Computational Design of Hyperstable, *De Novo* Miniproteins Targeting PD-1

Cassie M. Bryan

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

David Baker, Chair

Peter Brzovic

Wesley Van Voorhis

Program Authorized to Offer Degree:

Biochemistry

© Copyright 2017

Cassie M. Bryan

University of Washington

Abstract

Computational Design of Hyperstable, *De Novo* Miniproteins Targeting PD-1

Cassie M. Bryan

Chair of the Supervisory Committee:
Professor David Baker
Department of Biochemistry

Computational protein design has recently advanced to a new era with the *de novo* design of stable proteins targeting native protein ligands. In this dissertation, I will present the first *de novo* protein binder with an all beta interface targeting the T cell receptor, programmed cell death protein 1 (PD-1). Expressed on activated T cells, PD-1 inhibits T cell function and proliferation to prevent an excessive immune response. Tumor cells often take advantage of this pathway by over-expressing one of the ligands of PD-1, PD-L1 or PD-L2, to evade immune destruction. Additionally, impairment of the PD-1 pathway through a variety of mechanisms can lead to autoimmunity. Using a combination of computational design and experimental approaches, we have developed a *de novo* miniprotein that specifically binds PD-1 at the ligand interface. This protein binds murine PD-1 at a K_d of approximately 1 μM on yeast. The apo crystal structure shows that the binder folds as designed with a backbone RMSD of 1.3 Å to the design model. The 4.5 kDa protein proved to be very stable by chemical denaturation in GuHCl likely due to its three disulfide bonds. Over the years, I have identified several other binders using the canonical method of yeast surface display that ultimately had to be abandoned because of their inability to be produced as soluble proteins. I hypothesize this results from the use of the highly expressed native yeast protein Aga2p for display of the protein-of-interest (POI) on the cell surface. Here, I present a new method that replaces the Aga2p fusion with a minimal tag for covalent capture of the secreted protein and takes advantage of the natural yeast quality control pathways to discriminate between misfolded and well-folded proteins. This novel secretion capture system is a powerful tool for the screening, optimization, and production of well-expressed, functional proteins. Improved high-throughput methods for screening and optimizing stable, functional proteins enable

generation of *de novo* binders that are more readily amenable to a variety of applications. The small, hyperstable PD-1 binding domain presented here has potential use in a variety of cancer and autoimmune therapy platforms.

Acknowledgements

I have had the privilege of seeing the UW Institute for Protein Design (IPD) grow from its inception with the Baker lab moving into the Molecular Engineering & Sciences building to the highly collaborative environment that contains four labs and spans three buildings. This has given my graduate mentor, David Baker, an increasing amount of responsibility, and I appreciate all of the time he has taken to meet with me one-on-one. My project was quite a rollercoaster ride and David always had the optimism and foresight to keep me going.

Several people in the Baker lab and IPD have contributed to my two projects over the years and I'd like to give them a big thank you here. David Younger has worked with me on the Secretion Capture project from the beginning. We had several very helpful brainstorming sessions to generate new ideas and troubleshoot problems along the way and he has done a lot of the initial strain engineering and protein reagent cloning work. Stephanie Berger designed a stable, high affinity Bfl-1 binder that was used to test ligand binding in the SecCap system. Gabe Rocklin is a fantastic protein designer and not only designed the four hyperstable proteins I used to develop SecCap but also first designed my PD-1 binder, GR918. Matt Bick collected the diffraction data set for my protein crystal and taught me how to do the refinement of the final structure. Scott Boyken, George Ueda, Jorge Fallas, and Neil King have designed highly useful protein oligomers that I am using to make multivalent versions of my PD-1 binder. Inna Goreshnik transformed some of my yeast libraries. Lauren Carter, Stephen Rettie, and Alex Kang in the IPD Protein Core have produced and tested some of my proteins over the years. I'd also like to thank a few past members of the Baker lab – Matt Smith, Summer Thyme, Jeremy Mills, and Chris Bahl – that have been great mentors and fostered an excitement for protein design. It's difficult to thank every lab mate by name that has lent a hand or a word of advice over the years, but I'm grateful I was able to spend my graduate career in such an experienced lab.

I have been lucky to have great collaborators over the years. Steve Almo's lab at Albert Einstein College has sent me soluble PD-1 protein to use in my binding assays. Gauri Bhise in Mohamed Oukka's lab as well as Anne Stephens at the Seattle Children's Research Institute have been performing the T cell assays, and I look forward to continuing to work with them during this exciting time as we proceed with the functional testing of my binders.

Thanks to the UW Biochemistry department and Erin Kirschner, in particular, for creating a well-organized and intellectually stimulating environment that made my time in graduate school highly fulfilling. I want to give a big thank you to my Supervisory Committee – Roland Strong, Peter Brzovic, Alexey Merz, and Wes Van Voorhis. I was very lucky to be able to put together such a diverse committee covering a wide range of expertise. Not only did all four of them make almost every committee meeting, but they also contributed time to one-on-one meetings when I needed extra guidance. I want to give a special shout-out to Wes and Peter who contributed valuable feedback to this written dissertation.

My PhD work has been funded by several sources over the years. Initially, the PD-1 project was started by a grant with Steve Almo's lab, NIH 31537J AM-04, until June 2015. I was also supported through the Molecular Medicine Training Program for my first two years. In the recent years, I've been funded by a generous gift from two wonderful people, Rocky and Genie Higgins, whom I've had the pleasure getting to know personally. The excitement and genuine interest they took in me and my work gave me a special motivation to persevere through the many challenges I faced throughout my project.

Thanks to the many friends outside of work that helped maintain a level of fun during the long grad school road. My grad school cohort was quite small – only two others (Henry and YuRu) in my Biochemistry program and four (Amanda, Dan, James, and Shane) in the sister program BPSD – so we became very close. There was even a large portion of us that had birthdays around the same time, and the Epic March Birthday Bash became an annual time to catch up, especially in later years when we all got busy with our own projects. I'm glad we've all stayed in touch and wish them all the best.

Last but not least, I want to thank my family. My mom originally got me interested in science when she took me to libraries and bookstores to research the rare form of cancer she ultimately beat when I was in middle school. I know she didn't realize at the time that it would result in a scientist in the family but I will forever be grateful for her example and sacrifices for me. I particularly have loved having my brother so close. He'll often call me from the road to tell me he's coming to go to a concert with me and crash at my place for the weekend. I have really loved those random visits. My little nephew, Zeke, has been a great source of energy and smiles every chance I have to see him. I was fortunate to have my Aunt Julie living in Seattle while I was in grad school. I could not have maintained sanity without our regular happy hour venting sessions. When I was at the end of my second year in graduate school, I met my boyfriend Daniel

who is also a graduate student in another program. It's been so important having someone that could identify with the hardships of grad school, especially on nights and weekends when experiments kept me from getting home. I'm lucky that Daniel is a very talented computer scientist and has been super helpful and patient with me every time I bugged him with naïve programming questions. Thanks for sticking with me and can't wait to see what we do next!

Table of Contents

List of Figures

Acronyms and Abbreviations

APC	Antigen-presenting cell
BSA	Bovine serum albumin
CD	Circular dichroism spectroscopy
DDG	$\Delta\Delta G$, or binding energy
ERAD	ER-associated degradation pathway
FACS	Fluorescence-activated cell sorting
FITC	Fluorescein isothiocyanate
HEK	Human embryonic kidney
HT	High throughput
IgV	Immunoglobulin variable domain
IMAC	Immobilized metal affinity chromatography
K_d	Dissociation constant
MHC	Major histocompatibility complex
NGS	Next-generation sequencing
PBS	Phosphate-buffered saline
PBSF	Phosphate-buffered saline (PBS plus 1% BSA)
PD-1	Programmed cell death protein11
PD-L1	Programmed cell death protein-1 ligand 1
PD-L2	Programmed cell death protein-1 ligand 2
PE	Phycoerythrin
PEAR	Paired-end read merger
PEG	Polyethylene glycol
POI	Protein-of-interest
RFU	Relative fluorescence units
RMSD	Root mean square deviation
SA	Streptavidin

SASA	Solvent-accessible surface area
scFv	Single-chain variable fragment
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SSM	Site-saturation mutagenesis
TCR	T cell receptor
TIL	Tumor-infiltrating lymphocytes
T_m	Melting temperature

Introduction

Since its inception with the design of specificity switches[1] in 2004, the field of computational protein-protein interface design has advanced at a rapid rate. When I joined the Baker lab in 2012, the main protocol for the design of novel protein binders involved repurposing unrelated natural proteins by grafting in known interacting fragments from native complexes with the intended target. However, most natural proteins have not evolved to be incredibly stable and are not tolerant to surface mutations. This can lead to proteins that are successful in binding the intended target but have expression and stability problems. Additionally, this approach is limited by the number of folds that exist in nature whose structure has been solved to a high enough resolution to be tractable to design. These natural protein scaffolds are not necessarily well-suited to their new function, particularly for the case of protein-protein interactions where shape complementarity has been shown to be crucial to optimal binding[2]. The ability to customize backbones to fit each target has revolutionized the field of protein interface design.

Throughout my PhD, I have seen computational protein design advance from the reliance on native proteins as starting scaffolds to fully *de novo* design of fold and function at once. Several successes over the last decade demonstrate that proteins can now be generated from scratch with almost any fold and incredible stabilities never before observed in nature using the Rosetta software suite[3–10]. No longer limited to natural protein space, there are vast combinatorial possibilities of topologies and interface sizes to sample that can be tailored to individual targets. However, in order to test an increased number of designs, there also needed to be advancements in experimental screening techniques. New array-based oligonucleotide gene synthesis technologies[11,12] and multiplex assembly protocols[13] have enabled the cost effective generation of thousands of genes in a single pool. FACS binding selections and next-generation sequencing enables screening the whole pool in a single yeast surface display library[14]. Today, high-throughput design and testing of thousands of proteins simultaneously is the standard in computational protein-protein interface design.

The protein target I chose to focus on for my Ph.D. work is an immune receptor called Programmed Cell Death Protein-1, or PD-1, primarily expressed on activated T cells. T cells become activated when they first recognize an antigen presented through an MHC molecule on an antigen-presenting cell (APC). This

sends a stop signal through the T cell receptor (TCR) that stabilizes the immune synapse, allowing interaction of the costimulatory molecule CD28 with one of its ligands, B7-1 or B7-2, on the APC. A subsequent phosphorylation cascade activates the T cell, resulting in cell proliferation, cytokine production, differentiation into other mature T cell subtypes, and cytotoxicity (Figure 0.1A). Also after activation of the T cell, PD-1 is expressed on the surface of the cell. Upon engagement with one of its ligands, PD-L1 or PD-L2, the PD-1 molecules are clustered at the immune synapse. This receptor clustering recruits the phosphatase Shp2 that dephosphorylates the activation molecules and leads to T cell tolerance, exhaustion, apoptosis, and loss of effector function[15] (Figure 0.1B).

PD-1 consists of an extracellular Ig Variable binding domain, a transmembrane domain, and a cytoplasmic tail responsible for signaling by recruiting the Shp2 phosphatase[16]. PD-L1 and PD-L2 are also both IgV-type domains and have very similar backbone structures with an RMSD of 0.582 Å. Crystal structures of both the ligands in complex with PD-1 show that the interfaces are created by sheet-to-sheet contacts (Figure 0.2A-B). A large pocket on PD-1 is filled by nonpolar, aromatic residues on an edge strand of the ligands that was identified as the main binding fragment, and the rest of the ligand beta sheet interacts with a flat, hydrophobic surface on PD-1. The beta sheet interface in PD-L1 contains four strands while PD-L2 contains three and a very short fourth strand that curves around PD-1 giving this molecule a higher degree of shape complementarity as compared to PD-L1. This is reflected in the higher binding affinity of PD-L2 at 2.3 μM compared to PD-L1 at 8.2 μM[17]. The affinities of both 1:1 complexes are relatively weak as they rely on avidity for signaling. PD-1 is highly conserved across species. The murine and human orthologs are incredibly similar with a backbone RMSD of 1.3 Å, and 14 of the 17 binding residues are conserved (Figure 0.2C). The three variable positions interact exclusively with the motif residues of the ligands that are 100% conserved between these two species (Figure 0.2D). Human and murine PD-1 cross-react with both human and murine PD-L1 and PD-L2[18]. The high conservation of the PD-1 interface together with the cross-reactivity of the ligands suggests that a *de novo* protein binder could be engineered to bind both murine and human PD-1. This would be incredibly useful for animal studies in mice followed by development as a human therapeutic.

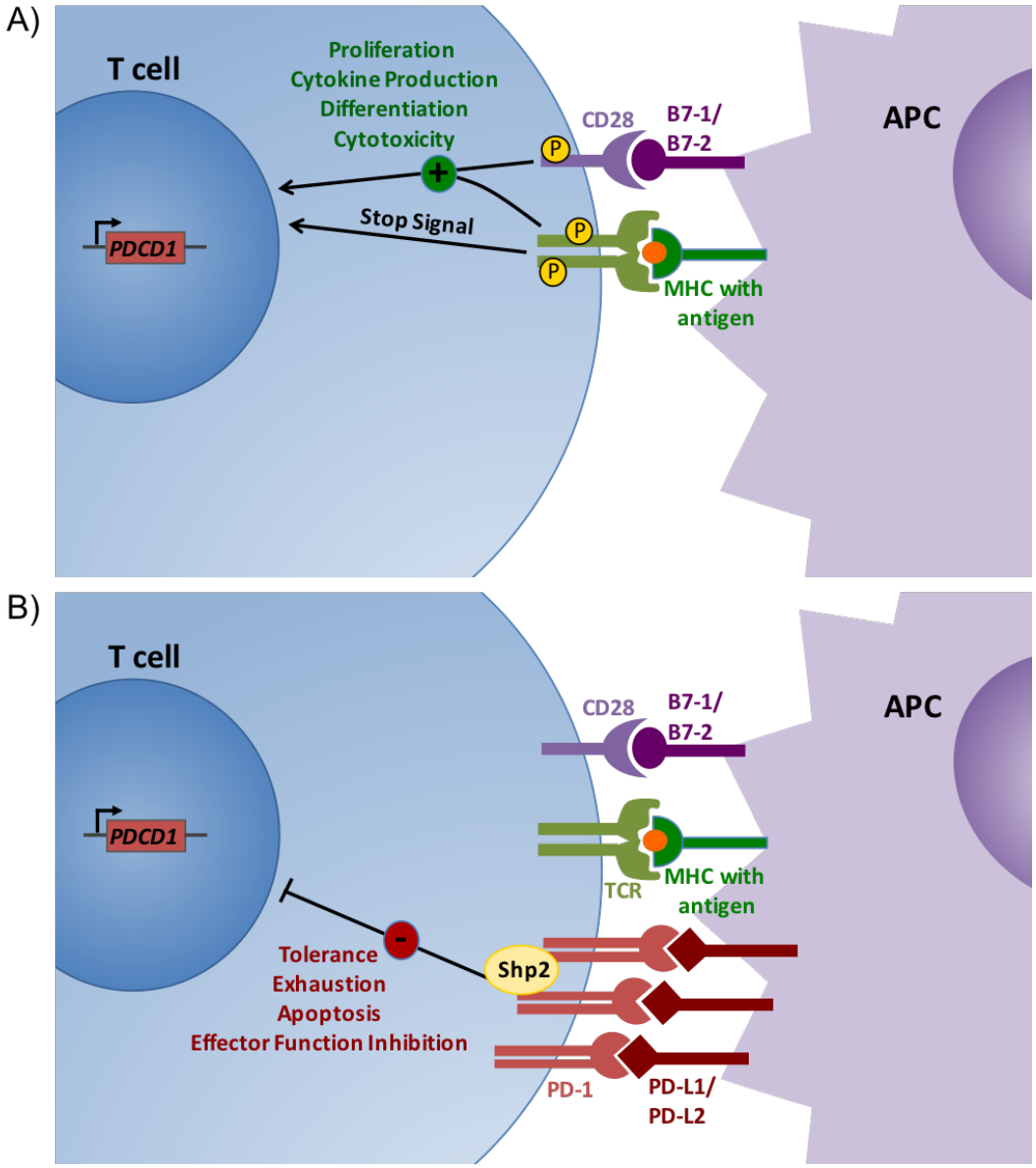

Figure 0.1: Role of PD-1 in T Cell Signaling. A) Naïve T cells become activated through a stop signal sent by interaction of the T cell receptor with an antigen displayed through an MHC molecule on an antigen presenting cell (APC) plus a costimulatory signal sent by the interaction of CD28 on the T cell with either B7-1 or B7-2 on the APC. A phosphorylation cascade causes activation of the T cell, resulting in cell proliferation, cytokine production, differentiation, and cytotoxicity. B) PD-1 is expressed on the surface of activated T cells and becomes clustered at the immune synapse upon engagement with one of its ligands, PD-L1 or PD-L2. The phosphatase Shp2 is recruited that dephosphorylates the activation molecules and deactivates the T cell, leading to tolerance, exhaustion, apoptosis, and effector function inhibition.

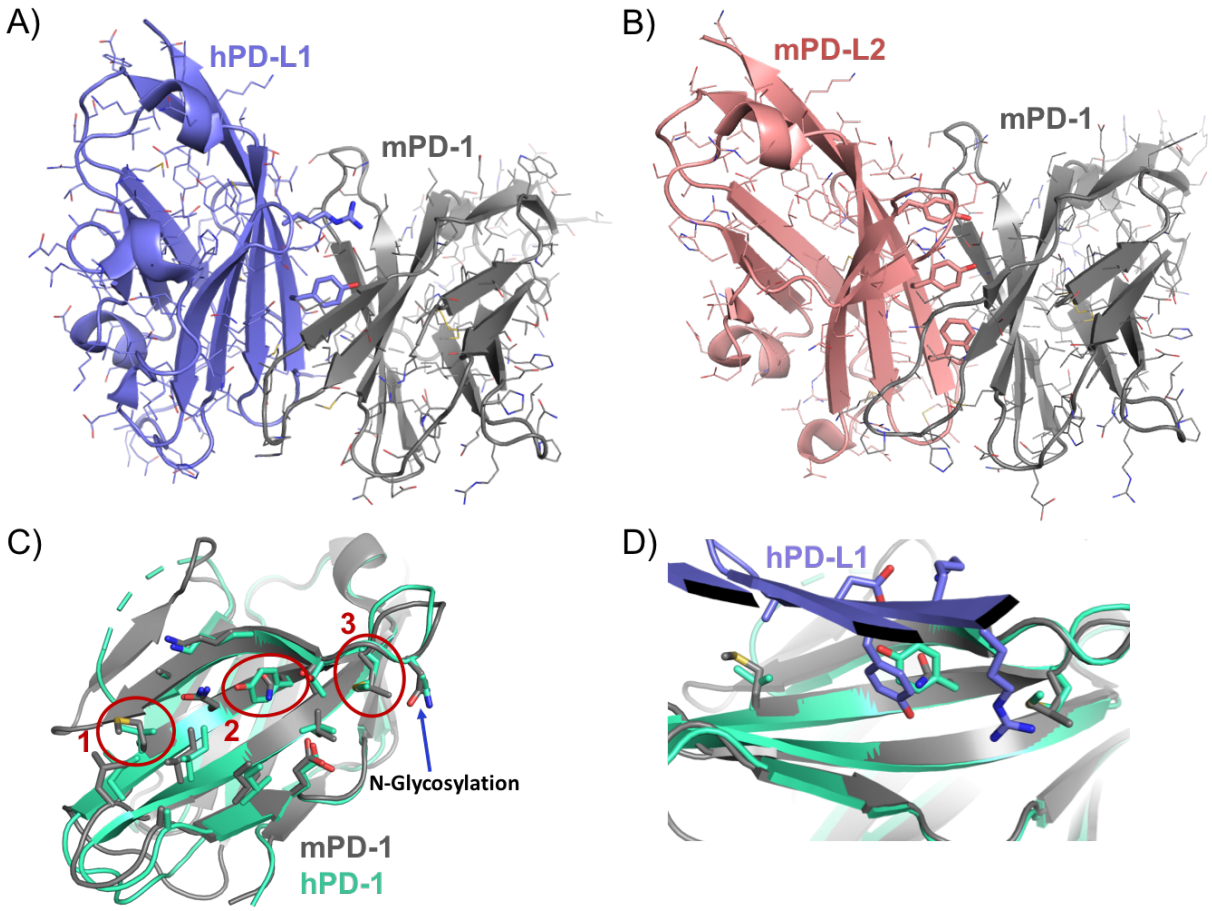

Figure 0.2: Structure of PD-1 and Its Ligands, PD-L1 and PD-L2. A) Crystal structure at 2.65 Å resolution of mPD-1 (gray) in complex with hPD-L1 (purple) (PDB 3BIK[19]). Both are IgV-type folds and the interface is formed by an edge strand fragment on PD-L1, "AYR" highlighted in sticks, that fills a pocket on PD-1 and the rest of the PD-L1 sheet interacts with a flat, hydrophobic surface on PD-1. B) Crystal structure at 1.80 Å resolution of mPD-1 (gray) in complex with mPD-L2 (pink) (PDB 3BP5[20]). PD-L2 has a very similar IgV fold and binding mode to PD-L1, and the edge strand binding motif in PD-L2 is "WYY" highlighted on the structure in sticks. C) Alignment of murine (gray) and human (green) orthologs of PD-1 with a backbone RMSD of 1.3 Å. The interface is completely conserved except for the three residues circled in red. The first position is a valine in hPD-1 and methionine in mPD-1, the second is tyrosine in hPD-1 and an asparagine in mPD-1, and the third is a methionine in hPD-1 and a leucine in mPD-1. There is also a conserved glycosylation site near the interface at the indicated asparagine. D) The three variable interface positions between mouse (gray) and human (green) PD-1 interact solely with the ligand binding motifs – shown here as hPD-L1 (purple) – that is completely conserved between mouse and human ligands.

As one of the major checkpoint pathways of T cells, PD-1 function is critical in gaining tolerance to self-antigens during immune system development and maintaining a healthy immune system by balancing T cell activation[21]. This pathway is exploited by cancer which overexpresses the ligands of PD-1, PD-L1 and PD-L2, to evade immune detection[22]. PD-1 also plays a role in chronic infection by converting T cells to an exhausted state if the pathogen persists[23]. Alternatively, malfunction of the pathway can lead to autoimmune symptoms[24]. The goal of my Ph.D. was to computationally design small, stable *de novo* proteins that would bind PD-1 at the ligand interface for potential use as cancer or autoimmune therapeutics.

Over the years, I have had several successful PD-1 binders as determined by yeast surface display but had difficulty obtaining soluble protein for *in vitro* characterization. This was likely because many of my binder designs contained very hydrophobic, all beta interfaces which led to trouble folding and susceptibility to oligomerization and aggregation. Unfortunately, the traditional yeast display system[25] does not discriminate well between proteins of different stabilities which leads to a lot of time wasted following up on false leads. This led me to my second project developing a new high-throughput method that could select for ligand binding and protein folding simultaneously. In the following chapters, I will describe the design of a hyperstable, *de novo* miniprotein that binds PD-1 specifically, as well as a novel secretion capture method for the screening, optimization, and production of proteins from yeast.

Chapter 1: Computational Design of a Hyperstable Miniprotein Targeting PD-1

INTRODUCTION

A healthy immune system is maintained by a dynamic and delicate balance of various stimulatory and inhibitory receptors that are responsible for activating an appropriate response when a foreign antigen is encountered and then returning the system to stasis to prevent a chronic, over-activation that could damage healthy tissue. Programmed cell death protein-1 (PD-1) is an inhibitory receptor that is expressed on most immune cells[26], however its function on activated T cells is most well-studied. Upon engagement with one of its ligands, programmed cell death-1 ligand 1 (PD-L1) or programmed cell death-1 ligand 2 (PD-L2), on the surface of another cell, PD-1 receptors cluster at the immune synapse. This clustering recruits the phosphatase Shp2 which dephosphorylates signaling molecules responsible for T cell activation. Ultimately T cell function is suppressed leading to decreased production of pro-inflammatory cytokines, inhibition of effector function, and apoptosis[15].

Many tumors use the PD-1 pathway to evade the immune system by over-expressing PD-L1 and/or PD-L2 to deactivate tumor-infiltrating lymphocytes (TILs). To date, two antibodies targeting PD-1 (nivolumab, pembrolizumab) and three antibodies targeting PD-L1 (atezolizumab, avelumab, durvalumab) have been approved by the FDA for the treatment of several types of cancer[27]. These antibodies, often referred to as checkpoint inhibitors, work by blocking the interaction between PD-1 and its ligands and thereby reactivating T cells to attack the tumor cells. However, full length antibodies often have low tissue penetrance due to their large size which can limit their effectiveness on some solid tumors[28]. A small PD-1 mimetic with increased affinity to PD-L1 was found to have increased tumor penetrance in mice compared to an anti-PD-L1 antibody[29], demonstrating that a small peptide may be beneficial for treatment of large, solid tumors. Currently, hundreds of combination therapies are being explored that combine anti-PD-1 and anti-PD-L1 antibodies with all types of radio-, chemo- and immunotherapies[30]. As relevant targets are identified, it may be beneficial to not only co-administer different therapeutic molecules, but also be able to combine them into a single molecule for multi-specific targeting of the tumor microenvironment. For this purpose, small proteins would be advantageous as modular binding domains that could be easily fused together using a linker or oligomeric scaffold[31]. Here, we present a fully de novo disulfide miniprotein[8] of 41

residues in length and containing 3 disulfide bonds that specifically binds murine PD-1 at a K_d of approximately 1μM on yeast.

RESULTS

Computational Design of *De Novo* PD-1 Binding Proteins

Two computational design protocols were employed to generate several thousand small designs for high-throughput testing. We first utilized the crystal structures of murine PD-1 in complex with human PD-L1[19] and murine PD-L2[20] to identify a five-residue beta-strand binding fragment responsible for the majority of the binding energy of these molecules. In both cases, this fragment was an edge strand containing mostly hydrophobic residues at positions 1, 3, and 5 that fit in a pocket on PD-1 and conserved polar residues at positions 2 and 4 that hydrogen bond with the backbone of the target. For the higher-affinity ligand, PD-L2, we used residues B110 through B114 from PDB 3BP5[20] which contained 'WDYKY' (Figure 1.1A). We used the PD-L1 positions A121-A125 containing 'ADYKR' from PDB 3BIK[19] (Figure 1.1B). While the tryptophan at the first position in the PD-L2 motif filled the pocket on PD-1 better than PD-L1's alanine at this position, PD-L1 seemed to have the more ideal amino acid at the fifth position due to arginine's more extensive contribution to a hydrogen bond network with the target. Therefore, in addition to the 'ADYKR' PD-L1 binding motif and the 'WDYKY' PD-L2 binding motif, we also tried using a hybrid motif of 'WDYKR' (Figure 1.1C).

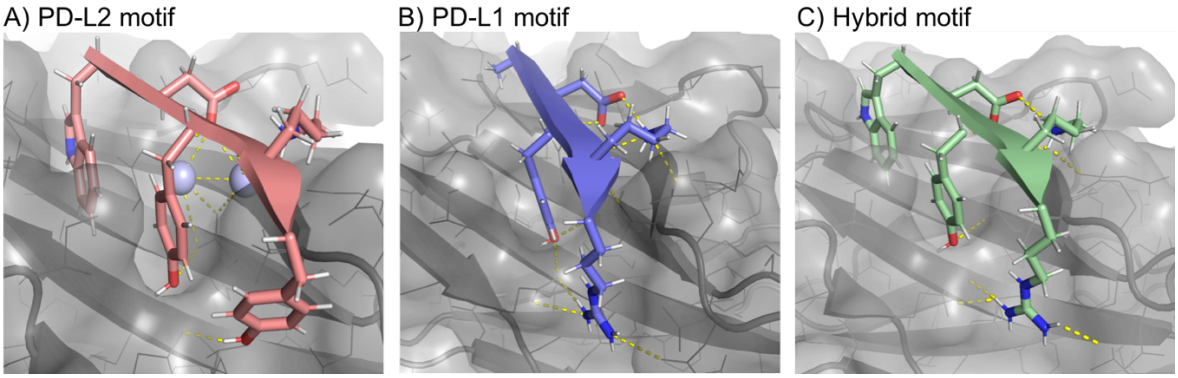

Figure 1.1: Binding Motifs. A) 'WDYKY' binding motif from PD-L2 (PDB 3BP5). B) 'ADYKR' binding motif from PD-L1 (PDB 3BIK). C) 'WDYKR' hybrid binding motif. Murine PD-1 shown in grey. Water molecules placed in crystal structure shown as light blue spheres. Hydrogen bonds shown as yellow dashed lines.

The three identified binding motifs were grafted onto a set of previously characterized, *de novo* scaffolds using the Motifgraft protocol[32]. A scaffold set of 34,840 computationally designed and validated miniproteins were employed, all containing 3 beta strands and a single alpha helix with varying connections. All scaffolds were 35-41 residues in length and contained 1-4 disulfide bonds for stability. The design of these disulfide-constrained peptides has been previously described[8]. Each motif fragment was aligned and grafted into a scaffold if the backbone alignment between the motif and the original scaffold was less than 0.7 Å RMSD and the scaffold did not clash with the target molecule. Multiple grafts were allowed for the same backbone if there was good alignment in multiple positions along the peptide chain. A total of 25,709 grafts were generated for the next phase of design. The binding motif and all disulfide bonds were constrained and then all positions at the interface were designed using Rosetta Monte Carlo sequence design[10] to optimize interactions with murine PD-1. Finally, designs were filtered for interface size (SASA) above 1100 Å, shape complementarity above 0.6 (Figure 1.4A), and binding energy (DDG) below -6 REU (Figure 1.4B) to yield a total of 1,058 Motifgraft designs to test.

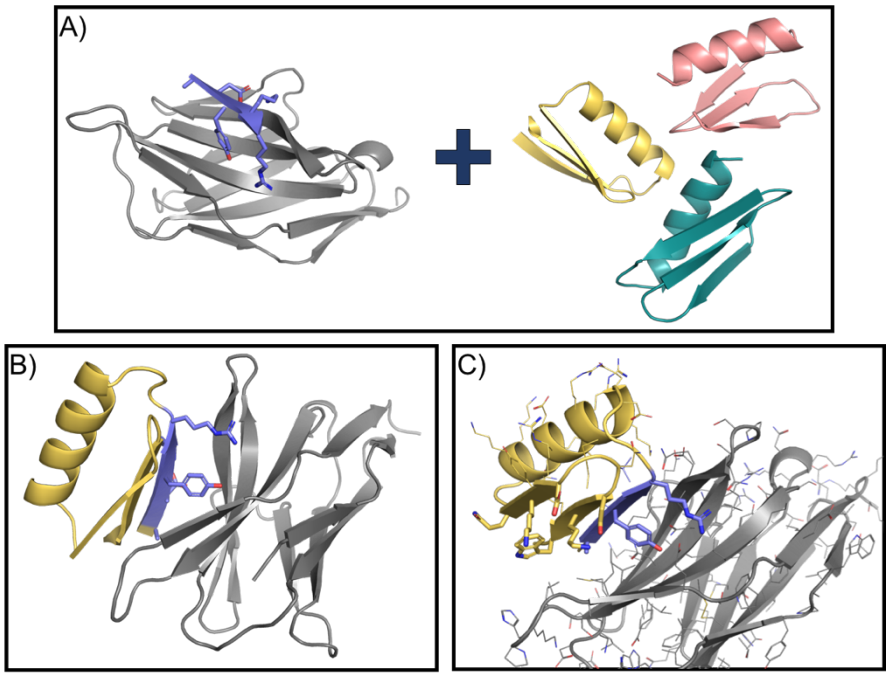

Figure 1.2: Motifgraft Computational Design Protocol. A) A binding motif docked to PD-1 is aligned to a set of *de novo* miniprotein scaffolds. B) A graft is considered successful if both the backbone RMSD between the motif and the scaffold backbone is less than 0.7 Å and the rest of the scaffold does not clash with the target. C) Motif residues are constrained and the rest of the interface is designed to optimize interactions with the target.

For the second set of designs, the binding motif was used as a seed around which to generate *de novo* backbones, a protocol is referred to as Fold-From-Loops[33]. In contrast to the Motifgraft design protocol, we picked just the PD-L2 binding motif as it had the higher binding affinity of the two native ligands (Figure 1.3A). A 'blueprint' file was created to designate the length and identity of each secondary structure element to be constructed on either end of the motif. For our purposes, we wanted to construct the interface out of a beta sheet containing two 5-residue strands in addition to the motif strand and also add a 14-residue alpha helix to stabilize the scaffold. Monte Carlo-based fragment assembly was used to generate thousands of backbones around the motif for this specified topology (Figure 1.3B). Because of the small size of these proteins, cysteines were added by scanning for pairs of positions that were geometrically compatible to formation of a disulfide bond (Figure 1.3C). The mPD-1 chain was added by superimposing the backbones using the motif strand to the PD-1/PD-L2 crystal structure, and then the backbones were filtered for clashes with the target. The binding motif and disulfide cysteines were constrained and the rest of the backbone sequence designed to stabilize the protein fold and optimize interactions across the target interface (Figure 1.3D). Designs were filtered for DDG below -10 REU and for several monomer metrics, including backbone geometry, disulfide scores, and core packing, resulting in 3,999 final designs for this set.

Genes encoding all 5,057 designs – 1,058 Motifgraft designs and 3,999 Fold-From-Loops designs – were ordered as an oligonucleotide array synthesis pool and transformed into yeast with a linearized yeast display vector. Yeast displaying the library of designs were first sorted for the displaying population by collecting all $FITC^+$ yeast cells that had been labeled with anti-cMyc-FITC. We tested for binding to both human and mouse PD-1 as the molecules are slightly different by incubating two separate samples of the library with 1 μM mPD1-Fc or 1 μM hPD1-Fc and collecting cells with binding signal above background. Background signal was determined using a negative control sample incubated with only the secondary fluorophore labels. To distinguish true binders from those that just bound the Fc domain in the PD-1 construct we used for screening, we also did a binding sort with 1 μM of the Fc isotype control. DNA was extracted from the yeast in the unsorted library as well as all four selected pools. The gene inserts were amplified and a unique barcode for each pool added using qPCR. Illumina next-generation sequencing (NGS) was used to sequence all five barcoded pools (unsorted, displaying, mPD-1 binding, hPD-1 binding,

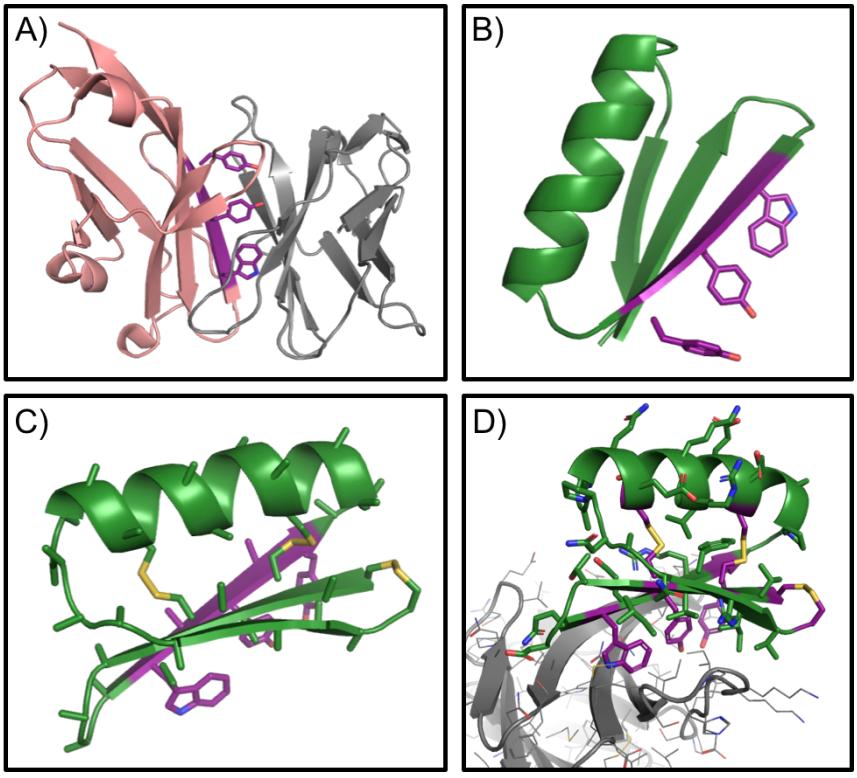

Figure 1.3: Fold-From-Loops Computational Design Protocol. A) A primary binding fragment (shown in purple) is extracted from the crystal structure of mPD-1 (gray) in complex with mPD-L2 (pink) (PDB 3BP5). B) *De novo* backbones, shown in green, are generated around the PD-L2 binding fragment. C) Cysteines are incorporated at geometrically compatible positions for disulfide bond formation. D) Residues in the binding motif and disulfide bonds are first constrained, and then Rosetta sequence design is performed to stabilize the fold and increase interactions with the target.

and Fc binding) simultaneously. Raw sequencing reads were filtered by quality score, and then the count and frequency for each sequence in each pool was calculated. Enrichment ratios were generated by dividing the frequency of a given sequence in one of the binding selections by its frequency in the display reference sort. Binding hits were identified as designs that were enriched in either the murine or human PD-1 selections (enrichment ratio above 2) and depleted in the Fc binding selection (enrichment ratio less than 2).

One Motifgraft design and three Fold-From-Loops designs were found to bind mPD-1 specifically and one Motifgraft design was found to bind hPD-1 specifically in this high-throughput screen. As shown in Figures 1.4A and 1.4B, the biggest difference in the designs generated by each protocol was that the Fold-From-Loops protocol generally yielded larger interfaces. However, of the filters used, there was no single

metric that delineated working designs from non-working designs, and further analysis must be done to identify other metrics that may help explain what makes these five designs successful binders.

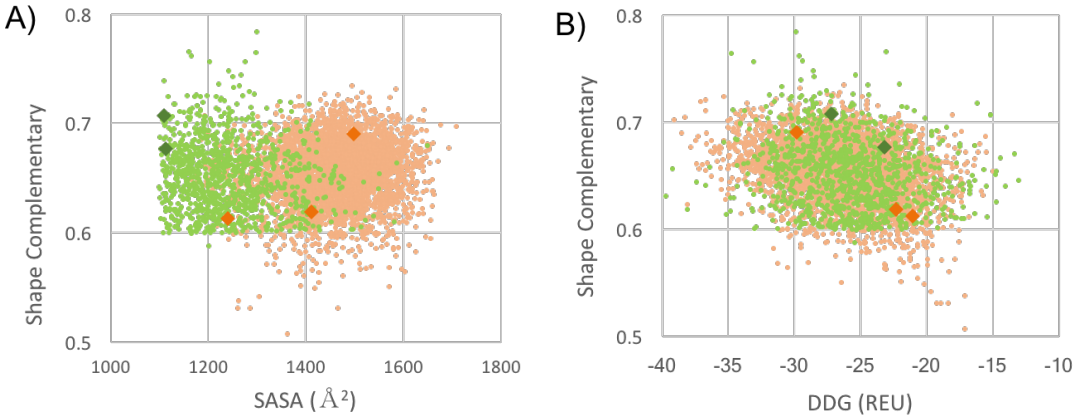

Figure 1.4: High-Throughput Design & Screening Results. Sequences generated by the Motifgraft protocol are in green. Sequences generated by the Fold-From-Loops protocol are in orange. Large, darker dots are successful designs as determined by yeast surface display. A) Shape complementarity and interface size (SASA) of tested sequences. B) Shape complementarity and binding energy (DDG) distribution of designed interfaces of tested sequences.

Affinity Maturation

One of the Fold-From-Loops designs, GR918, was chosen for further optimization as this design had the highest mPD-1 enrichment in the high-throughput screen. The original design was confirmed to bind mPD-1 on yeast at a K_d of ~5 μM (Figure 1.5A, red), but showed no affinity for hPD-1. While murine and human PD-1 are very similar at 70% sequence identity and the natural ligands bind to both with slightly different affinities[18], lack of binding to hPD-1 was understandable as the binding of the original design to mPD-1 was relatively weak. Because we planned to test the protein in a mouse model after affinity maturation, we decided to focus on increasing the affinity of GR918 to mPD-1 first. Due to the high error rate of the gene synthesis technology employed, the initial high-throughput screen also included several variants of each design ordered. We identified 13 single point mutants that had higher enrichments than the ordered GR918 sequence in the mPD-1 selected pool. Each mutation was confirmed individually on yeast. Several simply increased the level of display of the protein, but some of them also increased the binding affinity. Combinations of these mutations were tested and a double mutant, GR918-D7T-Y38E

(Figure 1.5B, orange), was identified to give the best mPD-1 binding. This second generation of GR918, GR918.2, had a K_d of ~4 μM for mPD-1 (Figure 1.5A, green).

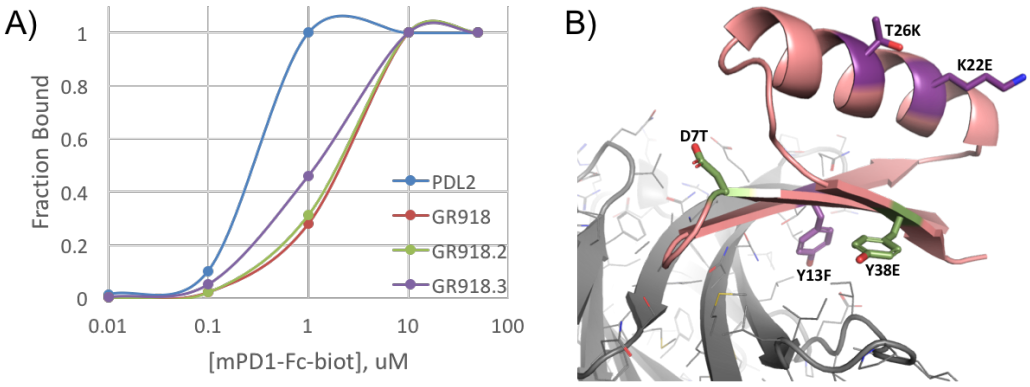

Figure 1.5: Affinity Maturation of GR918. A) mPD-1 binding titration of PD-L2 (blue), GR918 (red), GR918.2 (green), and GR918.3 (purple) on yeast. B) Design model of GR918 (blue) bound to mPD-1 (gray). GR918.2 mutations are indication on structure in orange. GR918.3 are indicated on structure in green.

A site saturation mutagenesis (SSM) library containing every possible single point mutant of GR918.2 was screened for increased binding to mPD-1 on yeast (Figure 1.6C). Combinations of four mutations identified in the SSM found to be beneficial to binding were tested and a triple mutant, referred to here as GR918.3, consisting of Y13F, K22E, and T26K was found to be optimal (Figure 1.5B, green). Position 13 is the middle residue of the binding motif and is a tyrosine in both PD-L1 and PD-L2. Both crystal structures show hydrogen bonding of this tyrosine to an asparagine and a glutamate residue on PD-1 (Figure 1.1A, 1.1B). However, if this residue is not placed perfectly to form these hydrogen bonds, an enalthalpic cost will have to be paid upon binding for displacing waters that would be hydrogen-bonding to the tyrosine in the unbound state. In that case, a fully non-polar amino acid would be preferred at this position which might explain the phenylalanine at position 13 in GR918.3. We also see threonine being replaced by lysine which is more preferred in a helix, and subsequently the neighboring lysine mutates to a glutamate to balance the charges on the helix. This final generation of GR918 will be referred to as GR918.3 and has a K_d of ~1 μM for mPD-1 (Figure 1.5A, purple), which is approximately two-fold higher than mPD-L2, the higher affinity of PD-1's native ligands, on yeast (Figure 1.5A, blue).

Miniprotein Binds at PD-L1/2 Interface on mPD-1

The beta sheet of GR918 was designed to bind the ligand interface on PD-1 to block PD-L1/2 binding (Figure 1.6A). The GR918.2 SSM data was consistent with GR918 binding to PD-1 as anticipated. Calculating the Shannon entropy[34] of each position along the peptide (Figures 1.6B, 1.6C), interface β-sheet residues are highly conserved while non-interface, solvent-exposed residues on the helix are highly variable. This demonstrates that the beta strands of GR918.2 are forming the interface with PD-1 as designed. Furthermore, all cysteine residues involved in the three designed disulfide bonds are highly conserved indicating that the protein is folding correctly.

To verify experimentally that GR918 is binding the correct site on PD-1, we performed a competition assay on yeast using soluble hPD-L1. Yeast cells displaying GR918.3 were first incubated with 0.1 μM soluble, biotinylated mPD-1-Fc followed by the secondary label streptavidin-phycoerythrin (SA-PE) to give a signal for mPD-1 binding by flow cytometry (Figure 1.6E, blue). Another sample of the GR918.3-displaying yeast cells were then incubated with 0.1 μM mPD-1-Fc-biot plus 1 μM unbiotinylated hPD-L1. If hPD-L1 is binding the same interface on PD-1 as GR918.3, it will compete for binding to the mPD-1 molecules in the sample. This will lead to a decrease in the number of mPD-1-Fc-biot molecules that are bound to GR918.3 on the yeast cell and therefore fewer number of SA-PE molecules bound as well (Figure 1.6D). This will be reflected in a decrease in mPD-1 binding signal by flow cytometry. A large decrease in signal was indeed seen with 10-fold excess of hPD-L1 (Figure 1.6E, green) providing evidence that GR918.3 binds the intended region of PD-1.

Binder Folds as Designed & Is Hyperstable

After the first round of affinity maturation, soluble GR918.2 was expressed using the Daedalus mammalian system[35] to confirm the fold, stability, and monomeric state of the protein in solution prior to the final round of affinity maturation. The apo crystal structure at 1.16 Å resolution confirms the protein folds as designed with a backbone RMSD of 1.33 Å between the crystal structure and the computational model (Figure 1.7A). All three disulfides are clearly seen between the correct pairs of cysteines and the rest of the core residues are packed as designed. The disulfide bond between C7 and C30 is forming at a different angle than we had modeled it, causing the helix to shift slightly. While this structure represents only a single snapshot into the conformational space of the unbound protein, we still wanted to analyze if the interface

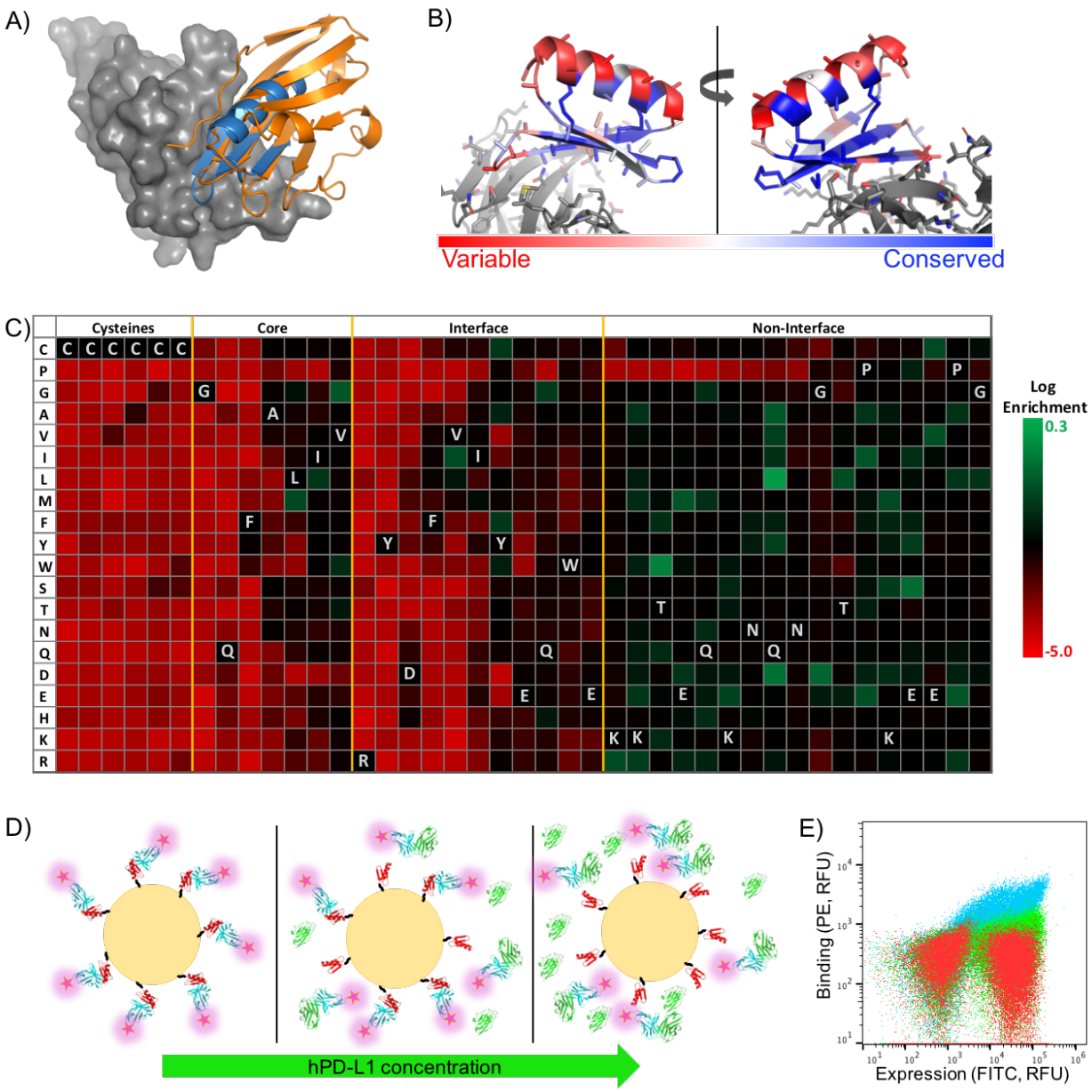

Figure 1.6: Binding Mode Confirmation of GR918. A) Crystal structure of mPD-1/mPD-L2 (PDB 3PB5) overlaid with the design model of GR918. B) The Shannon Entropy of each position derived from the SSM data is mapped onto the model of GR918 ranging from conserved positions in blue with a minimum value of 0.07 to variable positions in red with a maximum value of 3.11. C) Heat map representing the log enrichments for the GR918.2 SSM library selected with 2 μM mPD-1-Fc. Enriched mutations are shown in green and depleted in red. D) Schematic of the Yeast Display competition assay. With increasing concentrations of hPD-L1 (green) added, it will compete with GR918.3 (red) displayed on the yeast for binding to biotinylated mPD-1 (blue) if the binding sites are similar. E) Flow data of the hPD-L1 yeast competition assay. The red sample was incubated with 0.01 mg/ml anti-cmyc-FITC, 0.01 mg/ml SA-PE only. The blue sample was incubated with 100

nM mPD1-biot, washed, and then labeled with 0.01 mg/ml anti-cmyc-FITC, 0.01 mg/ml SA-PE. The green sample was incubated with 100 nM mPD1-biot, 1 μM hPD-L1, washed, and then labeled with 0.01 mg/ml anti-cmyc-FITC, 0.01 mg/ml SA-PE.

residues in the crystal structure adopted their designed rotamers to favor binding. Unfortunately, the protein molecules were packed very tightly in the crystal structure, leading the Trp-Tyr-Tyr motif residues to flip almost a full 180˚ from our bound model to prevent clashing with the same motif on a neighboring molecule (Figure 1.7B). This had a domino effect on the Phe on the neighboring beta strand which is also flipped from the design model. The last strand contains all polar residues which we'd expect to be highly flexible in the absence of PD-1. While the correct structure of the monomer can be confirmed, nothing can be inferred about binding from this crystal structure.

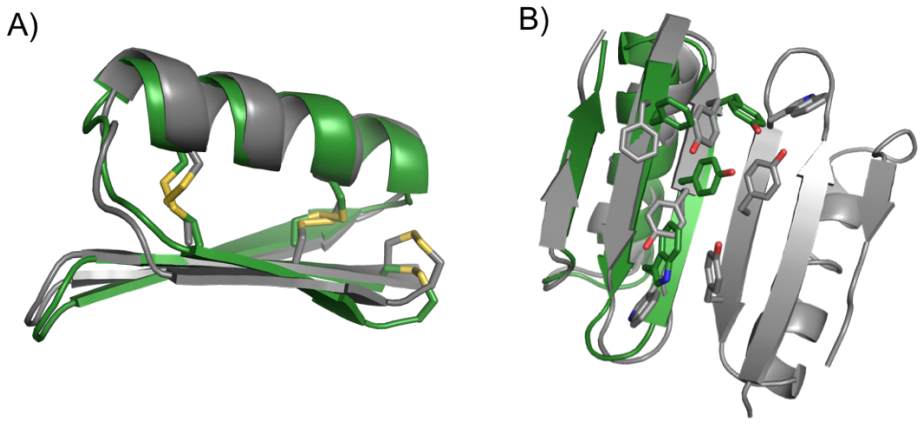

Figure 1.7: GR918.2 Crystal Structure. A) GR918.2 crystal structure (gray) aligned to the design model (green) with a backbone RMSD of 1.33 Å. All three disulfide bonds were confirmed. B) Crystal packing causes interface rotamers to flip to prevent clashing with neighboring molecule.

GR918.2 was found to be hyperstable to both thermal and chemical denaturation. Thermal denaturation was performed by heating the protein from 25˚ C to 95˚ C and the secondary structure was monitored using circular dichroism spectroscopy (CD) at 222 nm. The protein gave a CD spectra characteristic of an alpha-beta fold and only ~20% of the signal was lost at 95˚ C and fully regained upon cooling back down to 25˚ C (Figure 1.8A). Chemical denaturation was performed in a similar manner, and the protein only lost ~35% of the CD signal at 222 nm in 7 M GuHCl (Figure 1.8B). This incredible stability

is likely due to the three disulfide bonds and would not be expected to be retained upon reduction of these bonds.

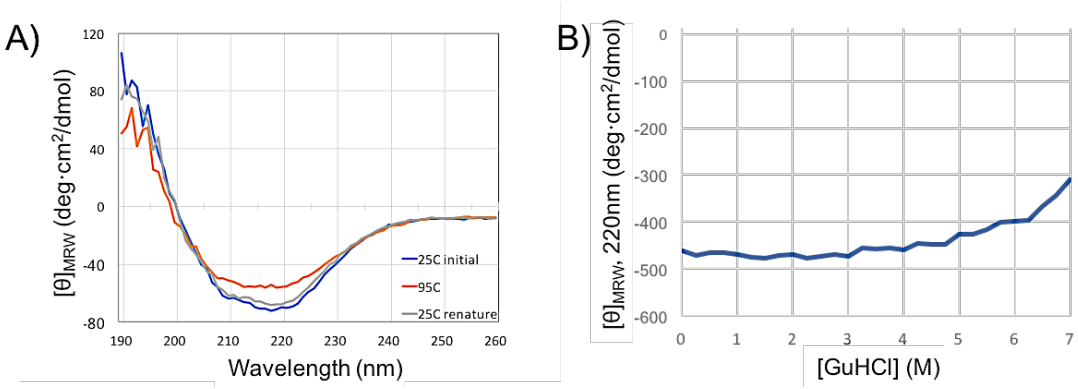

Figure 1.8: Stability of GR918.2. A) CD spectra at an initial temperature of 25˚ C (blue), after heating to 95˚ C (orange), and then after cooling back down to 25˚ C (gray). B) Chemical denaturation with GuHCl, monitoring CD signal at 220 nm.

Generation of Oligomeric PD-1 Binders

As further attempts to optimize binding to PD-1 on yeast failed, we next turned to oligomerization to increase the affinity of GR918.3. The Spycatcher/Spytag system[36] was used to covalently attach GR918.3 to various stable *de novo* oligomers with different orientations (Figure 1.9A). When mixed in solution, a lysine on Spycatcher forms a covalent isopeptide bond with an aspartic acid on the Spytag peptide. This system allows irreversible ligation of any proteins the two peptide tags are attached to. A Spycatcher domain was genetically fused to three homotrimers (1na0C3_int2 (unpublished), 1na0C3_int2_rd4 (unpublished), and 2L6HC3_13[3]) and the trimeric subunit of a two-component icosahedral cage (I53-50[37]). These Spycatcher-oligomer fusions were expressed and purified from bacteria. GR918.3 was expressed with a Spytag and a flexible linker in the Daedalus mammalian system. Spytag-GR918.3 was mixed with each Spycatcher-oligomer at equimolar concentrations to form a covalent conjugation product that could be confirmed as a shift in size by SDS-PAGE (Figure 1.9B). Size exclusion chromatography (SEC) was used to confirm the conjugated proteins could still form the correct oligomeric state. This method works by separating a protein sample by size and comparing the volumes at which each species elutes to a set of standards with known molecular weights. The Spytag-GR918.3 protein alone yielded a single major peak whose elution volume around 18 ml corresponded to the expected monomer

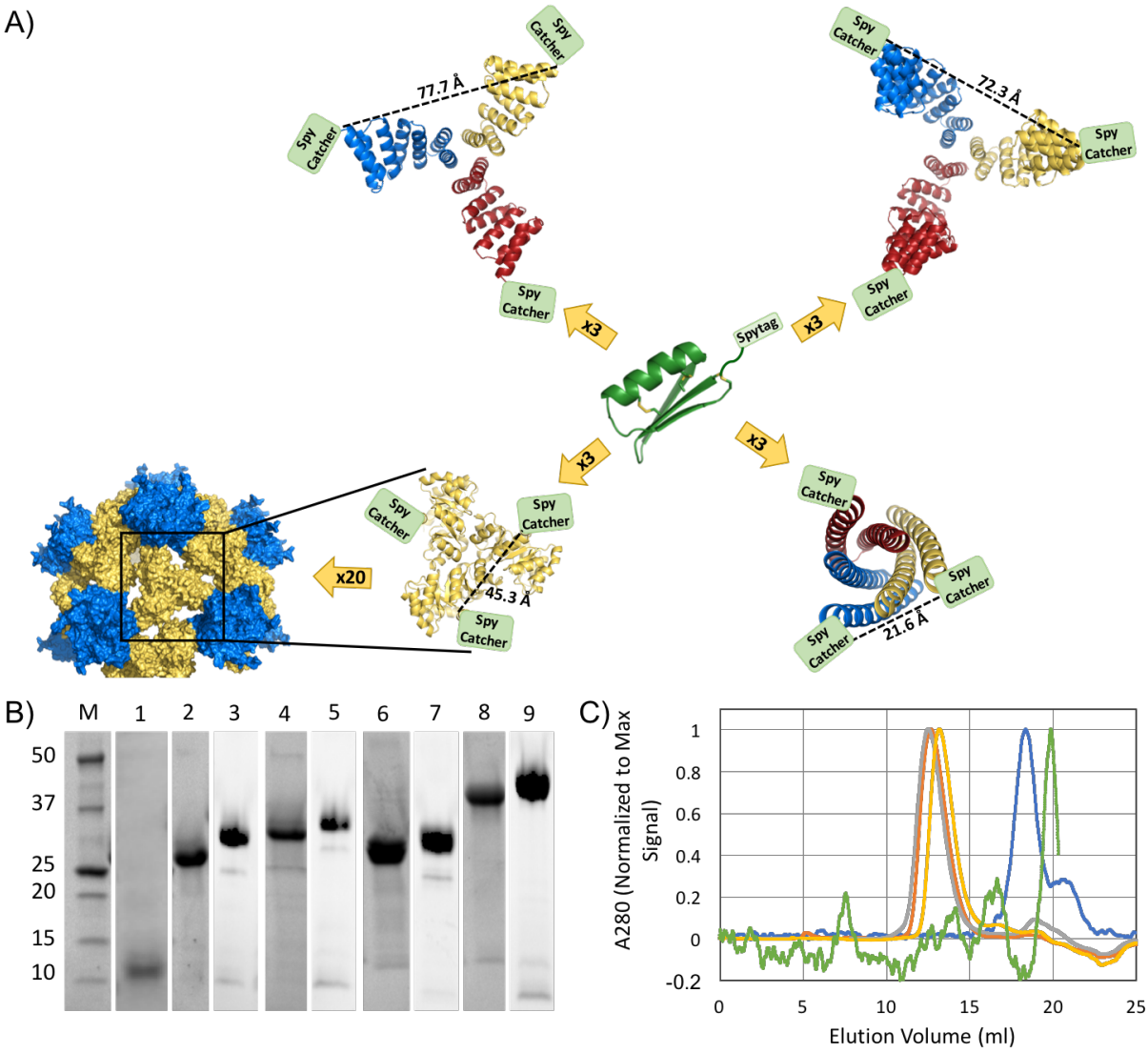

Figure 1.9: Generation of GR918.3 Oligomers. A) Schematic of oligomerization of PD-1 binder via fusion to three designed homo-trimers and a two-component cage. Each trimer subunit is expressed with a Spycatcher domain which covalently binds a Spytag peptide on GR918.3 allowing display of three copies of the binder at the specified distances. The trimer component of the I53-50 cage was also expressed with a Spycatcher domain for conjugation of three GR918.3 binding domains. 20 conjugated trimers then assemble with 12 copies of the second pentamer component to form the icosahedral cage with a total of 60 GR918.3 domains. B) SDS-PAGE of Spytag-GR918.3 conjugation to Spycatcher-oligomers. M=Precision Plus Dual Xtra Protein Ladder (labels in kDa), 1=Spytag-linker-GR918.3 (7.9 kDa), 2=1na0C3_int2_Spycatcher (26.9 kDa), 3= 1na0C3_int2_Spycatcher + Spytag-linker-GR918.3 (conjugated product = 34.8kDa), 4= 1na0C3_int2_r4_Spycatcher (34.9 kDa), 5=1na0C3_int2_r4_Spycatcher + Spytag-linker-GR918.3 (conjugated product = 42.8 kDa), 6=Spycatcher_2L6HC3_13 (24.6 kDa), 7= Spycatcher_2L6HC3_13 + Spytag-linker-GR918.3 (conjugated product = 32.5 kDa), 8=

Spycatcher_I53-50A (37.2 kDa), 9=Spycatcher_I53-50A + Spytag-linker-GR918.3 (conjugated product = 45.1 kDa) C) SEC chromatograms of GR918.3 monomer (blue, 7.9 kDa expected), 1na0C3_int2-GR918.3 trimer (orange, 105 kDa expected), 1na0C3_int2_rd4-GR918.3 trimer (gray, 129 kDa expected), 2L6HC3_13-GR918.3 trimer (yellow, 98 kDa expected), and I53_50-GR918.3 cage (green, 3,798 kDa expected). All samples were run on a Superdex 200 Increase 10/300 GL column. Column standards: 667 kDa=9.00 ml, 158 kDa=12.50 ml, 67 kDa=14.15 ml, 44 kDa=15.23 ml, 17 kDa=17.25 ml, 1.3 kDa=20.53 ml.

size of 7.9 kDa (a 17 kDa standard eluted at 17.25 ml and a 1.3 kDa standard eluted at 20.53 ml) (Figure 1.9C, blue). All three conjugated timers eluted in a single peak around 12.5 ml which corresponds to the expected sizes of 97.5-128.4 kDa (standards: 667 kDa=9.00 ml, 158 kDa=12.50 ml, 67 kDa=14.15 ml) (Figure 1.9C, orange, gray, yellow). Only a small portion of the conjugated cage assembled correctly as seen by the small peak eluting in the void volume of the column around 7.5 ml as is expected for the 3,798 kDa cage (667 kDa standard=9.00 ml) (Figure 1.9C, green). The other peaks in this sample correspond to unassembled components. The fully assembled cage was purified by collecting only the SEC fractions in the 7.5 ml peak.

Designed Miniprotein Specifically Binds Murine PD-1 on Mammalian Cells

The human embryonic kidney cell line, 293T, was used to test specific binding of the 2L6HC3_13-GR918.3 trimer to PD-1 on mammalian cells. 293T cells do not endogenously express PD-1 which enabled confirmation of specific binding by comparing binding signal of the proteins to untransfected cells to cells transfected with the gene for murine PD-1 (*PDCD1*). Binding of the two cell populations was tested against the biotinylated, trimerized binder, 2L6HC3_13-GR918.3-biot, as well as biotinylated, unconjugated trimer, Spycatcher_2L6HC3_13-biot, to ensure any binding signal measured is caused by the GR918.3 binding domain and not the trimer or spycatcher domains. After incubation of the cells with one of the biotinylated trimers, the cells were labeled with SA-PE to produce a binding signal by flow cytometry. Only a background level of signal (<1%) was seen for binding of the unconjugated trimer to either cell line (Figure 1.10A-B). The trimerized binder also showed a background level of signal to untransfected 293T cells not expressing PD-1 (Figure 1.10C). When 293T cells expressing murine PD-1 were incubated with the trimerized binder, a strong binding signal was seen by flow cytometry for 40.3% of the cells (Figure 1.10D). This experiment

yielded the same results with two different batches of protein, confirming that GR918.3 specifically binds murine PD-1 on mammalian cells.

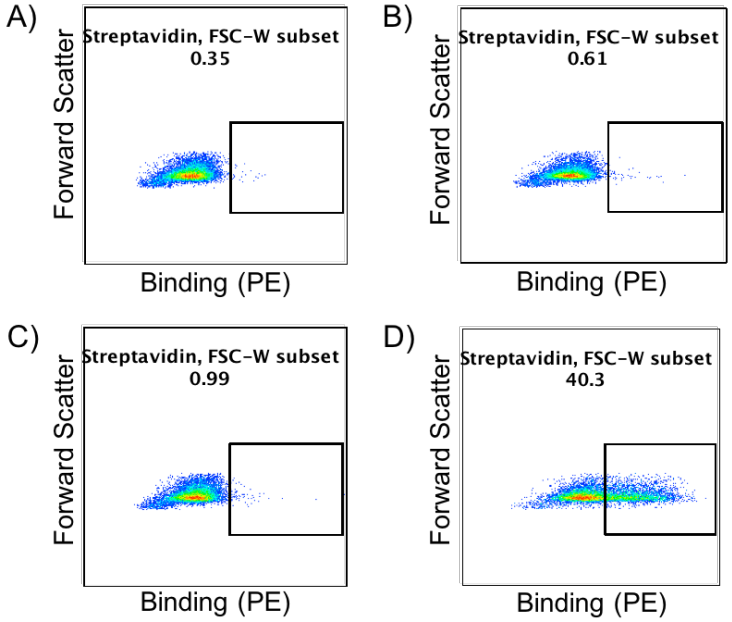

Figure 1.10: Mammalian Cell Binding. A) Flow cytometry results for untransfected 293T cells incubated with 3.7 μM unconjugated, biotinylated Spycatcher_2L6HC3_13 trimer followed by secondary labeling with SA-PE. 0.35% of cells were PE^+, indicated by the rectangular gate. B) Flow cytometry results for 293T cells transfected with the gene for murine PD-1 incubated with 3.7 μM unconjugated, biotinylated Spycatcher_2L6HC3_13 trimer and then SA-PE. 0.61% of these cells were PE^+. C) 0.99% of untransfected 293T cells incubated with 2.1 μM biotinylated 2L6HC3_13-GR918.3 trimer followed by SA-PE were PE^+. D) 40.3% of mPD1-transfected 293T cells incubated with 2.1 μM biotinylated 2L6HC3_13-GR918.3 trimer followed by SA-PE were PE^+.

DISCUSSION

GR918 represents the successful design of the first small, non-antibody protein that binds PD-1 to our knowledge. The crystal structure confirms that it folds into the designed globular alpha-beta protein with all three intramolecular disulfide bonds forming between the correct pairs of cysteines. These disulfide bonds give the miniprotein incredible stability against thermal and chemical denaturation. GR918.2 interface residues were found to be highly conserved in the SSM screen and the binder competed with PD-L1 binding. Together, this data provides evidence that GR918 binds as designed with its beta sheet at the

ligand interface on PD-1. Initial mammalian cell binding assays have been performed that demonstrate that GR918 does indeed bind specifically to murine PD-1 expressed on mammalian cells.

Several experiments are ongoing to further characterize the GR918.3 monomer and various oligomeric versions of the binder. We'll next be doing a binding titration to determine the binding affinity of the trimerized binder as well as testing for binding to stimulated primary murine T cells. While the GR918.3 monomer displayed on yeast does not show any binding to human PD-1, this may be due to the low affinity of the binder. Murine PD-1 is 70% identical to the human ortholog with only three binding residues different, and the native ligands have substantial cross reactivity with both mouse and human orthologs[18]. Therefore, we hypothesize that the trimerized binder may have a sufficient affinity boost from the increase in avidity of the oligomer to display binding to human PD-1, and this will be tested soon using 293T cells transfected with human *PDCD1*.

While we know that the 2L6HC3_13-GR918.3 trimer binds specifically to mPD-1 on mammalian cells, we do not yet know what effect that binding will have on PD-1 and ultimately its function on T cells. All of the current bivalent anti-PD-1 antibodies inhibit the receptor by blocking its interaction with PD-L1 and PD-L2. However, when coupled to a bead, these same antibodies showed some degree of agonistic properties[38] likely due to receptor clustering as is known to occur during natural PD-1 signaling[15]. We hypothesize that monomeric GR918.3 and lower oligomers will likely inhibit PD-1 while higher order oligomers may agonize the receptor. PD-1 agonism may not only depend on the valency of the binding molecule but also the geometry of the fused binding domains that bring the receptors together in specific orientations. We are currently generating several oligomeric versions of GR918.3 with different valencies and geometries for testing in functional T cell assays. GR918.3 could potentially be used to create both inhibitors to treat cancer with very different properties than the full-length antibodies currently used, as well as the first PD-1 agonists that would be very powerful for the treatment of autoimmune disorders.

MATERIALS & METHODS

Computational Design & Gene Synthesis

The ROSETTA protein structure prediction and design software suite[10] was used for all design calculations using the Talaris2013 energy function[39]. The Motifgraft[32] and Fold-From-Loops[33] computational

protocols were performed as previously described using the MotifGraft and BluePrintBDR movers respectively. The Remodel mover was used to add disulfide bonds to *de novo* backbones. Two rounds of design and all-atom minimization using the PackRotamersMover and TaskAwareMinMover were performed to generate final sequences for the Motifgraft designs. The Fold-From-Loops backbones were sequence designed using three rounds of the FastDesign mover followed by 15,000 rounds of the GenericSimulatedAnnealer to optimize the SSPrediction, CavityVolume, SSShapeComplementarity, PackStat, hbond_lr_bb/residue, dslf_fa13/residue, TotalHydrophobic/TotalSasa_Hydrophobic (buried nonpolars), Holes, Ddg, and Rmsd Rosetta filter scores. Genes encoding all 5,057 initial designs to be tested were codon optimized for *S. cerevisiae*[40] and ordered as a single oligo pool from CustomArray. SSM genes were amplified from an Agilent oligo array pool. Genes for all individual mutants to be tested were constructed through oligo assembly using IDT oligos.

Yeast Surface Display

Oligo pools were amplified and transformed into yeast as previously described[14]. Initial testing of all designs and affinity maturation of GR918 were performed using yeast surface display[41]. For the high-throughput screen, yeast displaying designs were labeled with either 1 μM mPD1-Fc, 1 μM hPD1-Fc, or 1 μM IgG2a Fc (expressed from HEK 293F cells) as well as 5 μM biotinylated protein ZZ[42] (expressed from *E. coli* BL21 cells), which specifically binds Fc and acts as an intermediate label for 1 hour at 23˚ C. Cells were then washed and labeled with 0.01 mg/ml SA-PE (ThermoFisher) and 0.01 mg/ml anti-cmyc-FITC (Immunology Consultants Laboratory) as a control for display of the proteins on ice for 10 minutes under aluminum foil. Two consecutive rounds of FACS were performed under these labeling conditions and all $FITC^+/PE^+$ cells collected using a Sony SH800 cell sorter. A single reference sort of all $FITC^+$ displaying cells was also done on the naïve library. Sequencing of the $FITC^+$ pool and both binding selections was performed on an Illumina MiSeq.

The GR918.2 SSM library went through two subsequent rounds of FACS. The first was labeled with 2 μM mPD1-Fc, 10 μM ZZ-biot followed by secondary labeling with SA-PE and anti-cmyc-FITC, and all $FITC^+/PE^+$ cells were collected. For the second sort, cells were labeled under avid conditions with 3 μM biotinylated mPD1-Fc (expressed from HEK 293F cells), 0.75 μM SA-PE, 0.01 mg/ml anti-cmyc-FITC and

the top 20% of $FITC^+/PE^+$ cells collected. Both mPD-1 selections were deep sequenced along with a $FITC^+$ reference sort.

For the competition assay, yeast displaying GR918.3 were incubated with 1 μM mPD1-Fc-biot alone or co-incubated with 10 μM or 50 μM hPD-L1 (expressed in *E. coli*) for 1 hour at 23˚ C. After secondary labeling with SA-PE and anti-cmyc-FITC, the cells were analyzed using an Accuri C6 flow cytometer.

Crystallography

Soluble GR918.2 was produced using the Daedalus mammalian expression system[35] by the Molecular Design & Therapeutics core at the Fred Hutchinson Cancer Research Center. Crystal screens were set up using the sitting drop vapor diffusion method by mixing 2.25 mg/ml protein solution 2:1, 1:1, or 1:2 with reservoir solution from the 96-well Morpheus Crystallization Screen. Crystals grew in several drops (Morpheus D1, E1, F1, H1) within 3 weeks at 20˚ C. Optimization screens were set up with each of these 4 conditions at 1:4, 1:2, 1:1, 2:1, and 4:1 mixtures with 9.85 mg/ml protein solution. Crystals were seen in most of the drops after ~24 hours at 20˚ C. Because the Morpheus screen conditions already contain cryoprotectant, the crystals were looped and immediately flash frozen in liquid nitrogen. The crystal that resulted in a full data set for structure determination was grown in 0.1 M MES/Imidazole pH 6.5, 0.03 M Diethylene Glycol, 0.03 M Triethylene Glycol, 0.03 M Tetraethylene Glycol, 0.03 M Pentaethylene Glycol, 10% PEG 20,000, 20% PEG MME 550.

The X ray diffraction data set was collected at the Advanced Photon Source, Northeastern Collaborative Access Team, Beamline ID-C. Images were processed to 1.16 Å in space group C121. The structure was solved by molecular replacement using PHASER[43] in the PHENIX software suite[44] and the helix (residues 18-31) from the GR918 design model as a search model. Two copies of GR918 were placed in the asymmetric unit. Coot[45] and PHENIX.refine[46] were used for several rounds of manual building and refinement. Data collection and refinement statistics are listed in Figure 1.11.

Circular Dichroism

All CD experiments were conducted using an AVIV Model 420 Circular Dichroism Spectrometer. Thermal denaturation was conducted by first performing a wavelength scan from 190-260 nm, sampling every 1 nm, at 23° C using 0.23 mg/ml GR918.2 in PBS in a 1 mm quartz cuvette. Then, the CD signal at

Wavelength	0.9791
Resolution range	25.15 - 1.159 (1.2 - 1.159)
Space group	C 1 2 1
Unit cell	51.027 26.746 51.669 90 99.645 90
Total reflections	68154 (3883)
Unique reflections	22603 (1818)
Multiplicity	3.0 (2.1)
Completeness (%)	93.59 (76.41)
Mean I/sigma(I)	8.85 (1.66)
Wilson B-factor	14.87
R-merge	0.04697 (0.3456)
R-meas	0.05593 (0.4421)
R-pim	0.02974 (0.2718)
CC1/2	0.999 (0.826)
CC*	1 (0.951)
Reflections used in refinement	22605 (1817)
Reflections used for R-free	1999 (161)
R-work	0.1499 (0.2211)
R-free	0.1621 (0.2431)
CC(work)	0.969 (0.902)
CC(free)	0.963 (0.902)
Number of non-hydrogen atoms	807
macromolecules	720
solvent	87
Protein residues	86
RMS(bonds)	0.008
RMS(angles)	1.53
Ramachandran favored (%)	97.56
Ramachandran allowed (%)	2.44
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	6.58
Clashscore	8.68
Average B-factor	19.80
macromolecules	18.47
solvent	30.77

Figure 1.11: Data Collection & Refinement Statistics for GR918.2.

222 nm was monitored while heating the sample to 95° C, taking a data point every 2° C after incubating the sample for 30 seconds at the new temperature. Another wavelength scan was taken at 95° C and again after cooling the sample back down to 23° C. All measurements were also taken with PBS for buffer subtraction.

Chemical denaturation was performed in a similar manner. Initially, a wavelength scan from 190-260 nm was taken on 0.05 mg/ml GR918.2 in PBS in the absence of denaturant in a 1 cm quartz cuvette at 23° C. A solution of 8M GuHCl with 0.05 mg/ml GR918.2 was then titrated into the cuvette under constant

volume and with stirring, and the CD signal at 220 nm was recorded every 0.25 M up to 7 M. A final wavelength scan was performed of the protein solution with 7 M GuHCl.

Oligomer Conjugations

Design and *in vitro* characterization of 1na0C3_int2 and 1na0C3_int2_r4, not yet published, were performed by George Ueda following protocols published previously for very similar cyclic oligomers[47], and the crystal structures for both have been solved confirming the designed trimers. Genes encoding 1na0C3_int2 and 1na0C3_int2_r4 were cloned with a C-terminal Spycatcher[36] domain and His-tag into the *E. coli* expression vector pET28b(+) (EMD Millipore). 2L6HC3_13 was cloned with an N-terminal Spycatcher domain, Avi-tag, and His-tag into the *E. coli* expression vector pET29b(+) (EMD Millipore). The trimeric component A of I53-50 was cloned with an N-terminal His-tag, Spycatcher domain, and 8-residue GS linker into pET29b(+). The pentameric component B of I53-50 was cloned with a C-terminal His-tag into pET29b(+). All proteins were expressed in the *E. coli* expression strain Lemo21(DE3) (New England BioLabs) and purified via Ni-NTA IMAC followed by gel filtration using a Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences). The construct Spytag[36]-(GDGGKGSDGGKGSGGSSG)-GR918.3 was expressed and purified from the Daedalus mammalian system[35] by the Molecular Design & Therapeutics core at the Fred Hutchinson Cancer Research Center.

Spycatcher/Spytag conjugations were performed by incubating equimolar concentrations (>35 μM) of Spycatcher-oligomer with Spytag-GR918.3 at 23° C for >1 hour. Conjugated product was separated from unconjugated components via gel filtration using a Superdex 200 Increase 10/300 GL column. Purified, conjugated products were analyzed for oligomeric state by SEC-MALS. The I53-50 cage was assembled by incubating 16 μM I53-50A conjugated to GR918.3 with 16 μM I53-50B at 23° C for 1 hour. Assembled cage was then purified from unassembled components by gel filtration using a Superose 6 Increase 10/300 GL column (GE Healthcare Life Sciences).

Mammalian Cell Assays

Gauri Bhise in Mohamed Oukka's lab at Seattle Children's Research Institute performed the mammalian cell assays. Two cell populations were used for this assay – wild type human 293T cells and 293T cells transfected with the gene for murine PD-1 (*PDCD1*). Binding was tested for both Spycatcher_2L6HC3_13-biot (unconjugated trimer control) and 2L6HC3_13-GR918.3-biot (trimerized

binder). Cells were incubated with either 3.7 μM biotinylated, unconjugated trimer or 2.1 μM biotinylated, trimerized binder for 30 minutes at 37° C. The cells were then washed and incubated with SA-PE for 10 minutes on ice under aluminum foil. Cells were analyzed for their PE fluorescence by flow cytometry.

Chapter 2: A Novel Yeast Secretion Capture Method for Selection of Folded, Functional Proteins

INTRODUCTION

Yeast surface display has become a highly valuable tool since it was originally developed by Dane Wittrup's lab for the selection of single-chain variable fragments, or scFv's, with increased antigen binding affinities in 1997[48]. Since then, it has been expanded to engineer full-length antibodies, protein-protein interactions, protein-small molecule interactions, and enzymes[49]. In traditional yeast display, the protein-of-interest (POI) is fused to the C-terminus of the *Saccharomyces cerevisiae* mating protein Aga2p (Figure 2.1A). This protein is highly expressed and easily accessible on the cell surface, making it well-suited to display the fused protein for effective labeling. After induction of the displayed proteins, the cells can be labeled using a fluorophore-conjugated antibody against a c-myc tag on the C-terminus of the fusion protein to assess the level of display using a flow cytometer. Displayed proteins can be further labeled with a biotinylated ligand molecule and streptavidin conjugated to the orthogonal fluorophore PE to test binding (Figure 2.1B). When combined with fluorescence-activated cell sorting (FACS) and Illumina deep sequencing, yeast surface display allows for the high-throughput screening of millions of potential binding proteins simultaneously.

While this is an extremely powerful method, one crucial drawback is the possible display of misfolded or unstable proteins[50–52]. One possible reason for this is the use of the Aga2p fusion. Some terminally misfolded proteins that would be otherwise degraded by the ER-associated degradation pathway (ERAD) are able to be successfully secreted from yeast when genetically fused to a soluble secretory protein[53]. This makes well-folded proteins indistinguishable from misfolded proteins in the current yeast display system, leading to many false-positive "hits" coming out of screening or evolution of a binder to a less stable yet higher-affinity variant. Typically, protein hits screened or evolved as Aga2p fusions in the yeast display system are then cloned into an expression system to generate soluble protein for *in vitro* characterization. At this point, many of the proteins selected in yeast display are unable to be expressed solubly on their own. These proteins have to be stabilized by individually testing sometimes dozens of point mutations and then combinations thereof for increased soluble expression[54]. If a stable variant is not able

to be achieved through low-throughput expression screens, more yeast display hits must be tested to identify one with sufficient expression and stability. The design testing pipeline could be drastically improved through development of an assay that could select for both function and stability.

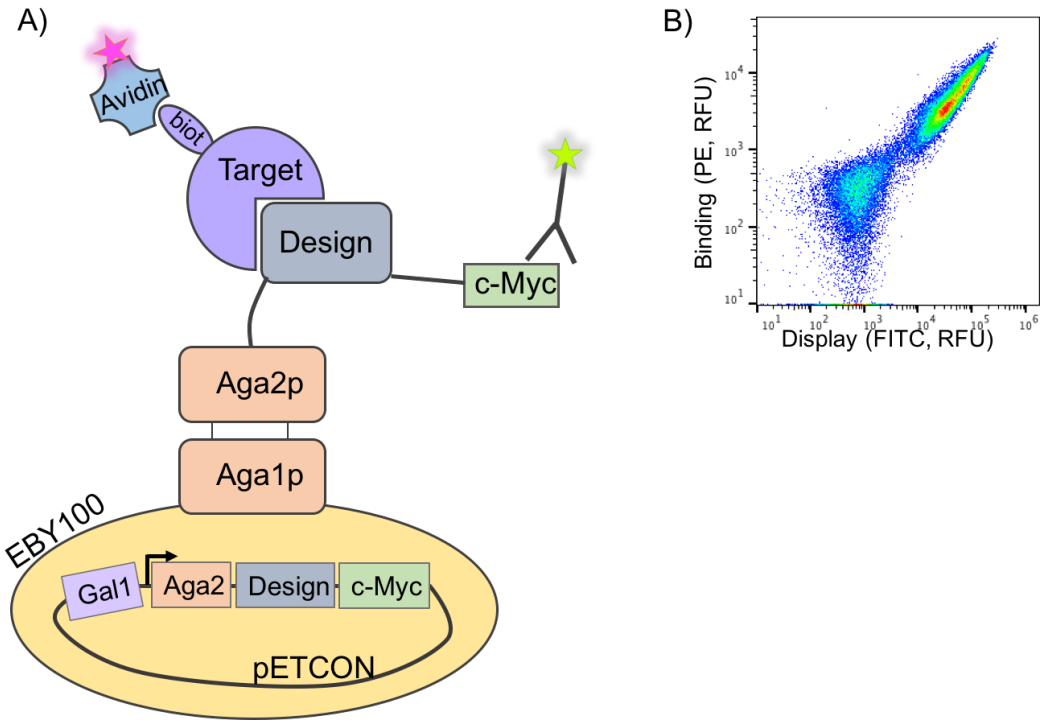

Figure 2.1: Traditional Yeast Surface Display System. A) Schematic of the canonical yeast surface display binding assay developed by Dane Wittrup in 1997. The protein design to be tested is displayed on the cell surface through an Aga2p fusion. This can be labeled via a C-terminal c-Myc tag to give a display signal or labeled with a biotinylated target protein and streptavidin-PE to give a binding signal by flow cytometry. B) An example of flow cytometry results in the yeast display system. Yeast displaying a POI are labeled with a FITC-conjugated antibody targeting the displayed construct's cMyc tag which gives a "Display Signal" by flow, depicted on the X axis of this plot. The right $FITC^+$ population are yeast cells displaying the POI. These cells are also incubated with a biotinylated target molecule that binds the displayed POI and then labeled with a secondary SA-PE reagent that gives an orthogonal binding signal (Y axis) to just the displaying cells, seen here as the $FITC^+PE^+$ population.

Assessing protein expression and stability in a high-throughput manner remains a challenge and several methods have been put forward to try to solve this problem. Some of these take advantage of the Aga2p yeast surface display system and use either heat[55] or proteases[56] to remove unstable or misfolded proteins prior to selection. However, heat selections are limited in that they rely on utility of a highly

conformationally-specific ligand in order to recognize proteins that retain the correct fold after incubation at elevated temperatures. Misfolded proteins often expose hydrophobic patches that are very sticky and can bind nonspecifically[57], giving a misleading "binding signal" that would be interpreted as the protein being folded. Additionally, this assay does not discern between folded, stable proteins and molten globules with a high thermostability[52]. The protease assay does not use binding as a readout for stability and therefore does not have these limitations. Instead it uses a C-terminal tag to detect protein that is uncleaved and therefore retained on the surface of the yeast cell. However, if the POI contains disulfide crosslinks, it can still be covalently attached to the cell surface even if the protein is unstructured and sensitive to proteolysis. Reducing agent cannot be used in this assay because the Aga2p protein is connected to the yeast surface through disulfide bonds with the transmembrane protein Aga1p. Addition of reducing agent therefore results in loss of the displayed protein[58]. Due to these limitations, a general protocol is still needed to select for protein stability and function.

Rather than look at the stability of the protein after it gets displayed on the cell surface, another approach would be to take advantage of the yeast quality control pathway's ability to discern folded from misfolded proteins before they are displayed for screening. This could be done by replacing the Aga2p fusion protein with a short, unstructured peptide tag that would not be expected to have any effect on the folding or stability of the protein it's attached to. To be used for directed evolution, there would have to be a way to couple the secretion phenotype to the genotype of the successfully secreted protein. This could be done by capturing the protein on the surface of the cell it was secreted from, resulting in a cell displaying the secreted protein in a manner similar to the canonical yeast surface display for downstream assays.

Several methods tailored to specific applications have been developed using this concept. One, termed the Cell Surface Secretion Assay (CeSSA), was developed for the purposes of enhanced protein production and involves labeling yeast cells with a known ligand of the POI and then secreting the protein with a small detection tag for selection of cells with increased secretion[59]. Unfortunately, because cells are captured using a known ligand, this method cannot be used for screening novel binders. Multiple approaches have been described to secrete and capture full-length antibodies[60,61] but are not applicable to non-antibody proteins. A more general method, called the Secretion and Capture Technology or SECANT[62], captures secreted biotinylated proteins on the cell surface that has been labeled with avidin. However,

yeast display assays most commonly utilize biotinylated target molecules to detect binding[49] which would be unusable in this system due to cross-reactivity with the avidin-labeled cells. Additionally, none of these methods have been demonstrated to probe differences in the degree of foldedness or stability of proteins.

To this end, we sought to develop a general yeast secretion capture assay, we call SecCap, that could be used to analyze and select for protein stability and function simultaneously. Our hypothesis is that only well-folded proteins get past the yeast quality control machinery to be secreted, and only these will be captured covalently on the cell surface. The yeast cells with captured protein can then be labeled with a fluorophore-conjugated antibody against a small, unstructured tag in the secreted protein construct to give a secretion capture signal by flow cytometry to identify well-folded proteins or incubated with a target ligand to assess binding. Additionally, we found that the system can be used with FACS and deep sequencing to robustly screen or evolve proteins for improved secretion from yeast. This same system can also be used to make a small amount of protein for in vitro characterization without having to transfer the gene to another expression system. The novel SecCap system presented here is a complete method for screening, optimizing, and producing proteins readily secreted from yeast.

RESULTS

Development of the SecCap System

The SecCap system consists of three main parts: 1) a yeast strain that constitutively expresses a cell surface protein to be covalently labeled (Figure 2.2A), 2) an exogenously produced capture reagent (Figure 2.2B), and 3) a plasmid that allows secretion of the POI with two short N-terminal tags for capture and detection of the secreted peptide (Figure 2.2C). Addition of an exogenous capture reagent is critical as it prevents fusion of the POI with the display protein inside the cell prior to secretion, which would result in a construct similar to the Aga2p fusion in standard yeast display. For cell surface labeling and capture, we employed two orthogonal pairs of protein ligation tags, Spycatcher/Spytag[36] and Snoopcatcher/ Snooptag[63] that allow irreversible covalent linkage between the proteins they are attached to.

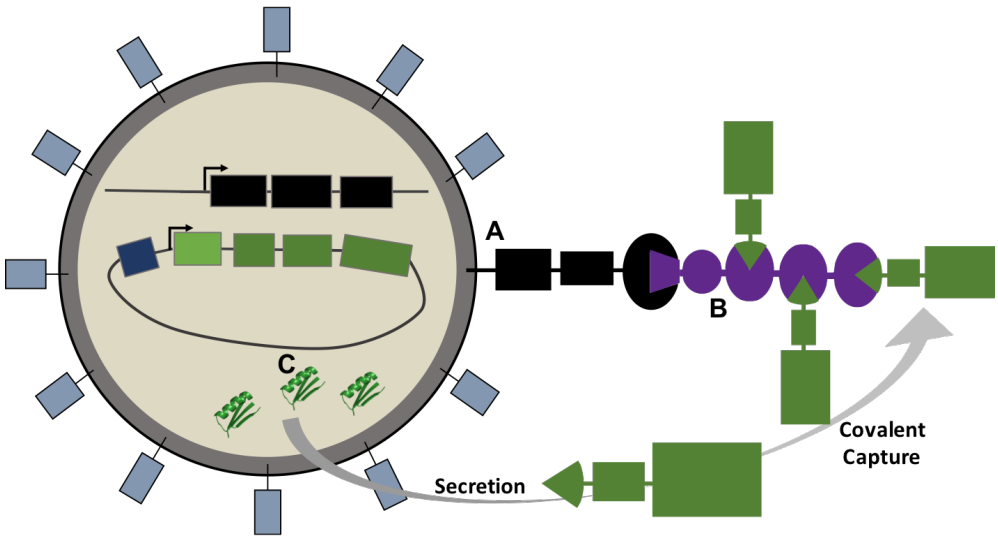

Figure 2.2: Main Components of the SecCap Assay. A) The *S. cerevisiae* strain ySecCap displays a protein domain (black) to be covalently labeled. B) Cells are covalently labeled with an exogenous capture reagent (purple). C) Secreted protein (green), encoded in the plasmid pSecCap, is covalently captured on cell surface.

A *Saccharomyces cerevisiae* strain, ySecCap, was engineered to constitutively surface display a Snoopcatcher domain via fusion to Aga2p. Two identical cassettes containing Aga2p, a short flexible linker, and a Snoopcatcher domain were introduced into the yeast genome at two different chromosomal loci. To confirm display of the Snoopcatcher construct, I labeled ySecCap cells with a FITC-conjugated antibody against a c-Myc tag contained in the Aga2p/Snoopcatcher construct. The signal distribution of the anti-cMyc-FITC labeled cells (Figure 2.3A, blue) was shifted to a higher FITC fluorescence compared to the unlabeled cells (Figure 2.3A, red) as measured by flow cytometry. Linearized plasmid containing the engineered secretion signal sequence app8[64] followed by a 12-residue unstructured Spytag, 8-residue polar Flag-tag, and cloning site for the POI under a GAL1 promoter with a TRP1 auxotrophic selectable marker, termed pSecCap, was transformed along with DNA fragments containing the genes to be tested and adaptors complementary to the plasmid cloning site for homologous recombination in ySecCap[65]. After growth of the transformed yeast cells in selective media, the cultures were transferred to selective galactose induction media for a pre-induction step to prime the expression machinery. Cells were then covalently labeled with a Capture Reagent consisting of a 12-residue Snooptag and three Spycatcher domains with a 10x(GS) linker between each. This Capture Reagent was recombinantly expressed and purified in-house

from *E. coli*. A biotinylated form of the Capture Reagent was used to confirm ligation of the Snooptag to the Snoopcatcher domain displayed on ySecCap. A sample of cells labeled with 5 μM biotinylated Capture Reagent for one hour as well as a sample of unlabeled cells were incubated with the secondary reagent SA-PE. 84.5% of the Capture Reagent-labeled cells (Figure 2.3B, blue) had a positive PE signal above background as determined by the unlabeled control (Figure 2.3B, red), which can be seen from the large shift in PE fluorescence of the labeled cells. These labeling conditions were found to be sufficient for the SecCap assay. Capture Reagent-labeled cells were then moved to a galactose medium to induce expression of the secreted protein. The captured protein could be detected by incubating the yeast cells with a PE-conjugated antibody against a Flag tag included in the secreted protein construct. Flow cytometry gave a Secretion Capture signal as seen by a shift in PE fluorescence of secreting cells (Figure 2.3C, blue) compared to non-secreting cells (Figure 2.3C, red), both of which had been labeled with Capture Reagent prior to induction. Therefore, the Secretion Capture signal obtained using the anti-Flag-PE antibody was specific to the protein secreted and covalently captured on the yeast surface.

A method had to be devised to ensure that protein being secreted from one cell in a culture containing a library of proteins to be tested was solely captured by the cell it was secreted from in order to couple the phenotype being tested to the correct genotype. Rather than induce the cells in a liquid culture with shaking as is typically done, we instead suspended the cells in a highly viscous induction media containing 30% 8 kDa PEG and plated a thin layer of ~1 mm as a stationary culture. 1 μM Spycatcher domain was added to the media as a blocking agent to quench any protein that did not get immediately captured by the cell from which it was. Additionally, 1 μM Spycatcher is added to the recovery buffer when the plate inductions are harvested to prevent capture of unquenched protein in the media when the cells are mixed.

To test the method and determine an optimal plate induction time, a set of four characterized, stable proteins[56], referred to here as SC1-4, and a set of four unstable proteins, SC5-8, were used (Figure 2.4). All eight of these proteins display to a high level as Aga2p fusions in the traditional yeast surface display system, as seen by the high percentage of anti-cMyc-FITC$^+$ cells compared to the unlabeled cell controls (Figure 2.5A-H, histograms on left). High yields of soluble SC1-4 could be expressed and purified via nickel immobilized metal affinity chromatography (IMAC) from bacteria as SUMO fusions, seen by the large bands

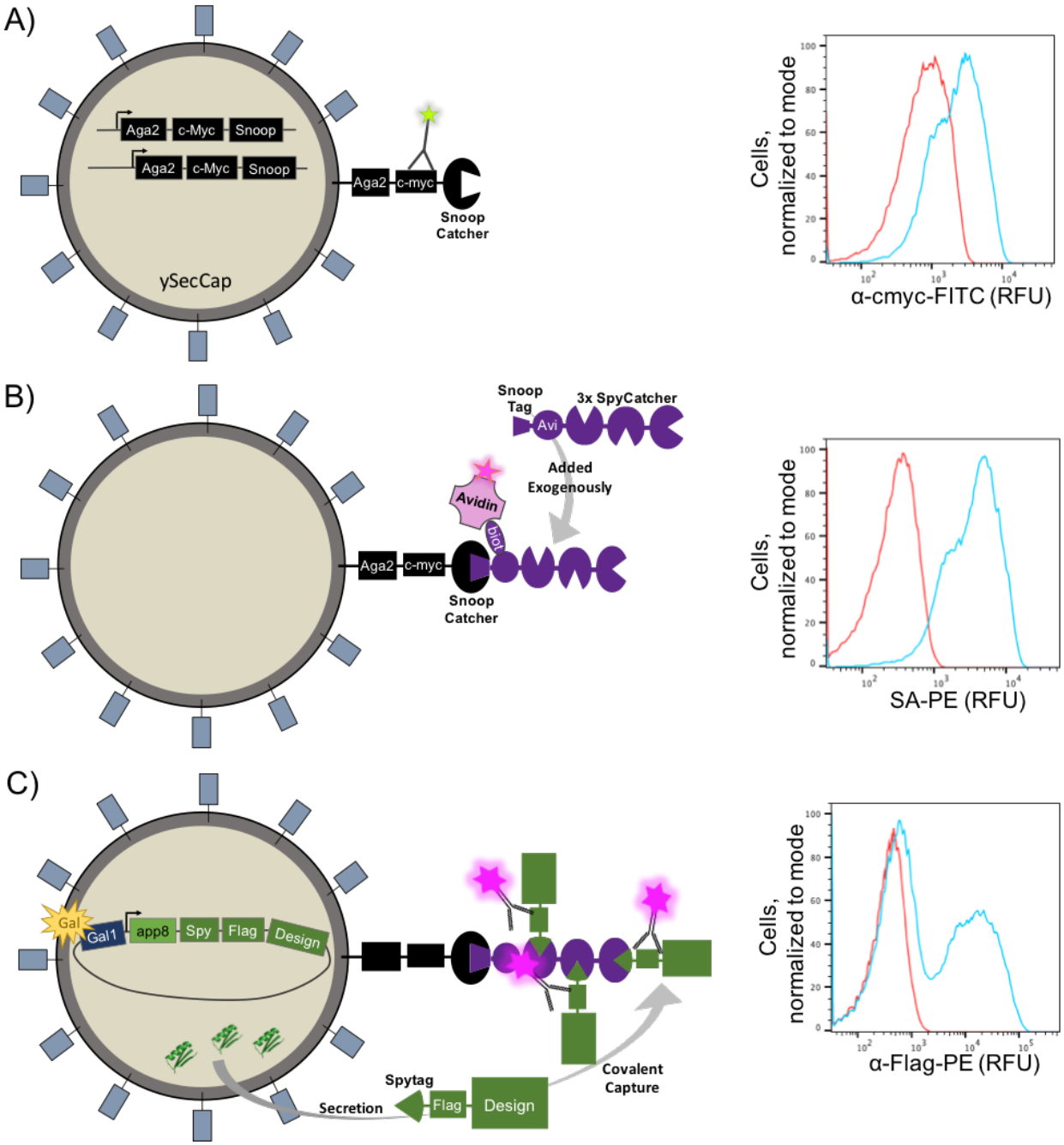

Figure 2.3: Validation of SecCap Components. A) Display of Aga2p-cmyc-Spycatcher is confirmed by flow cytometry seen by the shift in FITC fluorescence of ySecCap cells labeled with anti-cmyc-FITC (blue histogram) compared to unlabeled cells (red histogram). B) ySecCap cells are labeled with biotinylated Capture Reagent followed by the secondary label SA-PE. A control sample to identify any nonspecific binding of SA-PE to unlabeled cells was used to determine the background level of PE fluorescence by flow (red histogram). 84.5% of cells labeled with Capture Reagent had a PE signal above background and these cells show a large shift in fluorescence (blue histogram). C) Capture of secreted protein is detected using a PE-conjugated antibody

against a Flag tag in the secreted construct, giving a SecCap signal by flow cytometry. Capture Reagent-labeled, secreting cells show a shift in PE fluorescence (blue histogram) compared to non-secreting cells also labeled with Capture Reagent (red histogram) after labeling both with anti-Flag-PE.

around 18 kDa in the total lysate ("T") and the IMAC elution ("E") fractions by SDS-PAGE (Figure 2.5A-D, gel images on right). The 5 kDa cleaved proteins ("C" on gel) were found to be folded and stable by Circular Dichroism spectroscopy (CD). SC7-8 yielded no soluble expression as SUMO fusions, seen by the lack of overexpressed proteins at the 18-21 kDa in the total lysate ("T"), soluble lysate ("S"), and IMAC elution ("E") fractions by SDS-PAGE (Figure 2.5G-H, gel images on right). SC5-6 had a high level of bacterial expression ("T") when fused to a SUMO domain, however most of this protein was found in the insoluble lysate fraction ("IS"). A small amount of soluble fusion protein could be purified (lane "E") that ran around the expected 20-21 kDa sizes on a gel (Figure 2.5E-F, gel images on right). However, this protein was highly aggregation-prone upon cleavage of the fusion partner and only background levels of the 7-8 kDa cleaved protein ("C") could be observed by SDS-PAGE after substantial concentration of the sample. Due to the similarities in yeast display levels but the drastic differences in soluble expression level, this set of proteins was ideal to determine if SecCap could discern between well-folded, stable proteins and misfolded, inexpressible proteins.

Clone	Topology	Length	Disulfides
SC1	EHEE	40 residues	0
SC2	EHEE	40 residues	0
SC3	HHH	43 residues	0
SC4	EHEE	40 residues	0
SC5	EEHEHE	67 residues	0
SC6	EEHEHE	67 residues	0
SC7	unknown	66 residues	0
SC8	EEHE	40 residues	2

Figure 2.4: Protein Set Used in Development of SecCap. Topology, length, and number of disulfides of SC1-8 proteins. E=beta strand, H=helix. The topology of SC7 is unknown as this is a chimera of two different designs resulting from multiplex assembly[13] of an oligo array pool.

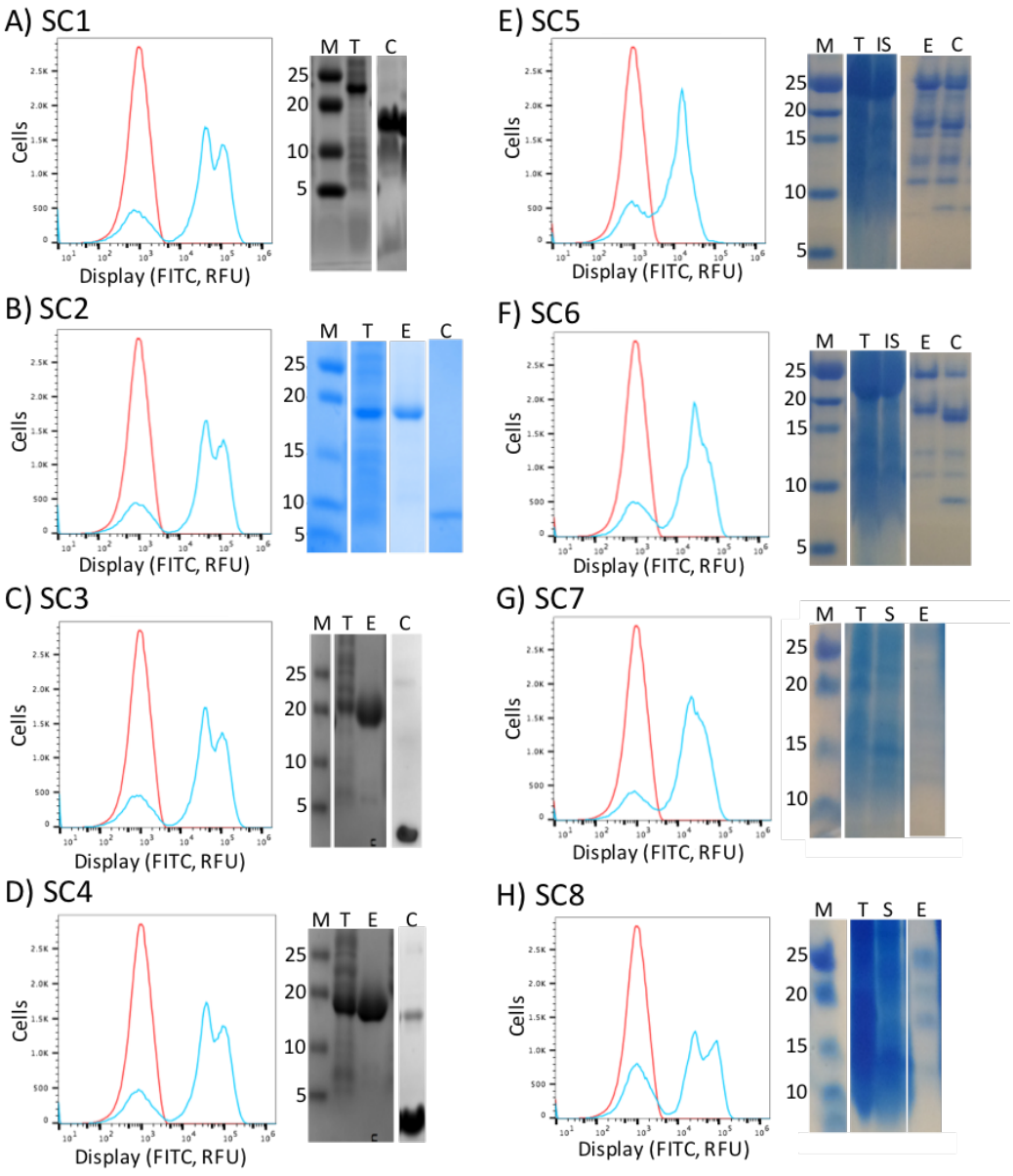

Figure 2.5: Yeast Surface Display & Expression of Protein Set. Left) Flow plots of the display level in the standard yeast display system with unlabeled cells in red and cells labeled with anti-cmyc-FITC in blue. A high percentage of displaying cells is seen for all eight proteins as seen by the large peak in the blue histograms that is shifted to a higher level of fluorescence compared to the red histograms. Right) SDS-PAGE of bacterial expression, purification, and cleavage fractions. M=protein ladder, T=total expression, S=soluble lysate, E=Ni IMAC elution, C=cleaved protein. A) SC1=18.0 kDa fusion, 5.0 kDa cleaved. B) SC2=18.0 kDa fusion, 5.0 kDa cleaved. C) SC3=17.9 kDa fusion, 4.9 kDa cleaved. D) SC4=17.9 kDa fusion, 4.9 kDa cleaved. D) SC5=20.4 kDa fusion, 7.5 kDa cleaved. E) SC6=21.4 kDa fusion, 8.4 kDa cleaved. F) SC7=20.8 kDa fusion, 7.9 kDa cleaved. G) SC8=17.7 kDa fusion, 4.7 kDa cleaved.

A time course of the SecCap assay was performed using SC1-8 to identify the optimal amount of plate induction time for maximum secretion capture signal. Cultures were grown, pre-induced, labeled, and 8 wells of plate inductions were started for each clone. Time points were taken at 0, 6, 8, 10, 12, 14, 16, and 24 hours by recovering one of the 8 induced wells of each clone, labeling with α-flag-PE, and measuring the secretion capture signal by flow cytometry. Because the optimal *S. cerevisiae* growth temperature is 30° C, plate inductions were initially incubated at this temperature for up to 24 hours. The plate induction time identified to give the highest secretion capture signal at this temperature was 8-12 hours (Figure 2.6A). This would make the assay very difficult from a practical perspective as the pre-induced cells would have to undergo one-hour labeling with Capture Reagent, transferred to a plate in induction media, and then recovered after 8-12 hours for immediate downstream testing. Therefore, the same timecourse was then conducted at 23° C to slow the rate of protein production. At this temperature, the induction time that gave the largest difference in mean PE signal between the stable SC1-4 clones and the unstable SC5-8 clones was indeed shifted to around 16 hours (Figure 2.6B), and an overnight plate induction at 23° C was chosen for all further SecCap assays.

We next tested the assay for diffusion of secreted protein to neighboring cells under the conditions we had developed. To do this, we mixed equal parts of two cell strains: 1) a ySecCap culture that had been transformed with pSecCap containing SC-2, the highest secreted gene tested in the timecourse, and 2) an untransformed ySecCap containing a chromosomally integrated, constitutively expressed mTurquoise gene (Figure 2.6C). This mixture of cells was labeled with capture reagent, plated in PEG induction media with 1 µM Spycatcher, and incubated at 23° C. After 17 hours, the cells were recovered with PBSF containing 1 µM Spycatcher, labeled with α-flag-PE, and analyzed by flow cytometry. Approximately 20% of the total yeast cells had captured secreted protein, seen by the PE^+ signal in the upper left and right quadrants of Figure 2.6D. 90% of the total PE^+ cells were $mTurquoise^-$ (upper left quadrant compared to the sum of the upper left and right quadrants), meaning these cells came from the SC2/pSecCap-transformed strain and could be actively secreting protein, while only about 10% were $mTurquoise^+$, non-secreting cells (upper right quadrant compared to the sum of the upper left and right quadrants). Additionally, the mean PE signal of the $mTurquoise^-PE^+$ population (upper left quadrant) at approximately 12,000 RFU was much higher than

that of the mTurquoise^{+}PE^{+} population (upper right quadrant) at about 1,000 RFU. Of all the mTurquoise^{+} cells observed, only 6% had any PE signal above background (upper right quadrant compared to the sum of the upper and lower left quadrants). This experiment shows that only a very small amount of protein is captured on cells it wasn't directly secreted from. Therefore, the assay conditions proved to be sufficient to prevent the majority of cells from capturing foreign protein.

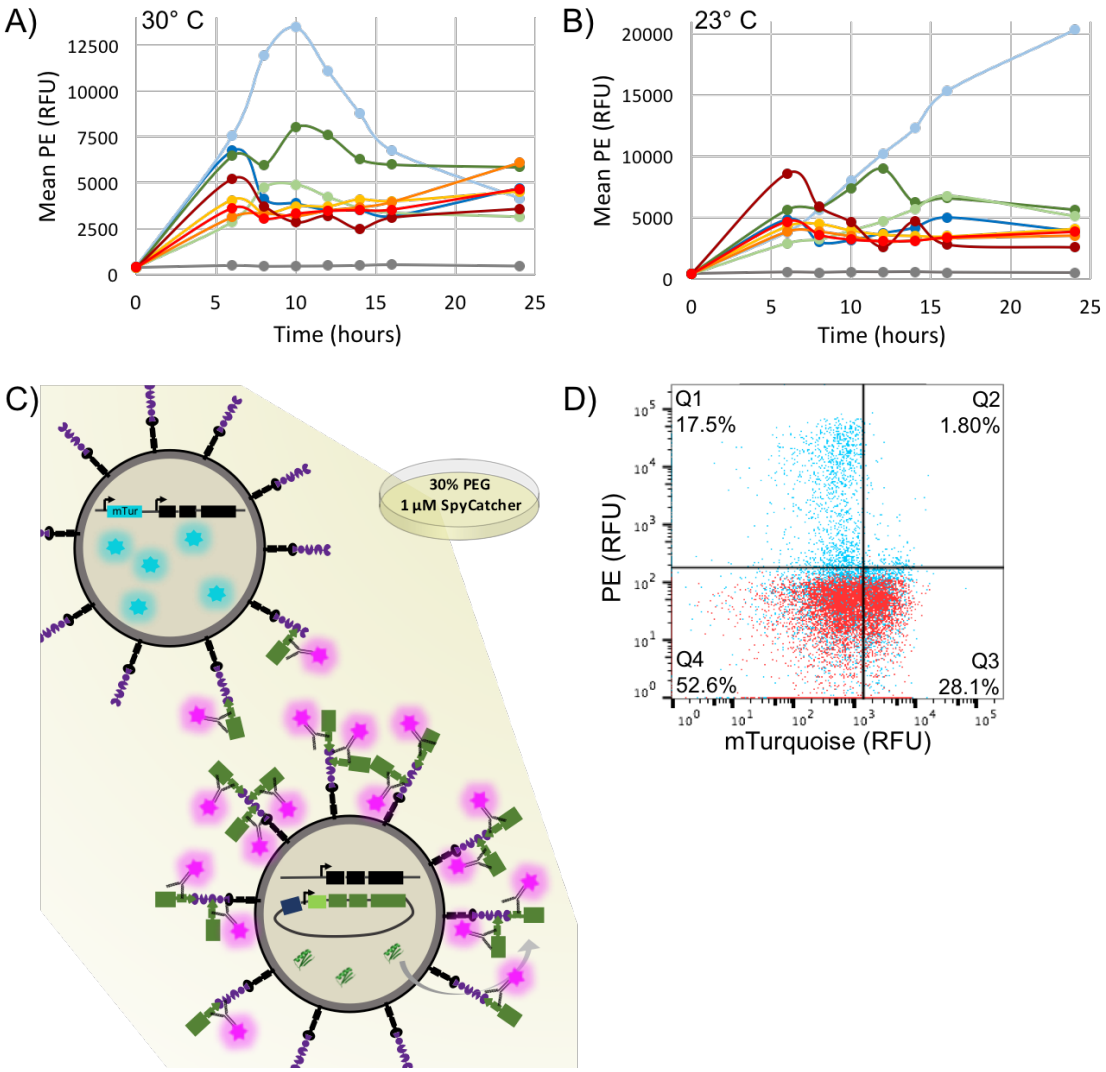

Figure 2.6: SecCap Timecourse & Diffusion. A) SecCap timecourse at 30° C. B) SecCap timecourse at 23° C. Mean fluorescence of PE^{+} population is plotted from 0-24 hours. Nonsecreting cells are in gray, SC1 in dark blue, SC2 in light blue, SC3 in dark green, SC4 in light green, SC5 in yellow, SC6 in orange, SC7 in dark red, SC8 in red. C) Schematic of SecCap diffusion experiment. mTurquoise^{+}, nonsecreting yeast are mixed with mTurquoise^{-}, secreting yeast and induced with 30% PEG and 1 µM Spycatcher. Anti-Flag-PE is used to label captured protein on cells. D) Flow plot with unlabeled cells in red and cells labeled with anti-Flag-PE in blue.

Stability is Well Correlated with SecCap Signal for SC1-8

The SecCap system was successful in separating well-folded proteins from unstable proteins in the small set of clones tested (Figure 2.7A). The lowest average SecCap signal for the well-folded proteins was SC4 at 9,144 RFU while the highest of the unstable set was SC7 at 5,928 RFU (Figure 2.7B). This difference was statistically significant as all four stable proteins (SC1-4) were more than two standard deviations higher than the mean of the unfolded proteins, SC5-8, at 5,463 RFU. While this is only a small set of proteins with a limited range of properties, the difference in SecCap signals between stable and unstable proteins is a promising result for developing the SecCap system as a general tool to select for well-folded proteins and we next sought to expand the applications of the assay.

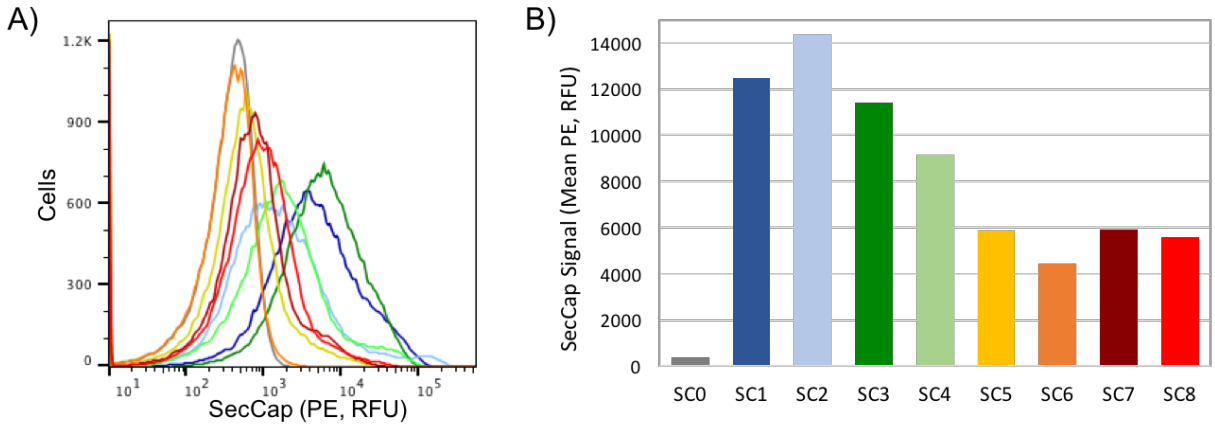

Figure 2.7: SecCap Assay Results for SC1-8. A) Distribution of SecCap signal for SC1-8. Nonsecreting cells (SC0) are in gray, SC1 in dark blue, SC2 in light blue, SC3 in dark green, SC4 in light green, SC5 in yellow, SC6 in orange, SC7 in dark red, SC8 in red. B) The SecCap signal as defined by the mean fluorescence of the PE^+ population for each clone. SC1-4 have SecCap signals that are significantly higher than SC5-8 as determined by the number of standard deviations between the SC1-4 signals and the mean of SC5-8 (5463 ± 703 RFU). The number of standard deviations was calculated by subtracting the mean of the signals of the unstable clones SC5-8 (5463 RFU) from the signal for each individual well-folded clone (SC1=12480 RFU, SC2=14375 RFU, SC3=11416 RFU, SC4=9144 RFU), and then dividing this difference by the standard deviation of SC5-8 (703 RFU). The SecCap signal of a well-folded clone was identified as significantly greater than the unstable clones if it was greater than two standard deviations higher than the mean of SC5-8. Number of standard deviations above mean of unstable clones: SC1=9.98, SC2=12.68, SC3=8.47, SC4=5.24.

SecCap to Measure Target Binding

After secreted protein is captured on the yeast cell surface, the assay should work similarly to traditional yeast display for downstream functional assays. To test this, we used a computationally designed Bfl-1 binding protein, called αBFL1[66]. This protein was secreted and captured on ySecCap in the manner described above and labeled with α-flag-PE to confirm secretion capture. Cells with captured αBFL1 were then incubated with several concentrations of biotinylated Bfl-1. SA-PE was used as a secondary label to give a binding signal by flow (Figure 2.8A). A concentration-dependent binding signal of Bfl1-biot to αBFL1 captured on ySecCap was observed (Figure 2.8B) and titration of Bfl-1 binding to αBFL1 in the SecCap system gave a K_d of ~3 nM (Figure 2.8C) which is in accordance with the *in vitro* measured K_d of αBFL1 in solution of 1.0 ± 0.6 nM via BLI. The SecCap system was demonstrated to be an effective platform for ligand binding assays.

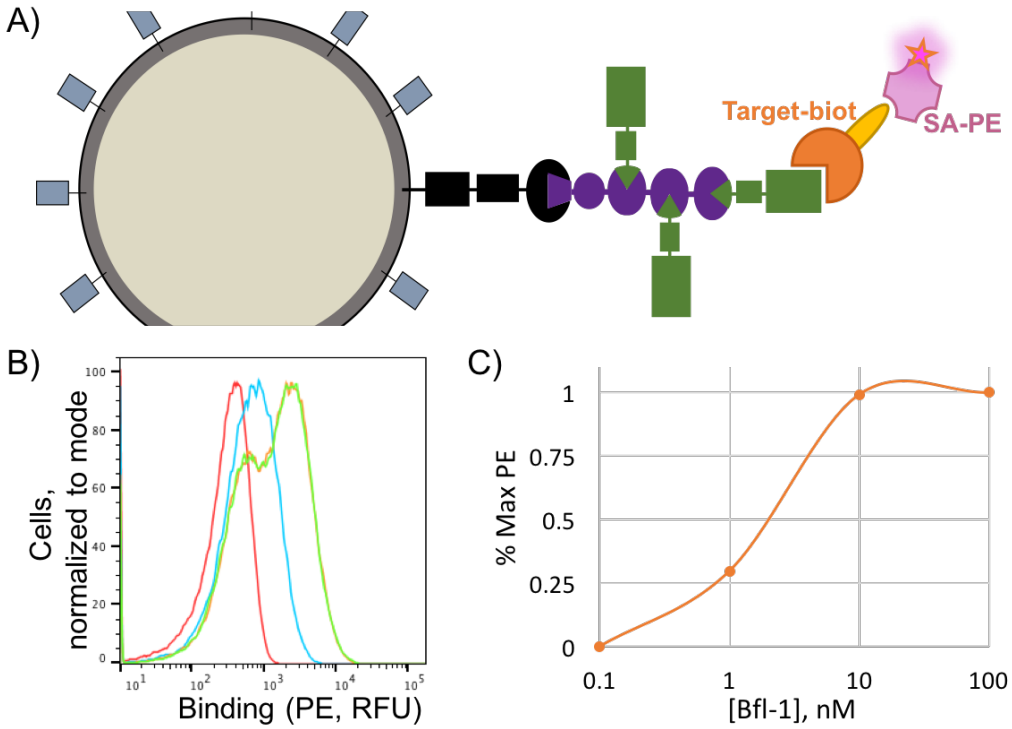

Figure 2.8: SecCap Ligand Binding Assay. A) Schematic of ligand binding assay using SecCap system. B) Binding signal of ySecCap cells secreting αBFL1 labeled with 0 nM Bfl1-biot (red), 1 nM Bfl1-biot (blue), 10 nM Bfl1-biot (orange), 100 nM Bfl1-biot (green). C) Binding curve of Bfl-1 to αBFL1 displayed in SecCap system. K_d is approximately 3 nM.

Protein Production in SecCap

An additional feature of the SecCap system is that it can be used as a protein production platform as well. Instead of labeling with Capture Reagent and performing the stationary plate inductions using ySecCap transformed with pSecCap containing a POI, cultures are induced in a traditional liquid culture without prior Capture Reagent labeling (Figure 2.9A). Secreted Spytagged, Flag-tagged protein can be purified by using α-Flag resin to extract the protein from the spent media (Figure 2.9B). A Tev cleavage site was included between the Flag-tag and the cloning site in pSecCap (shown in Figure 2.9A) to allow easy cleavage of the Spytag and Flag-tag, generating soluble tagless protein. Depending on the secretion level of the protein, the yield is ~0.3 mg per liter culture but could likely be improved for individual proteins. This is only a small amount of protein compared to other expression systems yet is enough to do preliminary *in vitro* characterization to identify which proteins to scale up in other systems. This saves a huge amount of time in cloning and transforming initial hits into another system and is also beneficial in that the same cell line that was used to screen and optimize protein candidates can also be used to generate soluble protein for downstream purposes.

SecCap as a High-Throughput Selection

SecCap was successful in distinguishing stable from unstable proteins on a small set of proteins with known expression and stability properties. We next wanted to see if it would function as a high-throughput screen of protein stability by testing a pool of several thousand genes in parallel. This pool was divided between native proteins with a solved structure in the Protein Data Bank (PDB), scrambled controls of these native proteins, mutants of native protein domains with published thermal stabilities[67–77], previously characterized designed scaffolds[56], these same scaffolds with a small number (1-3) of surface hydrophobic mutations, and these scaffolds with a large number (4-10) of surface hydrophobic mutations (Figure 2.10A). We hypothesized that the native proteins, which had to be generated in a soluble form in some way in order to generate an NMR or crystal structure, would yield a better SecCap signal than their scrambled counterparts that would not be expected to have a single stable conformation by random chance. Because thermal stability is generally correlated with foldedness, we predicted SecCap signal would correlate with *in vitro* stability for sets of well-studied mutants. Finally, we wanted to determine if there was a limit to the

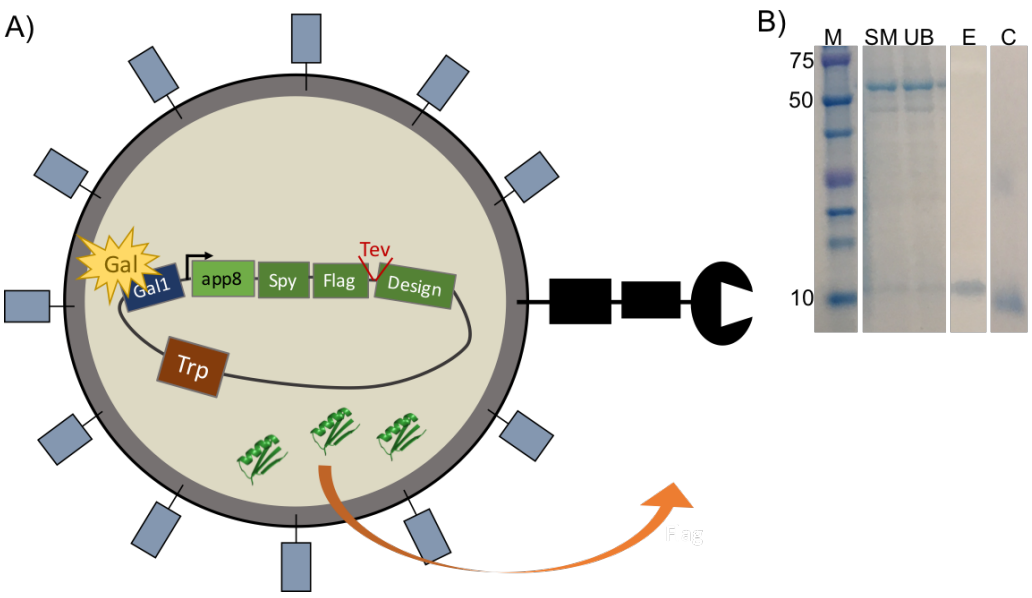

Figure 2.9: Protein Production in SecCap. A) Schematic of protein production via secretion from unlabeled ySecCap. The Tev cleavage site included in pSecCap between the Flag tag and cloning site for the POI is depicted. B) SDS-PAGE of expression and purification fractions of SC2 produced in SecCap. Uncleaved SC2 is 8.5 kDa seen by the faint lower band in the spent media and unbound fractions as well as the single band in the anti-flag elution fraction. The cleaved protein is 5.2 kDa seen by the single band in the cleavage sample. The >50 kDa band in the spent media and unbound fractions is BSA that was added to the induction media to prevent nonspecific binding of the secreted protein to the yeast cell that would lead to a decrease in yield of the purified protein. M=protein ladder, SM=spent media, UB=unbound fraction, E=anti-flag elution, C=cleavage reaction.

number of hydrophobics that could be loaded onto the surface of a stable scaffold protein before it would lose its ability to fold and if this could be determined in our SecCap assay.

Genes encoding all 21,692 sequences were ordered as a single oligonucleotide array synthesis pool and transformed into ySecCap with linearized pSecCap twice to generate two independent biological replicates. After preinduction, the libraries were labeled with Capture Reagent, and plate-induced overnight at 23° C. Recovered cells were labeled with α-flag-PE, and all cells with secretion capture (PE) signal above background were collected via FACS. Four more sequential rounds of sorting were preformed, with increasing sort stringency each time by increasing the PE cutoff for collected cells. Both replicate libraries were each sorted twice for a total of four independent series of sorts. Library genes from the unsorted

libraries and all selected pools were extracted and deep sequenced. Sequencing results for all 4 replicates had very good agreement (R^2=0.72-0.87, Figure 2.10B), and a single score and standard error for each sequence was generated using the Enrich2 method[78], a higher Enrich2 score correlating with a higher level of secretion capture. While there was a large spread of scores between -7 and 1 (Figure 2.10C), we were not able to identify any specific property that correlated with SecCap score. There did not appear to be any correlation between the published stabilities of the mutants we tested and their SecCap score (R^2=0.002, Figure 2.10D). Additionally, all subpools tested showed a similar range of scores, averaging between -4.25 and -4.5 (Figure 2.10E). Because the assay gives consistent results between several replicates for the same sequences, there is likely a parameter(s), perhaps something biological, that is determining secretion level other than protein stability as was anticipated. We are currently using a machine learning approach to identify patterns in the linear amino acid sequences that may correlate with SecCap score with David Gifford's group at MIT.

MATERIALS & METHODS

Yeast Surface Display

SC1-8 were tested in the traditional yeast surface display system by cloning the gene for each protein into pCTCON2 and transforming this plasmid into the *S. cerevisiae* strain EBY100[41]. Cultures were grown from a fresh colony in C-Trp, 2% glucose at 30° C for 24 hours. Expression of the Aga2p-POI construct was induced by transferring cells to SGCAA (2.0% Galactose, 0.67% Yeast nitrogen base, 0.5% Casamino acids, 0.54% Disodium phosphate, 0.86% Monosodium phosphate) for 16 hours at 30° C. Induced cells were labeled in 0.01 mg/ml anti-cmyc-FITC at a cell density of 50 million cells per ml for 10 minutes on ice under aluminum foil. FITC signal was measured on an Accuri C6 flow cytometer in the FL1 channel for 100,000 events per sample.

Bacterial Expressions

E. coli codon-optimized genes for SC1-8 were obtained as gBlocks from IDT and then cloned via Gibson Assembly[79] into the expression vector pCDB24[80]. Proteins were expressed as fusions to the yeast SUMO protein Smt3 in *E. coli* BL21*(DE3) cells (Invitrogen) by first growing a 5 ml starter culture in LB plus 50 µg/ml carbenicillin overnight at 37° C. These starter cultures were used to inoculate 500 ml Studier

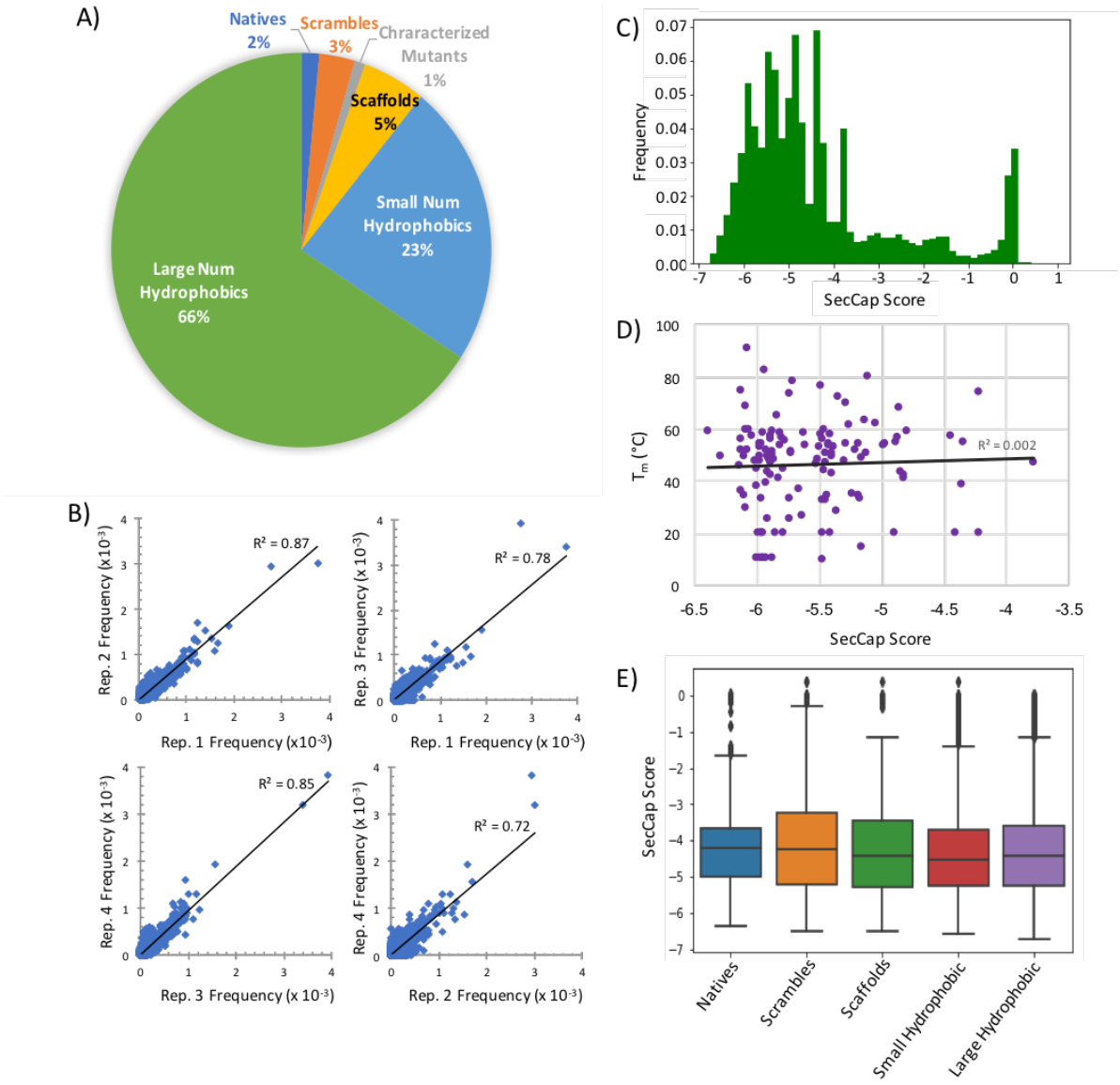

Figure 2.10: SecCap High-Throughput Screen. A) Distribution of library sequences to be tested in SecCap HT screen. B) Pair-wise agreements of sequence frequencies in the sort 2 pools between four replicates. C) Distribution of SecCap scores. D) Correlation between thermal stability of previously characterized mutants and SecCap score. E) SecCap score distributions of subpools.

autoinduction media[81] and incubated with shaking at 30° C for ~24 hours. Cells were harvested by centrifugation at 4000g for 10 minutes, resuspended in 25 ml Lysis Buffer (20 mM Tris pH 8.0, 500 mM NaCl, 30 mM Imidazole, 1 mg/ml lysozyme, 0.1 mg/ml Dnase, 1 mM PMSF), and then lysed using a microfluidizer. Insoluble components were pelleted at 18,000g for 30 minutes at 4° C. Cleared lysate was incubated on a rotator with 5 ml 50% equilibrated Ni-NTA resin at 4° C for 30 minutes and then added to a

gravity flow column for washing and elution in 10 ml Elution Buffer (20 mM Tris pH 8.0, 500 mM NaCl, 1 M Imidazole). Soluble, purified proteins were dialyzed into Cleavage Buffer (150mM NaCl, 20mM NaPi pH 7.4) prior to cleavage with the SUMO protease Ulp1 at a 1:50 mass ratio for 1 hour at 23° C. SDS-PAGE using a Criterion 16.5% Tris Tricine Precast Gel (BioRad) was run on the expression and purification fractions alongside Precision Plus Dual Xtra Prestained Protein Standards (BioRad).

ySecCap Strain

Starting from the EBY100 *S. cerevisiae* strain[41], two identical Snoopcatcher surface expression cassettes were chromosomally integrated in order to achieve strong and consistent expression – one in the ARS314 cassette with a KanMX marker and the other in the YCR043 cassette with a BleoMX marker. Both cassettes contained pGPD, a strong constitutive promoter, followed by Aga2p, a c-myc tag, and a Snoopcatcher[63] domain, as well as a short GS linker between each element. The Gal1 inducible promoter that regulates expression of Aga1 in EBY100 was also replaced with pGPD using a NatMX selection marker to generate the new *S. cerevisiae* strain termed ySecCap.

pSecCap Plasmid

pSecCap was engineered by replacing the DNA fragment between the EcoRI and BglII restriction sites that contains the gene for Aga2p, an HA-tag, cloning site, and cmyc-tag in pCTCON2[41] with the secretion signal app8[64] followed by a Spytag (AHIVMVDAYKPT), Flag-tag (DYKDDDDK), TEV cleavage site (ENLYFQG), and a cloning site for the POI. The plasmid still contains a TRP1 selectable marker and the cloned ORF is under galactose induction via the GAL1 promoter. NdeI and XhoI restriction sites are included around the POI site for easy cloning.

Protein Reagents

The Capture Reagent was generated by cloning a single Snooptag domain followed by three Spycatcher domains with a 10x(GS) linker between each domain along with C-terminal Avi and polyHis tags into pET29b(+) (EMD Millipore). Soluble Spycatcher domain to be used as a blocking agent in the SecCap assay was also cloned into pET29b(+) with a C-terminal polyHis tag. Both proteins were expressed via IPTG induction in Lemo21(DE3) (New England BioLabs) *E. coli* cells at 18° C for ~16 hours. Ni-NTA IMAC followed by size exclusion chromatography (SEC) was used to purify the proteins. Following this protocol, ~6 mg/L of pure, soluble Capture Reagent and ~2 mg/L Spycatcher blocking

agent could be readily made. Both proteins were stored in PBS, 5% glycerol at -80° C and small aliquots thawed as needed. All expressions and purifications were performed by the UW Institute for Protein Design (IPD) Protein Core.

SecCap Assay

Genes for SC1-8 and αBFL1 were introduced into ySecCap along with linearized pSecCap via chemical transformation[82]. Transformed ySecCap cultures were shaken in C-Trp selective yeast media with 2% glucose for 24-48 hours at 30° C. Cultures were then transferred to C-Trp supplemented with 2% galactose at an OD600 of 1.0 and shaken at 30° C for 16-24 hours for the pre-induction step. Pre-induced cells were washed in PBSF (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 2.7 mM KCl, 137 mM NaCl, 1% [w/v] BSA, pH 7.4) before labeling with 5 μM Capture Reagent at a cell density of 50 million cells/ml at 23° C for 1 hour with rotation. After washing in PBSF, labeled cells were resuspended in YPG, 30% [w/v] 8 kDa PEG, 0.025% [w/v] BSA with 1 μM Spycatcher blocking agent at a cell density of OD600 0.1. Resuspended cells were plated at 0.13 ml/cm^2 which resulted in a layer of ~1 mm in depth for adequate aerobic conditions. Varying numbers of cells could be plated by utilizing different sterile culture plates (10 cm petri dish = ~30 million cells, 6-well plate with 34.8 mm wells = ~3.75 million cells). For the timecourse, plate inductions were incubated without disturbance at 30° C or 23° C for up to 24 hours. For all other SecCap experiments, plate inductions were incubated at 23° C for 15-17 hours. Cells were recovered by carefully adding 1.5x volume of 1 μM Spycatcher in PBSF to plated culture while keeping the plate completely still, incubating for 5 minutes, and then transferring solution to a sterile conical tube. Plates were washed with an additional 0.8x volume 1 μM Spycatcher, PBSF, added to same tube, and then inverted to homogenize solution. An equal volume of PBSF was added to further reduce the viscosity of the solution. Cells were pelleted by centrifugation at 4500g for 3 minutes. Recovered cells were washed two additional times in 1 ml PBSF to ensure all the PEG was removed. After resuspension in an adequate volume of PBSF, the OD600 of the solution was measured to determine the number of cells recovered. To measure the SecCap signal, cells were incubated with 1:100 α-Flag-PE (Miltenyi cat #130-101-576) at 50 million cells/ml for 10 minutes on ice under aluminum foil and then analyzed on an Accuri C6 flow cytometer.

Diffusion Experiment

An mTurquoise^{+} cell line was created by integrating a gene for mTurquoise preceded by the GPD constitutive yeast promoter into the genome of ySecCap. These cells were used without pSecCap transformation, and a culture was grown in YPD. A culture of ySecCap transformed with SC2/pSecCap was grown separately in C-Trp, 2% glucose. 15 million cells of each were pooled and transferred to 1 ml C-Trp, 2% galactose. The combined preinduction culture was shaken at 30° C for 16 hours. Preinduced cells were labeled with 5 μM Capture Reagent as described above and plated in YPG, 30% [w/v] 8 kDa PEG, 0.025% [w/v] BSA with 1 μM Spycatcher. Cells were carefully recovered after 17 hours at 23° C using Recovery Buffer consisting of 1 μM Spycatcher in PBSF and labeled with α-Flag-PE. mTurquoise and PE fluorescence was measured on a MACSQuant flow cytometer.

Binding Experiment

After performing the SecCap assay as above, aliquots of ySecCap with captured αBFL1 were incubated in 1 nM, 10 nM and 100 nM biotinylated Bfl-1 at 50 million cells/ml at 23° C for 1 hour. Cells were washed with 1 ml PBSF and then a secondary labeling was performed with 0.01 mg/ml SA-PE (Thermo Fisher) for 10 minutes on ice under aluminum foil. PE fluorescence was measured on an Accuri C6 flow cytometer to analyze the binding signal.

High-Throughput Experiment

An oligonucleotide array pool (Agilent) containing the library of 21,692 sequences to be tested were amplified, and two independent transformations were conducted with linearized pSecCap via electroporation[14] to yield two replicate libraries. Library 1 had a transformation efficiency of 1.65 million, and library 2 had a transformation efficiency of 5.5 million which should be sufficient to cover the entire library. The SecCap assay was conducted on the libraries as described in the “SecCap Assay” section above. Five consecutive sorts were conducted on a Sony SH800 Cell Sorter, growing up, preinducing, and plate inducing the collected cells between each sort. All cells above the following PE fluorescence cutoffs were collected for each sort: sort 1 = 400 RFU, sort 2 = 600 RFU, sort 3 = 1000 RFU, sort 4 = 3000 RFU, and sort 5 = 7000 RFU. Two series of sorts were conducted for each library, giving four independent sets of replicates.

DNA was extracted and prepared for deep sequencing for each naïve library and selected pool. Approximately 100 million cells for each pool were lysed by incubating with zymolyase (Zymo Research) for 4 hours at 37° C and performing two freeze-thaw cycles. Plasmids were isolated using a Qiagen DNA Spin Column and subsequent digestion with Exonuclease I and Lambda Exonuclease (New England BioLabs) to remove any remaining genomic DNA. Genes were extracted and adaptors and pool indexes added by two subsequent qPCR steps with custom primers. The gene pools were sequenced by performing two paired-end 150-cycle reads and a 6-cycle index read on an Illumina Miseq. Reads were assembled using PEAR[83] and sequence counts for each pool calculated. Algorithms from the Enrich2 software suite[78] were used to calculate individual scores for each replicate based on the enrichment of a given sequence in each selection and then combine the replicate scores into a single final score and standard error for each sequence.

SecCap Protein Production

A 500 ml ySecCap culture was grown to saturation in C-Trp, 2% glucose by shaking at 30° C for ~24 hours. Protein expression was induced by transferring the cells to 500 ml YPG, 0.025% BSA and shaking at 30° C for 24 hours. Cells were harvested by centrifugation at 3000g for 3 minutes. Media was transferred to a fresh container and incubated with 1 ml 50% equilibrated α-Flag resin (Bimake cat #B23101) on a rotator at 4° C for 2 hours. Resin was separated by centrifugation at 1500g for 3 minutes. The media was carefully transferred without disturbing the settled resin to a fresh container as the “Unbound” fraction. The resin was transferred to a gravity flow column for washing with TBS (20 mM Tris pH 8.0, 150 mM NaCl). Bound protein was eluted in 1.5 ml 20 mM Tris pH 8.0, 150 mM NaCl, 100 µg/ml Flag peptide (Sigma, cat. #F3290). “Spent Media,” “Unbound,” “Elution,” and “Cleaved” fractions were run alongside Precision Plus Dual Xtra Prestained Protein Standards (BioRad) on a Criterion 16.5% Tris Tricine Precast Gel (BioRad) by SDS-PAGE. Approximately 0.3 mg eluted protein was dialyzed into 50mM Tris pH 8.0, 0.5mM EDTA, 1mM DTT prior to cleavage with 30 µg Tev protease (made in-house by the UW IPD Protein Core) at 4° C for 19 hours.

DISCUSSION

While SecCap has proven to be a robust, reproducible assay, it remains unclear what property it is selecting for. SC1-8 could be separated by their known stabilities in the assay, but this pattern did not appear to hold true in the high-throughput assay. This is most clear with the mutants of natural protein domains that were included in the pool. Published thermal stabilities ranged from melting temperatures around 50° C for the wild type proteins to hyperstable mutants up to a T_m of 90.6° C. Some mutations were so destabilizing that soluble protein was not able to be generated for characterization. We hypothesized that a protein with very low stability *in vitro* would have more trouble folding and be recognized and destroyed by the ERAD pathway in the yeast cell as is known to happen with terminally misfolded proteins[84,85]. This would lead to very little protein being secreted and would result in a lower score in our SecCap assay. Contrarily, a highly thermostable protein would be more likely to be found in the folded state and sorted for cellular export in the ER[53], leading to a high SecCap signal. However, we did not observe any correlation between thermal stability and secretion from yeast in our assay. Furthermore, scrambled sequences performed no worse than native proteins with known structures. Because the assay consistently gives very similar levels of secretion capture for the sequences tested, some of which give a very low signal and some very high, we hypothesize that there could be factors other than conformational stability that contribute to protein secretion from yeast.

Some possible factors that may contribute to signal in the SecCap assay are the presence of protease cleavage sites, overall charge, membrane sequestration, or even a more complex stability component than initially anticipated. Kex2 is a major protease in the yeast secretory pathway and was recently found to play a particularly significant role in removing proteins containing its cleavage sequence in the yeast surface display system[86]. However, sequences containing the Kex2 cleavage site were found to have the same distribution of scores as the sequences that did not contain the site (Figure 2.11A-B). While Kex2 did not explain the SecCap scores seen in the high-throughput pool, there are other proteases in the yeast secretory pathway that could play a role. Net charge has been found to be a factor in secretion efficiency in other systems[87]. Another possibility is the presence of particularly hydrophobic peptides that might interact with membranes and sequester the protein from being secreted. These hypotheses are currently being examined using a machine learning approach to look for patterns in the dataset that

correlate with SecCap score. If a significant correlation can be found between SecCap score and some identifiable protein property, the SecCap assay is potentially a very powerful and complete platform for the screening, optimization, and production of secreted proteins from yeast.

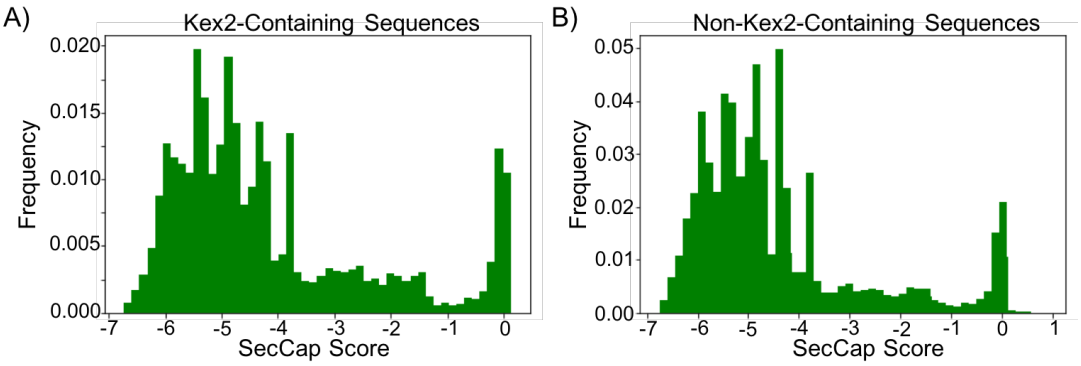

Figure 2.11: Kex2 Analysis of SecCap Dataset. Distribution of SecCap scores for sequences containing a Kex2 cleavage site (A) and those without (B).

Conclusion

While there have been a few examples of *de novo* binders with alpha helical interfaces[14], this is the first *de novo* binder with an all beta protein interface to my knowledge. Beta sheets are particularly difficult because of their natural propensity to aggregation[88]. This is expounded by the fact that designed interfaces tend to be very hydrophobic due to the difficulty of incorporating correct hydrogen bonds and the inability to place explicit water molecules during design[89]. An interface that is mostly hydrophobic is potentially more destabilizing to beta interfaces than to helical interfaces. Due to the alternating geometry of beta strands, every other residue of the strand is already hydrophobic as it is involved in core packing. If this strand is also involved in binding, the solvent exposed residues will also likely be hydrophobic, making every residue along the peptide for the beta strand hydrophobic. In the absence of polar residues to guide the initial chain collapse and formation of secondary structure[90], folding in homo-nonpolar polypeptide regions would be completely reliant on ideal core packing[91] and the secondary structure propensity of each amino acid[92] to drive folding towards the intended structure. In accordance with these observations, I found core stabilization through the use of disulfide bonds to be very important to obtaining a soluble PD-1 binder that folds as designed.

Several of my binders identified through yeast surface display over the years have yielded proteins with low or no soluble expression that were very prone to aggregation. This highlights the need for better methods that could select for not only functional but also well-folded and stable proteins. I set out to develop a method, SecCap, to do just this. While initial development of the system using a small set of proteins of known stabilities proved to be successful by providing a flow cytometry signal that correlated with foldedness of the proteins, this correlation did not hold true in a high-throughput experiment of over 21,000 with both known and unknown stabilities. Using a cellular expression readout may include more variables than just protein stability and developing a high-throughput selection for this property continues to be an unsolved problem. Until a method of simultaneous selection for protein function and stability is developed, binders identified in high-throughput yeast display assays must continue to be monitored closely for soluble expression and stability in a separate system as was done with the designed PD-1 binder, GR918.

There are potentially many applications for a small PD-1 inhibitor in the field of cancer immunotherapy. On its own, GR918 would likely have improved tissue penetrance compared to the current antibody inhibitors[29], which could result in stronger potency against solid tumors as even solid tumors have been shown to contain tumor-infiltrating lymphocytes (TIL)[93]. A targeted therapeutic could be made by fusing the modular anti-PD-1 domain with a protein domain that binds a tumor cell marker. This might lead to fewer toxic side effects caused by general over-stimulation of the immune system. Checkpoint inhibitors are currently being explored for use in combination with thousands of other immuno- and chemotherapies[30,94]. Multi-specific compounds could be generated by fusing GR918 with other small binding domains against targets that prove to be synergistic. A variety of cell-based platforms are being explored as the next generation of immunotherapeutics[95–97]. These platforms have the benefit of being self-sustaining and are targetable to any tumor antigen. As has been done previously with CAR T cells[98], the gene for GR918 could be introduced to the engineered immune cell with an inducible promoter, allowing specific expression and secretion only upon target activation in the tumor microenvironment. This would have a two-fold effect of the engineered cell response as well as reactivation of the endogenous tumor-specific T cells through PD-1. At only 4.5 kDa, any GR918 that diffuses away from the tumor site, will be cleared very quickly by the kidneys[99], making it ideal for a targeted cancer immunotherapeutic. However, this could act against its usefulness if used in monotherapy, and ways to reduce its clearance from plasma, such as polyethylene-glycol binding, may need to be explored to make it stable in circulation.

PD-1 is also known to play a role in several autoimmune disorders caused by a decrease in function of the pathway[100]. To restore this function, a molecule would need to act as an agonist to signal through the receptor. Currently, there are no known agonists and even soluble forms of the natural ligands act as inhibitors[26]. Despite the current excitement surrounding this pathway due to its success as a target for cancer immunotherapy, little has been done to identify the mechanism of signaling. One study has shown that signaling is caused by clustering of the receptors at the immune synapse resulting from the multivalent interaction between PD-1 on the T cell and PD-L1/2 on another cell[15]. Additionally, anti-PD-1 antibodies that inhibit the receptor in solution have agonistic properties when coupled to a bead[38]. Therefore, I hypothesize that an agonist could be created by generating a multivalent PD-1 binder. The small, modular GR918 binding domain can be easily combined in different numbers and orientations using computationally

designed stable oligomers of varying geometries and oligomeric states. Studies of the GR918 monomer and various oligomers is currently underway to determine their function on T cells. As an ultra-stable, modular PD-1 binding domain, GR918 has limitless applications to the creation of new therapeutics for cancer and autoimmunity.

Bibliography

1. Kortemme, T. & Baker, D. Computational design of protein-protein interactions. *Curr. Opin. Chem. Biol.* **8,** 91–7 (2004).
2. Li, Y., Zhang, X. & Cao, D. The Role of Shape Complementarity in the Protein-Protein Interactions. *Sci. Rep.* **3,** 3271 (2013).
3. Boyken, S. E. *et al.* De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352,** 680–687 (2016).
4. Lin, Y.-R. *et al.* Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci.* **112,** E5478–E5485 (2015).
5. Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491,** 222–227 (2012).
6. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302,** 1364–8 (2003).
7. Marcos, E. *et al.* Principles for designing proteins with cavities formed by curved β sheets. *Science* **355,** 201–206 (2017).
8. Bhardwaj, G. *et al.* Accurate de novo design of hyperstable constrained peptides. *Nature* **538,** 329–335 (2016).
9. Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346,** 481–485 (2014).
10. Leaver-Fay, A. *et al.* Rosetta3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487,** 545–574 (2011).
11. Ghindilis, A. L. *et al.* CombiMatrix oligonucleotide arrays: Genotyping and gene expression assays employing electrochemical detection. *Biosens. Bioelectron.* **22,** 1853–1860 (2007).
12. Saaem, I., Ma, K. S., Marchi, A. N., LaBean, T. H. & Tian, J. In situ Synthesis of DNA Microarray on Functionalized Cyclic Olefin Copolymer Substrate. *ACS Appl. Mater. Interfaces* **2,** 491–497 (2010).
13. Klein, J. C. *et al.* Multiplex pairwise assembly of array-derived DNA oligonucleotides. *Nucleic Acids Res.* **44,** e43 (2016).
14. Chevalier, A. A. *et al.* Massively parallel de novo protein design for targeted therapeutics. *Nat. Publ. Gr.* **550,** 74–79 (2017).
15. Yokosuka, T. *et al.* Programmed cell death 1 forms negative costimulatory microclusters that directly inhibit T cell receptor signaling by recruiting phosphatase SHP2. *J. Exp. Med.* **209,** 1201–17 (2012).
16. Zhang, X. *et al.* Structural and functional analysis of the costimulatory receptor programmed death-1. *Immunity* **20,** 337–47 (2004).
17. Cheng, X. *et al.* Structure and interactions of the human programmed cell death 1 receptor. *J. Biol. Chem.* **288,** 11771–85 (2013).
18. Lázár-Molnár, E. *et al.* Structure-guided development of a high-affinity human Programmed Cell Death-1: Implications for tumor immunotherapy. *EBioMedicine* **17,** 30–44 (2017).
19. Lin, D. Y.-W. *et al.* The PD-1/PD-L1 complex resembles the antigen-binding Fv domains of antibodies and T cell receptors. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 3011–6 (2008).
20. Lázár-Molnár, E. *et al.* Crystal structure of the complex between programmed death-1 (PD-1) and its ligand PD-L2. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 10483–8 (2008).
21. Fife, B. T. *et al.* Interactions between PD-1 and PD-L1 promote tolerance by blocking the TCR-

induced stop signal. *Nat. Immunol.* **10,** 1185–92 (2009).

22. Iwai, Y. *et al.* Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. *Proc. Natl. Acad. Sci.* **99,** 12293–12297 (2002).

23. Hofmeyer, K. A., Jeon, H. & Zang, X. The PD-1/PD-L1 (B7-H1) pathway in chronic infection-induced cytotoxic T lymphocyte exhaustion. *Journal of Biomedicine and Biotechnology* **2011,** 451694 (2011).

24. Tocheva, A. S. & Mor, A. Checkpoint Inhibitors: Applications for Autoimmunity. *Curr. Allergy Asthma Rep.* **17,** 72 (2017).

25. Kieke, M. C., Cho, B. K., Boder, E. T., Kranz, D. M. & Wittrup, K. D. Isolation of anti-T cell receptor scFv mutants by yeast surface display. *Protein Eng.* **10,** 1303–10 (1997).

26. Dai, S., Jia, R., Zhang, X., Fang, Q. & Huang, L. The PD-1/PD-Ls pathway and autoimmune diseases. *Cellular Immunology* **290,** 72–79 (2014).

27. Abril-Rodriguez, G. & Ribas, A. SnapShot: Immune Checkpoint Inhibitors. *Cancer Cell* **31,** 848–848.e1 (2017).

28. Lee, C. M. & Tannock, I. F. The distribution of the therapeutic monoclonal antibodies cetuximab and trastuzumab within solid tumors. *BMC Cancer* **10,** 255 (2010).

29. Maute, R. L. *et al.* Engineering high-affinity PD-1 variants for optimized immunotherapy and immuno-PET imaging. *Proc. Natl. Acad. Sci.* **112,** E6506–E6514 (2015).

30. Qiao, M., Jiang, T., Ren, S. & Zhou, C. Combination Strategies on the Basis of Immune Checkpoint Inhibitors in Non-small- cell Lung Cancer: Where do we stand? *Clin. Lung Cancer* 1–11 (2017). doi:10.1016/j.cllc.2017.06.005

31. Alvarez-Cienfuegos, A. *et al.* Intramolecular trimerization, a novel strategy for making multispecific antibodies with controlled orientation of the antigen binding domains. *Scientific Reports* **6**, 28643 (2016).

32. Correia, B. E. *et al.* Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope. *Structure* **18,** 1116–1126 (2010).

33. Correia, B. E. *et al.* Proof of principle for epitope-focused vaccine design. *Nature* **507**, 201-206 (2014).

34. Bywater, R. P. Prediction of protein structural features from sequence data based on Shannon entropy and kolmogorov complexity. *PLoS One* **10,** e0119306 (2015).

35. Bandaranayake, A. D. *et al.* Daedalus: A robust, turnkey platform for rapid production of decigram quantities of active recombinant proteins in human cell lines using novel lentiviral vectors. *Nucleic Acids Res.* **39,** e143 (2011).

36. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci.* **109,** E690–E697 (2012).

37. Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science (80-.).* **353,** 389–394 (2016).

38. Bennett, F. *et al.* Program Death-1 Engagement Upon TCR Activation Has Distinct Effects on Costimulation and Cytokine-Driven Proliferation: Attenuation of ICOS, IL-4, and IL-21, But Not CD28, IL-7, and IL-15 Responses. *J. Immunol.* **170,** 711–718 (2003).

39. Bazzoli, A., Kelow, S. P. & Karanicolas, J. Enhancements to the Rosetta energy function enable improved identification of small molecules that inhibit protein-protein interactions. *PLoS One* **10**, e0140359 (2015).

40. Hoover, D. M. & Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30,** e43 (2002).

41. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1,** 755–768 (2006).

42. Chen, C., Huang, Q.-L., Jiang, S.-H., Pan, X. & Hua, Z.-C. Immobilized protein ZZ, an affinity tool for immunoglobulin isolation and immunological experimentation. *Biotechnol. Appl. Biochem.* **45,** 87 (2006).

43. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

44. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

45. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

46. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68,** 352–367 (2012).

47. Fallas, J. A. *et al.* Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9,** 353–360 (2016).

48. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15,** 553–557 (1997).

49. Cherf, G. M. & Cochran, J. R. in *Applications of Yeast Surface Display for Protein Engineering. Methods in Molecular Biology (Clifton, N.J.)* **1319**, 155-175 (2015).

50. Esteban, O. & Zhao, H. Directed evolution of soluble single-chain human class II MHC molecules. *J. Mol. Biol.* **340,** 81–95 (2004).

51. Kim, Y., Bhandari, R., Cochran, J. R., Kuriyan, J. & Wittrup, K. D. Directed evolution of the epidermal growth factor receptor extracellular domain for expression in yeast. *Proteins Struct. Funct. Bioinforma.* **62,** 1026–1035 (2006).

52. Park, S. *et al.* Limitations of yeast surface display in engineering proteins of high thermostability. *Protein Eng. Des. Sel.* **19,** 211–217 (2006).

53. Suyama, K., Hori, M., Gomi, K. & Shintani, T. Fusion of an intact secretory protein permits a misfolded protein to exit from the endoplasmic reticulum in yeast. *Biosci. Biotechnol. Biochem.* **78,** 49–59 (2014).

54. Procko, E. *et al.* A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell* **157,** 1644–1656 (2014).

55. Traxlmayr, M. W. & Shusta, E. V. in *Methods in molecular biology (Clifton, N.J.)* **1575,** 45–65 (2017).

56. Rocklin, G. J. *et al.* Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168-175 (2017).

57. James, L. C. & Tawfik, D. S. The specificity of cross-reactivity: Promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. *Protein Sci.* **12,** 2183–2193 (2009).

58. Blazic, M. *et al.* Yeast surface display for the expression, purification and characterization of wild-type and B11 mutant glucose oxidases. *Protein Expr. Purif.* **89,** 175–180 (2013).

59. Rakestraw, J. A., Baskaran, A. R. & Wittrup, K. D. A Flow Cytometric Assay for Screening Improved Heterologous Protein Secretion in Yeast. *Biotechnology Progress* **22,** 1200–1208 (2006).

60. Rhiel, L. *et al.* REAL-select: Full-length antibody display and library screening by surface capture on yeast cells. *PLoS One* **9,** e114887 (2014).

61. Shaheen, H. H. *et al.* A Dual-Mode Surface Display System for the Maturation and Production of

Monoclonal Antibodies in Glyco-Engineered Pichia pastoris. *PLoS One* **8,** e70190 (2013).

62. Rakestraw, J. A., Aird, D., Aha, P. M., Baynes, B. M. & Lipovšek, D. Secretion-and-capture cell-surface display for selection of target-binding proteins. *Protein Eng. Des. Sel.* **24,** 525–530 (2011).

63. Veggiani, G. *et al.* Programmable polyproteams built using twin peptide superglues. *Proc. Natl. Acad. Sci.* **113,** 1202–1207 (2016).

64. Rakestraw, J. A., Sazinsky, S. L., Piatesi, A., Antipov, E. & Wittrup, K. D. Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in Saccharomyces cerevisiae. *Biotechnol. Bioeng.* **103,** 1192–1201 (2009).

65. Ma, H., Kunes, S., Schatz, P. J. & Botstein, D. Plasmid construction by homologous recombination in yeast. *Gene* **58,** 201–216 (1987).

66. Berger, S. *et al.* Computationally designed high specificity inhibitors delineate the roles of BCL2 family proteins in cancer. *Elife* **5,** e20352 (2016).

67. Jiang, X. I. N., Kowalski, J. & Kelly, J. W. Increasing protein stability using a rational approach combining sequence homology and structural alignment: Stabilizing the WW domain. *Protein Sci.* **10,** 1454–1465 (2001).

68. Araya, C. L. *et al.* A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc. Natl. Acad. Sci.* **109,** 16858–16863 (2012).

69. Xiao, S., Bi, Y., Shan, B. & Raleigh, D. P. Analysis of core packing in a cooperatively folded miniature protein: The ultrafast folding villin headpiece helical subdomain. *Biochemistry* **48,** 4607–4616 (2009).

70. Xiao, S. *et al.* Rational modification of protein stability by targeting surface sites leads to complicated results. *Proc. Natl. Acad. Sci.* **110,** 11337–11342 (2013).

71. Neuweiler, H. *et al.* Downhill versus Barrier-Limited Folding of BBL 2: Mechanistic Insights from Kinetics of Folding Monitored by Independent Tryptophan Probes. *J. Mol. Biol.* **387,** 975–985 (2009).

72. Jäger, M., Dendle, M. & Kelly, J. W. Sequence determinants of thermodynamic stability in a WW domain - An all-β-sheet protein. *Protein Sci.* **18,** 1806–1813 (2009).

73. Hatami, A., Monjazeb, S., Milton, S. & Glabe, C. G. Familial Alzheimer's disease mutations within the amyloid precursor protein alter the aggregation and conformation of the amyloid-β peptide. *J. Biol. Chem.* **292,** 3172–3185 (2017).

74. Caine, J. *et al.* Alzheimer's Aβ fused to green fluorescent protein induces growth stress and a heat shock response. *FEMS Yeast Res.* **7,** 1230–1236 (2007).

75. Kim, W. & Hecht, M. H. Sequence determinants of enhanced amyloidogenicity of Alzheimer A??42 peptide relative to A??40. *J. Biol. Chem.* **280,** 35069–35076 (2005).

76. Villar-Piqué, A. & Ventura, S. Protein aggregation propensity is a crucial determinant of intracellular inclusion formation and quality control degradation. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833,** 2714–2724 (2013).

77. Nair, S., Traini, M., Dawes, I. W. & Perrone, G. G. Genome-wide analysis of Saccharomyces cerevisiae identifies cellular processes affecting intracellular aggregation of Alzheimer's amyloid-β42: importance of lipid homeostasis. *Mol. Biol. Cell* **25,** 2235–49 (2014).

78. Rubin, A. F. *et al.* A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* **18,** 150 (2017).

79. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).

80. Rocklin, G. J. *et al.* Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science (80-.).* **357,** 168–175 (2017).

81. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).

82. Bergkessel, M. & Guthrie, C. Chemical transformation of yeast. *Methods Enzymol.* **529,** 311–320 (2013).

83. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30,** 614–620 (2014).

84. Kostova, Z. & Wolf, D. H. For whom the bell tolls: Protein quality control of the endoplasmic reticulum and the ubiquitin-proteasome connection. *EMBO Journal* **22,** 2309–2317 (2003).

85. Whyteside, G. *et al.* Native-State stability determines the extent of degradation relative to secretion of protein variants from Pichia pastoris. *PLoS One* **6,** e22692 (2011).

86. Li, Q. *et al.* Profiling Protease Specificity: Combining Yeast ER Sequestration Screening (YESS) with Next Generation Sequencing. *ACS Chem. Biol.* **12,** 510–518 (2017).

87. Widmaier, D. M. & Voigt, C. A. Quantification of the physiochemical constraints on the export of spider silk proteins by Salmonella type III secretion. *Microb. Cell Fact.* **9,** 78 (2010).

88. Bellesia, G. & Shea, J.-E. Effect of beta-sheet propensity on peptide aggregation. *J. Chem. Phys.* **130,** 145103 (2009).

89. Benjamin Stranges, P. & Kuhlman, B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **22,** 74–82 (2013).

90. Udgaonkar, J. B. Polypeptide chain collapse and protein folding. *Archives of Biochemistry and Biophysics* **531,** 24–33 (2013).

91. Rami, B. R., Krishnamoorthy, G. & Udgaonkar, J. B. Dynamics of the core tryptophan during the formation of a productive molten globule intermediate of barstar. *Biochemistry* **42,** 7986–8000 (2003).

92. Fujiwara, K., Toda, H. & Ikeguchi, M. Dependence of α-helical and β-sheet amino acid propensities on the overall protein fold type. *BMC Struct. Biol.* **12,** 18 (2012).

93. Kovacsovics-Bankowski, M. *et al.* Detailed characterization of tumor infiltrating lymphocytes in two distinct human solid malignancies show phenotypic similarities. *J. Immunother. Cancer* **2,** 38 (2014).

94. Bu, X., Mahoney, K. M. & Freeman, G. J. Learning from PD-1 Resistance: New Combination Strategies. *Trends in Molecular Medicine* **22,** 448–451 (2016).

95. Gill, S., Maus, M. V. & Porter, D. L. Chimeric antigen receptor T cell therapy: 25 years in the making. *Blood Reviews* **30,** 157–167 (2016).

96. Chiang, C. L.-L., Balint, K., Coukos, G. & Kandalaft, L. E. Potential approaches for more successful dendritic cell-based immunotherapy. *Expert Opin. Biol. Ther.* **15,** 569–582 (2015).

97. Li, Y. *et al.* NK cell-based cancer immunotherapy: from basic biology to clinical application. *Science China Life Sciences* **58,** 1233–1245 (2015).

98. Chmielewski, M., Hombach, A. A. & Abken, H. Of CARs and TRUCKs: Chimeric antigen receptor (CAR) T cells engineered with an inducible cytokine to modulate the tumor stroma. *Immunol. Rev.* **257,** 83–90 (2014).

99. Tibbitts, J., Canter, D., Graff, R., Smith, A. & Khawli, L. A. Key factors influencing ADME properties of therapeutic proteins: A need for ADME characterization in drug discovery and development. *mAbs* **8,** 229–245 (2016).

100. Francisco, L., Sage, P. & Sharpe, A. H. The PD-1 Pathway in Tolerance and Autoimmunity. *Immunol. Rev.* **236,** 219–242 (2010).