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

LIST OF FIGURES

CHAPTER 1

LIST OF TABLES

CHAPTER 1

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 "druggablilty"[1] 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.[2] 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.[3, 4] 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.[5] 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.[5, 6] 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.[7, 8] 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 ($K_d \approx 10^{-14}$ 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.

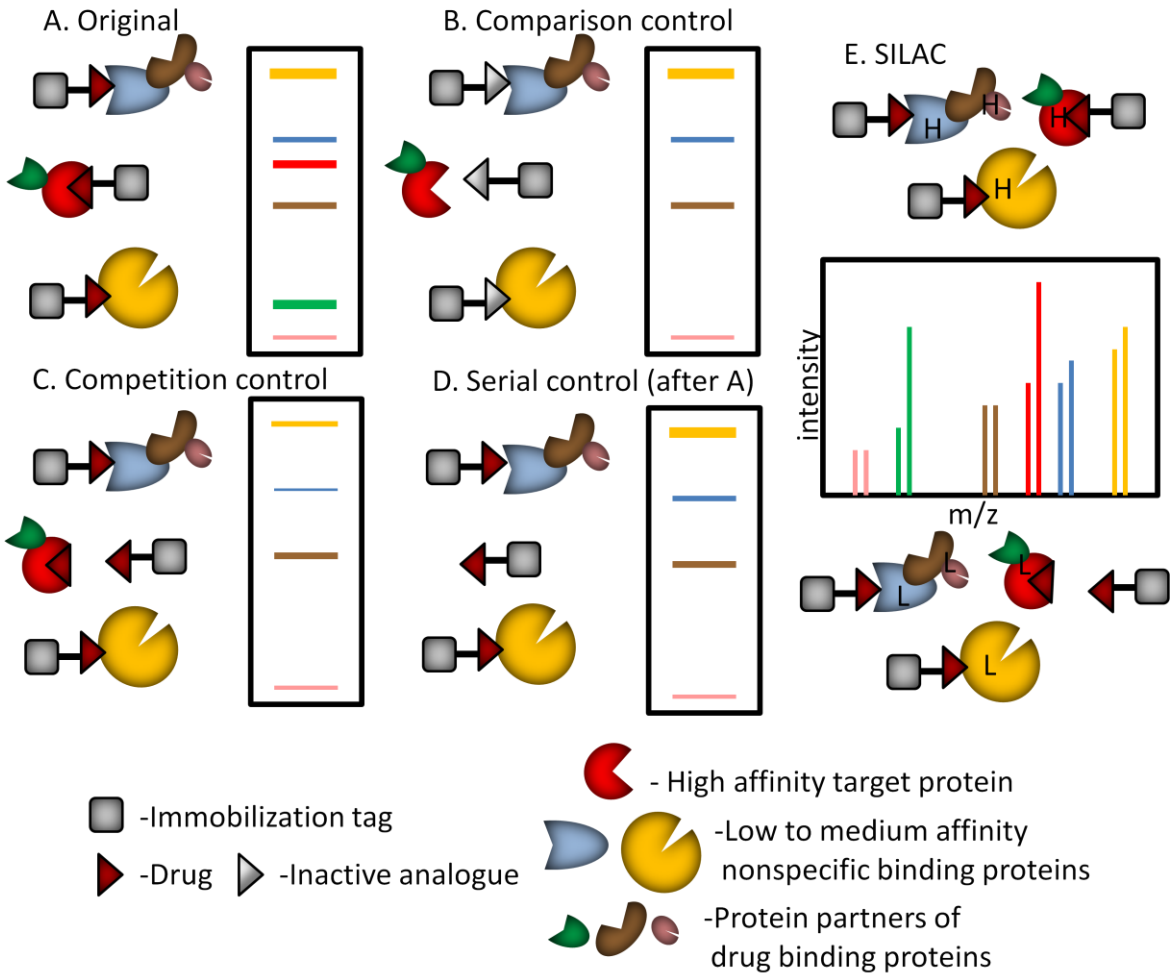


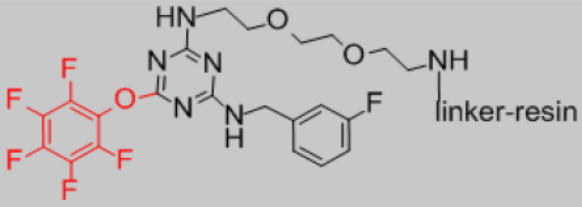
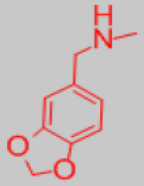
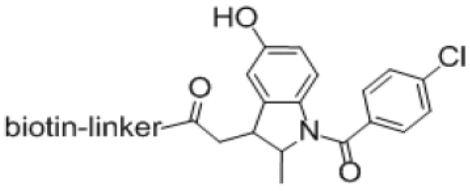
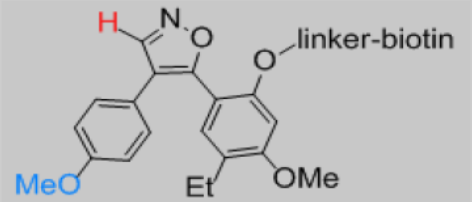
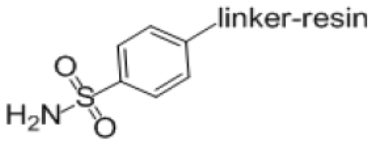
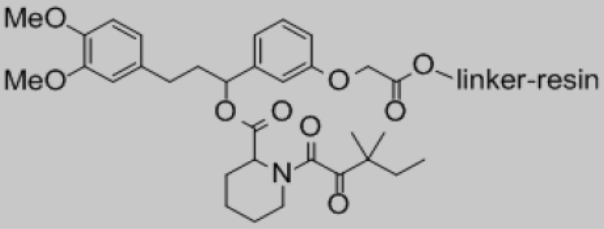
Figure 1.1. Schematic representation of affinity chromatography outcomes in which the drug is tethered to a solid support resin, usually directly or through a biotin tag (which binds to a streptavidin fused matrix). Proteins still attached to the drug after washing are eluted and investigated by SDS-PAGE (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.[9] 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.

Drug (Target, ref.)	Affinity Probe	Control
Melanogenin (Prohibitin, Snyder, J. R. *et al*. J. *Chem. Biol.* **2005**, *12*, 477-484.)		Comparison: 
Indomethacin (Glyoxalase 1, Sato, S. et *al*. *J. Am. Chem. Soc.* **2007**, *129*, 873-880.)	 Pre-bound to Neutravidin-Avagorse resin	Competition
KRIBB3 (Hsp27, Shin, K. D. *et al*. *J. Biol. Chem.* **2005**, *280*, 41439-41448.)	 Free in cell lysate	Comparision: OMe, H
Benzene-sulfonamide (Carbonic Anhydrase II, Yamamoto, K. *et al*. *Anal. Biochem.* **2006**, *352*, 15-23.)		Serial
AP1497 (FKBP1A, Ong, S. E. *et al*. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4617-4622.)		SILAC

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 then 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.[10] 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.[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.[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.

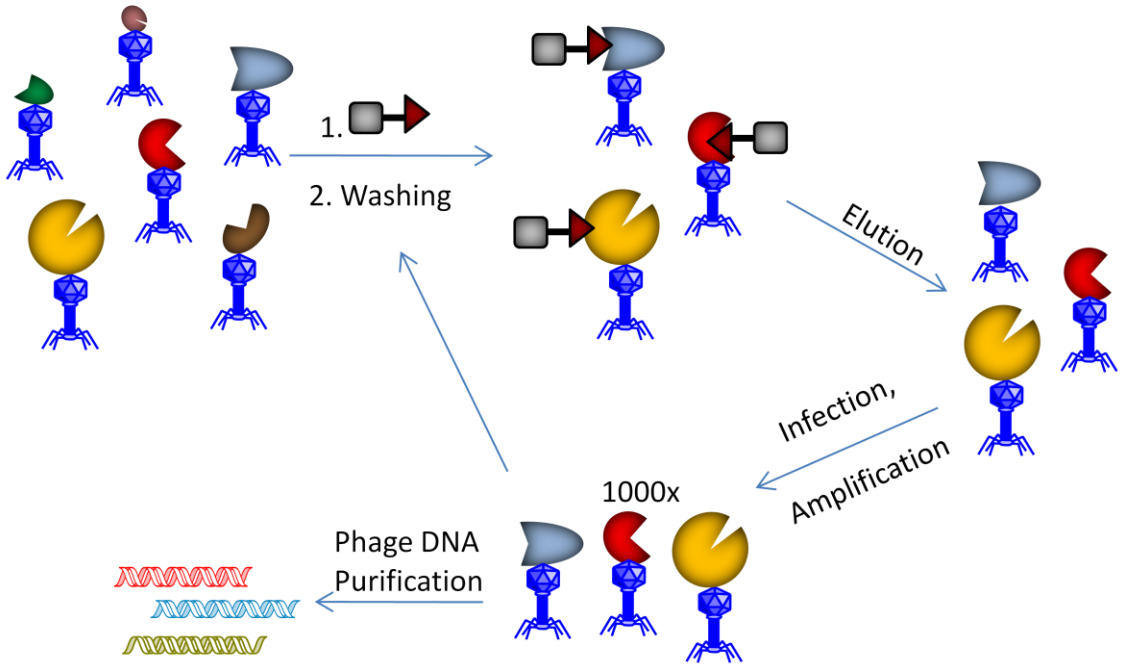


Figure 1.2. Schematic representation of a phage display assay of a cloned protein library. Proteins are fused to a phage coat protein and assayed with affinity chromatography. Eluted proteins can be amplified by infection of bacteria. Proteins can be identified with SDS-PAGE/MS or through sequencing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

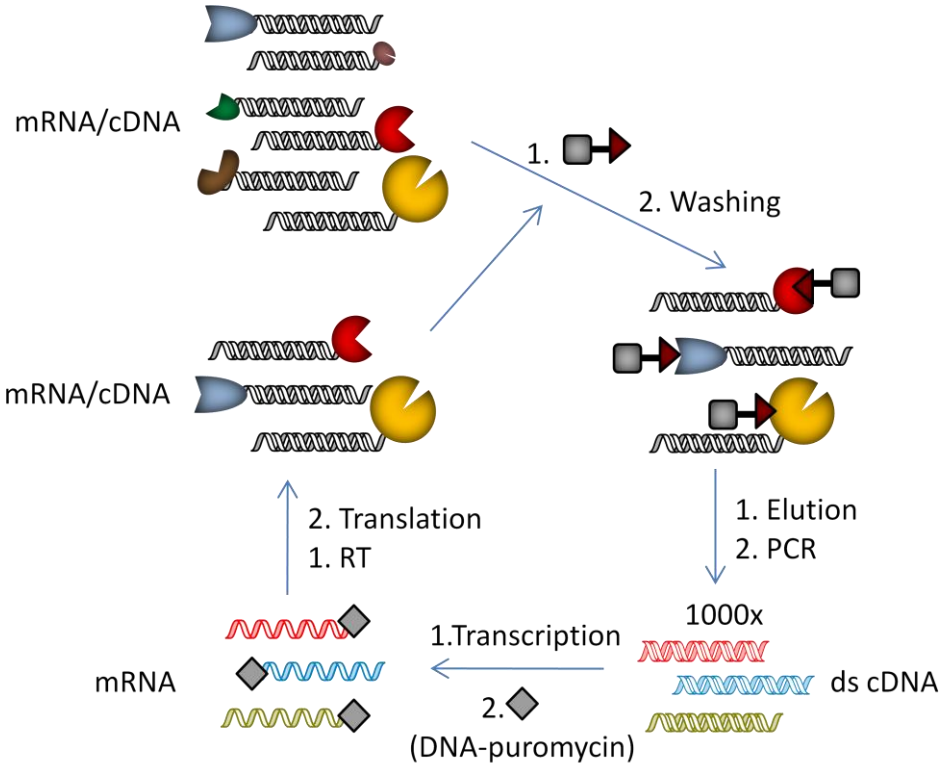


Figure 1.3. Schematic representation of an mRNA display assay, in which cDNA/mRNA-protein constructs are screened with affinity chromatography. Eluted constructs may be identified by sequencing of cDNA or amplified by PCR and transcription, creating an enriched mRNA pool for repeating the procedure.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

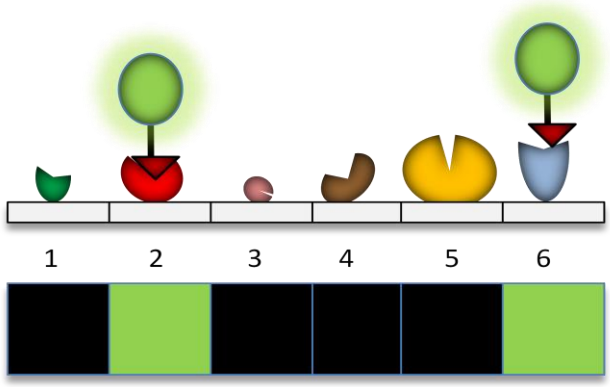


Figure 1.4. Schematic representing a protein microarray outcome in which screened protein library is spotted individually on a derivitized surface. After binding, unbound drug is washed from the array. The drug is labeled to be able to give a localized signal for identifying its binding partner(s).

A variation on the microarray platform indirectly spots the protein library by expressing individual proteins in localized, adherent mammalian cell populations.[20] 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.[21,22] Drug Affinity Response Target Stability (DARTS)[23] aims to take advantage of the resistance to proteolysis[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,[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,[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.

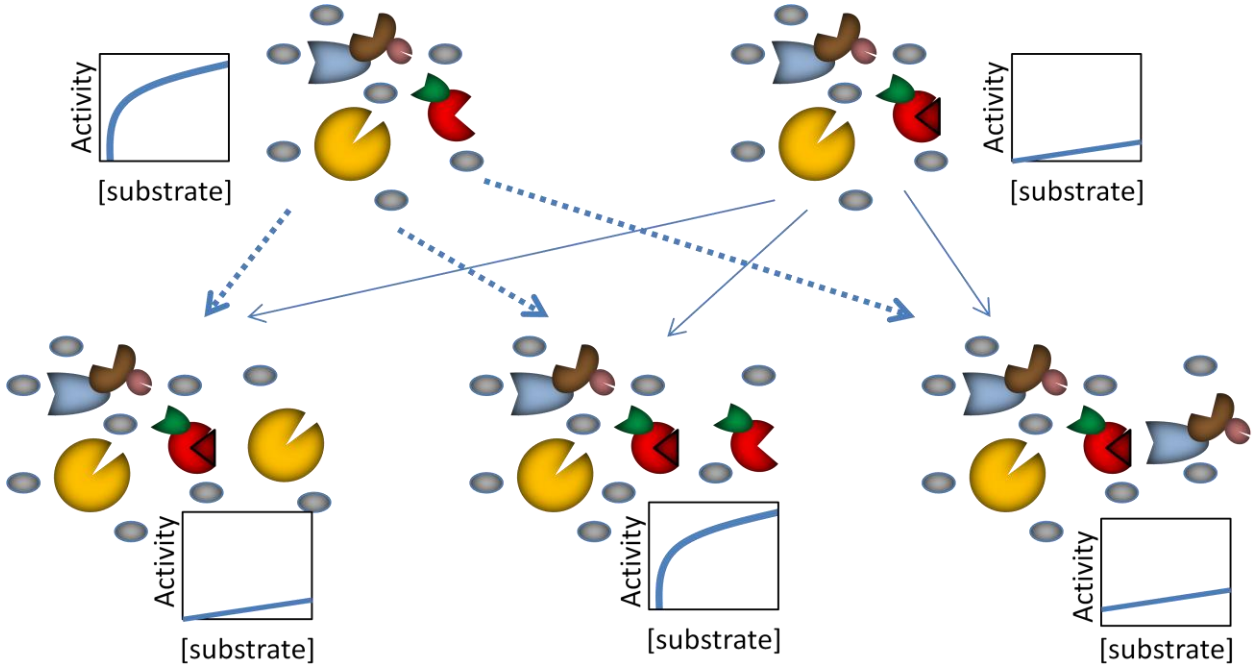


Figure 1.5. Schematic depicting a biochemical suppression strategy that uses an in vitro activity assay. For simplicity, only unbound target produces activity. Aliquots of the drug inhibited sample (upper right) are 'treated' with fractions of the untreated sample (upper left). Only fractions with drug 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.[27] Yeast have proven themselves invaluable tools for understanding mammalian cellular processes and diseases,[28] 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 then 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,[29] many of

which have known human homologs. Even before this, yeast have been used to discover the molecular targets of drugs such as rapamycin,[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.[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,[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).[34] Inherently this method can only apply to drugs able to impact yeast growth.

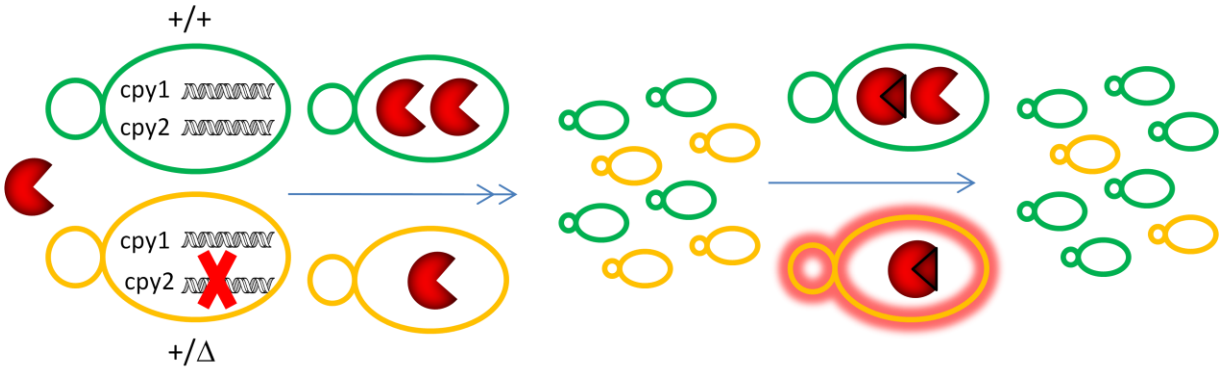


Figure 1.6. Schematic representation of a drug-induced haploinsufficiency competition assay of an essential target gene. In this example, diploid yeast with two copies of the target gene (green) produces twice as much target protein as the strain possessing only one gene (orange). Untreated yeast strains 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.[35] 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.[36] 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.[37] 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.

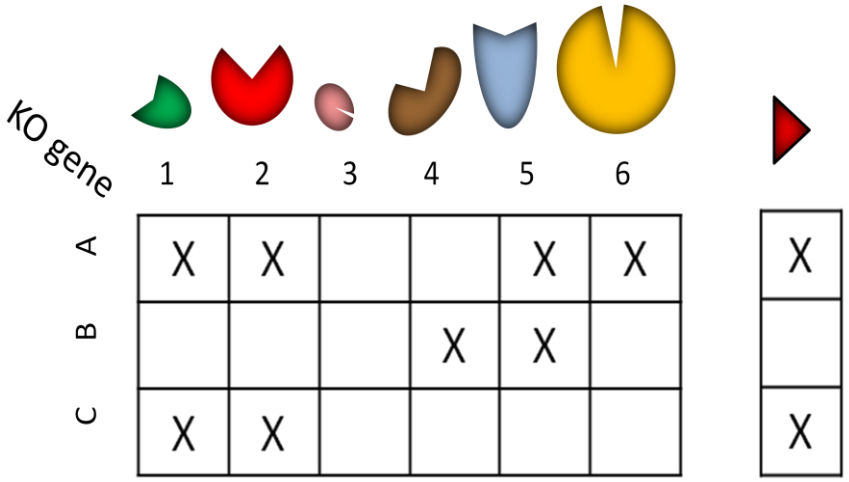


Figure 1.7. Schematic depicting a small synthetic lethality profile of 6 possible target proteins (1-6) and drug treatment (right), based on three random nonessential genes (A, B, C). Each square in the main grid catalogues the viability of the DKO strain of intersecting genes (*e.g.* A and 1), and on 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).

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,[38] 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 *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 *GAL1* and *GAL10*.[39] Therefore, cloning the promoters for either *GAL1* or *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 *E. coli*.[40] 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 *GAL4* domains.[41]

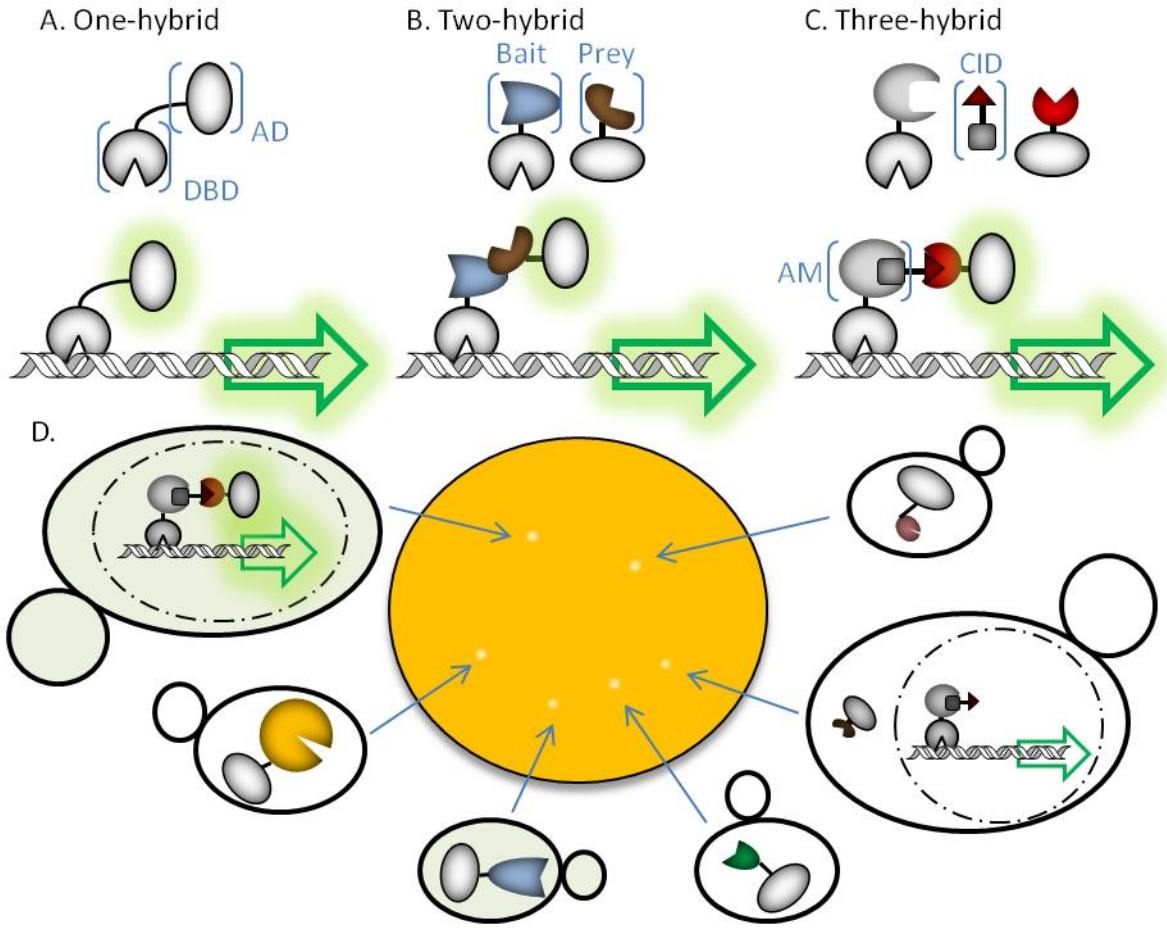
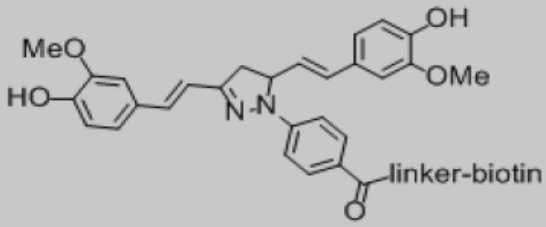
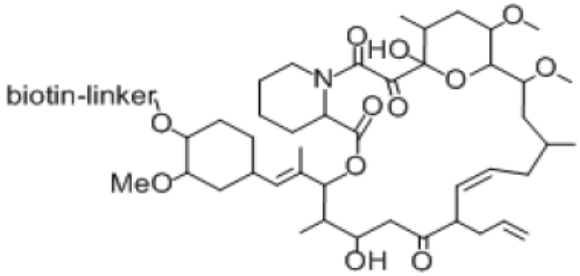
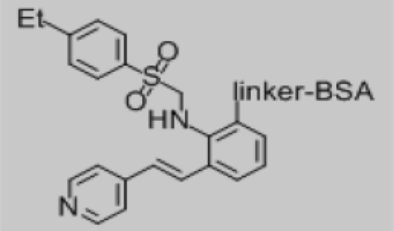
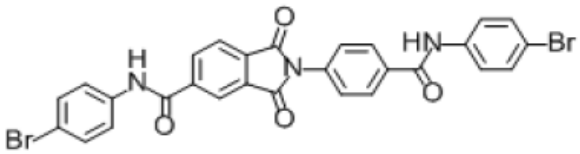
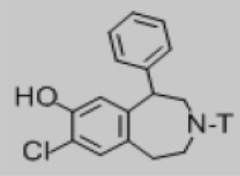


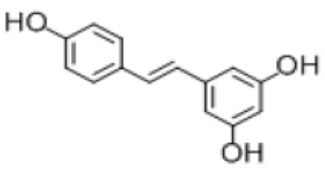
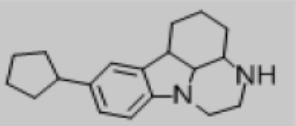
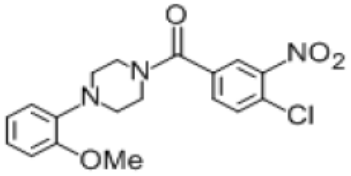
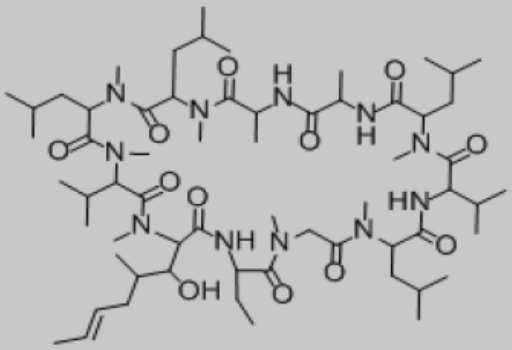
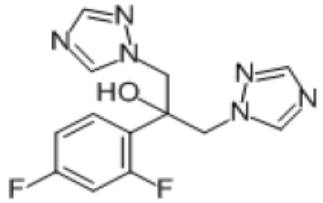
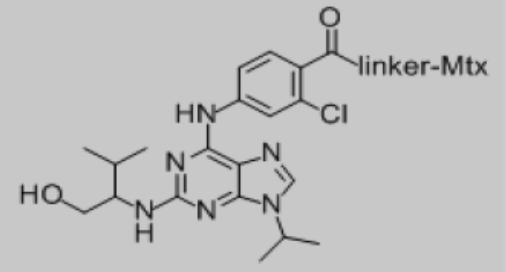
Figure 1.8. Schematic depiction of yeast one- (A), two- (B), and three- (C) hybrid systems. Reporter expression is activiated by localizing the activation domain (AD) to the promoter of the reporter gene (green arrow). The DNA binding domain (DBD) localizes the system to the promoter, which is directly 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 (tethored to the anchor moeity, AM) to activate reporter expression.

A derivative of this assay used to study protein-small molecule interactions, termed the yeast three-hybrid system,[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),[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.[44]

Table 1.2. Selected examples of miscellaneous target discovery strategies.

Method	Drug (Target, ref)	Probe
Phage display	HBC (Ca2+-calmodulin, Shim, J. S. *et al*. *Chem. Biol.* **2004,** *11*, 1455-1463.)	
mRNA display	FK506 (FKBP12, McPherson, M. *et al*. *Chem. Biol.* **2002**, *9*, 691-698.)	
Drug western	HMN-154 (NF-YB, thymosin β-10, Tanaka, H. *et al*. *Mol. Pharmacol.* **1999**, 55, 356-363.)	
Protein Microarray	SMIR4 (Nir1p, Huang, J. *et al*. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101*, 16594-16599.)	
Transfected Cell Microarrays	SCH23390 (Dopamine receptor, Ziauddin, J.; Sabatini, D. M. *Nature* **2001,** *411*, 107-110.)	

DARTS	Resveratrol (Tif1, Lomenick, B. *et al*. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106*, 21984-21989.)	
Biochemical suppression	Pirl1 (CDC42-RhoGDI, Giaever, G. *et al*. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 793-798.)	
HIP	4130-1278 (SEC14, Hoon, S. *et al*. *Nat. Chem. Biol.* **2008**, *4*, 498-506.)	
mRNA profiling	Cyclosporin A (CPH1, Hughes, T. R. *et al*. *Cell* **2000**, *102*, 109-126.)	
Chemogenetic Profiling	Fluconazole (ERG11, Parsons, A. B. *et al*. *Nat. Biotechnol.* **2004**, *22*, 62-69.)	
Yeast three-hybrid	Puvalanol B (CDK1/5/6,FYN,PCTK1/2..., Becker, F. *et al*. *Chem. Biol.* **2004**, *11*, 211-223.)	

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

1. Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **2002,** 1, 727-730.
2. Sams-Dodd, F. Target-based drug discovery: is something wrong? *Drug Discov. Today* **2005,** 10, 139-147.
3. Norton, J. T.; Titus, S. A.; Dexter, D.; Austin, C. P.; Zheng, W.; Huang, S. Automated high-content screening for compounds that disassemble the perinucleolar compartment. *J. Biomol. Screen* **2009,** 14, 1045-1053.

4. Wagner, B. K.; Clemons, P. A. Connecting synthetic chemistry decisions to cell and genome biology using small-molecule phenotypic profiling. *Curr. Opin. Chem. Biol.* **2009,** 13, 539-548.

5. Hart, C. P. Finding the target after screening the phenotype. *Drug Discov. Today* **2005,** 10, 513-519.

6. Terstappen, G. C.; Schlupen, C.; Raggiaschi, R.; Gaviraghi, G. Target deconvolution strategies in drug discovery. *Nat. Rev. Drug Discov.* **2007,** 6, 891-903.

7. Sato, S.; Murata, A.; Shirakawa, T.; Uesugi, M. Biochemical target isolation for novices: affinity-based strategies. *Chem. Biol.* **2010,** 17, 616-623.

8. Leslie, B. J.; Hergenrother, P. J. Identification of the cellular targets of bioactive small organic molecules using affinity reagents. *Chem. Soc. Rev.* **2008,** 37, 1347-1360.

9. Ong, S. E.; Schenone, M.; Margolin, A. A.; Li, X.; Do, K.; Doud, M. K.; Mani, D. R.; Kuai, L.; Wang, X.; Wood, J. L.; Tolliday, N. J.; Koehler, A. N.; Marcaurelle, L. A.; Golub, T. R.; Gould, R. J.; Schreiber, S. L.; Carr, S. A. Identifying the proteins to which small-molecule probes and drugs bind in cells. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** 106, 4617-4622.

10. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **2006,** 5, 993-996.

11. Kwok, B. H.; Koh, B.; Ndubuisi, M. I.; Elofsson, M.; Crews, C. M. The anti-inflammatory natural product parthenolide from the medicinal herb Feverfew directly binds to and inhibits IkappaB kinase. *Chem. Biol.* **2001,** 8, 759-766.

12. Wulff, J. E.; Siegrist, R.; Myers, A. G. The natural product avrainvillamide binds to the oncoprotein nucleophosmin. *J. Am. Chem. Soc.* **2007,** 129, 14444-14451.

13. Savinov, S. N.; Austin, D. J. The cloning of human genes using cDNA phage display and small-molecule chemical probes. *Comb. Chem. High Throughput Screen.* **2001,** 4, 593-597.

14. Sche, P. P.; McKenzie, K. M.; White, J. D.; Austin, D. J. Display cloning: functional identification of natural product receptors using cDNA-phage display. *Chem. Biol.* **1999,** 6, 707-716.

15. McPherson, M.; Yang, Y.; Hammond, P. W.; Kreider, B. L. Drug receptor identification from multiple tissues using cellular-derived mRNA display libraries. *Chem. Biol.* **2002,** 9, 691-698.

16. Mattheakis, L. C.; Bhatt, R. R.; Dower, W. J. An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** 91, 9022-9026.

17. Tanaka, H.; Ohshima, N.; Hidaka, H. Isolation of cDNAs encoding cellular drug-binding proteins using a novel expression cloning procedure: drug-western. *Mol. Pharmacol.* **1999,** 55, 356-363.

18. Zhu, H.; Snyder, M. Protein chip technology. *Curr. Opin. Chem. Biol.* **2003,** 7, 55-63.

19. Jona, G.; Snyder, M. Recent developments in analytical and functional protein microarrays. *Curr. Opin. Mol. Ther.* **2003,** 5, 271-277.

20. Ziauddin, J.; Sabatini, D. M. Microarrays of cells expressing defined cDNAs. *Nature* **2001,** 411, 107-110.

21. Makowski, L.; Rodi, D. J.; Mandava, S.; Minh, D. D.; Gore, D. B.; Fischetti, R. F. Molecular crowding inhibits intramolecular breathing motions in proteins. *J. Mol. Biol.* **2008,** 375, 529-546.

22. Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007,** 450, 964-972.

23. Lomenick, B.; Olsen, R. W.; Huang, J. Identification of direct protein targets of small molecules. *ACS Chem. Biol.* **2011,** 6, 34-46.

24. Markus, G. Protein substrate conformation and proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **1965,** 54, 253-258.

25. Lomenick, B.; Hao, R.; Jonai, N.; Chin, R. M.; Aghajan, M.; Warburton, S.; Wang, J.; Wu, R. P.; Gomez, F.; Loo, J. A.; Wohlschlegel, J. A.; Vondriska, T. M.; Pelletier, J.; Herschman, H. R.; Clardy, J.; Clarke, C. F.; Huang, J. Target identification using drug affinity responsive target stability (DARTS). *Proc. Natl. Acad. Sci. U. S. A.* **2009,** 106, 21984-21989.

26. Peterson, J. R.; Lebensohn, A. M.; Pelish, H. E.; Kirschner, M. W. Biochemical suppression of small-molecule inhibitors: a strategy to identify inhibitor targets and signaling pathway components. *Chem. Biol.* **2006,** 13, 443-452.

27. Armour, C. D.; Lum, P. Y. From drug to protein: using yeast genetics for high-throughput target discovery. *Curr. Opin. Chem. Biol.* **2005,** 9, 20-24.

28. Hartwell, L. H. Yeast and cancer. *Biosci. Rep.* **2004,** 24, 523-544.

29. Winzeler, E. A.; Shoemaker, D. D.; Astromoff, A.; Liang, H.; Anderson, K.; Andre, B.; Bangham, R.; Benito, R.; Boeke, J. D.; Bussey, H.; Chu, A. M.; Connelly, C.; Davis, K.; Dietrich, F.; Dow, S. W.; El Bakkoury, M.; Foury, F.; Friend, S. H.; Gentalen, E.; Giaever, G.; Hegemann, J. H.; Jones, T.; Laub, M.; Liao, H.; Liebundguth, N.; Lockhart, D. J.; Lucau-Danila, A.; Lussier, M.; M'Rabet, N.; Menard, P.; Mittmann, M.; Pai, C.; Rebischung, C.; Revuelta, J. L.; Riles, L.; Roberts, C. J.; Ross-MacDonald, P.; Scherens, B.; Snyder, M.; Sookhai-Mahadeo, S.; Storms, R. K.; Veronneau, S.; Voet, M.; Volckaert, G.; Ward, T. R.; Wysocki, R.; Yen, G. S.; Yu, K.; Zimmermann, K.; Philippsen, P.; Johnston, M.; Davis, R. W. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **1999,** 285, 901-906.

30. Heitman, J.; Movva, N. R.; Hall, M. N. Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. *Science* **1991,** 253, 905-909.

31. Giaever, G.; Shoemaker, D. D.; Jones, T. W.; Liang, H.; Winzeler, E. A.; Astromoff, A.; Davis, R. W. Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat. Genet.* **1999,** 21, 278-283.

32. Giaever, G.; Flaherty, P.; Kumm, J.; Proctor, M.; Nislow, C.; Jaramillo, D. F.; Chu, A. M.; Jordan, M. I.; Arkin, A. P.; Davis, R. W. Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** 101, 793-798.

33. Luesch, H.; Wu, T. Y.; Ren, P.; Gray, N. S.; Schultz, P. G.; Supek, F. A genome-wide overexpression screen in yeast for small-molecule target identification. *Chem. Biol.* **2005,** 12, 55-63.

34. Ben-Aroya, S.; Coombes, C.; Kwok, T.; O'Donnell, K. A.; Boeke, J. D.; Hieter, P. Toward a comprehensive temperature-sensitive mutant repository of the essential genes of Saccharomyces cerevisiae. *Mol. Cell* **2008,** 30, 248-258.

35. Hughes, T. R.; Marton, M. J.; Jones, A. R.; Roberts, C. J.; Stoughton, R.; Armour, C. D.; Bennett, H. A.; Coffey, E.; Dai, H.; He, Y. D.; Kidd, M. J.; King, A. M.; Meyer, M. R.; Slade, D.; Lum, P. Y.; Stepaniants, S. B.; Shoemaker, D. D.; Gachotte, D.; Chakraburtty, K.; Simon, J.; Bard, M.; Friend, S. H. Functional discovery via a compendium of expression profiles. *Cell* **2000,** 102, 109-126.

36. Parsons, A. B.; Brost, R. L.; Ding, H.; Li, Z.; Zhang, C.; Sheikh, B.; Brown, G. W.; Kane, P. M.; Hughes, T. R.; Boone, C. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat. Biotechnol.* **2004,** 22, 62-69.

37. Tong, A. H.; Evangelista, M.; Parsons, A. B.; Xu, H.; Bader, G. D.; Page, N.; Robinson, M.; Raghibizadeh, S.; Hogue, C. W.; Bussey, H.; Andrews, B.; Tyers, M.; Boone, C. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **2001,** 294, 2364-2368.

38. Fields, S.; Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **1989,** 340, 245-246.

39. Johnston, M. A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. *Microbiol. Rev.* **1987,** 51, 458-476.

40. Gyuris, J.; Golemis, E.; Chertkov, H.; Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **1993,** 75, 791-803.

41. Lodi, T.; Donnini, C.; Ferrero, I. Catabolite repression by galactose in overexpressed GAL4 strains of Saccharomyces cerevisiae. *J. Gen. Microbiol.* **1991,** 137, 1039-1044.

42. Licitra, E. J.; Liu, J. O. A three-hybrid system for detecting small ligand-protein receptor interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** 93, 12817-12821.

43. Lin, H.; Abida, W. M.; Sauer, R. T.; Cornish, V. W. Dexamethasone−Methotrexate: An Efficient Chemical Inducer of Protein Dimerization In Vivo. *J. Am. Chem. Soc.* **2000,** 122, 4247-4248.

44. Becker, F.; Murthi, K.; Smith, C.; Come, J.; Costa-Roldan, N.; Kaufmann, C.; Hanke, U.; Degenhart, C.; Baumann, S.; Wallner, W.; Huber, A.; Dedier, S.; Dill, S.; Kinsman, D.; Hediger, M.; Bockovich, N.; Meier-Ewert, S.; Kluge, A. F.; Kley, N. A three-hybrid approach to scanning the proteome for targets of small molecule kinase inhibitors. *Chem. Biol.* **2004,** 11, 211-223.

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).[1] 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.[2] 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 chomophores may be linked with galactose, creating a chromogenic β-gal substrate to detect *L*acZ 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 spectra[8] (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 O_2, 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.

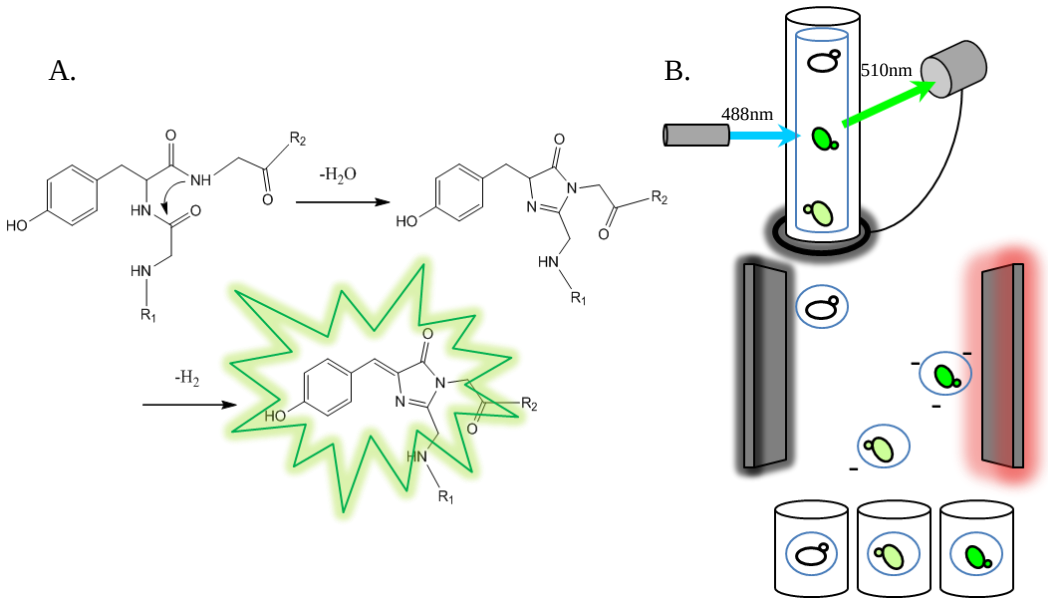


Figure 2.1. Maturations of EGFP fluorophore(A) takes place inside the β-barrel structure and requires no cofactors or post-translational modifications. Live cells can be sorted based on the fluorescence intensity of expressed EGFP with FACS (depicted in B).

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).[13-15] This technology, depicted in Figure 2.1B, uses microfluidics to push a cell culture into a single file stream that passes through a 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.[16] The vector used for this reporter contains eight LexA binding sites and the *GAL1-10* divergent promoter, taken from the commercial *LacZ* reporter pSH18-34 (Invitrogen) and inserted into the leucine selectable plasmid pBC103.[17] The resulting plasmid, pDCLryEGFP,[2] 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).

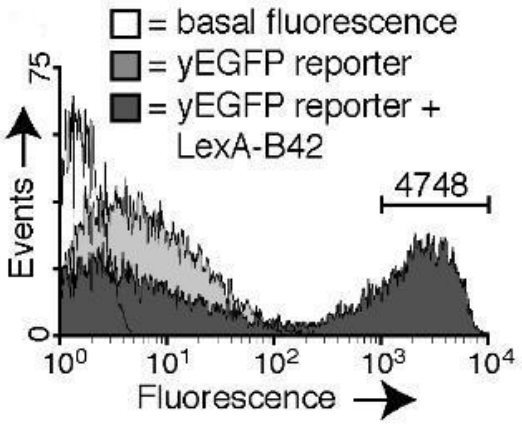


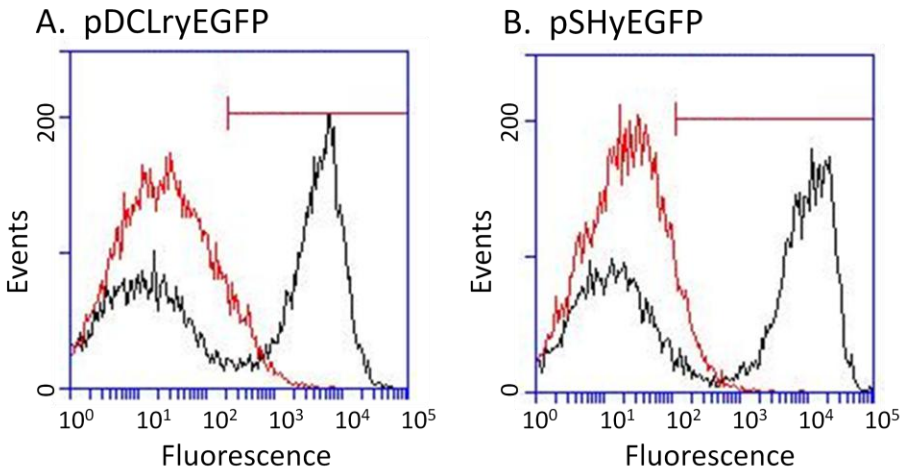
Figure 2.2. Overlay of flow-cytometric histograms demonstrating fluorescence from pDCLryEGFP reporter expression. Each histogram represents 15,000 events, with a gate set to quantify the number of highly fluorescent cells from each experiment. Basal fluorescence was quantified with yeast transformants 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,[2] 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-activited 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 (Figure2.3). This effect was almost equally the result of decreased background (45%) and increased fluorescent signal (161%).



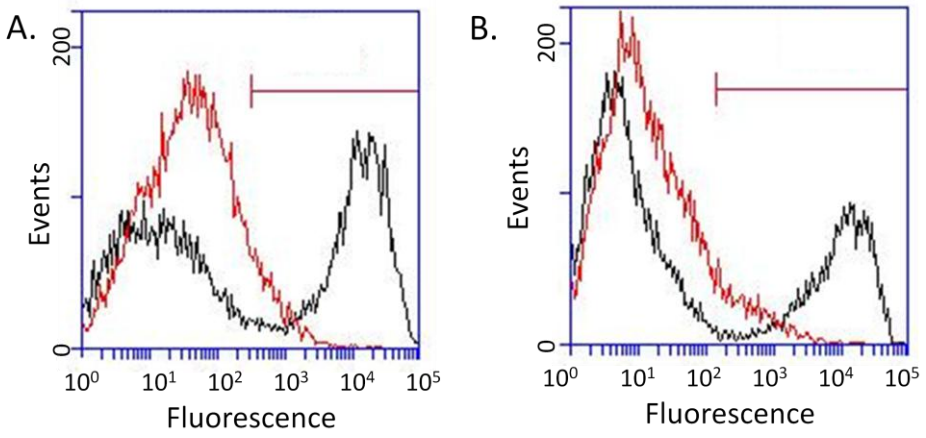
C.

pDCLryEGFP		pSHyEGFP
21,611	Background	11,893
501,209	Med. Signal	1,310,969
83.3	(CV, %)	80.3
23.192	S/N	**110.23**

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).

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.



C.

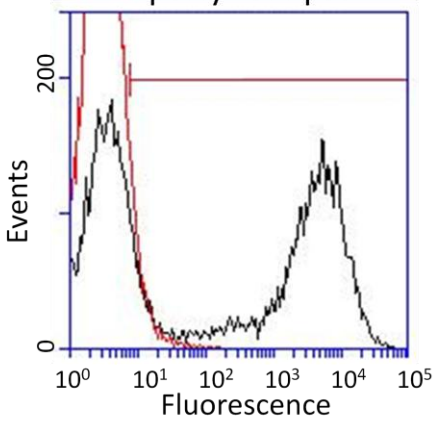
1. reporter 2. LexA-B42	Order	1. LexA-B42 2. reporter
31,623	Background	13,895
1,416,891	Med. Signal	1,162,313
83.2	CV,%	81.7
44.8	S/N	**83.6**

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 *yEGFP* 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.

A.



B.

pSHyEGFPpst	
Background	730.0
Med. Signal	362,661
(CV, %)	114.3
S/N	**496.8**

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.

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.

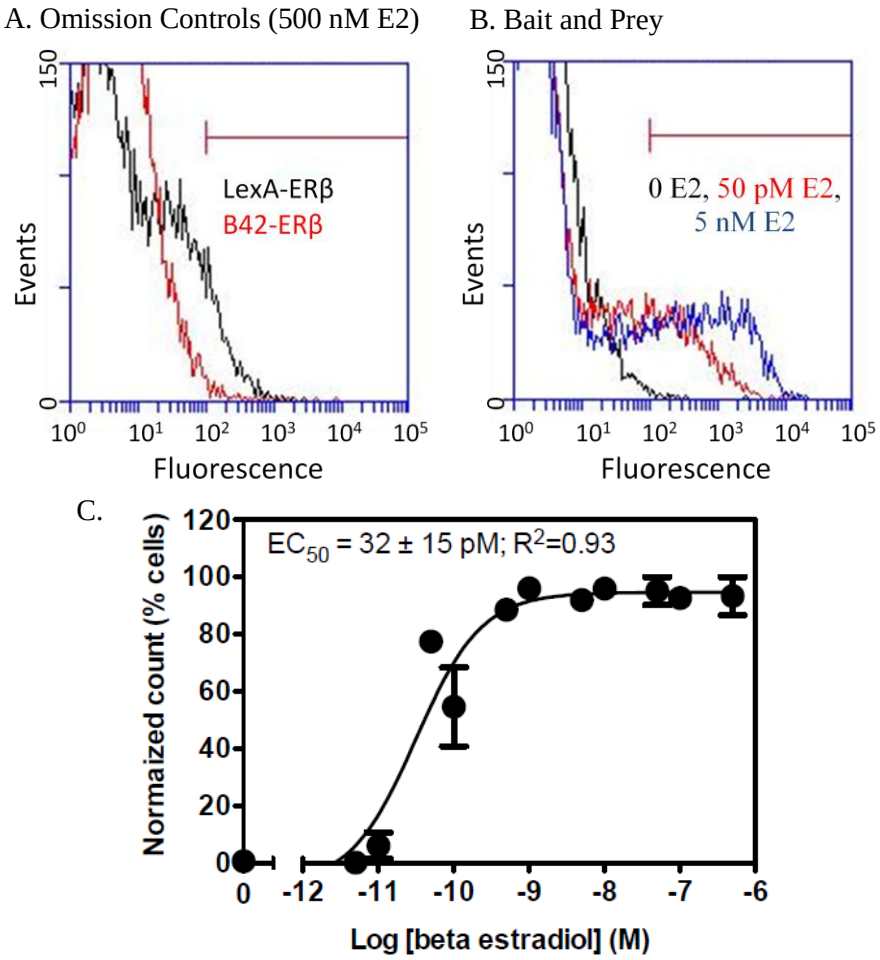


Figure 2.6. Overlay of flow-cytometric histograms demonstrating E2 dependent dimerization of ERβ(255-509) monomers. Each histogram represents 10,000 events with gate set to quantify highly fluorescent cells (>10^2 rfu) used in analysis. A) Bait and prey omission controls, treated with 500 nM E2, show 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 dihydrotestoterone (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.[18] Yeast have been used to study mammalian steroid receptors through ligand-dependent transcriptional activation,[19-23] 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.[24, 25] The TAF-2 domain of ERβ (between residues 535 and 545) was not included in the fragment used for ER dimerization assays.

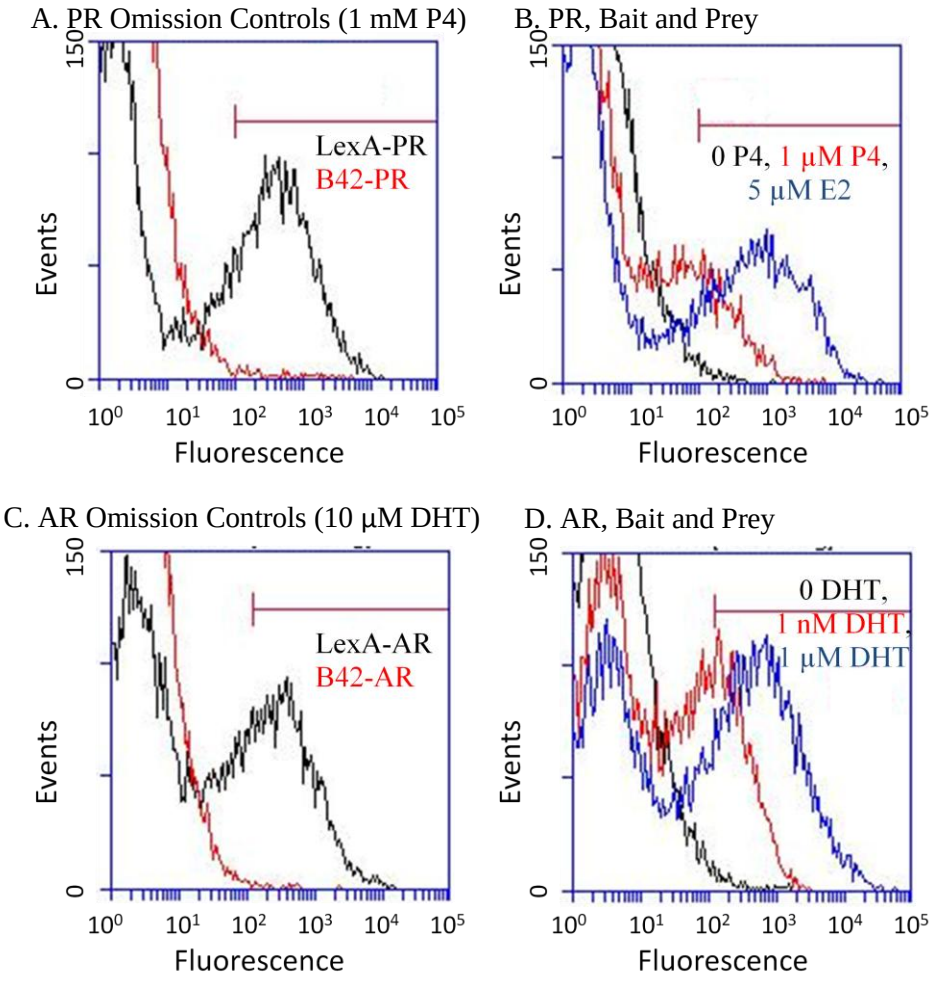


Figure 2.7. Overlays of flow-cytometric histograms from PR (P4 CID) and AR (DHT CID) yeast ligand dependant two-hybrid assays. Each histogram represents 10,000 events with gate of highly fluorescent cells (>100 rfu). A,C) Omission controls with maximum CID treatments and (B,D) representative histograms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

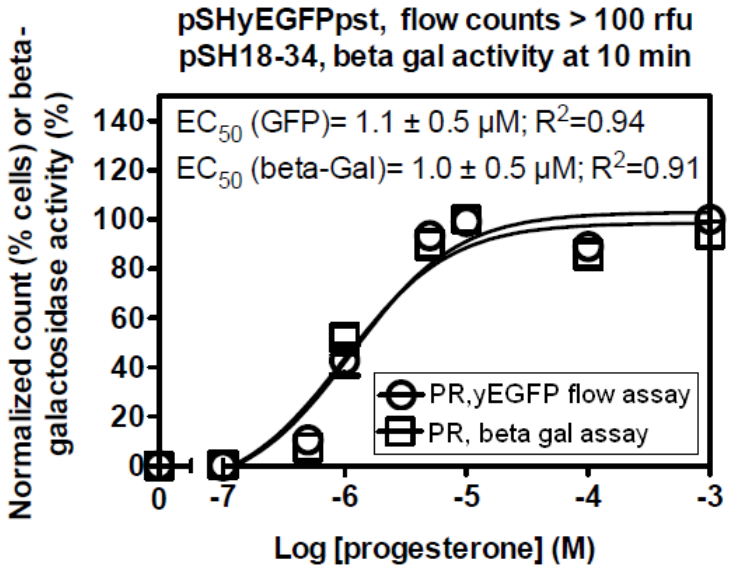
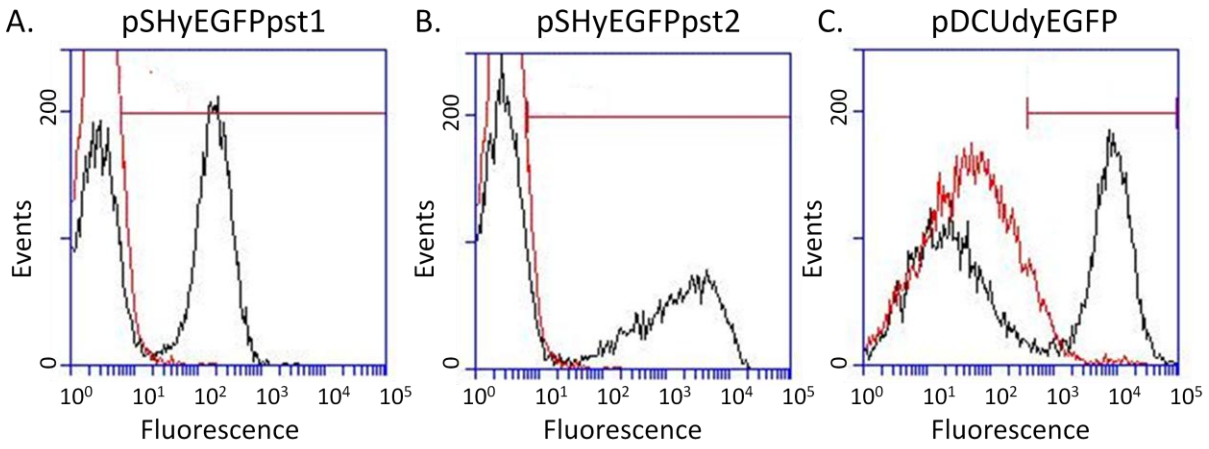


Figure 2.8. Overlay of dose-response curves of ligand dependent PR yeast two-hybrid assays treated with progesterone. The yEGFP curve is based on counts of cells emitting fluorescence >100 rfu. The LacZ curve is based on calculated β-gal activity using a CPRG substrate assay (10 minutes after substrate 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.[14, 26] Plasmids used for the described yeast genetic systems, including reporter pSHyEGFPpst, 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.[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 *yEGFP* copy number, yielding a more precise and consistent signal intensity.[28] This was achieved by integrating *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 capabilitiy), transforming it into a yeast integration plasmid (yIP), termed pSHIPyEGFPpst. Integration at the genomic *URA3* locus was performed by restriction digest of pSHIPyEGFPpst at single site within its *URA3* selection marker (*SapI*). The linearized plasmid was then transformed into yeast FY250. Successfully integrated yeast strains should contain *yEGFP* and a functional *URA3* selection marker. A small set of these yeast colonies were separately screened by expression of *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.



D.

	FY250yEGFPpst1 (A)	FY250yEGFPpst2 (B)	pDCUdyEGFP (C)
Background	594	636	38,841
Med. Signal	17,372	142,618	873,041
CV,%	80.0	123.1	78.3
S/N	**29.2**	**224**	**22.477**

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.

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.[29] An N-terminal deletion in the *URA3* promoter, termed *URA3-d*,[29] 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 pDCLryEGFP. 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.[30] 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 (*MATα, 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[31], in part because necessary restriction sites in the vector were not available. The vector backbone and upstream control elements was amplified from pSH18-34[30, 32] using PCR primers (pSH fwd: CATAACGCAGGAAAGAACATGTGAGCA; pSH rvs: GCACTTTTCGGCCAATGGTCTTG). Reporter gene *yEGFP* and yADHt were amplified from pDCLryEGFP[2] using PCR primers (VHR-yEGFP fwd: GACCATTGGCCGAAAAGTGC ATGTCTAAAGGTGAAGAATTATTCACTG; VHR-ADHt-yEGFP: ATGTTCTTTCCTGCGTTATG GCATGCCGGTAGAGGTGT) that flank the insert with 20 bp homology to the vector. Cloning was performed following the protocol as reported,[31] with scaled down 30 min exonuclease chew-back reactions using approximately 500 μg vector and insert PCR products in 10 μL reactions.

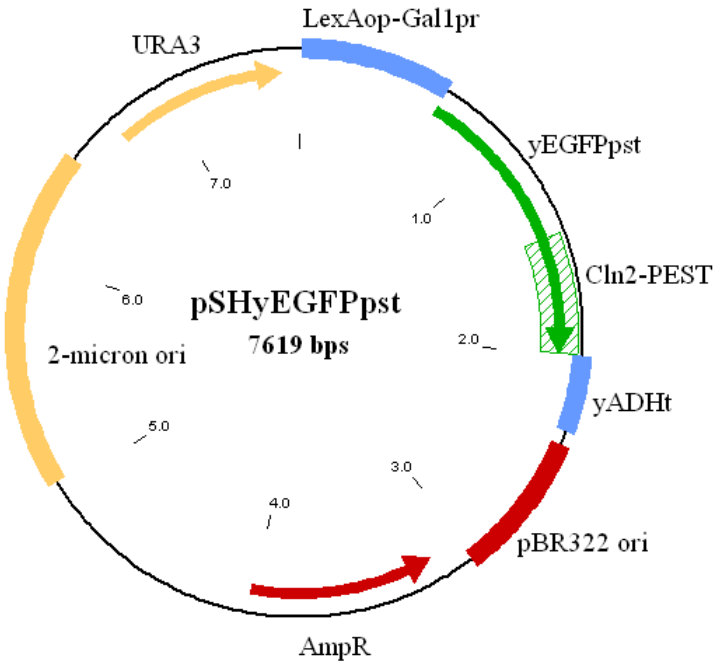


Figure 2.10. Plasmid map of pSHyEGFPpst. The 2-micron ori and *URA3* selection marker, shown in gold, maintain plasmid copies in yeast. The pBR322 ori and *AmpR* selection marker, shown in red, maintain plasmid copies for propagation in *E. coli*. Plasmid pSHyEGFP (7091 bps) lacks the Cln2-PEST (green, 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: CCCATGGTATGGATGAATTG GCATCCAACTTGAACATTTCGAG; VHR-PEST rvs: GGCGAAGAAGTCCAAAGCTTGTCGAC CTATATTACTTGGGTATTGCCCATAC) that flanked the insert with 20 bp vector homology. The PCR fragment was cloned into XhoI digested pSHyEGFP using SLIC.

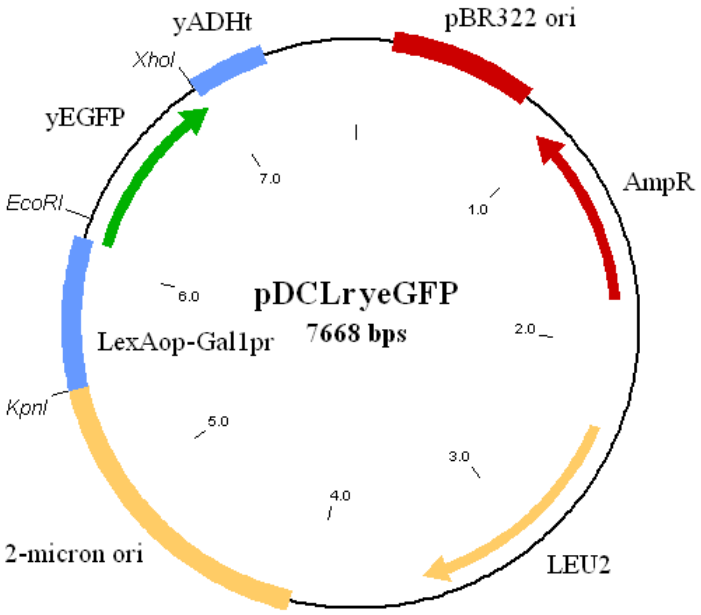


Figure 2.11. Plasmid map of pDCLryEGFP. In pDCUdyEGFP (6994 bps) the *LEU2* selection marker (gold arrow) is replaced with a *URA3-d* selection marker, but otherwise is identical to pDCLryEGFP.

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: GGCCCTTTCGTCTTCAAGAA ATAACCCAACTGCACAGAACAA; URA3d rvs: TGTCAAACATGAGAATTAAT CCCGGGTAATAACTGATATAATTAAATTGAAG) 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:

TCGTCTTCAAGAATTCTCATGTTTGACAGCTTATCATCGA; pSHIPyEGFP rvs: GTCAAACATGAGAATTCTTGAAGACGAAAGGGCCT) 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).[33] Plasmid pAM423 LexA-B42 (previously constructed)[33] 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[34] 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 reporer 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[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)[36] using the "Quick and Easy" LiAc/SS carrier DNA/PEG method[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.[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 pSHyEGFPpst. 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).[38] 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

1. Çakar, Z. P.; Sauer, U.; Bailey, J. E. Principles and practice of recombinant DNA research with yeast. In *The molecular biology of the yeast Saccharomyces cerevisiae.*, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 1982; pp 607-636.
2. Clark, D. D.; Peterson, B. R. Fluorescence-based cloning of a protein tyrosine kinase with a yeast tribrid system. *Chembiochem* **2005,** 6, 1442-1448.

3. de Felipe, K. S.; Carter, B. T.; Althoff, E. A.; Cornish, V. W. Correlation between ligand-receptor affinity and the transcription readout in a yeast three-hybrid system. *Biochemistry (Mosc.)* **2004,** 43, 10353-10363.

4. Nir, R.; Yisraeli, Y.; Lamed, R.; Sahar, E. Flow cytometry sorting of viable bacteria and yeasts according to beta-galactosidase activity. *Appl. Environ. Microbiol.* **1990,** 56, 3861-3866.

5. Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y. Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* **1995,** 20, 448-455.

6. Niedenthal, R. K.; Riles, L.; Johnston, M.; Hegemann, J. H. Green fluorescent protein as a marker for gene expression and subcellular localization in budding yeast. *Yeast* **1996,** 12, 773-786.

7. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **1994,** 263, 802-805.

8. Morise, H.; Shimomura, O.; Johnson, F. H.; Winant, J. Intermolecular energy transfer in the bioluminescent system of Aequorea. *Biochemistry (Mosc.)* **1974,** 13, 2656-2662.

9. Cormack, B. P.; Valdivia, R. H.; Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **1996,** 173, 33-38.

10. Mateus, C.; Avery, S. V. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast* **2000,** 16, 1313-1323.

11. Wang, H.; Peters, G. A.; Zeng, X.; Tang, M.; Ip, W.; Khan, S. A. Yeast two-hybrid system demonstrates that estrogen receptor dimerization is ligand-dependent in vivo. *J. Biol. Chem.* **1995,** 270, 23322-23329.

12. Cormack, R. S.; Hahlbrock, K.; Somssich, I. E. Isolation of putative plant transcriptional coactivators using a modified two-hybrid system incorporating a GFP reporter gene. *Plant J.* **1998,** 14, 685-692.

13. Starling, A. L.; Ortega, J. M.; Gollob, K. J.; Vicente, E. J.; Andrade-Nobrega, G. M.; Rodriguez, M. B. Evaluation of alternative reporter genes for the yeast two-hybrid system. *Genet. Mol. Res.* **2003,** 2, 124-135.

14. De Wulf, P.; Brambilla, L.; Vanoni, M.; Porro, D.; Alberghina, L. Real-time flow cytometric quantification of GFP expression and Gfp-fluorescence generation in Saccharomyces cerevisiae. *J. Microbiol. Methods* **2000,** 42, 57-64.

15. Cormack, B. Green fluorescent protein as a reporter of transcription and protein localization in fungi. *Curr. Opin. Microbiol.* **1998,** 1, 406-410.

16. Cormack, B. P.; Bertram, G.; Egerton, M.; Gow, N. A.; Falkow, S.; Brown, A. J. Yeast-enhanced green fluorescent protein (yEGFP)a reporter of gene expression in Candida albicans. *Microbiology* **1997,** 143 (Pt 2), 303-311.

17. Peterson, B. R.; Sun, L. J.; Verdine, G. L. A critical arginine residue mediates cooperativity in the contact interface between transcription factors NFAT and AP-1. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** 93, 13671-13676.

18. Danielian, P. S.; White, R.; Lees, J. A.; Parker, M. G. Identification of a conserved region required for hormone dependent transcriptional activation by steroid hormone receptors. *EMBO J.* **1992,** 11, 1025-1033.

19. Gaido, K. W.; Leonard, L. S.; Lovell, S.; Gould, J. C.; Babai, D.; Portier, C. J.; McDonnell, D. P. Evaluation of chemicals with endocrine modulating activity in a yeast-based steroid hormone receptor gene transcription assay. *Toxicol. Appl. Pharmacol.* **1997,** 143, 205-212.

20. Metzger, D.; Losson, R.; Bornert, J. M.; Lemoine, Y.; Chambon, P. Promoter specificity of the two transcriptional activation functions of the human oestrogen receptor in yeast. *Nucleic Acids Res.* **1992,** 20, 2813-2817.

21. Schena, M.; Yamamoto, K. R. Mammalian glucocorticoid receptor derivatives enhance transcription in yeast. *Science* **1988,** 241, 965-967.

22. Bovee, T. F.; Helsdingen, R. J.; Hamers, A. R.; van Duursen, M. B.; Nielen, M. W.; Hoogenboom, R. L. A new highly specific and robust yeast androgen bioassay for the detection of agonists and antagonists. *Anal. Bioanal. Chem.* **2007,** 389, 1549-1558.

23. De Boever, P.; Demare, W.; Vanderperren, E.; Cooreman, K.; Bossier, P.; Verstraete, W. Optimization of a yeast estrogen screen and its applicability to study the release of estrogenic isoflavones from a soygerm powder. *Environ. Health Perspect.* **2001,** 109, 691-697.

24. Meyer, M. E.; Quirin-Stricker, C.; Lerouge, T.; Bocquel, M. T.; Gronemeyer, H. A limiting factor mediates the differential activation of promoters by the human progesterone receptor isoforms. *J. Biol. Chem.* **1992,** 267, 10882-10887.

25. Bevan, C. L.; Hoare, S.; Claessens, F.; Heery, D. M.; Parker, M. G. The AF1 and AF2 domains of the androgen receptor interact with distinct regions of SRC1. *Mol. Cell. Biol.* **1999,** 19, 8383-8392.

26. Nacken, V.; Achstetter, T.; Degryse, E. Probing the limits of expression levels by varying promoter strength and plasmid copy number in Saccharomyces cerevisiae. *Gene* **1996,** 175, 253-260.

27. Hjortso, M. A.; Bailey, J. E. Plasmid stability in budding yeast populations: dynamics following a shift to nonselective medium. *Biotechnol. Bioeng.* **1984,** 26, 814-819.

28. Chen, J.; Zhou, J.; Bae, W.; Sanders, C. K.; Nolan, J. P.; Cai, H. A yEGFP-based reporter system for high-throughput yeast two-hybrid assay by flow cytometry. *Cytometry A* **2008,** 73, 312-320.

29. Loison, G.; Vidal, A.; Findeli, A.; Roitsch, C.; Balloul, J. M.; Lemoine, Y. High level of expression of a protective antigen of schistosomes in Saccharomyces cerevisiae. *Yeast* **1989,** 5, 497-507.

30. Ausubel, F. M.; Brent, R.; Kingston, R. E.; Moore, D. D.; Seidman, J. G.; Smith, J. A.; Struhl, K. *Short Protocols in Molecular Biology*. Fifth ed.; John Wiley & Sons, Inc: 2002.

31. Li, M. Z.; Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **2007,** 4, 251-256.

32. Gyuris, J.; Golemis, E.; Chertkov, H.; Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **1993,** 75, 791-803.

33. Clark, D. D.; Peterson, B. R. Analysis of protein tyrosine kinase inhibitors in recombinant yeast lacking the ERG6 gene. *Chembiochem* **2003,** 4, 101-107.

34. Gietz, R. D.; Schiestl, R. H. Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc* **2007,** 2, 1-4.

35. Gietz, R. D.; Schiestl, R. H. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat.. Protoc.* **2007,** 2, 35-37.

36. Hussey, S. L.; Muddana, S. S.; Peterson, B. R. Synthesis of a beta-estradiol-biotin chimera that potently heterodimerizes estrogen receptor and streptavidin proteins in a yeast three-hybrid system. *J. Am. Chem. Soc.* **2003,** 125, 3692-3693.

37. Muddana, S. S.; Peterson, B. R. Fluorescent cellular sensors of steroid receptor ligands. *Chembiochem* **2003,** 4, 848-855.

38. Muddana, S. S.; Peterson, B. R. Facile synthesis of cids: biotinylated estrone oximes efficiently heterodimerize estrogen receptor and streptavidin proteins in yeast three hybrid systems. *Org. Lett.* **2004,** 6, 1409-1412.

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 ($K_d \approx 10^{-14}$ 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 wass 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 β-strands with short connecting loops, and then folds into a classic up and down β-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[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,[3] found in chicken egg white. Chicken avidin actually has a slightly higher binding affinity,[8] but is notorious for nonspecific binding due to its native glycosolation 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.[9]

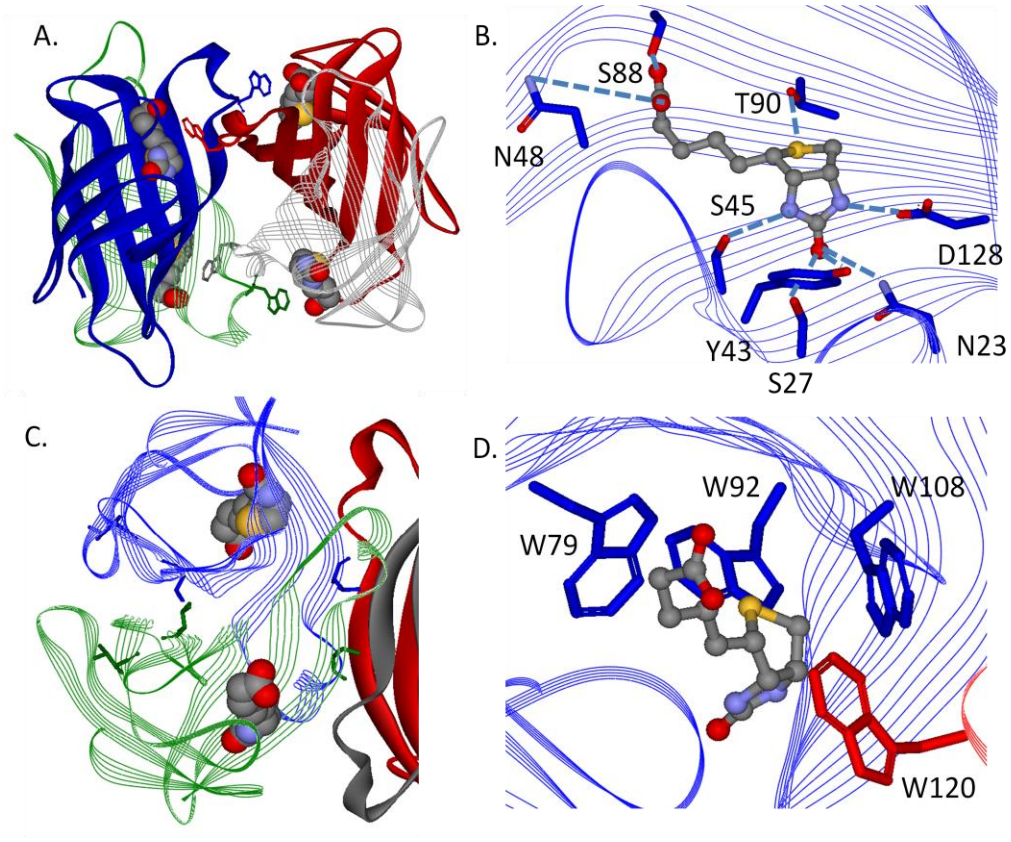


Figure 3.1. Selected views of X-ray crystal structure of tetrameric core streptavidin complexed with biotin (PDB: 1SWE).[10] A) Within a functional dimer (solid red and blue) two W120 residues (top, shown in sticks) enhance biotin affinity. C) Structural dimer (blue and green) interfaces are more extensive. 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.[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 (Figure3.1B).[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).[12] In the absence of biotin, five water molecules mimic biotin,[5] but biotin binding induces conformational changes that stabilize a complete streptavidin.[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 ($K_d \approx 1.3 \times 10^{-8}$ 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 tetrmerization through an H127A mutation and a loop (G113-W120, between β-strands 7 and 8) deletion.[18] These mutations did result in a soluble dimeric streptavidin, but biotin affinity was again dramatically reduced ($K_d \approx 1.5 \times 10^{-7}$ 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.[21] 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-fluoroscein unexpectedly maintained a high affinity ($K_d \approx 1.2 \times 10^{-10}$ M for both wild-type streptavidin and the SCD).

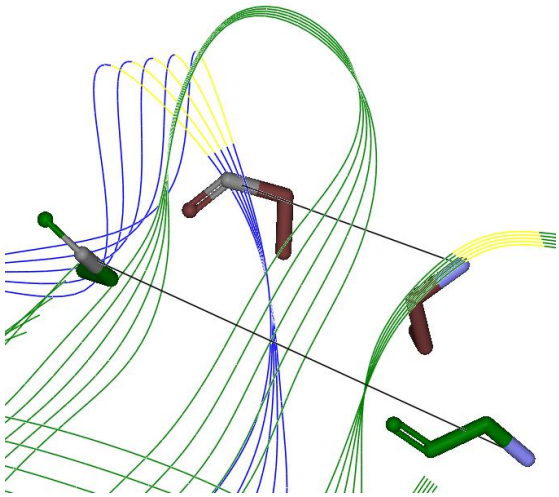


Figure 3.2. View of streptavidin X-ray crystal structure (PDB: 1SWE) with linkers in SCD streptavidin represented by black lines. The GGGS linker (bottom) between the native N- and C- termini (shown in green stick) allows direct fusion between two monomers (top) between S69 and T115 (shown in brown stick). 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.[22] 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.[22] This design was taken further by fusing duplicates of dcAvd with a 12 peptide serine and glycine rich linker.[23] 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.[24] Homotetramers of the "dead" mutant yield "negligible biotin binding" (app. $K_d=1.2 \times 10^{-3}$ M).[24] 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 chromotrography.[24] 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, Figure3.3).[21] 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 ($cpSA_{87}$ and $cpSA_{54}$; $cpSAnb_{87}$ and $cpSAnb_{54}$). 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}).

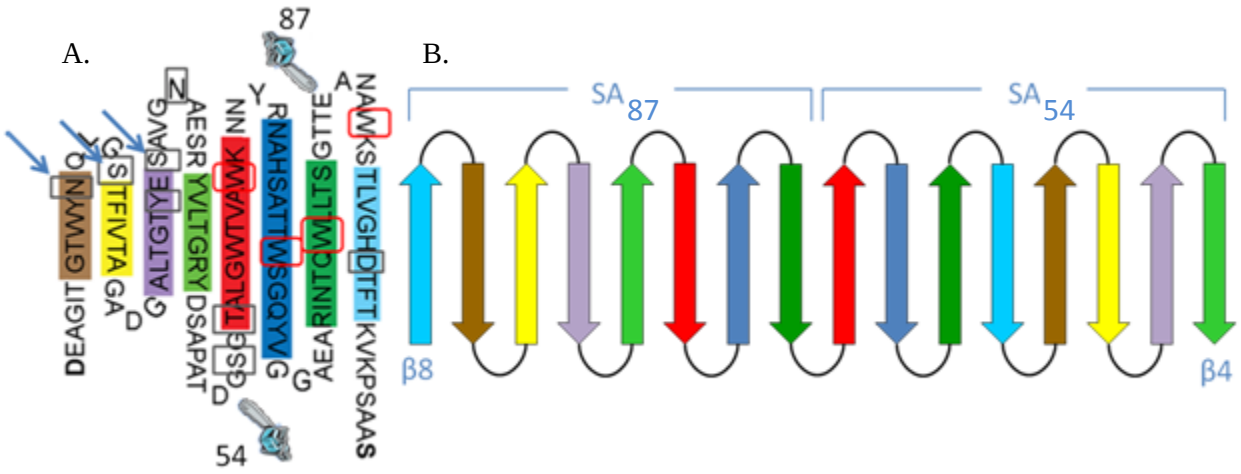


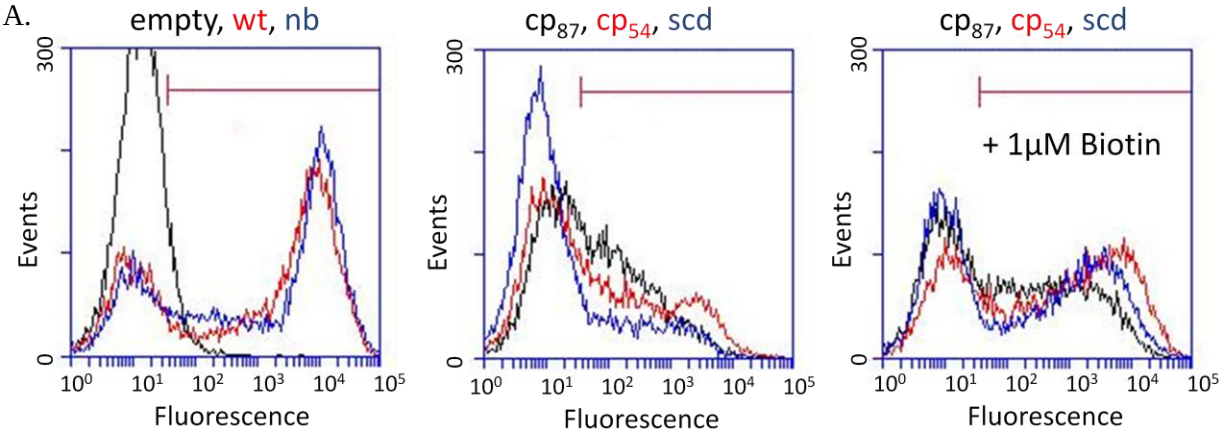
Figure 3.3. Topology map of core streptavidin monomer (A) and SCD (B), with identical coloring of β-strands. Residues involved in biotin binding are identified (H-bonding: grey boxes, Hydrophobic contacts: red boxes), and the three residues mutated to create SAnb (blue arrows). Chainsaws indicate where 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 oligmerizaiton, 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[25]) 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 $cpSAnb_{87}$. 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).

A.



B.

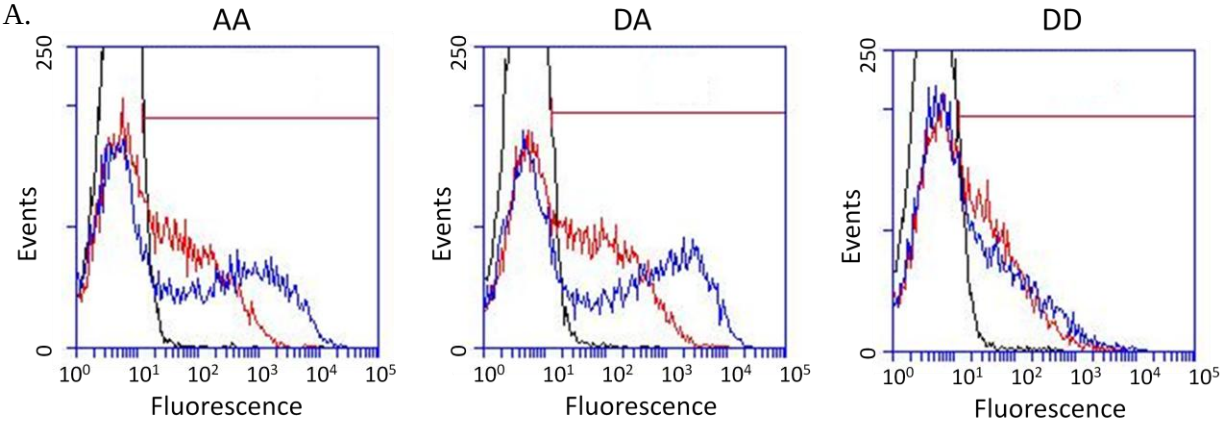
bait	wt	nb	cp_{87}	cp_{54}	scd
10 nM Biotin	6,616	6,934	3,412	3,494	1,779
1.01 μM Biotin	6,638	7,485	4,480	6,159	4,981

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_{DD} dimerization was not enhanced by excess biotin due to its negligible biotin affinity.

A.



B.

bait	AA	DA	DD
10 nM Biotin	1,956	2,500	1,220
1.01 µM Biotin	3,502	4,061	1,661

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 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[26] 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 streptavdin 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.[27] 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 (*MATα, 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]

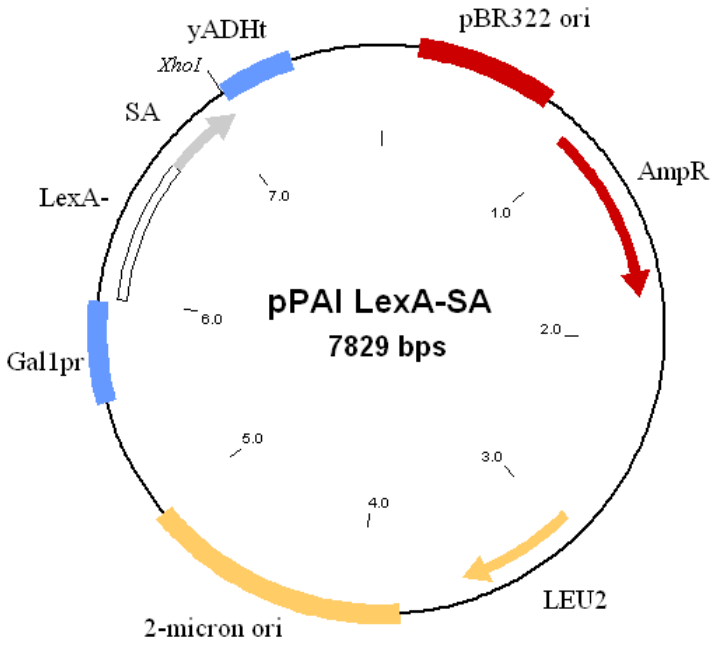


Figure 3.6. Plasmid map of pPA1 LexA-SA. All streptavidin LexA- fusions were cloned into pPA1 LexA, replacing SAwt.

To make all mutations in one round of mutagenesis, approximately half of pPA1 LexA-SA was amplified with PCR primers (SAtko fwd: CGTGACTGCAGGCGCCGACGGCGCCCTGACCGGAACCTACGAGGCTGCCGTCGGCAA; SAtko w/fwd: CGGTATCATTGCAGCACTGG) 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[28] to yield pPA1 LexA-SAnb.

Circular permutations $cpSA_{87}$, $cpSA_{54}$ fused to LexA were made with a two insert SLIC reaction. For permutation $cpSA_{87}$, core streptavidin residues 116-139 were amplified by PCR primers (SC1 fwd: CGCAACGGCGACTGGCTGGAATTC GAGGCCAACGCCTGGAAGTCC; SC1 rvs: GCCGGCCTCGTCGGATCCACCACC GGAGGCGGCGGACGGCTT) 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: GGTGGTGGATCC GACGAGGCCGGCATCACCGGCACCTGGT; SC2 rvs: GAAGTCCAAAGCTTCTCGAG TCA TCCGGAGGTGGTGCCGGAGGTCAGCAGCCACTG) 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 LexAstop with SLIC. Similarly for $cpSA_{54}$, streptavidin 69-139 were amplified with PCR primers (SC3 fwd: CGCAACGGCGACTGGCTGGAATTC TCCGGAACCGCCCTCGGTTGGACGGTGGCC; SC3 rvs: GGATCCACCACC GGAGGCGGCGGACGGCTTCACCTT) that flanked the product 5' 20-bp vector homology + BspEI site + 3'GGGS linker. Residues 1-68 were amplified with PCR primers (SC4 fwd: TCCGCCGCCTCCGGTGGTGGATCC GACGAGGCCGGCATCACCGGCACC; SC4 rvs: CGAAGAAGTCCAAAGCTTCTCGAG TCA GCCGTCGGTGGCCGGGGCGCTGTC) 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 LexAstop with SLIC. Circular permutations $cpSAnb_{87}$ and $cpSAnb_{54}$ 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 $cpSA(nb)_{54}$ downstream of $SA(nb)_{87}$ in pPA1 LexA. Constructs $cpSA_{54}$ and $cpSAnb_{54}$ were amplified with sequencing plasmid PCR primers (5'pPA1 LexA: CATTGAAGGGCTGGCGGTT; 3'pPA1 LexA: GACAACCTTGATTGGAGACTTGACC). The PCR product was digested with BspEI, XhoI and ligated into BspEI, XhoI digested pPA1 LexA-$cpSA_{87}$ or pPA1 LexA-$cpSAnb_{87}$.

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).[29]

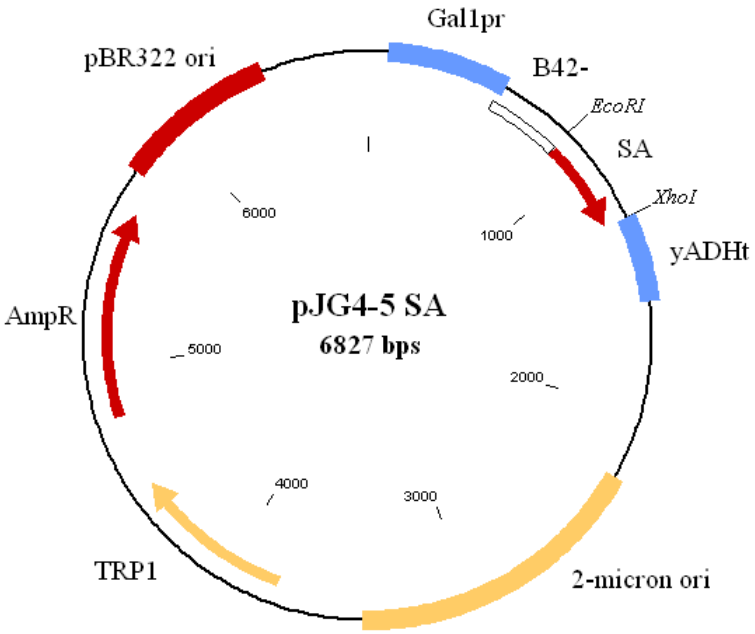


Figure 3.7. Plasmid map of pJG4-5 SA, used to create new B42- streptavidin fusions (SAnb, dSA_{DD}).

Yeast Three Hybrid Assays with Flow-Cytometry: Frozen competent yeast FY250 were first transformed with reporter plasmid pSHyEGFPpst 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

1. Green, N. M. Avidin and streptavidin. *Methods Enzymol.* **1990,** 184, 51-67.
2. Gonzalez, M.; Bagatolli, L. A.; Echabe, I.; Arrondo, J. L.; Argarana, C. E.; Cantor, C. R.; Fidelio, G. D. Interaction of biotin with streptavidin. Thermostability and conformational changes upon binding. *J. Biol. Chem.* **1997,** 272, 11288-11294.
3. Wilchek, M.; Bayer, E. A.; Livnah, O. Essentials of biorecognition: the (strept)avidin-biotin system as a model for protein-protein and protein-ligand interaction. *Immunol. Lett.* **2006,** 103, 27-32.
4. Diamandis, E. P.; Christopoulos, T. K. The biotin-(strept)avidin system: principles and applications in biotechnology. *Clin. Chem.* **1991,** 37, 625-636.
5. Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R. Structural origins of high-affinity biotin binding to streptavidin. *Science* **1989,** 243, 85-88.
6. Hendrickson, W. A.; Pahler, A.; Smith, J. L.; Satow, Y.; Merritt, E. A.; Phizackerley, R. P. Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** 86, 2190-2194.

7. Chilkoti, A.; Tan, P. H.; Stayton, P. S. Site-directed mutagenesis studies of the high-affinity streptavidin-biotin complex: contributions of tryptophan residues 79, 108, and 120. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** 92, 1754-1758.

8. Pazy, Y.; Kulik, T.; Bayer, E. A.; Wilchek, M.; Livnah, O. Ligand exchange between proteins. Exchange of biotin and biotin derivatives between avidin and streptavidin. *J. Biol. Chem.* **2002,** 277, 30892-30900.

9. Marttila, A. T.; Laitinen, O. H.; Airenne, K. J.; Kulik, T.; Bayer, E. A.; Wilchek, M.; Kulomaa, M. S. Recombinant NeutraLite avidin: a non-glycosylated, acidic mutant of chicken avidin that exhibits high affinity for biotin and low non-specific binding properties. *FEBS Lett.* **2000,** 467, 31-36.

10. Freitag, S.; Le Trong, I.; Klumb, L.; Stayton, P. S.; Stenkamp, R. E. Structural studies of the streptavidin binding loop. *Protein Sci.* **1997,** 6, 1157-1166.

11. Laitinen, O. H.; Hytonen, V. P.; Nordlund, H. R.; Kulomaa, M. S. Genetically engineered avidins and streptavidins. *Cell. Mol. Life Sci.* **2006,** 63, 2992-3017.

12. DeChancie, J.; Houk, K. N. The origins of femtomolar protein-ligand binding: hydrogen-bond cooperativity and desolvation energetics in the biotin-(strept)avidin binding site. *J. Am. Chem. Soc.* **2007,** 129, 5419-5429.

13. Williams, D. H.; Stephens, E.; Zhou, M. Ligand binding energy and catalytic efficiency from improved packing within receptors and enzymes. *J. Mol. Biol.* **2003,** 329, 389-399.

14. Laitinen, O. H.; Nordlund, H. R.; Hytonen, V. P.; Kulomaa, M. S. Brave new (strept)avidins in biotechnology. *Trends Biotechnol.* **2007,** 25, 269-277.

15. Wu, S. C.; Wong, S. L. Engineering soluble monomeric streptavidin with reversible biotin binding capability. *J. Biol. Chem.* **2005,** 280, 23225-23231.

16. Malmstadt, N.; Hyre, D. E.; Ding, Z.; Hoffman, A. S.; Stayton, P. S. Affinity thermoprecipitation and recovery of biotinylated biomolecules via a mutant streptavidin-smart polymer conjugate. *Bioconjug. Chem.* **2003,** 14, 575-580.

17. Qureshi, M. H.; Wong, S. L. Design, production, and characterization of a monomeric streptavidin and its application for affinity purification of biotinylated proteins. *Protein Expr. Purif.* **2002,** 25, 409-415.

18. Sano, T.; Vajda, S.; Smith, C. L.; Cantor, C. R. Engineering subunit association of multisubunit proteins: a dimeric streptavidin. *Proc. Natl. Acad. Sci. U. S. A.* **1997,** 94, 6153-6158.

19. Laitinen, O. H.; Nordlund, H. R.; Hytonen, V. P.; Uotila, S. T.; Marttila, A. T.; Savolainen, J.; Airenne, K. J.; Livnah, O.; Bayer, E. A.; Wilchek, M.; Kulomaa, M. S. Rational design of an active avidin monomer. *J. Biol. Chem.* **2003,** 278, 4010-4014.

20. Pazy, Y.; Eisenberg-Domovich, Y.; Laitinen, O. H.; Kulomaa, M. S.; Bayer, E. A.; Wilchek, M.; Livnah, O. Dimer-tetramer transition between solution and crystalline states of streptavidin and avidin mutants. *J. Bacteriol.* **2003,** 185, 4050-4056.

21. Aslan, F. M.; Yu, Y.; Mohr, S. C.; Cantor, C. R. Engineered single-chain dimeric streptavidins with an unexpected strong preference for biotin-4-fluorescein. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** 102, 8507-8512.

22. Nordlund, H. R.; Laitinen, O. H.; Hytonen, V. P.; Uotila, S. T.; Porkka, E.; Kulomaa, M. S. Construction of a dual chain pseudotetrameric chicken avidin by combining two circularly permuted avidins. *J. Biol. Chem.* **2004,** 279, 36715-36719.

23. Nordlund, H. R.; Hytonen, V. P.; Horha, J.; Maatta, J. A.; White, D. J.; Halling, K.; Porkka, E. J.; Slotte, J. P.; Laitinen, O. H.; Kulomaa, M. S. Tetravalent single-chain avidin: from subunits to protein domains via circularly permuted avidins. *Biochem. J.* **2005,** 392, 485-491.

24. Howarth, M.; Chinnapen, D. J.; Gerrow, K.; Dorrestein, P. C.; Grandy, M. R.; Kelleher, N. L.; El-Husseini, A.; Ting, A. Y. A monovalent streptavidin with a single femtomolar biotin binding site. *Nat. Methods* **2006,** 3, 267-273.

25. BD Bionutrients Technical Manual. In 3rd ed.; B.D.: 2006; p 64.

26. Athavankar, S.; Peterson, B. R. Control of gene expression with small molecules: biotin-mediated acylation of targeted lysine residues in recombinant yeast. *Chem. Biol.* **2003,** 10, 1245-1253.

27. Ausubel, F. M.; Brent, R.; Kingston, R. E.; Moore, D. D.; Seidman, J. G.; Smith, J. A.; Struhl, K. *Short Protocols in Molecular Biology*. Fifth ed.; John Wiley & Sons, Inc: 2002.

28. Li, M. Z.; Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **2007,** 4, 251-256.

29. Gyuris, J.; Golemis, E.; Chertkov, H.; Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **1993,** 75, 791-803.

30. Gietz, R. D.; Schiestl, R. H. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2007,** 2, 35-37.

APPENDIX A. Plasmid List

Section 1. Empty Plasmids

Name	Source	Expressed Gene	Restriction Sites	Plasmid Type	Bacterial Resistance	Yeast Marker	Seq. Primers
pBP7-7	BRPeterson	6x His tag	T7pr-XbaI-MRGSHHHHHHP-BamHI-NcoI-KpnI-MfeI-EcoRI-STOP-XhoI-SalI-STOP-HindIII	Bacterial Expression Vector, 2μ Ori	Amp	N/A	5' pLM, 3' pLM
pJG4-5	Brent Lab	SV40NLS-B42 AD-HA	Gal1pr- HindIII-SV40 NLS-B42 AD-HA Tag-EcoRI-Stuffer-XhoI-HindIII-ADH1t	Yeast Expression Vector, YEp	Amp	Trp 1	5' pJG4-5, 3' ADH1
pPA1 LexA stop	S Athavankar	HA-SV40NLS-LexA	KpnI-Gal1pr-EcoRI KO-HA SV40 NLS-LexA -EcoRI-XhoI-stop-ADH1t-BamHI-SalI-PstI	Yeast Expression Vector, YEp	Amp	LEU2	5' Gal1, 3' ADHI

Section 2. Reporter Constructs

Name	Source	Expressed Gene	Restriction Sites	Plasmid Type	Bacterial Resistance	Yeast Marker	Seq. Primers
pDCLr yEGFP	D. Clark	yEGFP	KpnI-8x LexAop-Gal1-10pr region-EcoRI-yEGFP-XhoI-ADH1t-BamHI	Yeast reporter plasmid (LexA DBD),YEp	Amp	LEU2	5' Gal1, 3' ADH1
pDCUd yEGFP	K. Bailey	yEGFP	KpnI-8x LexAop-Gal1-10pr region-EcoRI-yEGFP-XhoI-ADH1t-BamHI	Yeast reporter plasmid (LexA DBD), YEp	Amp	URA3-d	5' Gal1, 3' ADH1
pSH18-34	Brent Lab	beta-Gal reporter	N/A	Yeast reporter plasmid (LexA DBD),YEp	Amp	Ura 3	5' Gal1, 3' ADH1
pSH yEGFP	K. Bailey	yEGFP	BamHI-8xLexAop-Gal1-10pr region-yEGFP-XhoI-ADH1t-SphI	Yeast reporter plasmid (LexA DBD),YEp	Amp	URA3	5' Gal1, 3' ADH1
pSH yEGFPpst	K. Bailey	yEGFP-Cln2PEST	BamHI-8xLexAop-Gal1-10pr region-yEGFP-Cln2PEST-SalI-ADH1t-SphI	Yeast reporter plasmid (LexA DBD),YEp	Amp	URA3	5' Gal1, 3' ADH1
pSHIP yEGFPpst	K. Bailey	yEGFP-Cln2PEST	BamHI-8xLexAop-Gal1-10pr region-yEGFP-Cln2PEST-SalI-ADH1t-SphI	Yeast reporter plasmid (LexA DBD),YIp	Amp	URA3	5' Gal1, 3' ADH1

Section 3. Hormone Receptor Constructs

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

Section 4. Streptavidin Constructs

Name	Source	Expressed Gene	Restriction Sites	Plasmid Type	Bacterial Resistance	Yeast Marker	Seq. Primers
pJG4-5 dSA_{AA}	K. Bailey	SV40 NLS-B42 AD-HA -$cpSA_{87}$-$cpSA_{54}$	Gal1pr- HindIII-SV40 NLS-B42 AD-HA Tag-EcoRI-$cpSA_{87}$-BspEI-$cpSA_{54}$-XhoI-HindIII-ADH1t	Yeast Expression Vector, YEp	Amp	Trp 1	5' pJG4-5, 3' ADH1t
pJG4-5 $cpSAnb_{54}$	K. Bailey	SV40NLS-B42 AD-HA Tag-$cpSAnb_{54}$	Gal1pr- HindIII-SV40 NLS-B42 AD-HA Tag-EcoRI-$cpSAnb_{54}$-XhoI-HindIII-ADH1t	Yeast Expression Vector, YEp	Amp	Trp 1	5' pJG4-5, 3' ADH1t

pPA1 LexA $cpSA_{54}$	K. Bailey	HA-SV40NLS-LexA-$cpSA_{54}$	KpnI-Gal1pr-EcoRI KO-HA SV40 NLS-LexA -EcoRI-BspEI-$cpSA_{54}$-XhoI-ADH1t-SalI-PstI	Yeast Expression Vector, YEp	Amp	LEU2	5' Gal1, 3' ADH1t
pPA1 LexA $cpSA_{87}$	K. Bailey	HA-SV40NLS-LexA-$cpSA_{87}$	KpnI-Gal1pr-EcoRI KO-HA SV40 NLS-LexA -EcoRI-$cpSA_{87}$-BspEI-stop-XhoI-ADH1t-SalI-PstI	Yeast Expression Vector, YEp	Amp	LEU2	5' Gal1, 3' ADH1t
pPA1 LexA $cpSAnb_{54}$	K. Bailey	HA-SV40NLS-LexA-$cpSAnb_{54}$	KpnI-Gal1pr-EcoRI KO-HA SV40 NLS-LexA -EcoRI-BspEI-$cpSAnb_{54}$-XhoI-ADH1t-SalI	Yeast Expression Vector, YEp	Amp	LEU2	5' Gal1, 3' ADH1t
pPA1 LexA $cpSAnb_{87}$	K. Bailey	HA-SV40NLS-LexA-$cpSAnb_{87}$	KpnI-Gal1pr-EcoRI KO-HA SV40 NLS-LexA -EcoRI-$cpSAnb_{87}$-BspEI-stop-XhoI-ADH1t-SalI	Yeast Expression Vector, YEp	Amp	LEU2	5' Gal1, 3' ADH1t
pPA1 LexA dSA_{AA}	K. Bailey	HA-SV40NLS-LexA-$cpSA_{87}$-$cpSA_{54}$	KpnI-Gal1pr-EcoRI KO-HA SV40 NLS-LexA-$cpSA_{87}$-BspEI-$cpSA_{54}$-XhoI-ADH1t-SalI-PstI	Yeast Expression Vector, YEp	Amp	LEU2	5' Gal1, 3' ADH1t
pPA1 LexA dSA_{AD}	K. Bailey	HA-SV40NLS-LexA-$cpSA_{87}$-$cpSAnb_{54}$	KpnI-Gal1pr-EcoRI KO-HA SV40 NLS-LexA-$cpSAnb_{87}$-BspEI-$cpSAnb_{54}$-XhoI-ADH1t-SalI	Yeast Expression Vector, YEp	Amp	LEU2	5' Gal1, 3' ADH1t
pPA1 LexA dSA_{DD}	K. Bailey	HA-SV40NLS-LexA-$cpSAnb_{87}$-$cpSAnb_{54}$	KpnI-Gal1pr-EcoRI KO-HA SV40 NLS-LexA-$cpSAnb_{87}$-BspEI-$cpSAnb_{54}$-XhoI-ADH1t-BamHI-SalI	Yeast Expression Vector, YEp	Amp	LEU2	5' Gal1, 3' ADH1t
pPA1 LexA SA	K. Bailey	HA-SV40NLS-LexA--SA	KpnI-Gal1pr-EcoRI KO-HA SV40 NLS-LexA -SA-EcoRI-XhoI-stop-ADH1t-BamHI-SalI-PstI	Yeast Expression Vector, YEp	Amp	LEU2	5' Gal1, 3' ADH1t
pPA1 LexA SAnb	K. Bailey	HA-SV40NLS-LexA-SA(N23A,S27D,S45A)	KpnI-Gal1pr-EcoRI KO-HA SV40 NLS-LexA -SA(N23A,S27D,S45A)-EcoRI-XhoI-stop-ADH1t-BamHI-SalI-PstI	Yeast Expression Vector, YEp	Amp	LEU2	5' Gal1, 3' ADH1t

pSA1 T7SAwt B42	S Hussey	T7SAwt-NLS-B42 AD-HA tag-stop	KpnI-Gal1pr-T7 SAwt-NLS-B42-HA tag-EcoRI-XhoI-stop-ADH1t	Yeast Expression Vector, YEp	Amp	TRP1	5' Gal1, 3' ADH1t

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 tct aaa ggt gaa gaa tta ttc act ggt   361  aat aga atc gaa tta aaa ggt att gat ttt
  1     M   S   K   G   E   E   L   F   T   G  121     N   R   I   E   L   K   G   I   D   F

  31  gtt gtc cca att ttg gtt gaa tta gat ggt   391  aaa gaa gat ggt aac att tta ggt cac aaa
 11     V   V   P   I   L   V   E   L   D   G  131     K   E   D   G   N   I   L   G   H   K

  61  gat gtt aat ggt cac aaa ttt tct gtc tcc   421  ttg gaa tac aac tat aac tct cac aat gtt
 21     D   V   N   G   H   K   F   S   V   S  141     L   E   Y   N   Y   N   S   H   N   V

  91  ggt gaa ggt gaa ggt gat gct act tac ggt   451  tac atc atg gct gac aaa caa aag aat ggt
 31     G   E   G   E   G   D   A   T   Y   G  151     Y   I   M   A   D   K   Q   K   N   G

 121  aaa ttg acc tta aaa ttt att tgt act act   481  atc aaa gtt aac ttc aaa att aga cac aac
 41     K   L   T   L   K   F   I   C   T   T  161     I   K   V   N   F   K   I   R   H   N

 151  ggt aaa ttg cca gtt cca tgg cca acc tta   511  att gaa gat ggt tct gtt caa tta gct gac
 51     G   K   L   P   V   P   W   P   T   L  171     I   E   D   G   S   V   Q   L   A   D

 181  gtc act act ttc ggt tat ggt gtt caa tgt   541  cat tat caa caa aat act cca att ggt gat
 61     V   T   T   F   G   Y   G   V   Q   C  181     H   Y   Q   Q   N   T   P   I   G   D

 211  ttt gcg aga tac cca gat cat atg aaa caa   571  ggt cca gtc ttg tta cca gac aac cat tac
 71     F   A   R   Y   P   D   H   M   K   Q  191     G   P   V   L   L   P   D   N   H   Y

 241  cat gac ttt ttc aag tct gcc atg cca gaa   601  tta tcc act caa tct gcc tta tcc aaa gat
 81     H   D   F   F   K   S   A   M   P   E  201     L   S   T   Q   S   A   L   S   K   D

 271  ggt tat gtt caa gaa aga act att ttt ttc   631  cca aac gaa aag aga gac cac atg gtc ttg
 91     G   Y   V   Q   E   R   T   I   F   F  211     P   N   E   K   R   D   H   M   V   L

 301  aaa gat gac ggt aac tac aag acc aga gct   661  tta gaa ttt gtt act gct gct ggt att acc
101     K   D   D   G   N   Y   K   T   R   A  221     L   E   F   V   T   A   A   G   I   T
                                                                                    XhoI
 331  gaa gtc aag ttt gaa ggt gat acc tta gtt   691  cat ggt atg gat gaa ttg tac aaa tag ctc gag
111     E   V   K   F   E   G   D   T   L   V  231     H   G   M   D   E   L   Y   K   -
```

SV40NLS-B42-HA: Used for B42 AD fusion proteins

Gyuris, J. *et al*. H.; Brent, R. *Cell* **1993**, *75*, 791.

```
  1  atg ggt gct cct cca aaa aag aag aga aag  181  gcc ggg ccg ccg atc cag cct gac tgg ctg
 1     M   G   A   P   P   K   K   K   R   K  61     A   G   P   P   I   Q   P   D   W   L

 31  gta gct ggt atc aat aaa gat atc gag gag  211  aaa tcg aat ggt ttt cat gaa att gaa gcg
11     V   A   G   I   N   K   D   I   E   E  71     K   S   N   G   F   H   E   I   E   A

 61  tgc aat gcc atc att gag cag ttt atc gac  241  gat gtt aac gat acc agc ctc ttg ctg agt
21     C   N   A   I   I   E   Q   F   I   D  81     D   V   N   D   T   S   L   L   L   S

 91  tac ctg cgc acc gga cag gag atg ccg atg  271  gga gat gcc tcc tac cct tat gat gtg cca
31     Y   L   R   T   G   Q   E   M   P   M  91     G   D   A   S   Y   P   Y   D   V   P
                                                                      EcoRI
121  gaa atg gcg gat cag gcg att aac gtg gtg  301  gat tat gcc tct ccc gaa ttc
41     E   M   A   D   Q   A   I   N   V   V 101     D   Y   A   S   P   E   F

151  ccg ggc atg acg ccg aaa acc att ctt cac
51     P   G   M   T   P   K   T   I   L   H
```

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

Gyuris, J. *et al*. H.; Brent, R. *Cell* **1993**, *75*, 791.

```
  1  atg gcc tcc tac cct tat gat gtg cca gat  361  gaa ggt cat tat cag gtc gat cct tcc tta
 1     M   A   S   Y   P   Y   D   V   P   D 121     E   G   H   Y   Q   V   D   P   S   L

 31  tat gcc tct ccc gaa ttg gcc ccc aag aaa  391  ttc aag ccg aat gct gat ttc ctg ctg cgc
11     Y   A   S   P   E   L   A   P   K   K 131     F   K   P   N   A   D   F   L   L   R

 61  aag aga aag gtg gaa ttt atg aaa gcg tta  421  gtc agc ggg atg tcg atg aaa gat atc ggc
21     K   R   K   V   E   F   M   K   A   L 141     V   S   G   M   S   M   K   D   I   G

 91  acg gcc agg caa caa gag gtg ttt gat ctc  451  att atg gat ggt gac ttg ctg gca gtg cat
31     T   A   R   Q   Q   E   V   F   D   L 151     I   M   D   G   D   L   L   A   V   H

121  atc cgt gat cac atc agc cag aca ggt atg  481  aaa act cag gat gta cgt aac ggt cag gtc
41     I   R   D   H   I   S   Q   T   G   M 161     K   T   Q   D   V   R   N   G   Q   V

151  ccg ccg acg cgt gcg gaa atc gcg cag cgt  511  gtt gtc gca cgt att gat gac gag gtt acc
51     P   P   T   R   A   E   I   A   Q   R 171     V   V   A   R   I   D   D   E   V   T

181  ttg ggg ttc cgt tcc cca aac gcg gct gaa  541  gtt aag cgc ctg aaa aaa cag ggc aat aaa
61     L   G   F   R   S   P   N   A   A   E 181     V   K   R   L   K   K   Q   G   N   K

211  gaa cat ctg aag gcg ctg gca cgc aaa ggc  571  gtc gaa ctg ttg cca gaa aat agc gag ttt
71     E   H   L   K   A   L   A   R   K   G 191     V   E   L   L   P   E   N   S   E   F

241  gtt att gaa att gtt tcc ggc gca tca cgc  601  aaa ccg att gtc gta gat ctt cgt cag cag
81     V   I   E   I   V   S   G   A   S   R 201     K   P   I   V   V   D   L   R   Q   Q

271  ggg att cgt ctg ttg cag gaa gag gaa gaa  631  agc ttc acc att gaa ggg ctg gcg gtt ggg
91     G   I   R   L   L   Q   E   E   E   E 211     S   F   T   I   E   G   L   A   V   G
                                                                                  EcoRI
301  ggg ttg ccg ctg gta ggt cgt gtg gct gcc  661  gtt att cgc aac ggc gac tgg ctg gaa ttc
101    G   L   P   L   V   G   R   V   A   A 221     V   I   R   N   G   D   W   L   E   F

331  ggt gaa cca ctt ctg gcg caa cag cat att
111    G   E   P   L   L   A   Q   Q   H   I
```

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

Salama, S. R.; Hendricks, K. B.; Thorner, J. *Mol Cell Biol* **1994**, *14*, 7953.

```
  1  gca tcc aac ttg aac att tcg aga aag ctt  271  aat aac aaa caa agg atc tac gaa caa atc
 1     A   S   N   L   N   I   S   R   K   L  91     N   N   K   Q   R   I   Y   E   Q   I

 31  acc ata tca acc cca tca tgc tct ttc gaa  301  acg ggt cct aac agc aat aac gca acc aat
11     T   I   S   T   P   S   C   S   F   E 101     T   G   P   N   S   N   N   A   T   N

 61  aat tca aat agc aca tcc att cct tcg ccc  331  gat tat att gat ttg cta aac cta aat gag
21     N   S   N   S   T   S   I   P   S   P 111     D   Y   I   D   L   L   N   L   N   E

 91  gct tcc tca tct caa agc cac act cca atg  361  tct aac aag gaa aac caa aat ccc gca acg
31     A   S   S   S   Q   S   H   T   P   M 121     S   N   K   E   N   Q   N   P   A   T

121  aga aac atg agc tca ctc tct gat aac agc  391  gcg cat tac ctc aat ggg ggc cca ccc aag
41     R   N   M   S   S   L   S   D   N   S 131     A   H   Y   L   N   G   G   P   P   K

151  gtt ttc agc cgg aat atg gaa caa tca tca  421  aca agc ttc att aac cat gga atg ttc ccc
51     V   F   S   R   N   M   E   Q   S   S 141     T   S   F   I   N   H   G   M   F   P

181  cca atc act cca agt atg tac caa ttt ggt  451  tcg cca act ggg acc ata aat agc ggt aaa
61     P   I   T   P   S   M   Y   Q   F   G 151     S   P   T   G   T   I   N   S   G   K

211  cag cag cag tca aac agt ata tgt ggt agc  481  tct agc agt gcc tca tct tta att tct ttt
71     Q   Q   Q   S   N   S   I   C   G   S 161     S   S   S   A   S   S   L   I   S   F

241  acc gtt agt gtg aat agt ctg gtg aat aca  511  ggt atg ggc aat acc caa gta ata
81     T   V   S   V   N   S   L   V   N   T 171     G   M   G   N   T   Q   V   I
```

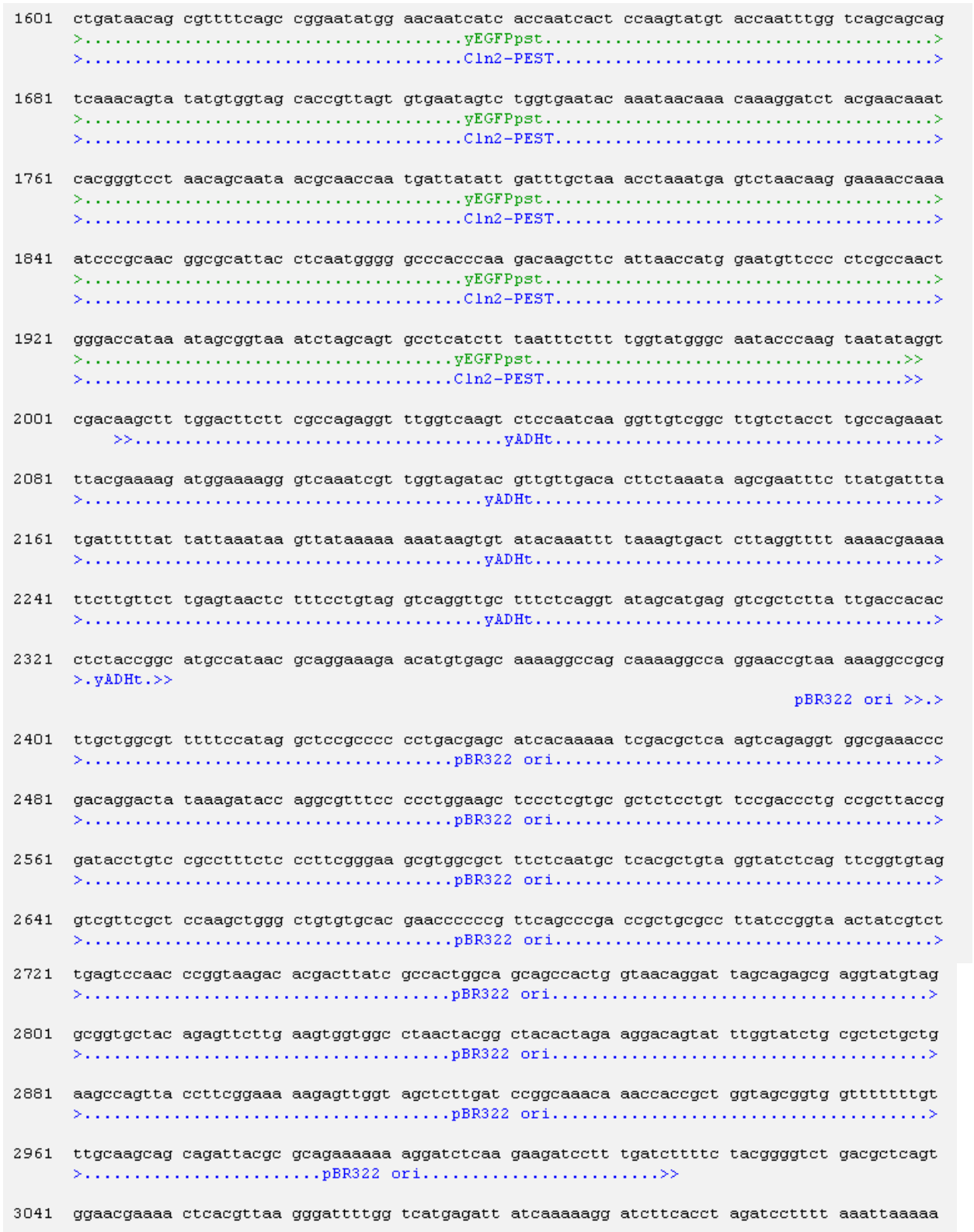
SA(13-139): Core streptavidin(13-139) used as template and benchmark for all SA variants.

Sano, T. *et al*. *J Biol Chem* **1995**, *270*, 28204.

```
  1  gac gag gcc ggc atc acc ggc acc tgg tac  211  tac cgc aac gcc cac tcc gcg acc acg tgg
 1     D   E   A   G   I   T   G   T   W   Y  71     Y   R   N   A   H   S   A   T   T   W

 31  aac cag ctc ggc tcg acc ttc atc gtg acc  241  agc ggc cag tac gtc ggc ggc gcc gag gcg
11     N   Q   L   G   S   T   F   I   V   T  81     S   G   Q   Y   V   G   G   A   E   A

 61  gcg ggc gcc gac ggc gcc ctg acc gga acc  271  agg atc aac acc cag tgg ctg ctg acc tcc
21     A   G   A   D   G   A   L   T   G   T  91     R   I   N   T   Q   W   L   L   T   S

 91  tac gag tcg gcc gtc ggc aac gcc gag agc  301  ggc acc acc gag gcc aac gcc tgg aag tcc
31     Y   E   S   A   V   G   N   A   E   S 101     G   T   T   E   A   N   A   W   K   S

121  cgc tac gtc ctg acc ggt cgt tac gac agc  331  acg ctg gtc ggc cac gac acc ttc acc aag
41     R   Y   V   L   T   G   R   Y   D   S 111     T   L   V   G   H   D   T   F   T   K

                                                                        EcoRI       XhoI
151  gcc ccg gcc acc gac ggc agc ggc acc gcc  361  gtg aag ccg tcc gcc gcc tcc gaa ttc tga ctc gag
51     A   P   A   T   D   G   S   G   T   A 121     V   K   P   S   A   A   S   E   F   -

181  ctc ggt tgg acg gtg gcc tgg aag aat aac
61     L   G   W   T   V   A   W   K   N   N
```

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

```
   1  ggatccccag cttggaattc gacaggttat cagcaacaac acagtcatat ccattctcaa ttagctctac cacagtgtgt
                 >>.........................LexAop-Gal1pr.............................>

  81  gaaccaatgt atccagcacc acctgtaacc aaaacaattt tagaagtact ttcactttgt aactgagctg tcatttatat
      >.....................................LexAop-Gal1pr.....................................>

 161  tgaattttca aaaattctta ctttttttt ggatggacgc aaagaagttt aataatcata ttacatggca ttaccaccat
      >.....................................LexAop-Gal1pr.....................................>

 241  atacatatcc atatacatat ccatatctaa tcttacctcg actgctgtat ataaaaccag tggttatatg tacagtactg
      >.....................................LexAop-Gal1pr.....................................>

 321  ctgtatataa aaccagtggt tatatgtaca gtacgtcgac tgctgtatat aaaaccagtg gttatatgta cagtactgct
      >.....................................LexAop-Gal1pr.....................................>

 401  gtatataaaa ccagtggtta tatgtacagt acgtcgaggg atgataatgc gattagtttt ttagccttat ttctggggta
      >.....................................LexAop-Gal1pr.....................................>

 481  attaatcagc gaagcgatga tttttgatct attaacagat atataaatgc aaaaactgca taaccacttt aactaatact
      >.....................................LexAop-Gal1pr.....................................>

 561  ttcaacattt tcggtttgta ttacttctta ttcaaatgta ataaaagtat caacaaaaaa ttgttaatat acctctatac
      >.....................................LexAop-Gal1pr.....................................>

 641  tttaacgtca aggagaaaaa actataatga ctaaatctca ttcagaagaa gtgattgtac ctgagttcaa ttctagcgca
      >......LexAop-Gal1pr......>>
                                 >>.......................yEGFPpst.........................>

 721  aaggaattac caagaccatt ggccgaaaag tgcatgtcta aaggtgaaga attattcact ggtgttgtcc caattttggt
      >.....................................yEGFPpst.....................................>

 801  tgaattagat ggtgatgtta atggtcacaa attttctgtc tccggtgaag gtgaaggtga tgctacttac ggtaaattga
      >.....................................yEGFPpst.....................................>

 881  ccttaaaatt tatttgtact actggtaaat tgccagttcc atggccaacc ttagtcacta ctttcggtta tggtgttcaa
      >.....................................yEGFPpst.....................................>

 961  tgttttgcga gatacccaga tcatatgaaa caacatgact ttttcaagtc tgccatgcca gaaggttatg ttcaagaaag
      >.....................................yEGFPpst.....................................>

1041  aactattttt ttcaaagatg acggtaacta caagaccaga gctgaagtca agtttgaagg tgatacctta gttaatagaa
      >.....................................yEGFPpst.....................................>

1121  tcgaattaaa aggtattgat tttaaagaag atggtaacat tttaggtcac aaattggaat acaactataa ctctcacaat
      >.....................................yEGFPpst.....................................>

1201  gtttacatca tggctgacaa acaaaagaat ggtatcaaag ttaacttcaa aattagacac aacattgaag atggttctgt
      >.....................................yEGFPpst.....................................>

1281  tcaattagct gaccattatc aacaaaatac tccaattggt gatggtccag tcttgttacc agacaaccat tacttatcca
      >.....................................yEGFPpst.....................................>

1361  ctcaatctgc cttatccaaa gatccaaacg aaaagagaga ccacatggtc ttgttagaat ttgttactgc tgctggtatt
      >.....................................yEGFPpst.....................................>

1441  acccatggta tggatgaatt ggcatccaac ttgaacattt cgagaaagct taccatatca accccatcat gctctttcga
      >.....................................yEGFPpst.....................................>
                           >>.........................Cln2-PEST.........................>

1521  aaattcaaat agcacatcca ttccttcgcc cgcttcctca tctcaaagcc acactccaat gagaaacatg agctcactct
      >.....................................yEGFPpst.....................................>
      >....................................Cln2-PEST.....................................>
```


1601 ctgataacag cgttttcagc cggaatatgg aacaatcatc accaatcact ccaagtatgt accaatttgg tcagcagcag
>.....yEGFPpst.....>
>.....Cln2-PEST.....>
1681 tcaaacagta tatgtggtag caccgttagt gtgaatagtc tggtgaatac aaataacaaa caaaggatct acgaacaaat
>.....yEGFPpst.....>
>.....Cln2-PEST.....>
1761 cacgggtcct aacagcaata acgcaaccaa tgattatatt gatttgctaa acctaaatga gtctaacaag gaaaaccaaa
>.....yEGFPpst.....>
>.....Cln2-PEST.....>
1841 atcccgcaac ggcgcattac ctcaatgggg gcccacccaa gacaagcttc attaaccatg gaatgttccc ctcgccaact
>.....yEGFPpst.....>
>.....Cln2-PEST.....>
1921 gggaccataa atagcggtaa atctagcagt gcctcatctt taatttcttt tggtatgggc aatacccaag taatataggt
>.....yEGFPpst.....>>
>.....Cln2-PEST.....>>
2001 cgacaagctt tggacttctt cgccagaggt ttggtcaagt ctccaatcaa ggttgtcggc ttgtctacct tgccagaaat
>>.....yADHt.....>
2081 ttacgaaaag atggaaaagg gtcaaatcgt tggtagatac gttgttgaca cttctaaata agcgaatttc ttatgattta
>.....yADHt.....>
2161 tgatttttat tattaaataa gttataaaaa aaataagtgt atacaaattt taaagtgact cttaggtttt aaaacgaaaa
>.....yADHt.....>
2241 ttcttgttct tgagtaactc tttcctgtag gtcaggttgc tttctcaggt atagcatgag gtcgctctta ttgaccacac
>.....yADHt.....>
2321 ctctaccggc atgccataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg
>.yADHt.>>
pBR322 ori >>.>
2401 ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc
>.....pBR322 ori.....>
2481 gacaggacta taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg
>.....pBR322 ori.....>
2561 gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcaatgc tcacgctgta ggtatctcag ttcggtgtag
>.....pBR322 ori.....>
2641 gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct
>.....pBR322 ori.....>
2721 tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag
>.....pBR322 ori.....>
2801 gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgctg
>.....pBR322 ori.....>
2881 aagccagtta ccttcggaaa aagagttggt agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt
>.....pBR322 ori.....>
2961 ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt
>.....pBR322 ori.....>>
3041 ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct agatcctttt aaattaaaaa

```
4961  gagcgctttt gaaaaccaaa agcgctctga agacgcactt tcaaaaaacc aaaaacgcac cggactgtaa cgagctacta

5041  aaatattgcg aataccgctt ccacaaacat tgctcaaaag tatctctttg ctatatatct ctgtgctata tccctatata
          >>..............................2-micron ori..............................>

5121  acctacccat ccacctttcg ctccttgaac ttgcatctaa actcgacctc tacatttttt atgtttatct ctagtattac
      >..................................2-micron ori..................................>

5201  tctttagaca aaaaaattgt agtaagaact attcatagag tgaatcgaaa acaatacgaa aatgtaaaca tttcctatac
      >..................................2-micron ori..................................>

5281  gtagtatata gagacaaaat agaagaaacc gttcataatt ttctgaccaa tgaagaatca tcaacgctat cactttctgt
      >..................................2-micron ori..................................>

5361  tcacaaagta tgcgcaatcc acatcggtat agaatataat cggggatgcc tttatcttga aaaaatgcac ccgcagcttc
      >..................................2-micron ori..................................>

3121  tgaagtttta aatcaatcta aagtatatat gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat
                                                          <<...........AmpR...........<

3201  ctcagcgatc tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg gagggcttac
      <.....................................AmpR.....................................<

3281  catctggccc cagtgctgca atgataccgc gagacccacg ctcaccggct ccagatttat cagcaataaa ccagccagcc
      <.....................................AmpR.....................................<

3361  ggaagggccg agcgcagaag tggtcctgca actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt
      <.....................................AmpR.....................................<

3441  aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc attgctgcag gcatcgtggt gtcacgctcg tcgtttggta
      <.....................................AmpR.....................................<

3521  tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc cccatgttgt gcaaaaaagc ggttagctcc
      <.....................................AmpR.....................................<

3601  ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct
      <.....................................AmpR.....................................<

3681  tactgtcatg ccatccgtaa gatgcttttc tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc
      <.....................................AmpR.....................................<

3761  gaccgagttg ctcttgcccg gcgtcaacac gggataatac cgcgccacat agcagaactt taaaagtgct catcattgga
      <.....................................AmpR.....................................<

3841  aaacgttctt cggggcgaaa actctcaagg atcttaccgc tgttgagatc cagttcgatg taacccactc gtgcacccaa
      <.....................................AmpR.....................................<

3921  ctgatcttca gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa
      <.....................................AmpR.....................................<

4001  taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat tattgaagca tttatcaggg ttattgtctc
      <.............AmpR.............<<

4081  atgagcggat acatatttga atgtatttag aaaaataaac aaataggggt tccgcgcaca tttccccgaa aagtgccacc

4161  tgacgtctaa gaaaccatta ttatcatgac attaacctat aaaaataggc gtatcacgag gccctttcgt cttcaagaat

4241  tctgaaccag tcctaaaacg agtaaatagg accggcaatt cttcaagcaa taaacaggaa taccaattat taaaagataa

4321  cttagtcaga tcgtacaata aagctttgaa gaaaaatgcg ccttattcaa tctttgctat aaaaaatggc ccaaaatctc

4401  acattggaag acatttgatg acctcatttc tttcaatgaa gggcctaacg gagttgacta atgttgtggg aaattggagc

4481  gataagcgtg cttctgccgt ggccaggaca acgtatactc atcagataac agcaatacct gatcactact tcgcactagt
```

```
5441  gctagtaatc agtaaacgcg ggaagtggag tcaggctttt tttatggaag agaaaataga caccaaagta gccttcttct
      >.....................................2-micron ori.....................................>

5521  aaccttaacg gacctacagt gcaaaaagtt atcaagagac tgcattatag agcgcacaaa ggagaaaaaa agtaatctaa
      >.....................................2-micron ori.....................................>

5601  gatgctttgt tagaaaaata gcgctctcgg gatgcatttt tgtagaacaa aaaagaagta tagattcttt gttggtaaaa
      >.....................................2-micron ori.....................................>

5681  tagcgctctc gcgttgcatt tctgttctgt aaaaatgcag ctcagattct ttgtttgaaa aattagcgct ctcgcgttgc
      >.....................................2-micron ori.....................................>

5761  atttttgttt tacaaaaatg aagcacagat tcttcgttgg taaaatagcg ctttcgcgtt gcatttctgt tctgtaaaaa
      >.....................................2-micron ori.....................................>

5841  tgcagctcag attctttgtt tgaaaaatta gcgctctcgc gttgcatttt tgttctacaa aatgaagcac agatgcttcg
      >.....................................2-micron ori.....................................>

5921  ttaacaaaga tatgctattg aagtgcaaga tggaaacgca gaaaatgaac cggggatgcg acgtgcaaga ttacctatgc
      >.....................................2-micron ori.....................................>

6001  aatagatgca atagtttctc caggaaccga aatacataca ttgtcttccg taaagcgcta gactatatat tattatacag
      >.....................................2-micron ori.....................................>

6081  gttcaaatat actatctgtt tcagggaaaa ctcccaggtt cggatgttca aaattcaatg atgggtaaca agtacgatcg
      >.....................................2-micron ori.....................................>

6161  taaatctgta aaacagtttg tcggatatta ggctgtatct cctcaaagcg tattcgaata tcattgagaa gctgcagcgt
      >.....................................2-micron ori.....................................>

6241  cacatcggat aataatgatg gcagccattg tagaagtgcc ttttgcattt ctagtctctt tctcggtcta gctagtttta
      >.....................................2-micron ori.....................................>

6321  ctacatcgcg aagatagaat cttagatcac actgcctttg ctgagctgga tcaatagagt aacaaaagag tggtaaggcc
      >.....................................2-micron ori.....................................>

6401  tcgttaaagg acaaggacct gagcggaagt gtatcgtaca gtagacggag tatactagta tagtctatag tccgtggaat
      >.....................................2-micron ori.....................................>

6481  tctcatgttt gacagcttat catcgataag cttttcaatt caattcatca ttttttttt attcttttt ttgatttcgg
      >........2-micron ori........>>

6561  tttctttgaa attttttga ttcggtaatc tccgaacaga aggaagaacg aaggaaggag cacagactta gattggtata

6641  tatacgcata tgtagtgttg aagaaacatg aaattgccca gtattcttaa cccaactgca cagaacaaaa acctgcagga

6721  aacgaagata aatcatgtcg aaagctacat ataaggaacg tgctcctact catcctagtc ctgttgctgc caagctattt

6801  aatatcatgc acgaaaagca aacaaacttg tgtgcttcat tggatgttcg taccaccaag gaattactgg agttagttga
      >.......................................URA3.......................................>

6881  agcattaggt cccaaaattt gtttactaaa aacacatgtg gatatcttga ctgatttttc catggagggc acagttaagc
      >.......................................URA3.......................................>

6961  cgctaaaggc attatccgcc aagtacaatt ttttactctt cgaagacaga aaatttgctg acattggtaa tacagtcaaa
      >.......................................URA3.......................................>

7041  ttgcagtact ctgcgggtgt atacagaata gcagaatggg cagacattac gaatgcacac ggtgtggtgg gcccaggtat
      >.......................................URA3.......................................>
```

```
7121  tgttagcggt ttgaagcagg cggcagaaga agtaacaaag gaacctagag gccttttgat gttagcagaa ttgtcatgca
      >.....URA3.....>

7201  agggctccct atctactgga gaatatacta agggtactgt tgacattgcg aagagcgaca aagattttgt tatcggcttt
      >.....URA3.....>

7281  attgctcaaa gagacatggg tggaagagat gaaggttacg attggttgat tatgacaccc ggtgtgggtt tagatgacaa
      >.....URA3.....>

7361  gggagacgca ttgggtcaac agtatagaac cgtggatgat gtggtctcta caggatctga cattattatt gttggaagag
      >.....URA3.....>

7441  gactatttgc aaagggaagg gatgctaagg tagagggtga acgttacaga aaagcaggct gggaagcata tttgagaaga
      >.....URA3.....>

7521  tgcggccagc aaaactaaaa aactgtatta taagtaaatg catgtatact aaactcacaa attagagctt caatttaatt
      >......URA3......>>

7601  atatcagtta ttacccggg
```

SAnb: Core streptavidin with negligible biotin affinity due to N23A, S27D, S45A mutations (underlined). Silent PstI restriction site (blue) added for screening.

```
  1  gac gag gcc ggc atc acc ggc acc tgg tac  211  tac cgc aac gcc cac tcc gcg acc acg tgg
 1     D   E   A   G   I   T   G   T   W   Y  71     Y   R   N   A   H   S   A   T   T   W

 31  gct cag ctc ggc gat acc ttc atc gtg act  241  agc ggc cag tac gtc ggc ggc gcc gag gcg
11     A   Q   L   G   D   T   F   I   V   T  81     S   G   Q   Y   V   G   G   A   E   A

     PstI                                     271  agg atc aac acc cag tgg ctg ctg acc tcc
 61  gca ggc gcc gac ggc gcc ctg acc gga acc  91     R   I   N   T   Q   W   L   L   T   S
21     A   G   A   D   G   A   L   T   G   T
                                              301  ggc acc acc gag gcc aac gcc tgg aag tcc
 91  tac gag gct gcc gtc ggc aac gcc gag agc 101     G   T   T   E   A   N   A   W   K   S
31     Y   E   A   A   V   G   N   A   E   S
                                              331  acg ctg gtc ggc cac gac acc ttc acc aag
121  cgc tac gtc ctg acc ggt cgt tac gac agc 111     T   L   V   G   H   D   T   F   T   K
41     R   Y   V   L   T   G   R   Y   D   S                                 EcoRI       XhoI
                                              361  gtg aag ccg tcc gcc gcc tcc gaa ttc tga ctc gag
151  gcc ccg gcc acc gac ggc agc ggc acc gcc 121     V   K   P   S   A   A   S   E   F   -
51     A   P   A   T   D   G   S   G   T   A

181  ctc ggt tgg acg gtg gcc tgg aag aat aac
61     L   G   W   T   V   A   W   K   N   N
```

$cpSA_{87}$: Circular permutation of core streptavidin beginning with E116 (between 7th and 8th β-strands). BamHI restriction site (blue) embedded in GGGS linker (blue) for screening. Mutations N23A (gct), D27A (gat), and S45A are made for $cpSAnb_{87}$ (underlined, silent PstI made but not shown). C-terminal SG added with embedded BspEI (blue) for dSA cloning.

```
     EcoRI
  1  gaa ttc gag gcc aac gcc tgg aag tcc acg
  1    E   F   E   A   N   A   W   K   S   T

 31  ctg gtc ggc cac gac acc ttc acc aag gtg
 11    L   V   G   H   D   T   F   T   K   V

                                    BamHI
 61  aag ccg tcc gcc gcc tcc ggt ggt gga tcc
                            linker
 21    K   P   S   A   A   S   G   G   G   S

 91  gac gag gcc ggc atc acc ggc acc tgg tac
 31    D   E   A   G   I   T   G   T   W   Y

121  aac cag ctc ggc tcg acc ttc atc gtg acc
 41    N   Q   L   G   S   T   F   I   V   T

151  gcg ggc gcc gac ggc gcc ctg acc gga acc
 51    A   G   A   D   G   A   L   T   G   T

181  tac gag tcg gcc gtc ggc aac gcc gag agc
 61    Y   E   S   A   V   G   N   A   E   S

211  cgc tac gtc ctg acc ggt cgt tac gac agc
 71    R   Y   V   L   T   G   R   Y   D   S

241  gcc ccg gcc acc gac ggc agc ggc acc gcc
 81    A   P   A   T   D   G   S   G   T   A

271  ctc ggt tgg acg gtg gcc tgg aag aat aac
 91    L   G   W   T   V   A   W   K   N   N

301  tac cgc aac gcc cac tcc gcg acc acg tgg
101    Y   R   N   A   H   S   A   T   T   W

331  agc ggc cag tac gtc ggc ggc gcc gag gcg
111    S   G   Q   Y   V   G   G   A   E   A

361  agg atc aac acc cag tgg ctg ctg acc tcc
121    R   I   N   T   Q   W   L   L   T   S

                          BspEI       XhoI
391  ggc acc acc tcc gga tga ctc gag
131    G   T   T   S   G   -
```

$cpSA_{54}$: Circular permutation of core streptavidin beginning with S69 (between 4th and 5th β-strands). BamHI restriction site (blue) embedded in GGGS linker (blue) for screening. Mutations N23A (gct), D27A (gat), and S45A are made for $cpSAnb_{54}$ (underlined, silent PstI made but not shown). N-terminal SG codons mutated to embed silent BspEI (blue) for dSA cloning.

```
          BspEI
     EcoRI
  1  gaa ttc tcc gga acc gcc ctc ggt tgg acg
  1    E   F   S   G   T   A   L   G   W   T

 31  gtg gcc tgg aag aat aac tac cgc aac gcc
 11    V   A   W   K   N   N   Y   R   N   A

 61  cac tcc gcg acc acg tgg agc ggc cag tac
 21    H   S   A   T   T   W   S   G   Q   Y

 91  gtc ggc ggc gcc gag gcg agg atc aac acc
 31    V   G   G   A   E   A   R   I   N   T

121  cag tgg ctg ctg acc tcc ggc acc acc gag
 41    Q   W   L   L   T   S   G   T   T   E

151  gcc aac gcc tgg aag tcc acg ctg gtc ggc
 51    A   N   A   W   K   S   T   L   V   G

181  cac gac acc ttc acc aag gtg aag ccg tcc
 61    H   D   T   F   T   K   V   K   P   S

211  gcc gcc tcc ggt ggt gga tcc gac gag gcc
                    linker
 71    A   A   S   G   G   G   S   D   E   A

241  ggc atc acc ggc acc tgg tac aac cag ctc
 81    G   I   T   G   T   W   Y   N   Q   L

271  ggc tcg acc ttc atc gtg acc gcg ggc gcc
 91    G   S   T   F   I   V   T   A   G   A

301  gac ggc gcc ctg acc gga acc tac gag tcg
101    D   G   A   L   T   G   T   Y   E   S

331  gcc gtc ggc aac gcc gag agc cgc tac gtc
111    A   V   G   N   A   E   S   R   Y   V

361  ctg acc ggt cgt tac gac agc gcc ccg gcc
121    L   T   G   R   Y   D   S   A   P   A

                                  XhoI
391  acc gac ggc tga ctc gag
131    T   D   G   -
```

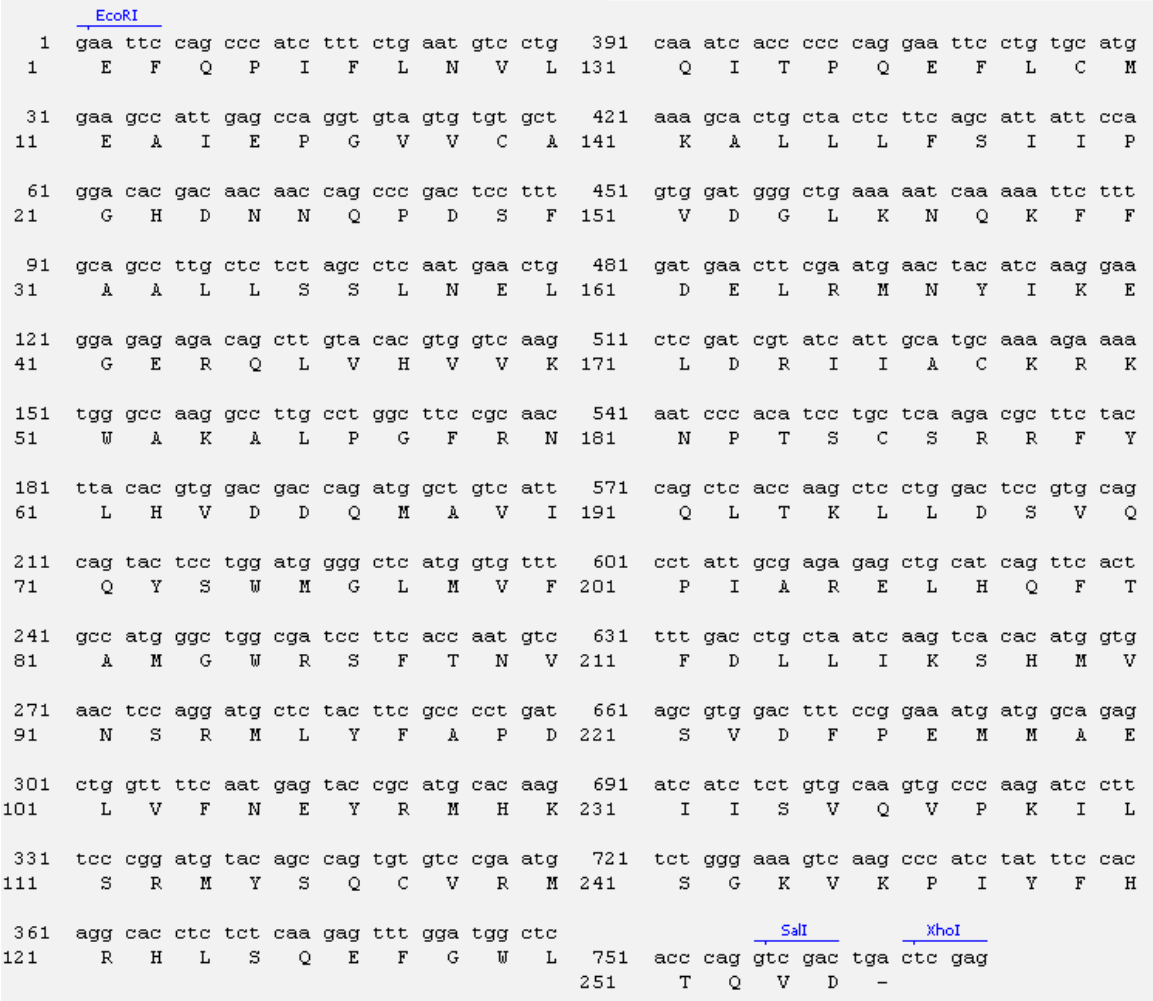
ERβ(255-509): Human estrogen receptor beta ligand binding domain (255-509) used in ERβ homodimerization yeast three-hybrid assays.
Ogawa, S. *et al*. M. *Biochem Biophys Res Commun* **1998**, *243*, 122.

```
      EcoRI
   1  gaa ttc gtg cgg gag ctg ctg ctg gac gcc   421  caa cac aaa gaa tat ctc tgt gtc aag gcc
 1      E   F   V   R   E   L   L   L   D   A  141     Q   H   K   E   Y   L   C   V   K   A

  31  ctg agc ccc gag cag cta gtg ctc acc ctc   451  atg atc ctg ctc aat tcc agt atg tac cct
11      L   S   P   E   Q   L   V   L   T   L  151     M   I   L   L   N   S   S   M   Y   P

  61  ctg gag gct gag ccg ccc cat gtg ctg atc   481  ctg gtc aca gcg acc cag gat gct gac agc
21      L   E   A   E   P   P   H   V   L   I  161     L   V   T   A   T   Q   D   A   D   S

  91  agc cgc ccc agt gcg ccc ttc acc gag gcc   511  agc cgg aag ctg gct cac ttg ctg aac gcc
31      S   R   P   S   A   P   F   T   E   A  171     S   R   K   L   A   H   L   L   N   A

 121  tcc atg atg atg tcc ctg acc aag ttg gcc   541  gtg acc gat gct ttg gtt tgg gtg att gcc
41      S   M   M   M   S   L   T   K   L   A  181     V   T   D   A   L   V   W   V   I   A

 151  gac aag gag ttg gta cac atg atc agc tgg   571  aag agc ggc atc tcc tcc cag cag caa tcc
51      D   K   E   L   V   H   M   I   S   W  191     K   S   G   I   S   S   Q   Q   Q   S

 181  gcc aag aag att ccc ggc ttt gtg gag ctc   601  atg cgc ctg gct aac ctc ctg atg ctc ctg
61      A   K   K   I   P   G   F   V   E   L  201     M   R   L   A   N   L   L   M   L   L

 211  agc ctg ttc gac caa gtg cgg ctc ttg gag   631  tcc cac gtc agg cat gcg agt aac aag ggc
71      S   L   F   D   Q   V   R   L   L   E  211     S   H   V   R   H   A   S   N   K   G

 241  agc tgt tgg atg gag gtg tta atg atg ggg   661  atg gaa cat ctg ctc aac atg aag tgc aaa
81      S   C   W   M   E   V   L   M   M   G  221     M   E   H   L   L   N   M   K   C   K

 271  ctg atg tgg cgc tct att gac cac ccc ggc   691  aat gtg gtc cca gtg tat gac ctg ctg ctg
91      L   M   W   R   S   I   D   H   P   G  231     N   V   V   P   V   Y   D   L   L   L

 301  aag ctc atc ttt gct cca gat ctt gtt ctg   721  gag atg ctg aat gcc cac gtg ctt cgc ggg
101     K   L   I   F   A   P   D   L   V   L  241     E   M   L   N   A   H   V   L   R   G

 331  gac agg gat gag ggg aaa tgc gta gaa ggg                                  SalI
111     D   R   D   E   G   K   C   V   E   G   751  tgc aag tcc tcc atc acg ggg gtc gac tga
                                                251     C   K   S   S   I   T   G   V   D   -
 361  att ctg gaa atc ttt gac atg ctc ctg gca
121     I   L   E   I   F   D   M   L   L   A          XhoI
                                                781  ctc gag act ggc
 391  act act tca agg ttt cga gag tta aaa ctc   261
131     T   T   S   R   F   R   E   L   K   L
```

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.
Misrahi, M *et al*. *Biochem Biophys Res Commun* **1987**, *143*, 740.

```
      EcoRI
      ┌──┐
    1 gaa ttc cag ttg att cca cca ctg atc aac
        E   F   Q   L   I   P   P   L   I   N

   31 ctg tta atg agc att gaa cca gat gtg atc
        L   L   M   S   I   E   P   D   V   I

   61 tat gca gga cat gac aac aca aaa cct gac
        Y   A   G   H   D   N   T   K   P   D

   91 acc tcc agt tct ttg ctg aca agt ctt aat
        T   S   S   S   L   L   T   S   L   N

  121 caa cta ggc gag agg caa ctt ctt tca gta
        Q   L   G   E   R   Q   L   L   S   V

  151 gtc aag tgg tct aaa tca ttg cca ggt ttt
        V   K   W   S   K   S   L   P   G   F

  181 cga aac tta cat att gat gac cag ata act
        R   N   L   H   I   D   D   Q   I   T

  211 ctc att cag tat tct tgg atg agc tta atg
        L   I   Q   Y   S   W   M   S   L   M

  241 gtg ttt ggt cta gga tgg aga tcc tac aaa
        V   F   G   L   G   W   R   S   Y   K

  271 cac gtc agt ggg cag atg ctg tat ttt gca
        H   V   S   G   Q   M   L   Y   F   A

  301 cct gat cta ata cta aat gaa cag cgg atg
        P   D   L   I   L   N   E   Q   R   M

  331 aaa gaa tca tca ttc tat tca tta tgc ctt
        K   E   S   S   F   Y   S   L   C   L

  361 acc atg tgg cag atc cca cag gag ttt gtc
        T   M   W   Q   I   P   Q   E   F   V

  391 aag ctt caa gtt agc caa gaa gag ttc ctc
        K   L   Q   V   S   Q   E   E   F   L

  421 tgt atg aaa gta ttg tta ctt ctt aat aca
        C   M   K   V   L   L   L   L   N   T

  451 att cct ttg gaa ggg cta cga agt caa acc
        I   P   L   E   G   L   R   S   Q   T

  481 cag ttt gag gag atg agg tca agc tac att
        Q   F   E   E   M   R   S   S   Y   I

  511 aga gag ctc atc aag gca att ggt ttg agg
        R   E   L   I   K   A   I   G   L   R

  541 caa aaa gga gtt gtg tcg agc tca cag cgt
        Q   K   G   V   V   S   S   S   Q   R

  571 ttc tat caa ctt aca aaa ctt ctt gat aac
        F   Y   Q   L   T   K   L   L   D   N

  601 ttg cat gat ctt gtc aaa caa ctt cat ctg
        L   H   D   L   V   K   Q   L   H   L

  631 tac tgc ttg aat aca ttt atc cag tcc cgg
        Y   C   L   N   T   F   I   Q   S   R

  661 gca ctg agt gtt gaa ttt tca gaa atg atg
                                P906S
        A   L   S   V   E   F   S   E   M   M

  691 tct gaa gtt att gct gca caa tta ccc aag
        S   E   V   I   A   A   Q   L   P   K

  721 ata ttg gca ggg atg gtg aaa ccc ctt ctc
        I   L   A   G   M   V   K   P   L   L

                        SalI        XhoI
                        ┌──┐        ┌──┐
  751 ttt cat aaa aag gtc gac tga ctc gag act
        F   H   K   K   V   D   -

  781 ccg
```

AR(670-919): Human androgen receptor ligand binding domain (670-919) used in AR homodimerization yeast three-hybrid assays.
Laitinen, O. H. *et al*. *Trends Biotechnol* **2007**, *25*, 269.




```
      EcoRI
   1  gaa ttc cag ccc atc ttt ctg aat gtc ctg   391  caa atc acc ccc cag gaa ttc ctg tgc atg
 1     E   F   Q   P   I   F   L   N   V   L 131     Q   I   T   P   Q   E   F   L   C   M

  31  gaa gcc att gag cca ggt gta gtg tgt gct   421  aaa gca ctg cta ctc ttc agc att att cca
11     E   A   I   E   P   G   V   V   C   A 141     K   A   L   L   L   F   S   I   I   P

  61  gga cac gac aac aac cag ccc gac tcc ttt   451  gtg gat ggg ctg aaa aat caa aaa ttc ttt
21     G   H   D   N   N   Q   P   D   S   F 151     V   D   G   L   K   N   Q   K   F   F

  91  gca gcc ttg ctc tct agc ctc aat gaa ctg   481  gat gaa ctt cga atg aac tac atc aag gaa
31     A   A   L   L   S   S   L   N   E   L 161     D   E   L   R   M   N   Y   I   K   E

 121  gga gag aga cag ctt gta cac gtg gtc aag   511  ctc gat cgt atc att gca tgc aaa aga aaa
41     G   E   R   Q   L   V   H   V   V   K 171     L   D   R   I   I   A   C   K   R   K

 151  tgg gcc aag gcc ttg cct ggc ttc cgc aac   541  aat ccc aca tcc tgc tca aga cgc ttc tac
51     W   A   K   A   L   P   G   F   R   N 181     N   P   T   S   C   S   R   R   F   Y

 181  tta cac gtg gac gac cag atg gct gtc att   571  cag ctc acc aag ctc ctg gac tcc gtg cag
61     L   H   V   D   D   Q   M   A   V   I 191     Q   L   T   K   L   L   D   S   V   Q

 211  cag tac tcc tgg atg ggg ctc atg gtg ttt   601  cct att gcg aga gag ctg cat cag ttc act
71     Q   Y   S   W   M   G   L   M   V   F 201     P   I   A   R   E   L   H   Q   F   T

 241  gcc atg ggc tgg cga tcc ttc acc aat gtc   631  ttt gac ctg cta atc aag tca cac atg gtg
81     A   M   G   W   R   S   F   T   N   V 211     F   D   L   L   I   K   S   H   M   V

 271  aac tcc agg atg ctc tac ttc gcc cct gat   661  agc gtg gac ttt ccg gaa atg atg gca gag
91     N   S   R   M   L   Y   F   A   P   D 221     S   V   D   F   P   E   M   M   A   E

 301  ctg gtt ttc aat gag tac cgc atg cac aag   691  atc atc tct gtg caa gtg ccc aag atc ctt
101    L   V   F   N   E   Y   R   M   H   K 231     I   I   S   V   Q   V   P   K   I   L

 331  tcc cgg atg tac agc cag tgt gtc cga atg   721  tct ggg aaa gtc aag ccc atc tat ttc cac
111    S   R   M   Y   S   Q   C   V   R   M 241     S   G   K   V   K   P   I   Y   F   H

 361  agg cac ctc tct caa gag ttt gga tgg ctc                 SalI       XhoI
121    R   H   L   S   Q   E   F   G   W   L   751  acc cag gtc gac tga ctc gag
                                               251     T   Q   V   D   -
```

APPENDIX C. Microbial Stains

Name	Cell Type	Source	Genotype	Comments
DH5α	*E. coli*	Invitrogen	F- F50d*lac*ZDM15 D(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(rk-mk+) *pho*|A supE44 l-thi-1 *gyr*A96 *rel*A1	General subcloning
FY250	*S. cerevisiae*	M. Ptashne	*MATα, ura3-52, hisD200, leu2D1, trp1D63*	General yeast strain
FY250yEGFPpst	*S. cerevisiae*	K. Bailey	*MATα, ura3-52, URA3::yEGFPpst, hisD200, leu2D1, trp1D63*	yEGFPpst integrated reporter strain