast plate and spot briefly onto the grid plate before placing each tip into a well of the PCR strip containing the Zymolase solution.
5. Incubate the grid plate at 30°C until colonies have grown to the desired size, usually 1 day.

6. Carefully shake the PCR strip by hand with the tips still in it for 1 minute to displace the yeast cells from the ends of the pipette tips.
7. Carefully remove the pipette tips from the PCR strip, ensuring that no cells remain on the tips and taking care not to cross-contaminate adjacent wells. (I use a multi-channel pipette to do this.)
8. Apply a cap strip to the top of the PCR strip and incubate the PCR strip for 15 minutes at 37°C.
9. Vortex the PCR strip for 10 seconds at full speed and boil the PCR strip for 15 minutes at 97°C
10. While the PCR strip is boiling, make up a PCR master mix according to the following recipe:

Component:	For 1 sample:	For 96 samples:
Sterile water	18.25 uL	2025 uL
10X Taq buffer	2.5 uL	240 uL
5 uM fwd primer	1 uL	---
5 uM rev primer	1 uL	---
100 uM fwd primer	---	5 uL
100 uM rev primer	---	5 uL
10 mM dNTPs	1 uL	100 uL
5 U/uL Taq Pol	0.25 uL	25 uL
Total	24 uL	2400 uL

11. If using the pETCON vector, use the primers below for the PCR reaction.

>colony_outer_forward (COF)
TGACAACTATATGCGAGCAAATCCCCTCAC

>colony_outer_reverse (COR)
GTACGAGCTAAAAGTACAGTGGGAAC

12. Following boiling, vortex the PCR strip again for another 10 seconds and briefly spin the strip to pellet the intact cells.

13. Add 24 uL of the PCR master mix to the desired number of wells of a 2nd PCR strip. (I use the repeater pipette to do this.)
14. Add 1 uL of the boiled cells' supernatant to the corresponding well of the 2nd PCR strip. (I use a multi-channel pipette to do this.)
15. Affix a cap strip to the top of the 2nd PCR strip and place the strip in a thermocycler.
16. Run the following PCR protocol:

Step:	Temperature:	Time:
1	94°C	5 minutes
2	94°C	30 seconds
3	65°C	30 seconds
4	72°C	60 seconds
5	Go to Step 2	9 times
6	94°C	30 seconds
7	60°C	30 seconds
8	72°C	60 seconds
9	Go to Step 6	9 times
10	94°C	30 seconds
11	55°C	30 seconds
12	72°C	60 seconds
13	Go to Step 10	24 times
14	72°C	10 minutes
15	4°C	forever

17. Optional: If desired, pour a 1% agarose gel with SYBR safe gel stain (Life Technologies, Carlsbad, California, USA) and run 5 uL each of a random selection of PCR reactions, each mixed with 1 uL 6X DNA loading dye (New England Biolabs, Ipswich, Massachusetts, USA), on the gel at 100 V to confirm amplification of the gene from the plasmid DNA.
18. Make up the ExoSap master mix according to the following recipe:

Component:	For 1 sample:	For 96 samples:
Exonuclease I (ExoI)	0.5 uL	50 uL
Shrimp Alkaline Phosphatase (SAP)	1.0 uL	100 uL
Total	1.5 uL	150 uL

19. Once the PCR reaction has finished, add 1.5 uL of the ExoSap solution to each well. (I use a multi-channel pipette to do this.)

20. Affix a cap strip to the PCR strip and place in a thermocycler.

21. Run the following thermocycler protocol:

Step:	Temperature:	Time:
1	37°C	15 minutes
2	80°C	15 minutes
3	4°C	forever

22. During the Exosap step, make up the sequencing master mix according to the following recipe:

Component:	For 1 sample:	For 96 samples:
5 uM forward primer	4 uL	20 uL
Sterile water	9 uL	1280 uL
Total	13 uL	1300 uL

23. Use one of the following sequencing primers:

> GS_forward

GGACAATAGCTCGACGATTGAAGGTAGATACCCATA

> cMyc_reverse

CAAGTCCTCTTCAGAAATAAGCTTTTGTTC

24. For each sample, place 13 uL of the sequencing master mix and 2 uL of each Exosap product into a well of a labeled PCR strip.

25. Cap the PCR strip, wrap it in parafilm, and send it for Sanger sequencing.

8.2 E. coli colony PCR

PCR directly from E. coli colonies allows more rapid identification of the gene sequences contained therein. This protocol is a modified version of a protocol adapted by Michelle Scalley-Kim that was originally developed in the Daniel Geraghty lab at the Fred Hutchinson Cancer Research Center. In short, an E. coli colony is diluted into a PCR reaction and boiled. Then the gene contained therein is amplified by PCR and the PCR product is treated with Exonµease I (Epicentre, Madison, Wisconsin, USA) and Shrimp Alkaline Phosphatase (Affymetrix, Santa Clara, California, USA) prior to Sanger sequencing. The protocol can easily be run in 96-well format. A detailed protocol is given below:

1. Prepare a new LB-agar plate, supplemented with the antibiotic of choice, with a numbered grid, with a square for each colony you want to pick, if you desire to save samples of each clone.
2. Add 100 uL of the sterile water to each of the desired number of wells of a PCR strip. (I use a repeater pipette to do this.)
3. Using pipette tips, pick colonies from an E.coli plate and spot briefly onto the grid plate before placing each tip into a well of the PCR strip containing the sterile water.
4. Incubate the grid plate at 37°C until colonies have grown to the desired size, usually 1 day.
5. Carefully shake the PCR strip by hand with the tips still in it for 1 minute to displace the yeast cells from the ends of the pipette tips.
6. Carefully remove the pipette tips from the PCR strip, ensuring that no cells remain on the tips and taking care not to cross-contaminate adjacent wells. (I use a multi-channel pipette to do this.)

7. Make up a PCR master mix according to the following recipe:

Component:	For 1 sample:	For 96 samples:
Sterile water	7.15 uL	715 uL
10X Taq buffer	2.5 uL	250 uL
5 uM T7_fwd primer	1 uL	---
5 uM T7_rev primer	1 uL	---
100 uM T7_fwd primer	---	5 uL
100 uM T7_rev primer	---	5 uL
5 M Betaine	5 uL	500 uL
10 mM dNTPs	1 uL	100 uL
5 U/uL Taq Pol	0.25 uL	25 uL
Total	21 uL	2100 uL

8. If using the pET29b vector, use the primers below for the PCR reaction.

>T7_fwd

TAATACGACTCACTATAGGG

>T7_rev

GCTAGTTATTGCTCAGCGG

9. Add 21 uL of the PCR master mix to the desired number of wells of a 2nd PCR strip. (I use the repeater pipette to do this.)

10. Add 4 uL of cell solution to the corresponding well of the 2nd PCR strip. (I use a multi-channel pipette to do this.)

11. Affix a cap strip to the top of the 2nd PCR strip and place the strip in a thermocycler.

12. Run the following PCR protocol:

Step:	Temperature:	Time:
1	94°C	5 minutes
2	94°C	20 seconds
3	65°C	30 seconds
4	72°C	45 seconds
5	Go to Step 2	9 times
6	94°C	20 seconds
7	55°C	30 seconds
8	72°C	45 seconds
9	Go to Step 6	24 times
10	72°C	5 minutes
11	4°C	forever

13. Optional: If desired, pour a 1% agarose gel with SYBR safe gel stain (Life Technologies, Carlsbad, California, USA) and run 5 uL each of a random selection of PCR reactions, each mixed with 1 uL 6X DNA loading dye (New England Biolabs, Ipswich, Massachusetts, USA), on the gel at 100 V to confirm amplification of the gene from the plasmid DNA.
14. Make up the ExoSap master mix according to the following recipe:

Component:	For 1 sample:	For 96 samples:
Exonuclease I (ExoI)	0.5 uL	50 uL
Shrimp Alkaline Phosphatase (SAP)	1.0 uL	100 uL
Total	1.5 uL	150 uL

15. Once the PCR reaction has finished, add 1.5 uL of the ExoSap solution to each well. (I use a multi-channel pipette to do this.)
16. Affix a cap strip to the PCR strip and place in a thermocycler.
17. Run the following thermocycler protocol:

Step:	Temperature:	Time:
1	37°C	15 minutes
2	80°C	15 minutes
3	4°C	forever

18. During the Exosap step, make up the sequencing master mix according to the following recipe:

Component:	For 1 sample:	For 96 samples:
5 uM forward primer	4 uL	20 uL
Sterile water	9 uL	1280 uL
Total	13 uL	1300 uL

19. Use either the T7_fwd or the T7_rev primer for sequencing.
20. For each sample, place 13 uL of the sequencing master mix and 2 uL of each Exosap product into a well of a labeled PCR strip.
21. Cap the PCR strip, wrap it in parafilm, and send it for Sanger sequencing.

8.3 Chemically competent yeast transformation

Genes for the p53 trans-activation helix and MB1-12 were synthesized and cloned into pETCON by Genescript (Piscataway, New Jersey, USA). Genes for MB13-26 were synthesized and cloned into pETCON by IDT (San Jose, California, USA). Genes for the VH9 control and the MB17 supercharge variants MB17.23-24 and MB17.43-44 were synthesized by IDT as linear double-stranded DNA fragments (gBlocks). Genes for the MB17 supercharge variants MB17.25-28 and MB17.45-48 were synthesized by making point mutations to MB17.23-24 and MB17.43-44 using the Overlap PCR mutagenesis and assembly protocol. The remaining Mdmx binding proteins were synthesized by Gen9 (Cambridge, Mass, USA) as linear double-stranded DNA fragments. The cloned genes were transformed directly into EBY100 yeast while the linear gene fragments were transformed along with linearlized pETCON. Linearized pETCON vector was prepared by digestion with NdeI, NheI, XhoI, and BamH1 and purified by gel extraction from a 1% agarose gel.

The gene for the N-terminal EED-binding helix of Ezh2 was synthesized and cloned into pETCON by Genescript. Genes for EB15-22 and the 1le2 and 3lf9 controls were synthesized by IDT as linear double-stranded DNA fragments (gBlocks). The cloned genes were transformed directly into EBY100 yeast (125) while the linear gene fragments were transformed along with linearlized pETCON.

A detailed protocol is given below:

1. Preparing chemically competent yeast:

1. Innoculate 5 mL YPD media with a single EBY100 yeast colony and grow overnight at 30°C with shaking at 250 rpm.
2. Measure the absorbance at 600 nm of 500 uL of a 1:10 dilution of the overnight culture using 500 uL plain YPD as a blank.
3. Dilute the culture up to 50 mL YPD at an absorbance of 0.3 AU.
4. Grow at 30°C until the absorbance reaches 1.2, about 4 hours.
5. Harvest the cells by spinning at 4000 g for 5 minutes.
6. Resuspend the cells in 25 mL sterile water and pellet the cells again as above.
7. Resuspend the cells in 1 mL 100 mM LiOAc.
8. Pellet the cells at 3000 g for 5 minutes.
9. Resuspend the cells in 350 uL 100 mM LiOAc and then add 150 uL 50 % glycerol.
10. Make 70 uL aliquots of cells in eppendorf tubes and store at -80°C until needed. Do not snap-freeze the cells.

2. Transforming chemically competent yeast:

1. Boil 10 uL salmon sperm DNA at 97°C for 10 minutes.
2. Thaw 1 tube (70 uL) chemically competent yeast cells at 26°C.

3. Combine the following reagents, in order:

Component:	Volume:
Thawed cells	70 uL
50% PEG	240 uL
1 M LiOAc	36 uL
Sterile water	14 uL
Total	360 uL

4. Vortex the mixture hard to mix.

5. Add 50 uL of the mixture to 0.5 - 5 uL DNA to be transformed.

6. Vortex hard to mix.

7. Incubate at 30°C for 30 minutes.

8. Heat shock the cells by incubating at 42°C for 20 minutes.

9. Pellet the cells by spinning at 20000 g for 30 seconds.

10. Resuspend the cells in 100 uL sterile water and let stand for 5 minutes

11. Plate the cells onto C-Trp-Ura plates in 3 volumes: 1 uL, 10 uL, and 100 uL.

12. Incubate the plates for 2-3 days at 30°C.

13. Alternatively, resuspend the cells in 2 mL CD-UT media with 8 uL antibiotic cocktail (12.5 mg/mL Kanamycin sulfate, 12.5 mg/mL Streptomycin sulfate, and 50 mg/mL Carbenicillin in sterile water).

14. Incubate the culture for 2 days at 30°C with shaking at 250 rpm.

15. After 2 days, add 1 mL of saturated yeast culture and 600 uL 50% glycerol to an eppendorf tube, mix well, and store at -80°C

Section 9 (Appendix)

Experimental methods specific to the design of proteins to inhibit the interaction of Mdmx-p53

9.1 Biotinylation and size-exclusion chromatography of Mdmx and Mdm2

Soluble Mdmx and Mdm2 N-terminal domains (residues 25-111) were expressed, purfied, and biotinylated at St. Jude Children's Research Hospital and further purified at the University of Washington. Some batches of soluble Mdm2 were expressed and purified at St. Jude's but biotinylated and further purified at the University of Washington.

Soluble Mdmx and Mdm2 were expressed in E. coli, purified over Nickel-NTA resin and desalted into 20 mM Hepes, 250 mM Potassium Glutamate, pH 7.5, 1 mM TCEP. The proteins were concentrated to about 0.8 mg/mL (60 uM) and enzymatically biotinylated using the BirA biotin ligase kit (Avidity LLC, Aurora, Colorado, USA). Mdmx and Mdm2 were separated from the BirA enzyme by flowing over Nickel-NTA resin (Qiagen, Venlo, Limburg, Netherlands), desalted into 25 mM Tris, pH 8.0, 200 mM NaCl, 2 mM BME, concentrated to 1.2 - 1.8 mg/mL (89 -134 uM), and snap frozen in liquid nitrogen for shipment. Upon reciept, the proteins were thawed on ice, centrifuged to remove precipitates and further purified by size exclusion chromatography over a Superdex75 column into 25 mM Tris, pH 8.0, 200 mM NaCl, 1 mM TCEP. The proteins were concentrated to 0.1 - 1.3 mg/mL (10 - 25 uM), aliquoted, and snap frozen in liquid nitrogen for storage. Thawed aliquots of Mdmx and Mdm2 were centrifuged to remove precipitates prior to incubation with yeast cells. A detailed protocol for those batches of Mdm2 biotinylated at the University of Washington is given below. All steps except the biotinylation reaction were carried out at 4°C.

1. Thaw 2500 uL Mdm2 on ice.
2. Equilibrate a PD-10 desalt column with HBKG (20 mM hepes, 250 mM potassium glutamate, 1 mM TCEP, pH 7.5).
3. Spin Mdm2 for 1 minute at 20000 g and remove supernatant to a new tube.
4. Buffer exchange the Mdm2 into HBKG using the PD-10 desalt column, taking 0.5 mL fractions of the flow-through, elution, and post-elution steps.
5. Spec the fractions at 280 nm on the Nanodrop spectophopmeter (Thermo Scientific, Waltham, Massachusetts, USA).
6. Combine fractions with significant A280 signal and concentrate to 20 uM.
7. Set up the biotinylation reaction as follows, mix, and incubate at 26°C for 4 hours.

Component:	Volume:
10X Biomix A	100 uL
10X Biomix B	100 uL
10X Biotin	100 uL
57 uM Mdm2	700 uL
BirA	16 uL
Total	1016 uL

8. Attach the Superdex 75 size exclusion column to the FPLC and run 35 mL ddH20 at 0.9 mL/minuteto remove the 20% EtOH.
9. Equilibrate the Superdex 75 column with 35 mL non-imidazole buffer (20 mM Hepes, 150 mM NaCl, 5% v/v glycerol, 1 mM TCEP, pH 7.5) at 0.9 mL/minute.
10. Prepare a 1 mL wide bore Ni column and equilibrate with 5 mL low-imidazole buffer (20 mM Hepes, 150 mM NaCl, 5% v/v glycerol, 10 mM imidazole, 1mM TCEP, pH 7.5) just before loading sample.
11. After the 4 hours are completed, dilute the biotinylation reaction up to 1 mL with low-imidazole buffer and load over the Ni column 10 times.

12. Wash the column with 10 mL low-imidazole buffer.
13. Elute the column with 5 mL high-imidazole buffer (20 mM Hepes, 150 mM NaCl, 5% v/v glycerol, 500 mM imidazole, 1 mM TCEP, pH 7.5) in 5 x 1 mL steps with 1-minute pauses in between steps.
14. Concentrate the eluate down to 500 uL.
15. Do size exclusion chromatography on the eluate and collect 0.5 mL fractions.
16. Spec those fractions that show significant 280 nM signal on the FPLC using the Nanodrop spec at 280 nm.
17. Combine those fractions that correspond to the Mdm2 peak and concentrate to > 10 uM.
18. Spec the concentrated Mdm2 at 280 nm using the Nanodrop spec and snap-freeze in 10 uL aliquots.

<u>9.2 Small volume yeast surface titration of Mdmx binders</u>

Due to the limited quantities of soluble biotinylated Mdmx and Mdm2 available, the yeast surface titration method of Chao, et al. was modified to work with smaller labeling volumes (125). The detailed protocol is given below:

1. For each clone to be titrated, inoculate a single yeast colony, a ~30 uL stab of frozen cell stock, or ~10 uL of saturated yeast culture into 1 mL CD-UT media with 4 uL antibiotic cocktail in a sterile 14 mL polypropylene culture tube.
2. Grow the cells at 30 °C for 16 hours with shaking at 250 rpm.
3. After 16 hours, the cell density will be about 6*10^7 cells/mL. Add 200 uL of the 1 mL culture (about 1.2*10^7 cells) to a sterile eppendorf tube.
4. Spin down the cells in the eppendorf tube at 20000 g for 30 seconds.

5. Pour the media out of the tube. Fear not, the cells will stay adhered to the bottom of the eppendorf tube.
6. Prepare a new sterile 14 mL culture tube with 1 mL SGCAA + 4 uL antibiotic cocktail.
7. Resuspend the cells in the eppendorf tube using the SGCAA solution from the new 14 mL culture tube and add the cell solution back to the culture tube.
8. Induce the cells at 30°C for 16 hours (or 22°C for 20 hours for difficult-to-express proteins) with shaking at 250 rpm. Alternatively, to speed up the protocol, the cells in step 1 can be immediately induced in 1 mL SGCAA instead of first growing overnight in CD-UT. If inducing at 30°C this results in a comparable amount of displayed protein/cell, but can result in fewer total cells if inducing directly from a stab of cell stock.
9. The following day, thaw out an aliquot of the soluble biotinylated target protein on ice and centrifuge it at 20000 g for 2 minutes to remove aggregates.
10. Use the Nanodrop spectophopmeter (Thermo Scientific, Waltham, Massachusetts, USA). to determine the protein's concentration if needed. The target protein should be kept on ice at all times.
11. Freshly sterile filter 50 mL PBSF (PBS + 0.1% Bovine Serum Albumin (BSA)) into a 50 mL polypropylene tube and keep on ice. PBSF left over at the end of each day should be re-filtered each new day to remove adventitious cells that enter the solution the previous day and expand overnight, even if stored at 4°C.
12. Spin down the induced cells at 20000 g for 30 seconds. The cell density is about 6*10^7 cells/mL.
13. Pour the media out of the tube.
14. Re-suspend the cells in 1 mL ice cold PBSF.

15. Spin down the cells again at 20000 g for 30 seconds.
16. Pour the media out of the tube.
17. Resuspend the cells in 300 uL ice cold PBSF. The cell concentration is about 2*10^8 cells/mL. These cells with displayed protein should be kept on ice and can be stored/used for up to 7 days, depending on the stability of the displayed protein.
18. Preblock tubes and wellplate wells that will contain protein at concentrations below 10 nM by filling them with PBSF and incubating at 26°C for 20 minutes.
19. Once the 20-minute pre-blocking period ends, remove all the pre-blocking solution and save for later use (such as for washes and resuspending samples prior to reading on the flow cytometer). Centrifuge the tubes to collect the residual pre-blocking solution and remove it as well.
20. For each target protein concentration to be titrated, make up a primary labeling master mix with the desired concentration of soluble biotinylated target protein. Include 10% BSA at a final concentration of 1% in master mixes having target protein concentrations above 30 nM. The increased concentration of BSA is intended to prevent nonspecific binding at higher protein concentrations. Use PBSF to make up any remaining volume in the master mix. An example titration setup is given in the following table (all volumes are given in uL):

Component	0	3 uM	1 uM	300 nM	100 nM	30 nM	10 nM	3 nM	1 nM	300 pM	100 pM	30 pM	Total needed
25 uM Mdmx		7.92	2.64										10.6
625 nM Mdmx (1:40 dilution)				31.7	10.6	3.80	3.59	3.48	3.51	3.51	3.51	3.51	67.1
10% BSA	3.96	3.96	3.96	3.96	3.96	5.28	19.8						44.9
PBSF	35.6	27.7	33.0	4.0	25.1	43.7	175	696	2161	7283	21882	73098	105464
Total	40	40	40	40	40	53	198	700	2165	7286	21886	73102	
Number of samples	12	12	12	12	12	12	12	12	12	12	12	12	
Add to each reaction	3	3	3	3	3	4	15	53	164	552	1658	5538	
Individual Reaction Volume	5	5	5	5	5	6	17	55	166	554	1660	5540	
Volume of cells in reaction	2	2	2	2	2	2	2	2	2	2	2	2	
Labeling time (minutes)	NA	0.2	0.5	1.7	5	17	50						
Labeling time (hours)								3	8	20	32	32	

The minimum labeling time used here is 3*Tau, which is the time required by the incubation reaction to reach 95% of equilibrium. 3*Tau is calculated according to the following formula:

3 Tau = 3 * (K_{on} (in Molar^-1, seconds^-1) * [target protein] (in Molar)) + K_{off})^-1

If the on-rate (K_{on}) and off-rate (K_{off}) are not known for the interaction to be titrated, an on-rate of 1*10^5/Molar/second may be used and the off-rate can be estimated according to the following formula:

K_{off} = $K_{dissociation}$ (in Molar) * K_{on} (in Molar^-1, seconds^-1)

The labeling time must be long enough to allow equilibrium to be reached, but no longer so as to avoid the TCEP-mediated reduction of the disulfide bonds adhering the displayed protein to the yeast surface.

21. For each clone to be titrated, place 2 uL of the induced cells into the wells of a polypropylene V-bottom 96-well plate, 1.7 mL eppendorf tube, or 15 mL polypropylene tube as indicated in the table above. This corresponds to 4*10^5 cells per labeling reaction.

22. For each concentration of soluble target protein and sample of cells, add the volume of the master mix indicated in the table above to each drop of cells and mix well.

23. Cover the plate with a clear plastic adhesive seal and/or cap the tubes and rotate the plate and/or tubes at 26 °C for the length of time indicated in the table above. For cases where different concentrations must be labeled for different lengths of time, simply assemble the labeling reaction corresponding to the longest-time point first and then sequentially assemble labeling reactions corresponding to shorter-time points such that all of the labeling reactions complete their respective incubation times at the same time. The higher-concentration points cannot be labeled for longer than the indicated length of time because the soluble biotinylated Mdmx and Mdm2 must be stored in 1 mM TCEP. At 26 °C, concentrations of TCEP above 10 nM or so will rapidly reduce the disulfide bonds tethering the Aga2 yeast display apparatus to the surface of the yeast cells, releasing the displayed proteins to the surrounding labeling solution.
24. Once the labeling/incubation has completed, centrifuge samples with volumes greater than 5 uL and remove the labeling solution. For samples labeled in the 96-well plate, centrifuge at 4000 g for 5 minutes in a plate centrifuge and remove the solution by removing the plastic cover and shaking the plate upside down over a sink hard 3 - 4 times. The cells will remain adhered to the bottom of the plate no matter how hard you shake it and harder shaking removes the wash more completely. For samples in eppendorf tubes, centrifuge for 30 seconds at 20000 g and pipette out the solution. For samples in 15 mL polypropylene tubes, centrifuge at 4000 g for 5 minutes in a swing-bucket centrifuge and remove the solution using a serological pipette.
25. Add 200 uL of PBSF to re-suspend and wash the cells in each of the eppendorf and 15 mL tubes transfer them to corresponding empty wells of the 96-well plate.

26. Add 200 uL of PBSF to each well of the 96-well plate that contains cells incubated in the plate and replace the plastic seal.
27. Pellet the cells in the plate by centrifugation at 4000 g for 5 minutes using the plate rotor in the tabletop centrifuge.
28. Remove the plastic seal and shake the plate upside down over a sink hard 3 - 4 times to remove the wash supernatant.
29. Make up the secondary labeling solution according to the following recipe:

Component:	Volume for 1 sample:	For 96 samples:
1 mg/mL Steptavidin-R-Phycoerythrin conjugate (SAPE)	0.2 uL	20 uL
1 mg/mL FITC-conjugated Chicken-anti-Myc Antibody (FITC)	0.2 uL	20 uL
PBSF	4.6 uL	460 uL
Total	5 uL	500 uL

30. Add 5 uL to the cells in each well of the 96-well plate and incubate on ice for 30 minutes.
31. Once the 30-minute secondary incubation has completed, add 200 uL PBSF to each well of the plate, replace the plastic seal, and centrifuge for 5 minutes at 4000 g.
32. Remove the plastic seal and shake the plate upside down over a sink hard 3 - 4 times to remove the wash supernatant.
33. Leave the pelleted, labeled cells on ice until right before reading each individual sample by flow cytometry.
34. Immediately before reading on the flow cytometer, Re-suspend a column of 8 cell pellets with 40 uL PBSF/well using a multichannel pipettor and transfer to a clean PCR tube strip. This allows you to quickly read 8 samples one after the other, greatly speeding up throughput during reading on the flow cytometer.

35. Be sure to also read an unlabeled sample and a sample labeled with only FITC-conjugated Chicken-anti-Myc Antibody to set display gates and compensation values on the flow cytometer.

9.3 Kunkel mutagenesis

A library of single site mutants (SSM) were generated using a modified version of the method of Kunkel, et al (126). In brief, single-stranded, uracilated DNA of the plasmid of interest is amplified using mutagenic oligos and a thermophilic polymerase. Next the mutagenic plasmid is electoporated into E. coli cells which degrade the wild-type template. The plasmid is then extracted from the E. coli, amplified by PCR, and electroporated into yeast. A detailed protocol is given below. The protocol is a modified version of a protocol received from Jasmine Gallaher at the University of Washington.

1. Transforming the plasmid into CJ236 cells:

1. In the late PM, transform the double-stranded plasmid into chemically competent CJ236 cells (New England Biolabs, Ipswich, Massachusetts, USA). In order to make ssDNA, the plasmid to be transformed must carry the f1 origin. CJ236 cells are used here because they lack uracil deglycosylase and hence produce DNA containing occasional uracil bases in place of thymine bases (dUssDNA). This allows to template DNA to be destroyed by the uracil deglycosylase present in the bacterial cell line receiving the template-mutant DNA hydrid.
2. Thaw 100 uL of CJ236 cells and add them to 2 uL of plasmid DNA in a 15 mL culture tube.
3. Incubate the culture tube on ice for 30 minutes.

4. Heat shock the cells at 42°C for 45 seconds.
5. Recover the cells on ice for 3 minutes.
6. Add 250 uL of LB media to the cells and incubate at 37°C for 1 hour with shaking at 250 RPM.
7. After 1 hour, plate 40 uL of the transformation mixture on an agarose plate supplemented with the appropriate antibiotic (Carbenicillin, in the case of pETCON).
8. Incubate the plate overnight at 37°C.
9. The next day, pick 6 colonies from the plate and add them to 3 mL LB media supplemented with the appropriate antibiotic in a 15 mL culture tube.
10. Incubate the culture tube for 6 hours at 37°C with shaking at 250 RPM.
11. After 6 hours, add 3 uL of M13K07 helper phage to the culture.
12. Continue shaking the culture at 37°C for 1hr.
13. Expand the culture by diluting 1mL into 50 mL LB media supplemented with the appropriate antibiotic in a baffled 250 mL Erlenmeyer flask. (Even if your plasmid does not confer kanamycin resistance, add it at this stage because the helper phage genome bears Kanamycin resistance. Kanamycin selects for bacterial cells that have been infected with M13 phage).
14. Grow the culture overnight at 30°C with shaking at 250 RPM. Growing at 30°C instead of 37°C appears to yield more dUssDNA.

2. Harvesting the M13 phage from the culture media:

1. Spin down the bacteria from the overnight culture in a 50 mL polypropylene tube at 10000 g for 20 minutes at 4°C in a fixed-angle rotor.
2. Transfer the supernatant to a clean 50 mL polypropylene tube and add 10 mL of

autoclaved 20% PEG, 2.5M NaCl and mix thoroughly. Phages are soluble in media and most buffers. They are harvested by repeated cycles of PEG/NaCl precipitation and centrifugation.

3. Incubate the 50 mL tubes on ice for 45 minutes. The precipitating phage may make the media slightly cloudy, but if not, don't be alarmed.
4. Spin down the phage at 10000 g for 20 minutes in at 4°C. The phage pellet should be visible, with a white streak down the outside edge.
5. Immediately after centrifuging, carefully pour off the media supernatant. Let the tube stand upright for a few minutes and carefully pipette off the remaining liquid.
6. Re-suspend the phage pellet in 2 mL PBS by vortexing or pipetting up and down.
7. Transfer the 2 mL of re-suspended phage into two 1.5 mL eppendorf tubes with 1 mL in each tube. Centrifuge the tubes at 20000 g for 5 minutes. This pellets the residual bacteria while the phage remains in solution.
8. Transfer the 2 mL of supernatant to two fresh eppendorf tubes, each with 300 uL PEG/NaCl solution.
9. Vortex the tubes and incubate them at room temperature for 10 minutes.
10. Spin down the tubes at 20000 g for 2 minutes. The white pellet of phage should be visible.
11. Pipette off all of the supernatant. Do a second quick spin to collect residual liquid and pipette it off.
12. Re-suspend the pellet in 1 mL PBS. Use that 1 mL to re-suspend the other pellet.
13. Spin down at max speed for 5min.
14. Transfer the supernatant containing the phage to a new eppendorf tube.

3. Harvesting dUssDNA from Phage:

This step uses the Qiagen Spin M13 kit (Qiagen, Venlo, Limburg, Netherlands) but a modified version of the vendor protocol.

1. Add 1 mL phage solution from the previous step to 11 uL Qiagen buffer MP in a 1.5 mL eppendorf tube and vortex the tube. Incubate the tube at room temp for 10 minutes to precipitate the phage.
2. Pipette 500 uL of precipitated phage to the top of a blue Qiaprep column.
3. Spin the column at 11500 g for 30 seconds and discard the flow through. Do not spin the column faster than 11500 g.
4. Add the remaining 500 uL to the column and spin again. Discard the flow through.
5. Add 700 uL Qiagen buffer MLB to the column. Spin the column for 20 seconds at 11500 g and discard the flow through.
6. Add another 700 uL of Qiagen buffer MLB to the column and let stand for 1 minute. Spin the column for 20 seconds at 11500 g and discard the flow through.
7. Add 700 uL of Qiagen buffer PE to the column. Spin the column for 20 seconds at 11500 g and discard the flow through.
8. Add a second 700 uL of Qiagen buffer PE to the column. Spin the column for 20 seconds at 11500 g and discard the flow through.
9. Spin for the column for 2 minutes at 20000 g to dry out the column.
10. Transfer blue column to a labeled eppendorf tube.
11. Add 150 uL of Qiagen buffer EB to the column and let stand for 1 minute. Spin the column at 20000 g for 1 minute to elute the dUssDNA.

12. Measure the concentration of dUssDNA using a Nanodrop spectophopmeter (Thermo Scientific, Waltham, Massachusetts, USA).
13. Run a sample of the dUssDNA on a 1% agarose gel with to check quality. A good preparation runs as a single band approximately 1/3 the molecular weight of the corresponding dsDNA plasmid. Higher molecule weight species are usually M13K07 helper phage DNA contamination.

4. Treat the oligos with Polynucleotide Kinase:

For an SSM library, it's best to mutate each amino acid position in its own reaction and then combine reactions immediately prior to the PCR cleanup. I saw about a 15-fold difference between the least and most represented variant when I incorporated an NNK codon at 80 positions in the MB17 gene using 45-mer oligos. Wild type contamination in the resulting library is due to annealing of truncated oligo products (the 3' end) that contaminate your oligo prep if it isn't page-purified. Thus, longer oligos (60-mers) result in less bias within a given position but more wild type variants, while shorter oligos (45-mers) give more bias within a given position but fewer wild type variants.

1. Calculate what volume of DNA elution buffer would be required to give each lyophilized oligo a final concentration of 100 uM according to the following formula:

   uL Qiagen buffer EB to add = nmol of lyophyized oligo x 10 uL/nmol
2. Bring up each lyophilized oligo in sufficient Qiagen buffer EB to give a final concentration of 100 uM.
3. Make up the kinase reaction master mix according to the following recipe:

Component:	For 1 reaction:	For 80 reactions:
10X kinase buffer	1 uL	88 uL
10 uM ATP	1 uL	88 uL
T4 polynucleotide kinase (New England Biolabs)	1 uL	88 uL
100 mM oligo	7 uL	---
Total	10 uL	264 uL

4. Add 7 uL of each oligo and 3 uL of the kinase master mix to a corresponding well of a PCR strip and mix.
5. Cap the PCR strips and incubate them for 1 hour in a 37°C water bath or incubator

5. Anneal the oligos to the template:

1. Dilute the kinase reaction mix 1:20 with autoclaved ddH2O (10 uL kinase rxn + 190 uL autoclaved ddH20) to give an oligo concentration of 3.5 uM. If using multiple oligos in a single Kunkel reaction, combine them here (The sum concentration of all oligos annealing to a single given position should be 3.5 uM at this point.
2. Make up the annealing reaction master mix in according to the following recipe. The final concentration of all oligos at a single given position will be 350 nM. If mutagenizing N positions in the reaction, then the total concentration of oligos in the reaction will be N*350 nM. This recipe results in a 38:1 oligo:template ratio. The template used here was at 9.2 nM, giving an oligo:template ratio of 38:1, which quantitatively converted dUssDNA to dsDNA, giving a 33% wild type rate.

Component:	For 1 reaction:	For 80 reactions:
10X ligase buffer	1 uL	88 uL
Diluted kinase reaction mix	1 uL (6.9 ng/reaction at 350 nM)	---
37 ng/uL dUssDNA template	5.4 uL (200 ng/reaction at 9.2 nM)	475 uL
Sterile water	2.6 uL	229 uL
Total	10 uL	792 uL

3. Add 1 uL of each diluted kinase reaction and 9 uL of the annealing master mix to a corresponding well of a PCR strip and mix.
4. Cap the PCR strip, place it in a thermocycler, and run the following protocol:

Step:	Temperature:	Time:
1	90°C	2 minutes
2	50°C	3 minutes
3	20°C	5 minutes
4	20°C	forever

6. Polymerize the oligo-template DNA hybrid:

1. Make up the polymerase reaction master mix according to the following recipe in the same PCR tube used to anneal the oligos to the template:

Component:	For 1 reaction:	For 80 reactions:
10X Ligase Buffer	1.5 uL	132 uL
25 mM dNTPs	1 uL	88 uL
10 mM ATP	1 uL	88 uL
T7 DNA Polymerase (New England Biolabs)	1 uL	88 uL
T4 DNA Ligase (New England Biolabs)	1 uL	88 uL
Annealed template from previous step	10 uL	---
Total	15.5 uL	484 uL

2. Add the other components to the annealed template already in the tube and mix.
3. Let the tube sit at room temperature for 1.5 hours. This will result in 100% conversion of ssDNA to dsDNA with an approximate 33% WT rate if you use 45-mer oligos.
4. Combine the annealing reaction(s) into 5 pools of 16 positions each and remove the enzymes and unincorporated oligos using the QIAquick PCR purification kit (Qiagen, Venlo, Limburg, Netherlands). This step dramatically increases transformation efficiency in the subsequent step. If you must pause before doing the PCR cleanup, store your annealing reactions at -80°C as the polymerase begins to degrade the PCR products once it runs out of DNA to polymerize.

7. Transform the mutagenic plasmid into E. coli:

1. Electroporate 100 ng of the DNA from each PCR cleaned-up pool of polymerization reaction products each into 45 uL of ElectroMAX DH10B cells (Life Technologies, Carlsbad, California, USA) according to the vendor protocol.
2. Plate 20 uL of each of the transformation reaactions onto and LB-agar plate supplemented with 100 ug/mL Carbenicillin and place the rest of the cells from each reaction into a corresponding 3 mL of LB media supplemented with 100 ug/mL Carbenicillin in a 15 mL polypropylene culture tube.
3. Incubate the plate and liquid cultures at 37°C for 16 hours.
4. After 16 hours, count the colonies on the plates to ensure that enough transformants were obtained to cover the whole library at least 100-fold.
5. Use the E. coli colony PCR protocol to Sanger sequence 48 colonies from the agar plates (9-10 per plate).
6. Analyze the sequences to ensure minimal mutational bias and wild type contamination in the single-site mutant library.

8. Extract and amplify the mutagenic plasmids:

1. Extract the plasmid DNA from the 3 mL overnight cultures of ElectroMAX cells using the QIAprep Spin Miniprep Kit (Qiagen, Venlo, Limburg, Netherlands) and following the vendor protocol.
2. Quantify the miniprepped DNA using a Nanodrop spectophopmeter (Thermo Scientific, Waltham, Massachusetts, USA).
3. Make up the amplification PCR master mix according to the following recipe:

Component:	For 1 reaction:	For 5 reactions:
5X Phusion HF buffer	20 uL	102 uL
20 mM dNTPs	1 uL	5.1 uL
20 mM forward amp primer	2.5 uL	12.75 uL
20 mM reverse amp primer	2.5 uL	12.75 uL
Template DNA from miniprep	1 uL	---
2U/uL Phusion polymerase	1 uL	5.1 uL
Sterile water	72 uL	367.2 uL
Total	100 uL	505 uL

4. Use the following primers for the amplification reaction if using the pETCON vector:

> GS_forward
GGACAATAGCTCGACGATTGAAGGTAGATACCCATA

> cMyc_reverse
CAAGTCCTCTTCAGAAATAAGCTTTTGTTC

5. Add 1 uL of template from each of the miniprepped library pools to 99 uL of the amplification PCR master mix in a well of a PCR strip and mix.

6. Cap the PCR strip, place it in a thermocycler, and run the following protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	10 seconds
3	55°C	30 seconds
4	72°C	30 seconds
5	Go to step 2	24 times
6	72°C	10 minutes
7	4°C	forever

7. Remove the enzymes and unincorporated oligos from the completed amplification reactions using the QIAquick PCR purification kit.

8. Quantify the cleaned up DNA using a Nanodrop spectophopmeter according to the manufacturer protocol.

9. Desalt the eluted DNA fragments by aliquoting the DNA solution onto nitrocellulose membranes suspended above a petri dish filled with sterile water and letting stand for 20 minutes at room temperature.
10. Electroporate the DNA fragments with linearlized pETCON vector into EBY100 yeast using the protocol of Benatuil, et al.

9.4 Labeling MB17 SSM library for FACS selection

The MB17 single site mutant and combinatorial libraries were FACS (127) sorted under 5 different selection conditons: 1. maximal affinity for Mdmx, 2. maximal affinity for Mdm2, 3. maximal specificity for Mdmx over Mdm2, 4. maximal specificity for Mdm2 over Mdm2, and 5. maximal but equal affinity for Mdmx and Mdm2 (pan-secificity). Variants with maximal affinity were isolated by labeling the library with the lowest possible concentration of soluble biotinylated Mdmx or Mdm2. Variants with maximal specificity for Mdmx over Mdm2 were isolated by labeling the library with a low concentration of Mdmx pre-conjugated to one fluorophore and a high concentration of Mdm2 pre-conjugated to a second, orthologous fluorophore and gating those variants that had the maximal Mdmx signal and minimal Mdm2 signal. Variants with maximal specificity for Mdm2 over Mdmx were likewise isolated using a low concentration of Mdm2 and a high concentration of Mdmx. Variants with maximal but equal affinity for Mdmx and Mdm2 were isolated by labeling the library with equally low concentrations of Mdmx and Mdm2. p53, which binds both Mdmx and Mdm2 with roughly equal affinity, was also labeled at concentrations equal to it's published Kds for Mdmx and Mdm2. The labeled library cells were gated in the same region of the Mdmx-binding vs. Mdm2-

binding FACS plot where the labeled p53 cells appeared, thus selecting for cells having equal affinities for Mdmx and Mdm2.

Concurrently, the libraries were also sorted for a number of other criteria, including minimal off-rate and minimal nonspecific binding. Variants with minimal off-rate were isolated by labeling the libraries at saturating levels of Mdmx or Mdm2, washing away the unbound Mdmx or Mdm2, and incubating respectively with a high concentration of soluble MB17.41 or MB17.21 competitor for lengthy periods at 37°C. The MB17.41 and MB17.21 competitors are intended to respectively bind any molecules of Mdmx or Mdm2 that dissociate from the yeast surface and prevent their re-binding. Clones with minimal nonspecific binding were isolated by alternating between 3 different secondary fluorophores used for each round of sorting: streptavidin-R-phycoerythrin conjugate (SAPE), streptavidin-allophycocyanin-Alexa-fluor-750-conjugate (SA750), and neutravidin-R-phycoerythrin conjugate (NAPE)(Life Technologies, Carlsbad, California, USA). A detailed protocol for inducing and labeling the MB17 libraries is given below:

1. Induce a sufficient number of yeast cells to achieve 10-fold coverage of the library diversity at each round of sorting by incubating the cells in either 1 mL CG-UT (128) or SGCAA (125) supplemented with 50 mg/mL kanamycin, 50 mg/mL streptomycin, and 100 mg/mL Carbenicillin overnight at 30°C. Note that while CG-UT media is useful for selecting against contaminating yeast species and against display of overly sticky proteins, some genes cannot be displayed in CG-UT that otherwise could be displayed in SGCAA.
2. Pellet the induced cells by centrifugation for 30 seconds at 20000 g and discard the media.

3. Wash the induced cells by adding 900 uL PBS supplemented with 1 mg/mL BSA (PBSF), centrifuging again, and discarding the wash solution. Keep the induced, washed cells on ice.
4. Measure the absorbance of the cells at 600 nm and calculate the concentration of cells according to the following formula:

   1 absorbance unit = 2*10^7 cells/mL

Labeling cells for maximal affinity and minimal off-rate:

1. Prepare the 1° labeling solutions according to the following table. The recipes given below are intended to label 1*10^7 cells from the MB17 SSM or combinatorial libraries for maximal affinity for either Mdmx or Mdm2 under non-avid conditions and minimal off-rate. The volume of 1° labeling solution must be large enough to provide at least 10 molar equivalents of Mdmx or Mdm2 for every displayed molecule in the surface of the yeast cells, assuming 5*10^4 molecules/cell for well-displaying cells and taking into account the fraction of cells displaying protein. The MB17.41 and MB17.21 solutions are used to execute the off-rate selection. All volumes are given in uL.

	Mdmx Affinity	Mdm2 Affinity	SAPE Mix	MB17.41 competitor	MB17.21 competitor	
Component	**10 nM**	**5 nM**		**1 uM**	**1 uM**	**Total needed**
1 uM Mdm4	4.15					4.15
1 uM Mdm2		4.16				4.16
14 uM MB17.41				7.37		7.37
17 uM MB17.21					5.82	5.82
1 mg/mL SAPE			5.08			5.08
1 mg/mL Fitc			6.67			6.67
100 mg/mL BSA	41.5	83.1	20.0	10.0	10.0	165
PBSF	369	744	168	90.0	90.0	1461
Total	**415**	**831**	**200**	**100**	**100**	

2. Add to an eppendorf tube a sufficient volume of cells from each library provide 10-fold coverage of the library diversity.

3. Centrifuge the cells and remove the supernatant.
4. Add the Mdmx Affinity or Mdm2 Affinity 1° labeling solution to the cells.
5. Incubate the cells for 60 minutes at 4°C.
6. Following the 1° labeling, wash the cells with 900 uL PBSF, centrifuge the cells at 20000 g for 30 seconds, and discard the wash solution.
7. Add the MB17.41 competitor to those cells previously labeled with the Mdmx Affinity solution and the the MB17.21 competitor to those cells previously labeled with the Mdm2 Affinity solution and incubate the cells for 4 hours on the tube rotator in the 37°C room. This step selects for variants with the slowest off-rates.
8. Following the off-rate selection incubation, wash the cells with 900 uL PBSF, centrifuge the cells at 20000 g for 30 seconds, and discard the wash solution.
9. Add 100 uL of the SAPE 2° solution to each pool of cells and incubate on ice for 30 minutes.
10. Following the 2° labeling, wash the cells with 900 uL PBSF, centrifuge the cells at 20000 g for 30 seconds, and discard the wash solution. Save the labeled cells as pellets on ice until ready to sort.
11. Place the labeled cells in a 5 mL FACS tube on the FACS machine and sort them according to established procedures.

Labeling cells for maximal specificity:

1. Prepare the 1°/2° labeling solutions according to the following table. Omit the 50 uM biotin when initially preparing the master mixes. The recipes given below are intended to label 1*10^7 cells from the MB17 SSM or combinatorial libraries for maximal specificity for Mdmx over Mdm2 under avid conditions or for maximal specificity for Mdm2 over

Mdmx under avid conditions. Avidity is used in the specificity selection conditions to allow Mdmx-specific libraries that bind Mdm2 very weakly to give usable levels of Mdm2 signal during sorting and to allow Mdm2-specific libraries that bind Mdmx very weakly to give usable levels of Mdmx signal during sorting. The volume of 1° labeling solution must be large enough to provide at least 10 molar equivalents of Mdmx or Mdm2 for every displayed molecule in the surface of the yeast cells, assuming 5*10^4 molecules/cell for well-displaying cells and taking into account the fraction of cells displaying protein. All measurements are in uL.

	Mdm4_Speci_A	Mdm4_Speci_B	Mdm2_Speci_A	Mdm2_Speci_B	
Component	**10 nM**	**1 uM**	**1 uM**	**60 nM**	**Total needed**
10 uM Mdm4			4.00		4.00
1 uM Mdm4	4.16				4.16
10 uM Mdm2		41.6			41.6
1 uM Mdm2				6.00	6.00
1 mg/mL SAPE				7.03	7.03
100 ug/mL SAPE	3.05				3.05
1 mg/mL SA750		16	3.95		20.4
1 mg/mL Fitc	1.67	1.67	1.67	1.67	6.67
100 mg/mL BSA		20.8	5.00		25.8
PBSF	178	107	30.4	30.3	346
50 uM Biotin	20.8	20.8	5.00	5.00	51.6
Total	**208**	**208**	**50**	**50**	

2. Pre-incubate the master mixes for 30 minutes on ice to allow the Mdmx and Mdm2 to bind their cognate fluorophores.
3. Add the 50 uM biotin to each pre-incubation reaction and incubate the tubes on ice for 15 minutes. The biotin binds any fluorophores that have not bound Mdmx or Mdm2, preventing the fluorophores from cross-labeling the non-cognate target protein once the 2 pre-incubation reactions are combined.
4. Add to an eppendorf tube a sufficient volume of cells from each library provide 10-fold coverage of the library diversity.
5. Centrifuge the cells and remove the supernatant.

6. Combine the Mdmx Specificity A and B 1°/2° master mixes and add to the cells to be sorted for Mdmx specificity.
7. Combine the Mdm2 Specificity A and B 1°/2° master mixes and add to the cells to be sorted for Mdm2 specificity.
8. Incubate the cells for 2 hours on the tube rotator in the cold room.
9. Following the 1°/2° labeling, wash the cells with 900 uL PBSF, centrifuge the cells at 20000 g for 30 seconds, and discard the wash solution.
10. Place the labeled cells in a 5 mL FACS tube on the FACS machine and sort them according to established procedures.

Labeling cells for pan-specificity:

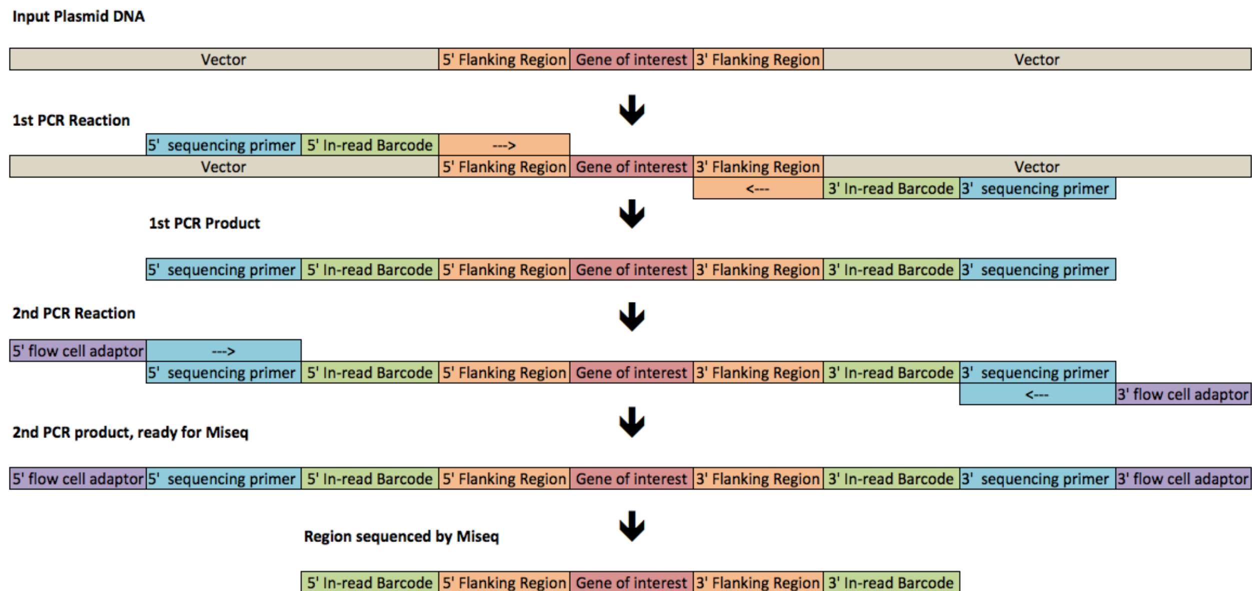
1. Prepare the 1°/2° labeling solutions according to the following table. Omit the 50 uM biotin when initially preparing the master mixes. The recipes given below are intended to label 1*10^7 cells from the MB17 combinatorial libraries for maximal but equal affinity for Mdmx and Mdm2 under non-avid conditions (pan-specificity) and 1*10^6 p53-displaying cells at the published Kds of p53 for Mdmx and Mdm2. The volume of 1° labeling solution must be large enough to provide at least 10 molar equivalents of Mdmx or Mdm2 for every displayed molecule in the surface of the yeast cells, assuming 5*10^4 molecules/cell for well-displaying cells and taking into account the fraction of cells displaying protein. All measurements are in uL.

Component	Pan_Speci_A 125 nM	Pan_Speci_B 125 nM	p53_control_A 210 nM	p53_control_B 140 nM	Total needed
2 uM Mdm4	6.3		2.10		8.4
1 uM Mdm2		12.5		2.80	15.3
1 mg/mL SAPE	14.6		4.92		19.6
500 ug/mL SA750		15.8		3.54	19.4
1 mg/mL Fitc	1.67	1.67			1.67
200 ug/mL Fitc			1.00	1.00	2.00
100 mg/mL BSA	5.00	5.00	0.98	1.66	12.6
PBSF	17.4	10.0			27.5
50 uM Biotin	5.00	5.00	1.00	1.00	12.0
Total	50	50	10	10	

2. Pre-incubate the master mixes for 30 minutes on ice to allow the Mdmx and Mdm2 to bind their cognate fluorophores.
3. Add the 50 uM biotin to each pre-incubation reaction and incubate the tubes on ice for 15 minutes. The biotin binds any fluorophores that have not bound Mdmx or Mdm2, preventing the fluorophores from cross-labeling the non-cognate target protein once the 2 pre-incubation reactions are combined.
4. Add to an eppendorf tube a sufficient volume of cells from each library provide 10-fold coverage of the library diversity.
5. Centrifuge the cells and remove the supernatant.
6. Combine the Pan-Specificity A and B 1°/2° master mixes and add to the cells to be sorted for Pan-specificity.
7. Combine the p53 control A and B 1°/2° master mixes and add to the p53-displaying cells.
8. Incubate the cells for 2 hours on the tube rotator in the cold room.
9. Following the 1°/2° labeling, wash the cells with 900 uL PBSF, centrifuge the cells at 20000 g for 30 seconds, and discard the wash solution.
10. Place the labeled cells in a 5 mL FACS tube on the FACS machine and sort them according to established procedures.

9.5 Extraction and high-throughput sequencing of MB17 gene pools from yeast

We extracted the plasmid DNA from cells saved from the unselected library and from successive rounds of sorting for each of the selection conditions. We amplified the MB17 gene in 2 150 base-pair halves. We sequenced the resulting gene fragments on an Illumina MiSeq v2 (Illumina, San Diego, CA, USA) with paired-end 2 x 259 base pair reads, obtaining roughly 7-15 million paired-end reads after filtering. This DNA extraction method is a modification of a method developed by Aaron Chevalier and Tim Whitehead. Portions of this procedure originally come from the Andy Scharenberg and Maitreya Dunham labs at the University of Washington. The protocol produces about 100 ng template per 1*10^7 yeast cells containing low-copy number plasmid with a Cen6 ori, which quantification shows gives 1-3 plasmid copies per cell. Briefly, 1*10^7 yeast cells are stored as pellets in -80°C freezer. The cell wall is removed using zymolase and the cells are lysed by a freeze-thaw cycle followed by an alkaline lysis miniprep procedure. Sheared genomic DNA and ssDNA are partially cleaned up from the plasmid DNA by an exonuclease processing step and then a PCR step amplifies the gene and appends pool-specific barcodes. A 2nd PCR step appends Ilumina flow-cell adaptors. PCR products are purified by extraction from an agarose gel and quantified using a Qubit fluorimeter (Life Technologies, Carlsbad, CA, USA). The order of the PCR steps and structure of the template, primers, and product is shown the following schematic. Note that the relative length of DNA regions is not to scale.



A detailed protocol is given below:

1. Yeast Miniprep (protocol developed in the Andy Scharenberg lab):

Uses the Yeast Plasmid Miniprep II Kit (Zymo Research, Irvine, California, USA) and the QIAprep Spin Miniprep Kit (Qiagen, Venlo, Limburg, Netherlands).

1. Pellet 4*10^7 yeast cells by centrifugation at 20000 g for 30 seconds.
2. Resuspend the cells in 200 uL Yeast Plasmid Miniprep II Kit Solution 1.
3. Add 5µl Zymolyase (5U/uL) to the cells
4. Incubate the cells at 37°C for 4 hours, mixing once per hour.
5. Perform 1 freeze-thaw cycle in a dry ice/EtOH bath and at 42°C.
6. Add 200µl Yeast Plasmid Miniprep II Kit Solution 2, mix well, and let sit 3-5 min.
7. Add 400µl Yeast Plasmid Miniprep II Kit Solution 3, mix well, and centrifuge at 20000 g for 5 min (I generally have to remove supernatant and spin a second time to remove all of the flocculent).
8. Transfer the supernatant to a QIAprep Spin Miniprep Kit column.
9. Spin the column for 1 minute at 20000 g.

10. Add 700 uL QIAprep Spin Miniprep Kit PB buffer and spin for 30 seconds at 20000 g.
11. Add 700 uL QIAprep Spin Miniprep Kit PE buffer and spin for 30 seconds at 20000 g.
12. Repeat the PE wash.
13. Pour out supernatant and spin for 1 minute at 20000 g to dry out the column.
14. Add 32 uL sterile water to elute, let the column stand for 1 minute, and spin for 1 minute at 20000 g.
15. Reload column with the eluate and spin again.

2. Further Purification of the yeast plasmid (protocol developed in the Maitreya Dunham lab):

The DNA prep is dirty and needs to be cleaned up before any large scale PCR amplification. This step is used to degrade any interfering genomic DNA. We saw inhibition with PCR and qPCR before adding this step. Store 15 uL of the elution from the previous step at -20°C and proceed with remaining 15 uL. This step requires Exonuclease I (Epicentre, Madison, Wisconsin, USA) that catalyzes the removal of nucleotides from single-stranded DNA in the 3' to 5' direction and Lamda exonuclease (New England Biolabs, Ipswich, Massachusetts, USA) that cleaves mononucleotides from duplex DNA 5' to 3'.

1. Make up the following recipe:

Component:	Volume:
Miniprepped plasmid DNA	15 uL
ExoI	2 uL
Lambda	1 uL
Lambda buffer 10X	2 uL
Final volume	20 uL

2. On a thermocycler, incubate the mixture for 90 minutes at 30°C and then inactivate the enzymes for 20 min at 80°C.

3. Qiagen PCR cleanup:

Follow the standard kit procedure and elute in 32 uL sterile water. The yield will be 30 uL. Store 15 uL of the elution at -20°C and proceed with remaining 15 uL.

4. First large Scale PCR Amplification to append barcodes:

1. Phusion high fidelity polymerase (New England Biolabs, Ipswich, Massachusetts, USA) is used to amplify up the template and add the pool-specific barcodes.
2. Make up the 1st PCR master mix as follows:

Component:	For 1 reaction:	For 16 reactions:
Pasmid DNA from PCR cleanup	15 uL	-
5X Phusion HF Buffer	10 uL	170 µl
5 uM Fwd primer	2 uL	-
5 uM Rev primer	2 uL	-
10 mM dNTPs	1 uL	17 uL
sterile water	19.5 uL	331.5 uL
Phusion Polymerase	0.5 uL	8.5 uL
Total	50 uL	527 uL

3. Use the following amplification primers in the 1st PCR step to append unique barcodes to the genes from each experimental pool:

>MB17_PCR_1_Front_Fwd_0

ACGACGCTCTTCCGATCTACCAGGCGCTGGGGTGGTAGTGGAGGTATG

>MB17_PCR_1_Front_Fwd_1

ACGACGCTCTTCCGATCTGAGGCCTTGGCCGGTGGTAGTGGAGGTATG

>MB17_PCR_1_Front_Fwd_2

ACGACGCTCTTCCGATCTCTTTAAAATATAGGTGGTAGTGGAGGTATG

>MB17_PCR_1_Front_Fwd_3

ACGACGCTCTTCCGATCTTGACTTGCACATGGTGGTAGTGGAGGTATG

>MB17_PCR_1_Front_Fwd_4

ACGACGCTCTTCCGATCTGGTCGCGGACCAGGTGGTAGTGGAGGTATG

>MB17_PCR_1_Front_Fwd_5

ACGACGCTCTTCCGATCTCCGGTTCCGGAGGGTGGTAGTGGAGGTATG

>MB17_PCR_1_Front_Fwd_6

ACGACGCTCTTCCGATCTATATAAAATTTCGGTGGTAGTGGAGGTATG

>MB17_PCR_1_Front_Fwd_7

ACGACGCTCTTCCGATCTTACACGTTCAGTGGTGGTAGTGGAGGTATG

>MB17_PCR_1_Front_Rev_8

ACGTGTGCTCTTCCGATCTTGGTCCGCGACCTCAATACCTTCAGAAGCC

>MB17_PCR_1_Front_Rev_9

ACGTGTGCTCTTCCGATCTCTCCGGAACCGGTCAATACCTTCAGAAGCC

>MB17_PCR_1_Front_Rev_10

ACGTGTGCTCTTCCGATCTGAAATTTTATATTCAATACCTTCAGAAGCC

>MB17_PCR_1_Front_Rev_11

ACGTGTGCTCTTCCGATCTACTGAACGTGTATCAATACCTTCAGAAGCC

>MB17_PCR_1_Front_Rev_12

ACGTGTGCTCTTCCGATCTCCAGCGCCTGGTTCAATACCTTCAGAAGCC

>MB17_PCR_1_Front_Rev_13

ACGTGTGCTCTTCCGATCTGGCCAAGGCCTCTCAATACCTTCAGAAGCC

>MB17_PCR_1_Front_Rev_14

ACGTGTGCTCTTCCGATCTTATATTTTAAAGTCAATACCTTCAGAAGCC

>MB17_PCR_1_Front_Rev_15

ACGTGTGCTCTTCCGATCTATGTGCAAGTCATCAATACCTTCAGAAGCC

>MB17_PCR_1_Back_Fwd_8

ACGACGCTCTTCCGATCTTGGTCCGCGACCCTATGCTGATATTTTCTCTGTT

>MB17_PCR_1_Back_Fwd_9

ACGACGCTCTTCCGATCTCTCCGGAACCGGCTATGCTGATATTTTCTCTGTT

>MB17_PCR_1_Back_Fwd_10

ACGACGCTCTTCCGATCTGAAATTTTATATCTATGCTGATATTTTCTCTGTT

>MB17_PCR_1_Back_Fwd_11

ACGACGCTCTTCCGATCTACTGAACGTGTACTATGCTGATATTTTCTCTGTT

>MB17_PCR_1_Back_Fwd_12

ACGACGCTCTTCCGATCTCCAGCGCCTGGTCTATGCTGATATTTTCTCTGTT

>MB17_PCR_1_Back_Fwd_13

ACGACGCTCTTCCGATCTGGCCAAGGCCTCCTATGCTGATATTTTCTCTGTT

>MB17_PCR_1_Back_Fwd_14

ACGACGCTCTTCCGATCTTATATTTTAAAGCTATGCTGATATTTTCTCTGTT

>MB17_PCR_1_Back_Fwd_15

ACGACGCTCTTCCGATCTATGTGCAAGTCACTATGCTGATATTTTCTCTGTT

>MB17_PCR_1_Back_Rev_0

ACGTGTGCTCTTCCGATCTACCAGGCGCTGGCCTCCTTCTTCAACTTCTT

>MB17_PCR_1_Back_Rev_1

ACGTGTGCTCTTCCGATCTGAGGCCTTGGCCCCTCCTTCTTCAACTTCTT

>MB17_PCR_1_Back_Rev_2

ACGTGTGCTCTTCCGATCTCTTTAAAATATACCTCCTTCTTCAACTTCTT

>MB17_PCR_1_Back_Rev_3

ACGTGTGCTCTTCCGATCTTGACTTGCACATCCTCCTTCTTCAACTTCTT

>MB17_PCR_1_Back_Rev_4

ACGTGTGCTCTTCCGATCTGGTCGCGGACCACCTCCTTCTTCAACTTCTT

>MB17_PCR_1_Back_Rev_5

ACGTGTGCTCTTCCGATCTCCGGTTCCGGAGCCTCCTTCTTCAACTTCTT

>MB17_PCR_1_Back_Rev_6

ACGTGTGCTCTTCCGATCTATATAAAATTTCCCTCCTTCTTCAACTTCTT

>MB17_PCR_1_Back_Rev_7

ACGTGTGCTCTTCCGATCTTACACGTTCAGTCCTCCTTCTTCAACTTCTT

4. Add 31 uL of the master mix, 15 uL of the PCR cleanup product, and 2 uL each of 5 uM pool-specific forward and reverse primers to the wells of a PCR strip.

5. Mix the PCR reactions and cap the PCR strip.

6. Place the PCR strip in a thermocycler and run the following protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	30 seconds
3	58°C	30 seconds
4	72°C	1 minute
5	Go to Step 2	24 times
6	72°C	10 minutes
7	4°C	forever

7. Pour a 2% agarose gel with with SYBR safe gel stain (Life Technologies, Carlsbad, California, USA).

8. Add the entire completed 1st PCR reaction to 10 uL of 6X DNA loading dye (New England Biolabs, Ipswich, Massachusetts, USA) and run on the gel at 100 V.
9. Visualize the DNA bands and use a razor blade to cut out the band corresponding in size to the amplified gene.
10. Extract the DNA from the gel slice using a QIAquick Gel Extraction Kit, following the manufacturer's protocol and eluting in 50 uL Qiagen Buffer EB.
11. Quantify the amplified DNA using a Qubit fluorimeter and the Qubit dsDNA HS Kit, following the manufacturer's protocol.

5. Second large Scale PCR Amplification to append Illumina adaptors:

1. Phusion high fidelity polymerase (New England Biolabs, Ipswich, Massachusetts, USA) is used to append the Illumina adaptor regions to the 1st PCR product.
2. Make up the 2nd PCR master mix as follows

Component:	For 1 reaction:	For 16 reactions:
Pasmid DNA from PCR cleanup	1 uL	-
5X Phusion HF Buffer	10 uL	170µl
5 uM Fwd primer	2 uL	34 uL
5 uM Rev primer	2 uL	34 uL
10 mM dNTPs	1 uL	17 uL
sterile water	33.5 uL	569.5 uL
Phusion Polymerase	0.5 uL	8.5 uL
Total	50 uL	833 uL

3. Use the following amplification primers in the 2nd PCR step to append the Illumina adaptor regions to the 1st PCR product.

>MB17_PCR_2_Fwd

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA
TCT

>MB17_PCR_2_Rev

CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCG

ATCT

4. Add 31 uL of the master mix, 15 uL of the PCR cleanup product, and 2 uL each of 5 uM pool-specific forward and reverse primers to the wells of a PCR strip.
5. Mix the PCR reactions and cap the PCR strip.
6. Place the PCR strip in a thermocycler and run the following protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	30 seconds
3	66°C	30 seconds
4	72°C	1 minute
5	Go to Step 2	24 times
6	72°C	10 minutes
7	4°C	forever

7. Pour a 2% agarose gel with with SYBR safe gel stain (Life Technologies, Carlsbad, California, USA).
8. Add the entire completed 2nd PCR reaction to 10 uL of 6X DNA loading dye (New England Biolabs, Ipswich, Massachusetts, USA) and run on the gel at 100 V.
9. Visualize the DNA bands and use a razor blade to cut out the band corresponding in size to the amplified gene.
10. Extract the DNA from the gel slice using a QIAquick Gel Extraction Kit, following the manufacturer's protocol and eluting in 50 uL Qiagen Buffer EB.
11. Quantify the amplified DNA using a Qubit fluorimeter and the Qubit dsDNA HS Kit, following the manufacturer's protocol.

6. High-throughput sequencing using the Illumina Miseq:

Sequence the DNA from each of the pools on an Illumina Miseq (Illumina, San Diego, CA, USA) using a paired-end run according to the manufacturer's protocol. The barcodes will be the first bases read during the forward and reverse reads. The Miseq must read equal amounts of all 4 bases during the first cycles to properly identify clusters and calibrate the instrument. The following barcode sets are an expanded version of a barcode set obtained from Dr. Rijhu Das at Stanford University:

Barcode_Set_A: (Forward)

Barcode_00: ACCAGGCGCTGG

Barcode _01: GAGGCCTTGGCC

Barcode _02: CTTTAAAATATA

Barcode _03: TGACTTGCACAT

Barcode_Set_B: (Reverse)

Barcode _04: GGTCGCGGACCA

Barcode _05: CCGGTTCCGGAG

Barcode _06: ATATAAAATTTC

Barcode _07: TACACGTTCAGT

Barcode_Set_C: (Forward complement)

Barcode _08: TGGTCCGCGACC

Barcode _09: CTCCGGAACCGG

Barcode _10: GAAATTTTATAT

Barcode _11: ACTGAACGTGTA

Barcode_Set_D: (Reverse complement)

Barcode _12: CCAGCGCCTGGT

Barcode _13: GGCCAAGGCCTC

Barcode _14: TATATTTTAAAG

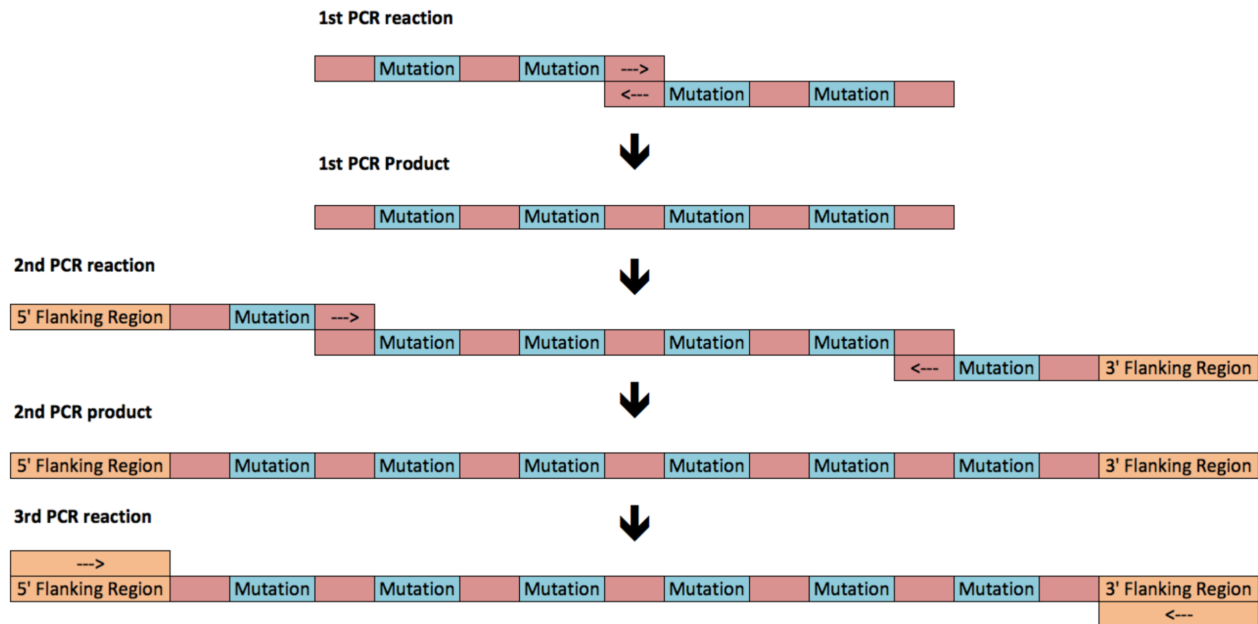
Barcode _15: ATGTGCAAGTCA

Each of the barcodes within a set of 4 should be used together in equal molar amounts to ensure that the Miseq will see equal amounts of each base at during each of the first 12 cycles of sequencing. As the barcodes are contained within the forward and reverse reads, there is no need to include an index read, freeing up more sequencing cycles for the forward and reverse reads. If there are more than 4 pools to sequence, simply use additional sets of barcodes. If the number of pools to sequence is not a factor of 4, use 2 barcodes per pool for larger pools until the number of barcodes is a factor of 4. Barcodes sets B, C, and D are simply the reverse, forward complement, and reverse complement of barcode set A, respectively. In order to avoid mispriming during PCR amplification and hairpin formation within the PCR products, certain sets of barcodes should not be used together in the forward and reverse primers for a given pool during the PCR amplification steps. Barcode set A should be paired with set B and set C should be paired with set D. Barcode set A should not be paired with sets C or D, and so forth.

9.6 MB17 Combinatorial library construction via ultramer assembly

The combinatorial libraries were each constructed by assembling 4 ultramers that contained degenerate codons at 9-11 positions in the target gene and that included the flanking sequences from pETCON to allow for recombination into the vector. Each degenerate codon was carefully selected to code for the wild type amino acid plus a small number of target mutations

while coding for as few off-target mutations as possible (see **Section 3.4** in the Appendix). The target diversity of the library was 1*10^6 theoretical variants. The ultramers used to assemble any given library were between 120 and 180 bases in length overlapped one another with ~ 30 base pair overlaps to allow for assembly. The library assembly strategy is given in the schematic below. The inner two ultramers were first assembled using overlap PCR and then gel purified. Next the outer ultramers were added to the gel purified inner fragment in a traditional PCR amplification step and the resulting product was also gel purified. The complete gene fragment was amplified with short outer primers and gel purified a third time in preparation for electroporation into yeast. The N-terminal 2 ultramers and amplification primer should be sense while the C-terminal 2 ultramers and amplification primer should be antisense in order for this to work.



A detailed protocol is given below:

1. Re-suspend 4 nmol of each ultramer in 400 uL of sterile water to give a final ultramer concentration of 10 uM.

2. For each library, make up the 1st PCR reactions in the wells of a PCR strip according to the following recipes. It's important to try a range of ultramer concentrations when assembling the inner ultramers to optimize purity and yield. Experience has shown that the optimal ultramer concentration varies from library to library.

Component:	20 nM	50 nM	100 nM	200 nM	500 nM	1000 nM
5X Phusion HF buffer	10 uL	10 uL	10 uL	10 uL	10 uL	10 uL
1 uM forward inner ultramer	1 uL	2.5 uL	5 uL	---	---	---
1 uM reverse inner ultramer	1 uL	2.5 uL	5 uL	---	---	---
10 uM forward inner ultramer	---	---	---	1 uL	2.5 uL	5 uL
10 uM reverse inner ultramer	---	---	---	1 uL	2.5 uL	5 uL
10 mM dNTPS	1 uL	1 uL	1 uL	1 uL	1 uL	1 uL
Sterile water	36.5 uL	33.5 uL	28.5 uL	36.5 uL	33.5 uL	28.5 uL
2U/uL Phusion Pol	0.5 uL	0.5 uL	0.5 uL	0.5 uL	0.5 uL	0.5 uL
Total	50 uL	50 uL	50 uL	50 uL	50 uL	50 uL

3. Place the PCR strip in a thermocycler and run the following PCR protocol:

Step:	Temperature:	Time:
1	98°C	3 minutes
2	98°C	30 seconds
3	70°C	30 seconds
4	72°C	1 minute
5	Go to step 2	4 times
6	72°C	10 minutes
7	4°C	forever

4. Run each entire 1st PCR reaction with 10 uL of 6X DNA loading dye (New England Biolabs, Ipswich, Massachusetts, USA) on a 2% agarose gel supplemented with SYBR safe gel stain (Life Technologies, Carlsbad, California, USA) at 100 V.

5. Visualize the gel and cut out the bands corresponding to the expected size of the inner ultramer assembly product. Choose the reaction condition(s) that give(s) the best yield of the desired product with a minimal amount of undesired products.

6. Extract the DNA fragments from the agarose gel using the QIAquick Gel Extraction Kit (Qiagen, Venlo, Limburg, Netherlands) according to the vendor protocol and elute in Qiagen buffer EB.
7. Quantify the gel extracted DNA using a Qubit fluorimeter according to the manufacturers protocol.
8. Make up the 2nd PCR reactions in the wells of a PCR strip according to the following recipes. In the case of MB17_Affi_Mdmx and MB17_Affi_Mdm2, the products from the 500 and 1000 nM conditions in the 1st PCR were each combined and used in respective 2nd PCR reactions. In the case of MB17_Speci_Mdmx and MB17_Speci_Mdm2, the 50 nM condition from the 1st PCR was used in eight 2nd PCR reactions each.

Component:	Affi_Mdmx 500 nM	Affi_Mdm2 500 nM	Speci_Mdmx 50 nM	Speci_Mdm2 50 nM
5X Phusion HF buffer	10 uL	10 uL	80 uL	80 uL
10 uM forward outer ultramer	2.5 uL	2.5 uL	20 uL	20 uL
10 uM reverse outer ultramer	2.5 uL	2.5 uL	20 uL	20 uL
1st PCR product	1.98 uL	1.14 uL	16 uL	16 uL
10 mM dNTPS	1 uL	1 uL	8 uL	8 uL
Sterile water	32 uL	32 uL	252 uL	252 uL
2U/uL Phusion Pol	0.5 uL	0.5 uL	4 uL	4 uL
Total	50 uL	50 uL	400 uL	400 uL

9. Place the PCR strip in a thermocycler and run the following PCR protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	10 seconds
3	70°C	30 seconds
4	72°C	30 seconds
5	Go to step 2	24 times
6	72°C	10 minutes
7	4°C	forever

10. Run each entire 2nd PCR reaction with 10 uL of 6X DNA loading dye on a 2% agarose gel supplemented with SYBR safe gel stain at 100 V.
11. Visualize the gel and cut out the bands corresponding to the expected size of the inner ultramer assembly product. Choose the reaction condition that gives the best yield of the desired product with a minimal amount of undesired products.
12. Extract the DNA fragments from the agarose gel using the QIAquick Gel Extraction Kit (Qiagen, Venlo, Limburg, Netherlands) according to the vendor protocol and elute in Qiagen buffer EB.
13. Quantify the gel extracted DNA using a Qubit fluorimeter according to the manufacturers protocol.
14. Make up the 3rd PCR reactions according to the following recipes. Each library was amplified in 8 respective 3rd PCR reactions.

Component:	MB17 Affi Mdmx:	MB17 Affi Mdm2:	MB17 Speci Mdmx:	MB17 Speci Mdm2:
5X Phusion HF buffer	80 uL	80 uL	80 uL	80 uL
100 uM forward amp primer	6 uL	6 uL	4 uL	4 uL
100 uM reverse amp primer	6 uL	6 uL	4 uL	4 uL
10 mM dNTPS	24 uL	24 uL	16 uL	16 uL
2nd PCR product	4.40 uL	15.4 uL	20 uL	20 uL
Sterile water	276 uL	265 uL	272 uL	272 uL
2U/uL Phusion Pol	4 uL	4 uL	4 uL	4 uL
Total	400 uL	400 uL	400 uL	400 uL

15. The amplification primers used for the 3rd PCR are given below:

>Fwd_amp

GCTAGTGGTGGAGGAGGCTCTG

>Rev_amp

GTTGTTATCAGATCTCTATTACAAGTCCTCTTCAGA

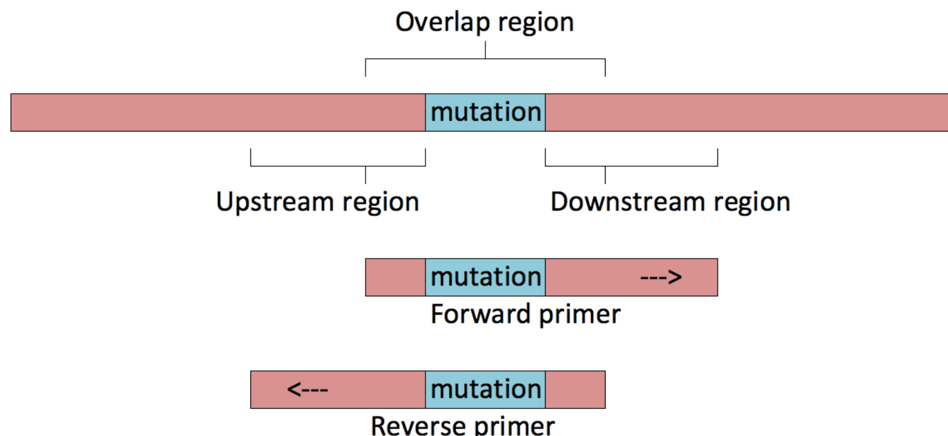
16. Aliquot 50 uL of the 3rd PCR reaction into each of 8 wells of a PCR strip, place the PCR strip in a thermocycler, and run the following PCR protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	10 seconds
3	68°C	30 seconds
4	72°C	30 seconds
5	Go to step 2	24 times
6	72°C	10 minutes
7	4°C	forever

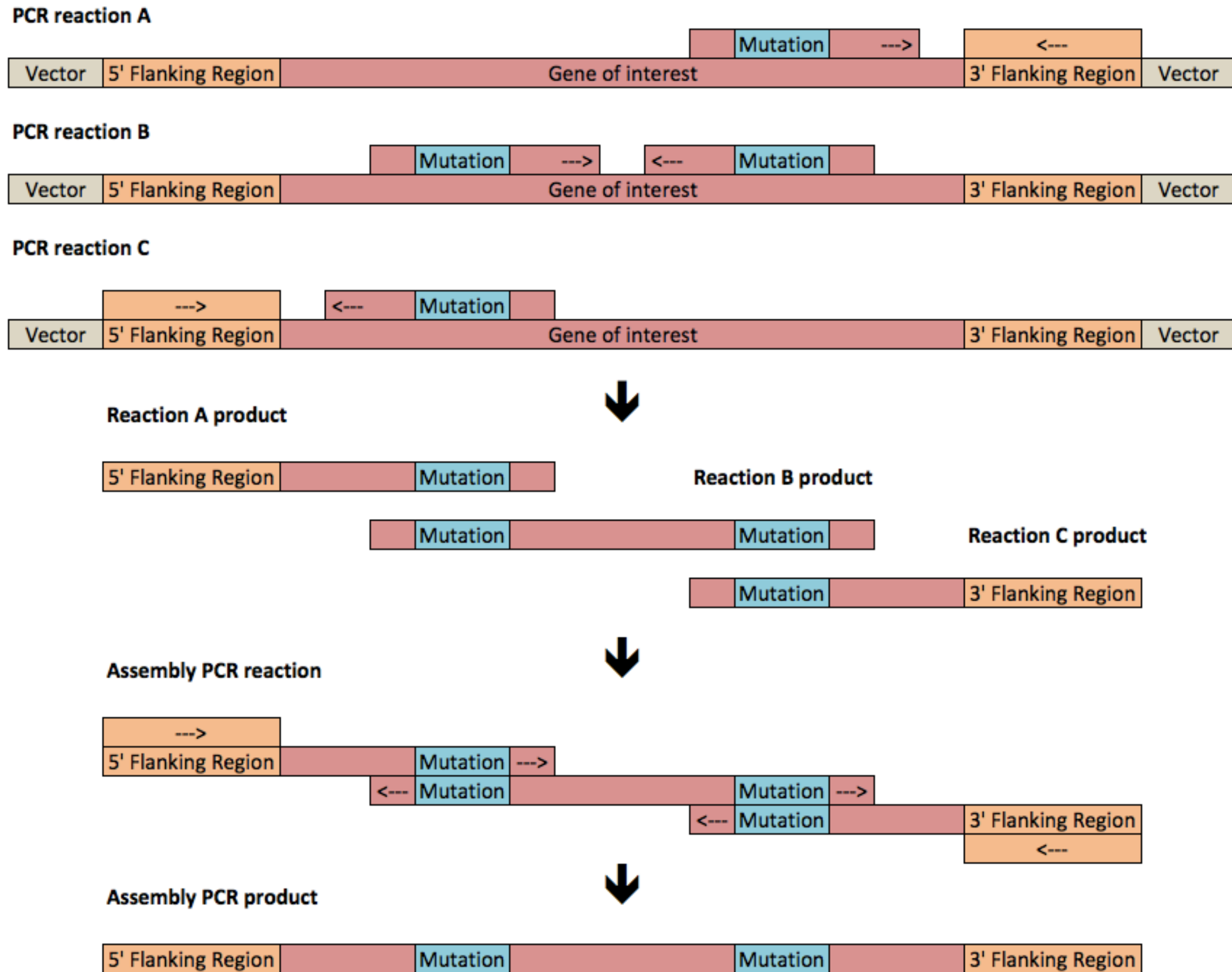
17. Run each entire 3rd PCR reaction with 10 uL of 6X DNA loading dye on a 2% agarose gel supplemented with SYBR safe gel stain at 100 V.
18. Visualize the gel and cut out the bands corresponding to the expected size of the inner ultramer assembly product. Choose the reaction condition that gives the best yield of the desired product with a minimal amount of undesired products.
19. Extract the DNA fragments from the agarose gel using the QIAquick Gel Extraction Kit (Qiagen, Venlo, Limburg, Netherlands) according to the vendor protocol and elute in Qiagen buffer EB.
20. Quantify the gel extracted DNA using a Qubit fluorimeter according to the manufacturers protocol.
21. Desalt the eluted DNA fragments by aliquoting the DNA solution onto nitrocellulose membranes suspended above a petri dish filled with sterile water and letting stand for 20 minutes at room temperature.
22. Electroporate the DNA fragments with linearlized pETCON vector into EBY100 yeast using the protocol of Benatuil, et al.

9.7 Overlap PCR mutagenesis and assembly

This protocol is a modification of the “Improved methods for site-directed mutagenesis using Gibson Assembly Master Mix” protocol (New England BioLabs, Inc, Ipswich, Massachusetts, USA). Primers for the PCR reactions are designed as described below and as depicted in the following schematic:



For each site to be mutated, a region of at least 18 bp surrounding that site is identified such that the altered nucleotides are centered in the region and the Phusion melting temperature of the region is between 65°C and 68°C. Next a region of at least 15 bp immediately upstream of the altered nucleotides is identified such that Phusion melting temperature of the region is between 65°C and 68°C. Likewise, a region of at least 15 bp immediately downstream of the altered nucleotides is identified such that Phusion melting temperature of the region is between 65°C and 68°C. The reverse primer for each mutagenic site covers from the 5’end of the upstream region to the 3’ end of the overlap region. The forward primer for each mutagenic site covers from the 5’end of the overlap region to the 3’ end of the downstream region. Forward primers are sense and reverse primers are antisense. The PCR steps are set up as described below and as illustrated in the following schematic:



During the initial PCR step, the forward primer covering the 5' flanking region of the gene is paired with the 1st reverse mutagenic primer downstream of it. The 1st forward mutagenic primer is paired with the 2nd reverse mutagenic primer, the 2nd forward mutagenic primer is paired with the 3rd reverse mutagenic primer, and the 3rd forward mutagenic primer is paired with the primer covering the 3' flanking region of the gene. The fragments generated in each of the initial PCR steps are combined with the primers covering the 5' and 3' flanking regions of the gene and assembled in an assembly PCR reaction. A detailed laboratory protocol is given below:

1. Mutagenic primers are synthesized by a commercial vendor (IDT, San Jose, California, USA) and are brought up in Qiagen buffer EB to a concentration of 100 uM.

2. Purify gene of interest (in pETCON) from EBY100 yeast using the Yeast Plasmid Miniprep II Kit (Zymo Research, Irvine, California, USA) according the vendor protocol.
3. Transform the miniprepped gene of interest into chemically competent XL1-Blue E. coli cells (Agilent Technologies, Santa Clara, California, USA), plate on LB agar supplemented with 50 mg/mL carbenicillin, and incubate the plate at 37°C for 16 hours.
4. Pick 4 colonies from the plate and grow in 3 mL LB media supplemented with 50 mg/mL carbenicillin for 16 hours at 37°C.
5. Extract the amplified plasmid DNA from the cell using the QIAprep spin miniprep kit (Qiagen, Venlo, Limburg, Netherlands) according to the vendor protocol.
6. Sanger sequence the miniprepped plasmid DNA and select a preparation having no mutations and a clean sequencing trace. It's best to use a sequenced clone of the gene of interest as the template for the initial PCR reaction as some fraction of the gBlock molecules contain deletions and point mutations that would then be incorporated into the assembled product.
7. Make up the initial PCR reaction master mix according to the following recipes:

Component:	For 1 reaction:	For 3 reactions:
Serile water	36.5 uL	113 uL
5X Phusion HF Buffer	10 uL	31 uL
4 ng/uL template	1 uL	3.1 uL
100 uM forward primer	0.5 uL	---
100 uM reverse primer	0.5 uL	---
10 mM dNTPs	1 uL	3.1 uL
2U/uL Phusion Pol	0.5 uL	1.55 uL
Total	50 uL	152 uL

8. The 5' and 3' flanking primers used here are given below:

>colony_outer_forward (COF, 5')

TGACAACTATATGCGAGCAAATCCCCTCAC

>colony_outer_reverse (COR, 3')

GTACGAGCTAAAAGTACAGTGGGAAC

9. For each mutagenic fragment to be generated, add 0.5 uL of the 100 uM appropriate forward primer, 0.5 uL of the 100 uM appropriate reverse primer and 49 uL of the PCR master mix to a well of a PCR strip and mix.

10. Cap the PCR strip, place in a thermocycler, and run the following PCR protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	30 seconds
3	67°C	30 seconds
4	72°C	1 minute
5	Go to Step 2	24 times
6	72°C	10 minutes
7	4°C	forever

11. Run 5 uL of each PCR reaction with 1 uL 6X DNA loading dye (New England Biolabs, Ipswich, Massachusetts, USA) on a 2% agarose gel supplemented with SYBR safe gel stain (Life Technologies, Carlsbad, California, USA) at 100 V.

12. Visualize the gel to verify that a single DNA fragment of the expected size has been obtained from each initial PCR reaction.

13. Make up the assembly PCR reaction in a well of a PCR strip according to the following recipe:

Component:	Volume:
Sterile water	36.5 uL
5X Phusion HF Buffer	10 uL
100 uM COF primer	0.5 uL
100 uM COR primer	0.5 uL
10 mM dNTPs	1 uL
2U/uL Phusion Pol	0.5 uL
Total	50 uL

14. Cap the PCR strip, place in a thermocycler, and run the following PCR protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	30 seconds
3	67°C	30 seconds
4	72°C	1 minute
5	Go to Step 2	24 times
6	72°C	10 minutes
7	4°C	forever

15. Run all of the assembly PCR reaction with 10 uL 6X DNA loading dye on a 2% agarose gel supplemented with SYBR safe gel stain at 100 V.
16. Visualize the gel to verify that a single DNA fragment of the expected size has been obtained from the assembly PCR reaction.
17. Cut out the gel band of the correct size for the assembly PCR product and extract the DNA fragment using the QIAquick Gel Extraction Kit (Qiagen, Venlo, Limburg, Netherlands) according to the vendor protocol and elute in Qiagen buffer EB.
18. Quantify the gel extracted DNA using a Qubit fluorimeter according to the manufacturers protocol.
19. Transform the DNA fragments with linearlized pETCON vector into chemically competent EBY100 yeast using the Yeast transformation protocol.
20. Use the Yeast colony PCR protocol to sequence 4-8 clones from from each gene that was mutagenized.
21. Analyze the sequencing data for genes that contain all of the desired mutations and no undesired mutations.

Section 10 (Appendix)

Experimental methods specific to the design of proteins to inhibit the interaction of EED-Ezh2

10.1 Biotinylation and size-exclusion chromatography of EED3

Soluble human EED3 was expressed and purified at Harvard University and biotinylated and further purified at the University of Washington (UW). Soluble human EED3 was expressed in E. coli, purified over Nickel-NTA resin, and desalted into PBS buffer without magnesium or calcium (10 mM disodium phosphate, 1.8 mM monopotassium phosphate, pH 7.4, 137 mM NaCl, 2.7 mM KCl). The EED was then concentrated, supplemented with glycerol to a final concentration of 10% v/v, and snap frozen in liquid nitrogen prior to shipment to the UW. At the UW, the EED was thawed and desalted into 20 mM Hepes, 250 mM Potassium Glutamate, pH 7.5, 1 mM TCEP. The proteins were concentrated to 3.1 mg/mL (57 uM) and enzymatically biotinylated with BirA biotin ligase using the kit from Avidity (Avidity LLC, Aurora, Colorado, USA). EED3 was separated from the BirA enzyme by re-purification over Nickel-NTA resin (Qiagen, Venlo,, Limburg, Netherlands), concentrated to a volume of 500 uL and further purified by size exclusion chromatography over a Superdex200 column into 20 mM Hepes, 150 mM NaCl, 5% v/v glycerol, 1 mM TCEP, pH 7.5. The proteins were concentrated to 0.54 mg/mL (10 uM), aliquoted, and snap frozen in liquid nitrogen for storage at -80°C. Thawed aliquots of EED3 were centrifuged to remove insoluble precipitates prior to incubation with yeast cells. A detailed protocol for the biotinylation and purification steps carried out at the University of Washington is given below. All steps except the biotinylation reaction were carried out at 4°C.

Buffers used:

Non-Imidazole Buffer: 20 mM HEPES, 150 mM NaCl, 5% v/v glycerol, 1 mM TCEP, pH 7.5

HBKG: 20 mM HEPES, 250 mM potassium glutamate, 1 mM TCEP, pH 7.5

Lo Imidazole Buffer: 20 mM Hepes, 150 mM NaCl, 5% v/v glycerol, 10 mM imidazole, 1mM TCEP, pH 7.5

Hi Imidazole Buffer: 20 mM HEPES, 150 mM NaCl, 5% v/v glycerol, 500 mM imidazole, 1 mM TCEP, pH 7.5

1. Thaw 900 uL EED3.
2. Equilibrate a PD-10 desalt column with 25 mL HBKG.
3. Spin EED3 1 min at 20000 g to remove insoluble aggregates and transfer the supernatant to a new tube.
4. Buffer exchange the EED3 into HBKG using the PD-10 desalt column, following the vendor protocol and taking 0.5 mL fractions of the flow-through, elution, and clean steps.
5. Spec the fractions at 280 nm on the Nanodrop spectophopmeter (Thermo Scientific, Waltham, Massachusetts, USA).
6. Combine the desalted fractions with significant A280 signal and concentrate to $\geq$ 57 uM using an amnicon concentrator with a 10,000 Da molecular weight cutoff.
7. Set up the biotinylation reaction as follows, mix, and incubate at room temperature for 4 hours.

Component:	Volume:
10X Biomix A	44 uL
10X Biomix B	44 uL
10X Biotin	44 uL
57 uM EED3	300 uL
BirA	10 uL
Total	442 uL

8. Attach the Superdex 200 size exclusion column to the Akta FPLC (GE Healthcare) and run 35 mL ddH20 at 0.9 mL/min to remove the 20% EtOH.
9. Equilibrate the Superdex 200 column with 35 mL Non Imidazole Buffer at 0.9 mL/min.

10. Prepare a 1 mL wide bore Ni column and equilibrate with 5 mL Lo Imidazole Buffer just before loading sample.
11. After the 4-hour biotinylation reaction is completed, dilute the reaction up to 1 mL with Lo Imidazole Buffer and load over the Ni column 10 times.
12. Wash the column with 10 mL Lo Imidazole Buffer.
13. Elute the column with 5 mL Hi Imidazole Buffer in 5 x 1 mL steps with 1-minute pauses in between steps.
14. Concentrate the eluted EED3 protein down to 500 uL.
15. Load the eluted, concentrated EED3 onto the Superdex 200 column and flow Non-Imidazole Buffer at 0.5 mL/minute for 24 mL, collecting 0.5 mL fractions.
16. Spec those fractions that show significant 280 nM signal on the FPLC using the Nanodrop at 280 nm.
17. Combine those fractions that correspond to the EED3 peak and concentrate to ≥ 10 uM.
18. Spec the concentrated EED3 at 280 nm using the Nanodrop and snap freeze in liquid nitrogen in 10 uL aliquots.

10.2 Small volume yeast surface titration of EED binders

Due to the limited quantities of soluble biotinylated EED3 available, the yeast surface titration method of Chao, et al. was modified to work with smaller labeling volumes (125). The detailed protocol is given below:

36. For each clone to be titrated, inoculate a single yeast colony, a ~30 uL stab of frozen cell stock, or ~10 uL of saturated yeast culture into 1 mL CD-UT (128) media with 4 uL antibiotic cocktail in a sterile 14 mL polypropylene culture tube.

37. Grow the cells at 30 °C for 16 hours with shaking at 250 rpm.
38. After 16 hours, the cell density will be about 6*10^7 cells/mL. Add 200 uL of the 1 mL culture (about 1.2*10^7 cells) to a sterile eppendorf tube.
39. Spin down the cells in the eppendorf tube at 20000 g for 30 seconds.
40. Pour the media out of the tube. Fear not, the cells will stay adhered to the bottom of the eppendorf tube.
41. Prepare a new sterile 14 mL culture tube with 1 mL SGCAA + 4 uL antibiotic cocktail.
42. Resuspend the cells in the eppendorf tube using the SGCAA solution from the new 14 mL culture tube and add the cell solution back to the culture tube.
43. Induce the cells at 30°C for 16 hours (or 22°C for 20 hours for difficult-to-express proteins) with shaking at 250 rpm. Alternatively, to speed up the protocol, the cells in step 1 can be immediately induced in 1 mL SGCAA instead of first growing overnight in CD-UT. If inducing at 30°C this results in a comparable amount of displayed protein/cell, but can result in fewer total cells if inducing directly from a stab of cell stock.
44. The following day, thaw out an aliquot of the soluble biotinylated EED3 on ice and centrifuge it at 20000 g for 2 minutes to remove aggregates.
45. Use the Nanodrop spectophopmeter (Thermo Scientific, Waltham, Massachusetts, USA) to determine the protein's concentration if needed. The target protein should be kept on ice at all times.
46. Freshly sterile filter 50 mL PBSF (PBS + 0.1% Bovine Serum Albumin (BSA)) into a 50 mL polypropylene tube and keep on ice. PBSF left over at the end of each day should be re-filtered each new day to remove adventitious cells that enter the solution the previous day and expand overnight, even if stored at 4°C.

47. Spin down the induced cells at 20000 g for 30 seconds. The cell density is about 6*10^7 cells/mL.
48. Pour the media out of the tube.
49. Re-suspend the cells in 1 mL ice cold PBSF.
50. Spin down the cells again at 20000 g for 30 seconds.
51. Pour the media out of the tube.
52. Re-suspend the cells in 300 uL ice cold PBSF. The cell concentration is about 2*10^8 cells/mL. These cells with displayed protein should be kept on ice and can be stored/used for up to 7 days, depending on the stability of the displayed protein.
53. Pre-block tubes and well-plate wells that will contain protein at concentrations below 10 nM by filling them with PBSF and incubating at 26°C for 20 minutes.
54. Once the 20-minute pre-blocking period ends, remove all the pre-blocking solution and save for later use (such as for washes and re-suspending samples prior to reading on the flow cytometer). Centrifuge the tubes to collect the residual pre-blocking solution and remove it as well.
55. For each target protein concentration to be titrated, make up a primary labeling master mix with the desired concentration of soluble biotinylated target protein. Include 10% BSA at a final concentration of 1% in master mixes having target protein concentrations above 30 nM. The increased concentration of BSA is intended to prevent nonspecific binding at higher protein concentrations. Use PBSF to make up any remaining volume in the master mix. An example titration setup is given in the following table (all volumes are given in uL):

Component	0	3 uM	1 uM	300 nM	100 nM	30 nM	10 nM	3 nM	1 nM	300 pM	100 pM	30 pM	Total needed
10 uM EED3		19.8	6.60	1.98									28.4
200 nM EED3 (1:50 dilution)					33.0	11.9	11.2	10.9	11.0	11.0	11.0	11.0	111
10% BSA	3.96	3.96	3.96	3.96	3.96	5.28	19.8						44.9
PBSF	35.6	15.8	29.0	33.7	2.64	35.6	167	689	2154	7275	21875	73091	105403
Total	40	40	40	40	40	53	198	700	2165	7286	21886	73102	
Number of samples	12	12	12	12	12	12	12	12	12	12	12	12	
Add to each reaction	3	3	3	3	3	4	15	53	164	552	1658	5538	
Individual Reaction Volume	5	5	5	5	5	6	17	55	166	554	1660	5540	
Volume of cells in reaction	2	2	2	2	2	2	2	2	2	2	2	2	
Labeling time (minutes)	NA	0.2	0.5	1.7	5	17	50						
Labeling time (hours)								3	8	20	32	32	

The minimum labeling time used here is 3*Tau, which is the time required by the incubation reaction to reach 95% of equilibrium. 3*Tau is calculated according to the following formula:

3 Tau = 3 * (K_{on} (in Molar^-1, seconds^-1) * [target protein] (in Molar)) + K_{off})^-1

If the on-rate (K_{on}) and off-rate (K_{off}) are not known for the interaction to be titrated, an on-rate of 1*10^5/Molar/second may be used and the off-rate can be estimated according to the following formula:

K_{off} = $K_{dissociation}$ (in Molar) * K_{on} (in Molar^-1, seconds^-1)

The labeling time must be long enough to allow equilibrium to be reached, but no longer so as to avoid the TCEP-mediated reduction of the disulfide bonds adhering the displayed protein to the yeast surface.

56. For each clone to be titrated, place 2 uL of the induced cells into the wells of a polypropylene V-bottom 96-well plate, 1.7 mL eppendorf tube, or 15 mL polypropylene tube as indicated in the table above. This corresponds to 4*10^5 cells per labeling reaction.

57. For each concentration of soluble target protein and sample of cells, add the volume of the master mix indicated in the table above to each drop of cells and mix well.
58. Cover the plate with a clear plastic adhesive seal and/or cap the tubes and rotate the plate and/or tubes at 26 °C for the length of time indicated in the table above. For cases where different concentrations must be labeled for different lengths of time, simply assemble the labeling reaction corresponding to the longest-time point first and then sequentially assemble labeling reactions corresponding to shorter-time points such that all of the labeling reactions complete their respective incubation times at the same time. The higher-concentration points cannot be labeled for longer than the indicated length of time because the soluble biotinylated Mdmx and Mdm2 must be stored in 1 mM TCEP. At 26 °C, concentrations of TCEP above 10 nM or so will rapidly reduce the disulfide bonds tethering the Aga2 yeast display apparatus to the surface of the yeast cells, releasing the displayed proteins to the surrounding labeling solution.
59. Once the labeling/incubation has completed, centrifuge samples with volumes greater than 5 uL and remove the labeling solution. For samples labeled in the 96-well plate, centrifuge at 4000 g for 5 minutes in a plate centrifuge and remove the solution by removing the plastic cover and shaking the plate upside down over a sink hard 3 - 4 times. The cells will remain adhered to the bottom of the plate no matter how hard you shake it and harder shaking removes the wash more completely. For samples in eppendorf tubes, centrifuge for 30 seconds at 20000 g and pipette out the solution. For samples in 15 mL polypropylene tubes, centrifuge at 4000 g for 5 minutes in a swing-bucket centrifuge and remove the solution using a serological pipette.

60. Add 200 uL of PBSF to re-suspend and wash the cells in each of the eppendorf and 15 mL tubes transfer them to corresponding empty wells of the 96-well plate.
61. Add 200 uL of PBSF to each well of the 96-well plate that contains cells incubated in the plate and replace the plastic seal.
62. Pellet the cells in the plate by centrifugation at 4000 g for 5 minutes using the plate rotor in the tabletop centrifuge.
63. Remove the plastic seal and shake the plate upside down over a sink hard 3 - 4 times to remove the wash supernatant.
64. Make up the secondary labeling solution according to the following recipe:

Component:	Volume for 1 sample:	For 96 samples:
1 mg/mL Steptavidin-R-Phycoerythrin conjugate (SAPE)	0.2 uL	20 uL
1 mg/mL FITC-conjugated Chicken-anti-Myc Antibody (FITC)	0.2 uL	20 uL
PBSF	4.6 uL	460 uL
Total	5 uL	500 uL

65. Add 5 uL to the cells in each well of the 96-well plate and incubate on ice for 30 minutes.
66. Once the 30-minute secondary incubation has completed, add 200 uL PBSF to each well of the plate, replace the plastic seal, and centrifuge for 5 minutes at 4000 g.
67. Remove the plastic seal and shake the plate upside down over a sink hard 3 - 4 times to remove the wash supernatant.
68. Leave the pelleted, labeled cells on ice until right before reading each individual sample by flow cytometry.
69. Immediately before reading on the flow cytometer, Re-suspend a column of 8 cell pellets with 40 uL PBSF/well using a multichannel pipettor and transfer to a clean PCR tube strip. This allows you to quickly read 8 samples one after the other, greatly speeding up throughput during reading on the flow cytometer.

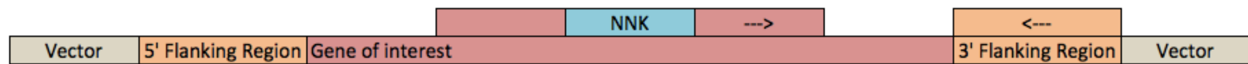
70. Be sure to also read an unlabeled sample and a sample labeled with only FITC-conjugated Chicken-anti-Myc Antibody to set display gates and compensation values on the flow cytometer.

10.3 Overlap PCR mutagenesis and assembly

This protocol has been adapted from Procko, et al. and is illustrated in the schematic below (129). In one PCR reaction, a long internal site-specific forward primer containing a degenerate NNK codon and a common reverse primer are used to amplify a C-terminal fragment of the gene of interest from the mutation site to the C-terminus of the gene. In a separate, concurrent PCR reaction, a short internal site-specific reverse primer and a common forward primer are used to amplify a second, N-terminal fragment of the gene from the N-terminus to just before the mutation site. In a third PCR reaction, the N-terminal fragment and C-terminal fragment are joined and further amplified using common forward and reverse primers that flank the gene of interest. The product is the full-length gene with an NNK degenerate codon at the desired position. This process is carried out in parallel for every residue to be mutated in the gene. The joined products from the 3rd PCR reaction are pooled and purified through gel extraction. The mutagenic genes fragments are then electroporated along with linearized vector into yeast. The NNK codon is only contained in the long internal site-specific forward primer. The forward and reverse internal primers for each position to be mutated are automatically designed from the parent gene sequence using a Python script developed by Eva Strauch (Ref for Eva's script). The assembly protocol is a modified version of a protocol developed by Eva Strauch.

Input Plasmid DNA

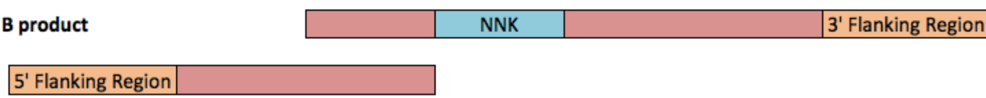
PCR reaction A



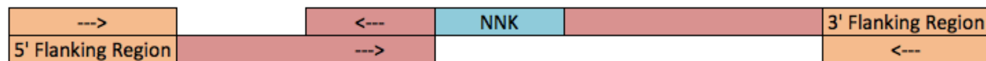
PCR reaction B

PCR reaction A product

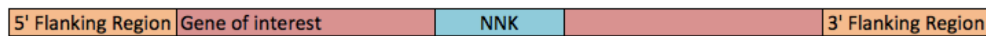
PCR reaction B product



Joining PCR reaction



Joining PCR reaction product



A detailed protocol is given below:

22. Internal primers are synthesized by a commercial vendor (IDT, San Jose, California, USA) and shipped in a 384-well plate in 150 uL sterile water at a concentration of 20 uM.
23. Purify gene of interest (in pETCON) from EBY100 yeast using the Yeast Plasmid Miniprep II Kit (Zymo Research, Irvine, California, USA) according the vendor protocol.
24. Transform the miniprepped gene of interest into chemically competent XL1-Blue E. coli cells (Agilent Technologies, Santa Clara, California, USA), plate on LB agar supplemented with 100 mg/mL carbenicillin, and incubate the plate at 37°C for 16 hours.
25. Pick 4 colonies from the plate and grow in 3 mL LB media supplemented with 100 mg/mL carbenicillin for 16 hours at 37°C.

26. Extract the amplified plasmid DNA from the cell using the QIAprep spin miniprep kit (Qiagen, Venlo, Limburg, Netherlands) according to the vendor protocol.
27. Sanger sequence the miniprepped plasmid DNA and select a preparation having no mutations and a clean sequencing trace.
28. Make up the C-terminal and N-terminal PCR reaction master mixes according to the following recipes:

Component:	For EB15 C-term:	For EB15 N-term:	For EB22 C-term:	For EB22 N-term:
Sterile water	2600 uL	2600 uL	2045 uL	2045 uL
5X Phusion HF Buffer	750 uL	750 uL	590 uL	590 uL
28 ng/uL EB15	13.4 uL	13.4 uL	10.5 uL	10.5 uL
100 uM COR primer	30 uL	---	23.6 uL	---
100 uM COF primer	---	30 uL	---	23.6 uL
10 mM dNTPs	187.5 uL	187.5 uL	147.5 uL	147.5 uL
2U/uL Phusion Pol	18.8 uL	18.8 uL	18.8 uL	18.8 uL
Total	3600 uL	3600 uL	2832 uL	2832 uL

29. Use the following primers as the common outer forward and reverse primers for genes in pETCON:

>colony_outer_forward (COF)

TGACAACTATATGCGAGCAAATCCCCTCAC

>colony_outer_reverse (COR)

GTACGAGCTAAAAGTACAGTGGGAAC

30. Add 24 uL of the appropriate PCR master mix to 1 uL of each forward or reverse internal primer in a well of a PCR plate and mix.
31. Cap the PCR plate with adhesive metallic foil, place in a thermocycler, and run the following PCR protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	30 seconds
3	66°C	30 seconds
4	72°C	1 minute
5	Go to Step 2	24 times
6	72°C	10 minutes
7	4°C	forever

32. Make up the joining PCR master mix according to the following recipe:

Component:	For 1 reaction:	For 271 reactions:
Sterile water	34.9 uL	9444 uL
5X Phusion HF Buffer	10 uL	2710 uL
100 uM COF primer	0.2 uL	54.2 uL
100 uM COR primer	0.2 uL	54.2 uL
10 mM dNTPs	2.5 uL	678 uL
2U/uL Phusion Pol	0.25 uL	67.8 uL
Total	48 uL	13008 uL

33. Make up the joining PCR master mix according to the following recipe:

Component:	For 1 reaction:	For 271 reactions:
Sterile water	34.9 uL	9444 uL
5X Phusion HF Buffer	10 uL	2710 uL
100 uM COF primer	0.2 uL	54.2 uL
100 uM COR priemr	0.2 uL	54.2 uL
10 mM dNTPs	2.5 uL	678 uL
2U/uL Phusion Pol	0.25 uL	67.8 uL
Total	48 uL	13008 uL

34. Add 1 uL of each forward PCR product and the corresponding reverse PCR product to 48 uL of the joining PCR master mix to a well of a PCR strip.

35. Cap the PCR strip, place in a thermocycler, and run the above PCR protocol.

36. Pour 256 lanes of 2% agarose gel (3 x 96-well gels) with SYBR safe gel stain (Life Technologies, Carlsbad, California, USA).

37. Run 5 uL of each joining PCR reaction with 1 uL of 6X DNA loading dye (New England Biolabs, Ipswich, Massachusetts, USA) on the agarose gel at 100 V.
38. Visualize the gel to validate that a PCR product of the correct size was obtained for each joining reaction.
39. Combine 4 uL of each EB15 joining PCR reaction, mix, combine with 6X DNA loading dye, and run on a new 2% agarose gel with SYBR safe gel stain. Do the same with 4 uL of each EB22 joining PCR reaction. For PCR reactions having less product, use 8 uL. Do not combine the EB15 and EB22 libraries together.
40. Cut out the gel band of the correct sizes for the EB15 and EB22 joining PCR products and extract the DNA fragments using the QIAquick Gel Extraction Kit (Qiagen, Venlo, Limburg, Netherlands) according to the vendor protocol and elute in Qiagen buffer EB.
41. Quantify the gel extracted DNA using a Qubit fluorimeter according to the manufacturers protocol.
42. Desalt the eluted DNA fragments by aliquoting the DNA solution onto nitrocellulose membranes suspended above a petri dish filled with sterile water and letting stand for 20 minutes at room temperature.
43. Electroporate the DNA fragments with linearlized pETCON vector into EBY100 yeast using the protocol of Benatuil, et al (Ref for Benatuil protocol).

10.4 Labeling EB15 and EB22 libraries for FACS

The EB15 and EB22 single site mutant and combinatorial libraries were simultaneously FACS (125, 127) sorted for a number of criteria, including maximal affinity for EED3, maximal on-rate, minimal off-rate, maximal stability, and minimal nonspecific binding. Variants with

maximal affinity were isolated by labeling the libraries with the lowest possible concentration of soluble biotinylated EED3 under non-avid conditions. Clones with maximal on-rate were isolated by labeling the libraries for short periods of time on ice. Clones with maximal stability were isolated by labeling the libraries for short lengths of time at 42°C. Note that the conditions for maximal stability and maximal on-rate conflict and could be used for the same library in a single sort round. The EB15 library was selected for maximal on-rate while the EB22 library was selected for maximal stability. Clones with minimal off-rate were isolated by labeling the libraries at saturating levels of EED3, washing away the unbound EED3, and incubating with a high concentration of soluble Ezh2 competitor for lengthy periods at 37°C. The Ezh2 competitor is intended to bind any molecules of EED3 that dissociate from the yeast surface and prevent their re-binding. Clones with minimal nonspecific binding were isolated by alternating between 3 different secondary fluorophores used for each round of sorting: streptavidin-R-phycoerythrin conjugate (SAPE), streptavidin-allophycocyanin-Alexa-fluor-750-conjugate (SA750), and neutravidin-R-phycoerythrin conjugate (NAPE)(Life Technologies, Carlsbad, CA, USA). A detailed protocol for inducing and labeling yeast libraries is given below:

1. Induce a sufficient number of yeast cells to achieve 10-fold coverage of your library diversity at each round of sorting by incubating the cells in either 1 mL CG-UT (128) or SGCAA (125) supplemented with 50 mg/mL kanamycin, 50 mg/mL streptomycin, and 100 mg/mL Carbenicillin overnight at 30°C. The yeast cells should be induced at a density of no greater than 2*10^7 cells/mL.
2. Pellet the induced cells by centrifugation for 30 seconds at 20000 g and discard the media.

3. Wash the induced cells by adding 900 uL PBS supplemented with 1 mg/mL BSA (PBSF), centrifuging again, and discarding the wash solution. Keep the induced, washed cells on ice.
4. Measure the absorbance of the cells at 600 nm and calculate the concentration of cells according to the following formula:

   1 absorbance unit = 2*10^7 cells/mL
5. Prepare the 1° labeling solutions according to the following table. The recipes given below are intended to label 1*10^7 cells from each of the EB15 and EB22 libraries at 30 pM EED3. The volume of 1° labeling solution must be large enough to provide at least 10 molar equivalents of EED3 for every displayed molecule in the surface of the yeast cells, assuming 5*10^4 molecules/cell for well-displaying cells and taking into account the fraction of cells displaying protein. The Ezh2 solution is only needed if executing the off-rate selection. The 30 pM solutions are each made up in 2 x 50 mL polypropylene tubes. All volumes are given in uL.

	EB15_EED3	EB22_EED3	SAPE Mix	Ezh2	
Component	**30 pM**	**30 pM**		**10 uM**	**Total needed**
1 uM EED3	2.99	2.99			5.99
1 mg/mL SAPE			2.63		2.63
44 uM Ezh2				49.0	49.0
1 mg/mL Fitc			6.67		6.67
100 mg/mL BSA			20.0	20.0	40
PBSF	99997	99997	171	131	200296
Total	100000	100000	200	200	

6. Add to an eppendorf tube a sufficient volume of cells from each library provide 10-fold coverage of the library diversity.
7. Centrifuge the cells and remove the supernatant.
8. Add the cells to the 1° labeling solution.

On-rate selection:

9. Incubate the cells on ice for 30 minutes.

Thermal stability selection:

10. Incubate the cells for 30 minutes at 42°C.

Off-rate selection:

11. Following the 1° labeling, wash the cells with 900 uL PBSF, centrifuge the cells at 20000 g for 30 seconds, and discard the wash solution.
12. Add the Ezh2 solution the cells and incubate for 4 to 20 hours on the tube rotator in the 37°C room.

2° labeling:

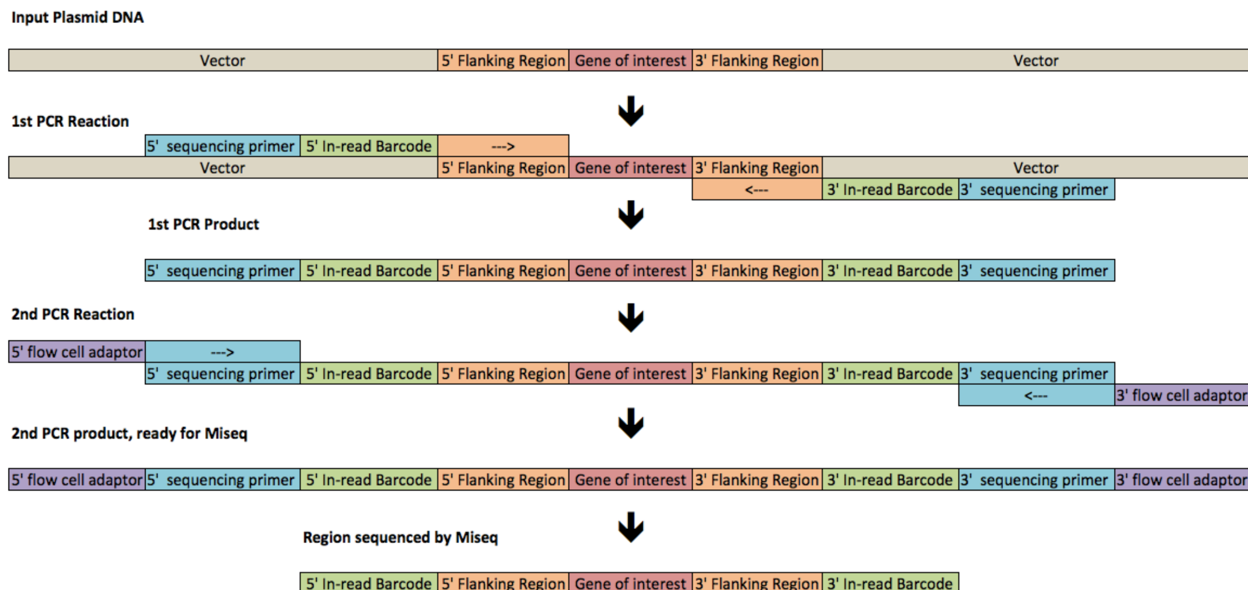
13. Following the off-rate selection incubation, wash the cells with 900 uL PBSF, centrifuge the cells at 20000 g for 30 seconds, and discard the wash solution.
14. Add 100 uL of the SAPE 2° solution to the cells from each library and incubate on ice for 30 minutes.
15. Following the 2° labeling, wash the cells with 900 uL PBSF, centrifuge the cells at 20000 g for 30 seconds, and discard the wash solution. Save the labeled cells as pellets on ice until ready to sort.
16. For each sample to be sorted, assemble a 25 mL Econo-Pac Chromatography Column (Bio Rad, Hercules, California, USA) by placing the upper frit into the column and pushing it down until it sits flush on top of the lower frit.
17. Equilibrate the frits by running 5 mL of PBSF over them.
18. Immediately before sorting each library, re-suspend the cell pellet in 2 mL PBSF and vortex the cells for 10 seconds at full power.

19. Apply the 2 mL of resuspended cells to the frits and collect the flowthrough in a FACS tube. The polyethylene sinter frits have a nominal pore size of 30 um and this step is intended to remove any particulate matter that could clog the 70 um opening on the FACS nozzle tip.
20. Add another 1 mL of PBSF to the frits to wash out any remaining cells and collect it in the same tube.
21. Place the labeled, filtered cells on the FACS machine and sort them according to established procedures.

10.5 Extraction and high-throughput sequencing of EB15 and EB22 gene pools from yeast

We extracted the plasmid DNA from cells saved from the unselected library and from successive rounds of FACS selection. We amplified the EB15 and EB22 genes each as single fragments. We sequenced the resulting gene fragments on an Illumina MiSeq v2 (Illumina, San Diego, CA, USA) with paired-end 2 x 259 base pair reads, obtaining roughly 4 million paired-end reads after filtering. This extraction method is a modification of a method developed by Aaron Chevalier and Tim Whitehead. Portions of this procedure originally come from the Andy Scharenberg and Maitreya Dunham labs at the University of Washington. The protocol produces about 100 ng template per 1*10^7 yeast cells containing low-copy number plasmid with a Cen6 ori, which quantification shows gives 1-3 plasmid copies per cell. Briefly, 1*10^7 yeast cells are stored as pellets in -80°C freezer. The cell wall is removed using zymolase and the cells are lysed by a freeze-thaw cycle followed by an alkaline lysis miniprep procedure. Sheared genomic DNA and ssDNA are partially cleaned up from the plasmid DNA by an exonuclease processing step and then a PCR step amplifies the gene and appends pool-specific barcodes. A 2nd PCR step

appends Ilumina flow-cell adaptors. PCR products are purified by extraction from an agarose gel and quantified using a Qubit fluorimeter. The order of the PCR steps and structure of the template, primers, and product is shown the following schematic. Note that the relative length of DNA regions is not to scale.



A detailed protocol is given below:

1. Yeast Miniprep (protocol developed in the Andy Scharenberg lab):

Uses the Yeast Plasmid Miniprep II Kit (Zymo Research, Irvine, California, USA) and the QIAprep Spin Miniprep Kit (Qiagen, Venlo, Limburg, Netherlands).

16. Pellet 4*10^7 yeast cells by centrifugation at 20000 g for 30 seconds.
17. Resuspend the cells in 200 uL Yeast Plasmid Miniprep II Kit Solution 1.
18. Add 5µl Zymolyase (5U/uL) to the cells
19. Incubate the cells at 37°C for 4 hours, mixing once per hour.
20. Perform 1 freeze-thaw cycle in a dry ice/EtOH bath and at 42°C.
21. Add 200µl Yeast Plasmid Miniprep II Kit Solution 2, mix well, and let sit 3-5 min.

22. Add 400μl Yeast Plasmid Miniprep II Kit Solution 3, mix well, and centrifuge at 20000 g for 5 min (I generally have to remove supernatant and spin a second time to remove all of the flocculent).
23. Transfer the supernatant to a QIAprep Spin Miniprep Kit column.
24. Spin the column for 1 minute at 20000 g.
25. Add 700 uL QIAprep Spin Miniprep Kit PB buffer and spin for 30 seconds at 20000 g.
26. Add 700 uL QIAprep Spin Miniprep Kit PE buffer and spin for 30 seconds at 20000 g.
27. Repeat the PE wash.
28. Pour out supernatant and spin for 1 minute at 20000 g to dry out the column.
29. Add 32 uL sterile water to elute, let the column stand for 1 minute, and spin for 1 minute at 20000 g.
30. Reload column with the eluate and spin again.

2. Further Purification of the yeast plasmid (protocol developed in the Maitreya Dunham lab):

The DNA prep is dirty and needs to be cleaned up before any large scale PCR amplification. This step is used to degrade any interfering genomic DNA. We saw inhibition with PCR and qPCR before adding this step. Store 15 uL of the elution from the previous step at -20°C and proceed with remaining 15 uL. This step requires Exonuclease I (Epicentre, Madison, Wisconsin, USA) that catalyzes the removal of nucleotides from single-stranded DNA in the 3' to 5' direction and Lamda exonuclease (New England Biolabs, Ipswich, Massachusetts, USA) that cleaves mononucleotides from duplex DNA 5' to 3'.

3. Make up the following recipe:

Component: Volume:

Miniprepped plasmid DNA	15 uL
ExoI	2 uL
Lambda	1 uL
Lambda buffer 10X	2 uL
Final volume	20 uL

4. On a thermocycler, incubate the mixture for 90 minutes at 30°C and then inactivate the enzymes for 20 min at 80°C.

3. Qiagen PCR clean-up:

Follow the standard kit procedure and elute in 32 uL sterile water. The yield will be 30 uL. Store 15 uL of the elution at -20°C and proceed with remaining 15 uL.

4. First large Scale PCR Amplification to append barcodes:

1. Phusion high fidelity polymerase (New England Biolabs, Ipswich, Massachusetts, USA) is used to amplify up the template and add the pool-specific barcodes.
2. Make up the 1st PCR master mix as follows:

Component:	For 1 reaction:	For 8 reactions:
Pasmid DNA from PCR cleanup	15 uL	---
5X Phusion HF Buffer	10 uL	90µl
5 uM Fwd primer	2 uL	---
5 uM Rev primer	2 uL	---
10 mM dNTPs	1 uL	9 uL
Sterile water	19.5 uL	175.5 uL
Phusion Polymerase	0.5 uL	4.5 uL
Total	50 uL	279 uL

3. Use the following amplification primers in the 1st PCR step to append unique barcodes to the genes from each experimental pool:

>EB15PCR_1_Fwd_0

ACGACGCTCTTCCGATCTACCAGGCGCTGGGGGTCGGCTAGCCATATG

>EB15_PCR_1_Fwd_1

ACGACGCTCTTCCGATCTGAGGCCTTGGCCCGGGTCGGCTAGCCATATG

>EB15_PCR_1_Fwd_2

ACGACGCTCTTCCGATCTCTTTAAAATATATCGGGTCGGCTAGCCATATG

>EB15_PCR_1_Fwd_3

ACGACGCTCTTCCGATCTTGACTTGCACATTTCGGGTCGGCTAGCCATATG

>EB22_PCR_1_Fwd_4

ACGACGCTCTTCCGATCTGGTCGCGGACCAGGGTCGGCTAGCCATATG

>EB22_PCR_1_Fwd_5

ACGACGCTCTTCCGATCTCCGGTTCCGGAGCGGGTCGGCTAGCCATATG

>EB22_PCR_1_Fwd_6

ACGACGCTCTTCCGATCTATATAAAATTTCTCGGGTCGGCTAGCCATATG

>EB22_PCR_1_Fwd_7

ACGACGCTCTTCCGATCTTACACGTTCAGTTTCGGGTCGGCTAGCCATATG

>EB15_PCR_1_Rev_8

ACGTGTGCTCTTCCGATCTTGGTCCGCGACCGATCCACCACCCTCGAG

>EB15_PCR_1_Rev_9

ACGTGTGCTCTTCCGATCTCTCCGGAACCGGTGATCCACCACCCTCGAG

>EB15_PCR_1_Rev_10

ACGTGTGCTCTTCCGATCTGAAATTTTATATATGATCCACCACCCTCGAG

>EB15_PCR_1_Rev_11

ACGTGTGCTCTTCCGATCTACTGAACGTGTAGATGATCCACCACCCTCGAG

>EB22_PCR_1_Rev_12

ACGTGTGCTCTTCCGATCTCCAGCGCCTGGTGATCCGCCACCCTCGAG

>EB22_PCR_1_Rev_13

ACGTGTGCTCTTCCGATCTGGCCAAGGCCTCTGATCCGCCACCCTCGAG

>EB22_PCR_1_Rev_14

ACGTGTGCTCTTCCGATCTTATATTTTAAAGATGATCCGCCACCCTCGAG

>EB22_PCR_1_Rev_15

ACGTGTGCTCTTCCGATCTATGTGCAAGTCAGATGATCCGCCACCCTCGAG

4. Add 31 uL of the master mix, 15 uL of the PCR cleanup product, and 2 uL each of 5 uM pool-specific forward and reverse primers to the wells of a PCR strip.
5. Mix the PCR reactions and cap the PCR strip.
6. Place the PCR strip in a thermocycler and run the following protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	30 seconds
3	66°C	30 seconds
4	72°C	1 minute
5	Go to Step 2	24 times
6	72°C	10 minutes
7	4°C	forever

7. Use 1 uL of the 1st PCR reaction as the template for the 2nd PCR reaction.

5. Second large Scale PCR Amplification to append Illumina adaptors:

12. Phusion high fidelity polymerase (New England Biolabs, Ipswich, Massachusetts, USA) is used to append the Illumina adaptor regions to the 1st PCR product.
13. Make up the 2nd PCR master mix as follows:

Component:	For 1 reaction:	For 8 reactions:
Pasmid DNA from PCR cleanup	15 uL	---
5X Phusion HF Buffer	10 uL	90µl
5 uM Fwd primer	2 uL	18 uL
5 uM Rev primer	2 uL	18 uL
10 mM dNTPs	1 uL	9 uL
sterile water	19.5 uL	301.5 uL
Phusion Polymerase	0.5 uL	4.5 uL
Total	50 uL	441 uL

14. Use the following amplification primers in the 2nd PCR step to append the Illumina adaptor regions to the 1st PCR product.

>Illum_PCR_2_Fwd

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA

TCT

>Illum_PCR_2_Rev

CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCG

ATCT

15. Add 31 uL of the master mix, 15 uL of the PCR cleanup product, and 2 uL each of 5 uM pool-specific forward and reverse primers to the wells of a PCR strip.

16. Mix the PCR reactions and cap the PCR strip.

17. Place the PCR strip in a thermocycler and run the following protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	30 seconds
3	66°C	30 seconds
4	72°C	1 minute
5	Go to Step 2	24 times
6	72°C	10 minutes
7	4°C	forever

18. Pour a 2% agarose gel with with SYBR safe gel stain (Life Technologies, Carlsbad, California, USA).
19. Add the entire completed 2nd PCR reaction to 10 uL of 6X DNA loading dye (New England Biolabs, Ipswich, Massachusetts, USA) and run on the gel at 100 V.
20. Visualize the DNA bands and use a razor blade to cut out the band corresponding in size to the amplified gene.
21. Extract the DNA from the gel slice using a QIAquick Gel Extraction Kit, following the manufacturer's protocol and eluting in 50 uL Qiagen Buffer EB.
22. Quantify the amplified DNA using a Qubit fluorimeter and the Qubit dsDNA HS Kit, following the manufacturer's protocol.

6. High-throughput sequencing using the Illumina Miseq:

Sequence the DNA from each of the pools on an Illumina Miseq using a paired-end run according to the manufacturer's protocol. The barcodes will be the first bases read during the forward and reverse reads. The Miseq must read equal amounts of all 4 bases during the first cycles to properly identify clusters and calibrate the instrument. The following barcode sets are an expanded version of a barcode set obtained from Dr. Rijhu Das at Stanford University:

Barcode_Set_A: (Forward)

Barcode_00: ACCAGGCGCTGG

Barcode _01: GAGGCCTTGGCC

Barcode _02: CTTTAAAATATA

Barcode _03: TGACTTGCACAT

Barcode_Set_B: (Reverve)

Barcode _04: GGTCGCGGACCA

Barcode _05: CCGGTTCCGGAG

Barcode _06: ATATAAAATTTC

Barcode _07: TACACGTTCAGT

Barcode_Set_C: (Forward complement)

Barcode _08: TGGTCCGCGACC

Barcode _09: CTCCGGAACCGG

Barcode _10: GAAATTTTATAT

Barcode _11: ACTGAACGTGTA

Barcode_Set_D: (Reverse complement)

Barcode _12: CCAGCGCCTGGT

Barcode _13: GGCCAAGGCCTC

Barcode _14: TATATTTTAAAG

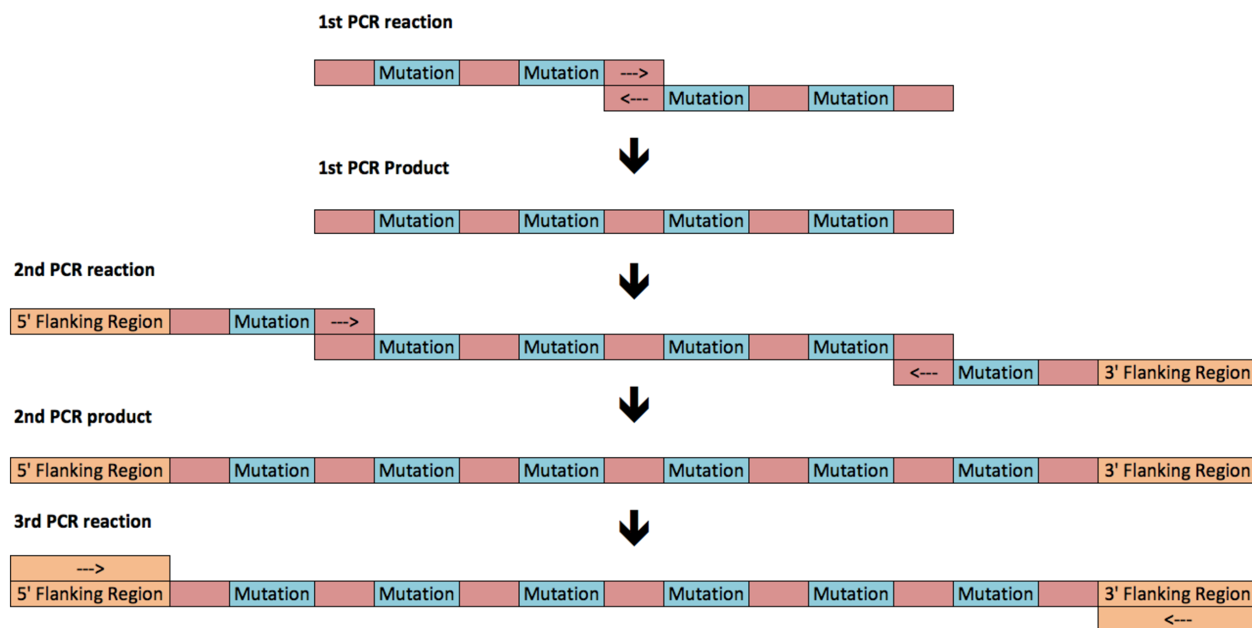
Barcode _15: ATGTGCAAGTCA

Each of the barcodes within a set of 4 should be used together in equal molar amounts to ensure that the Miseq will see equal amounts of each base at during each of the first 12 cycles of sequencing. As the barcodes are contained within the forward and reverse reads, there is no need to include an index read, freeing up more sequencing cycles for the forward and reverse reads. If there are more than 4 pools to sequence, simply use additional sets of barcodes. If the number of pools to sequence is not a factor of 4, use 2 barcodes per pool for larger pools until the number of barcodes is a factor of 4. Barcodes sets B, C, and D are simply the reverse, forward complement, and reverse complement of barcode set A, respectively. In order to avoid mispriming during PCR

amplification and hairpin formation within the PCR products, certain sets of barcodes should not be used together in the forward and reverse primers for a given pool during the PCR amplification steps. Barcode set A should be paired with set B and set C should be paired with set D. Barcode set A should not be paired with sets C or D, and so forth.

10.6 EB15 and EB22 combinatorial library construction via ultramer assembly

The combinatorial libraries were each constructed by assembling 4 ultramers that contained degenerate codons at 7-8 positions in the target gene and that included the flanking sequences from pETCON to allow for recombination into the vector. Each degenerate codon was carefully selected to code for the wild type amino acid plus a small number of target mutations while coding for as few off-target mutations as possible (see **Section 4.4** in the Appendix). The target diversity of the library was 5*10^5 theoretical variants. The ultramers used to assemble any given library were between 120 and 180 bases in length overlapped one another with ~ 30 base pair overlaps to allow for assembly. The library assembly strategy is given in the schematic below. The inner two ultramers were first assembled using overlap PCR and then gel purified. Next the outer ultramers were added to the gel purified inner fragment in a traditional PCR amplification step and the resulting product was also gel purified. The complete gene fragment was amplified with short outer primers and gel purified a third time in preparation for electroporation into yeast. The N-terminal 2 ultramers and amplification primer should be sense while the C-terminal 2 ultramers and amplification primer should be antisense in order for this to work.



A detailed protocol is given below:

1. Re-suspend 4 nmol of each ultramer in 400 uL of sterile water to give a final ultramer concentration of 10 uM.

2. For each library, make up the 1st PCR reactions in the wells of a PCR strip according to the following recipes. It's important to try a range of ultramer concentrations when assembling the inner ultramers to optimize purity and yield. Experience has shown that the optimal ultramer concentration varies from library to library.

Component:	20 nM	50 nM	100 nM	200 nM	500 nM	1000 nM
5X Phusion HF buffer	10 uL	10 uL	10 uL	10 uL	10 uL	10 uL
1 uM forward inner ultramer	1 uL	2.5 uL	5 uL	---	---	---
1 uM reverse inner ultramer	1 uL	2.5 uL	5 uL	---	---	---
10 uM forward inner ultramer	---	---	---	1 uL	2.5 uL	5 uL
10 uM reverse inner ultramer	---	---	---	1 uL	2.5 uL	5 uL
10 mM dNTPS	1 uL	1 uL	1 uL	1 uL	1 uL	1 uL
Sterile water	36.5 uL	33.5 uL	28.5 uL	36.5 uL	33.5 uL	28.5 uL
2U/uL Phusion Pol	0.5 uL	0.5 uL	0.5 uL	0.5 uL	0.5 uL	0.5 uL
Total	50 uL	50 uL	50 uL	50 uL	50 uL	50 uL

3. Place the PCR strip in a thermocycler and run the following PCR protocol:

Step:	Temperature:	Time:
1	98°C	3 minutes
2	98°C	30 seconds
3	70°C	30 seconds
4	72°C	1 minute
5	Go to step 2	4 times
6	72°C	10 minutes
7	4°C	forever

4. Run each entire 1st PCR reaction with 10 uL of 6X DNA loading dye (New England Biolabs, Ipswich, Massachusetts, USA) on a 2% agarose gel supplemented with SYBR safe gel stain (Life Technologies, Carlsbad, California, USA) at 100 V.
5. Visualize the gel and cut out the bands corresponding to the expected size of the inner ultramer assembly product. Choose the reaction condition(s) that give(s) the best yield of the desired product with a minimal amount of undesired products.
6. Extract the DNA fragments from the agarose gel using the QIAquick Gel Extraction Kit (Qiagen, Venlo, Limburg, Netherlands) according to the vendor protocol and elute in Qiagen buffer EB.
7. Quantify the gel extracted DNA using a Qubit fluorimeter according to the manufacturers protocol.
8. Make up the 2nd PCR reactions in the wells of a PCR strip according to the following recipes. In the case of EB15 and EB22, the products from the 200, 500, and 1000 nM conditions in the 1st PCR were each used in respective 2nd PCR reactions.

Component	EB15 200 nM	EB15 500 nM	EB15 1000 nM	EB22 200 nM	EB22 500 nM	EB22 1000 nM
5X Phusion HF buffer	10 uL	10 uL	10 uL	10 uL	10 uL	10 uL
10 uM forward outer ultramer	2.5 uL	2.5 uL	2.5 uL	2.5 uL	2.5 uL	2.5 uL
10 uM reverse outer ultramer	2.5 uL	2.5 uL	2.5 uL	2.5 uL	2.5 uL	2.5 uL
1st PCR product	0.5 uL	0.5 uL	0.5 uL	0.5 uL	0.5 uL	0.5 uL
10 mM dNTPS	1 uL	1 uL	1 uL	1 uL	1 uL	1 uL
Sterile water	36.5 uL	33.5 uL	28.5 uL	36.5 uL	33.5 uL	28.5 uL
2U/uL Phusion Pol	0.5 uL	0.5 uL	0.5 uL	0.5 uL	0.5 uL	0.5 uL
Total	50 uL	50 uL	50 uL	50 uL	50 uL	50 uL

9. Place the PCR strip in a thermocycler and run the following PCR protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	10 seconds
3	70°C	30 seconds
4	72°C	30 seconds
5	Go to step 2	24 times
6	72°C	10 minutes
7	4°C	forever

10. Run each entire 2nd PCR reaction with 10 uL of 6X DNA loading dye on a 2% agarose gel supplemented with SYBR safe gel stain at 100 V.

11. Visualize the gel and cut out the bands corresponding to the expected size of the inner ultramer assembly product. Choose the reaction condition that gives the best yield of the desired product with a minimal amount of undesired products.

12. Extract the DNA fragments from the agarose gel using the QIAquick Gel Extraction Kit (Qiagen, Venlo, Limburg, Netherlands) according to the vendor protocol and elute in Qiagen buffer EB.

13. Quantify the gel extracted DNA using a Qubit fluorimeter according to the manufacturers protocol.

14. Make up the 3rd PCR reaction master mix according to the following recipe. In the case of EB15 and EB22, the products from the 500 nM conditions of the 1st PCR gave the cleanest 2nd PCR products and were used in respective 3rd PCR reactions.

Component:	For 1 reaction:	For 10 reactions: (5 for each library)
5X Phusion HF buffer	10 uL	101 uL
100 uM forward amp primer	1.25 uL	13 uL
100 uM reverse amp primer	1.25 uL	13 uL
10 mM dNTPS	3 uL	30 uL
Sterile water	33 uL	333 uL
2U/uL Phusion Pol	0.5 uL	5 uL
Total	49 uL	495 uL

15. The amplification primers used for the 3rd PCR are given below:

>Fwd_amp

ATTGAAGGTAGATACCCATACGACGTTC

>Rev_amp

GTTGTTATCAGATCTCTATTACAAGTCCTCTTCAG

16. For each library, add 1 uL of the 2nd PCR product to 49 uL of the 3rd PCR master mix to each of 5 wells of a PCR strip, place the PCR strip in a thermocycler, and run the following PCR protocol:

Step:	Temperature:	Time:
1	98°C	30 seconds
2	98°C	10 seconds
3	68°C	30 seconds
4	72°C	30 seconds
5	Go to step 2	24 times
6	72°C	10 minutes
7	4°C	forever

17. Run each entire 3rd PCR reaction with 10 uL of 6X DNA loading dye on a 2% agarose gel supplemented with SYBR safe gel stain at 100 V.

18. 10. Visualize the gel and cut out the bands corresponding to the expected size of the inner ultramer assembly product. Choose the reaction condition that gives the best yield of the desired product with a minimal amount of undesired products.
19. Extract the DNA fragments from the agarose gel using the QIAquick Gel Extraction Kit (Qiagen, Venlo, Limburg, Netherlands) according to the vendor protocol and elute in Qiagen buffer EB.
20. Quantify the gel extracted DNA using a Qubit fluorimeter according to the manufacturers protocol.
21. Desalt the eluted DNA fragments by aliquoting the DNA solution onto nitrocellulose membranes suspended above a petri dish filled with sterile water and letting stand for 20 minutes at room temperature.
22. Electroporate the DNA fragments with linearlized pETCON vector into EBY100 yeast using the protocol of Benatuil, et al (Ref for Benatuil protocol).

10.7 Sub-cloning, expressing, and purifying top EB15 and EB22 combinatorial variants

Buffers Used:

Lysis Buffer: 20 mM HEPES, pH 7.5, 500 mM NaCl, 30 mM Imidazole, 1 mM Phenyl-methyl-sulfonyl-fluoride (PMSF), 1 mM Benzamadine, 1 mg/mL lysozyme, 1 m/mL deoxyribonuclease.

Imidazole Wash buffer: 20 mM HEPES, pH 7.5, 500 mM NaCl, 30 mM Imidazole

Salt Wash Buffer: 20 mM HEPES, pH 7.5, 1000 mM NaCl

Nickel Elution Buffer: 20 mM Tris, pH 7.5, 500 mM NaCl, 250 mM Imidazole

Desalting Buffer: 20 mM Tris, pH 7.5, 50 mM NaCl

Ion Exchange Elution Buffer: 20 mM Tris, pH 7.5, 1000 mM NaCl

1. Sub-cloning the gene of interest:

1. Use the Yeast colony PCR protocol and specific subcloning primers to amplify a Gibson-ready fragment of the gene of interest out of the pETCON vector contained within the cells of the appropriate yeast colony.
2. Use Gibson assembly to integrate 25 fmol the Gibson-ready fragment into 25 fmol of linearlized, gel-purified pET29b vector.
3. Transform 5 uL of the 20 uL Gibson reaction into 70 uL of chemically competent BL21Star cells (and plate on LB-Agar supplemented with 30 ug/mL Kanamycin.
4. Allow the plate to incubate at 37*C for 16 hours.
5. Use the E. coli colony PCR protocol and the T7_Fwd and T7_Rev primers to PCR the genes out of the pET29b vector contained within the cells of the appropriate E. coli colony.
6. Sanger sequence the purified PCR products using the T7_Fwd primer.
7. Select a colony that contains the gene of interest and lack deletions, insertions, and mutations.

2. Expressing the protein in E. coli:

For EB22 variants:

1. Inoculate 1 colony into 32 mL autoclaved LB media supplemented with 30 ug/mL Kanamycin in a 250 mL baffled Erlenmeyer flask.
2. Grow 16 hours at 37*C with shaking at 250 RPM.
3. Add 1 mL of the overnight culture to 600 uL of sterile filtered 50% glycerol in water in an eppendorf tube and mix well.
4. Snap freeze the eppendorf tube by immersion in liquid nitrogen and store at -80*C.

5. Add 28 mL of the overnight culture to 1 L of autoclaved TB media supplemented with 30 ug/mL Kanamycin in a 2 L baffled Erlenmeyer flask.
6. Grow about 3 hours at 37*C with shaking at 220 RPM in the refrigerated incubator-shaker until the O.D. reaches 0.6.
7. Once the O.D. reaches 0.6, turn the temperature down to 18*C and grow for another 30 minutes.
8. Add 130 uL of freshly-prepared 1 M isopropyl-thio-galacto-pyranoside (IPTG) to the 1 L culture to a final concentration of 0.13 mM to induce expression of the protein.
9. Grow the induced 1 L cultures for an additional 20 hours at 18*C with shaking at 220 RPM.

For EB15 variants:

1. Inoculate 1 colony into 32 mL autoclaved LB media supplemented with 30 ug/mL Kanamycin in a 250 mL baffled Erlenmeyer flask.
2. Grow 16 hours at 37*C with shaking at 250 RPM.
3. Add 1 mL of the overnight culture to 600 uL of sterile filtered 50% glycerol in water in an eppendorf tube and mix well.
4. Snap freeze the eppendorf tube by immersion in liquid nitrogen and store at -80*C.
5. Add 28 mL of the overnight culture to 1 L of autoclaved TB media supplemented with 30 ug/mL Kanamycin in a 2 L baffled Fernbach flask.
6. Grow about 3 hours at 37*C with shaking at 220 RPM in the refrigerated incubator-shaker until the O.D. reaches 0.9.
7. Add 150 uL of freshly-prepared 1 M isopropyl-thio-galacto-pyranoside (IPTG) to the 1 L culture to a final concentration of 0.15 mM to induce expression of the protein.

8. Grow the induced 1 L cultures for an additional 4.5 hours at 37*C with shaking at 220 RPM.

3. Extracting the expressed protein from the E. coli:

1. Collect the induced E. coli cells by centrifugation at 4000 g for 20 minutes at 4*C in a swing-bucket rotor.
2. Pour off the media and re-suspend the cell pellet in 15 mL of Imidazole Wash Buffer using a vortexer.
3. Transfer the resuspended cells to a 50 mL polypropylene tube and in the -80*C freezer for 30 minutes to 16 hours.
4. Thaw the frozen cell pellets in an ice bath with occasional vortexing.
5. Add to a final concentration of 1 mM, Phenyl-methyl-sulfonyl-fluoride (PMSF), 1 mM, benzamdine, 1 mg/mL, lysozyme, and 1 mg/mL, deoxyribonuclease (DNAse) and mix by inverting the tube.
6. Sonicate the thawed cell pellets using 15 cycles of 20 seconds on and 20 seconds off at 70% amplitude for a total of 5 minutes of sonication time. The tube should be placed in wet ice during sonication and the sonicator tip should be washed with ethanol and then water between successive samples.
7. Transfer the sonicated cells to an SS34 centrifuge and add Imidazole Wash Buffer to a final volume of no more than 35 mL. Ensure that successive pairs of tubes are balanced relative to each other.
8. Spin the cells for 40 minutes at 20500 g and 4*C in a fixed-angle rotor.
9. Filter the supernatant through a 0.45 uM syringe filter into a clean 50 mL polypropylene tube.

4. Purifying the extracted protein from the bacterial lysate:

These steps were completed at 4*C in a cold room.

1. Cap the bottom to an empty 30 mL plastic chromatography column and add 25 mL sterile water.
2. Add 2 mL of well-mixed Qiagen Nickel-NTA-Superflow resin slurry to the sterile water in the column.
3. Allow the resin to settle to the bottom of the column.
4. Add a frit to the top of the column and ensure that it is level and that there is no air in or under it.
5. Push the frit down unto the resin using a 5 mL serological pipette, taking care not to compress the resin to any degree.
6. Remove the bottom cap and allow the sterile water to drain out of the bottom end of the column.
7. Add 5 mL of Imidazole Wash Buffer to the top of the column and allow it to flow through the nickel resin and out of the bottom of the column.
8. Add the filtered supernatant to the top of the column and allow it to flow through the nickel resin and out of the bottom of the column. Repeat this step 2 times more for a total of 3 times over the column.
9. Add 25 mL of Imidazole Wash Buffer to the top of the column and allow it to flow through the nickel resin and out of the bottom of the column.
10. Add 30 mL of Salt Wash Buffer to the top of the column and allow it to flow through the nickel resin and out of the bottom of the column.

11. Add another 25 mL of Imidazole Wash Buffer to the top of the column and allow it to flow through the nickel resin and out of the bottom of the column.
12. Add 1 mL of Nickel Elution Buffer to the top of the column and allow it to flow through the nickel resin and out of the bottom of the column into a clean collection tube.
13. Let the column stand for 5 minutes.
14. Add another 1.5 mL of Nickel Elution Buffer to the top of the column and allow it to flow through the nickel resin and out of the bottom of the column into the same collection tube.
15. Desalt the 2.5 mL of protein eluted from the nickel column into Desalting Buffer using a PD-10 desalting column and following the vendor protocol. The PD-10 desalting columns can be reused if washed with 50 mL of water and re-equilibrated with 25 mL of desalting buffer.
16. Quantify the protein using the absorbance at 280 nm on the Nanodrop spectophopmeter (Thermo Scientific, Waltham, Massachusetts, USA).
17. Run 15 uL of the protein on a 4-20% gradient Tris-Glycine SDS PAGE gel and estimate the purity of the protein prep

Section 11

References

11.1 References for the introduction

1. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA, Baker D. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. **Science**. 2011 May 13;332(6031):816-21.

2. Fleishman SJ, et al. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. **J Mol Biol**. 2011 Nov 25;414(2):289-302.

3. Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, Baker D. Principles for designing ideal protein structures. **Nature**. 2012 Nov 8;491(7423):222-7. doi: 10.1038/nature11600.

4. Bonneau R, Strauss CE, Baker D. Improving the performance of Rosetta using multiple sequence alignment information and global measures of hydrophobic core formation. **Proteins**. 2001 Apr 1;43(1):1-11.

5. Roberts RW. Totally in vitro protein selection using mRNA-protein fusions and ribosome display**. Curr Opin Chem Biol**. 1999 Jun;3(3):268-73. Review.

6. Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH. Evolving strategies for enzyme engineering. **Curr Opin Struct Biol**. 2005 Aug;15(4):447-52. Review.

7. Seelig B. mRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. **Nat Protoc**. 2011 Apr;6(4):540-52.

8. Seelig B, Szostak JW. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. **Nature**. 2007 Aug 16;448(7155):828-31.

9. King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, André I, Gonen T, Yeates TO, Baker D. Computational design of self-assembling protein nanomaterials with atomic level accuracy. **Science**. 2012 Jun 1;336(6085):1171-4.

10. King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, Yeates TO, Baker D. Accurate design of co-assembling multi-component protein nanomaterials. **Nature**. 2014 Jun 5;510(7503):103-8.

11. Tinberg CE, et al. Computational design of ligand-binding proteins with high affinity and selectivity. **Nature**. 2013 Sep 12;501(7466):212-6.

12. Procko E, et al. A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. **Cell**. 2014 Jun 19;157(7):1644-56.

13. Strauch EM, Fleishman SJ, Baker D. Computational design of a pH-sensitive IgG binding protein. **Proc Natl Acad Sci** U S A. 2014 Jan 14;111(2):675-80.

14. Mills JH, et al. Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy. **J Am Chem Soc**. 2013 Sep 11;135(36):13393-9.

15. Procko E, et al. Computational design of a protein-based enzyme inhibitor. **J Mol Biol**. 2013 Sep 23;425(18):3563-75.

16. Fleishman SJ, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. **Science**. 2011 May 13;332(6031):816-21.

17. Correia BE, et al. Proof of principle for epitope-focused vaccine design. **Nature**. 2014 Mar 13;507(7491):201-6.

18. Röthlisberger D, et al. Kemp elimination catalysts by computational enzyme design. **Nature**. 2008 May 8;453(7192):190-5.

19. Siegel JB, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. **Science**. 2010 Jul 16;329(5989):309-13.

20. Jiang L, et al. De novo computational design of retro-aldol enzymes. **Science**. 2008 Mar 7;319(5868):1387-91.

21. Bjelic S, et al. Computational design of enone-binding proteins with catalytic activity for the Morita-Baylis-Hillman reaction. **ACS Chem Biol**. 2013 Apr 19;8(4):749-57.

22. Richter F, et al. Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. **J Am Chem Soc**. 2012 Oct 3;134(39):16197-206.

23. Khare SD, et al. Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis. **Nat Chem Biol**. 2012 Feb 5;8(3):294-300.

24. Cherny I, et al. Engineering V-type nerve agents detoxifying enzymes using computationally focused libraries. **ACS Chem Biol**. 2013 Nov 15;8(11):2394-403.

25. Rajagopalan S, et al. Design of activated serine-containing catalytic triads with atomic-level accuracy. **Nat Chem Biol**. 2014 May;10(5):386-91.

11.2 References for the design of proteins to inhibit the interaction of Mdmx-p53

26. Vousden KH1, Lu X. Live or let die: the cell's response to p53. **Nat Rev Cancer**. 2002 Aug;2(8):594-604.

27. Oliner JD, Kinzler KW, Meltzer PS, George DL, and Vogelstein B. Amplification of a gene encoding a p53-associated protein in human sarcomas. **Nature** (1992) 358, 80-83.

28. Danovi D, et al. Amplification of Mdmx (or Mdm4) directly contributes to tumor formation by inhibiting p53 tumor suppressor activity. **Mol Cell Biol**. 2004 Jul;24(13):5835-43.

29. Laurie NA, et al. Inactivation of the p53 pathway in retinoblastoma. **Nature**. 2006 Nov 2;444(7115):61-6.

30. Popowicz GM, Czarna A, Wolf S, Wang K, Wang W, Dömling A, Holak TA. Structures of low molecular weight inhibitors bound to MDMX and MDM2 reveal new approaches for p53-MDMX/MDM2 antagonist drug discovery. **Cell Cycle**. 2010 Mar 15; 9(6):1104-11.

31. Blackburn TJ, et al. Diaryl- and triaryl-pyrrole derivatives: inhibitors of the MDM2-p53 and MDMX-p53 protein-protein interactions. **Medchemcomm**. 2013 Sep 21;4(9):1297-1304.

32. Zak K, et al. Mdm2 and MdmX inhibitors for the treatment of cancer: a patent review (2011-present). **Expert Opin Ther Pat**. 2013 Apr;23(4):425-48.

33. Graves B, et al. Activation of the p53 pathway by small-molecule-induced MDM2 and MDMX dimerization. **Proc Natl Acad Sci** U S A. 2012 Jul 17;109(29):11788-93.

34. Chang YS, et al. Stapled α-helical peptide drug development: a potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. **Proc Natl Acad Sci** U S A. 2013 Sep 3;110(36):E3445-54.

35. Yu GW, Vaysburd M, Allen MD, Settanni G, Fersht AR. Structure of human MDM4 N-terminal domain bound to a single-domain antibody. (2009) **J.Mol.Biol**. 385: 1578

36. Popowicz GM, Czarna A, Rothweiler U, Szwagierczak A, Krajewski M, Weber L, Holak TA. Molecular basis for the inhibition of p53 by Mdmx. (2007) **Cell Cycle** 6: 2386-2392

37. Czarna A, Popowicz GM, Pecak A, Wolf S, Dubin G, Holak TA. High affinity interaction of the p53 peptide-analogue with human Mdm2 and Mdmx (2009) **Cell Cycle** 8: 1176-1184

38. Phan J, et al. Structure-based design of high affinity peptides inhibiting the interaction of p53 with MDM2 and MDMX. (2010) **J.Biol.Chem.** 285: 2174-2183

39. Li C, et al. Systematic mutational analysis of peptide inhibition of the p53-MDM2/MDMX interactions. (2010) **J.Mol.Biol**. 398: 200-213

40. Pazgier M, et al. Structural basis for high-affinity peptide inhibition of p53 interactions with MDM2 and MDMX. (2009) **Proc.Natl.Acad.Sci**. USA 106: 4665-4670

41. Kallen J, et al. Crystal Structures of Human MdmX (HdmX) in Complex with p53 Peptide Analogues Reveal Surprising Conformational Changes. (2009) **J.Biol.Chem**. 284: 8812-8821

42. Popowicz GM, Czarna A, Holak TA. Structure of the human Mdmx protein bound to the p53 tumor suppressor transactivation domain. (2008) **Cell Cycle** 7: 2441-2443

43. Zhang Z, et al. Discovery of potent and selective spiroindolinone MDM2 inhibitor, RO8994, for cancer therapy. **Bioorg Med Chem**. 2014 Aug 1;22(15):4001-9.

44. Huang X, Yu M, Rew Y, Sun, D, Berger MD, Medina JC, Olson SJ. Discovery of Potent and Simplified Piperidinone-Based Inhibitors of the MDM2-p53 Interaction (2014) **Acs Med.Chem.Lett.** (in press)

45. Zhang Z, et al. Discovery of Potent and Orally Active p53-MDM2 Inhibitors RO5353 and RO2468 for Potential Clinical Development. (2014) **Acs Med.Chem.Lett**. 5: 124-127

46. Vaupel A, et al. Tetra-substituted imidazoles as a new class of inhibitors of the p53-MDM2 interaction. (2014) **Bioorg.Med.Chem.Lett**. 24: 2110-2114

47. Gonzalez AZ, et al. Novel Inhibitors of the MDM2-p53 Interaction Featuring Hydrogen Bond Acceptors as Carboxylic Acid Isosteres. (2014) **J.Med.Chem**. 57: 2963-2988

48. Gonzalez AZ, et al. Selective and Potent Morpholinone Inhibitors of the MDM2-p53 Protein-Protein Interaction. (2014) **J.Med.Chem**. 57: 2472-2488

49. Sun D, et al. Discovery of AMG 232, a Potent, Selective, and Orally Bioavailable MDM2-p53 Inhibitor in Clinical Development. (2014) **J.Med.Chem**. 57: 1454-1472

50. Bista M, et al. Transient protein states in designing inhibitors of the MDM2-p53 interaction. (2013) **Structure** 21: 2143-2151

51. Fry DC, et al. Deconstruction of a Nutlin: Dissecting the Binding Determinants of a Potent Protein-Protein Interaction Inhibitor. (2013) **Acs Med.Chem.Lett**. 4: 660-665

52. Anil B, Riedinger C, Endicott JA, Noble ME. The structure of an MDM2-Nutlin-3a complex solved by the use of a validated MDM2 surface-entropy reduction mutant. (2013) **Acta Crystallogr**.,Sect.D 69: 1358-1366

53. Ding Q, et al. Discovery of RG7388, a Potent and Selective p53-MDM2 Inhibitor in Clinical Development. (2013) **J.Med.Chem**. 56: 5979-5983

54. Miyazaki M, et al. Synthesis and evaluation of novel orally active p53-MDM2 interaction inhibitors. (2013) **Bioorg.Med.Chem**. 21: 4319-4331

55. Gonzalez-Lopez de Turiso F, et al. Rational design and binding mode duality of MDM2-p53 inhibitors. (2013) **J.Med.Chem**. 56: 4053-4070

56. Vu B, et al. Discovery of RG7112: A Small-Molecule MDM2 Inhibitor in Clinical Development. (2013) **Acs Med.Chem.Lett. (in press)**

57. Tovar C, et al. MDM2 small-molecule antagonist RG7112 activates p53 signaling and regresses human tumors in preclinical cancer models. (2013) **Cancer Res**. 73: 2587-2597

58. Miyazaki M, et al. Lead optimization of novel p53-MDM2 interaction inhibitors possessing dihydroimidazothiazole scaffold. **Bioorg.Med.Chem.Lett**. 23: 728-732

59. Zhan C, et al. An ultrahigh affinity d-peptide antagonist Of MDM2. (2012) **J.Med.Chem**. 55: 6237-6241

60. Rew Y, et al. Structure-based design of novel inhibitors of the MDM2-p53 interaction. (2012) **J.Med.Chem**. 55: 4936-4954

61. Furet P, Chene P, De Pover A, Valat TS, Lisztwan JH, Kallen J, Masuya K. The central valine concept provides an entry in a new class of non peptide inhibitors of the p53-MDM2 interaction. (2012) **Bioorg.Med.Chem.Lett**. 22: 3498-3502

62. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. (1996) **Science** 274: 948-953

63. Allen JG, et al. Discovery and optimization of chromenotriazolopyrimidines as potent inhibitors of the mouse double minute 2-tumor protein 53 protein-protein interaction. (2009) **J.Med.Chem**. 52: 7044-7053

64. Michelsen K, et al. Ordering of the N-terminus of human MDM2 by small molecule inhibitors. (2012) **J.Am.Chem.Soc**. 134: 17059-17067

65. Baek S, Kutchukian PS, Verdine GL, Huber R, Holak TA, Lee KW, Popowicz GM. Structure of the stapled p53 peptide bound to Mdm2. (2012) **J.Am.Chem.Soc**. 134: 103-106

66. Huang Y, Wolf S, Koes D, Popowicz GM, Camacho CJ, Holak TA, Domling A. Exhaustive fluorine scanning toward potent p53-Mdm2 antagonists. (2012) **Chem.med.chem** 7: 49-52

67. Liu M, et al. D-peptide inhibitors of the p53-MDM2 interaction for targeted molecular therapy of malignant neoplasms. (2010) **Proc.Natl.Acad.Sci**.USA 398: 200-213

68. Liu M, Pazgier M, Li C, Yuan W, Li C, Lu W. A left-handed solution to peptide inhibition of the p53-MDM2 interaction. (2010) **Angew.Chem.Int.Ed.Engl**. 49: 3649-3652

69. Li C, Pazgier M, Liu M, Lu WY, Lu W. Apamin as a template for structure-based rational design of potent peptide activators of p53. (2009) **Angew.Chem.Int.Ed.Engl**. 48: 8712-8715

70. Kallen J, et al. Crystal Structures of Human MdmX (HdmX) in Complex with p53 Peptide Analogues Reveal Surprising Conformational Changes. **J.Biol.Chem**. 284: 8812-8821

71. Popowicz GM, Czarna A, Rothweiler U, Szwagierczak A, Krajewski M, Weber L, Holak TA. Molecular basis for the inhibition of p53 by Mdmx. (2007) **Cell Cycle** 6: 2386-2392

72. Sakurai K, Schubert C, Kahne D. Crystallographic analysis of an 8-mer p53 peptide analogue complexed with MDM2. (2006) **J.Am.Chem.Soc**. 128: 11000-11001

73. Fasan R, Dias RL, Moehle K, Zerbe O, Obrecht D, Mittl PR, Robinson JA. Structure-activity studies in a family of beta-hairpin protein epitope mimetic inhibitors of the p53-HDM2 protein-protein interaction. (2006) **Chem.biochem** 7: 515-526

74. Grasberger BL, et al. Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. (2005) **J.Med.Chem**. 48: 909-912

75. Vassilev LT, et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. (2004) **Science** 303: 844-848

76. Ji Y, et al. In vivo activation of the p53 tumor suppressor pathway by an engineered cyclotide. (2013) **J.Am.Chem.Soc**. 135: 11623-11633

77. Uhrinova S, et al. Structure of free MDM2 N-terminal domain reveals conformational adjustments that accompany p53-binding. (2005) **J.Mol.Biol**. 350: 587-598

78. Fry DC, Emerson SD, Palme S, Vu BT, Liu CM, Podlaski F. NMR structure of a complex between MDM2 and a small molecule inhibitor. (2004) **J.Biomol.Nmr** 30: 163-173

79. Fleishman SJ, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. **Science**. 2011 May 13;332(6031):816-21.

80. Fleishman SJ, et al. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. **PLoS One**. 2011;6(6):e20161.

81. Leaver-Fay A, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. **Methods Enzymol**. 2011;487:545-74.

82. Cooper S, et al. Predicting protein structures with a multiplayer online game. **Nature**. 2010 Aug 5;466(7307):756-60.

83. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. **Nat Protoc**. 2006;1(2):755-68.

84. Walkinshaw MD, et al. Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. (2002) **Mol.Cell** 9: 187-194

85. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, Fields S. High-resolution mapping of protein sequence-function relationships. **Nat Methods**. 2010 Sep;7(9):741-6.

86. Whitehead TA, et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. **Nat Biotechnol**. 2012 May 27;30(6):543-8.

87. Kunkel TA. Rapid and efficient site-specific mutagenesis without phenotypic selection. **Proc Natl Acad Sci** U S A. 1985 Jan;82(2):488-92.

88. Hulett HR, Bonner WA, Barrett J, Herzenberg LA. Cell sorting: automated separation of mammalian cells as a function of intracellular fluorescence. **Science**. 1969 Nov 7;166(3906):747-9.

89. Van Dilla MA, Trujillo TT, Mullaney PF, Coulter JR. Cell microfluorometry: a method for rapid fluorescence measurement. **Science**. 1969 Mar 14;163(3872):1213-4.

90. Procko E, et al. A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. **Cell**. 2014 Jun 19;157(7):1644-56.

11.3 References for the design of proteins to inhibit the interaction of EED-Ezh2

91. Yamaguchi H1, Hung MC. Regulation and Role of EZH2 in Cancer. **Cancer Res Treat**. 2014 Jul;46(3):209-22.

92. Ciferri C, Lander GC, Maiolica A, Herzog F, Aebersold R, Nogales E. Molecular architecture of human polycomb repressive complex 2. **Elife** (Cambridge). 2012 Oct 30;1:e00005.

93. Han Z, Xing X, Hu M, Zhang Y, Liu P, Chai J. Structural basis of EZH2 recognition by EED. **Structure**. 2007 Oct;15(10):1306-15.

94. Kim W, Bird GH, Neff T, Guo G, Kerenyi MA, Walensky LD, Orkin SH. Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer. **Nat Chem Biol**. 2013 Oct;9(10):643-50.

95. Fleishman SJ, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. **Science**. 2011 May 13;332(6031):816-21.

96. Karanicolas J, et al. A de novo protein binding pair by computational design and directed evolution. **Mol Cell**. 2011 Apr 22;42(2):250-60.

97. Azoitei ML, et al. Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. **Science**. 2011 Oct 21;334(6054):373-6.

98. Azoitei ML, et al. Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. **J Mol Biol**. 2012 Jan 6;415(1):175-92.

99. Leaver-Fay A, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. **Methods Enzymol**. 2011;487:545-74.

100. Wilson C, Mau T, Weisgraber KH, Wardell MR, Mahley RW, Agard DA. Salt bridge relay triggers defective LDL receptor binding by a mutant apolipoprotein. (1994) **Structure** 2: 713-718

101. Margueron R, et al. Role of the polycomb protein EED in the propagation of repressive histone marks. (2009) **Nature** 461: 762-767

102. Xu C, et al. Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). (2010) **Proc.Natl.Acad.Sci**.USA 107: 19266-19271

103. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. **Nat Protoc**. 2006;1(2):755-68.

104. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, Fields S. High-resolution mapping of protein sequence-function relationships. **Nat Methods**. 2010 Sep;7(9):741-6.

105. Whitehead TA, et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. **Nat Biotechnol**. 2012 May 27;30(6):543-8.

106. Hulett HR, Bonner WA, Barrett J, Herzenberg LA. Cell sorting: automated separation of mammalian cells as a function of intracellular fluorescence. **Science**. 1969 Nov 7;166(3906):747-9.

107. Correia BE, et al. Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope. (2010) **Structure** 18: 1116-1126

11.4 References for computational methods

108. Leaver-Fay A, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. **Methods Enzymol**. 2011;487:545-74.

109. Cooper S, et al. Predicting protein structures with a multiplayer online game. **Nature**. 2010 Aug 5;466(7307):756-60.

110. Fleishman SJ, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. **Science**. 2011 May 13;332(6031):816-21.

111. Fleishman SJ, et al. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. **PLoS One**. 2011;6(6):e20161.

112. Phan J, et al. Structure-based design of high affinity peptides inhibiting the interaction of p53 with MDM2 and MDMX. (2010) **J.Biol.Chem**. 285: 2174-2183

113. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. (1996) **Science** 274: 948-953

114. Hoover DM, Lubkowski J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. **Nucleic Acids Res**. 2002 May 15;30(10):e43.

115. Bonneau R, Strauss CE, Baker D. Improving the performance of Rosetta using multiple sequence alignment information and global measures of hydrophobic core formation. **Proteins**. 2001 Apr 1;43(1):1-11.

116. Walkinshaw MD, et al. Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. (2002) **Mol.Cell** 9: 187-194

117. Fowler DM, Araya CL, Gerard W, Fields S. Enrich: software for analysis of protein function by enrichment and depletion of variants. **Bioinformatics**. 2011 Dec 15;27(24):3430-1.

118. Han Z, Xing X, Hu M, Zhang Y, Liu P, Chai J. Structural basis of EZH2 recognition by EED. **Structure**. 2007 Oct;15(10):1306-15.

119. Azoitei ML, et al. Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. **Science**. 2011 Oct 21;334(6054):373-6.

120. Azoitei ML, et al. Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. **J Mol Biol**. 2012 Jan 6;415(1):175-92.

121. Correia BE, et al. Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope. (2010) **Structure** 18: 1116-1126

122. Saikrishnan K, Kalapala, SK, Varshney U, Vijayan M. X-ray structural studies of Mycobacterium tuberculosis RRF and a comparative study of RRFs of known structure. Molecular plasticity and biological implications. (2005) **J.Mol.Biol**. 345: 29-38

123. Huang PS, Ban YE, Richter F, Andre I, Vernon R, Schief WR, Baker D. RosettaRemodel: a generalized framework for flexible backbone protein design. **PLoS One**. 2011;6(8):e24109.

124. Nivón LG, Bjelic S, King C, Baker D. Automating human intuition for protein design. **Proteins**. 2014 May;82(5):858-66.

11.5 References for experimental methods

125. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. **Nat Protoc**. 2006;1(2):755-68.

126. Kunkel TA. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci U S A. 1985 Jan;82(2):488-92.

127. Hulett HR, Bonner WA, Barrett J, Herzenberg LA. Cell sorting: automated separation of mammalian cells as a function of intracellular fluorescence. **Science**. 1969 Nov 7;166(3906):747-9.

128. Benatuil L, Perez JM, Belk J, Hsieh CM. An improved yeast transformation method for the generation of very large human antibody libraries. **Protein Eng Des Sel**. 2010 Apr;23(4):155-9.

129. Procko E, et al. Computational design of a protein-based enzyme inhibitor. **J Mol Biol**. 2013 Sep 23;425(18):3563-75.