High-Throughput Characterization of Protein-Protein Interactions by Reprogramming Yeast Mating

David A. Younger

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

David Baker, Chair

Eric Klavins, Chair

Wendy Thomas

Program Authorized to Offer Degree:
Bioengineering

University of Washington

Abstract

High-Throughput Characterization of Protein-Protein Interactions by Reprogramming Yeast Mating

David A. Younger

Chairs of the Supervisory Committee:

Professor David Baker
Department of Biochemistry

Professor Eric Klavins
Department of Electrical Engineering

High-throughput methods for screening protein-protein interactions enable the rapid characterization of engineered binding proteins and interaction networks. While existing approaches are powerful, none allow quantitative library-on-library characterization of protein interactions in a modifiable extracellular environment. Here, we show that sexual agglutination of *S. cerevisiae* can be reprogrammed to link interaction strength with mating efficiency using synthetic agglutination (SynAg). Validation of SynAg with 89 previously characterized interactions shows a log-linear relationship between mating efficiency and protein binding strength for interactions with K_D's ranging from below 500 pM to above 300 μM. Using induced chromosomal translocation to pair barcodes representing binding proteins, thousands of distinct interactions can be screened in a single pot. We demonstrate the ability to characterize protein interaction networks in a modifiable environment by introducing a soluble peptide that selectively disrupts a subset of interactions in a representative network by up to 800-fold. SynAg enables the high-throughput, quantitative characterization of protein-protein interaction networks in a fully-defined extracellular environment at a library-on-library scale.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This work would not have been possible without the support of many brilliant, thoughtful, patient, and generous people. I would first like to thank my advisors, Eric and David, for allowing me to learn from them and for giving me the freedom to work on a project of my choosing. My confidence as a researcher today is largely due to their constant trust and encouragement. Among countless other things, Eric has taught me the value of setting ambitious goals, whether they be scientific, business, or at the climbing gym. David has taught me the importance of combining focus with flexibility in order to pursue challenging problems most effectively. I would also like to thank my committee members, Maitreya Dunham, Barry Lutz, and Wendy Thomas, for their thoughtful advice and encouragement. In particular, I want to thank Maitreya for sharing her yeast genetics knowledge, which was critical for shaping the initial concept of synthetic agglutination.

When I began my graduate work, I had no hands-on molecular biology experience. I was able to make progress in the lab because of the patience and generosity of an incredible group of scientific mentors who provided me with all of the necessary tools: Possu Huang, Kyle Havens, Nick Bolten, Miles Gander, and Ben Groves. I also want to extend an enormous thank you to Stephanie Berger and Cassie Bryan, who I worked with on the synthetic agglutination and secretion capture projects, respectively. Both Stephanie and Cassie taught me that working with incredible and dedicated people is far more enjoyable and fulfilling than trying to do everything alone. Throughout graduate school, I had the opportunity to learn from other extraordinary colleagues (and friends) who made me excited to go to lab each day: Laura Adam, Tileli Amimeur, Leandra Brettner, Alberto Carignano, Yuan Chen, Arjun Khakhar, Marc Lajoie, Randolph Lopez, Sergii Pochekailov, Sunny Rao, Stephen Rettie, Alex Rosenberg, Chris Takahashi, Willy Voje, Justin Vrana, and Yaoyu Yang. I especially want to thank Michelle Parks for holding the Klavins lab together over the past many years as well as Cami Cordray and the other phenomenal people who keep everything running smoothly.

I had no idea what I was getting myself into when I started graduate school, but I certainly did not imagine it would be so much fun. I am incredibly thankful for my amazing friends who have kept me sane despite the stress and provided a constant reminder to maintain a

work-life balance. In particular, I want to thank Anna Blakney, Blake Bluestein, Bob Lamm, Jason Stevens, Jon Tsui, and Andrew Wang.

Finally, I want to thank my family for their never-ending encouragement and support. Returning to Seattle for graduate school allowed me to spend more time with my parents, which has been a highlight of the past five years. As the years go by, I realize more and more just how amazing they are and how much they have influenced me to become a better and stronger person. I also want to thank my older brother, Noah, who has been an invaluable role model and friend. Without my incredible family, this work would not have been possible.

Chapter 1
INTRODUCTION

1.1 TOP-DOWN SYNTHETIC BIOLOGY

Biological organisms are capable of extraordinary behaviors, including many that can provide substantial utility for academic or industrial applications. However, gaps in understanding of even the most basic biological phenomenon make engineering biological systems challenging. A common synthetic biology workflow involves selecting a model organism or cell strain and adding genetic parts until a desired behavior is achieved. The cell "background" provides all of the components for basic cellular function and growth, but is viewed as a blank canvas into which genetic parts are added. This approach can only be used when the underlying biology of the desired system is sufficiently understood. Successful applications of this approach include synthetic gene regulatory networks made from host-orthogonal transcription factors and simple metabolic pathways transferred from one organism to another.

An alternative approach is to start with a fully functional complex biological phenomenon and rework a small and well-characterized component to generate a useful tool. This strategy enables the engineering of a greater complexity and diversity of cellular behaviors by not requiring a complete understanding of the underlying biology. For example, metabolic engineers commonly choose a host strain that produces a close precursor of a desired product. A current limitation with this approach is that many organisms are not easily engineerable. However, even among model organisms, there exist an enormous number of behaviors and metabolic processes that can be engineered for useful applications. Here, we present a new top-down engineering challenge. Our goal is to build a synthetic biological assay for the multiplexed characterization of protein-protein interactions in an extracellular environment. Building such a system from the ground up is far beyond the capabilities of modern synthetic biology, but fortunately there is an existing natural system in an easily engineerable model organism that can be repurposed: sexual agglutination of *Saccharomyces cerevisiae*.

1.2 BIOLOGICAL BACKGROUND OF YEAST AGGLUTINATION AND MATING

The life cycle of budding yeast includes asexual reproduction of both haploid and diploid forms. Haploid *S. cerevisiae* cells consist of two mating types, MATa and MATα, which fuse to form a single diploid cell, a process called mating.

Mating of *S. cerevisiae* in an aerated liquid culture depends critically on an intercellular protein-protein interaction that drives agglutination between MATa and MATα haploid cells[1,2]. In response to the detection of mating factor secreted from a cell of the opposite mating type, haploid cells begin to express mating type specific sexual agglutinin proteins[3]. Prior to induction, both the a- and α-agglutinins are present on the cell surface at $0\text{-}10^4$ molecules per cell. In the presence of the appropriate mating factor, expression of the sexual agglutinin proteins on the yeasts' surface increases to $10^4\text{-}10^5$ molecules per cell[2]. The sexual agglutinin proteins consist of the MATa sexual agglutinin subunits, Aga1 and Aga2, and the MATα sexual agglutinin, Sag1. Aga1 and Sag1 form glycosylposphatidylinositol (GPI) anchors with the yeast cell wall and extend into the extracellular space with a glycosylated stalk. Aga2 is secreted by MATa cells and forms a disulfide bond with Aga1 (Figure 1.1)[4]. The interaction between Aga2 and Sag1 is strong (K_D = 2-5nM) and multiple interactions occur between pairs of haploid cells, resulting in irreversible binding even when strong shear forces from a turbulent liquid culture are exerted[5]. Cellular agglutination and mating is highly efficient, occurs in a matter of hours, and each mating event forms a stable and propagating diploid strain[6].

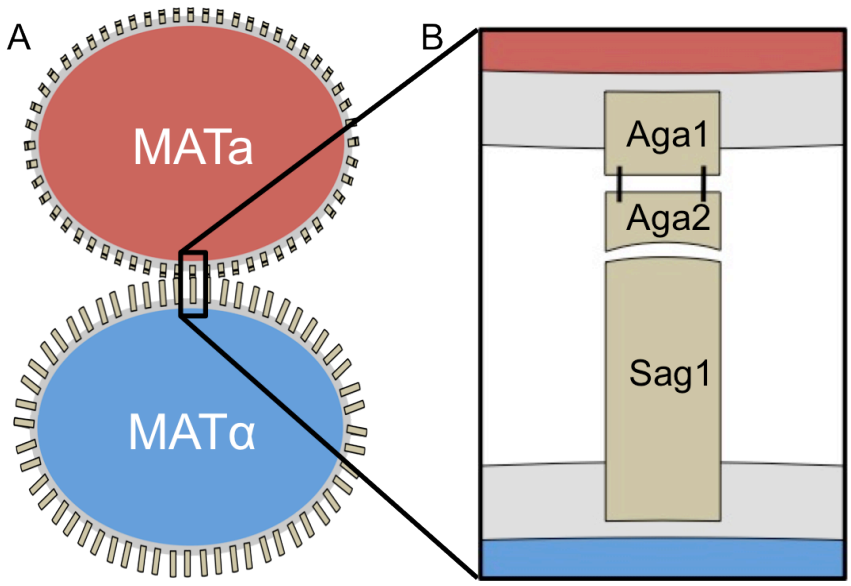


Figure 1.1: A cartoon depiction of yeast sexual agglutination. **(**A**)** In a turbulent liquid culture, MATa and MATα haploid cells adhere to one another due to the highly avid interaction of proteins expressed on their surfaces. **(**B**)** The sexual agglutination interaction occurs between the MATa sexual agglutinin subunit, Aga2, and the MATα sexual agglutinin, Sag1. Aga2 is attached to the yeast surface with a disulfide bond to Aga1. Both Aga1 and Sag1 are GPI-anchored to the yeast cell wall and extend into the extracellular space with a glycosylated stalk.

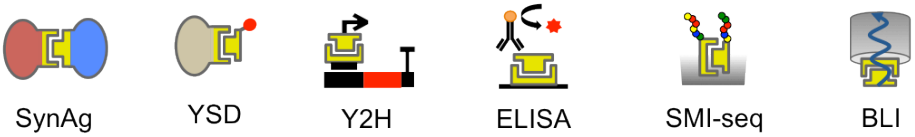
The role of the sexual agglutinin proteins is limited to cell adhesion. Both Aga2 and Sag1 are essential for mating in liquid culture. However, when grown on solid culture, mating efficiency is unaffected by the knockout of the sexual agglutinin proteins[7,8]. Additionally, since they are cell wall anchored, Aga2 and Sag1 cannot directly transduce any signals into the cell upon binding[2]. Their simple and well-defined behavior makes the sexual agglutinin proteins ideal candidates for engineering. Furthermore, the sexual agglutinin proteins exhibit a far greater diversity across yeast species than most proteins[9]. A lack of conservation is indicative of a large possible design space, which is another favorable trait for engineering.

Yeast mating is an enormously complex biological process. Despite being one of the best-studied biological phenomenon in one of the best-studied model organisms, much of the underlying biology is not fully understood. For example, the underlying biological mechanics and regulatory elements involved in cellular fusion during mating are still being investigated[10]. Engineering a process as complex as cellular fusion from the bottom-up is a distant goal for

synthetic biologists. However, with a top-down engineering approach, the complexity of yeast agglutination and mating can be reduced to the simplicity of a single protein-protein interaction.

1.3 EXISTING APPROACHES FOR PROTEIN INTERACTION CHARACTERIZATION

Many powerful methods have been developed for the high-throughput screening of protein-protein interactions (Figure 1.2). Phage display[11] and yeast surface display[12] (YSD) have enabled the high-throughput binding characterization of large protein libraries, but can only screen binding against a limited number of targets due to the spectral resolution of existing fluorescent reporters[13]. Both approaches also require the expression and purification of recombinant target proteins, which is limiting if the target is unstable or expresses poorly. Yeast two-hybrid (Y2H)[14] can be used to intracellularly screen pairwise protein interactions and has been extended to the screening of large interaction networks using next generation sequencing[15–17]. However, intracellular assays are limited by an inability to control the binding environment and suffer from frequent false-positives and false-negatives[18,19]. Enzyme linked immunosorbent assays (ELISAs) can be used for quantitative protein interaction characterization, but require purified proteins, costly reagents, and a separate analysis of each protein interaction. SMI-seq can be used to characterize whole protein-protein interaction networks in a cell-free environment, but requires the use of purified proteins and a dedicated flow cell for the analysis of each network and condition[20]. Finally, biolayer interferometry (BLI)[21] produces accurate affinity measurements, but cannot be multiplexed, making it slow and expensive for the characterization of large protein interaction networks. While each approach expands screening capabilities, only synthetic agglutination allows for cell-based, quantitative, library-on-library protein interaction characterization in an extracellular environment that can be modified as desired. This makes yeast synthetic agglutination (SynAg) uniquely suited for many academic and industrial applications.



	SynAg	YSD	Y2H	ELISA	SMI-seq	BLI
One pot / library-on-library	✓	✗	✓	✗	✓	✗
Parallelizable	✓	✓	✓	✓	✗	✗
Controllable environment	✓	✓	✗	✓	✓	✓
Quantitative	✓	✓	✗	✓	✓	✓
Easy to iterate	✓	✓	✓	✗	✗	✗
Cost / equipment needs	Low	Low	Low	Low	High	High

Figure 1.2: A comparison of synthetic agglutination (SynAg) with other approaches for characterizing protein-protein interactions. Synthetic agglutination is the only quantitative and parallelizable approach for one-pot library-on-library screening.

1.4 Overview of Thesis Contributions

This dissertation presents a novel synthetic biology platform for the quantitative and multiplexed characterization of protein-protein interactions using yeast synthetic agglutination. We have demonstrated that by replacing the native sexual agglutinin proteins, Aga2 and Sag1, with arbitrary proteins expressed on the surface of MATa and MATα haploid cells, we can recover mating efficiency in a turbulent liquid culture. This finding strongly supports the hypothesis that the role of the sexual agglutinin proteins is limited to binding and provides a novel platform for controlling yeast mating efficiencies, which can be used to study the biology of yeast mating and reproductive ecology.

Furthermore, we can quantitatively determine the affinity of a protein pair by measuring mating efficiency. Specifically, we see a strong log-linear relationship (R^2 = 0.89) between mating efficiency and affinity for protein pairs with K_D's ranging from below 500 pM to above 300 μM. As a test case, we use natural and engineered proteins and peptides from the BCL2 family of apoptosis regulating proteins that have been fully characterized with biolayer interferometry. The tested proteins range from 26 to 206 amino acids, indicating a large engineerable space for synthetic agglutination.

We then extend the assay for the multiplexed characterization of protein interactions by incorporating a recombination site and barcode flanking each surface expression cassette. Following a batched mating containing MATa and MATα libraries, a chromosomal translocation is induced to pair the barcodes from each haploid cassette onto the same chromosome and next generation sequencing is used to count the frequency of each pair, which is indicative of interaction strength. We find that there is again a strong log-linear relationship (R^2 = 0.87) between diploid frequency and affinity for protein pairs with K_D's ranging from below 500 pM to above 300 μM. This approach is validated with the characterization of site-saturation mutagenesis libraries, consisting of 7,000 possible protein-protein interactions, in a single tube.

One exciting industrial application of yeast synthetic agglutination is for the characterization of compounds that inhibit protein-protein interactions. Instead of building and characterizing an unknown protein interaction network, this involves the recapitulation of known networks in the presence of an uncharacterized soluble compound. By comparing the network in the presence and absence of the compound, a quantitative protein interaction disruption profile can rapidly be generated for each pairwise interaction within the network. This platform may enable the multiplexed characterization of drug off-target effects on protein interactions. As a proof of concept, we have tested the effect of adding a fully characterized pro-apoptotic BH3 peptide to an interaction network consisting of pro-survival BCL2 homologues and *de novo* binding proteins. The disruptive effects on protein interactions in this network perfectly match the expectation based on the binding profile of the pro-apoptotic peptide.

Throughout the development of the yeast synthetic agglutination platform, our primary goal was to develop a functional tool that can be used to solve real world problems in academic and industrial settings. To that effect, we aimed to develop a very simple workflow that can be adapted to a wide variety of applications. Specifically, we have two main workflows: one for characterizing a single protein interaction with a flow cytometry output and the other for characterizing whole library-on-library protein interaction networks using a next generation sequencing output. Both workflows consist of three steps: strain construction, mating assay, and data analysis (Figure 1.3). Detailed information about both workflows is provided in subsequent chapters.

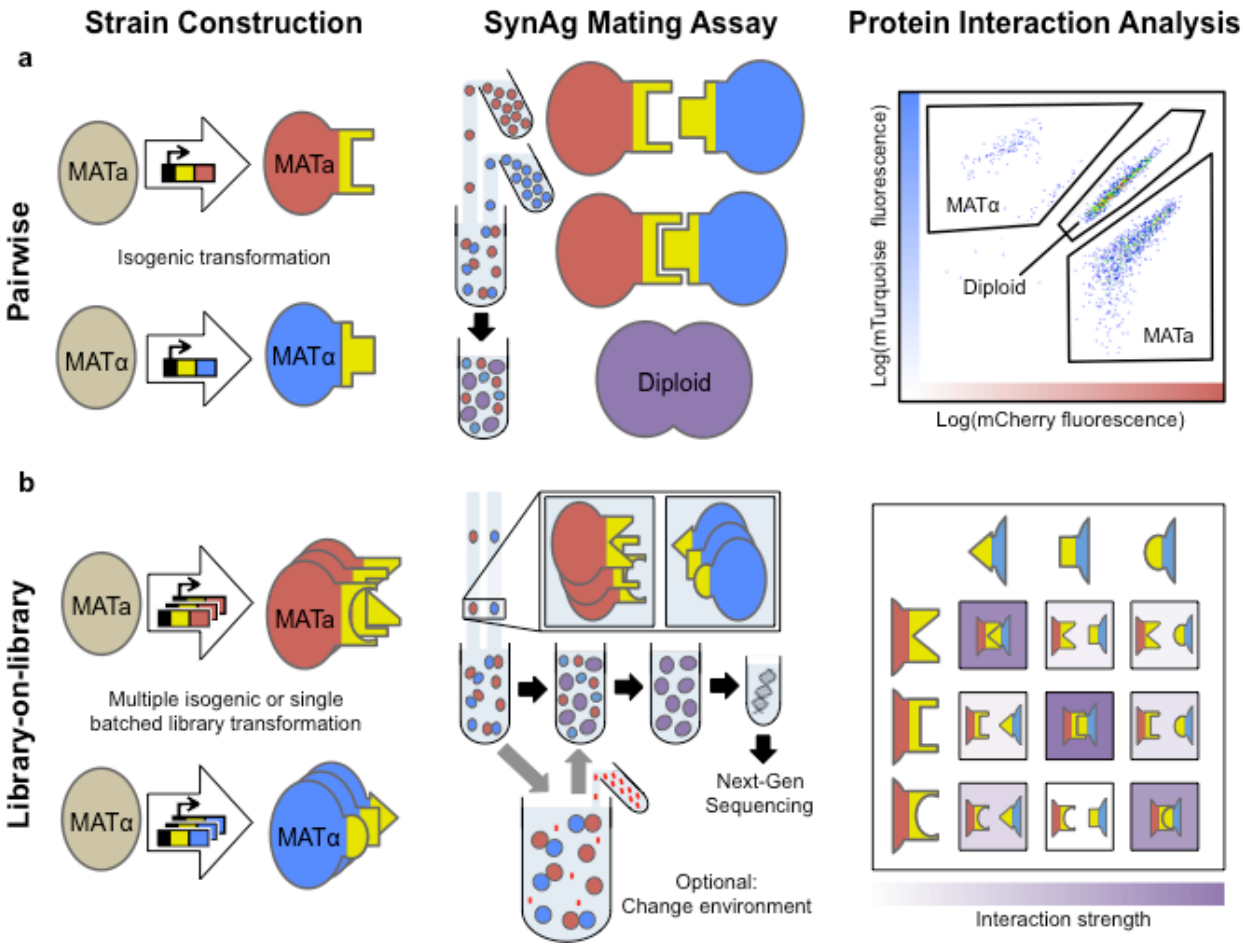


Figure 1.3: Yeast synthetic agglutination workflow. (A) Pairwise protein interaction characterization involves isogenic yeast transformations followed by mating and flow cytometry. (B) Library-on-library protein interaction network characterization involves strain library construction followed by mating, diploid DNA isolation, and NGS. Manipulation of the mating environment prior to mating is optional for the characterization of protein interactions that respond to environmental changes.

While the application and final system is novel, yeast synthetic agglutination is built from a strong foundation of previous work. The expression of arbitrary binding proteins on the surface of haploid cells is accomplished with yeast surface display[12]. By utilizing an extremely well validated approach, we can be confident that synthetic yeast agglutination can be applied with a wide diversity of properly folded proteins including single chain antibodies, enzymes, growth factors, and cell-surface receptors[22–25]. The CRE/lox recombination system is used for controlled chromosomal translocation in order to combine barcodes from two different haploids onto the

same chromosome[26]. Again, we benefit greatly from previous work that demonstrated biased lox-site variants for the maximization of translocation efficiencies[27]. In order to build large yeast libraries with integrated cassettes, we adapted a nuclease assisted chromosomal integration method to damage the DNA at the integration site. This approach was previously used to improve yeast transformation efficiencies by many orders of magnitude[28].

Chapter 2

REPROGRAMMING SEXUAL AGGLUTINATION

2.1 INTRODUCTION

Sexual agglutination is a natural process by which *S. cerevisiae* haploid cells adhere to one another in a turbulent liquid culture in order to mate. This process is governed by a single protein interaction between Aga2, expressed on the surface of MATa haploid cells, and Sag1, expressed on the surface of MATα haploid cells. Here we replace the native Aga2-Sag1 interaction with arbitrary binding proteins to demonstrate that sexual agglutination can be reprogrammed to characterize protein interactions. We begin by confirming that both sexual agglutinin proteins are essential for wild type yeast to mate in liquid culture and that mating efficiency is highly dependent on nutrient availability and carbon source. These characteristics of yeast agglutination and mating informed the design of an assay to measure mating efficiency with the surface display of arbitrary protein binding pairs. We then accurately distinguish between binders and non-binders using a synthetic agglutination assay.

2.2 PARENT STRAINS AND PLASMIDS FOR YEAST SURFACE DISPLAY

Genetic differences between common laboratory yeast strains can have a dramatic effect on phenotypes. Therefore, all mating characterization was performed with a parent strain that is suitable for the final application of synthetic agglutination. Specifically, we began with the strain EBY100 (**a** *GAL1-AGA1::URA3 trp leu pep4::HIS2 prb1Δ1.6R can1 GAL*), which has been optimized for efficient surface expression of recombinant proteins[29]. This strain contains a second integrated copy of Aga1 under the expression of pGAL1 and knockouts of pep4 and prb1, two proteases that decrease surface expression. We generated a MATα variant of EBY100 using a mating, sporulation, and tetrad selection (Figure 2.1). Positive selection was performed for canavanine resistance, as well as histidine, leucine, and uracil synthesis. Negative selection was performed for lysine synthesis. Mating type was screened using a halo growth assay[30].

We will refer to the parent yeast strains used for all subsequent assays and strain constructions as EBY100a and EBY100α. In addition to the difference in mating type, EBY100a

synthesizes lysine but not leucine while EBY100α synthesizes leucine but not lysine. A mating between these strains produces diploid cells that synthesize both lysine and leucine, allowing for diploid selection with media lacking both amino acids. A transformation with pETCON2[12] is used to express recombinant proteins on the surface of either haploid strain. This centromeric plasmid contains a TRP marker and a GAL inducible surface expression cassette with a myc-tag to measure surface expression level.

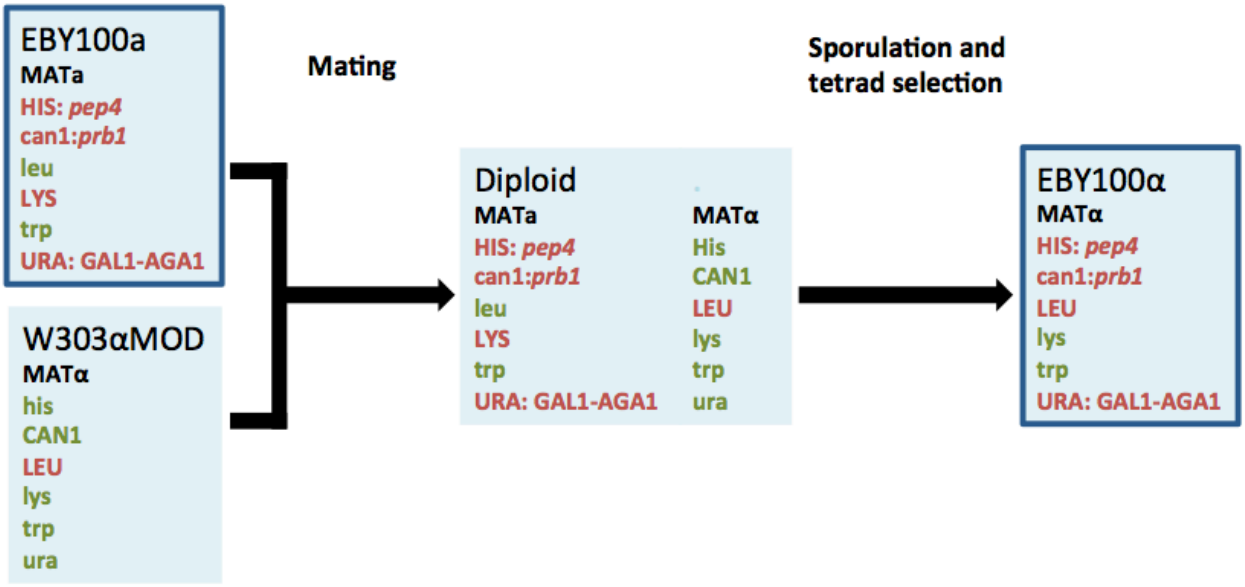


Figure 2.1: Construction of EBY100α. EBY100α was constructed from EBY100a and a W303α variant by mating, sporulation, and tetrad selection.

2.3 Experimental Characterization of Yeast Mating

Wild type mating: When mixed in liquid culture, *S. cerevisiae* haploid cells undergo agglutination that leads to mating and the formation of diploid cells. We confirmed that agglutination and mating occurred as expected upon the mixing of EBY100a and EBY100α. Both haploid strains were grown separately from isogenic colonies to saturation in synthetic complete (SC) media. Fresh MATa and MATα cultures were inoculated with 100 µL of the saturated cultures in 3 mL SC media. Additionally, a co-culture was inoculated with 50 µL of both saturated cultures. Five hours after inoculation of the co-culture, agglutination resulted in the formation of large cell aggregates that rapidly settled. When removed from the shaker, the

co-culture settled after 5 minutes, while a culture containing only one mating type remained in suspension (Figure 2.2A,B).

5 hours after inoculation of the co-culture, matings had occurred between EBY100a and EBY100α cells. Mating was detected by transferring cells to lysine and leucine deficient media and observing growth after a 24 hour incubation, which indicated the presence of mated diploid cells. Combining 5 μL from separately grown MATa and MATα cultures generated no growth in selective media, while transferring 5 μL from a 5-hour co-culture of MATa and MATα cells grew to saturation after 24 hours in selective media (Figure 2.2C,D).

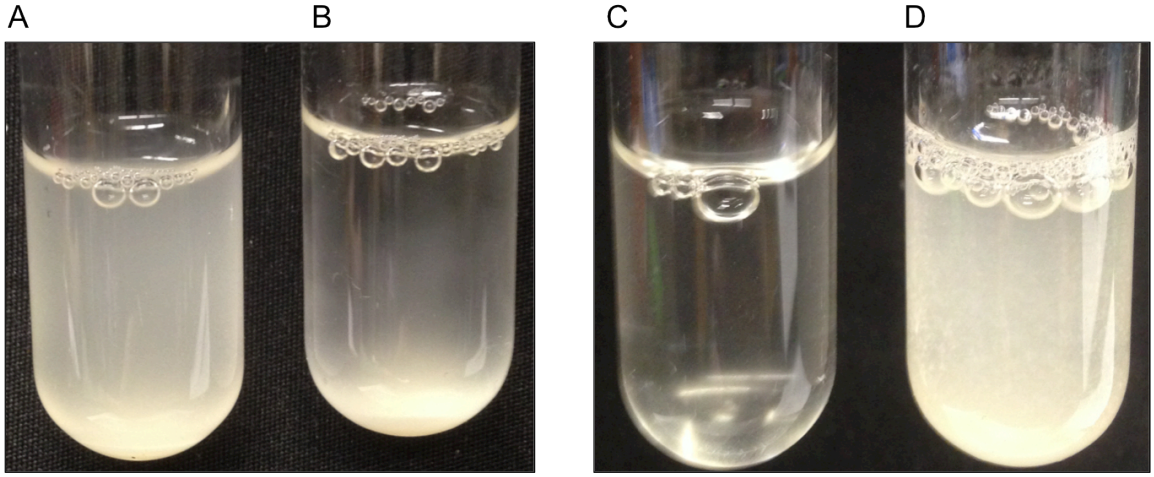


Figure 2.2: Characterization of agglutination and mating. EBY100a (A) or a co-culture of EBY100a and EBY100α (B) was grown for 5 hours in a shaker incubator and then allowed to settle for 5 minutes. Separately grown EBY100a and EBY100α (C) or a co-culture (D) was transferred to lysine and leucine deficient media after 5 hours of growth in SC media and incubated for 24 hours.

Aga2 and Sag1 knockouts: Both Aga2 and Sag1 are required for the mating of *S. cerevisiae* in liquid culture. For both EBY100a and EBY100α, both agglutinin proteins were separately knocked out and a mating assay was performed for each strain combination. A pair of strains were mixed in SC media for 5 hours, washed twice with water, and struck onto a plate lacking lysine and leucine. Colony growth after 48 hours was used to indicate that mating had occurred, since only mated diploid cells contained genes for the synthesis of both lysine and leucine. No mating occurred either when Aga2 was knocked out of the MATa strain or when Sag1 was knocked out of the MATα strain (Table 2.1). This result is consistent with previous work describing the mating type specific roles of the sexual agglutinin proteins[6].

Table 2.1: Mating assays with wild type (WT) and agglutinin knockout strains (-Aga2 or -Sag1).

	EBY100a WT	EBY100a-Aga2	EBY100a-Sag1
EBY100α WT	Growth	No Growth	Growth
EBY100α-Aga2	Growth	No Growth	Growth
EBY100α-Sag1	No Growth	No Growth	No Growth

Nutrient availability: Mating only occurred when all required nutrients were available for both haploid strains, indicating that haploid growth is necessary for mating to occur (Table 2.2). Therefore, all future mating assays must be designed so that selection is not required during the mating step. Instead, if diploid isolation is required, a mating must be conducted in non-selective media and then transferred to selective media after matings have occurred.

Table 2.2: Growth after plating saturated haploid pairs on complete (SC) or selective media.

	SC	SC-lys-leu	SC-lys	SC-leu
EBY100a WT & EBY100α WT	Growth	No Growth	No Growth	No Growth
EBY100a-Aga2 & EBY100α-Sag1	Growth	No Growth	No Growth	No Growth

Yeast surface display traditionally uses a centromeric plasmid, pETCON2, containing an auxotrophic selection marker (TRP) and surface expression cassette. After mating, a diploid cell must maintain the surface expression cassettes from each haploid. Gene retention in diploids will be critical for later assays that characterize interactions between mixed libraries of yeast strains. In order to prevent the spontaneous dropping of plasmids from diploids, dual selection is required. However, since nutrient limitation prevents mating, dual selection cannot be used during the mating step. One possible solution is to conduct matings without selection and then transfer the cells to selective media to isolate diploids that have maintained both plasmids. However, a substantial diploid population would likely have dropped at least one plasmid over the course of a 17-hour mating, resulting in the isolation of only a subset of the total diploid population that was formed from matings. Instead, chromosomal integration of the surface expression cassette is preferable, since this will eliminate the risk of gene dropping in non-selective media.

Low mating efficiency in nutrient limited conditions likely means that metabolic activity is required for the initiation of mating. To test this hypothesis, we evaluated the mating of

haploid cells added to nutrient limited media at different phases of growth. We observed that cells mixed in log phase but not at saturation were able to mate in nutrient limited conditions on a plate (Table 2.3). Therefore, when using a secondary selection to determine whether or not mating has occurred in liquid culture, it is critical that the cells are in stationary phase prior to plating in order to prevent false-positives from post-plating diploid formation. This observation also means that dual selection for plasmid maintenance during mating may be possible. However, there are still two major challenges. First, it would be difficult to ensure that many strains are in an identical growth phase at a given time. The addition of cells at saturation simplifies assay preparation considerably, which likely improves consistency. Second, dual selection would still limit mating to a very short time window, since haploids cells will rapidly exit log phase in nutrient limitation. Due to these challenges, the use of chromosomal integration of the expression cassettes still seems preferable to the use of centromeric plasmids.

Table 2.3: Growth after plating haploid cells on diploid selective media.

	Plating from saturation	Plating from log growth
EBY100a WT & EBY100α WT	No Growth	Growth
EBY100a-Aga2 & EBY100α-Sag1	No Growth	Growth

Carbon source: The use of galactose as a carbon source dramatically reduced the mating efficiency of EBY100a and EBY100α. Yeast surface display traditionally has the surface expression cassette under the control of the pGAL1 promoter, so that growth in 2% galactose is used to induce surface expression[12]. Since matings are typically performed in 2% dextrose media, we tested the effect of switching the carbon source to galactose. Mating assays have two liquid culture growth steps. First, haploid strains are grown separately overnight to saturation. Second, haploid strains are mixed in a single culture for 17 hours. Cultures are then plated on media lacking lysine and leucine and grown for 48 hours for diploid selection. Growth in galactose during either step decreased mating efficiency significantly (Figure 2.3). This strongly suggests that assays requiring induction by growth in galactose are not suitable for a mating assay. Instead, a constitutive promoter, such as pGPD[31], could be used for the expression of Aga1 and the Aga2 fusion. Characterization is required to determine how changing promoters effects surface expression level.

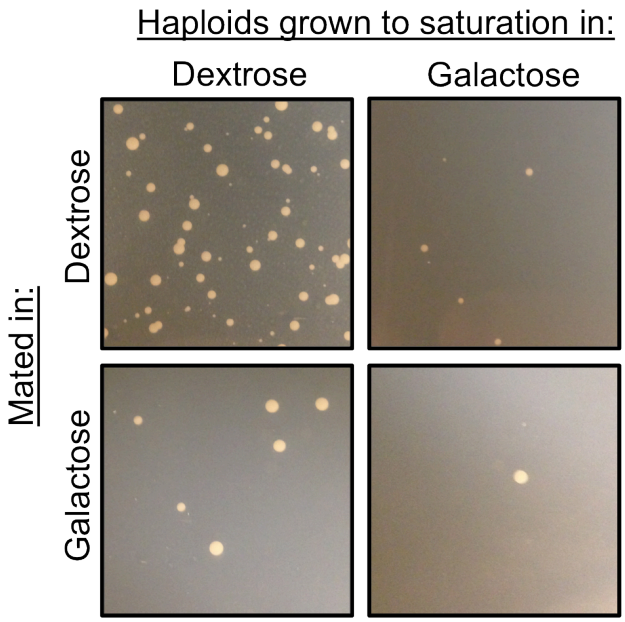


Figure 2.3: Mating efficiency is reduced with galactose induction. Plate images from a carbon source mating test are shown. Colony growth indicates diploid formation from the mating of haploid cells individually grown in dextrose or galactose and then mated in dextrose or galactose.

2.4 STRAIN CONSTRUCTION AND VALIDATION

In order to co-opt yeast mating for probing protein-protein interactions, we genetically replaced the native sexual agglutination proteins with arbitrary binding proteins using yeast surface display. Based on the assay requirements for agglutination and yeast mating described above, many modifications to the standard surface display system were required. Specifically, native agglutination had to be knocked out, surface display had to be constitutive rather than galactose inducible, and all genetic components had to be chromosomally integrated. To eliminate wild type agglutination, we knocked out the MATα sexual agglutinin protein, Sag1, from EBY100α. Sag1 was chosen rather than Aga2 because an Aga2 fusion is used for yeast surface display. It is possible that Aga2 fused to a protein of interest would retain some ability to interact with Sag1. Next we replaced two pGAL1 promoters from the standard yeast surface display system with one of the strongest constitutive promoters native to *S. cerevisiae*, pGPD[31]. This included a promoter driving Aga1 expression from the yeast chromosome and a promoter driving the surface expression cassette from the centromeric plasmid, pETCON2. These genetic modifications are

represented in the yeast strains ySYNAGa and ySYNAGα and a modified surface display plasmid, pETCON2[pGPD]. Finally, we further modified pETCON2[pGPD] for chromosomal integration by adding chromosomal homology and restriction digest sites.

The yeast strains ySYNAGa and ySYNAGα were tested for proper surface display using the surface expression of streptavidin as a control. Display was tested by incubating freshly saturated cells with FITC conjugated anti-myc antibody and measuring cellular fluorescence intensity with the FL.1A channel on an Accuri C6 flow cytometer. Fluorescence intensity profiles of EBY100a and EBY100α populations were nearly identical (Figure 2.4). We compared our modified surface display system in ySYNAGa and ySYNAGα with integrated expression cassettes to conventional display in EBY100a and EBY100α using pETCON2-streptavidin and galactose induction. Constitutive expression and integration improved surface display and reduced bimodality. Both observations were unsurprising given the strength of the pGPD promoter and the elimination of gene dropping due to plasmid loss.

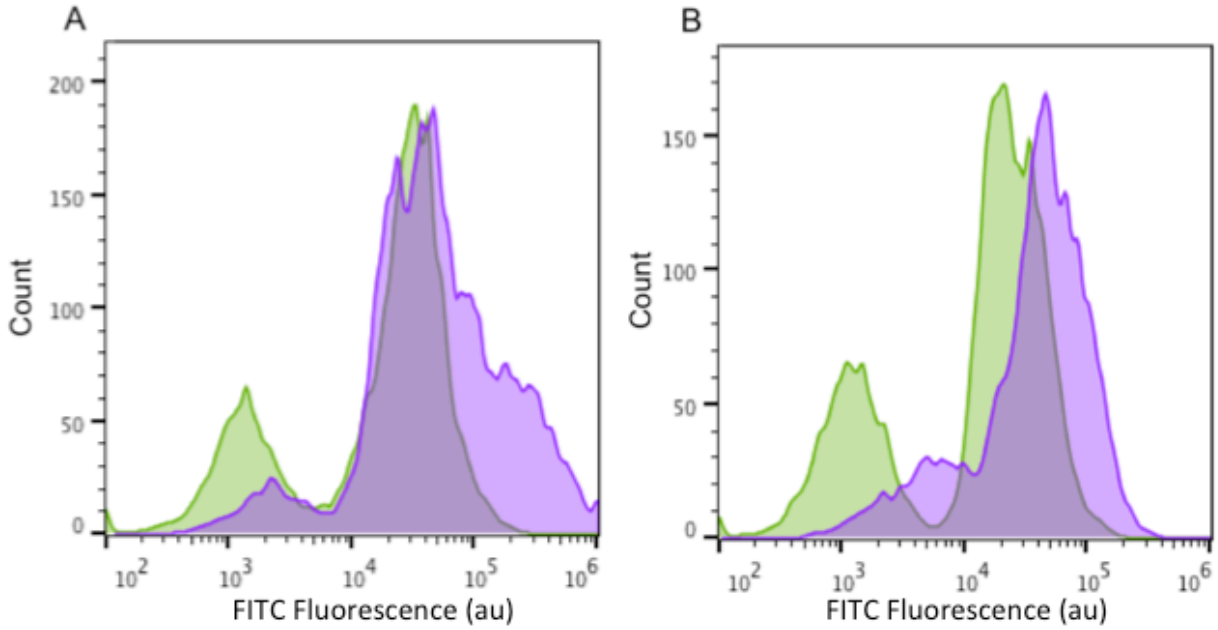


Figure 2.4: Surface expression analysis of ySYNAGa and ySYNAGα. MATa (A) and MATα (B) strains were tested for surface display by labeling with FITC conjugated anti-myc antibody. For both mating types, the EBY100 display system (green) was compared with the SYNAG system (purple).

Despite improved surface expression and reduced bimodality, there are two disadvantages of the SYNAG system that will become important for later library building and screening. First, constitutive expression of the surface display cassette means that differences in metabolic load or protein toxicity can generate sustained growth differences between strains. For mixed yeast libraries, even subtle growth differences can generate substantial population biases over time. This means that it will be critical to thoroughly screen naïve libraries and characterize population distributions in order to account for differences in growth rates. Second, chromosomal integrations have low transformation efficiency compared to plasmid transformations. This will pose a challenge for later construction of large libraries and will necessitate a new approach for improving transformation efficiency.

2.5 REPROGRAMMING SEXUAL AGGLUTINATION

Using ySYNAGa and ySYNAGα, we were able to recover mating efficiency in liquid culture after knocking out Sag1 by expressing complementary binding proteins on the surface of a haploid pair to generate synthetic agglutination. We transformed ySYNAGa with a surface expression cassette containing Bfl-1, a human pro-survival BCL2 homologue that is a target for cancer therapeutics and part of a complex protein interaction network that regulates apoptosis. We transformed ySYNAGα with 13 different proteins, including pro-survival and pro-apoptotic BCL2 family members and *de novo* binding proteins. Liquid culture matings between ySYNAGa:Bfl-1 and all 13 MATα variants were performed along with negative controls lacking one or both surface expression cassettes (Table 2.4). Mating cultures were then plated on diploid selective media to assess mating efficiency.

Table 2.4: Matings between yeast strains displaying complementary (green) or non-complementary (white) binding proteins.

Plate	ySYNAGa	ySYNAGα	Affinity[32]
1	Bfl-1	Bfl-1	>25 μM
2	Bfl-1	Bcl-B	>25 μM
3	Bfl-1	Bcl-xL	>25 μM
4	Bfl-1	BHRF1	>25 μM
5	Bfl-1	Bcl-2	>25 μM
6	Bfl-1	Mcl-1	>25 μM
7	Bfl-1	Bcl-w	>25 μM
8	Bfl-1	Bim.BH3	<10 nM
9	Bfl-1	BINDI-N62S	>2 μM
10	Bfl-1	αMCL1	>25 μM
11	Bfl-1	2CDP06	>25 μM
12	Bfl-1	FECM04	<10 nM
13	Bfl-1	BECM01	>2 μM
14	Bfl-1	None	NA
15	None	None	NA

Of the 15 matings, two contained a haploid pair expressing complementary binding proteins. Bfl-1 had previously been shown to bind to both Bim.BH3 and FECM04 with low nanomolar affinities[32]. No other protein pairs were expected to interact with sufficient strength to recover mating efficiency. Colony growth on diploid selection plates revealed that the two complementary haploid pairs formed an order of magnitude more diploid cells than all non-complementary haploid pairs, indicating that synthetic agglutination can recover mating efficiency (Figure 2.5).

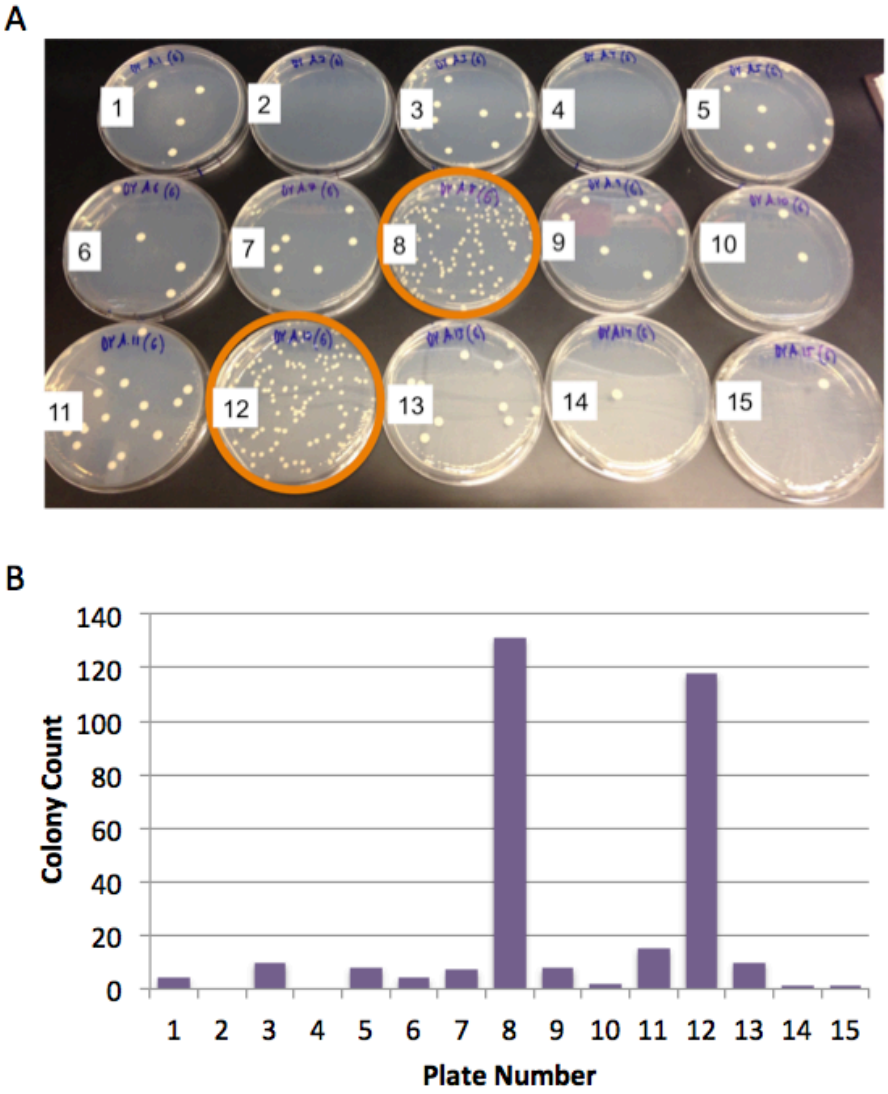


Figure 2.5: Mating efficiency recovery with synthetic agglutination. (A) Diploid selection plates from a liquid culture mating assay. The highlighted plates, 8 and 12, were matings of haploid strains expressing complementary binding proteins. (B) Colony counts for each plate.

2.6 DISCUSSION

We have demonstrated an ability to reprogram sexual agglutination with the surface expression of yeast-foreign binding proteins. Cell adhesion is necessary and sufficient for replacing the function of the agglutinin proteins Aga2 and Sag1, indicating that their role is limited to binding, as was previously suggested[5]. Synthetic agglutination provides a novel platform for studying cell-cell adhesion using an easily engineerable model organism. This could

be useful for many applications, such as the study of proteins involved in human cell-cell adhesion[33–35], bacterial adhesion[36,37], viral capsid adhesion[38,39], or gamete adhesion[40,41].

The need to plate cells for diploid isolation is a major limitation of this method. With a plate based assay, only diploid populations can be quantified and the resolution is severely limited by the need to count distinguishable colonies. A secondary growth requirement also extends the assay time by multiple days and introduces a possibility of false-positives due to post-plating diploid formation. There are many possible strategies for improving the resolution and throughput of a synthetic agglutination mating recovery assay, such as using fluorescent reporters to differentiate between cell types and analyzing mixed populations with flow cytometry.

Chapter 3

PAIRWISE PROTEIN INTERACTION CHARACTERIZATION

3.1 INTRODUCTION

A selection-based assay for measuring diploid populations provides only a qualitative measurement of mating efficiency, is limited in resolution, requires separate platings for each condition, and has a high probability of false positives. To overcome these challenges, we developed a dual-channel flow cytometry assay for the measurement of haploid and diploid populations directly from a mixed mating culture. With flow cytometry, we can individually count haploid and diploid cells in a mixed population, which enables quantitative mating efficiency measurements and improved mating efficiency resolution. We can also eliminate the need for a secondary growth step for diploid selection, which decreases assay time and the probability of false positives. Following assay construction and optimization, we used pairwise matings as a quantitative tool for measuring protein interaction strength by incorporating synthetic agglutination. We found a strong log-linear relationship between mating efficiency and affinity across over five orders of magnitude of K_D. Synthetic agglutination coupled with flow cytometry provides a powerful new approach for the rapid and quantitative characterization of protein-protein interaction strength.

3.2 IDENTIFYING A SUITABLE PAIR OF FLUORESCENT REPORTERS

To distinguish between cell populations, distinct fluorescent reporters were constitutively expressed from the pGPD promoter in EBY100a and EBY100α. In addition to being spectrally resolvable with one another, resolvability with FITC (excitation: 495 nm, emission 519 nm) was also required, since a FITC conjugated anti-myc antibody was used to validate and characterize surface expression of each displayed binding protein[42]. A red florescent protein, mCherry (excitation: 587 nm, emission 610 nm), was chosen as the first marker due to previous experience using this protein with FITC. A blue florescent protein, mTurquoise, was chosen as the second marker[43,44]. Three closely related reporters in the blue/cyan spectrum were considered due to having an excitation wavelength of 380-440 nm and an emission wavelength of 450-475

nm, making all three clearly distinguishable from FITC and mCherry. mTurquoise was found to give the highest signal intensity of the three considered proteins (Figure 3.1).

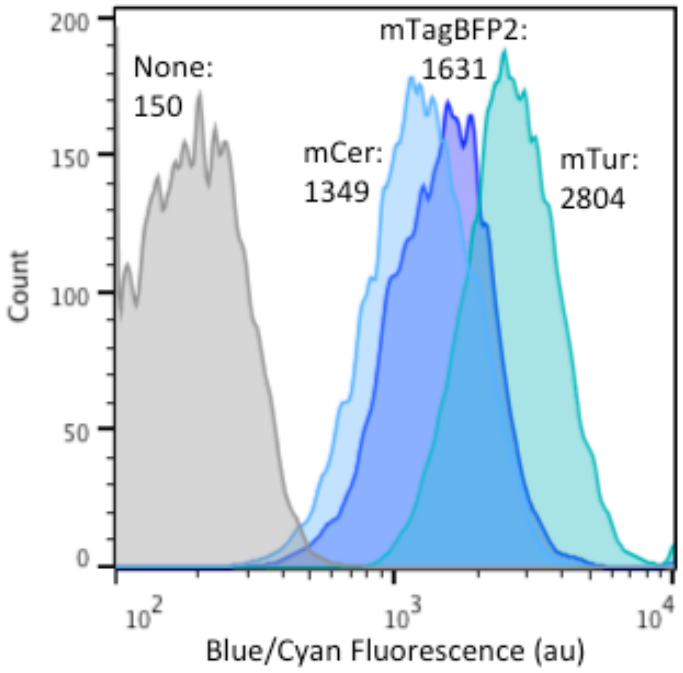


Figure 3.1: Fluorescent intensity histograms for three candidate proteins in the blue/cyan spectrum: mCerulean, mTagBFP2, and mTurquoise. Mean fluorescence (au) in the V1.A channel is given for each.

3.3 Counting Yeast Populations with Dual-Channel Cytometry

Using a dual-channel flow cytometry assay, we were able to clearly differentiate between MATa, MATα, and diploid populations in a mixed culture. We began by constitutively expressing mCherry in EBY100a and EBY100a-Aga2 and constitutively expressing mTurquoise in EBY100α and EBY100α-Sag1. All four strains were grown to saturation. Fresh EBY100a:mCherry and EBY100α:mTurquoise cultures were then grown for 17 hours along with a mixed culture with both agglutinin positive strains and a mixed culture with both agglutinin knockout strains. All four cultures were then diluted and run on a Miltenyi MACSQuant flow cytometer where mCherry intensity was monitored with the Y2 channel and mTurquoise intensity was monitored with the V1 channel (Figure 3.2). Both haploid monocultures gave the expected fluorescent profiles, with MATa haploids showing strong mCherry fluorescence and MATα haploids showing strong mTurquoise fluorescence. Three distinct populations were

identified from a 17-hour co-culture of EBY100a:mCherry and EBY100α:mTurquoise, including a large population of diploid cells showing a strong signal for both mCherry and mTurquoise fluorescence. No obvious diploid population was observed from a 17-hour co-culture of the agglutinin knockout strains. This result is consistent with previous observations of liquid culture mating inhibition with the knockout of the sexual agglutinin proteins and validates the dual-channel cytometry assay as a method for measuring cell populations and determining whether or not mating has occurred.

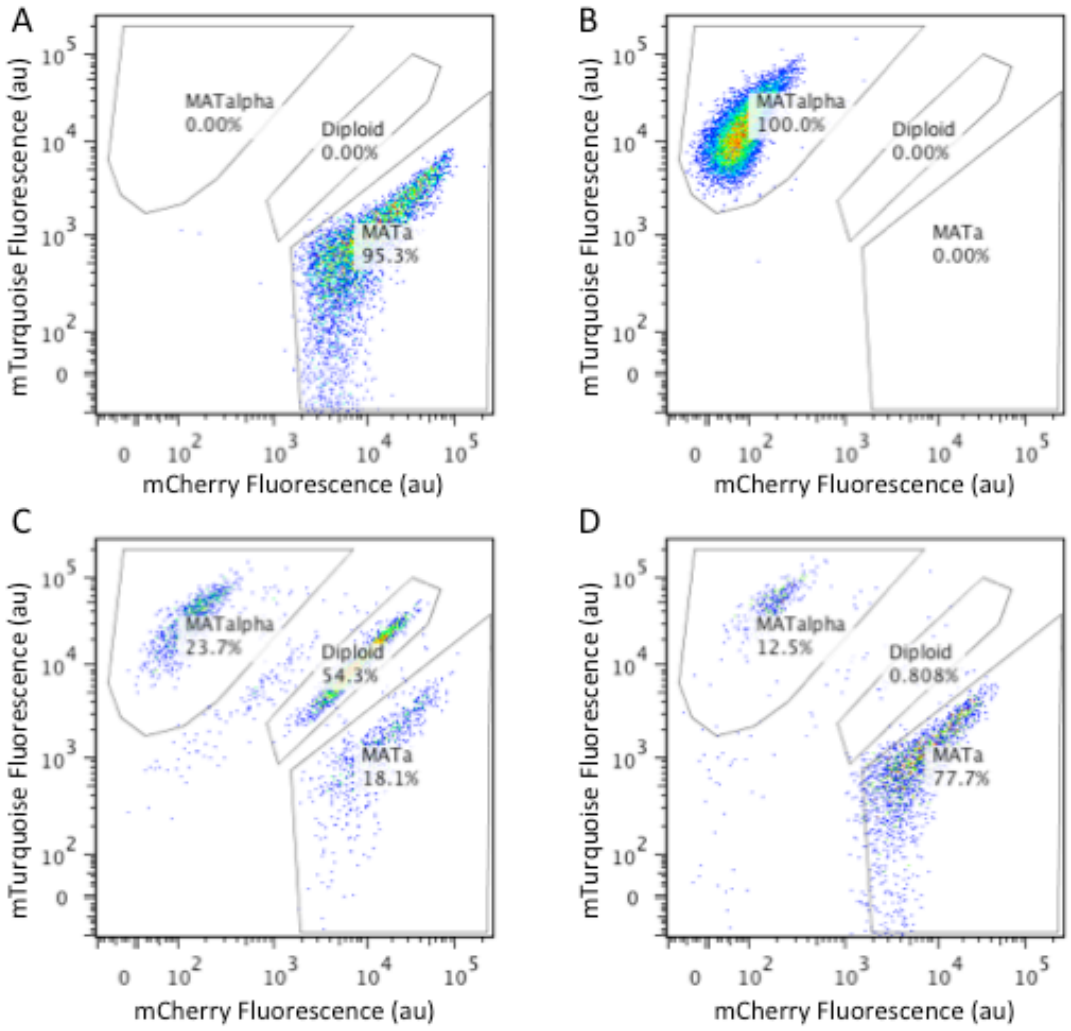


Figure 3.2: Dual-channel flow cytometry analysis. Flow cytometry was performed on monocultures of (A) mCherry expressing MATa haploids and (B) mTurquoise expressing MATα haploids along with 17-hour co-cultures of (C) EBY100a and EBY100α and (D) EBY100a-Aga1 and EBY100α-Sag1. Fluorescence gates were added to analyze population distributions of MATa, MATα, and diploid cells.

In addition to differentiating between haploid and diploid cells, it is useful that the dual channel cytometry assay can also differentiate between haploid cells of different mating types. We found that MATa cells grow faster than MATα cells when co-cultured, even though their growth rates are nearly identical when grown separately. In a co-culture, both cell types are exposed to mating factor secreted by cells of the opposite mating type, which activates the mating pathway. Many genes are differentially expressed after exposure to mating factor including genes that cause growth arrest in order to prime the cells for mating[45]. We found that MATα cells undergo a stronger growth arrest than MATa cells, leading to an uneven population distribution of the two mating types throughout a 17-hour mating. To correct for the discrepancy in growth effects, we added twice as many MATα cells as MATa cells, which resulted in similar counts of MATa and MATα cells after a 17-hour mating. Unless otherwise noted, all future mating assays begin with the inoculation of a 3 mL culture with 2.5 μL of saturated MATa cells and 5 μL of saturated MATα cells.

3.4 Measuring Mating Efficiency with Synthetic Agglutination

Dual-channel flow cytometry was used to correctly characterize the recovery of mating efficiency with synthetic agglutination. As an initial validation, previously characterized BCL2 family apoptosis regulating proteins[46] were expressed on the surface of MATa and MATα haploid cells and tested for their ability to recover mating efficiency in agglutinin knockout strains. Specifically, strong (Bfl-1 & Noxa.BH3), medium (Bcl-2 & Bmf.BH3), and weak (Bcl-W & Bik.BH3) pairs of binding proteins were tested, as well as a non-binding pair (Bcl-2 & Noxa.BH3)[47]. An isogenic yeast strain was generated for the display of each individual protein by integrating a surface expression cassette into the yeast genome. Pairs of surface expressing haploid cells were co-cultured in non-selective liquid media for 17 hours to allow agglutination-dependent mating. Flow cytometry was performed to differentiate between mCherry expressing MATa haploids, mTurquoise expressing MATα haploids, and mated diploids that expressed both fluorescent markers. Diploid percent was used as a metric for mating efficiency to quantitatively characterize the interaction strength between a given MATa and MATα strain.

We again validated that complementary binding proteins expressed on the surface of yeast are necessary and sufficient to replace the function of the native sexual agglutinin proteins, Aga2 and Sag1. Wild-type EBY100 *S. cerevisiae* haploid cells mated with an efficiency of 63.6% ± 3.1% in standard laboratory conditions and a knockout of Sag1 in the MATα haploid eliminated mating with wild-type MATa (Figure 3.3A). In the Sag1 knockout, expression of an interacting SynAg protein pair recovered mating efficiency to 51.6% ± 7.9%, while expression of a non-interacting SynAg protein pair showed no observable recovery (Figure 3.3B). SynAg-dependent recovery of mating occurred with a variety of natural and engineered proteins ranging from 26 to 206 amino acids, indicating a large engineerable space for synthetic agglutination. Future testing of proteins with diverse properties will be required to determine the limitations of synthetic agglutination.

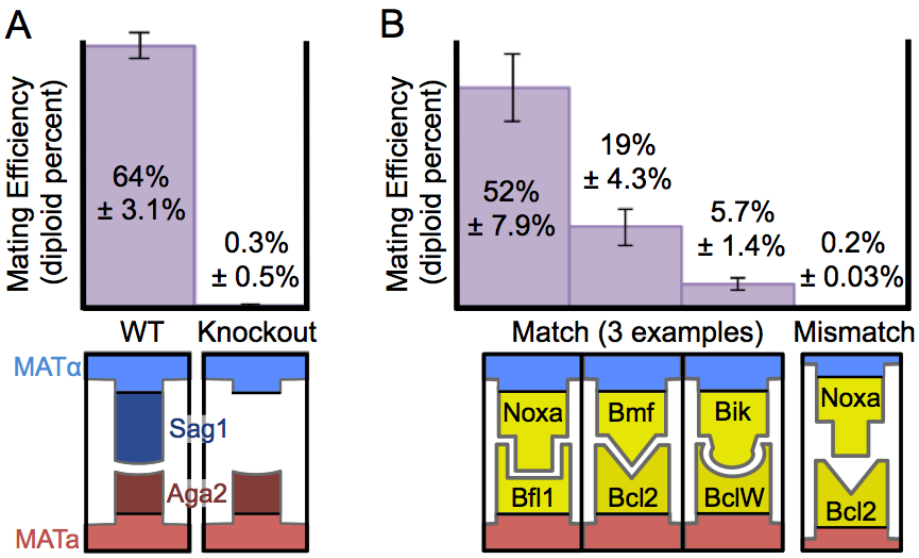


Figure 3.3: A demonstration of mating efficiency recovery with synthetic agglutination. Mating efficiency is given as the percent of diploid cells for representative haploid pairs expressing (A) wild type sexual agglutinin proteins and (B) synthetic agglutinin proteins. Error bars represent standard deviation from three replicates.

3.5 Population Dynamics of Yeast Mating

Dual-channel flow cytometry can track haploid and diploid populations over time in order to characterize population dynamics. By growing MATa, MATα, and diploid cells in

isolation, we determined that the growth rate for each cell type is very similar and consistent with literature values (Figure 3.4A). All cell types displayed a substantial lag phase before entering exponential growth, which is expected because all cells began in stationary phase from a saturated overnight culture. Each condition was inoculated with approximately 100 cells/μL in 3 mL of yeast peptone dextrose (YPD) media. We also characterized population dynamics of two matings with synthetic agglutination in which 100 cells/μL of a MATa strain and 200 cells/μL of a MATα strain were mixed in a single 3 mL YPD culture (Figure 3.4B,C). Differences in starting cell concentration are due to uneven growth arrest of MATa and MATα cells when mixed, as was discussed previously. The first mating included a haploid pair expressing strong complementary binding proteins, Bfl-1 and Bim.BH3 (K_D <10 nM)[47]. Diploids were first detected at 9 hours after mixing and a sizable diploid population was detected at 11 hours after mixing. The second mating included a haploid pair expressing weakly interacting binding proteins, Bfl-1 and BINDI.N62S (K_D = 4,000 nM +/- 2,000 nM). For this pair, a diploid population was not detected until 13 hours after mixing. Since we aim to detect weak interactions, a mating time of at least 13 hours is required for future assays. Unless otherwise noted, all future mating assays were conducted for 17 hours.

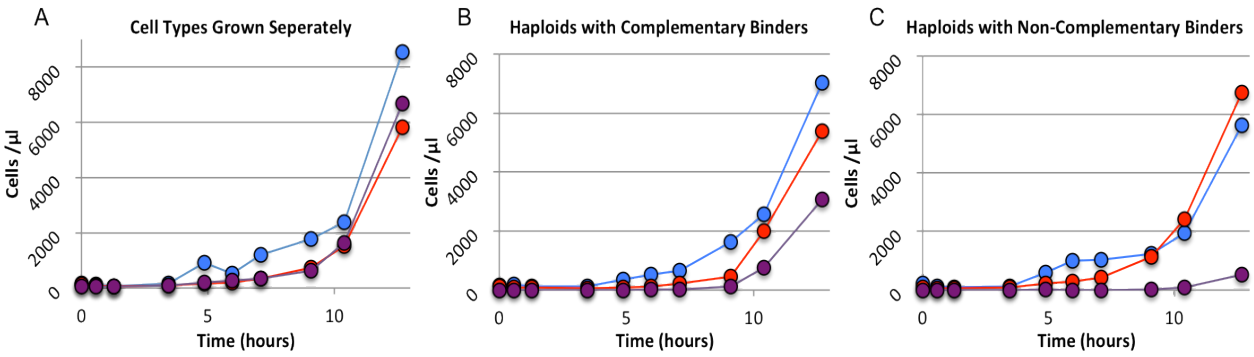


Figure 3.4: Haploid and diploid growth dynamics during mating. Growth dynamics of MATa haploid cells (blue), MATα haploid cells (red), and diploid cells (purple) were analyzed in YPD media. (A) Growth curves for MATa, MATα, and diploid cells grown in separate YPD cultures. (B,C) Growth curves for each cell type from a mating consisting of haploids expressing (B) a strong binding pair or (C) a weak binding pair.

3.6 Growth Conditions for Yeast Agglutination and Mating

Growth conditions, such as starting cell concentrations and media type, have a large effect on mating efficiency. To test the effect of various growth conditions, we measured the mating efficiency of ySYNAGa:Bfl-1 with four ySYNAGα variants representing two strong binding pairs (ySYNAGα:Bim.BH3, FECM04) and two non-binding pairs (ySYNAGα:2CDP06, BECM04). We began by testing different starting concentrations of haploid cells to determine an optimal starting cell concentration. For simplicity, we measured starting concentration by the volume of saturated haploid culture added to a 3 mL mating. We found that adding approximately 10 µL of each mating type provides the greatest difference in mating efficiency between binding and non-binding pairs (Figure 3.5). This provided an optimal order of magnitude for the cell concentration. Upon further analysis, we found optimal differentiation between binding strength with 2.5 µL of MATa and 5 µL of MATα.

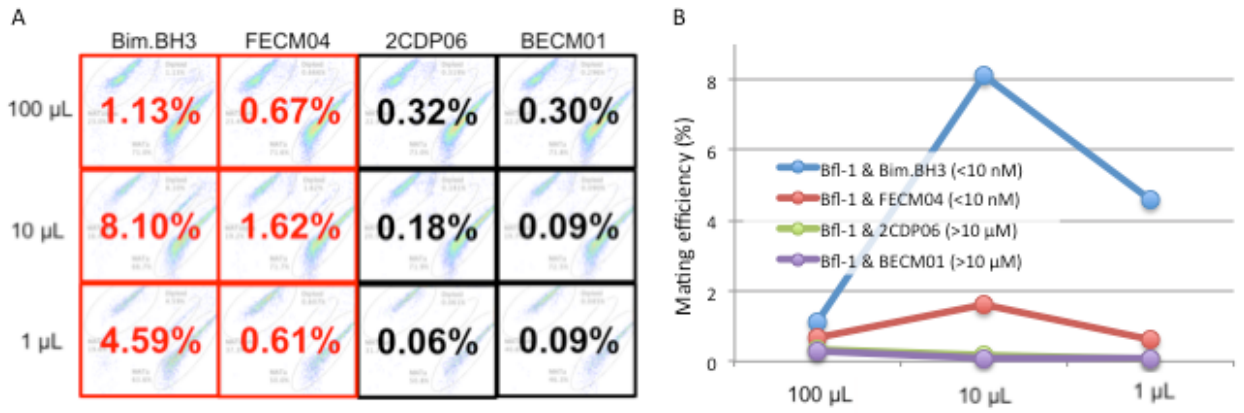


Figure 3.5: Optimization for starting haploid concentrations. (A) Binding pairs (red) and non-binding pairs (black) were mixed at different concentrations and mating efficiency was measured. (B) Adding 10 µL of each mating type maximizes the mating efficiency difference between binders and non-binders.

Next, we characterized mating efficiency in synthetic complete (SC) fully defined minimal media with dextrose and yeast peptone dextrose (YPD) rich media. We found that mating in SC media resulted in far higher mating efficiencies, but also high background (Figure 3.6). In SC media, both non-binding pairs mated with an efficiency of greater than 5%. Switching from SC media to YPD media resulted in a 70-85% reduction in mating efficiency for

strong binders and a 96-99% reduction in mating efficiency for non-binders. Therefore, despite a decrease in signal, YPD media gave considerably better resolution between binders and non-binders.

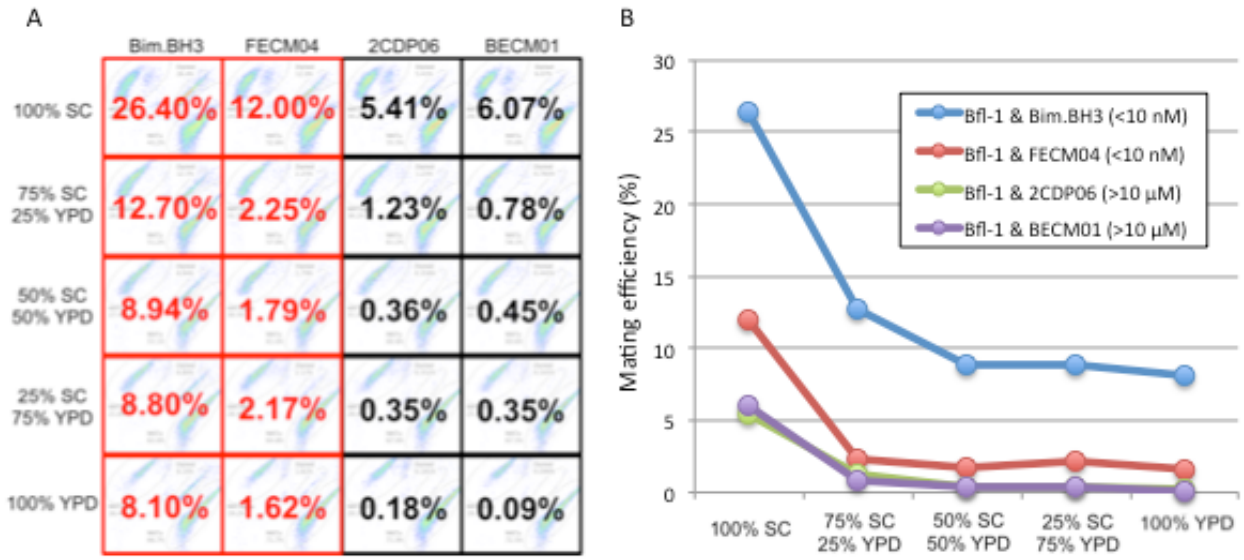


Figure 3.6: Optimization of media type. (A) Binding pairs (red) and non-binding pairs (black) were mated in different ratios of synthetic complete (SC) media and yeast peptone dextrose (YPD) media and mating efficiency was measured. (B) YPD media reduces the mating efficiency of binders, but also reduces non-specific mating of non-binders.

We hypothesized that the high mating efficiencies observed in SC media were due to a lack of blocking reagent in the media to prevent non-specific protein interactions. For many binding assays, bulk protein or surfactant is added to compete with non-specific interactions[48]. Unlike SC, which is a fully defined minimal media, YPD contains yeast extract. This ingredient includes a random assortment of proteins that would function as a blocking reagent. We confirmed this hypothesis by testing the mating efficiency of the same haploid pairs in SC media supplemented with different concentrations of bovine serum albumin (BSA) (Figure 3.7) or Tween-20 (Figure 3.8). The addition of either BSA or Tween-20 reduced the mating efficiencies for binding and non-binding pairs to levels resembling mating in YPD. Based on these results, we can conclude that YPD contains sufficient blocking reagent to reduce non-specific interactions that enable mating between a pair of haploid cells expressing non-binding synthetic agglutinin proteins. YPD media is used exclusively for all future mating assays described here.

However, there are many possible future applications in which a fully defined minimal media would be useful. For such applications, blocking reagent should be added to prevent non-specific binding.

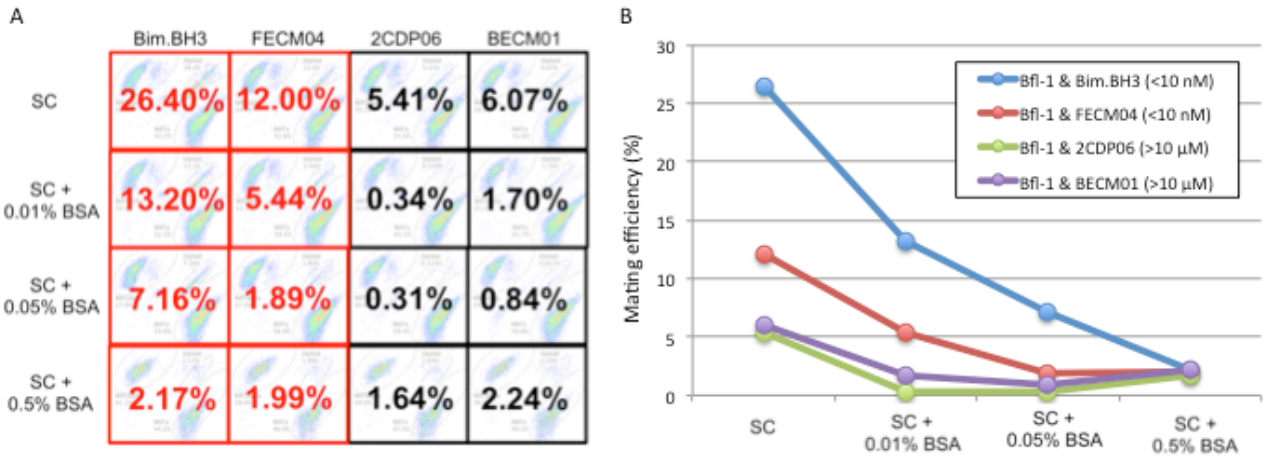


Figure 3.7: Characterization of bovine serum albumin (BSA) blocking in synthetic complete (SC) media. (A) Binding pairs (red) and non-binding pairs (black) were mated in SC media with different concentrations of BSA and mating efficiency was measured. (B) BSA addition reduced the mating efficiency of binders, but also reduced non-specific mating of non-binders.

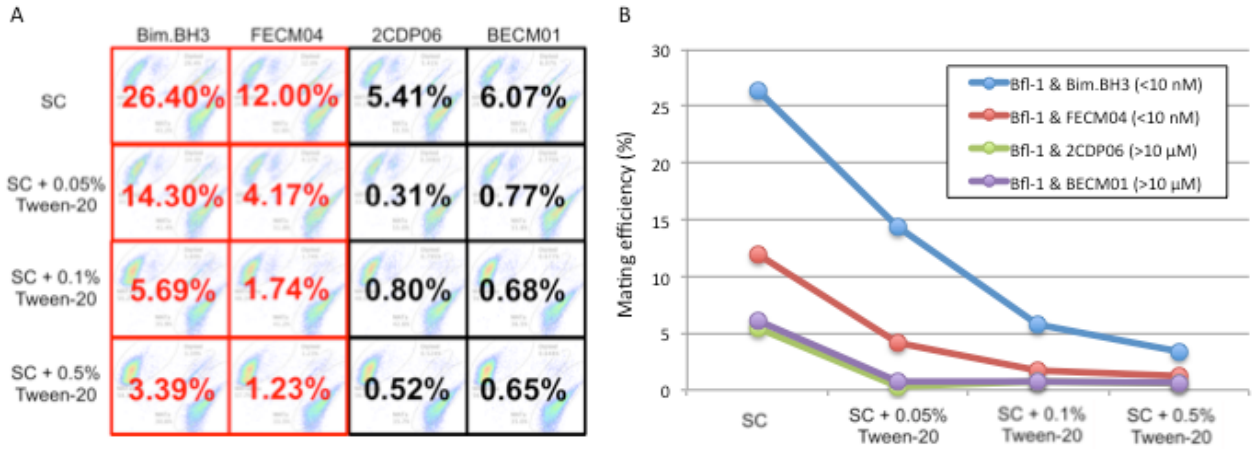


Figure 3.8: Characterization of Tween-20 blocking in synthetic complete (SC) media. (A) Binding pairs (red) and non-binding pairs (black) were mated in SC media with different concentrations of Tween-20 and mating efficiency was measured. (B) Tween-20 addition reduced the mating efficiency of binders, but also reduced non-specific mating of non-binders.

3.7 Relationship Between Mating Efficiency and Affinity

Early synthetic agglutination assays suggested that mating recovery is not binary. Instead, co-cultures of haploid cells expressing strong binders result in a higher mating efficiency than co-cultures of haploid cells expressing weak binders. This relationship suggests that mating efficiency may enable a quantitative or semi-quantitative characterization of binding strength. To determine the relationship between mating efficiency and binding affinity, we used human apoptosis regulatory proteins in the BCL2 family that were previously characterized with biolayer interferometry (BLI)[32,49]. Six pro-survival BCL2 homologues (Bcl-2, Bfl-1, Bcl-B, Bcl-w, Bcl-xL, and Mcl-1) were expressed on MATa cells. Seven pro-apoptotic peptides (Bim.BH3, Noxa.BH3, Puma.BH3, Bad.BH3, Bik.BH3, Hrk.BH3, and Bmf.BH3) and nine engineered binding proteins (αBFL1, FECM04, αBCLB, BCDP01, BECM01, αBCL2, 2CDP06, XCDP07, and αMCL1) representing a broad range of affinities for the BCL2 homologues were expressed on MATα cells.

From pairwise matings of each MATa and MATα strain, we determined that mating efficiency and affinity are related log-linearly ($R^2 = 0.89$) for protein interactions across over five orders of magnitude of K_D (Figure 3.9). We tested proteins with binding affinities ranging from below 500 pM to above 300 μM, which gave mating efficiencies of up to 35.4% and down to below 0.2%. None of the 14 tested pairs with a K_D above 25 μM resulted in a recovery of mating above 0.4%. The weakest interaction showing a detectable mating recovery had a K_D of 12.5 μM and a mating efficiency of 0.6%.

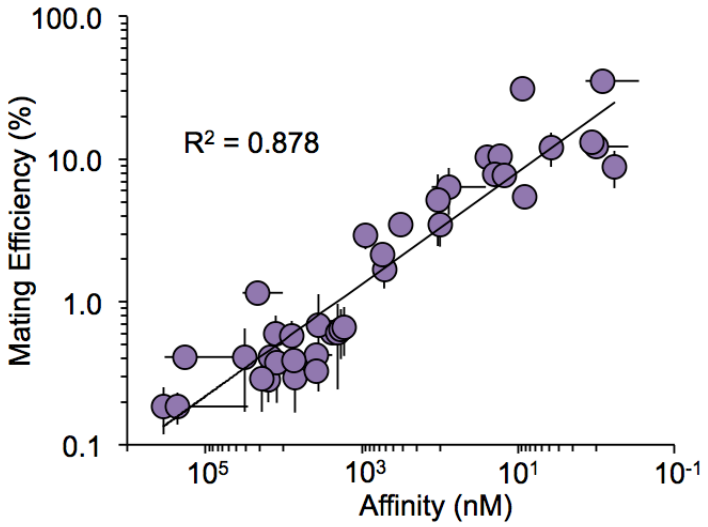


Figure 3.9: Relationship between synthetic agglutination mating efficiency and affinity. Mating efficiencies for synthetic agglutinin expressing haploid pairs with a K_D between 500 pM and 300 μM show a log-linear relationship with binding affinity. Error bars represent standard deviation from three replicates.

The strong log-linear relationship between mating efficiency and affinity over multiple orders of magnitude contradicted our expectation of avidity as the main driving force for yeast agglutination[5,8]. We expected that upon the formation of a single interaction between cells, newly localized protein pairs would rapidly bind, making off-rate largely irrelevant. However, both on- and off-rate showed a correlation with mating efficiency, and neither provided as good a fit as K_D (Figure 3.10). Most simple binding systems involve soluble molecules that only require diffusion to be considered as a mechanism for dissociation. However, for cell-cell attachment in liquid culture, the shear force pulling apart two cells must be considered. In this case, it is likely that individual protein interactions are constantly being broken and reformed, meaning that both on-rate and off-rate are involved in the dissociation of two cells.

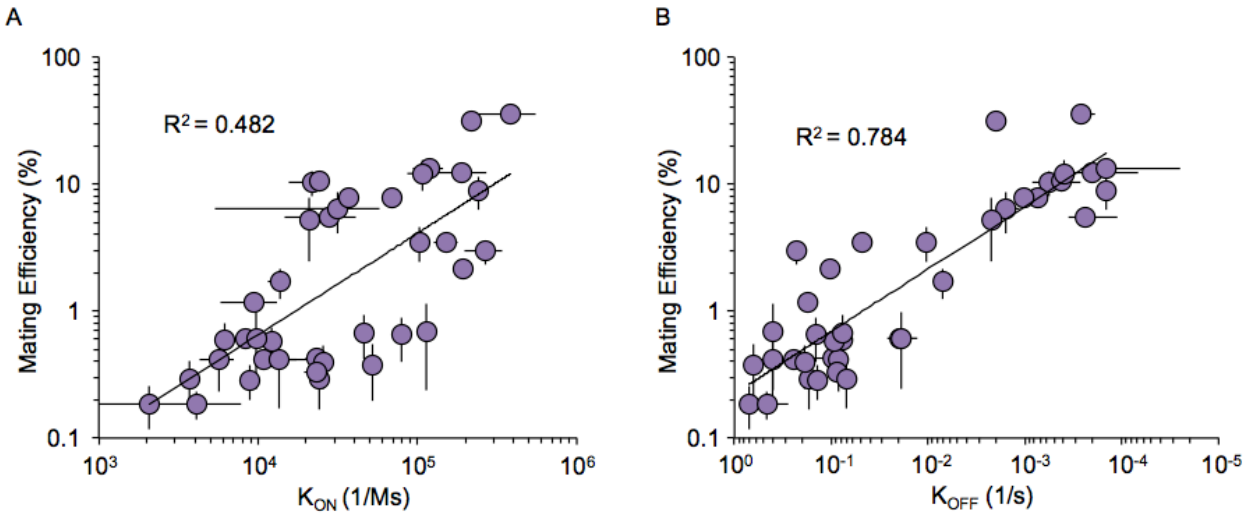


Figure 3.10: Relationship between mating efficiency and binding kinetics. Mating efficiency percent plotted against (A) on-rate and (B) off-rate, as measured with BLI. Error bars represent standard deviation from three replicates.

3.8 Characterization of Surface Expression

Mating recovery is dependent on the amount of synthetic agglutinin proteins displayed on the surface of both haploid cells. Prior to mating, each synthetic agglutinin-expressing yeast strain was tested for surface expression level by labeling with FITC-conjugated anti-myc antibody and measuring fluorescent intensity with an Accuri C6 flow cytometer[12]. One BCL2 homologue, Mcl-1, showed a very low surface expression and subsequently minimal recovery of mating efficiency regardless of its mating partner. We hypothesized that a threshold surface expression level is critical for the recovery of agglutination. However, based on the strong relationship between mating efficiency and affinity for all BCL2 homologues other than Mcl-1 without adjusting for surface display levels, it seems as though there is a saturating effect. Above some threshold surface expression level, mating efficiency is no longer effected by perturbations in expression. A semi-functional truncation of Mcl-1 (151-321) improved surface expression and enabled affinity-dependent mating[50] (Figure 3.11). BLI affinity measurements for the Mcl-1 truncation are not available, so mating efficiency was plotted against the affinity with full length Mcl-1. This explains why the apparent relationship between affinity and mating efficiency is different for all interactions involving the truncated protein.

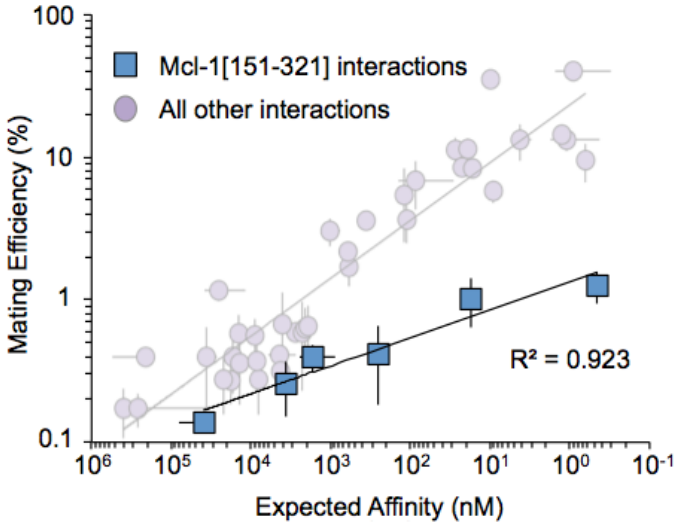


Figure 3.11: Synthetic agglutination mating efficiencies with a truncation of Mcl-1 [151-321]. Since BLI data is not available for the truncated variant, mating efficiency is plotted against the full length Mcl-1 affinity, which gives a strong log-linear relationship (r^2=0.923). The fit, however, is different from the bulk data that includes fully characterized protein variants (faded purple). Error bars represent standard deviation from three replicates.

To test the relationship between surface expression strength and mating efficiency, we constructed a Bcl-2 surface expression cassette with an inducible promoter, pZ4, containing a binding site for a zinc finger transcription factor. The Z4EV transcription factor was transformed into ySYNAGa for constitutive expression from the ACT1 promoter. Z4EV contains a zinc finger binding site for the Z4 promoter, an estradiol-binding domain for nuclear localization, and VP16 for recruitment of transcription machinery and activation of gene expression[51].

Strains were grown in a range of β-estradiol (βE) concentrations, and surface display was measured (Figure 3.12). The inducible strain grown in 0 nM βE showed no detectable surface display compared to non-expressing cells, indicating minimal leak of the inducible promoter. We found that surface expression strength was highly sensitive to βE at low concentrations, but saturation was reached by approximately 100 nM βE induction. At this concentration, surface expression levels were still lower than for a strain constitutively expressing Bcl-2 from the pGPD promoter. To maximize the range of surface display levels tested for mating efficiency, we included an inducible strain grown in βE concentrations ranging from 0 to 250 nM, and a constitutive Bcl-2 expressing strain.

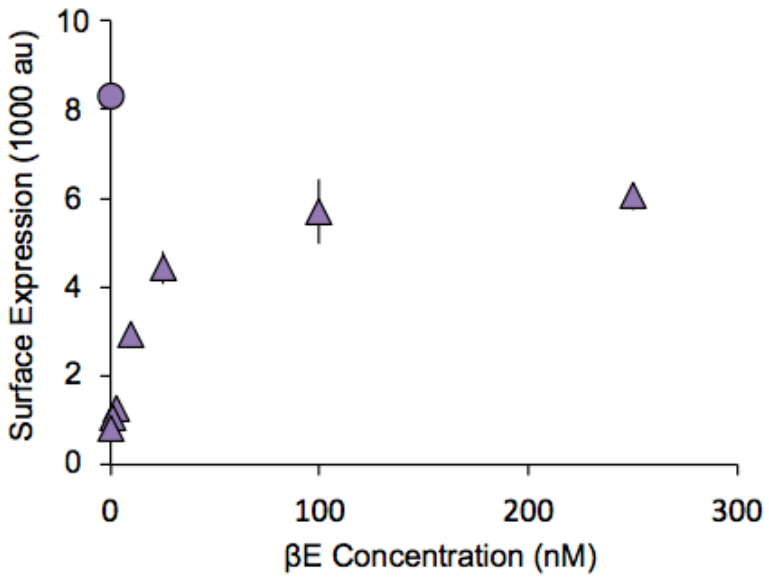


Figure 3.12: Surface expression strength in different βE concentrations. A yeast strain expressing Bcl-2 from an inducible promoter (triangles) or a constitutive promoter (circle) gave a wide range of surface concentrations of the binding protein. Error bars represent standard deviation from two replicates.

Mating efficiency is highly dependent on surface expression level only for low display concentrations (Figure 3.13A). A change in Bcl-2 surface display level from about 1,000 AU to about 8,000 AU, gave an over 80-fold difference in mating efficiency with αBCL2. However, saturation of mating efficiency is reached at a surface display level of approximately 4,000 AU. A change in display level from about 4,000 AU to about 8,000 AU gave a less than 2-fold difference in mating efficiency. This strongly supports our hypothesis that surface expression levels only have a major impact on mating efficiency at low levels of display, which means that a threshold should be applied for all new display strains. Mcl-1 was the only protein tested that gave a surface expression level below 4,000 AU, but a truncation was able to improve surface expression strength (Figure 3.13B).

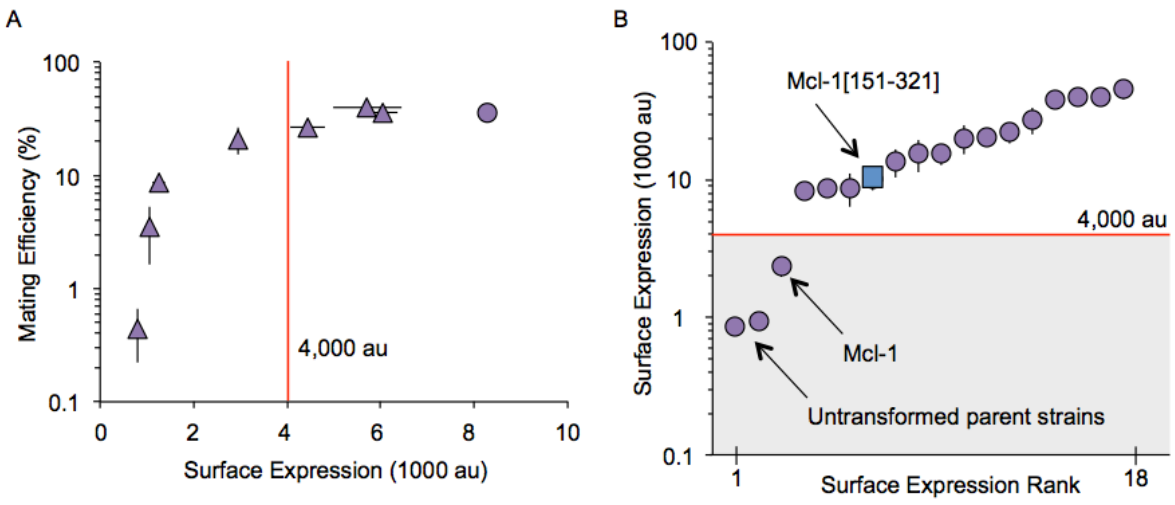


Figure 3.13: Surface expression strength characterization and effect on mating efficiency. (A) A yeast strain expressing Bcl-2 from an inducible promoter in different induction conditions (triangles) or a constitutive promoter (circle) show that mating efficiency is highly dependent on surface expression strength only at low expression levels. All matings were conducted with αBCL2. Error bars represent standard deviation from two replicates. (B) Surface expression characterization of ySYNAG strains. Only Mcl-1 has a surface expression strength below the threshold of 4,000 au. However, a semi-functional truncation, 151-321, improves surface expression strength.

3.9 DISCUSSION

Synthetic agglutination coupled with dual-channel flow cytometry provides a powerful new approach for rapidly characterizing the interaction strength of a pair of proteins without requiring costly reagents or recombinant protein purification. To date, a limited number of proteins have been tested with this system, many from the same protein family. Therefore, we cannot say with certainly that this approach can be generalized for use with all protein interactions. However, the protein family tested does contain considerable diversity. The tested proteins range from 26 to 206 amino acids and include highly stable *de novo* three helix bundles and unstructured BH3 domain peptides. Based on the biological mechanism of sexual agglutination, we expect that this approach will be capable of screening any binding protein that will properly display on the surface of yeast. Any interaction that induces agglutination will probably recover mating efficiency.

The BCL2 protein network was chosen to demonstrate and validate SynAg due to previous characterization that showed a wide range of interaction strengths. We expect any of the diverse classes of proteins that can be functionally expressed on the surface of yeast[52] to be compatible with SynAg. Some proteins do not functionally display on the yeast surface, and would therefore not be compatible with SynAg. For example, the detection of interactions requiring specific post-translational modifications may not be possible[53]. Additionally, SynAg is likely ill-suited for the screening of homodimer libraries. Oligomeric proteins are known to display on the yeast surface as functional assemblies[54], which means that homodimers would already be bound and not accessible for agglutination with a neighboring yeast cell expressing the same protein. Further studies are required to investigate these and other possible restrictions to SynAg.

An advantage of synthetic agglutination over intracellular assays, like yeast two-hybrid, is the ability to change the binding environment. Some binding environment effects were characterized here in the context of assay optimization. For example, we were able to show how specific and non-specific binding interactions are unevenly affected by changing concentrations of bulk media protein or surfactant. These experiments demonstrate that environmental effects can be quantitatively measured using synthetic agglutination. Based on these results, we expect that synthetic agglutination may have applications for screening dynamic protein interactions that are affected by environmental changes, such as pH or small-molecule dependent binders.

Chapter 4

LIBRARY-ON-LIBRARY CHARACTERIZATION OF PROTEIN INTERACTION NETWORKS

4.1 INTRODUCTION

Synthetic agglutination coupled with dual-channel flow cytometry provides a powerful tool for quantitatively characterizing the interaction strength of proteins without requiring costly reagents or the purification of recombinant proteins. However, only a single protein pair can be characterized per mating. This approach is feasible for the analysis of individual protein pairs of interest or small protein interaction networks, but does not easily scale for the analysis of hundreds or thousands of interactions. For screening large networks, such as mutagenic libraries or cellular interactomes, a multiplexed approach is required. Here, we describe a next-generation sequencing output for synthetic agglutination that enables the multiplexed characterization of thousands of protein interactions from a single batched mating (Figure 4.1). This approach was validated with the natural and *de novo* BCL2 interaction network, introduced previously. We then used synthetic agglutination to correctly identify affinity and specificity enhancing mutations from site-saturation mutagenesis (SSM) libraries, in which thousands of protein-protein interactions were characterized in a single batched mating. This highly multiplexed assay dramatically improves capabilities for characterizing proteins at a library-on-library scale and introduces new applications for synthetic agglutination.

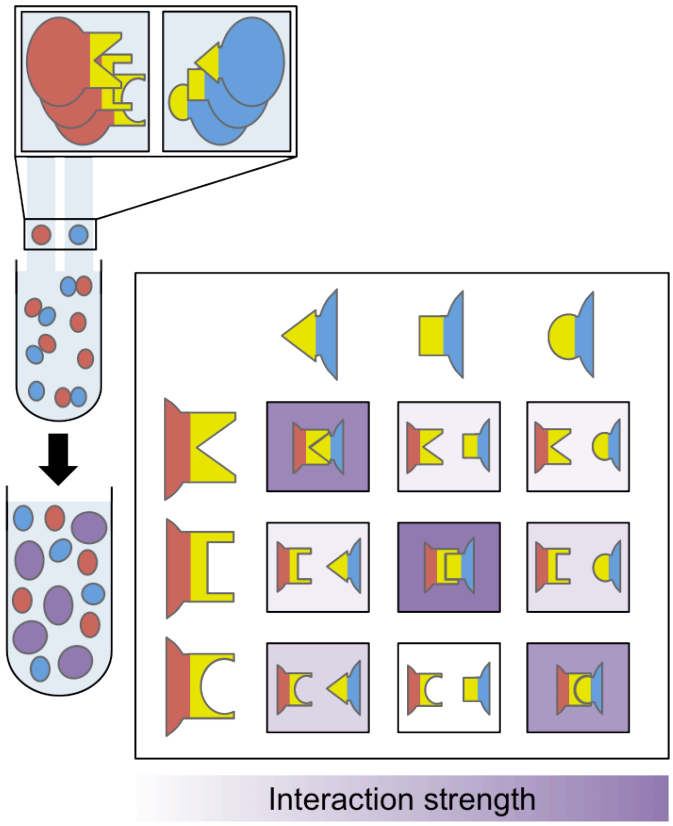


Figure 4.1: A synthetic agglutination batched mating. Multiple MATa and MATα strains expressing distinct synthetic agglutinin proteins on their surfaces are mixed in a single tube. Post-mating analysis with next generation sequencing is used to characterize the strength of each pairwise protein-protein interaction.

4.2 BATCHED MATING PROOF-OF-CONCEPT

After a batched mating, the diploid population distribution is proportional to the relative synthetic agglutination strength of each haploid pair. We previously demonstrated that the yeast sexual agglutinin proteins, Aga2 and Sag1, can be replaced with arbitrary binding proteins on the surface of yeast haploid cells and that mating recovery in liquid culture is dependent on the affinity of pairs of synthetic agglutinin proteins expressed on opposite mating types. It follows that when multiple MATa and MATα strains are mixed in a single tube, the frequency of mating for any pair of haploid cells depends on the interaction strength of their synthetic agglutinin proteins. To test this hypothesis, we performed a small batched mating with three MATa strains and three MATα strains with pairwise affinities ranging from 1 nM to 25 μM. Strain of each mating type were pooled with equal cell counts and then combined in a single culture. As before,

a total of 2.5 µL of pooled MATa culture and 5 µL of pooled MATα culture were mixed. Following a 17-hour mating, diploid isolation was achieved by streaking the culture on a double selection plate lacking both lysine and leucine. 22 colonies were then randomly selected. Colony PCRs were performed to isolate genomic DNA, which was then sequenced to determine the identity of both haploid synthetic agglutination proteins. We found that 18 of the 22 chosen diploids were formed from a mating of cells expressing the strongest binding pair (Figure 4.2). This supports the hypothesis that population distributions after a batched mating will reflect relative agglutination strengths. However, far more diploids would need to be sequenced in order to determine a relationship between diploid frequency and interactions strength.

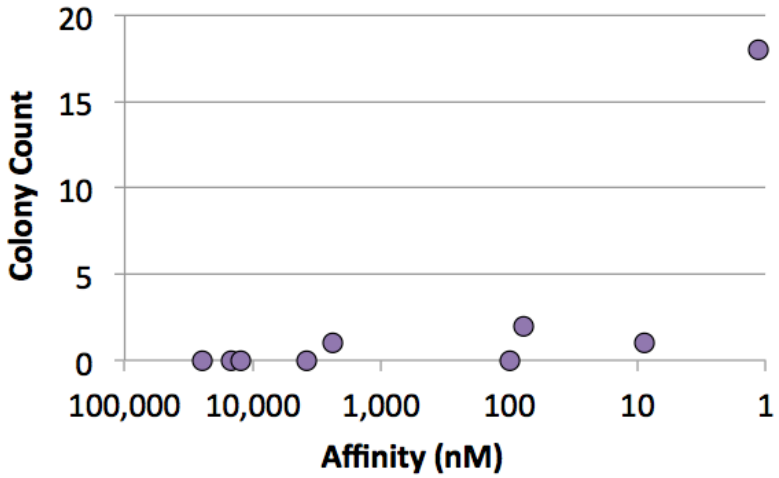


Figure 4.2: Diploid distribution by synthetic agglutinin affinity. The synthetic agglutinin protein pairs from 22 isogenic diploid strains formed during a batched mating were characterized by Sanger sequencing and plotted against binding affinity.

4.3 Barcoding and Recombination for High-Throughput Analysis

Sanger sequencing is a time consuming and expensive approach for characterizing a diploid population. Streaking a mating on a double selection plate for isogenic colony isolation adds days to the assay. Each chosen colony then requires two PCRs, one to amplify the surface expression cassette from each haploid, and both must be sequenced. The cost for just these steps is approximately $20 for each single colony. Instead, next-generation sequencing (NGS) can be used to screen millions of yeast strains at a time and achieve high-resolution characterization of a

diploid population. However, preparation for sequencing is not trivial. For each diploid cell, DNA from two different chromosomes must be sequenced and associated. Since the MATa and MATα haploid cells each contributed a surface expression cassette during mating, the identity of the MATa and MATα synthetic agglutinin proteins is split across two chromosomes in each diploid cell. To prepare a sample for sequencing, DNA must be isolated and amplified, which requires cell lysis. After lysis of a newly mated diploid population, it would no longer be possible to associate the two synthetic agglutinin genes that were present in a given cell. To solve this problem, one option was to use an emulsion ligation PCR to lyse each individual cell in isolation and combine DNA from two different chromosomes onto the same fragment[55]. However, this approach would be time consuming and would require expensive microfluidics equipment.

Instead, we developed a method for inducible chromosomal translocation that combines barcodes representing MATa and MATα synthetic agglutinin proteins into close proximity on the same chromosome in a live diploid cell. Our method uses a similar approach to previous work describing plasmid-based barcode recombination[16]. We began by constructing MATa and MATα parent strains, ySYNAGa and ySYNAGα, into which pools of barcoded surface display cassettes were transformed (Figure 4.3A). These strains include complementary lysine and leucine auxotrophic markers for diploid selection and express CRE recombinase[26,56] after mating and induction with β-Estradiol (βE)[57]. For small libraries, surface expression cassettes were assembled with isothermal assembly[58] in one of two standardized vectors, pSYNAGa or pSYNAGα, for integration into the corresponding parent yeast strain. In addition to a barcoded surface expression cassette, each vector backbone contains a mating type specific lox recombination site[59], Lox66 for MATa and Lox71 for MATα, and primer binding site, G1T1 for MATa and T2G2 for MATα. Sanger sequencing[60] was used during strain construction to match barcodes with their corresponding synthetic agglutinin protein. The construction of SynAg libraries is comparable in time and cost to the construction of yeast two-hybrid or yeast surface display libraries and identical methods can be used for DNA preparation and transformation.

Unidirectional CRE induced chromosomal translocation[61] in diploid cells results in the combining of barcodes representing two interacting SynAg proteins onto the same chromosome (Fig. 4.3A). After recombination, interacting SynAg proteins are identified from a mixed culture using Illumina next-generation sequencing (Figure 4.3B)[62]. To test the approach, haploid cells

containing SynAg cassettes were mated, induced with β-Estradiol, and lysed. The yeast lysate was used as a template for a PCR using mating type-specific primers. Amplification indicated that the primer binding sites had been combined onto one contiguous DNA strand, and hence that recombination had occurred. Sanger sequencing of the amplicon confirmed that recombination resulted in the expected chromosomal translocation.

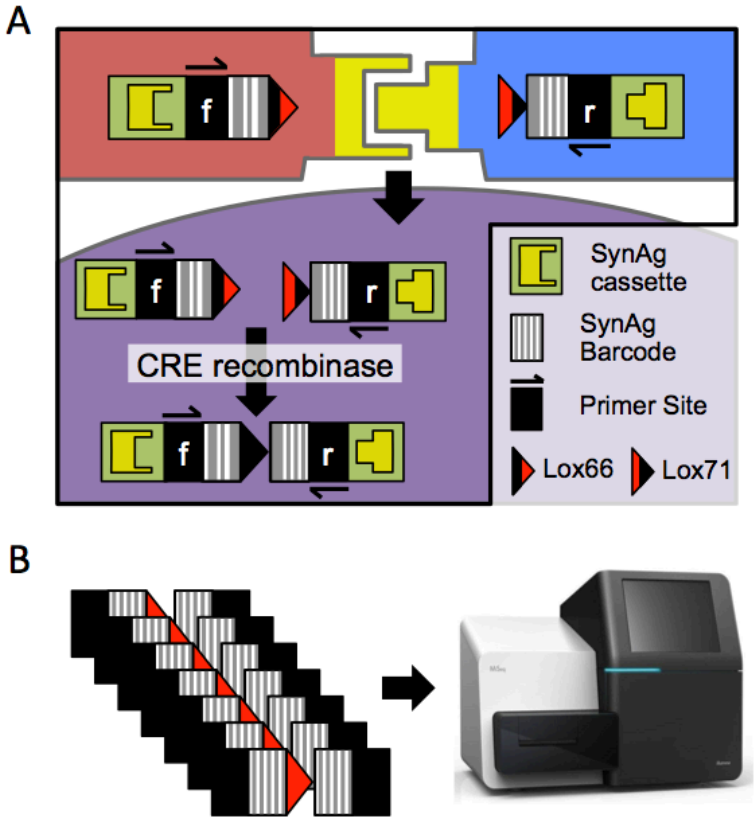


Figure 4.3: A recombination scheme to prepare samples for next-generation sequencing. (A) Each surface expression cassette is flanked by a unique barcode, a mating type specific primer binding site, and a lox recombination site. CRE recombinase expression in diploid cells causes a chromosomal translocation that pairs MATa and MATα barcodes on the same chromosome. (B) The mating type specific primers are used to amplify a fragment library containing both barcodes, which is analyzed with next-generation sequencing.

4.4 One-Pot Characterization of Manually Mixed Libraries

The frequency with which pairs of barcodes corresponding to interacting synthetic agglutinin proteins appear in diploid lysate following a batched mating was observed to be log-

linear with BLI affinity measurements. We constructed barcoded surface expression cassettes for six BCL2 family pro-survival proteins and nine engineered binders and measured the relative interaction frequencies of each possible interaction in a batched mating using next-generation sequencing. We observed a strong log-linear relationship ($R^2 = 0.87$) between mating efficiency and affinity across over five orders of magnitude of K_D (Figure 4.4A). The binding affinities ranged from below 500 pM to above 300 μM, which led to a more than a 500-fold difference in batched mating percent. In order to visualize the protein interaction network, we represented the same batched mating data with a sankey diagram (Figure 4.4B).

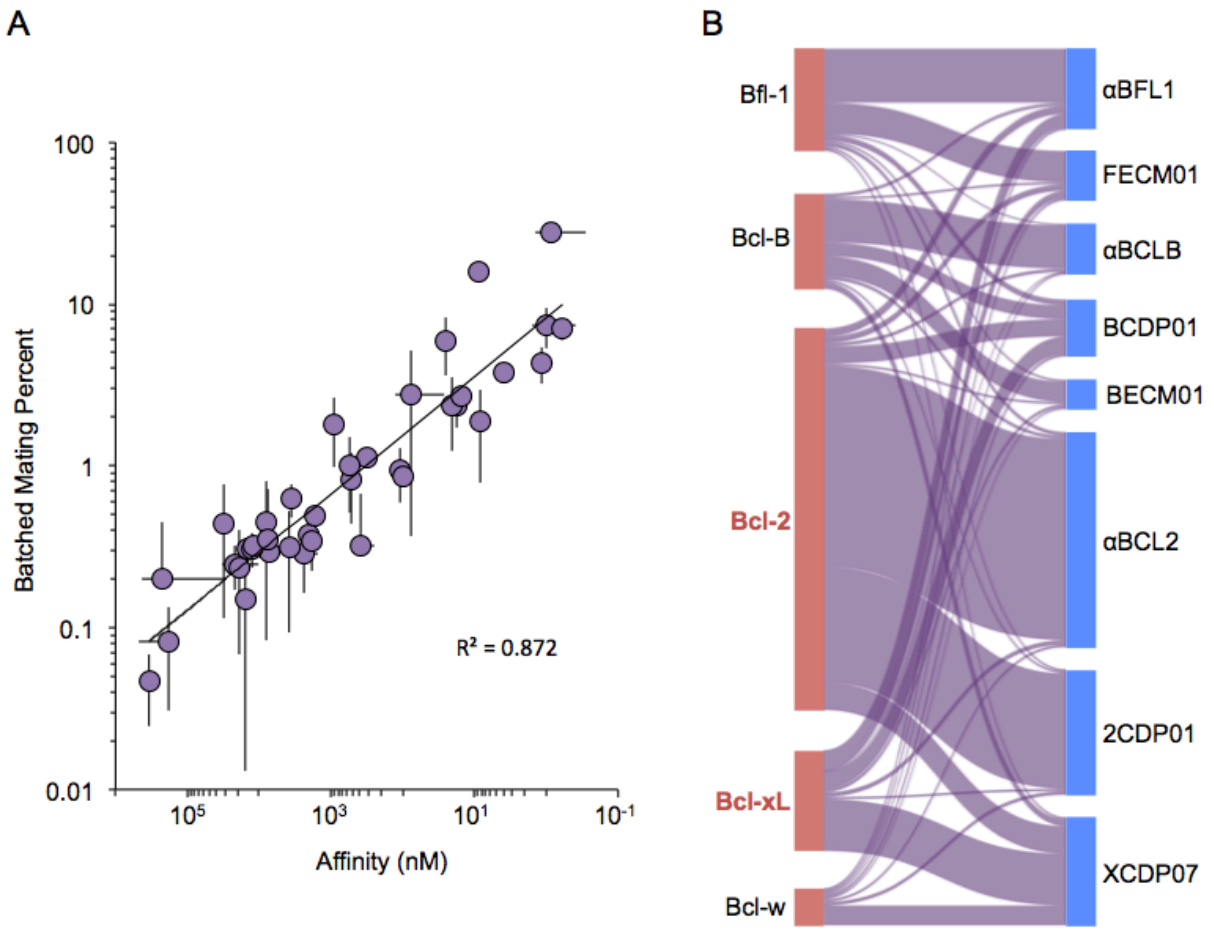


Figure 4.4: Validation of library-on-library synthetic agglutination. (A) A one-pot batched mating assay gave a strong log-linear relationship between the batched mating percent and affinity for synthetic agglutinin proteins with a K_D between 500 pM and 25 μM. Error bars represent standard deviation from two replicates. (B) The same batched mating was represented with a sankey diagram, where the height of each purple bar connecting two synthetic agglutinin proteins represents the relative number of matings that a particular protein interaction generated.

A comparison of the library-on-library and pairwise synthetic agglutination methods showed a near perfect 1:1 agreement. Qualitatively, the two methods revealed the same overall protein interaction network topology (Figure 4.5). To compare the two approaches quantitatively, pairwise mating efficiency was normalized so that the mating efficiency of all tested pairs summed to 100. In doing so, we converted pairwise mating efficiency to relative mating percent (Figure 4.6). A paired two-sided T-test of relative mating percent and batched mating percent gave a p-value of 0.80, indicating no statistically significant difference between the two methods.

		Bfl-1 AGTAGATCGT	**Bcl-B** TTATTACCAT	**Bcl-2** TCTGAATCAA	**Bcl-xL** GGTTCTATAA	**Bcl-w** CTCACGTGTG	**Mcl1 [151-321]** AATCCAACGA
αBFL1 GTCAACTATT	Pairwise (%)	12 ± 2.0	0.41 ± 0.24	3.5 ± 0.40	2.9 ± 0.62	0.39 ± 0.15	0.39 ± 0.08
	Batched (%)	7.4 ± 2.1	0.44 ± 0.32	1.1 ± 0.15	1.8 ± 0.84	0.35 ± 0.36	0.05 ± 0.04
FECM04 ATACCTGTAC	Pairwise (%)	13 ± 0.85	0.43 ± 0.04	3.5 ± 1.1	2.2 ± 0.04	0.33 ± 0.06	0.42 ± 0.24
	Batched (%)	4.3 ± 1.1	0.31 ± 0.22	0.86 ± 0.08	1.0 ± 0.50	0.32 ± 0.36	0.05 ± 0.04
αBCLB TGTAACTTGT	Pairwise (%)	0.17 ± 0.07	10 ± 2.3	0.17 ± 0.06	0.19 ± 0.10	0.20 ± 0.08	0.25 ± 0.06
	Batched (%)	0.16 ± 0.05	5.9 ± 2.3	0.46 ± 0.00	0.12 ± 0.02	0.25 ± 0.27	0.03 ± 0.03
BCDP01 GACTACGGGG	Pairwise (%)	1.7 ± 0.45	5.5 ± 0.92	11 ± 1.9	7.9 ± 0.52	0.61 ± 0.36	1.2 ± 0.31
	Batched (%)	0.81 ± 0.38	1.9 ± 1.1	2.4 ± 0.64	2.4 ± 1.10	0.37 ± 0.12	0.28 ± 0.03
BECM01 GCTATTCTGT	Pairwise (%)	0.61 ± 0.08	6.4 ± 2.3	0.60 ± 0.21	0.58 ± 0.16	0.29 ± 0.12	0.26 ± 0.11
	Batched (%)	0.28 ± 0.12	2.8 ± 2.4	0.31 ± 0.07	0.44 ± 0.36	0.23 ± 0.17	0.02 ± 0.00
αBCL2 CCGTAAGGCT	Pairwise (%)	0.19 ± 0.09	0.26 ± 0.09	35 ± 3.4	0.69 ± 0.45	0.65 ± 0.25	0.26 ± 0.08
	Batched (%)	0.52 ± 0.15	0.45 ± 0.10	27 ± 3.7	0.63 ± 0.14	0.34 ± 0.12	0.06 ± 0.04
2CDP06 GGGTGAGGTG	Pairwise (%)	0.18 ± 0.03	0.41 ± 0.18	31 ± 2.6	0.37 ± 0.18	0.67 ± 0.25	0.14 ± 0.02
	Batched (%)	0.29 ± 0.05	0.30 ± 0.01	16 ± 1.2	0.32 ± 0.06	0.49 ± 0.04	0.04 ± 0.03
XCDP07 TGTGGTAATG	Pairwise (%)	0.29 ± 0.13	5.1 ± 2.7	12 ± 3.3	8.8 ± 2.5	7.8 ± 0.82	1.0 ± 0.38
	Batched (%)	0.29 ± 0.04	0.94 ± 0.35	3.8 ± 0.01	7.1 ± 1.1	2.7 ± 0.15	0.16 ± 0.04
αMCL1 GGCGGGTGCG	Pairwise (%)	0.41 ± 0.04	1.2 ± 0.17	0.29 ± 0.09	0.19 ± 0.07	0.19 ± 0.05	4.6 ± 1.5
	Batched (%)	0.08 ± 0.05	0.25 ± 0.08	0.15 ± 0.14	0.05 ± 0.02	0.20 ± 0.25	0.63 ± 0.28

Pairwise Mating %	10	1	0.1	
Batched Mating %	10	1	0.1	0.01

Figure 4.5: Pairwise and batched mating percent for protein interactons involving 6 BCL2 homologues and 9 *de novo* binding proteins. For each interaction, the pairwise mating percent is given on top with an error of one standard deviation (n=3). The batched mating percent is given on the bottom with an error of one standard deviation (n=2). Shading provides a qualitative comparison between the two methods. The unique 10 bp barcodes used to represent each SynAg protein for next-generation sequencing are also shown.

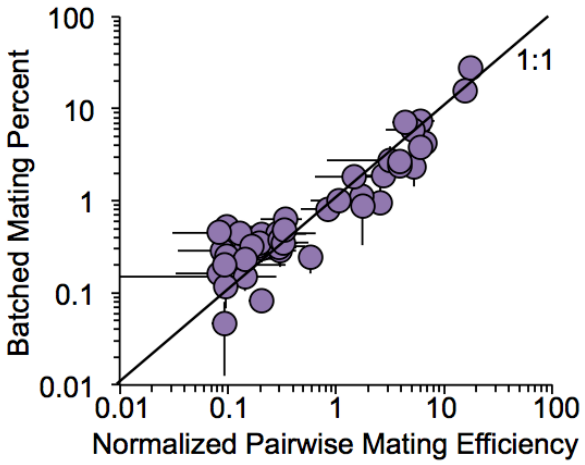


Figure 4.6: A comparison of pairwise and batched (library-on-library) SynAg approaches. Pairwise mating efficiency was converted to a percent by normalizing the total mating efficiency sum to 100. A line showing a 1:1 relationship is given for reference. Horizontal error bars represent standard deviation from 3 replicates. Vertical error bars represent standard deviation from 2 replicates.

In addition to the *de novo* binding proteins, seven pro-apoptotic BH3-only peptides with diverse binding profiles were added to a batched mating[63]. The interaction profile between these peptides and the five pro-survival homologues was consistent with previous work (Figure 4.7)[47]. For example, Noxa.BH3 was confirmed to bind Bfl-1 with high specificity (Figure 4.7B) and Puma.BH3 was confirmed to bind nonspecifically to Bcl-w, Bcl-xL, Bcl-2, and Bfl-1 (Figure 4.7C). Even Bad.BH3, which had been observed to interact the least overall, gave the expected interaction profile: relatively strong binding to Bcl-xL and Bcl-2, weak binding to Bcl-w, and minimal binding to Bcl-B and Bfl-1 (Figure 4.7D).

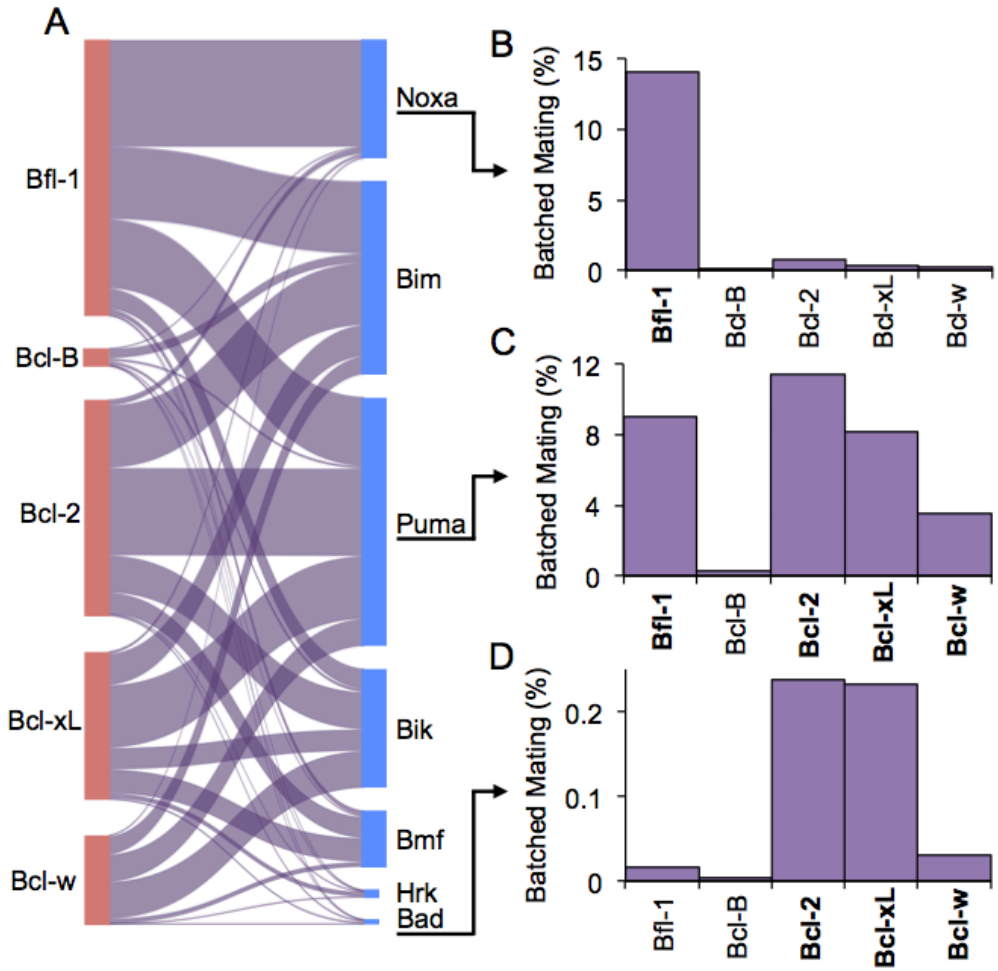


Figure 4.7: Batched mating characterization of native BCL2 protein interactions. (A) All interactions between five BCL2 pro-survival homologues and seven pro-apoptotic BH3-only peptides were characterized in a library-on-library mating assay. (B,C,D) Interaction profiles for (B) Noxa.BH3, (C) Puma.BH3, and (D) Bad.BH3 are shown in greater detail, with bolded targets indicating an affinity below 1 μM.

As before, the library-on-library and pairwise synthetic agglutination methods were qualitatively consistent in their characterization of all relative protein-protein interactions strengths (Figure 4.8).

		Bfl-1 AGTAGATCGT	Bcl-2 TCTGAATCAA	Bcl-xL GGTTCTATAA	Bcl-w CTCACGTGTG	Mcl1 [151-321] AATCCAACGA
Bim.BH3 GAGAGTACGG	Pairwise (%)	22 ± 2.9	25 ± 8.8	19 ± 8.0	12 ± 2.0	3.7 ± 0.77
	Batched (%)	9.6	8.3	4.2	2.4	1.9
Noxa.BH3 TCGTAAAGCG	Pairwise (%)	52 ± 7.9	0.25 ± 0.03	0.23 ± 0.07	0.25 ± 0.05	0.82 ± 0.04
	Batched (%)	14	0.18	0.37	0.23	0.09
Puma.BH3 AGGTGATCAT	Pairwise (%)	24 ± 1.4	31 ± 3.5	20 ± 5.0	8.3 ± 1.6	3.6 ± 0.44
	Batched (%)	9.2	12	8.3	3.6	0.57
Bad.BH3 CAGTTTTGTG	Pairwise (%)	0.24 ± 0.14	5.5 ± 1.4	8.7 ± 0.2	0.67 ± 0.24	0.47 ± 0.09
	Batched (%)	0.02	0.24	0.24	0.03	0.00
Bik.BH3 AGCTTGACAA	Pairwise (%)	11 ± 0.8	25 ± 11	5.8 ± 1.5	5.7 ± 1.4	1.2 ± 0.53
	Batched (%)	2.7	4.9	2.9	4.9	0.23
Hrk.BH3 GTAATGTACT	Pairwise (%)	3.6 ± 1.2	5.5 ± 1.1	11 ± 0.97	0.55 ± 0.10	0.42 ± 0.20
	Batched (%)	0.18	0.08	0.66	0.01	0.00
Bmf.BH3 TATCGAGTAT	Pairwise (%)	5.6 ± 2.0	19 ± 4.3	9.9 ± 2.3	3.1 ± 0.96	0.63 ± 0.15
	Batched (%)	0.68	2.8	3.2	0.66	0.04

Figure 4.8: Pairwise and batched mating percent for protein interactions involving 5 pro-survival BCL2 homologues and 7 pro-apoptotic BH3 only peptides. For each interaction, the pairwise mating percent is given on top with an error of one standard deviation (n=3). The batched mating percent is given on the bottom (n=1). Shading provides a qualitative comparison between the two methods. The unique 10 bp barcodes used to represent each SynAg protein for next-generation sequencing are also shown.

4.5 Library Construction with High-Efficiency Integrations

Integration of surface expression cassettes into the chromosomes of haploid yeast strains is essential to eliminate the need for plasmid selection during mating, but also decreases transformation efficiency. In order to construct large libraries, a nuclease assisted chromosomal integration approach[28] was used to achieve over 10,000 integrants in a single transformation and enabled efficient multi-fragment homologous recombination integrations for the construction of site-saturation mutagenesis libraries (Figure 4.9). Multi-fragment integrations are preferred for building surface display libraries because the protein of interest, barcode, and genomic homology regions are constructed separately. At the surface expression cassette integration locus, both ySYNAGa and ySYNAGα were transformed with a temporary cassette consisting of a pGAL promoter driving the expression of SceI endonuclease and a Cyc1 terminator flanked by SceI cut sites. Galactose induction prior to transformation with fragment libraries resulted in DNA nicking at the site of integration. By damaging the chromosome at the site of integration, efficiency was dramatically improved[64]. Next-generation sequencing of genomic DNA extracted

from yeast libraries was then used to pair each SynAg protein variant to its distinct barcode and to count relative barcode frequencies in the naïve library prior to mating.

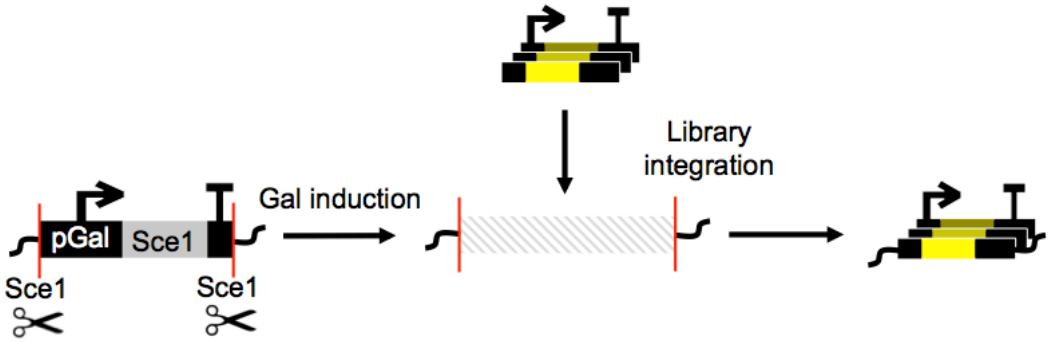


Figure 4.9: Nuclease assisted chromosomal integration for improved integration efficiency. The efficiency of chromosomal integration was increased with a galactose (Gal) inducible SceI endonuclease expression cassette flanked with SceI cut sites. Prior to transformation, competent yeast cells were grown in galactose to induce DNA damage at the integration locus, resulting in improved integration efficiency.

We tested the nuclease assisted chromosomal integration approach with single-fragment and four-fragment genomic integrations. Yeast competent cells containing the SceI endonuclease cassette were prepared with a 6-hour growth in 2% galactose media to induce DNA nicking. Control cells were prepared in dextrose media. We found that galactose induction improved the transformation efficiency of both the one and four fragment transformations by multiple orders of magnitude (Figure 4.10).

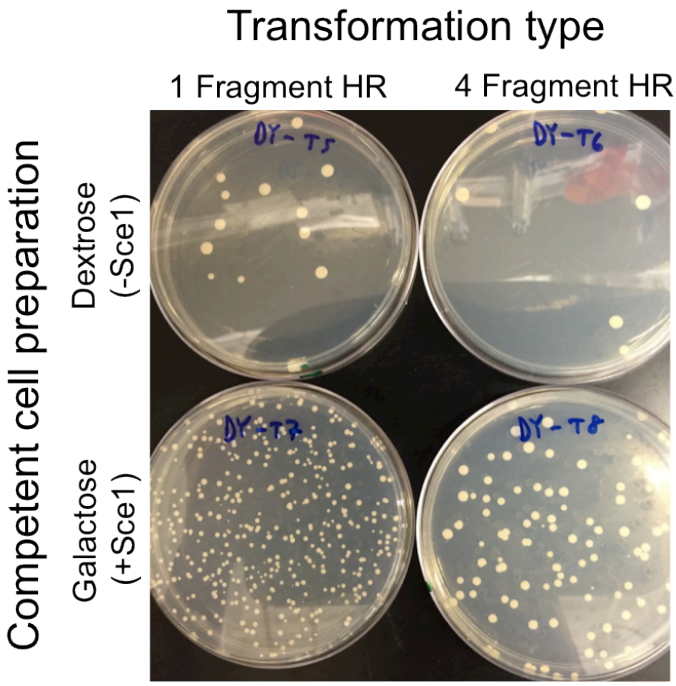


Figure 4.10: A comparison of one- and four-fragment homologous recombination efficiencies with and without induction of SceI endonuclease. Yeast competent cells were prepared in dextrose media or galactose media and then transformed with either a single fragment or four fragments with overlapping homologues for chromosomal integration. Transformations were diluted and spread onto a plate for selection.

The transformation efficiency improvement with the nuclease assisted chromosomal integration approach was sufficient for the construction of partial site-saturation mutagenesis libraries with over 1,000 members. Dilution spotting after transforming a library showed approximately 10,000-100,000 transformants (Figure 4.11). Applying the Clarke Carbon formula[65], in order to attain 95% coverage of a 1,000 member library, approximately 3,000 transformants are required.

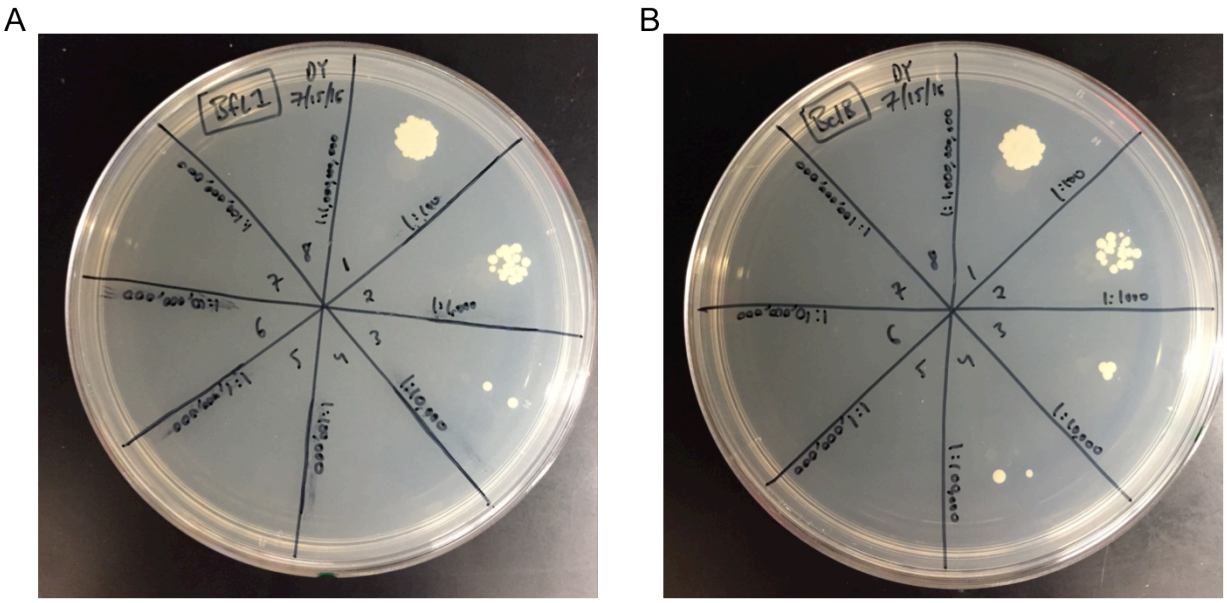


Figure 4.11: Example dilution spotting of high-efficiency transformations to measure the approximate number of transformants. Transformation of site-saturation mutagenesis libraries for (A) FCDP01 and (B) BCDP01 were diluted and spotted onto a plate to select for integrations. Both libraries gave approximately 10,000-100,000 transformants.

4.6 One-Pot Characterization of Mutagenic Libraries

A single-pot batched mating was used to characterize 7,000 distinct protein interactions. A partial site-saturation mutagenesis library of XCDP07[32] consisting of 1,400 distinct variants was characterized for interactions with five pro-survival BCL2 homologues, including the intended binding partner of XCDP07, Bcl-xL. For each variant, interaction strength (the number of times a particular variant was observed to have mated with Bcl-xL divided by the number of times that variant was observed in the naïve library) and specificity (the percent of observed matings with Bcl-xL minus the percent of observed matings with the next highest pro-survival BCL2 homologue) was determined. As a proof of principal, interactions involving variants with premature stop codons were analyzed (Figure 4.12). Only 8 of 55 premature stop codons included in the library resulted in even a single mating and only 6 resulted in more than 2 matings. All six variants contained stop codons at residue 93 or later, which leaves the central binding helix intact. Two variants, with stop codons at residues 113 and 114, showed improved interaction strength and specificity. These early terminations resulted in the removal of the C-

terminal myc tag from the 116-residue full-length protein, which may have negatively impacted binding.

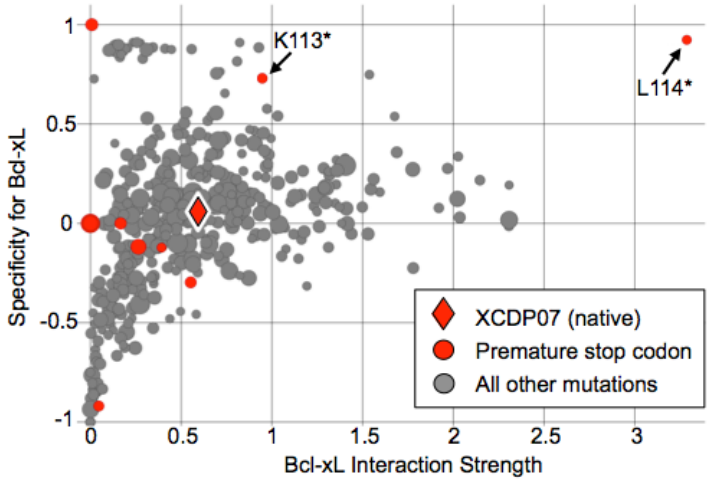


Figure 4.12: Stop codon analysis of an XCDP07 partial site-saturation mutagenesis library. Interaction strength versus specificity is plotted with highlighted premature stop codon variants. The diameter of each point is a function of its representation in the naïve library, which is used as a measure of confidence.

Favorable mutations from a yeast surface display library were correctly identified using library-on-library SynAg, but with additional information about relative binding affinities and specificities (Figure 4.13). In particular, two mutations at the interface periphery, L47R and A48T, were found to be favorable for interaction strength with Bcl-xL. Both mutations were enriched by fluorescence-activated cell sorting of an XCDP07 site-saturation mutagenesis surface display library incubated with fluorescently labeled Bcl-xL and unlabeled competitor homologues[32]. Unlike a traditional one-sided yeast surface display assay, SynAg provided detailed information about binding affinities and specificities to each target (Table 4.1). We observed moderately improved on target specificity for L47R, mostly through relative weakening of the interactions with Bcl-w and Bcl-B. We observed that A48T more dramatically weakened all off target interactions with a 16.5% increase of on-target binding.

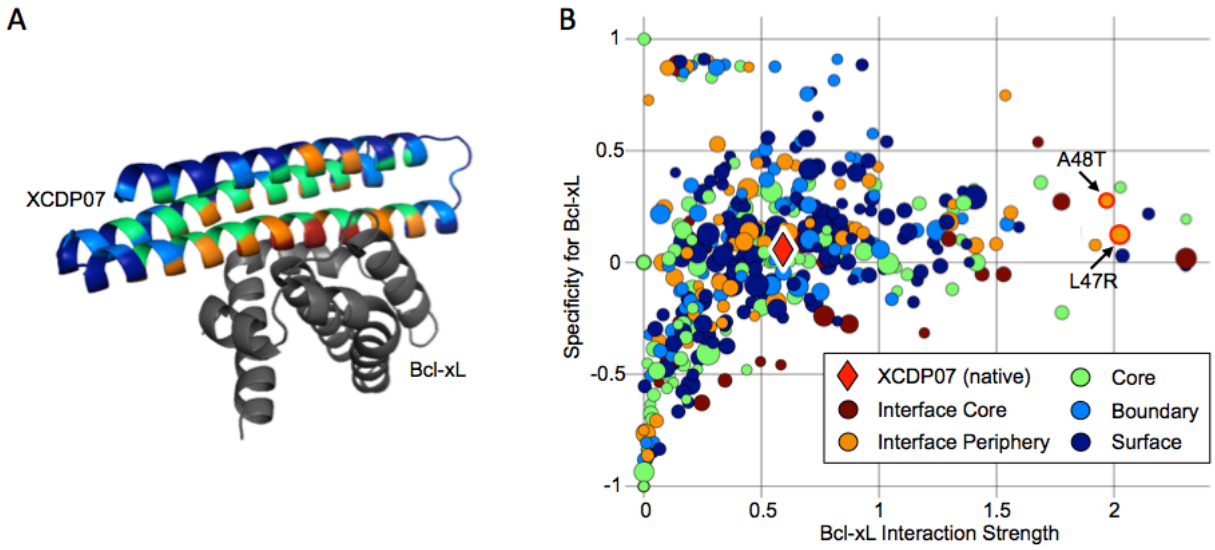


Figure 4.13: XCDP07 binding characterization. (A) A cartoon model of XCDP07 bound to its intended target, Bcl-xL. XCDP07 is colored by region. (B) Interaction strength and specificity of the XCDP07 partial site-saturation mutagenesis library is plotted with two confirmed affinity and specificity improving single amino acid mutations, L47R and A48T, highlighted.

Table 4.1: XCDP07 site-saturation mutagenesis variant comparisons. The mating percent for wild type XCDP07 and two single amino acid mutants with known enrichment from a one-sided yeast surface display screen, L47R and A48T, with each BCL2 homologue is listed.

	Bfl-1 %	Bcl-B %	Bcl-2 %	**Bcl-xL %**	Bcl-w %
WT	0.7	5.5	36.1	**42.1**	15.5
L47R	1.1	3.3	35.1	**47.8**	12.7
A48T	0.1	0.5	30.9	**58.6**	9.8

A single-pot batched mating was repeated with a site-saturation mutagenesis library of the 117-residue Bcl-2 binder A2CDP06. This library is less interesting for analysis with synthetic agglutination because the starting protein already exhibits a very high binding specificity for Bcl-2. However, a stop codon analysis confirmed that the method is again able to correctly identify enriching and non-enriching mutations (Figure 4.14). We were also able to find two affinity-improving mutations, K24R and G107R, that were previously identified with one-sided yeast surface display enrichment (Figure 4.15).

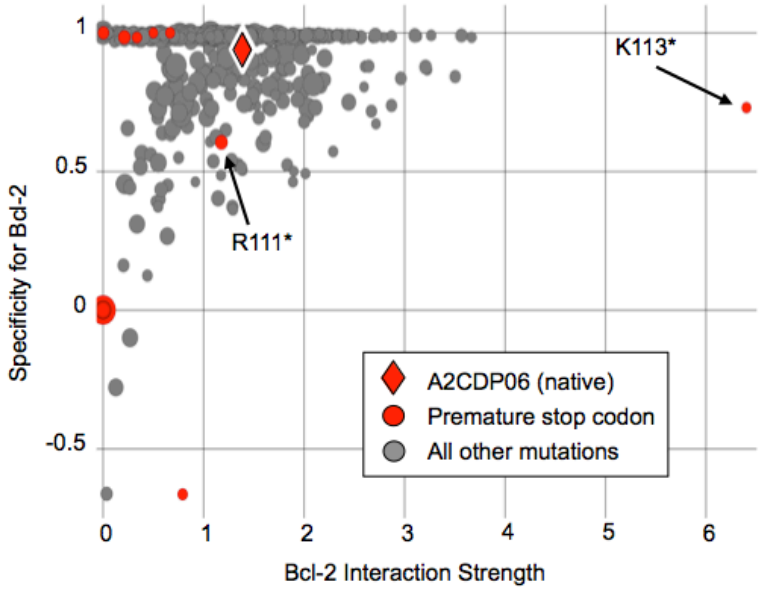


Figure 4.14: Stop codon analysis of an A2CDP06 partial site-saturation mutagenesis library. Interaction strength versus specificity is plotted with highlighted premature stop codon variants. The diameter of each point is a function of its representation in the naïve library, which is used as a measure of confidence.

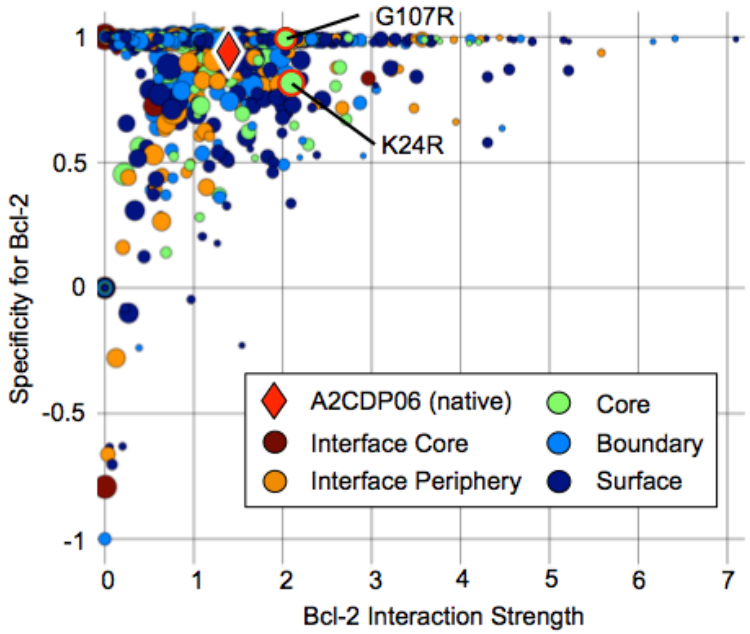


Figure 4.15: A2CDP07 binding characterization. Interaction strength and specificity of an A2CDP07 partial site-saturation mutagenesis library is plotted with two confirmed affinity-improving single amino acid mutations, K24R and G107R, highlighted.

4.7 DISCUSSION

Coupling synthetic agglutination with a next-generation sequencing output enables the quantitative analysis of thousands of protein-protein interactions in a single tube. This approach combines the throughput of a cell-based assay with the versatility and accuracy of a cell-free system. To date, a limited number of protein interactions have been characterized with this system, but we expect synthetic agglutination to work with any protein that can be functionally displayed on the surface of yeast. Besides the screening of engineered protein interaction networks and mutagenic libraries, synthetic agglutination may be a valuable tool for characterizing natural protein interaction networks and for the discovery of new interactions. Existing cell-based platforms, such as yeast two-hybrid, have largely been unsuccessful with the identification of extracellular interactions[19]. While the natural interactions included in this study were all intracellular, we expect that synthetic agglutination is particularly well suited for the characterization of extracellular proteins, since all proteins are being displayed on the surface of yeast cells.

In addition to protein interaction characterization, synthetic agglutination provides a unique ecological model for studying pre-zygotic genetic isolation. Previous work described the large diversity in sexual agglutination proteins across yeast species and suggested that co-evolution of these proteins may drive speciation by genetically isolating haploid pairs[9]. With synthetic agglutination, we have created a fully engineerable pre-zygotic barrier that can be used as a model to study complex ecological phenomena such as speciation and sexual selection, similar to the use of engineered *E. coli* for modeling predator-prey dynamics[66]. With a next-generation sequencing output, the population dynamics of libraries of mating type variants can be precisely tracked over generations.

When coupled with a next-generation sequencing output, synthetic agglutination provides a high-throughput platform for screening environment-responsive protein interactions. Engineered proteins that respond to environmental changes, such as pH, are valuable for the development of biosensors[67] and for drug delivery systems[68]. Synthetic agglutination may enable the rapid identification of functional variants using one-pot screening of design libraries rather than the current workflow of individually testing protein pairs. The same approach may be used to characterize the effect of arbitrary compounds on a network of protein interactions, such as

drug candidates that may enhance or inhibit interactions. Protein interaction inhibition is a powerful therapeutic strategy that has already been successfully employed for the treatment of cancers[69] and inflammation[70]. Yeast synthetic agglutination can be used to streamline pre-clinical drug testing workflows with multiplexed screening, which may enable simultaneous evaluation of drug efficacy and specificity.

Chapter 5

APPLYING SYNTHETIC AGGLUTINATION FOR DRUG CHARACTERIZATION

5.1 INTRODUCTION

The primary application of synthetic agglutination has been the characterization of protein interactions or protein interaction networks. Previously, we showed that media composition during a mating assay, such as bulk protein or surfactant concentration, has a dramatic effect on mating efficiency and that synthetic agglutination can be used to quantitatively characterize the effect of environmental changes. This principle can be extended with the addition of arbitrary compounds to the media during a mating, which may selectively enhance or inhibit specific protein interactions (Figure 5.1). As an example, we return to the interaction network consisting of pro-survival BCL2 homologues and a panel of *de novo* binders to demonstrate that the addition of a fully characterized peptide competitor preferentially disrupts the expected interactions. This platform has a potential commercial application for screening the effects of drug candidates on protein-protein interactions at a library-on-library scale.

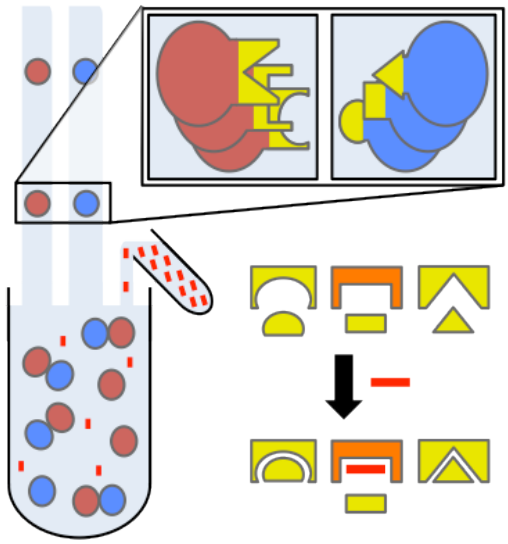


Figure 5.1: A cartoon depiction of mating environment manipulation. The library-on-library mating environment can be altered by adding an arbitrary molecule (red) that disrupt interactions involving particular proteins (orange) but not others (yellow).

5.2 MULTIPLEXED SCREENING OF PROTEIN INTERACTION INHIBITION

To demonstrate the characterization of a protein interaction network in a new extracellular environment, we added a soluble competitive binder at the start of a batched mating, which selectively inhibited certain interactions up to 800-fold. In the interaction network consisting of the pro-survival BCL2 homologues and their natural and *de novo* binding partners, one pro-survival peptide, Bad.BH3, bound predominantly to Bcl-xL and Bcl-2, weakly to Bcl-w, weaker still to Bfl-1, and minimally to Bcl-B (Figure 5.2). We added this peptide at a concentration of 100 nM to a batched mating consisting of these five pro-survival BCL2 homologues and eight *de novo* binding proteins and expected that protein interaction disruption would be proportional to Bad.BH3 interaction strength. Since Bcl-B showed no detectable interaction with Bad.BH3, batched matings with and without 100 nM Bad.BH3 were normalized to one another with the assumption that interactions involving Bcl-B were not affected by on-target binding. This normalization accounted for differences in total sequencing reads between conditions.

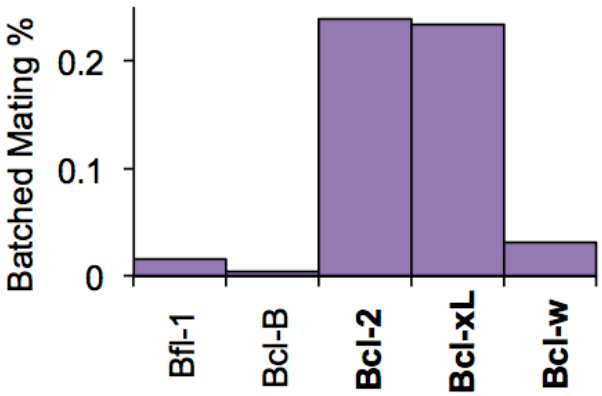


Figure 5.2: Interaction strength between Bad.BH3 and 5 pro-survival BCL2 homologues as measured with library-on-library SynAg.

The addition of 100 nM Bad.BH3 resulted in specific inhibition of interactions involving its expected binding partners: Bcl-2 and Bcl-xL (Figure 5.3). No change was observed for all strong protein-protein interactions involving pro-survival homologues that weakly interact with Bad.BH3: Bfl-1, Bcl-B, and Bcl-w. Weak protein-protein interactions involving these three homologues showed reduced mating efficiency in the presence of Bad.BH3, which can be

attributed to an increased concentration of bulk protein in the media that serves to block non-specific interactions. This was previously observed when bovine serum albumin (BSA) was added to the media during matings. Pairwise interactions involving Bcl-2 and Bcl-xL, however, were inhibited by at least 16-fold and up to 800-fold. Considered together, all protein interactions involving Bcl-xL and Bcl-2 were strongly inhibited, with normalized mating percent fold changes of 209 and 162, respectively (Figure 5.4). The weaker Bad.BH3 binders, Bcl-w and Bfl-1, displayed a normalized mating percent fold change of 2.6 and 1.5, respectively. All aggregate fold changes were consistent with previous characterization of Bad.BH3 interactions with the five pro-survival homologues and with previous work[47].

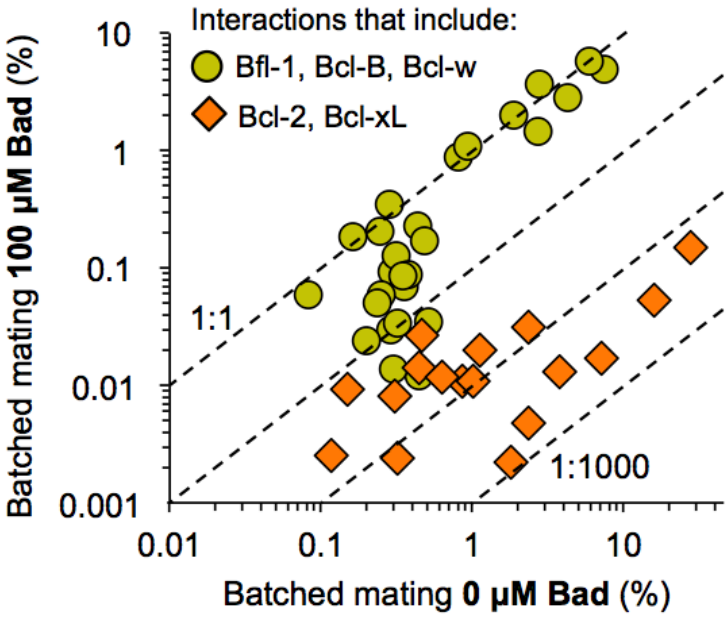


Figure 5.3: Selective inhibition with the addition of Bad.BH3 to a library-on-library SynAg mating. The addition of a competing peptide, Bad.BH3, to a mating between five pro-survival BCL2 homologues and a panel of *de novo* binders results in the isolated disruption of interactions involving Bcl-2 and Bcl-xL. Dashed lines representing 1, 10, 100, and 1000-fold differences between conditions are included for reference.

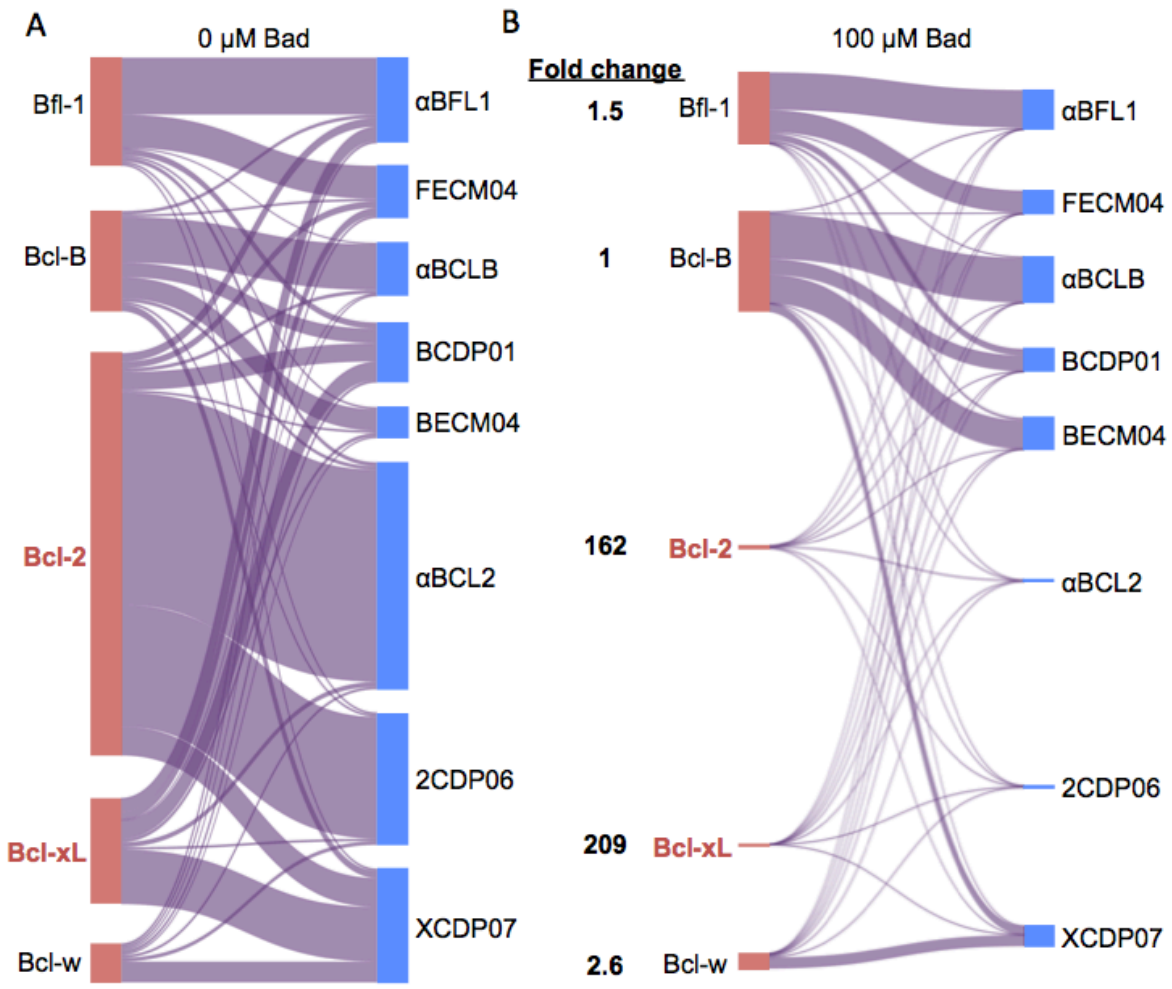


Figure 5.4: The effect of adding soluble Bim.BH3 to a library-on-library SynAg mating. Visualization of pro-survival BCL2 homologue – *de novo* binder interaction strengths (A) without and (B) with the addition of 100 μM Bad.BH3 to the media during a batched mating is shown. Fold changes are given for the aggregate interactions of each pro-survival BCL2 homologue.

5.3 Commercialization Potential

Introduction: Following the successful characterization of multiplexed protein interaction inhibition with Bim.BH3, we realized that synthetic agglutination could be applied for preclinical pharmaceutical drug screening. Specifically, a batched mating of a clinically relevant protein interaction network with a drug candidate added to the media can be used to characterize the effect on thousands of protein interactions in a single pot to screen for possible toxicity of a drug

candidate. For commercial development applications, we call our platform AlphaSeq. This platform technology will enable pharmaceutical companies to rapidly screen their drug candidates for off-target effects on protein interactions that may otherwise not present themselves until late into clinical trials or even after a drug is on the market.

Problem: The failure rate for a new drug entering clinical trials is about 90%, which has driven the average cost for developing a new drug to over $2.5 billion[71]. This enormous cost is passed on to patients in the form of costly medical bills and inflated insurance premiums. In order to minimize the risk of failure, pharmaceutical companies use *in vitro* toxicity assays to measure off-target effects on cellular receptors, enzymes, and protein interactions[72]. It has been suggested by representatives at pharmaceutical companies that comprehensive toxicity screening could identify 75% of the current failures before clinical trials[73]. However, due to the low-throughput and high cost of current screening technologies, only a small number of "high-value" off-target effects are selected for preclinical screening (Figure 5.5). For example, Eurofins, a leading contract research organization (CRO) for preclinical drug characterization, offers a "safety screen" package, which includes analysis of 87 possible off-target effects. This service costs approximately $80,000 for one drug candidate and has a one-month turnaround. With the full human protein interactome consisting of thousands of protein interactions, the approaches used by CROs today are too expensive and slow to enable comprehensive coverage[74,75]. The result is a major gap in preclinical drug characterization, leading to the frequent discovery of toxicity during or after clinical trials.

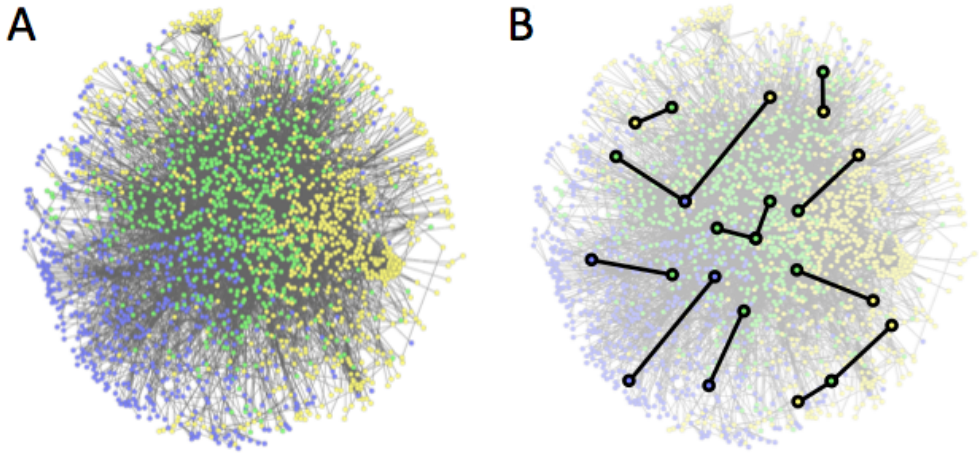


Figure 5.5: Visualization of the human protein interactome. (A) The complete interactome and (B) a representative subset of interactions that are currently considered during pre-clinical off-target screening are shown.

The development of all new pharmaceuticals would benefit from a protein interaction toxicity screen, but the need is particularly severe for an expanding class of drugs that function by selectively inhibiting disease-causing protein-protein interactions. Most human proteins have common structural motifs that they share with other closely related proteins. Therefore, by designing a drug to target a particular structural element on the surface of a protein, pharmaceutical companies must ensure that all related protein interactions are unaffected. The lack of a multiplexed off-target screening platform has contributed to the slow development of protein interaction inhibiting drugs despite known druggable targets for cancers, autoimmune diseases, infectious diseases, and more[76].

Solution: We have developed AlphaSeq, the first technology to enable comprehensive screening of protein interactions for preclinical drug toxicity testing. Using synthetic biology methods, we have reprogrammed yeast cells to reduce the challenging task of determining a protein interaction network to the ease of sequencing DNA. Put simply, we attach proteins to the outside of yeast cells and observe how the cells stick to each other, which allows us to infer protein interaction strength. When a drug candidate is added to the media, we can quantitatively determine the effect on each protein interaction. With AlphaSeq, pharmaceutical companies will no longer need to compromise with preclinical characterization of off-target protein interactions.

This disruptive technology will enable screening for toxic drug effects that would otherwise not be discovered until late in clinical trials or even after a drug is on the market.

Market: The *in vitro* toxicology testing market for small molecules is nearly $10B and growing at 14.7% annually, according to BCC research. If biologics are included, the market nearly doubles. According to clinicaltrials.gov there are currently over 120,000 registered clinical trials for drugs, and each requires multiple drug candidates to be screened for toxicity. There is also a sizable academic market for compound off-target characterization as well as for characterizing new protein interaction networks.

AlphaSeq provides an additional benefit for a growing class of drugs that function by deliberately disrupting protein interactions. To date, most approved drugs are small molecules that target receptors or enzymes and allow for simple experiments to screen for efficacy. Drugs that act by inhibiting protein interactions have been referred to as the "holy grail" of drug discovery, due to therapeutic possibilities for numerous diseases[77]. For this particular class of drugs, AlphaSeq can be used to simultaneously screen for efficacy and toxicity, making it an ideal platform for an emerging market.

Competition: AlphaSeq provides a strong advantage over competing technologies for the screening of protein interaction inhibition by drug candidates. Current methods used by pharmaceutical companies and CROs for probing protein interactions are ill-suited for screening whole interaction networks. Biolayer interferometry (BLI)[21] is the current gold standard, which requires purified recombinant proteins, expensive antibodies, and a separate assay for each protein pair. AlphaSeq requires no recombinant protein purification, low cost reagents, and can be used to screen thousands of interactions in a single tube. Additionally, we have demonstrated a comparable accuracy to BLI (Figure 5.6).

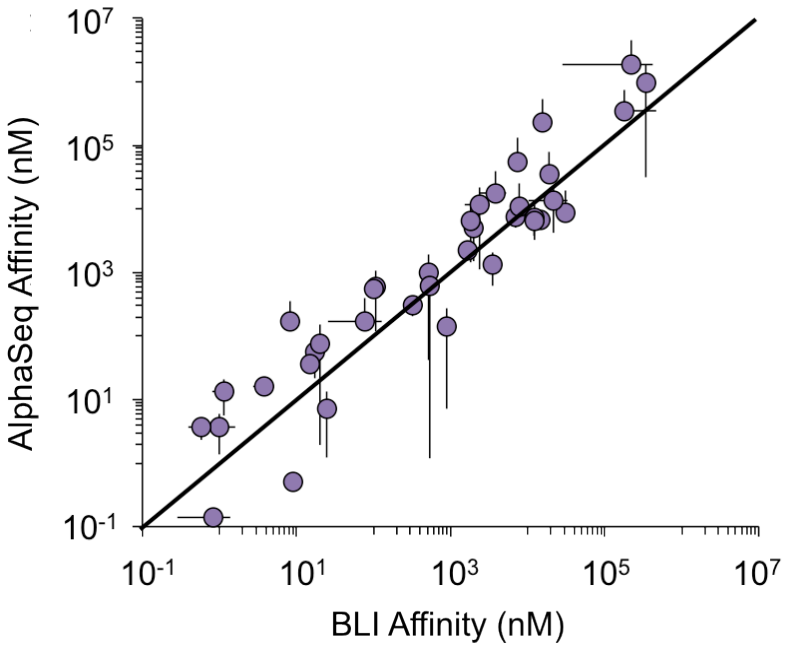


Figure 5.6: Comparing AlphaSeq to the best in class competitor, biolayer interferometry (BLI).

Possible competitors include CROs like Eurofins, ProteinLinks, and Fluofarma who are contracted by pharmaceutical companies for drug screening services, including off-target toxicity screens for receptors, enzymes, and protein interactions. Today, pharmaceutical companies are paying CROs approximately $80,000 to test 8 drug candidates against 12 protein interaction targets, each with a 10-point dilution series. With current technologies, the cost of comprehensive screening against thousands of protein interactions is prohibitively expensive and time consuming. With AlphaSeq, the same 8 drug candidates can be screened against thousands of protein interactions for less than half the cost. The improvement in throughput provided by AlphaSeq brings comprehensive preclinical drug screening within reach. While we will be competing with the services currently offered by CROs, we hope to partner with these companies and have them offer AlphaSeq as part of their recommended toxicology-screening pipeline.

Proposed business model: Our proposed business model is to offer a kit/service hybrid. Pharmaceutical companies would purchase an AlphaSeq kit, consisting of a 96-well plate pre-loaded with yeast libraries, and have their trusted CRO load their drug candidates in parallel with other toxicology assays. After an overnight incubation, the CRO would return the kit to us where we would perform all downstream processing and return a detailed report to the pharmaceutical

company. We would either charge a flat rate for each drug candidate screened or negotiate a price for a drug development project.

Intellectual property: A patent was filed through UW in JAN 2017 (USPTO # 15/407,215).

Chapter 6

MATERIALS, METHODS, AND SUPPLEMENTARY INFORMATION

6.1 MATERIALS AND METHODS

DNA construction: Isogenic fragments for yeast transformation or plasmid assembly were PCR amplified from existing plasmids or yeast genomic DNA with Kapa polymerase (Kapa Biosystems), gel extracted from a plasmid digest (Qiagen), or synthesized by a commercial supplier (Integrated DNA Technologies). All plasmids were constructed with isothermal assembly[58] and verified with Sanger sequencing[60]. Sequencing was typically performed for at least the open reading frame, unless a functional screen (such as for binding or fluorescence) could be performed following yeast strain construction. MATa and MATα surface expression cassette plasmids were assembled using a four-piece assembly, including two backbone fragments, a surface expression cassette fragment, and a barcode-containing fragment (Figure 6.1). Site-saturated mutagenesis (SSM) library DNA was prepared with overlap PCR[32] using Phusion polymerase and custom NNK primers for each codon.

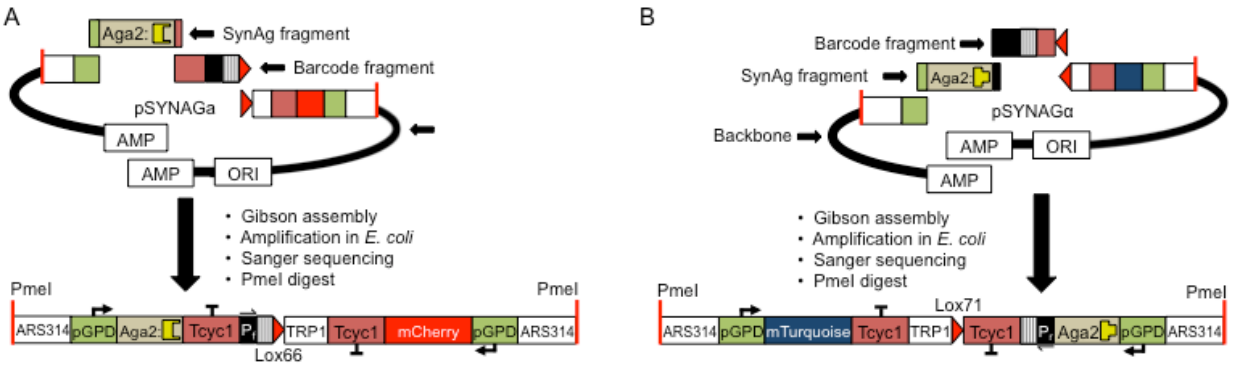


Figure 6.1: Construction strategy for SYNAG plasmids. (A) pSYNAGa and (B) pSYNAGα are each constructed with a four fragment Gibson assembly. The fragments include a synthetic agglutinin protein (SAP) fragment with standard overhangs, a barcode containing fragment amplified with a degenerate primer, and two backbone fragments. Following transformation into *E. coli*, plasmid open reading frames and barcodes are sequenced. Verified plasmids are digested with the restriction enzyme PmeI and transformed into ySYNAGa or ySYNAGα.

In total, over 700 fragments and over 500 plasmids were constructed for this study. Table 6.1 lists the plasmids used for the construction of key control and experimental yeast strains. All fragment and plasmid descriptions are available in the Aquarium software, described below.

Table 6.1: Plasmids used for constructing key control and experimental yeast strains. Protein truncations are indicated in parenthesis.

Plasmid Name	Gene Cassette(s)	Marker	Integration Locus
pMOD_NatMX_HIS_ pGPD-Aga1	pGPD-Aga1	NatMX	HIS
pMOD_BleoMX_LTR2_ pGPD-mChe	pGPD-mCherry	BleoMX	LTR2
pMOD_BleoMX_LTR2_ pGPD-mTur	pGPD-mTurquoise	BleoMX	LTR2
pYMOD_URA_SAG1_KO	Knockout	URA	SAG1
pYMOD_KanMX_YCR043_pZ4-CRE	pZ4-CRE	KanMX	YCR043
pYMOD_KanMX_YCR043_pACT1-Z4EV	pACT1-Z4EV	KanMX	YCR043
pYMOD_BleoMX_ARS314_ pGAL-Sce1	pGAL-Sce1	BleoMX	ARS314
pSYNAGa_Bfl-1 (1-153)	pGPD-mCherry & pGPD-Aga2-Bfl-1	TRP	ARS314
pSYNAGa_Bcl-B (1-165)	pGPD-mCherry & pGPD-Aga2-Bcl-B	TRP	ARS314
pSYNAGa_Bcl-2	pGPD-mCherry & pGPD-Aga2-Bcl-2	TRP	ARS314
pSYNAGa_Bcl-w (1-182)	pGPD-mCherry & pGPD-Aga2-Bcl-w	TRP	ARS314
pSYNAGa_Bcl-xL	pGPD-mCherry & pGPD-Aga2-Bcl-xL	TRP	ARS314
pSYNAGa_Mcl-1 (172-327)	pGPD-mCherry & pGPD-Aga2-Mcl-1	TRP	ARS314
pSYNAGa_Mcl-1(151-321)	pGPD-mCherry & pGPD-Aga2-Mcl-1[151-321]	TRP	ARS314
pSYNAGa[pZ4]_Bcl-2	pGPD-mCherry & pZ4-Aga2-Bcl-2	TRP	ARS314
pSYNAGα_Bim.BH3 (141-166)	pGPD-mTurquoise & pGPD-Aga2-Bim.BH3	TRP	ARS314
pSYNAGα_Noxa.BH3 (18-46)	pGPD-mTurquoise & pGPD-Aga2-Noxa.BH3	TRP	ARS314
pSYNAGα_Puma.BH3 (129-158)	pGPD-mTurquoise & pGPD-Aga2-Puma.BH3	TRP	ARS314
pSYNAGα_Bad.BH3 (103-131)	pGPD-mTurquoise & pGPD-Aga2-Bad.BH3	TRP	ARS314
pSYNAGα_Bik.BH3 (50-78)	pGPD-mTurquoise & pGPD-Aga2-Bik.BH3	TRP	ARS314
pSYNAGα_Hrk.BH3 (26-54)	pGPD-mTurquoise & pGPD-Aga2-Hrk.BH3	TRP	ARS314
pSYNAGα_Bmf.BH3 (126-154)	pGPD-mTurquoise & pGPD-Aga2-Bmf.BH3	TRP	ARS314
pSYNAGα_αBFL1	pGPD-mTurquoise & pGPD-Aga2-αBFL1	TRP	ARS314
pSYNAGα_FECM04	pGPD-mTurquoise & pGPD-Aga2-FECM04	TRP	ARS314
pSYNAGα_αBCLB	pGPD-mTurquoise & pGPD-Aga2-αBCLB	TRP	ARS314
pSYNAGα_BCDP01	pGPD-mTurquoise & pGPD-Aga2-BCDP01	TRP	ARS314
pSYNAGα_BECM01	pGPD-mTurquoise & pGPD-Aga2-BECM01	TRP	ARS314
pSYNAGα_αBCL2	pGPD-mTurquoise & pGPD-Aga2-αBCL2	TRP	ARS314
pSYNAGα_2CDP06	pGPD-mTurquoise & pGPD-Aga2-2CDP06	TRP	ARS314
pSYNAGα_XCDP07	pGPD-mTurquoise & pGPD-Aga2-XCDP07	TRP	ARS314
pSYNAGα_αMCL1	pGPD-mTurquoise & pGPD-Aga2-αMCL1	TRP	ARS314

Yeast Methods: Unless otherwise noted, yeast transformations were performed with a standard lithium acetate transformation[78] using approximately 300 ng of plasmid digested with PmeI. Yeast Peptone Dextrose (YPD), Yeast Peptone Galactose (YPG), and Synthetic Drop Out (SDO) medium supplemented with 80 mg/mL adenine were made according to standard protocols. Saturated yeast cultures were prepared by inoculating 3 mL of YPD from a freshly struck plate and growing for 24 hours at 30°C.

Yeast Strain Construction: A MATα variant of the EBY100[12] strain was constructed with mating, sporulation, tetrad dissection, and screening with selectable markers[79]. EBY100a was mated with a leucine prototroph W303α variant. Following sporulation, positive selection was performed for HIS, LEU, and URA and replica plating was used to identify MATα haploids auxotrophic for lys and trp. Numerous selections and transformations were then performed to construct all strains used for this study (Figure 6.2). Plating on 5-FOA was used to select strains with URA3 inactivating mutations[80]. Final strains were constructed with many rounds of chromosomal integrations, each consisting of a single transformation, auxotrophic or antibiotic selection, and PCR to verify integration into the expected locus. One final transformation of ySYNAGa or ySYNAGα with pSYNAGa or pSYNAGα, respectively, was used to build final experimental strains for SynAg assays (Figure 6.3).

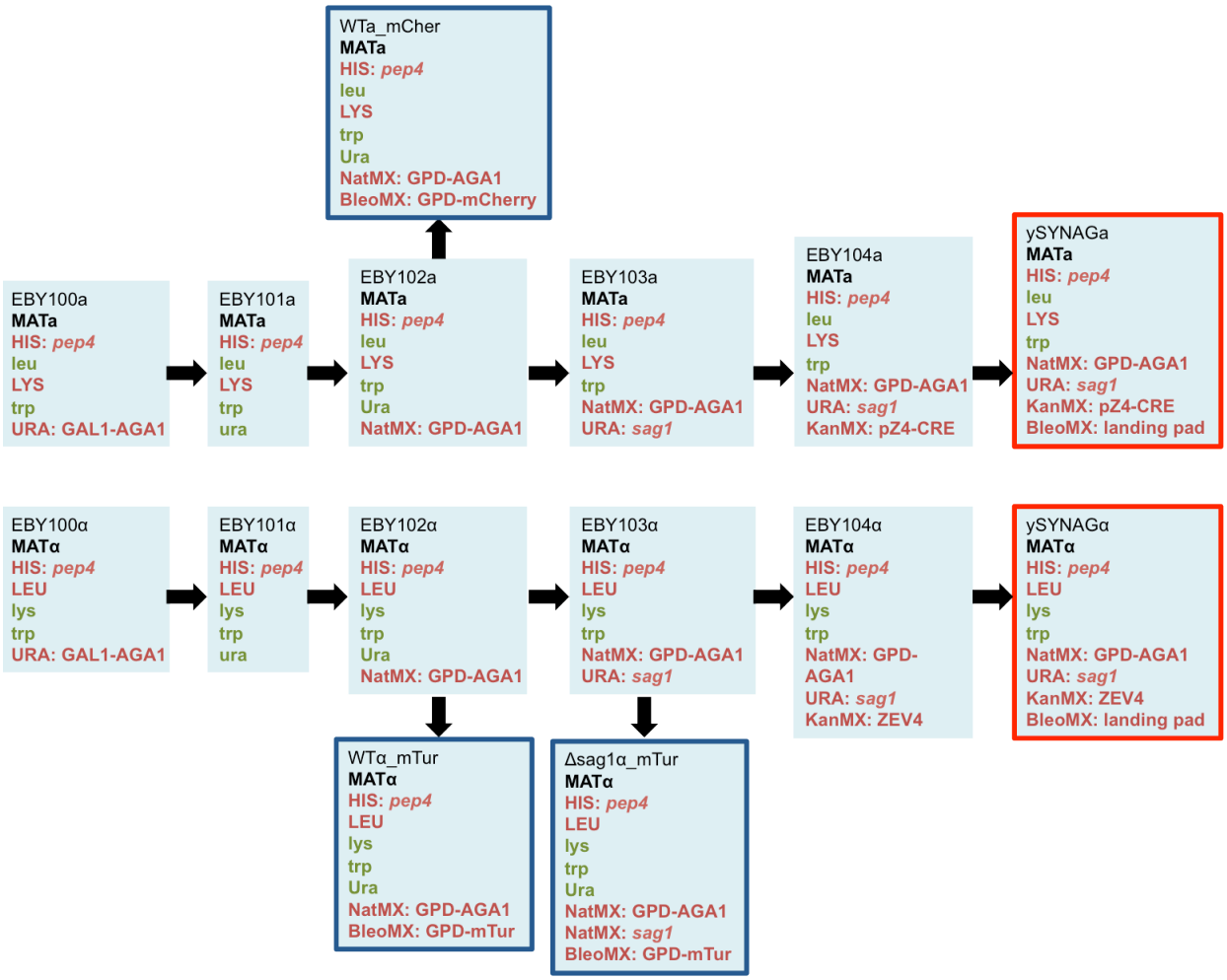


Figure 6.2: Yeast strain construction lineage flowchart. The distinct genetic modifications for construction of primary experimental and parent strains are shown. Strains used for positive and negative mating controls are outlined in blue. A mating with both wild type (WT) strains was used as a positive control for agglutination. A mating with WT MATa and a Sag1 deficient MATα was used as a negative control for agglutination. Parent strains for synthetic agglutination are outlined in red.

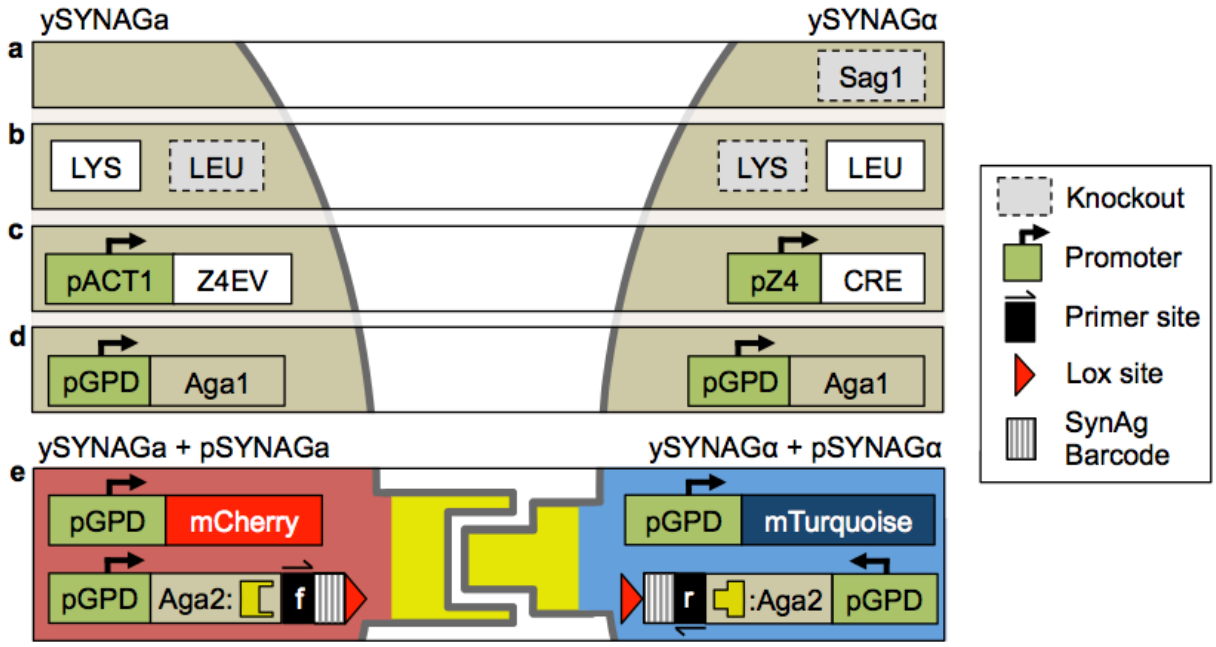


Figure 6.3: Yeast strain genetic components. (A) The α-agglutinin, Sag1, is knocked out to eliminate wild-type agglutination. (B) ySYNAGa and ySYNAGα have complementary lysine and leucine markers for diploid selection. (C) ySYNAGa cells constitutively express a β-Estradiol inducible transcription factor, Z4EV, which activates the pZ4 promoter for CRE recombinase expression in diploid cells[51]. (D) ySYNAGa and ySYNAGα constitutively express Aga1 for yeast surface display. The strongest *S. cerevisiae*-native constitutive promoter, pGPD, was chosen to maximize expression. (E) A transformation with pSYNAGa or pSYNAGα adds a constitutively expressed fluorescent reporter and SynAg protein fused to Aga2. A recombination site, barcode, and primer binding site flank the SynAg expression cassettes.

Site-saturation mutagenesis libraries were transformed into yeast using nuclease assisted chromosomal integration. Prior to transformation, parent yeast strains were grown in YPG media for five hours. Growth in galactose media induced SceI expression and caused DNA damage at the integration site. 100 µL of cell pellet, rather than 10 µL for a standard transformation, was used for each library transformation and all other reagents were scaled up accordingly. Four fragments, approximately 2 µg of each, were added to each transformation. The fragments included two mating type specific adaptor fragments, a synthetic agglutination protein site-saturation mutagenesis library fragment, and a barcode library containing fragment (Figure 6.4). Following transformation, cells were washed in 5 mL YPD and resuspended in YPD to a total volume of 5 mL. 100 µL were immediately removed and a dilution series was plated on SDO-trp to quantify the total number of transformants in the library. The remaining culture was grown

for 5 hours, washed twice with 5 mL SDO-trp, and grown in 20 mL SDO-trp overnight to select for transformants. 2 mL 25% glycerol aliquots were then prepared for later use.

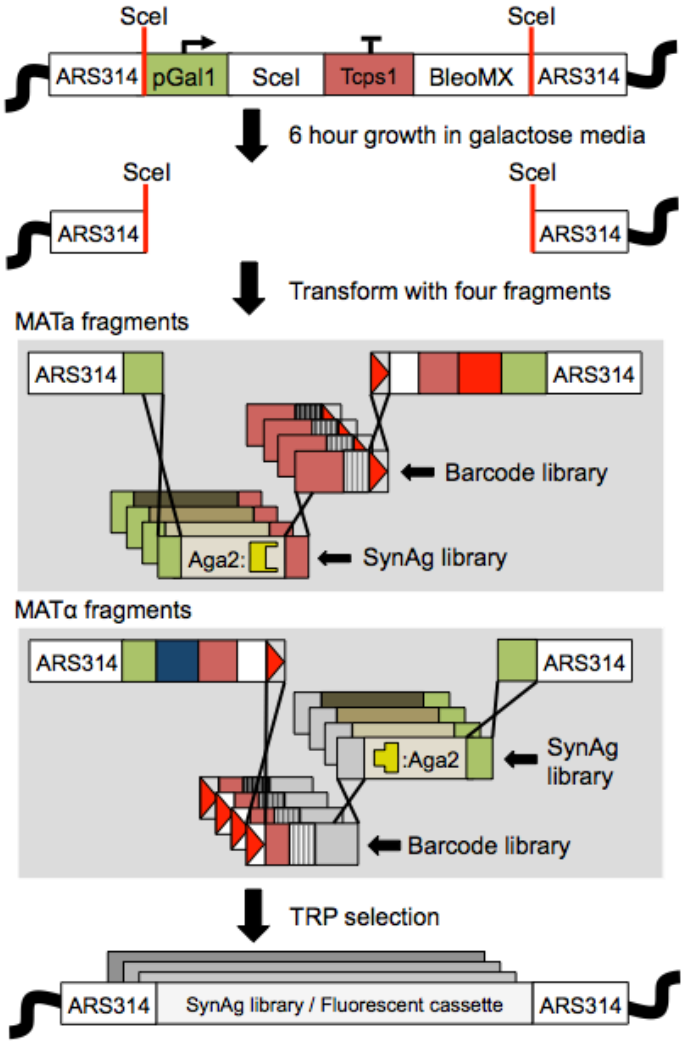


Figure 6.4: Construction strategy for surface expression ySYNAG yeast strains. For library integrations, ySYNAGa and ySYNAGα were first grown for 6 hours in GAL media to induce SceI expression causing DNA damage at the integration site. Cells were then transformed with four mating type dependent fragments, which assemble with homologous recombination and are selected with TRP.

Table 6.2: Key control and experimental yeast strains.

Strain Name	Description	Parent	Transformant
EBY100a	Yeast surface display optimized strain		
W303αMOD	MATα for generation of EBY100α		
EBY100α	MATα version of yeast surface display strain	Mating and sporulation: EBY100a and W303αMOD	
EBY101a	URA knockout with 5-FOA selection	EBY100a	
EBY101α	URA knockout with 5-FOA selection	EBY100α	
EBY102a	Constitutive expression of Aga1	EBY101a	pMOD_NatMX_HIS_pGPD-Aga1
EBY102α	Constitutive expression of Aga1	EBY101α	pMOD_NatMX_HIS_pGPD-Aga1
WTa_mCher	MATa, Constitutive mCherry with WT SAG1	EBY102a	pMOD_BleoMX_LTR2_pGPD-mChe
WTα_mTur	MATα, Constitutive mTurquoise with WT SAG1	EBY102α	pMOD_BleoMX_LTR2_pGPD-mTur
EBY103α	MATα, Sag1 knockout	EBY102α	pYMOD_URA_KO_SAG1
Δsag1α_mTur	MATα, Constitutive mTurquoise with SAG1 KO	EBY103α	pMOD_BleoMX_LTR2_pGPD_mTur
EBY104a	MATa, Constitutive Z4EV transcription factor	EBY102a	pYMOD_KanMX_YCR043_pACT1-Z4EV
EBY104α	MATα, Inducible CRE recombinase	EBY103α	pYMOD_KanMX_YCR043_pZ4-CRE
ySYNAGa	Final MATa parent strain, with Sce1 cassette	EBY104a	pYMOD_BleoMX_ARS314_pGAL-Sce1
ySYNAGα	Final MATα parent strain, with Sce1 cassette	EBY104α	pYMOD_BleoMX_ARS314_pGAL-Sce1
ySYNAGa_Bfl-1	MATa haploids expressing pro-survival BCL2 homologues used in pairwise and batched mating assays	ySYNAGa	pSYNAGa_Bfl-1
ySYNAGa_Bcl-B		ySYNAGa	pSYNAGa_Bcl-B
ySYNAGa_Bcl-2		ySYNAGa	pSYNAGa_Bcl-2
ySYNAGa_Bcl-w		ySYNAGa	pSYNAGa_Bcl-w
ySYNAGa_Bcl-xL		ySYNAGa	pSYNAGa_Bcl-xL
ySYNAGa_Mcl-1		ySYNAGa	pSYNAGa_Mcl-1
ySYNAGa_Mcl-1[151-321]		ySYNAGa	pSYNAGa_Mcl-1[151-321]
ySYNAGα_Bim.BH3	MATα haploids expressing pro-apoptotic peptides or *de novo* binding proteins used in pairwise and batched mating assays	ySYNAGα	pSYNAGα_Bim.BH3
ySYNAGα_Noxa.BH3		ySYNAGα	pSYNAGα_Noxa.BH3
ySYNAGα_Puma.MH3		ySYNAGα	pSYNAGα_Puma.BH3
ySYNAGα_Bad.BH3		ySYNAGα	pSYNAGα_Bad.BH3
ySYNAGα_Bik.BH3		ySYNAGα	pSYNAGα_Bik.BH3
ySYNAGα_Hrk.BH3		ySYNAGα	pSYNAGα_Hrk.BH3
ySYNAGα_Bmf.BH3		ySYNAGα	pSYNAGα_Bmf.BH3
ySYNAGα_αBFL1		ySYNAGα	pSYNAGα_αBFL1
ySYNAGα_FECM04		ySYNAGα	pSYNAGα_FECM04
ySYNAGα_αBCLB		ySYNAGα	pSYNAGα_αBCLB
ySYNAGα_BCDP01		ySYNAGα	pSYNAGα_BCDP01
ySYNAGα_BECM01		ySYNAGα	pSYNAGα_BECM01
ySYNAGα_αBCL2		ySYNAGα	pSYNAGα_αBCL2
ySYNAGα_2CDP06		ySYNAGα	pSYNAGα_2CDP06
ySYNAGα_XCDP07		ySYNAGα	pSYNAGα_XCDP07
ySYNAGα_αMCL1		ySYNAGα	pSYNAGα_αMCL1
ySYNAGα_XCDP07 [SSM_library]	MATα haploid site-saturation mutagenesis (SSM) library for XCDP07	ySYNAGα	4 piece homologous recombination
ySYNAGα_2CDP06 [SSM_library]	MATα haploid site-saturation mutagenesis (SSM) library for A2CDP06	ySYNAGα	4 piece homologous recombination

Peptide construction and purification: DNA encoding the BH3 domain of Bad (Bcl-2 agonist of cell death protein; residues 103-131) was synthesized by Integrated DNA Technologies and inserted into a modified pMAL-c5x vector resulting in an N-terminal fusion to maltose binding protein and a C-terminal 6-histidine tag. The vector was transformed into BL21(DE3)* *E.*

coli (NEB) for protein expression. Protein was purified from soluble lysate first with nickel affinity chromatography (NiNTA resin from Qiagen), then by size exclusion chromatography (Superdex 75 10/300 GL; GE). Purified protein was concentrated via centrifugal filter (Millipore), snap-frozen in liquid nitrogen, and stored at -80°C.

Surface expression screening: Yeast strains were grown separately in 3 mL YPD media from a fresh plate for 24 hours. 10 µL were washed with 1 mL PBSF, incubated in 50 µL PBSF with 1 µg FITC-anti-myc antibody (Immunology Consultants Laboratory, Inc.) for 1 hour at 22°C, washed with 100 µL PBSF, and read with the FL1.A channel on an Accuri C6 cytometer. For testing inducible surface expression, an additional liquid culture growth step was added to ensure that equilibrium surface expression was reached. The inducible strain was grown in 3 mL YPD media + βE (variable concentration) from a fresh plate for 24 hours. 5 µL were then transferred to a fresh 3 mL YPD culture containing the same βE concentration, grown for an additional 24 hours, and prepared for surface expression screening as before.

Colony count mating assays: Yeast strains were grown separately in 3 mL synthetic complete (SC) media from a fresh plate for 24 hours. 15 µL of a MATa strain and 15 µL of a MATα strain were combined in 3 mL SC media and grown for 5 hours. 5 µL of the mixed culture was struck onto an SC-Lys-Leu plate and allowed to grow for 48 hours.

Pairwise (Two-Strain) Mating Assays: Yeast strains were grown separately in 3 mL YPD media from a fresh plate for 24 hours. 2.5 µL of a saturated MATa culture and 5 µL of a saturated MATα culture were combined in 3 mL of YPD media and incubated at 30°C and 275 RPM for 17 hours. 5 µL from the mixed culture were diluted in 1 mL of water and cellular expression of mCherry and mTurquoise was characterized with a Miltenyi MACSQuant VYB cytometer using channels Y2 and V1, respectively. A standard yeast gate was applied to all cytometry data and Flowjo was used for analysis and visualization.

Yeast Library Preparation: Pre-characterized yeast libraries were prepared by combining individually transformed isogenic yeast strains with validated surface expression and known barcodes, determined with Sanger sequencing (Table 6.3). Yeast strains were grown separately

in 3 mL YPD media from a fresh plate for 24 hours. Strains of the same mating type were pooled with equal cell counts of each isogenic strain. Cell counts were measured with an Accuri c6 flow cytometer.

Table 6.3: Yeast strains used in pairwise matings and next-generation sequencing. Strain ID numbers and barcode sequences are listed.

Strain name	Mating type	Strain ID number	Barcode
WTa_mCher	MATa	3555	
WTα_mTur	MATalpha	3556	
Δaga2a_mCher	MATa	3557	
Δsag1α_mTur	MATalpha	3558	
ySYNAGa_Bfl1	MATa	11743	AGTAGATCGT
ySYNAGa_BclB	MATa	11744	TTATTACCAT
ySYNAGa_Bcl2	MATa	11745	TCTGAATCAA
ySYNAGa_BclW	MATa	11748	GGTTCTATAA
ySYNAGa_BclXL	MATa	16314	CTCACGTGTG
ySYNAGa_Mcl1[151-321]	MATa	16076	AATCCAACGA
ySYNAGα_FINDI-F21	MATalpha	11339	GTCAACTATT
ySYNAGα_FINDI-F30D	MATalpha	11340	ATACCTGTAC
ySYNAGα_BINDI-B+	MATalpha	11342	TGTAACTTGT
ySYNAGα_BINDI-BCDP01	MATalpha	11343	GACTACGGGG
ySYNAGα_BINDI-B40A	MATalpha	11344	GCTATTCTGT
ySYNAGα_2INDI-2+	MATalpha	11345	CCGTAAGGCT
ySYNAGα_2INDI-4LVT	MATalpha	11346	GGGTGAGGTG
ySYNAGα_WINDI-aBclW	MATalpha	11351	TGTGGTAATG
ySYNAGα_XINDI-XCDP07	MATalpha	11352	GGCGGGTGCG
ySYNAGα_Bim.BH3	MATalpha	11331	GAGAGTACGG
ySYNAGα_Noxa.BH3	MATalpha	11333	TCGTAAAGCG
ySYNAGα_Puma.MH3	MATalpha	11334	AGGTGATCAT
ySYNAGα_Bad.BH3	MATalpha	11335	CAGTTTTGTG
ySYNAGα_Bik.BH3	MATalpha	11336	AGCTTGACAA
ySYNAGα_Hrk.BH3	MATalpha	11337	GTAATGTACT
ySYNAGα_Bmf.BH3	MATalpha	11338	TATCGAGTAT

Uncharacterized yeast site-saturation mutagenesis libraries were constructed with nuclease assisted chromosomal integration and large volume transformation, as described above. Prior to mating, libraries were characterized using next-generation sequencing to map each library variant with its 10 bp barcode and to determine relative counts of each variant in the naïve library population. One 2 mL glycerol stock was thawed, washed once with 1 mL YPD, and grown in 50 mL YPD for 24 hours. Genomic DNA was then prepared for next-generation sequencing.

Library Mating Assays: 2.5 μL of a MATa library and 5 μL of a MATα library were combined in 3 mL of YPD and incubated at 30°C and 275 RPM for 17 hours. When characterizing interactions in the presence of Bad.BH3, the peptide was added at a concentration of 100 nM to the 3 mL YPD culture. Following the 17-hour incubation, 1 mL was washed twice in 1 mL SDO-lys-leu and transferred to 50 mL SDO-lys-leu with 100 nM β-estradiol (βE) for diploid selection and induction of CRE recombinase. After 24 hours of growth, genomic DNA was prepared for next-generation sequencing.

Preparation for Next-generation sequencing: 50 mL yeast cultures were harvested by centrifugation and lysed by heating to 70°C for 10 min in 2 mL 200 mM LiOAc and 1% SDS[81]. Cellular debris was removed with centrifugation and the supernatant was incubated at 37°C for 4 hours with 0.05 mg/mL RNase A. An ethanol precipitation was performed to purify and concentrate the genomic DNA and a 2% agarose gel was run to verify genomic DNA extraction. Two rounds of qPCR were performed to amplify a fragment pool from the genomic DNA and to add standard Illumina sequencing adaptors and assay specific index barcodes. For the primary PCR, different primers were used for naïve library characterization and post-mating characterization. An index barcode was added in the secondary PCR with the reverse primer. Both PCRs were terminated before saturation in order to minimize PCR bias. The first PCR was run for 25-30 cycles, and the second PCR was run for 5-7 cycles. The final amplified fragment was gel extracted, quantified with a Qubit and sequenced with a MiSeq sequencer (Illumina). A 600-cycle v3 reagent kit was used for naïve library characterization and a 150-cycle v3 reagent kit was used for post-mating characterization.

Sequence analysis: Pre-mated site-saturation mutagenesis libraries were sequenced in order to match each variant with a 10 bp barcode and to determine the relative population size of each variant. All sequences were first filtered for quality by requiring a perfect match for 15 bp in a constant region immediately before and after the mutated gene. Forward and reverse reads were stitched together and full site-saturation mutagenesis coding regions were translated to amino acid sequences. Sequences were then grouped by their 10 bp barcode and a consensus amino acid sequence was determined for each group. Only groups with zero or one amino acid mutation were kept. Groups representing the same amino acid mutation were then pooled. The number of

sequences in each pooled group provided naïve library counts for each site-saturation mutagenesis variant and the barcodes attributed to each pooled group were used for later matching of mated diploids to their mutation.

Post-mating sequences were first filtered for quality by requiring a perfect match for 10 bp in a constant region immediately before and after both barcodes. Barcodes from forward and reverse reads were then isolated and replaced with the protein variant they were previously found to represent. A dataframe with interaction counts for every possible pairwise interaction was generated. Python was used for all site-saturation mutagenesis library analysis and visualization.

6.2 DATA

All library-on-library screening data is available on NCBI:

BioProject Accession number: PRJNA380247
BioSample Accession numbers: SAMN06642476, SAMN06642477, SAMN06642478, SAMN06642479, SAMN06642480, SAMN06642481, SAMN06642482, SAMN06642483, SAMN06642484, SAMN06642485

6.3 CODE

All code for library-on-library analysis is fully available on GitHub:
https://github.com/dyounger/yeast_synthetic_agglutination
Sankey diagrams were generated with sankeyMATIC (http://sankeymatic.com/)

6.4 MODULAR PLASMID CLONING SCHEME

A novel modular plasmid cloning scheme, called pYMOD, was developed in order to make many of the plasmids described above. While other yeast destination vectors had been developed, we needed a vector that was compatible with two expression cassettes, could contain any marker, and could be targeted to any locus. Since so many genetic modifications were required for the construction of the SYNAG yeast strains, we would have quickly run out of

available selectable markers using another cloning scheme. Additionally, two-cassette integrations considerably reduce the cloning time.

pYMOD consists of eleven completely modular components that can be cloned as separate fragments and combined in to a plasmid with isothermal assembly[58] (Figure 6.5). Many markers have been built into the pYMOD system (URA, TRP, HIS, LEU, KanMX, BleoMX, NatMX, HygMX) and we have shown that arbitrary loci can be target with two 500-nucleotide homology components. PME1 restriction sites flanking the homology component are included to linearize the plasmid for integration. A convergent dual cassette architecture separated by the selection marker component was chosen based on experimental validation. Standard linkers are included between each component for fragment modularity and ease of assembly. If only a single cassette is desired, the downstream homology part can be amplified with the MCT2 linker to eliminate cassette 2. For genomic knockouts, the upstream homology part can be amplified with the T1MC linker to eliminate cassette 1 and the downstream homology part can be amplified with the MCT2 linker to eliminate cassette 2.

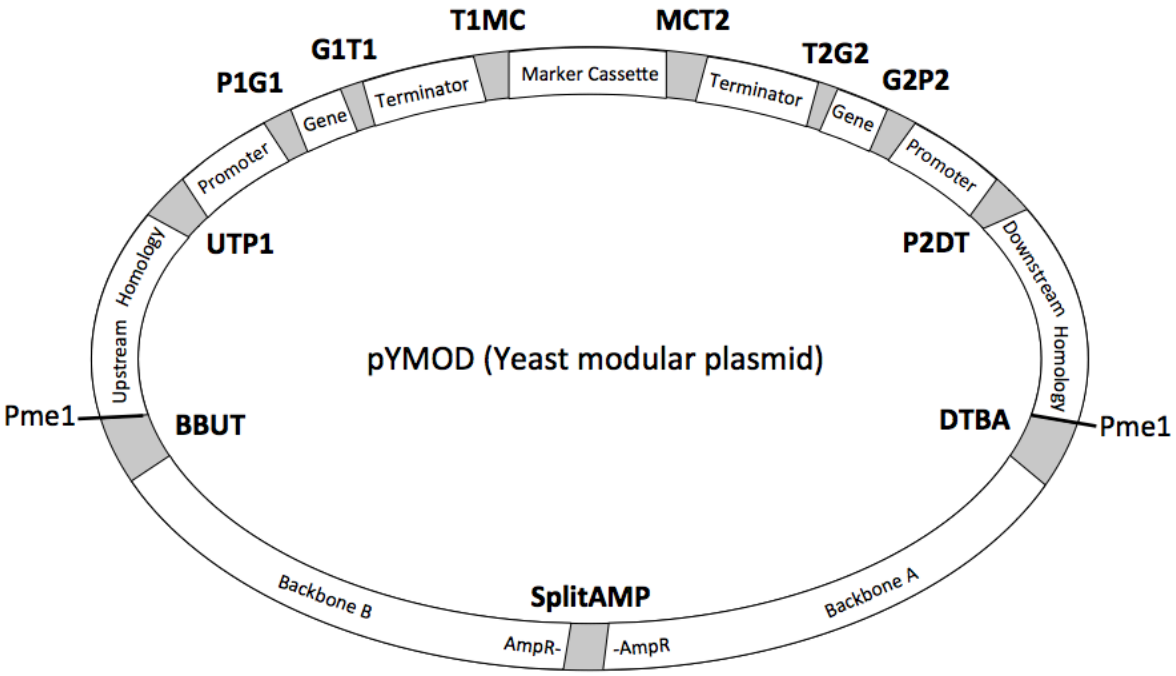


Figure 6.5: The pYMOD yeast modular plasmid architecture. The plasmid contains two sites for gene cassettes in a convergent orientation separated by a marker cassette. Standard linkers, optimized for Gibson assembly, flank each component.

The linker sequences, except for SplitAMP, were designed with the r2oDNA designer[82] (Table 6.4). Each linker is 25 bp, has a T_m of 72 °C, has no more than 10 bp of homology to the genome of *S. cerevisiae*, has a minimum intra-molecular folding energy of -4 kcal/mol, and has a minimum inter-molecular folding energy of -9 kcal/mol. These specifications make the linkers ideal for Gibson assembly[58].

Table 6.4: Linkers used for the pYMOD yeast modular plasmid cloning scheme.

Linker Name	Sequence
BBUT	GTCGGCGGGACCAGGGAGTTTAAAC
UTP1	GCCGATACGAAGGTTTTCTCCAGCG
P1G1	GGGACCGTCAACCCTGAACCACAAA
G1T1	TGAGCAGGCATCGAGTGAAGTCAAC
T1MC	GCTTCAATAAAGGAGCGAGCACCCG
MCT2	CAGAAGCGAGGCGAATAAAGGTGGC
T2G2	CGATACCTGGTTGTGGGCTCTCTCA
G2P2	TTTGTCTGACAACCGTTCGCAGAGC
P2DT	GTCCCTGAAAACCACTGAGTTGCCC
DTBA	CATGGTCATAGCTGTTTCCTGTGT
SplitAMP_F	GTTCGCCAGTTAATAGTTTGCGCAACG
SplitAMP_R	CTCGAGGGGGGCGGATCC

This modular plasmid construction scheme has been used to successfully construct the majority of plasmids used in this work. Even with Gibson-optimized linkers, an assembly with 11 components rarely produces enough plasmid for *E.coli* transformation. However, enough assembled plasmid is generated for amplification with PCR, which can be used to combine multiple components and repeat a Gibson assembly with fewer parts.

6.5 AQUARIUM

Much of the cloning work used to construct the plasmids and yeast strains described here was done through “Aquarium,” a lab automation system developed in the Klavins lab. With this system, all cloning protocols from ordering primers to transforming yeast are coded in Krill, a formalized programming language for laboratory protocols. This code is interpreted by the

Aquarium software, which displays step-by-step instructions on touch screen monitors to undergraduate and staff laboratory technicians. The software generates and automated laboratory notebook that contains all information about what was done in the lab in order to produce a new sample and tracks the inventory of every item.

Chapter 7
YEAST SECRETION CAPTURE

7.1 INTRODUCTION

The design of *de novo* proteins is complicated by a frequent tradeoff between function and stability. Engineered binding proteins, for example, are often designed with non-polar amino acids at the binding site to increase the energy gap between a bound and unbound state[83,84]. However, this design feature generates an unstable monomeric state with exposed hydrophobic residues. Given this tradeoff, it is unsurprising that function-based directed evolution assays that ignore stability often enrich highly unstable proteins that are not suitable for their desired application[85–87]. Previous work has shown that secretion of a biotinylated protein and capture on the cell surface can be used to dramatically reduce the display of unstable library variants[88,89]. However, this approach requires *in vivo* biotinylation, which makes the system incompatible with standard yeast surface display binding assays that use biotinylated target protein. Here, we introduce secretion capture (SecCap), a yeast surface display assay for simultaneous selection of function and stability that uses covalent attachment to the cell surface. Once a suitable design is selected with SecCap, the yeast strain displaying that construct can immediately be used for the production of soluble protein, which dramatically reduces the time lag between the initial identification of an optimized design and further *in vitro* characterization.

Yeast surface display (YSD) is frequently used to enrich high affinity binders from a design library by sorting for target-bound cells. Traditional surface display involves the fusion of a protein of interest to Aga2, a large and stable protein domain that coordinates trafficking and cell surface attachment. Proper display does require some degree of design stability, which has been suggested as a method for eliminating unstable variants[90]. However, fusion to Aga2, like fusion to maltose binding protein or other highly stable proteins, creates artificial stabilization. After multiple rounds of enrichment with YSD, affinity optimized binders are often unable to be expressed when separated from Aga2 due to insufficient stability. Frequent enrichment of unstable designs is a major roadblock for *de novo* protein optimization.

To solve this problem, we developed a secretion capture (SecCap) yeast surface display system in which monomer stability is required for the display of a design on the cell surface

(Figure 7.1A,B). Instead of an Aga2 fusion, a 13 residue unstructured peptide, SpyTag, is attached to the design[91]. In order to be displayed on the cell surface, the design must avoid degradation as it is translated and trafficked outside of the cell before being covalently captured by an exogenously added capture reagent. The capture reagent attaches to the cell surface with SnoopTag, an orthogonal covalent attachment system[92]. The addition of exogenous capture reagent instead of co-expression of the design and capture reagent serves three purposes. First, it necessitates that the design is fully translated and trafficked through the secretory pathway prior to covalent attachment to the capture reagent, which eliminates the possibility of artificial stabilization before the degradation of unstable designs. Second, it reduces the metabolic load associated with expressing and secreting multiple proteins. Third, a SecCap strain can be used to express soluble protein for *in vitro* characterization by leaving out the capture reagent (Figure 7.1C). Protein can be purified from the bulk media after secretion with FLAG purification[93] and the SpyTag peptide can be cleaved with TEV protease[94]. Here, we outline the development of the SecCap assay and demonstrate the ability to differentiate between stable and unstable proteins in a context that would enable the screening of design libraries.

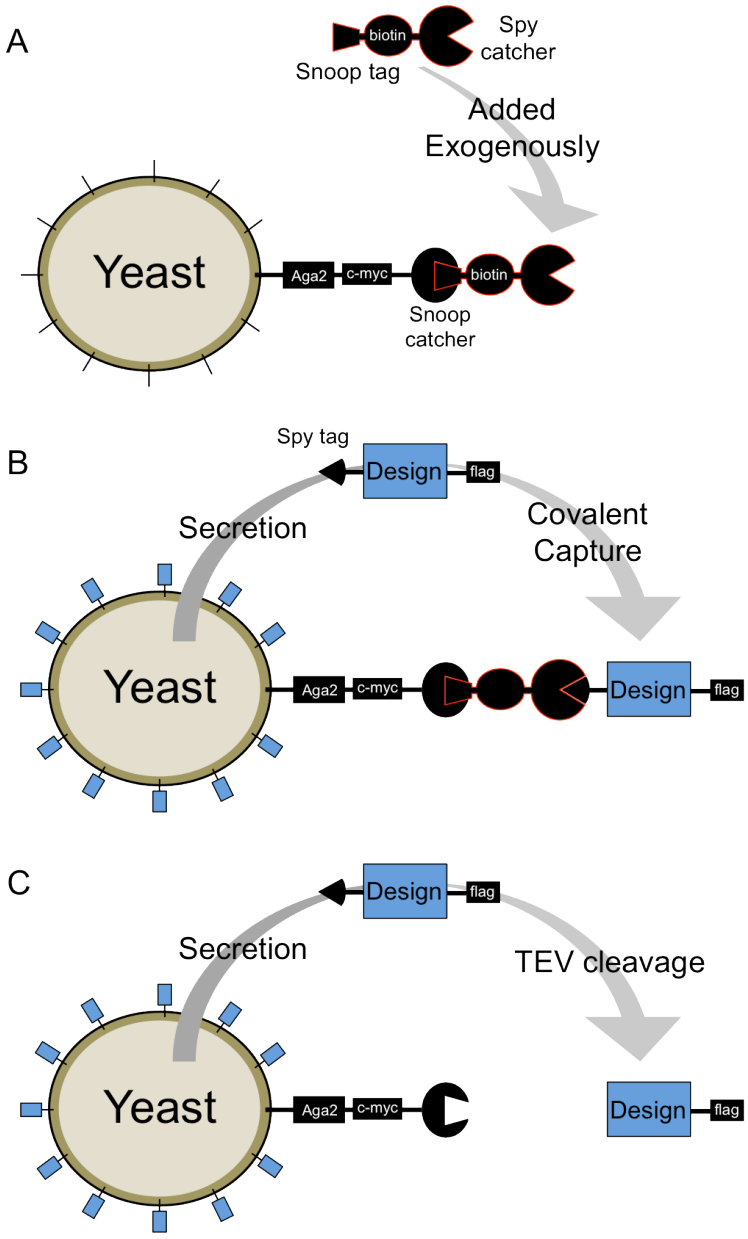


Figure 7.1: An overview of the secretion capture system. (A) Yeast cells constitutively surface expressing SnoopCatcher are incubated with "capture reagent," which covalently attaches to the cell surface with a SnoopCatcher/SnoopTag interaction. The capture reagent also contains an optional biotinylation site and a SpyCatcher domain. (B) A SpyTagged design is secreted by the cell and covalently captured on the cell surface with a SpyCatcher/SpyTag interaction. Misfolded proteins are degraded prior to secretion. (C) A selected strain can immediately be used to produce soluble protein for *in vitro* characterization with flag-tag purification and TEV cleavage to remove the SpyTag domain.

7.2 Strain and Reagent Construction and Validation

The secretion capture system requires that four distinct components properly assemble for the simultaneous selection of design stability and binding affinity. First, SnoopCatcher must be displayed on the cell surface as a fusion to Aga2. Second, the capture reagent must be covalently bound to SnoopCatcher. Third, the design protein must be secreted and covalently bound to the capture reagent. Fourth, a target protein must be screened for binding to the design. For assay development and library screening, each component contains a domain that can be fluorescently labeled for detection with flow cytometry (Figure 7.2). For the development of the SecCap assay, we individually verified and optimized the assembly of each component.

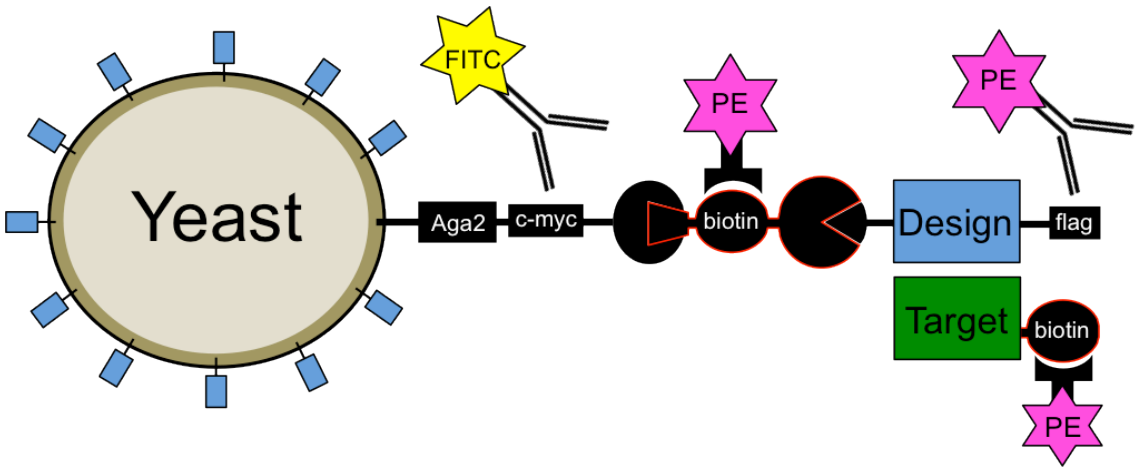


Figure 7.2: Optional labels for strain and assay verification. Each separate component of the secretion capture system contains a domain that enables antibody or streptavidin binding for the verification of proper attachment to the cell surface. The surface expression protein and secreted design contain a myc or flag tag, respectively, that allow for detection with an antibody conjugated to a fluorescent reporter. The capture reagent and target can optionally be biotinylated for detection with streptavidin phycoerythrin (SAPE).

We began with a variant of EBY100, an *S. cerevisiae* strain optimized for surface expression[12]. This strain typically requires galactose induction for display. However, for SecCap, we use galactose induction for the expression and secretion of the design, so we replaced the inducible promoter, pGAL1, with a strong constitutive promoter, pGPD[31]. Yeast surface display traditionally uses a centromeric plasmid, pETCON2, which contains a surface expression

cassette that is driven by pGAL1. In order to make surface display constitutive, we also replaced this promoter with pGPD. The use of a centromeric plasmid enables the construction of large libraries. However, the SecCap system contains only a single protein, SnoopCatcher, which is surface displayed as an Aga2 fusion for all cell strains. To eliminate the need to maintain plasmid selection, we modified the plasmid for chromosomal integration by adding chromosomal homology and restriction digest sites. Variants of the new secretion capture vector, pYMOD_SnoopCatcher, were constructed with different selectable markers and chromosomal homologies.

We found that SnoopCatcher was weakly displayed on the cell surface when expressed from a single cassette. However, by integrating a second copy of the SnoopCatcher expression cassette, better surface display was achieved (Figure 7.3). Surface display was measured by growing cells from a fresh plate in yeast peptone dextrose (YPD) media to saturation, incubating with FITC-anti-myc antibody, and measuring fluorescence with the FL1.A channel on an Accuri C6 cytometer. Many approaches for optimization were tested such as the display of SnoopCatcher repeat proteins, codon optimization, and transformation with a third expression cassette. Double and triple integrations produced a similar SnoopCatcher density, which was higher than with any other approach tested. The final yeast patent strain, ySecCap, contains two SnoopCatcher integrations at the URA and ARS314 loci.

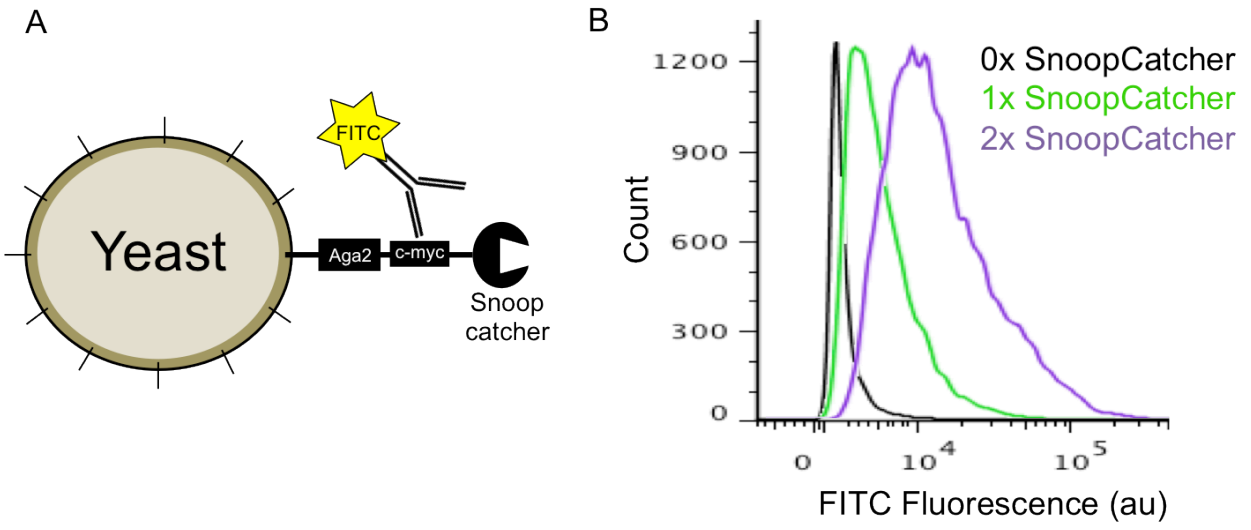


Figure 7.3: Surface display validation of SnoopCatcher. (A) FITC conjugated anti-myc antibody is used to detect the surface display of SnoopCatcher. (B) Surface expression with 0, 1, or 2 chromosomally integrated expression cassettes.

Capture reagent was found to efficiently adhere to the cell surface through a covalent SnoopCatcher-SnoopTag interaction. Capture reagent attachment was detected by incubating cells expressing SnoopCatcher with biotinylated capture reagent, labeling with streptavidin phycoerythrin (SAPE), and measuring fluorescence with the Y1.A channel on a Miltenyi MACSQuant flow cytometer (Figure 7.4). To achieve optimal secretion capture resolution, we aimed to maximize the density of SpyCatcher domains on the cell surface. We tested capture reagents with one, two, or three SpyCatcher repeats, and all showed comparable signal for capture. It had previously been shown that SpyCatcher is functional at either terminus or internally in a protein fusion[95]. Therefore, a reagent with three repeats was chosen to triple the total number of functional SpyCatcher domains on the cell surface.

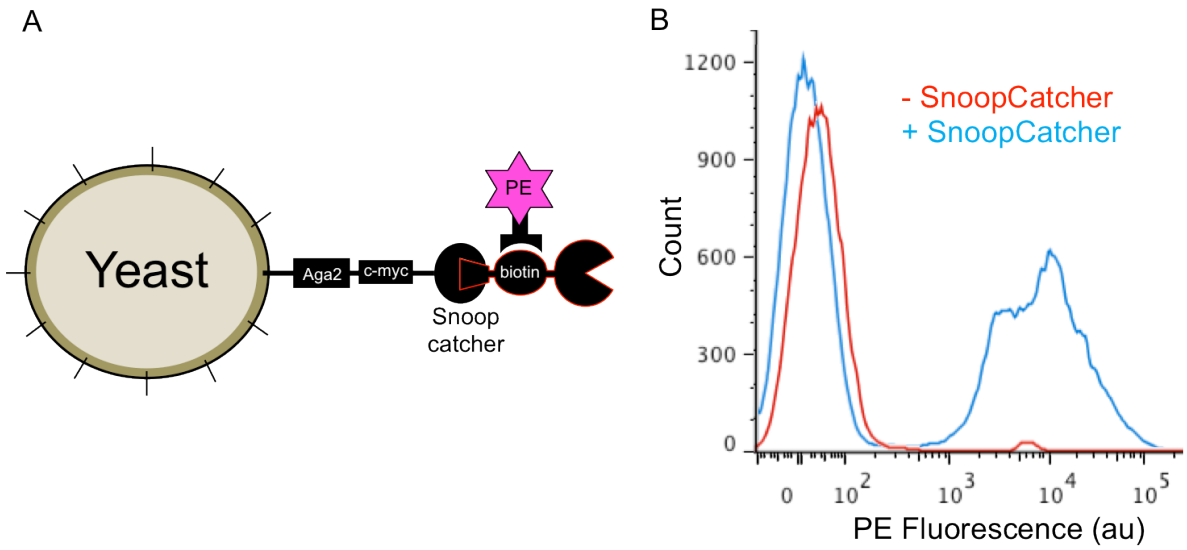


Figure 7.4: Validation of capture reagent attachment for SecCap. (A) SAPE is used to detect biotinylated capture reagent attached to the cell surface. (B) Capture signal with and without the expression of SnoopCatcher.

A centromeric plasmid, pSecCap, fuses a protein of interest to the yeast prepro-alpha-factor leader region for yeast secretion[96]. The open reading frame also contains SpyTag for surface attachment and flag tag for labeling and purification. NdeI and XhoI restriction sites flank the cloning site for the protein of interest. The fusion protein is expressed from the Gal1 promoter. A pSecCap plasmid containing a design is transformed into ySecCap and sustained

with TRP selection. The secretion capture assay consists of yeast preinduction in galactose media, labeling with capture reagent, and growth in galactose media with 30% PEG. Cells are then quenched with excess SpyTag to prevent any additional covalent attachments, washed, and resuspended in PBSF for labeling.

Secreted proteins fused to SpyTag covalently attach to the surface of cells labeled with capture reagent. Known stable proteins were added to the pSecCap plasmid and separately transformed into ySecCap. A secretion capture assay was conducted for each, where the preinduction and 30% PEG growth steps were performed in media containing either 2% dextrose or 2% galactose. Following secretion capture, cells were labeled with PE conjugated anti-FLAG antibody (Figure 7.5). All cells grown in 2% dextrose showed only background fluorescence, while all cells grown in 2% galactose contain a population of cells with high PE fluorescence. Bimodality is expected for secretion capture signal. As yeast cells divide, mother cells retain the cell wall and all bound proteins. Therefore, all cell divisions after labeling with the capture reagent produce a daughter cell that contains no capture reagent and will therefore show no secretion capture signal.

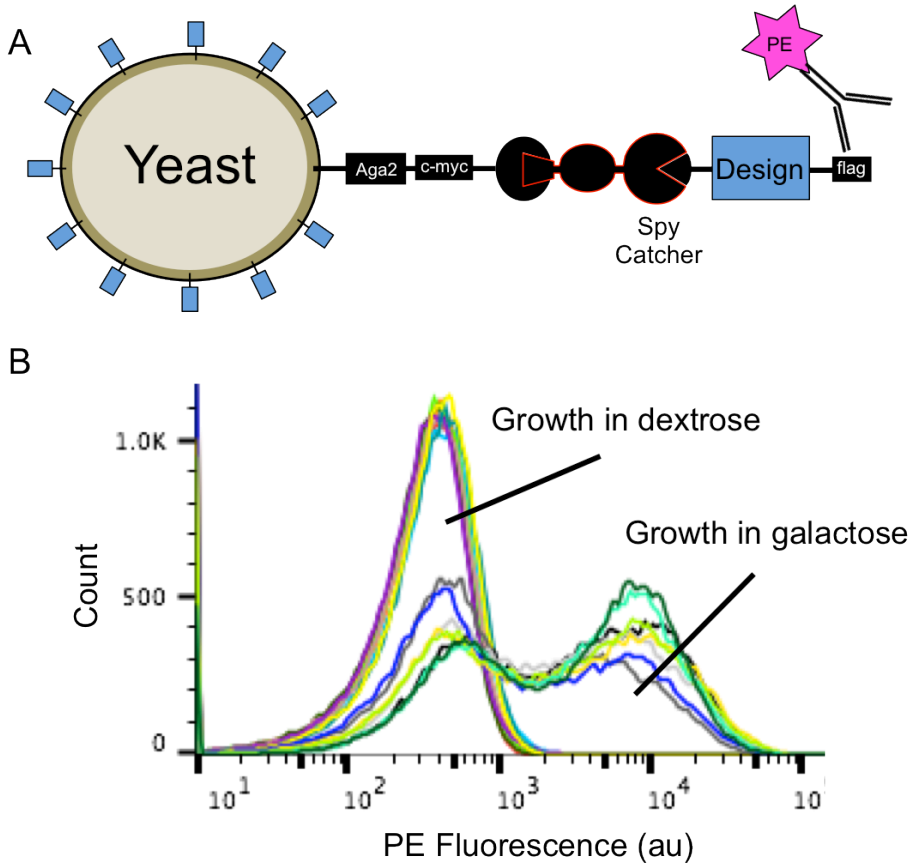


Figure 7.5: Secretion capture validation for many designs. (A) PE conjugated anti-flag antibody is used to detect the capture of SnoopTagged design proteins on the cell surface. (B) Secretion capture signal for many designs that were grown either in 2% dextrose or in 2% galactose media.

Yeast cells prepared with secretion capture can be used for functional binding assays (Figure 7.6). To validate binding, a well-characterized and highly stable binding protein, αBFL1, was attached to the yeast surface with secretion capture. αBFL1 had previously been shown to bind to Bfl-1, a human pro-survival BCL2 homologue, with a K_D of 1 nM ± 0.6 nM using biolayer interferometry[32]. Cells were then incubated with 1 nM, 10 nM, or 100 nM biotinylated Bcl-2 and labeled with SAPE. As expected given the interaction affinity, PE fluorescence saturates at approximately 10 nM Bfl-1.

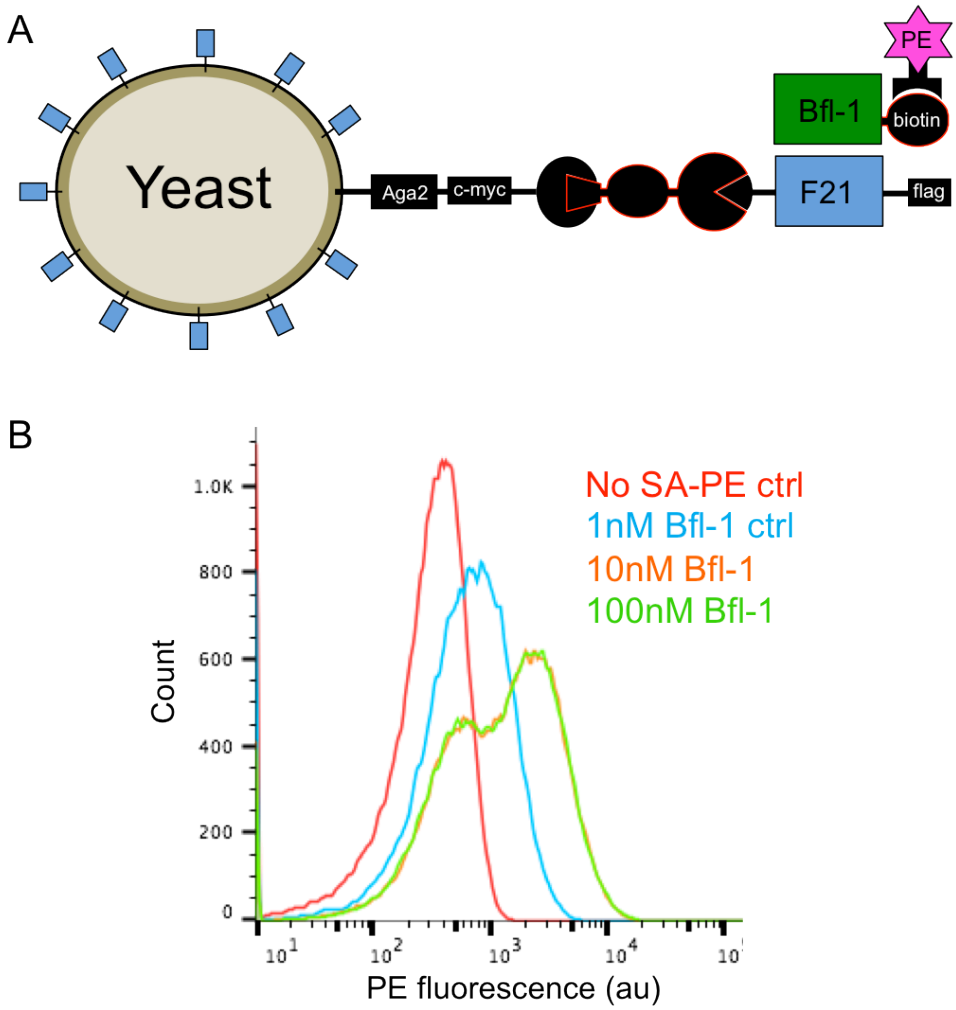


Figure 7.6: Binding validation with secretion capture. (A) SAPE is used to detect biotinylated target protein bound to the cell surface. (B) Bfl-1 binding signal at different concentrations of target protein for a strain secreting and capturing αBFL1.

7.3 System Validation

The secretion capture system correctly differentiated between proteins of known stability. Two proteins were chosen that exhibited high thermal stability with circular dichroism and were monomeric by SEC-MALS. Four proteins were chosen that displayed well in traditional yeast surface display, but failed all attempts at soluble expression. All six proteins were tested with secretion capture at 30°C and 22°C (Figure 7.7). SecCap signal was measured between 6 and 12 hours after labeling with capture reagent and being transferred to 30% PEG galactose media. For all time points, both stable proteins gave a higher secretion capture signal than any of the

unstable proteins. Optimal differentiation between stable and unstable proteins occurred between 6 and 10 hours for growth at 30°C and at 12 hours for growth at 22°C.

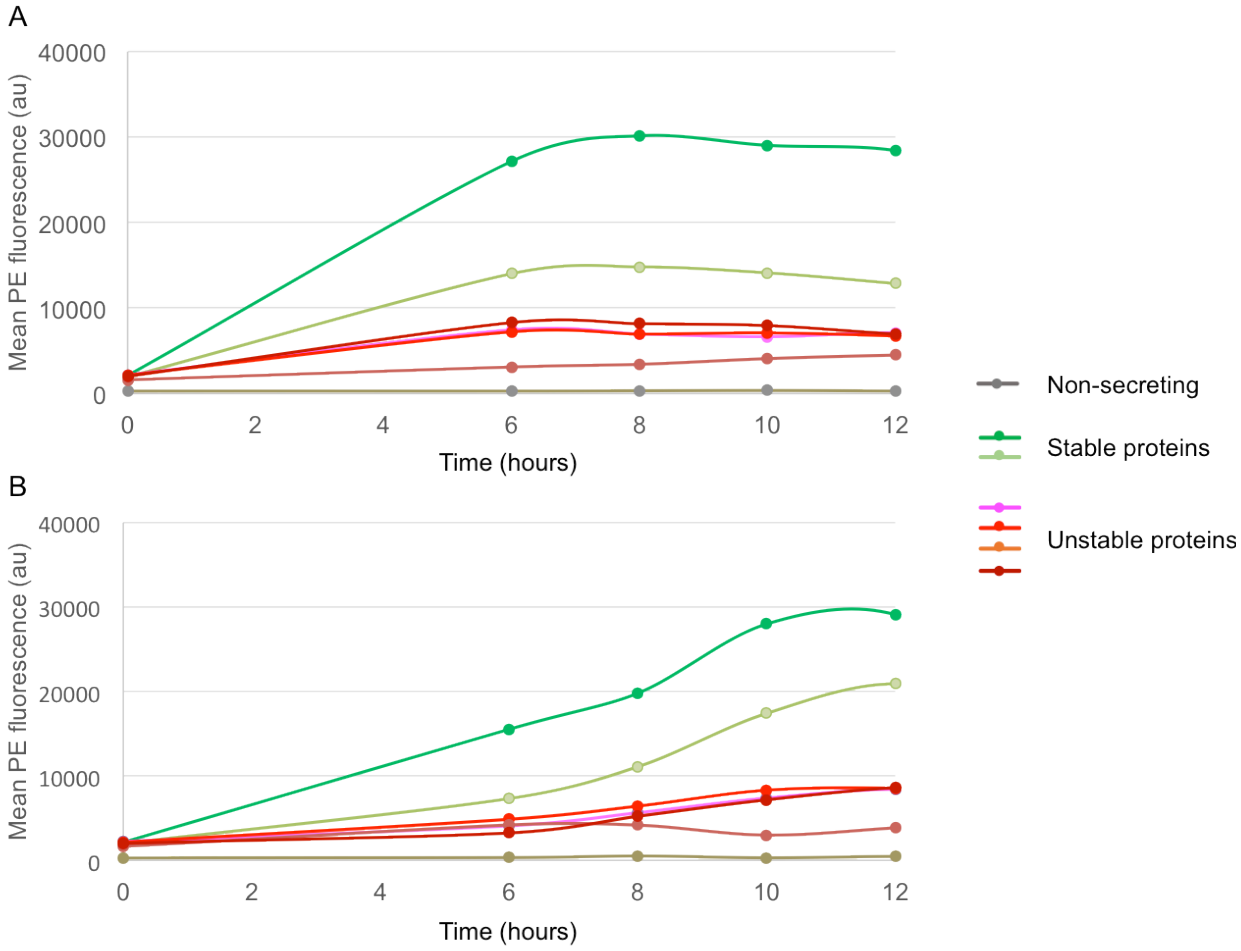


Figure 7.7: Secretion capture signal for control proteins. Cells secreting and capturing known stable or unstable proteins were grown at (A) 30°C or (B) 22°C and screened for secretion capture signal between 6 and 12 hours after induction.

When screening libraries with secretion capture, it is essential that cells exclusively capture their own secreted protein to maintain a link between protein sequence and secretion capture level (Figure 7.8). To prevent cross binding between cells, the secretion capture assay is conducted with a low cell concentration and without mixing in a high-viscosity media containing 30% PEG. These conditions ensure that cells capture only their own designs during incubation, but there is still a possibility of cross binding when cells are harvested and washed. To prevent

any binding from occurring during this step, excess SpyTag is added during harvesting to quench covalent attachment.

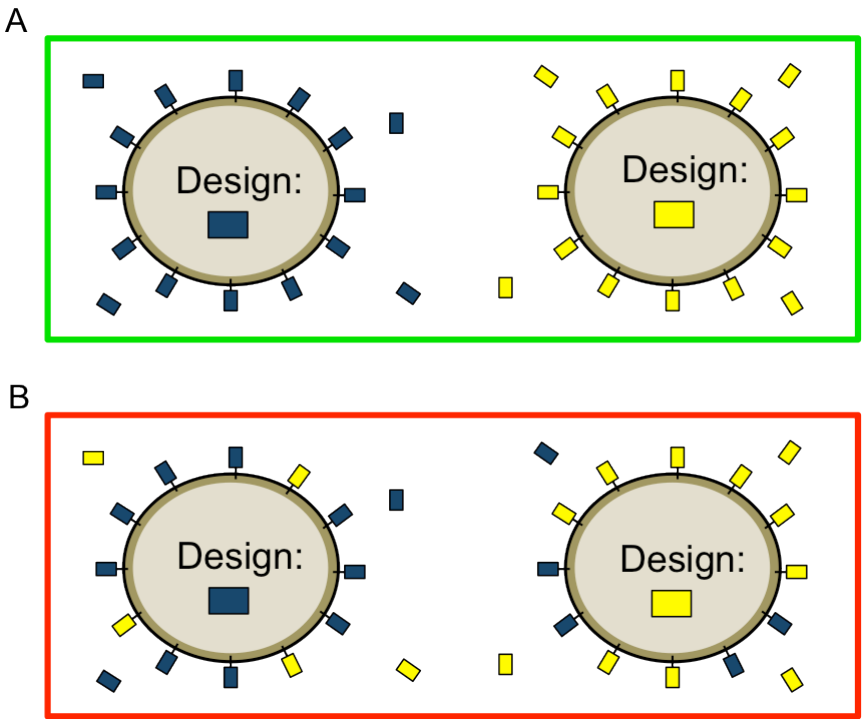


Figure 7.8: A cartoon describing the requirement for self-attachment. (A) A functioning assay in which all designs being secreted by a particular cell type are captured by the same cell type. (B) A non-functioning system where designs are diffusing and being captured by other cell types, resulting in a heterogeneous surface capture.

A library screening scenario was simulated by mixing a secreting strain and a non-secreting strain. The non-secreting yeast strain was transformed with a constitutive mTurquoise expression cassette, so that the two populations could be distinguished with flow cytometry. The secreting strain showed a large population shift for secretion capture signal, while the non-secreting strain showed no shift (Figure 7.9). This indicates that self-attachment is far more prevalent than cross binding.

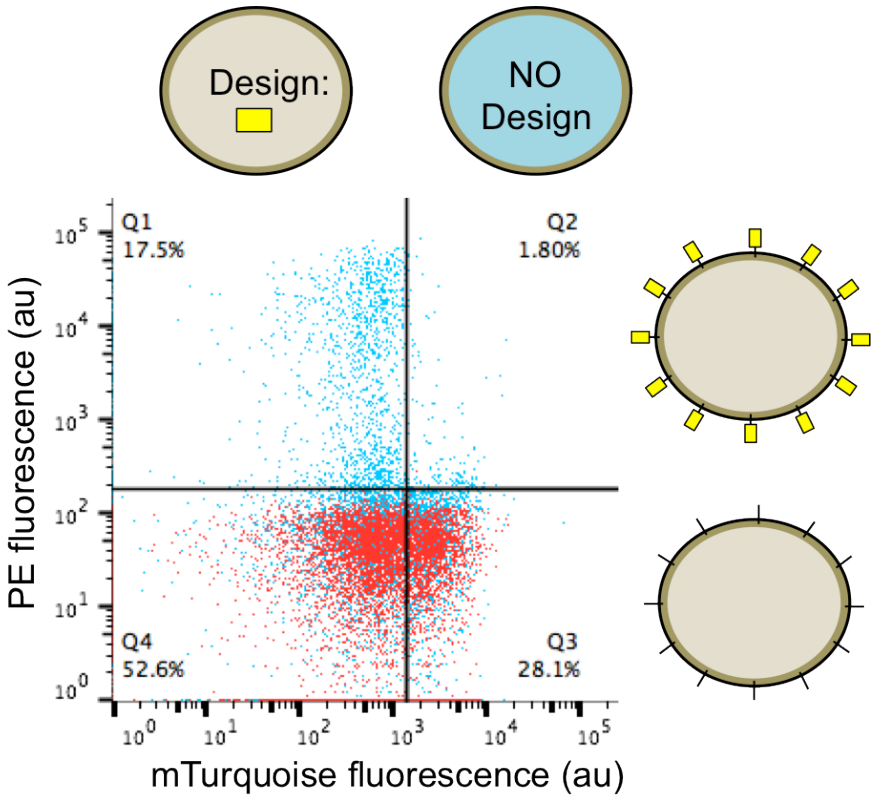


Figure 7.9: Validation of SecCap self-attachment. Cells secreting a design or not secreting a design are differentiated with cytosolic mTurquoise expression. Secretion capture signal is measured with PE conjugated anti-flag antibody (Blue). A non-labeled control is also included (red).

7.4 DISCUSSION

Secretion capture is a powerful platform for the simultaneous selection of function and stability. We expect that the requirement for a design to be expressed and trafficked without a stabilizing fusion will result in fewer enriched designs that are discarded due to low stability. Since secretion capture generates a yeast cell population with cell surface immobilized designs, this approach should be suitable for all functional selections that have been demonstrated with yeast surface display. Additionally, SecCap enables a rapid transition from selection to soluble protein production for further *in vitro* design characterization. As a next step, the SecCap system will be tested with large protein libraries. Initially, proteins with known stabilities will be screened in order to validate and characterize the performance of the system.

Chapter 8
CONCLUSION

Here we have described the development of two novel cell-based platforms that address major challenges for common computational protein design workflows. Yeast synthetic agglutination, or SynAg, is a system for rapidly, quantitatively, and controllably measuring protein-protein interaction strength that can easily be expanded for the characterization of interactions at a library-on-library scale. Yeast secretion capture, or SecCap, is a system for simultaneous selection of stability and function that can also be used to rapidly produce soluble protein for initial *in vitro* characterization once enriched designs have been isolated. In addition to applications for computational protein design, possible applications and future directions, such as multiplexed screening of drug toxicity and modeling reproductive ecology, were discussed in previous chapters.

At the core of synthetic biology is the idea that biological systems are fundamentally engineerable by programming cells at the DNA level. As for other engineering disciplines, the development of complex systems requires that specific behaviors can be abstracted as "parts" and that parts are sufficiently understood and characterized so that they can be systematically combined to achieve new predictable behaviors. Biological parts are often discussed in the limited context of transcriptional regulators for intracellular signal processing. However, these are a small fraction of the available parts. For the development SynAg and SecCap, many established synthetic biological parts were combined in order to develop novel systems. These parts included: natural and synthetic transcription factors, promoters, and terminators; cell surface display systems; fluorescent reporters; peptide secretion signals; recombinases, endonucleases, and proteases; and covalent protein-peptide attachment systems. Throughout the development of both platforms, most tested parts turned out to be functional, even in the context of a complex system. With the parts that are currently available, the synthetic biology community has the potential to make an enormous impact on science, medicine, and industry by utilizing the strengths of biological systems for the development of practical tools.

BIBLIOGRAPHY

1. Roy, A., Lu, C. F., Marykwas, D. L., Lipke, P. N. & Kurjan, J. The AGA1 product is involved in cell surface attachment of the Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin. *Mol. Cell. Biol.* **11,** 4196–206 (1991).
2. Dranginis, A. M., Rauceo, J. M., Coronado, J. E. & Lipke, P. N. A Biochemical Guide to Yeast Adhesins: Glycoproteins for Social and Antisocial Occasions. *Microbiol. Mol. Biol. Rev.* **71,** 282–294 (2007).
3. Lipke, P. N. & Kurjan, J. Sexual agglutination in budding yeasts: structure, function, and regulation of adhesion glycoproteins. *Microbiol Rev* **56,** 180–194 (1992).
4. Chen, E. H., Grote, E., Mohler, W. & Vignery, A. Cell-cell fusion. *FEBS Letters* **581,** 2181–2193 (2007).
5. Zhao, H., Shen, Z., Kahn, P. C. & Lipke, P. N. Interaction of a-Aglutinin and a-Agglutinin, Saccharomyces cerevisiae Sexual Cell Adhesion Molecules. *J. Bacteriol.* **183,** 2874–2880 (2001).
6. Herskowitz, I. Life cycle of the budding yeast Saccharomyces cerevisiae. *Microbiol. Rev.* **52,** 536 (1988).
7. Chenevert, J., Valtz, N. & Herskowitz, I. Identification of genes required for normal pheromone-induced cell polarization in Saccharomyces cerevisiae. *Genetics* **136,** 1287–1296 (1994).
8. Guo, B., Styles, C. a, Feng, Q. & Fink, G. R. A Saccharomyces gene family involved in invasive growth, cell-cell adhesion, and mating. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 12158–12163 (2000).
9. Xie, X., Qiu, W. G. & Lipke, P. N. Accelerated and adaptive evolution of yeast sexual adhesins. *Mol. Biol. Evol.* **28,** 3127–3137 (2011).
10. Stein, R. A., Smith, J. A. & Rose, M. D. An amphiphysin-like domain in Fus2p is required for Rvs161p interaction and cortical localization. *G3 Genes, Genomes, Genet.* **6,** 337–349 (2016).
11. Smith, G. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science (80-.).* **228,** 1315–1317 (1985).
12. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15,** 553–557 (1997).
13. Perfetto, S. P., Chattopadhyay, P. K. & Roederer, M. Innovation: Seventeen-colour flow cytometry: unravelling the immune system. *Nat. Rev. Immunol.* **4,** 648–655 (2004).
14. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. [Yeast two hybrid]. *Nature* **340,** 245–246 (1989).
15. Yu, H. *et al.* Next-generation sequencing to generate interactome datasets. *Nat. Methods* **8,** 478–480 (2011).
16. Hastie, A. R. & Pruitt, S. C. Yeast two-hybrid interaction partner screening through in vivo Cre-mediated Binary Interaction Tag generation. *Nucleic Acids Res.* **35,** (2007).
17. Yachie, N. *et al.* Pooled-matrix protein interaction screens using Barcode Fusion Genetics. *Mol. Syst. Biol.* **12,** 863–863 (2016).
18. Chen, Y.-C., Rajagopala, S. V., Stellberger, T. & Uetz, P. Exhaustive benchmarking of the yeast two-hybrid system. *Nat. Methods* **7,** 667–668 (2010).

19. Braun, P. *et al.* An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6,** 91–97 (2009).
20. Gu, L. *et al.* Multiplex single-molecule interaction profiling of DNA-barcoded proteins. *Nature* **515,** 554–557 (2014).
21. Concepcion, J. *et al.* Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization. *Comb. Chem. High Throughput Screen.* **12,** 791–800 (2009).
22. Daugherty, P. S., Chen, G., Olsen, M. J., Iverson, B. L. & Georgiou, G. Antibody affinity maturation using bacterial surface display. *Protein Eng.* **11,** 825–832 (1998).
23. Wen, F., Sun, J. & Zhao, H. Yeast surface display of trifunctional minicellulosomes for simultaneous Saccharification and fermentation of cellulose to ethanol. *Appl. Environ. Microbiol.* **76,** 1251–1260 (2010).
24. Chao, G., Cochran, J. R. & Wittrup, K. D. Fine epitope mapping of anti-epidermal growth factor receptor antibodies through random mutagenesis and yeast surface display. *J Mol Biol* **342,** 539–550 (2004).
25. Boder, E. T., Bill, J. R., Nields, A. W., Marrack, P. C. & Kappler, J. W. Yeast surface display of a noncovalent MHC class II heterodimer complexed with antigenic peptide. *Biotechnol. Bioeng.* **92,** 485–491 (2005).
26. Sauer, B. & Henderson, N. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc. Natl. Acad. Sci.* **85,** 5166–5170 (1988).
27. Albert, H., Dale, E. C., Lee, E. & Ow, D. W. Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J.* **7,** 649–659 (1995).
28. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4,** 975–986 (2015).
29. Boder, E. T. & Wittrup, K. D. Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol.* **328,** 430–444 (2000).
30. Julius, D., Blair, L., Brake, A., Sprague, G. & Thorner, J. Yeast α factor is processed from a larger precursor polypeptide: The essential role of a membrane-bound dipeptidyl aminopeptidase. *Cell* **32,** 839–852 (1983).
31. Hubmann, G., Thevelein, J. M. & Nevoigt, E. Natural and modified promoters for tailored metabolic engineering of the yeast Saccharomyces cerevisiae. *Methods Mol Biol* **1152,** 17–42 (2014).
32. Berger, S. *et al.* Computationally designed high specificity inhibitors delineate the roles of BCL2 family proteins in cancer. *Elife* **5,** (2016).
33. Farahani, E. *et al.* Cell adhesion molecules and their relation to (cancer) cell stemness. *Carcinogenesis* **35,** 747–759 (2014).
34. Kratzer, A. *et al.* Endothelial cell adhesion molecule CD146: implications for its role in the pathogenesis of COPD. *J. Pathol.* **230,** 388–98 (2013).
35. Toret, C. P., D'Ambrosio, M. V., Vale, R. D., Simon, M. A. & Nelson, W. J. A genome-wide screen identifies conserved protein hubs required for cadherin-mediated cell-cell adhesion. *J. Cell Biol.* **204,** 265–279 (2014).
36. Foster, T. J., Geoghegan, J. A., Ganesh, V. K. & Höök, M. Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus. *Nat.Rev.Microbiol.* **12,** 49–62 (2014).
37. Jensen, H. *et al.* Role of Lactobacillus reuteri cell and mucus-binding protein A (CmbA) in adhesion to intestinal epithelial cells and mucus in vitro. *Microbiology* **160,** 671–681

(2014).
38. Bhella, D. The role of cellular adhesion molecules in virus attachment and entry. *Phil. Trans. R. Soc. B* **370,** 20140035 (2015).
39. Xiong, X. *et al.* Receptor binding by an H7N9 influenza virus from humans. *Nature* **499,** 496–9 (2013).
40. Aagaard, J. E., Springer, S. A., Soelberg, S. D. & Swanson, W. J. Duplicate Abalone Egg Coat Proteins Bind Sperm Lysin Similarly, but Evolve Oppositely, Consistent with Molecular Mimicry at Fertilization. *PLoS Genet.* **9,** (2013).
41. Melcher, K. Structural biology: When sperm meets egg. *Nature* **534,** 484–485 (2016).
42. Bishop, B. B., Raymond, K., Rieger, S., Held, P. & Instruments, B. Fluorescent Proteins. *Biotek Appl. Guid.* 1–9 (2013). doi:10.1016/B978-0-12-803581-8.00987-5
43. Kremers, G., Gilbert, S. G., Cranfill, P. J., Davidson, M. W. & Piston, D. W. Fluorescent proteins at a glance. *J. Cell Sci.* **124,** 157–160 (2011).
44. Chudakov, D., Matz, M., Lukyanov, S. & Lukyanov, K. Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiol. Rev.* **90,** 1103–1163 (2010).
45. Chang, F. & Herskowitz, I. Identification of a gene necessary for cell cycle arrest by a negative growth factor of yeast: FAR1 is an inhibitor of a G1 cyclin, CLN2. *Cell* **63,** 999–1011 (1990).
46. Lee, E. F. *et al.* The functional differences between pro-survival and pro-apoptotic b cell lymphoma 2 (Bcl-2) proteins depend on structural differences in their Bcl-2 homology 3 (BH3) domains. *J. Biol. Chem.* **289,** 36001–36017 (2014).
47. Chen, L. *et al.* Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. *Mol. Cell* **17,** 393–403 (2005).
48. Steinitz, M. Quantitation of the blocking effect of tween 20 and bovine serum albumin in ELISA microwells. *Anal. Biochem.* **282,** 232–238 (2000).
49. Procko, E. *et al.* A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell* **157,** 1644–1656 (2014).
50. Day, C. L. *et al.* Solution structure of prosurvival Mcl-1 and characterization of its binding by proapoptotic BH3-only ligands. *J. Biol. Chem.* **280,** 4738–4744 (2005).
51. McIsaac, R. S. *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* **41,** (2013).
52. Gai, S. A. & Wittrup, K. D. Yeast surface display for protein engineering and characterization. *Current Opinion in Structural Biology* **17,** 467–473 (2007).
53. Wadle, A. *et al.* Serological identification of breast cancer-related antigens from a Saccharomyces cerevisiae surface display library. *Int. J. Cancer* **117,** 104–113 (2005).
54. Van Den Beucken, T. *et al.* Affinity maturation of Fab antibody fragments by fluorescent-activated cell sorting of yeast-displayed libraries. *FEBS Lett.* **546,** 288–294 (2003).
55. Williams, R. *et al.* Amplification of complex gene libraries by emulsion PCR. *Nat. Methods* **3,** 545–550 (2006).
56. Smith, A. J. *et al.* A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination. *Nat. Genet.* **9,** 376–385 (1995).
57. Louvion, J.-F., Havaux-Copf, B. & Picard, D. Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene* **131,** 129–134 (1993).
58. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).

59. Araki, K., Araki, M. & Yamamura, K. Site-directed integration of the cre gene mediated by Cre recombinase using a combination of mutant lox sites. *Nucleic Acids Res.* **30,** e103 (2002).
60. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* **74,** 5463–5467 (1977).
61. Oberdoerffer, P. Unidirectional Cre-mediated genetic inversion in mice using the mutant loxP pair lox66/lox71. *Nucleic Acids Res.* **31,** 140e–140 (2003).
62. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).
63. Van, M. & Huang, D. How the Bcl-2 family of proteins interact to regulate apoptosis. *Cell Res.* **16,** 203–13 (2006).
64. Wingler, L. M. & Cornish, V. W. Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. *Proc. Natl. Acad. Sci.* **108,** 15135–15140 (2011).
65. Clarke, L. & Carbon, J. A colony bank containing synthetic CoI EI hybrid plasmids representative of the entire E. coli genome. *Cell* **9,** 91–99 (1976).
66. Balagaddé, F. K. *et al.* A synthetic Escherichia coli predator–prey ecosystem. *Mol. Syst. Biol.* **4,** (2008).
67. Strauch, E.-M., Fleishman, S. J. & Baker, D. Computational design of a pH-sensitive IgG binding protein. *Proc. Natl. Acad. Sci.* **111,** 675–680 (2014).
68. Lee, E. J., Lee, N. K. & Kim, I. S. Bioengineered protein-based nanocage for drug delivery. *Advanced Drug Delivery Reviews* **106,** 157–171 (2016).
69. Ivanov, A. A., Khuri, F. R. & Fu, H. Targeting protein-protein interactions as an anticancer strategy. *Trends in Pharmacological Sciences* **34,** 393–400 (2013).
70. Yusuf-Makagiansar, H., Anderson, M. E., Yakovleva, T. V., Murray, J. S. & Siahaan, T. J. Inhibition of LFA-1/ICAM-1 and VLA-4/VCAM-1 as a therapeutic approach to inflammation and autoimmune diseases. *Medicinal Research Reviews* **22,** 146–167 (2002).
71. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **47,** 20–33 (2016).
72. Bowes, J. *et al.* Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. *Nat. Rev. Drug Discov.* **11,** 909–922 (2012).
73. Redfern, W. S. *et al.* Safety pharmacology--a progressive approach. *Fundam. Clin. Pharmacol.* **16,** 161–173 (2002).
74. Rual, J. F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437,** 1173–1178 (2005).
75. Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159,** 1212–1226 (2014).
76. Arkin, M. R. & Wells, J. a. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.* **3,** 301–317 (2004).
77. Hamon, V. & Morelli, X. Druggability of protein–protein interactions. *Underst. Exploit. protein–protein Interact. as drug targets* 18–31 (2013).
78. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc / SS carrier DNA / PEG method. *Nat. Protoc.* **2,** 31–35 (2008).
79. Spencer, J. F. T., Spencer, D. M. & Bruce, I. J. *Yeast genetics: a manual of methods*. (Springer Science & Business Media, 2012).
80. Boeke, J. D., La Croute, F. & Fink, G. R. A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance.

MGG Mol. Gen. Genet. **197,** 345–346 (1984).
81. Lööke, M., Kristjuahan, K. & Kristjuhan, A. Extraction of Genomic Dna From Yeasts for Pcr- Based Applications. *Biotechniques* **50,** 325–328 (2011).
82. Casini, A. *et al.* R2oDNA designer: Computational design of biologically neutral synthetic DNA sequences. *ACS Synth. Biol.* **3,** 525–528 (2014).
83. Karanicolas, J. *et al.* A De Novo Protein Binding Pair By Computational Design and Directed Evolution. *Mol. Cell* **42,** 250–260 (2011).
84. Hackel, B. J., Kapila, A. & Dane Wittrup, K. Picomolar Affinity Fibronectin Domains Engineered Utilizing Loop Length Diversity, Recursive Mutagenesis, and Loop Shuffling. *J. Mol. Biol.* **381,** 1238–1252 (2008).
85. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **4,** (2008).
86. Foit, L. *et al.* Optimizing Protein Stability In Vivo. *Mol. Cell* **36,** 861–871 (2009).
87. Xu, L. *et al.* Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* **9,** 933–942 (2002).
88. Rakestraw, J. A., Baskaran, A. R. & Wittrup, K. D. A Flow Cytometric Assay for Screening Improved Heterologous Protein Secretion in Yeast - Rakestraw - 2008 - Biotechnology Progress - Wiley Online Library. *Biotechnol. Prog.* **22,** 1200–8 (2006).
89. Rakestraw, J. A., Aird, D., Aha, P. M., Baynes, B. M. & Lipovšek, D. Secretion-and-capture cell-surface display for selection of target-binding proteins. *Protein Eng. Des. Sel.* **24,** 525–530 (2011).
90. Shusta, E. V, Kieke, M. C., Parke, E., Kranz, D. M. & Wittrup, K. D. Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *J. Mol. Biol.* **292,** 949–956 (1999).
91. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci.* **109,** E690–E697 (2012).
92. Veggiani, G. *et al.* Programmable polyproteams built using twin peptide superglues. *Proc. Natl. Acad. Sci.* **113,** 1202–1207 (2016).
93. Einhauer, A. & Jungbauer, A. The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. *Journal of Biochemical and Biophysical Methods* **49,** 455–465 (2001).
94. Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. & Dougherty, W. G. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal. Biochem.* **216,** 413–7 (1994).
95. Li, L., Fierer, J. O., Rapoport, T. A. & Howarth, M. Structural analysis and optimization of the covalent association between SpyCatcher and a peptide tag. *J. Mol. Biol.* **426,** 309–317 (2014).
96. Zsebo, K. M. *et al.* Protein secretion from Saccharomyces cerevisiae directed by the prepro-alpha-factor leader region. *J. Biol. Chem.* **261,** 5858–5865 (1986).

VITA

David Younger is from Seattle, Washington and graduated from Interlake High School in Bellevue. He earned a Bachelor of Science in Bioengineering and a Bachelor of Arts in Economics from Rice University in Houston, Texas. He returned to Seattle in 2012 to begin his graduate work in Bioengineering at the University of Washington.