IMPROVING THE YEAST THREE-HYBRID SYSTEM FOR HIGH-THROUGHPUT TARGET DISCOVERY

By

Kyle D. Bailey

Submitted to the graduate degree program in Medicinal Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science.

______________________________
Chairperson Dr. Blake R. Peterson

______________________________
Dr. Gary Grunewald

______________________________
Dr. Jeff Krise

Date Defended: April 14, 2011
The Thesis Committee for Kyle D. Bailey

certifies that this is the approved version of the following dissertation:

IMPROVING THE YEAST THREE-HYBRID SYSTEM FOR HIGH-THROUGHPUT TARGET DISCOVERY

________________________________
Chairperson Dr. Blake R. Peterson

Date approved: April 23, 2011
Abstract

High-throughput screening and rational design can be used to create bioactive compounds with high affinity for selected therapeutic targets. However, a significant challenge in preclinical drug development is the identification of off-target proteins that contribute to phenotypic effects. This limits the understanding of the molecular basis of such effects, thus subverting rational drug design and hindering the identification of new therapeutic targets. Most previous strategies for proteome-wide target identification (target ID) have involved incubating cell lysates with compound-conjugated affinity resins. Despite their simplicity, such approaches can subject the proteome to conditions that prevent the detection of small molecule-protein interactions. The yeast three-hybrid system is an attractive alternative that uses genetic tools to screen for protein-small molecule interactions in cellulo. This thesis describes efforts to improve the utility of the yeast three-hybrid system to screen for drug targets. The proposed improvements utilize 1) the native fluorescence of green fluorescent protein (GFP) to identify interactions by flow-cytometry and fluorescence-activated cell sorting (FACS) and 2) the extreme affinity of streptavidin to search the mammalian proteome with biotinylated probes. The first objective required improvements to the sensitivity and dynamic range of a reporter vector encoding a popular GFP spectral variant. A new reporter vector was constructed and shown to exhibit better fluorescent properties compared to an existing reporter in a yeast one-hybrid assay. This reporter was also used to detect ligand dependent dimerization of the estrogen receptor β and progesterone receptor proteins. The second goal involved efforts to create a reduced valency streptavidin to enhance sensitivity for detection of biotinylated molecules in yeast three-hybrid systems. Circular permutations of wild-type and a low affinity mutant of streptavidin were constructed and fused to create dimeric streptavidins with variable valency. These constructs were tested with yeast three-hybrid assays using a GFP variant reporter, and shown to have altered profiles in fluorescence-based assays.
Acknowledgements

The author would like to thank individuals that have supported the work described in this Thesis. Thanks to the Peterson lab members past and present, specifically Dr. Kalai Selvi Shanmugam, Mohammad Almishwat, and David Hymel.

Funding for this work was provided by Dr. Blake Peterson and the Specialized Chemistry Center.

I would like to thank my committee, Dr. Gary Grunewald and Dr. Jeff Krise, and my research advisor Dr. Blake Peterson. Additional thanks to Dr. Peterson for guidance, patience, and support over the past years.
# TABLE OF CONTENTS

ABSTRACT.................................................................................................................................................. iii

ACKNOWLEDGEMENTS................................................................................................................................. iv

TABLE OF CONTENTS....................................................................................................................................... v

LIST OF FIGURES........................................................................................................................................... vii

LIST OF TABLES............................................................................................................................................... x

CHAPTER 1  Strategies for Identification of Protein Targets of Small Molecules................................. 1

   Introduction.................................................................................................................................................. 1

   Affinity Chromatography of Cell Lysates ................................................................................................. 1

   Affinity Assays of Expressed cDNA Libraries ....................................................................................... 6

   Tag-free Target ID Strategies ................................................................................................................... 10

   Yeast Genetic Systems ............................................................................................................................ 11

   Conclusions............................................................................................................................................... 19

   References................................................................................................................................................. 19

CHAPTER 2  Improving the Dynamic Range of a Yeast Fluorescent Reporter .................................... 25

   Introduction............................................................................................................................................... 25

   Design and Evaluation of GFP reporter constructs ................................................................................. 29

   Conclusions............................................................................................................................................... 39

   Experimental............................................................................................................................................. 39
LIST OF FIGURES

CHAPTER 1

Figure 1.1  Schematic representation of affinity chromatography outcomes in which the drug is tethered to a solid support resin, usually directly or through a biotin tag (which binds to a streptavidin fused matrix). Proteins still attached to the drug after washing are eluted and investigated by SDS-PAGE (A-D) or mass spectroscopy (E). In a typical affinity assay (A), many non-specific interacting proteins or partners are eluted with the target. To identify these false positives, a separate chromatography assay is carried out with a biologically inactive analogue of the drug (B), excess untagged drug (C), or a second batch of the affinity probe with washed extract (D). Alternatively, cell lysates of the original and control assay are composed of heavy (H) and light (L) isotopically labeled proteins (D), respectively. The elutants from these assays are pooled, digested with trypsin, and then compared using mass spectroscopy. .................................................. 3

Figure 1.2  Schematic representation of a phage display assay of a cloned protein library. Proteins are fused to a phage coat protein and assayed with affinity chromatography. Eluted proteins can be amplified by infection of bacteria. Proteins can be identified with SDS-PAGE/MS or through sequencing of host phage DNA (colored DNA fragments). .................................... 7

Figure 1.3  Schematic representation of an mRNA display assay, in which cDNA/mRNA-protein constructs are screened with affinity chromatography. Eluted constructs may be identified by sequencing of cDNA or amplified by PCR and transcription, creating an enriched mRNA pool for repeating the procedure. Proteins can be identified directly with SDS-PAGE/MS. .......................................................... 8

Figure 1.4  Schematic representing a protein microarray outcome in which screened protein library is spotted individually on a derivitized surface. After binding, unbound drug is washed from the array. The drug is labeled to be able to give a localized signal for identifying its binding partner(s). .......................................................... 9

Figure 1.5  Schematic depicting a biochemical suppression strategy that uses an in vitro activity assay. For simplicity, only unbound target produces activity. Aliquots of the drug inhibited sample (upper right) are ‘treated’ with fractions of the untreated sample (upper left). Only fractions with the drug target should restore assay activity (center). Iterative fractionation (not shown) and assays are used to enrich the target for identification. ..... 11

Figure 1.6  Schematic representation of a drug-induced haploinsufficiency competition assay of an essential target gene. In this example, diploid yeast with two copies of the target gene (green) produces twice as much target protein as the strain possessing only one gene (orange). Untreated yeast strains grow at equal rates. Drug treatment fully inhibits the product of one gene copy, stifling growth in the mutant single copy strain. Inhibition of the wild-type strain is reduced due to increased target protein content, and therefore growth is relatively unaffected. .................................................................................. 13
Figure 1.7 Schematic depicting a small synthetic lethality profile of 6 possible target proteins (1-6) and drug treatment (right), based on three random nonessential genes (A, B, C). Each square in the main grid catalogues the viability of the DKO strain of intersecting genes (e.g., A and 1), and on the right viability of drug treated KO strain (A, B, C). An X indicates the strain did not grow in rich media. Target gene deletion (2) and drug treatment both induce lethality in A and C KO stains (as does gene 1)..........................

Figure 1.8 Schematic depiction of yeast one- (A), two- (B), and three- (C) hybrid systems. Reporter expression is activated by localizing the activation domain (AD) to the promoter of the reporter gene (green arrow). The DNA binding domain (DBD) localizes the system to the promoter, which is directly fused to the AD in the one-hybrid system. The two-hybrid system depends on bait and prey dimerization for reporter activation, which is mediated by a small-molecule (CID) in the three-hybrid system. D) Schematic of yeast colonies in three-hybrid screen in which each colony possesses identical CID and bait protein. Each colony expresses a different prey, which must bind the drug (tethered to the anchor moiety, AM) to activate reporter expression. ............................................................... 16

CHAPTER 2

Figure 2.1 Maturation of the EGFP fluorophore (A) takes place inside the β-barrel structure and requires no cofactors or post-translational modifications. Live cells can be sorted based on the fluorescence intensity of expressed EGFP with FACS (depicted in B).......................... 27

Figure 2.2 Overlay of flow-cytometric histograms demonstrating fluorescence from pDCLryEGFP reporter expression. Each histogram represents 15,000 events, with a gate set to quantify the number of highly fluorescent cells from each experiment. Basal fluorescence was quantified with yeast transformants harboring empty plasmids (white). Background fluorescence produced with only the yEGFP reporter (light gray) and maximal fluorescence with reporter activated by a LexA-B42 fusion (dark gray). Taken with permission from (Clark, D. D.; Peterson, B. R. Chembiochem 2005, 6, 1442)..........28

Figure 2.3 Overlays of flow-cytometric histograms to compare fluorescence generated by reporters pDCLryEGFP (A) and pSHyEGFP (B). Each histogram represents 10,000 events with a gate for analysis of highly fluorescent cells. The gate begins at the 90th percentile of induced yeast lacking LexA-B42 (red histogram). C) Table of relevant statistics taken from histograms. The median signal and coefficient of variation (CV) is taken from the gated population of induced cells bearing the complete system (black histogram). The signal to noise ratio (S/N) is calculated as the quotient of the signal median and the background threshold (90th percentile of the omission control)................................................. 30

Figure 2.4 Overlays of flow-cytometric histograms comparing pSHyEGFP one-hybrid assays with different plasmid transformation sequences (A,B). Cells were assayed approximately 8 days after the first transformation and 4 days after the second. Each histogram represents 10,000 events with gates set as described in Figure 2.3. Red and black histograms represent uninduced and induced reporter expression, respectively. C) Table of relevant statistics (described in Figure 2.3).................................................................. 31

Figure 2.5 (A) Overlay of flow-cytometric histograms from reporter pSHyEGFPpst. Background fluorescence from induced yeast lacking LexA-B42 (red) was used to determine the background threshold. (B) Table of relevant statistics. ................................................................. 32
Figure 2.6 Overlay of flow-cytometric histograms demonstrating E2 dependent dimerization of ERβ(255-509) monomers. Each histogram represents 10,000 events with gate set to quantify highly fluorescent cells (>10^5 rfu) used in analysis. A) Bait and prey omission controls, treated with 500 nM E2, show that each component is necessary for high levels of fluorescence. B) Complete three hybrid-system incubated with three representative E2 concentrations. C) Dose-response curve for three-hybrid assay. Estradiol treatment concentration is plotted against normalized total count (100% ≈ 2,000 or 3,000) of cells emitting fluorescence >100 rfu. ........................... 34

Figure 2.7 Overlays of flow-cytometric histograms from PR (P4 CID) and AR (DHT CID) yeast ligand dependent two assays. Each histogram represents 10,000 events with gate of highly fluorescent cells (>100 rfu). A,C) Omission controls with maximum CID treatments and (B,D) representative histograms of complete system with added agonist shown. ................................................................. 35

Figure 2.8 Overlay of dose-response curves of ligand dependent PR yeast two-hybrid assays treated with progesterone. The yEGFP curve is based on counts of cells emitting fluorescence >100 rfu. The LacZ curve is based on calculated β-gal activity using a CPRG substrate assay (10 minutes after substrate addition). Curves are normalized based on maximum assay readouts (app. 3000 cells/β-gal activity = 100%). ................................................................. 36

Figure 2.9 Overlays of flow-cytometric histograms of integrated reporter strains FY250yEGFPpst1 (A) and 2 (B), and plasmid reporter pDCUdyEGFP (C). Background fluorescence from induced yeast lacking LexA-B42 (red) was used to determine the background threshold, as described previously. (D) Table of relevant statistics. ............................................................................ 38

Figure 2.10 Plasmid map of pSHyEGFPpst. The 2-micron ori and URA3 selection marker, shown in gold, maintain plasmid copies in yeast. The pBR322 ori and AmpR selection marker, shown in red, maintain plasmid copies for propagation in E. coli. Plasmid pSHyEGFP (7091 bps) lacks the Cln2-PEST (green, lined box) region of the reporter gene, but otherwise is identical. Plasmid pSHIPyEGFPpst (5386 bps) lacks the 2-micron ori, but is otherwise identical to pSHyEGFPpst................................................................. 41

Figure 2.11 Plasmid map of pDCLryEGFP. In pDCUdyEGFP (6994 bps) the LEU2 selection marker (gold arrow) is replaced with a URA3-d selection marker, but otherwise is identical to pDCLryEGFP. ............................................................................ 42

CHAPTER 3

Figure 3.1 Selected views of X-ray crystal structure of tetrameric core streptavidin complexed with biotin (PDB: 1SWE) (Gonzalez, M. et al. J. Biol. Chem. 1997, 272, 11288-11294.). A) Within a functional dimer (solid red and blue) two W120 residues (top, shown in sticks) enhance biotin affinity. C) Structural dimer (blue and green) interfaces are more extensive. Selected residues that have been mutated to disrupt oligimerization are shown in sticks. The biotin binding pocket involves a hydrogen bonding network (B) and a hydrophobic pocket (D). ................................................................. 52
Figure 3.2 View of streptavidin X-ray crystal structure (PDB: 1SWE) with linkers in SCD streptavidin represented by black lines. The GGGS linker (bottom) between the native N- and C- termini (shown in green stick) allows direct fusion between two monomers (top) between S69 and T115 (shown in brown stick). New N- and C- termini (E116 and G68, highlighted in yellow) are created.................................................................55

Figure 3.3 Topology map of core streptavidin monomer (A) and SCD (B), with identical coloring of β-strands. Residues involved in biotin binding are identified (H-bonding: grey boxes, Hydrophobic pocket: red boxes), and the three residues mutated to create SAnb (blue arrows). Chainsaws indicate where new termini are created in the circular permutations.57

Figure 3.4 A) Overlays of flow-cytometric histograms of three-hybrid assays with SA variants (bait) and wtSA (prey). Untreated assays contain 10 nM biotin present in media. B) Table of total counts of cells generating EGFP fluorescence greater than 100 rfu. ......................58

Figure 3.5 A) Overlays of flow-cytometric histograms of three-hybrid assays with dimeric SA variants (bait) and variant dSA_{AA} (prey). Untreated cells (red) contain 10 nM biotin, treated cells (blue) are spiked with 1 µM biotin. B) Table of total counts of cells that generated GFP fluorescence greater than 100 rfu.................................................................59

Figure 3.6 Plasmid map of pPA1 LexA-SA. All streptavidin LexA- fusions were cloned into pPA1 LexA, replacing SAwt. .................................................................61

Figure 3.7 Plasmid map of pJG4-5 SA, used to create new B42- streptavidin fusions (SAnb, dSA_{DD}).63

LIST OF TABLES

CHAPTER 1

Table 1.1 Selected example of affinity chromatography assays for proof-of-principle or target discovery.................................................................5

Table 1.2 Selected examples of miscellaneous target discovery strategies. .................................17
CHAPTER 1. Strategies for Identification of Protein Targets of Small Molecules

Introduction

Modern drug discovery programs are largely target oriented. Biomolecules (usually proteins) shown to be influential, if not essential, in disease progression are carefully chosen and attacked with a choice of iterative strategies aimed at safely modifying function (usually inhibition) and stifle the resulting ailment or resulting symptoms. Although direct, these strategies often underestimate the promiscuous nature of many small molecules. Consequently, toxic off-target effects may be missed until much later in the drug development process. These efforts are totally dependent on the “druggability” and disease relevance of the chosen target, and likewise are almost totally wasted if target validation is errant or incomplete. These factors perhaps explain why the optimistic adoption of target-based drug discovery, largely replacing physiology-based methods in the 90s, has coincided with a decline in productivity in the pharmaceutical industry. This assessment has led many to readopt more systematic approaches that probe biological outcomes in early drug discovery with phenotypic screens of cellular or whole organismic model systems. Yet even with a return to holism in early drug discovery, there is still much to be gained from a mechanistic understanding of phenotypic outcomes, particularly target identification. A well understood target is still the basis for rational drug design, which can greatly enhance optimization of phenotypically chosen compounds with structure activity relationship (SAR) studies and structure-based drug design. Target ID efforts could also lead to the discovery of new target-based drug discovery strategies.

Affinity Chromatography of Cell Lysates

The classical method used to discover targets of bioactive small molecules is to attach an immobilizing functional group so that the molecule can “catch” binding proteins from a mixed suspension of possible targets, usually a cell lysate or protein extract. The immobilization tag allows non-interacting proteins to be ‘washed’ away using an aqueous buffer and discarded, thus enriching binding
partners that are subsequently purified using protein-denaturing conditions or excess unmodified drug. In drug affinity chromatography, the drug is effectively transformed into a target purification tool. Separation of interacting proteins is achieved by SDS-PAGE, as depicted in Figure 1.1A. Identification of gel-extracted protein bands utilizes mass-spectroscopy based protein sequencing and/or immunoprecipitation.

As the oldest and still most prolific method for target ID, affinity chromatography has undergone much iteration. The simplest version directly links the compound of interest to a solid support matrix (e.g. agarose beads) to perform affinity chromatography. Many vendors (e.g. Biorad, Pierce, Sigma) sell these supports, or resins, chemically activated (e.g. with NHS esters) to bond covalently with a particular functional group (e.g. amine), which can be added to the drug by the user if necessary. The drug is usually tethered to the resin with a long, hydrophilic linker to minimize influence the resin might have on binding properties of the drug. Even so, many investigators have instead chosen to tag compounds with biotin (aka biotinylation) and incubate the derivative with cell lysate. In these experiments the protein streptavidin is immobilized to the support matrix, and can be added to the cell lysate before or after probe binding. The biotin tag binds streptavidin with virtually irreversible affinity (K_d ≈ 10^{-14} M), providing a noncovalent platform for target enrichment. If the tagged small molecule is suitably cell permeable, the biotinylated compound can be incubated with live cells and then lysed for chromatography. This allows for ligand-protein interactions to occur in their native cellular environment. Biotinylation also allows the modified probe to be tested to ensure the tag does not disrupt biological activity and target affinity.
Figure 1.1. Schematic representation of affinity chromatography outcomes in which the drug is tethered to a solid support resin, usually directly or through a biotin tag (which binds to a streptavidin fused matrix). Proteins still attached to the drug after washing are eluted and investigated by SDS-PAGE (A-D) or mass spectroscopy (E). In a typical affinity assay (A), many non-specific interacting proteins or partners are eluted with the target. To identify these false positives, a separate chromatography assay is carried out with a biologically inactive analogue of the drug (B), excess untagged drug (C), or a second batch of the affinity probe with washed extract (D). Alternatively, cell lysates of the original and control assay are composed of heavy (H) and light (L) isotopically labeled proteins (D), respectively. The elutants from these assays are pooled, digested with trypsin, and then compared using mass spectroscopy.

In the typical chromatography experiment, many proteins bind to the matrix through nonspecific interactions that are maintained throughout the washing step. Eluting with excess untagged compound may help reduce nonspecific binding proteins from eluting with the target(s), especially those that interact with the support matrix itself. In addition, several control experiments have been reported to distinguish genuine targets from nonspecific binding proteins. These controls use subtle variations of the standard
protocol, in which they run in parallel. The SDS-PAGE gels of the eluted proteins from the control and original assay are compared to find protein bands that warrant identification. A so called comparison control uses a probe that is a biologically inactive analogue of the drug in question. The chromatography protocol is otherwise identical, and therefore the same set of proteins should be eluted (Figure 1.1B), except target protein(s) whose affinity is disrupted by the modification. In a competition control, depicted in Figure 1.1C, excess untagged drug is incubated with the cell lysate along with the probe. These conditions favor target binding to the free drug over the affinity probe, and therefore impede binding to the resin. Target proteins bound to the untagged drug are washed with other free proteins. The absence or lower intensity of resulting PAGE bands, compared to the outcome of original assay, indicates the interaction is specific. Another variation, the serial control, incubates the washed cell lysate of the original assay with a second batch of affinity matrix (Figure 1.1D). Specifically bound proteins are expected to yield more prominent bands in the PAGE gel of the first batch, whereas nonspecific binding proteins should produce roughly equivalent intensity bands.

A recently reported affinity chromatography method uses a competition control with stable isotope protein labeling with amino acids in culture (SILAC) to streamline analysis of eluted protein. Two separate cultures are grown, one under normal conditions and the other in media containing heavy isotopically labeled arginines that are incorporated into cellular proteins. Each cell culture is lysed and incubated with drug-conjugated matrix, with the competition control applied to the natural isotope lysate (Figure 1.1E, this is expected to reduce target protein binding to the matrix). After incubation, the lysates are pooled, washed, eluted, and proteolytically digested for MS analysis of peptide fragments. Homologous fragments from the original and control assays can be readily differentiated and quantified. Target proteins are expected to yield high heavy-to-light fragment ratios (H/L), due to the competitive displacement of these proteins in the natural isotope lysate. It seems feasible that SILAC could also be applied to a comparison control (using an inactive analogue as described previously).
Table 1.1. Selected example of affinity chromatography assays for proof-of-principle or target discovery.

<table>
<thead>
<tr>
<th>Drug (Target, ref.)</th>
<th>Affinity Probe</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indomethacin (Glyoxalase I, Sato, S. et al. J. Am. Chem. Soc. 2007. 129, 873-880.)</td>
<td><img src="image_url" alt="Indomethacin Affinity Probe" /></td>
<td>Competition</td>
</tr>
</tbody>
</table>

A significant criticism of affinity chromatography is that it places a high value on equilibrium binding kinetics, which is not always indicative of biological activity. These experiments make the implicit assumptions that drugs bind their targets with significantly higher affinity than the remaining proteome, and that targets have ample cellular concentrations for detection. However, many drug targets have low cellular concentrations and bind their drug ligand with moderate affinity. These properties put
affinity chromatography at a disadvantage for discovering such targets. Low protein concentration may be compensated for by maximizing the total lysate used in a given assay. To address affinity, photoreactive groups have been attached to drug affinity probes to facilitate cross-linking to target proteins.\textsuperscript{11, 12}

**Affinity Assays of Expressed cDNA Libraries**

A more assertive strategy to overcome poor drug-target kinetics is to clone the genes of proteins to be screened into recombinant systems where expression can be tightly controlled. Expression-cloning target discovery methods offer the additional benefit of maintaining a traceable link between gene and gene product, allowing for selective amplification of potential target genes and identification with DNA sequencing. Also, with overexpression it is generally faster and cheaper to prepare large amounts of protein necessary for binding assays. These advantages come at the cost of expressing screened proteins in a foreign environment and/or at exogenous concentrations, resulting in the loss of native protein characteristics that may be important for probe affinity. For example, post-translational modifications of proteins in the mammalian cell are often essential for molecular recognition, but these modifications may not be mimicked with heterologous or in vitro expression. These strategies usually fuse participating proteins to a common protein or protein domain necessary for the assay, which can also alter binding.

Phage display is a platform that expresses cloned mammalian proteins on the head of a virus to present them to the probe.\textsuperscript{13,14} Mammalian genes are cloned into separate phage genomes, each fused to the gene of a phage coat protein. The expressed protein localizes to the exterior of the viral coat, where it can interact with an immobilized affinity probe. After binding and washing, eluted proteins remain fused to host bacteriophages that encode the supposed target’s gene. Bacteria are infected with these phages for amplification of genes, which may be expressed for iterative chromatography runs (Figure 1.2). Targets may be identified directly from protein (as described earlier) or by sequencing the gene embedded in the phage genome.
**Figure 1.2.** Schematic representation of a phage display assay of a cloned protein library. Proteins are fused to a phage coat protein and assayed with affinity chromatography. Eluted proteins can be amplified by infection of bacteria. Proteins can be identified with SDS-PAGE/MS or through sequencing of host phage DNA (colored DNA fragments).

In mRNA display, an in vitro alternative to phage display, screened proteins are fused almost directly to the genes that encode them. A cDNA library is first transcribed in vitro to produce a mixed mRNA population, followed by ligation of a puromycin-DNA cassette to the 3’ end of each transcript.\textsuperscript{15,16} The mRNA is fully translated in vitro, but the cassette prevents translational termination and triggers the formation of a C-terminal amide bond with the puromycin amine. The mRNA-protein fusion is purified, reverse transcribed, and incubated with the affinity probe. Following washing, eluted cDNA serves as a template for polymerase chain reaction (PCR) amplification (Figure 1.3). The amplified double-stranded cDNA can be transcribed to mRNA for additional rounds of screening. Like phage display, enriched proteins can be identified either directly or via their genetic precursors. To my knowledge, this method has only been reported in proof-of-concept studies.
Figure 1.3. Schematic representation of an mRNA display assay, in which cDNA/mRNA-protein constructs are screened with affinity chromatography. Eluted constructs may be identified by sequencing of cDNA or amplified by PCR and transcription, creating an enriched mRNA pool for repeating the procedure. Proteins can be identified directly with SDS-PAGE/MS.

In most target ID strategies, including those already discussed, the immobilized drug serves as the anchor for affinity-based target enrichment. In contrast, several assays immobilize expressed proteins and use a tagged drug to report the interaction. For example, drug-westerns immobilize expressed proteins, from localized *E. coli* colonies, onto nitrocellulose paper that is incubated with the affinity probe.\(^{17}\) Unbound probe is washed away, and a localized signal (*e.g.* chemiluminescence, fluorescence) is produced by the tag. The location of the signal can then be traced back to the *E. coli* expression host for amplification and identification. In a similar manner, microarray technologies, traditionally used for DNA hybridization assays, have also been employed for target ID.\(^{18,19}\) Each expressed fusion protein is immobilized, or spotted, to a small section of a functionalized surface (*e.g.* microscope slide) using a protein tag (*e.g.* GST), as depicted in Figure 1.4. The surface is then incubated with the labeled drug,
washed and assayed. The bound probe leaves a binding profile of the drug for the entire library, which can easily be compared to other compounds using the same array and assay conditions. Unfortunately, this technique requires time intensive purification of individual participating proteins. The concentrated immobilization of proteins in these assays may lead to greater interference with probe binding than with other expression-based strategies.

![Figure 1.4](image)

**Figure 1.4.** Schematic representing a protein microarray outcome in which screened protein library is spotted individually on a derivitized surface. After binding, unbound drug is washed from the array. The drug is labeled to be able to give a localized signal for identifying its binding partner(s).

A variation on the microarray platform indirectly spots the protein library by expressing individual proteins in localized, adherent mammalian cell populations. These living microarrays are created by spotting the corresponding cDNA to the array, then overlaying the slide with transfection reagent, and finally incubating the slide in detached cell culture which naturally adhere themselves to the surface. Prepared slides can then be incubated with the cell permeable probe, washed, and analyzed to obtain an in cellulo protein binding profile for the compound. However, inconsistent transfection and protein expression hinder assay reproducibility and therefore drug profile comparisons.
Tag-free Target ID Strategies

Drug labeling is a necessary evil for all target ID methods discussed thus far, and no universal tagging strategy exists. Discovering modifications that do not alter drug pharmacology can be a bottleneck requiring significant time and expertise. Most “tagless” strategies employ functional assays that are more abstract and require confirmation with more concrete binding studies.

Recently a novel binding assay was reported that recognizes that higher levels of protein structure are stabilized in a ligand-bound state.\textsuperscript{21,22} Drug Affinity Response Target Stability (DARTS)\textsuperscript{23} aims to take advantage of the resistance to proteolysis\textsuperscript{24} afforded by this stabilization. This simple assay comprises of equivalent protease treatment of cell lysates treated with varying concentrations of drug. Bound proteins are negatively selected by proteolysis, and are identified by variable band intensity on an SDS-PAGE gel. Higher concentrations of drug theoretically enhance resistance to proteolysis in drug targets by shifting the kinetics to a stabilized, ligand-bound state. The assay has been used to confirm several drug targets from cell lysates,\textsuperscript{25} including EF-1α for didemnin B. Detection sensitivity is a concern because resistant proteins can’t be amplified, but DARTS is unique in its claim to identify drug targets with a binding assay that does not require modification of the drug or proteins being screened.

Biochemical suppression is a functional but tagless method that attempts to identify drug targets through iterative protein fractionation. This strategy requires that the drug induces notable inhibition of activity in an in vitro assay (e.g. actin assembly) of a protein extract. Fractions of untreated extract are added to aliquots of unfractionated extract that has been drug inhibited,\textsuperscript{26} as depicted in Figure 1.5. Fractions containing target proteins are expected to ‘treat’ inhibition, restoring assay activity at least in part. Theoretically this is a consequence of an increase in drug binding sites from these fractions, shifting kinetics towards an increase in unbound target. From here iterative fractionation of the suppressing fractions can be carried out until the target(s) is isolated and identified. Unfortunately, suppressing fractions may instead (or additionally) contain protein(s) which functionally complement the target’s
inhibition (e.g. a downstream component of the target’s biochemical pathway), thereby confounding interpretation of results.

**Figure 1.5.** Schematic depicting a biochemical suppression strategy that uses an in vitro activity assay. For simplicity, only unbound target produces activity. Aliquots of the drug inhibited sample (upper right) are ‘treated’ with fractions of the untreated sample (upper left). Only fractions with drug target should restore assay activity (center). Iterative fractionation (not shown) and assays are used to enrich the target for identification.

**Yeast Genetic Systems**

The model eukaryote *Saccharomyces cerevisiae*, has been used to discover drug targets. Yeast have proven themselves invaluable tools for understanding mammalian cellular processes and diseases, in part because they are much cheaper, faster, and easier to maintain than many other model organisms or mammalian cell lines. Genetic manipulation of yeast, both chromosomally and through the introduction of exogenous plasmids, has proved to be much easier in yeast than higher organisms. Their utility has been enhanced with the sequencing of the entire genome in 1996. Since then, mutant strain collections have been produced in which each of the nearly 6,000 genes have been individually deleted, many of
which have known human homologs. Even before this, yeast have been used to discover the molecular targets of drugs such as rapamycin,\textsuperscript{30} using targeted yeast mutations.

Target discovery using yeast can take advantage of modern computational tools that enable cataloguing, comparing and deconvoluting large biological data sets. These strategies operate on a model where the drug inhibits its target in manner that mimics the reduction or complete absence of that protein from the cell. One such strategy monitors the competitive growth of pairs of drug treated diploid yeast strains. One assayed strain possesses a single functional copy of a screened essential gene.\textsuperscript{31,32} The decrease in gene copy number theoretically reduces the resulting protein’s concentration. If this protein is a drug target, the protein concentration differential may confer a drug induced competitive growth disadvantage (possible lethality) to the single copy strain, as depicted in Figure 1.6. This phenotypic difference should be dose dependent, requiring a certain concentration to sufficiently inhibit the target concentration in the single copy strain, but that is insufficient to inhibit the increased target concentration in the wild-type strain. Confirmation of the phenotypic influence of that protein can be achieved by overexpression,\textsuperscript{33} which should reverse the effect and perhaps result in a competitive advantage. Like other functional strategies, it is tenuous to conclude from these assays whether an indicated gene product is the target of the drug assayed, or a protein bearing some functional relationship with the target. Because not all gene copy reductions confer this type of protein reduction, other methods to reduce protein concentrations have been devised (e.g. protein destabilization, siRNA).\textsuperscript{34} Inherently this method can only apply to drugs able to impact yeast growth.
**Figure 1.6.** Schematic representation of a drug-induced haploinsufficiency competition assay of an essential target gene. In this example, diploid yeast with two copies of the target gene (green) produces twice as much target protein as the strain possessing only one gene (orange). Untreated yeast strains grow at equal rates. Drug treatment fully inhibits the product of one gene copy, stifling growth in the mutant single copy strain. Inhibition of the wild-type strain is reduced due to increased target protein content, and therefore growth is relatively unaffected.

Another strategy compares gene expression profiles of drug treated wild-type haploid yeast to profiles of untreated gene knockout (KO) strains. Since many drugs act by functionally “knocking out” their target(s), mRNA levels (quantified with cDNA microarrays) of functionally associated proteins may reflect this. In theory, targets of new compounds could be identified by profiling mRNA expression of drug treated yeast and comparing results to catalogued profiles of KO strains.

This manner of pattern matching is not limited to mRNA expression. A blunter example that has been reported is the comparison of drug induced synthetic lethality with genetic synthetic lethality. For this strategy, a collection of nonessential gene deletion strains is first arrayed. To create a synthetic lethal profile, a common second nonessential gene is knocked out in all strains, and the effect on cell viability is catalogued. If the products of the two nonessential genes interact or participate in a common pathway, a double knockout (DKO) is more likely to induce lethality. This understanding has been used in yeast and other model organisms to study protein-protein interactions. Similarly, if a drug functionally knocks out the product of a nonessential gene, it should reduce viability in KO strains where a functionally associated protein is absent. The drug induced lethality on each KO strain is catalogued and compared to
the viability of DKO strains. Theoretically, the genetic and drug-induced lethality profile of a target gene should bear strong similarity, as depicted in Figure 1.7.

![Figure 1.7. Schematic depicting a small synthetic lethality profile of 6 possible target proteins (1-6) and drug treatment (right), based on three random nonessential genes (A, B, C). Each square in the main grid catalogues the viability of the DKO strain of intersecting genes (e.g. A and 1), and on the right viability of drug treated KO strain (A, B, C). An X indicates the strain did not grow in rich media. Target gene deletion (2) and drug treatment both induce lethality in A and C KO stains (as does gene 1).](image)

As previously discussed, these functional assays are rarely sufficient to identify a physical drug-target interaction with the confidence of binding assays. However, yeast affinity-based reporter systems have been developed for detecting molecular interactions in cellulo. These systems were first employed for the detection of protein-protein interactions, in what is known as the yeast two-hybrid system.

The yeast two-hybrid system essentially works by coupling yeast transcriptional machinery to the interaction in question. Specifically, a pair of interacting proteins is required to initiate the expression of a tailored reporter gene. In a simple transcription model (Figure 1.8A), a monomeric transcription factor contains two domains, a DNA binding domain (DBD) and an activation domain (AD). The DBD has sequence specific DNA affinity which localizes the transcription factor to the gene or genes to be activated, while the AD recruits the transcriptional machinery which ultimately initiates proper RNA
polymerase function. In a yeast two-hybrid system (Figure 1.8B), the DBD and AD are fused to two different mammalian proteins, creating two hybrid proteins. Proper function of the transcription factor can only be restored with suitable affinity between these hybridized mammalian proteins. For the purpose of the assay, the mammalian protein fused to the DBD is called the bait (DBD-bait), and the AD is fused to the prey (AD-prey). The recognition sequence of the DBD employed is manipulated to localize the system and express the desired reporter gene. In the first yeast two-hybrid systems, fragments of the yeast gene \textit{GAL4} were cloned to make the DBD and AD. The Gal4p is a transcription factor of the yeast galactose metabolism pathway, initiating over 1,000 fold expression of both \textit{GAL1} and \textit{GAL10}. Therefore, cloning the promoters for either \textit{GAL1} or \textit{GAL10} upstream of a suitable gene creates a dynamic reporter for the interaction in question. The galactose pathway is still the biological basis for most yeast hybrid systems, including the conditional expression of proteins DBD-bait, AD-prey in many experiments. However, the Gal4 DBD and AD have for many investigators been replaced by DBDs and ADs from other organisms, commonly the LexA DBD and B42 AD from \textit{E. coli}. A LexA-based yeast two-hybrid system is sometimes referred to as a yeast interaction trap. These changes were made in part to prevent the system from disturbing normal cellular function due to overexpression of \textit{GAL4} domains.
Figure 1.8. Schematic depiction of yeast one- (A), two- (B), and three- (C) hybrid systems. Reporter expression is activated by localizing the activation domain (AD) to the promoter of the reporter gene (green arrow). The DNA binding domain (DBD) localizes the system to the promoter, which is directly fused to the AD in the one-hybrid system. The two-hybrid system depends on bait and prey dimerization for reporter activation, which is mediated by a small-molecule (CID) in the three-hybrid system. D) Schematic of yeast colonies in three-hybrid screen in which each colony possesses identical CID and bait protein. Each colony expresses a different prey, which must bind the drug ( tethered to the anchor moiety, AM) to activate reporter expression.

A derivative of this assay used to study protein-small molecule interactions, termed the yeast three-hybrid system,\textsuperscript{42} requires only the addition of a small-molecule to mediate bait and prey ‘dimerization’. This compound is thus termed a chemical inducer of dimerization (CID). The CID serves as a third hybrid: a chimera of two moieties, one with affinity to the bait and the other with affinity to the prey (Figure 1.8C). For target screening (Figure 1.8D), this hybrid consists of the drug tagged by an established high affinity ligand for the bait or prey (by convention usually the bait). This high affinity
interaction creates a platform, or anchor moiety, for screening cloned proteins against the tethered query compound. An example of a validated anchor moiety is the protein-ligand duo dihydrofolate reductase and methotrexate (DHFR-Mtx)\textsuperscript{43} due in part to its picomolar affinity. A system using DHFR-Mtx has been used to successfully screen the mammalian proteome for targets of kinase inhibitors.\textsuperscript{44}

**Table 1.2.** Selected examples of miscellaneous target discovery strategies.

<table>
<thead>
<tr>
<th>Method</th>
<th>Drug (Target, ref)</th>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phage display</td>
<td>HBC (Ca\textsuperscript{2+}-calmodulin, Shim, J. S. et al. <em>Chem. Biol.</em> 2004, 11, 1455-1463.)</td>
<td><img src="image1.png" alt="HBC probe" /></td>
</tr>
<tr>
<td>Transfected Cell Microarrays</td>
<td>SCH23390 (Dopamine receptor, Ziauddin, J.; Sabatini, D. M. <em>Nature</em> 2001, 411, 107-110.)</td>
<td><img src="image5.png" alt="SCH23390 probe" /></td>
</tr>
<tr>
<td>Method</td>
<td>Compound Name</td>
<td>References</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
Conclusions

Target discovery is critical for understanding the mechanism of action of any small molecule. Since the majority of known drug targets are proteins, rational drug discovery and development can be dramatically improved with the understanding of specific proteins that directly interact with drugs. This knowledge can be used to improve drug potency and/or reduce off-target effects, enabling in vitro and in vivo assays in which a specific drug-protein interaction is monitored and compared with a variety of analogues. Many affinity chromatography methods have been designed to enrich target proteins from a normal cell lysate, which are subsequently identified by mass spectrometry-based sequencing. However, many drug targets are known to be expressed in low concentrations and possess only moderate target affinity. This can be problematic when using the protein detection techniques employed in these assays. Molecular biology allows for the selective, controlled overexpression of mammalian proteins in a number of different organisms or in vitro, allowing some of the limitations of classic affinity chromatography to be overcome.

The yeast three-hybrid system is a valuable tool that uses of the genetic flexibility of yeast to present fused mammalian proteins to tagged drug targets in cellulo. Research directed at improving the sensitivity and ease of detection in these assays is described in this thesis.

References


CHAPTER 2. Improving the Dynamic Range of a Yeast Fluorescent Reporter

Introduction

To date, most commercially available yeast hybrid systems do not feature fluorescent reporters such as GFP, nor is their use in the literature pervasive. One of the more common reporter types employs an essential gene that has been knocked out in the assay strain, and thus reporter activation is required for yeast viability. Genes that impart prototrophy (the ability to synthesize a specific compound required for growth) are most suitable for yeast hybrid systems, allowing the gene knockout to be compensated by growth on appropriate media while the genetic system is being prepared. For example, a yeast strain lacking HIS3 is grown in histidine supplemented media to maintain viability until the DBD-bait, AD-prey, and a reporter vector bearing an AD inducible HIS3 open reading frame are transformed into cells. After the hybrid proteins are expressed, cells are then switched to defined media lacking histidine, requiring activation of the HIS3 expressing reporter for selection. Several other validated genes involved in amino acid synthesis have been used in a similar manner (e.g. LEU2, TRP1, etc). Plasmids for these assays typically use the same set of genes for positive selection of transformants, and care must be taken to avoid overlap between plasmid markers and reporters. Auxotrophy in theory reduces false positives because reporter expression is required for cell viability. However, these types of selective reporters do not allow for transient expression of Y3H components, which may be preferable to prevent possible toxicity from constitutive overexpression of heterologous proteins. The most prolific non-selective reporter used in yeast hybrid assays is the E. coli gene LacZ, which encodes the enzyme β-galactosidase (β-gal). This protein catalyzes hydrolysis of the β-glycosidic bond of galactose linked carbohydrates (β-glycosides). Advantageously, particular chromophores may be linked with galactose, creating a chromogenic β-gal substrate to detect LacZ reporter expression. One such substrate, X-gal (bromo-chloro-indolyl-galactopyranoside), upon β-gal cleavage releases a substituted indole that strains the yeast colony blue. If transient expression is desired, a liquid assay of permeabilized cultures from individual colonies is necessary. The most common liquid assay substrate, ONPG (o-nitrophenyl β-D-galactopyranoside),
releases a yellow ortho-nitrophenol-derived chromophore that can be detected by absorbance measurements at 420 nm. Proof of concept experiments with liquid assays have demonstrated that this activity is loosely correlated to small molecule-prey affinity in a Y3H assays, although with a low dynamic range. ³ A more sensitive liquid assay uses a chlorophenol red releasing substrate termed CPRG (chlorophenolred-β-D-galactopyranoside). Unfortunately this substrate is significantly more expensive than ONPG. Fluorogeneic β-gal substrates have also been reported (e.g. fluorescein-di-beta-D-galactopyranoside).⁴ LacZ is not an ideal reporter for proteomic scale screens since no effective method exists to separate expressing and non-expressing cells in mixed cultures, and reporter quantification requires cell lysis and time-sensitive, substrate consuming assays. The use of multiple reporters (such as HIS3 and LacZ) is not uncommon, as it is an effective strategy to reduce false positives, and the total number of colonies required for more labor intensive confirmatory liquid assays.

Green fluorescent protein (GFP) from Aequorea victoria, along with its many variants, has become the default tool to observe gene expression and protein localization in live cells of various organisms (reviewed in ⁵), including yeast.⁶,⁷ The subject of the 2008 Nobel Prize, the 700 bp gene encoding this protein can be cloned into most open reading frames often with little consequence to the gene products native properties, except the addition of GFP’s 395/510 absorbance and emission spectra⁸ (488/510 for the popular spectral variant EGFP). Unlike HIS3 and LacZ, GFP requires no special media or substrate for signal detection. The tripeptide fluorophore (Ser-Tyr-Gly in wtGFP, Gly-Tyr-Gly in EGFP),⁹ is induced by higher level structural elements to rearrange, cyclize, and dehydrate to form an imidazolin-5-one heterocycle. This is followed by the oxidation of the Tyrosine alpha-beta carbon by O₂, conjugating the ring system to the tyrosine phenyl ring (Figure 2.1A). The mature fluorophore is stabilized and protected in an iconic β-barrel structure, which has a half life of around 7 h in cells.¹⁰ Variants of GFP have been used in yeast two-hybrid assays to detect the interactions of proteins,¹¹-¹³ validating its potential in similar yeast three-hybrid systems.
Figure 2.1. Maturations of EGFP fluorophore (A) takes place inside the β-barrel structure and requires no cofactors or post-translational modifications. Live cells can be sorted based on the fluorescence intensity of expressed EGFP with FACS (depicted in B).

For high throughput use, a fluorescent reporter such as GFP has a major advantage over other reporters due to its compatibility with fluorescence-activated cell sorting (FACS).\textsuperscript{13-15} This technology, depicted in Figure 2.1B, uses microfluidics to push a cell culture into a single file stream that passes through an excitation laser. Emitted fluorescence from each cell can be measured, rather than a batch culture measurement. A charge is placed on each cell that is directly proportional to its measured fluorescence intensity. Cells are then deflected into different containers based on the associated charge. If a yeast three-hybrid system could differentially trigger expression of a fluorescent reporter protein based on interaction of components in cells, this technology would allow live cells to be separated based on the strength of a protein-ligand interaction. From here, viable cells could be isolated through growth on solid media. Hits can be lysed and their protein encoding plasmids isolated. From here the mammalian gene, theoretically the target of the query affinity probe, could be sequenced and/or cloned into other plasmids for further validation.
The Peterson Lab previously used a reporter encoding EGFP to successfully screen a mammalian cDNA library and identify tyrosine kinases using a variant of the yeast two-hybrid system known as the yeast tribrid system. The employed reporter gene, \( yEGFP \), is a variant of GFP that encodes a mutated fluorophore to improve fluorescence quantum yield and was codon optimized for translation in yeast. The vector used for this reporter contains eight LexA binding sites and the \( GALI-10 \) divergent promoter, taken from the commercial \( LacZ \) reporter pSH18-34 (Invitrogen) and inserted into the leucine selectable plasmid pBC103. The resulting plasmid, pDCLryEGFP, contains a 2-micron origin of replication, an ADH1 transcription terminator (downstream of GFP), and a \( LEU2 \) selection marking to maintain a high plasmid copy number (50-100 copies/cell) in yeast grown in media lacking the essential amino acid leucine (see Figure 2.11).

**Figure 2.2.** Overlay of flow-cytometric histograms demonstrating fluorescence from pDCLryEGFP reporter expression. Each histogram represents 15,000 events, with a gate set to quantify the number of highly fluorescent cells from each experiment. Basal fluorescence was quantified with yeast transformants harboring empty plasmids (white). Background fluorescence produced with only the \( yEGFP \) reporter (light gray) and maximal fluorescence with reporter activated by a LexA-B42 fusion (dark gray). Adapted from Clark, D.D. and Peterson, B. R. *ChemBioChem* **2005**, 6, 1442-1448.

Even though pDCLryEGFP proved adequate in a yeast tribrid assay to identify the kinase Fyn from a human T-cell library, the limited 20-fold dynamic range made data interpretation difficult. Flow-cytometry histograms in a one-hybrid assay (Figure 2.1, dark grey), where the “affinity” between the
DBD and AD is functionally infinite, shows bimodal EGFP expression. Approximately a third of cells express background levels of EGFP, and are likely the result of leaky reporter expression (compare with yEGFP reporter alone, light grey). This background fluorescence, produced by copies of a non-activated plasmid reporter, is significantly greater then basal cell fluorescence (white) and produces cells with fluorescence intensity within a decade of LexA-B42 (DBD-AD) activated reporter expression. This low dynamic range would likely contribute to many false positives and/or negatives in a screen using components from this system.

**Design and Evaluation of GFP reporter constructs**

To best take advantage of FACS screening, the fluorescence profiles of control one-hybrid assays (as in Figure 2.2) should be improved. To improve the dynamic range, the major goal was to reduce the leaky reporter expression. An initial hypothesis posited that cryptic vector enhancer sites of pDCLryEGFP induced non-activated EGFP expression. To test this, yEGFP (along with ADHt) was cloned into the commercial yeast reporter pSH18-34, replacing the LacZ open reading frame. Plasmid pSH18-34 has shown much greater dynamic ranges in reporter expression, and because of its widespread use is compatible with many existing commercial yeast two-hybrid systems (and therefore potential three-hybrid systems). The resulting plasmid, labeled pSHyEGFP (see Figure 2.10), produced almost a 5 fold increase in signal to noise ratio (S/N) in a LexA-B42 one-hybrid assay (Figure 2.3). This effect was almost equally the result of decreased background (45%) and increased fluorescent signal (161%).
Figure 2.3. Overlays of flow-cytometric histograms to compare fluorescence generated by reporters pDCLryEGFP (A) and pSHyEGFP (B). Each histogram represents 10,000 events with a gate for analysis of highly fluorescent cells. The gate begins at the 90\textsuperscript{th} percentile of induced yeast lacking LexA-B42 (red histogram). C) Table of relevant statistics taken from histograms. The median signal and coefficient of variation (CV) is taken from the gated population of induced cells bearing the complete system (black histogram). The signal to noise ratio (S/N) is calculated as the quotient of the signal median and the background threshold (90\textsuperscript{th} percentile of the omission control).

Hoping to further improve fluorescence profiles, a series of protocol alterations were designed and tested with pSHyEGFP, changing growth and/or induction conditions (data not shown). Although these assays helped streamline the protocol, these changes did not yield consistent improvements in dynamic range, and unfortunately highlighted problems with reproducibility for these assays. To further investigate, the temporal correlation with background fluorescence was specifically monitored. Sequential transformations of reporter and the LexA-B42 transcription factor confirmed that increased background fluorescence correlated with the time between reporter transformation and the assay itself (Figure 2.4). Whereas it was suspected that this background was likely a result of leaky EGFP expression, it was not
initially suspected that EGFP significantly accumulated over extended periods in cells were grown in the sugar dextrose, which should repress expression of the reporter protein. This finding led to a change in strategy. Rather than attempting to reduce leaky EGFP expression, perhaps the resulting fluorescence could be limited by increasing the turnover of the protein product.

![Graph A](image1.png) \( \text{Fluorescence} \) \( \text{Events} \)

![Graph B](image2.png) \( \text{Fluorescence} \) \( \text{Events} \)

<table>
<thead>
<tr>
<th>1. reporter</th>
<th>1. LexA-B42</th>
<th>Order</th>
<th>1. LexA-B42</th>
<th>2. reporter</th>
</tr>
</thead>
<tbody>
<tr>
<td>31,623</td>
<td>Background</td>
<td></td>
<td>13,895</td>
<td></td>
</tr>
<tr>
<td>1,416,891</td>
<td>Med. Signal</td>
<td></td>
<td>1,162,313</td>
<td></td>
</tr>
<tr>
<td>83.2</td>
<td>CV,%</td>
<td></td>
<td>81.7</td>
<td></td>
</tr>
<tr>
<td><strong>44.8</strong></td>
<td>S/N</td>
<td></td>
<td><strong>83.6</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.4.** Overlays of flow-cytometric histograms comparing pSHyEGFP one-hybrid assays with different plasmid transformation sequences (A,B). Cells were assayed approximately 8 days after the first transformation and 4 days after the second. Each histogram represents 10,000 events with gates set as described in Figure 2.3. Red and black histograms represent uninduced and induced reporter expression, respectively. C) Table of relevant statistics (described in Figure 2.3).

Mateus and Avery previously reported that by fusing the carboxy terminus of CLN2, a tightly regulated yeast G(1) cyclin, the half life of GFP in yeast could be reduced from 7 h to 30 min.\(^{10}\) The fused 178 residues are known to contain a number of PEST sequences, which destabilize proteins by expediting their polyubiquitination and result in their destruction via the proteosome. I cloned this fragment onto the C-terminal end of \(y\)EGFP in pSHyEGFP, creating pSHyEGFPpst.
Plasmid pSHyEGFPpst was tested with simultaneous transformation for a direct comparison with other reporters. Whereas the median signal was reduced to levels below pDCLryEGFP, the background fluorescence was reduced substantially (> 90%) compared to pSHyEGFP and pDCLryEGFP, resulting in almost another 5 fold increase in S/N (Figure 2.5). This improvement can be explained by the intended reduction in half-life of both the background and induced expressed EGFP-PEST protein. Compared to untagged EGFP, the rate of degradation of EGFP-PEST is much closer to the rate of constitutive expression of background EGFP(-PEST). However, the increased degradation of Lex-B42 activated EGFP is a detrimental consequence of this increased degradation rate, resulting in the reduction in signal fluorescence intensity. The shorter EGFP-PEST half-life also means that optimal timing between LexA-B42 induction and flow-cytometric evaluation is increasingly important. Reporter expression induced by the LexA-B42 transcription factor produces a temporary spike in the EGFP-PEST expression rate, but the “EGFP boom” degrades shortly after the expression surge has ended.

Figure 2.5. (A) Overlay of flow-cytometric histograms from reporter pSHyEGFPpst. Background fluorescence from induced yeast lacking LexA-B42 (red) was used to determine the background threshold. (B) Table of relevant statistics.

<table>
<thead>
<tr>
<th>pSHyEGFPpst</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>730.0</td>
</tr>
<tr>
<td>Med. Signal</td>
<td>362.661</td>
</tr>
<tr>
<td>(CV, %)</td>
<td>114.3</td>
</tr>
<tr>
<td>S/N</td>
<td>496.8</td>
</tr>
</tbody>
</table>

To demonstrate the usefulness of the new reporter to detect protein-ligand interactions, pSHyEGFPpst was applied to detect β-estradiol (E2) induced dimerization of human estrogen receptor β
ligand binding domains (ERβ(255-509)). The native ligand β-estradiol is known to have nanomolar affinity (0.4 nM) and has been shown to promote ligand dependent dimerization in a yeast two-hybrid system.\(^{11}\) Reporter pSHyEGFPpst expression was activated in a sigmoidal E2 concentration dependent manner over a 5 pM-500 nM concentration range, with a picomolar EC\(_{50}\) (Figure 2.5). The high signal variability of these three-hybrid assays led to significantly greater CV’s than in the one-hybrid assay, suggesting that median fluorescence may not be the most appropriate statistic for sample comparison. Instead, total counts of cells exceeding an arbitrary threshold of 100 rfu was quantified in each sample. Plotted against E2 concentration, the total counts yield a dose-response curve with a greater dynamic range than could be obtained by quantification of signal medians.
Figure 2.6. Overlay of flow-cytometric histograms demonstrating E2 dependent dimerization of ERβ(255-509) monomers. Each histogram represents 10,000 events with gate set to quantify highly fluorescent cells (>10^2 rfu) used in analysis. A) Bait and prey omission controls, treated with 500 nM E2, show that each component is necessary for high levels of fluorescence. B) Complete three hybrid-system incubated with three representative E2 concentrations. C) Dose-response curve for three-hybrid assay. Estradiol treatment concentration is plotted against normalized total count (100% ≈ 2,000 or 3,0000) of cells emitting fluorescence >100 rfu.

We also ran similar yeast two-hybrid assays with ligand binding domains (LBDs) of human androgen receptor (AR(670-919)) using the ligand dihydrotestosterone (DHT), as well as progesterone receptor (PR(682-933)) using progesterone (P4). These assays showed dose dependent reporter expression, but omission control assays showed that prey domains were unnecessary to induce reporter expression. The B42 independent reporter activation is likely due to a C-terminal transcription activation factor domain, termed TAF-2, known to require hormone binding for activity. We have been used to study mammalian steroid receptors through ligand-dependent transcriptional activation, but typically
these assays use known receptor response elements upstream of the reporter. A conserved hydrophobic domain located in the LBD of AR (between residues 890 and 900) and PR (between 905 and 915) has been shown to engage transcriptional machinery in yeast.\textsuperscript{24, 25} The TAF-2 domain of ER\( \beta \) (between residues 535 and 545) was not included in the fragment used for ER dimerization assays.

![Flow-cytometric histograms](image)

**Figure 2.7.** Overlays of flow-cytometric histograms from PR (P4 CID) and AR (DHT CID) yeast ligand dependent two-hybrid assays. Each histogram represents 10,000 events with gate of highly fluorescent cells (>100 rfu). A,C) Omission controls with maximum CID treatments and (B,D) representative histograms of complete system with added agonist shown.

Dose dependent reporter activation with PR LBDs was used to compare pShyEGFPpst and the commercial LacZ reporter pSH18-34 (Figure 2.8). LacZ reporter expression was quantified by measuring
β-galactoside activity (CRPG colorimetric assay). There was strong correlation between β-gal activity and highly fluorescent cell counts over all progesterone concentrations tested, and very similar EC\textsubscript{50} values were calculated. Beta-galactosidase activity was measured at various times after substrate addition. At initial times (10, 20 min), the dynamic range (background from untreated transformants) of both reporters was very similar (Figure 2.8), with about 50 fold activation at the highest concentrations. At later times, β-gal activity rose to 200-250 fold at the highest treatment concentrations. Both reporters were expressed when the prey was omitted, as expected due to the presence of the TAF-2 domain in this construct.

![Graph](image)

**Figure 2.8.** Overlay of dose-response curves of ligand dependent PR yeast two-hybrid assays treated with progesterone. The yEGFP curve is based on counts of cells emitting fluorescence >100 rfu. The LacZ curve is based on calculated β-gal activity using a CPRG substrate assay (10 minutes after substrate addition). Curves are normalized based on maximum assay readouts (app. 3,000 cells/units of β-gal activity = 100%).

Not all reporter modifications produced the desired reporter improvements. Several studies have correlated cellular fluorescence intensity of GFP with gene copy number.\textsuperscript{14, 26} Plasmids used for the described yeast genetic systems, including reporter pSHyEGFP\textsubscript{pst}, contain a 2-micron origin and are therefore are autonomously replicating episomal plasmids (yEP). The plasmids replicate at a much higher rate than chromosomal DNA, but are not under mitotic control during cellular division. Therefore
episomal plasmids can reach cellular copy numbers up to 100, but cultures have a high cell to cell copy number variability. This instability makes selection even in pure cultures essential.\textsuperscript{27} Fortunately several other genetic platforms exist where plasmids can replicate with greater stability.

It has been reported that more favorable fluorescence profiles in a yeast two-hybrid system result from controlling \textit{yEGFP} copy number, yielding a more precise and consistent signal intensity.\textsuperscript{28} This was achieved by integrating \textit{yEGFP} (and its control elements) into the yeast genome. I integrated pSHyEGFPpst into the genome of FY250 to yield a rigorously stable copy number per cell. Unfortunately directed integrations are only designed to yield a single copy of the insert, although multiple integrations are possible. The majority of plasmid pSHyEGFPpst was integrated by first deleting its 2-micron origin (removing its autonomous replication capabiliti), transforming it into a yeast integration plasmid (yIP), termed pSHIPyEGFPpst. Integration at the genomic \textit{URA3} locus was performed by restriction digest of pSHIPyEGFPpst at single site within its \textit{URA3} selection marker (\textit{SapI}). The linearized plasmid was then transformed into yeast FY250. Successfully integrated yeast strains should contain \textit{yEGFP} and a functional \textit{URA3} selection marker. A small set of these yeast colonies were separately screened by expression of \textit{LexA-B42} (Figure 2.5A,B). Integrations maintained the low background of episomal pSHyEGFPpst. However, the integrated reporter produced, at best, 40\% of the median signal yielded by its episomal version. Discrete colony to colony variation was likely a result of multiple integration events.
Figure 2.9. Overlays of flow-cytometric histograms of integrated reporter strains FY250yEGFPpst1 (A) and 2 (B), and plasmid reporter pDCUdyEGFP (C). Background fluorescence from induced yeast lacking LexA-B42 (red) was used to determine the background threshold, as described previously. (D) Table of relevant statistics.

<table>
<thead>
<tr>
<th></th>
<th>FY250yEGFPpst1 (A)</th>
<th>FY250yEGFPpst2 (B)</th>
<th>pDCUdyEGFP (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>594</td>
<td>636</td>
<td>38,841</td>
</tr>
<tr>
<td>Med. Signal</td>
<td>17,372</td>
<td>142,618</td>
<td>873,041</td>
</tr>
<tr>
<td>CV, %</td>
<td>80.0</td>
<td>123.1</td>
<td>78.3</td>
</tr>
<tr>
<td>S/N</td>
<td>29.2</td>
<td>224</td>
<td>22.477</td>
</tr>
</tbody>
</table>

In a separated strategy, it was hypothesized that signal intensity might be increased by increasing yEGFP copy number. It has been reported that higher copy numbers (up to 150) could be maintained in 2-micron plasmids with deficient selection markers. An N-terminal deletion in the URA3 promoter, termed URA3-d, yields decreased expression rates of the selection gene, and in turn lower concentrations of the protein product orotidine 5-phosphate decarboxylase. The result is a reduction in de novo synthesis rates of uracil per gene, and thus yeast cells are under selective pressure to have high copy numbers of URA3-d to compensate. I replaced the LEU2 marker in pDCLryEGFP with URA3-d and tested the plasmid, named pDCUdyEGFP, with the LexA-B42 one-hybrid assay (Figure 2.5C). A 74% increase in the median fluorescence intensity was gained from switching from LEU2 to Ura3-d, but still only 67% of the median signal of pSHyEGFP. Unfortunately the background threshold increased by almost 80%, to give a slightly
reduced S/N compared to pDCLrEGFP. It should be emphasized that this strategy was carried out on the plasmid reporter lacking the PEST ubiquitination tag.

**Conclusions**

The results presented here demonstrate destabilization of EGFP has a positive effect on signal to noise ratios in yeast hybrid assays. This effect is mainly seen in a significant reduction in background fluorescence caused by constitutive leaky expression of the reporter. This improved reporter may be useful for more complex yeast reporter assays, as demonstrated by the β-estradiol dependent dimerization of estrogen receptor β. To my surprise yeast copy number did not have a significant positive effect on results of the described yeast one-hybrid assay.

This reporter should be tested in control yeast two- and three-hybrid screens with a cDNA library. It is yet to be seen whether a threshold value can be assigned that effectively distinguishes target protein(s) from the rest of a library, which is essential for selection using FACS. As is common with β-galactosidase reported screens, the simultaneous use of a prototrophy restoring reporter gene (*e.g.* HIS3) could be investigated.

**Experimental**

**General:** Standard techniques for plasmid construction and microbial maintenance/manipulation were employed.\(^3^0\) PCR for plasmid construction was preformed with Phusion Flash PCR Master Mix (New England Biolabs (NEB)) following the manufacturer’s instructions. Reagents for plasmid construction and PCR were purchased from the following sources: DNA oligonucleotides (Integrated DNA Technologies (IDT)); restriction endonucleases ,T4 DNA ligase, T4 DNA polymerase and Phusion DNA polymerase (NEB); and DNA isolation and purification kits (Qiagen and Zymo Research). Microbiological media (Luria broth (LB), Brent Supplement Mixtures (BSM), yeast nitrogen base, carbon sources, yeast extract, and peptone) were from Difco and qBiogene. Prepoured solid yeast selection media
was purchased from Teknova. DNA sequences were confirmed by automated dideoxynucleotide sequencing at KU Medical Center Biotech Support Facility or ACGT.

**Bacterial and Yeast Strains:** *Escherichia coli* DH5-α (Invitrogen) was employed for plasmid construction. *Saccharomyces cerevisiae* FY250 (MATa, ura3-52, his3Δ200, leu2Δ1, trp1Δ63) (a gift of Prof. M. Ptashne (Memorial Sloan-Kettering) was employed for yeast hybrid assays.

**Plasmid Construction:** Reporter plasmid pSHyEGFP is a *URA3* selectable, episomal plasmid containing 8 LexA binding sites (LexAop) and *GAL1-GAL10* (Gal1pr) divergent promoter upstream of *yEGFP* and the yeast transcription termination sequence *yADHt* (see Figure 2.10). This plasmid was constructed using the SLIC technique reported by Li and Elledge, in part because necessary restriction sites in the vector were not available. The vector backbone and upstream control elements was amplified from pSH18-34 using PCR primers (pSH fwd: CATAACGCAGAAAGAACATGTGAGCA; pSH rvs: GCACTTTTTCGCCAATGGTCTTGG). Reporter gene *yEGFP* and *yADHt* were amplified from pDCLryEGFP using PCR primers (VHR-yEGFP fwd: GACCATTGGCCGAAAATGCTAAATGGAAGAAATTATTCACTG; VHR-ADHt-yEGFP: ATGTTCTTCTCTGGGTTATG GCATGCCGTTAGAGGTGT) that flank the insert with 20 bp homology to the vector. Cloning was performed following the protocol as reported, with scaled down 30 min exonuclease chew-back reactions using approximately 500 µg vector and insert PCR products in 10 µL reactions.
Figure 2.10. Plasmid map of pSHyEGFPpst. The 2-micron ori and URA3 selection marker, shown in gold, maintain plasmid copies in yeast. The pBR322 ori and AmpR selection marker, shown in red, maintain plasmid copies for propagation in E. coli. Plasmid pSHyEGFP (7091 bps) lacks the Cln2-PEST (green, lined box) region of the reporter gene, but otherwise is identical. Plasmid pSHIPyEGFPpst (5386 bps) lacks the 2-micron ori, but is otherwise identical to pSHyEGFPpst.

Plasmid pSHyEGFPpst is a variant of pSHyEGFP with a C-terminal fusion of a 178 residue fragment from Cln2. The CLN2 fragment was amplified from purified yeast genomic DNA using PCR primers (VHR-PEST fwd: CCCATGTGATGGATGAATTGGCATCCAACTTGAACATTTCGAG; VHR-PEST rvs: GGCGAAGAAGTCCAAAGCTTGCACCTATATTACTTGATGGATGAATTGCTATATTACTTGATGGATGAATTGCTATATTACTTG) that flanked the insert with 20 bp vector homology. The PCR fragment was cloned into XhoI digested pSHyEGFP using SLIC.
Figure 2.11. Plasmid map of pDCLryEGFP. In pDCUdyEGFP (6994 bps) the LEU2 selection marker (gold arrow) is replaced with a URA3-d selection marker, but otherwise is identical to pDCLryEGFP.

Plasmid pDCUdyEGFP is a derivative of pDCLryEGFP (see Figure 2.11) with a deficient URA3 selection marker (URA3-d). Plasmid pDCLryEGFP minus its LEU2 selectable marker (and controlling elements) was amplified by PCR with primers (pDCLr-LEU2 fwd: ATTAATTCTCATGTTTGACAGCTTATCATCGG; pDCLr-LEU2 rvs: TTCTTGAAGACGAAAGGGCCTCGT). Amplification of the Ura3-d selectable marker was carried out with PCR primers (Ura3d fwd: GGCCCTTTTCGTTTCAAGAA ATAACCCAACTGCACAGAAAC; URA3d rvs: TGTCACGACGAAATTAAT CCCGGTTAAATACTGAATAATTAAATTGAAG) that flanked each end of the product with 20 bp of homology to the ends of PCR product pDCLryEGFP-LEU2. The two PCR products were cloned together using SLIC.

Plasmid pSHIPyEGFPpst was constructed using an intramolecular SLIC reaction to delete its 2-micron origin. Amplification of pSHyEGFPpst minus the entire 2-micron origin sequence was carried out using PCR primers (pSHIPyEGFP fwd:
TCGTCTTTCAAGAATTCTCATGTTTGACAGCTTATCATCGA; pSHIPyEGFP

GTAAACATGAGAATTCTTGAAGACGAAAGGGCCT) that flanked the produce with 20 bp of homology to allow sequence specific circularization following T4 polymerase treatment.

**Yeast Control One-Hybrid Assays with Flow-Cytometry:** Yeast one-hybrid assays are typically used to discover new protein-DNA interactions. For this study, this system was employed for the quantitative comparison of reporter constructs using the heterologous transcription factor *LexA-B42* (previously inserted into the *HIS3* selectable expression vector pAM423). Plasmid pAM423 LexA-B42 (previously constructed) was cotransformed with a reporter plasmid (pDCLryEGFP, pSHyEGFP, pSHyEGFPpst, pDCUdyEGFP, or SapI cut pSHIPyEGFPpst) using the LiAc/SS carrier DNA/PEG into frozen competent yeast on appropriate solid yeast selection media (SD-His-Leu + 2% glucose or SD-His-Ura + 2% glucose).

From selective plates bearing both components of the one-hybrid system, multiple (2-8) large colonies were picked and grown (30 °C incubation with shaking, 300rpm) to saturation (app. 18 h) in selective 2% glucose media (3 mL). Colonies were pooled for episomal reporters, but for the yIP reporter pSHIPyEGFPpst, picked colonies were kept separate, due to possible variability of integration. Aliquots of saturated cultures were pelleted and washed once with PBS, then diluted 1:4 in selective 1% raffinose, 2% galactose induction media (400 µL) and grown for 10 h before quantification. Resultant reporter gene fluorescence intensity was measured by flow-cytometry. Cell viability was monitored concurrently with nuclear staining (using 10 nM propidium iodide). Flow-cytometry was performed on an Accuri C6 Flow-Cytometer excited with a 488 nm laser and collected with a 530 +/-15 nm emission filter (GFP) and a >670 nm emission filter (propidium iodide). These assays were preformed in duplicate from two different sets of colonies which produced similar results. Histograms and statistics are shown from one of these duplicates.
For sequential transformations, either pSHyEGFP or pAM423 LexA-B42 was transformed into frozen competent yeast and plated onto appropriate selection media. Multiple pooled colonies were inoculated into appropriate liquid yeast selection media (3 mL; SD-His-Ura + 2% glucose) and incubated with shaking (30 °C, app. 18 h) to saturation. The resulting cultures were then transformed with the other plasmid using the Quick and Easy LiAc/SS carrier DNA/PEG method\textsuperscript{35} and plated on appropriate selection media.

**Yeast Three-Hybrid Assays with Flow-Cytometry:** Frozen competent yeast FY250 were first transformed with reporter plasmid pSHyEGFP and plated on selective solid media (SD-ura + 2% glucose). Multiple large colonies were pooled and inoculated into appropriate liquid selection media and grown to saturation (3 mL). Saturated cultures were diluted 1:6 in liquid selection media (10 mL) and grown to mid-log phase (app. 4 h) and resuspended in 200 µL sterile water. An aliquot of yeast concentrate (100 µL) was then transformed with previously constructed pAM423 LexA-ERβ(255-509) and pJG45 ERβ(255-509)\textsuperscript{36} using the “Quick and Easy” LiAc/SS carrier DNA/PEG method\textsuperscript{35} and plated on yeast solid selective media (SD-His-Trp-Ura + 2% glucose). From this plate multiple large colonies were pooled and grown to saturation in appropriate liquid selection media. Aliquots (100 µL) were pelleted and washed once with PBS, then diluted 1:4 in selective 1% raffinose, 2% galactose induction media. These cultures were grown for 4 h and then treated with the various concentrations of β-Estradiol (0, 5 pM, 10 pM, 50 pM, 500 pM, 1 nM, 5 nM, 10 nM, 50 nM, 100 nM, 500 nM). Resultant reporter gene fluorescence intensity was measured by flow-cytometry (described previously). These assays were preformed in triplicate from different sets of colonies that produced similar results. Histograms and statistics are shown from one run.

Cells for the androgen and progesterone receptor homodimerization assays were prepared similarly to ER dimerization assays. Plasmids pAM423 AR(670-919), pAM423 PR(682-933), pJG4-5 AR(670-919), and pJG4-5 PR(682-933) were previously constructed.\textsuperscript{37} Sequencing (after running the assays) revealed an unexpected and unintended P906S mutation in the progesterone receptor (PR)
fragment used. For PR dimerization, cells were treated with various concentrations of progesterone (0, 0.1 µM, 0.5 µM, 1 µM, 5 µM, 10 µM, 100 µM, and 1 mM). For AR dimerization, cells were treated with various concentrations of dihydrotestosterone (0, 0.1 nM, 0.5 nM, 1 nM, 5 nM, 10 nM, 50 nM, 100 nM, 500 nM, 1 µM, 5 µM, 10 µM). These assays were preformed in duplicate from the same set of colonies that produced similar results. Histograms and statistics are shown from one run.

**Yeast Three-Hybrid Assays with Microtiterplate β-galactosidase Assays:** Cells were prepared, induced, and treated just as cells for flow-cytometry, except reporter plasmid pSH18-34 was used rather then pSHyEGFPst. The resulting treated cultures (400 µL, SD-Trp-Leu-Ura + 1% raffinose + 2% galactose) were harvested, resuspended in Z lysis buffer (38) (additionally containing 0.1% SDS, 200 µL), and transferred to a 96-well plate. The absorbance at 570 nm (OD570cells) was measured to determine cell density, followed by the addition of the substrate chlorophenol red-b-D-galactopyranoside (15 mM, 30 µL) in sodium phosphate buffer (0.1 M, pH 7.5). The plate was shaken at 30 °C with periodic (10 min, 20 min, 40 min, and 60 min) absorbance measurements at 570 nm (OD570). β-Galactosidase activity was calculated as follows: Activity = 1000*(OD570-OD570cells-Blank1)/(TIME*(OD570cells-Blank2)). The Blank1 value represents the absorbance at 570 nm of chlorophenol-red-b-galactopyranoside alone (30 µL). The Blank2 value corresponds to the absorbance at 570 nm of Z lysis buffer alone (150 µL). The TIME value was expressed in minutes

**References**


CHAPTER 3. Investigations into Modifying the Valency of Streptavidin for an Improved Yeast Three-Hybrid Platform

Introduction

The extreme affinity of streptavidin for its natural ligand biotin ($K_d \approx 10^{-14} \text{ M}$)\(^1\)\(^2\) has found utility in many different life science applications, and serves as a model for noncovalent protein-ligand interactions.\(^3\) This interaction is stable under a variety of typically disruptive conditions, such as heat, denaturation, pH, and proteolysis.\(^3\) Biotin can easily be tagged to a variety of molecules, often with little effect on their biological properties.\(^4\) Streptavidin can also be fused to other proteins or conjugated to lower MW molecules\(^4\), although its large size may be problematic for some applications. These features make streptavidin-biotin suitable for a number of detection and separation techniques. Detection typically entails the biotinylation of the compound of interest and the conjugation of streptavidin to the labeling system (an enzyme, fluorophore, etc). For separation or purification of biotinylated molecules, streptavidin is often immobilized to a solid support (e.g. agarose) and captures the biotin moiety from solution. The work presented here was based on the hypothesis that the unmatched affinity of biotin-streptavidin might be used to improve existing yeast three-hybrid systems. A biotin-streptavidin anchor moiety would also be attractive for its compatibility with the many affinity chromatography methods that use biotinylated small molecules.

Streptavidin is an app. 60-70 kDa\(^4\) homotetrameric, tetravalent protein secreted by the bacteria *Streptomyces avidinii*,\(^1\) possibly as a natural defense to deprive competing organisms of the cofactor biotin. Each subunit of streptavidin consists of an essential 126 residues\(^5\) that secondarily form eight antiparallel $\beta$-strands with short connecting loops, and then folds into a classic up and down $\beta$-barrel.\(^6\) Each subunit is said to bind one biotin, but more accurately, a pair of subunits share two biotins. Each monomer largely encapsulates a single biotin towards the end of its barrel, and contributes all of the hydrogen bonds with that ligand. However the conserved W120 from the loop of a neighboring subunit
contributes significantly to the hydrophobic pocket\textsuperscript{7} (Figure 3.1A), enhancing biotin binding by up to 8 orders of magnitude. A functional dimer is formed by an almost perpendicular pair of monomers that use each other’s W120 residues. A structural dimer of dimers completes the quaternary structure, orchestrated in an antiparallel fashion and interfaced by only a handful of residues (Figure 3.1C). Streptavidin is structurally and functionally very similar to the protein avidin,\textsuperscript{3} found in chicken egg white. Chicken avidin actually has a slightly higher binding affinity,\textsuperscript{8} but is notorious for nonspecific binding due to its native glycosylation pattern (mannose and N-acetylglucosamine) and a pI $\approx 10^4$. Avidin also contains structural disulfide bonds that are not likely to be stable in cytoplasm. Therefore streptavidin is preferred for most applications, although an acidic (pI $\approx 4.7$), non-glycosolated variant of avidin has been constructed to overcome some of avidin’s limitations.\textsuperscript{9}
Figure 3.1. Selected views of X-ray crystal structure of tetrameric core streptavidin complexed with biotin (PDB: 1SWE).\textsuperscript{10} A) Within a functional dimer (solid red and blue) two W120 residues (top, shown in sticks) enhance biotin affinity. C) Structural dimer (blue and green) interfaces are more extensive. Selected residues that have been mutated to disrupt oligimerization are shown in sticks. The biotin binding pocket involves a hydrogen bonding network (B) and a hydrophobic pocket (D).

The biotin binding pocket of native streptavidin is highly optimized, and all reported ligand pocket mutants have reduced biotin affinity.\textsuperscript{11} At the binding site, there a total of eight hydrogen bond contacts with streptavidin residues Asn23, Tyr43, Ser27, Ser45, Asn49, Ser88, Thr90 and Asp128 (Figure 3.1B).\textsuperscript{5} These hydrogen bonds cover all the non-carbon atoms of biotin, concentrated at the ureido ring. Paradoxically, biotin is relatively hydrophobic and likewise a hydrophobic binding pocket is maintained by four conserved tryptophan side chains (W79, W92, W108, and W120 from a neighboring monomer, Figure 3.1D).\textsuperscript{12} In the absence of biotin, five water molecules mimic biotin,\textsuperscript{5} but biotin binding induces conformational changes that stabilize a complete streptavidin.\textsuperscript{13}
Despite the utility of the streptavidin-biotin interactions, many modifications have been made to improve its compatibility with a specific application (reviewed in 11, 14). For some precipitation or delivery strategies, a lower affinity streptavidin (e.g. T76R and V125R) allows reversible biotin binding under mild conditions. 15, 16 An E51K, N118K mutant has been conjugated to polymers that change conformation in response to specific stimuli (pH change, light, temperature), creating steric hindrance that induces the capture or release of biotin. 16 However, most modifications aim to adjust the valency of streptavidin. For some detection systems, tetravalency serves to amplify the signal produced by streptavidin fusions. 4 For other systems, tetravalency causes detrimental aggregation of fused proteins. Investigations into altering streptavidin typically aim to disrupt tetramerization through interface mutations, or to cross link monomers together so that they may be individually engineered. A monomeric (and therefore monovalent) streptavidin should effectively tackle aggregation problems, and have the additional benefit of decreasing the size of a streptavidin tag by 75%, reducing the risk of steric disturbance of other parts of the assay. A fully monomeric streptavidin has been reported, but unfortunately its biotin affinity is substantially reduced (K_d \approx 1.3 \times 10^{-8} \text{ M}). This was achieved by knocking out two biotin hydrogen bond contacts (T90A, D128A). 17 The same group later achieved effective monomerization with a set of monomer interface mutations (T76R, V125R, V55T, and L109T), but with further reduced biotin affinity. 15 Attempts have also been made to preserve the functional dimers, but disrupt tetramerization through an H127A mutation and a loop (G113-W120, between \beta-strands 7 and 8) deletion. 18 These mutations did result in a soluble dimeric streptavidin, but biotin affinity was again dramatically reduced (K_d \approx 1.5 \times 10^{-7} \text{ M}). Interestingly, biotin was required for dimerization of this mutant. Reduced binding affinity in these mutants likely results from the disruption of the hydrophobic binding pocket afforded, in part, by a neighboring subunits W120 on the native tetramer. Similar efforts with chicken avidin have yielded analogous results. 19, 20

Another approach to produce a monovalent streptavidin requires combining the subunits of streptavidin into a single polypeptide chain, encoded on a single gene. This would allow individual
monomers to be modified through directed mutations, and create a pseudo-heterotetramer of monomers with different biotin affinities. Unfortunately neither the native N- or C-termini of any monomer fold into positions near subunit interfaces. Therefore the design of a linker connecting native subunit termini’s has never been considered. However, the native termini do fold into a proximity and orientation that allows a short, flexible peptide linker to form a β-turn between the native first and last strand of each subunit. From here, new termini could possibly be designed which allow fusion with minimal effect on protein folding.

Thus far, the only successful design of a single chain dimer of streptavidin uses a GGGS native termini linker, and fuses the 7/8 strand loop of one subunit to 4/5 loop of its partner. The engineered polypeptide’s N-terminus is residue 116 of the native monomer and ends C-terminally with residue 68 of a second subunit. A new amide bond between residues 115 and 69 of the linked monomers is created (Figure 3.2). This design fuses monomers of structural dimer, whose subunits do not rely on each other’s W120 for biotin affinity. This single-chain dimer (SCD) also includes a W5K (W120K in the native topology) to prevent functional dimerization (pseudotetramerization). In theory this would mean that neither biotin of the ‘dimer’ would possess a neighboring W120 necessary for extreme biotin affinity. However, even with this mutation, pseudotetramerization was highly favored in the presence of ligand. Unsurprisingly, biotin affinity was significantly reduced ($K_d \approx 6.7 \times 10^{-6}$ M), but the ligand derivative biotin 4-fluorescein unexpectedly maintained a high affinity ($K_d \approx 1.2 \times 10^{-10}$ M for both wild-type streptavidin and the SCD).
Figure 3.2. View of streptavidin X-ray crystal structure (PDB: 1SWE) with linkers in SCD streptavidin represented by black lines. The GGGS linker (bottom) between the native N- and C- termini (shown in green stick) allows direct fusion between two monomers (top) between S69 and T115 (shown in brown stick). New N- and C- termini (E116 and G68, highlighted in yellow) are created.

Similar aims with chicken avidin have been pursued with greater success. Circular permutations of avidin establish new termini between at the 4/5 loop and the 5/6 loop. The 4/5 subunit permutation was shown to maintain biotin affinity, while the 5/6 permutations forfeited a slight affinity reduction. These permutations were fused with an SGG linker, creating dcAvd. This design was taken further by fusing duplicates of dcAvd with a 12 peptide serine and glycine rich linker. The resulting scAvd subunits maintain wild-type like biotin affinity. To demonstrate the usefulness of a single-chain ‘tetramer,’ biotin affinity was weakened in one of the biotin binding domains to generate a dual-affinity streptavidin (dadcAvd). This scheme has yet to be replicated in streptavidin, but due to their structural similarities, this inspires confidence that a single chain streptavidin is possible. Because of internal disulfide bonds in avidin, the use of avidin in yeast genetic systems is problematic.

In the context of yeast three-hybrid systems, a reduced streptavidin valency might improve assay sensitivity and dynamic range for the majority of the mammalian proteome. This is based on the hypothesis that protein aggregation caused by streptavidin’s tetravalency would sterically hinder certain ligand-prey interactions, thus preventing reporter expression. The reduced biotin affinity of existing
monovalent and divalent streptavidin mutants negates the major potential advantage of biotin-streptavidin three-hybrid systems over existing anchor moieties (*e.g.* Methotrexate-DHFR).

**Design and Evaluation of New Streptavidin Constructs**

To reduce the valency of streptavidin tetramers, my goal was to convert the four subunit complex from a homotetramer of high affinity biotin binding monomers to a heterotetramer of subunits with normal biotin affinity and those with negligible biotin affinity. Howarth *et al.* have reported the creation of a triple mutant N23A, S27D, S45A subunit that maintained the quaternary structure of wild-type streptavidin. Homotetramers of the “dead” mutant yield “negligible biotin binding” (*app. K_d=1.2 \times 10^{-3} M*). I recreated this triple mutation in core wild-type streptavidin (SAwt) with a single round of PCR mutagenesis, terming it SAnb (streptavidin “non-binding”).

Howarth *et al.* combined the purified protein subunits of wild-type and mutant streptavidin, and were able to differentially elute the different combinations of tetramers with affinity chromatography. Unfortunately streptavidin is not yeast cell permeable and yeast hybrid assays require LexA fusions to be present in the nucleus of living cells. The idea of co-expressing wild-type and LexA fused mutant streptavidin monomers was entertained. However this strategy would generate a statistical mixture of mono- to tretravalent streptavidins, each with a different number of LexA DBDs.

I instead decided to try to combine subunits of SAwt and SAnb, first into a dimer, using the circular permutation method described by Aslan *et al.* (described earlier, Figure 3.3). Since maintaining biotin affinity in the active subunits was of utmost importance, the W120K mutation was omitted. I recreated the described 5>4 and 8>7 circular permutations (cp) with SAwt and SAnb, labeling them according their new N- and C-terminal β-sheets (cpSA87 and cpSA54; cpSAnb87 and cpSAnb54).

Permutations of ‘alive’ (A) and ‘dead’ (D) monomers were then fused in different combinations to yield single-chain dimers that are theoretically unable to bind biotin (dSA_{DD}), bind a single biotin (dSA_{DA}), or bind two biotin ligands (dSA_{AA}).
Figure 3.3. Topology map of core streptavidin monomer (A) and SCD (B), with identical coloring of β-strands. Residues involved in biotin binding are identified (H-bonding: grey boxes, Hydrophobic contacts: red boxes), and the three residues mutated to create SAnb (blue arrows). Chainsaws indicate where new termini are created in the circular permutations.

To test these constructs, I employed a yeast three-hybrid system with the modified EGFP reporter pSHyEGFPpst. All created streptavidin mutants are designed to have tetramerization (or pseudo-tetramerization) properties similar to wild-type streptavidin, and therefore should form tetramers with each other and/or wild-type streptavidin monomers. A yeast three-hybrid system was proposed to confirm this interaction by expressing a fluorescent protein that can be quantified easily by flow-cytometry. Up to two versions of streptavidin were tested against one another, fusing one construct to the DBD and another to the AD of a split transcription factor.

Using this system, the LexA-SAwt was tested separately against B42 fused SAnb, circular permutations cpSA_{87}, cpSA_{54}, and the fusion of these permutations dSA_{AA} (Figure 3.4). Each variant tested produced significant fluorescent populations above the set threshold, indicting tetramer formation in each case. Unfortunately, the smaller fluorescent populations in the circular permutations also suggest reduced affinity for oligmerizaition, more so with dimeric fusion of the two. Since it is known that biotin
enhances tetramerization of streptavidin, the induction media (containing 10 nM biotin) was spiked with 1 μM biotin. Excess biotin had little effect on SAwt and SANb (not shown), but significantly increased EGFP expression with bait cpSA_{87}, cpSA_{54} and dSA_{AA}. This result also suggests that the circular permutations have reduced affinity for biotin. To further test this, cpSANb_{54} was fused to B42 and assayed with LexA-fused cpSA_{87} and cpSANb_{57}. These combinations failed to yield highly fluorescent counts of even 1,000 (10 nM biotin), and the addition of 1 μM biotin failed to increase the activated populations (histograms not shown).

**Figure 3.4.** A) Overlays of flow-cytometric histograms of three-hybrid assays with SA variants (bait) and wtSA (prey). Untreated assays contain 10 nM biotin present in media. B) Table of total counts of cells generating EGFP fluorescence greater than 100 rfu.

We later assayed the single-chain dimers for their ability to homodimerize (forming a pseudotetramer) after fusing dSA_{AA}, dSA_{DA} and dSA_{DD} to LexA and dSA_{AA} to B42 (Figure 3.5). Similarly, histograms indicted that homodimers were forming and in AA and DA that was enhanced by
excess biotin, but with even less affinity then the monomers tested previously. As expected, dSA<sub>DD</sub> dimerization was not enhanced by excess biotin due to its negligible biotin affinity.

![Graph](https://via.placeholder.com/150)

**Figure 3.5.** A) Overlays of flow-cytometric histograms of three-hybrid assays with dimeric SA variants (bait) and variant dSA<sub>AA</sub> (prey). Untreated cells (red) contain 10 nM biotin, treated cells (blue) are spiked with 1 µM biotin. B) Table of total counts of cells that generated GFP fluorescence above 100 rfu.

Construction of a single-chain streptavidin was attempted and a few permutations were completed. These genes fuse two dSA with a 12 residue SG rich linker. Unfortunately, tests of these constructs with a previously described yeast tribrid system<sup>26</sup> yielded inconclusive histograms.

**Conclusions**

The results described here indicate that the streptavidin circular permutations first reported by Aslan *et al.* possess reduced biotin affinity, even without the reported W120K mutation. This is inferred
because increased concentrations of biotin are necessary to produce EGFP reporter fluorescence comparable to wild-type streptavidin in otherwise identical yeast genetic systems. Fusions of these permutations were used to create single-chain dimeric streptavdins with theoretical valencies of 0, 1 and 2. Control yeast three-hybrid assays suggest that oligimerization affinity of non-valent dimeric streptavidin is reduced compared to monovalent and divalent single-chain dimers, as increased biotin concentrations marginally increase reporter fluorescence with mono- and divalent streptavdins, but not with non-valent dimers.

In order to create a monovalent, monomeric streptavidin, novel pairs of circular permutations should be designed and investigated. At least one of these permutations should have wild-type biotin affinity, and the permutations should have the ability to be fused, possibly with peptide linkers, in a way that does not alter secondary or tertiary structure. Theoretically two different fusions will need to be made, one to create a single-chain dimer and the other to link homodimers of this construct, thus creating a single-chain tetramer.

**Experimental**

**General:** Standard techniques for plasmid construction and microbial maintenance/manipulation were employed. PCR for plasmid construction was preformed with Phusion Flash PCR Master Mix (New England Biolabs (NEB)) following the manufacturer’s instructions. Reagents for plasmid construction and PCR were purchased from the following sources: DNA oligonucleotides (Integrated DNA Technologies (IDT)); restriction endonucleases ,T4 DNA ligase, T4 DNA polymerase and Phusion DNA polymerase (NEB); and DNA isolation and purification kits (Qiagen and Zymo Research). Microbiological media (Luria broth (LB), Brent Supplement Mixtures (BSM), yeast nitrogen base, carbon sources, yeast extract, and peptone) were from Difco and qBiogene. Prepoured solid yeast selection media was purchased from Teknova. DNA sequences were confirmed by automated dideoxynucleotide sequencing at KU Medical Center Biotech Support Facility or ACGT.
**Bacterial and Yeast Strains:** *Escherichia coli* DH5-α (Invitrogen) was employed for plasmid construction. *Saccharomyces cerevisiae* FY250 (MATa, ura3-52, his3Δ200, leu2Δ1, trp1Δ63) (a gift of Prof. M. Ptashne) was employed for yeast hybrid assays.

**Plasmid Construction:** Streptavidin triple mutant N23A, S27D, S45A was made by PCR mutagenesis from pPA1 LexA-SA, a yeast expression plasmid containing a 2-micron origin of replication, a *LEU2* selection marker and a *GAL1* promoter (see Figure 3.6). Plasmid pPA1 LexA-SA was constructed by ligation of EcoRI, XhoI digested pPA1 LexA and MfeI, XhoI SA(13-139) digested PCR fragment, amplified from pPA1 T7SA-LexA.26

![Figure 3.6. Plasmid map of pPA1 LexA-SA. All streptavidin LexA- fusions were cloned into pPA1 LexA, replacing SAwt.](image)

To make all mutations in one round of mutagenesis, approximately half of pPA1 LexA-SA was amplified with PCR primers (SAtko fwd: CGTGACTGCAGGCGCCGACGGC GCCCTGACCGGAACCTACGAGGCTGCCGTCGGAA; SAtko w/fwd: CGGTATCATTGCAGCAGCAGCTGG) designed to create mutations S27D and S45A. The other half
to the plasmid was amplified with PCR primers (SAtko rvs: CGTCGGCGCCTGCAGTCACGATGAAGGTATCGCCGAGCTGAGCGTACCAGGTGCCGGTGA and SAtko w/rvs: CCAGTGCTGCAATGATACCG) designed to create mutations N23A and S27D and flank the product with 20-bp homology to the other PCR product. These primers also create a silent PstI restriction site to facilitate screening. The PCR products were joined using SLIC\textsuperscript{28} to yield pPA1 LexA-SAnb.

Circular permutations cpSA\textsubscript{87}, cpSA\textsubscript{54} fused to LexA were made with a two insert SLIC reaction. For permutation cpSA\textsubscript{87}, core streptavidin residues 116-139 were amplified by PCR primers (SC1 fwd: CGCAACGGCGACTGCTGGAATTC GAGGCAACCGCTGGAAGTCC; SC1 rvs: GCCGGCCTCGTCCGATCCACCACC GGAGGC GGCCCGAGCG CGCTT) that flanked the product with 5’ 20-bp homology to vector pPA1 LexA and 3’ GGGS linker + SA residues 1-4. Residues 1-115 were amplified by PCR primers (SC2 fwd: GACGAGGGCCCGCATCACC GGCCACCTG GT; SC2 rvs: GAAGTCCAAAGCTTCTCGAG TCA TCCGGAGGTGGTGC CGCGAGGTCAGCAGCCACTG) that flanked the product with 5’ GGGS linking and 3’ BspEI site + stop codon (tca) + 20-bp homology to vector pPA1 LexA. These two inserts were cloned into EcoRI, XhoI cut vector pPA1 LexAst with SLIC. Similarly for cpSA\textsubscript{54}, streptavidin 69-139 were amplified with PCR primers (SC3 fwd: CGCAACGGCGACTGCTGGAATTC GCCGGACCGCATCACC GGCCACCTG GT; SC3 rvs: GGATCCACCACC GGAGGCGCGCGAGCGTTTCACCT) that flanked the product with 5’ 20-bp vector homology + BspEI site + 3’GGGS linker. Residues 1-68 were amplified with PCR primers (SC4 fwd: TCCGGCCGCCTCGTGGTGATCC GACGAGGGCCCGCATCACC GGCCACCTG GT; SC4 rvs: CGAAGAAGTCCAAAGCTTCTCGAG TCA GCCGTCGGTGCCCCGGCGTGTC) that flanked the product with 5’ SA residues 136-139 + GGGS linker and 3’ 20-bp vector homology. These two inserts were cloned into EcoRI/XhoI cut vector pPA1 LexAst with SLIC. Circular permutations cpSAnb\textsubscript{87} and cpSAnb\textsubscript{54} were made from template pPA1 LexA-SAnb using the same primers and protocol.
Streptavidin single-chain dimers (dSA) fused to LexA were created by inserting cpSA(nb)\textsubscript{54} downstream of SA(nb)\textsubscript{87} in pPA1 LexA. Constructs cpSA\textsubscript{54} and cpSA\textsubscript{87} were amplified with sequencing plasmid PCR primers (5’pPA1 LexA: CATTGAAGGGCTGCGGTT; 3’pPA1 LexA: GACAACCTTGAGAGACTTGACC). The PCR product was digested with BspEI, XhoI and ligated into BspEI, XhoI digested pPA1 LexA-cpSA\textsubscript{87} or pPA1 LexA-cpSA\textsubscript{87}.

Converting the streptavidin constructs from LexA fused bait to B42 fused prey was performed by ligating the EcoRI, XhoI digested construct (in pPA1 LexA) with EcoRI + XhoI digested pJG4-5 (see Figure 3.7).\textsuperscript{29}

**Figure 3.7.** Plasmid map of pJG4-5 SA, used to create new B42- streptavidin fusions (SAnb, dSA\textsubscript{DD}).

**Yeast Three Hybrid Assays with Flow-Cytometry:** Frozen competent yeast FY250 were first transformed with reporter plasmid pSHyEGFP\textsubscript{pst} and plated on selective solid media (SD-ura + 2% glucose). Multiple large colonies were pooled and inoculated into appropriate liquid selection media and grown to saturation (3 mL). Saturated culture was diluted 1:6 in liquid selection media (10 mL), grown to mid-log phase (app. 4 h), and resuspended in 200 µL sterile water. An aliquot of yeast concentrate (100
μL) was then transformed with the specified bait and prey streptavidin constructs (on vectors pPA1 LexA and pJG 4-5, respectively) using the “Quick and Easy” LiAc/SS carrier DNA/PEG method[30] and plated on yeast solid selective media (SD-His-Trp + 2% glucose). From this plate multiple large colonies were pooled and grown to saturation in appropriate liquid selection media. Aliquots (100 μL) were pelleted and washed once with PBS, then diluted 1:4 in selective 1% raffinose, 2% galactose induction media. Cultures were then grown for 8 hours. For biotin treated assays, cultures were induced for 4 h before D-biotin (1 μM) treatment. Resultant reporter gene fluorescence intensity was measured by flow-cytometry (described previously).

References


# APPENDIX A. Plasmid List

## Section 1. Empty Plasmids

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Expressed Gene</th>
<th>Restriction Sites</th>
<th>Plasmid Type</th>
<th>Bacterial Resistance</th>
<th>Yeast Marker</th>
<th>Seq. Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBP7</td>
<td>BR Peterson</td>
<td>6x His tag</td>
<td>T7pr-Xbal-MRGSSHHHHHHHP-BamHI-NcoI-KpnI-MfeI-EcoRI-STOP-Xhol-Sall-STOP-HindIII</td>
<td>Bacterial Expression Vector, 2µ Ori</td>
<td>Amp</td>
<td>N/A</td>
<td>5' pLM, 3' pLM</td>
</tr>
<tr>
<td>pJG4-5</td>
<td>Brent Lab</td>
<td>SV40NLS-B42 AD-HA</td>
<td>Gal1pr-HindIII-SV40 NLS-B42 AD-HA Tag-EcoRI-Stuffer-Xhol-HindIII-ADH1t</td>
<td>Yeast Expression Vector, YEp</td>
<td>Amp</td>
<td>Trp 1</td>
<td>5' pJG4-5, 3' ADH1</td>
</tr>
<tr>
<td>pPA1 LexA stop</td>
<td>S Athavankar</td>
<td>HA-SV40NLS-LexA</td>
<td>KpnI-Gal1pr-EcoRI KO-HA SV40 NLS-LexA -EcoRI-Xhol-stop-ADH1t-BamHI-Sall-PstI</td>
<td>Yeast Expression Vector, YEp</td>
<td>Amp</td>
<td>LEU2</td>
<td>5' Gal1, 3' ADH1</td>
</tr>
</tbody>
</table>

## Section 2. Reporter Constructs

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Expressed Gene</th>
<th>Restriction Sites</th>
<th>Plasmid Type</th>
<th>Bacterial Resistance</th>
<th>Yeast Marker</th>
<th>Seq. Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDCLr yEGFP</td>
<td>D. Clark</td>
<td>yEGFP</td>
<td>KpnI-8x LexAop-Gal1-10pr region-EcoRI-yEGFP-Xhol-ADH1t-BamHI</td>
<td>Yeast reporter plasmid (LexA DBD), YEp</td>
<td>Amp</td>
<td>LEU2</td>
<td>5' Gal1, 3' ADH1</td>
</tr>
<tr>
<td>pDCUd yEGFP</td>
<td>K. Bailey</td>
<td>yEGFP</td>
<td>KpnI-8x LexAop-Gal1-10pr region-EcoRI-yEGFP-Xhol-ADH1t-BamHI</td>
<td>Yeast reporter plasmid (LexA DBD), YEp</td>
<td>Amp</td>
<td>URA3-d</td>
<td>5' Gal1, 3' ADH1</td>
</tr>
<tr>
<td>pSH18-34</td>
<td>Brent Lab</td>
<td>beta-Gal reporter</td>
<td>N/A</td>
<td>Yeast reporter plasmid (LexA DBD), YEp</td>
<td>Amp</td>
<td>Ura 3</td>
<td>5' Gal1, 3' ADH1</td>
</tr>
<tr>
<td>pSH yEGFP</td>
<td>K. Bailey</td>
<td>yEGFP</td>
<td>BamHI-8xLexAop-Gal1-10pr region-yEGFP-Xhol-ADH1t-SphI</td>
<td>Yeast reporter plasmid (LexA DBD), YEp</td>
<td>Amp</td>
<td>URA3</td>
<td>5' Gal1, 3' ADH1</td>
</tr>
<tr>
<td>pSH yEGFPpST</td>
<td>K. Bailey</td>
<td>yEGFP-Cln2PEST</td>
<td>BamHI-8xLexAop-Gal1-10pr region-yEGFP-Cln2PEST-Sall-ADH1t-SphI</td>
<td>Yeast reporter plasmid (LexA DBD), YEp</td>
<td>Amp</td>
<td>URA3</td>
<td>5' Gal1, 3' ADH1</td>
</tr>
<tr>
<td>pSHIP yEGFPpST</td>
<td>K. Bailey</td>
<td>yEGFP-Cln2PEST</td>
<td>BamHI-8xLexAop-Gal1-10pr region-yEGFP-Cln2PEST-Sall-ADH1t-SphI</td>
<td>Yeast reporter plasmid (LexA DBD), Ylp</td>
<td>Amp</td>
<td>URA3</td>
<td>5' Gal1, 3' ADH1</td>
</tr>
</tbody>
</table>
### Section 3. Hormone Receptor Constructs

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Expressed Gene</th>
<th>Restriction Sites</th>
<th>Plasmid Type</th>
<th>Bacterial Resistance</th>
<th>Yeast Marker</th>
<th>Seq. Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAM423 LexA AR</td>
<td>S. Muddana</td>
<td>HA-SV40NLS-LexA-AR (670-919)</td>
<td>Gal1pr-HA-NLS-LexA-EcoRI-AR (670-919)-SalI-Stop-XhoI-ADH1t</td>
<td>Yeast Expression Vector, YEp</td>
<td>Amp</td>
<td>His 3</td>
<td>5' Gal1, 3' ADH1</td>
</tr>
<tr>
<td></td>
<td>S. Muddana</td>
<td>HA-SV40NLS-LexA-PR (682-933)</td>
<td>KpnI-Gal1pr-HA Tag-SV40-NLS-LexA-EcoRI-hER (305-595)-ADH1t-BamHI-XbaI-NotI-SphI-PstI</td>
<td>Yeast Expression Vector, YEp</td>
<td>Amp</td>
<td>His 3</td>
<td>5' Gal1, 3' ADH1</td>
</tr>
<tr>
<td>pJG4-5 AR</td>
<td>S Muddana</td>
<td>SV40 NLS-B42 AD-HA-AR (670-919)</td>
<td>SacI-KpnI-Gal1pr-HindIII-SV40NLS-B42 AD-HA-EcoRI-AR (670-919)-XhoI-ADH1t-SphI-BamHI-SalI</td>
<td>Yeast Expression Vector, YEp</td>
<td>Amp</td>
<td>Trp1</td>
<td>5' pJG45, 3' ADH</td>
</tr>
<tr>
<td>pJG4-5 hERβ</td>
<td>S Muddana</td>
<td>SV40 NLS-B42 AD-HA-ERβ (255-509)</td>
<td>SacI-KpnI-Gal1pr-HindIII-SV40NLS-B42 AD-HA-EcoRI-hER (305-595)-ADH1t-SphI-BamHI-SalI</td>
<td>Yeast Expression Vector, YEp</td>
<td>Amp</td>
<td>Trp1</td>
<td>5' pJG45, 3' ADH</td>
</tr>
<tr>
<td>pJG4-5 PR</td>
<td>S. Muddana</td>
<td>SV40 NLS-B42 AD-HA-PR (682-933)</td>
<td>SacI-KpnI-Gal1pr-HindIII-SV40NLS-B42 AD-HA-EcoRI-PR (682-933)-SalI-ADH1t-SphI-BamHI-SalI</td>
<td>Yeast Expression Vector, YEp</td>
<td>Amp</td>
<td>Trp1</td>
<td>5' pJG45, 3' ADH</td>
</tr>
</tbody>
</table>

### Section 4. Streptavidin Constructs

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Expressed Gene</th>
<th>Restriction Sites</th>
<th>Plasmid Type</th>
<th>Bacterial Resistance</th>
<th>Yeast Marker</th>
<th>Seq. Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJG4-5 dSA&lt;sub&gt;AA&lt;/sub&gt;</td>
<td>K. Bailey</td>
<td>SV40 NLS-B42 AD-HA -cpSA&lt;sub&gt;47&lt;/sub&gt;-cpSA&lt;sub&gt;54&lt;/sub&gt;</td>
<td>Gal1pr-HindIII-SV40 NLS-B42 AD-HA Tag-EcoRI-cpSA&lt;sub&gt;47&lt;/sub&gt;-BspEI-cpSA&lt;sub&gt;47&lt;/sub&gt;-XhoI-HindIII-ADH1t</td>
<td>Yeast Expression Vector, YEp</td>
<td>Amp</td>
<td>Trp 1</td>
<td>5' pJG4-5, 3' ADH1t</td>
</tr>
<tr>
<td>pJG4-5 cpSA&lt;sub&gt;Nb&lt;/sub&gt;&lt;sub&gt;54&lt;/sub&gt;</td>
<td>K. Bailey</td>
<td>SV40NLS-B42 AD-HA Tag-cpSA&lt;sub&gt;Nb&lt;/sub&gt;&lt;sub&gt;54&lt;/sub&gt;</td>
<td>Gal1pr-HindIII-SV40 NLS-B42 AD-HA Tag-EcoRI-cpSA&lt;sub&gt;Nb&lt;/sub&gt;&lt;sub&gt;54&lt;/sub&gt;-XhoI-HindIII-ADH1t</td>
<td>Yeast Expression Vector, YEp</td>
<td>Amp</td>
<td>Trp 1</td>
<td>5' pJG4-5, 3' ADH1t</td>
</tr>
<tr>
<td>Plasmid</td>
<td>Construct</td>
<td>Expression Vector, YEps</td>
<td>Amp</td>
<td>LEU2</td>
<td>ADH1t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------------</td>
<td>-------------------------</td>
<td>-----</td>
<td>------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pPA1 LexA</td>
<td>HA-SV40NSL-LexA-cpSA54</td>
<td>Yeast Expression Vector, YEps</td>
<td>Amp</td>
<td>LEU2</td>
<td>5' Gal1, 3' ADH1t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cpSA54</td>
<td>K. Bailey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pPA1 LexA</td>
<td>HA-SV40NSL-LexA-cpSA97</td>
<td>Yeast Expression Vector, YEps</td>
<td>Amp</td>
<td>LEU2</td>
<td>5' Gal1, 3' ADH1t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cpSA97</td>
<td>K. Bailey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pPA1 LexA</td>
<td>HA-SV40NSL-LexA-cpSAnb54</td>
<td>Yeast Expression Vector, YEps</td>
<td>Amp</td>
<td>LEU2</td>
<td>5' Gal1, 3' ADH1t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cpSAnb54</td>
<td>K. Bailey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pPA1 LexA</td>
<td>HA-SV40NSL-LexA-cpSAnb57</td>
<td>Yeast Expression Vector, YEps</td>
<td>Amp</td>
<td>LEU2</td>
<td>5' Gal1, 3' ADH1t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cpSAnb57</td>
<td>K. Bailey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pPA1 LexA</td>
<td>HA-SV40NSL-LexA-cpSAnb54</td>
<td>Yeast Expression Vector, YEps</td>
<td>Amp</td>
<td>LEU2</td>
<td>5' Gal1, 3' ADH1t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dSA97</td>
<td>K. Bailey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pPA1 LexA</td>
<td>HA-SV40NSL-LexA-cpSAnb54</td>
<td>Yeast Expression Vector, YEps</td>
<td>Amp</td>
<td>LEU2</td>
<td>5' Gal1, 3' ADH1t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dSA54</td>
<td>K. Bailey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pPA1 LexA</td>
<td>HA-SV40NSL-LexA-cpSAnb54</td>
<td>Yeast Expression Vector, YEps</td>
<td>Amp</td>
<td>LEU2</td>
<td>5' Gal1, 3' ADH1t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dSA54</td>
<td>K. Bailey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pPA1 LexA</td>
<td>HA-SV40NSL-LexA-cpSAnb54</td>
<td>Yeast Expression Vector, YEps</td>
<td>Amp</td>
<td>LEU2</td>
<td>5' Gal1, 3' ADH1t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dSA54</td>
<td>K. Bailey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pPA1 LexA</td>
<td>HA-SV40NSL-LexA-cpSAnb54</td>
<td>Yeast Expression Vector, YEps</td>
<td>Amp</td>
<td>LEU2</td>
<td>5' Gal1, 3' ADH1t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dSA54</td>
<td>K. Bailey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**APPENDIX B. Protein Sequences**

*yEGFP*: Yeast enhanced variant of EGFP used as fluorescent reporter gene

Cormack, B. P. *e. Microbiology* 1997, 143 (Pt 2), 303.

<table>
<thead>
<tr>
<th>pSA1</th>
<th>T7SAwt</th>
<th>S Hussey</th>
<th>T7SAwt-NLS-B42</th>
<th>AD-HA tag-stop</th>
<th>KpnI-GalIpr-T7SAwt-NLS-B42-HA tag-EcoRI-XhoI-stop-ADH1t</th>
<th>Yeast Expression Vector, YEp</th>
<th>Amp</th>
<th>TRPI</th>
<th>5' Gal1, 3' ADH1t</th>
</tr>
</thead>
</table>

```
APPENDIX B. Protein Sequences

yEGFP: Yeast enhanced variant of EGFP used as fluorescent reporter gene
Cormack, B. P. e. Microbiology 1997, 143 (Pt 2), 303.

| 1 | atg | tac | amm | ggt | gas | gas | tta | ttc | act | gyt | 361 | aat | aga | atc | gas | tta | amm | ggt | att | gat | ttt |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 2 | M   | K   | N   | K   | G   | E   | E | L   | E | F | F | T | G | 121 | N | S | I | E | L | K | G | T | D | F |
| 31 | gtt | gtc | cca | att | tgg | gtt | gas | tta | gat | gyt | 391 | aam | gas | gat | gtt | amm | act | tta | gtt | ccc | amm |
| 31 | V   | V   | P   | I   | L   | V   | E   | L   | D   | G | 131 | K   | E   | D   | G   | N   | I   | L   | G   | H   | K   |
| 31 | gtt | gtt | aat | gtt | ccc | ccc | tcc | tcc | tgg | ccc | 421 | tgg | gaa | tcc | aac | tta | aac | tcc | ccc | aat | ggt |
| 31 | D   | V   | N   | G   | K   | F   | S   | V   | S   | V | 141 | L   | E   | Y   | N   | Y   | N   | S   | H   | N   | V   |
| 31 | gtt | gta | gtt | gaa | ggt | gat | gat | act | taa | tcc | 451 | tgg | gaa | tcc | aac | gaa | ccc | aag | aat | ggt |
| 31 | G   | G   | G   | E   | G   | E   | G   | D   | A   | T   | Y   | G   | 151 | Y   | I   | M   | A   | D   | Q   | E   | N   | G   |
| 121 | amm | tgg | acc | tta | amm | ttt | amm | att | tgg | act | aat | 451 | aac | amm | gtt | amm | tcc | amm | att | amm | aag | ccc | amm |
| 121 | K   | L   | L   | K   | P   | I   | C   | T   | T | 161 | I   | K   | V   | N   | F   | K   | I   | R   | P   | H   | N   |
| 181 | gtt | amm | ttc | cca | gtt | cca | tgg | cca | gtt | cca | 511 | aat | gas | gat | gtt | ttt | gtt | cca | tta | gtt | gas |
| 181 | G   | K   | L   | F   | Y   | P   | U   | P   | F   | L   | 271 | I   | E   | D   | G   | S   | V   | Q   | L   | A   | B   |
| 181 | gtt | gtt | act | ttc | gtt | ttt | gtt | ttt | gtt | ccc | 541 | ctt | ttt | cca | con | ctt | ccc | ctt | ccc | att | gtt | gtt |
| 181 | V   | T   | T   | F   | G   | Y   | G   | V   | Q   | C   | 101 | H   | Y   | Q   | C   | N   | T   | P   | I   | G   | D   |
| 211 | ctt | ggt | gta | cca | gtt | ccc | gat | cta | amm | ccc | 571 | ggt | cca | gtt | ccc | gtt | ccc | gat | cct | cca | gaa | ccc |
| 211 | F   | A   | R   | Y   | P   | D   | B   | A   | K   | Q   | 191 | G   | P   | U   | L   | F   | D   | H   | N   | H   | N   |
| 241 | ctt | ggt | cct | cca | gaa | gaa | act | att | tcc | tcc | 601 | ccc | amm | gaa | amm | gaa | ccc | gaa | gaa | ccc | gaa | ccc |
| 241 | H   | D   | F   | F   | K   | S   | A   | K   | P   | E   | 201 | L   | S   | T   | S   | A   | L   | S   | A   | D   | S   |
| 271 | gtt | ttt | gtt | cca | gaa | gaa | act | ttt | ttt | tcc | 631 | cca | amm | gaa | amm | gaa | amm | gaa | ccc | gaa | ccc | gaa |
| 271 | G   | Y   | V   | Q   | E   | R   | T   | I   | F   | P   | 211 | P   | N   | E   | K   | R   | D   | H   | M   | V   | L   |
| 301 | amm | gaa | ggt | amm | taa | amm | gaa | amm | gaa | gaa | 661 | taa | gaa | ttt | gtt | gtt | gtt | gtt | gtt | att | amm | aag |
| 101 | E   | D   | D   | G   | N   | Y   | E   | T   | R   | A   | 221 | L   | E   | F   | V   | T   | A   | G   | T   | E   | L   |
| 331 | gaa | gtt | amm | taa | gaa | ggt | amm | taa | gtt | gtt | 691 | ctt | gtt | agt | gas | ccc | tta | ccc | ccc | ccc | att | amm | ccc |
| 331 | E   | V   | X   | F   | E   | G   | D   | T   | L   | V   | 231 | H   | G   | M   | D   | E   | L   | Y   | E   | -   |
SV40NLS-B42-HA: Used for B42 AD fusion proteins

<table>
<thead>
<tr>
<th>1</th>
<th>agc gct gcc ctc tac ctc tac tac ggt ccc gat</th>
<th>36</th>
<th>gag ggt ctc tac ctc tac tac ggt ccc gat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M S Y P Y D V D P L</td>
<td>36</td>
<td>E G H T H Q V D P S L</td>
</tr>
<tr>
<td>31</td>
<td>gaa gaa gaa gaa gaa ccc tgg gaa ccc tgg gaa</td>
<td>36</td>
<td>gag gaa gaa gaa gaa ccc tgg gaa ccc tgg gaa</td>
</tr>
<tr>
<td>31</td>
<td>Y A S P E L A P K</td>
<td>36</td>
<td>F K P N A D F L L R</td>
</tr>
<tr>
<td>61</td>
<td>gag aag aag aag gaa gaa ccc gag aag gag ccc</td>
<td>42</td>
<td>gtc aag aag aag gaa gaa ccc gag aag gag ccc</td>
</tr>
<tr>
<td>61</td>
<td>Y A S P E L A P K</td>
<td>42</td>
<td>F K P N A D F L L R</td>
</tr>
<tr>
<td>91</td>
<td>gag aag aag aag gaa gaa ccc gag aag gag ccc</td>
<td>48</td>
<td>att aag gaa gaa gaa ccc gag aag gag ccc</td>
</tr>
<tr>
<td>91</td>
<td>T A R Q Q E V F D L</td>
<td>48</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>121</td>
<td>aag gaa gaa gaa gaa gaa ccc gag aag gag ccc</td>
<td>51</td>
<td>aag aag gaa gaa gaa gaa ccc gag aag gag ccc</td>
</tr>
<tr>
<td>121</td>
<td>T A R Q Q E V F D L</td>
<td>51</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>151</td>
<td>aag aag gaa gaa gaa gaa ccc gag aag gag ccc</td>
<td>59</td>
<td>aag aag gaa gaa gaa gaa ccc gag aag gag ccc</td>
</tr>
<tr>
<td>151</td>
<td>T A R Q Q E V F D L</td>
<td>59</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>181</td>
<td>gaa gaa gaa gaa gaa gaa ccc gag aag gag ccc</td>
<td>62</td>
<td>gaa gaa gaa gaa gaa gaa ccc gag aag gag ccc</td>
</tr>
<tr>
<td>181</td>
<td>T A R Q Q E V F D L</td>
<td>62</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>211</td>
<td>gaa gaa gaa gaa gaa gaa ccc gag aag gag ccc</td>
<td>67</td>
<td>gaa gaa gaa gaa gaa gaa ccc gag aag gag ccc</td>
</tr>
<tr>
<td>211</td>
<td>T A R Q Q E V F D L</td>
<td>67</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>241</td>
<td>aag aag gaa gaa gaa gaa ccc gag aag gag ccc</td>
<td>71</td>
<td>aag aag gaa gaa gaa gaa ccc gag aag gag ccc</td>
</tr>
<tr>
<td>241</td>
<td>T A R Q Q E V F D L</td>
<td>71</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>271</td>
<td>aag aag gaa gaa gaa gaa ccc gag aag gag ccc</td>
<td>76</td>
<td>aag aag gaa gaa gaa gaa ccc gag aag gag ccc</td>
</tr>
<tr>
<td>271</td>
<td>T A R Q Q E V F D L</td>
<td>76</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>301</td>
<td>aag aag gaa gaa gaa gaa ccc gag aag gag ccc</td>
<td>81</td>
<td>aag aag gaa gaa gaa gaa ccc gag aag gag ccc</td>
</tr>
<tr>
<td>301</td>
<td>T A R Q Q E V F D L</td>
<td>81</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>331</td>
<td>aag aag gaa gaa gaa gaa ccc gag aag gag ccc</td>
<td>86</td>
<td>aag aag gaa gaa gaa gaa ccc gag aag gag ccc</td>
</tr>
<tr>
<td>331</td>
<td>T A R Q Q E V F D L</td>
<td>86</td>
<td>I M D G D L L A V H</td>
</tr>
</tbody>
</table>

HA-SV40NLS-LexA: Used for LexA DBD fusion proteins

<table>
<thead>
<tr>
<th>1</th>
<th>atg gct ctc tac ctc tac tac ggt ccc gat</th>
<th>36</th>
<th>gag ggt ctc tac ctc tac tac ggt ccc gat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M S Y P Y D V D P L</td>
<td>36</td>
<td>E G H T H Q V D P S L</td>
</tr>
<tr>
<td>31</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
<td>36</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
</tr>
<tr>
<td>31</td>
<td>Y A S P E L A P K</td>
<td>36</td>
<td>F K P N A D F L L R</td>
</tr>
<tr>
<td>61</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
<td>42</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
</tr>
<tr>
<td>61</td>
<td>Y A S P E L A P K</td>
<td>42</td>
<td>F K P N A D F L L R</td>
</tr>
<tr>
<td>91</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
<td>48</td>
<td>att gag ggt ctc tac ctc tac tac ggt ccc gat</td>
</tr>
<tr>
<td>91</td>
<td>T A R Q Q E V F D L</td>
<td>48</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>121</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
<td>51</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
</tr>
<tr>
<td>121</td>
<td>T A R Q Q E V F D L</td>
<td>51</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>151</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
<td>59</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
</tr>
<tr>
<td>151</td>
<td>T A R Q Q E V F D L</td>
<td>59</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>181</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
<td>62</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
</tr>
<tr>
<td>181</td>
<td>T A R Q Q E V F D L</td>
<td>62</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>211</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
<td>67</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
</tr>
<tr>
<td>211</td>
<td>T A R Q Q E V F D L</td>
<td>67</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>241</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
<td>71</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
</tr>
<tr>
<td>241</td>
<td>T A R Q Q E V F D L</td>
<td>71</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>271</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
<td>76</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
</tr>
<tr>
<td>271</td>
<td>T A R Q Q E V F D L</td>
<td>76</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>301</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
<td>81</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
</tr>
<tr>
<td>301</td>
<td>T A R Q Q E V F D L</td>
<td>81</td>
<td>I M D G D L L A V H</td>
</tr>
<tr>
<td>331</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
<td>86</td>
<td>gag ggt ctc tac ctc tac tac ggt ccc gat</td>
</tr>
<tr>
<td>331</td>
<td>T A R Q Q E V F D L</td>
<td>86</td>
<td>I M D G D L L A V H</td>
</tr>
</tbody>
</table>

72

72

72

72
Cln(PEST): Protein destabilization tag used to reduce yEGFP half-life.


SA(13-139): Core streptavidin(13-139) used as template and benchmark for all SA variants.

pSHyEGFPpst: Full sequence of improved plasmid used for streptavidin three-hybrid assays.

```
1  ggatcccg cttggaattc gcagagttat agcaaacac acagccat cccatctnca caaaccttcct ccaagcttgt
   >>............................................LexAop-Gal1pr...................................
61  gcacccagt gatccacc accgccaa accggccccaccacatcatt gcagagttat tccactttgt gcagagctga tccactttat
   >............................................LexAop-Gal1pr...................................
151  tgaattctc aaaaactttcttctctct ttgatgacg aagaaagtta aaaaactttcttctctct aaaaactttcttctctct
   >............................................LexAop-Gal1pr...................................
241  atcatcatc tttctcttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttc
SAnb: Core streptavidin with negligible biotin affinity due to N23A, S27D, S45A mutations (underlined). Silent PstI restriction site (blue) added for screening.
**cpSA<sub>87</sub>:** Circular permutation of core streptavidin beginning with E116 (between 7<sup>th</sup> and 8<sup>th</sup> β-strands). BamHI restriction site (blue) embedded in GGGS linker (blue) for screening. Mutations N23A (gct), D27A (gat), and S45A are made for cpSA<sub>n87</sub> (underlined, silent PstI made but not shown). C-terminal SG added with embedded BspEI (blue) for dSA cloning.

**cpSA<sub>54</sub>:** Circular permutation of core streptavidin beginning with S69 (between 4<sup>th</sup> and 5<sup>th</sup> β-strands). BamHI restriction site (blue) embedded in GGGS linker (blue) for screening. Mutations N23A (gct), D27A (gat), and S45A are made for cpSA<sub>n54</sub> (underlined, silent PstI made but not shown). N-terminal SG codons mutated to embed silent BspEI (blue) for dSA cloning.
ERβ(255-509): Human estrogen receptor beta ligand binding domain (255-509) used in ERβ homodimerization yeast three-hybrid assays.

PR(682-933): Human progesterone receptor ligand binding domain (682-933) used in PR homodimerization yeast three-hybrid assays. The unintended P906S mutation is shown in blue.
AR(670-919): Human androgen receptor ligand binding domain (670-919) used in AR homodimerization yeast three-hybrid assays.

### APPENDIX C. Microbial Stains

<table>
<thead>
<tr>
<th>Name</th>
<th>Cell Type</th>
<th>Source</th>
<th>Genotype</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH5α</td>
<td>E. coli</td>
<td>Invitrogen</td>
<td>F- F50lacZDM15 D(lacZYA-argF) U169 recA1 endA1 hisd17(k-m) phoA supE44 thi-1 gyrA96 relA1</td>
<td>General subcloning</td>
</tr>
<tr>
<td>FY250</td>
<td>S. cerevisiae</td>
<td>M. Ptashne</td>
<td>MATa, ura3-52, hisD200, leu2D1, trp1D63</td>
<td>General yeast strain</td>
</tr>
<tr>
<td>FY250yEGFPpst</td>
<td>S. cerevisiae</td>
<td>K. Bailey</td>
<td>MATa, ura3-52, URA3::yEGFPpst, hisD200, leu2D1, trp1D63</td>
<td>yEGFPpst integrated reporter strain</td>
</tr>
</tbody>
</table>