DISSERTATION

PROTEIN RESURFACING TO IDENTIFY MACROMOLECULAR ASSEMBLIES

Submitted by

Alex Michael Chapman

Department of Chemistry

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2016

Doctoral Committee:

Advisor: Brian McNaughton

Alan Van Orden
Tomislav Rovis
Nick Fisk
Olve Peersen

ABSTRACT

PROTEIN RESURFACING TO IDENTIFY MACROMOLECULAR ASSEMBLIES

Protein engineering is an emerging discipline that dovetails modern molecular biology techniques with high-throughput screening, laboratory evolution technologies, and computational approaches to modify sequence, structure, and in some cases, function and properties of proteins. The ultimate goal is to develop new proteins with improved or designer functions for use in biotechnology, medicine and basic research. One way to engineer proteins is to change their solvent exposed regions through focused or random 'protein resurfacing'. Here, I describe several approaches towards the development of synthetic proteins with new properties and function, including resistance to aggregation, increased solubility, and potent and selective macromolecule recognition. The first part of this thesis describes the use of protein supercharging to develop a split-superpositive GFP reassembly assay that is more efficient, faster, and more robust than previously described variants, largely due to increased resistance to aggregation. The second part of this thesis describes the use of shape complementarity, protein resurfacing, and high-throughput screening to evolve the first potent and selective protein-based inhibitor of the oncoprotein gankyrin. Concomitant with this work, I also describe a protein grafting strategy to identify a soluble mimic of S6 ATPase, which is subsequently used to characterize the S6 ATPase/gankyrin interaction by isothermal titration calorimetry.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Professor Brian McNaughton. There is no way I can thank him enough for his dedication to ensure my success as a graduate student, as well as my future as a chemist. His passion and excitement regarding scientific discovery, as well as his tireless work ethic, was nothing short of inspirational. I believe that I discovered my true passion for chemical biology during my time under Brian's mentorship, which will no doubt shape the rest of my scientific career.

I would like to acknowledge my committee members, Prof. Rovis, Prof. Van Orden, Prof. Fisk, and Prof. Peersen, for pushing me during both my preliminary and final examinations. I thank Prof. Jun Fujita providing detailed insight into gankyrin biochemistry. I would also like to thank Prof. Ron Raines (Wisconsin) and Prof. David Liu (Harvard) for providing images for Figure 1.7 and 1.8. I thank Dr. Muneera Beach for providing detailed insight into much of the ITC data in this dissertation.

I would like to thank the members of the McNaughton lab, both past and present, who have contributed to my scientific growth over the last five years, mainly through long, detailed discussions at the white board. In particular, I would like to acknowledge Dr. Brett Blakeley for his mentorship, as well as countless hours of intense scientific discussion that helped shape my approach to research.

Finally, I would like to acknowledge family and friends for their support. I thank my mom Diane, my dad Jeff, and my sister Ryan for their continual, unwavering support. I would also like to thank my girlfriend Suzie for her support and for always being there to lift me up when grad school was the most difficult.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One

Chapter Two

CHAPTER ONE

Resurfacing Proteins to Endow New Properties and Function

Adapted from:
Chapman, A.M.; McNaughton, B.R. *Cell Chemical Biology*, **2016**, *23*, 543.

1.1 Introduction

Proteins possess properties that make them well suited for applications in basic research and clinical settings. Their large size, well-defined and complex structure, and functional group diversity (by virtue of the proteinogenic amino acids) often allows proteins to perform complex processes, foremost among them being catalysis, and molecular recognition of small-molecule, peptide, protein, or nucleic acid binding partners. With the advent of modern molecular biology came the ability to alter the genes that encode proteins. The fusion of molecular biology techniques to control or alter protein sequence, and high-throughput screening or selection methods to identify those few proteins with a desired function, has led to an explosion in the number of non-natural proteins with tailored properties. Alteration of protein sequence in an effort to generate improved or new function can be divided into three general areas: structure guided and focused engineering of a few amino acids within a catalytic pocket or of known function; randomized extensive mutagenesis of the entire protein sequence; or focused extensive mutagenesis of residues that reside on the protein surface. Here, we generally refer to the last approach (which sometimes precedes optimization by randomized mutagenesis) as *protein resurfacing* (**Figure 1.1**). As the reader will find below, protein resurfacing can endow a number of important properties and functions, including potent and selective recognition of macromolecules, increased stability,

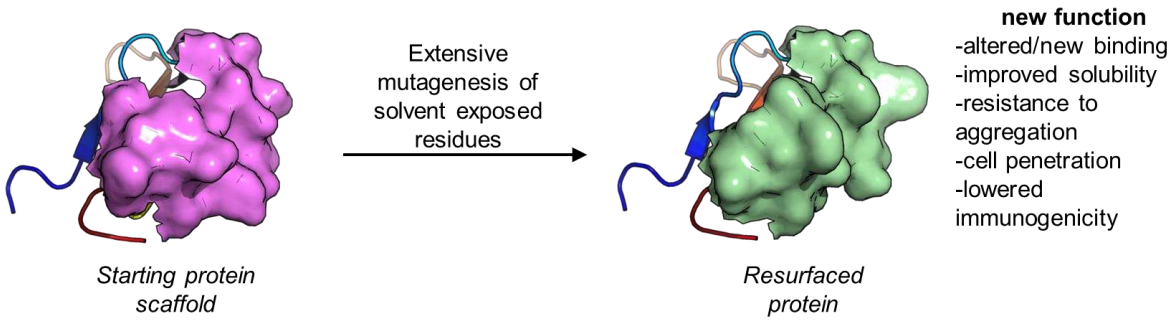


Figure 1.1 Protein resurfacing, extensive mutagenesis of surface residues, can lead to dramatically altered properties and functions. Brazzein PDB code: 4HE7

solubility, and expression in non-natural hosts, resistance to aggregation, penetration of mammalian cells, and reduced immunogenicity.

1.2 Resurfacing Proteins to Endow New Recognition

Recognition is central to biology, and thus central to *controlling* biology. Historically, small-molecules (molecular weight <800 Da) have been used to modulate cell function and fate through potent and selective recognition of a macromolecule – often a protein. However, upwards of 90% of the proteome is deemed 'undruggable' by traditional small-molecules, by virtue of the fact that most proteins do not possess exposed hydrophobic small-molecule binding pockets.[1] Proteins offer a possible solution to recognition and modulation in areas of the proteome that stymie small-molecules. With this in mind, as well as the desire to better understand the molecular dictates of protein-protein interactions (PPIs), researchers have established research programs to identify, optimize, and interrogate these macromolecular assemblies.

When initiating an effort to develop a new PPI, the first question one needs to answer is "*what protein scaffold should I resurface in order to achieve new recognition?"* A probable first response is an antibody. After all, our immune system uses antibodies as a molecular recognition framework to seek out, recognize, and facilitate the destruction of invaders.[2] However, antibodies

and their fragments are not necessarily the best answer. First, anyone who has ever run a Western blot knows that not all antibodies are the same. Some have good target affinity, while others are relatively poor and promiscuous binders. Given the diversity of structure within the proteome, it is unlikely that one solution (sequence optimization of loop residues within the Fab fragment of an antibody) is a one-size-fits-all solution to potent and selective recognition. In addition, full length antibodies and antibody fragments generally express poorly in *E. coli* and are aggregation-prone, leaving researchers searching for alternatives.[3–5]

Minimalized antibodies and antibody mimics that—like full length antibodies—rely on optimization of loop residues to achieve new affinity include fibronectin-based domains (monobodies[6]), camelid-derived nanobodies[7], kunitz domains[8], and lipocalins[9]. An excellent review has been written on the topic of minimalized immunoglobulin and non-immunoglobulin protein scaffolds that are evolvable (amenable to extensive resurfacing to achieve new recognition).[10] In cases where loop optimization (through the use of antibodies and their structural mimics), computational design, or principles such as shape complementarity do not define a starting point, scaffolds highly amenable to mutagenesis, which display structured features, can be utilized. One such scaffold is a small three α-helix bundle protein derived from the B domain of staphylococcal protein A, marketed as an affibody.[11] Resurfaced affibodies have been generated to bind a diverse array of protein surfaces. As shown in **Figure 1.2A**, researchers often resurface two of the three α-helices on an affibody scaffold. In one representative example, researchers used laboratory evolution based resurfacing to identify an affibody that binds a previously untargeted epitope on domain III of human epidermal growth factor receptor 2 (HER2), a protein overexpressed in ~30% of all breast and ovarian cancers—with exceptional affinity (K_D ~20 pM, **Figure 1.2B**).[12,13]

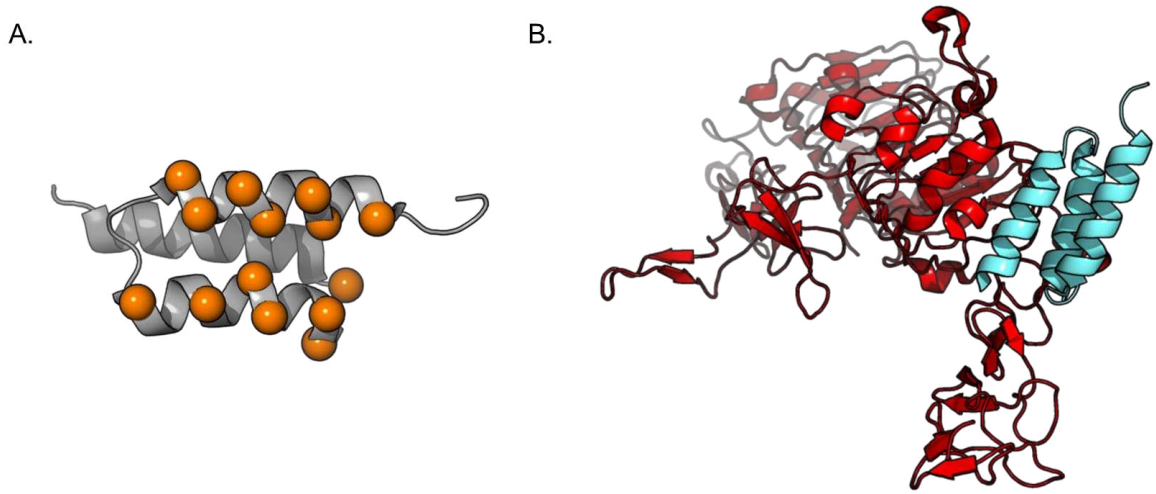


Figure 1.2 (*A*) An affibody scaffold. Residues optimized during resurfacing by in laboratory evolution are highlighted in orange spheres. (*B*) A recently reported affibody (teal) that binds HER2 (red), a breast cancer biomarker. Affibody-HER2 PDB code: 3MZW.

Randomization and optimization of binding loops, within antibodies, antibody fragments, or non-immunoglobulin antibody mimics, is a reasonable starting point when little is known about the structure of the biopolymer being targeted. However, in other cases, structural features of the target, or efforts to optimize an existing low affinity complex, can guide or simplify selection of the starting protein scaffold. In these cases, researchers are given freedom to move beyond the (potential) confines of the antibody or non-immunoglobulin antibody mimic paradigm. While the literature is full of excellent case studies, we have selected representative examples that showcase resurfacing of 'privileged' protein scaffolds, using existing structural features, shape-complementarity, or optimization of a low affinity complex between the scaffold and target, as a design principle and starting point rationale. While the majority of cases, in the literature and our selected examples, use high-throughput screening or laboratory evolution to endow new recognition through protein resurfacing, computational approaches play an increasingly significant role in this field, and representative examples are included in this chapter.

There is wide interest in the molecular dictates of helical assembly, controlling such assemblies, and their role in various biological processes—including disease-relevant processes. Thus, methods for displaying and optimizing the surface of a helix, through resurfacing, is a focus of current protein science and chemical biology. The Kim[14] and Schepartz[15] labs have previously reported *chemically synthesized* miniature proteins or fragments with binding residues displayed on the surface of a helix. Proper positioning of residues on an existing helical structure, or optimization of the surface by high-throughput screening, can endow new recognition of these synthetic structures. More recently, our lab has used protein resurfacing (of a helix) to develop a protein that inhibits a 'trimer-of-dimers' helical assembly formed by gp41, which is required for HIV-1 infection of mammalian cells (**Figure 1.3A and B)**. In particular, we used 'helix-grafted display', wherein many solvent exposed residues on a helix within the Pleckstrin Homology domain (PH domain) are mutated to mimic the HIV-1 gp41 C-peptide **(Figure 1.3C)**. These helix resurfaced proteins selectively bind to a mimic of prefusogenic HIV-1 gp41.[16]

It should go without saying that the molecular requirements for generating a potent and selective PPI is incredibly complex, and difficult to determine *de novo*. However, if you know something about the structure of the target you are going after, it makes sense that shape-complementarity be a principal component of strategy to develop a binder. After all, developing new PPIs is fundamentally about generating a surface or binding pocket that complements the target. Many labs, including ours, have explored the virtues of shape-complementarity as a design principle for developing new PPIs, and representative examples are presented below, as well as discussed in detail in chapters three and five.

Kuhlman and co-workers used shape-complementarity and computational design to generate a small, helical bundle protein that binds a hydrophobic region on p21-activated kinase 1

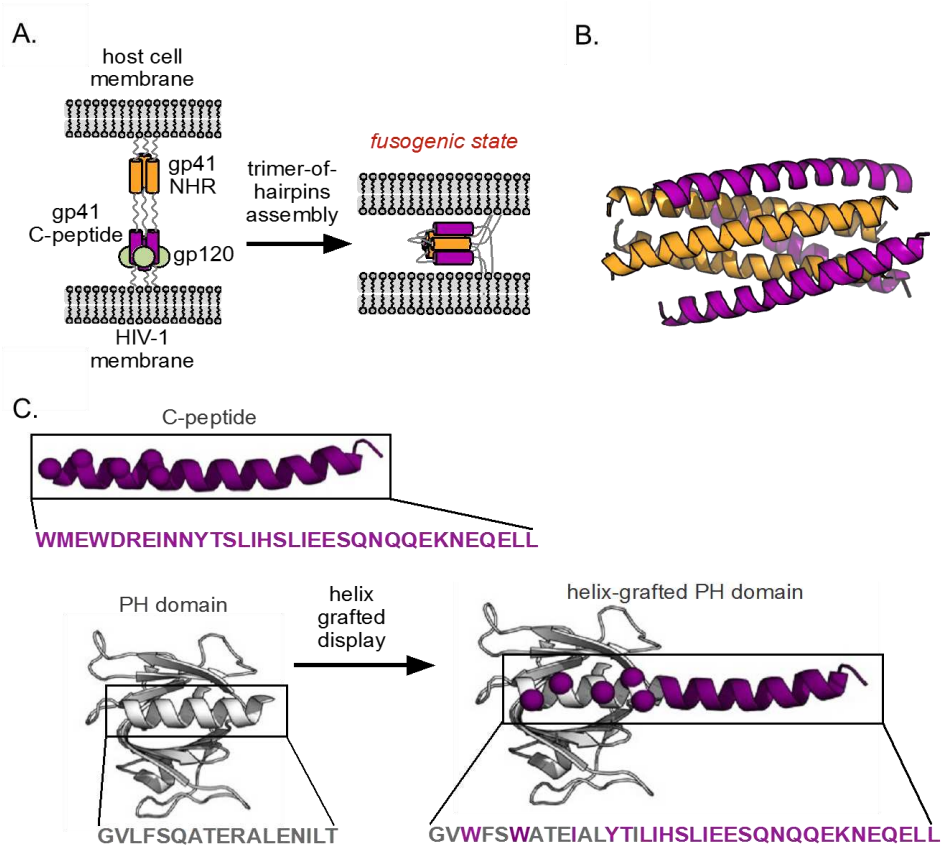


Figure 1.3 Protein resurfacing via helix-grafted display. (*A*) HIV-1 gp41-mediated membrane fusion of HIV-1 with a host cell. Binding of a C-peptide (purple) to an N-terminal coiled coil (NHR, orange) drags viral and cell membranes into close proximity, promoting fusion. (*B*) Crystal structure of the C-peptide (purple)/NHR (orange) helical assembly. (*C*) Helix-grafted display strategy for generating stable C-peptide mimics on a Pleckstrin Homology (PH) domain. The native ligand helix (purple) is overlaid on that of the scaffold protein (grey), and solvent-exposed scaffold residues are mutated to those of the ligand to be displayed (purple spheres). gp41 helical assembly PDB code: 1AIK, PH domain PDB code: 2CAY

(PAK1, **Figure 1.4,** hydrophobic region is highlighted in blue).[17] The human hyperplastic discs protein (HYP) scaffold they chose possessed the general size and shape to fit the desired cleft on PAK1. DDMI interface software was used to dock HYP onto PAK1, followed by protein resurfacing (highlighted in red), guided by Rosetta[18], to computationally optimize the complex. One protein (termed 'Spider Roll') bound the targeted hydrophobic region on PAK1 with modest affinity (K_D ~100 μM **Figure 1.4**), representing one of the first examples of a *de novo* designed PPI based on shape complementarity.

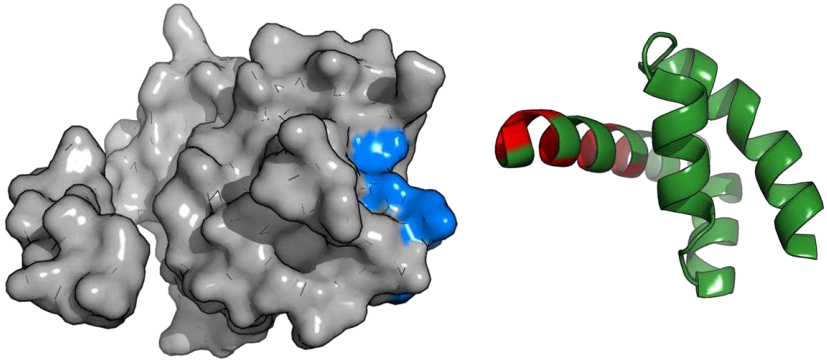


Figure 1.4 The design principal of shape-complementarity, and a computational tool (Rosetta), were used to resurface a helix of human hyperplastic discs protein (HYP, resurfaced region is highlighted red) that binds p21-activated kinase 1 (PAK1, targeted hydrophobic region is highlighted in blue). PAK1 PDB code: 1F3M, HYP PDB code: 1I2T.

In a conceptually similar effort, Baker, Karanicolas, and coworkers recently co-engineered a PPI by combining shape-complemented proteins, computational design, and directed evolution.[19] After choosing an *in silico* designed ankyrin repeat protein (this structural class is described in detail in chapter three), which they call Pdar, they used PatchDock[20] to screen 37 protein scaffolds for shape-complementarity to the largely hydrophobic binding face of the target. These researchers then used *in silico* protein resurfacing (Rosetta[18,21]) of the putative binding face. The best resurfaced protein, called Prb, binds Pdar with excellent affinity (K_D ~140 nM). Subsequent resurfacing (by laboratory evolution) on the anticipated ankyrin repeat binding surface of Prb, and the concave face of Pdar, resulted in a very high affinity complex (K_D ~180 pM).

The above examples largely rely on targeting a hydrophobic surface on a protein, which is less complex than achieving recognition of a surface that is more polar, and as such likely requires the generation of well-defined hydrogen binding, ion-pair, and salt-bridge interactions. Toward the goal of achieving a new PPI that does not principally rely on hydrophobic effects to drive assembly, Baker and coworkers dovetailed computational docking (PatchDock[20] and Rosetta Dock[18,21]) of shape-complementary scaffolds and directed evolution-based protein resurfacing to

identify a binding partner that recognizes a complex and highly charged surface on hen egg lysozyme (HEL). The resulting protein tightly binds HEL (K_D ~3 nM), and inhibits lysosomal activity.[22]

A majority of this thesis describes the use of shape complementarity as a protein design principle to target large, hydrophilic surfaces of a protein class (ankyrin repeats) that is extremely challenging from a molecular recognition standpoint. This work is discussed in chapters three and five.

1.3 Protein Resurfacing to Alter Target Specificity and Reduce Binding Promiscuity

Towards the goal of better understanding PPIs, protein resurfacing methods have been used extensively to alter (as opposed to merely broadening) target selectivity and reduce binding promiscuity. A few seminal examples are discussed below.

Prolactin is a human growth hormone (hGH) homolog that has very low affinity for the hGH receptor. In an early and seminal contribution to the field of protein resurfacing to endow new recognition, scientists at Genentech integrated (and later optimized) 9 residues from wild-type hGH—that are critical to hGH receptor recognition—into the prolactin scaffold. As a result of resurfacing the hGH binding face, this team of researchers (led by Jim Wells) generated an hGH binding prolactin mutant that bound the hGH receptor ~10,000-fold better than wild-type prolactin (K_D ~ 2 nM) , and only ~6-fold lower than hGH.[23]

In other efforts, protein resurfacing has been used to decrease target promiscuity and direct recognition to a single target. For example, in a 2003 study, Shifman and Mayo used a structure-based computational rotamer optimization method (ORBIT) to narrow down and optimize peptide recognition for calmodulin (CaM) – a messenger protein that promiscuously binds a multitude of

peptide sequences. In particular, these researchers focused on surface and boundary CaM residues to design a resurfaced CaM that potently and selectively binds smooth muscle myosin light chain kinase (smMLCK). Starting from a previously described variant[24], iterative rounds of design and analysis resulted in improved selectivity. While both the wild-type and resurfaced CaM proteins bind smMLCK peptide with excellent affinity (K_D ~ 2 nM for each), the evolved protein binds smMLCK with ~7-fold selectivity over other CaM targets, by virtue of the dramatically resurfaced binding face that was optimized for this specific ligand (**Figure 1.5)**.[25]

More recently, Reynolds and co-workers employed computational methods (EGAD[26]) to design a β-lactamase inhibitor protein (BLIP) with increased affinity and specificity for SHV β-lactamase. BLIP can bind to and inhibit several β-lactamases, but has the lowest affinity for SHV-1 (K_D ~1.7 μM). Comparing the crystal structures of BLIP/TEM-1 (K_D ~1.3 nM) and BLIP/SHV-1, the researchers focused on cluster 2 (C2), and found three mutations that strongly favor SHV-1 binding (E73M, S130K, S146M). The BLIP variant increased affinity for SHV-1 almost 400-fold (K_D ~4.6 nM), while reducing affinity for TEM-1 greater than 20-fold (K_D ~27 nM).[27]

1.4 Protein Resurfacing to Improve Protein Expression, Stability, Solubility, and Generate Resistance to Aggregation

Perhaps the most widely used protein over the last few decades is Green Fluorescent Protein (GFP), and is an excellent example of how protein resurfacing can optimize expression, stability and solubility. GFP is a 238 amino acid protein that forms a β-barrel structure (**Figure 1.6**).[28] Almost everyone is familiar with work (principally done by Tsien and coworkers) to alter spectroscopic properties of the GFP chromophore – resulting in the fruit-based color palette of fluorescent proteins we currently enjoy.[29] However, before this Nobel prize winning achievement,

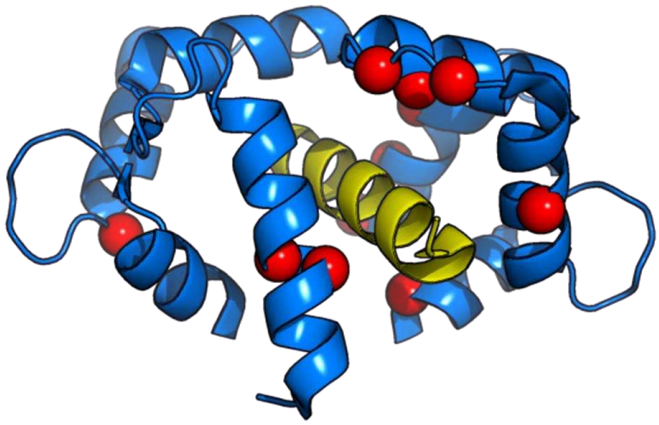


Figure 1.5 Structure-guided analysis and computational tools were used to resurface much of the binding pocket of calmodulin (blue), resulting in selective recognition of smooth muscle myosin light chain kinase (yellow). The position of mutations generated from iterative rounds of design and analysis are highlighted in red. Calmodulin-smMLCK PDB code: 1CDM.

research on GFP was hindered by challenges with its expression and aggregation, largely due to homodimerization and instability at high concentrations.[28] Coupling mutagenesis and fluorescence-activated cell sorting (FACS), researchers at Affymax discovered two solvent exposed mutations, F99S and M153T, which resulted in a GFP variant with significantly improved solubility (referred to as 'cycle 3 mutations, **Figure 1.6, blue**).[30] In 2006, Waldo and coworkers used mutagenesis and flow cytometry screening to identify a variant with optimized folding (which they linked to fluorescence intensity). Interestingly, this mutant, which they call 'superfolder GFP' (sfGFP) contains five solvent exposed mutations (S30R, Y39N, N105T, I171V, A206V, **Figure 1.6, red**), and this protein resurfacing led to a GFP mutant that only exists in monomeric form. Further underlining the importance of protein resurfacing, the robust folding kinetics of superfolder GFP is thought to derive from both an increase in surface stabilizing interactions and inhibition of dimer formation.[31]

In an intramolecular context, the delicate balancing act called protein folding is principally driven by the collapse of hydrophobic residues to form a hydrophobic core.[32] However, in an intermolecular context, the same basic process of hydrophobic driven assembly can result in

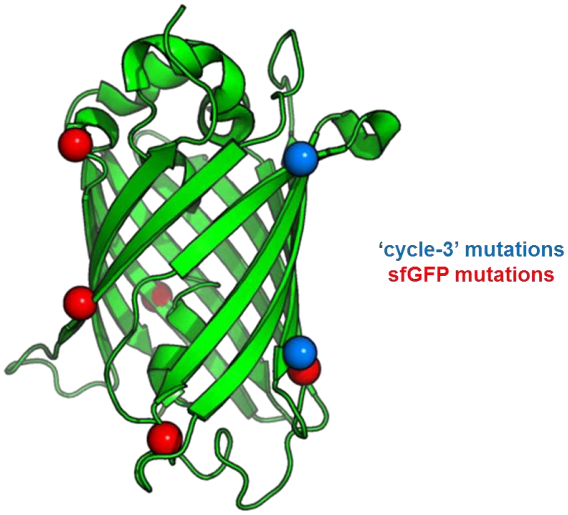


Figure 1.6 Protein resurfacing has led to new Green Fluorescent Proteins with dramatically improved stability and expression as a monomer in *E. coli.* Superfolder GFP PDB code: 2B3P

aggregation – a major issue for protein therapeutics, and a culprit in some human diseases, including Alzheimers, amyloidosis, and prion diseases.[33] Researchers have developed methods—founded in protein resurfacing—that seek to generate proteins resistant to aggregation. A detailed discussion of our lab's contribution in this field is discussed in chapter two.

Liu and coworkers recognized a well-established relationship between the solubility of natural proteins and their overall net charge. Using a process they call 'protein supercharging'—extensive mutagenesis of solvent exposed residues to either positively charged (lysine or arginine) or negatively charged (aspartic or glutamic acid)—they hypothesized that resistance to aggregation could be created. Starting from superfolder GFP (-7 theoretical charge, termed starting GFP or stGFP), they selected residues for mutation using a computational method (average neighboring atoms per side chain atom, AvNAPSA), resulting in GFP mutants with a theoretical net charge that ranged +15 to +48, and –25 to –30 (**Figure 1.7A**). Amazingly, supercharged GFP mutants with a theoretical net charge of +36 or -30 resist aggregation and regain fluorescence after heat denaturation at 100 °C followed by slow cooling. In contrast, the stGFP scaffold aggregates under

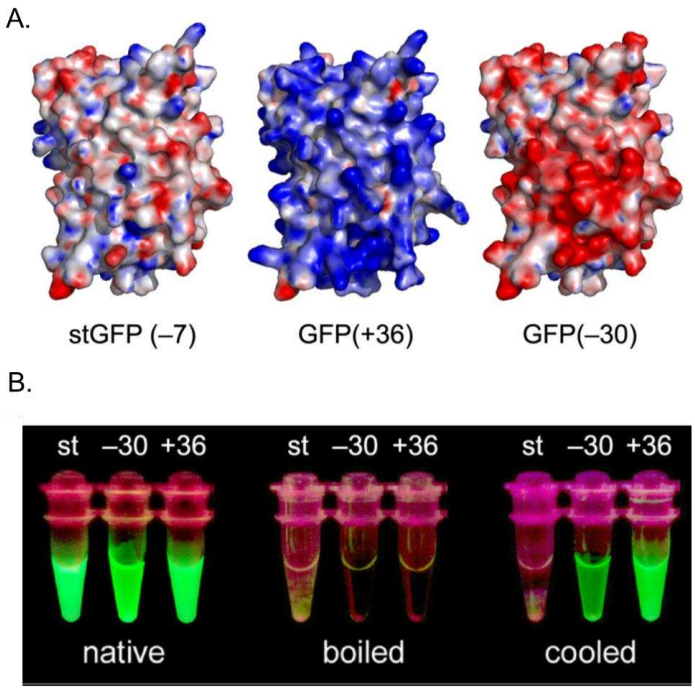


Figure 1.7 Protein 'supercharging', extensive mutagenesis of solvent exposed residues to positively charged (arginine or lysine) or negatively charged (aspartic acid or glutamic acid) endows resistance to aggregation. (*A*) Electrostatic surface potentials of the starting GFP (stGFP), GFP(+36), and GFP(–30). Blue = cationic; red = anionic. (*B*) UV-illuminated samples of purified GFP variants ("native"), those samples heated 1 min at 100 °C ("boiled"), and those samples subsequently cooled for 2 h at 25 °C ("cooled").

identical conditions (**Figure 1.7B**). Two additional proteins, streptavidin and glutathione-S-transferase (GST), were also shown to tolerate supercharging and resist aggregation to some degree.[34]

Recently, Ellington and coworkers applied supercharging to aggregation-prone short chain antibody variable fragments (scFv).[35] Starting from an anti-MS2 scFv (+7.5 theoretical net charge), the researchers computationally designed supercharged variants ranging from -19.5-+33.5 theoretical net charge. One supercharged mutant, K-pos-1 (+15.5 theoretical net charge), generated improved resistance to heat denaturation and aggregation, retaining ~70% binding activity after being heated at 70 °C, while the wild-type completely lost binding activity. Additionally K-pos-1

was less prone to aggregation than the wild-type protein when stored at 4 °C in phosphate-buffered saline (PBS), as determined by dynamic light scattering (DLS). Despite extensive mutagenesis, as a result of supercharging, K-pos-1 retained affinity for MS2, and actually bound this protein ~36 times tighter than the wild-type protein (K_D ~1 nM and 36 nM, respectively), perhaps due to stabilization of the protein complex.

Supercharging has been used to solubilize membrane-bound proteins, which are notoriously difficult to express in a soluble form, due to solvent-exposed hydrophobic patches that normally take refuge in the lipid bilayer membrane. Researchers in the Weiss lab used a combinatorial supercharging and phage display to optimize residues in the hydrophobic patch. This resulted in a supercharged and soluble mutant of the membrane protein caveolin, which expresses in large quantities in *E. coli* and retains affinity for three binding partners of the wild-type protein.[36]

1.5 Protein Resurfacing to Endow Mammalian Cell Penetration

One major challenge to the broader use of proteins in biomedical applications is their general inability to efficiently cross the lipid bilayer of mammalian cells and access the cytosol or nucleus. Thus, most current protein drugs and basic research tools target infectious reagents, secreted receptors, and disease-relevant receptors that reside on the surface of the cell or the extracellular matrix. Efforts to unlock the full potential of proteins in biomedical applications by enabling potent and functional cell penetration have been a major focus of modern biologics research. Incorporation of polycationic linkages—such as polyarginine and the trans-activating transcriptional activator peptide (TAT) from HIV-1—have previously been described as a means to enable cell penetration of various cargo, including proteins.[37–39] More recently, researchers have

used protein resurfacing to generate polycationic features on the protein surface. For example, Raines and Fuchs reported that 'arginine grafting' of GFP—mutagenesis of clustered solvent exposed amino acids to arginine—enables cellular uptake (**Figure 1.8A**).[40] While the arginine grafted GFP is internalized, it appears as punctate foci in fluorescence microscopy images – suggesting encapsulation of the internalized protein within endosomes. While such encapsulation might limit the utility of most internalized proteins, high turnover enzymes might be an outlier, since only a small amount of the protein needs to access the cytosol to alter cell function and fate. In the Raines lab, arginine grafting has been used to endow uptake of proteins with possible therapeutic utility. In particular, effort was focused on generating resurfaced 'arginine grafted' variants of ribonuclease A (RNase A), which, if delivered to the cytoplasm is cytotoxic to mammalian cells. Installation of two arginine mutations (E49R, D53R), distant from the catalytic site on RNase A, resulted in cell penetration and a 3-fold increase in cytotoxicity in K-562 leukemia cells.[41]

Onconase, an amphibian ribonuclease with cytotoxicity for tumor cells, enters cells due to a lysine-decorated surface. Previous research showed that arginine is more effective as a transporter across the cell membrane, compared to lysine.[37,39] In an effort to improve cell uptake potency, Raines and coworkers created an arginine grafted variant, wherein all ten solvent-exposed lysines are mutated to arginine. Arginine-grafted Onconase (R-Onc) was internalized ~3-fold more efficiently compared to the wild-type protein; however, no appreciable increase in cytotoxicity was observed, potentially due to decreased stabilization and increased proteolytic degradation.[42]

In addition to endowing resistance to aggregation, protein supercharging (in the positive direction) has been used to endow mammalian cell penetration. Researchers in the Liu lab found that resurfaced 'supercharged' GFPs with a theoretical net charge of +15, +25, and +36 potently

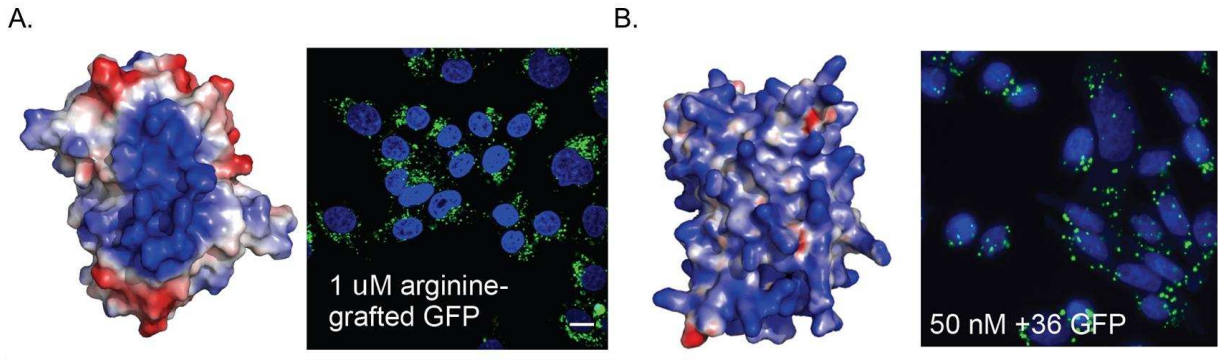


Figure 1.8 (*A*) Raines' arginine grafted GFP. Blue = cationic; red = anionic. Arginine grafted GFP penetrates mammalian cells, and largely appears as punctate foci, suggesting encapsulation within endosomes.(*B*) Liu's supercharged +36 GFP. Blue = cationic; red = anionic. +36 GFP potently penetrates mammalian cells, and largely appears as punctate foci, suggesting encapsulation within endosomes. In microscopy images, nuclei are stained blue with DAPI.

penetrate a variety of mammalian cell lines, in a charge-dependent manner. Similar to arginine grafted GFP, internalized protein appears as punctate foci in fluorescence miscroscopy images – suggesting that most of the internalized protein is encapsulated in endosomes (**Figure 1.8B**).[43] Nonetheless, not only does +36 GFP enter mammalian cells, the polycationic protein forms a complex with siRNA or plasmid DNA, and this complex resulted in the delivery of bound nucleic acids to mammalian cells.[44] Liu and coworkers showed that +36 GFP can drag a fused protein into mammalian cells. Of particular note, when fused to +36 GFP, appreciable levels of Cre-recombinase is delivered to various mammalian cells or mouse retinal tissue, resulting in Cre-dependent expression of DsRed2 or β-galactosidase, respectively.[43] More recently, Liu and coworkers have shown that resurfaced polyanionic GFP (-30 theoretical net charge), and fusions thereof, can be delivered to the interior of cells when pre-mixed with commercially available cationic lipids.[45] Using up to a 1,000-fold lower dosage than +36 GFP fusions, these researchers were able to delivery various genome editing proteins into mammalian cells, including transcription activator-like nucleases (TALENs). Intracellular delivery of TALENs specific for the

neurotrophin-3 gene (NTF3) resulted in a ~7-fold increase in protein expression in just 4 hours. Additionally, -30 GFP-Cre fusions could be delivered to mouse inner ear cells, resulting in Cre-dependent recombination of tdTomato in ~90% of outer hair cells.

While these polycationic resurfacing methods endow potent cell penetration, a major challenge to their broader application is the lack of established and broadly applicable guidelines for this extensive mutagenesis. While computational tools can guide protein supercharging, relatively little is known about how to dramatically resurface a protein with a polycationic feature in a manner that does not dramatically alter or abolish its utility and/or function (stability, target affinity, expression in *E. coli*). In our experience, even structurally similar proteins respond differently to such extensive mutagenesis, and many proteins of therapeutic interest were not amenable to polycationic resurfacing. Perhaps a simpler approach is to focus effort on developing a single resurfaced polycationic, cell-penetrating, protein scaffold that is stable, expresses in *E. coli*, maintains the function of the original protein, but can be evolved to bind virtually any disease-relevant intracellular target.

In the context of this challenge, our lab very recently reported polycationic resurfaced cell-penetrating nanobodies.[46] The researcher team used three previously reported nanobodies that binds GFP[47], HER2[48] or β-lactamase[49] (referred to as NB1, NB2, and NB3, respectively). Solvent-exposed residues within the framework region—that are distinct from the complementary determining regions (CDR loops), where binding occurs, were mutated to either lysine or arginine (**Figure 1.9A, blue spheres)**. When fused to GFP, all three polycationic resurfaced nanobodies (pcNB1-3) drag GFP into the cell, in a concentration-dependent manner. Interestingly, in contrast to arginine grafted or supercharged GFP, which appear as punctate foci, resurfaced nanobodies do not appear as such (**Figure 1.9B-D**), suggesting that appreciable amounts of these internalized

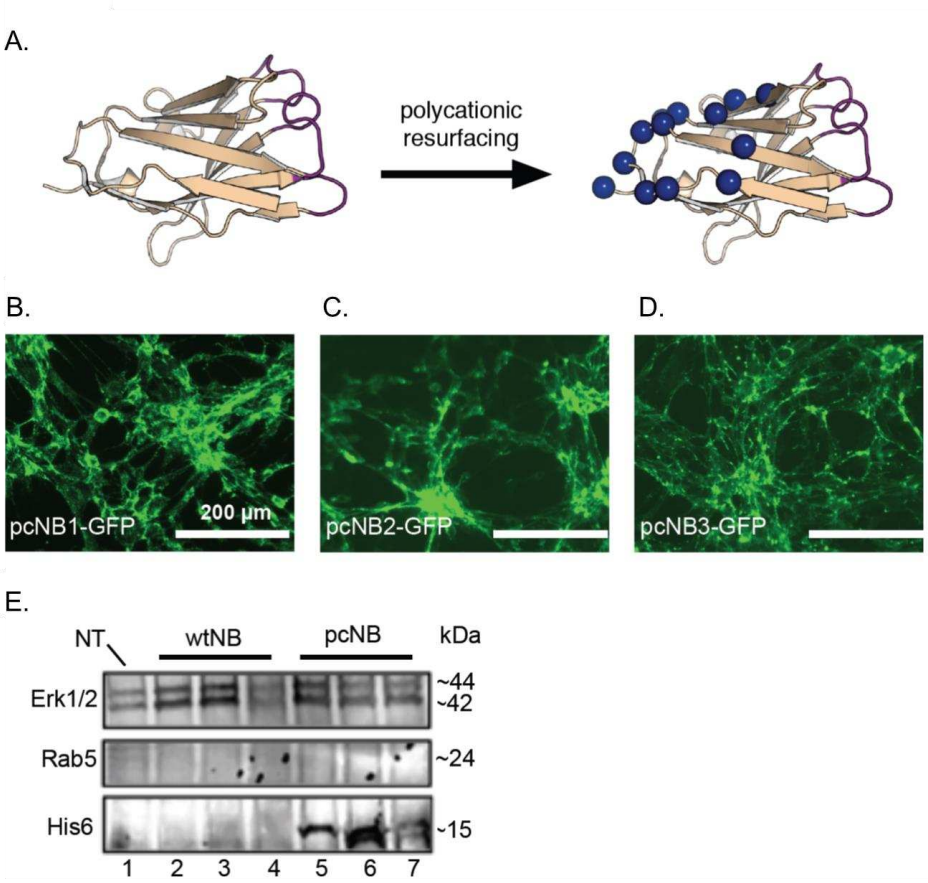


Figure 1.9 (*A*) Protein resurfacing of a previously reported nanobody with arginine or lysine (highlighted with blue spheres) to generate a polycationic resurfaced cell-penetrating nanobody. (*B-D*) Fluorescence microscopy images (endosome marker) for 3T3 cells following treatment with 250 nM resurfaced nanobody-GFP fusions. (*E*) Western blot analysis of digitonin cell lysate for Erk1/2 (cytosolic marker), His_{6x} (internalized resurfaced nanobody), or Rab5 (endosome marker). Lane 1= no treatment; lane 2= wild-type NB1; lane 3 = NB2; lane 4 = NB3; lanes 5-7 = polycationic resurfaced NB1, NB2, or NB3, respectively. Nanobody PDB code: 3OGO

nanobodies might access the cytosol. This important aspect of cell uptake was further analyzed using a previously described lysis assay and Western blot.[50] No appreciable amount of wild-type His_{6x} labelled nanobody is found within the cytosolic extraction (**Figure 1.9E**, lanes 2-4). In contrast, internalized resurfaced His_{6x}-nanobodies appear in the cytosol (**Figure 1.9E**, lanes 5-7). Thus, the polycationic resurfaced protein is capable of dragging another protein (GFP) into the cytosol of a mammalian cell.[46]

1.6 Genetic or Chemical Protein Resurfacing to Modulate Immunogenicity

Non-human proteins are often taken up by antigen-presenting cells (APCs), which in turn process and display small peptidic regions of that protein (9-15 amino acids).[51] Some of these segments (referred to as T-cell epitopes) are recognized by major histocompatibility complex class II (MHC II) as 'non-self', and can lead to the production of neutralizing antibodies.[52] These antibodies are problematic, since they can diminish the concentration or existence of the biological therapeutic, or even worse, initiate a full on immune response that leads to potentially dangerous side effects.[53] Thus, resurfacing as a means of removing T-cell epitopes to reduce immunogenicity is a major effort in the development and implementation of protein drugs.

Removing B-cell epitopes (antibody-binding epitopes), although a potential companion strategy with T-cell epitope removal[54–56], remains challenging, since it is requires knowing how anti-drug antibodies bind its target. In contrast, resurfacing to remove T-cell epitopes is independent of antibody recognition and is an established method for reducing protein immunogenicity. Critically, computational tools that scan for MHC II-binding T-cell epitopes in a protein sequence are available, and successful examples of T-cell epitope removal by dovetailing computational tools and extensive mutagenesis at the protein surface are provided below.

Bacterial proteins are almost always highly immunogenic (elicit an immune response). Staphylokinase, a protein secreted from strains of *Staphylococcus aureus*, has potent thrombolytic activity in humans, and thus is of interest as a method to dissolve blot clots, but elicits an immune response. Warmerdam and co-workers identified six highly immunogenic regions within the protein, and employed *in silico* alanine-scanning of overlapping epitopes regions to focus mutation. Impressively, the immunogenicity of a variant with only four mutations reduced donors staphylokinase-specific T-lymphocytes to undetectable levels, compared to a positive response

with the wild-type proteins. This was the first example highlighting the ability to deimmunize a bacterial protein using T-cell epitope elimination.[57,58]

Similarly, researchers in the Fiering, Bailey-Kellogg, and Griswold labs applied a structure-based deimmunization algorithm (EpiSweep[59]) to resurface Lysostaphin, a metalloprotease that binds and cleaves the cell wall of *Staphylococcus aureus* (**Figure 1.10**).[59–61] Using computation as a guide, these researchers identified a lysostaphin variant with 13 mutations that reduced anti-lysostaphin IgG titers after injection in transgenic DR4 mice up to ~1000-fold after two weeks, compared to treatment with the wild-type protein. Additionally, this mutant exhibited increased efficacy in the same mice challenged with a strain of methicillin-resistant *Staphylococcus aureus* (MRSA), highlighting the correlation between decreased immunogenicity and increased efficacy.

Finally, T-cell epitope removal has been used to deimmunize potent toxins, which are regularly combined with antibody fragments for targeted delivery in cancers (often referred to as immunotoxins). Perhaps most notably, the Pastan lab, in collaboration with several other labs, combined Rosetta computational design and guided alanine mutagenesis to deimmunize a 38 kDa toxin derived from *Pseudomonas* exotoxin A. Using high-resolution mapping, predicted epitope content was lowered by 93% by removing eight T-cell epitopes with just six mutations, resulting in a ~90% decrease in T-cell response, while retaining cytotoxicity.[62–64]

The above examples rely on altering the sequence of solvent exposed amino acids in order to suppress immunogenicity. An alternative strategy is to use chemical methods to modify the solvent exposed protein surface, to endow or improve function. In the context of suppressing immunogenicity, covalent attachment of polyethylene glycol to the protein surface (referred to as 'PEGylation', **Figure 1.11**) has been used extensively.[65–68] First discovered in the 1970's,

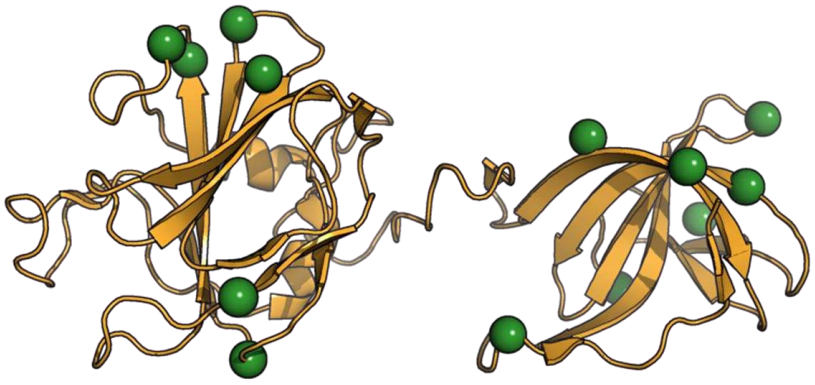


Figure 1.10 A computational tool (EpiSweep) was used to identify T-cell epitopes in lysostaphin (dark yellow), a metalloprotease that binds and cleaves the cell wall of *Staphylococcus aureus.* Computational prediction was then used to alter many amino acids in these regions (highlighted in green), without altering protein folding and/or stability. The resurfaced protein was significantly less immunogenic compared to wild-type lysostaphin. Lysostaphin PDB code: 4LXC.

PEGylation can be used to render proteins 'invisible' to the immune system, presumably by blocking recognition between immune system components and immunogenic solvent exposed protein surfaces. PEGylation can also dramatically reduce protease degradation and increase serum lifetime. Demonstrating the value of this chemical resurfacing method, a large number of FDA-approved protein drugs are sold as PEGylated variants, including asparaginase (Oncaspar®), granulocyte colony-stimulating factor (Neulasta®), and interferon variants (Pegasys® as one example).[69] More recently, researchers aimed to expand the value of covalent protein resurfacing as a means to endow new properties, including improved suppression of immunogenicity. One promising strategy is N-linked glycan conjugation, which is based on a natural defense mechanism used by certain viruses to evade the immune system.[70] As an example, researchers in the Sanders lab recently 'immunosilenced' immunogenic trimerization domains by incorporation of four N-linked glycans.[71]

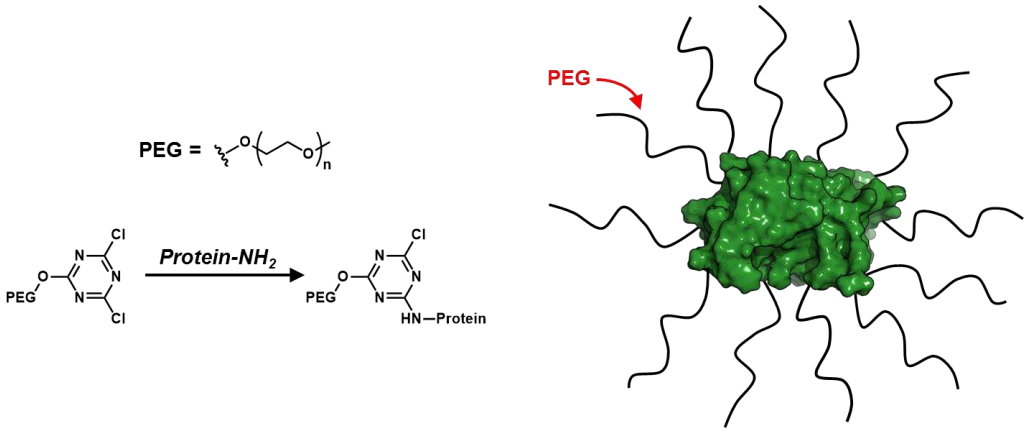


Figure 1.11 Proteins can be chemically resurfaced with polyethylene glycol (PEG), which endows a number of properties, including improved solubility, improved *in vitro* and/or *in vivo* stability, and decreased immunogenicity. Many chemical methods for protein PEGylation have been reported in the literature, and one example is shown above. Interferon alpha 2a PDB code: 1ITF.

1.7 Conclusion

Proteins have properties (large size and associated ability to bind a large surface area, well-defined structure, functional group diversity) that make them well suited as basic research tools and drugs, particularly in spaces that challenge small-molecule centered discovery. Moreover, modern molecular biology, and high-throughput screening and laboratory evolution technologies enable researchers to screen billions of proteins for a desired function – something that is virtually impossible in the small-molecule world. The truth is that using biologics in clinical applications is on the rise (half of the current top 20 selling drugs are biologics), and most of the proteome is not amenable to potent and selective modulation by small-molecules. However, using biologics in basic research and clinical settings introduces a host of challenges. Protein resurfacing has, and will, continue to play a pivotal role in generating novel proteins with a desired function, such as potent and selective recognition and modulation of a disease-relevant receptor, and can be used to improve properties relevant to biomedical applications. In this chapter, we showed the reader that protein resurfacing—extensive amino acid mutagenesis or covalent modification—can be done to

improve selective macromolecular recognition, expression, solubility and stability, endow resistance to aggregation, enable penetration of mammalian cells, and modulate immunogenicity. Moreover, we have shown that protein resurfacing methods have the potential to be combined to achieve a wide range of favorable characteristics. Clearly, we have only begun to 'scratch the surface' of protein resurfacing. As more and more biologics (overwhelmingly proteins) enter the market as FDA approved drugs, protein resurfacing will clearly play a prominent role in generating new drugs with novel recognition and properties, required for use in clinical settings.

REFERENCES

(1) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. *Nat. Rev. Drug Discov.* **2006**, *5* (12), 993.

(2) Sliwkowski, M. X.; Mellman, I. *Science* **2013**, *341* (6151), 1192.

(3) Ducry, L. *Methods Mol. Biol.* **2012**, *899*, 489.

(4) Samaranayake, H.; Wirth, T.; Schenkwein, D.; Raty, J. K.; Yla-Herttuala, S. *Ann. Med.* **2009**, *41* (5), 322.

(5) Lee, Y. J.; Jeong, K. J. *J. Biosci. Bioeng.* **2015**, *120* (5), 483.

(6) Koide, A.; Koide, S. *Methods Mol. Biol.* **2007**, *352*, 95.

(7) Muyldermans, S. *Annu. Rev. Biochem.* **2013**, *82*, 775.

(8) Dennis, M. S.; Herzka, A.; Lazarus, R. A. *J. Biol. Chem.* **1995**, *270* (43), 25411.

(9) Skerra, A. *J. Mol. Recognit.* **2000**, *13* (4), 167.

(10) Binz, H. K.; Amstutz, P.; Plückthun, A. *Nat. Biotechnol.* **2005**, *23* (10), 1257.

(11) Lofblom, J.; Feldwisch, J.; Tolmachev, V.; Carlsson, J.; Stahl, S.; Frejd, F. Y. *FEBS Lett.* **2010**, *584* (12), 2670.

(12) Orlova, A.; Magnusson, M.; Eriksson, T. L. J.; Nilsson, M.; Larsson, B.; Höiden-Guthenberg, I.; Widström, C.; Carlsson, J.; Tolmachev, V.; Ståhl, S.; Nilsson, F. Y. *Cancer Res.* **2006**, *66* (8), 4339.

(13) Eigenbrot, C.; Ultsch, M.; Dubnovitsky, A.; Abrahmsén, L.; Härd, T. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (34), 15039.

(14) Eckert, D. M.; Kim, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (20), 11187.

(15) Kritzer, J. A.; Zutshi, R.; Cheah, M.; Ran, F. A.; Webman, R.; Wongjirad, T. M.;

Schepartz, A. *ChemBioChem* **2006**, *7* (1), 29.

(16) Walker, S. N.; Tennyson, R. L.; Chapman, A. M.; Kennan, A. J.; McNaughton, B. R. *ChemBioChem* **2015**, *16* (2), 219.

(17) Jha, R. K.; Leaver-Fay, A.; Yin, S.; Wu, Y.; Butterfoss, G. L.; Szyperski, T.; Dokholyan, N. V.; Kuhlman, B. *J. Mol. Biol.* **2010**, *400* (2), 257.

(18) Schueler-Furman, O.; Wang, C.; Baker, D. *Proteins Struct. Funct. Genet.* **2005**, *60* (2), 187.

(19) Karanicolas, J.; Corn, J. E.; Chen, I.; Joachimiak, L. a; Dym, O.; Peck, S. H.; Albeck, S.; Unger, T.; Hu, W.; Liu, G.; Delbecq, S.; Montelione, G. T.; Spiegel, C. P.; Liu, D. R.; Baker, D. *Mol. Cell* **2011**, *42* (2), 250.

(20) Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. *Nucleic Acids Res.* **2005**, *33* (Web Server issue), W363.

(21) Gray, J. J.; Moughon, S. E.; Kortemme, T.; Schueler-Furman, O.; Misura, K.; Morozov, A. V.; Baker, D. *Proteins Struct. Funct. Genet.* **2003**, *52* (1), 118.

(22) Procko, E.; Hedman, R.; Hamilton, K.; Seetharaman, J.; Fleishman, S. J.; Su, M.; Aramini, J.; Kornhaber, G.; Hunt, J. F.; Tong, L.; Montelione, G. T.; Baker, D. *J. Mol. Biol.* **2013**, *425* (18), 3563.

(23) Cunningham, B.; Henner, D.; Wells, J. *Science (80-.).* **1990**, *247* (4949), 1461.

(24) Shifman, J. M.; Mayo, S. L. *J. Mol. Biol.* **2002**, *323* (3), 417.

(25) Shifman, J. M.; Mayo, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13274.

(26) Chowdry, A. B.; Reynolds, K. A.; Hanes, M. S.; Voorhies, M.; Pokala, N.; Handel, T. M. *J. Comput. Chem.* **2007**, *28* (14), 2378.

(27) Reynolds, K. A.; Hanes, M. S.; Thomson, J. M.; Antczak, A. J.; Berger, J. M.; Bonomo,

R. A.; Kirsch, J. F.; Handel, T. M. *J. Mol. Biol.* **2008**, *382* (5), 1265.

(28) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67* (11), 509.

(29) Shaner, N. C.; Steinbach, P. A.; Tsien, R. Y. *Nat. Methods* **2005**, *2* (12), 905.

(30) Crameri, a; Whitehorn, E. a; Tate, E.; Stemmer, W. P. *Nat. Biotechnol.* **1996**, *14* (3), 315.

(31) Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nat. Biotechnol.* **2006**, *24* (1), 79.

(32) Dobson, C. M. *Nature* **2003**, *426* (5), 884.

(33) Dobson, C. M. *Nat. Rev. Drug Discov.* **2003**, *2* (2), 154.

(34) Lawrence, M. S.; Phillips, K. J.; Liu, D. R. *J. Am. Chem. Soc.* **2007**, *129* (33), 10110.

(35) Miklos, A. E.; Kluwe, C.; Der, B. S.; Pai, S.; Sircar, A.; Hughes, R. A.; Berrondo, M.; Xu, J.; Codrea, V.; Buckley, P. E.; Calm, A. M.; Welsh, H. S.; Warner, C. R.; Zacharko, M. A.; Carney, J. P.; Gray, J. J.; Georgiou, G.; Kuhlman, B.; Ellington, A. D. *Chem. Biol.* **2012**, *19* (4), 449.

(36) Hajduczki, A.; Majumdar, S.; Fricke, M.; Brown, I. a M.; Weiss, G. a. *ACS Chem. Biol.* **2011**, *6* (4), 301.

(37) Stanzl, E. G.; Trantow, B. M.; Vargas, J. R.; Wender, P. a. *Acc. Chem. Res.* **2013**, *46* (12), 2944.

(38) Vives, E.; Brodin, P.; Lebleu, B. *J. Biol. Chem.* **1997**, *272* (25), 16010.

(39) Mitchell, D. J.; Kim, D. T.; Steinman, L.; Fathman, C. G.; Rothbard, J. B. *J. Pept. Res.* **2000**, *56* (5), 318.

(40) Fuchs, S. M.; Raines, R. T. *ACS Chem. Biol.* **2007**, *2* (3), 167.

(41) Fuchs, S. M.; Rutkoski, T. J.; Kung, V. M.; Groeschl, R. T.; Raines, R. T. *Protein Eng. Des. Sel.* **2007**, *20* (10), 505.

(42) Sundlass, N. K.; Raines, R. T. *Biochemistry* **2011**, *50* (47), 10293.

(43) Cronican, J. J.; Thompson, D. B.; Beier, K. T.; McNaughton, B. R.; Cepko, C. L.; Liu, D. R. *ACS Chem. Biol.* **2010**, *5* (8), 747.

(44) McNaughton, B. R.; Cronican, J. J.; Thompson, D. B.; Liu, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (15), 6111.

(45) Zuris, J. A.; Thompson, D. B.; Shu, Y.; Guilinger, J. P.; Bessen, J. L.; Hu, J. H.; Maeder, M. L.; Joung, J. K.; Chen, Z.-Y.; Liu, D. R. *Nat. Biotechnol.* **2014**, *33* (1), 73.

(46) Bruce, V. J.; Lopez-Islas, M.; McNaughton, B. R. *Protein Sci.* **2016**.

(47) Kubala, M. H.; Kovtun, O.; Alexandrov, K.; Collins, B. M. *Protein Sci.* **2010**, *19* (12), 2389.

(48) Pruszynski, M.; Koumarianou, E.; Vaidyanathan, G.; Revets, H.; Devoogdt, N.; Lahoutte, T.; Zalutsky, M. R. *Nucl. Med. Biol.* **2013**, *40* (1), 52.

(49) Conrath, K. E.; Lauwereys, M.; Galleni, M.; Matagne, A.; Frère, J. M.; Kinne, J.; Wyns, L.; Muyldermans, S. *Antimicrob. Agents Chemother.* **2001**, *45* (10), 2807.

(50) Liao, X.; Rabideau, A. E.; Pentelute, B. L. *ChemBioChem* **2014**, *15* (16), 2458.

(51) Rudensky, Y. A.; Preston-Hurlburt, P.; Hong, S. C.; Barlow, A.; Janeway, C. A. *Nature* **1991**, *353* (6345), 622.

(52) Hunt, D. F.; Michel, H.; Dickinson, T. A.; Shabanowitz, J.; Cox, A. L.; Sakaguchi, K.; Appella, E.; Grey, H. M.; Sette, A. *Science* **1992**, *256* (5065), 1817.

(53) De Groot, A. S.; Scott, D. W. *Trends Immunol.* **2007**, *28* (11), 482.

(54) Onda, M.; Beers, R.; Xiang, L.; Nagata, S.; Wang, Q.-C.; Pastan, I. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (32), 11311.

(55) Onda, M.; Beers, R.; Xiang, L.; Lee, B.; Weldon, J. E.; Kreitman, R. J.; Pastan, I. *Proc.*

Natl. Acad. Sci. U. S. A. **2011**, *108* (14), 5742.

(56) Laroche, Y.; Heymans, S.; Capaert, S.; De Cock, F.; Demarsin, E.; Collen, D. *Blood* **2000**, *96* (4), 1425.

(57) Warmerdam, P. A. M.; Vanderlick, K.; Vandervoort, P.; De Smedt, H.; Plaisance, S.; De Maeyer, M.; Collen, D. *J. Immunol.* **2002**, *168* (1), 155.

(58) Warmerdam, P. A. M.; Plaisance, S.; Vanderlick, K.; Vandervoort, P.; Brepoels, K.; Collen, D.; De Maeyer, M. *Thromb. Haemost.* **2002**, *87* (4), 666.

(59) Parker, A. S.; Choi, Y.; Griswold, K. E.; Bailey-Kellogg, C. *J. Comput. Biol.* **2013**, *20* (2), 152.

(60) Blazanovic, K.; Zhao, H.; Choi, Y.; Li, W.; Salvat, R. S.; Osipovitch, D. C.; Fields, J.; Moise, L.; Berwin, B. L.; Fiering, S. N.; Bailey-Kellogg, C.; Griswold, K. E. *Mol. Ther. Methods Clin. Dev.* **2015**, *2* (April), 15021.

(61) Zhao, H.; Verma, D.; Li, W.; Choi, Y.; Ndong, C.; Fiering, S. N.; Bailey-Kellogg, C.; Griswold, K. E. *Chem. Biol.* **2015**, *22* (5), 629.

(62) Mazor, R.; Vassall, A. N.; Eberle, J. A.; Beers, R.; Weldon, J. E.; Venzon, D. J.; Tsang, K. Y.; Benhar, I.; Pastan, I. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (51), E3597.

(63) King, C.; Garza, E. N.; Mazor, R.; Linehan, J. L.; Pastan, I.; Pepper, M.; Baker, D. *Proc. Natl. Acad. Sci.* **2014**, *111* (23), 8577.

(64) Mazor, R.; Eberle, J. a; Hu, X.; Vassall, A. N.; Onda, M.; Beers, R.; Lee, E. C.; Kreitman, R. J.; Lee, B.; Baker, D.; King, C.; Hassan, R.; Benhar, I.; Pastan, I. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (23), 8571.

(65) Veronese, F. M. *Biomaterials* **2001**, *22* (5), 405.

(66) Abuchowski, A.; Mccoy, J. R.; Palczuk, N. C.; Van Es, T.; Davis, F. F. *J. Biol. Chem.*

1976, *252* (11), 3582.

(67) Roberts, M. J.; Bentley, M. D.; Harris, J. M. *Adv. Drug Deliv. Rev.* **2002**, *54* (4), 459.

(68) Veronese, F. M.; Pasut, G. *Drug Discov. Today* **2005**, *10* (21), 1451.

(69) Leader, B.; Baca, Q. J.; Golan, D. E. *Nat. Rev. Drug Discov.* **2008**, *7* (1), 21.

(70) Vigerust, D. J.; Shepherd, V. L. *Trends Microbiol.* **2007**, *15* (5), 211.

(71) Sliepen, K.; Van Montfort, T.; Melchers, M.; Isik, G.; Sanders, R. W. *J. Biol. Chem.* **2015**, *290* (12), 7436.

CHAPTER TWO

Split-Superpositive Green Fluorescent Protein Reassembly is a Fast, Efficient, and Robust Method for Detecting Protein-Protein Interactions in Living Cells

Adapted from:
Blakeley, B.D.; Chapman, A.M.; McNaughton, B.R., *Mol. BioSyst.*, **2012**, *8*, 2036.

In this work, led by third year graduate student Brett Blakeley, I assisted in gene construction and molecular cloning, as well as planning and execution of flow cytometry experiments.

2.1 Introduction

The development of methods that rapidly and accurately identify interactions between structurally diverse proteins and/or peptides is key to the development of new proteins with novel function, as well as proteomics. Techniques such as immunoprecipitation, mass spectrometry, affinity purification, and protein microarrays have been used to identify interactions involving proteins and/or peptides *in vitro*.[1] However, these approaches are laborious (typically requiring expression and purification steps for each protein or peptide studied), low-throughput, are often limited to high-affinity interactions, and can involve complicated and/or expensive equipment. Perhaps most importantly, these methods do not provide any strong information on the likelihood of identified interactions occurring *in vivo*.

Popular *in vivo* approaches to identify and study interactions involving proteins and/or peptides include two-hybrid screening and split-protein reassembly.[2,3] Perhaps the most common limitation of two-hybrid screening is the significant number of false positive results and the need for nuclear localization of the interacting protein or peptide. In addition, visualization of an interaction using two-hybrid methods requires subsequent transcription and translation of a reporter protein, which increases the overall length and complexity of the screen.

Split-protein reassembly, also known as a protein fragment complementation assay (PCA), offers an alternative to two-hybrid methods, and has been used to identify and study interactions involving proteins and/or peptides in prokaryotic and eukaryotic cells. In a split-reporter reassembly experiment, a reporter protein is split into two fragments and fused to possible interacting peptide and/or protein partners. In the absence of fused binding partners, the split-reporter fragments do not reassemble and reporter activity is not observed. However, if the fused proteins and/or peptides have affinity for one another, that interaction brings the split-reporter fragments into close proximity, those fragments associate and/or fold, and a functional reporter protein is generated.

Reporter proteins typically fluoresce, catalyze a colorimetric or fluorescent reaction, or endow the host cell with resistance to an exogenous toxin. Common split-reporter proteins used in *E. coli*, *S. cerevisiae*, and mammalian cells include β-lactamase[4], β-galactosidase[5], dihydrofolate reductase (DHFR)[6], ubiquitin[7], and Green Fluorescent Protein (GFP)[8]. In many ways, GFP is perfectly suited as a reporter in a PCA experiment. Unlike enzymatic reporter proteins that catalyze the formation of a fluorescent or colorimetric molecule from a precursor, or catalyze the degradation of an exogenous toxin, GFP does not require the addition of exogenous reagents in order to generate a signal. In addition, GFP expresses, folds, and fluoresces in a large number of

cell types and intracellular compartments, and is generally resistant to proteolytic degradation.[9] Since the formation of a fluorescent chromophore in GFP is irreversible, split-GFP reassembly can be used to examine weak interactions with dissociation constants (K_D) as high as 1 mM.[10]

Despite the operational simplicity of split-GFP reassembly, the susceptibility of GFP fragments to aggregation limits its more general use.[10] Aggregation of split-GFP fragments lowers the cellular concentration of competent split-GFP fusions. As a result, the probability of interacting pairs finding each other in a complex cellular environment is decreased and the overall number of possible interaction-dependent GFP reassembly events is lowered. Unsurprisingly, interaction-dependent split-GFP reassembly can be slow and inefficient (see below). Slow evolution of interaction-dependent GFP fluorescence increases the time required to identify an interaction and low total cell fluorescence increases the difficulty of isolating "hits" from a large library of possible interacting protein and/or peptide pairs.

Split-sg100 GFP, a split-GFP reporter prepared from an enhanced stability variant, is a commonly used system for identifying and studying interactions involving proteins and/or peptides.[8,10] Split-sg100 GFP reassembly typically requires 24 - 72 hours in order to generate visible levels of cellular GFP fluorescence.[10] Since the stability of GFP fragments (or most proteins and/or peptides, for that matter) typically decreases as a function of increased temperature, interaction-dependent reassembly screens using split-sg100 GFP fusions are often performed well below physiological temperature (typically 20 °C - 30 °C). Since binding interactions are often stabilized at lower temperatures, it is reasonable to anticipate that some interactions identified at lower temperatures may not take place at physiological temperature (37 °C). Therefore, the number of false positive results may be significant.

As discussed in detail in chapter one, researchers have noted a strong correlation between protein solubility and theoretical net charge[11], and have taken advantage of this correlation through a protein resurfacing process referred to as supercharging. Researchers in the Liu lab developed a positively supercharged variant of GFP (+36 GFP) with enhanced stability and resistance to aggregation, while maintaining similar fluorescence characteristics as wild-type GFP.[12] With this technology in mind, we hypothesized that supercharging could be applied to split-GFP reassembly to improve on the limitations discussed above. We postulated that the individual GFP fragments should be substantially more resistant to aggregation in comparison to non-supercharged counterparts, which have low theoretical net charge (-4 for each fragment), and that higher concentration of soluble split-GFP fragments would therefore be present in the cell. If this effect outweighs any inhibitory effect of the association rate of the GFP fragments due to electrostatic repulsion, faster and more efficient split-superpositive GFP (split-spGFP) reassembly would expected (summarized in **Figure 2.1**). In addition, we hoped that split-spGFP fragments might increase the solubility of fused proteins and peptides, potentially expanding the scope of split-GFP reassembly to proteins and peptides that are themselves susceptible to aggregation.

The theoretical net charge of sg100 GFP is -8. When sg100 GFP is split into N- and C-terminal fragments, each has a theoretical net charge of -4 (**Figure 2.1A**). In order to compare the efficiency of split-sg100 GFP reassembly to a split-supercharged GFP variant, we prepared N-terminal and C-terminal superpositive GFP (spGFP) fragments with net theoretical charges of +24 and +10, respectively (**Figure 2.1B**). This split-spGFP (based on +34 GFP) is similar to the supercharged GFP reported by Liu and coworkers.[12] We cloned the spGFP N- and C-terminal fragments into pET11a and pMRBAD plasmids, respectively, as fusions to high-affinity

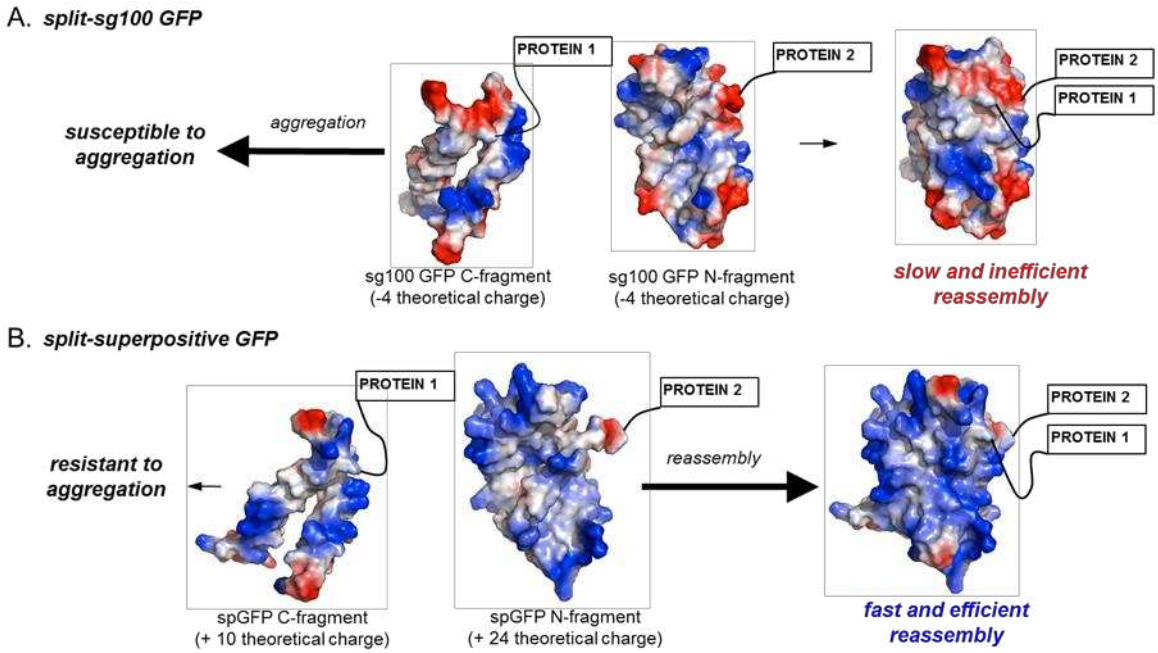


Figure 2.1 (*A*) sg100GFP fragments have low theoretical net charge (-4 for each fragment). (*B*) Superpositive GFP fragments have high theoretical net charge (+10 for the C-terminal fragment and +24 for the N-terminal fragment). We hypothesized that peptide and/or protein fusions connected to sg100GFP fragments would be more susceptible to aggregation, by virtue of the fact that each sg100 fragment is not well folded and has a low theoretical net charge. In contrast, each spGFP fragment has a high theoretical net charge, and should therefore be comparatively resistant to aggregation. Electrostatic surface potentials are colored from -25 kT/e (red) to +25 kT/e (blue).

antiparallel leucine zipper peptides, which have been used extensively as a positive control to measure interaction-dependent reassembly of split-sg100 GFP fragments.[8,10]

2.2 Split-spGFP Reassembly Produces a Brighter Fluorescence Signal

We co-transformed BL21 *E. coli* (DE3) with NGFP-Z pET11a and Z′-CGFP pMRBAD, where Z and Z′ are the positive control leucine zipper peptides and GFP is an N- or C-terminal fragment of either sg100GFP or spGFP. In order to compare the maximum cell fluorescence generated as a result of split-GFP reassembly, *E. coli* were induced to express the split-GFP-Z fusions with isopropyl β-D-1-thiogalactopyranoside (IPTG) and arabinose, incubated at 25 °C, and analyzed by flow cytometry. For both systems, maximum cell fluorescence was reached after 12

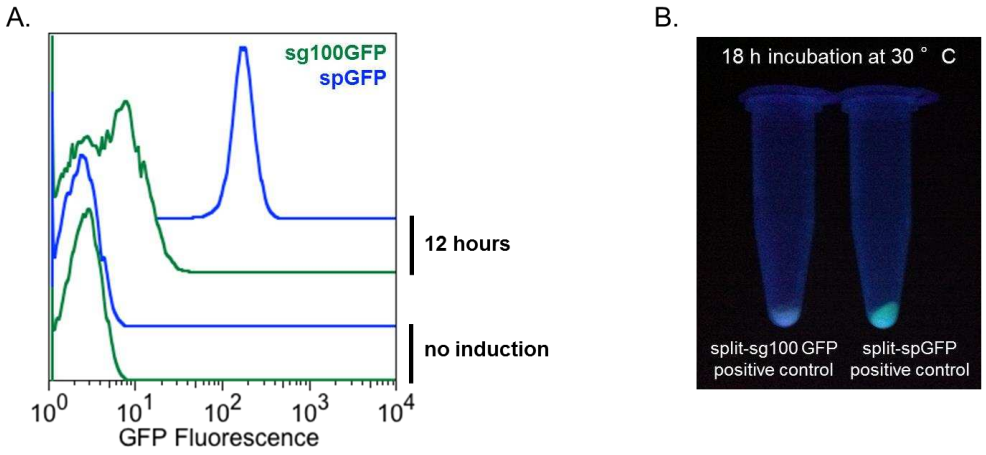


Figure 2.2 Efficiency of GFP reassembly for split-sg100GFP and split-spGFP leucine zipper positive controls in *E. coli* that was incubated for 12 hours at 25 °C. (*A*) Flow cytometry analysis showing GFP fluorescence levels result of either split-sg100GFP reassembly (green) or split-spGFP reassembly (blue). (*B*) Comparative GFP cell fluorescence split-GFP controls. *E. coli* was illuminated with a hand held 365 nm UV lamp.

hours. Consistent with our hypothesis, the mean cellular fluorescence in *E. coli* expressing the split-spGFP positive control fragments increased 75-fold compared to uninduced cells, while mean cellular fluorescence in *E. coli* expressing the split-sg100 GFP positive control fragments only increased 2.3-fold, and fluorescence distribution was broad (**Figure 2.2A**). When *E. coli* containing each set of positive control plasmids were incubated at 30 °C for 18 hours, conditions previously reported[10,13], much brighter cell fluorescence was observed visually in *E. coli* expressing the split-spGFP positive controls (**Figure 2.2B**). *E. coli* containing the split-sg100 GFP or split-spGFP constructs does not generate significant cell fluorescence in the absence of IPTG and arabinose induction reagents (**Figure 2.2A**). Taken together, these results clearly show that split-spGFP reassembly is much more efficient than split-sg100 GFP reassembly. This 72.8-fold increase in cell fluorescence drastically simplifies the identification of interacting pairs when using either flow cytometry or by picking fluorescent colonies.

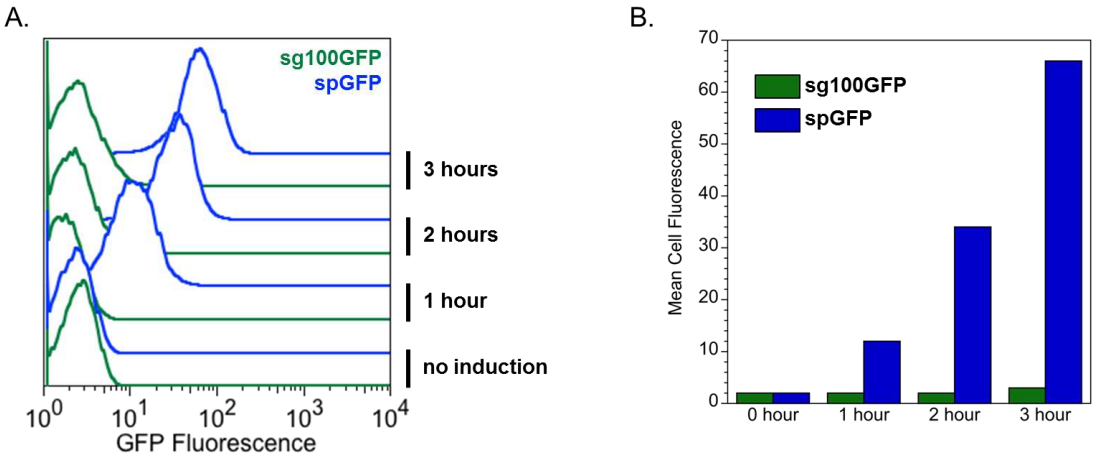


Figure 2.3 Rate of GFP reassembly for split-sg100GFP and split-spGFP leucine zipper fusion positive controls for *E. coli* incubated at 25 °C. (*A*) Flow cytometry analysis showing the rate of GFP fluorescence as a result of split-GFP reassembly at 25 °C. (*B*) Quantitated levels of mean cell fluorescence.

2.3 Split-spGFP Reassembly Facilitates Faster Signal Generation

Next, we measured the rate of split-GFP reassembly in *E. coli* expressing the split-sg100 GFP or split-spGFP positive controls. *E. coli* was induced, incubated at 25 °C, and cell fluorescence was measured by flow cytometry after 1, 2, and 3 hours. The results of these experiments are shown in **Figures 2.3A** and **2.3B**. One hour after incubation, no appreciable increase in cell fluorescence is observed in *E. coli* expressing the split-sg100 Z peptide positive control fusions. In contrast, after the same period of time, a 5-fold increase in mean cell fluorescence was observed in *E. coli* expressing the split-spGFP positive control fusions. After two hours, no appreciable change in fluorescence was observed in *E. coli* expressing the split-sg100 GFP leucine zipper fusions, while a 15-fold change was observed for the split-spGFP system. Three hours after induction, a very modest 1.2-fold increase in mean cell fluorescence was observed as a result of split-sg100 GFP reassembly. After the same period of time, a 28-fold increase in mean cell fluorescence was observed in *E. coli* expressing the split-spGFP positive controls. Taken together, these data clearly show that split-spGFP reassembly is much faster than

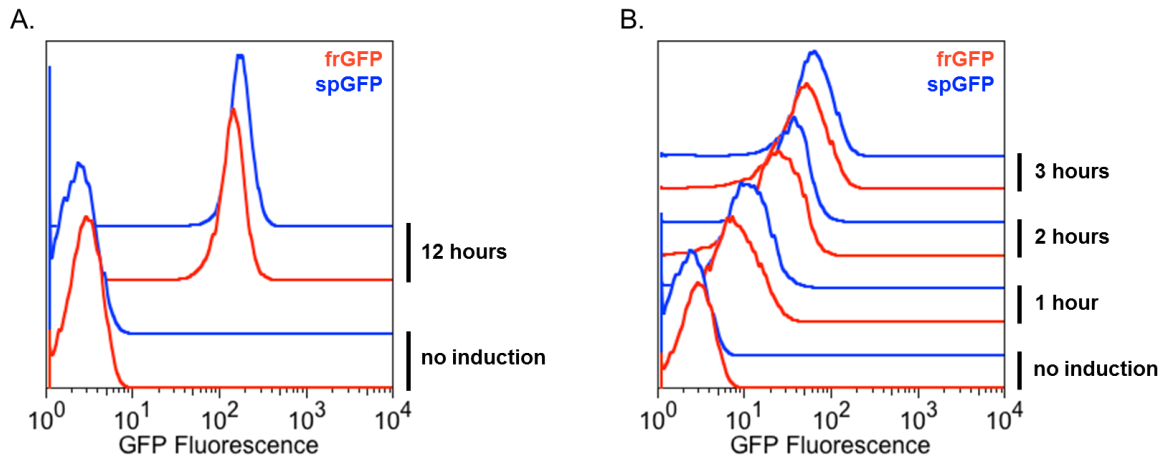


Figure 2.4 Overall efficiency and rate of GFP reassembly of split-frGFP (red) and split-spGFP (blue) leucine zipper positive controls for *E. coli* incubated at 25 °C as analyzed by flow cytometry. (*A*) Overall efficiency of GFP fluorescence as a result of split-GFP reassembly. (*B*) Rate of mean cell fluorescence evolution.

split-sg100 GFP reassembly. This increase in reassembly rate drastically shortens the experimental time needed to visualize and identify an *in cellulo* interaction. Interaction-dependent changes in cell fluorescence as a result of split-spGFP reassembly are easily observed by flow cytometry in as little as one hour.

2.4 Comparison of Split-spGFP and Split-frGFP Assembly Efficiency

Magliery and coworkers recently described split-folding reporter-GFP (split-frGFP), which is constructed from a GFP variant optimized to fold robustly.[13,14] When fused to antiparallel leucine zipper peptides, split-frGFP reassembly was qualitatively determined to be faster and more efficient than split-sg100 GFP reassembly. We compared the efficiency and rate of GFP reassembly for split-frGFP and split-spGFP by flow cytometry.

As shown in **Figure 2.4A**, when incubated at 25 °C for 12 hours, which is when maximum fluorescence is observed for both systems, *E. coli* expressing the split-frGFP and split-spGFP positive control leucine zipper fusions both exhibit high levels of cell fluorescence. However,

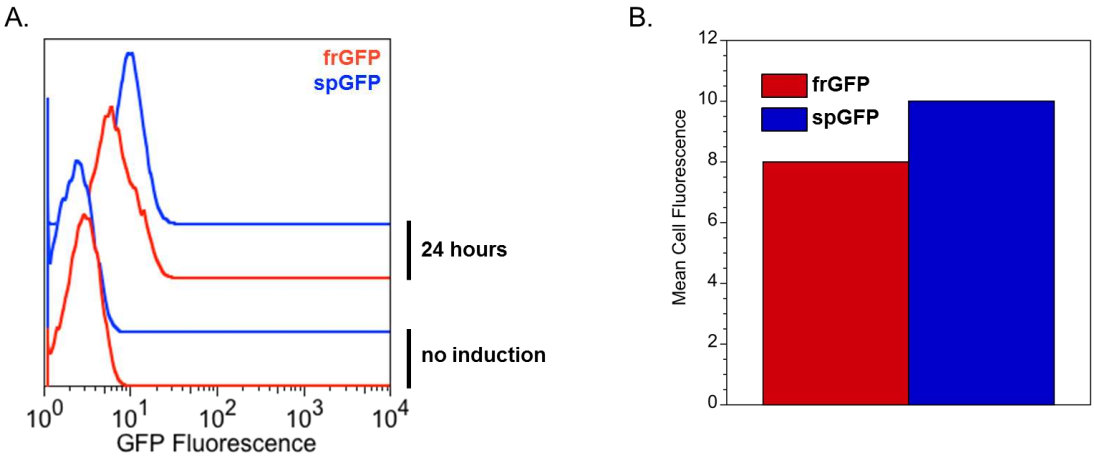


Figure 2.5 (*A*) Flow cytometry analysis of GFP reassembly of split-frGFP (red) and split-spGFP (blue) positive control leucine zipper fusions for *E. coli* incubated at 37 °C for 24 hours. (*B*) Quantitated levels of mean cell fluorescence.

maximum mean cell fluorescence was 22% brighter in *E. coli* expressing the split-spGFP positive controls (**Figure 2.4A**). As seen in **Figure 2.4B**, reassembly of the split-spGFP positive controls is faster than reassembly of split-frGFP. One hour after induction, *E. coli* expressing the split-spGFP positive control fusions was 34% more fluorescent than cells expressing the split-frGFP fusions.

2.5 Split-spGFP Reassembly is Brighter at Physiological Temperature

As stated previously, reassembly systems that operate efficiently at 37 °C are ideal, since interactions identified at this temperature are more likely to occur at physiological conditions. As shown in **Figures 2.5A** and **2.5B**, interaction-dependent split-spGFP reassembly is more efficient than split-frGFP reassembly at 37 °C. When incubated for 24 hours, *E. coli* expressing the split-spGFP positive control antiparallel leucine zipper fusions were 24% more fluorescent than cells expressing the split-frGFP fusions. The increase in split-spGFP reassembly efficiency at 37 °C

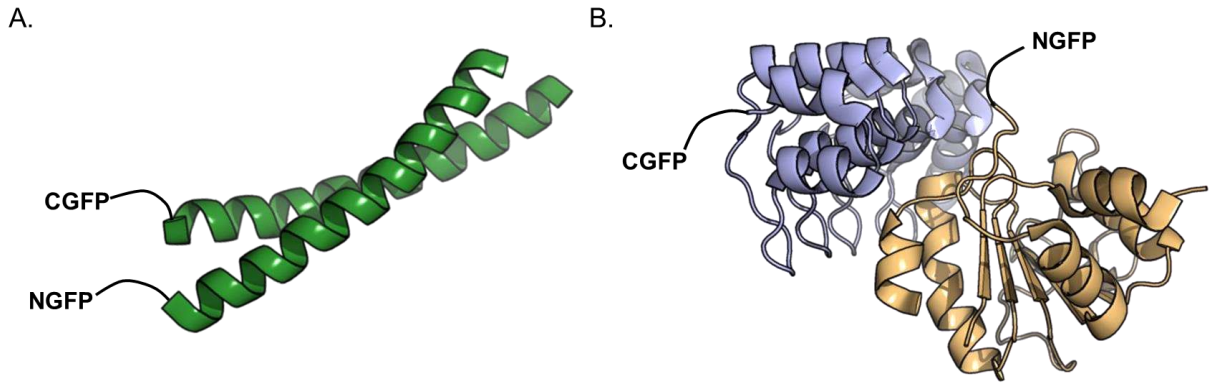


Figure 2.6 (*A*) Fusion placement of split-GFP fragments when fused to Z and Z' antiparallel leucine zippers. (*B*) Fusion placement of split-GFP fragments when fused to Pdar (blue) and Prb (tan). Antiparallel leucine zipper PDB code: 1SER, Pdar-Prb PDB code: 3Q9N.

demonstrates the robustness of this reporter system and its use to identify interactions at physiological conditions.

2.6 Split-GFP Reassembly is Orientation-Independent

Split-GFP reassembly is often used to study interactions involving relatively short helical peptides such as leucine zipper interactions or simple protein - protein interactions involving helical assembly. In many ways, an interaction involving two leucine zipper peptides is ideal for split-GFP reassembly because the antiparallel leucine zipper interaction perfectly aligns split-GFP fragments for reassembly (**Figure 2.6A**). In order to be used more generally, split-GFP reassembly systems need to efficiently detect an interaction involving proteins with additional structural complexity. In addition, unlike leucine zipper interactions, most interactions will not have perfectly aligned N- and C-termini between the interacting pairs. The ability to detect interactions that involve complex protein - protein interactions where the N- and C-termini of those proteins are not perfectly aligned for split-GFP reassembly is critical to more general use, as well as for use in high-throughput screening of protein libraries.

In order to compare the efficiency of protein - protein interaction detection using split-spGFP and split-frGFP, we fused their respective N- and C-terminal fragments to Pdar and Prb, a high-affinity (K_D ~180 pM) designed protein - protein interaction discussed in chapter one.[15] As seen in **Figure 2.6B**, the N- and C-termini of Pdar and Prb are not perfectly aligned for split-GFP reassembly. *E. coli* harboring either NspGFP-Prb/Pdar-CspGFP or NfrGFP-Prb/Pdar-CfrGFP plasmids were induced with IPTG and arabinose, and Pdar/Prb interaction-dependent cell fluorescence was monitored at 37 °C by flow cytometry. For both split-GFP systems, maximum cell fluorescence was reached after 12 hours. Cells expressing the split-spGFP Pdar/Prb fusions were 27% more fluorescent than cells expressing the split-frGFP Pdar/Prb fusions (**Figure 2.7**). This increase in cell fluorescence demonstrates the increased efficiency, utility, and robustness of split-spGFP reassembly. Based on these findings, we postulate that split-spGFP reassembly results from a high-effective molarity of complementary GFP fragments brought together by the PPI, rather than requiring an orientation-specific interaction to hold the fragments together (**Figure 2.8**). This reassembly model therefore expands the scope of split-spGFP to PPIs that are orientation-independent, as compared to many other PCA methods, which are generally limited to ideal binding-partner orientation.

One potential concern we had was that the individual GFP fragments would exist in such high cellular concentrations, due to decreased aggregation, that they may reassemble independent of interacting fused proteins. Towards this end, we fused a pair of non-interacting protein partners, the ankyrin repeat protein gankyrin (discussed in chapter three)[16] and Prb, to the C- and N-fragments of spGFP, respectively. Flow cytometry was measured after 24 hours, comparing induced and uninduced *E. coli*, resulting in no appreciable change in GFP fluorescence (**Figure**

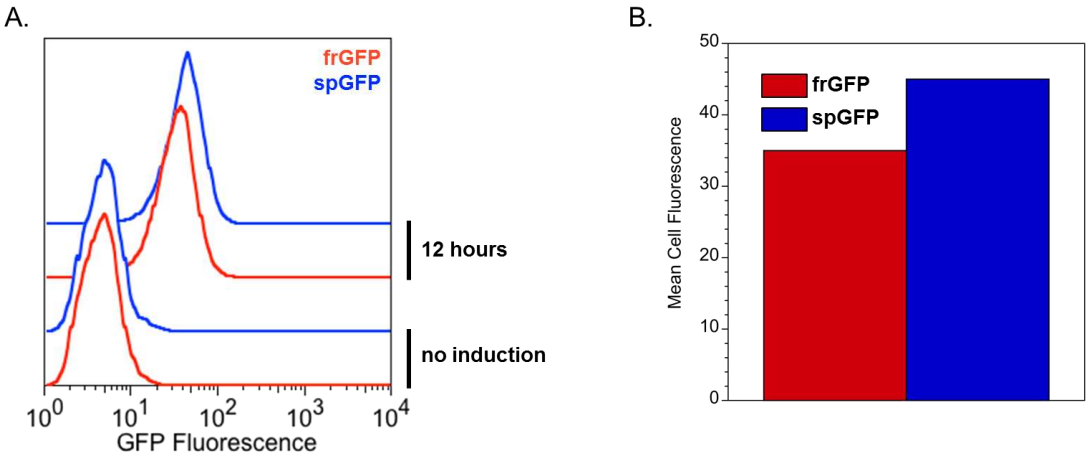


Figure 2.7 (*A*) Flow cytometry analysis of GFP reassembly of split-frGFP (red) and split-spGFP (blue) positive control Pdar/Prb fusions for *E. coli* incubated at 37 °C for 12 hours. (*B*) Quantitated levels of mean cell fluorescence.

S2.4). This finding demonstrates that split-spGFP fragments do not spontaneously reassemble in the absence of a PPI.

2.7 Conclusion

The susceptibility of split-GFP fragments to aggregation has been suspected to play a significant role in decreasing the efficiency and rate of interaction-dependent reassembly.[10] Supercharged resurfaced proteins, including superpositive GFPs, have been shown to be resistant to aggregation.[12] We constructed a split-superpositive GFP consisting of N- and C-terminal fragments that have theoretical net charges of +24 and +10, respectively. When fused to interacting antiparallel leucine zipper peptides, split-spGFP undergoes interaction-dependent reassembly faster and more efficiently than split-sg100 GFP and split-frGFP. Split-spGFP reassembly has been shown to efficiently detect formation of complexes involving interacting peptides or proteins at physiological temperature. Given its overall high efficiency and robustness, we anticipate that split-spGFP will find broad use in the proteomics, chemical biology, biochemistry, and biophysics

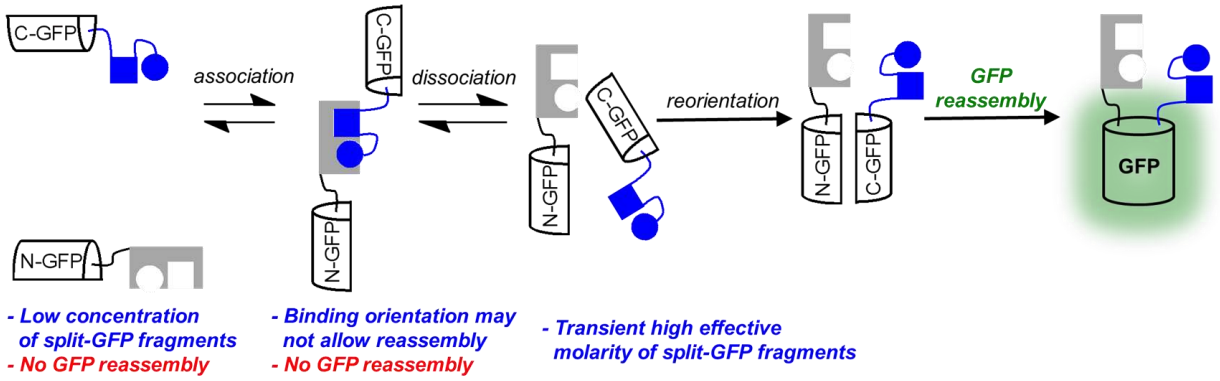


Figure 2.8 Proposed model for orientation-independent split-GFP reassembly. Interacting proteins fused to the N- and C-fragments of GFP bind in an orientation that may not allow GFP reassembly, but rather facilitates a high effective molarity of split-GFP fragments. Dissociation of the PPI allows for reorientation of the split-GFP fragments, resulting in functionally irreversible GFP reassembly.

communities, and may be useful for identifying and/or studying interactions involving proteins or peptides that are themselves susceptible to aggregation. Furthermore, we expect that split-spGFP will be particularly useful for high-throughput screening applications, largely due to the increased rate of fragment reassembly and minimized loss of potential library members to aggregation. Towards this end, application of split-spGFP as a high-throughput screening tool is discussed further in chapter three.

2.8 Methods

Plasmid Construction Nsg100GFP-Z pET11a (NGFP-Z) and Z′-Csg100GFP pMRBAD (Z′-CGFP) plasmids were generously provided from Professor Lynn Regan (Yale University). NfrGFP-Z pET11a and Z′-CfrGFP pMRBAD plasmids were generously provided from Professor Thomas Magliery (Ohio State University). The gene for spGFP was constructed (amino acid sequence of +36 GFP was kindly provided by Professor David Liu at Harvard University) using overlap polymerase chain reaction (PCR) and amplified (all primers purchased from Integrated DNA Technologies, IDT), and cloned into the NcoI and KpnI sites of a pET plasmid (all cloning

enzymes purchased from New England Biolabs, NEB). From this deoxyribonucleic acid (DNA), a DNA sequence encoding the N-terminal fragment of spGFP (encoding amino acids 1-157) was amplified by PCR. This amplicon was double digested with NdeI and BamH1 and cloned into NGFP-Z pET11a, replacing NGFP (sg100), to create the new construct NspGFP-Z pET11a. The C-terminal fragment (CspGFP) was amplified from spGFP pET. This amplicon was double digested with NcoI and BsrG1 and cloned into Z′-CGFP pMRBAD, replacing CGFP, to create the new construct Z′-CspGFP pMRBAD. All constructs used confirmed by GENEWIZ DNA sequencing.

Plasmids encoding Pdar and Prb were generously provided by Professor David Baker (University of Washington). Plasmid encoding gankyrin was purchased from Addgene. Pdar and gankyrin were amplified, double digested with NcoI and AatII and cloned into -CfrGFP pMRBAD or –CspGFP pMRBAD, replacing Z′. Prb was amplified double digested with XhoI and BamHI and cloned into NfrGFP- pET11a or NspGFP- pET11a, replacing Z.

Generation, Growth and Induction of E. coli Expressing Split-GFP Positive Control Plasmids To co-transform compatible pairs of plasmids (NGFP-Z pET11a and Z′-CGFP pMRBAD, NfrGFP-Z pET11a and Z′-CfrGFP pMRBAD, or NspGFP-Z pET11a and Z′-CspGFP pMRBAD), the construct containing the N-terminal fragment was transformed into chemically competent BL21(DE3) *E. coli* using a standard heat-shock protocol. These cells were then made electrocompetent by standard methods and transformed with the construct containing the respective complementary C-terminal fragment, using electroporation (1 mm cuvettes and 1.7 keV, Bio-Rad). Expression of split-GFP fragments was accomplished by first growing cells overnight at 250 rpm to saturation at 37 °C in LB supplemented with 100 μg/mL ampicillin and 35 μg/mL kanamycin. Overnight cultures were diluted 1:100 into fresh LB supplemented with

100 μg/mL ampicillin and 35 μg/mL kanamycin. Cultures were monitored by optical density and induced at OD_{600} = 0.6 using 100 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 0.2% (L)-arabinose. Cultures were induced under the appropriate conditions as specified.

Flow Cytometry Growth and induction was carried out as described above. Flow cytometry samples were incubated at a desired temperature and agitated at 250 rpm for the time period indicated. *E. coli* samples were prepared for flow cytometry by centrifugation (5000 rpm for 5 minutes) and resuspended in phosphate buffered saline (PBS). All samples were analyzed on a MoFlo (Dako Colorado, Inc.) flow cytometer using a solid-state iCyt 488 nm laser. Flow cytometry data were analyzed using FloJo software (Tree Star).

2.9 Proteins Used in This Work

NspGFP-Z

MGHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGEGKGDATRGKLTLKFIC
TTGKLPVPWPTLVTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKY
KTRAEVKFEGRTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRGGSGSGSS
ALKKELQANKKELAQLKWELQALKKELAQ*

Z′-CspGFP

MASEQLEKKLQALEKKLAQLEWKNQALEKKLAQTSGGSGKNGIKAKFKIRHNVKDGS
VQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMVLLEFVTAAGIKHGR
DERYK*

Nsg100GFP-Link

MASHHHHHHGASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIC
TTGKLPVPWPTLVTTLCYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNY
KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVPIMADKQGGSGSGSS

Link-Csg100GFP

MGTSGGSGKNGIKVNFKTRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALS
KDPNEKRDHMVLLEFVTAAGITHGMDELYN

NfrGFP-Link

MASHHHHHHGVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIC
TTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNY
KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQGGSGSGSS

Link-CfrGFP

MGTSGGSGKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALS
KDPNEKRDHMVLLEFVTAAGITHGMDELYK

Prb

MGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLEHGYDVYPVNPNYE
EVLGRKCYPSVLDIPDKIEVVDLFVNPAKAWRFVVYAIKKGAKVVWFQYNTYYPLAAR
QAKEAGLIIVANRCMMREHERLLGEK*

Pdar

MSDLGKKLLEAAAAGQDDEVRILMANGADVNATDDDGLTPLHLAAANGQLEIVEVLL
KNGADVNASDSAGITPLHLAAYDGHLEIVEVLLKHGADVNAYDRAGWTPLHLAALSG
QLEIVEVLLKHGADVNAQDALGLTAFDISINQGQEDLAEILQ*

Gankyrin

MEGCVSNIMICNLAYSGKLDELKERILADKSLATRTDQDSRTALHWACSAGHTEIVEFL
LQLGVPVNDKDDAGWSPLHIAASAGRDEIVKALLVKGAHVNAVNQNGCTPLHYAASK
NRHEIAVMLLEGGANPDAKDHYDATAMHRAAAKGNLKMVHILLFYKASTNIQDTEG
NTPLHLACDEERVEEAKFLVTQGASIYIENKEEKTPLQVAKGGLGLILKRLAEGEEASM*

2.10 Primers Used in This Work

All DNA primers are listed 5' to 3'

NspGFP FP: GGGAATTCCATATGGGTCATCACCACCACCATCACGG

NspGFP RP: GCGCTCGAGCCAGAGCCAGAGCCACCGCGTTTATCGGCCGTAATATAC

CscGFP FP: GCAGACGTCGGGTGGAAGCGGTAAGAATGGTATCAAGGC

CscGFP RP: AGTTGTACATTACTTGTAGCGTTCGTCGCGTCCGTGCTTAATG

Prb FP: CTGGCTCGAGCGGCAGCACCCGTCCGATTGATGGTCTG

Prb RP: GCCGGATCCTTACTATTTTTCGCCCAGCAGGCGTTCGTGC

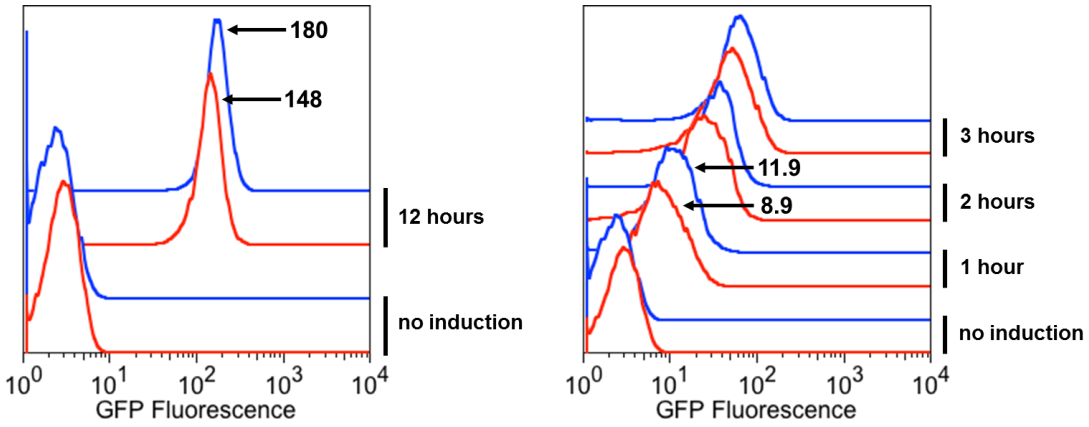
Pdar FP: CATGCCATGGCAAGCGATCTGGGTAAAAAGCTGC

Pdar RP: ACCCGACGTCCCTTGCAGGATCTCTGCCAGATCTTCC

Gankyrin FP: CATGCCATGGAGGGGTGTGTGTCTAACATAATGATC

Gankyrin RP: CACCCGACGTCCCCATAGAAGCCTCTTCACCTTCTGCTAGTCTC

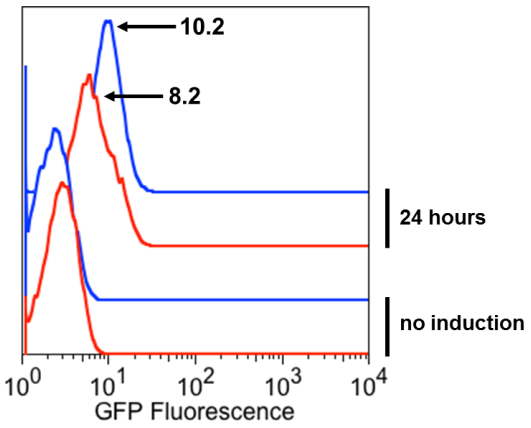
2.11 Supplemental Data



180 – 148 = 32
32/148 = 0.22
22% increase in spGFP cell fluorescence

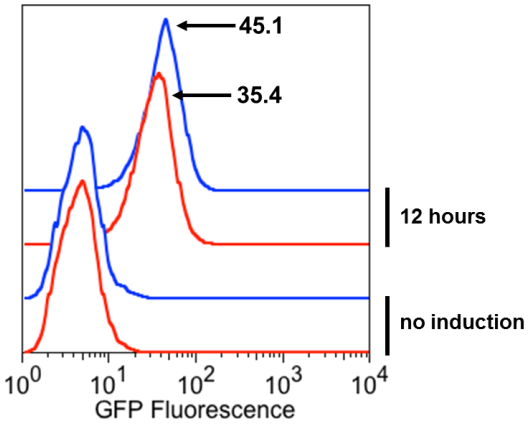
11.9 – 8.9 = 2
2/8.9 = 0.34
34% increase in spGFP cell fluorescence

Figure S2.1 Mean cell fluorescence values are provided for Figure 2.4. Cells were incubated at 25 °C for the indicated time.



10.2 – 8.2 = 2
2/8.2 = 0.24
24% increase in spGFP cell fluorescence

Figure S2.2 Mean cell fluorescence values are provided for Figure 2.5. Cells were incubated at 37 °C for the indicated time.



45.1 – 35.4 = 9.7
9.7/35.4 = 0.27
27% increase in spGFP cell fluorescence

Figure S2.3 Mean cell fluorescence values are provided for Figure 2.7. Cells were incubated at 37 °C for the indicated time.

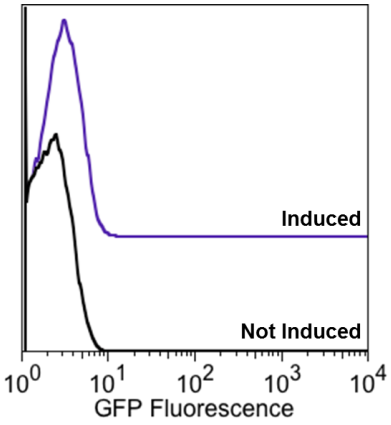


Figure S2.4 Flow cytometry data for *E. coli* expressing NspGFP-Prb and gankyrin-CspGFP 24 hours after induction.

REFERENCES

(1) Howell, J. M.; Winstone, T. L.; Coorssen, J. R.; Turner, R. J. *Proteomics* **2006**, *6* (7), 2050.

(2) Chien, C. T.; Bartel, P. L.; Sternglanz, R.; Fields, S. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (21), 9578.

(3) Shekhawat, S. S.; Ghosh, I. *Curr. Opin. Chem. Biol.* **2011**, *15* (6), 789.

(4) Galarneau, A.; Primeau, M.; Trudeau, L.-E.; Michnick, S. W. *Nat. Biotechnol.* **2002**, *20* (6), 619.

(5) Rossi, F.; Charlton, C. a; Blau, H. M. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (16), 8405.

(6) Pelletier, J. N.; Campbell-Valois, F. X.; Michnick, S. W. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (21), 12141.

(7) Johnsson, N.; Varshavsky, A. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (22), 10340.

(8) Ghosh, I.; Hamilton, A. D.; Regan, L. *J. Am. Chem. Soc.* **2000**, *122* (23), 5658.

(9) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67* (11), 509.

(10) Magliery, T. J.; Wilson, C. G. M.; Pan, W.; Mishler, D.; Ghosh, I.; Hamilton, A. D.; Regan, L. *J. Am. Chem. Soc.* **2005**, *127* (1), 146.

(11) Loeb, J. *J. Gen. Physiol.* **1921**, *4*, 547.

(12) Lawrence, M. S.; Phillips, K. J.; Liu, D. R. *J. Am. Chem. Soc.* **2007**, *129* (33), 10110.

(13) Sarkar, M.; Magliery, T. J. *Mol. Biosyst.* **2008**, *4* (6), 599.

(14) Waldo, G. S.; Standish, B. M.; Berendzen, J.; Terwilliger, T. C. *Nat. Biotechnol.* **1999**, *17* (7), 691.

(15) Karanicolas, J.; Corn, J. E.; Chen, I.; Joachimiak, L. a; Dym, O.; Peck, S. H.; Albeck, S.;

Unger, T.; Hu, W.; Liu, G.; Delbecq, S.; Montelione, G. T.; Spiegel, C. P.; Liu, D. R.; Baker, D. *Mol. Cell* **2011**, *42* (2), 250.

(16) Nakamura, Y.; Nakano, K.; Umehara, T.; Kimura, M.; Hayashizaki, Y.; Tanaka, A.; Horikoshi, M.; Padmanabhan, B.; Yokoyama, S. *Structure* **2007**, *15* (2), 179.

CHAPTER THREE

A Resurfaced Shape Complementary Protein That Selectively Binds the Oncoprotein Gankyrin

Adapted from:
Chapman, A.M.; McNaughton, B.R., *ACS Chem. Biol.*, **2014**, *9*, 2223.

3.1 Introduction

Ankyrin repeats are among the highest populated domains in Nature and play critical roles in various disease-relevant cellular processes. These domains almost exclusively participate in PPI's, and mutation or changes in expression level of many ankyrin repeats is directly linked to disease, including many cancers. Ankyrin repeats typically bind their targets on the extended concave face formed by a repeating helix-turn-helix-loop motif, and the surface area of many ankyrin-repeat PPI's is often much larger than the average of ~1400 $Å^2$ for a PPI (**Figure 3.1**).[1] While disruption of PPI's involving ankyrin repeats is a validated therapeutic strategy, doing so is a significant challenge, since these assemblies are often stabilized over large, discontinuous surfaces, mainly consisting of charged and polar residues. Thus, it should be unsurprising that recognition and modulation of ankyrin repeat-containing PPIs is a complex task for small-molecule reagents (<800 Da), which traditionally bind protein receptors in small, well-defined hydrophobic pockets. In an informative example, a fragment-based discovery approach recently identified small molecules that poorly bind the Notch-1 ankyrin repeat domain with a K_D of ~10 mM.[2]

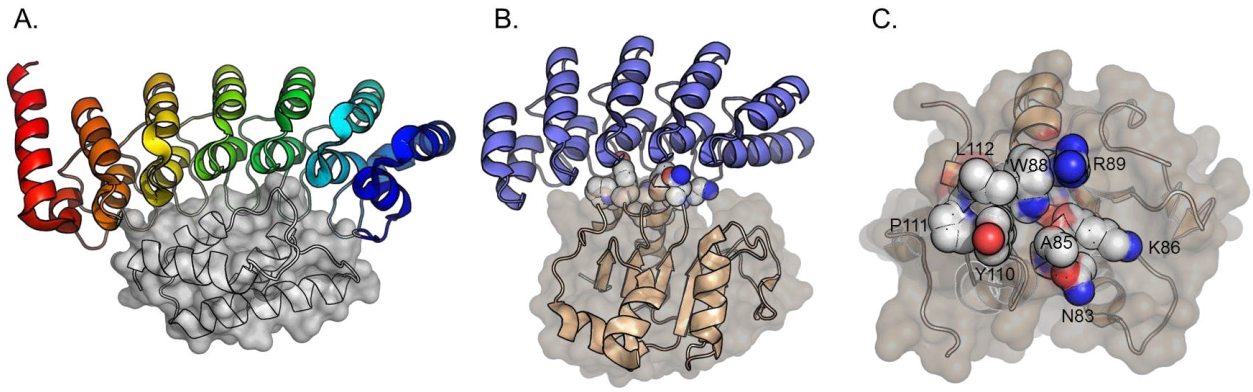


Figure 3.1 (*A*) Complex involving gankyrin (colored) and S6 ATPase (grey). (*B*) Complex involving Pdar (blue) and Prb (light brown). Binding face residues mutated in this work in highlighted. (*C*) View of the Pdar binding face of Prb. Residues randomized during library construction are highlighted and annotated. Gankyrin-S6 ATPase PDB code: 2DVW, Pdar-Prb PDB code: 3Q9N.

Gankyrin is a recently described oncoprotein whose overexpression has been directly linked to the onset, proliferation, and/or metastasis of a wide range of cancers, including breast[3,4], liver[5], oral[6], pancreatic[7], and colorectal cancers[8], as well as esophageal squamous cell carcinoma.[9] Additionally, gankyrin has also been shown to play a crucial role in Ras-mediated tumorigenesis, which is known to initiate ~30% of all cancers.[10] Structurally, gankyrin contains seven ankyrin repeat units, which generate a relatively featureless and large concave binding face. Gankyrin binds to a variety of physiological targets, including cyclin-dependent kinase 4 (CDK4)[11], the E3 ubiquitin ligase MDM2[12], and the C-terminal S6 ATPase subunit of the 26S proteasome (referred to as S6 ATPase throughout).[13] In forming a complex with both CDK4 and Rb, gankyrin drives phosphorylation of Rb (pRb), which leads to activation of E2F transcription factors.[11,14] By binding to MDM2, gankyrin regulates MDM2-dependent polyubiquitination of the tumor suppressor p53, resulting in downregulation of p53 and suppression of cellular apoptosis.[12] Overall, gankyrin overexpression leads to decreased genomic stability and the onset of oncogenic cellular phenotypes. For these reasons, identifying reagents that can bind to gankyrin and modulate gankyrin PPIs is potentially of great therapeutic interest.

The only available structural information for a PPI involving gankyrin (colored) is in complex with S6 ATPase (gray, **Figure 3.1A**). Although relatively little is known about the biological role of this PPI, the interaction highlights why inhibiting interactions featuring gankyrin may be especially challenging. S6 ATPase tightly binds to gankyrin on the concave binding face (a detailed quantitative analysis of the gankyrin-S6 ATPase interaction is discussed in chapter four), making discontinuous contacts with all seven ankyrin repeat units. As a result, the binding interface between gankyrin and S6 ATPase is ~2400 $Å^2$, which is much larger than the average PPI surface (~1400 $Å^2$).[15] It is therefore unlikely that gankyrin recognition and modulation is within the scope of traditional small-molecule reagents.

3.2 Prb as a Shape Complementary Protein Scaffold

As discussed in detail in chapter one, proteins represent an attractive alternative to small molecules for the potent and selective recognition of difficult macromolecular targets, principally due to their ability to take on complex, three-dimensional folds. Modern molecular biology tools have significantly simplified the process through which one can identify new proteins with novel function. Also discussed in chapter one to some length, innovative technologies now exist to allow the functional delivery of proteins into the interior of mammalian cells, opening up the possibility for the modulation of intracellular targets, such as gankyrin.[16–22] Even with these advances, we are still largely incapable of designing functional proteins from scratch. Perhaps the most sensible solution is thus one of semi-design: start with a stable protein with a privileged function, and apply resurfacing to bind a macromolecule of interest.[23]

Baker and coworkers recently used *in silico* design and *in vitro* evolution based on shape complementarity to generate a potent PPI involving a PH1109-derived protein called Prb (**Figure**

3.1B, light brown) and a synthetic thermostable ankyrin repeat called Pdar (**Figure 3.1B**, blue), which was discussed in chapter one.[24] PH1109 is a bacterial CoA-binding protein from the hyperthermophile *Pyrococcus horikoshii.* In contrast to many structurally characterized ankyrin repeat binding proteins, which are generally very large (>50 kDa) and/or unstable, Prb, and mutants thereof, are relatively small proteins (~16 kDa) that are thermostable and expresses exceptionally well in *E. coli*. Additionally, the scaffold has proven that it can reliably be mutated at the putative ankyrin repeat binding interface without fear of distorting the overall fold. All of these are valuable features when considering a scaffold for potentially generating new gankyrin-binding proteins.

The development of the Pdar-Prb complex highlights both the power and current limitations of *in silico* methods, as well as the utility of high-throughput screening and/or macromolecular evolution. For example, while *in silico* design of the complex provided a valuable starting point for its optimization, the reported crystal structure of this protein-protein complex is significantly different from the intended *in silico* design. Additionally, highest affinity complexes (K_Ds into the pM regime) were identified through the application of yeast display screening, a common method used for macromolecular evolution.[24]

Starting with shape complementarity and privileged scaffold resurfacing as design principles, we hypothesized that Prb-derived proteins could be generated to selectively recognize gankyrin. The amino acid backbone of gankyrin and Pdar align exceptionally well, with a backbone root-mean-square deviation (rmsd) value of 0.69 Å over all Pdar residues (**Figure 3.2A***)*. If a Prb-derived protein binds gankyrin in a manner that is similar to the Pdar-Prb complex, the binding face residues on the Prb-derived protein would likely engage large regions of gankyrin, and therefore might be able to compete with – or inhibit – disease-relevant complexes involving

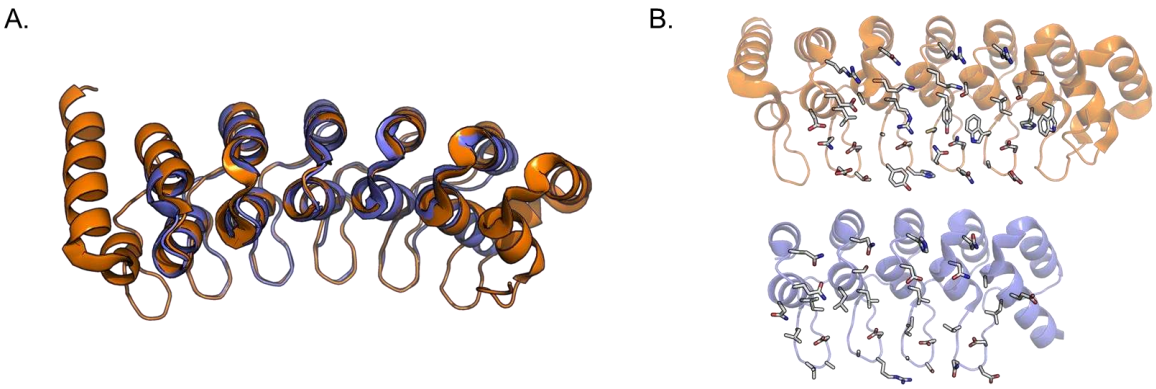


Figure 3.2 (*A*) Superposition of Pdar (blue) and gankyrin (orange) result in a rmsd of 0.69 Å (*B*) Only ~12% sequence homology exists for the residues on the concave binding face of gankyrin (top) and Pdar (bottom), over the region covered by Prb.

gankyrin. However, residues on the concave binding face and loop regions of Pdar and gankyrin are only ~12% sequence homologous (**Figure 3.2B**). Furthermore, analysis of the binding face on Pdar reveal an extensive hydrophobic patch, which facilitates interaction with Prb. In contrast, solvent exposed residues on the putative binding face of gankyrin are primarily polar or charged, suggesting that extensive resurfacing of Prb may be required to achieve selective and potent recognition of gankyrin. These observations indicate that the generation of a novel gankyrin-binding protein is a substantial molecular recognition challenge. Evaluation of the Prb-Pdar complex revealed eight Prb residues that directly engage, or are nearby, the surface of Pdar (N83, A85, K86, W88, R89, Y110, P111, and L112, **Figure 3.1C**). We reasoned that if a binding mode similar to that observed in the Pdar-Prb interaction is utilized, mutation of these residues might result in new proteins that selectively recognize gankyrin. The remaining question was how to best identify new gankyrin-binding proteins from this library.

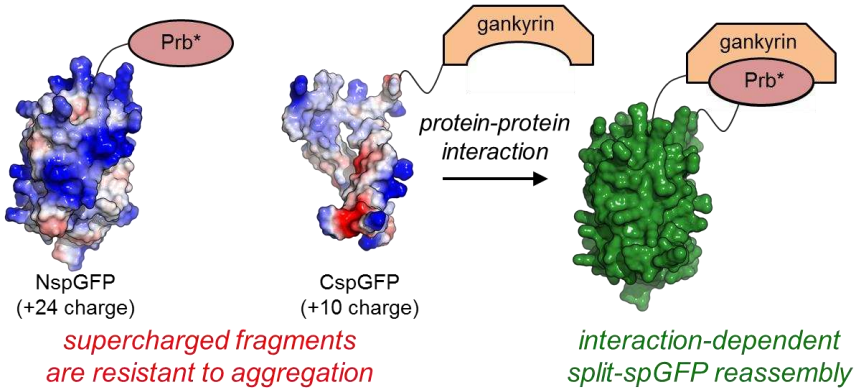


Figure 3.3 Interaction-dependent reassembly of split-superpositive GFP fragments to generate active (fluorescent) GFP.

3.3 Split-spGFP Screening Reveals New Gankyrin-Binding Proteins

In chapter two, we discussed the development of a split-spGFP reassembly assay for the identification of PPIs in living cells (**Figure 3.3**).[25] Split-spGFP reassembly is faster, more efficient, and more robust than previously described split-GFP methods. In contrast to previously reported split-GFP systems[26–28], we postulated that the enhanced properties of split-spGFP reassembly are likely due to a dramatic decrease in aggregation propensity of the individual GFP fragments by virtue of their high theoretical net charge.[29] Substantially decreasing the number of fragments lost to aggregation not only endows the properties described above, it also should increase the number of viable library members when employing split-spGFP in a high-throughput manner. Furthermore, unlike the majority of the more common high-throughput screening methods used by researchers (ie. phage display, yeast display, and mRNA display), which are performed either on the outside of cells or in a test tube, split-spGFP reassembly takes place inside of living cells. Screening for new PPIs in the context of a complex cellular milieu should aid in identifying physiologically-relevant interactions, as well as impart some degree of binding selectivity. For these reasons, we determined that split-spGFP, combined with fluorescence-activated cell sorting

(FACS), would be the ideal high-throughput screening tool to identify new gankyrin-binding proteins.

Using standard molecular biology techniques, we prepared a DNA library that encodes resurfaced Prb with eight randomized binding face residues (shown in **Figure 3.1C**). This Prb library was cloned into a pET plasmid as a C-terminal fusion to NspGFP. We also cloned gankyrin into a pBAD plasmid as an N-terminal fusion to CspGFP. These two plasmids were sequentially transformed into *E. coli*, generating a library of ~5 x 10^9 transformants. Sequencing ~50 library plasmids from our library suggested very efficient randomization of the Prb binding face, as we did not observe any duplicate sequences in this region. Doubly transformed *E. coli* were made to concomitantly express the NspGFP-Prb library fusion and gankyrin-CspGFP fusion proteins, and incubated at 30 °C for 6 hours. After such time, *E. coli* with the highest levels of GFP (indicating interaction-dependent GFP reassembly) were isolated by FACS. The resulting population (~800 cells) was re-sorted to separate real binders from potential false positives, based on the original gate (**Figure 3.4A**). Following two rounds of screening, we individually re-screened seven resurfaced shape complementary proteins, which bind gankyrin in living cells (*E. coli*, **Figure 3.4B**). These proteins are herein referred to as Gankyrin Binding Proteins 1-7 (GBP 1-7). While all seven of these resurfaced proteins bind gankyrin (as determined by split-spGFP reassembly), we focused on the five best performing proteins (GBPs 1, 3, 4, 5, and 7).

We further characterized binding by a lysate Enzyme-Linked Immunosorbant Assay (ELISA), which – in our hands – is more stringent than split-spGFP reassembly. As shown in **Figure 3.4C**, GBP 3, GBP 5 and GBP 7 appear to strongly bind gankyrin, while other GBPs are much weaker binders. Importantly, GBP 5 and GBP 7 do not appreciably bind off-target ankyrin repeats Pdar (green bars) and Notch-1 (red bars), which exhibit very high structural homology with

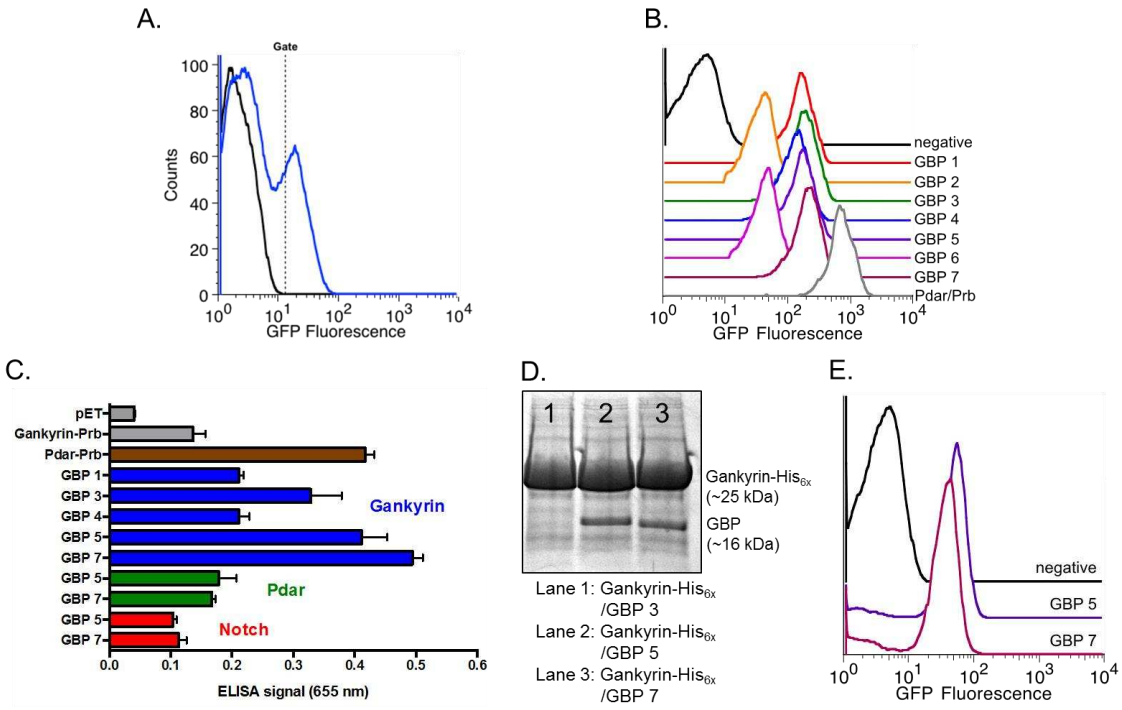


Figure 3.4 (*A*) FACS data showing the gate used to re-sort GFP-positive cells coexpressing gankyrin-CspGFP and NspGFP-Prb library. Negative control cells (gankyrin-CspGFP alone) shown in black, sorted population shown in blue. (*B*) Flow cytometry data showing GFP fluorescence in *E. coli* coexpressing gankyrin-CspGFP and NspGFP-GBP (1-7) or a Pdar/Prb positive control. (*C*) ELISA data. Targets of GBPs are colored as follows: gankyrin (blue); Pdar (green); Notch-1 ARD (red). (*D*) Coomassie-stained PAGE following copurification of gankyrin-His_{6x} and untagged GBP 3, 5, or 7. (*E*) Flow cytometry data showing GFP fluorescence in *E. coli* that express gankyrin-CspGFP/S6 ATPase and NspGFP-GBP 5 or 7. Error bars in panel C represent the standard deviation from three independent experiments.

gankyrin (backbone atom rmsd = 0.69 and 1.27 Å, respectively, **Figure 3.5A**), but differ dramatically with respect to the makeup of amino acids on their concave binding face. Pdar and Notch-1 exhibit ~12% and ~9% sequence homology, respectively, with the concave binding face of Gankyrin (**Figure 3.5B**).

Binding was further confirmed by measuring the amount of GBP that is co-purified with His_{6x}-tagged gankyrin from *E. coli* cell lysate.[30] *E. coli* was induced to co-express His_{6x}-tagged gankyrin and untagged GBP 3, 5 or 7 off of a pETDuet plasmid. Cleared cell lysate was incubated with Ni-NTA agarose, followed by washing steps and elution of His_{6x}-tagged gankyrin by addition of imidazole. Gankyrin, or gankyrin-GBP co-purified complexes were identified by SDS-PAGE

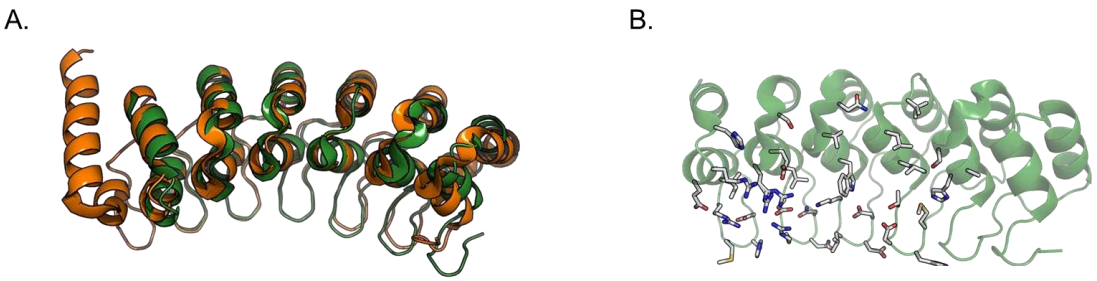


Figure 3.5 (*A*) Superposition of Notch-1 (green) and gankyrin (orange) result in a rmsd of 1.27 Å (*B*) Only ~9% sequence homology exists for the residues on the concave binding face of gankyrin and Notch-1. Notch-1 ARD PDB code: 1YYH.

and coomassie staining. As seen in **Figure 3.4D**, appreciable levels of co-purified GBP 5 and GBP 7 were observed, while essentially no GBP 3 was co-purified with gankyrin, suggesting that GBP 5 and 7 are the highest affinity GBPs, and warrant further study. The relative absence of other co-purified cellular proteins further demonstrates the high level of selectivity that is achieved in these newly identified PPIs.

As stated previously, S6 ATPase does not express independently as a soluble protein. The gankyrin-S6 ATPase complex is only generated by co-expressing these two proteins from a single pETDuet plasmid.[13] In order to determine if GBP 5 or GBP 7 bind gankyrin in the presence of S6 APTase – or inhibit this physiological interaction – we performed a modified split-spGFP experiment. We co-expressed gankyrin-CspGFP and S6 ATPase from pETDuet and NspGFP-GBP 5 or -GBP 7 from pBAD, in *E. coli.* Since gankyrin and S6 ATPase assemble when co-expressed, we reasoned that gankyrin-GBP interaction-dependent reassembly of the fused spGFP fragments would only occur if GBP 5 or GBP 7 bind gankyrin over S6 ATPase, or recognize a region of gankryrin that differs from S6 ATPase. We observed virtually identical amounts of gankyrin interaction-dependent GFP signal in *E. coli* that co-express gankyrin-CspGFP / S6 ATPase and NspGFP-GBP 5 or NspGFP-GBP 7 (**Figure 3.4E**).

3.4 Alanine-Scanning Mutagenesis of GBP 5 and GBP 7

To assess the contribution of each residue on the resurfaced region of GBP 5 and GBP 7, we performed pull-down experiments from *E. coli* cell lysate containing His_{6x}-tagged gankyrin and untagged alanine mutants of each GBP. In each pull-down experiment, a single residue that was randomized in construction of the protein library was mutated to alanine (with the exception of glycine 83 in GBP 5, which we viewed as a minor change unlikely to dramatically alter complex stability). Consistent with our ELISA data, gankyrin does not bind appreciable levels of Prb (**Figure 3.6A**, lane 1), but does co-purify with GBP 5 (**Figure 3.6A**, lane 2). Three mutations to the resurfaced region – R85A, N110A and W111A –decreased the amount of co-purified mutant GBP to varying levels (**Figure 3.6A**, lanes 3, 7, and 8, respectively), suggesting these residues are particularly critical for gankyrin recognition. For GBP 7, Y83A, I85A and W86A mutations resulted in significantly decreased levels of co-purified mutant GBP (**Figure 3.6B**, lanes 3, 4, and 5). For GBP 5 and GBP 7, mutations that result in significantly lower levels of co-purification are tightly grouped and different, suggesting unique recognition "hot spots" (**Figures 3.6C** and **3.6D**).

3.5 Biophysical Characterization of GBP 5 and GBP 7

In the majority of our data (split-spGFP reassembly, ELISA, His_{6x} co-purification) GBP 7 appears to have the highest affinity for gankyrin. We measured the K_D between gankyrin and GBP 7 by Isothermal Titration Calorimetry (ITC). This resurfaced shape complementary protein binds gankyrin with good affinity (K_D ~ 6 μM, **Figure 3.7A**). The observed change in enthalpy (ΔH) and entropy (ΔS) for this binding interaction were -2.78 kcal/mol and 14.6 cal/mol*K, respectively. Consistent with our previous data, GBP 5 binds gankryin, but with lower affinity. We observed an unsaturated binding isotherm under identical conditions that were used to measure

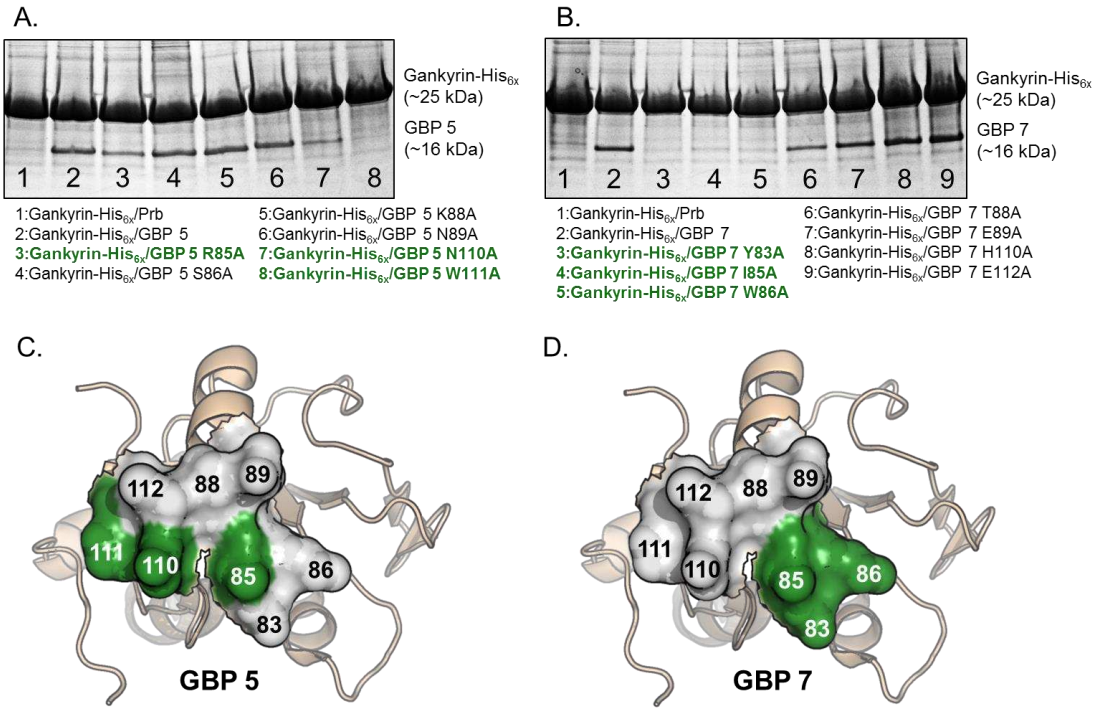


Figure 3.6 (*A*) Coomassie-stained PAGE following co-purification of gankyrin-His_{6x} and untagged Prb, gankyrin binding protein 5 (GBP 5), and alanine mutants thereof (stated below the gel). (*B*) Coomassie-stained PAGE following co-purification of gankyrin-His_{6x} and untagged Prb, gankyrin binding protein 7 (GBP 7), or alanine mutants thereof (stated below the gel). (*C*) Binding face of GBP 5, with key Gankyrin-binding residues highlighted in green. (*D*) Binding face of GBP 7, with key gankyrin-binding residues highlighted in green. Structures shown in (*C*) and (*D*) are of the putative binding face of Prb – which is the the starting point for our protein resurfacing. These representations are not intended to provide any information on structural features of GBP 5 or GBP 7 – or alanine mutants thereof – but rather to graphically represent where mutations deleterious to gankyrin binding reside on GBP 5 and GBP 7. Taken together, these depictions indicate where binding "hot spots" are on the resurfaced proteins GBP 5 and GBP 7, as determined by our pull-down data in (*A*) and (*B*).

the GBP 7 - gankyrin interaction (data not shown). Since GBP 5 and GBP 7 are derived from a protein natively expressed in the hyperthermophile *Pyrococcus horikoshii*, these proteins are likely to be very thermostable, a desired characteristic of protein reagents. We measured the thermostability of Prb and resurfaced gankyrin-binding mutants GBP 5 and GBP 7 by Differential Scanning Calorimetry (DSC). Impressively, Prb exhibits a very high melting temperature (T_m) of 91.1 °C. Despite extensive mutagenesis, both resurfaced mutants GBP 5 and GBP 7 retain excellent thermostability (T_m ~ 86.8 °C and 87.1 °C, respectively, **Figure 3.7B**).

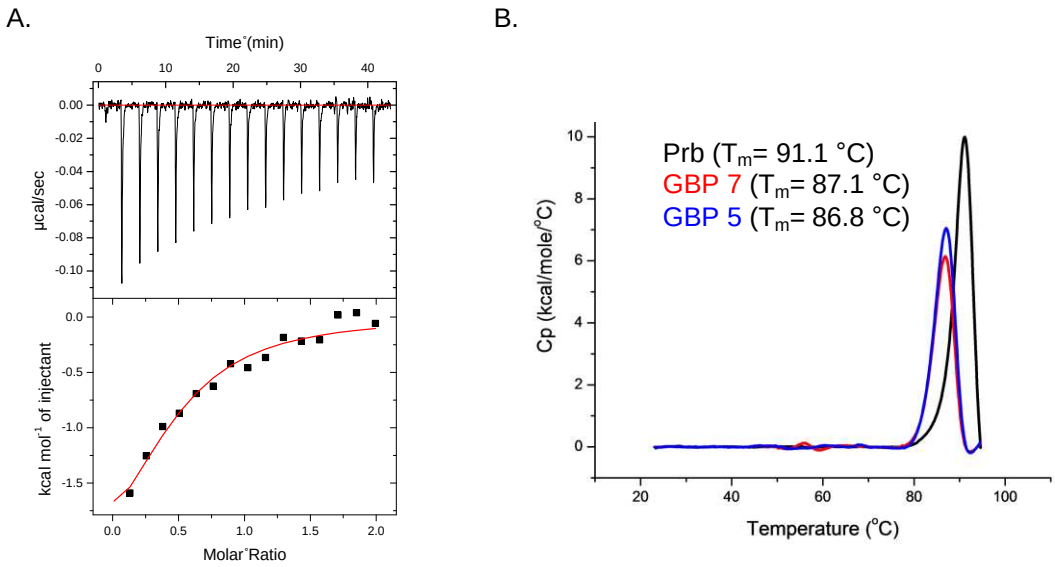


Figure 3.7 (*A*) Isothermal titration calorimetry (ITC) binding isotherm for gankyrin and GBP 7. (*B*) Differential scanning calorimetry (DSC) for ankyrin repeat-binding proteins Prb (black), GBP 7 (red), and GBP 5 (blue).

3.6 Conclusion

In conclusion, limitations to small molecule reagents and drug leads require fundamentally new approaches for the recognition of disease-relevant receptors. The size, electrostatic complexity, and relatively featureless surfaces associated with many protein-protein interactions involving disease-relevant ankyrin repeat domains present a particularly difficult challenge for small molecule reagents. Synthetic proteins offer a unique opportunity to recognize – and potentially modulate the activity of – challenging macromolecular targets such as ankyrin repeats. Here, we described novel synthetic proteins that selectively and tightly bind the oncoprotein ankyrin repeat gankyrin. Split-spGFP reassembly, ELISA, and cell lysate pull-down experiments suggest that these interactions occur in living cells and are highly selective. These new gankyrin-binding proteins are thermostable, express well in *E. coli* as soluble proteins, and represent the first synthetic proteins that recognize gankyrin *in vitro* and in complex cellular environments.

These proteins likely represent valuable starting points for further optimizing affinity to gankyrin, and modulating gankyrin-dependent oncogenic cell function and fate. Efforts toward this end are discussed in chapter five.

3.7 Methods

Library Preparation Plasmids containing Pdar and Prb were generously provided by Professor David Baker (University of Washington). The construct NspGFP-Prb was amplified using oligonucleotides (all primers purchased from Integrated DNA Technologies, IDT), digested with restriction enzymes NcoI and PacI, and ligated into pre-cut pETDuet vector to serve as the backbone for the library (all cloning enzymes purchased from New England Biolabs, NEB). NspGFP-Prb pETDuet was cut with restriction enzymes AatII and BglII for library insertion, resulting in 20 µg digested plasmid. The library insert was amplified using the saturation mutagenesis primers NNK (and complementary MNN), where N represents a 25% mix each of adenine, thymine, guanine, and cytosine nucleotides; and K represents a 50% mix each of thymine and guanine nucleotides, and M represents a 50% mix each of adenine and cytosine. NNK codons were used to remove 2 of 3 possible stop codons, while maintaining all 20 amino acids. The library insert was cut with AatII and BglII and ligated into pre-cut NspGFP-Prb pETDuet plasmid in its entirety, resulting in ~12 µg library plasmid DNA, which was subsequently purified via two rounds phenol/chloroform extraction and ethanol precipitation. The entire library was transformed in 48 batches by electroporation into 2.4 ml of electrocompetent 10 βs (NEB) using 1mm cuvettes at 1.7 keV. Cells were allowed to recover in 1L pre-warmed SOC for 1hr. Library size was calculated by serial dilution, resulting in a library of ~5.0×10^9. Carbenicillin was added to a working concentration of 100 µg/mL and culture was grown overnight at 37 ºC. Library plasmid DNA was

recovered the following day using a maxiprep kit (OMEGA). Gankyrin-CspGFP pBAD was prepared by amplifying gankyrin with oligonucleotides, digested with NcoI and AatII, and ligated into pre-cut pBAD containing the link-CspGFP construct.

Preparation of Electrocompetent *E. coli* The gankyrin-CspGFP pBAD construct was transformed into chemically competent BL21-Gold (DE3). Cells were made electrocompetent following standard procedures. Efficiencies of > $4.0x10^8$ cfu/µg DNA were achieved regularly using ~250 ng library DNA.

Fluorescence Activated Cell Sorting (FACS) 2 µg library DNA was transformed by electroporation into 400 µl BL21-Gold *E. coli* containting the gankyrin-CspGFP pBAD construct in 8 batches. Cells were rescued in 250 ml pre-warmed 2XYT and allowed to recover for 1hr at 37ºC. Kanamycin (50 µg/mL) and carbenicillin were added to the culture and grown until an OD_{600} of 0.5 was reached. Cells were brought to room temperature and induced with 0.2% arabinose and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Growth was continued for 6 hours at 30 °C, then cells were pelleted and washed with ice-cold PBS 3x and resuspended in PBS. Cells were sorted by FACS using a MoFlo Flow Cytometer and High Speed Cell Sorter with a solid state iCyt 488nm laser (CSU proteomics) at a rate of 9-11,000 events/second using single sort mode for 5 hours, setting the GFP-positive gate above the negative control (uninduced). ~800 GFP-positive cells were sorted into fresh 2XYT and allowed to recover for 12 hours at 37 °C before the addition of antibiotics, and growth was continued overnight. Cells were inoculated into fresh 2XYT containing antibiotics and induced (as described above) when OD_{600} reached 0.5. Growth was continued for 6 hours at 30 °C, then prepared for FACS as described above. Cells were sorted a

2nd time at a rate of ~5,000 events/second for 30 minutes, taking the top 10% of the GFP-positive population. Cells were rescued and grown as described above. Plasmid DNA was harvested the following day using a miniprep kit (OMEGA).

Identification of Library Members Plasmid DNA from the 2nd round sort was used as a template for a PCR. Prb mutants were amplified, digested with XhoI and PacI, and ligated back into a pre-cut pETDuet vector containing NspGFP. After transformation into chemically competent 5αs, individual colonies were picked, cultured, plasmid purified, and sequenced (GENEWIZ). Unique sequences were termed Gankyrin Binding Protein (GBP) 1-7 and separately transformed into BL21-Gold *E. coli* containing gankyrin-CspGFP pBAD via electroporation. Each unique GBP was verified for binding with split-spGFP reassembly.

Protein Purification Constructs were cloned into pETDuet using restriction enzymes BamHI and PacI, resulting in N-terminally His_{6x} tagged proteins and transformed into BL21s (DE3). Cells were grown in 2 L LB cultures containing carbenicillin at 37 °C to OD_{600} = 0.5 and induced with 1 mM IPTG at 25 °C overnight. Cells were then collected by centrifugation (5000 rpm, 10 minutes), resuspended in phosphate buffer (20 mM Sodium Phosphate pH 7.4, 150 mM NaCl, 2.5 mM 2-mercaptoethanol) and stored at -20 °C. Frozen pellets were thawed and sonicated for 2 minutes. The lysate was cleared by centrifugation (15000 rpm, 30 minutes) and the supernatant was mixed with 1 mL of Ni-NTA agarose resin for 1 hour at 4 °C. The resin was collected by centrifugation (5000 rpm, 5 minutes). The resin was washed with 50 mL of buffer and 20 mM imidazole. The protein was then eluted with 5 mL of buffer containing 400 mM imidazole. The

proteins were dialyzed against buffer and analyzed for purity by SDS-PAGE. Purified proteins were quantified using absorbance at 280 nm.

ELISA Separately, gankyrin, Pdar, and the Notch-1 ankyrin repeat domain were cloned into MCS2 of pETDuet with FLAG tags using restriction enzymes NdeI and PacI. Prb and GBPs were cloned into MCS1 of pETDuet using restriction enzymes BamHI and HindIII, resulting in N-terminal His_{6x}-tagged proteins. Completed constructs were transformed into BL21s (DE3). Cells containing the co-expressed pair were inoculated and induced as described previously. Cells were spun down (5000 rpm, 10 minutes) and resuspended in lysis buffer (100 mM potassium glutamate, 20 mM Hepes pH 7.5), lysed by sonication, and spun down (15000 rpm, 20 minutes) to remove cell debris. Cleared lysates were incubated on clear Ni-NTA coated plates for 1 hour at room temperature and washed 3x with 200 µL wash buffer (100 mM potassium glutamate, 20 mM Hepes pH 7.5, 0.05% Tween-20, 0.01 mg/mL BSA). Horseradish peroxidase (HRP)-conjugated mouse anti-DDDDK antibody in LiCor Blocking Buffer was incubated for 1 hour at room temperature (1:5000), followed by 4x 200 µL washes. Colorimetry was developed using TMB-One substrate and absorbance was measured at 655nm on a plate reader (BioTek).

Lysate Ni-NTA Pull-down Assay Ankyrin repeats (gankyrin and Pdar) were cloned into MCS 1 of pETDuet using restriction enzymes BamHI and HindII, resulting in N-terminal His_{6x}-tagged proteins. Prb and GBPs were cloned into MCS2 of pETDuet using the restriction enzymes NdeI and PacI. Completed constructs were transformed into BL21s (DE3). Cells containing the co-expressed pair were inoculated and induced as described previously. Cells were spun down (5000 rpm, 10 minutes) and resuspended in lysis buffer (100 mM potassium glutamate, 20 mM Hepes

pH 7.5), lysed by sonication, and spun down (15000 rpm, 20 minutes) to remove cell debris. Cleared lysate was incubated with 100 µL Ni-NTA agarose resin for 1 hour. Ni-NTA agarose was washed with 5 mL lysis buffer and 5 mL lysis buffer with 20 mM imidazole. Proteins were eluted with lysis buffer containing 400 mM imidazole. The pull-down was analyzed by SDS-PAGE.

Isothermal Titration Calorimetry (ITC) ITC experiments were performed in collaboration with GE Healthcare using a MicroCal iTC200 calorimeter maintained at 25 °C. All proteins were purified as described previously and dialyzed extensively in phosphate buffer (20 mM sodium phosphate pH 7.4, 150 mM NaCl, 2.5 mM 2-mercaptoethanol). Gankyrin was placed in the sample cell at a concentration of 20 µM, and 200 µM of GBP 5 or GBP 7 were titrated in 2 µL increments (16 injections total) at 160 sec intervals using a stirring speed of 750 rpm. Data were analyzed using Origin7.0 (MicroCal, iTC200) using a one set of sites binding model for fitting.

Differential Scanning Calorimetry (DSC) DSC experiments were performed in collaboration with GE Healthcare using a MicroCal VP-Capillary DSC system. All proteins were purified as described previously and dialyzed extensively in phosphate buffer (20 mM sodium phosphate pH 7.4, 150 mM NaCl, 2.5 mM 2-mercaptoethanol). DSC experiments were performed using 1 mg/mL of Prb, GBP 5, and GBP 7 in phosphate buffer containing 2-mercaptoethanol. Temperature scanning was performed from 20 to 95 °C at a scan rate of 60 °C/hr using passive feedback mode.

3.8 Proteins Used in This Work

Gankyrin

MEGCVSNIMICNLAYSGKLDELKERILADKSLATRTDQDSRTALHWACSAGHTEIVEFL
LQLGVPVNDKDDAGWSPLHIAASAGRDEIVKALLVKGAHVNAVNQNGCTPLHYAASK

NRHEIAVMLLEGGANPDAKDHYDATAMHRAAAKGNLKMVHILLFYKASTNIQDTEGN
TPLHLACDEERVEEAKFLVTQGASIYIENKEEKTPLQVAKGGLGLILKRLAEGEEASMG*

Prb

MGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLEHGYDVYPVNPNYE
EVLGRKCYPSVLDIPDKIEVVDLFVNPAKAWRFVVYAIKKGAKVVWFQYNTYYPLAGR
QAKEAGLIIVANRCMMREHERLLGEK*

GBP 3

MGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLEHGYDVYPVNPNYE
EVLGRKCYPSVLDIPDKIEVVDLFVLPNEAGEFVVYAIKKGAKVVWFQYNTYSNDAGR
QAKEAGLIIVANRCMMREHERLLGEK*

GBP 5

MGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLEHGYDVYPVNPNYE
EVLGRKCYPSVLDIPDKIEVVDLFVGPRSAKNFVVYAIKKGAKVVWFQYNTYNWLAGR
QAKEAGLIIVANRCMMREHERLLGEK*

GBP 7

MGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLEHGYDVYPVNPNYE
EVLGRKCYPSVLDIPDKIEVVDLFVYPIWATEFVVYAIKKGAKVVWFQYNTYHGEAGR
QAKEAGLIIVANRCMMREHERLLGEK*

Pdar

MASDLGKKLLEAAAAGQDDEVRILMANGADVNATDDDGLTPLHLAAANGQLEIVEVL
LKNGADVNASDSAGITPLHLAAYDGHLEIVEVLLKHGADVNAYDRAGWTPLHLAALSG
QLEIVEVLLKHGADVNAQDALGLTAFDISINQGQEDLAEILQG*

Notch-1 ARD

MDVNVRGPDGFTPLMIASCSGGGLETGNSEEEEDAPAVISDFIYQGASLHNQTDRTGET
ALHLAARYSRSDAAKRLLEASADANIQDNMGRTPLHAAVSADAQGVFQILIRNRATDL
DARMHDGTTPLILAARLAVEGMLEDLINSHADVNAVDDLGKSALHWAAAVNNVDAA
VVLLKNGANKDMQNNREETPLFLAAREGSYETAKVLLDHFANRDITDHMDRLPRDIAQ
ERMHHDIVRLLDLE*

3.9 Primers Used in This Work

All DNA primers are listed 5' to 3'

Prb Library FP:

CGGATAAAATTGAGGTCGTAGATCTGTTTGTG**NNK**CCG**NNKNNK**GCG**NNKNNK**

TTCGTTGTCTATGCCATCAAGAAAGGG

Prb Library RP:

TTCCTTGGCTTGACGTCCAGC**MNNMNNMNN**ATATGTGTTGTACTGAAA

CCATACCACTTTTGC

Gankyrin FP: CATGCCATGGAGGGGTGTGTGTCTAACATAATGATCTGTAACC

Gankyrin RP:

CCTTAATTAATTAGTGATGGTGGTGGTGATGACCCATAGAAGCCTCTTCACCTTCTG

CTA

Pdar FP: ATATACCATGGCAAGCGATCTGGGTAAAAAGCTGCTG

Pdar RP: CCTTAATTAATTACCCTTGCAGGATCTCTGCCAGATCTTCC

Notch-1 FP: CATGCCATGGACGTAAATGTCCGCGGG

Notch-1 RP: ATATAGACGTCCTCGAGGTCCAGCAGCCT

Prb FP (includes GBPs): CGCGGATCCGGGCAGCACCCGTCCGATT

Prb RP (includes GBPs): CCTTAATTAATTATTTTTCGCCCAGCAGGCGTTC

GBP 5 Alanine-Scanning Primers

R85A FP: GATCTGTTTGTGGGTCCGGCGAGTGCGAAGAATTTCG

R85A RP: CGAAATTCTTCGCACTCGCCGGACCCACAAACAGATC

S86A FP: GATCTGTTTGTGGGTCCGAGGGCTGCGAAGAATTTCG

S86A RP: CGAAATTCTTCGCAGCCCTCGGACCCACAAACAGATC

K88A FP: GTTTGTGGGTCCGAGGAGTGCGGCGAATTTCGTTGTCTATGC

K88A RP: GCATAGACAACGAAATTCGCCGCACTCCTCGGACCCACAAAC

N89A FP: GTTTGTGGGTCCGAGGAGTGCGAAGGCTTTCGTTGTCTATGC

N89A RP: GCATAGACAACGAAAGCCTTCGCACTCCTCGGACCCACAAAC

N110A FP: GGTATGGTTTCAGTACAACACATATGCTTGGTTGGCTGGACGTCAAGCC

N110A RP: GGCTTGACGTCCAGCCAACCAAGCATATGTGTTGTACTGAAACCATACC

W111A FP: GGTATGGTTTCAGTACAACACATATAATGCGTTGGCTGGACGTCAAGCC

W111A RP: GGCTTGACGTCCAGCCAACGCATTATATGTGTTGTACTGAAACCATACC

GBP 7 Alanine-Scanning Primers

Y83A FP: GAGGTCGTAGATCTGTTTGTGGCTCCGATTTGGGCGACGGAG

Y83A RP: CTCCGTCGCCCAAATCGGAGCCACAAACAGATCTACGACCTC

I85A FP: GATCTGTTTGTGTATCCGGCTTGGGCGACGGAGTTCG

I85A RP: CGAACTCCGTCGCCCAAGCCGGATACACAAACAGATC

W86A FP: GATCTGTTTGTGTATCCGATTGCGGCGACGGAGTTC

W86A RP: GAACTCCGTCGCCGCAATCGGATACACAAACAGATC

T88A FP: GTTTGTGTATCCGATTTGGGCGGCGGAGTTCGTTGTCTATGCC

T88A RP: GGCATAGACAACGAACTCCGCCGCCCAAATCGGATACACAAAC

E89A FP: GTATCCGATTTGGGCGACGGCGTTCGTTGTCTATGCC

E89A RP: GGCATAGACAACGAACGCCGTCGCCCAAATCGGATAC

H110A FP: GTTTCAGTACAACACATATGCTGGGGAGGCTGGACGTCAAG

H110A RP: CTTGACGTCCAGCCTCCCCAGCATATGTGTTGTACTGAAAC

E112A FP: GTACAACACATATCATGGGGCGGCTGGACGTCAAGCCAAGG

E112A RP: CCTTGGCTTGACGTCCAGCCGCCCCATGATATGTGTTGTAC

3.10 Supplemental Data

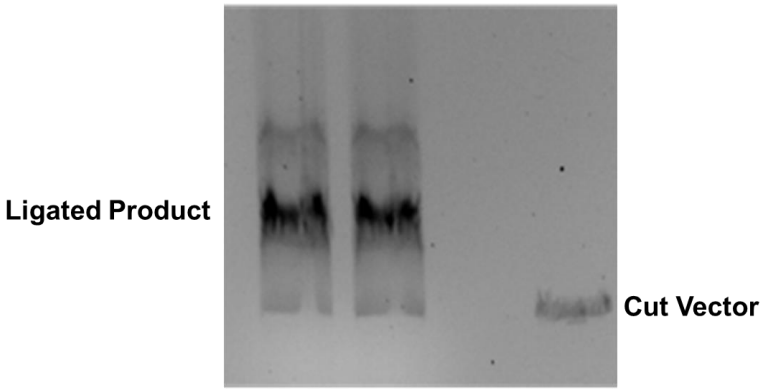


Figure S3.1 DNA agarose gel stained with ethidium bromide, showing a representative example of the Prb library ligation reaction after phenol/chloroform extraction (~ 90% ligation efficiency).

REFERENCES

(1) Li, J.; Mahajan, A.; Tsai, M.-D. *Biochemistry* **2006**, *45* (51), 15168.

(2) Abdel-Rahman, N.; Martinez-Arias, A.; Blundell, T. L. *Biochem. Soc. Trans.* **2011**, *39* (5), 1327.

(3) Kim, Y. H.; Kim, J. H.; Choi, Y. W.; Lim, S. K.; Yim, H.; Kang, S. Y.; Chung, Y. S.; Lee, G. Y.; Park, T. J. *Exp. Mol. Pathol.* **2013**, *94* (2), 360.

(4) Zhen, C.; Chen, L.; Zhao, Q.; Liang, B.; Gu, Y.-X.; Bai, Z.-F.; Wang, K.; Xu, X.; Han, Q.; Fang, D.; Wang, S.; Zhou, T.; Xia, Q.; Gong, W.-L.; Wang, N.; Li, H.-Y.; Jin, B.-F.; Man, J. *Oncogene* **2013**, *32* (29), 3452.

(5) Fu, X.-Y.; Wang, H.-Y.; Tan, L.; Liu, S.-Q.; Cao, H.-F.; Wu, M.-C. *World J. Gastroenterol.* **2002**, *8* (4), 638.

(6) Li, J.; Knobloch, T. J.; Kresty, L. a; Zhang, Z.; Lang, J. C.; Schuller, D. E.; Weghorst, C. M. *Anticancer Res.* **2011**, *31* (9), 2683.

(7) Meng, Y.; He, L.; Guo, X.; Tang, S.; Zhao, X.; Du, R.; Jin, J.; Bi, Q.; Li, H.; Nie, Y.; Liu, J.; Fan, D. *Cancer Lett.* **2010**, *297* (1), 9.

(8) Tang, S.; Yang, G.; Meng, Y.; Du, R.; Li, X.; Fan, R.; Zhao, L.; Bi, Q.; Jin, J.; Gao, L.; Zhang, L.; Li, H.; Fan, M.; Wang, Y.; Wu, K.; Liu, J.; Fan, D. *Cancer Biol. Ther.* **2010**, *9* (2), 88.

(9) Ortiz, C. M.; Ito, T.; Tanaka, E.; Tsunoda, S.; Nagayama, S.; Sakai, Y.; Higashitsuji, H.; Fujita, J.; Shimada, Y. *Int. J. Cancer* **2008**, *122* (2), 325.

(10) Man, J.; Liang, B.; Gu, Y.; Zhou, T.; Li, A.; Li, T.; Jin, B.; Bai, B. **2010**, *120* (8), 2829.

(11) Li, J.; Tsai, M.-D. *Biochemistry* **2002**, *41* (12), 3977.

(12) Higashitsuji, H.; Higashitsuji, H.; Itoh, K.; Sakurai, T.; Nagao, T.; Sumitomo, Y.; Sumitomo, H.; Masuda, T.; Dawson, S.; Shimada, Y.; Mayer, R. J.; Fujita, J. *Cancer Cell* **2005**, *8* (1), 75.

(13) Nakamura, Y.; Nakano, K.; Umehara, T.; Kimura, M.; Hayashizaki, Y.; Tanaka, A.; Horikoshi, M.; Padmanabhan, B.; Yokoyama, S. *Structure* **2007**, *15* (2), 179.

(14) Higashitsuji, H.; Itoh, K.; Nagao, T.; Dawson, S.; Nonoguchi, K.; Kido, T.; Mayer, R. J.; Arii, S.; Fujita, J. *Nat. Med.* **2000**, *6* (1), 96.

(15) Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (1), 13.

(16) Cronican, J. J.; Beier, K. T.; Davis, T. N.; Tseng, J.-C.; Li, W.; Thompson, D. B.; Shih, A. F.; May, E. M.; Cepko, C. L.; Kung, A. L.; Zhou, Q.; Liu, D. R. *Chem. Biol.* **2011**, *18* (7), 833.

(17) Cronican, J. J.; Thompson, D. B.; Beier, K. T.; McNaughton, B. R.; Cepko, C. L.; Liu, D. R. *ACS Chem. Biol.* **2010**, *5* (8), 747.

(18) DePorter, S. M.; Lui, I.; Bruce, V. J.; Gray, M. a; Lopez-Islas, M.; McNaughton, B. R. *Mol. Biosyst.* **2014**, *10* (1), 18.

(19) Deporter, S. M.; Lui, I.; Mohan, U.; McNaughton, B. R. *Chem. Biol.* **2013**, *20* (3), 434.

(20) Deshayes, S.; Morris, M. C.; Divita, G.; Heitz, F. *Cell. Mol. Life Sci.* **2005**, *62* (16), 1839.

(21) Fuchs, S. M.; Raines, R. T. *Protein Sci.* **2005**, *14* (6), 1538.

(22) Fuchs, S. M.; Raines, R. T. *ACS Chem. Biol.* **2007**, *2* (3), 167.

(23) Burton, D. R.; Weiss, R. A. *Science (80-.).* **2010**, *329* (5993), 770.

(24) Karanicolas, J.; Corn, J. E.; Chen, I.; Joachimiak, L. a; Dym, O.; Peck, S. H.; Albeck, S.; Unger, T.; Hu, W.; Liu, G.; Delbecq, S.; Montelione, G. T.; Spiegel, C. P.; Liu, D. R.; Baker, D. *Mol. Cell* **2011**, *42* (2), 250.

(25) Blakeley, B. D.; Chapman, A. M.; McNaughton, B. R. *Mol. Biosyst.* **2012**, *8* (8), 2036.

(26) Magliery, T. J.; Regan, L. *Methods Biochem. Anal.* **2006**, *47*, 391.

(27) Magliery, T. J.; Wilson, C. G. M.; Pan, W.; Mishler, D.; Ghosh, I.; Hamilton, A. D.; Regan, L. *J. Am. Chem. Soc.* **2005**, *127* (1), 146.

(28) Wilson, C. G. M.; Magliery, T. J.; Regan, L. *Nat. Methods* **2004**, *1* (3), 255.

(29) Lawrence, M. S.; Phillips, K. J.; Liu, D. R. *J. Am. Chem. Soc.* **2007**, *129* (33), 10110.

(30) Root, M. J. *Science (80-.).* **2001**, *291* (5505), 884.

CHAPTER FOUR

Characterization of the Binding Interaction Between the Oncoprotein Gankyrin and a Grafted S6 ATPase

Adapted from:
Chapman, A.M.; Rogers, B.E.; McNaughton, B.R., *Biochemistry*, **2014**, *53*, 6857.

In this work, I designed the FtsH-S6 ATPase grafting strategies, as well as carried out cloning and protein purification. I also carried out the design, execution, and analysis of isothermal titration calorimetry (ITC) experiments. Bryce Rogers, a first year graduate student, made contributions to the cloning, purification, and analysis of proteins used in this work, including running all circular dichroism (CD) experiments.

4.1 Introduction

As discussed in chapter three, overexpression of gankyrin is directly linked to the onset, proliferation, and/or metastasis of a large number of cancers.[1–7] After identifying GBP 7 (discussed in chapter three), which binds gankyrin well (K_D ~6 μM), we next sought to better understand the dictates of gankyrin-binding, as well as be able to potentially quantitatively compare our protein to known physiological gankyrin-binding partners. Gankyrin is reported to bind both cyclin-dependent kinase 4 (CDK4)[8] and MDM2[9], resulting in increased efficiency of pRb phosphorylation and p53 polyubiquitination and degradation, respectively. However, the structural basis for these interactions has not yet been reported, and neither the targets nor their putative gankyrin-binding domains express as soluble recombinant proteins in *E. coli* (**Figure 4.1**). A more

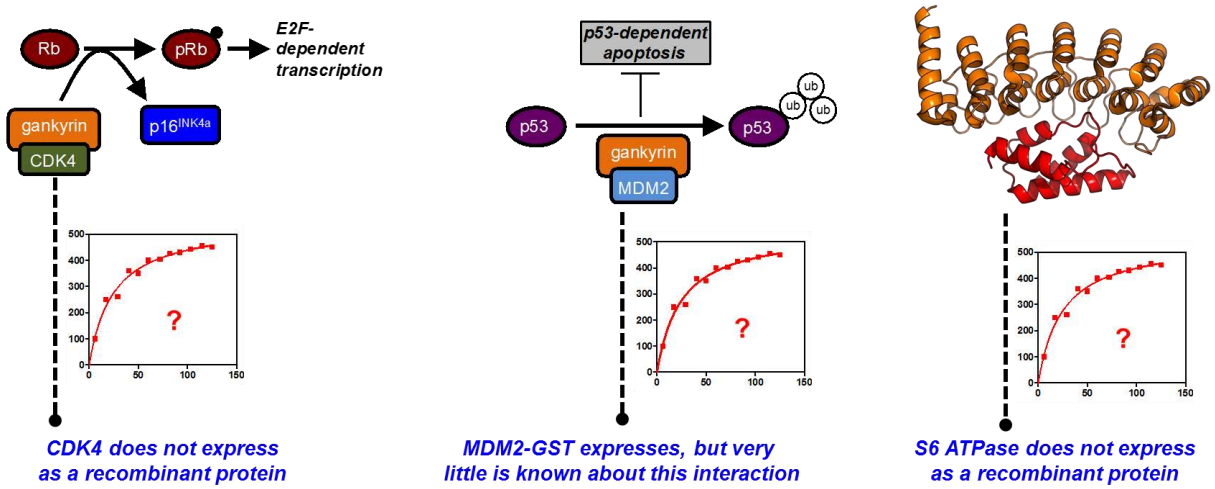


Figure 4.1 Characterizing interactions involving gankyrin is complicated by the poor biophysical properties of physiological binding partners and/or the complexity of binding interactions. Gankyrin-S6 ATPase PDB code: 2DVW.

promising venue for studying gankyrin–protein interactions is the co-crystal structure with a C-terminal portion of the S6 ATPase from the 26S proteasome, reported by Yokoyama and coworkers (**Figure 4.2A**).[10] Preliminary characterization of the recognition interface by Yokoyama and coworkers was achieved by a series of pull-down experiments with gankyrin-His_{6x} and S6 ATPase mutants (concomitantly expressed from a pETDuet plasmid), in which binding face residues believed to participate in complex stability were mutated to mostly alanine (R342A, R338A/R342A, R338A/R339A/R342A, E356A/E357A, D359A/D362A, and K397E in S6 ATPase and R41A, K116A, D39A/D71A, R41A/K116A, and E182A in gankyrin, highlighted in **Figure 4.2B**.

Efforts to directly probe the gankyrin–S6 ATPase complex are hampered by the tendency of the latter to form inclusion bodies when expressed in the absence of gankyrin. In our hands such material could not be refolded, and fusion to proteins commonly used to improve stability and solubility was likewise ineffective (**Figure 4.1**). An alternative strategy for display of folded and functional S6 ATPase is protein grafting. In this approach, a protein scaffold is identified that is

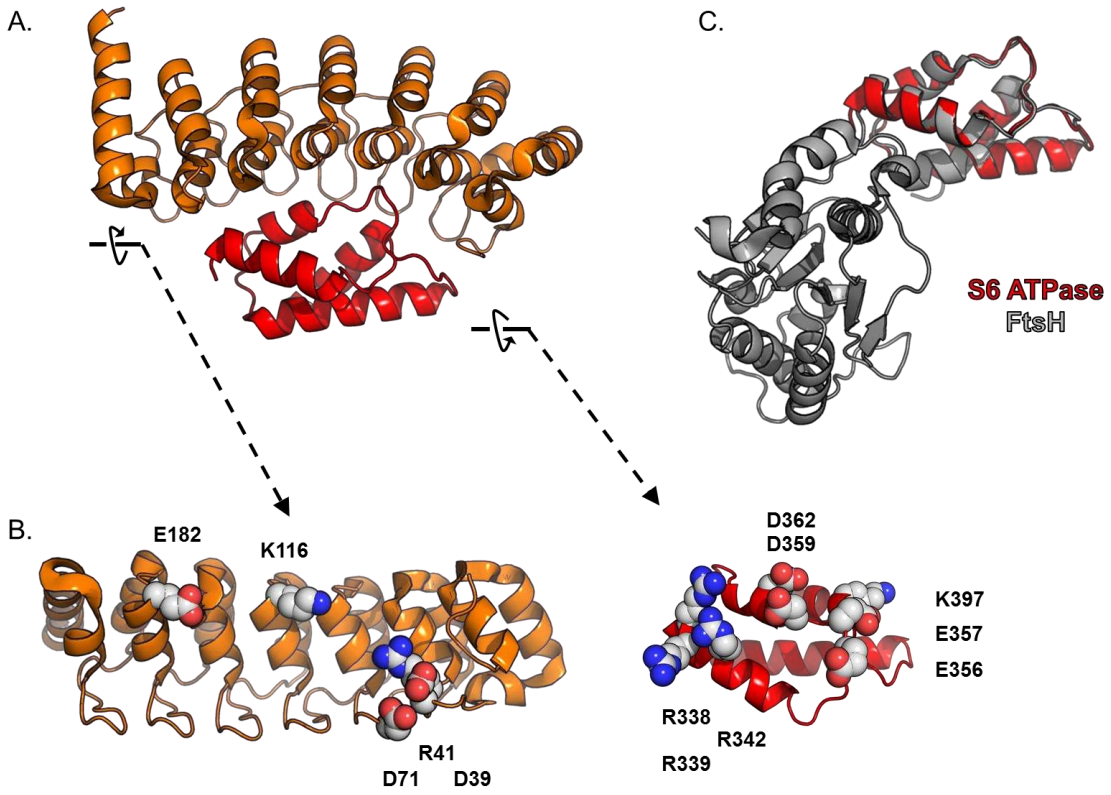


Figure 4.2 (*A*) Complex between gankyrin (orange) and S6 ATPase (red). The direction of the arrow next to each protein indicates the direction of a 90° rotation, which reveals the binding surfaces, as shown in *B*. (*B*) Gankyrin and S6 ATPase binding face residues critical to complex stabilization (and mutated in the work). (*C*) S6 ATPase (red) superimposed on the C-terminal domain of FtsH (grey). Gankyrin-S6 ATPase PDB code: 2DVW, FtsH PDB code ILV7.

stable, expresses well in *E. coli*, and contains a domain with excellent structural homology to S6 ATPase. If that protein is stable enough to tolerate replacement of the structurally homologous domain with S6 ATPase, it could serve as a generic platform for display of a folded and functional variant of this otherwise inaccessible protein.

4.2 Grafting FtsH-S6 ATPase Results in a Soluble, Folded Protein

Our initial efforts to identify such a scaffold relied on the recognition by Yokoyama and coworkers that, while the C-terminal portion of FtsH from *E. coli* has low sequence homology (~25%) with S6 ATPase, the two proteins have similar tertiary structures (rmsd of ~1.4 Å over 74

main chain residues, **Figure 4.2C**). Expanding on this finding, we set out to determine if a grafted protein–in which the C-terminal ATPase domain of FtsH is replaced with S6 ATPase–expresses as a soluble protein in *E. coli* that mimics the native S6 ATPase/gankyrin interaction. To design the grafted protein, S6 ATPase was genetically fused to FtsH, starting after FtsH position 326 and beginning with S6 ATPase position 338. The resulting protein is referred to as FtsH-S6 ATPase throughout.

To our satisfaction, grafted FtsH-S6 ATPase was expressed in *E. coli* as soluble protein, compared to insoluble S6 ATPase (Supplemental Data**, Figure S4.1**). Additionally, FtsH-S6 ATPase expressed and purified as a His_{6x}-tagged protein in a comparable yield and purity to the wild-type FtsH (wt-FtsH, Supplemental Data, **Figure S4.2**). Circular dichroism (CD) spectra of the two proteins show they are virtually identical (**Figure 4.3A**), suggesting no appreciable structural change to the FtsH scaffold or grafted S6 ATPase domain.

4.3 Alanine-Scanning Mutagenesis Reveals FtsH-S6 ATPase Faithfully Mimics S6 ATPase Binding

Affinity of this grafted protein to gankyrin was first assessed using a pull-down assay in *E. coli*. Binding face residues on gankyrin or FtsH-S6 ATPase were mutated to alanine (based on the findings of Yokoyama and coworkers using S6 ATPase), and their effect on complex stability was qualitatively assessed by measuring the amount of untagged FtsH-S6 ATPase co-purified with gankyrin-His_{6x}. Most notably, FtsH-S6 ATPase R338A/R339A/R342A (**Figure 4.3B**, lane 4), FtsH-S6 ATPase D359A/D362A (**Figure 4.3B**, lane 6), gankyrin R41A (**Figure 4.3C**, lane 2), gankyrin R41A/K116A (**Figure 4.3C**, lane 4), and gankyrin D39A/D71A (**Figure 4.3C**, lane 5) appear to form significantly less stable complexes than the native proteins. This is in contrast to

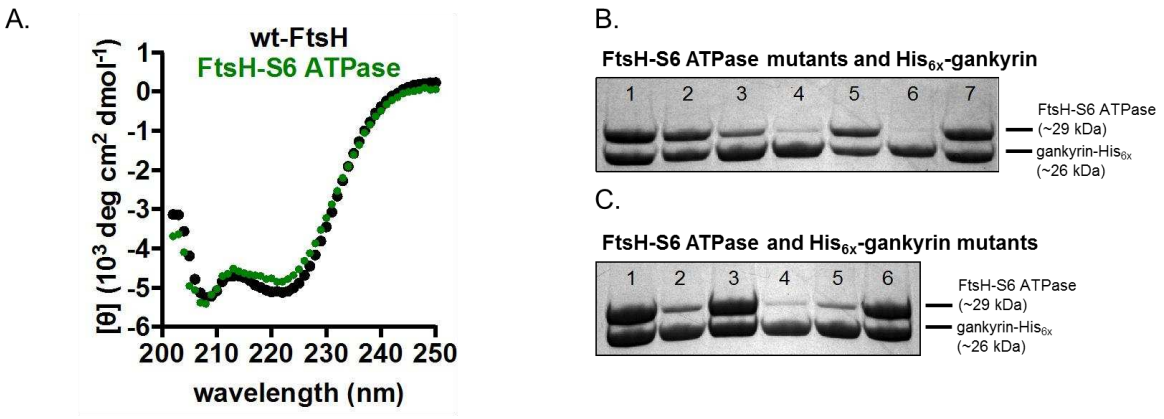


Figure 4.3 (*A*) Circular dichroism spectra of wild-type FtsH (wt-FtsH, black) and FtsH-S6 ATPase (green). (*B*) Co-purification of wild-type gankryin-His_{6x} and FtsH-S6 ATPase mutants: wt-FtsH-S6 ATPase (lane 1), R342A (lane 2), R338A/R342A (lane 3), R338A/R339A/R342A (lane 4), E356A/E357A (lane 5), D359A/D362A (lane 6), and K397E (lane 7). (*C*) Co-purification of gankryin-His_{6x} mutants and wt-FtsH-S6 ATPase mutants: wt-gankyrin (lane 1), R41A (lane 2), K116A (lane 3), R41A/K116A (lane 4), D39A/D71A (lane 5), and E182A (lane 6).

Yokoyama's original pull-down, in which all mutants but R342A S6 ATPase did not appreciably co-purify with gankyrin-His_{6x}. This highlights a potential virtue of our grafting approach. It is unclear if mutations to this unstable form of S6 ATPase appreciably modulate, or abolish, gankyrin/S6 ATPase complex stability, or simply further decrease structure and stability of the C-terminal S6 ATPase fragment.

4.4 Characterizing the FtsH-S6 ATPase-Gankyrin Interaction by ITC

While the FtsH scaffold displays S6 ATPase in a manner that faithfully mimics the native protein (facilitates binding to gankyrin), no information on the exact differences in binding energies can be obtained using the pull-down assay. Moreover, mutational effects that do not dramatically lower, or completely abolish, complex stability cannot be probed using this assay. Only through the described grafting strategy are able to create a soluble and stable mimic of S6 ATPase, which permits the use of more sensitive biophysical methods to probe this important binding interaction.

We used Isothermal Titration Calorimetry (ITC) to obtain the full thermodynamic signature (ΔH, –TΔS, ΔG) and stoichiometry (*N*-value) of this interaction, as well as characterize mutational effects on complex stability. Perhaps not surprising, we found that gankyrin binds the grafted FtsH-S6 ATPase with a dissociation constant (K_D) of ~ 67 nM (**Table 4.1**, entry 1). The observed change in enthalpy (ΔH) and entropy (-TΔS) for this binding interaction were –28.7 kcal/mol and 19.0 kcal/mol, respectively. Gankyrin does not bind wt-FtsH with any appreciable affinity (Supplemental Data, **Figure S4.3**), which is unsurprising, given that the S6 and FtsH ATPase subdomains share only ~25% sequence homology. These results confirmed our prediction that gankyrin and S6 ATPase form a high affinity complex, due to the fact that S6 ATPase makes contact with all seven repeating ankyrin units of gankyrin.

Alanine mutation of S6 ATPase R342, which engages gankyrin through a salt bridge with gankyrin E182, modestly lowers complex stability (K_D = 216.6±25.8 nM, **Table 4.1**, entry 2). Double (R338A/R342A) and triple (R338A/R339A/R342A) mutation of a positively charged patch on the S6 ATPase face, which disrupts a salt bridge between S6ATPase R342 and gankyrin E182, dramatically lower complex stability (K_D = 2.5±0.4 μM and 7.5±0.2 μM, **Table 4.1**, entries 3 and 4, respectively). Interestingly, both of these mutations result in favorable binding entropy (-TΔS = -1.5±0.8 and -4.7±0.1 kcal/mol, respectively, compared to -TΔS = 19.0±0.6 for the native interaction). While the molecular mechanism for this dramatic change is unclear, one possible rationale is a lower energy of desolvation for the alanine mutants, compared to the native protein. While Yokoyama's original pull-down data suggest a significant role for S6 ATPase E356/E357 in complex stability, double alanine mutation did not appreciably lower binding affinity (**Table 4.1**, entry 5). Conversely, removal of a negatively charged patch on the S6 ATPase binding face (D359A/362A, **Table 4.1**, entry 6) completely abolished binding. An E182A mutation in

Table 4.1 Analysis of Binding Interactions between Gankyrin and FtsH-S6 ATPase by ITC[a]

Entry	gankyrin	FtsH-S6 ATPase	K_D (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	−TΔS (kcal/mol)
1	wild-type	wild-type	67.3±5.7	-9.8±0.1	-28.7±0.5	19.0±0.6
2	wild-type	R342A	216.6±25.8	-9.1±0.1	-22.0±0.8	12.9±0.8
3	wild-type	R338A/R342A	2549±353	-7.6±0.1	-6.1±0.7	-1.5±0.8
4	wild-type	R338A/R339A/R342A	7471±301	-7.0±0.1	-2.2±0.1	-4.7±0.1
5	wild-type	E356A/E357A	71.8±5.9	-9.8±0.1	-27.3±0.9	17.5±0.8
6	wild-type	D359A/D362A	no binding	---	---	---
7	wild-type	K397E	95.2±12.2	-9.7±0.2	-25.6±2.5	15.9±2.7
8	R41A	wild-type	313.3±17.6	-8.1±1.2	-17.1±1.7	9.0±2.8
9	K116A	wild-type	71.3±15.5	-9.7±0.2	-24.0±0.9	14.3±1.1
10	D39A/D71A	wild-type	93.0±5.6	-9.7±0.2	-25.0±0.7	15.4±0.8
11	R41A/K116A	wild-type	3633±404	-7.4±0.1	-4.9±0.6	-2.5±0.7
12	E182A	wild-type	140.6±9.7	-9.4±0.1	-28.2±2.1	18.8±2.0

[a]All errors represent the standard deviation of three separate experiments. ITC conditions were as follows: 20 mM sodium phosphate, 150 mM NaCl, and 2.5 mM 2-mercaptoethanol (pH 7.4) at 25 °C.

gankyrin—which further probes a salt brigde with S6 ATPase 342—was found to modestly lower bindng affinity (K_D = 140.6±9.7 nM, **Table 4.1**, entry 12), further suggesting a relatively minor role of this interaction in complex stability. Residue K397 in S6 ATPase makes a salt bridge with gankyrin D39/D71. However, a mutant that reverses the ionic nature of this residue (K397E) binds gankyrin with a similar affinity to the native protein (K_D = 95.2±12.2 nM, **Table 4.1**, entry 7), suggesting a relatively minor role of this particular salt bridge in complex stability. While gankyrin mutations K116A and D39A/D71A mutations had minimal effects on binding affinity (**Table 4.1**, entries 9 and 10, respectively), an R41A mutation significantly decreased affinity (K_D = 313.3±17.6 nM, **Table 4.1**, entry 8). While the single K116A mutation had a minimal effect on binding, an R41A/K116A double mutation—which is designed to test the role of a larger hydrogen bond / salt-bridge network—dramatically lowers affinity (K_D = 3.6±0.4 μM, **Table 4.1**, entry 11)**.** The R41A/K116A mutant, however, binds gankrin with a favorable binding entropy (-TΔS = -2.5±0.7 kcal/mol), possibly due to a lower energy of desolvation for the alanine mutants, compared

to the native protein. The binding stoichiometry (N) for each interaction was found to be ~1 (N = 0.91-1.02, Supplemental Data, **Table S4.1**).

4.5 Conclusion

Collectively, our findings represent the first quantitative assessment of the binding interaction—and binding thermodynamics—of a physiologically-relevant complex involving the oncoprotein gankyrin. These data also potentially establish a target affinity for therapeutic reagents designed to inhibit gankyrin-dependent protein-protein interactions[11], including any gankyrin-binding protein variants[12], as well as a tool for assessing competitive binding. FtsH-S6 ATPase's value as a comparative tool is discussed further in chapter five.

4.6 Methods

Protein Purification Grafted FtsH-S6 ATPase was overlapped and amplified by PCR using oligonucleotides (all primers purchased from Integrated DNA Technologies, IDT), and cloned into a pET plasmid using restriction enzymes BamHI and PacI, resulting in a N-terminally His_{6x} tagged construct (all cloning enzymes purchased from New England Biolabs, NEB), which was confirmed by DNA sequencing (all constructs in this manuscript were confirmed by GENEWIZ, South Plainfield, NJ). FtsH-S6 ATPase mutants were made using site-directed mutagenesis. These constructs were transformed into BL21s (DE3). Gankyrin was cloned into a pET plasmid using restriction enzymes NcoI and PacI, resulting in a C-terminally His_{6x} tagged construct, and transformed into BL21s (DE3). Gankyrin mutants were made using site-directed mutagenesis, and transformed into BL21s (DE3). Cells were grown in 1-2.5 L LB cultures containing 100 μg/mL carbenicillin at 37 °C to OD_{600} = ~0.6 and induced with 1 mM isopropyl β-D-1-

thiogalactopyranoside (IPTG) at 25 °C for 8 hours. Cells were then collected by centrifugation, resuspended in phosphate buffer (20 mM sodium phosphate pH 7.4, 150 mM NaCl) with protease inhibitor tablets (Roche) and stored at -20 °C. Frozen pellets were thawed and sonicated (Branson) for 2 minutes. The lysate was cleared by centrifugation (15000 rpm, 30 minutes) and the supernatant was mixed with 1 mL of Ni-NTA agarose resin for 1 hour. The resin was collected by centrifugation (5000 rpm, 5 minutes). The resin was washed with 50 mL of buffer containing 20 mM imidazole, followed by 10 mL with 50 mM imidazole. The protein was then eluted with 5 mL buffer containing 400 mM imidazole. The proteins were dialyzed against phosphate buffer and analyzed for purity by SDS-PAGE and coomassie staining. Purified proteins were quantified using absorbance at 280nm and confirmed with a modified Lowry Assay.

Lysate Ni-NTA Pulldown Assay Gankyrin variants were cloned into MCS1 of pETDuet using restriction enzymes BamHI and HindIII, resulting in N-terminal His_{6x}-tagged constructs. FtsH-S6 ATPase constructs were cloned into MCS2 of pETDuet using the restriction enzymes NdeI and PacI. Completed constructs were transformed into BL21s (DE3). Cells containing the co-expressed pair were inoculated and induced as described previously. Cells were spun down and resuspended in lysis buffer (20 mM sodium phosphate pH 7.4, 150 mM NaCl, 1 mM DTT) lysed by sonication, and spun down to remove cell debris. Cleared lysate was incubated with 100 μL Ni-NTA agarose resin for 1 hour at 4 °C. Ni-NTA agarose was washed with 5mL lysis buffer and 5mL lysis buffer with 20 mM imidazole. Proteins were eluted with lysis buffer containing 400 mM imidazole. The pulldown was analyzed by SDS-PAGE and coomassie staining.

Isothermal Titration Calorimetry Isothermal titration calorimetry was performed using a MicroCal iTC200 calorimeter maintained at 25 °C. All proteins were purified as described previously and dialyzed extensively in phosphate buffer (20 mM sodium phosphate pH 7.4, 150 mM NaCl, 2.5 mM 2-mercaptoethanol). FtsH-S6 ATPase variants were placed in the sample cell at concentrations ranging from 15-18 μM (30 μM for FtsH-S6 ATPase R338A/R339A/R342A), and 150-180 μM (500 μM for FtsH-S6 ATPase R338A/R339A/R342A interaction) of gankyrin variants were titrated in 2.49 μL increments (16 injections total), with an initial injection of 0.2 μL, at 180 second intervals using a stirring speed of 750 rpm. Heats of dilution were measured in the same manner described above, separately titrating buffer into buffer and gankyrin into buffer. Data were analyzed using Origin7.0 (MicroCal, iTC200) using a one set of sites binding model for fitting. All data were reference subtracted by subtracting the mean heat of dilution from each data point. All data were performed in triplicate

Circular Dichroism Proteins were purified as described above. Separately, each protein was diluted to 5 μM in phosphate buffer (20 mM sodium phosphate pH 7.4, 150 mM NaCl, 2.5 mM 2-mercaptoethanol) and placed in a quartz cuvette with a pathlength of 0.2 cm. Data were collected on an Aviv model 202 circular dichroism spectrometer. Wavelength data were taken from scans of 250 nm to 200 nm in 1 nm steps at 25 °C.

4.7 Proteins Used in This Work

Gankyrin

MEGCVSNIMICNLAYSGKLDELKERILADKSLATRTDQDSRTALHWACSAGHTEIVEFL
LQLGVPVNDKDDAGWSPLHIAASAGRDEIVKALLVKGAHVNAVNQNGCTPLHYAASK
NRHEIAVMLLEGGANPDAKDHYDATAMHRAAAKGNLKMVHILLFYKASTNIQDTEGN

TPLHLACDEERVEEAKFLVTQGASIYIENKEEKTPLQVAKGGLGLILKRLAEGEEASMGH
HHHHH*

Gankyrin R41A

MEGCVSNIMICNLAYSGKLDELKERILADKSLATRTDQDS**A**TALHWACSAGHTEIVEFL
LQLGVPVNDKDDAGWSPLHIAASAGRDEIVKALLVKGAHVNAVNQNGCTPLHYAASK
NRHEIAVMLLEGGANPDAKDHYDATAMHRAAAKGNLKMVHILLFYKASTNIQDTEGN
TPLHLACDEERVEEAKFLVTQGASIYIENKEEKTPLQVAKGGLGLILKRLAEGEEASMGH
HHHHH*

Gankyrin K116A

MEGCVSNIMICNLAYSGKLDELKERILADKSLATRTDQDSRTALHWACSAGHTEIVEFL
LQLGVPVNDKDDAGWSPLHIAASAGRDEIVKALLVKGAHVNAVNQNGCTPLHYAAS**A**
NRHEIAVMLLEGGANPDAKDHYDATAMHRAAAKGNLKMVHILLFYKASTNIQDTEGN
TPLHLACDEERVEEAKFLVTQGASIYIENKEEKTPLQVAKGGLGLILKRLAEGEEASMGH
HHHHH*

Gankyrin R41A/K116A

MEGCVSNIMICNLAYSGKLDELKERILADKSLATRTDQDS**A**TALHWACSAGHTEIVEFL
LQLGVPVNDKDDAGWSPLHIAASAGRDEIVKALLVKGAHVNAVNQNGCTPLHYAAS**A**
NRHEIAVMLLEGGANPDAKDHYDATAMHRAAAKGNLKMVHILLFYKASTNIQDTEGN
TPLHLACDEERVEEAKFLVTQGASIYIENKEEKTPLQVAKGGLGLILKRLAEGEEASMGH
HHHHH*

Gankyrin D39A/D71A

MEGCVSNIMICNLAYSGKLDELKERILADKSLATRTDQ**A**SRTALHWACSAGHTEIVEFL
LQLGVPVNDKD**A**AGWSPLHIAASAGRDEIVKALLVKGAHVNAVNQNGCTPLHYAASK
NRHEIAVMLLEGGANPDAKDHYDATAMHRAAAKGNLKMVHILLFYKASTNIQDTEGN
TPLHLACDEERVEEAKFLVTQGASIYIENKEEKTPLQVAKGGLGLILKRLAEGEEASMGH
HHHHH*

Gankyrin E182A

MEGCVSNIMICNLAYSGKLDELKERILADKSLATRTDQDSRTALHWACSAGHTEIVEFL
LQLGVPVNDKDDAGWSPLHIAASAGRDEIVKALLVKGAHVNAVNQNGCTPLHYAASK
NRHEIAVMLLEGGANPDAKDHYDATAMHRAAAKGNLKMVHILLFYKASTNIQDTEGN
TPLHLACD**A**ERVEEAKFLVTQGASIYIENKEEKTPLQVAKGGLGLILKRLAEGEEASMG
HHHHHH*

wt FtsH

MGSSHHHHHHSQDPLTEDQIKTTFADVAGCDEAKEEVAELVEYLREPSRFQKLGGKIPK
GVLMVGPPGTGKTLLAKAIAGEAKVPFFTISGSDFVEMFVGVGASRVRDMFEQAKKAA
PCIIFIDEIDAVGRQRGAGLGGGHDEREQTLNQMLVEMDGFEGNEGIIVIAATNRPDVLD
PALLRPGRFDRQVVVGLPDVRGREQILKVHMRRVPLAPDIDAAIIARGTPGFSGADLAN
LVNEAALFAARGNKRVVSMVEFEKAKDKIMMGA*

FtsH-S6 ATPase

MGSSHHHHHHSQDPLTEDQIKTTFADVAGCDEAKEEVAELVEYLREPSRFQKLGGKIPK
GVLMVGPPGTGKTLLAKAIAGEAKVPFFTISGSDFVEMFVGVGASRVRDMFEQAKKAA
PCIIFIDEIDAVGRQRGAGLGGGHDEREQTLNQMLVEMDGFEGNEGIIVIAATNRPDVLD
PALLRPGRFDRQVVVGLPDRRGKRQIFSTHTSKMNLSEEVDLEDYVARPDKISGADINSI
CQESGMLAVRENRYIVLAKDFEKAYKTVIKKDEQEHEFYK*

FtsH-S6 ATPase R342A

MGSSHHHHHHSQDPLTEDQIKTTFADVAGCDEAKEEVAELVEYLREPSRFQKLGGKIPK
GVLMVGPPGTGKTLLAKAIAGEAKVPFFTISGSDFVEMFVGVGASRVRDMFEQAKKAA
PCIIFIDEIDAVGRQRGAGLGGGHDEREQTLNQMLVEMDGFEGNEGIIVIAATNRPDVLD
PALLRPGRFDRQVVVGLPDRRGK**A**QIFSTHTSKMNLSEEVDLEDYVARPDKISGADINSI
CQESGMLAVRENRYIVLAKDFEKAYKTVIKKDEQEHEFYK*

FtsH-S6 ATPase R338A/R342A

MGSSHHHHHHSQDPLTEDQIKTTFADVAGCDEAKEEVAELVEYLREPSRFQKLGGKIPK
GVLMVGPPGTGKTLLAKAIAGEAKVPFFTISGSDFVEMFVGVGASRVRDMFEQAKKAA
PCIIFIDEIDAVGRQRGAGLGGGHDEREQTLNQMLVEMDGFEGNEGIIVIAATNRPDVLD
PALLRPGRFDRQVVVGLPD**A**RGK**A**QIFSTHTSKMNLSEEVDLEDYVARPDKISGADINSI
CQESGMLAVRENRYIVLAKDFEKAYKTVIKKDEQEHEFYK*

FtsH-S6 ATPase R338A/R339A/R342A

MGSSHHHHHHSQDPLTEDQIKTTFADVAGCDEAKEEVAELVEYLREPSRFQKLGGKIPK
GVLMVGPPGTGKTLLAKAIAGEAKVPFFTISGSDFVEMFVGVGASRVRDMFEQAKKAA
PCIIFIDEIDAVGRQRGAGLGGGHDEREQTLNQMLVEMDGFEGNEGIIVIAATNRPDVLD
PALLRPGRFDRQVVVGLPD**AA**GK**A**QIFSTHTSKMNLSEEVDLEDYVARPDKISGADINSI
CQESGMLAVRENRYIVLAKDFEKAYKTVIKKDEQEHEFYK*

FtsH-S6 ATPase E356A/E357A

MGSSHHHHHHSQDPLTEDQIKTTFADVAGCDEAKEEVAELVEYLREPSRFQKLGGKIPK
GVLMVGPPGTGKTLLAKAIAGEAKVPFFTISGSDFVEMFVGVGASRVRDMFEQAKKAA
PCIIFIDEIDAVGRQRGAGLGGGHDEREQTLNQMLVEMDGFEGNEGIIVIAATNRPDVLD
PALLRPGRFDRQVVVGLPDRRGKRQIFSTHTSKMNLS**AA**VDLEDYVARPDKISGADINSI
CQESGMLAVRENRYIVLAKDFEKAYKTVIKKDEQEHEFYK*

FtsH-S6 ATPase D359A/D362A

MGSSHHHHHHSQDPLTEDQIKTTFADVAGCDEAKEEVAELVEYLREPSRFQKLGGKIPK
GVLMVGPPGTGKTLLAKAIAGEAKVPFFTISGSDFVEMFVGVGASRVRDMFEQAKKAA
PCIIFIDEIDAVGRQRGAGLGGGHDEREQTLNQMLVEMDGFEGNEGIIVIAATNRPDVLD
PALLRPGRFDRQVVVGLPDRRGKRQIFSTHTSKMNLSEEV**A**LE**A**YVARPDKISGADINSI
CQESGMLAVRENRYIVLAKDFEKAYKTVIKKDEQEHEFYK*

FtsH-S6 ATPase K397E

MGSSHHHHHHSQDPLTEDQIKTTFADVAGCDEAKEEVAELVEYLREPSRFQKLGGKIPK
GVLMVGPPGTGKTLLAKAIAGEAKVPFFTISGSDFVEMFVGVGASRVRDMFEQAKKAA

PCIIFIDEIDAVGRQRGAGLGGGHDEREQTLNQMLVEMDGFEGNEGIIVIAATNRPDVLD
PALLRPGRFDRQVVVGLPDRRGKRQIFSTHTSKMNLSEEVDLEDYVARPDKISGADINSI
CQESGMLAVRENRYIVL**A**EDFEKAYKTVIKKDEQEHEFYK*

4.8 Primers Used in This Work

All DNA primers are listed 5' to 3'

FtsH-S6 ATPase overlap primers

FP1: CATGCCATGGGCAGCAGCCATCACCATCATCACCACAGCC

RP1: GTCTTCGGTCAGCGGATCCTGGCTGTGGTGATGATGGTGA

FP2: AGGATCCGCTGACCGAAGACCAGATCAAAACCACNTTCGC

RP2: CGTCGCAACCAGCAACGTCAGCGAANGTGGTTTTGATCTG

FP3: TGACGTTGCTGGTTGCGACGAAGCTAAAGAAGAAGTTGCT

RP3: CGCAGGTATTCAACCAGTTCAGCAACTTCTTCTTTAGCTT

FP4: GAACTGGTTGAATACCTGCGTGAACCGTCTCGTTTCCAGA

RP4: CGGGATTTTACCACCCAGTTTCTGGAAACGAGACGGTTCA

FP5: AACTGGGTGGTAAAATCCCGAAAGGTGTTCTGATGGTTGG

RP5: TTTTACCGGTACCCGGCGGACCAACCATCAGAACACCTTT

FP6: TCCGCCGGGTACCGGTAAAACCCTGCTGGCTAAAGCTATC

RP6: GGAACTTTAGCTTCACCAGCGATAGCTTTAGCCAGCAGGG

FP7: GCTGGTGAAGCTAAAGTTCCGTTCTTCACCATCTCTGGTT

RP7: GAACATTTCAACGAAGTCAGAACCAGAGATGGTGAAGAAC

FP8: CTGACTTCGTTGAAATGTTCGTTGGTGTTGGTGCTTCTCG

RP8: GTTCGAACATGTCACGAACACGAGAAGCACCAACACCAAC

FP9: TGTTCGTGACATGTTCGAACAGGCTAAAAAAGCTGCTCCG

RP9: TCGTCGATGAAGATGATGCACGGAGCAGCTTTTTTAGCCT

FP10: TGCATCATCTTCATCGACGAAATCGACGCTGTTGGTCGTC

RP10: ACCCAGACCAGCACCACGCTGACGACCAACAGCGTCGATT

FP11: AGCGTGGTGCTGGTCTGGGTGGTGGTCACGACGAACGTGA

RP11: GCATCTGGTTCAGGGTCTGTTCACGTTCGTCGTGACCACC

FP12: ACAGACCCTGAACCAGATGCTGGTTGAAATGGACGGTTTC

RP12: ATGATACCTTCGTTACCTTCGAAACCGTCCATTTCAACCA

FP13: GAAGGTAACGAAGGTATCATCGTTATCGCTGCTACCAACC

RP13: CGGGTCGAGAACGTCCGGACGGTTGGTAGCAGCGATAACG

FP14: GTCCGGACGTTCTCGACCCGGCTCTGCTGCGTCCGGGTCG

RP14: CAACAACCTGACGGTCGAAACGACCCGGACGCAGCAGAGC

FP15: TTTCGACCGTCAGGTTGTTGTTGGTCTGCCGGACCGCCGC

RP15: GAGAAAATCTGTCTCTTCCCGCGGCGGTCCGGCAGACCAA

FP16: GGGAAGAGACAGATTTTCTCCACTCACACTAGCAAGATGA

RP16: GTCAACCTCCTCAGAGAGGTTCATCTTGCTAGTGTGAGTG

FP17: ACCTCTCTGAGGAGGTTGACTTGGAAGACTATGTGGCCCG

RP17: CTCCTGAAATCTTATCTGGCCGGGCCACATAGTCTTCCAA

FP18: GCCAGATAAGATTTCAGGAGCTGATATTAACTCCATCTGT

RP18: GCCAACATTCCACTCTCCTGACAGATGGAGTTAATATCAG

FP19: CAGGAGAGTGGAATGTTGGCTGTCCGTGAAAACCGCTACA

RP19: GAAGTCCTTGGCCAGGACAATGTAGCGGTTTTCACGGACA

FP20: TTGTCCTGGCCAAGGACTTCGAGAAAGCATACAAGACTGT

RP20: CCTGCTCGTCCTTCTTGATGACAGTCTTGTATGCTTTCTC

FP21: CATCAAGAAGGACGAGCAGGAGCATGAGTTTTACAAGTGA

RP21: CCTTAATTAATCACTTGTAAAACTCATGCT

FtsH-S6 ATPase mutant primers

R342A FP: CCGGACCGCCGCGGGAAGGCACAGATTTTCTCCACTCAC

R342A RP: GTGAGTGGAGAAAATCTGTGCCTTCCCGCGGCGGTCCGG

R338A/R342A FP:

GTTGTTGGTCTGCCGGACGCCCGCGGGAAGGCACAGATTTTCTCCACTCAC

R338A/R342A RP:

GTGAGTGGAGAAAATCTGTGCCTTCCCGCGGGCGTCCGGCAGACCAACAAC

R338A/R339A/R342A FP:

GTTGTTGGTCTGCCGGACGCCGCCGGGAAGGCACAGATTTTCTCCACTCAC

R338A/R339A/R342A RP:

GTGAGTGGAGAAAATCTGTGCCTTCCCGGCGGCGTCCGGCAGACCAACAAC

E356A/E357A FP: GCAAGATGAACCTCTCTGCGGCGGTTGACTTGGAAGACTATG

E356A/E357A RP: CATAGTCTTCCAAGTCAACCGCCGCAGAGAGGTTCATCTTGC

D359A/D362A FP: CCTCTCTGAGGAGGTTGCCTTGGAAGCCTATGTGGCCCGGCCAG

D359A/D362A RP: CTGGCCGGGCCACATAGGCTTCCAAGGCAACCTCCTCAGAGAGG

K397E FP: CTACATTGTCCTGGCCGAGGACTTCGAGAAAGC

K397E RP: GCTTTCTCGAAGTCCTCGGCCAGGACAATGTAG

Gankyrin mutant primers

R41A FP: CTAGAACTGATCAGGACAGCGCAACAGCTTTGCACTGGGCATG

R41A RP: CATGCCCAGTGCAAAGCTGTTGCGCTGTCCTGATCAGTTCTAG

K116A FP: CACTCCATTATGCAGCTTCGGCGAATAGGCATGAGATTGCTG

K116A RP: CAGCAATCTCATGCCTATTCGCCGAAGCTGCATAATGGAGTG

D39A FP: GCTACTAGAACTGATCAGGCCAGCAGAACAGCTTTGCAC

D39A RP: GTGCAAAGCTGTTCTGCTGGCCTGATCAGTTCTAGTAGC

D71A FP: GCCAGTGAATGATAAAGATGCCGCAGGTTGGTCTCCTCTTC

D71A RP: GAAGAGGAGACCAACCTGCGGCATCTTTATCATTCACTGGC

E182A FP: CACTTAGCCTGTGATGCAGAGAGAGTGGAAGAG

E182A RP: CTCTTCCACTCTCTCTGCATCACAGGCTAAGTG

Gankyrin FP: CATGCCATGGAGGGGTGTGTGTCTAACATAATGATCTGTAACC

Gankyrin RP:

CCTTAATTAATTAGTGATGGTGGTGGTGATGACCCATAGAAGCCTCTTCACCTTCTG

CTA

4.9 Supplemental Data

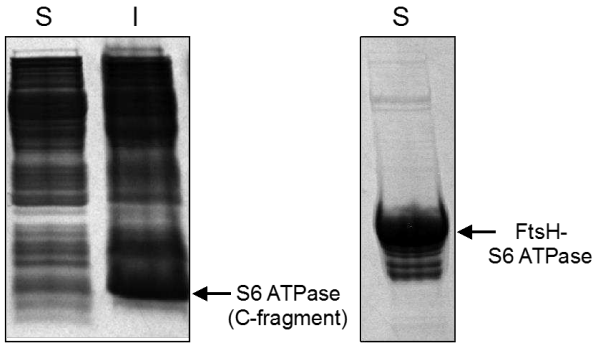


Figure S4.1 SDS-PAGE and coomassie staining of insoluble S6 ATPase (on the left) and soluble grafted FtsH-S6 ATPase (on the right). S = Soluble, I = Insoluble.

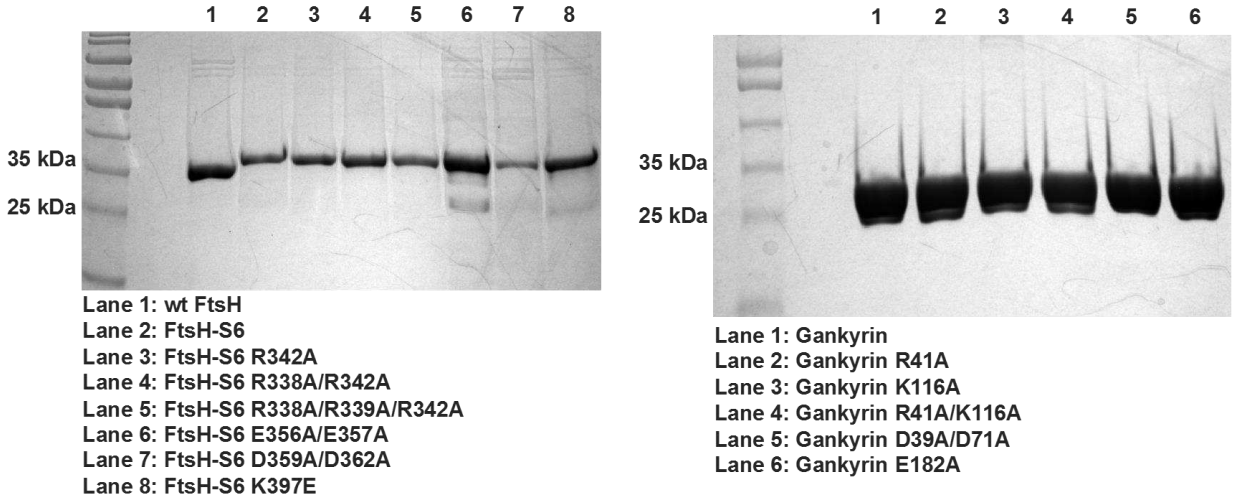


Figure S4.2 SDS-PAGE and coomassie staining of purified grafted FtsH-S6 ATPase and gankyrin, including mutants thereof, used for ITC analysis.

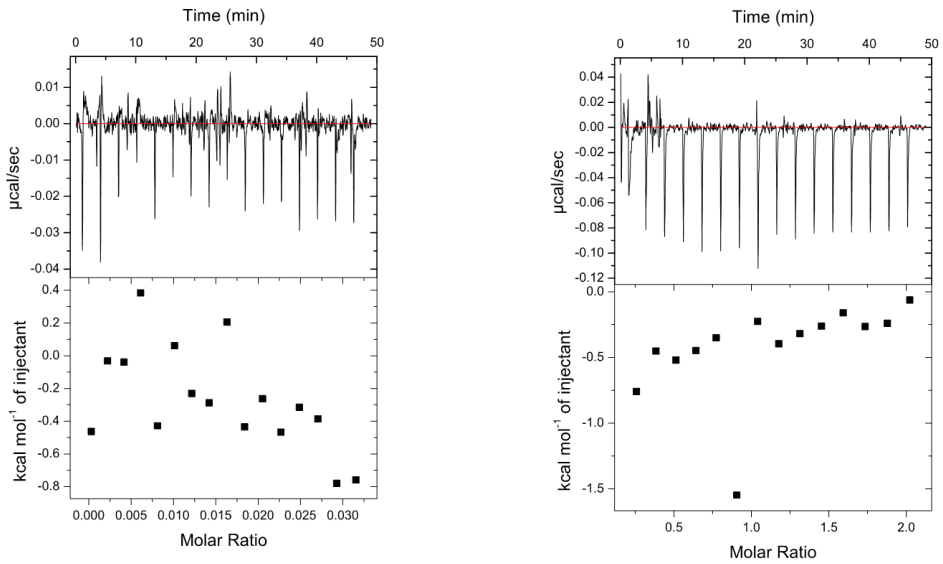


Figure S4.3 ITC evaluation of gankyrin titrated into buffer (left) and wt-FtsH titrated into gankyrin (right). No appreciable binding is observed.

Table 1, Entry 1

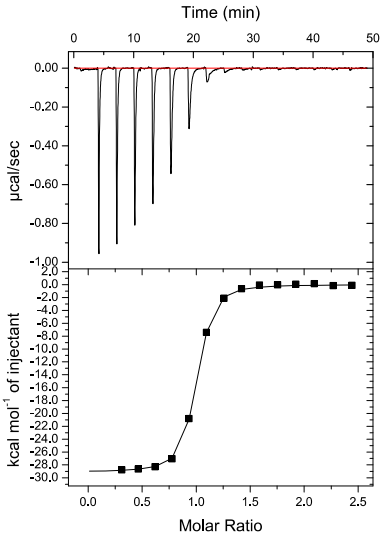


Table 1, Entry 2

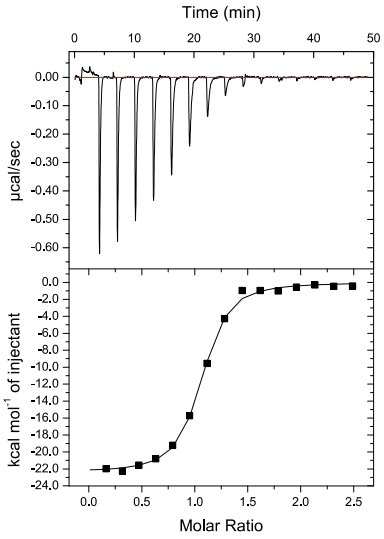


Table 1, Entry 3

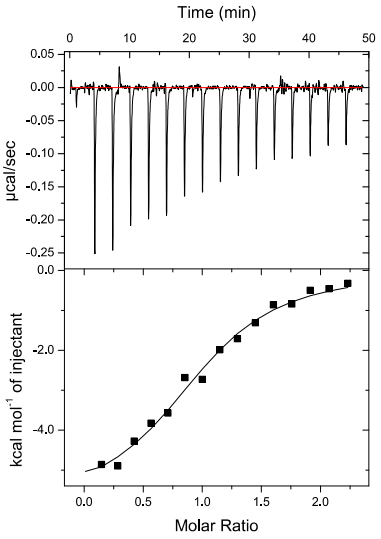


Table 1, Entry 4

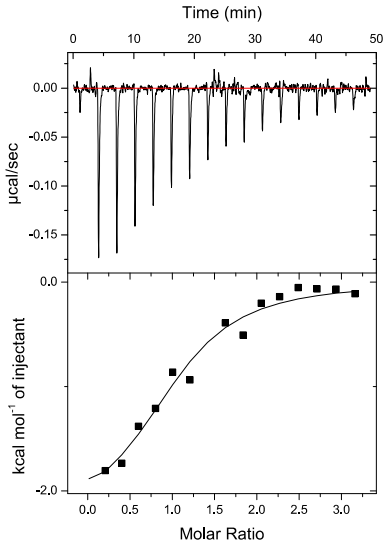


Table 1, Entry 5

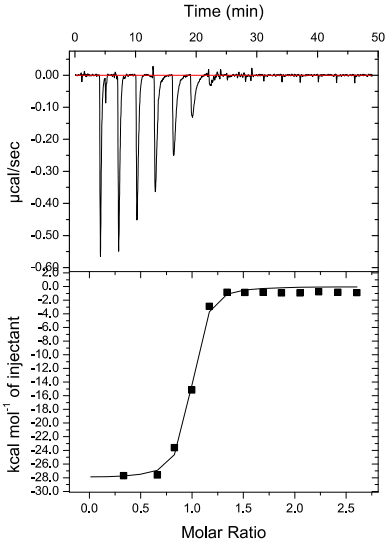
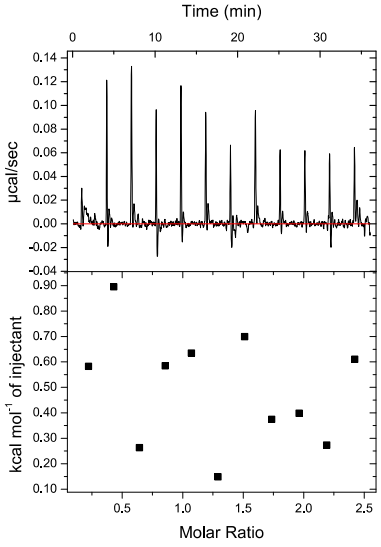


Table 1, Entry 6



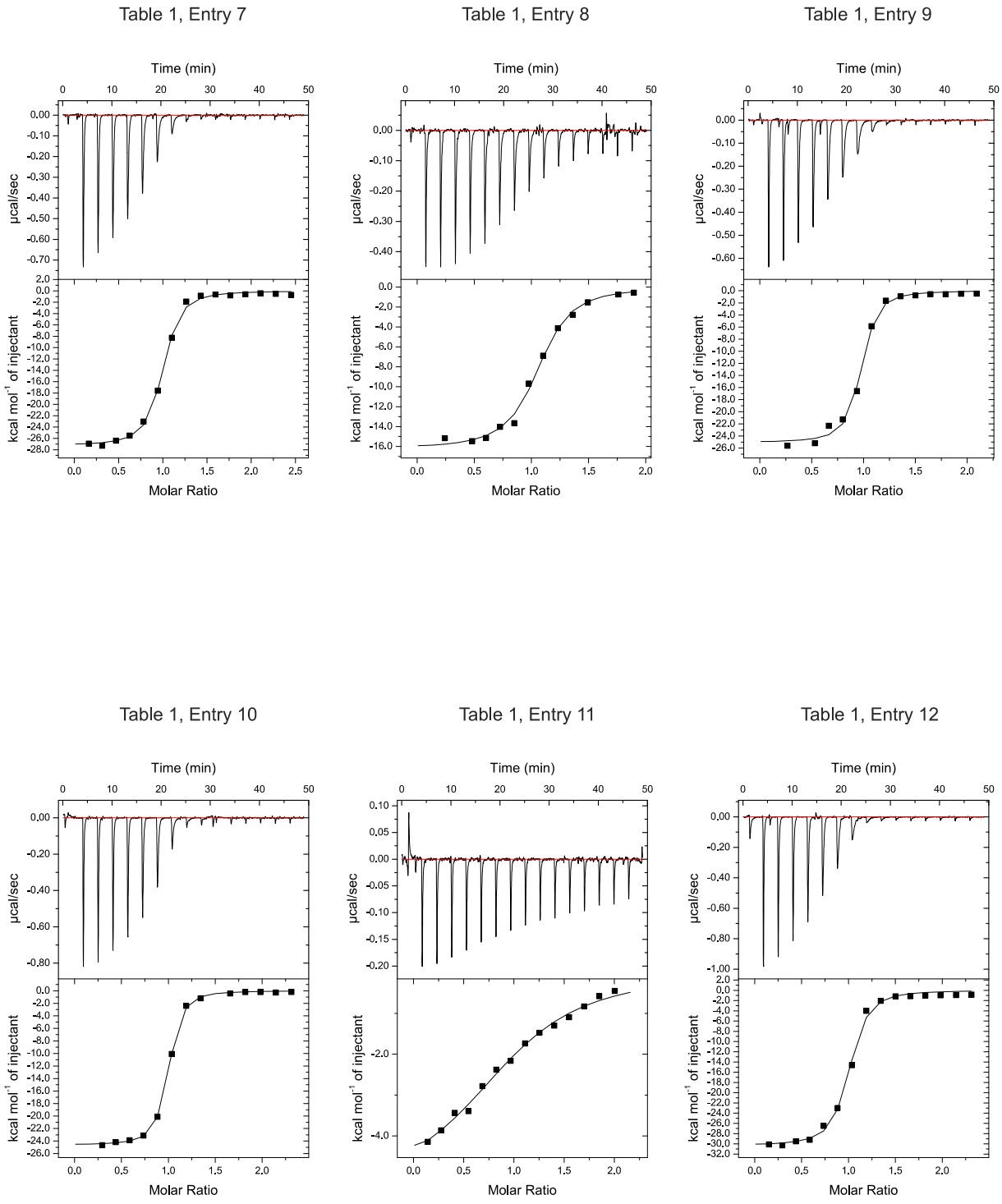


Figure S4.4 Representative ITC binding isotherms involving gankyrin and grafted FtsH-S6 ATPase, and specific mutants thereof. Data summarized in **Table 4.1**.

Table S4.1 N-values (binding stoichiometry) for each entry in Table 4.1.[a]

Entry	gankyrin	FtsH-S6 ATPase	N-value
1	wild-type	wild-type	0.93±0.01
2	wild-type	R342A	1.02±0.02
3	wild-type	R338A/R342A	0.99±0.03
4	wild-type	R338A/R339A/R342A	1.01±0.04
5	wild-type	E356A/E357A	0.93±0.02
6	wild-type	D359A/D362A	No binding
7	wild-type	K397E	0.94±0.01
8	R41A	wild-type	0.99±0.04
9	K116A	wild-type	0.91±0.01
10	D39A/D71A	wild-type	0.93±0.02
11	R41A/K116A	wild-type	0.97±0.05
12	E182A	wild-type	0.97±0.02

[a]All errors represent the standard deviation of three separate experiments. ITC conditions were as follows: 20 mM sodium phosphate, 150 mM NaCl, and 2.5 mM 2-mercaptoethanol (pH 7.4) at 25 °C.

REFERENCES

(1) Zhen, C.; Chen, L.; Zhao, Q.; Liang, B.; Gu, Y.-X.; Bai, Z.-F.; Wang, K.; Xu, X.; Han, Q.; Fang, D.; Wang, S.; Zhou, T.; Xia, Q.; Gong, W.-L.; Wang, N.; Li, H.-Y.; Jin, B.-F.; Man, J. *Oncogene* **2013**, *32* (29), 3452.

(2) Kim, Y. H.; Kim, J. H.; Choi, Y. W.; Lim, S. K.; Yim, H.; Kang, S. Y.; Chung, Y. S.; Lee, G. Y.; Park, T. J. *Exp. Mol. Pathol.* **2013**, *94* (2), 360.

(3) Fu, X.-Y.; Wang, H.-Y.; Tan, L.; Liu, S.-Q.; Cao, H.-F.; Wu, M.-C. *World J. Gastroenterol.* **2002**, *8* (4), 638.

(4) Li, J.; Knobloch, T. J.; Kresty, L. a; Zhang, Z.; Lang, J. C.; Schuller, D. E.; Weghorst, C. M. *Anticancer Res.* **2011**, *31* (9), 2683.

(5) Meng, Y.; He, L.; Guo, X.; Tang, S.; Zhao, X.; Du, R.; Jin, J.; Bi, Q.; Li, H.; Nie, Y.; Liu, J.; Fan, D. *Cancer Lett.* **2010**, *297* (1), 9.

(6) Tang, S.; Yang, G.; Meng, Y.; Du, R.; Li, X.; Fan, R.; Zhao, L.; Bi, Q.; Jin, J.; Gao, L.; Zhang, L.; Li, H.; Fan, M.; Wang, Y.; Wu, K.; Liu, J.; Fan, D. *Cancer Biol. Ther.* **2010**, *9* (2), 88.

(7) Man, J. H.; Liang, B.; Gu, Y. X.; Zhou, T.; Li, A. L.; Li, T.; Jin, B. F.; Bai, B.; Zhang, H. Y.; Zhang, W. N.; Li, W. H.; Gong, W. L.; Li, H. Y.; Zhang, X. M. *J. Clin. Invest.* **2010**, *120* (8), 2829.

(8) Li, J.; Tsai, M.-D. *Biochemistry* **2002**, *41* (12), 3977.

(9) Higashitsuji, H.; Higashitsuji, H.; Itoh, K.; Sakurai, T.; Nagao, T.; Sumitomo, Y.;

Sumitomo, H.; Masuda, T.; Dawson, S.; Shimada, Y.; Mayer, R. J.; Fujita, J. *Cancer Cell* **2005**, *8* (1), 75.

(10) Nakamura, Y.; Nakano, K.; Umehara, T.; Kimura, M.; Hayashizaki, Y.; Tanaka, A.; Horikoshi, M.; Padmanabhan, B.; Yokoyama, S. *Structure* **2007**, *15* (2), 179.

(11) Nanaware, P. P.; Ramteke, M. P.; Somavarapu, A. K.; Venkatraman, P. *Proteins Struct. Funct. Bioinforma.* **2014**, *82* (7), 1283.

(12) Chapman, A. M.; McNaughton, B. R. *ACS Chem. Biol.* **2014**, *9*, 2223.

CHAPTER FIVE

A Synthetic Protein that Potently and Selectively Binds the Oncoprotein Gankyrin, Disrupts Its Interaction with S6 ATPase, and Inhibits Gankyrin/MDM2-Dependent Ubiquitination of p53

Adapted from:
Chapman, A.M.; McNaughton, B.R., *ACS Chem. Biol.*, **2015**, *10*, 1880.

5.1 Introduction

In chapter three of this thesis, we described the use of an ankyrin repeat shape-complementary protein (Prb, **Figure 5.1A**, tan) as a scaffold for developing new gankyrin-binding proteins. Prb binds Pdar (**Figure 5.1A,** blue), an *in silico* designed thermostable ankyrin repeat with a largely hydrophobic binding face.[1] We randomized eight residues on the ankyrin repeat shape-complementary surface of Prb (**Figure 5.1B**) to all possible proteinaceous amino acids using standard molecular biology methods. We then performed two-rounds of split-superpositive GFP reassembly, and *in cellulo* screen we recently reported (chapter two)[2], to identify new proteins that bind gankyrin in a complex cellular environment (*E. coli* cells). The best protein we identified (GBP7, **Figure 5.1B)** binds gankyrin with moderate affinity (K_D ~6 μM). This work highlighted the utility of using protein shape complementarity as a starting design principle towards targeting ankyrin-repeat proteins.

Overexpression of the ankyrin repeat oncoprotein gankyrin (**Figure 5.2A**, orange) is directly linked to the onset, proliferation, and/or metastasis of many cancers, including breast[3,4], liver[5], oral[6], pancreatic[7], and colorectal cancers[8]. Additionally, gankyrin plays a prominent role in Ras-initiated tumorigenesis, which is operative in ~30% of all cancers[9]. Consistent with most

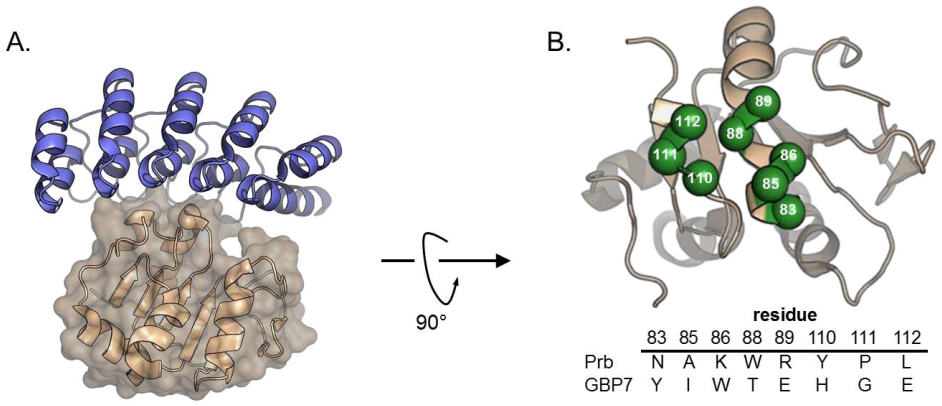


Figure 5.1 (*A*) Crystal structure of Pdar (light blue)/Prb (tan). (*B*) Residues on Prb (highlighted in green) that were mutated to generate GBP7 (chapter 3).

ankyrin repeat proteins[10], gankyrin does not have enzymatic activity. Disease-relevant processes originate from higher than normal cellular levels of gankyrin, resulting in abnormally high levels of protein-protein interactions (PPI's) involving this oncoprotein. For example, in cells that overexpress gankyrin, cyclin-dependent kinase 4 (CDK4) is bound by gankyrin at abnormally high levels. Through mechanisms that are not well understood, CDK4/gankyrin assembly increases the extent to which CDK4 phosphorylates retinoblastoma protein (Rb).[11,12] Increased cellular levels of phosphorylated Rb (pRb) leads to over-activation of E2F transcription factors and aberrant E2F-dependent transcription (**Figure 5.2B**).[11] Additionally, gankyrin can bind to the E3 ubiquitin ligase murine double minute 2 (MDM2), and in doing so, increases the extent to which p53 is ubiquitinated/polyubiquitinated (**Figure 5.2C**).[13,14] Increased ubiquitination/polybuiquitination of p53 by the MDM2/gankyrin complex ultimately leads to p53 degradation in the proteasome and suppression of p53-dependent apoptosis. Collectively, these gankyrin-dependent processes result in genome instability and cancer. Finally, gankyrin is known to function as a chaperone for the formation of the 26S proteasome[15], where it also associates with the S6 ATPase sub-unit (referred to as S6 ATPase herein, **Figure 5.2A**, grey).[12,16] It is believed that association between gankyrin

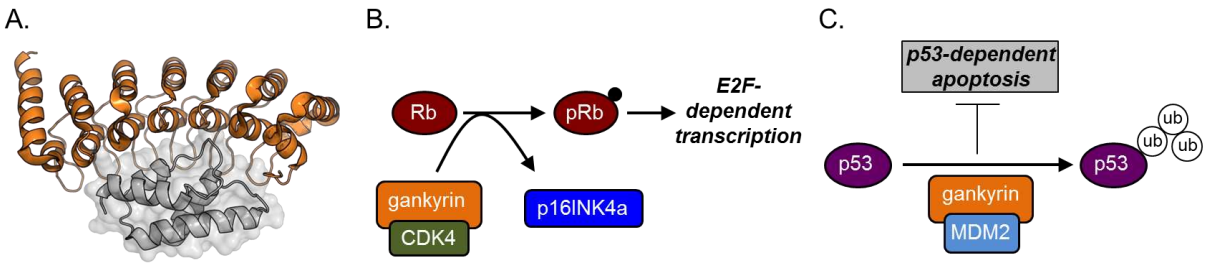


Figure 5.2 (*A*) Crystal structure of the gankyrin/S6 ATPase complex. Gankyrin is colored orange; S6 ATPase is colored gray. (*B*) Under normal conditions, p16INK4a binds cyclin-dependent kinase 4 (CDK4) and modulates phosphorylation of Rb. In gankyrin-overexpressing cells, gankyrin binds CDK4, which leads to increased cellular levels of phosphorylated Rb (pRb), which in turn leads to activation of E2F-dependent transcription. (*C*) Gankyrin forms a complex with the E3 ubiquitin ligase murine double minute 2 (MDM2), and this complex increases the extent to which p53 is polyubiquitinated , ultimately leading to p53 degradation and suppression of p53-dependent apoptosis.

and S6 ATPase facilitates delivery of polyubiquitinated p53 to the proteasome. Disease-relevant cellular processes resulting from increased cellular levels of PPIs that involve gankyrin make disruption of these assemblies an attractive therapeutic strategy.[17,18]

GBP7 represents the first example, to our knowledge, of a synthetic protein that binds gankyrin with good affinity (K_D ~6 μM).[18] However, it is likely that selective disruption of the PPIs involving gankyrin described above, as well as modulation of gankyrin-dependent oncogenesis, may require more potent recognition (K_D <100 nM) of this oncoprotein. Towards this goal, we hypothesized that GBP7 could be subjected to additional rounds of so-called 'affinity maturation' to yield more potent gankyrin-binding proteins. Our initial resurfaced Prb library, described in chapter three, focused on mutating (through saturation mutagenesis) 8 residues located on the putative ankyrin repeat binding face. Although we performed alanine scanning mutagenesis to identify which of these mutated residues were most important for gankyrin-binding, it is largely unclear how GBP7 contacts gankyrin, and therefore unclear which additional positions could be subjected to saturation mutagenesis. In the absence of more detailed structural characterization, it would be a somewhat daunting task to continue optimization of this scaffold in

a directed manner. With these limitations in mind, random mutagenesis of the entire GBP7 sequence (by error-prone PCR and DNA shuffling) represents the most attractive strategy for affinity maturation.

The original Prb resurfaced library was screened against gankyrin using our split-spGFP reassembly technology. Split-spGFP was the screening method of choice for the initial Prb library screen because it could be performed *in cellulo*, effectively building in binding-selectivity due to the presence of countless other macromolecule species, including other ankyrin repeat proteins. In addition, the screening process was relatively quick and simple; two rounds of screening over the span of roughly a week resulted in identifying our starting point, GBP7. However, split-spGFP has potential limitations when considering its use as a tool for further affinity maturation. Most notably, it is still somewhat unclear if there is a direct quantitative link between increased cellular GFP fluorescence levels and PPI binding affinity, a limitation that is fairly ubiquitous for *in cellulo* reporter-based protein-fragment complementation assays (PCAs).[19,20]

In contrast to PCAs, display-based high-throughput screening tools, including yeast display[21], phage display[22], and mRNA/ribosome display[23], have been utilized extensively to generate proteins/peptides that potently bind to their desired protein target. mRNA and ribosome display are cell-free techniques that can be used to screen remarkably large libraries ($>10^{10}$), but are operationally challenging. Phage display is the most widely used method for molecular evolution of proteins/peptides, including evolution of antibody fragments that resulted in FDA-approved humanized antibodies.[24] However, we chose to employ yeast display to screen our GBP7 random mutagenesis library for several reasons. First, yeast display is an operationally simple technique that allows for quantitative screening with the use of two-color labeling fluorescence-activated cell sorting (FACS, display and binding). Protein target concentration can be controlled

(lowered) over iterative rounds of screening to identify displayed proteins with increasingly tight binding affinity. Since fluorescence signal correlates directly with improved fitness, displayed proteins can be simultaneously screened for both increased stability (i.e. higher display signal) and binding affinity. In addition, PPI affinity can be directly measured using yeast display, eliminating the need for additional cloning and characterization steps. Although the potential library size is theoretically smaller using yeast display screening than with comparable methods (~10^7 vs. >10^9), the advantages stated above outweigh this fact when screening a random mutagenesis library.[25]

Starting from GBP7, we applied yeast display, error-prone PCR, DNA shuffling, and protein engineering to identify new gankyrin-binding proteins with dramatically improved affinity (K_D ~20 – 100 nM). Isothermal Titration Calorimetry (ITC) was used to provide the thermodynamic signature of these interactions, as well as measure the effect mutations in, near, or relatively far from the putative gankyrin-binding site have on binding affinity. The highest affinity protein binds gankyrin very tightly (K_D ~21 nM) and with exquisite selectivity in cell lysate. This protein is also able to modulate the PPI between gankyrin and S6 ATPase, and dramatically suppress gankyrin/MDM2-dependent polyubiquitination of p53. The proteins described in this chapter represent the tightest gankyrin-binding reagents known to date; the highest affinity proteins bind gankyrin ~3-fold tighter than S6 ATPase, a physiological binding partner of gankyrin. Additionally, these proteins represent the only known protein modulators of gankyrin function (p53 ubiquitination).

5.2 Yeast Display Screening Reveals Potent Synthetic GBPs

We began with GBP7, a synthetic protein we recently reported that binds gankyrin with moderate affinity (K_D ~6 μM, described in chapter three).[18] We performed rounds of yeast display

screening[21], error-prone PCR[26], DNA shuffling[27], and protein engineering to optimize this interaction. Other than the putative ankyrin repeat binding face, which we previously matured for gankyrin affinity by split-superpositive GFP reassembly, it was unclear which residues in GBP7 should be mutated to improve affinity. Thus, we started by incorporating mutations in an unbiased manner using error-prone PCR (ep-PCR), and identified the tightest binders by yeast display. Using a commercially available ep-PCR kit (GeneMorph II Random Mutagenesis Kit, Agilent) we prepared a library of GBP proteins. Sequencing 30 library members after a single round of ep-PCR indicated that approximately four nucleic acid mutations occurred per gene. Following standard methods, we performed yeast display to identify proteins with improved affinity. Briefly, a library of gankyrin-binding proteins (generated by ep-PCR) was displayed on EBY100 *saccharomyces cerevisiae.* EBY100 constitutively express and display the Aga1 mating protein. GBPs are cloned into a pcTCON2 vector as fusions with Aga2, which forms disulfide linkages with Aga1 when secreted from the yeast cell. In addition, displayed GBPs contain a C-terminal *myc* tag. Thus, cells can be incubated with a commercially available fluorescein isothiocyanate (FITC)-labelled anti-*myc* antibody and display efficiency of folded library members can be measured by flow cytometry. The yeast display library was incubated with varied concentrations of gankyrin that was first biotinylated with BirA using standard methods[28], then complexed with a streptavidin-phycoerythrin (PE) conjugate. Since the emission profiles of FITC and phycoerythrin are orthogonal, relative phycoerythrin:FITC levels can be measured by flow cytometry, and those yeast with the highest levels of bound gankyrin were sorted by FACS. Following the first round of yeast display and FACS, enriched yeast was grown to confluence over 3 days and plasmid DNA was extracted (Yeast Plasmid Miniprep II, Zymoprep). Library DNA was either used in further diversification or sequenced.

The library generation and screening protocol to optimize gankyrin-binding affinity and selectivity is shown in **Figure 5.3A**. In yeast display rounds 1-4, library generation was achieved by ep-PCR. In each round, the level of gankyrin was decreased (round 1: 1000 nM; round 2: 500 nM; round 3: 100 nM; round 4: 50 nM, described in **Table S5.1**) to ensure that only the tightest gankyrin-binding proteins were enriched to the next round of screening. In the fifth round, no diversification reaction was done; however, we challenged gankyrin-binding proteins by adding unlabeled off-target ankyrin repeat proteins. The enriched library of yeast displayed gankyrin-binding proteins were incubated with 50 nM gankyrin-phycoerythrin and 1 μM each of Pdar, a synthetic ankyrin repeat that is bound by the scaffold protein Prb[1], and p16INK4a, an ankyrin repeat protein that competes with gankyrin for binding to CDK4.[11] Thus, any yeast displayed protein that bound gankyrin in round 5 selectively did so in the presence of 20 equivalents each of unlabeled off-target protein. Flow cytometry data from each round is shown in **Figure 5.3B**; enriched cells are highlighted by encapsulation in the dashed box.

Following the first five rounds of yeast display screening, we sequenced ten clones and performed an initial assessment of gankyrin-binding by an Enzyme-Linked Immunosorbent Assay (ELISA). Briefly, biotinylated gankyrin was immobilized onto streptavidin-coated plates (Pierce). Solutions containing 25 nM C-terminally FLAG-tagged GBPs (GBP7.1-GBP7.10) were incubated with immobilized gankyrin, then washed 3 times. Following incubation with an anti-FLAG-HRP antibody (Abcam), and after subsequent washing steps, TMB-One HRP substrate (Promega) was added and relative amounts of complex in each well was measured by colorimetric analysis on a plate reader. As shown in **Figure 5.4A**, all of the proteins we tested exhibit appreciable affinity for gankyrin. However, two of the clones (GBP7.5 and GBP7.7) were the most potent gankyrin-binding proteins and thus were investigated further. To provide more quantitative feedback on the

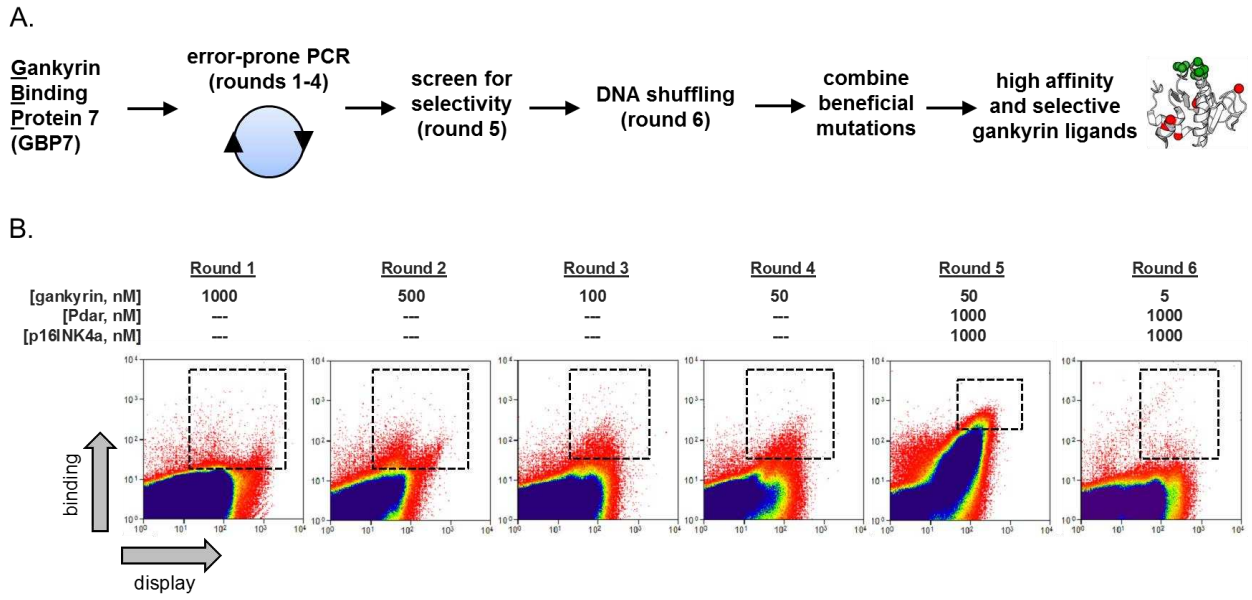


Figure 5.3 (*A*) Strategy for the development of potent (low nanomolar dissociation constant) GBPs from GBP7. GBP7, which binds gankyrin with moderate affinity (~6 μM dissociation constant), was diversified by error-prone PCR, and highest affinity library members were enriched by FACS, using yeast display. The enriched library was further diversified by a second round of error-prone PCR, and highest affinity binders were enriched by FACS. Error-prone PCR and FACS screening were done for a total of 4 rounds. Enriched library members from the fourth round were then screened by yeast display/FACS for selective recognition of gankyrin in a solution containing 20 equivalents of off-target ankyrin repeats (Pdar and p16INK4a). Following this round of yeast display/FACS, the enriched library was scrambled by DNA shuffling and the resulting protein library was screened for gankyrin affinity by yeast display/FACS. Following sequencing of enriched gankyrin-binding proteins, we combine common beneficial mutations in an effort to further optimize affinity for gankyrin. (*B*) Flow cytometry data from screening of error-prone PCR generated libraries (rounds 1-4), the screen for selective recognition of gankyrin (round 5), and screening of the DNA shuffling generated library (round 6).

effectiveness of our yeast display screening, the binding affinity, thermodynamic signature, and stoichiometry of these new protein-protein interactions were characterized by ITC. As shown in **Figure 5.4B**, GBP7.5 and GBP7.7 bind gankyrin with significantly improved affinities (K_D ~139 and ~125 nM, respectively), compared to our starting point GBP7 (K_D ~6 μM), representing a ~46-fold improvement in affinity.

To further optimize the complex, we performed DNA shuffling on the entire sub-library of clones that were enriched in round 5, and screened the shuffled protein library by yeast display. We sequenced 25 clones and performed an initial assessment of gankyrin-binding by ELISA. The

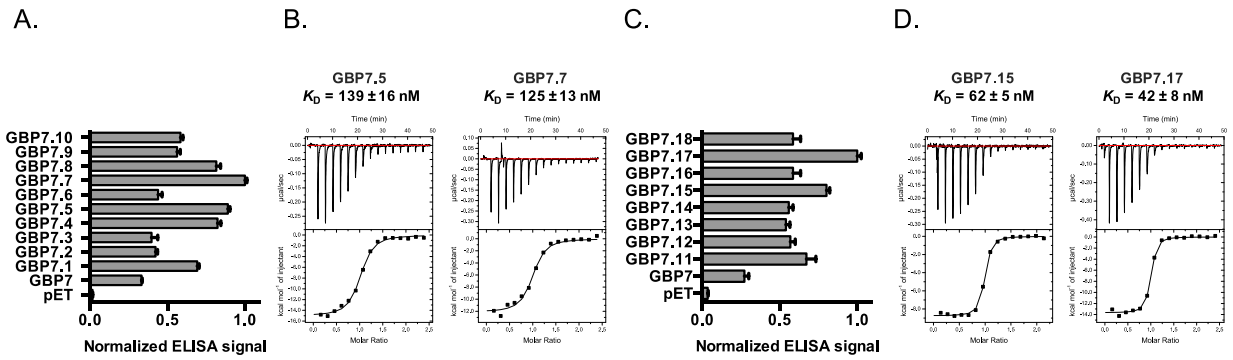


Figure 5.4 (*A*) Enzyme-Linked Immunosorbent Assay (ELISA) data for GBP7.1−7.10, which were generated by error-prone PCR and identified in screening rounds 1−5. (*B*) Isothermal titration calorimetry (ITC) analysis of the protein−protein interactions involving gankyrin and GBP7.5 or GBP7.7. (*C*) ELISA data for gankyrin-binding proteins 7.11−7.18, which were generated from DNA shuffling and identified in screening round 6. (*D*) ITC analysis of the protein−protein interactions involving gankyrin and GBP7.15 or GBP7.17. Stated K_D values are the mean of three independent experiments, with standard deviation error.

best performing clones are show in **Figure 5.4C**. Similar to the previous ELISA experiment, all of the proteins exhibited appreciable affinity for gankyrin; however, two of the clones (GBP7.15 and GBP7.17) were the most potent gankyrin-binding proteins, and thus the complex between these proteins and gankyrin was further characterized by ITC. As shown in **Figure 5.4D**, DNA shuffling generated proteins with significantly improved affinity. GBP7.15 and GBP7.17 bind gankyrin with low nanomolar dissociation constants (K_D ~62 and ~42 nM, respectively): a ~120-fold improvement from the initial complex involving GBP7. The observed changes in enthalpy (ΔH) and entropy (-TΔS) for the PPI involving gankyrin and GBP7.15 were -8.9 (±0.2) and -0.9 (±0.2) kcal/mol, respectively. The observed changes in enthalpy (ΔH) and entropy (-TΔS) for the PPI involving gankyrin and GBP7.17 were -13.5 (±0.5) and 3.5 (±0.5) kcal/mol, respectively.

PH1109, the precursor protein to Prb, is a Coenzyme A (CoA)-binding protein.[29] Although CoA is not involved in the binding interaction between Prb and Pdar, co-crystallization of the complex was only achieved in the presence of CoA, suggesting that CoA plays a role in stabilizing either Prb, or the complex in general.[1] One concern we had when identifying new GBPs was that

CoA, which binds adjacent to the resurfaced region on Prb, might still be able to bind to our synthetic proteins and disrupt the gankyrin-GBP interaction in mammalian cells, potentially complicating downstream biochemical applications (**Figure 5.5A**). In order to address this concern, we performed a modified ITC experiment with our tightest binding variant, GBP7.17. The modified experiment consisted of titrating gankyrin into a 1:1 mix of GBP7.17 and CoA. In the presence of CoA, gankyrin still tightly binds to GBP7.17 (K_D ~50 nM, **Figure 5.5B**), suggesting that CoA has little, if any, effect on the binding interaction. Additionally, we saw no binding isotherm when GBP7.17 was titrated into CoA, suggesting that CoA has no appreciable binding affinity for GBP7.17 (**Figure 5.5C**).

5.3 Mutational Analysis of New GBPs

Of the 25 clones we sequenced from the enriched gankyrin-binding library, three mutations, N55Y (which is near the putative ankyrin repeat binding face), D79G and R132C (both relatively distant from the putative ankyrin repeat binding face) were found in all 25 sequenced clones (**Figure 5.6A** and **Figure 5.6B**, boxed). In order to measure the importance of each mutation in gankyrin recognition, we prepared a library of reversion mutants and measured their affinity for gankyrin by ITC. Reverting the cysteine at position 132 back to arginine decreased affinity ~2.8-fold, compared to GBP7.17 (**Figure 5.6C**). The G79D reversion mutant bound gankyrin with no appreciable loss in affinity (**Figure 5.6D**). Interestingly, mutating tyrosine 55 back to asparagine had a much more dramatic effect, and resulted in ~22.5-fold lower affinity for gankyrin (**Figure 5.6E**). This makes sense, since residue 55 is relatively close to the putative gankyrin-binding face and might directly engage gankyrin.

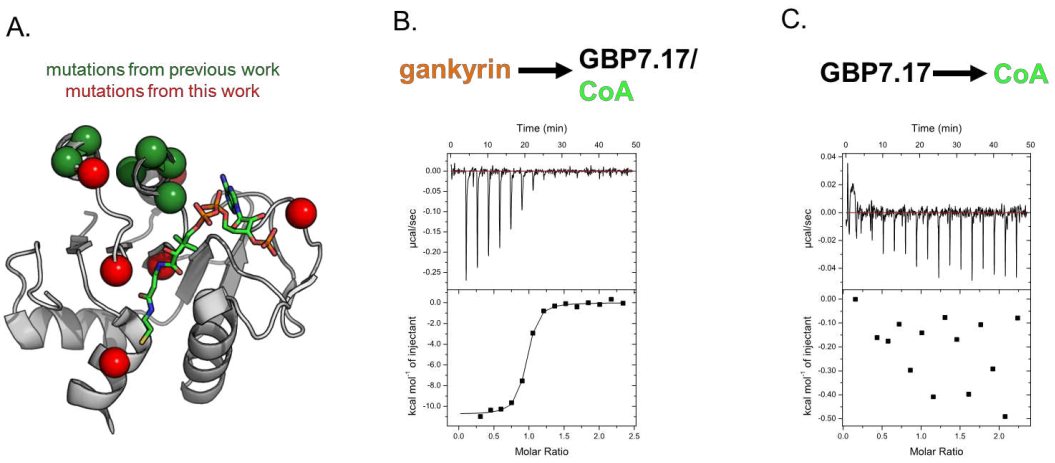


Figure 5.5 (*A*) CoA (stick model) binding to Prb (gray). Mutations for GBP7.17 are highlighted in green (split-spGFP screen) and red (affinity maturation). (*B*) ITC data for an experiment that involves titrating gankyrin into a 1:1 mix of GBP7.17 and CoA (K_D ~50 nM). (*C*) ITC data for an experiment that involves titrating GBP7.17 into CoA.

Among the residues that were initially optimized in our previous work to generate the starting protein GBP7, one (H110R) was mutated in GBP7.17 as a result of ep-PCR and DNA shuffling. Additionally, a Y93C mutation was found near the initially optimized binding face. To measure the importance of these residues, we made reversion mutants and measured their affinity for gankyrin by ITC. Interestingly, converting cysteine 93 back to tyrosine only lowered affinity for gankyrin by ~2.2-fold, compared to GBP7.17 (**Figure 5.6F**). Additionally, reverting arginine 110 back to histidine only lowered affinity by ~1.5-fold (**Figure 5.6G**).

5.4 Optimizing the Gankyrin-GBP PPI

Three mutations on or near the designated gankyrin-binding face differentiate the two best performing GBPs (GBP7.15 and GBP7.17). Both an E112V and Y93H mutation are found in GBP7.15, and these residues reside within the originally matured gankyrin-binding face. Slightly outside of this, we observe an E57K mutation in GBP7.15. As stated previously, a Y93C mutation was found in our best performing GBP (GBP7.19). However, reversion of this residue did not

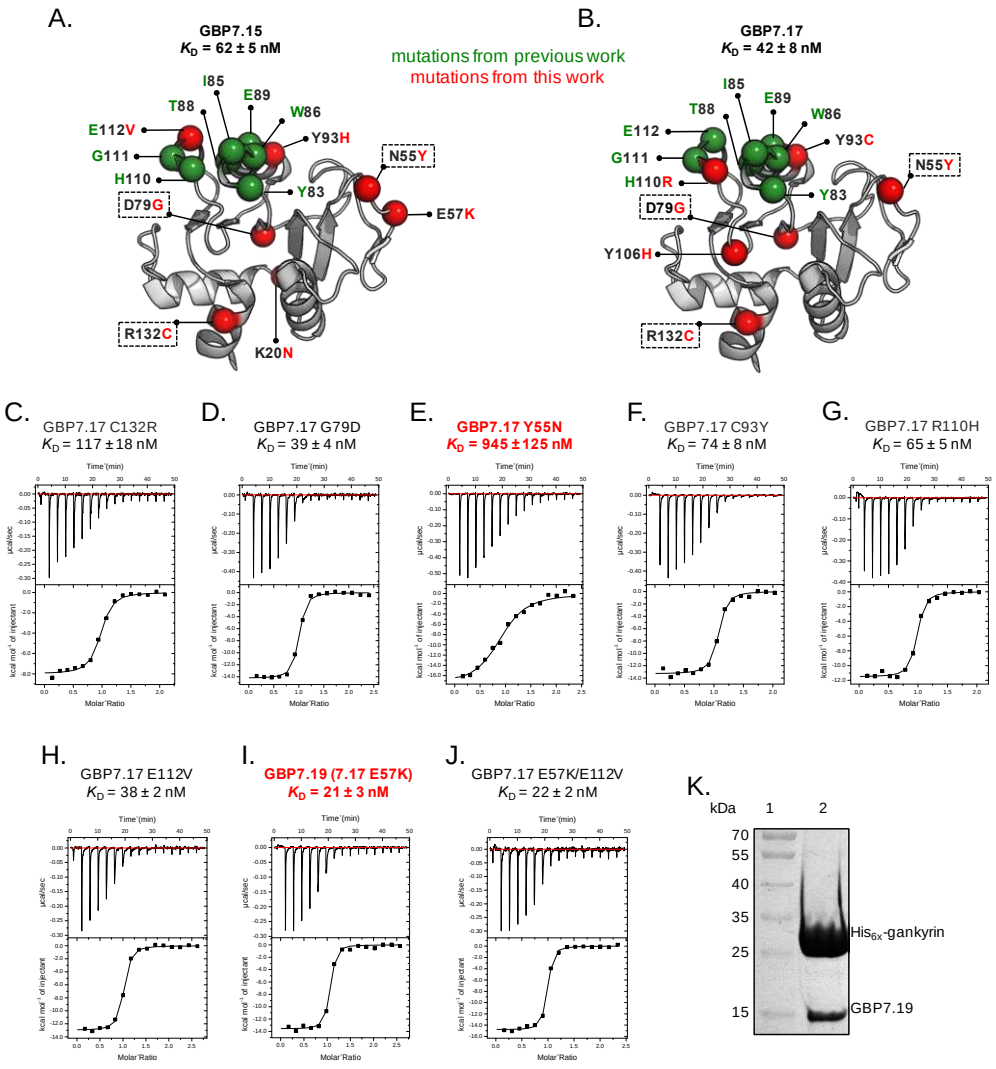


Figure 5.6 (*A* and *B*) GBP7.15 and GBP7.17, which were identified as the tightest gankyrin-binding proteins following screening of the DNA shuffled library. Green colored residues were generated in our earlier work as a result of split-superpositive GFP screening, and were thus found in the starting protein (GBP7). Red colored mutations were generated in this work, as a result of error-prone PCR and DNA shuffling. (*C-G*) ITC data for five reversion mutants of GBP7.17, which provided information on which residues are most critical to gankyrin recognition. (*H-J*) ITC data for GBP7.17 based proteins that contain mutations found in GBP7.15. Stated K_D values are the mean of three independent experiments, with standard deviation error. (*K*) Co-purification of His_{6x}-gankyrin and GBP7.19 from *E. coli* cell lysate.

dramatically lower binding, suggesting residue 93 plays a less prominent role in stabilizing a complex with between gankyrin and GBP7.17. Given this, we focused on integrating the E112V

and/or E57K mutations found in GBP7.15 into our best performing protein (GBP7.17), and measuring how each change alters affinity for gankyrin.

Replacing the glutamic acid at position 112 in GBP7.17 with the valine that is found in GBP7.15 does not appreciably improve affinity for gankyrin (**Figure 5.6H**). In contrast, mutating the negatively charged glutamic acid at position 57 in GBP7.17 with the positively charged lysine that is found in GBP7.15 results in a mutant with significantly improved affinity (K_D ~21 nM, **Figure 5.6I**). This protein is referred to as GBP7.19 herein. This is consistent with our earlier finding that the Y55N reversion mutation resulted in dramatically lowered affinity for gankyrin. Collectively, these two results suggest that the surface of GBP7.19 displaying residues 55 and 57 might directly engage gankyrin, and as a result, the chemical makeup of these residues can dramatically effect complex stability with gankyrin. The observed changes in enthalpy (ΔH) and entropy ($-T\Delta S$) for the PPI involving gankyrin and GBP7.19 were -13.8 (±0.3) and 3.3 (±0.2) kcal/mol, respectively. Addition of the E57K and E112V mutation to the GBP7.17 scaffold does not appreciably improve affinity, compared to GBP7.19 (K_D ~22 nM, **Figure 5.6J**). Owing to its affinity for gankyrin, we focused on measuring the selectivity of the GBP7.19/gankyrin interaction, and the ability of GBP7.19 to modulate a physiologically-relevant gankyrin-dependent PPI and a gankyrin-dependent and disease-relevant biochemical process.

GBP7.19 binds gankyrin with exquisite selectivity, when expressed in *E. coli* with His_{6x}-tagged gankyrin, the two proteins co-elute following nickel-NTA purification (**Figure 5.6K**, lane 2). Appreciable levels of other co-purified proteins is not observed, demonstrating the selectivity of the gankyrin/GBP7.19 binding interaction – even in a complex cellular milieu.

5.5 GBP7.19 Inhibits the Gankyrin/S6 ATPase PPI

Any therapeutic utility of gankyrin-targeted reagents requires tight and selective recognition of this oncoprotein. In this regard, the proteins described above are excellent candidates. However, modulation of disease-relevant gankyrin activity likely requires physical disruption or inhibition of physiological PPIs. As stated above, gankyrin is known to bind CDK4 and MDM2, leading to increased Rb phosphorylation and p53 polyubiquitination, respectively. Unfortunately, challenges exist when trying to express recombinant CDK4 or MDM2 in *E. coli*. Gankyrin is also known to bind the C-terminal portion of S6 ATPase, a proteasomal subunit (**Figure 5.1A**).

As described in chapter four, we recently grafted the C-terminal ATPase subdomain of S6 onto the C-terminal ATPase subdomain of FtsH, an *E. coli* derived protein, and showed that this new protein (FtsH-S6 ATPase, **Figure 5.7A**) expresses well in *E. coli,* is folded in solution, and binds gankyrin with excellent affinity (K_D ~67 nM, **Figure 5.7B**). Since FtsH-S6 ATPase expresses independent of gankyrin, this surrogate for the C-terminal subunit of S6 ATPase can be used to determine if GBP7.19 modulates the gankyrin/S6 ATPase interaction, and physically disrupts this complex.

The ability of our tightest gankyrin-binding protein (GBP7.19) to modulate the FtsH-S6 ATPase/gankyrin interaction was initially tested by ITC. As stated above, FtsH-S6 ATPase binds gankyrin with excellent affinity (K_D ~67 nM, **Figure 5.7B**). However, when FtsH-S6 ATPase is titrated into a gankyrin/GBP7.19 complex, we observe dramatically decreased affinity (**Figure 5.7C**). This finding suggests that GBP7.19 likely binds the concave face of gankyrin and therefore is in a position to block the gankyrin/S6 ATPase interaction. The dramatic change in enthalpy (ΔH =-29 kcal/mol for the titration of S6 ATPase into gankyrin; ΔH =-14 kcal/mol for the titration of

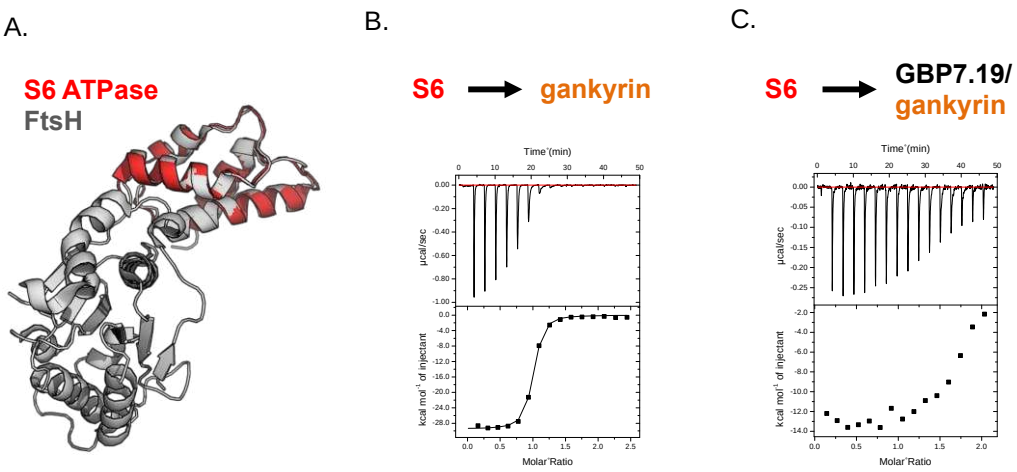


Figure 5.7 (*A*) S6 ATPase (red) superimposed on the C-terminal ATPase domain of FtsH. (*B*) ITC data for an experiment that involves titrating FtsH-S6 ATPase into gankyrin (K_D ~67 nM). (*C*) ITC data for an experiment that involves titrating FtsH-S6 ATPase into a preformed GBP7.19/gankyrin complex.

S6 ATPase into pre-complexed gankyrin/GBP7.19) also suggests an altered mode of binding between S6 ATPase and gankyrin, possibly due to GBP7.19 blocking a portion of the protein-protein interface found in the native S6 ATPase/gankyrin complex.

5.6 GBP7.19 Suppresses Gankyrin/MDM2-Dependent Ubiquitination of p53

We next set out to determine if GBP7.19 modulates gankyrin- and MDM2-dependent p53 ubiquitination – a principal disease-relevant role for this oncoprotein. As stated above, gankyrin binds the E3 ubiquitin ligase MDM2, and in doing so increases the extent to which p53 is polyubiquitinated. Ultimately, this leads to decreased cellular levels of p53 and suppression of p53-dependent apoptosis. We used a commercially available *in vitro* assay (MDM2/p53 Ubiquitination Kit, BostonBiochem) to measure p53 ubiquitination without gankyrin, with gankyrin, or with gankyrin and varied concentrations of GBP7.19. Briefly, p53 was mixed with the ubiquitination cocktail (reaction buffer, E1, E2, MDM2, Mg^{2+}-ATP, ubiquitin) and native p53 was measured by Western blot. As expected, only ~68% of native p53 is present following the

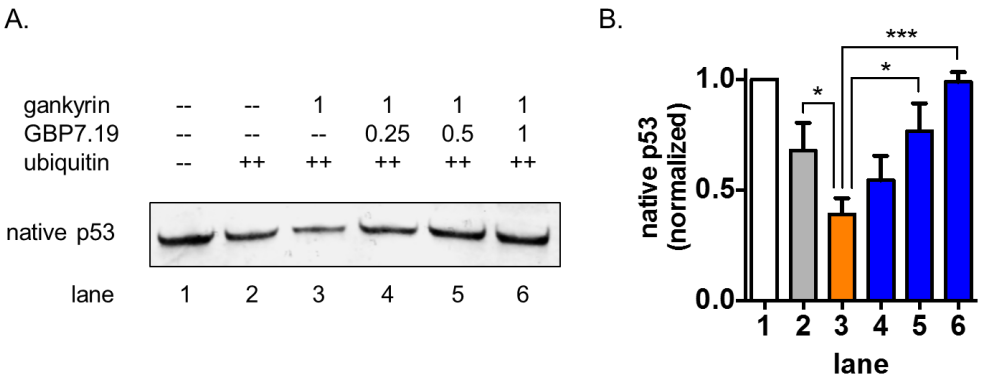


Figure 5.8 (*A and B*) Levels of native p53 following treatment with a ubiquitination cocktail. Lane 1, p53 that was treated with the ubiquitination cocktail that lacks ubiquitin (no ubiquitination occurs); lane 2, native p53 levels following ubiquitination under normal conditions; lane 3, native p53 levels following ubiquitination with 1 equivalents of gankyrin; lane 4, native p53 levels following ubiquitination with 1 equivalents of gankyrin and 0.25 equivalents of GBP7.19; lane 5, native p53 levels following ubiquitination with 1 equivalents of gankyrin and 0.5 equivalents of GBP7.19; lane 6, native p53 levels following ubiquitination with 1 equivalents of gankyrin and 1.0 equivalents of GBP7.19. Bars represent the average of three independent experiments. (*B*) Errors bars represent standard deviation; statistical analysis: unpaired parametric t test: *P < 0.05; ***P < 0.0005.

ubiquitination reaction (**Figure 5.8A** and **Figure 5.8B**, lane 2), compared to a control experiment where no ubiquitin is added (**Figure 5.8A** and **Figure 5.8B**, lane 1). Adding 1 equivalent of gankyrin to the ubiquitination cocktail (relative to MDM2) resulted in a significant reduction in the level of native (unubiquitinated) p53 (**Figure 5.8A** and **Figure 5.8B**, lane 3), compared to the ubiquitination reaction that lacks gankyrin (**Figure 5.8A** and **Figure 5.8B**, lane 2). Addition of GBP7.19 results in decreased ubiquitination of p53. When 0.25, 0.5, or 1 equivalents of GBP7.19 is added to the ubiquitination cocktail (relative to gankyrin and MDM2), we observe a concentration-dependent and dramatic increase in the levels of native p53 (**Figure 5.8A** and **Figure 5.8B**, lanes 4-6). One equivalent of GBP7.19 completely inhibits p53 ubiquitination. One possible explanation for this observation is that GBP7.19 binds the gankyrin/MDM2 complex (as opposed to displacing MDM2), and the ternary complex is unable to perform the ubiquitination reaction.

5.7 Conclusion

Gankyrin is a recently reported ankyrin repeat oncoprotein whose overexpression is directly linked to the onset, proliferation, and/or metastasis of a number of cancers. In forming a complex with MDM2 gankyrin increases the extent to which p53 is polyubiquitinated, leading to suppression of p53-dependent apoptosis and cell cycle deregulation in gankyrin overexpressing cells.

Collectively, the findings in this chapter can be summarized as follows: using error-prone PCR, DNA shuffling, yeast display screening, and protein engineering, we were able to dramatically optimize a PPI involving the oncoprotein gankyrin and synthetic gankyrin-binding proteins (GBPs). The most active protein we identified (GBP7.19) bind gankyrin very tightly (K_D ~21 nM), modulate a physiologically-relevant interaction involving gankyrin and S6 ATPase, and dramatically decrease gankyrin/MDM2-dependent ubiquitination of p53, a principal disease-relevant function of gankyrin.

The proteins reported in this chapter represent the most potent gankyrin-binding reagents known to date, establish GBP7.19 as a viable tool to study gankyrin-dependent cellular processes, and may represent an intermediate point in the development of a gankyrin-targeted protein therapeutic.

5.8 Methods

Protein Purification Gankyrin-binding proteins (GBPs) were amplified using oligonucleotides and cloned into a pET plasmid using restriction enzymes BamHI and PacI, resulting in N-terminal His_{6x}-tagged constructs, which were confirmed by DNA sequencing (all constructs in this manuscript were confirmed by GENEWIZ, South Plainfield, NJ). GBP mutants were made using

site-directed mutagenesis. These constructs were transformed into BL21s (DE3). Gankyrin was cloned into a pET plasmid using restriction enzymes NcoI and PacI, resulting in a C-terminally His_{6x} -tagged construct, and transformed into BL21s (DE3). A gankyrin construct with a C-terminal Avi-tag (termed gankyrin-Avi) was also made in this fashion. Cells were grown in 1-2.5 L LB cultures containing carbenicillin at 37 °C to OD_{600} ~0.5 and induced with 1 mM IPTG at 25 °C for 8 hours. Cells were then collected by centrifugation, resuspended in phosphate buffer (20 mM sodium phosphate pH 7.4, 150 mM NaCl) or HEPES buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 1mM DTT) with protease inhibitor tablets (Roche) and stored at -20 °C. Frozen pellets were thawed and sonicated for 2 minutes. The lysate was cleared by centrifugation (8000 rpm, 20 minutes) and the supernatant was mixed with 1 mL of Ni-NTA agarose resin for 1 hour. The resin was collected by centrifugation (4750 rpm, 5 minutes). The resin was washed with 50 mL of buffer containing 20 mM imidazole, followed by 10 mL buffer containing 50 mM imidazole. Proteins were then eluted with 5 mL buffer containing 400 mM imidazole. The proteins were dialyzed against buffer and analyzed for purity by SDS-PAGE. Purified proteins were quantified using absorbance at 280 nm and confirmed with a modified Lowry Assay.

Biotinylation Gankyrin-Avi (purified as shown above) was biotinylated using the BirA-biotin ligase kit (Avidity). Briefly, 40 μM gankyrin-Avi was biotinylated in a 1 mL reaction (containing Biomix A and Biomix B) at 30 °C for 45 minutes using 1.0 μL 3 mg/mL BirA. After extensive dialysis in HEPES buffer to remove biotin, complete biotinylation was confirmed by mass spectrometry (Colorado State University, Central Instrument Facility).

Protein Library Preparation EBY100 yeast (trp-, leu-, with the Aga1p gene stably integrated) and the pCTCON2 plasmid were generously provided by the Wittrup lab (MIT). The gene encoding gankyrin-binding protein 7 (GBP7) was amplified by PCR cloned into pCTCON2 in-frame with Aga2, an N-terminal HA-tag, and a C-terminal *myc* tag, using the restriction enzymes NheI and BamHI. When analyzed for display only, GBP7 displayed efficiently on EBY100 cells (~75 %). To construct the error-prone PCR library, homologous recombination (referred to as HR herein) primers were designed for pCTCON2 so that each primer sequence overlapped with ~40 bases on the plasmid. Random mutations were introduced into GBP7 using the Genemorph II random mutagenesis kit (Agilent). DNA shuffling was used in the final round as the diversification method. Briefly, 5 μg of amplified insert was digested with 1 μL DNase I (Sigma-Aldrich, diluted 1:1000 in 20 mM Tris-HCl pH 8.0, 5 mM $CaCl_2$, 50 % glycerol) in 50 mL 20 mM Tris-HCl pH 8.0 and 10 mM $MnCl_2$ for 2 minutes at 15 °C. The reaction was immediately transferred to 90 °C for 10 minutes to stop DNA cleavage, and then ran on a 3 % agarose gel. A smear from ~50-200 bases was extracted. A modified PCR was ran (10 cycles without primers, then 40 cycles with HR primers) using the digested fragments as a template, which generated the DNA shuffled library.

Yeast Display Screening Approximately 2 μg of double digested (NheI and BamHI) pCTCON2 plasmid and 5 μg amplified library (with HR overhangs) were mixed and transformed via electroporation into 50 μL of electrocompetent EBY100 using 2 mm cuvettes, and immediately rescued with 1 mL pre-warmed YPD for 2 hours at 30 °C. After rescue, yeast was centrifuged at 1,300 g for 1 minute and supernatant YPD was removed. Yeast was suspended in 50 mL fresh SD-CAA (5.4 g/L Na_2HPO_4, 8.6 g/L $NaH_2PO_4 \cdot H_2O$, 20 g/L dextrose, 6.7 g/L yeast nitrogen base w/o amino acids, 5 g/L casamino acids, 100 kU/L penicillin, 0.1 g/L streptomycin), plated by serial

dilution onto SD-CAA plates (SD-CAA with 182 g/L sorbitol and 15 g/L agar), and incubated at 30 °C. After 3 days, colonies were counted, resulting in libraries of 1-2 x 10^7. Yeast was sub-cultured to 0.5 x 10^7 cells/mL in 50 mL SD-CAA and grown to an OD_{600} = 2.0. Yeast were then sub-cultured at a concentration of 1 x 10^7 cells/mL in SG-CAA (galactose containing), and induced at 25 °C for 48 hours (shaking at 250 rpm). Yeast was washed 1 time with 4 °C PBS-BSA, and incubated with biotinylated gankyrin and competitors (Pdar and p16INK4a in rounds 5 and 6, concentrations are provided below) in PBS-BSA for 1 hour at 25 °C. After 30 minutes, a FITC-conjugated anti-*myc* antibody was added at 1:250, and incubation was continued. The yeast was washed 1 time, and then incubated with streptavidin-R-phycoerythrin (SAPE) at 1:100 for 1 hour on ice. After a final wash, yeast were resuspended and double-positive (FITC for display and PE for binding) yeast were sorted into 5 mL SD-CAA using a MoFlo Flow Cytometer (Beckman Coulter, CSU Proteomics and Metabolomics Facility). Compensation was done to assure no overlap in FITC and PE emission. Sort gates were set to take ~1% of the healthy cell population, in order to maintain maximum stringency in both display and binding. Gates were set to enrich for cells that display the best (indicating protein stability) and bind the tightest. Sorted yeast were transferred to 50 mL pre-warmed SD-CAA and incubated at 30 °C for 3 days at 250 rpm. Plasmid DNA was recovered from the sorted yeast using a Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research). DNA was either used as template for the next round of diversification, or transformed into 5α *E. coli* in order to sequence library members.

Enzyme-Linked Immunosorbent Assay (ELISA) GBPs were cloned into a pET plasmid with N-terminal His_{6x}-tags and C-terminal FLAG-tags (DYKDDDDK) and purified as described previously. 100 μL of biotinylated gankyrin was incubated at 25 °C for 2 hours on streptavidin-

coated plates (Pierce) at a concentration of 10 μg/mL in HEPES buffer. Wells were washed 3 times with 200 μL wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT, 0.05 % Tween-20, 0.01 mg/mL BSA). 100 μL of each GBP variant was incubated at 25 nM for 1 hour at 25 °C, and then washed 3 times. An HRP-conjugated anti-FLAG antibody (Abcam) was incubated at 1:5000 in Odyssey blocking buffer (Li-Cor) for 1 hour at 25 °C, and then washed 3 times. Colorimetry was developed using TMB-One substrate and absorbance was measured at 655 nm on a plate reader (SynergyMx BioTek).

Lysate Ni-NTA Pull-down Assay Gankyrin was cloned into MCS1 of pETDuet using restriction enzymes BamHI and HindIII, resulting in N-terminal His_{6x}-tagged gankyrin. GBP7.19 was cloned into MCS2 of pETDuet using the restriction enzymes NdeI and PacI. Completed constructs were transformed into BL21s (DE3). Cells containing the co-expressed pair were inoculated and induced as described previously. Cells were spun down and resuspended in lysis buffer (20 mM sodium phosphate pH 7.4, 150 mM NaCl, 1 mM DTT), lysed by sonication, and spun down to remove cell debris. Cleared lysate was incubated with 100 μL Ni-NTA agarose resin for 45 minutes. Ni-NTA agarose was washed with 5 mL lysis buffer and 5 mL lysis buffer with 20 mM imidazole. Proteins were eluted with lysis buffer containing 400 mM imidazole. The pull-down was analyzed by SDS-PAGE and coomassie staining.

Isothermal Titration Calorimetry Isothermal titration calorimetry (ITC) was performed using a MicroCal iTC200 calorimeter maintained at 25 °C. All proteins were purified by standard methods, using a His_{6x} tag. Purified proteins were dialyzed extensively in phosphate buffer (20 mM sodium phosphate pH 7.4, 150 mM NaCl, 2.5 mM 2-mercaptoethanol). GBP variants were placed in the

sample cell at concentrations ranging from 10-15 μM, and 100-150 μM of gankyrin was titrated in 2.49 mL increments (16 injections total), with an initial injection of 0.4 μL, at 180 second intervals using a stirring speed of 750 rpm. Heats of dilution were measured in the same manner described above, separately titrating buffer into buffer and gankyrin into buffer. CoA-binding studies were performed by titrating 100 μM of gankyrin into a 1:1 mix of GBP7.17 and CoA (10 μM each), or by titrating 100 μM GBP7.17 into the sample cell containing 10 μM CoA. Displacement experiments were performed by titrating 100 μM FtsH-S6 ATPase into a pre-formed 1:1 complex of gankyrin and GBP7.19 (10 μM each). Data were analyzed using Origin7.0 (MicroCal, iTC200) using a one set of sites binding model for fitting. All data were reference subtracted by subtracting the mean heat of dilution from each data point.

p53 Ubiquitination Assay Gankyrin-dependent ubiquitination of p53 was obtained, by modifying a commercially available assay (MDM2/p53 Ubiquitination Kit, BostonBiochem). Briefly, 1 μM gankyrin was added to standard 30 μL reactions (contents include reaction buffer, p53, E1, E2, MDM2, Mg^{2+}-ATP), along with varying concentrations of GBP7.19. Ubiquitin was added to start the reactions, and run at 25 °C for 1 hour. The reactions were terminated with 8 μL loading dye and 2 μL 1M DTT, and ran on 10 % Ready Gel precast gels (Biorad) at 160 V for 1 hour. The gel was transferred to a PVDF membrane with the iBlot gel transfer station (Invitrogen). The membrane was blocked for 30 minutes with 5 % milk in PBS, and then incubated with provided α-p53 anitbody (1:2000) in Odyssey Blocking Buffer (Li-Cor) for 1 hour at 25 °C. The membrane was then washed 3 times with PBS-Tween, rinsed with PBS, then incubated with an Alexa Fluor 790-conjugated anti-mouse secondary antibody (Abcam) for 1 hour at 25 °C, followed by extensive washing. Membrane was imaged using an Odyssey CLx Near IR Scanner (Li-Cor), and

densitometry was performed for each lane using ImageJ (NIH). Statistical analysis was performed with GraphPad.

5.9 Proteins Used in This Work

Gankyrin

MEGCVSNIMICNLAYSGKLDELKERILADKSLATRTDQDSRTALHWACSAGHTEIVEFL
LQLGVPVNDKDDAGWSPLHIAASAGRDEIVKALLVKGAHVNAVNQNGCTPLHYAASK
NRHEIAVMLLEGGANPDAKDHYDATAMHRAAAKGNLKMVHILLFYKASTNIQDTEGN
TPLHLACDEERVEEAKFLVTQGASIYIENKEEKTPLQVAKGGLGLILKRLAEGEEASMGH
HHHHH*

Gankyrin-Avi

MEGCVSNIMICNLAYSGKLDELKERILADKSLATRTDQDSRTALHWACSAGHTEIVEFL
LQLGVPVNDKDDAGWSPLHIAASAGRDEIVKALLVKGAHVNAVNQNGCTPLHYAASK
NRHEIAVMLLEGGANPDAKDHYDATAMHRAAAKGNLKMVHILLFYKASTNIQDTEGN
TPLHLACDEERVEEAKFLVTQGASIYIENKEEKTPLQVAKGGLGLILKRLAEGEEASMGH
HHHHHGGSGLNDIFEAQKIEWHE*

GBP7

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPNYEEVLGRKCYPSVLDIPDKIEVVDLFVYPIWATEFVVYAIKKGAKVV
WFQYNTYHGEAGRQAKEAGLIIVANRCMMREHERLLGEK*

GBP7.5

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYKKIALVGASPKLERDANIVMKYLLE
HGYDVYPVNPYYEEVLGRKCYPSVLDIPDEIEVVGLFVYPIWATEFVVYAIKKGAKVV
WFQYNTYHEEAGRQAKEAGLIIVANRCMMCEHERLLGEK*

GBP7.7

MGSSHHHHHHSQDPGSTRPIDGLTDEDISEILSRYRKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPYYEEVLGRKCYPSVLDIPDKIEVVGLFVYPIWATEFVVYAIKKGAKVV
WFQYNTYHGEAGRQAKEAGLIIVANRCMMCEHERLLGEK*

GBP7.15

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYNKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPYYKEVLGRKCYPSVLDIPDKIEVVGLFVYPIWATEFVVHAIKKGAKVV
WFQYNTYHGVAGRQAKEAGLIIVANRCMMCEHERLLGEK*

GBP7.17

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPYYEEVLGRKCYPSVLDIPDKIEVVGLFVYPIWATEFVVCAIKKGAKVV
WFQHNTYRGEAGRQAKEAGLIIVANRCMMCEHERLLGEK*

GBP7.17 G79D

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPYYEEVLGRKCYPSVLDIPDKIEVVDLFVYPIWATEFVVCAIKKGAKVV
WFQHNTYRGEAGRQAKEAGLIIVANRCMMCEHERLLGEK*

GBP7.17 Y55N

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPNYEEVLGRKCYPSVLDIPDKIEVVGLFVYPIWATEFVVCAIKKGAKVV
WFQHNTYRGEAGRQAKEAGLIIVANRCMMCEHERLLGEK*

GBP7.17 C93Y

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPYYEEVLGRKCYPSVLDIPDKIEVVGLFVYPIWATEFVVYAIKKGAKVV
WFQHNTYRGEAGRQAKEAGLIIVANRCMMCEHERLLGEK*

GBP7.17 C132R

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPYYEEVLGRKCYPSVLDIPDKIEVVGLFVYPIWATEFVVCAIKKGAKVV
WFQHNTYRGEAGRQAKEAGLIIVANRCMMREHERLLGEK*

GBP7.17 R110H

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPYYEEVLGRKCYPSVLDIPDKIEVVGLFVYPIWATEFVVCAIKKGAKVV
WFQHNTYHGEAGRQAKEAGLIIVANRCMMREHERLLGEK*

GBP7.17 E112V

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPYYEEVLGRKCYPSVLDIPDKIEVVGLFVYPIWATEFVVCAIKKGAKVV
WFQHNTYRGVAGRQAKEAGLIIVANRCMMCEHERLLGEK*

GBP7.19 (GBP7.17 E57K)

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPYYKEVLGRKCYPSVLDIPDKIEVVGLFVYPIWATEFVVCAIKKGAKVV
WFQHNTYRGEAGRQAKEAGLIIVANRCMMCEHERLLGEK*

GBP7.17 E57K/E112V

MGSSHHHHHHSQDPGSTRPIDGLTDEDIREILTRYKKIALVGASPKPERDANIVMKYLLE
HGYDVYPVNPYYKEVLGRKCYPSVLDIPDKIEVVGLFVYPIWATEFVVCAIKKGAKVV
WFQHNTYRGVAGRQAKEAGLIIVANRCMMCEHERLLGEK*

Pdar

MASDLGKKLLEAAAAGQDDEVRILMANGADVNATDDDGLTPLHLAAANGQLEIVEVL
LKNGADVNASDSAGITPLHLAAYDGHLEIVEVLLKHGADVNAYDRAGWTPLHLAALSG
QLEIVEVLLKHGADVNAQDALGLTAFDISINQGQEDLAEILQGHHHHHH*

p16INK4a

MGSSHHHHHHSQDPGEPSADWLATAAARGRVEEVRALLEAGALPNAPNSYGRRPIQV
MMMGSARVAELLLLHGAEPNCADPATLTRPVHDAAREGFLDTLVVLHRAGARLDVRD
AWGRLPVDLAEELGHRDVARYLRAAAGGTRGSNHARIDAAEGPSDIPD*

FtsH-S6

MGSSHHHHHHSQDPLTEDQIKTTFADVAGCDEAKEEVAELVEYLREPSRFQKLGGKIPK
GVLMVGPPGTGKTLLAKAIAGEAKVPFFTISGSDFVEMFVGVGASRVRDMFEQAKKAA
PCIIFIDEIDAVGRQRGAGLGGGHDEREQTLNQMLVEMDGFEGNEGIIVIAATNRPDVLD
PALLRPGRFDRQVVVGLPDRRGKRQIFSTHTSKMNLSEEVDLEDYVARPDKISGADINSI
CQESGMLAVRENRYIVLAKDFEKAYKTVIKKDEQEHEFYK*

5.10 Primers Used in This Work

All DNA primers are listed 5' to 3'

Prb (GBP) FP: CGCGGATCCGGGCAGCACCCGTCCGATT

Prb (GBP) RP: CCTTAATTAATTATTTTTCGCCCAGCAGGCGTTC

GBP7.17 Y55N FP: GTCTATCCTGTAAACCCGAATTATGAAGAAGTGCTGG

GBP7.17 Y55N RP: CCAGCACTTCTTCATAATTCGGGTTTACAGGATAGAC

GBP7.17 C132R RP: CCTTAATTAATTATTTTTCGCCCAGCAGGCGTTCGTGCTCACGCA

TCATACAGCG

GBP7.17 G79D FP: GATAAAATTGAGGTCGTAGATCTGTTTGTGTACCCGATTTG

GBP7.17 G79D RP: CAAATCGGGTACACAAACAGATCTACGACCTCAATTTTATC

GBP7.17 R110H FP: GTTTCAGCACAACACATATCATGGGGAGGCTGGACGTCAAG

GBP7.17 R110H RP: CTTGACGTCCAGCCTCCCCATGATATGTGTTGTGCTGAAAC

GBP7.17 C93Y FP: CGACGGAGTTCGTTGTCTATGCCATCAAGAAAGGGGC

GBP7.17 C93Y RP: GCCCCTTTCTTGATGGCATAGACAACGAACTCCGTCG

GBP7.17 E112V FP: GCACAACACATATCGTGGGGTGGCTGGACGTCAAGCCAAGG

GBP7.17 E112V RP: CCTTGGCTTGACGTCCAGCCACCCCACGATATGTGTTGTGC

GBP7.17 E57K FP: CCTGTAAACCCGTATTATAAAGAAGTGCTGGGCCGCAAG

GBP7.17 E57K RP: CTTGCGGCCCAGCACTTCTTTATAATACGGGTTTACAGG

Gankyrin FP NcoI: CATGCCATGGAGGGGTGTGTGTCTAACATAATGATCTGTAACC

Gankyrin RP PacI: CCTTAATTAATTAGTGATGGTGGTGGTGATGACCCATAGAAGCCT

CTTCACCTTCTG

HR FP NheI: CTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTCGGCTAGC

HR RP BamHI: CGAGCTATTACAAGTCCTCTTCAGAAATCAGCTTTTGTTCGGATCC

5.11 Supplemental Data

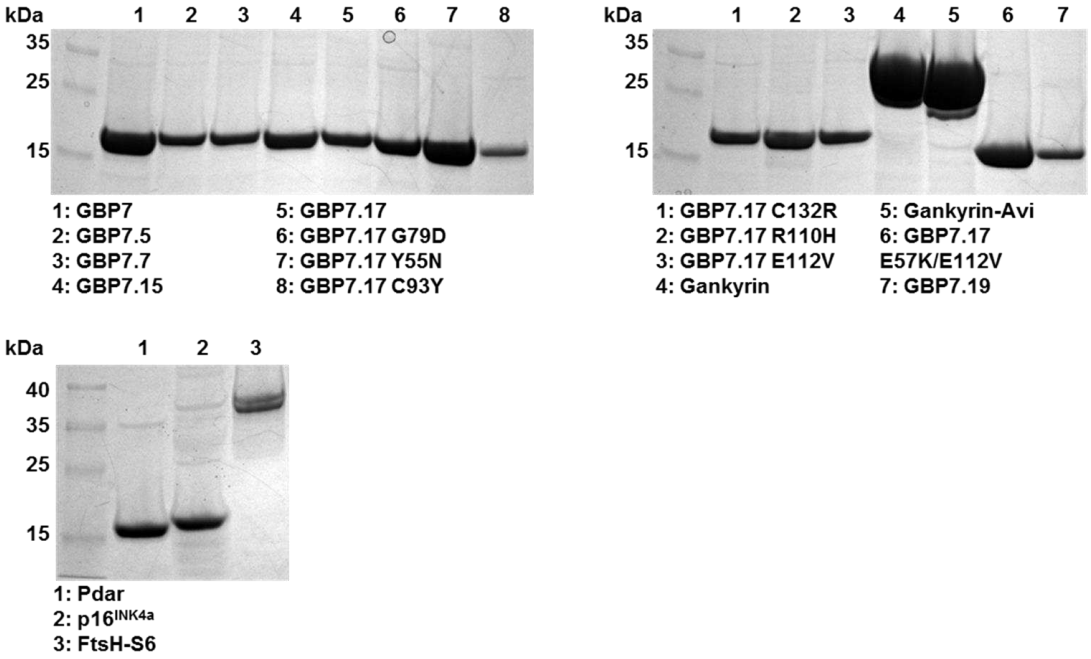


Figure S5.1 SDS-PAGE and coomassie staining of purified proteins used for ITC, yeast display, and p53 ubiquitination assays.

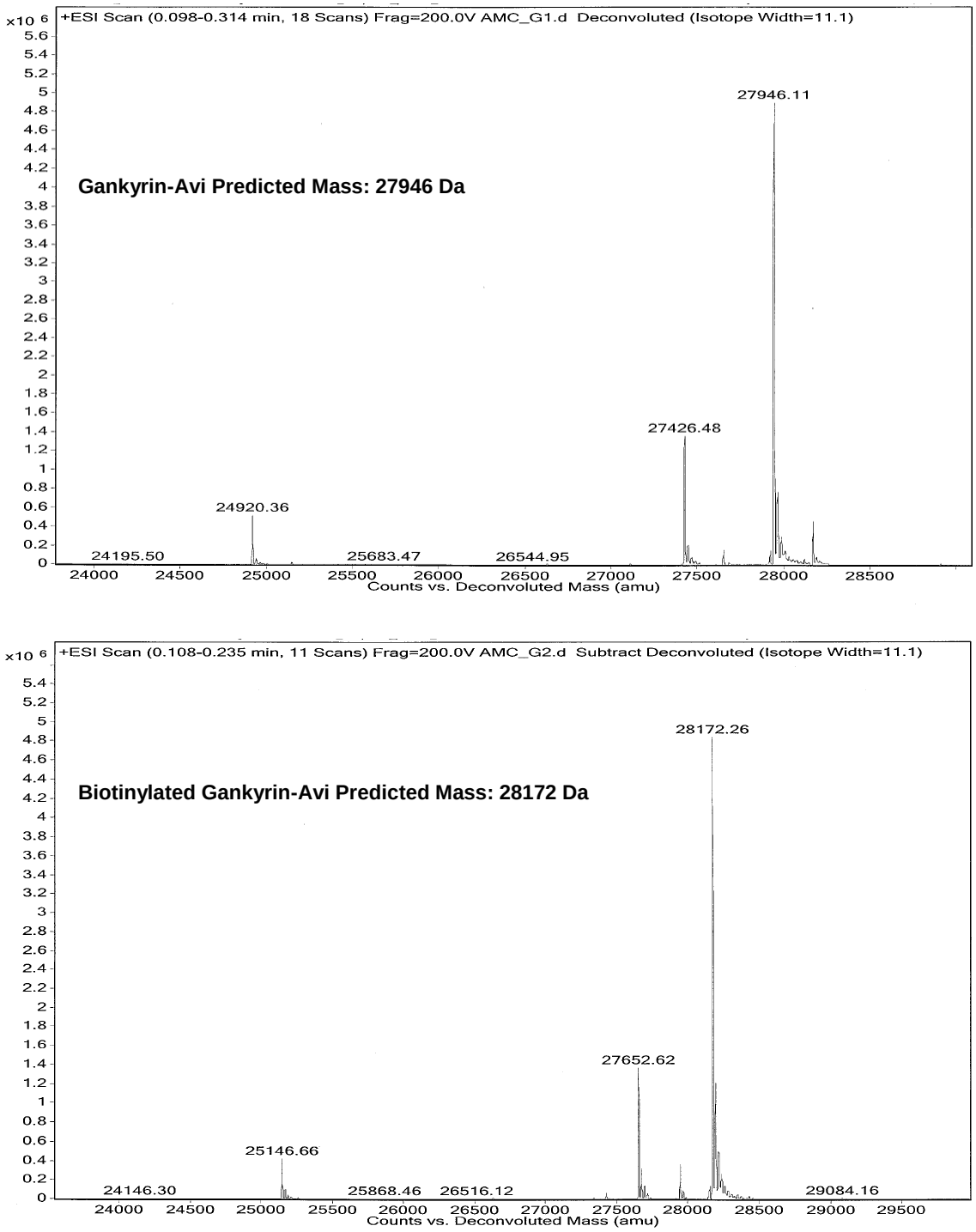


Figure S5.2 Confirmation of gankyrin-Avi biotinylation by mass spectrometry.

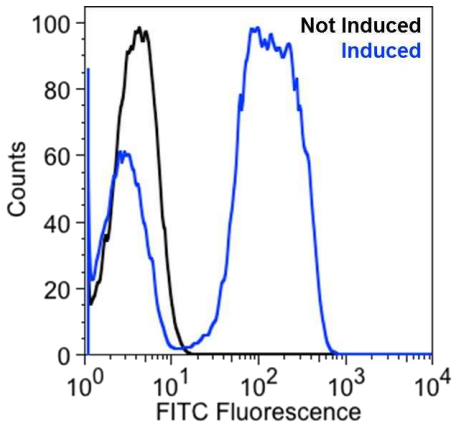


Figure S5.3 GBP7 with a C-terminal *myc*-tag displays efficiently on EBY100 yeast (~75% display), as confirmed using FITC-conjugated anti-*myc* antibody.

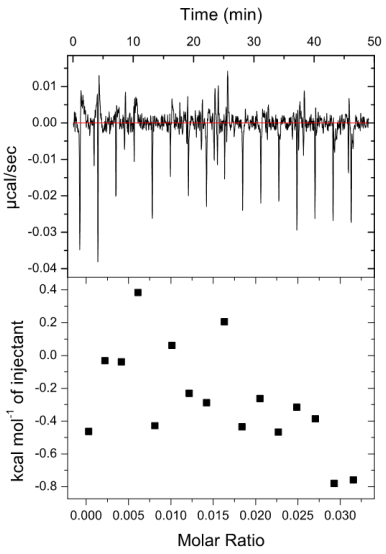


Figure S5.4 ITC evaluation of gankyrin titrated into buffer.

Table S5.1 Yeast display FACS parameters for each successive round.

	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6
[gankyrin, nM]	1000	500	100	50	50	5
[Pdar, nM]	---	---	---	---	1000	1000
[p16INK4a, nM]	---	---	---	---	1000	1000
Diversification	ep-PCR	ep-PCR	ep-PCR	ep-PCR	---	DNA Shuffling
# of Yeast Screened	3.6×10^7	1.0×10^7	1.5×10^7	9.5×10^6	1.2×10^6	1.0×10^6
# of Yeast Sorted	3.2×10^5	7.1×10^4	5.0×10^4	3.5×10^4	1.8×10^5	1.0×10^4

Table S5.2 Full thermodynamic profiles by ITC of gankyrin-binding for all GBPs studied.[a]

Entry	GBP Variant	K_D (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	-TΔS (kcal/mol)	N-value
1	GBP7.5	138.7 (±16.3)	-9.1 (±0.3)	-12.5 (±1.3)	3.4 (±1.0)	1.00 (±0.04)
2	GBP7.7	125.3 (±12.7)	-8.9 (±0.3)	-11.3 (±1.9)	2.4 (±0.7)	0.93 (±0.02)
3	GBP7.15	61.8 (±5.1)	-9.8 (±0.0)	-8.9 (±0.2)	-0.9 (±0.2)	0.94 (±0.00)
4	GBP7.17	41.8 (±8.0)	-10.0 (±0.1)	-13.5 (±0.5)	3.5 (±0.5)	0.94 (±0.01)
5	GBP7.17 C132R	117.0 (±18.0)	-9.5 (±0.1)	-7.9 (±0.3)	-1.6 (±0.2)	0.94 (±0.01)
6	GBP7.17 Y55N	945.3 (±125.3)	-8.1 (±0.1)	-18.2 (±0.5)	10.0 (±0.5)	0.94 (±0.04)
7	GBP7.17 C93Y	74.0 (±7.8)	-9.8 (±0.1)	-13.1 (±0.2)	3.4 (±0.2)	1.00 (±0.01)
8	GBP7.17 R110H	65.0 (±5.3)	-9.8 (±0.0)	-11.4 (±0.2)	1.6 (±0.2)	0.92 (±0.01)
9	GBP7.17 G79D	39.0 (±3.6)	-9.6 (±0.2)	-14.6 (±0.9)	5.1 (±0.8)	0.93 (±0.03)
10	GBP7.17 E112V	38.2 (±1.9)	-10.1 (±0.1)	-13.1 (±0.2)	2.9 (±0.2)	0.95 (±0.02)
11	GBP7.17 E57K/E112V	21.6 (±2.4)	-10.3 (±0.1)	-13.5 (±0.6)	3.2 (±0.6)	0.94 (±0.01)
12	GBP7.19 (GBP7.17 E57K)	20.9 (±2.5)	-10.5 (±0.1)	-13.8 (±0.3)	3.3 (±0.2)	0.95 (±0.03)

[a]All errors represent the standard deviation of three separate experiments. ITC conditions were as follows: 20 mM sodium phosphate, 150 mM NaCl, and 2.5 mM 2-mercaptoethanol (pH 7.4) at 25 °C.

REFERENCES

(1) Karanicolas, J.; Corn, J. E.; Chen, I.; Joachimiak, L. a; Dym, O.; Peck, S. H.; Albeck, S.; Unger, T.; Hu, W.; Liu, G.; Delbecq, S.; Montelione, G. T.; Spiegel, C. P.; Liu, D. R.; Baker, D. *Mol. Cell* **2011**, *42* (2), 250.

(2) Blakeley, B. D.; Chapman, A. M.; McNaughton, B. R. *Mol. Biosyst.* **2012**, *8* (8), 2036.

(3) Zhen, C.; Chen, L.; Zhao, Q.; Liang, B.; Gu, Y.-X.; Bai, Z.-F.; Wang, K.; Xu, X.; Han, Q.; Fang, D.; Wang, S.; Zhou, T.; Xia, Q.; Gong, W.-L.; Wang, N.; Li, H.-Y.; Jin, B.-F.; Man, J. *Oncogene* **2013**, *32* (29), 3452.

(4) Kim, Y. H.; Kim, J. H.; Choi, Y. W.; Lim, S. K.; Yim, H.; Kang, S. Y.; Chung, Y. S.; Lee, G. Y.; Park, T. J. *Exp. Mol. Pathol.* **2013**, *94* (2), 360.

(5) Fu, X.-Y.; Wang, H.-Y.; Tan, L.; Liu, S.-Q.; Cao, H.-F.; Wu, M.-C. *World J. Gastroenterol.* **2002**, *8* (4), 638.

(6) Li, J.; Knobloch, T. J.; Kresty, L. a; Zhang, Z.; Lang, J. C.; Schuller, D. E.; Weghorst, C. M. *Anticancer Res.* **2011**, *31* (9), 2683.

(7) Meng, Y.; He, L.; Guo, X.; Tang, S.; Zhao, X.; Du, R.; Jin, J.; Bi, Q.; Li, H.; Nie, Y.; Liu, J.; Fan, D. *Cancer Lett.* **2010**, *297* (1), 9.

(8) Tang, S.; Yang, G.; Meng, Y.; Du, R.; Li, X.; Fan, R.; Zhao, L.; Bi, Q.; Jin, J.; Gao, L.; Zhang, L.; Li, H.; Fan, M.; Wang, Y.; Wu, K.; Liu, J.; Fan, D. *Cancer Biol. Ther.* **2010**, *9* (2).

(9) Man, J. H.; Liang, B.; Gu, Y. X.; Zhou, T.; Li, A. L.; Li, T.; Jin, B. F.; Bai, B.; Zhang, H. Y.; Zhang, W. N.; Li, W. H.; Gong, W. L.; Li, H. Y.; Zhang, X. M. *J. Clin. Invest.* **2010**, *120* (8), 2829.

(10) Li, J.; Mahajan, A.; Tsai, M.-D. *Biochemistry* **2006**, *45* (51), 15168.

(11) Li, J.; Tsai, M.-D. *Biochemistry* **2002**, *41* (12), 3977.

(12) Dawson, S.; Apcher, S.; Mee, M.; Higashitsuji, H.; Baker, R.; Uhle, S.; Dubiel, W.; Fujita, J.; Mayer, R. J. *J. Biol. Chem.* **2002**, *277* (13), 10893.

(13) Higashitsuji, H.; Higashitsuji, H.; Itoh, K.; Sakurai, T.; Nagao, T.; Sumitomo, Y.; Sumitomo, H.; Masuda, T.; Dawson, S.; Shimada, Y.; Mayer, R. J.; Fujita, J. *Cancer Cell* **2005**, *8* (1), 75.

(14) Higashitsuji, H.; Liu, Y.; Mayer, R. J.; Fujita, J. **2005**, No. October, 1335.

(15) Kaneko, T.; Hamazaki, J.; Iemura, S. ichiro; Sasaki, K.; Furuyama, K.; Natsume, T.; Tanaka, K.; Murata, S. *Cell* **2009**, *137* (5), 914.

(16) Nakamura, Y.; Nakano, K.; Umehara, T.; Kimura, M.; Hayashizaki, Y.; Tanaka, A.; Horikoshi, M.; Padmanabhan, B.; Yokoyama, S. *Structure* **2007**, *15* (2), 179.

(17) Nanaware, P. P.; Ramteke, M. P.; Somavarapu, A. K.; Venkatraman, P. *Proteins Struct. Funct. Bioinforma.* **2014**, *82* (7), 1283.

(18) Chapman, A. M.; McNaughton, B. R. *ACS Chem. Biol.* **2014**, *9*, 2223.

(19) Michnick, S. W.; Ear, P. H.; Manderson, E. N.; Remy, I.; Stefan, E. *Nat. Rev. Drug Discov.* **2007**, *6* (7), 569.

(20) Shekhawat, S. S.; Ghosh, I. *Curr. Opin. Chem. Biol.* **2011**, *15* (6), 789.

(21) Boder, E. T.; Wittrup, K. D. *Nat. Biotechnol.* **1997**, *15* (6), 553.

(22) Kehoe, J.; Kay, B. *Chem. Rev.* **2005**, *3824*, 4056.

(23) Lipovsek, D.; Pluckthun, A. *J. Immunol. Methods* **2004**, *290* (1-2), 51.

(24) Nixon, A. E.; Sexton, D. J.; Ladner, R. C. *MAbs* **2014**, *6* (1), 73.

(25) Chao, G.; Lau, W. L.; Hackel, B. J.; Sazinsky, S. L.; Lippow, S. M.; Wittrup, K. D. *Nat.*

Protoc. **2006**, *1* (2), 755.

(26) Cadwell, R. C.; Joyce, G. F. *PCR Methods Appl.* **1992**, *2*, 28.

(27) Stemmer, W. P. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (22), 10747.

(28) Beckett, D.; Kovaleva, E.; Schatz, P. J. *Protein Sci.* **1999**, *8* (4), 921.

(29) Hiyama, T. B.; Zhao, M.; Kitago, Y.; Yao, M.; Sekine, S.-I.; Terada, T.; Kuroishi, C.; Liu, Z.-J.; Rose, J. P.; Kuramitsu, S.; Shirouzu, M.; Watanabe, N.; Yokoyama, S.; Tanaka, I.; Wang, B.-C. *J. Struct. Funct. Genomics* **2006**, *7* (3-4), 119.

ABBREVIATIONS

Aga	agglutinin protein
Ala/A	alanine
APC	antigen-presenting cell
Arg/R	arginine
Asn/N	asparagine
Asp/D	aspartic acid
ATP	adenosine triphosphate
AvNAPSA	average neighboring atoms per side chain atom
BLIP	β-lactamase inhibitor protein
bp	base pair
BSA	bovine serum albumin
CaM	calmodulin
CD	circular dichroism
CDK4	cyclin-dependent kinase 4
CDR	complementary determining region
Cfu	colony forming unit
CoA	coenzyme A
Cys/C	cysteine
Da	Dalton
DAPI	4′,6-diamidino-2-phenylindole
DHFR	dihydrofolate reductase

DLS	dynamic light scattering
DNA	deoxyribonucleic acid
DSC	differential scanning calorimetry
DTT	dithiothreitol
ELISA	enzyme-linked immunosorbent assay
ep-PCR	error-prone polymerase chain reaction
FACS	fluorescence-activated cell sorting
FDA	Food and Drug Administration
FITC	fluorescein isothiocyanate
frGFP	folding reporter green fluorescent protein
GBP	gankyrin-binding protein
GFP	green fluorescent protein
Gln/Q	glutamine
Glu/E	glutamic acid
Gly/G	glycine
HEL	hen egg lysozyme
hGH	human growth hormone
His/H	histidine
HIV	human immunodeficiency virus
HRP	horseradish peroxidase
HYP	human hyperplastic discs protein
Ile/I	isoleucine
IPTG	β-D-1-thiogalactopyranoside

ITC	isothermal titration calorimetry
K_D	dissociation constant
Leu/L	leucine
Lys/K	lysine
MCS	multiple cloning site
MDM2	murine double minute 2
Met/M	methionine
MHC	major histocompatibility complex
MRSA	methicillin-resistant *Staphylococcus aureus*
NHR	N-heptad repeat
NTA	nitrilotriacetic acid
NTF3	neurotrophin-3
PAGE	polyacrylamide gel electrophoresis
PAK1	p21-activated kinase 1
PBS	phosphate-buffered saline
PCA	protein-fragment complementation assay
PCR	polymerase chain reaction
PDB	protein data bank
PE	phycoerythrin
PEG	polyethylene glycol
PH Domain	pleckstrin homology domain
Phe/F	phenylalanine
PPI	protein-protein interaction

Pro/P	proline
pRb	phosphorylated retinoblastoma protein
RT	room temperature
RPM	revolutions per minute
SAPE	streptavidin-R-phycoerythrin
SDS	sodium dodecyl sulfate
Ser/S	serine
sfGFP	superfolder green fluorescent protein
SHV-1	sulfhydryl variable-1
smMLCK	smooth muscle myosin light chain kinase
spGFP	superpositive green fluorescent protein
TALEN	transcription activator-like nucleases
TAT	trans-activating transcriptional activator
TEM-1	temoneira-1
Thr/T	threonine
TMB	3, 3′, 5, 5′-tetramethylbenzidine
Trp/W	tryptophan
Tyr/Y	tyrosine
Val/V	valine