**The Dissertation Committee for Jimmy Dale Gollihar Jr. Certifies that this is the approved version of the following dissertation:**

# Methods in protein engineering and screening: From rational design to directed evolution and beyond

**Committee:**

Andrew D Ellington, Supervisor

Jeffrey Barrick

George Georgiou

Edward Marcotte

Boris Zemelman

# Methods in protein engineering and screening: From rational design to directed evolution and beyond

**by**

**Jimmy Dale Gollihar Jr., BS**

**Dissertation**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas at Austin**

**August 2017**

# Methods in protein engineering and screening methodologies: From rational design to directed evolution and beyond

Jimmy Dale Gollihar Jr., PhD

The University of Texas at Austin, 2017

Supervisor: Andrew D Ellington

Humans have engineered organisms for thousands of years. The domestication of plants and animals are early examples of how humans have changed genotypes of organisms for prized traits that persist today. The genotype to phenotype linkage is often the result of the expression or performance of proteins found within an organism. The modern era of biological engineering has seen the emergence of two fundamentally distinct philosophies on how best to alter the genetic code of an organism or sequence of a particular protein for desired phenotypic outcomes. Rational design and directed evolution are often seen as two competing methodologies working towards the same goal—the engineering of proteins and ultimately organisms. Here, I attempt to bridge the gap between the two engineering strategies and argue that a combination is needed in many cases. Starting with the rationalization of directed evolution experiments using software tools, we begin to understand the predictive power and limitations of rational computational approaches. We then explore the computational resurfacing of an enzyme and discover an unpredicted thermoswitchable phenotype. Next, we rationally design a number of fusion proteins, which allow us to create novel point-of-care diagnostics. We then turn our attention to the directed evolution of a number of DNA polymerases with novel functions. The amalgamation of rational design and directed evolution are then

extended to a eukaryotic organism, which enables us to engineer more complex proteins and gives us the ability to create novel drug screening and discovery platforms.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Current methods in protein engineering

"Slow though the process of selection may be, if feeble man can do much by his powers of artificial selection, I can see no limit to the amount of change, to the beauty and infinite complexity of the co-adaptations between all organic beings, one with another and with their physical conditions of life, which may be effected in the long course of time by nature's power of selection."

— Charles Darwin, *On the Origin of Species*

## INTRODUCTION

Natural selection is perhaps the most creative force known to humankind. For over a billion years, natural selection has acted on replication systems and organisms to give rise to the enormous diversity of life observed in extant and extinct organisms. Life on Earth is teeming, with organisms found in the most extreme environments—from thermal vents to glacial ice. The plasticity of life is profound when one considers the range of ecological niche spaces inhabited by Earth's creatures between these extremes. Regulatory and metabolic processes have been fine-tuned for these varying environments by genetic mutations that are passed from generation to generation. Adaptation to new environments often requires rare genetic mutations that allow variants of organisms to thrive in new and dynamic environments.

Similarly, humans have crafted variations of organisms on much faster timescales than natural selection by selectively breeding plants and animals. For example, the domestication of the dog is estimated to have happened over 10,000 years ago. The rich diversity seen in *Cannus familiaris* is largely attributed to the phenotypic selection by humans (Vilà et al., 1997, 1999), historically referred to as 'artificial selection'. Plants and other animals have been domesticated for a variety of reasons. The breeding of organisms for human consumption (crops and livestock) is

thought to have given rise to what we now consider modern civilization. The plants and animals consumed by humans today share little resemblance to their wild counterparts (Diamond, 2002).

Whether 'natural' or 'artificial', selection occurs at the level of the genes found within organisms (Dawkins, 2006). From a reductionist perspective, genotypes lead to phenotypes. This is a basic tenet of biology. The flow of information is largely unidirectional; DNA, the storage molecule, is transcribed to RNA, which is then translated into proteins (Mahmoudian, 2010). In short, genes encode the directions to make proteins. With minor exceptions, proteins then give rise to the phenotypes we observe. Taken together these facts suggest that selection often acts on the production or function of proteins.

Proteins make up a diverse set of macromolecules. By combining individual amino acids, the ribosome creates polypeptides which then fold and assemble into mature proteins. Proteins then carry out much of life's work, from catalysis to structural integrity to regulation (Eisenberg et al., 2000). Signaling is often the work of receptors and transcription factors, which bind to both extra- and intracellular molecules. Antibodies bind foreign objects and protect a host organism from disease or invasion. Enzymes are the chemists, carrying out the reactions needed to sustain life and keep an organism from reaching equilibrium (or death). The structure of cells and the movement of cargo between compartments is the work of a vast array of different proteins. The examples listed above are but a small fraction of the many functions proteins possess.

In the modern era, biological engineers attempt to modify or design proteins for a variety of applications. Today engineered proteins are used in: (1) agriculture and the food industry, (2) detergents, (3) diagnostic and pharmaceutical industries, (4) biopolymer production and nanotechnology, (5) environmental monitoring and bioremediation, (6) chemical syntheses of complex molecules, and (7) biofuel production—to name but a few (Turanli-Yildiz et al., 2012).

The goal of modern protein engineering is to improve or impart function on existing or novel protein scaffolds (Brannigan and Wilkinson, 2002). Protein engineering is a relatively young field in that its arrival was contingent upon the advent of recombinant DNA technologies and molecular cloning, collectively referred to as the Molecular Biology Revolution of the early 1980s. Historically, two distinct, but not mutually exclusive, strategies have been employed to engineer

proteins: (1) rational design and (2) directed evolution. Rational design of a protein requires that something is known of the structure, function, and mechanism of the target protein. Mutations to enhance or change the activity of the protein are made from prior structural and activity measurements (Cedrone et al., 2000). On the other hand, directed evolution does not require mechanistic knowledge of the protein to be engineered. Rather, it is limited by the throughput of screening or selection of new variants (Tobin et al., 2000). Below, we discuss the advances in each field and highlight their uses in modern interdisciplinary approaches.

## RATIONAL DESIGN

### Structure, function, and intuition

It has long been recognized that the function of a protein is predicated upon its properly folded 3-dimensional structure. This so-called 'structure-function relationship' is at the heart of modern molecular biology and biochemistry. How exactly structure informs function from the primary sequence of a protein can be probed exhaustively through the use modern computational and classical biochemical methods. With structural data from X-ray crystallography or NMR spectra, hypotheses about mechanism can be probed in exquisite detail.

The first two enzymatic activities examined in this fashion were a tyrosine aminoacyl-tRNA synthetase (TyrRS) and a β-lactamase in 1982 (Brannigan and Wilkinson, 2002). While these early experiments did not produce a gain of function, they were used to probe mechanistic questions that could not otherwise be answered by enzymologists at the time and laid the foundation for what would later evolve into the field of rational protein design. When combined with prediction, the construction of modified proteins based on their structures and biochemical characterization form the basis of rational design (Leatherbarrow et al., 1985). In a somewhat prophetic call to arms in as early as 1983, Ulmer perfectly anticipated what would become rational design and predicted the interplay with directed evolution techniques to form the field of modern protein engineering (Ulmer, 1983).

At the time, very little structural data were available for diverse protein families. The first membrane protein structure wasn't reported until 1985, and only 186 unique protein structures were reported (Berman et al., 2000). As late as 1991, only 648 structures were publicly available. Despite a significant lack of structural data from which to make predictions, rational protein design saw some early success. For example, from structural data and alanine scanning, human growth hormone variants were rationally designed to increase binding preference to either the growth hormone or prolactin receptors by factors of 34,000 and 150, respectively (Cunningham and Wells, 1991).

A number of examples of utilizing rational design to increase the thermostability of proteins have been reported. For example, the thermostability of α-amylase was increased with the replacement of asparagine sidechains by one group (Declerck et al., 2000). Another group introduced prolines to stabilize structural elements of the protein (Igarashi et al., 1999). Proline substitution was also successful in stabilizing d-xylose isomerase by introducing the amino acid into the turn of a random coil (Zhu et al., 1999). The engineering of two disulfide bonds in hen lysozyme greatly improved its thermotolerance (Ueda et al., 2000). Rational design of increased hydrophobic contacts stabilized 3-isopropylmalate dehydrogenase (Akanuma et al., 1999) and formate dehydrogenase (Rojkova et al., 1999). Still others have used sequence and structural information from thermophilic proteins to rationally design their mesophilic counterparts for increased stability (Kumar et al., 2000).

Beyond thermostability, other properties of proteins have been rationally designed. Oxidation-resistant proteins have been engineered through the removal of targeted sulfur containing sidechains—methionine and cysteine (Chelius et al., 2000; Ju et al., 2000; Slusarczyk et al., 2000). Using structural comparisons of two FAD-dependent oxidoreductases, van Berkel was able to switch the entantioselectivity of an (R)-specific vanillyl-alcohol oxidase to (S)-specific hydroxylation of 4-ethylphenol (Heuvel et al., 2000). Other examples of altering stereoselectivity through rational design involve the engineering of lipases (Scheib et al., 1998, 1999).

Often, enzymes require multiple steps along the reaction coordinate and produce any number of intermediates or additional side products that do not result in final product formation. In

order to increase the carboligase side product of pyruvate decarboxylase, a tryptophan residue in the substrate channel was mutated to either methionine or isoleucine, which increased the production of the carboligase reaction by nearly 5-fold (Pohl, 2000). PepC, an aminopeptidase, was transformed into an oligopeptidase by removing four carboxy-terminus amino acids, which were previously shown to interact with the active site (Mata et al., 1999). P450 enzymes are generally promiscuous (Guo et al., 2016) in nature. Efforts to extend their substrate range have proven successful using rational design. Variants of Cyp101 from *Pseudomonas putida* were engineered to contain two mutations at positions F87 and Y96 in a substrate-binding channel that conferred activity to polycyclic aromatic hydrocarbons (Harford-Cross et al., 2000). β-lactamases have also been extensively engineered to alter their substrate specificities. PER-1 was rationally designed to improve its activity with a number of cephalosporins (Bouthors et al., 1999).

New enzymatic activities have also been engineered. A tyrosine phenol-lyase was transformed into a dicarboxylic amino acid β-lyase, of which no known form exists in nature (Mouratou et al., 1999). In a landmark study two decades ago, Shokat and colleagues (Kenski et al., 2005) used the structural information of a protein kinase to create an orthogonal kinase and substrate to aid in identifying phosphorylated target proteins.

This list of rational design achievements described above is by no means comprehensive. Rational design of antibody fusions and CDR grafting are examples that deserve consideration (Jones et al., 1986; Liu et al., 2016; Sharma and Bagshawe, 2017). Changes in mechanism, substrate promiscuity, cofactor dependence, stability, and enantioselectivity have all been reported for a number of proteins using rational site-directed mutagenesis. See the references listed here for a full review (Cedrone et al., 2000; Harris and Craik, 1998; Jones and DeSantis, 1999).

**Computational protein design**

A subgenre in rational design is pure computational design of proteins for enhanced or novel function. Rather than give an exhaustive list of software packages that are used in the design process, this section will focus on the core components needed, the general framework used to

computationally design proteins, and examples of their usage. Computational protein design is born from the idea that we can understand and engineer structures for desired functions from first principles or approximations (Pantazes et al., 2011). Natural proteins are strings of amino acids, which tuned by evolutionary forces, fold into functional molecular entities. These proteins can range in size from tens to thousands of amino acid residues, each with varying degrees of freedom. The number of conformations a single sequence can adopt is enormous. The energy landscape a protein must traverse in order to achieve its mature folded state is complex. However, the native state must be sufficiently low in energy to thwart falling into local misfolded minima (Lippow and Tidor, 2007).

Due to the exponential increase of conformational and sequence space with each additional amino acid in a protein sequence ($20^N$), computational protein design is necessarily built on a number of assumptions and in many cases semi-empirical design rules or empirical data sets. A typical computer protein design protocol can be broken down into the following: (1) target objectives, (2) parametrization and component consideration, (3) the energy function, (4) search and sampling methods, and finally (5) the output of sequence(s) and 3-dimensional structure(s) (Samish et al., 2011).

From the outset, one must define what the program is targeting. This can range from a portion of a protein to the complete *de novo* design of a novel protein. In any case, any one or a combination of structure, function, or biophysical properties define the target objectives. For example, semi-artificial enzymes were previously designed to carry out the Kemp elimination reaction. To do so, quantum mechanical calculations of the ideal transition state were performed and amino acids were computationally inserted templated protein scaffolds to structurally stabilize the idealized state. While the rate enhancement without further engineering was a modest improvement over the uncatalyzed reaction, the artificial enzyme was functional and capable of performing multiple turnovers (Röthlisberger et al., 2008).

The degrees of freedom the backbone and sidechains are given is of particular importance when designing protein design protocols. While fixed backbone design strategies largely mitigate large conformational sequence space, they do not allow for main chain adjustments that may be

optimal for a targeted objective. Often proteins require some level of flexibility in backbone conformations for stability, activity, and solubility (Samish et al., 2011). For sidechain conformations, rotamer libraries are often employed, which dramatically reduce search space. Historically, these libraries were built from idealized bond lengths and dihedral angles derived from X-ray crystal structures. Two commonly used libraries reduce the degrees of freedom to 350 or less, dramatically decreasing the computational load when searching conformational space (Dunbrack, 2002; Dunbrack and Karplus, 1993; Lovell et al., 2000). However, more than 10% of residues in high-resolution structures are unable to be matched using these commonly used rotamer libraries. More recently, rotamer libraries and conformer libraries (bond lengths) have been constructed that more closely match empirical data from empircal structures. With the advent of cheap processing power, the need for smaller libraries has waned (Lassila et al., 2006; Porebski et al., 2016; Subramaniam and Senes, 2012).

Models are judged on scores derived from energy functions. While these can vary, the core energy function attempts to quantify energetic contributions from: Van der Waals interactions, electrostatics, solvation, hydrogen bonding, conformational strain, and a number of additional terms depending on the application. The potentials are generally parameterized in a way that naturally occurring protein structures represent local energy minima. Forcefields account for the parameter sets used in the potential energy calculations of atoms within the system. AMBER, CHARMM, and GROMOS are the most commonly used all-atom forcefields (Jo et al., 2017; Smith et al., 2015; Wang et al., 2017). Next, using negative design rules and folding criteria, sequences and structures are searched using a range of parameters. Sequence searches are initiated with either stochastic or deterministic approaches. For stochastic search methods, Monte Carlo methods employing a Boltzmann distribution are commonly used. Once the search criteria have been met, a number of output sequences and models are produced (Pantazes et al., 2011). These designs are then tested empirically using standard biochemical techniques.

Despite the enormous challenge, computational protein design has garnered success over the last 30+ years. An early win for the field included the computational design of the tightest-binding peptide (17 amino acid) inhibitor of calmodulin (DeGrado et al., 1985). In 1991, active

sites of predefined geometry could be designed using Hellinga's DEZYMER software package. A copper-binding site was successfully introduced into thioredoxin with the substitution of four amino acids predicted by the software (Hellinga et al., 1991). In 2003, the Baker lab designed the first α/β protein called Top7. This was the first novel fold of a protein designed and shown to possess angstrom-level accuracy (1.2 Å RMSD) (Kuhlman et al., 2003). More recently, a conglomerate of laboratories used computational protein design to engineer pseudo-viral epitopes for vaccine development. The designed epitopes were able to induce neutralizing antibody production, and one of the designs was bound by an existing antibody (Correia et al., 2014). The ability to engineer novel proteins or enhance existing proteins through computational design is gaining traction and will only improve with better tools and more computational power.

## DIRECTED EVOLUTION

### Introduction

Directed evolution of proteins attempts to mimic natural selection on much shorter timescales and has given rise to a number of therapeutically and industrially significant proteins. In contrast to rational design, detailed structure-function relationships are not needed in advance to screening or selection (Tobin et al., 2000). Through simulated Darwinian evolution, antibodies, enzymes, receptors, transcription factors, and components of biosynthetic pathways have all been crafted in the laboratory in practical time frames. In addition to functional or host optimization, entirely new properties and chemistries have been imparted on a number of proteins (Packer and Liu, 2015). The two main pillars of directed evolution are the generation of diversity and methods to select or screen variants for improved or novel function. Below we describe the current techniques in library design and selection.

### Creating library diversity

There are many ways to generate genetic diversity. However, each of the techniques falls into one of three main categories: (1) random mutagenesis, (2) focused mutagenesis, and (3)

recombination or sexual reproduction. Random mutagenesis can be achieved in a number of ways. First, chemical agents that promote random mutations during replication can be added supplied to an organism. The use of this technique in living organisms has declined in recent years because of the generation of off-target mutagenesis in the host or plasmid. Typically, random mutagenesis is achieved with 'mutagenic' or 'error-prone' polymerase chain reaction (PCR) amplification of the targeted gene (Cadwell and Joyce, 1994). DNA polymerases can be easily tuned to predefined error-rates with the addition of metals and base analogs to the PCR reaction (Wong et al., 2004). Other methods involving mutator strains exist, but they suffer from the same off-target effects mentioned above. Focused or targeted libraries are constructed through the use of degenerate codons during traditional oligonucleotide synthesis or more defined through the use of chip-based DNA synthesis (Tang et al., 2014). Here multiple positions can be mutated in a site-specific manner. This technique is particularly useful if some structure-function relationships are known. Finally, random recombination of related genes can be used in 'sexual' selection of protein variants. The most common variation of recombinatorial generation of library diversity is 'DNA Shuffling' (Stemmer, 1994). Through the use of DNAse- or sonication-fragmentation of DNA sequences, combinations of related genes can be randomly put together in novel ways. This technique resembles crossing over of chromosomes in sexually reproducing organisms. Variations on this theme exist. See RACHITT (Coco et al., 2001), NExT (Müller et al., 2005), StEP (Zhao et al., 1998), assembly PCR (Stemmer et al., 1995), ITCHY (Ness et al., 2002), and NRR (Zha et al., 2003). Over the course of a directed evolution experiment, it is common to use a combination of any the library generation methods mentioned.

## Screening and selecting mutant libraries

### *Traditional techniques*

After the library has been constructed, it must then be screened or selected in order to amplify variants of improved or novel function. Careful considerations must be made when designing screening or selections methods. Screens are typically performed by spatially separating

variants. By spatially separating individual members of the library, the coveted genotype to phenotype linkage is preserved. Often, protein variants are expressed recombinantly in model organisms and isolated on either solid media or individual culture wells (Forsburg, 2001). For bacterial screens, *Escherichia coli* is commonly used. *Saccharoymyces cerevisiae* is the preferred eukaryotic screening organism, although other yeasts, mammalian, and insect cells have been used (Singh et al., 2016). Screening is compatible with a several assay platforms where ~$10^3$-$10^4$ mutants can be measured in parallel or temporally close together. Commonly, genetic reporters that result in fluorescence (GFP), colorimetry (β-gal), or luminescence (luciferase) are used to screen functional variants. However, medium to highthroughput screens are compatible with NMR, HPLC, GC-MS, and LC-MS (Packer and Liu, 2015).

The most common and accessible screening platform is fluorescence-activated cell sorting (FACS). Fluorescence of individual cells is monitored, and those cells exhibiting fluorescence above an arbitrarily selected value or sorted and allowed to grow. Typically, enrichment factors of 10-200 are reported—more commonly near 20-fold enrichment (Aharoni et al., 2005). Yeast-display is another FACS-based technique that can be used to engineer protein-protein interactions as wells as tune enzymatic activities. Recently, yeast-display has been used extensively in the selection of antibodies (Wang et al., 2016a). By attaching a fluorophore to a target antigen, cells that display antibodies specific to a particular antigen are enriched.

The drawbacks to screening methods listed above are time and throughput. In traditional selections, libraries of over $10^8$ can be selected from at once. During selection, larger sequence space is sampled, and there is no requirement for spatially or temporally separating library members because selections are performed in bulk. The genotype to phenotype linkage can be preserved in many ways. Historically, the most common selection strategies rely on binding affinity or host survival. Affinity-based selections rely on a target protein to be displayed on surfaces of cells, phage (McCafferty et al., 1990), mRNA (Wilson et al., 2001), or ribosomes (Amstutz et al., 2002). Phage display is perhaps the most common and has been used to select for a number of protein and small-molecule binding proteins, most notably antibodies. In this regime, survivors are selected by passing library members over immobilized target reagents. Those library members that

do not bind are washed away. Binders are collected and propagated to the next round of mutagenesis and selection. Still other selections are coupled to survival of the host organism. A protein variant that confers a growth advantage or survival of its host under strong selective pressure (auxotrophic or antibiotic) is selected for and propagated. The linkage of genotype to phenotype is maintained in the context of the host cell expressing the variant that led to an increase in fitness. While a number of life/death selections have been successful, this selection strategy is limited. Often, cheaters arise in populations of these selections because there is a strong selective pressure for the organism to mutate something outside of the selection circuit in order to survive to the next round. This selection strategy can not be used for the evolution of toxic proteins or protein function under harmful conditions to host fitness (Packer and Liu, 2015; Tobin et al., 2000).

***In vitro compartmentalization***

Genotype to phenotype linkage can be maintained through the use of *in vitro* compartmentalization (IVC). Using this strategy, selection is performed within water-in-oil emulsions that are physically separated from one another. After assaying variants, the emulsions are broken, and DNA corresponding to survivors or superior performers is amplified for the next round of selection. In many cases, IVC removes the need to express protein variants in the context of cells, negating toxicity issues noted above. Another key feature is that substrates can be compartmentalized that would otherwise be inaccessible to living cells. This method was pioneered by Tawkfik in the engineering of a methyltransferase (Tawfik and Griffiths, 1998). Other enzymes have been engineered using IVC. For example, β-galactosidase activity has been imparted on a protein with unknown function.

In another instantiation of IVC, Holliger designed selections for the engineering of thermostable DNA polymerases, termed Compartmentalized Self-Replication (Ghadessy et al., 2001a). In CSR, DNA polymerase variants are expressed within bacterial cells and isolated within water-in-oil emulsions in PCR buffer with primers specific for the polymerase. Successful amplification of the target polymerase is only achieved by the function of a variant polymerase. In

a typical water-in-oil emulsion, up to $10^{10}$ compartments are formed. Thus, a library of $10^8$ can be readily selected from using a standard thermalcycler. CSR has been used to engineer novel properties of DNA polymerases (Ghadessy and Holliger, 2007) and is the basis for Chapter 5 of this dissertation.

Compartmentalized Partnered Replication, a directed evolution strategy developed in our laboratory, is the natural progression from CSR. Instead of selecting for polymerase variants in the emulsion PCR, CPR selects for biomolecules that lead to the production of the thermostable DNA polymerase (Ellefson et al., 2014). Selection circuits are generated such that the function of one biomolecule is tied to the production of the replicative DNA polymerase to be used in the emulsion PCR. The production of the 'partnered' gene and subsequent *in vivo* stage is short-lived in CPR. After induction of the partnered gene, its function is directly tied to the production of the thermostable DNA polymerase. That is, the more active a particular variant is, the more represented it will become after the emulsion PCR because more polymerase production leads to more amplification. In contrast to FACS-based screens, CPR enjoys enrichment factors in the $10^3$ to $10^4$ range due to the exponential amplification inherent to PCR (Maranhao and Ellington, 2017). Additionally, genetic circuits that lead to fluorescent proteins in traditional screens are easily converted to CPR circuits by replacing the fluorescent protein with a thermostable DNA polymerase, such as Taq or KOD. To date, CPR has been used to evolve: a panel of orthogonal T7 RNA polymerases (Meyer et al., 2014a), aminoacyl tRNA synthetase:tRNA pairs for unnatural amino acid incorportation (Ellefson et al., 2014; Maranhao and Ellington, 2017), repressors, and transcription factors. CPR forms the basis of Chapters 6 and 7 where we introduce CPR circuits into eukaryotic organisms for the first time.

***Continuous evolution***

Continuous evolution strategies attempt to break the laborious cycle of discrete library diversification and selection or screening. Many of these strategies rely on the fitness of an organism under continual dilution conditions. This technique has been successful in the creation of

improved antibiotic resistance markers and the evolution of chorismate mutase. Care must be taken that only the target gene is mutated, which is difficult to maintain in bacterial or yeast cells. Recently, a continuous evolution scheme has been adapted to phage. Phage-assisted continuous evolution (PACE) has been used to change the promoter preference of the T7 RNA polymerase. Other examples have used PACE to evolve proteases and DNA binding proteins (Packer and Liu, 2015). While limited success has been reported for continuous evolution, the adoption of these techniques has been slow due to the problematic nature of such selections—cheaters and contamination.

**Conclusions**

Up to this point, we have treated rational design and directed evolution as separate entities. In reality, it is now rare for successful protein engineers to employ only one of these strategies. Protein engineering is no longer binary in nature and is best thought to operate along a continuum of these complementary fields. While many labs possess more expertise in either of the two methodologies, optimal answers are seldom (if ever) found by utilizing one approach over the other. Recent years have seen the explosion of crystal structures, which now number in the tens of thousands. These structural data combined with bioinformatics inform both directed evolution and rational design. The advent of increased processing power, improved forcefields, and machine learning algorithms have increased the bandwidth that protein engineers can operate within. The next 35 years of protein engineering will see an unprecedented level of success in answering the challenges of modern society. The words of Ulmer and Darwin will echo along this path.

# Chapter 2: Molecular Modeling of Laboratory Evolution Experiments and Protein:Small Molecule Interactions

**ABSTRACT**

This chapter is the culmination of several successful collaborations during my tenure as a graduate student. The main theme presented in this chapter is molecular modeling and how computational tools can be applied to the *post facto* rationalization of laboratory evolution experiments (Sections 1-3). Examples of the predictive power of modern computational approaches are also discussed. In section one, we find that mutations observed in Citrate Synthase (CS) during the Lenski long-term evolution experiment (LTEE) occurred both before and after *Escherichia coli* evolved to utilize citrate aerobically as a carbon source. Detailed kinetic, molecular modeling, and metabolic flux analyses provide support of a model of evolution where mutations in CS were needed to first potentiate and then later refine central metabolic regulatory networks for the use of the newly available nutrient. In section two, we addict a diverse set of bacteria to noncanonical amino acids through engineering TEM-1 β-lactamase. The creation of rotamer libraries and simulations provide mechanistic clues to the addiction of the enzyme to two tyrosine analogs and may provide a framework for future engineering efforts of other enzymes. Extending upon our noncanonical amino acid work, section three rationalizes rifampicin resistance in *Escherichia coli* in the context of standard and extended genetic codes. Finally, section four uses homology modeling, molecular dynamics simulations, and docking studies in an attempt to explain differences in thiabendazole (TBZ) resistance in human and fungal β-tubulins. While covering seemingly disparate topics, the training in the use of various software packages provided me with a diverse toolkit that would aid in my development as a scientist and taught me to appreciate structure-function relationships at a much deeper level. Lessons learned in these four sections would help shape projects described hereafter.

## 2.1 FINE-TUNING CITRATE SYNTHASE FLUX POTENTIATES AND REFINES METABOLIC INNOVATION IN THE LENSKI EVOLUTION EXPERIMENT

### Introduction

Natural selection has provided the world with a vast array of diverse life forms, each with complex regulatory mechanisms and metabolic processes. Metabolism, life's chief use of chemistry, has been brilliantly crafted by evolution to enable organisms the ability to occupy niches over a staggering range of environments. The ability to adapt to a changing world, encoded by rare genetic mutations, gives rise to varying levels of fitness, which then in turn, leads to expansion of inherited traits and observable phenotypes (Ryall et al., 2012). In nature, this process rarely happens on observable timescales (Barrick and Lenski, 2013). However, the Lenski long-term evolution experiment (LTEE) has provided key insights into how metabolic and regulatory networks have evolved in *Escherichia coli* under laboratory conditions and gives clues as to how evolution works in nature. The LTEE has been an ongoing endeavor for more than 25 years and continues to enrich our understanding of evolutionary forces and provides us with a dataset from which to design and test new hypotheses (Lenski and Travisano, 1994; Lenski et al., 1991).

The ability of a single population to aerobically utilize citrate as a carbon source is perhaps the most widely regarded result of the LTEE to date. After roughly 31,500 generations (~15 years), only one of the original twelve populations evolved the ability to colonize this ecological niche, giving access to an abundant nutrient not previously accessible (Blount et al., 2008). After more than a decade since the observation of this phenotype, none of the other populations have gained this metabolic trait. Due to the rarity of this innovation, it is clear that a multi-step mutational pathway was necessary to occupy this new niche.

Three stages of evolution were required in the LTEE for robust citrate utilization: potentiation, actualization, and refinement. The first appearance of Cit+ cells is referred to as the actualization step and was caused by a promoter duplication event that allowed the citrate:succinate

antiporter (citT) to be expressed under aerobic conditions (Blount et al., 2012). This so called 'promoter capture' event is insufficient to explain the strong citrate utilization phenotype (Cit++) however. Additional refinement mutations are necessary to confer Cit++ cells, which can grow to density in a single 24 hour growth cycle (Blount et al., 2008). One pathway involves yet another promoter mutation that activates the C4-dicarboxylate:H+ symporter (dctA). Together these mutations confer strong growth on citrate in reconstructed strains, but this two-step pathway was apparently inaccessible at the time this trait evolved in the LTEE, giving rise to the notion that other 'potentiating' mutations were needed (Blount et al., 2008).

A phylogenetic study of the LTEE pointed to likely key potentiating and refinement mutations on the pathway to Cit++ (Blount et al., 2012). In the current work, we characterize Citrate Synthase (CS), encoded by the gltA gene, which was mutated both before and after the evolution of citrate utilization. CS catalyzes the aldol condensation of oxaloacetate (OAA) and acetyl-CoA to create citrate and is the first irreversible step in the Krebs cycle. Due to the enzyme's importance in central metabolism and appearance of multiple mutations in isolates of Cit++ strains, it has been postulated that CS mutations were largely refinement mutations in Cit+ strains (Blount et al., 2012).

Here we show that mutations in CS were important for potentiating the evolution of the Cit+ phenotype, and additional mutations in the enzyme were later needed to refine the Cit++ phenotype. Whole-genome sequencing of isolated clones from the Ara-3 LTEE population revealed a number of mutations to CS (Blount et al., 2012). In every Cit+ strain identified, CS A258T was present (Quandt et al., 2015). A258T was also present in Cit- backgrounds at this point. Thus, we interpret this to mean that this mutation was present before the weak citrate utilization phenotype. In addition to A258T, protein coding secondary CS mutations (A124T, V152A, or A162V) were found independently in isolated Cit++ strains.

Metagenomics analysis was performed by Erik Quandt and has been discussed at length (Quandt et al., 2015). Briefly, A258T was first detected at 25,000 generations but had dropped significantly by 30,000 generations. By 33,500 generations, A258T was found at nearly 100% frequency in the population. Coincident with the expansion of Cit++ subpopulations, secondary mutations arose in A258T, citT duplication, and dctA mutational backgrounds.

Biochemical characterization alongside computational modeling and metabolic flux analysis provide a clear picture of how these mutations affected cellular regulatory networks. Early mutations in CS increased activity of the enzyme by losing allosteric regulation of NADH. Later, refining mutations led to a decrease in enzyme activity and partial restoration of allosteric control. Differences in enzyme activities due to these mutations are in good agreement with flux modeling predictions, which are not shown here for brevity. Here, we zoom in to CS and discuss the kinetics and modeling of the various mutations observed in the LTEE.

## Results & Discussion

### *Most Secondary Citrate Synthase Mutations Reduce Enzyme Activity*

We characterized the enyzmatic activity of A258T alone and with the other secondary mutations found in CS and compared kinetic parameters against the wildtype enzyme. We found that the A258T mutation did not significantly change the kinetic parameters of the enzyme (**Table 2-1**). However, addition of either A162V or A124T was found to reduce activity. The A162V substitution increased the $K_m$ values for both substrates of the reaction, oxaloacetate (OAA) (143.4 ± 11.1 μM vs. 84.2 ± 10.4 μM) and acetyl-CoA (AcCoA) (374.1 ± 41.2 μM vs. 138.6 ± 17.6 μM). The A124T mutation resulted a doubling of the $K_m$ for acetyl-CoA (295.1 ± 52.6 μM vs. 138.6 ± 17.6 μM) and reduced catalytic efficiency ($k_{cat}$ = 33.9 ± 2.7 $s^{-1}$ vs. 51.1 ± 2.5 $s^{-1}$). We interpret these results in a framework where secondary mutations in CS likely impact flux through cellular metabolism in a similar manner (Quandt et al., 2015).

***The Initial A258T Mutation Loses Allosteric Inhibition by NADH***

CS is allosterically inhibited by the primary product of the Krebs cycle, NADH (Weitzman, 1966a, 1966b; Weitzman and Jones, 1968). Since we were unable to see differences in kinetic parameters between A258T and the wildtype enzyme, we hypothesized that A258T would affect regulation by NADH. To test this hypothesis, NADH binding to purified enzyme variants was measured by monitoring changes in NADH fluorescence that occur upon binding to the protein (Duckworth and Tong, 1976), and allosteric inhibition was measured by measuring kinetic parameters in the presence of varying concentrations of NADH.

In complete agreement with previous studies, the wildtype enzyme exhibited a $K_d$ and a $K_i$ for NADH of nearly 1 μM each (**Table 2-1** and **Figure 2-1**) (Anderson and Duckworth, 1988; Pereira et al., 1994; Stokell et al., 2003). In contrast, A258T was largely incapable of binding NADH at concentrations within the range of our the binding assay, which extends up to ~10 μM NADH (Dickinson, 1970). A258T also showed an enormous increase in *K*i for NADH, >30-fold relative to the wildtype enzyme. Each of the secondary mutations restored both NADH binding and allosteric inhibition to varying degrees. In fact, A162Vdisplayed near wild-type $K_d$ and $K_i$ values for NADH (**Table 2-1**).

To gain further insight into how these mutations affected allosteric inhibition by NADH, we performed 3D modeling and molecular dynamics (MD) simulations using the Molecular Operating Environment and NAMD, respectively. Mutations were first introduced into a computationally corrected structure of *E. coli* citrate synthase (Maurus et al., 2003). After energy minimization and ns-timescale MD in a 6 Å water sphere, we observed subtle differences in the predicted conformation of each enzyme variant, most notably around the NADH binding site. The most pronounced changes were in the orientation of H110 (**Figure 2-1**), even though this amino acid is distant from the A258T mutation. In the wild-type CS structure, H110 adopts an upward conformation that allows for accommodation of NADH in the binding pocket. The A258T mutation

is predicted to reorient the H110 side chain toward the binding pocket, presumably creating an unfavorable steric barrier to NADH binding. Addition of the secondary mutations (A162V or A124T) resulted in simulated structures with H110 oriented between these two extremes, potentially explaining the varying levels of binding and inhibition described above. We next used template-based docking of NADH to these mutant CS structures to determine relative binding energies. We found that these binding energies were correlated with the experimentally determined NADH binding affinities (**Tables 2-1and 2-2**).

### *Conclusions*

Together our results suggest that metabolism and regulation are tuned specifically to an ecological niche. However, as shown here, they can be refit to a new niche with the addition of only a few mutations to important enzymes and regulatory proteins. What we observed here may indeed be broadly used by evolution. Complex regulatory networks and epistatic relationships must be fine-tuned within the genetic and mutational landscape they find themselves if the organism carrying them is to survive and flourish in a dynamic environment.

## Methods & Materials

### *Enzyme Activity Measurements*

Citrate synthase activity of purified $His_6$-tagged GltA variants was measured using a 5′,5′-dithiobis-(2-nitrobenzoate) (DTNB) colorimetric assay (Srere, 1969). Readings at 412 nm were taken in 96-well plates at 25°C using a Synergy HT plate reader (BioTek). Standard CS assay buffer consists of 20 mM Tris•Cl (pH 7.8), 100 mM KCl, and 1 mM EDTA (Duckworth and Tong, 1976). Enzyme was present at a concentration at least 100-fold lower than both substrates in all assays. Under these conditions, *E.coli* CS has been shown to conform to the ordered bisubstrate mechanism (Anderson and Duckworth, 1988). Kinetic data was fit to the Ordered Bi Bi equation using SigmaPlot 10 (Systat Software, San Jose, CA).

### *NADH Binding and Inhibition Measurements*

NADH equilibrium binding assays were performed as previously described in CS buffer lacking KCl (Duckworth and Tong, 1976). Briefly, enzymes were equilibrated with varying concentrations of NADH (0–6.4 μM) for 1 h at 25°C. Fluorescent measurements were made with excitation at 340 nm and emission reading at 425 nm in a Tecan M200 plate reader. The observed changes in fluorescence versus NADH concentration were fit to a hyperbolic ligand-binding curve using SigmaPlot 10.

Inhibition assays were performed essentially as described elsewhere (Stokell et al., 2003). Varying concentrations of NADH were equilibrated with enzyme in CS buffer lacking KCl at 25°C for 1 h. Substrates OAA and AcCoA were both added at 100 μM and initial reaction rates were measured using the DTNB assay described above. Enzyme activities were normalized to the activity of wild-type GltA in the absence of NADH. Percent inhibition was plotted versus NADH concentration and fit to a hyperbolic model in SigmaPlot 10.

### *Molecular Modeling*

Wild-type and mutant citrate synthase structures were analyzed using the Molecular Operating Environment (MOE2013.08). Dimeric models of ligand-free and NADH-bound *E. coli* citrate synthase (PDB: 1NXE and 1NXG) (Maurus et al., 2003) were prepared for analysis by reverting alanine-383 to phenylalanine and processing with the Structure Preparation application within MOE. Each model was then protonated (37°C, pH 7.4, 0.1 salt) using Protonate3D. As further preparation for QM/MM analyses, non-bridging solvent molecules were removed, a 6 Å solvent sphere was added, and charges were neutralized by the addition of KCl. Lastly, amino acid substitutions in a given CS variant were added sequentially with energy minimization to an RMS gradient of $10^{-3}$ kcal/mol/Å$^2$ using the Amber12 force field with Extended Hückel Theory and R-field solvation electrostatics after each mutation. NADH binding energies for the resulting models were calculated using the Ligand Interactions subroutine.

Table 2-1. Kinetic and regulatory properties of citrate synthase variants

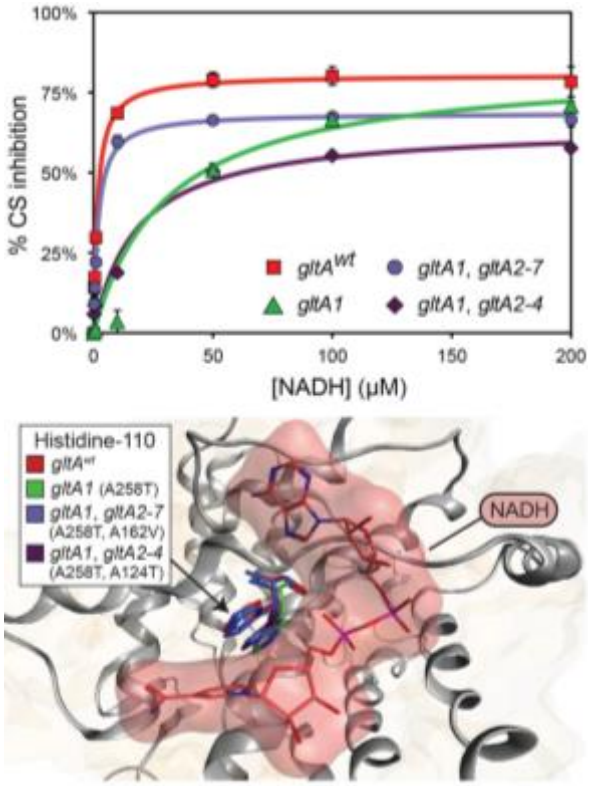
| Variant | $k_{cat}$ ($s^{-1}$) | $K_m$ OAA (μM) | $K_m$ AcCoA (μM) | $K_d$ NADH (μM) | $K_i$ NADH (μM) | Maximum % inhibition | Docking energy (kcal/mol) |
|---|---|---|---|---|---|---|---|
| Wildtype | 51.1 ± 2.5 | 84.2 ± 10.4 | 138.6 ± 17.6 | 1.2 ± 0.1 | 1.7 ± 0.1 | 80.5 ± 1.0 | –80.3 |
| A258T | 49.1 ± 5.3 | 90.3 ± 22.5 | 148.0 ± 41.1 | N.D. | 30.2 ± 4.6* | 83.2 ± 3.1 | –28.8 |
| A258T, A162V | 55.0 ± 3.0 | 143.4 ± 11.1* | 374.1 ± 41.2* | 1.5 ± 0.4 | 1.9 ± 0.1 | 68.7 ± 0.7* | –63.8 |
| A258T, A124T | 33.9 ± 2.7* | 88.1 ± 12.8 | 295.1 ± 52.6* | 9.5 ± 1.7* | 17.1 ± 6.2* | 64.6 ± 5.1* | –47.8 |

Data are represented as fit mean ± S.E. Significant differences from the wild-type enzyme are marked with an asterisk (two-tailed t-test, p-value < 0.05). N.D. indicates no detectable binding.

Table 2-2. Representative Binding Energy Evaluation of NADH and wildtype citrate synthase

| Interaction | Type | Distance | E (kcal/mol) |
|---|---|---|---|
| 1 | H-donor | 2.94 | -0.1 |
| 2 | H-donor | 2.63 | -3.8 |
| 3 | H-donor | 2.85 | -2.6 |
| 4 | H-acceptor | 2.69 | -2.6 |
| 5 | H-acceptor | 2.62 | -4.2 |
| 6 | H-acceptor | 2.74 | -15.1 |
| 7 | H-acceptor | 2.76 | -1.6 |
| 8 | H-acceptor | 2.84 | -1.3 |
| 9 | H-acceptor | 2.74 | -1.2 |
| 10 | H-acceptor | 2.97 | -7.4 |
| 11 | H-acceptor | 2.69 | -3.2 |
| 12 | H-acceptor | 2.72 | -8.6 |
| 13 | H-acceptor | 2.64 | -2.9 |
| 14 | H-acceptor | 2.66 | -0.7 |
| 15 | ionic | 2.74 | -6.4 |
| 16 | ionic | 2.97 | -4.7 |
| 17 | ionic | 3.73 | -1.1 |
| 18 | ionic | 3.18 | -3.4 |
| 19 | ionic | 3.64 | -1.4 |
| 20 | ionic | 2.72 | -6.7 |
| 21 | pi-H | 3.59 | -0.7 |
| 22 | pi-H | 4.03 | -0.6 |
| 23 | pi-pi | 3.9 | 0 |
| | | Total | -80.3 |

Figure 2-1. CS mutations alter allosteric regulation by NADH




CS mutations alter allosteric regulation by NADH (Top) NADH-mediated inhibition citrate synthase activity for the wild-type enzyme and evolved variants with combinations of gltA mutations (Bottom) Molecular modeling predicts that the observed changes in allosteric regulation in evolved CS sequences are primarily caused by how mutations affect the orientation of H110 in the NADH binding pocket. The A258T mutation is predicted to redirect this side chain into the binding pocket creating a steric clash with NADH (red surface). The two characterized variants of citrate synthase with an additional mutation are predicted to reorient H110 back toward the wild-type conformation. The degree of this predicted structural change correlates with the relative levels of NADH inhibition experimentally measured for these CS variants (**Table 2-1**).

## 2.2 Addiction of Diverse Bacteria by an Engineered Noncanonical Amino Acid Beta-Lactamase (ENABL)

### Introduction

The seemingly static genetic code has shaped life on Earth as we know it. The proteomes of all known organisms rely on the same 20 canonical amino acids. Naturally occurring expansion of the code has given rise to the use of additional rare amino acids, selenocysteine and pyrrolysine. Recent advances in engineering orthogonal translation systems (OTSs) comprised of tRNA and aminoacyl-tRNA synthetase (aaRS) pairs have enabled expansion of the genetic code using nonstandard amino acids (NSAAs), typically using Amber (UAG codon) suppression (Liu and Schultz). The biological and biochemical impact of NSAAs across the proteome are poorly understood because it has proven difficult to maintain new genetic codes in organisms. Often the OTS is completely lost or kept but inactivated, even when Amber suppression is required (Wang et al).

In an effort to overcome loss or inactivation, we reengineered the protein core of TEM-1 beta-lactamase to generate an antibiotic resistance marker that is dependent upon the incorporation of 3-nitro- and 3-iodo-tyrosine. TEM-1 β-lactamase confers resistance to β-lactam antibiotics across a wide range of Gram-negative bacteria. High-resolution crystal structures are also available, making it an ideal candidate for engineering NSAA-dependence across multiple species. In order to addict the protein to an unnatural amino acid, we sought core amino acids that would disrupt activity when replaced with 3-nitro-tyrosine (3nY) or 3-iodo-tyrosine (3iY). Using the Engineered Noncanonical Amino acid Beta-Lactamase (ENABL) variant, we could maintain a functional OTS and NSAA dependence in *E. coli* cells for more than 250 generations without detectable escape. In addition, ENABL allowed the retention of an OTS across a range of different bacterial species without modification of the host genomes (*Shigella flexneri, Salmonella enterica, Yersinia ruckeri,*

and *Acinetobacter baylyi*). Here, we describe the selection and modeling of ENABL and outline a methodology to addict diverse organisms to NSAAs.

## Results & Discussion

### *Design of Library Positions*

Visual inspection of the structure (PDB: 1XPB) revealed potential sites, which were then scored on solvent accessibility or orientation to the inner core of the enzyme. Naturally, we omitted sidechains directly involved or immediately adjacent to catalytic or substrate binding residues. Using these guideposts, we identified 3 possible sidechains: F66, L162, and T189 (**Figure 2-2**). Mutant enzymes with Amber codons at these positions were then tested using the standard *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase:tRNA (MjYRS/tRNACUA) pairs, which incorporate tyrosine, 3nY, and 3iY. L162X and T189X both abolished activity in the presence of Y, 3nY, and 3iY (Tack et al., 2016).

### *Selection of Addicted TEM-1 Variants*

Having confirmed loss of activity with tyrosine, 3nY, and 3iY, we next designed libraries based off of positions L162 and T189. The 6 closest residues to each suppressed position were saturated (**Figure 2-3**) and selected on solid media supplemented with carbenicillin and either 3nY or 3iY. The L162 library was shown to be the most robust, having more than 95% of isolated clones still requiring the unnatural amino acid with the orthogonal translation machinery still intact. To mitigate reversion, 23 clones from this population were tested with leucine and tyrosine replacing the Amber codon. Only two of the clones were shown to have reduced resistance with the canonical solutions (MIC <1000 μg/mL carbenicillin; Fig. 1a, TEM-1.B9 and TEM-1.F2). We next determined if any other canonical solutions might exist at this position and found that TEM-1.B9 activity could only be rescued by phenylalanine. It was subsequently shown that the OTS

machinery could be maintained for over 250 generations in *E. coli* and successfully inserted into a set of diverse bacterial species (Tack et al., 2016).

***Molecular Modeling of ENABL***

To gain insight into how 3nY and 3iY were incorporated into ENABL, we turned to molecular modeling and simulation. Using the Molecular Operating Environment (MOE) software package, we prepared the published, 1.9 Å structure for simulation by protonating and energy minimizing the structure. Our corrected structure was inspected for anomalies and aligned against the reference structure (0.433 RMSD), giving us confidence that we had not significantly altered the empirically derived coordinates.

Due to the fact that 3nY and 3iY are noncanonical amino acids and are not currently found in state-of-the-art modeling programs, it was necessary to build rotamer libraries *de novo*. To do so, we employed short ~1 pS timescale molecular dynamics simulations on individual noncanonical amino acids at constant temperature (LowModeMD). Capped and free termini were used in generating conformational searches. The AMBER:EHT forcefield was used with tethering to refine and score individual confirmations. Partial charges were scored using AM1-BCC. A total of 10,000 iterations were used in creating the initial conformational library. Duplicates were scored by alignment to a common backbone (<0.3 Å) and removed. At this stage, having the rotamer library in hand allowed us to mutate residues *in silica* in a physics-based manner and rationalize key positions within the protein that shaped the enzyme's addiction for the noncanonical amino acid.

We mutated F66, L162, and T189 in the context of the native residues to determine if any energy barriers might have been present in our initial experiments that would explain inactivity at those positions. This was also an interesting test case in that if modeling could predict the known site, then the methodology developed here might be more widely adopted for future addiction experiments. Consistent with our *in vivo* results, F66X was able to comfortably incorporate tyrosine

with a decrease (-15.2 kcal/mol) in potential energy of the system. Meanwhile, severe steric clashes were observed with 3nY and 3iY along with large increases in potential energy (+40 kcal/mol) of the system (**Table 2-3**). Additionally, the noncanonical amino acid substitutions resulted in the loss of a salt bridge among R43, E64, and the backbone carbonyl of R65 (**Figure 2-4**), likely destabilizing the protein. Potential Energy differences alone do not explain T189X/Y inactivity. Rather, it appears as though a key hydrogen bonding network among D157, T160, and T189 is disrupted (**Figure 2-5**). Finally, in complete agreement with our *in vivo* screening, L162X resulted in large perturbations in potential energy of the system upon incorporation of Y, 3nY, or 3iY (+35, +48, and +91 kcal/mol, respectively; **Table 2-4**) and were the least favorable substitutions in the context of native residues.

Given that our modeling at this stage was largely consistent with our experimental observations, we next sought to model ENABL in its final form. Using our corrected structure as a template, we made the following mutations to the enzyme: M68F, F72G, L139T, L152M, and L162X. Sidechains were repacked to allow for a robust conformational search of rotamers within the local environment and minimized in a 6 Å water sphere. In our final model of ENABL, which only deviated from our corrected structure by 0.627 Å RMSD (**Figure 2-6)**, M68F is oriented above the noncanonical amino acid and appears to make a Π-stacking interaction with the aromatic ring of the noncanonical amino acid, perhaps stabilizing the structure of the enzyme. F72G makes room for the hydroxyl of the tyrosine analog, which hydrogen bonds to the backbone carbonyl of F72G. L152M is a conservative mutation that likely shapes the core to allow for the bulkier noncanonical amino acids. Interestingly, L139T is predicted to participate in a newly formed hydrogen bonding network with a water molecule and the *orth*- substituent in 3nY. Together, the hydrogen bonding observed due to these changes contributes -5.4 kcal/mol (**Figure 2-7; Table 2-5**). A similar halogen bonding interaction may also be present in the 3iY variant.

To determine if potential energy predictions could account for activity of canonical solutions at position 162, we mutated X162 with each amino acid in the context of the surrounding ENABL mutations (**Table 2-6**). The 3iY substitution was shown to decrease the potential energy of the system more than 3nY (-15 kcal/mol), despite 3nY conferring a higher MIC in our experiments. The only canonical solution to rescue NSAA dependence, phenylalanine, slightly increased the potential energy by +1.5 kcal/mol. Tryptophan and tyrosine were unable to rescue dependence in our *in vivo* assays but were shown to possess some activity. Our simulations predicted that these two canonical solutions would lower the potential energy in the context of ENABL. X162R/Q/N/L were all shown to have lower predicted potential energies than 3nY. Given that some of these mutations were empirically shown not to confer resistance, we conclude that potential energy alone is not sufficient in predicting activity.

However, the change in stability ($\Delta G_s$) of each mutation in the context of surrounding residues, which is calculated as a change in Gibbs free energy after mutating the protein ($\Delta G_s=\alpha[\Delta E_{vdWf}+0.5(\Delta E_{coulf}+\Delta E_{solf})]+\beta E_{SS}+\gamma\Delta\Delta G_{solu}+\epsilon\Delta\delta_{gly}+\zeta\Delta\delta_{His}$), more accurately predicted activity. Using $\Delta G_s$ as a metric, we observed small differences between 3nY and 3iY at position 162 (0.77 kcal/mol). W, F, and Y are each within 2 kcal/mol of 3nY. All other amino acid substitutions, empirically shown to be nonfunctional solutions, were found to be far less stable (+3.4 to +8.9 kcal/mol) in the context of ENABL mutations. While the rank order did not perfectly match the empirical data, the values were clearly divided into functional and non-functional groups (**Figure 2-8**). Below a change of +2 kcal/mol of PE, variants are functional.

This was a somewhat surprising result in that molecular simulation was adequate in identifying functional solutions without empirical input. First, potential energy perturbations could predict candidate sites for selection. In the case of ENABL, both NSAAs were the most energetically unfavorable substitutions at position 162. Next, key contacts were identified in the

mutant enzyme that rationalized addiction at the molecular level. Finally, changes in stability of the enzyme upon mutation crudely predicted activity of enzymes without training data. Future *in silica* selections of neighboring residues adjacent to the NSAA may be useful in narrowing the number of potential enzyme variants to test or simplifying library design for traditional *in vivo* selections.

### *Conclusions*

Our ENABL selection allowed us to maintain OTSs over a diverse set of bacterial species using expanded genetic codes. We were unable to detect escape after hundreds of generations of evolution. In addition, modeling proved useful in rationalizing our 'blind' engineering efforts. Taken together, our results provide a framework that could be extended to other enzymes. Having trained our modeling regiment on empirical data, it should now prove possible to develop other resistance markers, perhaps even addicting industrially relevant bacterial strains or eukaryotic organisms to NSAAs. For example, KanMX is a commonly used marker that confers resistance to various aminoglycosides, kanamycin in bacteria and geneticin (G418) in eukaryotes. *Saccharomyces cerevisiae* is a model eukaryotic organism and can be engineered to suppress Amber codons using heterologous OTSs. By performing *in silica* scanning mutagenesis at residues distant from the catalytic or substrate binding residues in KanMX, candidate sites for selection should be identifiable by perturbations in potential energy of the system. A similar selection mechanism outlined here could then be applied allowing for the generation of the first 'addicted' eukaryotic organism to an expanded genetic code.

## Materials and Methods

### *Residue Selection and Library Design*

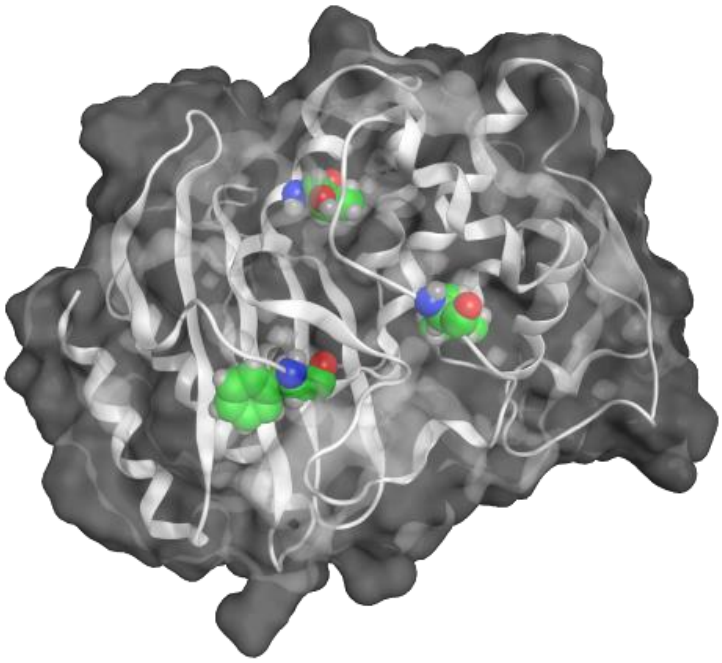
TEM-1 β-lactamase libraries were designed by manual inspection of the TEM-1 structure (PDB: 1XPB) using the molecular visualization program pymol (v1.3). For the L162 library, the six residues with side chains in closest proximity to the leucine side chain were selected for site-

saturation (NNS) mutagenesis. These corresponded to residues M68, F72, L139, L148, L152, and L169.

***Molecular Modeling of TEM-1 Library Variants***

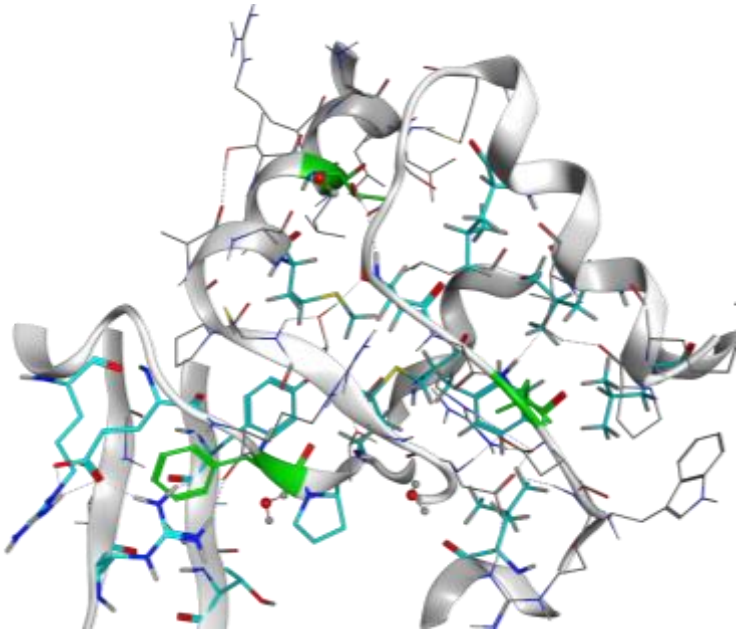
The wild-type structure of TEM-1 (PDB: 1XPB) was prepared for mutational analyses using the Molecular Operating Environment (MOE.09.2014) software package from Chemical Computing Group. The structure was inspected for anomalies and protonated/charged with the Protonate3D subroutine (310K, pH 7.4, 0.1 M salt). The protonated structure was then lightly tethered to reduce significant deviation from the empirically determined coordinates and minimized using the Amber10:EHT forcefield with R-field treatment of electrostatics to an RMS gradient of 0.1 kcal mol-1 Å-1. Next, we created rotamer libraries for each non-canonical amino acid using a low-mode molecular dynamics (LowModeMD) methodology. Library positions of representative variants from the selection were mutated and repacked within the local environment. Conformational analysis of the loop was evaluated with LowModeMD. Mutated structures were then solvated with water and counterions in a 6 Å sphere and minimized to an RMS gradient of 0.001 kcal mol-1 Å-1. 2-D contact maps of the non-canonical solutions were inspected with Ligand Interactions after arbitrarily setting the non-canonical amino acid to the ligand position. Measurements of stability and potential energy were scored within MOE.

Figure 2-2: Potential sites for NSAA dependence



Potential sites for NSAA dependence. Positions F66, L162, and T189 are highlighted. Each of these positions are solvent shielded and oriented away from catalytic and substrate binding residues.

Figure 2-3: Library positions for NSAA dependence

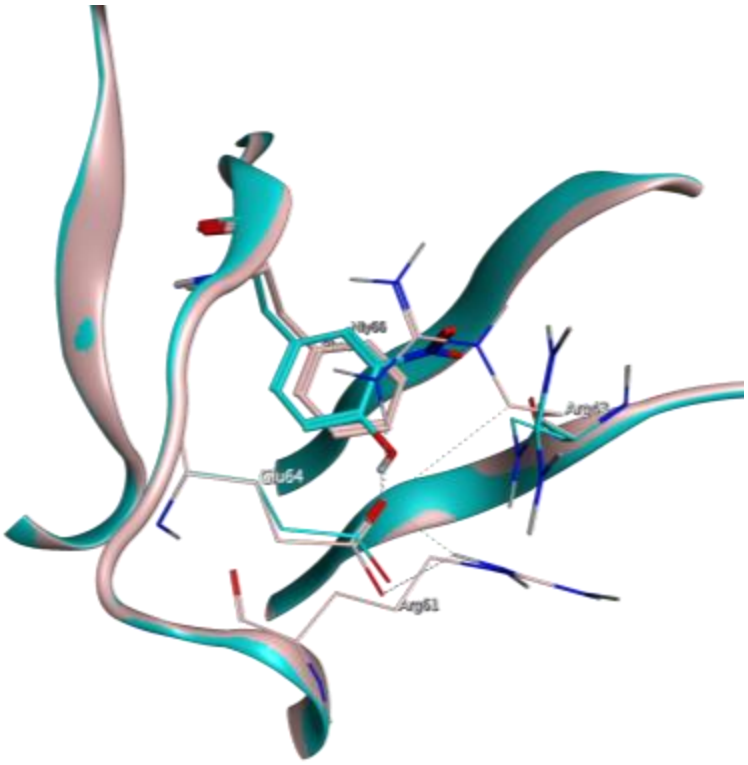


Library positions for NSAA. Amber codons were inserted individually into positions 66, 162, and 189 (green residues). The surrounding resiudes, highlighted in blue, were randomized using NNS codons.

Table 2-3: *In silica* analysis of position 66

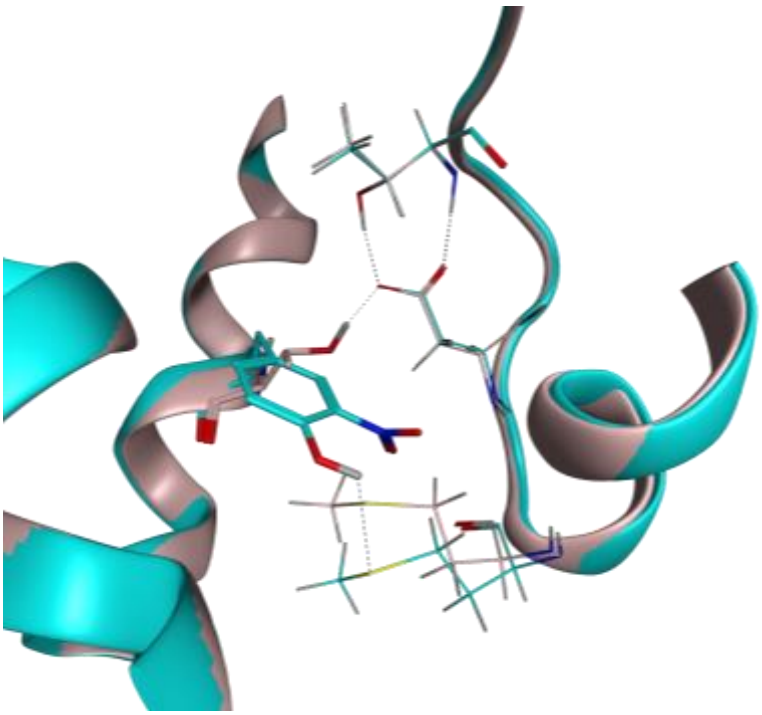
| Mutation | Stability | dStability | E | dE |
|---|---|---|---|---|
| 1:F66R | -9.49 | 4.19 | -7276.66 | -67.74 |
| 1:F66N | -9.20 | 4.48 | -7245.39 | -36.47 |
| 1:F66W | -12.78 | 0.90 | -7230.97 | -22.05 |
| 1:F66T | -9.79 | 3.89 | -7227.16 | -18.24 |
| 1:F66Y | -12.24 | 1.45 | -7224.11 | -15.19 |
| 1:F66Q | -9.75 | 3.93 | -7214.48 | -5.56 |
| 1:F66V | -10.74 | 2.94 | -7210.51 | -1.59 |
| 1:F66F | -13.68 | 0.00 | -7208.92 | 0.00 |
| 1:F66I | -10.51 | 3.17 | -7186.70 | 22.22 |
| 1:F66E | -8.78 | 4.90 | -7183.55 | 25.37 |
| 1:F66M | -10.84 | 2.84 | -7181.04 | 27.88 |
| 1:F66K | -8.41 | 5.27 | -7180.66 | 28.26 |
| 1:F66L | -11.03 | 2.65 | -7178.15 | 30.77 |
| 1:F66H | -5.91 | 7.77 | -7173.41 | 35.51 |
| 1:F66[(Niy)] | -13.43 | 0.25 | -7169.72 | 39.20 |
| 1:F66[(Ioy)] | -13.38 | 0.30 | -7168.05 | 40.87 |
| 1:F66S | -8.63 | 5.05 | -7167.85 | 41.07 |
| 1:F66C | -8.87 | 4.82 | -7164.51 | 44.41 |
| 1:F66D | -7.90 | 5.78 | -7153.32 | 55.60 |
| 1:F66G | -5.87 | 7.81 | -7151.37 | 57.55 |
| 1:F66A | -8.87 | 4.81 | -7149.83 | 59.09 |
| 1:F66P | -7.35 | 6.33 | -7118.13 | 90.79 |

Figure 2-4: F66X breaks a salt bridge




F66X breaks a salt bridge. Upon mutation, the increased size of the NSAAs cause a salt bridge to break between E64 and R43, depicted in blue. This potentially destabilizes the enzyme, causing it to become inactive.

Figure 2-5: T189X disrupts a hydrogen-bonding network

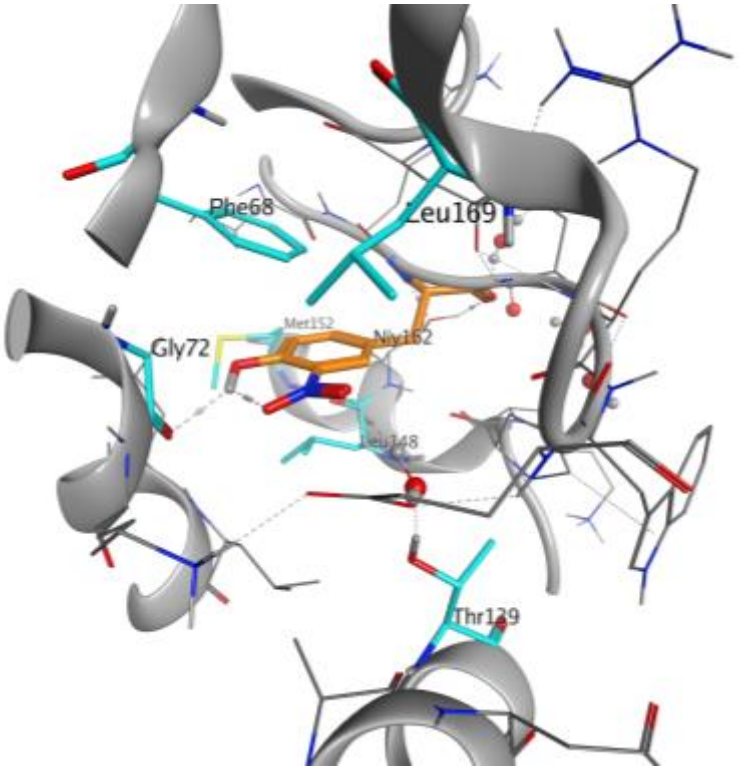


T189X disrupts a hydrogen-bonding network. The NSAA at position 189 disrupts bonding with an aspartic acid at the end of neighboring helix. This library was the most unstable in our experiments, highlighting the importance of this interaction.

Table 2-4: L162X destabilizes TEM-1

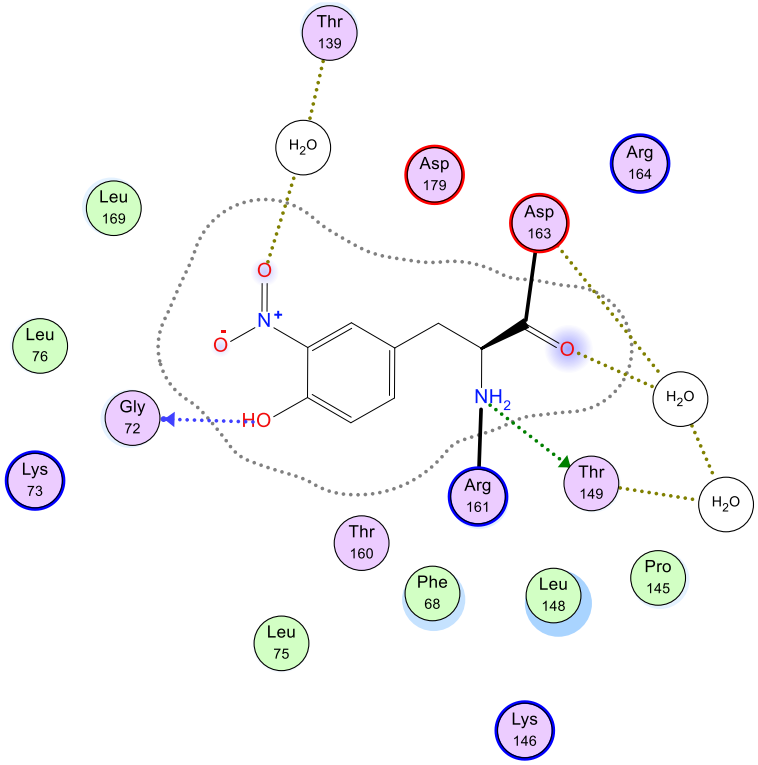
| Mutation | Stability | dStability | E | dE |
|---|---|---|---|---|
| 1:L162R | -11.31 | 2.01 | -7386.37 | -104.31 |
| 1:L162N | -10.35 | 2.97 | -7322.01 | -39.95 |
| 1:L162Q | -11.86 | 1.46 | -7312.36 | -30.30 |
| 1:L162V | -12.40 | 0.92 | -7290.84 | -8.78 |
| 1:L162K | -10.59 | 2.72 | -7289.09 | -7.03 |
| 1:L162M | -12.55 | 0.76 | -7286.09 | -4.03 |
| 1:L162T | -11.35 | 1.97 | -7285.62 | -3.56 |
| 1:L162L | -13.32 | 0.00 | -7282.06 | 0.00 |
| 1:L162G | -7.79 | 5.52 | -7277.13 | 4.93 |
| 1:L162I | -13.15 | 0.17 | -7276.27 | 5.79 |
| 1:L162E | -10.58 | 2.73 | -7276.18 | 5.88 |
| 1:L162S | -10.51 | 2.81 | -7268.30 | 13.76 |
| 1:L162C | -10.91 | 2.41 | -7262.32 | 19.74 |
| 1:L162F | -12.76 | 0.55 | -7257.22 | 24.84 |
| 1:L162A | -10.86 | 2.46 | -7250.96 | 31.10 |
| 1:L162D | -9.25 | 4.07 | -7247.74 | 34.32 |
| 1:L162Y | -11.28 | 2.03 | -7246.58 | 35.48 |
| 1:L162H | -8.51 | 4.81 | -7244.50 | 37.56 |
| 1:L162P | -9.59 | 3.72 | -7239.73 | 42.33 |
| 1:L162W | -13.37 | -0.05 | -7236.00 | 46.06 |
| 1:L162[(Niy)] | -13.10 | 0.22 | -7233.73 | 48.33 |
| 1:L162[(Ioy)] | -13.21 | 0.11 | -7190.76 | 91.30 |

Figure 2-6: Final ENABL model




Final ENABL model. Representative structure of ENABL produced by MOE. Library positions are shown in blue. A stabilized hydrogen-bonding network is observed between T139, 3nY162, and a water molecule. A total of -5.3 kcal/mol of stabilizing energy is added with this network. We hypothesize that a similar stabilizing interaction may be present with 3iY, presumably due to a halogen-bonding interaction.

Figure 2-7: 2D contact map of ENABL mutations




2D contact map of ENABL. By setting the NSAA to a ligand, we were able to measure interactions to the amino acids using the LigandInteractions subroutine. Hydrogen bonding between the hydroxyl of the NSAA and G72 is predicted. Further, the nitro substituent makes contact with a bridging water, which is stabilized by T139.

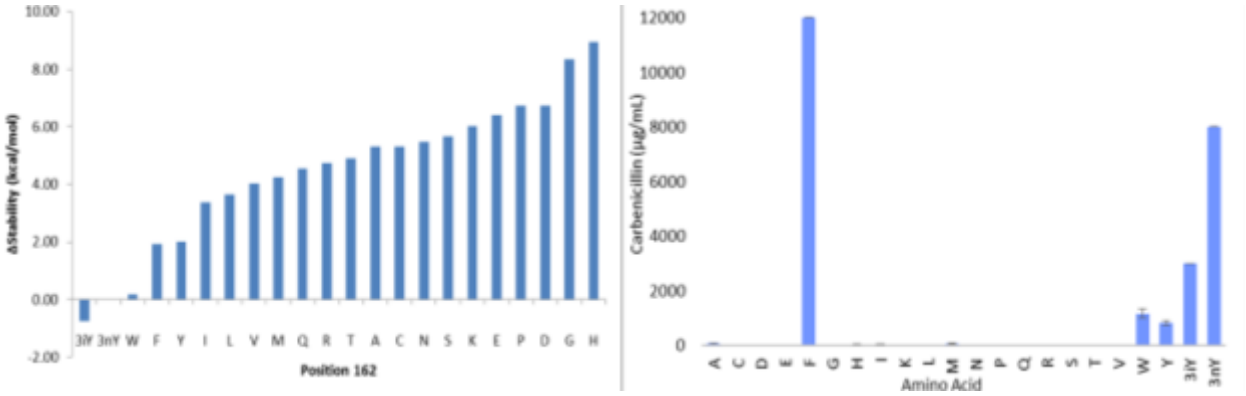
Table 2-5: ENABL stabilization interactions

| Interaction | Type | Distance (Å) | E (kcal/mol) |
|---|---|---|---|
| 1 | H-donor | 3.00 | -2.1 |
| 2 | H-donor | 2.87 | -1.2 |
| 3 | H-acceptor | 2.70 | -1.0 |
| 4 | H-acceptor | 3.04 | -1.0 |
| | | Total | **-5.3** |

Table 2-6: *In silica* analysis of position 162 in the context of surrounding ENABL mutations

| Mutation | Stability | dStability | E | dE |
|---|---|---|---|---|
| 1:(Niy)162[(Ioy)] | -16.90 | -0.77 | -88679.8 | -15.80 |
| 1:(Niy)162[(Niy)] | -16.13 | 0.00 | -88664 | 0.00 |
| 1:(Niy)162W | -15.96 | 0.17 | -88667 | -3.02 |
| 1:(Niy)162F | -14.21 | 1.91 | -88662.5 | 1.51 |
| 1:(Niy)162Y | -14.14 | 1.99 | -88675.9 | -11.93 |
| 1:(Niy)162I | -12.75 | 3.38 | -88651.1 | 12.93 |
| 1:(Niy)162L | -12.49 | 3.64 | -88664.5 | -0.49 |
| 1:(Niy)162V | -12.12 | 4.01 | -88644.8 | 19.22 |
| 1:(Niy)162M | -11.88 | 4.25 | -88631.5 | 32.50 |
| 1:(Niy)162Q | -11.58 | 4.55 | -88693.4 | -29.43 |
| 1:(Niy)162R | -11.38 | 4.74 | -88759.9 | -95.95 |
| 1:(Niy)162T | -11.23 | 4.90 | -88645.3 | 18.70 |
| 1:(Niy)162A | -10.83 | 5.30 | -88624.8 | 39.20 |
| 1:(Niy)162C | -10.82 | 5.31 | -88628.4 | 35.61 |
| 1:(Niy)162N | -10.65 | 5.48 | -88675.6 | -11.59 |
| 1:(Niy)162S | -10.48 | 5.65 | -88632.1 | 31.88 |
| 1:(Niy)162K | -10.10 | 6.03 | -88634.8 | 29.22 |
| 1:(Niy)162E | -9.74 | 6.39 | -88609.7 | 54.34 |
| 1:(Niy)162P | -9.41 | 6.72 | -88600.4 | 63.63 |
| 1:(Niy)162D | -9.39 | 6.74 | -88628.7 | 35.34 |
| 1:(Niy)162G | -7.79 | 8.34 | -88626.9 | 37.13 |
| 1:(Niy)162H | -7.19 | 8.94 | -88650.1 | 13.88 |

Figure 2-8: Change in stability of position 162 in the context of ENABL




Change in stability of position 162 in the context of ENABL mutations. We measured the energy differences for each amino acid at position 162 (left). Consistent with our *in vivo* characterization (right), NSAAs dominated stability and activity. The canonical solutions were within 2 kcal/mol of ΔStability. Each amino acid above +2 kcal/mol was shown to be nonfunctional.

## 2.3 Molecular Modeling of Rifampicin Resistance in Escherichia Coli Using Expanded Genetic Codes

### *Introduction*

In the previous section, we engineered a resistance marker to be dependent upon NSAA incorporation. Here, we investigate how genetic code expansion in *E. coli* influences evolution in the context of antibiotic resistance. Expansion of the genetic code with NSAAs changes not only the protein-coding potential of an organism but also the network topology of the code itself (Freeland and Hurst, 1998; Knight et al., 2001). For example, by recoding the Amber codon to a NSAA, a nonsense mutation is translated into a missense mutation. The connectivity of the code is fundamentally changed and in turn gives evolution more protein-coding options from which to explore. This property is in addition to the availability of a unique geometry or chemical potential of a particular NSAA. Taken together, it could be argued that NSAA expansion of the genetic code should make organisms more evolvable (Hammerling et al., 2014).

Rifampicin is a polyketide antibiotic, commonly prescribed to treat Mycobacterium infections. The broad spectrum antibiotic inhibits the bacterial RNA polymerase, disrupting elongation of DNA-dependent RNA synthesis. Resistance is conferred by single mutations along the β-subunit of RNAP, which inhibit rifampicin binding by disrupting hydrogen-bonding interactions or sterically occluding the antibiotic. Crystal structures of the *Thermophillus aquaticus* RNAP and more recently *E. coli* RNAP bound to rifampicin have shed light on the mechanism of inhibition and modes of resistance at the molecular level (Campbell et al. 2001; Molodtsov et al. 2013). Resistance has been well-characterized and shown to require a single codon mutation in amino acids located in three regions of rpoB, providing an easily tractable system for studying impacts of expanded genetic codes.

In collaboration with Dr. Michael Hammerling, we sought to explain resistance mutations at the molecular level in *E. coli* rpoB in the context of standard and expanded genetic codes (**Figure**

**2-9**). Several spontaneous mutations at position 513 in ropB were observed and conferred resistance in strains lacking native Amber codons, but capable of incorporating NSAAs with functional OTSs. Further, using recombineering at Region 1 of rpoB (residues 509-533) (Hammerling et al., 2016), we found an additional number of single- and double-Amber mutations using a number of NSAAs. The modeling and impacts of these mutations on the rifampicin binding pocket are discussed in detail.

## Results & Discussion

### *Molecular Modeling of RNAP Mutants*

We initially attempted to model the effects of 3-halo-substituted tyrosines on our computationally corrected structure by: (1) mutating Q513 to tyrosine, (2) adding the halogen substituent to the ring at the *ortho* position, and then (3) calculating the lowest conformational dihedral angles for optimal geometry of the newly formed noncanonical amino acid. We then superimposed rifampicin onto the mutated structure using coordinates from the crystal structure, optimized dihedral geometry again (**Figure 2-10**), and minimized the complex. We measured contact energies between rifampicin and the mutated structure and found positive energy values, representing unfavorable interactions or clashing between rifampicin and the sidechains. In contrast, the wildtype enzyme had -9.8 kcal/mol of contact energy (**Tables 2-7**). While this methodology proved useful in predicting binding or occlusion of the ligand in mutated structures (**Table 2-8**), the rotamer-by-dihedral methodology was slow and laborious; it was also unlikely to generate realistic sidechain conformations. Further, it was impossible to compare to native amino acids at that position using standard algorithms in a high-throughput manner.

Instead, using the same methodology discussed in Section 2.2, we added rotamer conformations for 3-substitueted halo-tyrosines, 4-azidophenylalanine (AzF) and 5-hyrdoxytrytophan (5HTP) to our nancanonical amino acid rotamer library. This allowed us to

quickly screen conformational space without visually inspecting dihedral angles and manually calculating contact or repulsion energies. It also allowed for physics-based solutions to conformational searches using validated forcefields alongside Extended Hückel Theory calculations. Having the additional rotamers also allowed us to use standard subroutines within MOE to make *in silica* measurements and comparisons across all sequence space.

With our rotamer library updated to include the additional NSAAs, we investigated all possible amino acids, standard and noncanonical, at position 513 in an effort to full characterize all evolutionary possibilities at this position. By examining changes in the stability of the antibiotic-bound complex in comparison to the energy minimized apo-reference structures, we were able to accurately predict all known resistance mutations from the literature and those found independently in the current study. Changes in the predicted stability of the RNAP-rifampicin complexes spanned nine orders of magnitude (**Table 2-9; Figure 2-11**). The larger values reflect the obviously insurmountable steric occlusion of the rifampicin binding pocket imposed by bulkier side chain substituents.

In our simulations, rifampicin-resistant mutations display positive values in $\Delta\Delta G$ of complex stability, denoting large energetic barriers to binding. On the other hand, amino acids that are known to display susceptible phenotypes result in more favorable negative $\Delta\Delta G$ values for complex formation. For 3-iodotyrosine incorporated at position 513, our modeling predicts that the iodine atom attached to the aromatic ring of the side chain at the *ortho* position projects directly into the rifampicin binding pocket, completely abolishing binding (**Figure 2-12**). The other halo- and nitro-substitueted NSAAs followed with similar geometries. These results are consistent with the growth advantage observed in the presence of rifampicin of all *ortho*-substituted nsAA resistance mutants and tryptophan over tyrosine at this position, and further imply that the rotamers

identified in our modeling experiments may be sufficient in explaining how these amino acids lead to faster-growing phenotypes (**Figure 2-13A**).

We next sought to understand how the NSAA contributed to resistance in our spontaneous double-Amber mutants. These mutations were shown to require the OTS and noncanonical amino acids in our *in vivo* assays. Again, we modeled the rifampicin resistant variant and compared the stability of the ligand:enzyme complex to reference structures lacking rifampicin. We observed strong complex destabilization upon addition of the NSAAs. In each case, both NSAAs were predicted to occlude binding to rifampicin (**Figure 2-13C, Table 2-10**). Template-based docking of rifampicin into a library of conformational variants at each of these positions failed to produce productive binding complexes, giving way to the notion that there may exist some epistatic relationship among the NSAAs in the double-mutants.

***Conclusions***

This section revealed the ability of NSAAs to impart rifampicin resistance on the *E. coli* RNAP. Expanded mutational connectivity, enabled by the expanded genetic code, allowed us to observe the potential role of NSAAs and additional canonical amino acids solutions, tyrosine and tryptophan in bacterial rifampicin resistance. The canonical solutions were previously unobserved due to the distance between the wildtype glutamine (CAG) codon and the tyrosine (TAC/TAT) or tryptophan (TGG) codons. In our recombineering experiments, we found a number of double-Amber mutations that lead to robust rifampicin resistance, which may point to so far unidentified epistatic relationships among these and other residues in the local environment. In both cases, our modeling accurately predicted known mutations at these positions and rationalized the molecular basis for rifampicin resistance in the context of standard and expanded genetic codes by measuring not only affinity of the mutated RNAPs but also complex stability.

## Materials & Methods

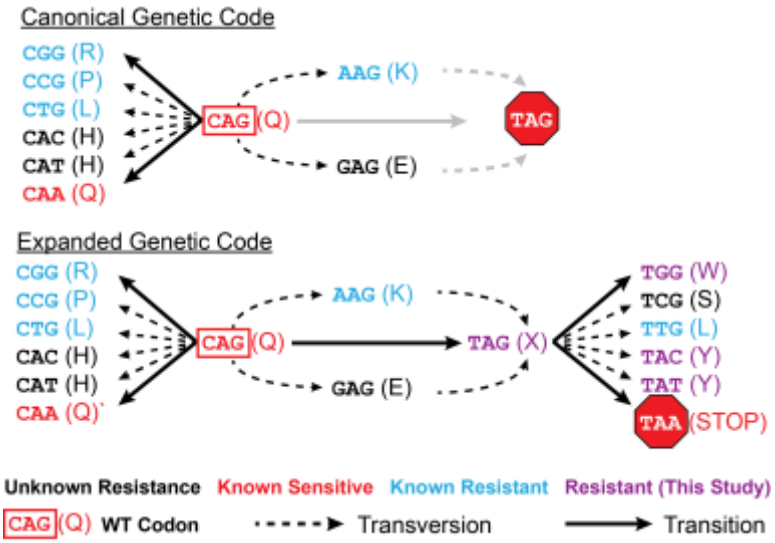
### *Molecular modeling*

The wild-type structure of *E. coli* RNA polymerase (PDB: 4KMU chains C and D) (Molodtsov et al., 2013) was prepared for mutational analyses using the Molecular Operating Environment (MOE.09.2014) software package from Chemical Computing Group. The structure was inspected for anomalies and protonated/charged with the Protonate3D subroutine (310K, pH 7.4, 0.1 M salt) (Labute, 2009). The protonated structure was then lightly tethered to reduce significant deviation from the empirically determined coordinates and minimized using the Amber10:EHT forcefield with R-field treatment of electrostatics to an RMS gradient of 0.1 kcal $mol^{-1}$ $Å^{-1}$. Next, we created rotamer libraries for each non-canonical amino acid using the low-mode molecular dynamics method (LowModeMD) (Labute, 2010a). Library positions of representative variants from the selection were mutated and repacked within the local environment. The final model for each variant was further refined by placing the protein within a 6 Å water sphere and minimizing the solvent enclosed structure to an RMS gradient of 0.001 kcal $mol^{-1}$ $Å^{-1}$. Models were evaluated by calculating Phi-Psi angles and superimposed against the reference structure to ensure that the native fold was maintained (Clark and Labute, 2007). Measurements of stability, potential energy, and contact interactions on a fixed-ligand structure for each variant were scored within MOE using an apo structure with each substitution and no Rif present as a reference.

Conformational variants of Rifampicin were created in 3-D within MOE. A database of conformations was then used to dock the small molecule to the models using induced fit and template similarity protocols using the empirically derived structure. Placement was scored with Triangle Matcher and rescored with London $\Delta G$. Poses were refined with the Amber10:EHT forcefield with GVBI/WSA $\Delta G$ scoring. Candidate poses were then identified by inspecting polar contacts. Geometry optimization was carried out with MOPAC 7.0 using AM1. Conformational analysis of the bound structure was evaluated with LowModeMD (Labute, 2010a). RMSD values

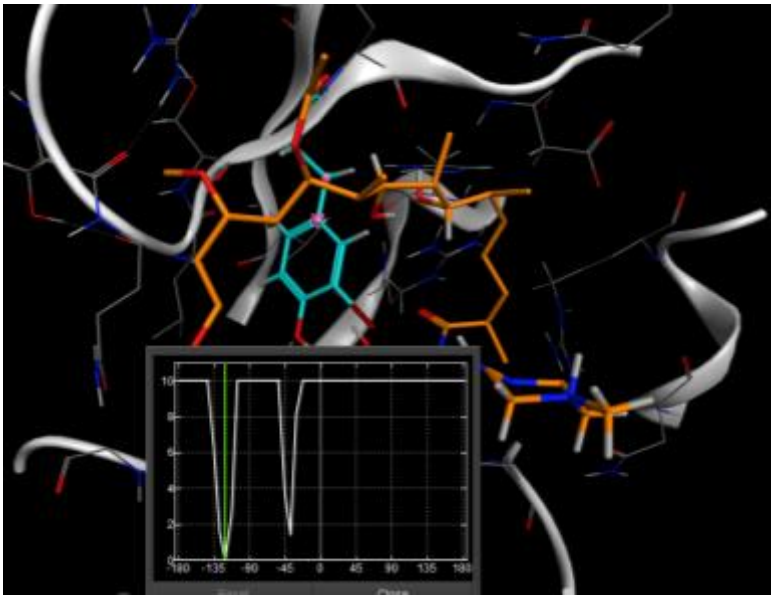
from the reference structure were calculated. 2-D contact maps of were created with Ligand Interactions (Clark and Labute, 2007).

Figure 2-9: Expanded genetic codes grant access to more protein coding solutions at Q513




Expanded genetic codes grant access to more protein coding solutions at Q513. (Top) In the canonical genetic code, the rifampicin-sensitive wildtype glutamine codon CAG (red) has 7 possible point-mutational routes to rifampicin resistance, with the CAG→TAG and the CAG→CAA mutations being inaccessible due to lethality and rifampicin sensitivity, respectively. Of the 7 available mutants, only 4 are reported to confer rifampicin resistance, and only 1 is a transition mutation. (Bottom) Incorporating NSAAs at the Amber codon alters the network topology surrounding the CAG codon. It provides an additional transition mutation, which confers resistance in the cases we studied. In addition, it creates new mutational routes to the codons directly adjacent to the TAG codon, several of which have been shown to be resistant in this study.

Figure 2-10: 3iY rotamer-by-dihedral example



3iY rotamer-by-dihedral example. We manipulated dihedral angles based on energy values to determine bond geometry. The NSAA is shown in blue. After adding the substituent to the 3 position of tyrosine, we determined the lowest energy conformation of each bond.

Table 2-7: 3iY Clashing observed using rotamer-by-dihedral method

| # | Chain | Atom A | | Chain | Atom B | Distance (Å) | Repulsion (kcal/mol) |
|---|---|---|---|---|---|---|---|
| 1 | 3:[4KMU.C]TYR | 513 | (CD1) | 7:[4KMU.C]RFP | 1401(O9) | 2.45 | 1.59 |
| 2 | 3:[4KMU.C]TYR | 513 | (CE1) | 7:[4KMU.C]RFP | 1401(C1) | 2.64 | 1.38 |
| 3 | 3:[4KMU.C]TYR | 513 | (CE1) | 7:[4KMU.C]RFP | 1401(C9) | 2.74 | 1.38 |
| 4 | 3:[4KMU.C]TYR | 513 | (HE1) | 7:[4KMU.C]RFP | 1401(O1) | 2.35 | 2.90 |
| 5 | 3:[4KMU.C]TYR | 513 | (HE1) | 7:[4KMU.C]RFP | 1401(C1) | 1.33 | 2.52 |
| 6 | 3:[4KMU.C]TYR | 513 | (HE1) | 7:[4KMU.C]RFP | 1401(C2) | 1.38 | 2.52 |
| 7 | 3:[4KMU.C]TYR | 513 | (HE1) | 7:[4KMU.C]RFP | 1401(N1) | 2.5 | 3.55 |
| 8 | 3:[4KMU.C]TYR | 513 | (HE1) | 7:[4KMU.C]RFP | 1401(C3) | 2 | 2.52 |
| 9 | 3:[4KMU.C]TYR | 513 | (HE1) | 7:[4KMU.C]RFP | 1401(C9) | 1.91 | 2.52 |
| 10 | 3:[4KMU.C]TYR | 513 | (HE1) | 7:[4KMU.C]RFP | 1401(C4) | 2.42 | 2.52 |
| 11 | 3:[4KMU.C]TYR | 513 | (HE1) | 7:[4KMU.C]RFP | 1401(C1) | 2.38 | 2.52 |
| 12 | 3:[4KMU.C]TYR | 513 | (CZ ) | 7:[4KMU.C]RFP | 1401(C9) | 2.8 | 1.38 |
| 13 | 3:[4KMU.C]TYR | 513 | (CZ ) | 7:[4KMU.C]RFP | 1401(C8) | 2.26 | 1.38 |
| 14 | 3:[4KMU.C]TYR | 513 | (CZ ) | 7:[4KMU.C]RFP | 1401(O2) | 2.04 | 1.59 |
| 15 | 3:[4KMU.C]TYR | 513 | (OH ) | 7:[4KMU.C]RFP | 1401(C9) | 2.15 | 1.59 |
| 16 | 3:[4KMU.C]TYR | 513 | (OH ) | 7:[4KMU.C]RFP | 1401(C8) | 1.09 | 1.59 |
| 17 | 3:[4KMU.C]TYR | 513 | (OH ) | 7:[4KMU.C]RFP | 1401(O2) | 0.87 | 1.83 |
| 18 | 3:[4KMU.C]TYR | 513 | (OH ) | 7:[4KMU.C]RFP | 1401(C7) | 2.1 | 1.59 |
| | | | | | | **Total** | **36.86** |

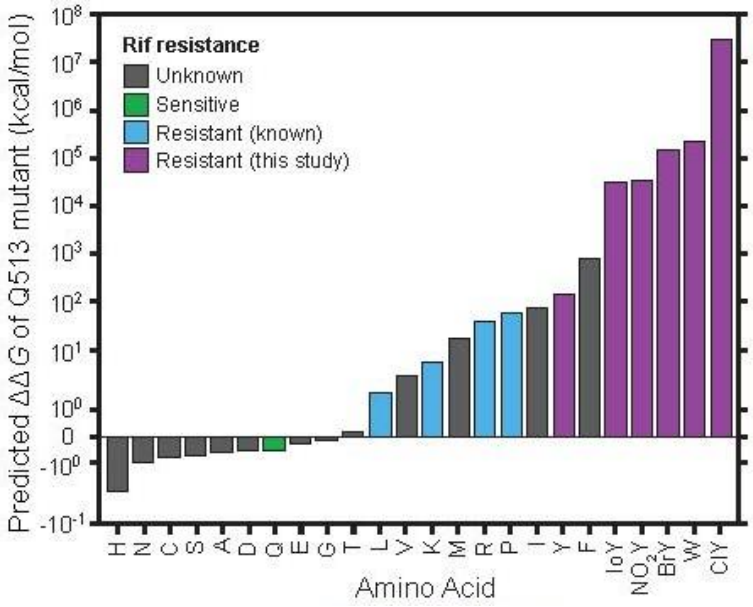
Table 2-8: Doubling rates and contact energies from rotamer-by-dihedral method

| | Doubling Rate | | Computed Energies (kcal/mol) | |
|---|---|---|---|---|
| **Amino Acid** | **Media** | **Rifampicin** | **Contact** | **PE** |
| 3-I-Y | 1.18 | 0.94 | 36.9 | -74286 |
| W | 1.03 | 1.03 | 29.6 | -74293 |
| 3-Cl-Y | 1.13 | 0.97 | 26.6 | -74286 |
| 3-Br-Y | 1.19 | 0.99 | 34.6 | -74283 |
| NitroY | 1.14 | 0.9 | 43.1 | -74282 |
| Y | 0.7 | 0.61 | 15.3 | -74298 |
| Q | 1.29 | n.d. | -9.8 | -74301 |

Table 2-9: Q513 *in silica* site saturation measurements

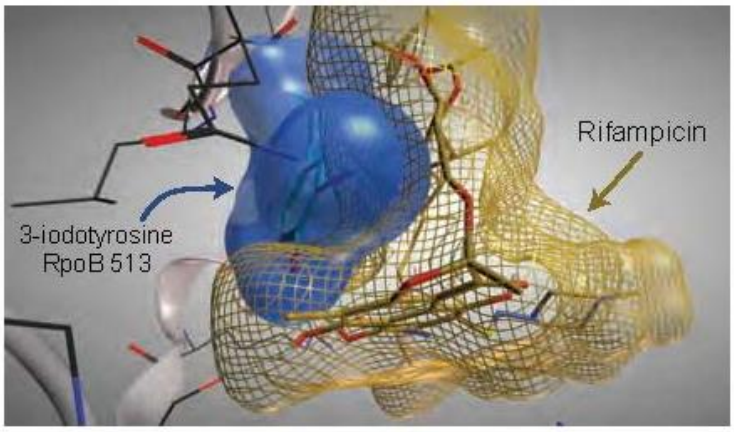
| | No Ligand | | Ligand | | | | |
|---|---|---|---|---|---|---|---|
| **Mutation** | **Stability** | **dStability** | **Affinity** | **dAffinity** | **Stability** | **dStability** | **ddStability** |
| 1:Q513Q | -5.94 | 0.00 | -11.81 | 0.00 | -6.44 | 0.00 | -0.50 |
| 1:Q513A | -5.80 | 0.13 | -11.03 | 0.79 | -6.40 | 0.03 | -0.60 |
| 1:Q513R | -6.61 | -0.67 | -6.85 | 4.96 | 31.06 | 37.50 | 37.67 |
| 1:Q513N | -5.50 | 0.44 | -11.59 | 0.23 | -6.46 | -0.02 | -0.97 |
| 1:Q513D | -4.39 | 1.54 | -11.35 | 0.46 | -4.90 | 1.54 | -0.51 |
| 1:Q513C | -5.57 | 0.37 | -11.18 | 0.63 | -6.38 | 0.06 | -0.81 |
| 1:Q513E | -4.99 | 0.95 | -11.26 | 0.55 | -5.29 | 1.14 | -0.30 |
| 1:Q513G | -3.22 | 2.71 | -10.79 | 1.02 | -3.35 | 3.09 | -0.13 |
| 1:Q513H | -2.17 | 3.77 | -11.49 | 0.33 | -4.27 | 2.17 | -2.10 |
| 1:Q513I | -7.48 | -1.54 | 28.35 | 40.17 | 65.64 | 72.08 | 73.12 |
| 1:Q513L | -7.30 | -1.37 | -10.18 | 1.64 | -5.68 | 0.76 | 1.63 |
| 1:Q513K | -4.77 | 1.17 | -9.63 | 2.19 | 0.89 | 7.32 | 5.65 |
| 1:Q513M | -6.68 | -0.74 | -11.33 | 0.48 | 10.84 | 17.27 | 17.52 |
| 1:Q513F | -8.24 | -2.30 | -10.66 | 1.15 | 788.47 | 794.90 | 796.71 |
| 1:Q513P | -4.76 | 1.18 | -11.25 | 0.56 | 54.06 | 60.50 | 58.82 |
| 1:Q513S | -5.36 | 0.57 | -11.15 | 0.67 | -6.09 | 0.35 | -0.73 |
| 1:Q513T | -5.96 | -0.02 | -10.60 | 1.21 | -5.81 | 0.62 | 0.15 |
| 1:Q513W | -9.62 | -3.68 | 155500.35 | 155512.16 | 229975.84 | 229982.28 | 229985.46 |
| 1:Q513Y | -8.01 | -2.07 | -10.01 | 1.80 | 141.08 | 147.52 | 149.09 |
| 1:Q513V | -6.79 | -0.85 | -10.39 | 1.42 | -3.99 | 2.44 | 2.80 |
| 1:Q513[(Bry)] | -9.21 | -3.27 | 7.47 | 19.28 | 146098.07 | 146104.50 | 146107.27 |
| 1:Q513[(Cly)] | -9.14 | -3.20 | 92.72 | 104.53 | 29136345.34 | 29136351.78 | 29136354.48 |
| 1:Q513[(Ioy)] | -9.49 | -3.55 | 115.20 | 127.01 | 31955.80 | 31962.24 | 31965.29 |
| 1:Q513[(Niy)] | -8.08 | -2.14 | 17.07 | 28.88 | 34162.07 | 34168.51 | 34170.15 |

Figure 2-11: Predicted changes in complex stability of Q513 site saturation



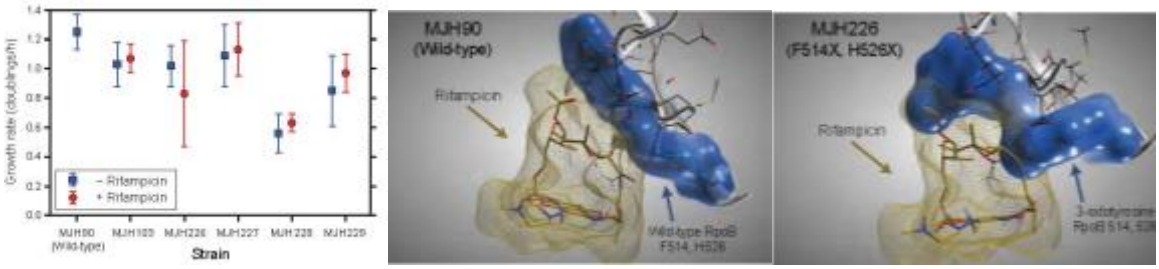

Predicted changes in complex stability of Q513 site saturation . Bars represent ΔΔ*G* values (kcal/mol) of protein stability calculated for the substitution of all 20 canonical amino acids and 4 nonstandard amio acids substituted at position 513 of *rpoB*. The wild-type amino acid (Q) is indicated in red, along with known resistance-conferring amino acids (cyan) and amino acids of unknown phenotype (black). 3-iodotyrosine (IoY), 3-nitrotrosine ($NO_2Y$), 3-bromotyrosine (BrY) and 3-chlorotyrosine (ClY) are modeled in addition to canonical amino acids. At the Q513 position, no amino acids with a negative ΔΔ*G* as calculated in MOE have been observed to confer rifampicin resistance, and are assumed to be rifampicin-sensitive. All resistant mutants have a positive value of ΔΔ*G*, with the nonstandard amino acids having the highest calculated values. Values are consistent with observed growth rates for tyrosine, tryptophan, and nsAA mutants at this position.

Figure 2-12: Q513X occludes rifampicin binding to rpoB




Q513X occludes rifampicin binding to rpoB. Modeling of Q513 3-iodotyrosine mutant (blue) in complex with a superimposed rifampicin molecule (gold). Space-filling models of the sidechain and rifampicin are depicted. A heavy steric clash is shown between the residue and antibiotic, presumably conferring resistance by occluding rifampicin from the complex.

Figure 2-13: Double-amber mutants confer robust rifampicin resistance




Double-amber mutants confer robust rifampicin resistance**. (**Left) Doubling times of $Rif^R$ *rpoB* mutants incorporating nsAAs in the absence (blue) or presence (red) of Rif. Error bars represent 95% confidence interval. The presence of Rif in the growth medium does not statistically impact the growth rate of any $Rif^R$ mutants. (Middle) Wild-type *rpoB* in complex with Rif (gold). Residues 514 and 526 are highlighted in blue. The orientation of the sidechains allows for a productive binding complex and inhibition of the enzyme. (Right) A double-amber mutant MJH227 (F514X, H526X) variant is shown with an overlay of Rif (gold) in the binding pocket. The occlusion of Rif from binding is achieved by steric occlusion involving the ortho-substituents of the nsAA. Rotamer exploration and docking failed to produce conformational variants that would allow for Rif: *rpoB* complex formation.

Table 2-10: Representative Double-Amber Mutation Binding Analysis

| Mutation | Affinity | dAffinity | Stability | dStability | E |
|---|---|---|---|---|---|
| 1:F514F | -11.4 | 0.0 | -7.6 | 0.0 | -21656.7 |
| 1:F514[(Ioy)] | 2.9 | 14.3 | 23.9 | 31.5 | -21484.4 |
| 1:H526H | -12.1 | 0.0 | 1.7 | 0.0 | -21614.8 |
| 1:H526[(Ioy)] | 2070.0 | 2082.1 | 118553899.4 | 118553897.6 | 645556290.0 |
| 1:F514F,1:H526H | -11.2 | 0.0 | -5.5 | 0.0 | -21606.9 |
| 1:F514F,1:H526[(Ioy)] | 2070.9 | 2082.1 | 118553892.1 | 118553897.6 | 645556290.0 |
| 1:F514[(Ioy)],1:H526H | 3.1 | 14.3 | 26.0 | 31.5 | -21433.6 |
| 1:F514[(Ioy)],1:H526[(Ioy)] | 2085.2 | 2096.5 | 118553923.5 | 118553929.1 | 645556420.0 |

## 2.4 CHARACTERIZATION OF A NOVEL ANGIOGENESIS INHIBITOR AND VASCULAR DISRUPTING AGENT

### *Introduction*

Evolutionarily conserved genetic traits often manifest in similar, obvious phenotypes. However, some phenotypes are nonobvious in distantly related organisms. Recently, the Marcotte laboratory developed a method that identifies conserved genetic modules among different species and found that lovastatin sensitivity in yeast and the regulation of blood vessel formation in vertebrates are controlled by conserved, distantly related genes (McGary et al., 2010). This somewhat surprising result gives credence to the notion that genes responsible for the maintenance of cell walls in yeast were later reconfigured to regulate vertebrate angiogenesis.

Angiogenesis refers to the formation of blood vessels and is essential for organismal development and tissue repair. Malfunctions in angiogenesis lead to major physiological defects (Carmeliet and Jain, 2011; Folkman, 2007). Cancer cells are also nourished by angiogenesis, and focused studies on anti-angiogenic therapeutics have rapidly increased in recent years (Carmeliet and Jain, 2011; Folkman, 2007; Heath and Bicknell, 2009; Kerbel, 2008). Vascular disrupting agents (VDAs) are a new class of anti-vascular drugs that have given promise for fighting developed tumors. However, VDAs have not been approved for human therapy, although a small number are currently undergoing clinical trials (Hinnen and Eskens, 2007; Tozer et al., 2005). In previous work, the Marcotte laboratory identified Thiabendazole (TBZ) as a novel angiogenesis inhibitor and vascular disrupting agent. Interestingly, TBZ has been approved by the FDA for treating various infections not only affecting humans but many other distantly related organisms. Moreover, the Marcotte laboratory demonstrated that TBZ also slows tumor growth and decreases intratumoral vessel density in mouse xenograft models (Cha et al., 2012).

Many VDAs bind endothelial tubulins, depolymerizing the microtubule cytoskeleton, and thus causing the vascular structure to collapse (Hinnen and Eskens, 2007; Tozer et al., 2005). TBZ

has been shown to work in a similar fashion on fungal infections of plants and animals. In fact, TBZ directly interacts with fungal tubulins and disturbs microtubule assembly (Davidse, 1986; Davidse and Flach, 1978). Thus, we hypothesized that TBZ might act in a similar or related fashion on human endothelial cells. Here we computationally assess TBZ binding to evolutionarily distant fungal and human tubulins. We present a model consistent with known resistance mutations in yeast and propose a molecular mechanism for anti-angiogenesis in TBZ-treated endothelial cells.

## Results & Discussion

### *Sequence Alignment of Human β-tubulins*

TBZ has previously been shown to directly interact with purified tubulins in the nematode *Ascaridia galli* and the fungus *Penicillium expansum* and perturbs polymerization of these proteins (Davidse and Flach, 1978; Dawson et al., 1984). The mitotic inhibition by TBZ observed in the worm and fungal studies was not replicated on tubulin assembly from mammalian brains (Dawson et al., 1984). Tubulin dynamics and interactions with other accessory proteins have been shown to be largely dependent upon its primary sequence (Asakawa et al., 2006; Paluh et al., 2004). For example, many TBZ-resistant mutations have been found in fungi and worms. Among those mutants, a number of studies have reported that one single amino acid mutation (F200Y) in the β-tubulin is sufficient to confer TBZ resistance in worm and fungus (Jung et al., 1992; Kotze et al., 2012; Lee et al., 2011; Paluh et al., 2004; Qiu et al., 2011; Schmidt et al., 2006). This conserved residue is positioned at the internal interface between two structural tubulin domains and is thus involved in inter-domain motion, which can cause conformational changes of tubulins and increase microtubule stability (Paluh et al., 2004). Notably, most β-tubulins in vertebrates have tyrosine at this position whereas some eukaryotes, (e.g. fungi, worm, and Antarctic fish) inherently have phenylalanine (Paluh et al., 2004).

Vertebrate species have many β-tubulin isoforms, and their usage and expression varies by cell and tissue type. In humans Tubb1 is expressed constitutively over a range of cells and tissues; however Tubb3 is exclusively found in the brain and neurons (Leandro-García et al., 2010; Sirajuddin et al., 2014). Interestingly, the mitotic inhibition by TBZ observed in the worm and fungal studies was not replicated on tubulin assembly from mammalian brains (Dawson et al., 1984). The specific roles of each tubulin isoform are not fully understood, but recent work show that sequence diversity seems to modulate binding affinity to tubulin-binding drugs, and shapes microtubule dynamics by interacting directly with molecular motors (Kavallaris, 2010; Sirajuddin et al., 2014). We hypothesized that some human tubulin isotypes might have the same amino acid as fungal tubulins at the position 200, and thus have increased susceptibility to TBZ. A multiple sequence alignment of human β-tubulin isotypes to wild type and mutant β-tubulin sequences of fission yeast indeed revealed that two human β-tubulin isotypes, *TUBB1* and *TUBB8*, have a phenylalanine at this position, while tyrosine is expressed in all other human β-tubulins at this position (**Figure 2-14**).

***Molecular Modeling Predictions of TBZ Binding to Fungal and Human β-tubulins***

Given the strong sequence conservation among β-tubulin isotypes and resistant phenotypes observed in some orthologs, we next sought to understand potential TBZ binding to both fission yeast and human isotypes at the molecular level. To do so, we employed molecular modeling and simulation. It was first necessary to create homology models, as high-resolution crystal structures of human and yeast versions of these proteins are not available. We searched the Protein Data Bank (PDB) for homologous structures and found 3UT5.B and 3N2G.D, corresponding to sheep brain β-tubulin, to be suitable reference structures. A total of 625 models were made for each isoform of β-tubulin and averaged to make each final model. Each model was: (1) encapsulated in a 6 Å

hydration sphere, (2) energy minimized using Amber10:EHT, (3) subjected to 2.5 ns NAMD molecular dynamics simulations, and (4) inspected for anomalies against the reference structure.

Using the *Schizosaccharomyces pombe* wildtype and the TBZ-restistant F200Y β-tubulin homology models, we employed MOE's Site Finder application to determine potential binding sites of TBZ to the energy minimized structures. A total of 20 sites were identified in the wildtype structure with the most probable site including F200 ((**Table 2-11; Figure 2-15**). In contrast, the single F200Y mutation, observed to be TBZ-resistant, greatly reduced the size of this potential site (**Table 2-12; Figure 2-16**). Next, we performed an induced fit docking protocol with TBZ at this site to see if we could observe differences in the susceptible wildtype protein and the resistant F200Y mutant. The induced fit protocol allowed us the most flexibility in sampling conformational space in both variants. As expected, TBZ fit comfortably within the proposed binding pocket of the wildtype β-tubulin but was incapable of adopting a productive binding conformation in the F200Y mutant (**Figure 2-17; Table 2-13**). Further, we found that the orientation of TBZ bound to the wildtype protein perfectly matched computational predictions on parasitic β-tubulins binding benzimidazoles (Aguayo-Ortiz et al., 2013a, 2013b) and recent crystal structures of benzimidazole-derivatives binding to bovine brain β-tubulins (McNamara et al., 2015). Template-based docking from the bovine structure yielded almost identical results to our previous induced fit docking experiments. In each case, the wildtype protein was shown to have productive binding energies, while F200Y was unable adopt a favorable conformation within the cleft and completely distorted the conformation of TBZ (**Tables 2-14 and 2-15**).

Next, we measured polar contacts and clashing energies of TBZ to both fungal tubulins. We found that the wildtype β-tubulin bound to TBZ with -9.9 kcal/mol of favorable contact energy, whereas F200Y was shown to contribute +27.6 kcal/mol of repulsion energy (**Figure 2-18**; **Table 2-16**). For the wildtype protein bound to TBZ, polar contacts are predicted to include E198 and

Q134. Arene-hydrogen interactions between the ligand and protein included contributions from F200, L250, and L253. For the F200Y mutant, repulsion was observed in our fixed ligand experiments, caused by TBZ making unfavorable contacts with Y200, F240, L250, and L253. From these data, it appears that the F200Y makes a hydrogen bond to E198 in the resistant mutant, thus constricting the pocket and occluding binding.

Confident in our docking results with the fungal tubulin, we next sought to model all human β-tubulin isoforms and predict TBZ binding energies (**Figure 2-19; Table 2-16**). Using a quantum mechanically corrected TBZ:wildtype fungal tubulin complex as a template, we docked TBZ into the same pocket in each of the human β-tubulins and measured interactions between ligand and protein in superimposed structures. In agreement with our primary sequence driven predictions for binding at position 200 (phenylalanine for susceptible and tyrosine for resistant), only Tubb1 and Tubb8 were shown to allow binding in our docking studies and displayed favorable binding energies, -1.8 and -8.3 kcal/mol, respectively. The large difference in contact energy among these two isoforms can be explained by differences at position 198. In Tubb1, position 198 is an alanine. In contrast, Tubb8 has a glutamate, which is predicted to heavily contribute to the binding energy in all of our simulations where F200 and E198 are present. Thus, we believe that TBZ binding is stabilized by hydrogen bonding interactions with residues Q134 and E198 in the context of F200, which allows TBZ into the binding pocket. Taken together, strong TBZ binding is predicted in isforms containing: Q134, E198, and F200. From this model, we conclude that TBZ strongly binds Tubb8, may only weakly bind Tubb1, and does not bind the rest of the human β-tubulin isoforms (**Table 2-16**).

***Conclusions***

The conserved residues at positions 134, 198 and 200, which seem to determine thiabendazole susceptibility from yeast to humans, have also been shown to be essential for the

activity of many benzimidazoles in fungus (Koenraadt, 1992; Yarden and Katan, 1993). Since some benzimidazoles are reported to be effective for cancer therapy (Hammond et al., 2001; Mukhopadhyay et al., 2002), our modeling suggests that the composition of tubulin isotypes in cancer cells may need to be considered before treatment. Microtubule targeting agents have long been used as classical cytotoxic cancer drugs, but more recently, a number of studies have reported that at low concentrations, many of them also have anti-angiogenic properties (Bocci et al., 2002; Pasquier et al., 2005; Pourroy et al., 2006). Although some studies suggest that microtubule targeting agents at low concentration might increase microtubule dynamic instability specifically in endothelial cells (Pasquier et al., 2005; Pourroy et al., 2006), the cellular and molecular mechanism—how these drugs have the selective activity on endothelial cells—is not fully understood. Encouragingly, we expect that specific effect of thiabendazole on the cellular component in endothelial cells and on different human tubulin isotypes can be employed as a model to understand the general mechanism of anti-angiogenic microtubule targeting agents. Understanding the detailed mechanism of repurposed drug is important for extending these drugs' therapeutic usage. This research provides a great example of using selective conservation of genes and modeling to determine the mode of action of a drug.

## Methods & Materials

### *Molecular Modeling of β-Tubulins*

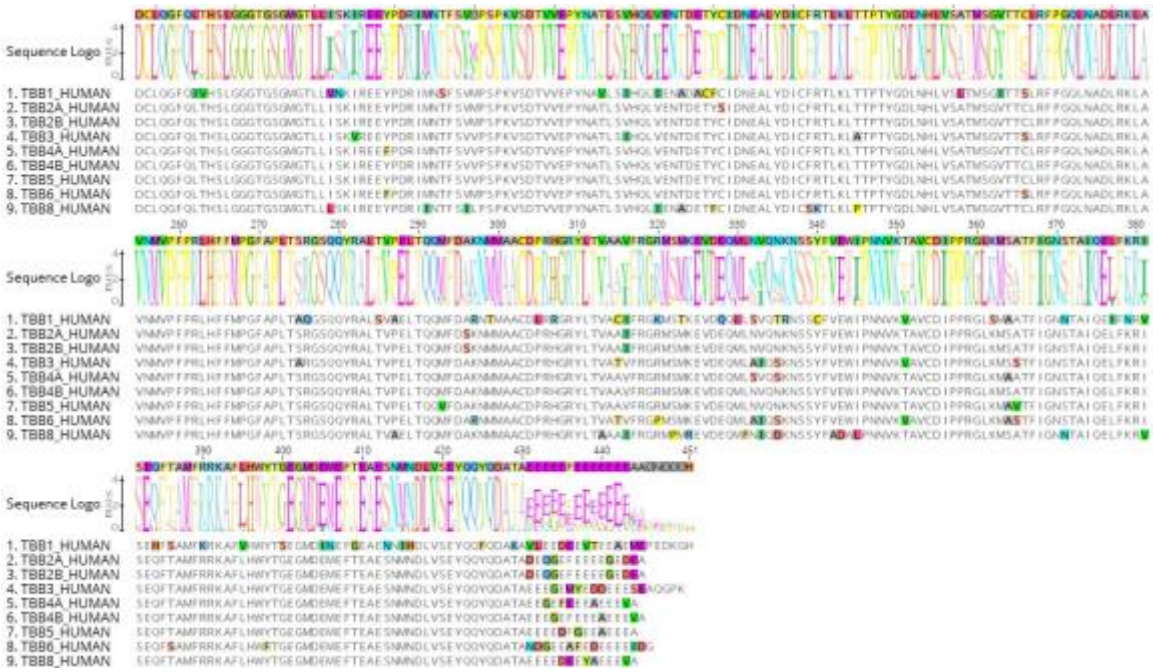
Homology models of human and yeast β-tubulins were constructed using two reference structures (PDB: 3UT5.B and 3N2G.D) (Barbier et al., 2010; Ranaivoson et al., 2012)**.** The template was prepared using the Molecular Operating Environment (MOE.09.2014) software package from Chemical Computing Group. The structure was inspected for anomalies and protonated/charged with the Protonate3D subroutine (310K, pH 7.4, 0.1 M salt). The protonated structure was then lightly tethered to reduce significant deviation from the empirically determined

coordinates and minimized using the Amber10:EHT forcefield with R-field treatment of electrostatics to an RMS gradient of 0.1 kcal $mol^{-1}$ $Å^{-1}$. Homology models of the wild-type yeast β-tubulin were prepared by creating 25 main chain models with 25 sidechain samples at 298K (625 total) within MOE. Intermediates were refined to an RMS gradient of 1 kcal $mol^{-1}$ $Å^{-1}$, scored with the GB/VI methodology, minimized again to an RMS gradient of 0.5 kcal $mol^{-1}$ $Å^{-1}$, and protonated. The final model for each variant was further refined by placing the protein within a 6 Å water sphere and minimizing the solvent enclosed structure to an RMS gradient of 0.001 kcal $mol^{-1}$ $Å^{-1}$. Models were evaluated by calculating Phi-Psi angles and superimposed against the reference structure.

***TBZ Docking***

Potential binding sites were evaluated using the Site Finder application and recent computational work on benzimidazole binding to parasitic β-tubulins **.** Conformational variants of TBZ were created in 3-D within MOE. A database of conformations was then used to dock TBZ to the wild-type homology model using induced fit and template similarity protocols. Placement was scored with Triangle Matcher and rescored with London dG. Poses were refined with the Amber10:EHT forcefield with GVBI/WSA dG scoring. Candidate poses were then identified by inspecting polar contacts. Geometry optimization was carried out with MOPAC 7.0 using AM1. Conformational analysis of the bound structure was evaluated with LowModeMD. 2-D contact maps of were created with Ligand Interactions.

Figure 2-14: Human β-tubulin Alignment

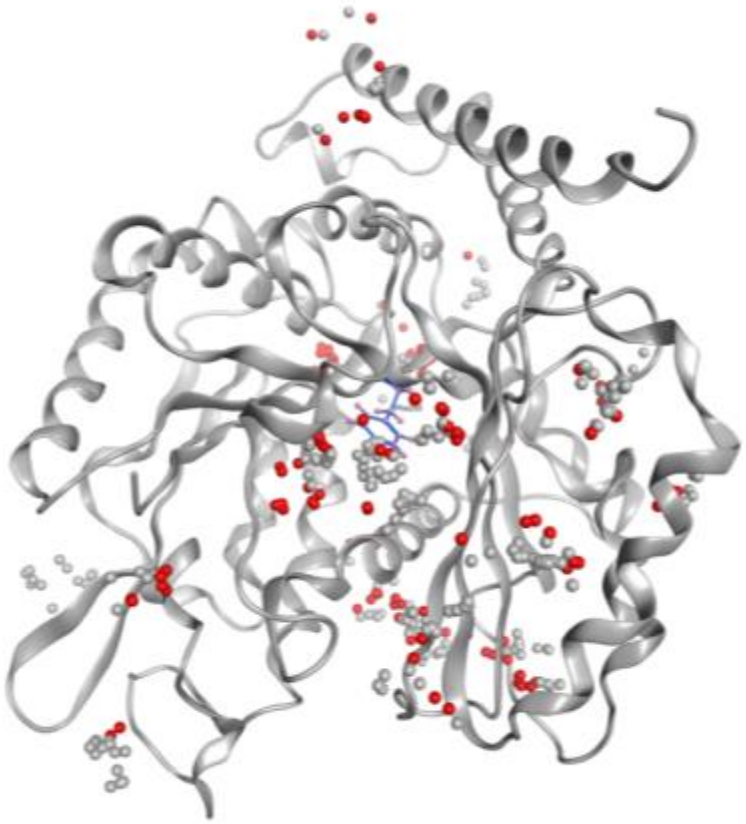



Human β-tubulin alignment. Strong sequence conservation is observed in all human isoforms. Key positions for this study are 134, 198, and 200. Two human tubulins, TUBB1 and TUBB8, have phenylalanine at position 200 similar to the wild type fungal tubulin, while all other human tubulins have tyrosine, which corresponds to the tubulin from TBZ-resistant yeast. TUBB1 also differs from fungal and other human tubulins at position 198 (E198A).

Table 2-11: Wildtype Yeast Site Finder Statistics

| Site | Size | PLB | Hyd | Side | Residues |
|---|---|---|---|---|---|
| 1 | 55 | 2.89 | 38 | 50 | 1:(ILE4 VAL49 TYR50 GLN134 ALA165 PHE167 GLU198 PHE200 VAL236 THR237 SER239 PHE240 LEU250 LEU253 ALA254 MET257 PHE266 ALA314 LEU316 ILE368) |
| 2 | 60 | 1.97 | 16 | 40 | 1:(GLN15 ALA18 ALA19 SER22 THR23 ASP223 ASN226 HIS227 SER230 PRO272 LEU273 ALA274 ALA275 ILE276 GLY277 SER278 SER279 SER280 LYS359 ASP360 LEU361) |
| 3 | 60 | 1.78 | 13 | 35 | 1:(GLN11 CYS12 GLY98 ASN99 SER138 LEU139 GLY140 GLY141 VAL169 ALA170 PRO171 ALA172 SER175 SER176 ASP177 THR178 GLU181 ASN204 TYR222) |
| 4 | 24 | 0.24 | 8 | 18 | 1:(ARG331 GLN334 SER338 PHE341 VAL342 GLU343 ILE345 PRO346 ASP347 ASN348 VAL349) |
| 5 | 21 | 0.01 | 6 | 14 | 1:(PRO243 GLY244 GLU245 VAL321 SER322 MET323 ASP327 LYS351 ALA352 VAL353 CYS354 SER355) |
| 6 | 11 | -0.06 | 6 | 15 | 1:(ALA172 PRO173 SER176 ASP177 THR178 VAL179 VAL180 GLU181 PRO182 MET388) |
| 7 | 19 | -0.07 | 12 | 19 | 1:(THR23 HIS227 SER230 ALA231 ALA234 PHE270 PRO358 LYS359 LEU361) |
| 8 | 17 | -0.1 | 11 | 15 | 1:(PRO272 SER279 SER280 PHE281 ALA283 VAL284 LEU361 LYS362 MET363) |
| 9 | 20 | -0.15 | 7 | 19 | 1:(GLY10 GLN11 ASP67 GLU69 THR72 GLY98 ASN99 SER138 GLY141 GLY142 THR143 GLY144) |
| 10 | 21 | -0.21 | 6 | 10 | 1:(ASN256 MET257 VAL313 ALA314 ILE345 PRO346 ASP347 ASN348 VAL349 LEU350) |
| 11 | 16 | -0.37 | 10 | 14 | 1:(GLU245 LEU253 ASN256 MET257 ALA314 LEU316 LEU350 LYS351 ALA352) |
| 12 | 13 | -0.4 | 7 | 12 | 1:(LYS174 SER175 SER208 SER219 PRO220 SER221 TYR222 LEU225) |
| 13 | 27 | -0.42 | 11 | 20 | 1:(ASP327 ILE330 ARG331 GLN334 ASP347 VAL349 LEU350 LYS351) |
| 14 | 15 | -0.52 | 8 | 16 | 1:(TYR36 HIS44 GLY57 TYR59) |
| 15 | 12 | -0.56 | 6 | 20 | 1:(HIS190 VAL193 GLU194 ARG262 GLU410 SER413 ASN414 ASP417) |
| 16 | 9 | -0.65 | 12 | 21 | 1:(THR290 GLN291 PHE294 ASP295 ARG306 VAL333 LYS336 ASN337) |
| 17 | 12 | -0.67 | 5 | 8 | 1:(VAL60 ASN83 LEU84 PHE85 ARG86 PRO87) |
| 18 | 11 | -0.74 | 5 | 12 | 1:(MET1 ARG2 GLU45 ARG46 ASN48 VAL49) |
| 19 | 6 | -0.86 | 4 | 4 | 1:(ILE210 ALA296 VAL301 ALA302 ALA303 ASP304) |
| 20 | 4 | -1.13 | 4 | 11 | 1:(ASP404 MET406 GLU407 THR409 GLU410) |

Figure 2-15: Wildtype yeast site finder locations

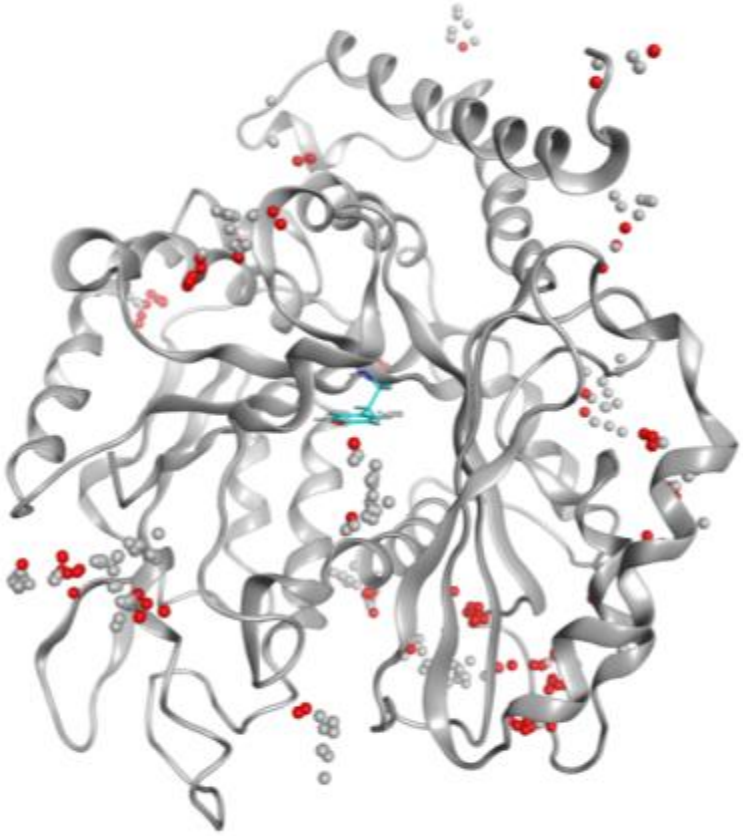


Wildtype yeast site finder locations. The SiteFinder application within MOE was used to assess potential binding sites in the wildtype fungal β-tubulin. Alpha centers, red and white spheres, are depicted. Red denote potential hydrophilic atoms, and white denote potential hydrophobic atoms. F200, neighboring the largest pocket, is shown in blue for context. This pocket was used in subsequent docking studies. Statistics for sites shown here can be found in Table 2-11.

Table 2-12: Yeast F200Y Site Finder Statistics

| Site | Size | PLB | Hyd | Side | Residues |
|---|---|---|---|---|---|
| 1 | 42 | 2.85 | 17 | 33 | 3:(THR23 HIS227 LEU228 SER230 ALA231 PHE270 PRO272 LEU273 ALA274 ALA275 ILE276 SER278 SER279 SER280 PRO358 LYS359 ASP360 LEU361) |
| 2 | 20 | 0.72 | 2 | 14 | 3:(ARG156 PRO160 ASP161 ARG162 VAL193 GLU194 ASN195 SER196 ASP197 ARG251 PRO261 ARG262 HIS264) |
| 3 | 19 | 0.69 | 12 | 21 | 3:(PRO272 SER280 PHE281 GLN282 ALA283 VAL284 ASP360 LEU361 LYS362 MET363) |
| 4 | 20 | 0.57 | 3 | 20 | 3:(GLY10 GLN11 CYS12 ASN99 SER138 GLY141 SER176 ASP177 THR178 GLU181 TYR222) |
| 5 | 24 | 0.56 | 8 | 21 | 3:(GLU3 ASN48 PHE51 ASN52 GLU53 ARG62 ALA126 CYS127 ASP128) |
| 6 | 18 | 0.45 | 12 | 17 | 3:(SER239 GLU245 LYS252 LEU253 ASN256 MET257 ALA314 LEU316 LEU350 LYS351 ALA352) |
| 7 | 7 | 0.29 | 10 | 19 | 3:(TRP101 TYR106 LEU187 HIS190 GLN191 GLU407 GLU410) |
| 8 | 23 | 0.25 | 5 | 19 | 3:(GLN15 ALA18 ALA19 SER22 THR23 ASP223 HIS227 SER230 LYS359) |
| 9 | 10 | 0.16 | 5 | 9 | 3:(PHE242 PRO243 GLY244 VAL353 CYS354 SER355 VAL356) |
| 10 | 17 | 0.14 | 9 | 14 | 3:(LEU215 PRO272 LEU273 ALA274 SER278 SER279 SER280 PHE281 VAL284 ASP360 LEU361) |
| 11 | 33 | 0.09 | 7 | 24 | 3:(MET1 ARG2 GLU45 ARG46 ASN48 VAL49 ASN247 SER248 ASP249 LYS252) |
| 12 | 19 | -0.13 | 12 | 18 | 3:(LEU68 TYR92 GLY93 GLN94 SER95 GLU108 GLU111 LEU112 ALA115) |
| 13 | 15 | -0.16 | 13 | 26 | 3:(GLN291 PHE294 ASP295 PRO305 ARG306 TYR310 VAL333 ASN337) |
| 14 | 15 | -0.19 | 5 | 6 | 3:(GLU205 ALA206 ILE210 ALA296 ASN297 VAL301 ALA302 ALA303 ASP304) |
| 15 | 10 | -0.51 | 4 | 19 | 3:(HIS307 GLY308 SER373 GLN375 GLU376 ARG379 GLN423 GLN426) |
| 16 | 7 | -1.15 | 6 | 7 | 3:(PHE389 ARG390 ARG391 LYS392) |
| 17 | 5 | -1.37 | 5 | 7 | 3:(GLN423 GLN424 GLU427 ALA428 ASP431 GLU432) |
| 18 | 4 | -1.57 | 4 | 11 | 3:(GLN291 GLN292 ASP295 ASN297 ASN298) |
| 19 | 5 | -1.69 | 8 | 12 | 3:(ALA271 PRO272 VAL284 GLU288 LEU289 GLN292 ASN298) |

Figure 2-16: Yeast F200Y Site Finder Locations



Yeast F200Y site finder locations. The SiteFinder application within MOE was used to assess potential binding sites in the F200Yfungal β-tubulin. Alpha centers, red and white spheres, are depicted. Red denote potential hydrophilic atoms, and white denote potential hydrophobic atoms. Y200, the position of TBZ-resistance, is shown for context. In contrast the wildtype, this pocket is considerably smaller and does not allow TBZ docking (Table 2-12, Figure 2-18).

Figure 2-17: Wildtype yeast (Nda3) 3D Model and 2D Contact Map

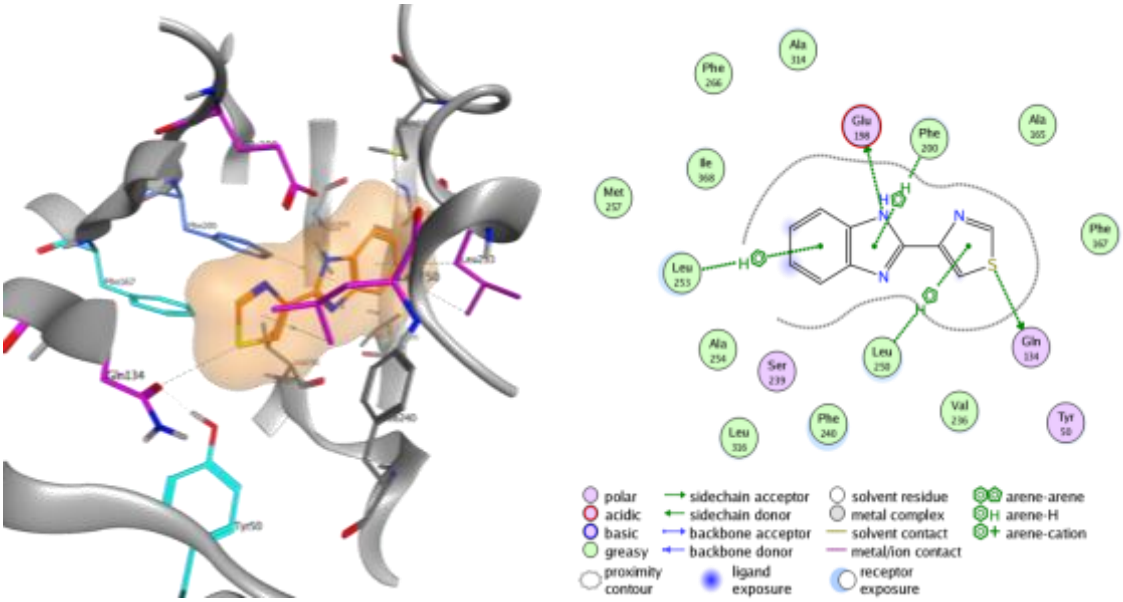



**Figure 2-17.** Left, wild-type yeast β-tubulin bound to TBZ. Right, 2-D contact map. Polar contacts contribute -9.9 kcal/mol.

Table 2-13: Wildtype yeast induced fit docking of TBZ

| Pose | S | rmsd | rmsd_refine | E_conf | E_place | E_score1 | E_refine | E_score2 |
|---|---|---|---|---|---|---|---|---|
| 1 | -7.21 | 0.20 | 0.78 | 20.35 | -65.04 | -9.78 | -27.90 | -7.21 |
| 2 | -7.15 | 5.53 | 0.69 | 19.77 | -74.50 | -10.09 | -25.25 | -7.15 |
| 3 | -7.11 | 5.62 | 1.04 | 21.83 | -65.20 | -9.54 | -26.64 | -7.11 |
| 4 | -7.08 | 5.52 | 0.87 | 21.03 | -67.68 | -9.52 | -24.44 | -7.08 |
| 5 | -6.95 | 1.58 | 0.58 | 22.12 | -65.14 | -9.56 | -25.61 | -6.95 |
| 6 | -6.87 | 5.84 | 0.95 | 19.70 | -57.47 | -9.50 | -23.99 | -6.87 |
| 7 | -6.87 | 1.90 | 1.28 | 19.83 | -58.92 | -9.48 | -24.49 | -6.87 |
| 8 | -6.85 | 5.74 | 0.79 | 21.92 | -57.94 | -9.48 | -22.60 | -6.85 |

Table 2-14: Wildtype yeast template-based docking of TBZ

| Pose | S | rmsd | rmsd_refine | E_conf | E_place | E_score1 | E_refine | E_score2 |
|---|---|---|---|---|---|---|---|---|
| 1 | -1.07 | 7.45 | 1.60 | 49.16 | -11.70 | -7.32 | 32.45 | -1.07 |
| 2 | -1.00 | 5.22 | 1.84 | 44.04 | -11.44 | -7.44 | 27.97 | -1.00 |
| 3 | -0.90 | 7.24 | 1.29 | 51.13 | -8.21 | -8.12 | 34.40 | -0.90 |
| 4 | -0.69 | 4.24 | 1.45 | 49.57 | -9.33 | -7.21 | 26.34 | -0.69 |
| 5 | 0.16 | 5.19 | 1.07 | 46.44 | -14.27 | -7.54 | 40.00 | 0.16 |
| 6 | 0.60 | 5.18 | 3.12 | 53.73 | -11.56 | -7.48 | 42.84 | 0.60 |
| 7 | 0.76 | 10.79 | 1.05 | 61.73 | -8.30 | -8.07 | 65.27 | 0.76 |
| 8 | 1.67 | 10.69 | 0.81 | 70.74 | -9.42 | -8.41 | 76.12 | 1.67 |
| 9 | 1.70 | 8.48 | 1.25 | 64.83 | -9.86 | -7.68 | 68.17 | 1.70 |
| 10 | 1.81 | 9.06 | 0.57 | 62.10 | -8.71 | -9.91 | 70.72 | 1.81 |
| 11 | 1.84 | 4.32 | 1.10 | 47.94 | -12.64 | -7.14 | 47.87 | 1.84 |
| 12 | 2.30 | 8.29 | 1.05 | 80.35 | -8.43 | -7.95 | 74.57 | 2.30 |
| 13 | 3.18 | 8.12 | 0.78 | 62.00 | -8.38 | -7.34 | 78.27 | 3.18 |
| 14 | 3.89 | 7.89 | 1.07 | 59.92 | -8.49 | -8.69 | 84.22 | 3.89 |
| 15 | 4.06 | 8.00 | 1.41 | 70.22 | -8.86 | -9.63 | 88.26 | 4.06 |
| 16 | 4.37 | 10.54 | 0.67 | 58.71 | -10.22 | -7.84 | 92.19 | 4.37 |
| 17 | 7.91 | 9.08 | 0.98 | 97.81 | -8.56 | -7.92 | 121.22 | 7.91 |

Table 2-15 F200Y Template Based Docking

| Pose | S | rmsd | rmsd_refine | E_conf | E_place | E_score1 | E_refine | E_score2 |
|---|---|---|---|---|---|---|---|---|
| 1 | 20.28 | 2.98 | 2.46 | 96.63 | -12.96 | -6.26 | 215.64 | 20.28 |
| 2 | 20.86 | 3.23 | 2.17 | 79.13 | -15.25 | -7.43 | 219.28 | 20.86 |
| 3 | 21.11 | 2.65 | 2.31 | 96.37 | -15.06 | -6.30 | 223.52 | 21.11 |
| 4 | 21.78 | 2.94 | 2.18 | 80.39 | -12.75 | -6.18 | 227.54 | 21.78 |
| 5 | 29.35 | 6.07 | 1.24 | 101.64 | -12.47 | -6.44 | 292.03 | 29.35 |
| 6 | 30.39 | 5.90 | 0.69 | 107.12 | -10.82 | -6.32 | 300.17 | 30.39 |
| 7 | 31.85 | 6.14 | 1.14 | 96.91 | -11.64 | -6.68 | 313.67 | 31.85 |
| 8 | 32.80 | 5.96 | 0.88 | 105.49 | -13.39 | -6.63 | 321.02 | 32.80 |

Figure 2-18: Yeast F200Y 3D Model and 2D Contact Map

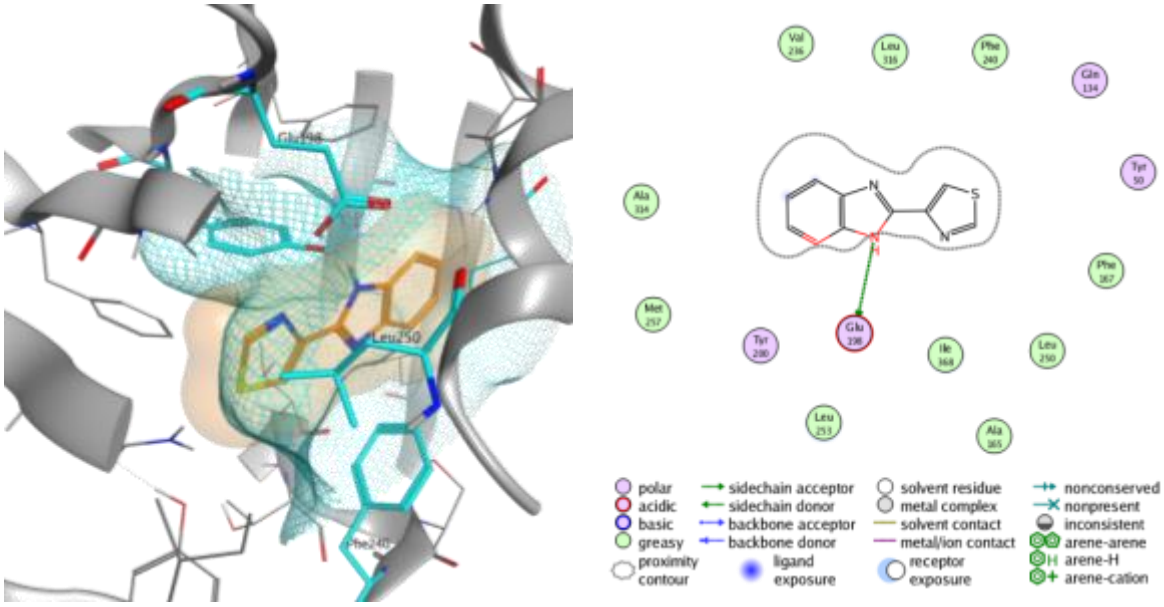



**Figure 2-18.** Left, F200Y yeast β-tubulin superimposed with TBZ. Clashing residues shown in cyan. Right, 2-D contact map. Constriction of the pocket leads to 27.6 kcal/mol of clashing energy to TBZ.

Table 2-16: TBZ contact energy and binding predictions to tubulin variants

| Model | Contacts | Predicted Binding |
|---|---|---|
| nda_3 | -9.9 | +++ |
| nda_3 F200Y | 27.6 | - |
| hTubb1 | -1.8 | + |
| hTubb2a | 38.3 | - |
| hTubb2b | 8.8 | - |
| hTubb3 | 11.4 | - |
| hTubb4a | 2.6 | - |
| hTubb4b | 18.3 | - |
| hTubb5 | 10.7 | - |
| hTubb6 | 12.3 | - |
| hTubb8 | -8.3 | +++ |

Figure 2-19: Human Tubb1 3D Model and 2D Contact Map

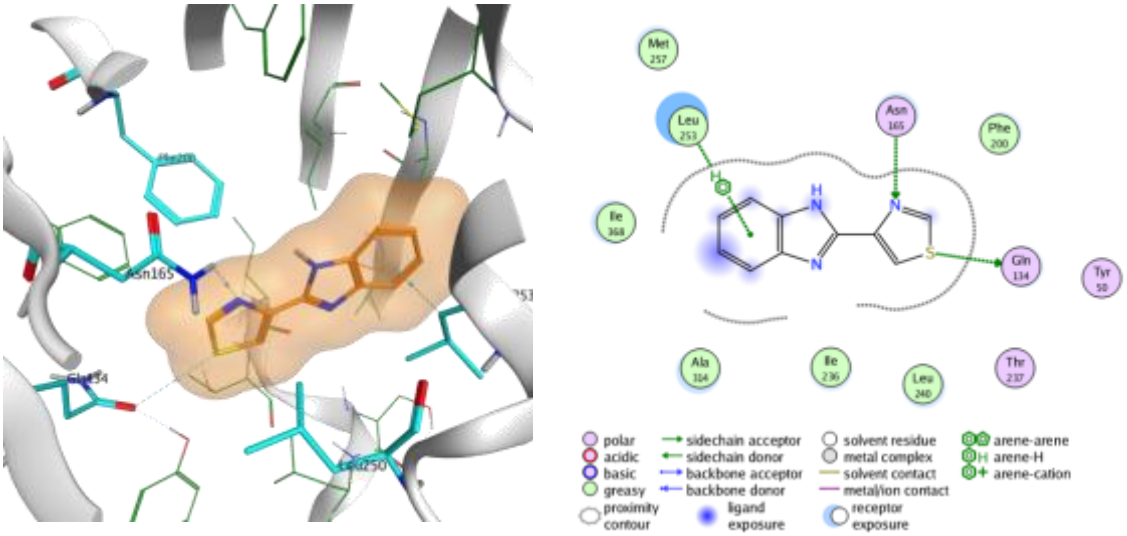



**Figure 2-19.** Left, Tubb1 bound to TBZ. Right, 2-D contact map. Contacts contribute -3.8 kcal/mol

Figure 2-20: Human Tubb2A 3D Model and 2D Contact Map

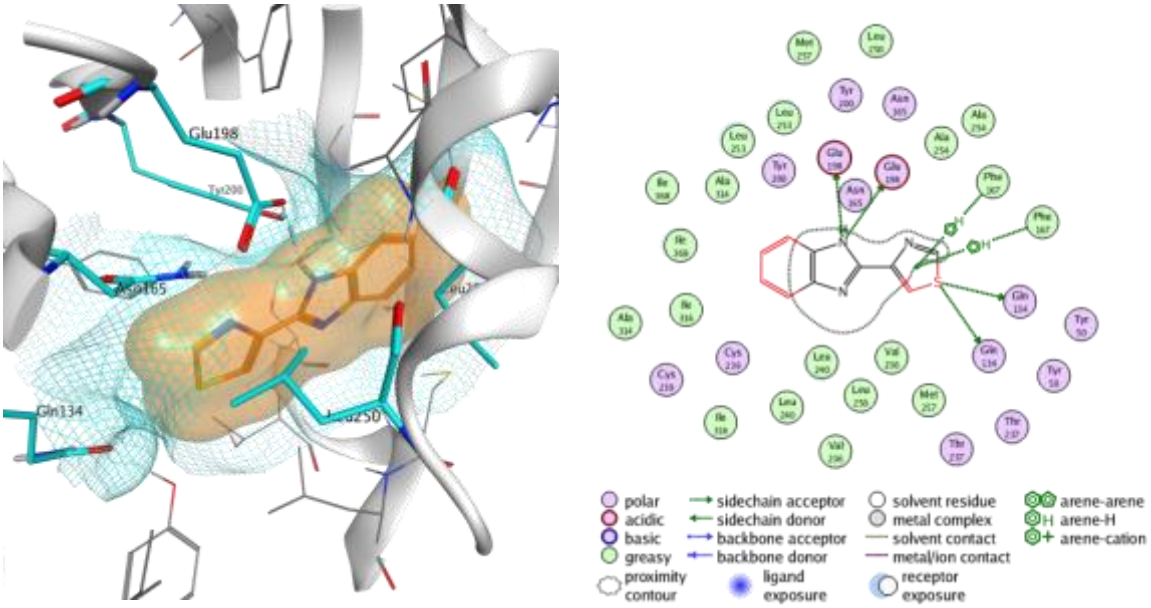



**Figure 2-20.** Left, Tubb2a superimposed with TBZ. Clashing residues shown in cyan. Right, 2-D contact map. Constriction of the pocket leads to 38.3 kcal/mol of clashing energy to TBZ.

Figure 2-21: Human Tubb2b 3D Model and 2D Contact Map

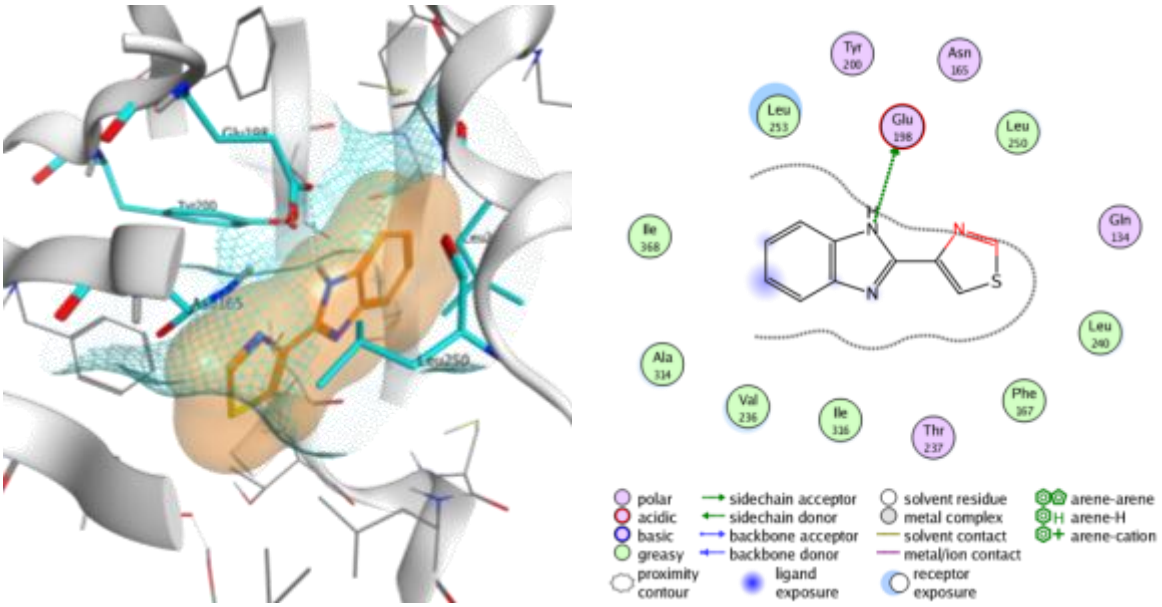



**Figure 2-21.** Left, Tubb2b superimposed with TBZ. Clashing residues shown in cyan. Right, 2-D contact map. Constriction of the pocket leads to 8.8 kcal/mol of clashing energy to TBZ.

Figure 2-22: Human Tubb3 3D Model and 2D Contact Map

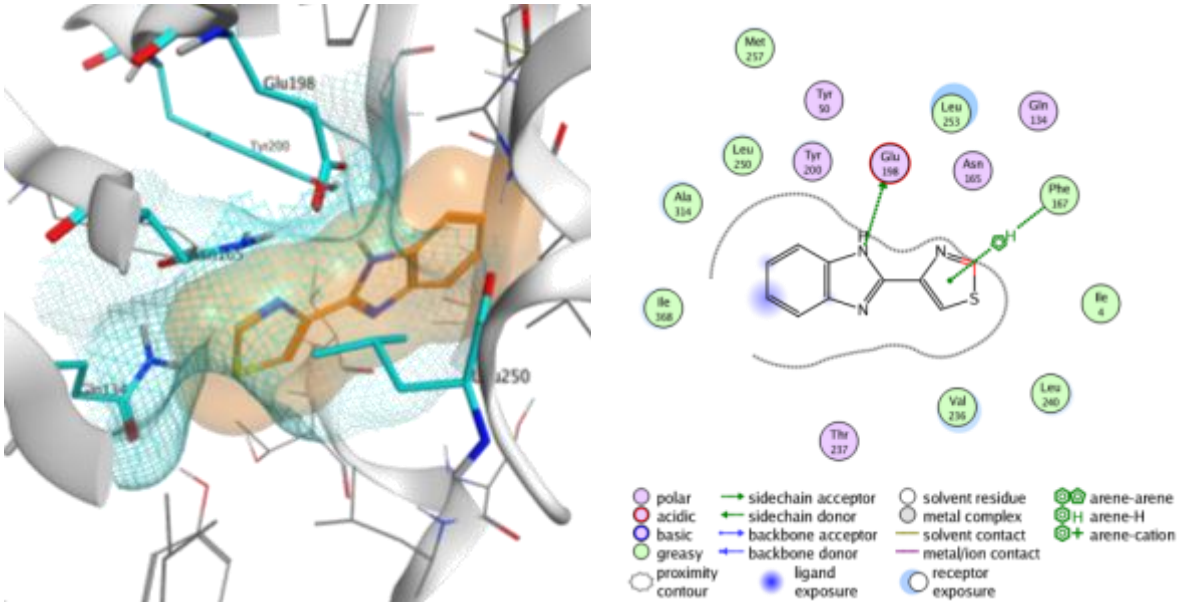



**Figure 2-22.** Left, Tubb3 superimposed with TBZ. Clashing residues shown in cyan. Right, 2-D contact map. Constriction of the pocket leads to 11.4 kcal/mol of clashing energy to TBZ.

Figure 2-23: Human Tubb4a 3D Model and 2D Contact Map

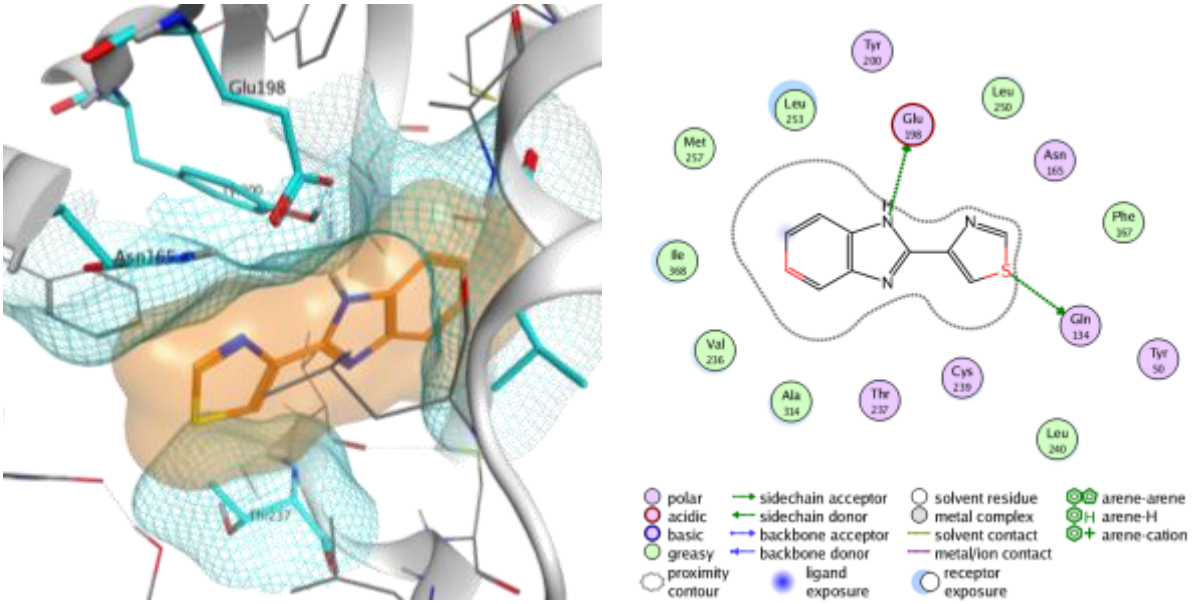
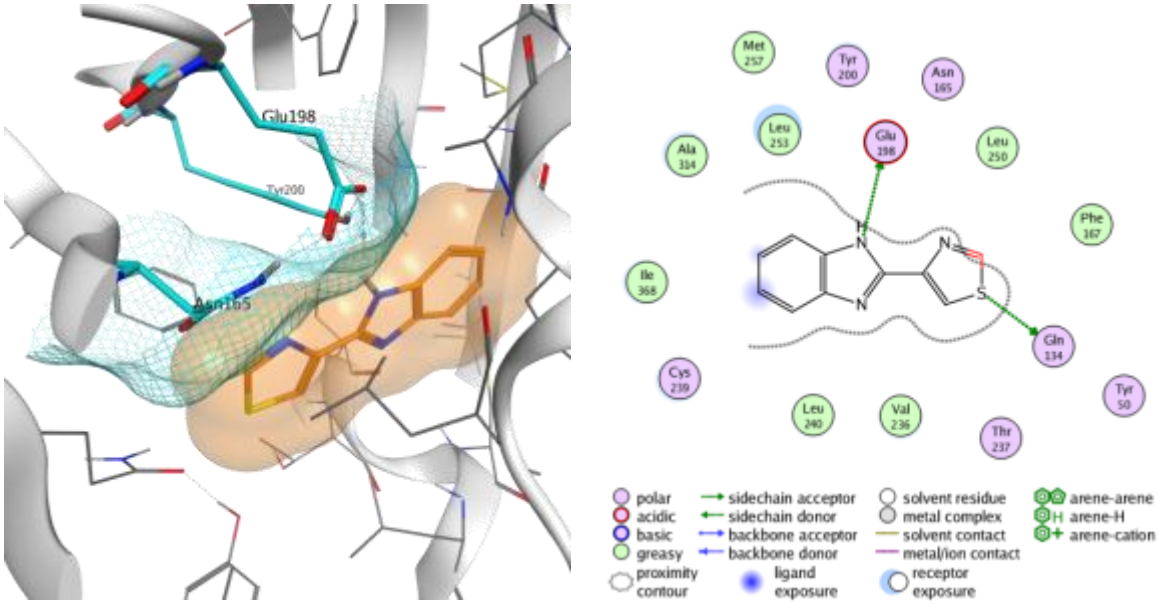



**Figure 2-23.** Left, Tubb4a superimposed with TBZ. Clashing residues shown in cyan. Right, 2-D contact map. Constriction of the pocket leads to 2.6 kcal/mol of clashing energy to TBZ.

Figure 2-24: Human Tubb4b 3D Model and 2D Contact Map




Left, Tubb4b superimposed with TBZ. Clashing residues shown in cyan. Right, 2-D contact map.

Figure 2-25: Human Tubb5 3D Model and 2D Contact Map

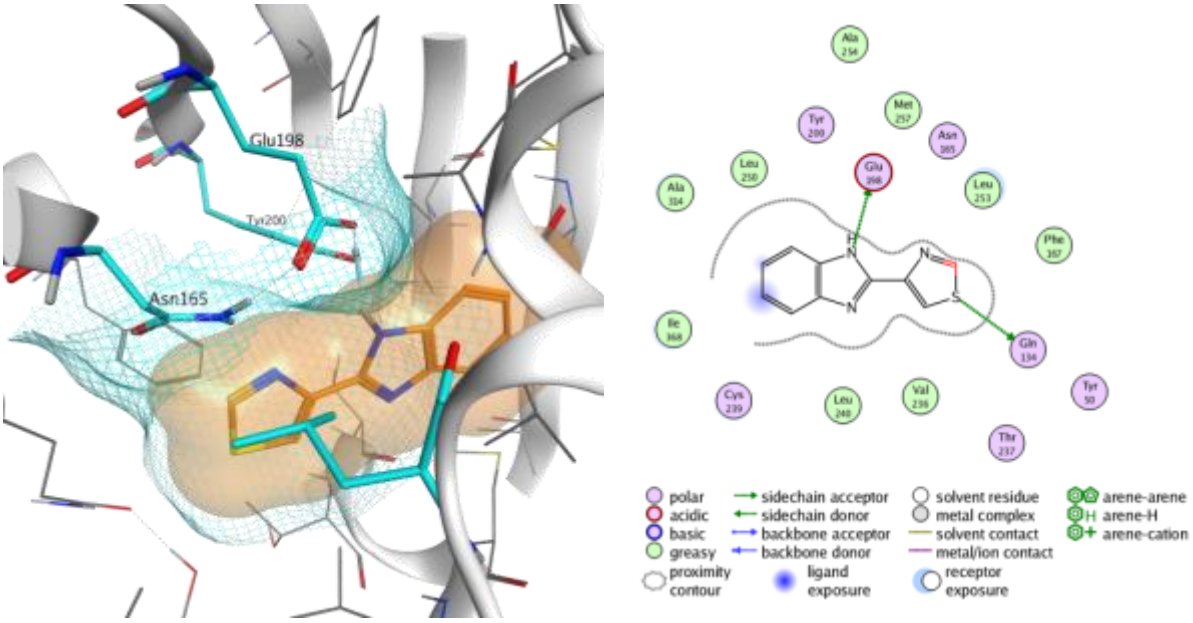



**Figure 2-25.** Left, Tubb5 superimposed with TBZ. Clashing residues shown in cyan. Right, 2-D contact map. Constriction of the pocket leads to 10.7 kcal/mol of clashing energy to TBZ.

Figure 2-26: Human Tubb63D Model and 2D Contact Map

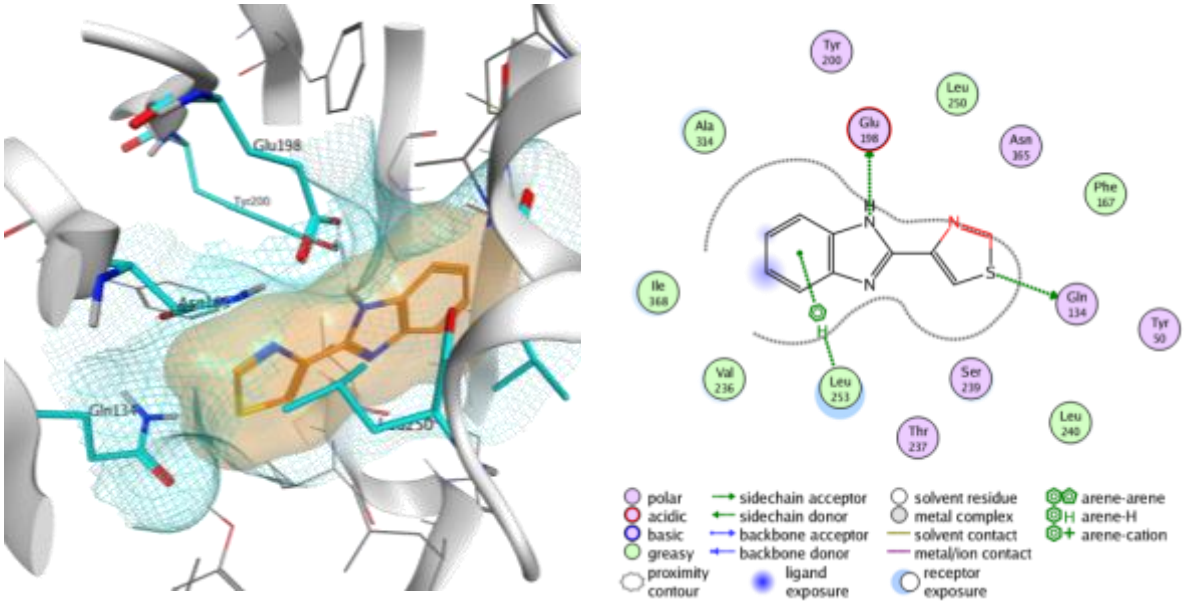



**Figure 2-26.** Left, human Tubb6 superimposed with TBZ. Clashing residues shown in cyan. Right, 2-D contact map. Constriction of the pocket leads to 12.3 kcal/mol of clashing energy to TBZ.

Figure 2-27: Human Tubb8 3D Model and 2D Contact Map

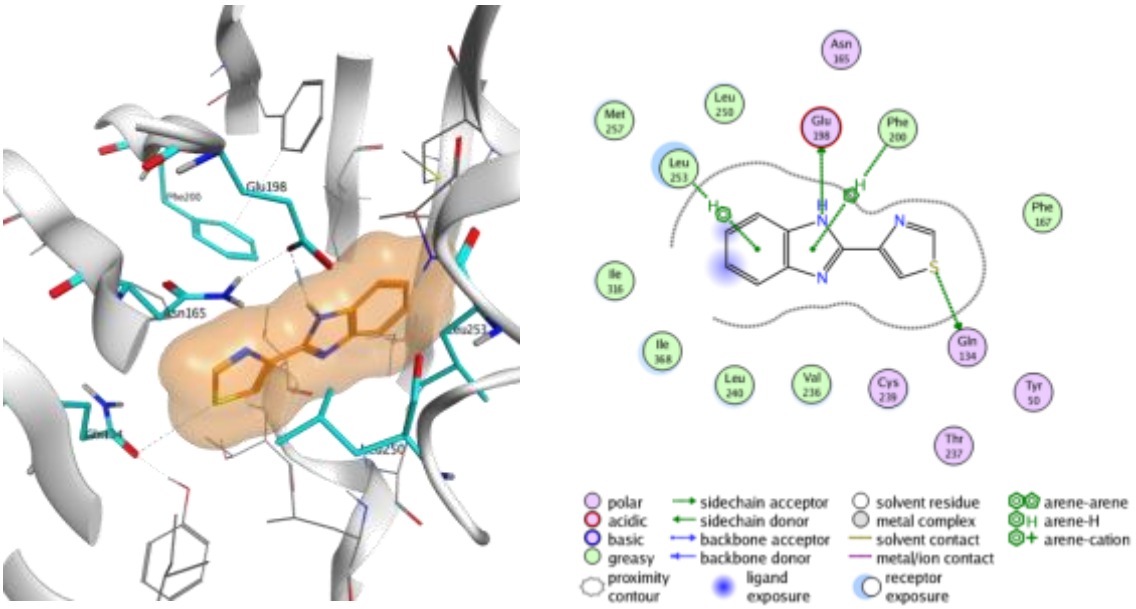



**Figure 2-27.** *Left*, human Tubb8 bound to TBZ. *Right*, 2-D contact map. Polar contacts contribute -8.3 kcal/mol to TBZ binding.

# Chapter 3: Resurfacing β-Glucuronidase Leads to Thermoswitchable Phenotypes in Bulk and Single-Molecule Experiments


## ABSTRACT

In this chapter, we sought to employ the Rosetta suite of protein design tools to redesign the surface of β-glucuornidase (GUS), a model enzyme, to include only one solvent exposed cysteine per active monomer in order to site-specifically label the enzyme or attach it to surfaces using thiol chemistry. While comparing the activity and stability of resurfaced designs to the wild-type enzyme and serine controls in bulk assays at room temperature and 37˚C, we found a pronounced thermoswitchable phenotype in the resurfaced variants. We also show that the activity of the redesigned variants could be modulated with temperature in single molecule assays using a novel platform (SiMoA) developed at Tufts University. One rationally designed serine variant (C262) was shown to be completely inactive at room temperature but had near wildtype levels of activity at 37˚C in both SiMoA and bulk reaction conditions. We show that Rosetta redesign variants are more prone to accessing inactive conformations after thermal pulsing, while serine containing variants retain elevated activities after a short temperature pulse. Taken together, we propose that the serine controls access the active conformation in an 'irreversible' manner, while the energy barrier between conformations in the Rosetta redesigns seems to favor the inactive state. This study highlights the significance of surface mutations on conformational dynamics and the activity of enzymes


## INTRODUCTION

It is difficult to overstate the significance of enzymology in the last half century. The study of nature's biological catalysts has contributed profoundly to our understanding of how life works at the molecular level and given us tools to probe deeper than early biochemists and biologists

could have possibly imagined (Zalatan and Herschlag, 2009). A profoundly important enzyme in this regard has been β-D-glucuronidase (EC 3.2.1.31). Though it catalyzes a seemingly simple reaction, the hydrolysis of a glycosidic bond, it has been used as a work horse in: plant molecular biology, microbiology, diagnostics, therapeutic design, and protein engineering for use in biotechnology. Perhaps now overshadowed by DNA/RNA polymerases(Broll, 2010), CRISPR/cas systems (Barrangou et al., 2007; Brouns et al., 2008), restriction enzymes, and green fluorescent protein (GFP), still β-glucuronidase (GUS) has been cited in over 10,000 PubMed articles. The impact this enzyme has had on molecular biology and biochemistry has been profound.

GUS is a member of family 2 in the glycoside hydrolase-A clan and is found over an enormous range of taxa (Jain et al., 1996). In humans, β-glucuronidase (hGUS) is localized in the lysosome and is produced from the gusB gene. Its primary purpose is to degrade sulfated glycosaminoglycans, namely chondroitin sulfate and hyaluronic acid. Loss of function or significantly reduced production of the enzyme leads to Sly syndrome, a buildup of mucopolysaccharides throughout the body (Sly et al., 1973). In some disease states, it can be found in the cytosol. *Escherichia coli* also possess a GUS form, encoded by gusA. Although the purpose of the enzyme differs drastically, it shares 50% sequence homology with hGUS. The bacterial form is used to breakdown glucronidated carbohydrates, which are then shuttled into central metabolism for carbon source utilization (Arul et al., 2008). It is the bacterial form of the enzyme that has been widely used in molecular biology and biotechnology.

**The Chemistry**

GUS catalyzes the generic reaction shown in **Figure 3-1**. Upon binding of a β-D-glucuronoside and water to the active site, the glucuronic acid moiety is cleaved, and an alcohol is produced. Products of the reaction retain their anomeric configuration (Wong et al., 1998). The two highly conserved glutamates participate in the reaction by a general acid/nucleophile mechanism. A

conserved tyrosine is suggested to recognize the carboxylic acid of the glucuronic acid through a hydrogen bonding network (Wallace et al., 2010). Early efforts of enzymologists to elucidate the reaction mechanism were thwarted by the lack of modern recombinant technologies and protein purification methods. However, these studies were not all in vein.

Wang and Touster, as early as 1971, were able to show that a general acid-nucleophile mechanism was in operation by studying electronic effects of substituted substrates using Hammett plots and comparing reaction rates in deuterated water (Wang and Touster, 1972a). In a follow-up study in the same year, the authors studied the pH dependence of the reaction. They were able to accurately determine pH 5.1 as the optimum of the liver lysosomal isoform. Wang and Touster also correctly identified carboxylate groups as the functional groups partly responsible for catalysis. However, their model was not entirely accurate. They suggested that an histidine was responsible for the nucleophilic attack of the substrate (Wang and Touster, 1972b).

In the modern era, sequence alignments of known hydrolases pointed to the conserved core as the catalytic residues (Henrissat and Bairoch, 1996). X-ray crystal structures(Jain et al., 1996; Wallace et al., 2010) and inhibition studies (Gebler et al., 1992) have also aided in our understanding of the two-step reaction mechanism (**Figure 3-2**). It has also been shown that an intact tetramer must be present in order for the enzyme to be active (Jain et al., 1996).

**Structural Characterization**

The first GUS structure solved was the human variant of the enzyme to 2.6 Å resolution (PDB 3LPF), using the multiple isomorphous replacement method (Jain et al., 1996). The structure was recently refined by Anumula to 1.7 Å (**Figure 3-3**). The initial structure revealed that human β-D-glucuronidase (hGUS) is a homotetramer, comprised of 4 identical 651 amino acid monomers, and exhibits ellipsoidal or asymmetric symmetry. The monomers of the enzyme are arranged in three recognizable domains: a barrel like structure with a jelly roll motif (22-223), an immunoglobulin-like domain (224-342), and the TIM barrel domain or $(\alpha/\beta)_8$.

The TIM barrel contains the active site of the enzyme, which is housed at the interface of opposing monomers. Each monomer of the enzyme has the catalytic core, meaning there are 4 active sites in each functional tetramer. hGUS also has a lysosomal targeting domain (179-199), which is thought to help localize the enzyme in the lysosome. Another feature of the structure, supporting lysosomal targeting, is the β-hairpin found between residues 184-204 (Jain et al., 1996).

In a study designed to probe the effects of glucuronidase inhibitors to alleviate toxicity from chemotherapeutic agents, Redinbo and coworkers solved the crystal structures of native and inhibitor-bound enzyme to 2.39 Å resolution (**Figure 3-4**) (Wallace et al., 2010).

Like hGUS, the *E. coli's* version has the three recognizable domains (TIM barrel, immunoglobulin-like domain, and jelly roll motif). Workers used molecular replacement,

based on hGUS to solve the structure. The bacterial and human forms of the enzyme share an astonishing 50% identity, including the catalytic core (2 glutamates and a tyrosine). A minor difference between the two forms is that each monomer of bacterial GUS is 603 amino acids in length, rather than 651. Another important distinction is that bacterial GUS lacks the lysosomal targeting domain.

Structural studies can aid in the understanding of underlying reaction methods. Careful examination of bound and unbound structures can give mechanistic insight in cases where mutagenic studies fail or rate constants are too fast or elusive to measure by direct methods. In bacterial GUS, we see a very small change in the structure of the enzyme upon substrate binding (**Figure 3-4**). Only careful analysis of the structure reveals any fluidity or change in the enzyme's conformation. The most drastic effects can be seen with the tyrosine residues in the active site. Beyond the ~14 Å shift of these residues, a mild shift (~7 Å) can be seen in a β-sheet distal from the active site. These small structural changes upon substrate binding may dampen signals generated in modern enzymological techniques.

## Molecular Biology & Biotechnology

### *Gus Reporter System*

A landmark in the field of plant molecular biology was the use of GUS as a reporter gene in higher plants. The GUS-fusion system, developed by Jefferson during his tenure as a graduate student (Jefferson et al., 1987), has been the most utilized tool in transgenic plant research to date (Miki and McHugh, 2004). Due to GUS's stability and relatively low toxicity levels, it has been a tried and true tool for plant molecular biologists since its inception. While other reporters have been developed, GUS-fusions remain in use today. A major advantage to the GUS system is that it does not require any specialized equipment. Instead, once GUS is co-expressed with its target promoter, a colorimetric substrate (5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid) can be seen with the naked eye (Xiong et al., 2012).

A major disadvantage to the GUS-fusion system is that, unlike previously thought, not all higher-ordered plants lack an active glucuronidase. This challenge has partly been overcome by engineering a GUS variant to withstand temperatures outside of the host's enzyme thermal tolerance (Xiong et al., 2011a). Unfortunately, this system cannot be used in most vertebrates or molluscs because these organisms have fully functional GUS forms (Jefferson, 1993).

### *Therapeutics*

hGUS has been involved in many studies involving therapeutic strategies for several diseases. Antibody dependent enzyme pro-drug therapy (ADEPT) is one example of an avenue that has been explored with the enzyme (Carter, 2001). This method involves the fusion of a monoclonal antibody to a recombinant form of GUS. Once the antibody has found its target, the attached enzyme is able to cleave the glucuronidated pro-drug into an active form (Haisma et al., 1998). In this approaches, the drug is inactive until the glucuronic acid is cleaved.

Another therapeutic strategy involving a similar mechanism is to use localized host hGUS at the site of a tumor. Due to hGUS's overproduction in some forms of cancers, which leads to it leaking out of the lysosome, again glucuronidated pro-drugs have been employed. This novel

method takes advantage of the overproduction because healthy cells do not generally have hGUS in the cytosol. An apparent advantage is that toxicity effects, generally seen in the parent drug, are largely nullified by the near target-specific manner in which the drug is delivered. Epirubicin-glucuronide was > 10,000 times less toxicity by glucuronidating a chemotherapeutic agent. However, there are limitations to this strategy. With some pro-drugs, including 9-amino-camptothecin, the pro-drug is cleaved by *E. coli* in the gut and toxicity levels increase to parent drug levels (Graaf et al., 2002).

***Diagnostics***

Bacterial GUS has been fused to monoclonal antibodies for use in enzyme-linked immuno sorbent assays (ELISA). In an ELISA colorimetric or fluorigenic substrates are used to determine whether or not the antibody is attached to its antigen. In several commercially available kits, GUS is used as the enzyme for the reaction. In others, GUS is the target, due to the increased GUS activity in several disease states. Falk and coworkers developed a GUS-based ELISA for the detection of 2-hydroxyesterone/16α-hydroxyesterone (Falk et al., 2000). Mackenzie also used GUS to develop an urinary immunoassay to detect a metabolite of chlorpyrifos (MacKenzie et al., 2000).

***Protein Engineering***

Directed evolution has proven to a useful tool for an array of biotechnological applications (Dalby, 2003; Tobin et al., 2000; Xiong et al., 2012). Once again, GUS has been utilized as a model enzyme in this regard. A significant portion of solved structures and known sequences of enzymes are built on the TIM barrel motif. Enzyme engineers have used GUS as a starting point to overcome various user-defined challenges by incrementally selecting variants from a library of mutants (typically $\sim 10^8$) in as little as weeks (Rowe et al., 2003). Protein engineers have evolved GUS to: be thermostable (Flores and Ellington, 2002; Xiong et al., 2011b); change its surface chemistry to resist glutaraldehyde and formaldehyde (Matsumura et al., 1999); and shift its pH profile and

change its substrate specificity (Geddie and Matsumura, 2004; Matsumura and Ellington, 2001). In the current work, we resurfaced GUS to allow for single-site conjugation using thiol chemistry.

## RESURFACING GUS FOR CONJUGATION TO SURFACES

In order to direct the orientation of bacterial GUS relative to surfaces, we sought to come up with a general method for identifying a single conjugation site in the enzyme. Thiol chemistry is well-known for site-specific labeling of proteins for use in a variety of downstream applications (1–3). The introduction of single cysteines into the surfaces of proteins has previously been largely empirical. To develop a more generalized method for generating single-cysteine variants, we sought to use the Rosetta suite of protein design tools (4) to remove all but one cysteine from the surface of a commonly used reporter protein, beta-glucuronidase (5). This was an especially interesting problem as there are six cysteines per monomer within the GUS tetramer. Additionally, serine only controls were constructed, expressed, and assayed for activity. When the variants were tested for activity at RT and 37°C, several variants demonstrated a significant increase in activity at 37°C in bulk and at the single molecule level. Single molecule experiments revealed reversible or irreversible thermal switching properties of a fraction of the population for these variants.

## RESULTS & DISCUSSION

### Rosetta computational resurfacing of β-glucuronidase

All six of the surface-exposed cysteines were identified by visual inspection of apo- and inhibitor-bound structures (PDBs: 3K4D & 3K46). Thus, in order to limit surface reactivity, five cysteines had to be removed. The systematic removal of the surface-exposed residues was accomplished by employing a fixed backbone design methodology using the Rosetta 3.4 protein modeling and structure prediction software package. The program was used to assess the impact of altering these residues on the overall stability of the enzyme. First, the crystal structure was relaxed, and sequence optimization simulations were performed on five of the six solvent-exposed cysteine residues at once to determine the most favorable residues to mutate. All combinations of

cysteine replacement were evaluated 100 times. We identified mutations at positions C133 and C262 to be the most destabilizing. Thus, designs were optimized using these two positions as single attachment points with all other residues allowed to adopt alternative identities and rotamer conformations during energy minimization. Designs were scored for their overall structural stability using the Rosetta all atom score function. Four of the designs exhibiting the most stable properties were chosen for expression and analysis (D1-D4).

In three of the four Rosetta designed GUS variants (Designs 1-3), Cys28, Cys133, Cys197, Cys253, and Cys527 were mutated while Cys262 was conserved in order to enable chemical conjugation (**Figures 3-5 and 3-6**). In the fourth re-designed GUS variant (Design 4), Cys28, Cys197, Cys253, Cys262, and Cys527 were mutated while Cys133 was maintained (Fig. 1C). Residues Cys443, Cys449, and Cys464 were deemed to be buried or mostly buried within GUS, and not considered to be able to undergo cysteine conjugation chemistry, so were therefore not candidates for mutation (**Table 3-1**).

Mutations in all four Rosetta re-designs were chosen to optimize overall GUS protein stability, as measured by the Rosetta energy score. All re-designs were predicted to be more stable than the wildtype GUS enzyme, and furthermore mutations in the designs were compatible with the overall surface properties of the GUS enzyme. In Designs 1 and 4, C28 was mutated to Lys, while in Designs 2 and 3 the residue was mutated to Leu. Both of these mutations extend existing patches of either positive charges or hydrophobic atoms that are adjacent to residue 28 on the GUS surface. In all designs where C133 and C197 were allowed to mutate, both residues were mutated to Asp, connecting to a slightly negative patch adjacent to each residue. (Fig 1E-F). Lastly, residues C252 and C527 were mutated to match a small, slightly hydrophobic surface patch in each Rosetta re-design.

**Bulk characterization of GUS variants reveals temperature switching activity**

We tested these designs for bulk activity using four commercially available substrates in endpoint assays (**Figure 3-7**). Activity was observed using each of the substrates with 4-

methylumbelliferylglucuronide (4-MUG) giving the most robust results. In each case, D4 was found to be the most active of the variants. Using 4-MUG, D4 exhibited nearly 60% of activity when compared to the wildtype enzyme, and we therefore decided to use 4-MUG in subsequent bulk characterization of the enzyme variants.

We next examined timecourse kinetics and observed a lag in activity of the Rosetta-designed variants at 22˚C (RT) (**Figure 3-8**). We hypothesized that this lag may be overcome by thermally perturbing the enzyme as it was unlikely that the resurfaced enzyme would now adopt a cooperative mechanism. Our rationale was that the designed enzymes may need kinetic energy to surmount an energetically unfavorable or slow conformation caused by the resurfacing mutations. Consistent with this prediction, we found that increasing the temperature abolished the lag in temperatures above 30˚C (**Figure 3-9**). We found that the optimal temperature for each of the designed variants matched the wildtype thermal preference (between 35-40˚C) (**Figure 3-10**).

We then compared these properties to a "dumb" all-Serine replacement re-design (S1-GUS) and two single cysteine variants, C133 and C262. S1-GUS is a rational re-design of the GUS enzyme surface, wherein all surface cysteine residues were mutated to serine, a rational re-design choice for cysteine replacement on the surface of a protein. C133 and C262 are also serine variants where a single cysteine is incorporated at the given position. Thus, we were able to compare GUS enzyme thermostability and activity (**Figure 3-11**) between computational re-designs and "rational" re-designs. All resurfaced GUS variants showed a decrease in thermostability of ~8˚C (**Table 3-2**). Looking at initial rates, we found that the wildtype exhibited a normal rate vs substrate curve that could be fit the Michaelis-Menten equation at 22˚C or 37˚C. Rosetta designs displayed a lag at 22˚C, but could be fit to the Hill equation. Suprisingly, C133 and C262 were both almost completely inactive at 22˚C but regained near wildtype activity at 37˚C (**Figure 3-12**). This suggests that the serine controls are more destabilized at room temperature, but are capable of surpassing an energy barrier to productive catalysis at the elevated temperature.

**Single-molecule experiments confirm temperature switching activity**

The Walt laboratory at Tufts University has pioneered the use of single-molecule arrays (SiMoA) in femtoliter reaction chambers (**Figure 3-13**). In these assays, low concentrations of enzyme and substrate are sealed and monitored with a detector through a fiber optic bundle. Product is incapable of diffusing through the sealant, and thus, through use of a Poisson distribution of enzyme, single enzyme activities can be measured. This technology has been applied to a number of single-molecule enzyme studies (Gorris and Walt, 2009; Rissin and Walt, 2006; Rissin et al., 2008).

The Walt laboratory then performed single molecule experiments for all of the GUS variants at 22˚C and 37˚C. Each enzyme was diluted to 1 pM and sealed within the optical fiber wells. Resorufin-glucuronide was used in these studies because the optical properties of SiMoA do no accommodate the use of 4-MUG, the substrate used in bulk experiments in the previous section. In each of these single-molecule experiments, 250 μM substrate was used in order to reduce background fluorescence of this molecule and determine reaction rates. Histograms of typical single molecule velocities at room temperature and 37˚C are shown in **Figure 3-14**. The single molecule velocities were largely in agreement with observations from bulk experiments. For each variant, a velocity increase of ~3-6 fold was observed at 37˚C with the exception of C262. For this variant, we observed roughly a 206-fold increase in the rate of the reaction **(Table 3-3)**. This result was consistent with our bulk measurements using 4-MUG as a substrate in our initial rate experiments where we did not observe any activity over background with this variant.

We next sought to compare activities of individual enzymes at room temperature and 37˚C in the context of shifting the temperature or briefly pulsing the enzymes with an elevated temperature (**Figure 3-15**). In the temperature shift experiments, activities of variants could be measured before and again after the target temperature of 37˚C was reached. Activity was measured

for 5 minutes at room temperature before ramping. We observed that the rates of individual enzymes increases for single molecules of the wildtype enzyme, although a significant correlation could not be determined (Spearman coefficient = 0.3, $p<0.005$) (**Figure 3-16**). We interpret these data to suggest that conformational changes were not needed to access a productive catalytic state as the number of active enzymes did not change after shifting the temperature.

In contrast, Rosetta designed variants showed a significant increase in activity after the shit as well as an increase in number of active enzymes. In particular, D2 and D4 increase single-molecule activities by 13.2 and 10.9-fold, respectively (**Figure 3-17**). For these variants, several inactive enzymes at room temperature also gained activity after the shift (plotted as <0 on the x-axis). Nearly 20% of D2 and D4 enzymes gain activity at the elevated temperature. This effect was more pronounced in the C133 and C262 controls where over 90% of enzymes gained activity after the temperature shift. We interpret these results in a framework that supports our conformational shift hypothesis from the lab observed in our single trace bulk experiments.

In order to determine if the temperature switchable phenotypes observed in our studies were reversible, we conducted thermal pulse experiments. Here, the increase to 37˚C is transient, lasting only 2 seconds. After the pulse, the temperature is cooled and held constant at room temperature. Measurements are made before and after the pulse. Surprisingly, many of the Rosetta enzymes lost activity after the heat pulse, which resulted in a reduced average activity. On the other hand, C133 and C262 gained and sustained activity after the pulse (**Figure 3-18**). These data suggest that the thermal switch behavior may be irreversible, or functionally so, in the timescales tested here. It could also be that once the active conformation of the enzyme is reached, it is energetically unfavorable to traverse back down the less active conformation.

**Transition state modeling of GUS variants reveals differences in complex stability**

In an effort to understand the rational and Rosetta designed mutations, we attempted to model the transition state with MOE using PDB: 3K4D, the wild-type GUS:inhibitor bound structure, as a starting point. We simulated the effects of mutations on the tetramer. In total 44 structural issues were identified and corrected in the initial structure file from the Protein Data Bank. These included: 2 alternate occupancy issues, 8 termini breaks or gaps, 24 atom clashes and 4 cis-amide issues. The break between residues 362 and 370 was built by homology modeling. The final prepared structure was protonated at pH 7.4 and 310K. A total of 1304 ionizable groups were protonated for optimal hydrogen bonding in the tetrameric structure. Inspection of the Phi-Psi plot indicated proper geometry without any outliers present, and all atom clashing was relieved upon minimization with the Amber10:EHT forcefield to an RMS gradient of 0.001 kcal/mol/Å$^2$. Mild perturbation from the uncorrected structure could be seen in the RMSD between structures with surface mutations present in the Rosetta and rational designs.

To model the transition state, we docked resorufin glucuronide into the active site of each monomer and formed a covalent adduct with the catalytic glutamate. A total of 100 poses were generated and evaluated with the binding of the inhibitor present in the wild-type structure. Final geometries were calculated using an all-atom QM minimization with Guassian09. The structure was then subjected to 10 ns of molecular dynamics using NAMD 2.9. The lowest energy structure was then used as a basis for *in silica* mutational analysis (Figure 3-19). *In silica* measurements of stability were then made by saturating each surface-exposed cysteine with all 20 canonical amino acids (**Table 3-4**). We summed each stability score for each of the variants. Interestingly, the all-serine mutant was shown to be the least stable (-46.1 kcal/mol), consistent with our thermal melt measurements in bulk. The wildtype enzyme was shown to have a sum of -61.7 kcal/mol of stability. Rosetta designed mutants were less stable (-52.6 to -53.4 kcal/mol) than the wildtype, but more stable than the all serine control. C133 was more stable than all of the Rosetta designs and

serine controls but less stable than the wildtype (**Table 3-5**). It may prove useful to build variants predicted by MOE to be the most stabilizing at each of the positions using the transition state model presented here.

**Conclusions**

Here we observed the effects of temperature on activity of resurfaced β-glucuronidase variants in bulk reactions as well as at the single molecule level using SiMoA, a novel platform developed by the Walt laboratory. Using two methodologies with two different fluorescent substrates, we were able to observe thermally switchable phenotypes in Rosetta and serine resurfaced mutants. The most pronounced changes in activity were observed with C262, the variant with a single cysteine on the surface at position 262 whereby all other cysteines had been mutated to serine. For the two serine variants, C133 and C262, the temperature switching in activity appeared to be irreversible on the timescales tested here. That is, the activity was sustained after pulsing. The Rosetta variants did not display this behavior. In fact, many of the Rosetta enzymes lost activity after the pulse.

These results suggest that the enzymes tested here occupy at least two distinct, thermally accessible states, active and inactive conformations—though a continuum of these extremes is likely. More detailed kinetic studies could potentially elucidate which of the steps along the reaction coordinate are affected by the surface mutations. This study also highlights the fact that amino acids at the surface of proteins contribute more than just a hydrophilic shell with which to interact with an aqueous environment. While substrate channeling on surfaces has been shown, we are unaware of any studies that have shown resurfacing to drastically change bulk and single-molecule activities of enzymes at moderate temperature changes. Single cysteine replacements may point to key positions that lead to the thermoswitchable phenotype. Further, it will be interesting to test if this is a general phenomenon within structurally-related enzymes. For example, β-

galactosidase (lacZ) is a closely related enzyme that has a comparable number of cysteines on the surface.

## MATERIALS AND METHODS

### Cloning and purification of GUS variants

*Escherichia coli* TOP10 and BL21 (DE3) (Invitrogen) strains were used for cloning and expression, respectively. Strains were maintained on either LB or 2XYT growth media. WtGUS, Rosetta-designed variants (D1-D4), an all-serine variant (S1), C133, and C262 were cloned into the pET28+ vector using the Gibson Assembly method. An N-terminal $His_6$-tag was fused to all enzymes for affinity purification. Overnight cultures of BL21 (DE3) harboring each of the GUS variants were grown in LB at 37°C. Cells were then diluted into Terrific broth. Protein production was induced with 1 mM IPTG at 18°C for 20 hrs. Cells were lysed in 50 mM phosphate, 20 mM imidazole, 500 mM NaCl, pH 7.4 and cleared cell lysate were purified over NiNTA resin (GE Healthcare). Proteins were eluted with 50 mM phosphate, 500 mM imidazole, 500 mM NaCl, pH 7.4 and further purified over a Sephadex 16/60 size exclusion column (GE Healthcare), concentrated, and dialyzed against 50 mM phosphate, 1 mM DTT, pH 7.4. Aliquots of each enzyme were flash-frozen in liquid nitrogen for downstream testing. Enzyme concentration was determined by the BCA method (Pierce).

### Rosetta protein redesign

The structure of the *E. coli* β-Glucoronidase protein (PDB code 3K4D) was analyzed for number and placement of surface cysteine residues, both visually and computationally using a nearest-neighbor count of the number of atoms surrounding each cysteine residue sidechain atom (reference supercharge paper). Six solvent- exposed, potentially reactive, surface cysteine residues were identified. Mutations to surface cysteine residues on GUS were probed using the Rosetta 3.4 protein modeling and structure prediction software package. The program was used to assess the

impact of altering these residues on the overall stability of the enzyme. The crystal structure was relaxed, and sequence optimization simulations were performed on five of the six solvent-exposed cysteine residues at once, with amino acid sidechains allowed to adopt alternative identities and neighboring side chains free to adopt new sidechain conformations. All sidechains were energy minimized. All combinations of five groups of cysteines redesigned at a time with one cysteine remaining unchanged were attempted 100 times. Designs were scored for their overall structural stability using the Rosetta all atom score function. Four of the designs exhibiting the most stable properties were chosen for expression and analysis. Rosetta flags: Rosetta relax flags: ./relax.[version][release] -database <path>/rosetta_database -in:file:s 3K4D.pdb -in:file:fullatom -relax:fast -nstruct 100 -ex1 -ex2 Rosetta fixed backbone design flags: ./fixbb.[version][release] -database <path>/rosetta_database -s <3K4D.pdb> -resfile <resfile> -ex1 -ex2 -nstruct 100 -minimize_sidechains

**Enzyme activity measurements**

Enzyme activities were measured in 96-well plate format in GUS buffer at 37°C in a Tecan M200 Pro plate reader. Initial rates of enzyme with varying concentration of 4-methylumbelliferyl glucuronide) were plotted using SigmaPlot to determine $k_{cat}$, $K_m$ and $k_{cat}/K_m$. Product formation was monitored by fluorescence (ex. 365/em. 455) and compared to a 4-MUG standard curve. Activity was also measured on a temperature controlled fluorimeter in 1 mL volumes from 15-65°C in 5°C increments and in pH increments of one unit (5 nM enzyme, 400 µM 4-MUG) to determine thermal and pH optima. To assess potential differences in glucuronide substrates, activities were also measured using pNP-glucuronide and resorufin-glucuronide (Sigma).

**Thermal melt measurements**

Thermal stability of each enzyme was assessed by incubating 2 $\mu$M enzyme in GUS buffer with TexasRed dye (Invitrogen). Enzyme/dye mixtures were heated rate of 1°C/s to 95°C using a LightCycler 96 thermocycler, while fluorescence was monitored. Data were analysed using Roche thermocycler software.

**Bulk experiments using resorufin-glucuronide**

Bulk experiments were performed on a microtiter plate reader to test for enzymatic activity. 100 µM of substrate solution was prepared in the reaction buffer (50 mM $Na_2PO_4$, 0.1% Triton X-100, 10 mM EDTA), and then 10 µL of 2 nM enzyme was mixed with 90 µL of 100 µM substrate solution. The fluorescence intensity ($\lambda_{ex}$=558 nm, $\lambda_{em}$=590 nm) was measured in the plate reader at 30 second intervals for 0.5 hours. All of the experiments were performed both at room temperature and at 37°C.

**SiMoA Assays**

Time lapse imaging of the fluorescent product generated as a result of enzyme catalysis, was performed with an upright Olympus BX61 microscope equipped with a short arc mercury lamp (Ushio, Tokyo, Japan) and a CMOS camera (ORCA-Flash 4.0, Hamamatsu, Japan), and a custom built automated sealing stage. A filter set with $\lambda_{ex}$ = 558 nm and $\lambda_{em}$ = 577 nm (Semrock, Rochester, NY) was used for imaging the resorufin product. Images of the distal end of the fiber were captured every 30 seconds for a period of 30 minutes or less. The images were analyzed using the MATLAB software Image Processing Toolbox.

Figure 3-1: β-glucuronidase reaction

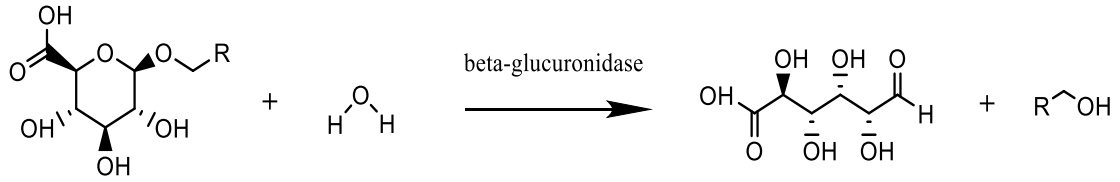



Figure 3-1. *β-glucuronidase reaction*. Depiction of the reaction catalyzed by β-glucuornidase. A β-D-glucuronoside is hydrolyzed by the enzyme to form a D-glucuronic acid and an alcohol. Figure created in *ChemDraw Pro*.

Figure 3-2: Two step reaction mechanism of GUS

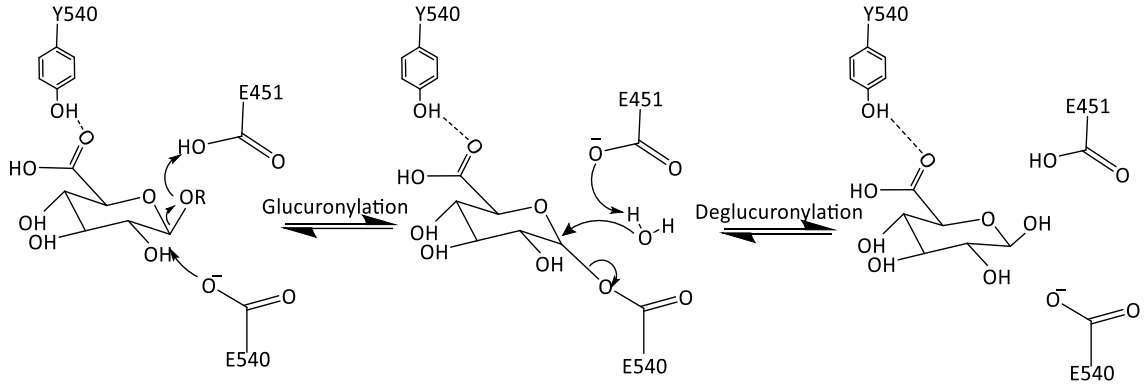



Figure 2: *Two Step Reaction Mechanism.* Proposed reaction mechanism for hGUS. Theoretical transition states have been omitted for clarity. Two conserved glutamate residues act as the nucleophile and electrophile, respectively. The conserved tyrosine is proposed to stabilize the carboxylic acid through a hydrogen bonding interaction. In the first step of the sequence, the OR group deprotonates E451, and E540 attacks the substrate. Next, E451 deprotonates a water molecule in the active site. The hydroxyl group attacks the anomeric carbon, releasing E540 from the substrate.

Figure 3-3: Human β-glucuronidase structure

Figure 3-3. Human β-glucuronidase structure. (PDB: 3HN3) with EVA bound ((2S,3R,4S,5R)-3,4,5-trihydroxy-6-oxopiperidine-2-carboxylic acid). Each monomer is depicted in a distinct color.

Figure 3-4: Overlay of Free and Inhibitor-Bound Bacterial GUS

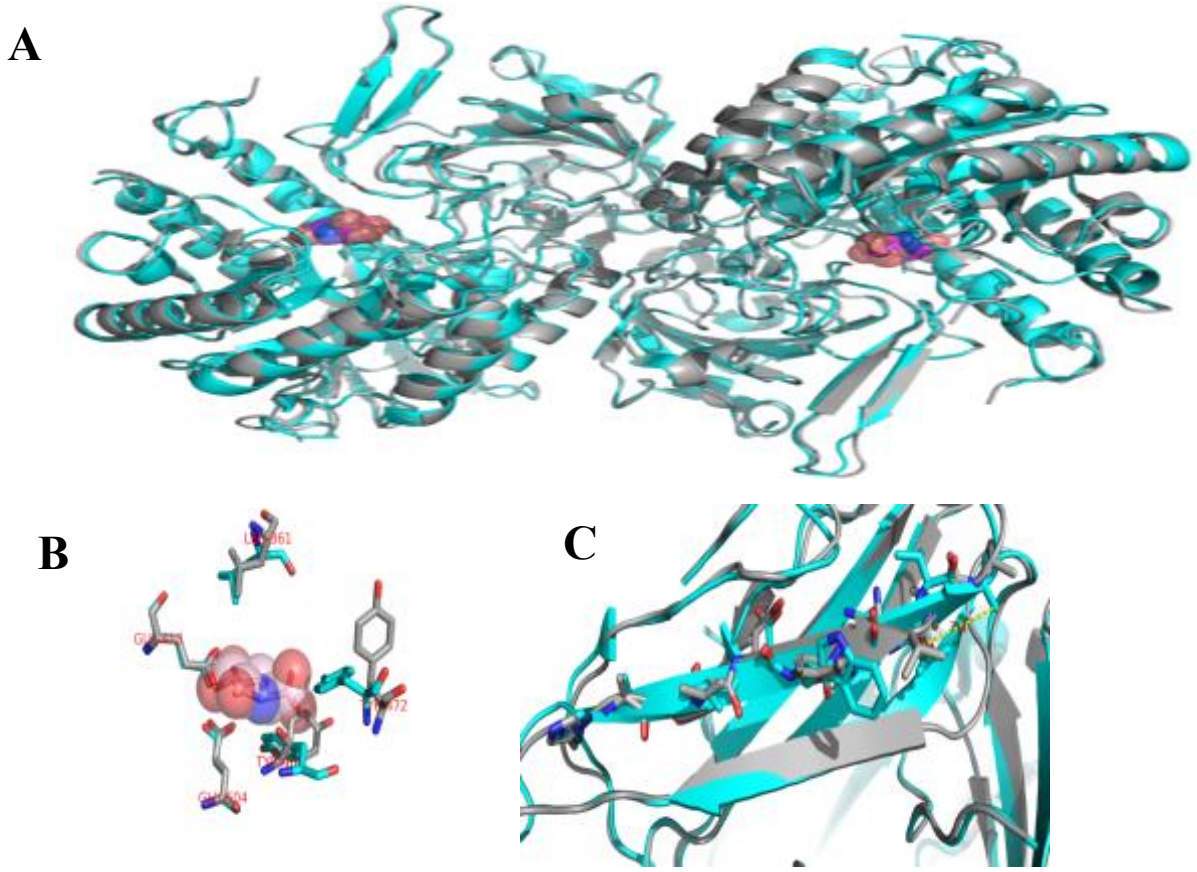



Figure 3-4. *Overlay of Free and Inhibitor-Bound Bacterial GUS.* A. Overlay. Native and inhibitor-bound structures with space-filling model of (2S,3R,4S,5R)-3,4,5-trihydroxy-6-oxopiperidine-2-carboxylic acid. The native structure is grey. B. Active Site Pertubation. Panel B shows one of the most obvious differences between the bound and unbound structures. Y472 moves drastically upon binding (~14 Å). C. Distal Pertubation. Depiction of a distal movement of a beta sheet. The measured shift in V206 is 7.4 Å. Figures generated in *Pymol* (PDBs 3K46 and 3K4D).

Figure 3-5: Bacterial GUS tetramer with position of cysteines highlighted

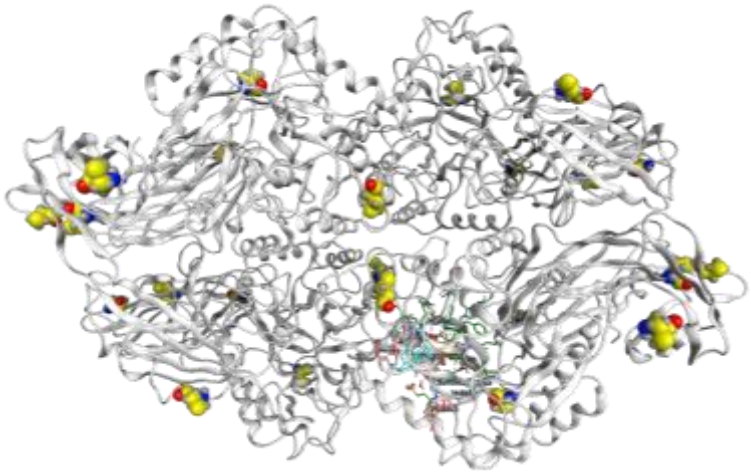


Figure 3-5: *Bacterial GUS tetramer with position of cysteines highlighted*. 3D representation of the GUS tetramer. Surface cysteines are highlighted in yellow. One active site is shown in the structure (blue sticks) for reference. None of the cysteines are near or involved in oligomerization, substrate binding, or catalysis.

Figure 3-6: Monomeric GUS with positions of cysteines highlighted

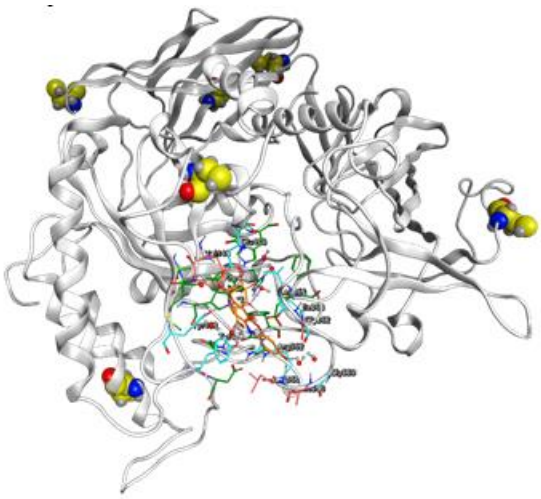


Figure 3-6: Monomeric GUS with positions of cysteines highlighted. Substrate is depicted in orange. Blue residues indicate substrate binding or catalytic residues.

Table 3-1: Rosetta and rational design mutations

| Variant | Mutations |
|---|---|
| D1 | 28K, 133D, 197D, 253V, 527M |
| D2 | 28L, 133D, 197D, 253V, 527M |
| D3 | 28L, 133D, 197D, 253I, 527M |
| D4 | 28K, 197D, 253I, 262Y, 527M |
| C133 | 28S, 197S, 253S, 262S, 527S |
| C262 | 28S, 133S, 197S, 253S, 527S |

Figure 3-7: Substrate optimization of GUS variants

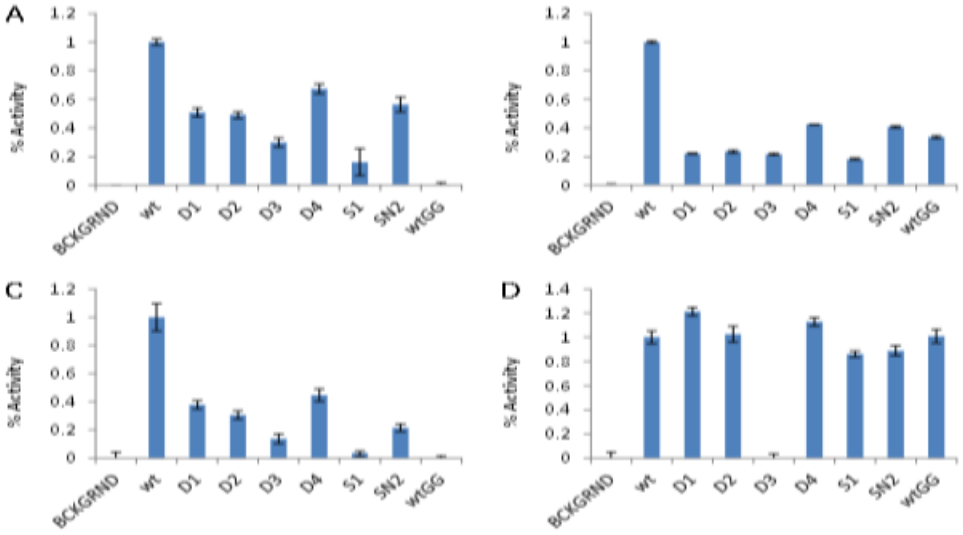



**Figure S2. Substrate Optimization.** Variants of GUS ([E] = 5 nM) were assayed with 400 µM substrate for 1 minute. The reactions were stopped with 2M $Na_2CO_3$ and read on a plate reader. Excitation, emission and absorbance wavelengths (nm) are shown in parenthesis. **A. 4-MUG** (365/455) **B. Resorufin** (550/610). **C. pNP** (405). **D. X-glucuronide** (650).

Figure 3-8: Timecourse kinetics of GUS variants

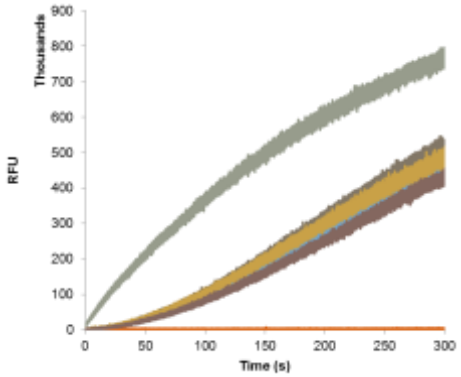



Figure 3-8: *Timecourse kinetics of GUS variants*. Variants were monitored for 300s. A lag in activity can be seen in each of the resurfaced variants. wtGUS (gray) displays normal activity, denoted by the hyperbolic rise in product.

Figure 3-9: Initial lag is overcome with temperature

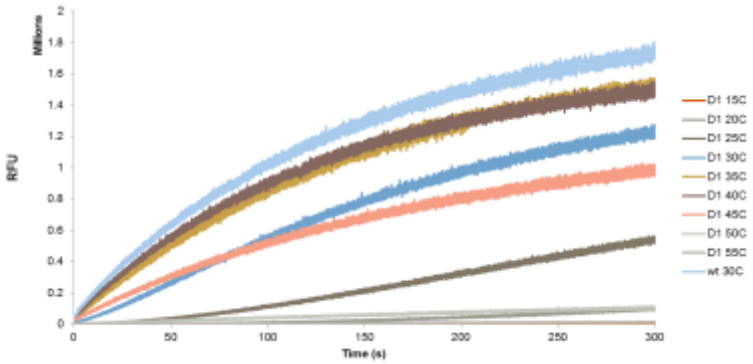



Figure 3-9: *Initial lag is overcome with temperature*. Representative traces of activity of resurfaced variants at various temperatures. The lag observed at room temperature is overcome at temperatures > 30˚C.

Figure 3-10: Thermal profile of D1

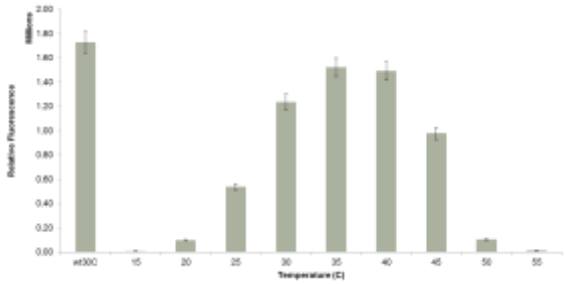


Figure 3-10: *Thermal profile of D1*. Representative plot to determine thermal optimum. Rosetta D1 is shown here. The optimal temperature for this and other variants was found to be between 35-40˚C, consistent with wtGUS.

Figure 3-11: Rosetta vs rational design endpoint assay

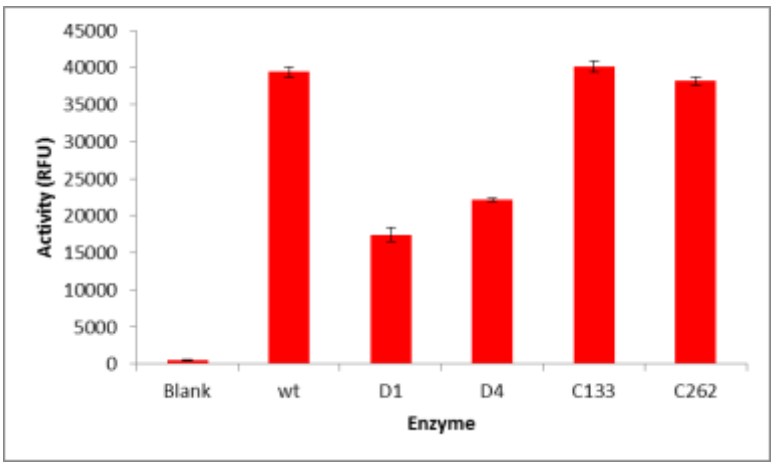



Figure 3-11: *Rosetta vs rational design endpoint assay.* Endpoint assay showing relative activities of the Rosetta and rational designs to wtGUS at 37˚C. Single cysteine controls display activity comparable to wtGUS.

Table 3-2: Thermal melt measurements of GUS variants

| Variant | wt | D1 | D4 | S1 | C133 | C262 |
|---|---|---|---|---|---|---|
| Tm | 68.1 | 60.7 | 60.8 | 55.4 | 62.4 | 61.5 |
| StdDev | 0.7 | 0.9 | 1.0 | 2.0 | 0.8 | 2.8 |

Figure 3-12: Initial rates of C133 and C262

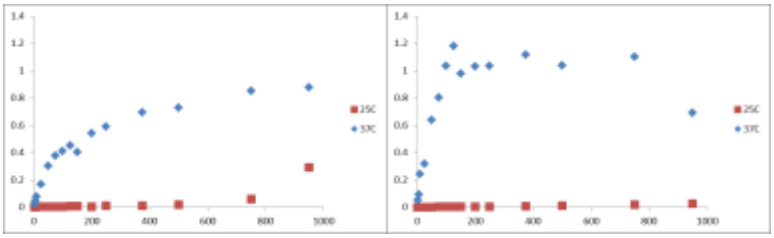


Figure 3-12: *Initial rates of C133 and C262*. Initial rates of C133 (left) and C262 (right) were compared at various substrate concentrations. Remarkably, both of the enzymes are essentially inactive at room temperature but regain robust activity at 37˚C.

Figure 3-13: SiMoA platform

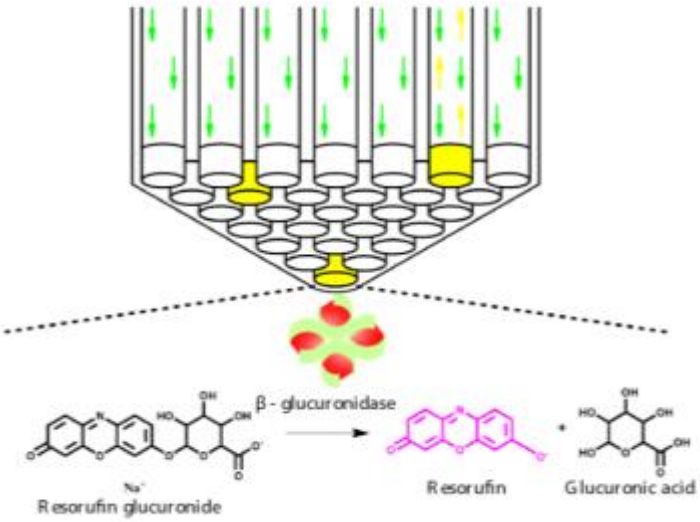



Figure 3-13: *SiMoA plaform*. 46 *f*L optical fiber microwells are generated at the distal end of the fiber bundle by etching the fibers for 115 seconds with 0.025 M HCl. The enzyme is diluted into a solution containing excess of the substrate (REG) such that the final concentration of the enzyme is 2 pM. This solution is immediately sealed into the array of 46fL wells. Using the Poisson distribution statistics $\boldsymbol{P_\mu(x) = e^{-\mu} X\mu^x/x!}$ where $\boldsymbol{P_\mu(x)}$ gives the probability of finding $x$ molecules per well, μ being the mean number of molecules per well, it can be determined that for a concentration of 2 pM, there is a very low probability of finding more than one enzyme per well. Hence every well contains either one or zero enzyme molecules. The turnover of REG to resorufin is used to detect the presence of the active species, as shown in the figure.

Figure 3-14: Histograms of single molecule activities at room temperature and 37˚C

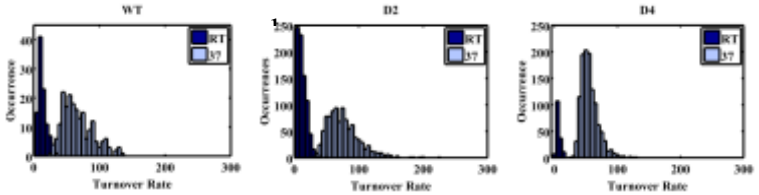



Figure 3-14: *Histograms of single molecule activities at room temperature and 37˚C* for WT, D2 and D4 (a, b and c respectively).

Table 3-3: Single molecule rates of GUS variants

| | RT | 37°C | |
|---|---|---|---|
| Enzyme | Velocity $s^{-1}$ | Velocity $s^{-1}$ | Fold Increase |
| wt | 48.8 | 144.4 | 3.0 |
| D1 | 33.4 | 158.8 | 4.8 |
| D2 | 46.1 | 145 | 3.1 |
| D3 | 28.2 | 160.6 | 5.7 |
| D4 | 43.7 | 150.2 | 3.4 |
| 133 | 30.21 | 113.7 | 3.8 |
| 262 | 0 | 206.3 | >206 |

Figure 3-15: Temperature profiles in SiMoA assays

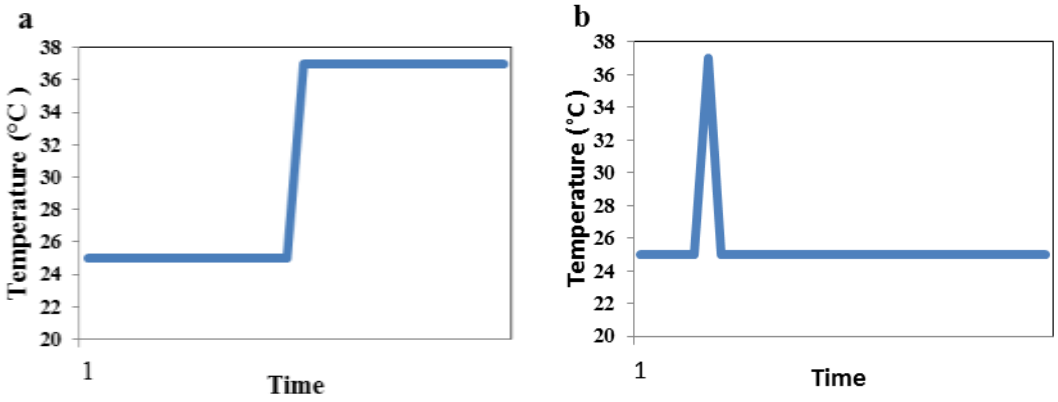



Figure 3-15: *Temperature profiles in SiMoA assays*. Temperature shift (a) and pulse (b) experiments.

Figure 3-16: wtGUS temperature shift experiment

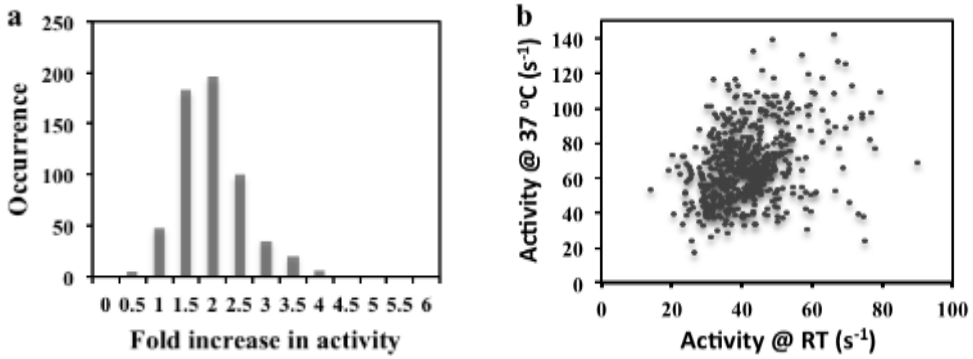



Figure 3-16: *wtGUS temperature shift experiment* (a) Fold increase in activity for single enzyme molecules (b) Scatter plot of the activities of the WT at RT and 37°C.

Figure 3-17: Temperature shift results of Rosetta resurfaced GUS variants

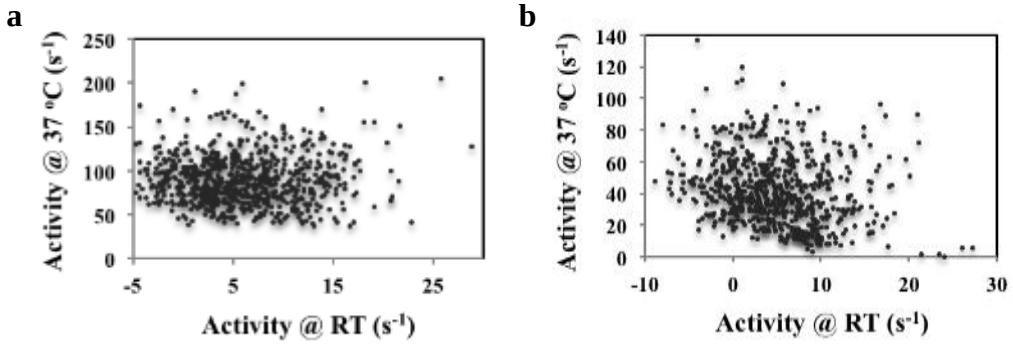



Figure 3-17: *Temperature shift results of Rosetta resurfaced GUS variants*. SiMoA assays were performed at RT and 37°C for D2 and D4. Several enzymes that were inactive at RT gain activity at 37°C.

Figure 3-18: Temperature pulse experiment of C133

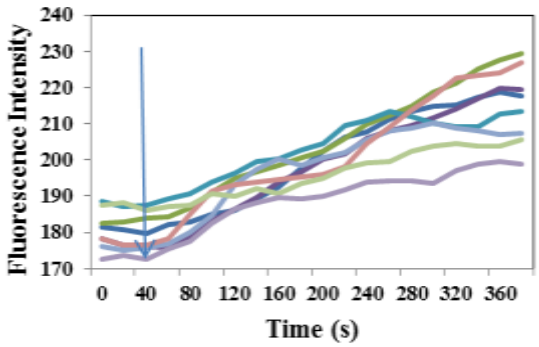



Figure 3-18: *Temperature pulse experiment of C133*. Enzymes were monitored at RT for 50 seconds and then pulsed to 37˚C for 2 seconds. Activities were measured for over 300 seconds following the pulse. Enzymes retained increased activity after the pulse.

Figure 3-19: Transition state model of GUS

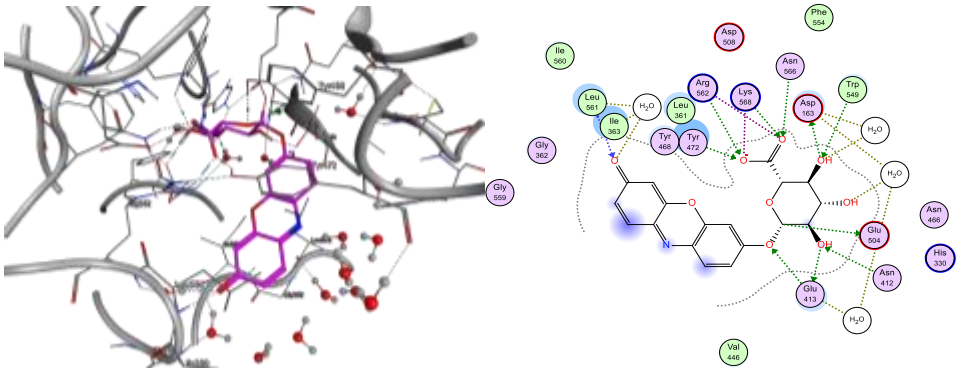



Figure 3-19: *Transition state model of GUS*. (Left) Using MOE and Guassian09, we made a transition state model in order to assess the effects of surface mutations on the tetrameric biomolecule. (Right) 2D contact map of resorufin glucuronide transition state model.

Table 3-4: *In silica* mutational analysis at surface-exposed cysteines

| C28 | | C133 | | C197 | | C253 | | C262 | | C443 | | C527 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mutation | Stability | Mutation | Stability | Mutation | Stability | Mutation | Stability | Mutation | Stability | Mutation | Stability | Mutation | Stability |
| 1:C28C | -8.1 | 1:C133I | -11.4 | 1:C197Y | -5.4 | 1:C253C | -10.6 | 1:C262C | -7.1 | 1:C443C | -10.9 | 1:C527C | -10.6 |
| 1:C28Y | -7.1 | 1:C133C | -11.1 | 1:C197S | -4.9 | 1:C253Y | -10.4 | 1:C262L | -6.7 | 1:C443I | -10.6 | 1:C527I | -9.7 |
| 1:C28P | -7.1 | 1:C133L | -10.4 | 1:C197N | -4.0 | 1:C253V | -10.1 | 1:C262Y | -6.3 | 1:C443M | -10.0 | 1:C527L | -9.2 |
| 1:C28R | -7.0 | 1:C133Y | -10.1 | 1:C197W | -3.8 | 1:C253I | -9.6 | 1:C262N | -6.1 | 1:C443V | -10.0 | 1:C527M | -9.1 |
| 1:C28V | -6.9 | 1:C133V | -10.0 | 1:C197L | -3.8 | 1:C253L | -9.4 | 1:C262V | -6.1 | 1:C443L | -9.7 | 1:C527Y | -8.9 |
| 1:C28L | -6.8 | 1:C133P | -9.6 | 1:C197V | -3.6 | 1:C253P | -9.1 | 1:C262S | -5.8 | 1:C443H | -9.3 | 1:C527V | -8.8 |
| 1:C28S | -6.5 | 1:C133M | -9.2 | 1:C197Q | -3.6 | 1:C253M | -9.1 | 1:C262E | -5.7 | 1:C443Y | -9.0 | 1:C527R | -8.5 |
| 1:C28N | -6.3 | 1:C133W | -9.1 | 1:C197R | -3.2 | 1:C253W | -8.8 | 1:C262R | -5.7 | 1:C443P | -9.0 | 1:C527N | -8.1 |
| 1:C28I | -5.8 | 1:C133N | -8.7 | 1:C197C | -3.2 | 1:C253F | -8.4 | 1:C262M | -5.4 | 1:C443R | -8.5 | 1:C527W | -8.0 |
| 1:C28A | -5.8 | 1:C133R | -8.7 | 1:C197K | -3.1 | 1:C253T | -8.1 | 1:C262D | -5.2 | 1:C443N | -8.1 | 1:C527P | -7.9 |
| 1:C28T | -5.5 | 1:C133F | -8.6 | 1:C197M | -2.8 | 1:C253R | -7.7 | 1:C262T | -5.2 | 1:C443T | -8.1 | 1:C527S | -7.4 |
| 1:C28M | -5.5 | 1:C133T | -8.1 | 1:C197D | -2.8 | 1:C253N | -7.7 | 1:C262I | -5.1 | 1:C443F | -7.9 | 1:C527T | -7.3 |
| 1:C28F | -5.5 | 1:C133H | -7.9 | 1:C197G | -2.7 | 1:C253A | -7.5 | 1:C262Q | -5.1 | 1:C443W | -7.9 | 1:C527A | -7.2 |
| 1:C28W | -5.4 | 1:C133A | -7.6 | 1:C197E | -2.6 | 1:C253S | -7.2 | 1:C262K | -4.9 | 1:C443K | -7.5 | 1:C527K | -7.0 |
| 1:C28K | -5.2 | 1:C133K | -7.6 | 1:C197P | -2.4 | 1:C253H | -7.2 | 1:C262W | -4.1 | 1:C443A | -7.3 | 1:C527F | -7.0 |
| 1:C28Q | -5.0 | 1:C133E | -7.6 | 1:C197I | -2.3 | 1:C253Q | -6.7 | 1:C262A | -4.1 | 1:C443S | -7.2 | 1:C527Q | -6.8 |
| 1:C28E | -4.9 | 1:C133S | -7.2 | 1:C197T | -2.2 | 1:C253E | -6.7 | 1:C262F | -3.9 | 1:C443E | -7.0 | 1:C527E | -6.7 |
| 1:C28D | -4.4 | 1:C133D | -7.0 | 1:C197A | -1.9 | 1:C253K | -6.7 | 1:C262P | -3.8 | 1:C443Q | -7.0 | 1:C527D | -6.5 |
| 1:C28H | -3.9 | 1:C133Q | -6.8 | 1:C197F | -1.4 | 1:C253D | -6.7 | 1:C262G | -3.5 | 1:C443D | -6.2 | 1:C527H | -5.3 |
| 1:C28G | -3.5 | 1:C133G | -4.6 | 1:C197H | 0.4 | 1:C253G | -4.4 | 1:C262H | -1.1 | 1:C443G | -3.9 | 1:C527G | -4.2 |

Table 3-5: Transition state stabilization scores

| Enzyme | Stability (kcal/mol) | ΔStability |
|---|---|---|
| Wildtype | -61.7 | 0.0 |
| Serine | -46.1 | 15.6 |
| Rosetta | -52.6 | 9.1 |
| Rosetta 3.4 | -53.4 | 8.3 |
| MOE | -61.3 | 0.4 |

# Chapter 4: Rational Design of Fusion Proteins for Diagnostic Applications


## ABSTRACT

The theme for this chapter is the rational design of fusion proteins. In the first section, we discuss a couple of proof-of-concepts where we fuse human and archaeal $O^6$-alklylguanine-DNA-alkyltransferases (AGTs) to antibodies. Doing so allows us to covalently and stoichiometrically label antibodies with oligonucleotides. In our first example, we attach a 93 base pair single-stranded DNA to an antibody, which then in turn serves as a template in an immunoPCR reaction. Using an immunosandwich assay, this strategy allows us to detect as little as 35 attomoles of interferon-gamma in a milliliter binding assay. Our second oligo-antibody fusion enables the use of DNA hybridization to programmably attach antibodies to surfaces. In collaboration with the Crooks laboratory at the University of Texas at Austin, we show electrochemical detection of human chorionic gonadotropin using oligo-antibody-silver nanoparticle fusions using their oSlip technology. In the second section, we extend the use of AGT fusions to pregnancy test strips. The use of an engineered hCG–SNAP fusion reporter protein (hCG-SNAP: human chorionic gonadotropin-O6- alkylguanine-DNA alkyltransferase) enabled loop-mediated amplification (LAMP)-to-hCG signal transduction on commercially available pregnancy test strips. Our novel diagnostic method reliably detected as little as 20 copies of Ebola virus templates in human serum and saliva. Further, we show that our strategy could be used to diagnose a common melanoma SNP allele (BRAF V600E) from the wild-type sequence. In the final section, we rationally design split nanoluciferase (nLuc) variants, which are then fused to AGTs to enable luminescent LAMP (Lumi-LAMP). We show the specific detection of HSV1 and β-actin using Lumi-LAMP and smartphones. We believe the methods described in this chapter are generalizable and will show great utility across a number of diagnostic applications.

## 4.1 SNAP-TAG LABELING OF ANTIBODIES ENABLES A PANOPLY OF DIAGNOSTIC TOOLS

### Introduction

The SNAP-tag protein labeling system allows specific, covalent attachment of virtually any molecule to a protein of interest. It is a two-step process requiring cloning and expression of the protein as a SNAP-tag fusion, and labeling of the fusion with the SNAP-tag probe of choice. SNAP-tag fusion proteins have been used extensively over a wide range of biotechnological applications. By expressing a single fusion, proteins can be functionalized to: fluorophores, surface-coupling linkers, biotin, oligonucleotides, and other molecules of interest (Kampmeier et al., 2009). The 20 kDa recombinant human $O^6$-alklylguanine-DNA-alkyltransferase (hAGT) has been engineered to use modified benzyl purine and pyrimidine derivatives for covalent attachment of molecules that can then be functionalized with a number of commercially available linkers (Keppler et al., 2004). For example, benzyl guanine (BG) with a NHS ester can be used to to allow for specific conjugation with 5' amine modified oligonucleotides (Gu et al., 2013), resulting in covalent attachment of oligonucleotides to proteins. The system has also been further optimized for both bacterial and mammalian expression systems, widening the range of potential protein fusions (Kampmeier et al., 2009). In addition to the commercially available SNAP-tag, an archael thermostable AGT from *Sulfolobus solfata*ricus has also been shown to be functionalizable with commercial SNAP-tag substrates (Perugino et al., 2012).

Here we discuss a number of proof-of-concepts with fusing the human and archael AGTs to antibodies. We begin with the development of immunoPCR reagents, whereby a PCR template is attached to an antibody. ImmunoPCR is the combination of two powerful diagnostics: PCR and immunodetection of analytes. For a full review on the power of immunoPCR, see references (Assumpção and da Silva, 2016; Chang et al., 2016; Ryazantsev et al., 2016). Using an

immunosandwich assay, we were able to detect as little as 35 attomoles of interferon-gamma, a cytokine used in the diagnosis of some cancers.

Next, we extend our use of oligo-tagged antibodies to the DNA-addressed attachment to silver nanoparticles. We found that an antibody that was incapable of binding antigen after passive adsorption to nanoparticles could be attached to the nanoparticle through nucleic acid hybridization while retaining binding to the target analyte using our fusions. This antibody-DNA conjugate was then used in the electrochemical detection of human chorionic gonadotropin. The limit of detection in this assay was 31.5 ng/mL. Together, these projects extend the many uses of SNAP-tag technology and allowed us to develop the techniques needed to design fusions for novel diagnostic strategies discussed in sections 2 and 3.

## Results & Discussion

### *Development of immunoPCR reagents*

In order to carry out immunoPCR, antibodies must be conjugated to DNA (**Figure 4-1)**, which is then in turn used as the template in a standard PCR or qPCR reaction. The use of SNAP-tag technology enables convenient, site-specific conjugation at a defined stoichiometry of 1 DNA/SNAP-tag. For example, an scFv-SNAP-tag fusion is labelled once, while a full-length IgG fusion to either the light or heavy chain is labeled twice, due to the dimeric nature of full-length antibodies. The SNAP-tag reacts specifically with a synthetic derivative of guanine, benzylguanine (BG),which then leads to irreversible covalent labeling of the SNAP-tag (Engin et al., 2010).

SNAP-tagged oligonucleotide labeling of antibodies is a two-step process that first requires cloning and expression of an antibody-AGT fusion, followed by labeling of the fusion with an appropriately modified DNA oligonucleotide. As a proof of concept, we engineered a construct in which the coding sequence of an anti-ricin single-chain fragment variable (scFv) antibody that binds specifically to ricin A chain was genetically fused to the SNAP-tag cassette, which allowed

site-specific conjugation of BG-modified substrates to the antibody, in this case an oligonucleotide. The fusion antibody was purified by nickel affinity and size exclusion chromatography before being used for site specific oligonucleotide conjugation. Binding assays indicated that SNAP fusion did not alter the binding of anti-ricin antibody (**Figure 4-2**). To attach the PCR template to the antibody, an oligonucleotide probe was first modified by conjugating BG-GLA-NHS (New England Biolabs, Ipswich, MA), an amine-reactive building block, with an amine functionalized oligonucleotide. Unreacted BG-GLA-NHS was removed by using NAP-5 desalting column (GE Healthcare life sciences, Piscataway, NJ). BG functionalized oligonucleotide were then mixed with SNAP-fusion anti-ricin antibodies in 3:1 ratio respectively, and final purification was carried out by using Hi-trap-Q sepharose anion exchange column (GE Healthcare life sciences, Piscataway, NJ). The formation of conjugate was validated by separating the products in a SDS polyacrylamide gel (**Figure 4-2C**).

Similarly, a preliminary sandwich-immunoPCR reaction, targeting IFN-ϒ, was performed using SNAP-tagged antibodies for both capture and detection. In this scheme, capture antibodies were covalently attached to benzylguanine-functionalized beads (NEB) via the SNAP-tag. Varying concentrations of antigen were then incubated with 20 μL beads in a 1 mL binding volume. We found that blocking the antibody-SNAP-tagged magnetic beads with tRNA greatly reduced nonspecific binding and background. After 3 washes by magnetic separation, a qPCR was initiated where the detection antibody was covalently attached to a 93mer oligonucleotide and subsequently used as the template in the PCR reaction. The lowest amount of antigen added (35 attomoles) was clearly detected in the assay (**Table 4-1**).

***Functionalization of silver nanoparticles with an oligonucleotide-conjugated antibody enables facile detection of a protein analyte***

The Crooks laboratory has previously published their oSlip PAD protocol, which has been used for paper-based immunoassay development using an electrochemical readout (**Cite Josephine**). The quantitative immunosandwich assay relies on detection of an analyte using antibodies on magnetic beads for capture, and antibody-coated silver nanoparticles (AgNPs) for electrochemical detection. In an effort to adapt the platform to pregnancy testing, we entertained the idea of using antibody-SNAPtag-Oligo fusions to programmably attach antibodies to AgNP surfaces *via* nucleic acid hybridization. This provided an interesting case because the detection antibody used in this reaction could not be passively adsorbed to AgNPs and remain functional. To determine oligo-antibody-AgNP coupling was possible, we expressed full-length IgG1 anti-human chorionic gonadotropin (hCG) antibody fusions and labeled them with short 22 base single-stranded DNA. Using thiolated oligos, we then functionalized silver nanoparticles (AgNPs) with the reverse compliment of this sequence. Oligo hybridization was then used to attach the antibodies to the AgNPs.

As the **Figure 4-3** shows, we were able to successfully attach the detection antibody to the AgNPs using oligo hybridization. Further, our unoptimized limit-of-detection was found to be near the 5 week mark of pregnancy (31.5 ng/mL). This project was abandoned due to the need to use $KMnO_4$ in the oxidation step, which made this incompatible with a new NoSlip variation of the technology. While we were not able to compete with commercial tests using this technology, we believe that attaching antibodies to surfaces using oligonucleotides will be of great utility that could be extended to a variety of surfaces that require antibody/protein adherence.

**Conclusions**

In addition to the antibodies shown above, we were able to show this to be generalizable to a number of other antibodies (**Table 4-2**). Over the course of these experiments, we learned a few lessons. First, while full-length IgG light-chain fusions did not appear to hinder binding, the SNAP-

tag did make the purification more difficult when expressed in the commercial pSNAPf vector. For full-length IgGs, we normally purify with Protein-A resin, typically a one-step process. However, the addition of the SNAP-tag, human or archael, prevented binding to the resin, resulting in lower yields. We were able to purify the light-chain fusions with nickel-NTA chromatography, using a histidine (6x) tag, followed by ion-exchange chromatography. Heavy-chain fusions did not suffer from this problem, and yields were typically much higher. Next, we learned that fusions in bacterial cells were problematic. In many cases, the fusions would express but contained truncations (**Figure 4-4**) that were difficult to purify out. We attributed the truncations to cryptic ribosomal binding and initiation sites. To mitigate truncations, it was much easier to express the fusions in yeast or mammalian cells (Figure XXX). Subsequently, we moved all of our expression of AGT-fusions to Expi293s, a common mammalian expression system. We also found that antibody fusion expression in pcDNA3.4 (Invitrogen) was far more robust and gave much higher yields, typically 10-20 mg from a 30 mL culture of Expi293s.The work here laid the foundation to what we call the "Antibody Production Pipeline" (**Figure 4-4**), which we used to supply reagents to a number of collaborators.

## Methods & Materials

### *Synthesis, expression, and purification of Antibody-SNAP fusion proteins.*

The DNA corresponding to antibody fragments were fused with the gene encoding either the human or thermophilic $O^6$-alkyl-guanine-DNA alkyltransferase from the thermophilic Archaea Sulfolobus solfataricus (SsOGT) (Perugino et al., 2012). Geneblocks were purchased from IDT. For mammalian expression, fusions were cloned into the pcDNA™ 3.4 vector from Invitrogen (Grand Island, NY, USA), which contains the human cytomegalovirus (CMV) immediate-early promoter/enhancer for high level gene expression (Gibson et al., 2009). Proteins were then either purified using Protein A, NiNTA, or ion-exchange chromatography. For bacterial expression,

fusions were cloned into modified pET21 vectors. Binding affinity was characterized by Enzyme-Linked Immunosorbent Assays (ELISA).

***Synthesis and purification of hCG-oligo probes***

Benzylguanine (BG) was conjugated to amine-functionalized oligonucleotides via the compound BG-GLA-NHS. Non aqueous DMSO was added to the BG-GLA-NHS vial (2 mg) with the help of a syringe to yield a final concentration of 20 mM. Amine- functionalized oligonucleotides were dissolved in water to a final concentration of 2 mM. Some 4 μL of oligonucleotide was mixed with 12 μL of BG-GLA-NHS along with 8 μL of 200 mM HEPES buffer (pH 8.3), and the mixture was shaken and incubated at room temperature for 2 h. Unconjugated BG was removed using a NAP-5 desalting column (GE Life Sciences, Pittsburgh, PA, USA). BG-conjugated oligonucleotides were stored at -20 ºC for further use. The conjugated probes were generated by combining BG-conjugated oligonucleotides with the purified hCG-stable SNAP in a 1:2.5 molar ratio in 1× TBS buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) followed by incubation at room temperature for 15-18 h. The unconjugated oligonucleotides and proteins were removed and the probes purified via fast protein liquid chromatography (FPLC) with a high trap Q Sepharose column (GE Life Sciences, Pittsburgh, PA, USA). Purified probes were concentrated on Amicon Ultra spin columns (Billerica, MA, USA) and purity was checked on a SDS-PAGE gel.

**Figure 4-1:** scFv:Snap Fusion

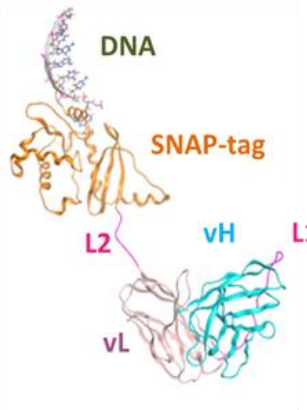



**Figure 4-1:** *scFv:Snap Fusion***.** Variable domains of the antibody are depicted in cyan and purple. L1 & L2 represent flexible linkers, shown in orange. Together these components are expressed as a single fusion protein. BG-functionalized oligonucleotides are then used to covalently modify the fusion protein with an oligonucleotide tag.

**Figure 4-2**: Binding profiles of anti-Ricin scFv and anti-Ricin-SNAP-tag fusion

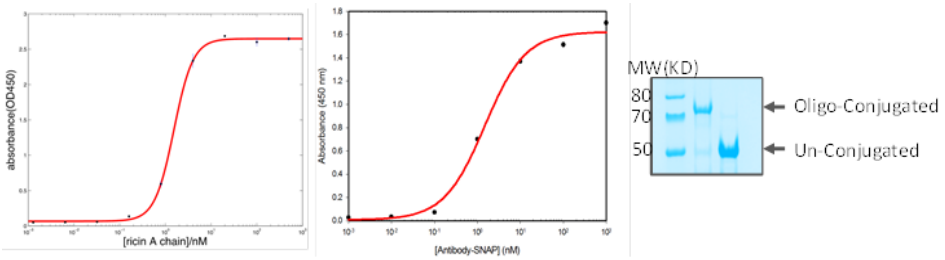



**Figure 4-2:** *Binding profiles of anti-Ricin scFv and anti-Ricin-SNAP-tag fusion***.** *Left,* ELISA showing binding of scFv. A non-linear regression fit revealed a Kd of 1.58 ± 0.2 nM. *Middle*, ELISA data of RCA43-SNAP-tag fusion. The Kd of the fusion was found to be 1.42 ± 0.3 nM and was within error of the scFv alone. *Right,* SDS-PAGE of anti-Ricin-pSNAP conjugated with 93mer oligonucleotide and purified by Hi-trap Q column. Control is unconjugated anti-ricin-SNAP fusion antibody.

**Table 4-1:** IFN-ϒ immunoPCR results

| Sample | Cq | Error | Slope |
|---|---|---|---|
| NTC | 17.16 | 0.07 | 2.25 |
| No Antigen | 16.56 | 0.11 | 2.22 |
| 35 picomoles | 2.54 | 0.02 | 0.11 |
| 3.5 picomoles | 5.73 | 0.15 | 2.31 |
| 350 attomoles | 10.98 | 0.27 | 2.24 |
| 35 attomoles | 12.31 | 0.30 | 2.21 |

**Table 4-2:** Constructed antibody fusions and Kd values

| Antibody | Type | Modifications | Kd (nM) |
|---|---|---|---|
| **Ricin** | IgG | StableSNAP | 1.52 |
| **IFN-ϒ** | IgG | StableSNAP | 0.13 |
| **GM-CSF** | IgG | StableSNAP | 0.10 |
| **TNF-α** | IgG | StableSNAP | 0.03 |
| **EpCAM** | IgG | StableSNAP | 7.0 |
| **P24** | IgG | StableSNAP | 30 |

**Figure 4-3:** oSLip electrochemical detection of anti-hCG-SNAP fusions

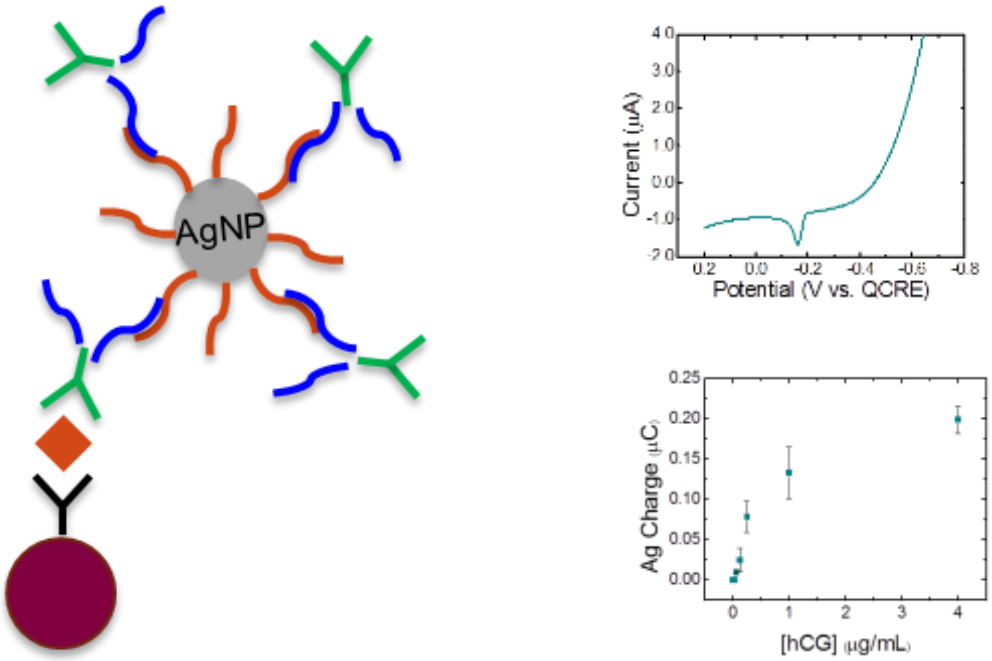



**Figure 4-3:** *oSlip electrochemical detection of anti-hCG-SNAP fusions*. (Left) Immunosandwich setup. Red ball indicates magnetic bead coupled to capture antibody. AgNPs are functionalized with a thiolated oligo, which contains the reverse compliment to an oligo attached to the antibody-SNAP fusion. (Right) Data collected using oSlip technology. Limit of detection (bottom right) was shown to be 31.25 ng/mL.

**Figure 4-4:** Antibody-oligo conjugation pipeline

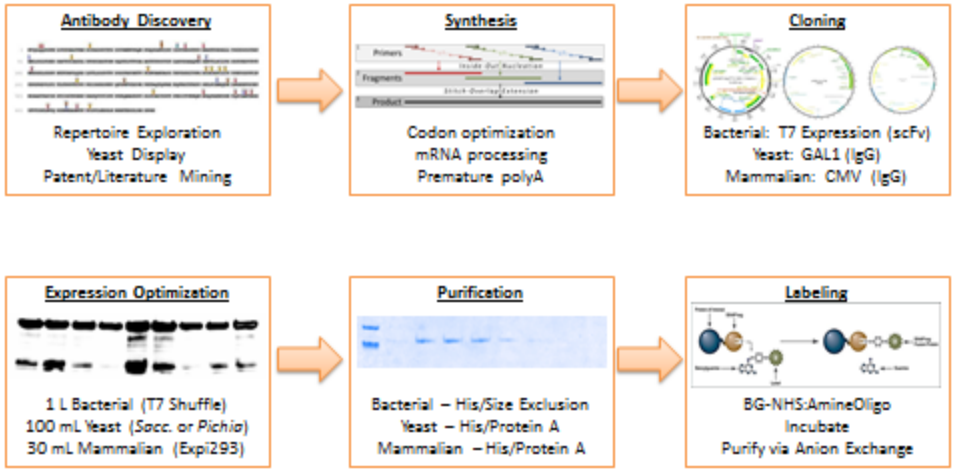



**Figure 4-4:** *Antibody-oligo conjugation pipeline*. Overview of the antibody production pipeline generated during this work.

## 4.2 Coupling Sensitive Nucleic Acid Amplification with Commercial Pregnancy Test Strips

### Introduction

With recent advancements in medical diagnostics, it has become increasingly more possible to quickly and accurately diagnose a variety of disease states using nucleic acid or protein biomarkers (Cheng et al., 2006; Vo-Dinh and Cullum, 2000; Wong, 2006). In order to increase the sensitivity of these methods, assays have been improved by the incorporation of target-based or signal-based amplification strategies (Giljohann and Mirkin, 2009). However, these methods are generally restricted to medical diagnostic laboratories, which require complex instrumentation and are geographically constrained. Low-cost assay methods that are accurate, reliable, and can be used in diverse or resource-limited locations are attractive alternatives for the early identification of disease causing agents (Urdea et al., 2006).

Point-of-care (POC) devices are low-cost and appropriate as a first line of defense against chemical and biological threats where rapid identification of exposure is essential (Daar et al., 2002; Tram et al., 2014; Yager et al., 2008; Zhu et al., 2014b). Despite many years of research, a limited number of POC products are commercially available. The most successful of these devices has been the personal glucose meter (PGM) (Clark and Lyons, 1962; Montagnana et al., 2009). Recently, the PGM was modified to quantify non-glucose targets by nucleic acid mediated release of invertase, which converts sucrose to glucose, from a functional DNA-invertase conjugate (Xiang and Lu, 2011). The PGM has also been adapted for aptamer- or DNAzyme-based detection of analytes (Gu et al., 2013; Lan et al., 2016; Wang et al., 2016b; Zhu et al., 2014a). Previously, our laboratory coupled the PGM to a loop-mediated isothermal amplification (LAMP), strand exchange signal transduction scheme for pathogen detection (Du et al., 2015; Jiang et al., 2015). However, usage of a glucometer in POC diagnostics is problematic for many biomarkers because targets exist in glucose-rich blood or urine samples. These limitations and other efforts (Liao and Li, 2014;

Wang et al., 2014) inspired our more generalizable approach to molecular diagnostics by transducing LAMP-mediated amplification to commercially available pregnancy test strips.

Utilizing pregnancy test strips in POC diagnostics is an attractive alternative, due to their low cost and ease of signal interpretation by end users (Marks et al., 2007). Here, we report the conversion of LAMP amplicons to commercially available pregnancy test strips, thereby combining molecular genomic diagnostics with a proven off-the-shelf device. This novel and modular technique is based on an hCG-SNAP-DNA fusion reporter protein that enables LAMP-induced signal transduction. The simplicity of our direct hybridization strategy ultimately resulted in a limit of detection of 20 copies of pathogenic Zaire Ebolavirus (ZEBOV), even in 10% human serum or 15% human saliva samples. To overcome the limitation of non-specific products of LAMP, we turned to strand exchange nucleic acid chemistry, which has been shown to mitigate false positives and can be used to probe and distinguish between highly similar amplicons (Du et al., 2015; Jiang et al., 2015). The toehold-mediated strand exchange reaction, termed one-step strand displacement (OSD), accurately distinguished single nucleotide polymorphisms (SNPs) in "YES-or-NO" detection of a mutant melanoma related biomarker BRAF allele (V600E) against the wild-type (WT) BRAF gene. Moreover, the novel toehold-mediated three-way junction strand exchange (3way-SE) strategy employed here can be readily adapted for signal-on detection of other nucleic acid biomarkers using pregnancy test strips and has great potential for use as a general signal transducer in other nucleic acid circuits.

## Results

### *Design of the hCG-SNAP:DNA fusion reporter*

In order to produce functional hCG-SNAP-tag fusions, capable of binding to pregnancy test strips, we first designed a series of N- and C-terminal hCGα- and hCGβ-thermostable SNAP-tag fusion proteins, expressed them in Expi293 cells, and tested their ability to bind a panel of

commercial pregnancy test strips. The C-terminal fusion of the thermostable SNAP-tag to hCGβ provided the most robust signal in our initial screening, and all further assay optimizations were conducted with this fusion (hCGβ-SNAP). Given that our hCGβ-SNAP fusion protein gave a robust signal during our initial pregnancy test strip screening, we wanted to determine if our fusion protein was negatively impacting binding to hCG antibodies. We used a standard ELISA and compared binding of our fusions against a commercially available preparation of hCG. Our hCG-SNAP protein had EC50 value of 1.0 nM , while the EC50 for the commercially obtained unmodified hCG was 45.5 nM, indicating that our hCG fusion did not negatively affect antibody binding (**Figure 4-5**). Next, we optimized conditions to label hCGβ-SNAP with a benzylguanine-labeled oligonucleotide, which would later serve as our LAMP probe, and found that we were able to label > 95% of hCGβ-SNAP. Further purification using anion exchange chromatography eliminated unlabeled fusions in our protein preparations for downstream testing (**Figure 4-6**).

**Adapting LAMP to a pregnancy test strip by direct hybridization**

The process of adapting LAMP to pregnancy test strips can be divided into three steps (**Figure 4-7**): (1) LAMP amplification; (2) direct hybridization of the LAMP amplicons and the specific hCG-DNA probe; (3) readout *via* a pregnancy strip. In this scheme, the LAMP reaction is carried out to produce complex, double-stranded concatameric amplicons, including four kinds of single strand loops positioning between F2-F1 (F-target), F2c-F1c (F-probe), B2-B1 (B-probe), and B2c-B1c (B-target), respectively. The LAMP amplicons are allowed to directly hybridize to the hCG-DNA probe that specifically recognizes one of the loop regions (here, the loop positioning between F2c-F1c) of the LAMP product. The hybridized products are then applied to the pregnancy strip. Preliminary experiments revealed that after the hCG-DNA probe hybridized with the large molecular weight LAMP amplicon, the complex was unable to flow up the strip, shown by a loss of

signal in the positive control and experimental lines. For the free loop regions that originated from off-target sequences, amplicons would not hybridize with the hCG-DNA probes or form complexes, allowing the hCG-DNA probe to migrate and produce a positive signal on the test strip. Together these results suggested that we were able to generate signal in the absence of the target sequence, thus creating a 'signal-off' detection method.

Encouraged by the results of our initial attempts at transducing LAMP products to pregnancy test strips, we then determined if we were able to reliably generate 'signal-off' detection of Zaire Ebolavirus (ZEBOV) and wilde-type BRAF sequences. Using our direct hybridization strategy, we were able to definitively detect each of the synthetic templates (**Figure 4-8**). Further, we were able to detect as few as 20 copies of the ZEBOV templates and 2,000 copies of WT-BRAF templates within 1.5 hrs. The reproducibility of the direct hybridization method developed here provided very sensitive YES-or-NO answers for both targets. Due to the high selectivity for both LAMP primers, we were also able to demonstrate that non-relevant LAMP products did not produce false positive responses**.** Next, we tested the stability of the hCG-SNAP-DNA probes. Ten aliquots of the probe were lyophilized,stored at room temperature, and assayed in 30 day intervals. The hCG-Ebola probe was robust and could be stored for at least 90 days without large signal deviations (**Figure 4-9)**.

Further demonstration of the performance of our diagnostic platform was shown by directly detecting ZEBOV template spiked into human saliva and human serum (**Figure 4-10**). The fresh collected human saliva and serum were used directly without any pre-treatments. The LAMP to pregnancy test strip method with the direct hybridization strategy showed true positives in both 15% saliva and 10% serum sample without signal or sensitivity loss.

***Adapting LAMP and OSD to pregnancy strip transduction to discriminate the WT- and SNP-BRAF gene.***

Detecting mutations that predispose or diagnose an individual to a specific cancer is of particular importance, and nucleic acid circuitry can be used to probe and distinguish between mutant and wild-type alleles. In previous work, we were able to demonstrate the utility of oligonucleotide-strand displacement (OSD) probes in SNP detection (Jiang et al., 2015). Here, we extend OSD probes to our newly developed pregnancy test strip assay (**Figure 4-7**). To do so, we rationally designed the OSD probes with a single-mismatch located in the toehold region. The hybridization of the reporter to the correct target loop is initiated at the toehold, which then proceeds by branch migration. This process ultimately leads to the release of the shorter signal strand.

It has been shown that the V600E mutation of the BRAF gene is one of the melanoma-associated nucleic acid biomarkers. We generated LAMP primers and hCG-labeled OSD reporters for both wild-type and V600E BRAF alleles. LAMP-OSD to hCG strip transduction was carried out with 0 and $2x10^6$ copies of both the WT- BRAF gene and SNP-BRAF gene (T to A transversion). When the hCG-B-WT reporter was introduced, no band appeared on test line for the WT-BRAF LAMP amplicons, as expected. Conversely, when the hCG-B-SNP reporter specific for the V600E allele was used, the mutant gene could be detected (Figure **4-11**). These results successfully validated that our method could accurately distinguish SNP- and WT-BRAF gene in LAMP amplicons with high signal amplitude ($\Delta I$).

***Three-way junction strand exchange (3way-SE) allows for 'Signal-on' transduction***

To produce "signal-on" detection of target sequences using pregnancy test strips (target indicated by typical positive pregnancy-test result), we designed a new three-way junction strand exchange (3way-SE) probe and immobilized it on streptavidin-modified magnetic beads (See **Figures 4-7** and **4-12** for details). To detect LAMP amplicons of ZEBOV and WT-BRAF using

this strategy, samples (30 μL) were mixed with 1 μL 3waySE-modified magnetic beads (0.5 mg/mL) and incubated for 30 min prior to magnetic separation. The supernatant was then run on individual pregnancy test strips (**Figure 4-12**). Using 3waySE, we were clearly able to distinguish between positive and negative signals on the test-strips. That is, typical positive signals on the test strips corresponded to the correct targeted amplicons. Astonishingly, as little as 20 copies of surrogate ZEBOV template produced easy-to-read positive signals on the test line of the pregnancy test strips. We were also able to repeat this process in LAMP reactions in 5% human serum (**Figure 4-12**).

***Discussion***

The detection of nucleic acid biomarkers for point-of-care (POC) diagnostics is currently limited by technical complexity, cost, and time constraints. To overcome these shortcomings, we developed a new nucleic acid detection method--combining loop-mediated isothermal amplification (LAMP), programmable nucleic acid circuitry (i.e. toehold-mediated strand exchange reaction), and standard pregnancy test strips. The incorporation an engineered hCG-SNAP fusion reporter protein has allowed for "LAMP-to-hCG" singal transduction, thus enabling the detection of nucleic acid biomarkers on low-cost, commercially available pregnancy test strips. Our assay reliably detected as little as 20 copies of ZEBOV-VP30 templates and was adapted to distinguish a common melanoma-associated SNP allele BRAF (V600E) from the wild-type sequence in both human serum and saliva.

The commercially available hCG strips we used here are 20 fold less expensive compared to custom made LFD strips, typically used in new fluorescence-based assays. Additionally, the DNA-hCG-SNAP fusion reporter, developed here, is stable at room temperature for months. This assay method is particularly amenable to further development for point-of-care applications and should be easily applicable to the detection of virtually any nucleic acid targets. These advances were

made possible through protein-engineering efforts that site-specifically and stoichiometrically labeled oligonucleotide probes to hCG. By focusing on molecular transduction, rather than on the development of a novel diagnostic platform, we were able to develop diverse assays that could be used to monitor public health issues.

## Methods & Materials

### *Chemicals and materials*

Commercial human chorionic gonadotropin (hCG) was obtained from Fitzgerald Industries International (Acton, MA, USA). BG-GLA-NHS, E.coli Strains DH-10B and Top 10, Gibson cloning master mix and all enzymes including Bst 2.0 DNA polymerase were obtained from New England Biolabs (Ipswich, MA, USA). The cloning vector pCDNA3.4 and Expi293™ expression system were from Invitrogen (Grand Island, NY, USA). Streptavidin-coated magnetic beads (1.5 μm in average diameter) were purchased from Bangs Laboratories Inc. (Fishers, IN, USA). The Amicon Ultra-2 mL 10 K was from Millipore Inc. (Billerica, MA, USA). Nickel affinity resin for affinity purification of the hCG-stable SNAP fusion protein was purchased from Thermo Fisher Scientific (Rockford, IL, USA). Prepacked High trap Q columns for hCG-oligo probe purification were from GE Life Sciences (Pittsburgh, PA, USA). The hCG protein standards were provided by Fitzgerald Industries International (Acton, MA, USA). All oligonucleotides used in these studies were purchased from Integrated DNA Technologies, Inc. (Coralville, IA, USA) and their sequences are listed in **Table 4-3**. The concentrations of DNA suspensions of oligonucleotides were measured by UV spectrophotometry using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Human serum was obtained from Sigma-Aldrich (St. Louis, MO, USA). To collect human saliva, a 1 mL sterile syringe (BD, Franklin Lakes, NJ, USA) with no needle was inserted under the tongue. Saliva was used immediately following collection. All other chemicals and solvents were purchased from Sigma-Aldrich, Inc. (St. Louis, MO, USA). The Blue Cross One

Step hCG Urine Pregnancy Test Strips (Blue Cross Bio-Medical Co, Ltd., Beijing, China) were obtained from Amazon.com.

***Synthesis, expression, and purification of hCG-stable SNAP fusion protein.***

The sequences of the alpha and beta subunits of hCG were synthesized by IDT (Morgan et al., 1975). The DNA corresponding to the hCG beta (hCGβ) chain was fused with the gene encoding a thermophilic $O^6$-alkyl-guanine-DNA alkyltransferase from the thermophilic Archaea Sulfolobus solfataricus (SsOGT); we refer to this herein as "stable SNAP" (Perugino et al., 2012) Geneblocks were purchased from IDT for the hCG alpha (hCGα) gene and the hCG beta (hCGβ)-stable SNAP fusion gene, with an intervening glycine serine rich linker (GGGGS GGGGS GGGGS). The beta fusions were constructed with an export tag (MEMFQGLLLLLLLS) at the N-terminus, and a 6xhistidine tag at the C-terminus. Both the hCGα gene and hCG beta (hCGβ)-stable SNAP fusion gene were inserted via Gibson cloning methods into the pcDNA™ 3.4 vector from Invitrogen (Grand Island, NY, USA), which contains the human cytomegalovirus (CMV) immediate-early promoter/enhancer for high level gene expression (Gibson et al., 2009). Sequence-verified expression constructs of both hCGα and hCGβ-stable SNAP were transiently transfected at a 1:1 ratio into suspension- adapted cells Expi293F™ cells. The hCGα:hCGβ-stable SNAP fusion protein was secreted into the media and purified via nickel affinity column chromatography. The typical yield of the full fusion protein was about 8 mg/30 mL culture, and its purity was routinely characterized by denaturing gel electrophoresis. Binding affinity was characterized by Enzyme-Linked Immunosorbent Assays (ELISA) (Bhadra et al., 2015) using anti-hCG IgG as the primary antibody and Goat Anti-Human IgG (Fcγ fragment specific) as the secondary antibody (Jackson ImmunoResearch Laboratpories Inc., West Grove, PA, USA).

***Synthesis and purification of hCG-oligo probe.***

Benzylguanine (BG) was conjugated to amine-functionalized oligonucleotides via the compound BG-GLA-NHS. Non aqueous DMSO was added to the BG-GLA-NHS vial (2 mg) with the help of a syringe to yield a final concentration of 20 mM. Amine- functionalized oligonucleotides were dissolved in water to a final concentration of 2 mM. Some 4 μL of oligonucleotide was mixed with 12 μL of BG-GLA-NHS along with 8 μL of 200 mM HEPES buffer (pH 8.3), and the mixture was shaken and incubated at room temperature for 2 h. Unconjugated BG was removed using a NAP-5 desalting column (GE Life Sciences, Pittsburgh, PA, USA). BG-conjugated oligonucleotides were stored at -20 $^{o}$C for further use. The conjugated probes hCG-P1, hCG-P2, hCG-P3, hCG-P4 and hCG-P6 were generated by combining BG-conjugated oligonucleotides with the purified hCG-stable SNAP in a 1:2.5 molar ratio in 1× TBS buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) followed by incubation at room temperature for 15-18 h. The unconjugated oligonucleotides and proteins were removed and the probes purified via fast protein liquid chromatography (FPLC) with a high trap Q Sepharose column (GE Life Sciences, Pittsburgh, PA, USA). Purified probes were concentrated on Amicon Ultra spin columns (Billerica, MA, USA) and purity was checked on a SDS-PAGE gel.

***Cloning of Zaire Ebolavirus VP30 and human v-raf murine sarcoma viral oncogene homolog B1 (BRAF) gBlocks and PCR amplification of transcription templates.***

The ZEBOV, wild-type BRAF (WT-BRAF) and V600E mutant single nucleotide polymorphisms (a T to A transversion) BRAF (SNP-BRAF) gBlocks were designed to include a T7 RNA polymerase promoter at their 5'-ends to enable subsequent transcription. The DNA polymerase chain reaction (PCR) amplification reactions were performed using AccuPrime™ Pfx DNA Polymerase (Thermo Scientific, Wilmington, DE, USA). The PCR products were purified from agarose gels using the Wizard SV gel and PCR purification system (Promega, Madison, WI, USA). The ZEBOV gBlocks were cloned into the pCR2.1-TOPO vector (Life Technologies,

Carlsbad, CA, USA) by Gibson assembly using the 2x mastermix (NEB) according to the manufacturer's instructions. Cloned plasmids were selected and maintained in an *E. coli* Top10 strain. Plasmid minipreps were prepared from these strains using the Qiagen miniprep kit (Qiagen, Valencia, CA, USA). All gBlock inserts were verified by sequencing at the Institute of Cellular and Molecular Biology Core DNA Sequencing Facility.

***Standard LAMP reaction***

The LAMP primers, composed of the outer primers F3 and B3 and the inner primers FIP and BIP, were generated by PrimerExplorer 4 (Eiken). Primer design was constrained to include at least a 30 bp gap between the F1 and F2 as well as between the B1 and B2 priming sites. Mixtures containing different copies of template (ZEBOV DNA plasmid or BRAF PCR products),1.6 μM each BIP and FIP, 0.4 μM each B3 and F3, 1 M betaine, 2 mM MgCl2, and 0.8 mM dNTPs in a total volume of 29 μL 1 × isothermal amplification buffer (20 mM Tris- HCl, 10 mM (NH4)2SO4, 10 mM KCl, 2 mM MgSO4, 0.1% Triton X-100, pH 8.8) were heated to 95°C and maintained for 2 min, followed by chilling on ice for 2 min. Then, 1μL (8 U) of Bst DNA polymerase 2.0 was added to initiate the LAMP reaction. The reactions (with a final volume of 30 μL) were incubated for 1.5 hour in a thermal cycler maintained at 65 oC. A 3.5 μL aliquot of the reaction mixed with 1.5 μL of 6 × DNA loading dye was then analyzed by electrophoresis through a 1% agarose gel containing ethidium bromide.

***End-point LAMP-hCG to pregnancy test strip transduction***

A series of tubes containing 10 μL aliquots of hCG-Ebola (18 nM) or hCG-Braf (8 nM) were mixed with 30 μL standard LAMP products with different starting concentration of templates. The hybridization reactions were allowed to proceed for 35 min at 25 °C. Subsequently, the

pregnancy testing strips were dipped into the reaction solutions and the pictures were taken after 2 min by using a cell phone (iphone 6 plus, Apple, CA, USA).

***Standard LAMP reaction in human serum and human saliva.***

Mixtures containing different copies of ZEBOV template, 1.6 μM each BIP and FIP, 0.4 μM each B3 and F3, 1 M betaine, 2 mM $MgCl_2$, 0.8 mM dNTPs and 10% human serum (or 15% human saliva) in a total volume of 29 μL 1 × isothermal amplification buffer were heated to 95°C for 7 min, followed by chilling on ice for 2 min. Then, 1 μL (8 U) of Bst DNA polymerase 2.0 was added to initiate the LAMP reaction. The reactions (with a final volume of 30 μL) were incubated for 1.5 hour in a thermal cycler maintained at 65 $^{o}$C.

***Stability tests for the lyophilized hCG-oligo probe (hCG-Ebola).***

15 μL of 18 nM hCG-Ebola probe was added into each 1.5 mL tubes and was lyophilized by using a freeze dryer (SP Scientific, NY, USA). The lyophilized probe was stocked at room temperature (about 25 $^{o}$C). The stability test was carried out every 30 days for a total of 90 days. Sixty μL ZEBOV negative and positive samples (2E4 copies) were added into the lyophilized hCG-Ebola probes, respectively. After 35 min hybridization reactions at 25 °C, the pregnancy testing strips were dipped into the reaction solutions and the pictures were taken after 2 min by using a cell phone (iphone 6 plus, Apple, CA, USA).

***Single nucleotide polymorphisms (SNP) test for WT- and SNP-BRAF genes through strand displacement (OSD) probe.***

BRAF OSD reporter, WT-reporter (or SNP-reporter), were prepared by mixing hCG-B-WT-P1 (or hCG-B-SNP-P1) and B-P2 with 1:1 molar ratio (final concentration of 200 nM) in 1× TBS buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.4) and incubating at 25 $^{o}$C for 1 h

hybridization. Then, 30 µL WT-BRAF (or SNP-BRAF) LAMP products negative and positive samples (2E6 copies) were added into 2 µL 200 nM hCG-B-WT reporter (or 3.5 µL 200 nM hCG-B-SNP reporter) and the mixture were incubated for 40 min at 25 ºC. The pregnancy testing strips were dipped into the final reaction solutions and the pictures were taken after 2 min by using a cell phone (iphone 6 plus, Apple, CA, USA).

***Signal-on strategy for detection of Ebola through three-way junction strand exchange (3way-SE) probe.***

Preparation of the 3way-SE/MB conjugate: Fifteen µL mixture of 30 µM biotinylated E-P2 and 31 µM E-P3 was heated to 95 ºC for 5 min, then cooled slowly to 25 ºC at a rate of 0.1 ºC /s. Two hundreds µL of 1 mg/mL streptavidin coated MBs were washed three times with 1× isothermal amplification buffer by using the magnetic rack and resuspended in 85 µL isothermal amplification buffer. Fifteen µL pre-hybridized E-P2-E-P3 was added into the 85 µL washed MBs and incubated on a vertical rotator for 30 min. The unbound E-P2/E-P3 was removed by washing the E-P2/E-P3/MBs in 200 µL 1× isothermal amplification buffer at least three times. Then the E-P2/E-P3/MBs were suspended in 70 µL 1× isothermal amplification buffer and incubated with 30 µL 1.56 mg/mL hCG-E-P1 on a vertical rotator for 1.5 h at 25 ºC. After eight times washes with 200 µL 1× isothermal amplification buffer, the hCG-E-P1/E-P2/E-P3/MBs (i.e. 3way-SE/MBs) were suspended in 200 µL SuperBlock™ (PBS) Blocking Buffer (Thermo Scientific, Wilmington, DE, USA) and incubated on a vertical rotator over night at 4 ºC. Finally, the hCG-E-P1/E-P2/E-P3/MBs were washed with 200 µL 1× isothermal amplification buffer at least eight times, suspended in 400 µL 1× isothermal amplification buffer with the final concentration of 0.5 mg/mL, and stored at 4 ºC for further use.

Detection of Ebola LAMP products with 3way-SE-to-prengncy test strip transduction: 1 µL aliquots of the 0.5 mg/mL 3way-SE/MB conjugates were incubated with 30 µL of standard LAMP (or LAMP in 10% human serum) products with different starting concentration of templates. The

3way-SE reactions were allowed to proceed for 30 min on a vertical rotator at 25 ℃. The MBs were separated using a magnetic rack and the resulting supernatant was transferred into separate tubes for pregnancy strips test. The pictures were taken after 2 min by using a cell phone (iphone 6 plus, Apple, CA, USA).

***Data analysis.***

The analysis of relative signal intensities was carried out using Adobe Photoshop CS5.1 (Adobe Systems Incorporated, CA, USA) and Image J.[8] The relative intensities were measured as follows by opening the image in Photoshop CS5.1; and converting it to greyscale, then transferring the image to Image J, where background subtractions were carried out. The relative intensity of signal response could be quantitatively calculated as: Relative Intensity = [intensity of the test band]/[intensity of control band]. The bar-graphs in maunscript present the quantification of inset picture using software, Image J. The error bars of the bar-graphs represent standard deviations obtained from two parallel assays. All the figures are fabricated likewise unless otherwise noted.

**Contributions**

J Gollihar designed and built the fusion proteins, aided in experimental design, analyzed data, and wrote the manuscript. A Pothukuchy and Y Du designed and performed all experiments and wrote the manuscript.

**Figure 4-5.** ELISA comparison of commercial and oligonucleotide-conjugated hCG to pregnancy test antibodies

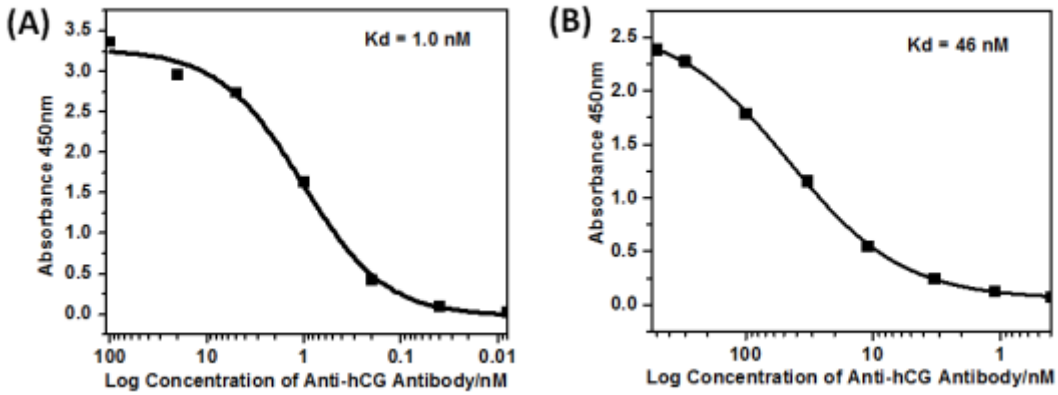



**Figure 4-5.** *ELISA comparison of commercial and oligonucleotide-conjugated hCG to pregnancy test antibodies*. Dose response curves were generated for the in-house hCG-oligonucleotide conjugate (A) and a commercially purchased hCG sample (B). The corresponding EC50 values were 1.0 nM and 46 nM, respectively.

**Figure 4-6** hCG-SNAP fusion design and purification

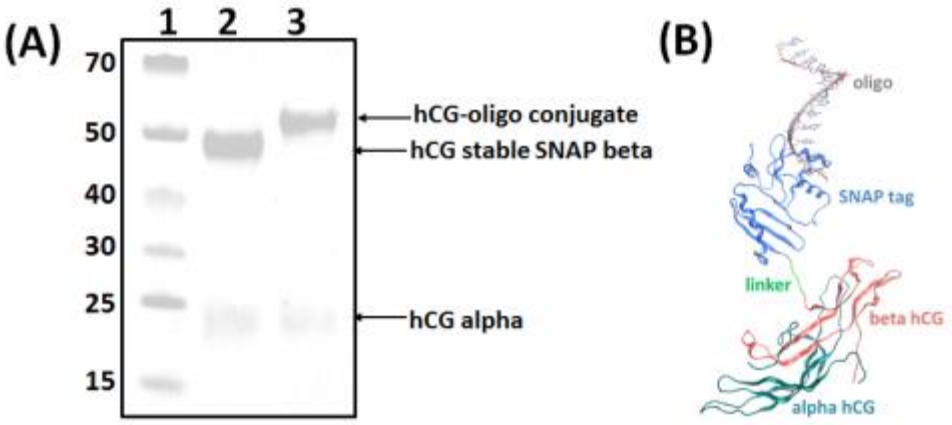



**Figure 4-6** *hCG-SNAP fusion design and purification* (A) SDS PAGE of hCG-SNAP fusion protein and hCG-oligo conjugate; (B) The structure of hCG-SNAP fusion protein.

**Figure 4-7**. Three strategies for detection of ZEBOV or wild-type BRAF (WT-BRAF) LAMP amplicons using pregnancy test strips

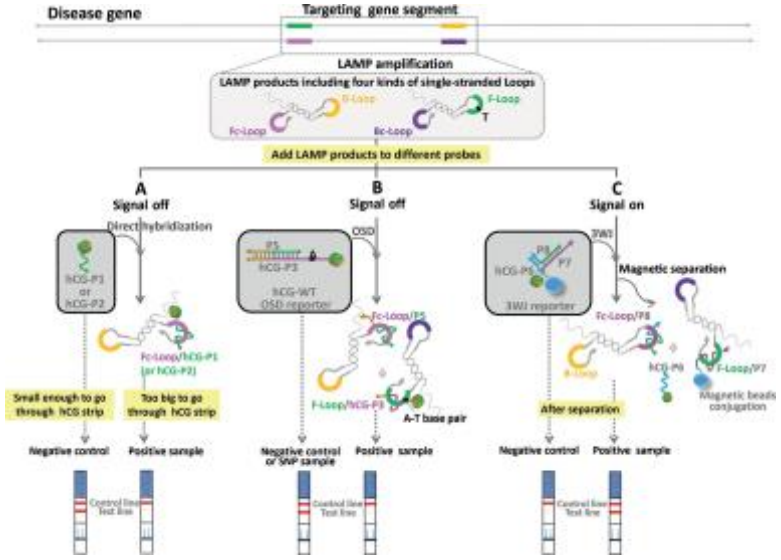



**Figure 4-7**. *Three strategies for detection of ZEBOV or wild-type BRAF (WT-BRAF) LAMP amplicons using pregnancy test strips*. A) "Signal-off" direct hybridization strategy. B) "Signal-off" OSD strategy for discriminating mismatches. C) "Signal-on" 3way-SE strategy.

**Figure 4-8** Sensitive and specific detection of synthetic ZEBOV and WT-BRAF

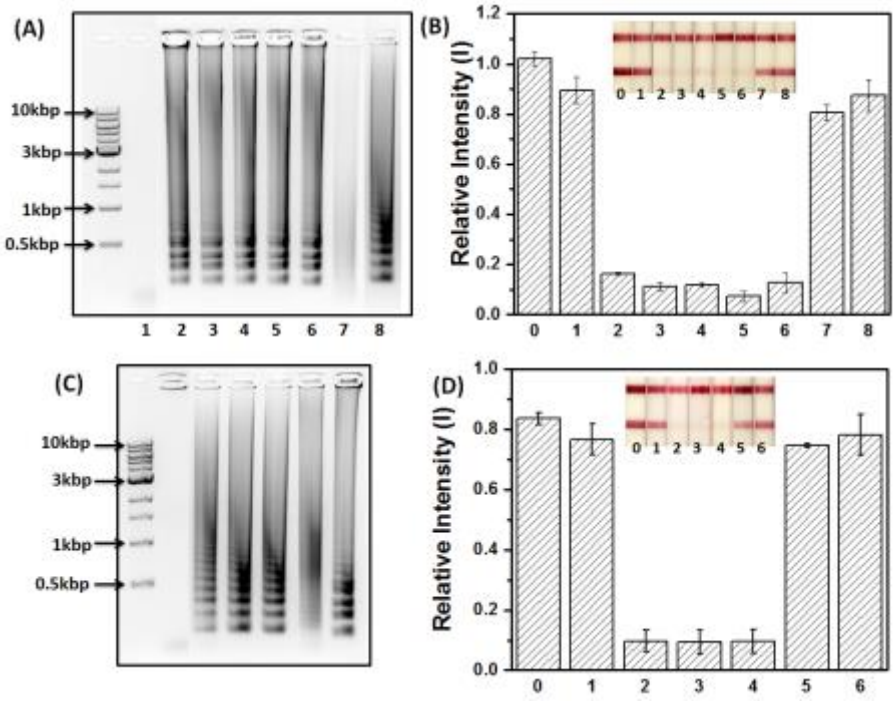



**Figure 4-8** *Sensitive and specific detection of synthetic ZEBOV (ZEBOV-VP30) and WT-BRAF DNA using LAMP to-hCG transduction*. The LAMP reaction was carried out at 65 $^{o}$C for 1.5 hours. (A) Agarose gel electrophoretic characterization of LAMP amplicons using ZEBOV-VP30 as a template. Lane 1: Negative control without template; lane 2: template 20 copies; lane 3: template 200 copies; lane 4: template 2,000 copies; lane 5: template 20,000 copies; lane 6: template 200,000 copies; lane 7: side products; lane 8: Non-target with WT-BRAF lamp products. (B) The hCG strip (inset) responses to both non-target (WT-BRAF) and different amounts of target (ZEBOV-VP30) templates, as mentioned in A. Strip 0 is hCG-E19 in 1X Iso buffer. (C) Agarose gel electrophoretic characterization of LAMP amplicons using WT-BRAF as a template. Lane 1: Negative control without template; lane 2: template 2,000 copies; lane 3: template 20,000 copies; lane 4: template 200,000 copies; lane 5: side products; lane 6: Non-target with ZEBOV-VP30 lamp products. (D) The hCG strip (inset) responses to both non-target (ZEBOV-VP30) and different amounts of target (WT-BRAF) templates, as mentioned in A. Strip 0 is hCG-Braf in 1X Iso buffer. The error bars represent standard deviations calculated from two parallel assays.

**Figure 4-9.** Time dependence of hCG strip responses to LAMP amplicons

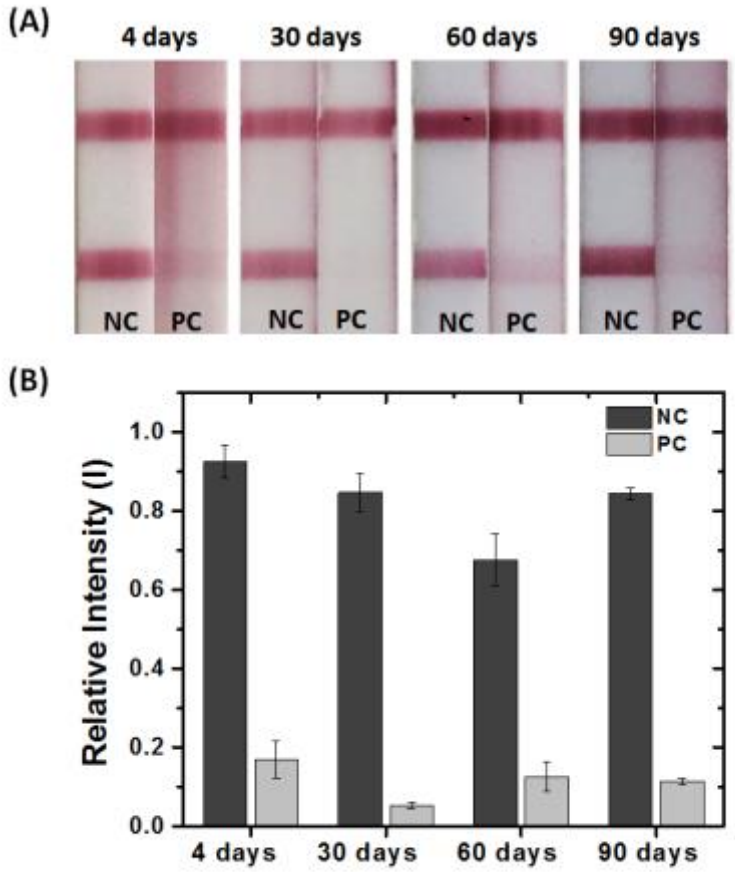



**Figure 4-9.** *Time dependence of hCG strip responses to LAMP amplicons*. Data generated from $2x10^4$ copies of ZEBOV-VP30 using the lyophilized hCG-Ebola probe. The bar-graph presents the quantification of inset pictures using Image J. The error bars represent standard deviations calculated from three parallel assays.

**Figure 4-10 10** Sensitive and selectivity of synthetic ZEBOV DNA

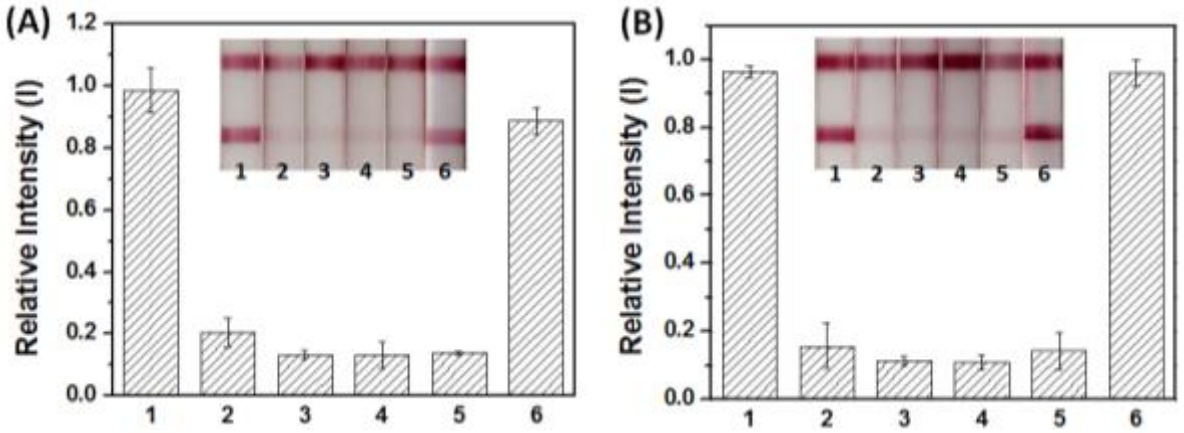



**Figure 4-10** Sensitive and selectivity of synthetic ZEBOV DNA (ZEBOV-VP30) in 15% human saliva (A) and in 10% human serum (B) using LAMP to-hCG transduction. The hCG strip (inset) responses to different amounts of target (ZEBOV-VP30) templates and off-target Braf templates. Strip 1: negative control; strip 2: 20 copies; strips 3: 200 copies; strip 4: 2,000 copies; strip 5: 20,000 copies; strip 6: 2,000,000 copies of Braf template. 2. The bar-graph presents the quantification of inset picture using software, Image J. The error bars represent standard deviations calculated from two parallel assays.

**Figure 4-11** Detection of single nucleotide polymorphisms (SNPs) using strand displacement (OSD) probes.

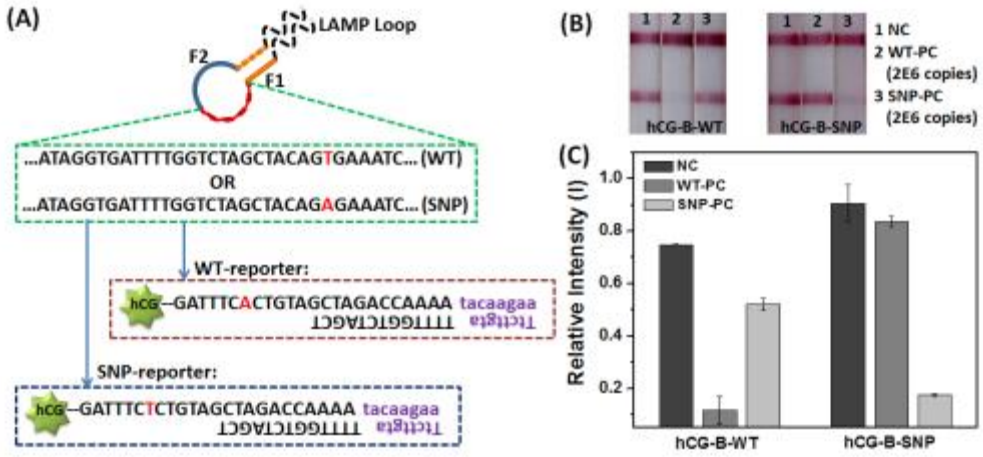



**Figure 4-11** *Detection of single nucleotide polymorphisms (SNPs) using strand displacement (OSD) probes*. (A) Sequence interactions between LAMP amplicons that represent either the wild-type (WT) BRAF gene or the allele V600E (SNP, which contains a single nucleotide difference) and the OSD probes. The discriminating nucleotide position is shown in red in both amplicons and the OSD probes (where it is found in the toehold region). (B) The hCG strip responses to both WT and SNP-Braf templates with either WT-hCG or SNP-hCG probe. (C) The bar-graph presents the quantification of (B) using Image J. The error bars represent standard deviations calculated from two parallel assays.

**Figure 4-12:** 3way-SE detection of ZEBOV and BRAF

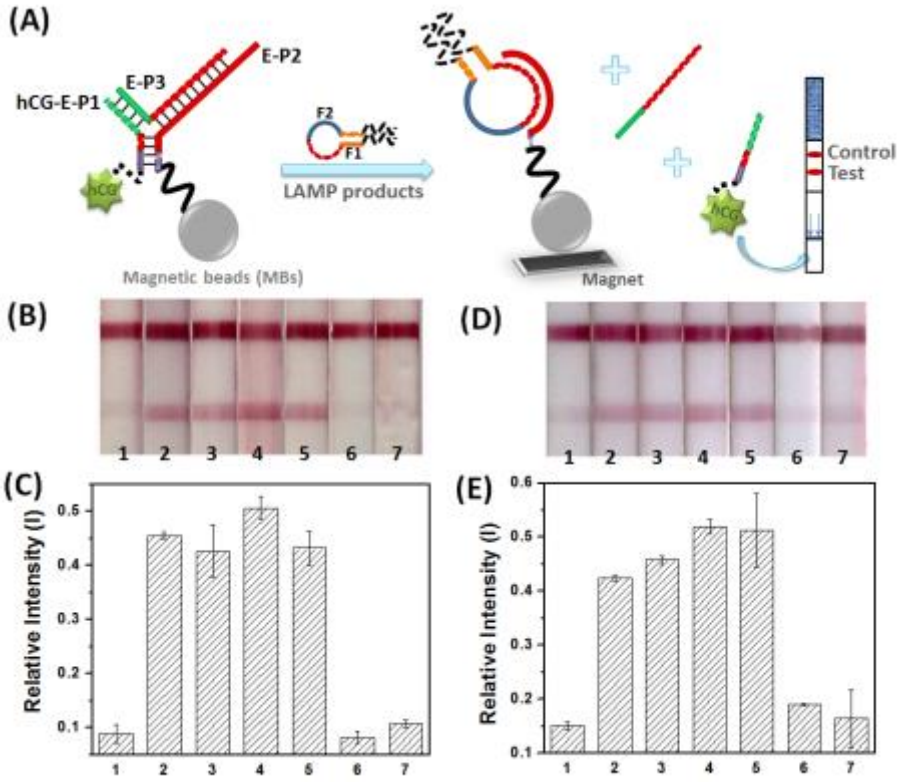



**Figure 4-12:** *3way-SE detection of ZEBOV and BRAF* (A) The scheme of "signal on" 3way-SE strategy for detecting of ZEBOV VP30 LAMP amplicons using LAMP to-hCG transduction. Sensitive and selectivity of synthetic ZEBOV DNA (ZEBOV-VP30) in buffer (B and C) and in 5% human serum (D and E) through (B) and (D) The hCG strip responses to different amounts of target (ZEBOV-VP30) templates, off-target Braf template, and non-target of Braf LAMP product. Strip 1: negative control; strip 2: 20 copies; strips 3: 200 copies; strip 4: 2,000 copies; strip 5: 20,000 copies; strip 6: 5,000,000 copies of Braf template with Ebola primers; strip 7: 5,000,000 copies of BRAF LAMP product. (C) and (E) The bar-graph presents the quantification of pictures B and D, respectively, using Image J. The error bars represent standard deviations calculated from two parallel assays.

**Table 4-3.** Sequence of oligonucleotides used for Pregnancy Lamp.

| **Name** | | **Sequence** 5'-3' |
|---|---|---|
| ZEBOV. 3.3 primer set | F3 | AGACAGCATTCAAGGGATG |
| | B3 | CCTTTTTTCAAGGTCGGACA |
| | FIP | CTCCTTGATTGACGGTACTCACCGACACGACCACCATGTTC |
| | BIP | CTCACAAGTGCGCGTTCCTAATGTCTTTAGGTGCTGGAG |
| BRAF primer set | F3 | CCACAGAGACCTCAAGAGT |
| | B3 | ACAGAACAATTCCAAATGCATAT |
| | FIP | ACTGATGGGACCCACTCCATAAGACCTCACAGTAAAAATA |
| | BIP | AACAGTTGTCTGGATCCATTTTGTGACATCTGACTGAAAGCTGTA |
| hCG-Ebola probe | | /5AmMC6/AATTCTCTCTGGATGATGA |
| hCG-Braf probe | | /5AmMC6/TAGGTGATTTTGGTCTAGC |
| Braf OSD reporter set (for SNP detection) | hCG-B-WT-P1 | /5AmMC6/GATTTCACTGTAGCTAGACCAAAATACAAGAA |
| | hCG-B-SNP-P1 | /5AmMC6/GATTTCTCTGTAGCTAGACCAAAATACAAGAA |
| | B-P2 | TTCTTGTATTTTGGTCTAGCT |
| Ebola 3 way junction reporter set (for signal on strategy) | hCG-E-P1 | /5AmMC12/ CTCTGCACAGTCGGAC |
| | E-P2 | GAGCACGATCATCATCCAGAGAGAATTATCGAGGTGCAGAGA AAAAAAAAAA /3BiodT/ |
| | E-P3 | GTCCGACTCTCGATAATTCTCTCTGGATGAT |
| | Mimic-T | TGCACCTCGATAATTCTCTCTGGATGATGATCGTGCTC |

## 4.3 Rational design of split nanoluciferase fusions enables luminiscent LAMP (Lumi-LAMP) and point-of-care smartphone diagnostics

### Introduction

Loop-mediated isothermal amplification (LAMP) is a powerful sequence amplification technology that relies on strand displacement DNA synthesis to generate $10^9$ to $10^{10}$ copies of nucleic acid targets within 1-2 hours (Notomi et al., 2000). Isothermal operation is ideally suited for adaptation to point-of-care diagnostic devices, as it does not require complex devices capable of thermal cycling (Yager et al., 2008). There are four primers in a typical LAMP reaction: two inner primers (FIP and BIP) and two outer primers (F3 and B3). These primers are specific to six consecutive blocks of target sequences designated in their order of appearance from the 5'-end as B3, B2, B1, F1c, F2c and F3c. The final LAMP product is comprised of large molecular weight concatemers that contain free loops between the F1 and F2, B1 and B2, F1c and F2c, and B1c and B2c regions that can be further exploited for detection and for speeding up amplification.

LAMP has frequently been limited by the accumulation of non-specific products and by the lack of versatile methods for unambiguous sequence distinction. To overcome this, we have engineered LAMP signal transduction modules called oligonucleotide strand displacement (OSD) probes that are triggered by LAMP loop sequences and that are essentially the equivalent of TaqMan probes for isothermal amplification. OSD probes are comprised of short DNA partial duplexes that undergo strand exchange reactions in the presence of specific loop sequences that terminate LAMP concatamers. The longer strand of the OSD probe duplex is complementary to the target-derived sequence in one of the LAMP loops.

Because of the versatility of OSD probes, we were recently able to transduce LAMP amplicons into glucose for detection by a commercial glucometer. For LAMP-to-glucose conversion the toehold-bearing OSD strand is immobilized on magnetic beads and pre-hybridized to the shorter, complementary strand that is covalently coupled to an in-house engineered

thermostable invertase. Specific LAMP amplicons initiate toehold-mediated strand displacement resulting in the release of invertase-bearing strands into the solution. Catalytic conversion of sucrose to glucose by isolated solution-phase invertase is quantified with a commercial glucometer. By combining the nucleic acid transducer circuit with a thermostable invertase we can use this high temperature isothermal amplification method in a one-tube format to directly transduce analytes such as ebolavirus templates into glucose signals, with a sensitivity to detect as few as 20-100 copies per microliter (Du et al., 2015). Moreover, as mentioned in the previous section, we were able to transduce LAMP amplicons to pregnancy test strips using an hCG-AGT fusion with an OSD reporter (Du et al., 2017).

In an effort to expand our use of OSD probes to commonly used devices, we endeavored to develop a smartphone-based point-of-care diagnostic. By engineering a split thermostable nanoluciferase (NLuc) into a sequence-dependent bioluminescent probe for LAMP amplicons, we were able to transduce LAMP amplicons to smartphones using a free application on an iPhone after 30 seconds of exposure. We believe that the preliminary experiments shown here will be of great utility in developing a panel of point-of-care diagnostics using our split nLuc.

**Results & Discussion**

nLuc is a small (19 kDa), single subunit, ATP-independent luciferase that produces a high intensity glow-type luminescence with a half-life >2 h (He et al., 2014). The specific activity of nLuc has been reported to be up to 150-fold greater than the commonly used firefly luciferases. nLuc does not possess disulfide bonds or any known post-translational modifications. The enzyme is stable and active at elevated temperatures. nLuc is also active over a wide pH range (5-9) and has been shown to remain active after being challenged with 8 M urea. Together these properties render nLuc an ideal candidate for engineering.

To minimize background luminescence and obviate the need to physically separate unused LAMP reporters, we sought to split nLuc into two non-luminescent fragments that would regain luminescence only upon proximity-dependent association. By utilizing LAMP loop sequences as scaffolds to facilitate split nLuc reconstitution, LAMP amplicons can be specifically transduced into bioluminescence (**Figure 4-13**).

Several split sites were selected based on the predicted NLuc structure (**Figure 4-14**). N- and C-terminal halves were fused with the SNAP-tag (**Figure 4-15**) and expressed in *E. coli.* Fusion proteins were purified using Ni-affinity chromatography and conjugated with benzylguanine-labeled oligonucleotides. The split reporters were then either incubated alone or in various equimolar combinations at 55 °C for 1 h. Luminescence upon adding Nano-Glo substrate was measured on a plate reader to measure background reconstitution. Duplicate fusion protein sets were incubated with equimolar amount of non-specific or a complementary target strand to foster adjacent hybridization of the oligonucleotides linked to the N and C-terminal halves of nLuc. Luminescence was measured after 1 h at 55 °C and compared to the activity in the absence of oligonucleotide hybridization. Our results demonstrate that a 122 amino acid N-terminal fragment ($N^{NLuc}$) and a 49 amino acid C-terminal fragment ($C^{NLuc}$) (N3:C3 pair in **Figure 4-16** reconstitute more efficiently in the presence of oligonucleotide hybridization (**Figure 4-17**).

Therefore we selected this pair of split proteins to create reporters for LAMP amplicons. We covalently conjugated the thermostable SNAP-tag of $N^{nLuc}$ fusion protein via benzylguanine to the 3'-end of single stranded oligonucleotide probe 1 (P1-$N^{NLuc}$) while the $C^{nLuc}$ fragment was conjugated to the 5'-end of the 3'-end blocked oligonucleotide probe 2 (P2-$C^{NLuc}$). A portion of the single stranded probes in P1-$N^{nLuc}$ and P2-$C^{nLuc}$ were subsequently converted to the OSD format (OSD1-$N^{nLuc}$ and OSD2-$C^{nLuc}$) by hybridization to the shorter complementary oligonucleotides. Probes: OSD 1 and 2 are designed to hybridize with adjacent sequences in target-derived LAMP

loop regions, thus juxtaposing the 3'-end $N^{nLuc}$ of probe 1 and the 5'-end $C^{nLuc}$ of probe 2. The ensuing split reporter reconstitution can be quantitated by applying its substrate furimazine (Promega) and measuring luminescence using a plate reader (**Figure 4-18**). In the absence of specific LAMP amplicons, the split NLuc reporter displays minimal reconstitution and luminescence. The bioluminescence triggered by the correct LAMP amplicons can also be visually distinguished from background (**Figure 4-19**) and can also be imaged using a smartphone camera without employing additional hardware (**Figure 4-20**).

**Conclusions**

To our knowledge, this work represents the first functional split of nanoluciferase. Additionally, we believe this to be the first example of oligo-mediated functionalization of this class of enzymes. Our rational design of nLuc enabled the facile detection of 2 LAMP amplicons in a sequence specific manner. This approach should prove generalizable to other targets and enable the use of smartphone in point-of-care diagnostics.

**Methods & Materials**

***Synthesis, expression, and purification of split nLuc-SNAPtag fusions***

The sequences of split nLuc designs were synthesized by IDT and ligated into a modified bacterial pET expression vector (**Figure 4-15**). Sequence-verified expression constructs of both were expressed in BL21 DE3 and purified via NiNTA column chromatography.

***Oligo-mediated activity check***

A total of 0.6 μM of each nLuc probe (N-nLuc and C-nLuc; Table 4-4) and a LAMP mimic were incubated in isothermal buffer (20 mM Tris-HCl, 10 mM ammonium sulfate, 50 mM KCl, 2

mM $MgSO_4$, 0.1% Tween20, pH 8.8) at 55˚C for 1 hour. 0.5 µL of Nano-Glo substrate was used in each reaction. Endpoint luminescence measurements were taken with a TECAN F500 platereader.

***Standard LAMP reaction***

The LAMP primers, composed of the outer primers F3 and B3 and the inner primers FIP and BIP, were generated by PrimerExplorer 4 (Eiken). Primer design was constrained to include at least a 30 bp gap between the F1 and F2 as well as between the B1 and B2 priming sites. Mixtures containing 10 pg template, 1.6 µM each BIP and FIP, 0.4 µM each B3 and F3, 1 M betaine, 2 mM MgCl2, and 0.8 mM dNTPs in a total volume of 29 µL 1 × isothermal amplification buffer (20 mM Tris- HCl, 10 mM (NH4)2SO4, 10 mM KCl, 2 mM MgSO4, 0.1% Triton X-100, pH 8.8) were heated to 95°C and maintained for 2 min, followed by chilling on ice for 2 min. Then, 1µL (8 U) of Bst DNA polymerase 2.0 was added to initiate the LAMP reaction. The reactions (with a final volume of 30 µL) were incubated for 1.5 hour in a thermal cycler maintained at 65 oC. A 3.5 µL aliquot of the reaction mixed with 1.5 µL of 6 × DNA loading dye was then analyzed by electrophoresis through a 1% agarose gel containing ethidium bromide.

***Synthesis and purification of nLuc-OSD probes***

Benzylguanine (BG) was conjugated to amine-functionalized oligonucleotides via the compound BG-GLA-NHS. Non aqueous DMSO was added to the BG-GLA-NHS vial (2 mg) with the help of a syringe to yield a final concentration of 20 mM. Amine- functionalized oligonucleotides were dissolved in water to a final concentration of 2 mM. Some 4 µL of oligonucleotide was mixed with 12 µL of BG-GLA-NHS along with 8 µL of 200 mM HEPES buffer (pH 8.3), and the mixture was shaken and incubated at room temperature for 2 h. Unconjugated BG was removed using a NAP-5 desalting column (GE Life Sciences, Pittsburgh, PA, USA). BG-conjugated oligonucleotides were stored at -20 ºC for further use. The conjugated probes generated by combining BG-conjugated oligonucleotides with the purified nLuc-stable SNAP in a 1:2.5 molar ratio in 1× TBS buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) followed

by incubation at room temperature for 15-18 h. The unconjugated oligonucleotides and proteins were removed and the probes purified via fast protein liquid chromatography (FPLC) with a high trap Q Sepharose column (GE Life Sciences, Pittsburgh, PA, USA). Purified probes were concentrated on Amicon Ultra spin columns (Billerica, MA, USA) and purity was checked on a SDS-PAGE gel.

***Real-time measurement of Lumi-LAMP***

LAMP reactions were incubated at 65˚C for 1 hr. Polymerase was then inactivated with a heating step of 80˚C for 20 minutes. Split nLuc probes were then added at 0.2 μM or 0.12 μM. A 1:100 dilution of Nano-Glo substrated was added, and the reaction was held at 55˚C. Measurements were taken at 2 minute intervals. For smartphone measurements, reactions were incabed for 15 minutes and exposed for 30 s using LongExpo, a freeware application found in the Apple App Store.

**Figure 4-13:** Split nLuc reporter (Lumi-LAMP) concept

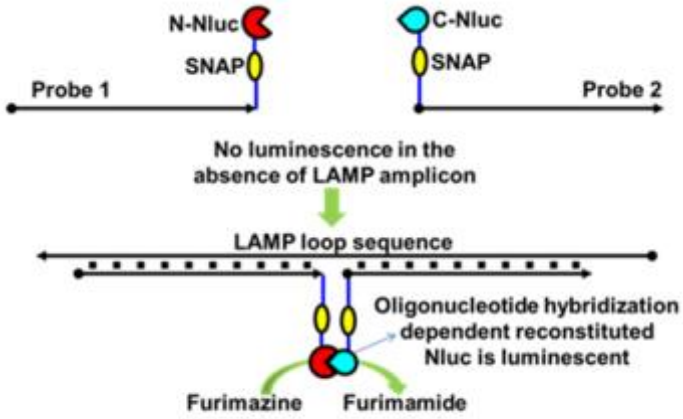



**Figure 4-13:** *Split nLuc reporter (Lumi-LAMP) concept.* In Lumi-LAMP, a split nanoluciferase becomes active via oligonucleotide hybridization. Functional nLuc is only produced by targeting a specific amplicon generated in a standard LAMP reaction.

**Figure 4-14:** Rational design of split nanoluciferase variants

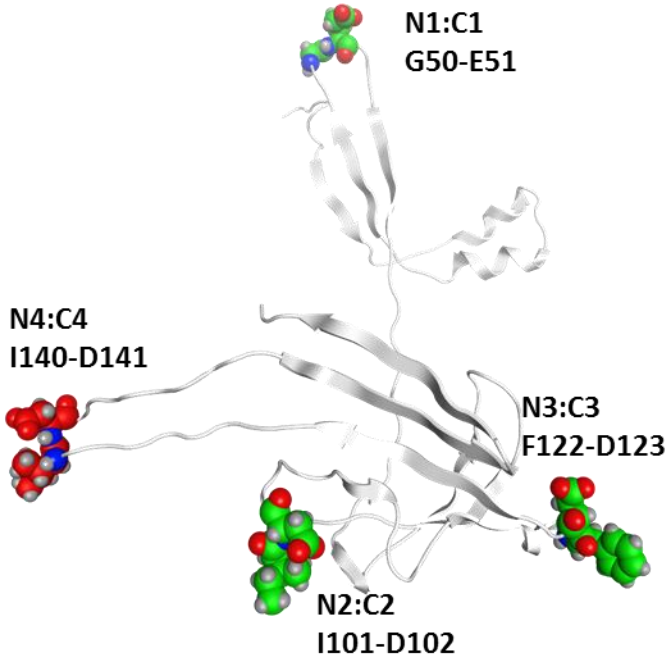



**Figure 4-14:** *Rational design of split nanoluciferase variants.* Four sites were chosen based on a homology model of nLuc. Sites were chosen based on secondary structure. Sites marked in green were shown to be functional in oligonucleotide-dependent activity assays. Site N4:C4, shown in red, was not functional.

**Figure 4-15:** nLuc primary sequence and AGT fusion strategy for split reporters.

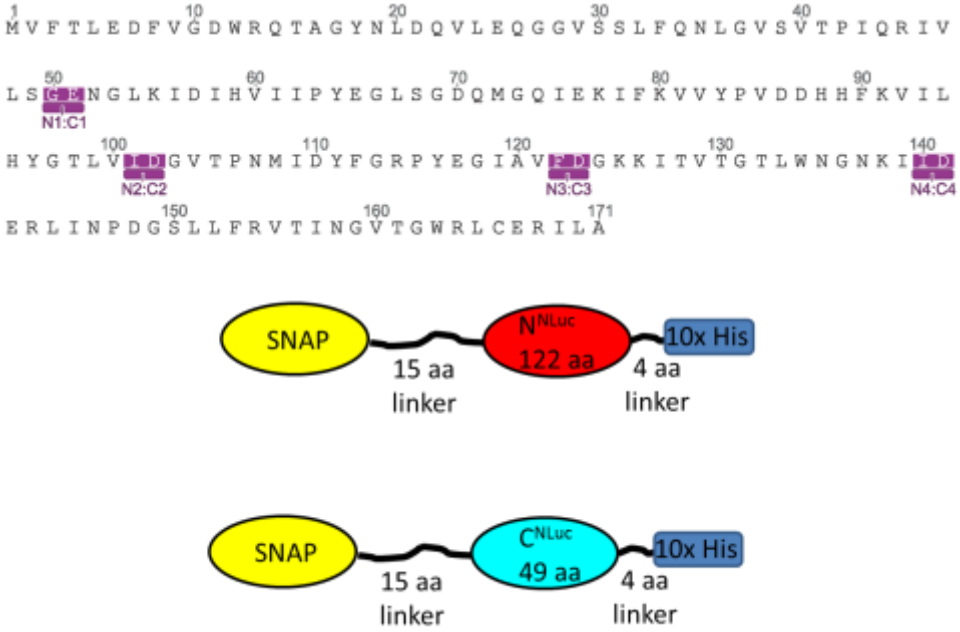



**Figure 4-15:** *nLuc primary sequence and AGT fusion strategy for split reporters*. Top, primary amino acid sequence of nLuc. Sites for split reporter are highlighted in purple. Bottom, fusion strategy for N- and C-termini of split nLuc variatns. In each case, the SNAPtag was fused to the N-terminus of each end of the split enzyme.

**Figure 4-16:** Reaction scheme for Lumi-LAMP proof-of-concept.

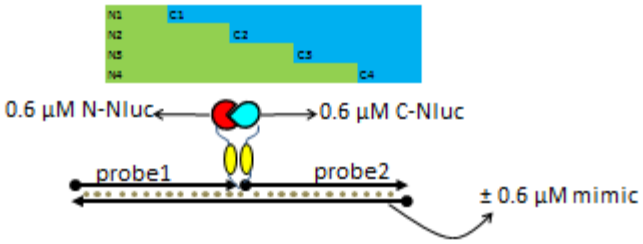



**Figure 4-16:** *Reaction scheme for Lumi-LAMP proof-of-concept.* The relative position of each of the split nLuc is shown at the top. For the reaction, 0.6 µM of each portion of the split nLuc was incubated with and without an equimolar amount of the LAMP probe mimic. Reactions are carried out at 55˚C.

**Figure 4-17:** Oligo-dependent activity of designed split nLucs.

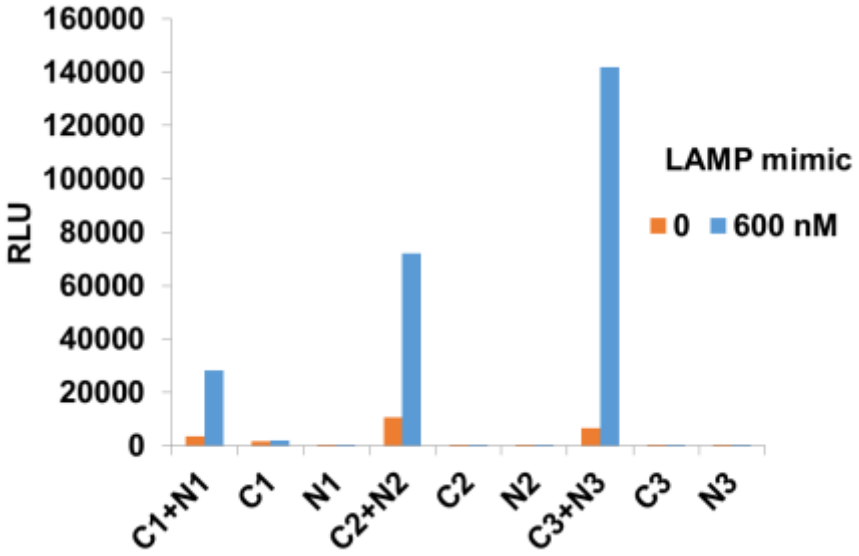



**Figure 4-17:** *Oligo-dependent activity of designed split nLucs*. Activity of split designs was measured using luminescence. In each case, oligo-dependent activity was observed with the addition of an equimolar amount of the LAMP mimic. Individual pieces of split enzymes were not reactive, and activity was dependent on the correct partner being present (not shown). C4:N4 as the only design that did not show oligo-dependent activity.

**Figure 4-18:** Real-time measurement of Lumi-LAMP products.

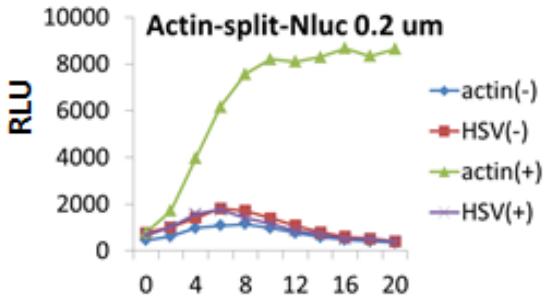



**Figure 4-18:** *Real-time measurement of Lumi-LAMP product*. Split nLuc variants with actin probes were incubated with actin LAMP products. Luminescence was measured every 2 minutes, revealing a large increase when both probes and LAMP product for actin were present.

**Figure 4-19:** Split nLuc OSD probes can distinguish between target and non-target LAMP amplicons.

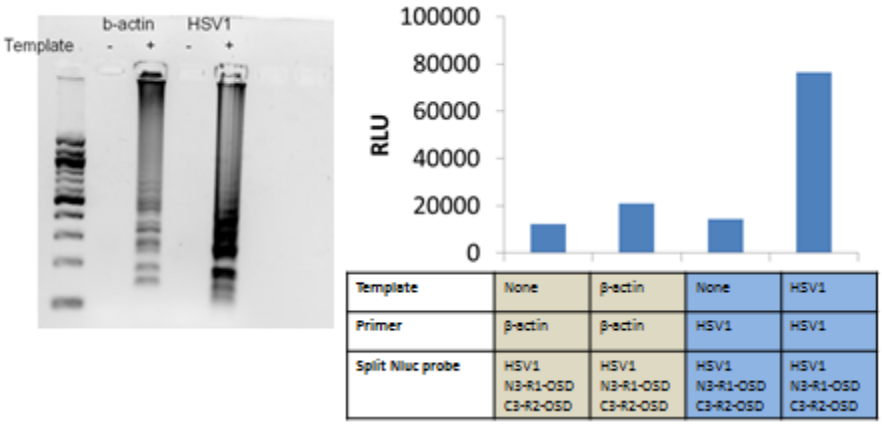



**Figure 4-19:** *Split nLuc OSD probes can distinguish between target and non-target LAMP amplicons.* Left, agarose gel of LAMP amplicons targeting β-actin or HSV1. Right, correct detection of HSV1 using HSV1 template and nLuc probes.

**Figure 4-20:** Lumi-LAMP to smartphone readout.

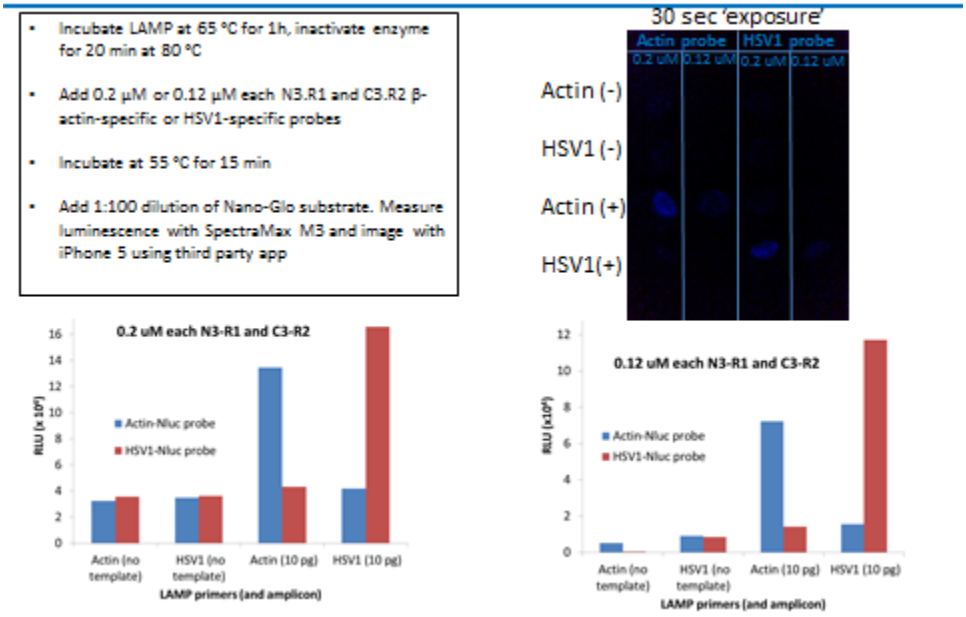



**Figure 4-20:** *Lumi-LAMP to smartphone readout. Top left,* assay conditions. *Bottom,* quantitative measurements of 0.2 µM vs 0.12 µM probes. LAMP reactions and controls are shown. Both actin and HSV1 could be distinguished on a smartphone (iPhone) after 30 seconds of exposure (bottom right).

**Table 4-4:** nLuc probe sequences

| Oligo Name | Sequence | NLuc conjugation partner | Target |
|---|---|---|---|
| HSV1.R1 | CCCCCAACATGACCCAGACAAAAAA/3AmMO/ | N3 | HSV1 |
| HSV1.R2 | /5AmMC6/AAAAAACGGCACCACCGACTCTCC/3InvdT/ | C3 | |
| Actin.R1 | GGAGGTGATAGCATTGCTAAAAAA/3AmMO/ | N3 | β-actin |
| Actin.R2 | /5AmMC6/AAAAAAGTGTAAATTATGTAATGC/3InvdT/ | C3 | |

# Chapter 5: Expanding the Reach of Compartmentalized Self-Replication


## ABSTRACT

The theme for this chapter is Compartmentalized Self-Replication. This directed evolution technique has been used by the scientific community to engineer thermostable DNA polymerases with a number of novel properties and functions. In the first section, we use DNA shuffling to create libraries of chimeric archaeal Family B DNA polymerases that we ultimately use in proof-of-concept selections. We find that shuffled polymerases are more active than parental enzymes. Further, we develop screening techniques that would allow us to perform multiple CSR directed evolution experiments in parallel. In the second section of this chapter, we develop a modified version of CSR, termed Reverse-Transcription Compartmentalized Self-Replication (RT-CSR). This novel selection strategy enables us to engineer the first known high-fidelity reverse transcriptase using a commonly used Family B DNA polymerase. With the aid of deep sequencing and modeling, we were able to halve the number of mutations needed to impart reverse transcriptase activity and re-engineered the polymerase to proofread mismatches on both DNA and RNA templates. Our selections at this point all produced error-prone polymerases. To combat this problem, we develop high-fidelity Compartmentalized Self-Replication (hifiCSR) in the third section. We also characterize inhibition of a model Family B DNA polymerase by chaotropes, heme, melanin, and lipoteichoic acids. We discuss the initiation of selections against these common inhibitors and provide sequencing data from each selection. Strong convergence is seen in many of the sequenced libraries after only five rounds of selection. In the final section, we extend the reach of CSR to a eukaryotic organism. The process by which we were able to perform CSR in *Saccharomyces cerevisiae* is discussed in detail. The ability to perform CSR in yeast provides the core technology needed to perform the more sophisticated directed evolution technique, Compartmentalized Partnered Replication (CPR), in a eukaryotic organism.

## 5.1 Preliminary Compartmentalized Self-Replication Directed Evolution Experiments

### Introduction

Compartmentalized Self-Replication (CSR) is a powerful directed evolution technique, originally developed in Phil Holliger's laboratory (Ghadessy et al., 2001b). In CSR, bacteria expressing individual polymerase variants are emulsified into individual water-in-oil emulsion bubbles. These bubbles can be thought of as individual reaction chambers or PCR 'tubes'. In the ideal case, the bubbles do not merge or mix with one another once they are formed. Emulsion PCR within these compartments in the presence of primers specific for the polymerase gene results in self-amplification of those genes that produce the most active polymerases. Amplified gene products are then re-cloned into plasmids for subsequent rounds of selection and amplification (**Figure 5-1**).

Using this selection strategy, thermostable DNA polymerases have been engineered to have novel functional properties using a variety of selection pressures. For example, heparin-resistant Taq variants were isolated in the first demonstration of CSR (Ghadessy et al., 2001b). Later, shuffling of Taq and other Family A polymerases led to variants that were resistant other environmental PCR inhibitors (Baar et al., 2011). Others have used CSR to train polymerases to use modified or unnatural bases during polymerization (Brown and Suo, 2011; Laos et al., 2014).

Our lab became interested in using CSR for the directed evolution of archaeal Family B polymerases. Thermostable family B polymerases are the high-fidelity enzymes commonly used in PCR reactions. In this section, we describe the shuffling of two commonly used Family B DNA polymerases and our first directed evolution experiments utilizing CSR. After only two round of selection, we identified two variants that were far more active than either of the parental enzymes. We also describe novel screening techniques that we developed in order to streamline CSR selections. The experiments described below would lay the foundation of future directed evolution experiments of DNA polymerases discussed in detail in later sections of this chapter.

## Results & Discussion

### *Library generation of DNA polymerases using gene shuffling*

To generate library diversity for selection of DNA polymerases with enhanced properties, two Family B DNA polymerases were chosen from thermostable archaeal organisms and shuffled. We chose KOD (*Thermococcus kodakaraensis*) and Pfu (*Pyrococcus furiosus)* because of their high level of fidelity and distinct functional properties. For example, KOD is more processive and faster (Takagi et al., 1997), and Pfu has been shown to generate longer amplicons in routine PCR reactions (Wang et al., 2004). We hypothesized that properties of each of these enzymes could be melded together to create novel, robust enzymes that would prove useful in a variety of biotechnological applications. To generate the libraries, the enzymes were codon optimized for both increased crossing-over during shuffling and expression in *Escherichia coli*. We reasoned that optimizing conditions for enhanced crossover would potentially lead to novel mutations not present in any of the parental enzymes. The polymerases were then synthesized in-house, fragmented by DNaseI digestion, and shuffled with equimolar ratios of fragments from each parental enzyme to reduce potential bias (Meyer et al., 2014b).

### *Selection of DNA polymerases using Compartmentalized Self-Replication*

In order to select novel polymerases, two independently created libraries ($> 10^7$) were subjected to two rounds of compartmentalized self-replication (CSR) (**Figure 5-1**). We chose minimal selective pressure in each round (1 min/kb) in an effort to decrease stringency and carry forward functional polymerases. We hypothesized that under these conditions, enhanced polymerases would begin to dominate the population over several cycles of PCR. To test for enhanced functionality after only two rounds of selection, the top variants were assayed again in emulsion PCR and purified to homogeneity for assays in bulk reactions. Target amplicons of 2.7 kb and 7 kb were amplified with each variant alongside wild-type controls. Two variants, A12 and PK6, were consistently better than KOD and Pfu under all conditions tested (**Figure 5-2**).

Additionally, A12 and PK6 were shown to be functional in qPCR using intercalating dyes, where a 1 second extension time was sufficient to amplify a 100 bp target (**Figure 5-3**). Moreover, A12 was able to amplify a 3 kbp target with only 15 seconds per kbp of extention time (**Figure 5-4**). We interpreted these results to mean that the evolved enzymes were indeed faster than the parental enzymes.

To elucidate possible mutations responsible for enhanced functionality in A12 and PK6, the polymerases were sequenced by the Sanger method and aligned. Interestingly, both variants greatly favored KOD, each sharing 94.5% identity with the parental enzyme (**Figure 5-5**). KOD and PK6 are both one amino acid shorter in length, highlighted by a gap at position 381. Each of the evolved polymerases acquired mutations that were not present in either of the wild-type enzymes (A12: E306G & V647A; PK6: E150G), possibly demarking sites of crossover. Homology modeling of these enzymes did not show any drastic changes in structure (**Figure 5-6**).

Given that we could select for polymerases with enhanced activity, we initiated selections under a variety of additional selection pressures: speed, processivity, and heme resistance (**Table 5-1**). To better address the need for faster, more robust DNA polymerases, we completed four additional rounds of CSR, using KOD, Pfu and Deep Vent as parental enzymes. We created libraries by shuffling random sequences generated from DNaseI treatment of the intact parental genes. The independently created libraries were then subjected to increasing amounts of selection pressure in three separate selections. To select for speed, the extension time was decreased in the CSR reaction from round to round. We started by allowing 1 min/kbp for the elongation step (~2.5 kb amplicon) in the emulsion PCR. This is a relatively small amount of selection pressure, and we anticipated only moving functional polymerases into the next round. The polymerases were only allowed 30 s/kbp in the second round, 10 s/kbp in round three, and 5 s/kbp in round 4. After

recovery of the 4th round, we sequenced the N- and C-termini (~420 AA of 776) to gauge the amount of diversity present in the population (**Figures 5-7 & 5-10**)

Processivity in polymerases generally refers to the number of bases polymerized after binding a primed target before the polymerase is released or "falls" off (Wang et al., 2004). Typically, Family B polymerases exhibit processivities between 20-200 bp in the absence of the proliferating cell nuclear antigen (PCNA) clamp, which allows these polymerases to processively synthesize tens of thousands of bases (Kitabayashi et al., 2002). In an attempt to select for more processive polymerases, we lowered the expression of the polymerase and extension time. Further, we forced the polymerase to ampify the full 7 kbp plasmid during the CSR reaction. After four rounds of selection, the survivors were able to elongate 7 kbp with only a 2.5 minute extension time per PCR cycle. Sequencing revealed that the populations were converging (**Figures 5-8 & 5-10**).

Heme-resistnace provided an interesting challenge. In initial testing, we found that parental enzymes were inhibited by heme ($IC_{50}$ = ~1-5 μM). Due to the large concentration of heme found in blood (~10 mM), we reasoned that evolving polymerases for heme-resistance would be a necessary first step in creating a polymerase to work in whole-blood (Akane et al., 1994; Al-Soud and R\a adström, 2001). We started at 10 μM heme and increased the concentration by an order of magnitude with each round. After round four, polymerases capable of handling physiologically relevant concentrations of heme (10 mM) were sequenced from the two starting libraries (**Figures 5-9 & 5-10)**.

Convergence for each of these libraries was observed after only 4 rounds of selection (**Figure 5-11**), giving us confidence that our selections were producing fully functional polymerase. However, it became apparent that hybrid polymerases would be difficult to tease apart mechanistically. Further, it was found that A12 and PK6 were error-prone, with error rates as high as 0.1%. This came as somewhat of a surprise because sequencing of these variants revealed that

the exonuclease, or proof-reading domain, was still intact. At this stage, we began to think about ways to do multiple selections in parallel and develop screening methodologies that would allow for multiplex selections. Given that KOD was the most robust Family B polymerase in emulsions and exhibited the highest fidelity, we turned our attention to evolving this polymerase in all subsequent selections.

***Development of screening technologies for CSR selections***

During our proofs-of-concept, we found that it took up to two weeks before a particular variant could be assayed after the selection was completed. The variant needed to be isolated and cloned into a suitable expression vector and then sequenced before purifying the enzyme using standard methods. The original pAK vector used for CSR selections did not produce enough polymerase for bulk assays using manageable volumes of culture (Ghadessy and Holliger, 2007). In fact, it required at least 500 mL of culture to produce enough enzyme to properly purify. For example, 25 mL of culture was not enough to reliably purify using NiNTA resin. However, once a variant was moved into a T7-based pET vector, 25 mL of culture was more than enough to purify usable quantities of polymerase (**Figure 5-12**). While this greatly reduced the cost and time of assaying a variant, this methodology was still slow. Fortunately, once a variant was cloned into the expression vector, we found that 0.5 μL of induced cells was enough to use in a typical PCR reaction (**Figure 5-12**). Further, cells could be stored as pellets at -20˚C or -80˚C before resuspension and testing. This allowed us to grow sub-milliliter quantities of cells in 96-well grow blocks. Further, cells could be screened in bulk using qPCR (**Figure 5-13**). This would allow us to screen up to 96 variants at once without having to run gels. We were also able to assay rounds within a day, rather than 6-8 variants over the course of two weeks.

***Conclusions***

The early experiments described above were important for a number of reasons. First, our laboratory learned how to perform rounds of Compartmentalized Self-Replication. Our first attempt

at CSR identified two highly active polymerases in A12 and PK6, and we would later use their apparent low-fidelity as a means to create error-prone libraries for future CSR and CPR evolution experiments. Many of the laboratories at the University of Texas at Austin still use A12 as a means of generating library diversity for directed evolution experiments. Secondly, we learned that we could change selection pressures and see convergence in our libraries. This gave us quite a bit of confidence in the technology. Finally, we streamlined the process of screening our libraries. Rather than waiting weeks for small numbers of active polymerases, we could screen hundreds in a single day. The techniques developed here would inform and shape selections described below.

## Methods & Materials

### *CSR selections and emulsion PCR*

Libraries were cloned into the pAK vector (Ghadessy and Holliger, 2007) and transformed into BL21s. Overnight library cultures were seeded at a 1:20 ratio into fresh 2xYT media supplemented with ampicillin and grown for 2 hours at 37◦C. Polymerases were induced with IPTG and incubated at 37◦C for 4 hours. Induced cells (400 µL total) were spun in a tabletop centrifuge at 3,000 x g for 8 minutes. The supernatant was discarded and the cell pellet was resuspended in 300 µL CSR mix: 1x Selection buffer (50 mM Tris-HCl (pH8.8), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$), 260 µM dNTPs, 530 nM of each of the forward and reverse primers, and 10-300 µM TMAC. The resuspended cells were placed into a 2 mL tube with a 1mL rubber syringe plunger and 1200 µL of oil mix (73% Tegosoft DEC, 7% AbilWE09 (Evonik), and 20% mineral oil (Sigma-Aldrich)). The emulsion was created by placing the cell and oil mix on a TissueLyser LT (Qiagen) with a program of 42Hz for 4 minutes. The emulsified cells were thermal-cycled for 20-25 cycles.

Emulsions were broken by spinning the reaction (20,000xg) for 15 minutes, removing the top oil phase, adding 750 µL chloroform, vortexing vigorously, and finally phase separating in a phase lock tube (5Prime). The aqueous phase was cleaned using a Promega PCR purification column. Overnight treatment with DpnI mitigated plasmid carryover. Subamplification with

recovery primers ensures that only polymerases were functional in the emulsion PCR amplified. Selection pressures are shown in **Table 5-1**. A representative CSR mastermix setup is shown in **Table 5-2**. A detailed protocol can be found in **Appendix I**.

***Purification of polymerase variants***

Overnight cultures of BL21 (DE3) harboring each of the variants were grown overnight in Superior broth at 37°C. Cells were then diluted 1:250, and protein production was induced with 1 mM IPTG during mid- log at 18°C for 20 hrs. Harvested cells were flash-frozen and lysed by sonication in 10 mM phosphate, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol, pH 7 (Buffer A). Cleared cell lysates were heated at 85˚C for 25 min, cooled on ice for 20 minutes, and filtered (0.2 μm). The filtrate was then passed over a NiNTA column and eluted with imidazole. Pooled fractions were then applied to a heparin column (GE Healthcare), concentrated, and dialyzed into storage buffer (50 mM Tris- HCl, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.1% Non-idet P40, 0.1% Tween20, 50% glycerol, pH 8.0). Working stocks were made at 0.2 mg/mL.

***Homology modeling of polymerase variants***

Homology models of the DNA polymerase variants A12 and PK6 were constructed using the parental enzymes, KOD and Pfu, as reference structures (PDBs: 4K8Z and 2JGU). To create the models, sequence identities were calculated, and RMS deviation values were used as guideposts in manipulating potential structures. Using the Amber12EHT force field with R-Field treatment of solvation electrostatics, an average of the 100 lowest energy conformations was used in the workup. Once the final homology model was complete, it was superimposed against the reference structures. Further evaluation of the models was checked by calculating Phi-Psi angles (Figure 6). Finally, the homology models were refined using fine grain energy minimization.

***Small scale purification for screening***

Overnight cultures of BL21 DE3 harboring polymerase variants in a pET21-derived expression vector were diluted 1:20 into 25 mL of 2XYT growth medium. Cells were grown for an additional 2 hours before inducing expression with 1 mM IPTG for 4 hours. Cells were then collected by centrifugation, washed in TBS, and resuspended in 2 mL of B-PER II cell lysis buffer (Thermo). The cells were then incubated at 80˚C for 20 min on a thermal block with light agitation. Debris was cleared by centriguation (20,000xg) for 15 minutes. Lysate was then incubated with 25 uL NiNTA resin and step eluted with imidazole. Purified protein was dialyzed into 50 mM Tris-HCl, 50 mM KCl, 0.1% Tween20, pH 8.0.

***Whole-cell PCR and qPCR screening***

Overnight cultures of BL21 DE3 harboring polymerase variants in a pET21-derived expression vector were diluted 1:20 into 500 μL or 4 mL of superior growth medium. Cells were grown for an additional 2 hours before inducing expression with IPTG (1 mM) for 4 hours. Cells were then collected by centrifugation, washed in TBS, and resuspended in 1X PCR buffer (Accuprime). Next, 0.5-10 μL cells were incubated with primers specific to a 550 bp piece of the pET21 vector. EvaGreen intercalating dye was used to monitor fluorescence in standard 30 μL reactions using a LightCycler96.

**Figure 5-1:** General CSR cycle

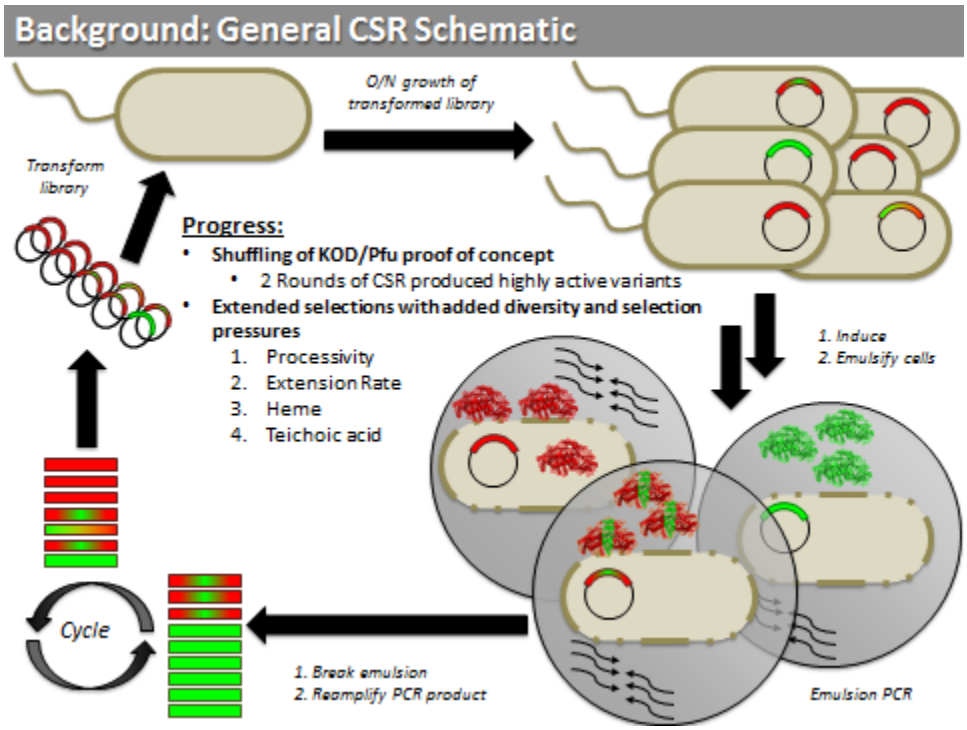



**Figure 5-1:** *General CSR cycle*. Depiction of the in vivo and in vitro stages selection using CSR. Once protein has been expressed within a cell, it is compartmentalized in a water-in-oil emulsion

**Figure 5-2.** Emulsion and Bulk PCR Results

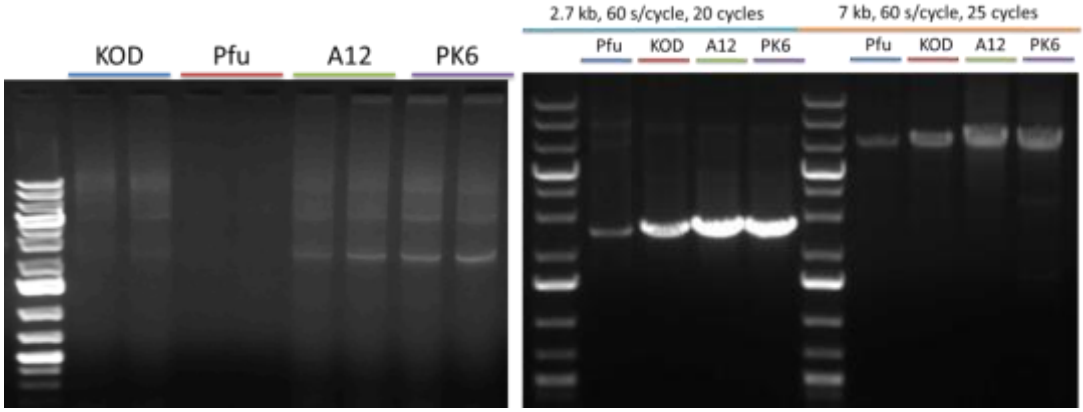
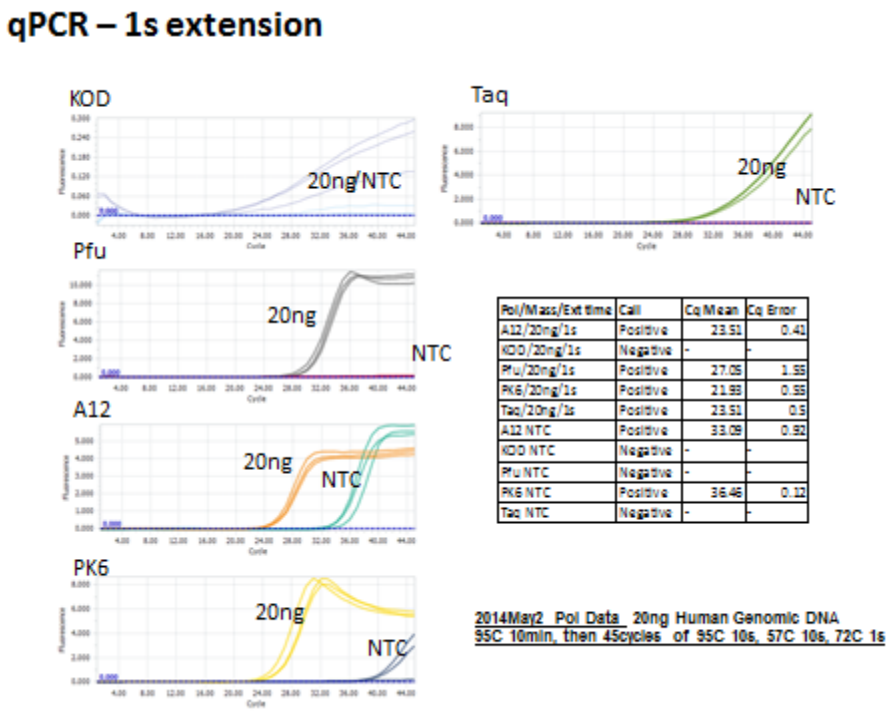



**Figure 5-2.** *Emulsion and Bulk PCR Results*. The left pane shows the initial screening of polymerases after 2 rounds of CSR. On the right, emulsion and bulk PCR results are shown. Evolved variants (A12 and PK6) are compared with their parental enzymes.

**Figure 5-3:** A12 and PK6 qPCR results




| Pol/Mass/Ext time | Call | Cq Mean | Cq Error |
|---|---|---|---|
| A12/20ng/1s | Positive | 23.51 | 0.41 |
| KOD/20ng/1s | Negative | - | - |
| Pfu/20ng/1s | Positive | 27.05 | 1.55 |
| PK6/20ng/1s | Positive | 21.93 | 0.55 |
| Taq/20ng/1s | Positive | 23.51 | 0.5 |
| A12 NTC | Positive | 33.09 | 0.92 |
| KOD NTC | Negative | - | - |
| Pfu NTC | Negative | - | - |
| PK6 NTC | Positive | 36.46 | 0.12 |
| Taq NTC | Negative | - | - |

**Figure 5-3:** *A12 and PK6 qPCR results.* qPCR results using a 1 second extension time for a 100 bp amplicon are shown. Cq values are shown in table. A12 and PK6 were both shown to be faster than either of the parental enzymes or the commonly used Taq polymerase.

**Figure 5-4:** A12 15s per kbp extension

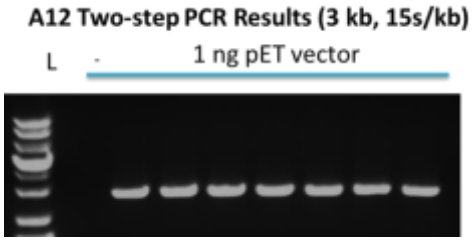



**Figure 5-4:** *A12 15s per kbp extension*. Replicates of a 3 kbp product are shown. A12 could robustly amplify DNA using a much shorter extension time than either of the parental enzymes, Pfu or KOD.

**Figure 5-5:** Sequence alignment and identity heat map of A12 and PK6


| | KOD | PFU | A12 | PK6 |
|---|---|---|---|---|
| KOD | | 79.7% | 94.5% | 94.5% |
| PFU | 79.7% | | 84.6% | 84.9% |
| A12 | 94.5% | 84.6% | | 93.7% |
| PK6 | 94.5% | 84.9% | 93.7% | |

**Figure 5-5:** *Sequence alignment and identity heat map of A12 & PK6*.

**Figure 5-6.** Homology models and calculated Phi-Psi plots of A12 and PK6

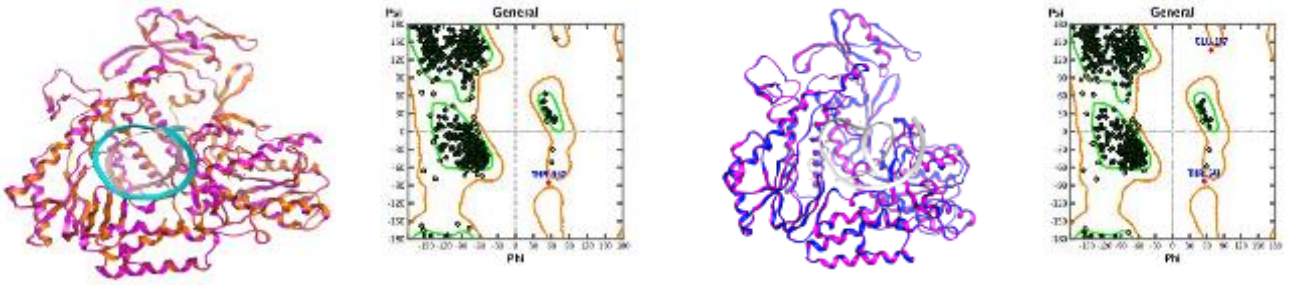



**Figure 5-6.** *Homology models and calculated Phi-Psi plots of A12 and PK6*. A.,C. Best-fit homology models of A12 and PK6, respectively. Final structures are overlaid with KOD, one of two parental enzymes used. B., D. Phi-Psi plots of A12 and PK6. A12 shows only one residue in the outlier region, while PK6 has two.

**Table 5-1:** Selection conditions for shuffled polymerases

| Selection | Rd 1 | Rd 2 | Rd 3 | Rd 4 |
|---|---|---|---|---|
| **Speed** | 1 min/kb | 30 s/kb | 15 s/kb | **7.5 s/kb** |
| **Heme** | 10 µM | 100 µM | 1 mM | **10 mM** |
| **Length** | 1 min/kb | 45 s/kb | 30 s/kb | **20 s/kb** |

**Figure 5-7.** N- and C-terminal heat maps of round 4 extension (speed) library variants

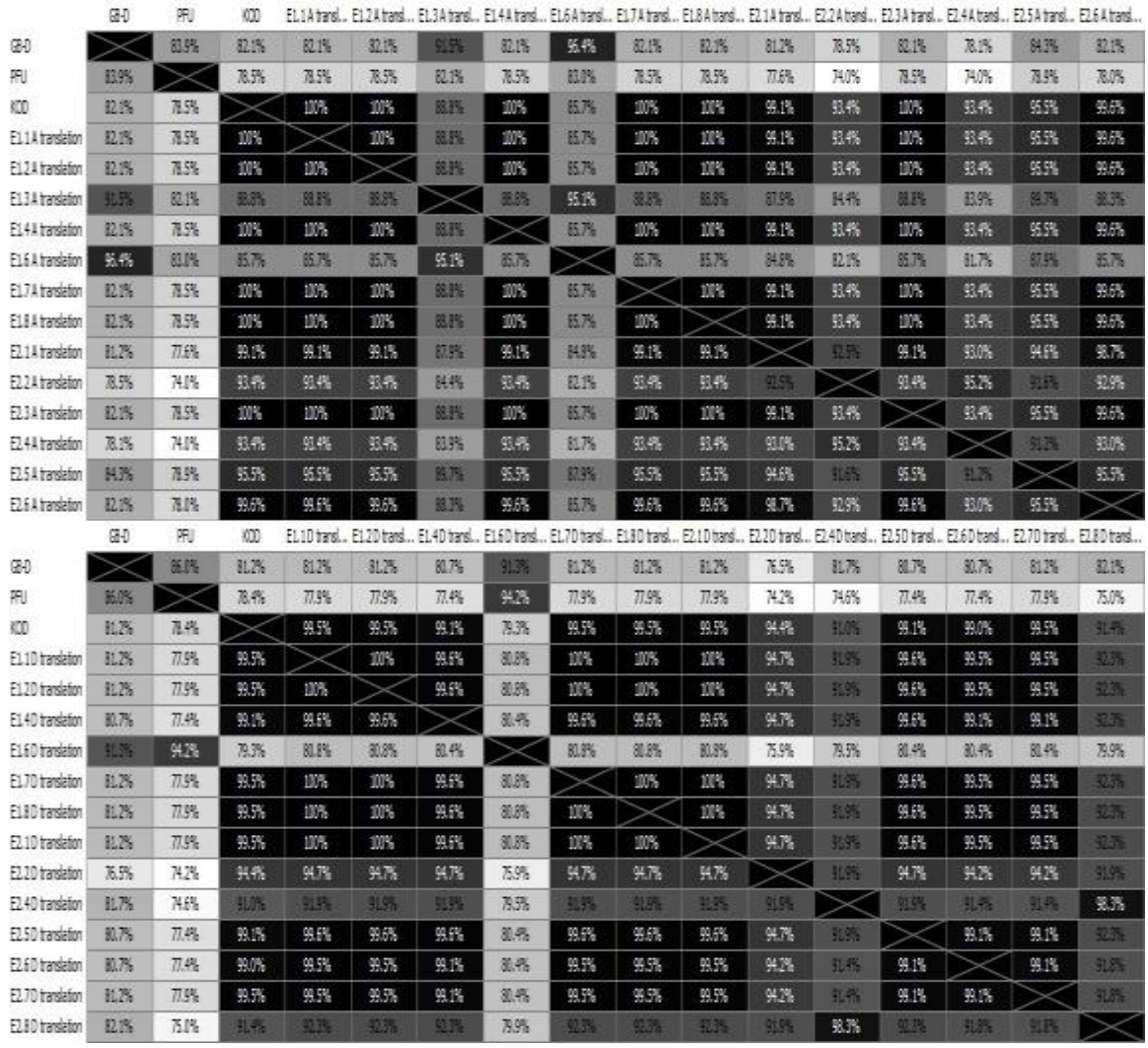



**Figure 5-7:** *N- and C-terminal heat maps of round 4 extension (speed) library variants***.** Heat map shows relationship of sequencing results of the polymerases through round four with the ancestral polymerases.

**Figure 5-9:** N- and C-terminal heat maps of round 4 length (processivity) library variants

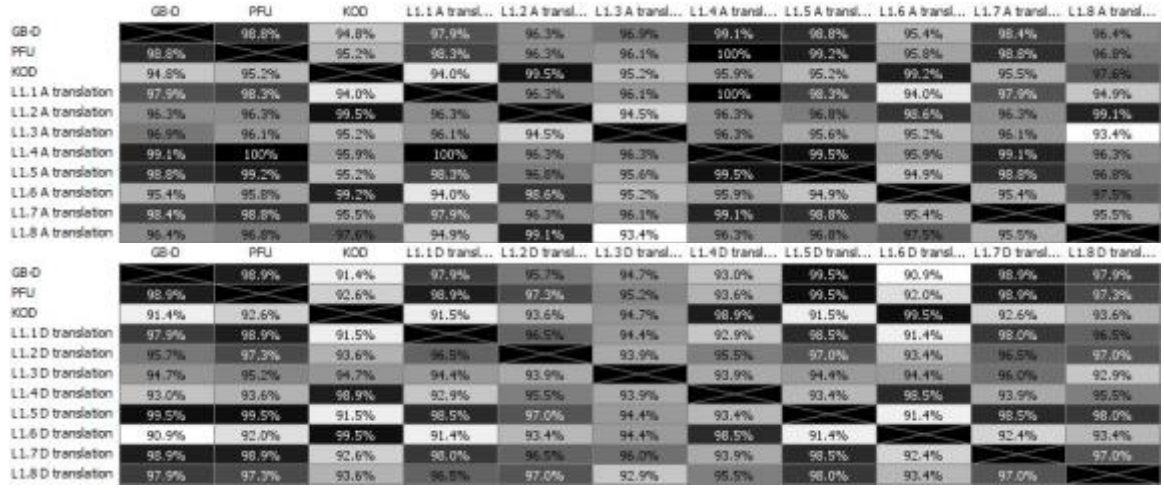



**Figure 5-9:** *N- and C-terminal heat maps of round 4 length (processivity) library variants***.** Heat map shows relationship of sequencing results of the polymerases through round four with the ancestral polymerases.

**Figure 5-10:** N- and C-terminal heat maps of round 4 heme library variants

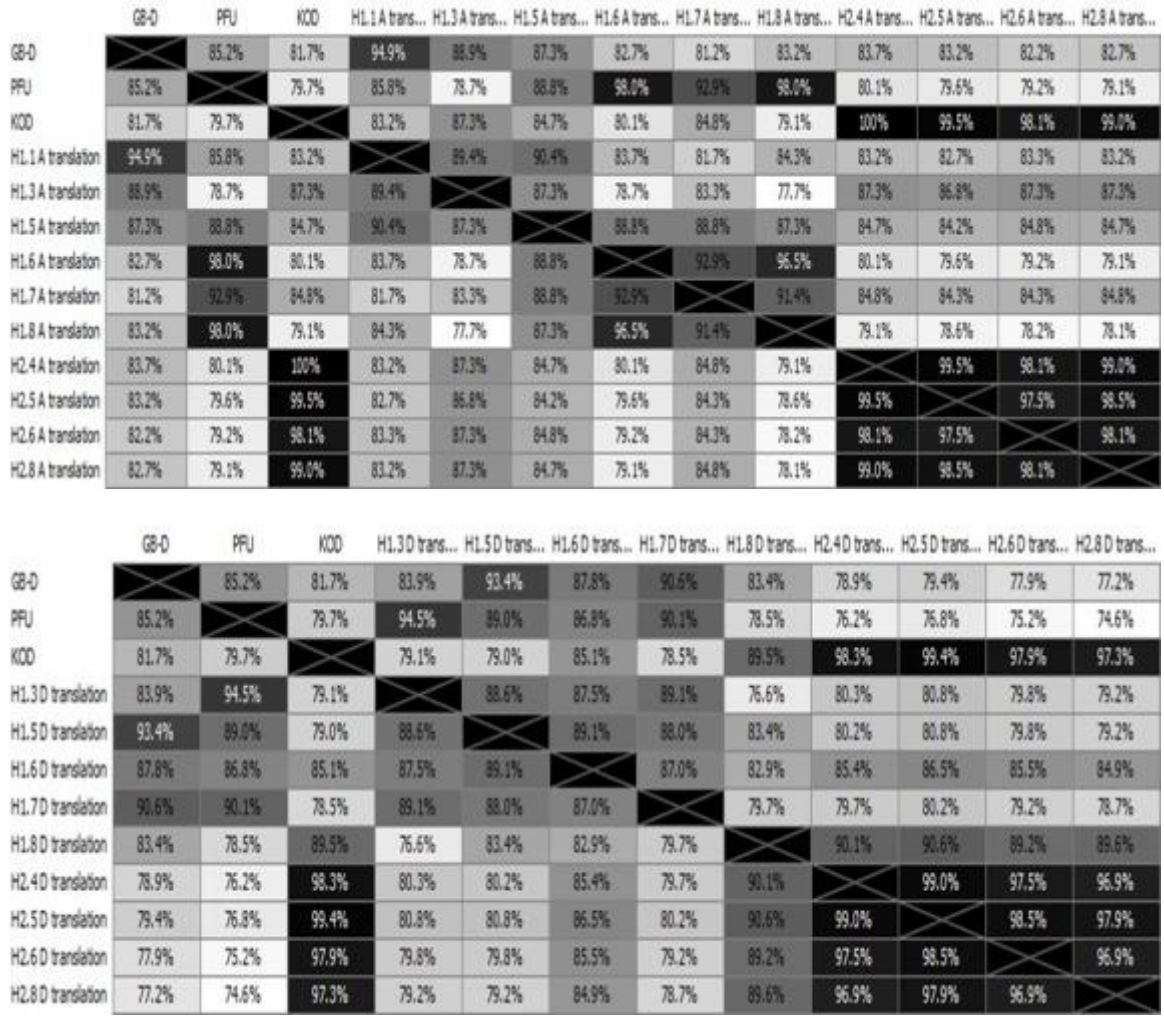



**Figure 5-10:** *N- and C-terminal heat maps of round 4 heme library variants*. Heat map shows relationship of sequencing results of the polymerases through round four with the ancestral polymerases.

**Figure 5-10:** Round 4 recovery of shuffled rounds

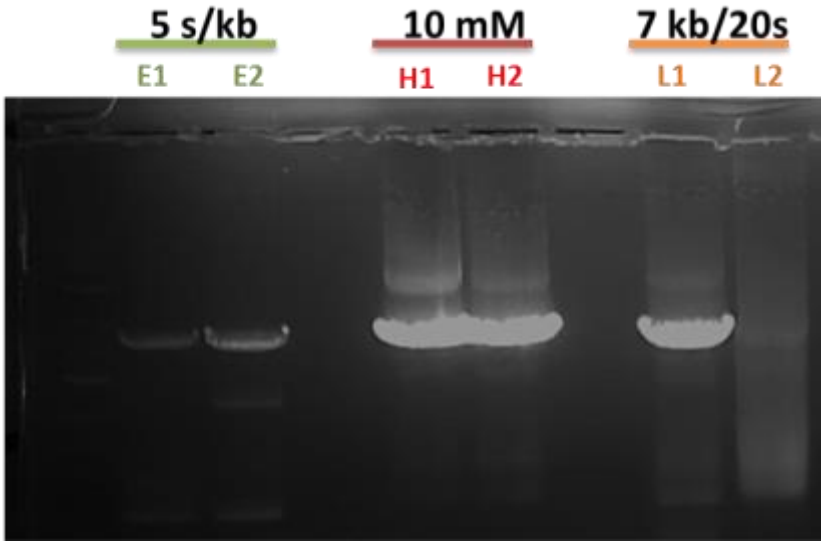



**Figure 5-10:** *Round 4 recovery of shuffled rounds*. After four rounds of selection, each round was recovered and recloned for sequencing analysis.

**Figure 5-11:** Round 4 convergence of processivity library

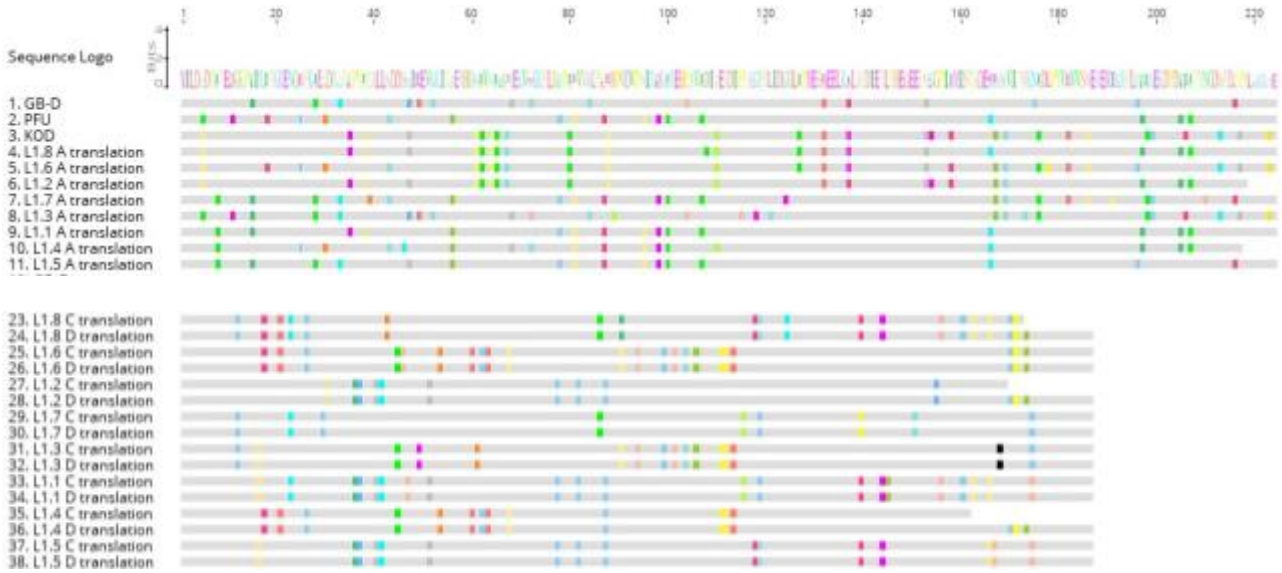



**Figure 5-11:** *Round 4 convergence of processivity library*. Convergence is shown by highlighting differences from the reference sequence. Parental enzymes are shown at the top for reference. A and D in the legends refer to the segment sequenced. A is the amino terminus, and D is the carboxy terminus.

**Figure 5-12.** Change in expression and activity in pET vector

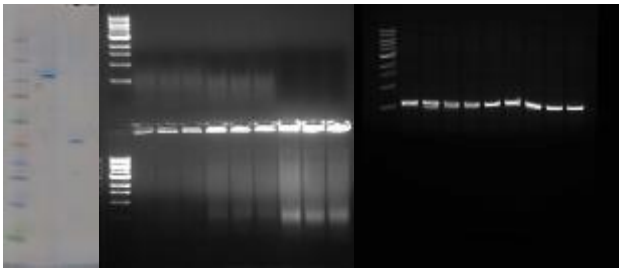
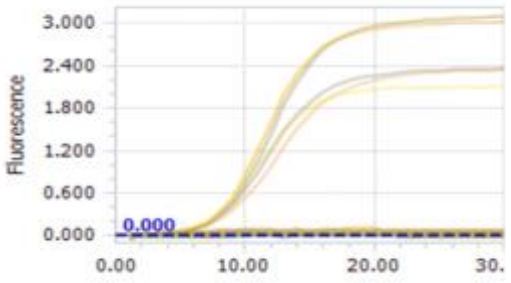


**Figure 5-12.** *Change in expression and activity in pET vector***. Left,** 25 mL cultures of KOD were induced for 4 hrs and purified by Ni-NTA chromatography. Lane 1 is a protein standard. Lane 2 shows production in pET28. Lane 3 is the pAK vector. No visible amount of polymerase could be seen when induced from pAK. **Middle,** varying amounts of cells (0.5-20 μL) were used in a 30 μL reaction using the original pAK selection vector for expression. Amplicons could not be seen in any of the lanes. **Right,** 1 μL of induced cells harboring the pET vector (biological replicates).

**Figure 5-13.** qPCR-based screen proof-of-concept




| Polymerase | Cq Mean | Cq Error |
|---|---|---|
| KOD-3 | 4.93 | 0.12 |
| KOD-2 | 4.9 | 0.03 |
| KOD-1 | 4.28 | 0.61 |
| KOD-1 NTC | - | - |
| KOD 2 NTC | - | - |
| KOD-3 NTC | - | - |

**Figure 5-13.** *qPCR-based screen proof-of-concept*. 1 µL of induced cells harboring the polymerase in a pET vector was used to detect 2.5 ng of an exogenous plasmid. Three independent clones were tested in triplicate (9 total). Cq values were all within error.

**Table 5-2:** Representative CSR reaction mix

| Component | Volume (µL) |
|---|---|
| **Buffer (10X)** | 37.5 |
| **dNTPS (4 mM)** | 18.75 |
| **F (20 µM)** | 7.5 |
| **R (20 µM)** | 7.5 |
| **RNase (0.5 U/µL)** | 1.875 |
| **TMAC (10 mM)** | 3.75 |
| **$H_2O$** | 298.125 |

## 5.2 DIRECTED EVOLUTION OF A THERMOSTABLE, PROOF-READING REVERSE TRANSCRIPTASE

### Introduction

In the previous section, we learned how to perform selections with Compartmentalized Self-Replication. We also developed methods to screen libraries, enabling us to perform multiple directed evolution experiments in parallel (see Section 3). Having these tools in our hands enabled us to pursue more ambitious projects. At this time, Dr. Jared Ellefson came up with the idea to evolve KOD into a thermostable reverse transcriptase using a novel selection platform. The technique employed here, Reverse-Transcription Compartmentalized Self-Replication (RT-CSR), is a modified version of CSR where hybrid primers, containing both RNA and DNA bases (**Figure 5-14**), are used in the CSR reaction. By forcing library variants to reverse transcribe the end of the selection primers, we were able to engineer KOD to function as both a DNA polymerase and a reverse transcriptase.

KOD was chosen as a scaffold for engineering due to its superior functional properties in standard and emulsion PCR reactions. In the previous section, we found that KOD was the only natural Family B polymerase that we tested capable of robustly amplifying DNA in water-in-oil emulsions. Moreover, our shuffled libraries almost always converged on sequences containing more KOD sequence than the other parental enzymes. Further, KOD is one of the most highly processive Family B polymerases known, and displays superior fidelity. Preliminary experiments using KOD and Pfu revealed that neither polymerase was capable of extending a template containing as little as 5 RNA bases. Modeling of these polymerases with DNA:RNA heteroduplexes revealed a vast number of contacts, and we thought it impractical to create focused or targeted libraries. Over 50 contacts were observed between the polymerase and duplex. Instead, we chose to create error-prone libraries of KOD with A12, the error-prone polymerase discovered in Section 1. Dr. Ellefson carried out 18 rounds of RT-CSR utilizing increasing selection pressure (10-176 RNA bases) and found a variant that was capable of robustly reverse-transcribing RNA. Further modeling and engineering led to the production of the first known high-fidelity,

thermostable reverse transcriptase (RTX). Here, we describe the modeling, engineering, and characterization of RTX.

## Results & Discussion

### *Selection of a thermostable reverse transcriptase*

At round 18, variants were challenged with 176 RNA bases during each thermal-cycle. Sequencing of random clones from the survivors of the final revealed a large number of mutations throughout the polymerase. We attributed the increased number of mutations to: (1) variants error-proning themselves during the selection and (2) over 1000 cycles of PCR having been performed throughout the course of the evolution experiment. Using techniques described in Section 1, we found one clone (B11) that was capable of reverse-transcribing 500 bp of RNA. Sequencing of this clone also indicated that the proofreading domain was nonfunctional, supporting our hypothesis that polymerases may be capable of mutating their sequences during RT-CSR. Transplantation of the exonuclease domain from the wildtype sequence restored a small amount proofreading activity (**Figure 5-15**). This led us to believe that a minimal set of mutations may be accessible that would retain reverse transcriptase activity and increase the fidelity of the enzyme by improving its proofreading ability. Thus, we sought to rationally design polymerase variants with a core, minimal set of mutations using deep sequencing and molecular modeling.

### *Deep sequencing and modeling provide molecular insight into RT activity*

In an effort to understand how KOD evolved to use RNA templates, rounds of RT-CSR were deep sequenced (Ellefson et al., 2016). A large number of mutations were identified along the polymerase-duplex interface. However, the most conserved mutations were found at the point where the polymerase encounters template bases. We believe that these mutations were part of molecular checkpoints that the wildtype enzyme uses to enforce strict DNA template utilization. We modeled these interactions using MOE (**Figures 5-16**).

Position R97 was identified near the entry site of the template-polymerase structure. Mutations at this position are known to affect 'uracil scanning' of template strands by Family B polymerases. A specialized pocket near this position has been shown to bind tightly to uracil, which arises from cytosine deamination, and stalls the polymerase until the base can be repaired (Connolly et al., 2003; Fogg et al., 2002). A number of mutations to R97 were observed in the sequencing data (>10%), which modeling predicted to abolish polar contacts between the residue and phosphate backbone of the template. This interaction is presumably used to 'flip' the template base into the checkpoint site.

Next, we rationalized the most conserved mutation from our RT-CSR evolution experiment, Y384H. This mutation appears to disrupt a hydrogen bonding network between the 2' hydroxyl of the RNA template and residues Y384 and Y494. The next most prevalent mutations (E664K, G711V, and E735K) appear to enhance binding to homo- and heteroduplexes in both A- and B-form duplex conformations. Our modeling is surprisingly supported by a previous study in which E664K was shown to greatly increase binding to RNA:DNA heteroduplexes (Cozens et al., 2012).

We then modeled a number of other conserved mutations along the polymerase-template interface (**Figure 5-17**). K118I appears to break a potential hydrogen bond on a RNA template, which may be involved in the checkpoint disrupted by R97M. Y493L is poised to allow for packing within the hydrophobic core adjacent to the heteroduplex. This movement may be necessary for binding to both A- and B-form duplexes. V389I adds a methyl group and pushes the neighboring beta sheet towards the heteroduplex to better orient the hydrogen-bonding network of T541 and D540 to both primer-template duplexes. G711V pushes two α-helices away from each other in the thumb domain, likely allowing for geometry differences in DNA:DNA and DNA:RNA duplexes (B- and A-form, respectively). N735K displays stronger binding by ~8 kcal/mol to the backbone of the template strand by making contact to the phosphate backbone.

Using our sequencing and modeling as guideposts, we then rationally designed four potential variants and tested their ability to perform reverse transcription and high-fidelity PCR. All four of the variants were capable of high-fidelity PCR. However, only two were shown to be robust in reverse transcription of a 1 kbp amplicon (**Figure 5-18**). The most promising variant, RTX, had

less than half of the mutations of the selected polymerase B11 without sacrificing catalytic efficiency or the Michaelis constant in single-nucleotide kinetics experiment (**Figure 5-19**). We next determined if RTX was capable of single-enzyme RT-PCR, where RTX performs reverse transcription and PCR in the same reaction. Using several RNA templates from both bacterial and human samples, RTX was capable of performing single-enzyme RT-PCR on RNAs over 5 kbp, potentially highlighting its processivity (**Figure 5-20**).

Modeling of RTX showed a preference for DNA templates to RNA templates (contact energies of -221.7 kcal/mol vs -204.8 kcal/mol). For KOD, our modeling predicted much less binding to both DNA and RNA templates, -182.62 kcal/mol and -167.28 kcal/mol, respectively (**Figure 5-21**). *In vitro* characterization of the binding of these enzymes to homo- and heteroduplexes was in excellent agreement with these predictions. RTX was shown to have a 2.3 ± 0.4 nM $K_d$ with DNA duplexes and a 46.0 ± 8.2 nM $K_d$ with DNA:RNA heteroduplexes. KOD displayed similar binding to both duplexes (25.5 ± 4.7 nM vs 33.0 ± 7.8 nM) (**Table 5-3**). These results gave us confidence that our modeling predictions were valid.

***RTX can proofread on DNA and RNA templates***

While our high-fidelity PCR results showed that RTX still had the ability to proofread mismatches, we were unsure if the enzyme would be able to proofread mismatches on a DNA-primed RNA template due to the A-form helical nature of heteroduplexes (Wang et al., 1982). Using primer extension analysis using matched and 3'-dideoxy mismatched primers, RTX was shown to elongate on mismatches using both RNA and DNA as templates. As expected, KOD was only able to extend on DNA templates. Interestingly though, KOD was shown to excise several bases in the primer when RNA was used as a template, presumably triggering proofreading due to the unique conformation of the heteroduplex. As expected, exonuclease deficient versions of RTX and KOD were unable to excise mismatches and extend the primer (**Figure 5-22**). To our knowledge, this marked the first demonstration of proofreading activity by a reverse transcriptase. Subsequent experiments revealed that RTX was capable of high-fidelity reverse transcription in a

novel fidelity assay. Interestingly, RTX was found to have a slightly altered mutational bias. This may be a result of RTX's ability to use DNA and RNA templates, which when primed with DNA are geometrically and structurally distinct.

### *Conclusions*

Using KOD as a scaffold, we were able to engineer the first known thermostable, high-fidelity reverse transcriptase. We believe that the mutations discovered here unlocked molecular checkpoints that the wildtype enzyme used to discriminate between templates. Throughout the course of RT-CSR, polymerases accumulated mutations that not only gave rise to RNA template utilization, but also mutations that inactivated the proofreading domain. This is a common feature of CSR type reactions because inactivating the proofreading domain typically increases the speed of polymerization. Our initial attempts to engineer proofreading into B11 were met with modest results, giving credence to the notion that other deleterious mutations were present—more than likely due to variants increasing mutational load throughout rounds of selection. By combining deep sequencing analysis of individual rounds and molecular modeling, we were able to rationally design variants with less than half of the mutations present in B11, the original reverse transcriptase found during our screening process. RTX was found to proofread on DNA and RNA templates, which was unexpected due to RNA:DNA heteroduplex geometry. This increase in fidelity should enable higher precision RNA-seq experiments in the future.

## Materials and Methods

### *Initial reverse transcription test for polymerases*

30 pmol of 5' fluorescein labeled primer (25FAM) were annealed with 30 pmol of template (TEMP.A.DNA/1RNA/5RNA) and 0.4 μg of polymerase by heat denaturation at 90°C for 1 minute and allowing to cool to room temperature. Reactions were initiated by the addition of "start" mix which contained (50mM Tris-HCl (pH8.4), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$ and

200 μM dNTPs. MMLV polymerase was treated according to manufacturer's recommendations (New England Biolabs). Reactions were incubated for 2 minutes at 68°C until terminated by the addition of EDTA to a final concentration of 25 mM. The labeled primer was removed from the template strand by heating sample at 75°C for 5 minutes in 1x dye (47.5% formamide, 0.01% SDS) and 1 nmol of unlabeled BLOCKER oligonucleotide (to competitively bind the template strand). Samples were run on a 20% (7 M urea) acrylamide gel.

Reverse Transcription CSR (RT-CSR)

***Library generation and RT-CSR***

KOD polymerase libraries were created through error prone PCR to have a mutation rate of ~1-2 amino acid mutations per gene. Libraries were cloned into tetracycline inducible vector and electroporated into DH10B *E. coli*. Library sizes were maintained with a transformation efficiency of at least $10^6$, but more typically $10^7$-$10^8$. Overnight library cultures were seeded at a 1:20 ratio into fresh 2xYT media supplemented with 100 μg / mL ampicillin and grown for 1 hour at 37°C. Cells were subsequently induced by the addition of anhydrotetracycline (typically at a final concentration of 200 ng / mL) and incubated at 37°C for 4 hours. Induced cells (200 μL total) were spun in a tabletop centrifuge at 3,000 x g for 8 minutes. The supernatant was discarded and the cell pellet was resuspended in RTCSR mix: 1x Selection buffer (50 mM Tris-HCl (pH8.4), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$), 260 μM dNTPs, 530 nM forward and reverse RNA containing primers. The resuspended cells were placed into a 2 mL tube with a 1mL rubber syringe plunger and 600 μL of oil mix (73% Tegosoft DEC, 7% AbilWE09 (Evonik), and 20% mineral oil (Sigma-Aldrich)). The emulsion was created by placing the cell and oil mix on a TissueLyser LT (Qiagen) with a program of 42Hz for 4 minutes. The emulsified cells were thermal-cycled with the program: 95°C - 3min, 20x (95°C- 30 sec, 62°C- 30 sec, 68°C- 2 min). Emulsions

were broken by spinning the reaction (10,000 x g - 5 min), removing the top oil phase, adding 150 µL of $H_2O$ and 750 µL chloroform, vortexing vigorously, and finally phase separating in a phase lock tube (5Prime). The aqueous phase was cleaned using a PCR purification column which results in purified DNA, including PCR products as well as plasmid DNA. Subamplification with corresponding outnested recovery primers ensures that only polymerases that reverse transcribed are PCR amplified. Typically this is achieved by addition of 1/10 the total purified emulsion using Accuprime Pfx (ThermoFisher) in a 20 cycle PCR, however challenging rounds of selection could require increasing the input DNA or cycle number to achieve desired amplification.

***Molecular Modeling of RT-CSR Mutations***

The wild-type structures of the family B polymerases from *Thermococcus kodakaraensis* (KOD) (PDB: 4K8Z) and *Thermococcus gorgonarius* (PDB: 2VWJ) were prepared for mutational analyses using the Molecular Operating Environment (MOE.10.2015) software package from Chemical Computing Group. The structures were inspected for anomalies and protonated/charged with the Protonate3D subroutine (341K, pH 8.0, 0.1 M salt). The protonated structures were then lightly tethered to reduce significant deviation from the empirically determined coordinates and minimized using the Amber12:EHT forcefield with Born solvation model to an RMS gradient of 0.1 kcal $mol^{-1}$ $Å^{-1}$. These structures were then used as templates to build homology models of RT-CSR mutations. Homology models of the variants were prepared by creating 25 main chain models with 25 sidechain samples at 341K (625 total) within MOE. Intermediates were refined to an RMS gradient of 1 kcal $mol^{-1}$ $Å^{-1}$, scored with the GB/VI methodology, minimized again to an RMS gradient of 0.5 kcal $mol^{-1}$ $Å^{-1}$, and protonated. The final model for each variant was further refined by placing the protein within a 6 Å water sphere and minimizing the solvent enclosed structure to an RMS gradient of 0.001 kcal $mol^{-1}$ $Å^{-1}$. Models were evaluated by calculating Phi-Psi angles and superimposed against the reference structures. RNA:DNA A-form duplexes were created and superimposed against the empirically derived coordinates for the DNA:DNA duplex. Models were then minimized and contact energies measured within MOE.

***Cloning and purification of polymerase variants***

*Escherichia coli* DH10B and BL21 (DE3) strains were used for cloning and expression, respectively. Strains were maintained on either Superior or 2XYT growth media. Polymerases were cloned into a modified pET21 vector using NdeI and BamHI sites. Overnight cultures of BL21

(DE3) harboring each of the variants were grown overnight in Superior broth at 37°C. Cells were then diluted 1:250, and protein production was induced with 1 mM IPTG during mid- log at 18°C for 20 hrs. Harvested cells were flash-frozen and lysed by sonication in 10 mM phosphate, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol, pH 7 (Buffer A). Cleared cell lysates were heated at 85˚C for 25 min, cooled on ice for 20 minutes, and filtered (0.2 μm). The filtrate was then passed over a DEAE column, immediately applied to an equilibrated heparin column, and eluted along a sodium chloride gradient. Polymerase fractions were collected and dialyzed into Buffer A. Enzymes were further purified using an SP column and again eluted along a salt gradient. Pooled fractions were then applied to a Sephadex 16/60 size exclusion column (GE Healthcare), concentrated, and dialyzed into storage buffer (50 mM Tris- HCl, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.1% Non-idet P40, 0.1% Tween20, 50% glycerol, pH 8.0). Working stocks were made at 0.2 mg/mL.

***PCR Proofreading Assay***

50 μL PCR reactions were set up with a final concentration of 1x Assay Buffer (60 mM Tris-HCl (pH8.4), 25 mM $(NH_4)_2SO_4$, 10 mM KCl), 200 μM dNTPs, 2 mM $MgSO_4$, 400 nM (PCRTest.F/PCRTest.R) or (PCRTest.DiDe.F/PCRTest.DiDe.R) forward and reverse primers, 20 ng of template plasmid and 0.2 μg polymerase. Reactions were thermal-cycled using the following program: 95˚C - 1min, 25x (95˚C- 30 sec, 55˚C- 30 sec, 68˚C- 2 min 30 sec).

***Primer Extension Assay***

10 pmol of 5' fluorescein labeled primer (RT.Probe or RT.Probe.3ddc) were annealed with 50 pmol of template RNA or DNA (RT.RNA.TEMP and RT.DNA.TEMP, respectively) and 0.4 μg of polymerase by heat denaturation at 80˚C for 1 minute and allowing to cool to room temperature. Reactions were initiated by the addition of "start" mix which contained (1x Assay Buffer, 2 mM

$MgSO_4$ and 200 μM dNTPs. Reactions were incubated for 10 minutes at 68°C until terminated by the addition of EDTA to a final concentration of 25 mM. The labeled primer was removed from the template strand by heating sample at 75°C for 5 minutes in 1x dye (47.5% formamide, 0.01% SDS) and 1 nmol of unlabeled RT.bigBlocker oligonucleotide (to competitively bind the template strand). Samples were run on a 20% (7 M urea) acrylamide gel.

***RT-PCR Assay***

50 μL reverse transcription PCR (RTPCR) reactions were set up on ice with the following reaction conditions: 1x Assay Buffer, 1 mM $MgSO_4$, 1 M Betaine (Sigma-Aldrich), 200 μM dNTPs, 400 nM reverse primer, 400 nM forward primer, 40 units RNasin Plus (Promega), 0.2 μg polymerase and 1 μg of Total RNA from Jurkat, Human Spleen or *E. coli* (Ambion). Primer sets used: PolR2A (PolII.R, PolII.F1/F2/F4), p532 (p532.R, p532.F1/F2/F5), rpoC (rpoC.R, rpoC.F1/F2/F4). Reactions were thermal-cycled according to the following parameters: 68°C - 30 min, 25x (95°C- 30 sec, 68°C (63°C for rpoC) - 30 sec, 68°C - 30 s/kb).

***Single nucleotide incorporation kinetics***

Duplexes (DNA:DNA or DNA:RNA) were assembled by combining equimolar amounts of a DNA 25-mer (5'-CCCTCGCAGCCGTCCAACCAACTCA-3') and DNA or RNA 36-mer (3'-GGGAGCGTCGGCAGGTTGGTTGAGTGCCTCTTGTTT-5') in 10 mM Tris-HCl, 0.1 mM EDTA (pH 8.0). Solutions were heated to 95°C for 5 min, slowly cooled to 60°C for 10 min, and then cooled to room temperature for 15 minutes. Reactions (100 μL) consisting of assay buffer, 1 mM MgSO4, and 500 nM duplex were initiated by variable amounts of α-P32-dCTP (0.003-400μM), which was diluted 1:400 in unlabeled dCTP. Reactions were allowed to proceed 3-14

minutes. 10 μL aliquots were quenched by the addition of EDTA (0.25 M final concentration) in 15-120s intervals. Aliquots (2 μL) were spotted on DE81 filter paper and washed 6 times in 5% NaH2PO4 (pH 7), 2 times in ddH2O and finally in 95% EtOH. Dried filter paper was exposed for 24 hrs and imaged on a STORM scanner. Initial rates were obtained by analysis using FiJi (Image J). Kinetic parameters were determined by non-linear regression using SigmaPlot10.

**Figure 5-14:** *RT-CSR Scheme*

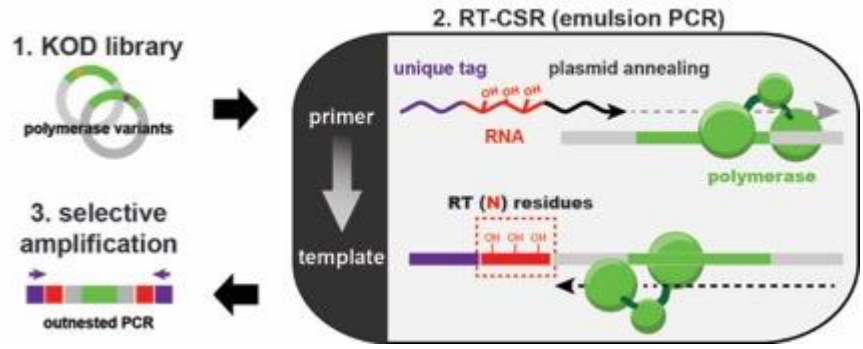



**Figure 5-14:** *RT-CSR Scheme***.** Directed evolution of thermostable reverse transcriptase using reverse transcription compartmentalized self replication (RT-CSR). Libraries of polymerase variants are created, expressed in *E. coli*, and *in vitro* compartmentalized. During emulsion PCR, primers flanking the polymerase enable self-replication, but are designed with a variable number of RNA bases separating the plasmid annealing portion from the unique recovery tag. Outnested PCR ensures that only polymerases with reverse transcriptase activity are selective amplified.

**Figure 5-15:** Characterization of B11

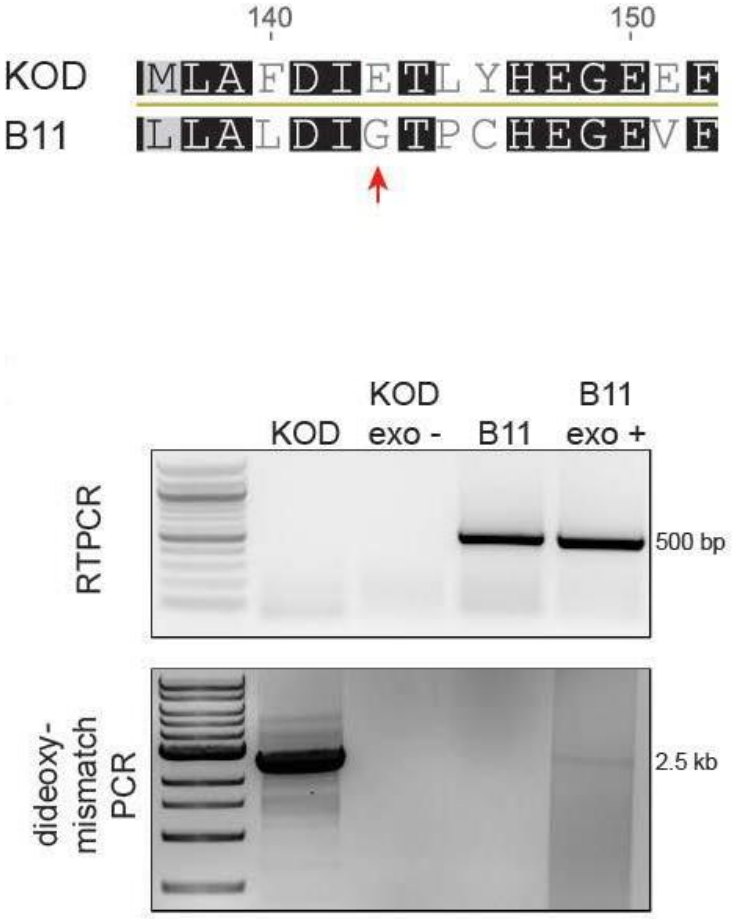



**Figure 5-15:** *Characterization of B11*. **a**, Mutations in the B11 polymerase (yellow) are mapped onto the KOD polymerase (grey with DNA primer:template duplex in blue). Thirty seven mutations were accumulated, many found in the proofreading domain. **b**, Examination of the 3'-5' exonuclease active site shows a mutation at glutamate 143 to glycine. **c**, Functional assays reveal B11 polymerase is capable of single enzyme RTPCR of a 500 base pair region of the *HSPCB* gene, as well as the B11 with grafted wildtype proofreading domain. Proofreading activity was qualitatively measured in a dideoxy-mismatch PCR, which requires removal of a 3' deoxy mismatch primer before polymerization occurs.

**Figure 5-16:** Molecular checkpoints involved in template recognition

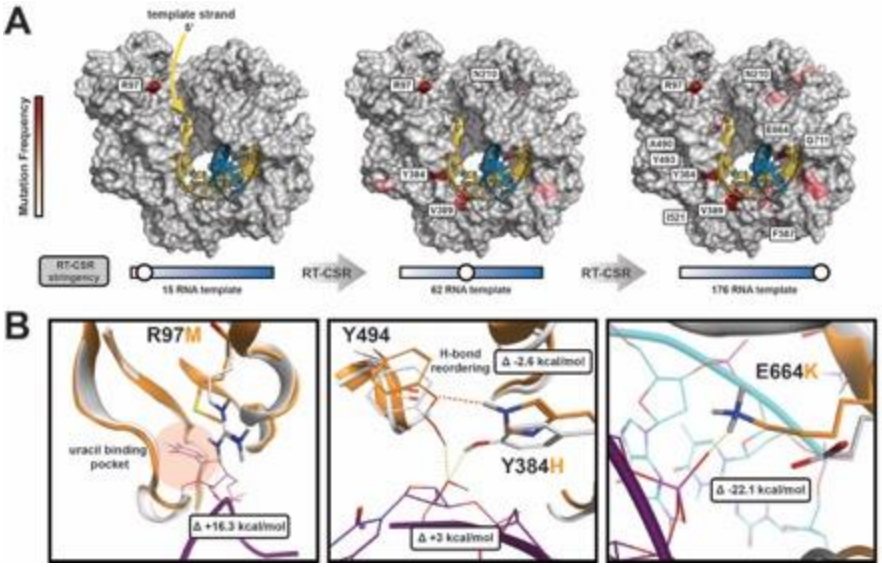



**Figure 5-16:** *Molecular checkpoints involved in template recognition***.** (**A**) Structural heat map of mutated residues over the RT-CSR process found by deep sequencing. Conserved mutations are colored incrementally darker shades of red to indicate frequency in the polymerase pool. Amino acid residues that were mutated in over 50% of the population were labeled. Figure was adapted from KOD structure PDB 4K8Z. (**B**) Computer modeling of KOD (grey) and RTX mutations (orange) at checkpoints responsible for DNA and RNA template recognition at R97, Y384, and E664. Free energy changes between wild-type KOD and RTX mutations are inlet displayed.

**Figure 5-17:** Modeling of conserved mutations along polymerase-template duplex

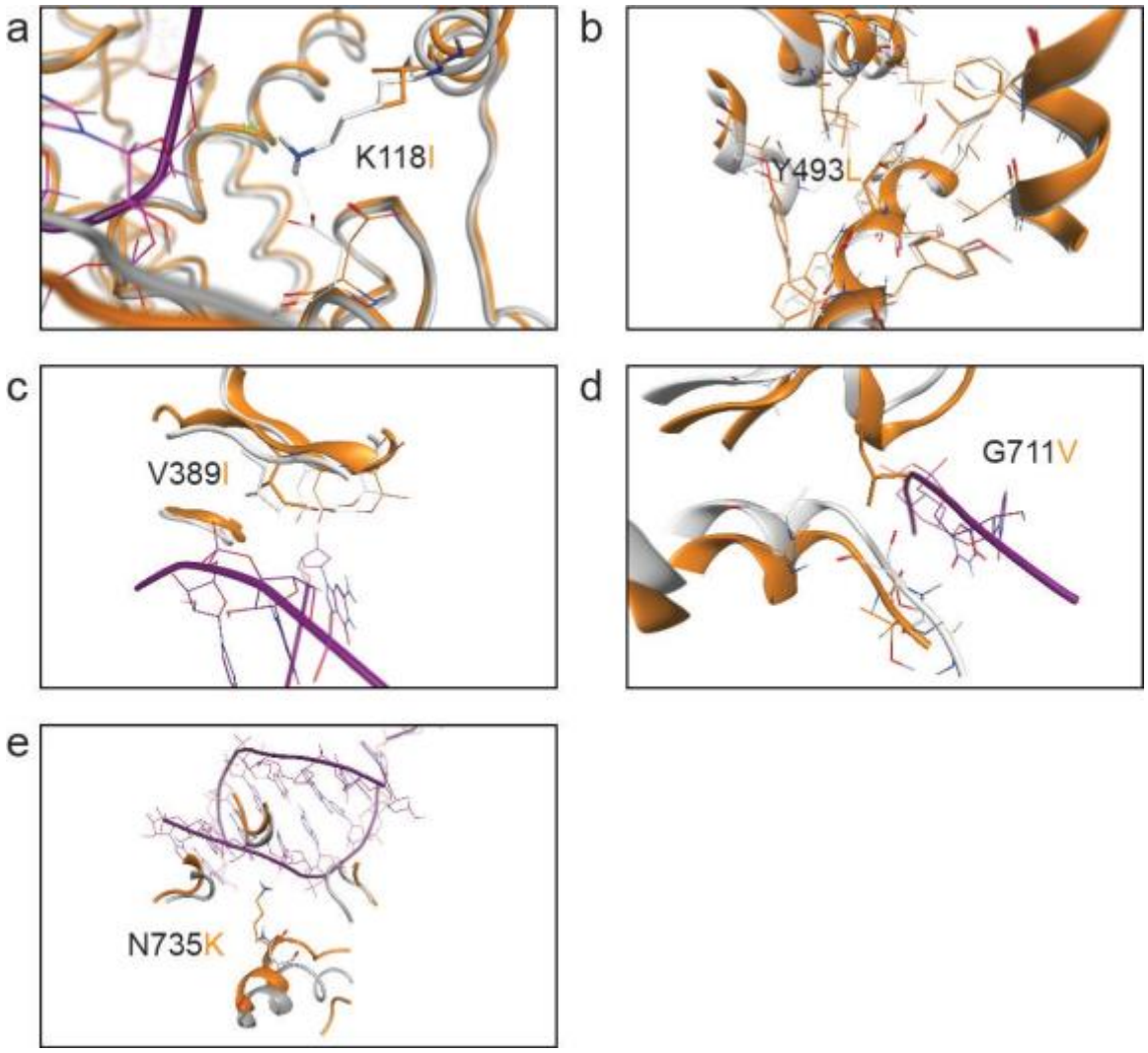
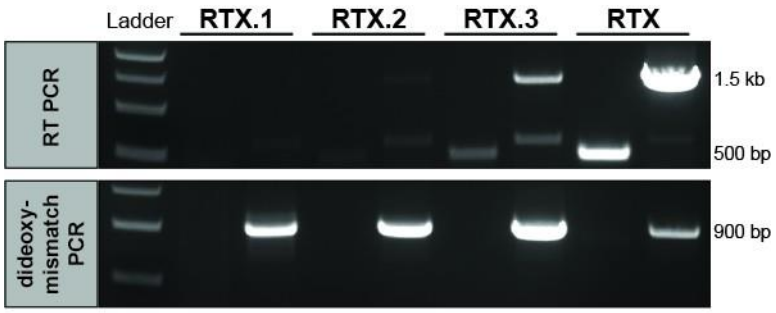



**Figure 5-17:** *Modeling of conserved mutations along polymerase-template duplex*. **a**, K118I breaks a potential hydrogen bond on a RNA template, which may be involved in the checkpoint disrupted by R97M. **b**. Y493L is poised to allow for packing within the hydrophobic core adjacent to the heteroduplex. This movement may be necessary for binding to both A- and B-form duplexes. **c**, V389I adds a methyl group and pushes the neighboring beta sheet towards the heteroduplex to better orient the hydrogen-bonding network of T541 and D540 to both primer-template duplexes. **d**, G711V pushes two α-helices away from each other in the thumb domain, likely allowing for geometry differences in DNA:DNA and DNA:RNA duplexes. **e**, N735K displays stronger binding by ~8 kcal/mol to the backbone of the template strand by making contact to the phosphate backbone. RTX mutations are shown in orange, and KOD wild- type residues are depicted in white. Models shown are in complex with a DNA:RNA heteroduplex.

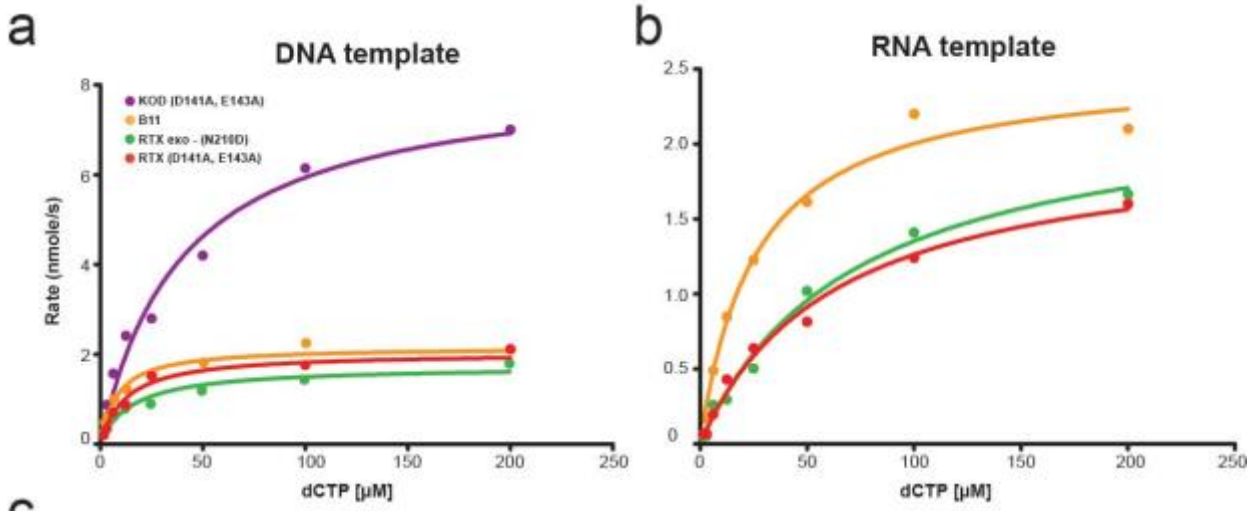
**Figure 5-18:** Sequence- and model-driven designs of RT variants




| Polymerase Variant | Mutations | Total Mutations |
|---|---|---|
| RTX.1 | R97M, Y384H, V389I, Y493L, F587L, E664K, G711V, W768R | 8 |
| RTX.2 | F38L, R97M, K118I, R381H, Y384H, V389I, Y493L, T514I, F587L, E664K, G711V, W768R | 12 |
| RTX.3 | F38L, R97M, K118I, M137L, R381H, Y384H, V389I, K466R, Y493L, T514I, F587L, E664K, G711V, W768R | 14 |
| RTX | F38L, R97M, K118I, M137L, R381H, Y384H, V389I, K466R, Y493L, T514I, I521L, F587L, E664K, G711V, N735K, W768R | 16 |

**Figure 5-18:** *Sequence- and model-driven designs of RT variants***.** The designed polymerases based on deep sequencing and modeling information were constructed and tested in an RTPCR assay (HSPCB loci) and proofreading assay. The mutations introduced into the wild-type KOD polymerase are shown for each of the designed reverse transcriptases.

**Figure 5-19.** Steady-state kinetics of polymerase variants




c

| | DNA | | | RNA | | |
|---|---|---|---|---|---|---|
| Enzyme | *k* cat | Km | *k* cat/Km | *k* cat | Km | *k* cat/Km |
| KOD | 160.4 | 39.7 | 4.0 | nd | nd | nd |
| B11 | 48.2 | 7.9 | 6.1 | 54.5 | 23.1 | 2.4 |
| RTX | 49.1 | 15.7 | 3.1 | 47.8 | 62.7 | 0.8 |
| RTX exo- | 56.8 | 12.1 | 4.7 | 55.9 | 71.5 | 0.8 |

**Figure 5-19.** *Steady-state kinetics of polymerase variants*. Initial rates of single nucleotide (dCTP) incorporation by exonuclease deficient polymerases were plotted against the concentration of dCTP using DNA (**a**) or RNA templates (**b**). **c**, Kinetic parameters were estimated by fitting the data to the Michaelis-Menten equation. Nucleotide addition could not be determined (nd) for KOD on RNA templates, due to low activity.

**Figure 5-20:** RTX can perform single-enzyme RT-PCR reactions

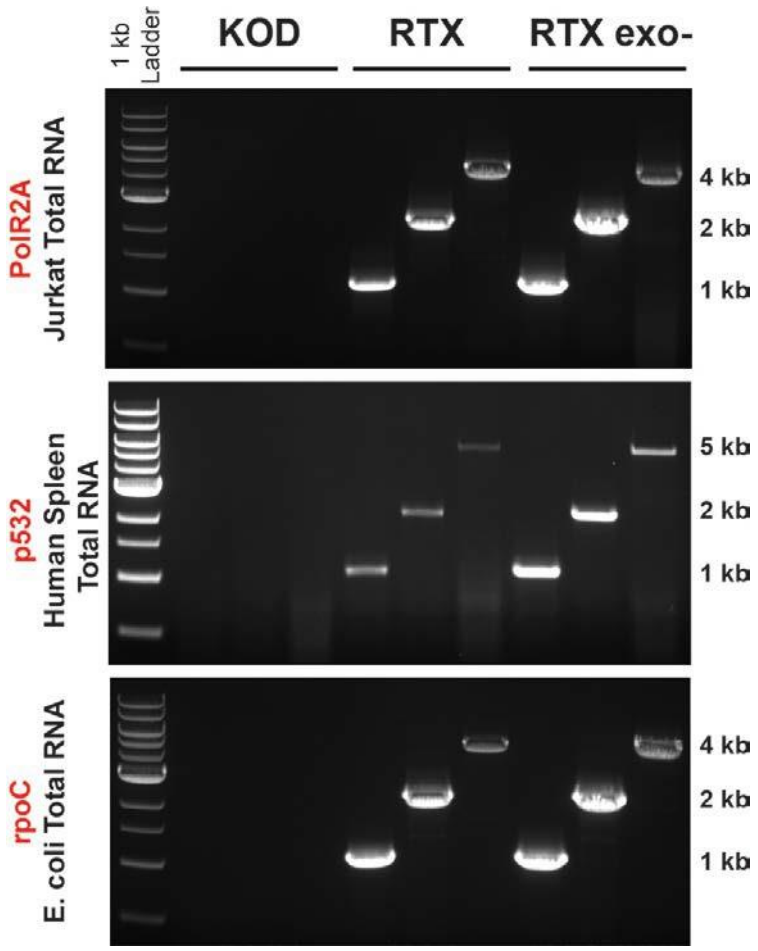



**Figure 5-20:** *RTX can perform single-enzyme RT-PCR reactions*. Reverse transcription PCR (RTPCR) was performed using KOD polymerase, RTX, and the proofreading deficient version of RTX (N210D; exo-). Various genes were amplified (red), two human genes, PolR2A and p532, and rpoC from *E. coli* from various RNA sources. Using gene specific forward and reverse primers, gene regions were amplified, demonstrating efficient single enzyme RTPCR.

**Figure 5-21:** RTX mutations are predicted to prefer DNA templates

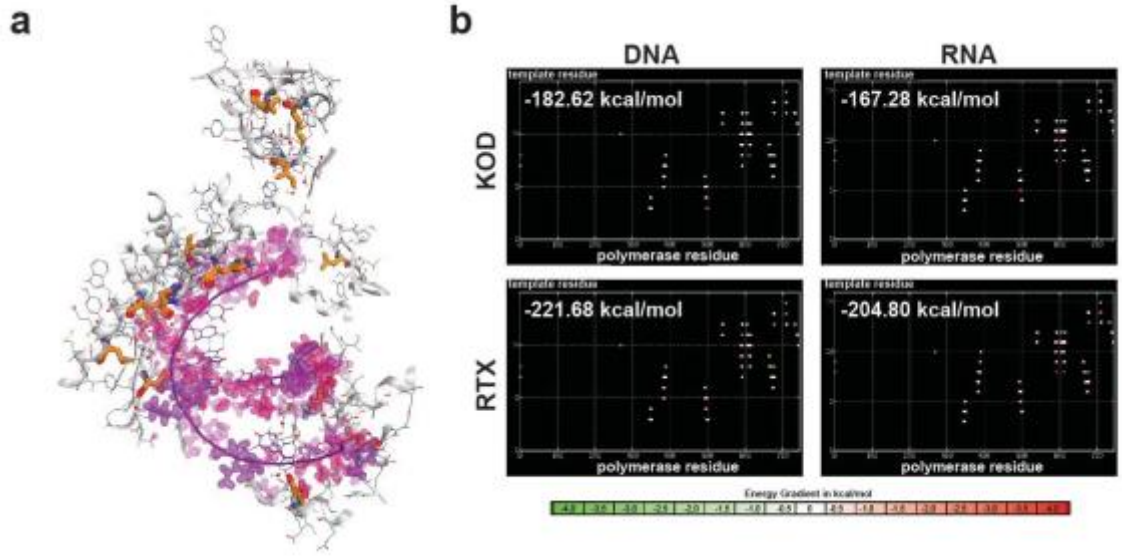



**Figure 5-21:** *RTX mutations are predicted to prefer DNA templates*. **a**, Positions of RTX mutations are depicted in orange. Surrounding residues are shown in gray stick format. Contact analysis was performed within MOE by setting primer-template duplexes as the ligand. Potential contact sites are shown in purple. **b**, Contacts were measured on A-form heteroduplexes and B-form DNA duplexes. Positions making contact are shown in along the x-axis relative to the position of the duplexes along the y-axis. Individual contributions are depicted by an energy gradient. Overall contact energies are displayed within each frame. Consistent with our kinetic data, RT-CSR mutations were found to exhibit better binding to both duplexes, with a small preference for DNA duplexes.

**Figure 5-22.** RTX polymerase proofreads during reverse transcription

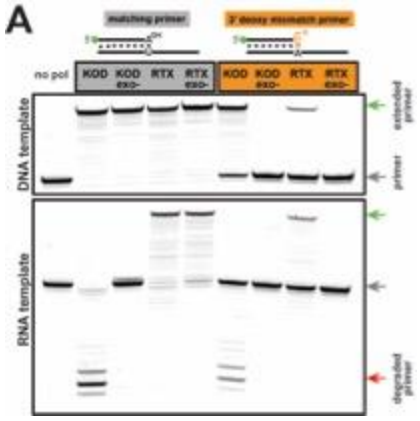



**Figure 5-22.** *RTX polymerase proofreads during reverse transcription*. (**A**) Primer extension reactions of KOD and RTX polymerases and their proofreading deficient counterparts (exo-), on both DNA and RNA templates. Extension reactions were performed with matched 3' primer:templates (grey) or a 3' deoxy mismatch (orange), which must be excised before extension can proceed. The primer is denoted by a gray arrow, extended product in green, and exonuclease degraded primer in red.

## 5.3 METHODS FOR THE SELECTION OF HIGH-FIDELITY AND INHIBITOR-RESISTANT POLYMERASES

### Introduction

Our previous CSR evolution experiments taught us that polymerase variants that survived selection would often lack 3'-5' exonuclease activity and become error-prone. Sequencing of individual clones would often lead us to find premature stop codons in polymerase sequences, which we attributed to errors generated by the polymerases themselves. Of course, the polymerases that survived the selection could not have too high of an error-rate. Error catastrophe would prevent those polymerases from dominating the population over prolonged selections. However, error-rates observed in variants isolated from our selections were several orders of magnitude higher than the parental enzymes.

We wanted our polymerases to retain high fidelity amplification during the course of the selection. However, the solution was not immediately obvious. Early ideas involved putting mismatches in primers over restriction sites. In this scheme, if the incorrect base was not cleaved, then we could use restriction enzymes to eliminate variants that did not make the correction before self-amplification. This seemed possible but messy and difficult to enforce. After several months of thinking about this problem, Dr. Jared Ellefson suggested that we add 3'-dideoxy mismatches to the CSR primers. This would force polymerases to recognize the mismatch and cleave the incorrect base before continuing polymerization during the PCR reaction. Using a di-deoxy base at the

end of the primer ensures that polymerization cannot occur, which is the reason these bases are often used in chain termination reactions.

At this time, we also became interested in producing polymerases for specific diagnostic or forensic applications. DNA sample preparation often included the use of chaotropes (Nagy, 2007; Wirtz, 2014; Wolfe et al., 2002), and so we wanted to understand how KOD would behave when challenged with these molecules. Inhibitors found within biological samples were also of concern. Heme is the principle inhibitor found in blood (Akane et al., 1994). Melanin, a pigment found in human hair and skin, has also been shown to inhibit polymerases (Eckhart et al., 2000). Lipoteichoic acids are found in Gram positive cell walls and are thought to be the chief reason that 'colony PCR' does not work these bacterial species. Many common nosocomial infections arise from Gram positive bacteria, and their diagnosis is slow due to the need to culture these organisms (Khan et al., 2015; Obritsch et al., 2005). The engineering of polymerases that would be resistant to chaotropes and these inhibitors would be of great utility. The ability to amplify nucleic acids with high fidelity enzymes with little to no sample preparation would ensure accurate diagnostic and forensic analyses. Below we describe our work in initiating selections to create high-fidelity, inhibitor-resistant polymerases.

## Results & Discussion

### *High-fidelity CSR*

Polymerases that do not contain active proofreading domains are incapable of exponential amplification of targets when using di-deoxy mismatches at the 3' end of PCR primers (**Figure 5-23**). Thus, we hypothesized that using di-deoxy mismatch

primers in CSR evolution experiments would strictly enforce polymerase variants to not only recognize the geometries of misaligned bases, but that it would force the 3'-5' exonuclease domain to remain intact and functional. To test this hypothesis, we constructed error-prone libraries of KOD and subjected them to 5 rounds of CSR using normal and mismatched primers. We added diversity between rounds using error-prone PCR and sequenced 8 survivors from each selection (**Figure 5-24**). The polymerase variants selected with normal primers were shown to have an average of 7.25 amino acid mutations after 5 rounds. In contrast, variants selected with 3'-dideoxy primers or 'hi-fi' primers had an average of 6.42 amino acid substitutions. While these numbers are not drastically different, this small difference after only 5 rounds of selection may support the notion that 'hi-fi' primers enforce functional exonuclease activity over the course of the evolution experiment. In the variants sequenced, we did not see the common mutations that lead to inactivation of the proofreading domain. Either our sample size was too small, or it may take more than 5 rounds to see clear differences in selection methods. A more rigorous test of this hypothesis is needed to before drawing too many conclusions. Deep sequencing of the final rounds or increasing the number of sequenced variants in the future will give a more accurate and detailed representation of the mutational load from each selection.

Future rounds of selection, described below, were performed using 'hifi' primers for two reasons. First, we believe that it enforces the fidelity checkpoint by the polymerase. Secondly, by using Taq polymerase in the recovery round, we significantly lower any background that may come through the selection due to primer and plasmid carryover from the emulsion reaction.

***Inhibitor and chaotrope screening to identify potential targets for selection***

Given our ability to select for polymerases with novel function, we wanted to create polymerase variants that would be resistant to common PCR inhibitors and ultimately be used over a variety of diagnostic applications. Before initiating selections, we needed to assess the level of inhibition of each particular compound. To determine $IC_{50}$ values of chaotropes and other inhibitors, we performed standard 50 μL PCR reactions using KOD. A summary of the results can be found in **Table 5-3**. For a representative gel and inhibition curve, see **Figure 5-25**.

Consistent with earlier work, KOD was inhibited by heme, the chief inhibitor present in blood, with an $IC_{50}$ of 5.3 μM. Lipoteichoic acid, a large component of Gram positive bacterial cell walls (Percy and Gründling, 2014), was also found to inhibit KOD (IC50 = 61 μg/mL). Melanin, found in human hair and skin, inhibited KOD with an $IC_{50}$ of 16.8 μM. Selections using these compounds are described below. Chaotropes inhibited KOD at varying concentrations hovering in the millimolar range. However, 20% SDS was unable to inhibit KOD, highlighting its potential use in sample preparation for point-of-care diagnostics using PCR. We also identified $Mg^{2+}$ as a strong inhibitor to KOD. While $Mg^{2+}$ is needed for the reaction, typically 1-2 mM, strong inhibition is seen beyond 2.3 mM $Mg^{2+}$.

***Initiation of inhibitor selections***

To select for inhibitor-resistant versions of KOD, we initiated CSR directed evolution experiments with 'hi-fi' primers using varying levels of lipoteichoic acid and melanin. For the lipoteichoic acid selections, we performed 5 rounds of CSR. In the end,

survivors of this selection were able to amplify themselves in the CSR reaction at final concentration of 2.67 mg/mL lipoteichoic acid, far above the $IC_{50}$ observed in our initial screening. Sequencing of 15 pre-screened clones from this selection revealed a number of strongly conserved mutations (**Figure 5-26**). Positions 97, 145, 228, 295, 468, 539, 580, 664, 740, and 753 appear to be under selection in this evolution experiment. For melanin resistance, we again performed 5 rounds of CSR. In the final round of this selection, polymerases were challenged with 4.16 mg/mL melanin. Of the 20 clones screened, only 6 were functional. In this selection, positions 28, 97, 256, 418, 539, 580, 664, and 759 appear to be under selection (**Figure 5-27**). Given the mutational overlap at positions 97, 539, 580, and 664 with our lipoteichic acid selection, we believe these mutations are general stabilizing mutations and do not affect resistance. Further characterization of these mutants is ongoing.

We also initiated selections with increasing concentrations of heme. However, after 5 rounds of selection, we were unable to observe the same level of convergence described in the selections above—possibly due to multiple routes to heme resistance by KOD. In an effort to find optimal mutations that would lead to resistance in whole-blood, we carried two libraries forward for an additional 5 rounds of selection in the presence of human whole-blood. At the end of the $10^{th}$ round of selection, survivors that were capable of amplification in 80% whole-blood during the selection were pooled and deep sequenced (**Tables 5-4 & 5-5**). As a control, we also deep sequenced a round 10 control selection where KOD was error-proned between rounds but was not challenged with an inhibitor (**Table 5-6**).

In each of the 3 selections, we see a number of conserved mutations. Common to all 3, is the truncation of the protein at the C-terminus. This is more than likely due to the fact that the polymerase was fused to a 6x histidine tag for purification of variants. There

is an apparently strong selection pressure to free the polymerase from this affinity tag during CSR.

There are also a number of silent mutations found across the libraries. We believe these may be present for 2 reasons. In the case of inhibitor selections, silent mutations may be beneficial if they codon optimize the polymerase for expression. Having more polymerase within a cell would increase the amount of inhibitor needed in the reaction. Secondly, silent mutations may be easier to amplify by reducing secondary structure of the template during amplification. KOD is not known to have strand-displacing activity, thus it is possible that reducing secondary structure may allow for faster, more robust amplification.

We also see strong pressure to mutate positions 97 and 664. These mutations seem to crop up in all of our CSR experiments. E664K likely stabilizes protein-nucleic acid binding. The effects of this mutation were modeled in the previous section. Interestingly, we did not observe any mutations that would likely or are known to disable the proofreading domain of KOD. We interpret these results as support of our hypothesis that our 'hi-fi' selection primers are indeed functional and prevent the polymerase from becoming error-prone during evolution.

Absent from the blood libraries are obvious mutations that lead to resistance. Again, this may be due to the likelihood of multiple mutational pathways to heme and blood resistance. More work is necessary to tease out necessary the necessary substitutions that ultimately allowed the variants to survive the selction. Highly functional variants have been isolated from whole-cell screening (**Figure 5-28**). However, these mutants have yet to be functionally validated with purified enzyme in the presence of whole-blood. The characterization of these and other variants is ongoing in our laboratory.

### *Conclusions*

We further expanded the reach of CSR in this section. By including di-deoxy mismatches in our selection primers, we no longer observed 3'-5' exonuclease debilitating mutations in any of our selections. While more characterization is needed, these results are promising. Beyond the apparent increase in fidelity, using dideoxy primers lowers the amount of background or carryover from round to round during a CSR evolution experiment.

We also characterized KOD in the presence of common chaotropes and inhibitors. Chaotropes generally inhibited KOD in the millimolar range. Future evolution experiments could attempt to increase resistance to these chemicals. Selections were initiated on lipoteichoic acids, melanin, and heme. These three inhibitors are immensely important for diagnostic and forensic applications. In each case, we were able to turn five rounds of selection. With the exception of the heme libraries, we observed convergence. Additional rounds of selection in whole blood did little to identify mutations that led to the resistant phenotype during the evolution experiment, possibly due to multiple mutational routes. Characterization of mutants identified here should shed light on optimal residues. It may also prove possible to combine mutations from each of the inhibitor selections and create a single polymerase that is resistant to all of the inhibitors.

## Methods & Materials

### *Characterization of inhibitors*

Fifty microliter PCR reactions were setup using Accuprime Pfx buffer, 400 nM F and R primers, and 10 ng template plasmid. Initial range finding experiment narrowed the search for inhibition. We then chose 10 concentrations for the final experiments. After 25

cycles of PCR, the entire reaction was observed on a 0.8% agarose gel. Intensities of each band were then measure with ImageJ. Percent inhibition was then calculated and graphed. We then fit the data using SigmaPlot10 to determine $IC_{50}$ values.

***CSR selections and emulsion PCR***

Libraries were cloned into pET21 vector and transformed into BL21 DE3 cells. Overnight library cultures were seeded at a 1:20 ratio into fresh 2xYT media supplemented with ampicillin and grown for 2 hours at 37◦C. Polymerases were induced with 1 mM IPTG and incubated at 37◦C for 4 hours. Induced cells (400 µL total) were spun in a tabletop centrifuge at 3,000 x g for 8 minutes. The supernatant was discarded and the cell pellet was washed in TBS and then resuspended in 300 µL CSR mix: 1x Selection buffer (50 mM Tris-HCl (pH8.8), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$), 260 µM dNTPs, 530 nM of each of the forward and reverse primers, 10-300 µM TMAC, and varying ranges of inhibitor concentrations. The resuspended cells were placed into a 2 mL tube with a 1mL rubber syringe plunger and 1200 µL of oil mix (73% Tegosoft DEC, 7% AbilWE09 (Evonik), and 20% mineral oil (Sigma-Aldrich)). The emulsion was created by placing the cell and oil mix on a TissueLyser LT (Qiagen) with a program of 42Hz for 4 minutes. The emulsified cells were thermal-cycled for 20-25 cycles.

Emulsions were broken by spinning the reaction (20,000xg) for 15 minutes, removing the top oil phase, adding 750 µL chloroform, vortexing vigorously, and finally phase separating in a phase lock tube (5Prime). The aqueous phase was cleaned using a Promega PCR purification column. Overnight treatment with DpnI mitigated plasmid carryover. Subamplification with recovery primers ensures that only polymerases were functional in the emulsion PCR amplified.

**Figure 5-23:** *Hifi PCR*

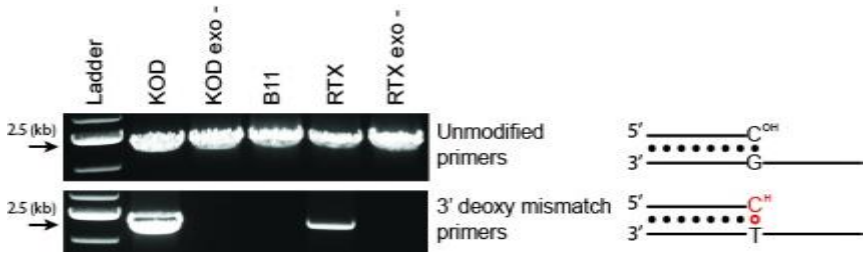



**Figure 5-23:** *Hifi PCR.* DNA polymerase activity was assessed by PCR using unmodified primers on a 2.5 kilobase fragment. Proofreading (3'-5' exonuclease activity) was assessed by use of 3' deoxy mismatch primers in the PCR. Inactivating the proofreading domain (N210D) in RTX prevents cleavage of the 3' deoxy primer and subsequently PCR.

**Figure 5-24:** Hifi vs normal sequencing results

**Table 5-3:** Common chaotropes and PCR inhibitors

| Inhibitor | IC50 | Error |
|---|---|---|
| SDS | 0 | 0 |
| Urea | 234.42 mM | 1.15 |
| Magnesium | 1.31 mM | 1.07 |
| Lithium Acetate | 10.69 mM | 0.11 |
| Guanidinium-Cl | 19.37 mM | 0.22 |
| Heme | 5.30 µM | 0.01 |
| Melanin | 5.35 ug/mL | 0.05 |
| Teichoic Acid | 60.98 ug/mL | 7 |

**Figure 5-25:** Representative inhibition gel and curve

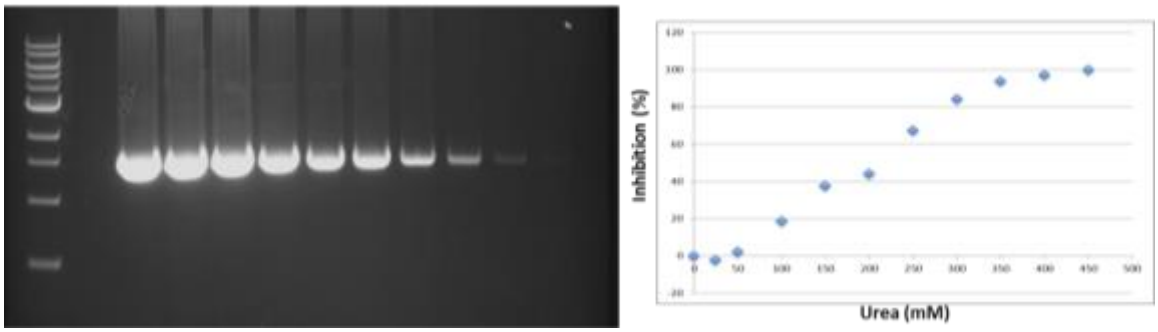



**Figure 5-25:** *Representative inhibition gel and curve*. Left, agarose gel of urea inhibition. 10 concentrations were tested (0-450 mM). Right, inhibition curve. Nonlinear regression was performed with SigmaPlot10. The IC50 was calculated to be 234.42 ± 1.15 mM urea.

**Figure 5-26:** Lipoteichoic acid round 5 sequencing results

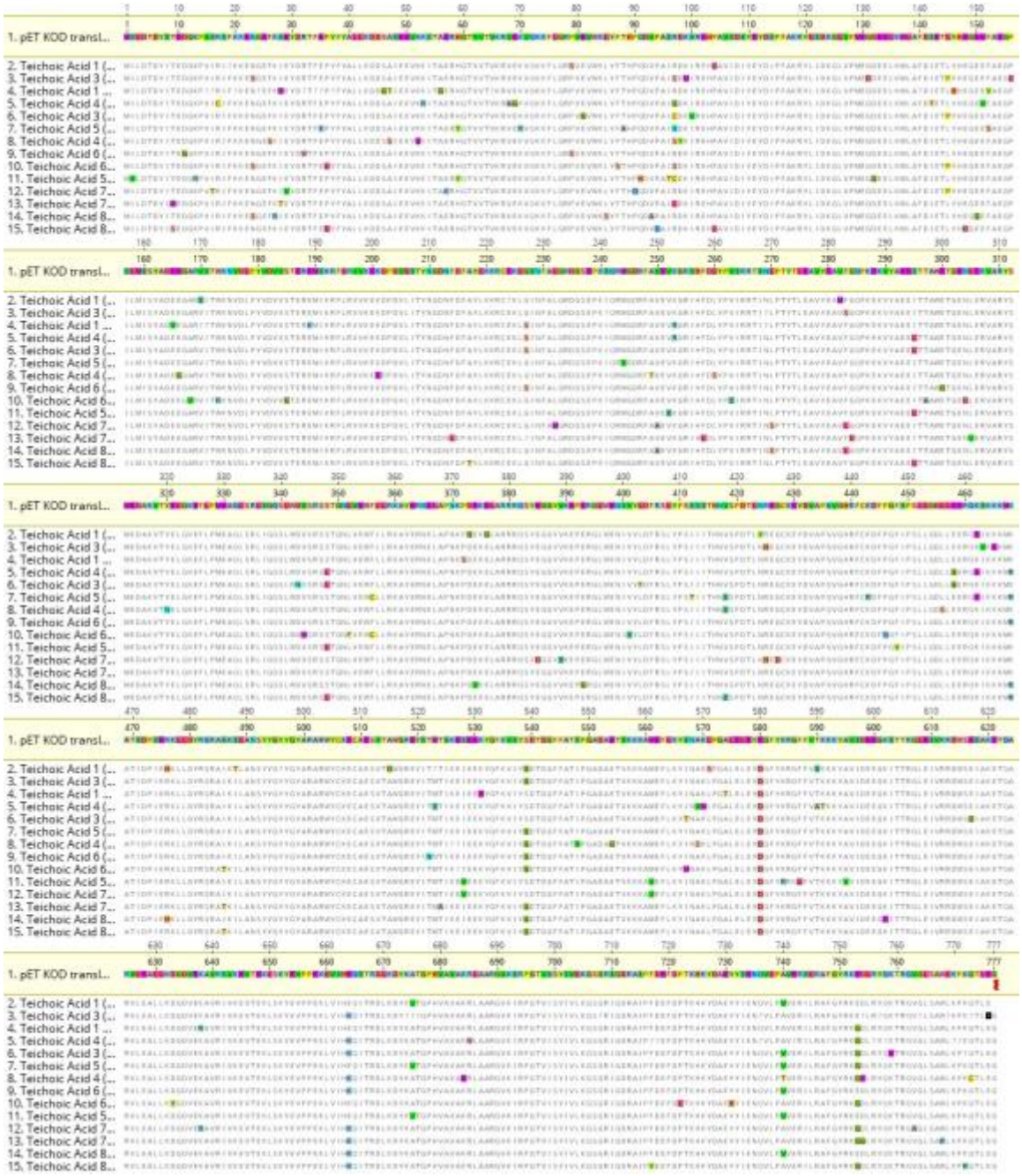


**Figure 5-27:** Melanin round 5 sequencing results

**Table 5-4:** Blood-A deep sequencing results

| Codon | Amino Acid | Variant Frequency | Change | Coverage | Protein Effect |
|---|---|---|---|---|---|
| 3 | | 99.70% | A -> T | 237,508 | None |
| 778 | | 99.50% | CC -> AG | 239,293 | Truncation |
| 2 | | 98.10% | A -> T | 234,416 | None |
| 777 | | 97.60% | GG -> TA | 246,429 | Truncation |
| 36 | P -> L | 41.30% | C -> T | 308,427 | Substitution |
| 741 | V -> A | 29.60% | T -> C | 331,662 | Substitution |
| 675 | A -> V | 28.60% | C -> T | 228,496 | Substitution |
| 758 | | 28.50% | C -> T | 283,297 | Truncation |
| 664 | E -> K | 22.10% | G -> A | 227,624 | Substitution |
| 768 | | 17.60% | G -> A | 260,934 | Truncation |
| 97 | R -> C | 14.10% | C -> T | 199,391 | Substitution |
| 97 | R -> H | 13.70% | G -> A | 199,670 | Substitution |
| 200 | E -> G | 11.10% | A -> G | 212,366 | Substitution |
| 726 | K -> E | 11.00% | A -> G | 378,356 | Substitution |
| 105 | | 10.40% | G -> A | 201,683 | None |
| 97 | R -> L | 9.60% | G -> T | 199,670 | Substitution |
| 155 | G -> D | 9.40% | G -> A | 208,301 | Substitution |
| 742 | E -> G | 9.10% | A -> G | 327,780 | Substitution |
| 59 | H -> R | 8.90% | A -> G | 363,891 | Substitution |
| 270 | | 8.90% | T -> C | 229,356 | None |
| 61 | | 8.80% | A -> T | 368,796 | None |
| 664 | E -> V | 8.40% | A -> T | 227,590 | Substitution |
| 46 | S -> N | 7.80% | G -> A | 330,271 | Substitution |
| 164 | D -> N | 7.80% | G -> A | 208,885 | Substitution |
| 632 | K -> Q | 7.70% | A -> C | 225,242 | Substitution |
| 59 | H -> Y | 7.60% | C -> T | 363,002 | Substitution |
| 743 | | 7.50% | C -> T | 323,101 | None |
| 759 | | 7.40% | A -> T | 281,000 | Truncation |
| 530 | | 7.40% | G -> A | 228,664 | None |
| 379 | R -> Q | 7.40% | G -> A | 228,534 | Substitution |
| 610 | I -> V | 7.20% | A -> G | 224,375 | Substitution |
| 503 | | 7.10% | G -> A | 228,136 | None |
| 343 | D -> G | 6.80% | A -> G | 237,152 | Substitution |
| 512 | | 6.70% | A -> G | 229,348 | None |
| 765 | | 6.60% | (T)3 -> (T)2 | 266,067 | Truncation |
| 27 | K -> E | 6.50% | A -> G | 284,044 | Substitution |
| 85 | | 6.50% | T -> C | 192,825 | None |

Table 5-4 Cont.

| | | | | | |
|---|---|---|---|---|---|
| 658 | | 6.40% | G -> A | 225,426 | None |
| 165 | | 6.40% | A -> G | 208,821 | None |
| 97 | R -> S | 6.40% | C -> A | 199,391 | Substitution |
| 373 | D -> N | 6.30% | G -> A | 229,628 | Substitution |
| 595 | | 6.30% | G -> T | 222,825 | None |
| 750 | | 6.20% | T -> C | 304,608 | None |
| 347 | | 6.10% | T -> C | 236,250 | None |
| 264 | I -> V | 6.00% | A -> G | 225,062 | Substitution |
| 638 | | 6.00% | A -> G | 223,390 | None |
| 753 | E -> G | 5.80% | A -> G | 294,827 | Substitution |
| 348 | | 5.80% | A -> G | 236,121 | None |
| 28 | I -> V | 5.70% | A -> G | 286,241 | Substitution |
| 378 | | 5.70% | T -> C | 228,004 | None |
| 644 | K -> E | 5.70% | A -> G | 224,499 | Substitution |
| 754 | D -> G | 5.60% | A -> G | 291,136 | Substitution |
| 242 | Q -> R | 5.60% | A -> G | 220,917 | Substitution |
| 146 | Y -> H | 5.60% | T -> C | 204,713 | Substitution |
| 524 | | 5.40% | G -> A | 228,603 | None |
| 662 | | 5.40% | T -> C | 227,236 | None |
| 522 | | 5.30% | C -> A | 228,478 | None |
| 235 | D -> N | 5.30% | G -> A | 219,463 | Substitution |
| 782 | H -> D | 5.30% | C -> G | 3,134 | Substitution |
| 741 | V -> G | 5.20% | T -> G | 331,662 | Substitution |
| 407 | | 5.10% | T -> C | 229,722 | None |
| 455 | D -> G | 5.10% | A -> G | 225,297 | Substitution |
| 237 | | 5.10% | A -> G | 219,732 | None |
| 93 | V -> A | 5.10% | T -> C | 196,672 | Substitution |
| 446 | P -> Q | 5.00% | C -> A | 225,802 | Substitution |
| 569 | | 5.00% | G -> T | 221,762 | None |
| 468 | K -> E | 5.00% | A -> G | 221,207 | Substitution |

**Table 5-5:** Blood-B deep sequencing results

| Codon | Amino Acid | Variant Frequency | Change | Coverage | Protein Effect |
|---|---|---|---|---|---|
| 3 | | 99.80% | A -> T | 229,616 | None |
| 778 | | 99.40% | CC -> AG | 290,402 | Truncation |
| 664 | E -> K | 98.50% | G -> A | 296,689 | Substitution |
| 2 | | 98.20% | A -> T | 226,630 | None |
| 777 | | 97.50% | GG -> TA | 298,162 | Truncation |
| 392 | | 58.00% | A -> G | 295,037 | None |
| 518 | | 44.30% | G -> A | 301,458 | None |
| 501 | R -> L | 22.70% | G -> T | 299,293 | Substitution |
| 93 | V -> A | 16.90% | T -> C | 231,198 | Substitution |
| 108 | | 14.80% | T -> C | 240,597 | None |
| 742 | E -> G | 12.00% | A -> G | 401,634 | Substitution |
| 768 | | 7.90% | G -> A | 314,145 | Truncation |
| 741 | V -> G | 7.20% | T -> G | 407,234 | Substitution |
| 753 | E -> G | 7.20% | A -> G | 357,834 | Substitution |
| 754 | D -> G | 6.80% | A -> G | 352,630 | Substitution |
| 116 | F -> I | 6.40% | T -> A | 244,900 | Substitution |
| 739 | | 6.30% | G -> A | 417,272 | None |
| 745 | L -> F | 6.10% | A -> T | 391,653 | Substitution |
| 242 | Q -> R | 5.80% | A -> G | 285,366 | Substitution |
| 551 | A -> V | 5.60% | C -> T | 291,017 | Substitution |
| 743 | R -> G | 5.50% | C -> G | 398,715 | Substitution |
| 698 | V -> G | 5.30% | T -> G | 560,458 | Substitution |

**Table 5-6:** KOD normal control selection deep sequencing results

| Codon | Amino Acid | Variant Frequency | Change | Coverage | Protein Effect |
|---|---|---|---|---|---|
| 3 | | 99.80% | A -> T | 200,985 | None |
| 778 | | 99.60% | CC -> AG | 214,727 | Truncation |
| 2 | | 98.20% | A -> T | 198,417 | None |
| 777 | | 97.70% | GG -> TA | 221,306 | Truncation |
| 580 | E -> D | 61.50% | A -> T | 214,686 | Substitution |
| 675 | A -> V | 38.60% | C -> T | 215,760 | Substitution |
| 741 | V -> A | 31.60% | T -> C | 295,179 | Substitution |
| 36 | P -> L | 29.50% | C -> T | 259,592 | Substitution |
| 760 | | 24.00% | (A)4 -> (A)3 | 247,659 | Frame Shift |
| 97 | R -> C | 22.30% | C -> T | 191,185 | Substitution |
| 97 | R -> H | 18.60% | G -> A | 191,667 | Substitution |
| 739 | | 14.50% | G -> A | 301,524 | None |
| 97 | R -> S | 13.90% | C -> A | 191,185 | Substitution |
| 765 | | 13.90% | (T)3 -> (T)2 | 237,139 | Truncation |
| 740 | A -> V | 12.50% | C -> T | 298,744 | Substitution |
| 649 | K -> R | 11.00% | A -> G | 213,763 | Substitution |
| 698 | V -> I | 11.00% | G -> A | 408,070 | Substitution |
| 763 | V -> G | 11.00% | T -> G | 241,735 | Substitution |
| 768 | | 9.20% | G -> A | 233,376 | Truncation |
| 587 | F -> L | 9.00% | T -> C | 213,638 | Substitution |
| 477 | | 7.80% | A -> G | 225,358 | None |
| 640 | V -> I | 7.60% | G -> A | 211,598 | Substitution |
| 352 | | 6.90% | A -> G | 222,223 | None |
| 614 | D -> N | 6.70% | G -> A | 212,799 | Substitution |
| 616 | | 6.70% | C -> T | 210,978 | None |
| 140 | F -> L | 6.60% | T -> C | 199,631 | Substitution |
| 304 | | 6.00% | C -> T | 223,762 | None |
| 168 | A -> V | 5.90% | C -> T | 204,448 | Substitution |
| 143 | | 5.80% | A -> G | 200,387 | None |
| 145 | L -> P | 5.70% | T -> C | 201,072 | Substitution |
| 629 | A -> V | 5.70% | C -> T | 210,525 | Substitution |
| 742 | E -> G | 5.50% | A -> G | 291,800 | Substitution |
| 754 | D -> G | 5.50% | A -> G | 258,613 | Substitution |
| 664 | E -> V | 5.40% | A -> T | 216,597 | Substitution |
| 57 | E -> K | 5.20% | G -> A | 308,922 | Substitution |
| 157 | | 5.10% | T -> C | 204,063 | None |
| 617 | E -> G | 5.10% | A -> G | 210,941 | Substitution |
| 779 | H -> N | 5.10% | C -> A | 2,872 | Substitution |

**Figure 5-28:** Blood selection survivors

## 5.4 THE DEVELOPMENT OF COMPARTMENTALIZED SELF-REPLICATION IN A EUKARYOTIC ORGANISM

### Introduction

In this section we adapt Compartmentalized Self-Replication to *Saccharomyces cerevisiae*. We refer to this as yCSR. This was an initial proof-of-concept to show that emulsion-based PCR selections would be compatible with yeast. Conventional wisdom on the third floor of MBB at the University of Texas at Austin stated that it would be impossible to perform emulsion PCR in yeast. For whatever reason, it was not thought possible. Here we describe: (1) the optimization of emulsion bubbles that would encapsulate single yeast cells; (2) the stabilization of emulsions to endure overcycling; (3) the process of expressing a thermostable DNA polymerase and amplifying its gene within the emulsion; and (4) the optimization of the buffering conditions to allow for in emulsion spheroplasting of cells. The protocols developed here are the basis for yCSR and yeast-based Compartmentalized Partnered Replication (yCPR). The adaptation of emulsion PCR-based selection methods to a eukaryotic organism enables the engineering of a number of protein classes not accessible to selections performed in bacterial systems.

### Results & Discussion

#### *Emulsion optimization*

In order to emulsify yeast, we first attempted the same protocols outlined in the sections above. Unfortunately, yeast cells are much larger than *E. coli*, and those conditions proved inadequate. Our previous work in optimizing emulsion conditions for bacterial cells led us to change the frequency of shaking by the tissuelyzer. We tested 35-

50 Hz in an effort to find a range where emulsion bubbles were uniform in size. We found that the ideal frequency laid somewhere between 35-40 Hz. Next, we determined the length of time (4-11 minutes) needed to produce large, uniform emulsions, capable of comfortably accommodating yeast cells. In the end, 37 Hz for 5 minutes was shown to be the optimal set of parameters to emulsify yeast in our water-in-oil emulsions. Brightfield images, processed by ImageJ, were used to determine the size of the emulsions, and we were able to accurately visualize yeast cells by constitutively expressing yeCitrine, a fluorescent reporter. Doing so would allow us to take images using fluorescent microscopy to verify that the cells were still intact after emulsification (**Figure 5-29**). To test for emulsion stability, we aliquoted the emulsion into PCR tubes and thermocycled them as we would in a typical CSR or CPR reaction. We used 35 cycles of PCR, as this would be the upper end of the number of cycles used in a directed evolution experiment. We optimized the amount of surfactant, oil, aqueous phase (buffer and cells), and mineral oil needed to produce stable emulsions, capable of enduring long cycling conditions. Stability was easily monitored at this stage. Unstable emulsions would phase separate, while stable emulsions would remain uniform (**Figures 5-30 & 5-31**). In the end, we found a number of optimal parameters that allowed us to emulsify yeast in thermostable emulsion bubbles (**Table 5-7**; **Figure 5-32**).

***Yeast CSR mastermix optimization***

To perform CSR in yeast, we built strains with yeast codon-optimized KOD under the control of the pGal1 promoter. After induction, we emulsified yeast in our standard CSR buffer and ran the products out on a gel. Using our normal CSR mastermix from our

bacterial cell selections, we were unable to see any amplification. We reasoned that this may be due to the cells not lysing under normal PCR cycling conditions. The yeast cell wall is impervious to heat, and colony PCR of yeast requires pretreatment with softening agents and typically zymolyase, an enzyme that rapidly degrades the cell wall. To test this hypothesis, we spheroplasted cells before the putting the cells in the emulsion. In spheroplasted cells, we were able to see product on an agarose gel. We next determined if we could spheroplast within the emulsion. That is, we wanted to add zymolyase to the mastermix and incubate cells at 37˚C before starting the normal cycling conditions. Spheroplasts are fragile, and we worried that cells would rupture during the emulsification process. Initial attempts at adding zymolyase failed. However, we found that if we softened the cells using sorbitol, DTT, and EDTA, we could get detectable levels of amplification (**Figure 5-33**).

Encouraged by these results, we then initiated a number of optimization experiments. We started our optimization by testing the amount of zymolyase needed. We found that as little as 80 uL of a 5 mg/mL stock was adequate in initial testing. There seemed to be diminishing returns above this value (**Figure 5-34**). Next, we varied the concentrations of dNTPs and primers in the mastermix. Commercial zymolyase preparations tend to have trace phosphatases and nucleases. We saw a modest improvement using 300 μM dNTPs (**Figure 5-35**) and 4 μM of each primer (**Figure 5-36**), much higher than a typical PCR reaction. We then determined if adding tetramethylammonium chloride (TMAC), a known PCR enhancer, had an effect on the emulsion PCR product. We found that increasing concentrations improved amplification

(10-300 μM TMAC). Finally, we determined that polymerases were more active in the CSR reaction in Tris-$SO_4$,pH 8.0 supplemented with 1 mM $MgSO_4$. Our original CSR mix was conducted at pH 8.8 for historical reasons, which may be a way to tune stringency in both CSR and CPR evolution experiments. The optimized concentrations of each component can be found in **Table 5-8**. A full protocol can be found in **Appendix II**.

***Yeast selection strain creation***

Homologous recombination is a benefit of using yeast over bacterial cells in performing CSR or CPR selections. By transforming libraries into the yeast chromosome, copy number variability found in plasmid-based selections is mitigated. In order to streamline the process, we created a number of yeast strains harboring IsceI cut sites and auxotrophic markers at the common Ura, Leu, and HO sites (Lee et al., 2015). For example, at the Ura integration site, we integrated a vector containing the IsceI cut site along with a histidine prototrophic marker. When transforming libraries with a uracil marker into this position along with a linear plasmid that transiently expressed IsceI nuclease, we saw an average of 7.5 x $10^7$ transformants, which is nearly as large as we see in bacterial cells (~$10^8$). By changing out prototrophic markers, we eliminate background during library construction. Plasmid maps for these vectors can be found in **Figures 5-37 – 5-40**.

***Future directons***

The development of long-read enzymes or PCR-based replication machinery has remained elusive due the size constraints in plasmid-based systems. In order to select for

highly processive, high-fidelity replication machinery, we have adapted CSR to *Saccharomyces cerevisiae* where integration of libraries into the chromosome is trivial, allowing for much longer amplification reactions within the emulsion PCR. It should now be possible to evolve an archaeal Family B DNA polymerase alongside its cognate proliferating cell nuclear antigen (PCNA) clamp. The successful selection of a DNA polymerase:PCNA pair will enable the amplification of much longer stretches of DNA.

The proliferating cell nuclear antigen (PCNA) is a clamp that is used to increase the processivity of DNA polymerases during replication in archaeal and eukaryotic organisms. Typically, the clamp is loaded during replication fork creation by the RFC clamp loader. The combination of the clamp and polymerase ensures that tens of thousands of bases are faithfully replicated without loss of the DNA polymerase. An engineered variant (E143K) of the proliferating cell nuclear antigen (PCNA) from Thermococcus kodakarensis (KOD), has been shown previously to be functional in PCR applications (Kitabayashi et al., 2002; Kranaster and Marx, 2010). In order to enable the amplification of long amplicons assembled in yeast, it should be possible to insert the PCNA and the Family B DNA polymerase (also from *T. kodakar*ensis) genes into our selection strains using pGal1 and pGal10 (Lee et al., 2015) and select long-read replication machinery (**Figure 5-41**).

***Conclusions***

In this section we adapted CSR to yeast. To do so, a number of optimizations were necessary. First, we needed to enlarge the emulsion bubbles to accommodate the larger organism. Next, we needed to find conditions to create stable emulsions. Through

iterative rounds of optimization, we found conditions that were stable after 35 cycles of PCR. Next, we optimized the mastermix conditions, improving our amplification yields more than 1,000-fold. We then created selection strains, which were capable of integrating libraries with nearly $10^8$ transformants with only a single electroporation. In addition to selecting for long-read replication systems, the methods developed here are also transferrable to the adaptation of Compartmentalized Partnered Replication (CPR) in yeast (see **Chapters 6 & 7**).

## Methods & Materials

### *Emulsion optimization*

BY4741 cells expressing yeCitrine were emulsified in PCR buffer. A range of concentrations of the following components were used in the development of stable emulsions: (1) light mineral oil, (2) Tegosoft, (3) Gransurf W9 surfactant, and (4) the aqueous phase containing live cells. Pictures of emulsions were taken with light and fluorescent microscopy. Stability was monitored after thermalcycling aliquoted emulsions. Phase separated emulsions were judged to be broken.

### *Yeast CSR optimization*

Geneblocks for KOD were purchased from IDT and cloned under the control of the pGal1 promoter in an integration vector targeting the URA3 locus using the Yeast Toolkit (Lee et al., 2015). Transformants were selected on SD-Ura. Overnight cultures were washed and induced with 2% galactosidase for 18 hours at 30˚C. Induced cells (500 µL- 3 mL total) were spun in a tabletop centrifuge at 3,000 x g for 5 minutes. The

supernatant was discarded and the cell pellet was washed in water and then resuspended in 1 M sorbitol, 25 mM EDTA, 50 mM DTT. Cells were collected again and washed in 1 M sorbitol. After another wash with water, cells were resuspended in 500 µL PCR mastermix. For optimal conditions, see **Table 5-7**. The resuspended cells were placed into a 2 mL tube with a 1mL rubber syringe plunger and 1200 µL of oil mix (876 Tegosoft DEC, 84 Gransurf W9, and 240 µL light mineral oil (Sigma-Aldrich)). The emulsion was created by placing the cell and oil mix on a TissueLyser LT (Qiagen) with a program of 37 Hz for minutes. The emulsified cells were incubated at 37˚C for 15 minutes, followed by 10 minutes at 95˚C, and then thermal-cycled for 35 cycles (**Appendix II**). Emulsions were broken by spinning the reaction (20,000xg) for 15 minutes, removing the top oil phase, adding 750 µL chloroform, vortexing vigorously, and finally phase separating in a phase lock tube (5Prime). The aqueous phase was cleaned using a Promega PCR purification column. Products of the yCSR amplification were visualized using agarose gel electrophoresis.

***Selection strain creation and transformation***

IsceI cutter and IsceI target sites were created using the Yeast Toolkit (Lee et al., 2015). For the IsceI cutter plasmid, the enzyme was put under the control of the pPGK1 promoter. See figure Figure 5-39. For target sites, we again employed the use of the standard Yeast Toolkit parts. We chose to integrate an IsceI cut site at each of the common integration sites (Ura, HO, and Leu) into standard BY4741 and BY4741ΔGal2. We selected transformants on drop out media at 30˚C. For the Ura locus, we integrated a His3 marker in addition to the IsceI site. At the Leu site, we inserted the recognition sequence and the URA3 marker. Finally, at the HO site, we integrated the Leu2 marker alongside the IsceI cut site. After validating IsceI cut sites within the chromosomes of targeted strains, we prepared electrocompetent cells. We electroporated NotI digested

fragments containing (1) a different auxotrophic marker at each of the homologous sites; and (2) the linearized IsceI cutter plasmid. Transformation efficiency was calculated by dilution plating.

**Figure 5-29:** Yeast emulsion images

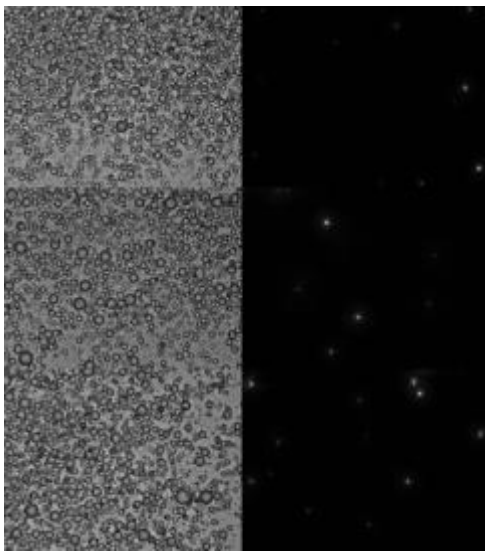


**Figure 5-29:** *Yeast emulsion images*. Left, brightfield images of representative emulsions. Right, fluorescent microscopy of intact cells within emulsions. Cells are not disrupted upon emulsification.

**Figure 5-30:** Stability of yeast emulsions example

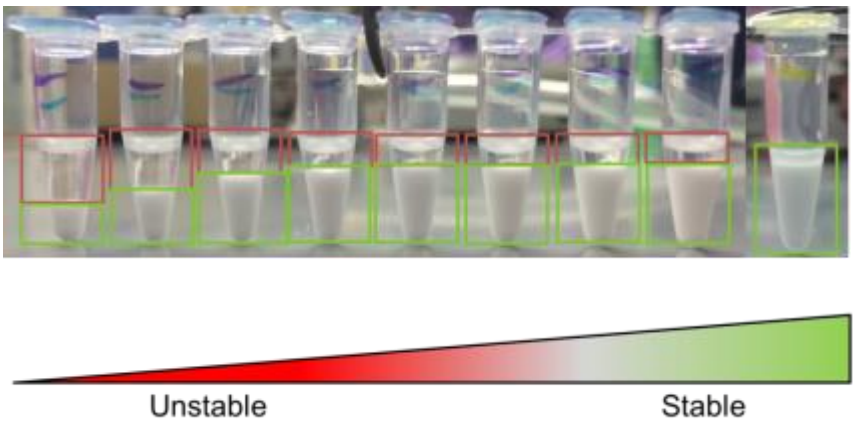



**Figure 5-30:** *Stability of yeast emulsions example*. Unstable emulsions phase separate. Shown is a gradient of stability observed.

**Figure 5-31:** Stable yeast emulsion optimization

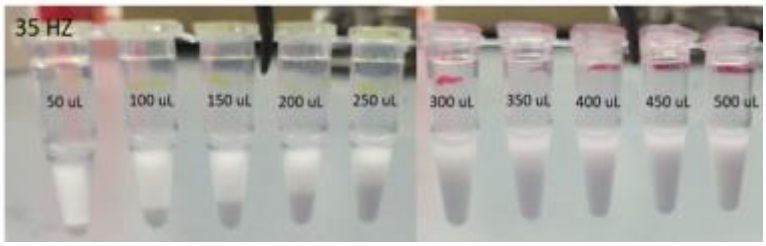



**Figure 5-31:** *Stable yeast emulsion optimization*. Increasing the amount of cells/aqueous phase in the setup of the water-in-oil emulsion stabilized the emulsions.

**Figure 5-32:** Final yeast emulsion stability

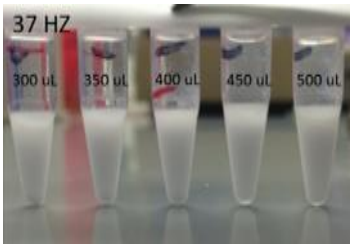



**Figure 5-32:** *Final yeast emulsion stability.* Final stability conditions tested at 37 Hz, 5 min in tissuelyzer. The 500 µL aqueous phase treatment showed the highest level of emulsion stability.

**Table 5-7:** Emulsion parameters

| Parameter | Original *(E. coli)* | Optimal Yeast |
|---|---|---|
| Surfactant | 84 uL Abil | 84 uL Gransurf |
| Frequency of Tissuelyzer (Hz) | 42 | 37 |
| Time (min) | 4 | 5 |
| Cells (uL) | 300 | 500 |
| Mineral oil:Tegosoft (uL) | 240:876 | 240:876 |

**Figure 5-33:** Original CSR conditions using spheroplasted cells

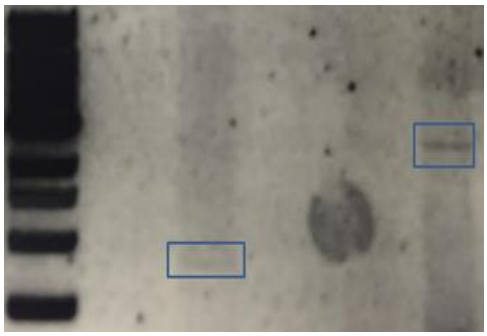


**Figure 5-33:** *Original CSR conditions using spheroplasted cells*. Yeast CSR-derived amplicons targeting a 700 bp fragment (left) of KOD and the full 2.5 kbp gene (right). This initial proof-of-concept was performed using pre-spheroplasted cells.

**Figure 5-34:** Zymolyase optimization

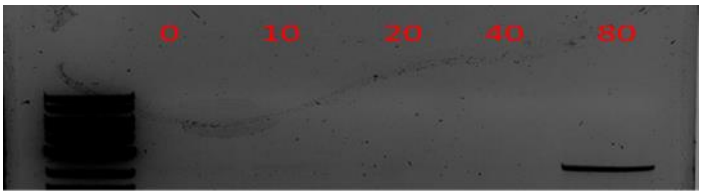



**Figure 5-34:** *Zymolyase optimization*. Varying amounts of zymolyase (0-80 μL) used in the optimization of yeast CSR.

**Figure 5-35:** dNTP optimization

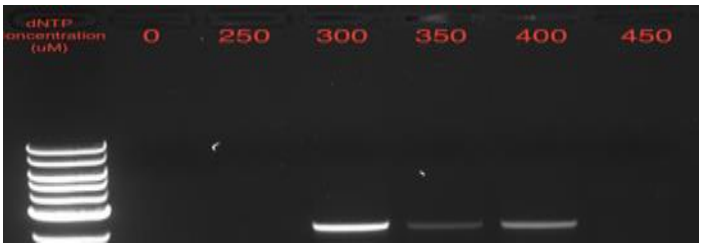



**Figure 5-35:** *dNTP optimization.* Varying concentrations of dNTPs (0-450 μM) used in the optimization of yeast CSR.

**Figure 5-36:** Primer optimization

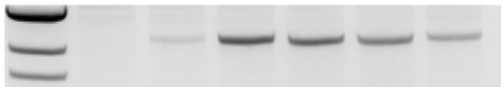


**Figure 5-36:** *Primer optimization*. Varying concentrations of primers (0-7.5 μM) used in the optimization of yeast CSR.

**Table 5-8:** Yeast CSR/CPR Mastermix

| Component | Volume (µL) |
|---|---|
| 10X Buffer (pH 8.0) | 50 |
| 4 mM Stock dNTPs | 37.5 |
| F Primer (20 µM) | 20 |
| R Primer (20 µM) | 20 |
| TMAC | 5 |
| Zymolyase (5 mg/mL) | 80 |
| RNAse | 1.75 |
| di$H_20$ | 285.75 |

**Figure 5-37:** Ura-His3 IsceI site

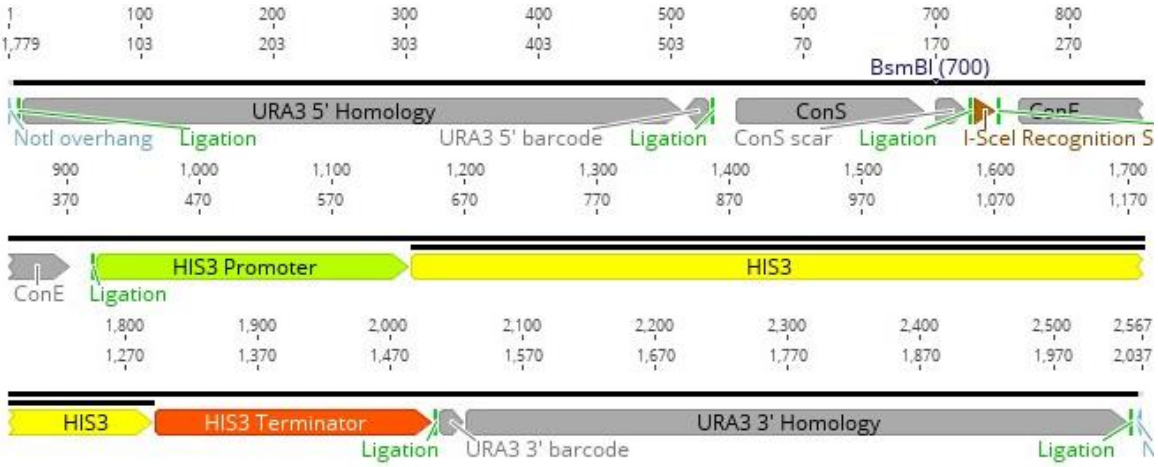



**Figure 5-37:** *Ura-His3 IsceI site*. Recognition site is shown between connectors S and E. His3 selection marker is used at the Ura3 site in targeted strains.

**Figure 5-38:** HO-Leu2 IsceI site

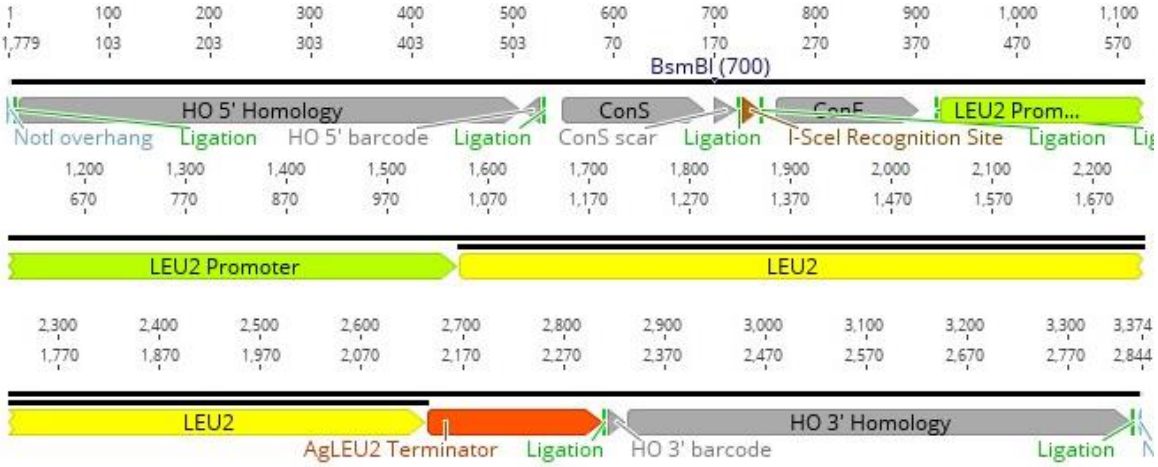



**Figure 5-38:** *HO-Leu2 IsceI site*. Recognition site is shown between connectors S and E. Leu2 selection marker is used at the HO site in targeted strains.

**Figure 5-39:** Leu-Ura3 IsceI site

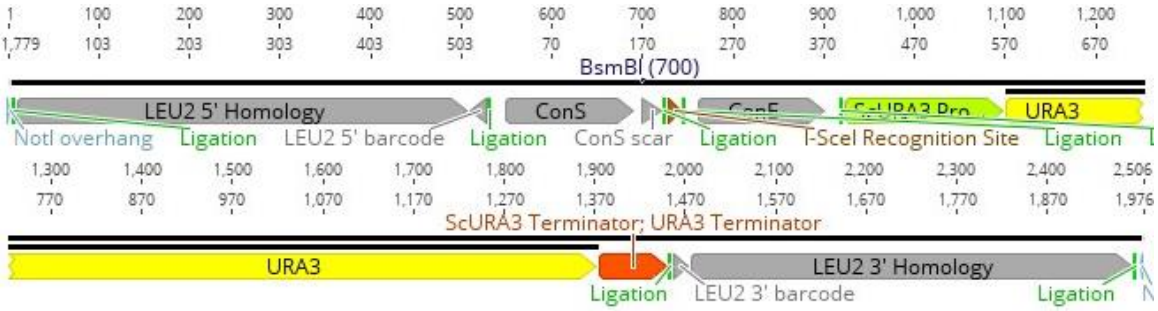



**Figure 5-39:** *Leu-Ura3 IsceI site*. Recognition site is shown between connectors S and E. URA3 selection marker is used at the Leu2 site in targeted strains.

**Figure 5-40:** IsceI cutter plasmid

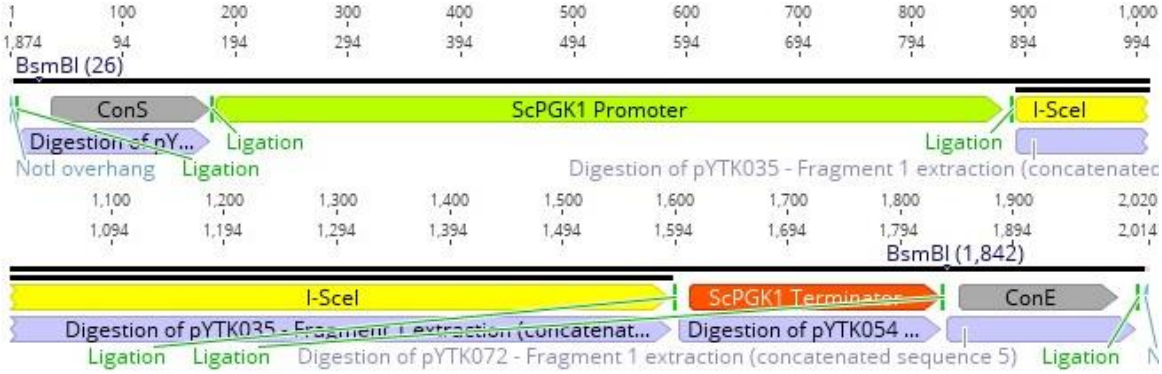



**Figure 5-40:** *IsceI cutter plasmid*. The IsceI linear fragment is shown. The enzyme is expressed under the PGK1 promoter/terminator pair. The expression of the enzyme is transient because there is no selectable marker and the piece is linear.

**Figure 5-41:** Long-read yCSR

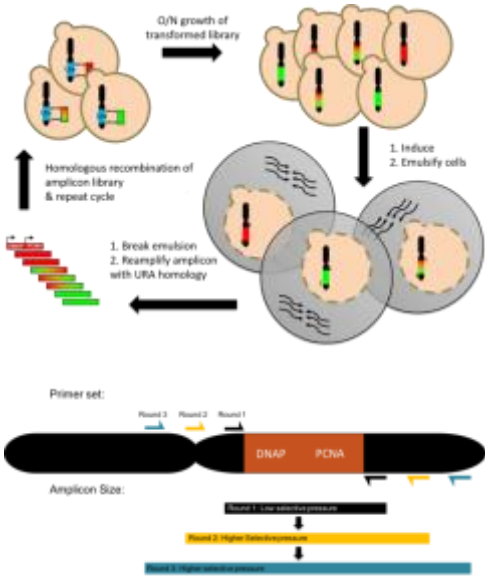



**Figure 5-41:** *Long-read yCSR*. A library of polymerase and PCNA variants are cloned into S. cerevisiae. The cells are emulsified after induction of the replication machinery. During emPCR, replication machinery variants are guided by primers to amplify individual variants. Variants of higher fitness that are capable of successfully amplifying themselves will dominant the recovered pool.

# Chapter 6: Functional expression of human GPCRs and plant cannabinoid biosynthetic pathways in yeast


## ABSTRACT

This chapter explores a number of proof-of-concepts. Having shown CSR to be functional in yeast in the previous chapter, we turned our attention to adapting Compartmentalized Partnered Replication (CPR) to our emulsion-based yeast platform. We start by creating *S. cerevisiae* strains for coupling to human G-protein coupled receptors (GPCRs). By disrupting a number of genes in the pheromone response pathway, we are able to show dose-dependent signal transduction of human serotonin receptor (5HT1A) and cannabinoid receptor subtype 2 (CB2) with the addition of their native ligands to yeast growth media. To create antagonism-based circuitry, we engineer and proof an inverter circuit. We also explore the production of the dietary cannabinoid β-caryophyllene and show autoactivation of an engineered strain expressing CB2. To our knowledge, this is the first example of a plant biosynthetic pathway activating a human receptor in the same cell. Next, we start the optimization of cannabidiol (CBD) producing yeast strains. We discuss the implications of yeast CPR and the many applications that are now afforded by coupling eukaryotic signal transduction pathways to CPR circuits in an amendable host organism.

## INTRODUCTION

G-protein coupled receptors (GPCRs) are integral membrane proteins that regulate a number of cellular processes in eukaryotic organisms (Vassilatis et al., 2003). It is thought that GPCRs date to the common ancestor of all eukaryotes (de Mendoza et al., 2014). These 7-transmembrane helical proteins respond to vast assortment of extracellular molecules, and signaling mechanisms are conserved across taxa. In humans, these receptors are involved in neurotransmission, growth, vision, olfaction and heart rate regulation (Vögler et al., 2008). Humans contain over 800 GPCRs with more than 400 dedicated to olfaction. Roughly 350 of the remaining receptors have been identified as potential drug targets (Tautermann, 2014). In fact, GPCRs make up the largest class of targeted proteins in the human proteome, making up an estimated 35% of current drug targets (Hutchings et al., 2010). The development of high-throughput screening technologies using these receptors would greatly increase the speed and development of pharmaceuticals (Zhang and Xie, 2012).

While screening ligands for druggable properties has been of great utility, engineering human GPCRs has also proven to be a fruitful enterprise. Orthogonal GPCR engineering was pioneered by neuroscientists to selectively depolarize neural subtypes. Early iterations of this include receptors activated solely by synthetic ligand (RASSLs), which utilized rationally designed mutants of the human kappa opioid receptor to engineer GPCR mutants that responded to the synthetic agonist spiradoline but not its native ligands (Coward et al., 1998). While these receptors were among the first chemogenetic engineered GPCRs, they were soon replaced by GPCRs engineered *via* a

platform that cleverly ported human GPCRs into *Saccharomyces cerevisiae*. By hijacking native pheromone signaling pathways, this system offered a means to screen mutagenized receptors. The chemogenetic receptors developed by this technology, termed DREADDS (designer receptors exclusively activated by designer drugs) were a far more powerful tool than their RASSL predecessors, as DREADDs offered a more generalizable scheme by which to evolve receptors to respond to a wide variety of drug-like molecules (Alexander et al., 2009). While various Gi, Gq, and Gs coupled DREADDs have been widely used by the neuroscience community for interrogations of neural function, this powerful yeast GPCR screening platform has been underutilized as a tool for drug discovery and moreover the autodetection of synthesized biomolecules in metabolic engineering projects. By adapting yeast-based GPCR screening to 1) a powerful directed evolution platform developed in our lab and 2) biosynthesis pathways in one yeast strain, it should be possible to create a highly sensitive and powerful means by which to evolve metabolic synthesis pathways for drug synthesis as well as to screen compounds for functional activity at these receptors.

Compartmentalized Partnered Replication (CPR) is a method in which genetic circuits are designed to express a PCR-ready, thermostable DNA polymerase, such as KOD, in response to an input. The utility of CPR is apparent when considered with respect to directed evolution of libraries. Here a portion of the genetic circuit is randomized, different cells contain different variants, and those variants that promote the production of the thermostable polymerase can be selectively amplified in an emulsion PCR. When the emulsion is broken, the productive partners are overrepresented relative

to unproductive partners. Multiple rounds of cloning, emulsion, and amplification will lead to the identification of those partners that are most effective in generating the DNA polymerase. These variants can have any of a variety of functions. Our group has previously shown that RNA polymerases, repressors, tRNA synthetases, and tRNAs can be optimized for function in the context of such circuits ((Ellefson and others; Ellefson et al., 2014; Meyer et al., 2014a).

Below we discuss the construction of agonist and antagonist CPR (**Figure 6-1**) and screening strains. We show functional coupling of receptor agonism to genetic circuits as well as autoactivation of the human CB2 receptor by cannabinoid production in a single strain. We conclude with the discussion of a number of technologies that are now possible from the groundwork performed here.

## RESULTS & DISCUSSION

### GPCR strain construction

The ability to manipulate native GPCR signaling in *S. cerevisiae* has proven to be a valuable tool in a number of high-throughput screening and directed evolution methodologies (Brown et al., 2000; Ladds et al., 2005; Nakamura et al., 2015; Reiländer and Weiβ, 1998; Yoshimoto et al., 2014). The deletion or disruption of Sst2p, Ste2p, and Far1p from the yeast genome allows downstream host signaling to be completely seized by heterologous receptor function. By producing chimeric and transplant $G_\alpha$ subunits, it has proven possible to couple receptor function to the production of auxotrophic markers, fluorescent proteins, and enzymatic reporters under the control of pFus1 or pFig1, the commonly used native pheromone response promoters that are activated by the MAPK signaling cascade. Transplanting the last 5 amino acids of the human subunit to the native yeast protein has proven sufficient to couple the heterologous expression and function of a foreign GPCR to the native pheromone response in yeast (**Table 6-1**). While strains exist with each of the necessary modifications, they are difficult to obtain, and here we built markerless versions using Cas9.

We used Benchling to identify potential sgRNAs for each of the targeted genes. We built 3 sgRNAs for each of the targets (**Table 6-2**) and tested the efficiency of knockouts (Ste2p, SSt2p, and Far1p), gene replacements (reporters), and site directed mutagenesis of the carboxy terminus of the native yeast $G_\alpha$ subunit (**Table 6-3**). For gene disruptions or knockouts, we designed repair DNA that inserted an in-frame stop codon

towards the 5' end of the gene along with a barcode that we could use with qPCR probes for multiplex detection of genomic alterations (**Figure 6-2**). For gene replacements at pFig1 or pFus1, we targeted the native gene and inserted either: (1) the archaeal Family B DNA polymerase KOD for use in agonist-driven CPR circuits; (2) the fluorescent reporter zsGreen1; or (3) the bacterial tetracycline repressor (tetR) to be used in inverter circuits for coupling antagonism of a GPCR to CPR circuits. To change the $G_{\alpha}$ subunits we changed the codon usage at the Cas9 cut site along with changing the terminal 5 amino acids at the carboxy terminus for coupling to heterologous GPCRs. Strains were constructed in BY4741ΔGal2 to enable dose-response activity of the pGal1 and pGal10 promoters (**Table 6-4**). In total, 49 strains were constructed for GPCR-dependent experiments. The sequence-validated strains constructed here can now be used for GPCR-CPR in agonist and antagonist mode as well as in FACS-based selections. Further, the zsGreen1 strains can be used in plate-based fluorescence assays. In addition, the tetR strains can be coupled with any reporter under the control of a hybrid pGal1-tetOx2 promoter (**Figure 6-3**). Each of the alterations were made without using common auxotrophic or antibiotic resistance markers, making them easier to manipulate, engineer, or customize.

**Functional expression of human 5HT1A**

The human serotonin receptor (5HT1A) has been previously shown to function in yeast using strains similar to those outlined above (Nakamura et al., 2015). To test our newly created strains, we constructed expression vectors containing human 5HT1A under the control of the strong, constitutive pTDH3 promoter. We tested a single integration

vector along with a high-copy plasmid (2 micron) in JG05 strains. We found that reporter strains using the pFus1 promoter had high levels of background. Strains utilizing pFig1 showed a good dose-response relationship with the 2 micron plasmid (high-copy), exhibiting far superior signal:noise (**Figure 6-4**).

Having shown functionality of the 5HT1A receptor, we constructed CPR strains (JG05-1.0) using this receptor. Initial coupling to CPR circuits showed a high level of background. However, we were able to mitigate background by increasing the pH of the CPR master mix to pH 8.8 and lowering the number of cycles from 35 to 20 during the emulsion PCR. This resulted in a 20-fold increase in signal:noise and marked the first example of a CPR circuit functioning in a eukaryotic organism (**Figure 6-5**). Further optimization of expression parameters should greatly increase the signal:noise. For example, we expect to see a much more pronounced difference when expressing the receptor from a high-copy plasmid. The experiments outlined here were also performed with cells growing with the ligand for 18 hours. By concentrating the cells and limiting the time of induction, we should see a decrease in our background signal. Optimization experiments are currently underway in our laboratory.

**Inverter circuit design, construction, and validation**

In order to couple antagonism to CPR, we devised an inverter circuit. In this scheme, activation of the GPCR triggers the expression of bacterial tetR, which then in turn inhibits the inducible expression of a modified Gal1 promoter containing 2 tet operator sequences downstream of the TATA box. The details of this hybrid promoter have been discussed elsewhere (Nevozhay et al., 2009). If the GPCR is inhibited, then

tetR is not produced, and the promoter is inducible with galactosidase. This circuit could also be used in the engineering of receptors. Often, GPCRs show a low level of constitutive activity. Using the inverter circuit described here, it should prove possible to select against constitutive activity or receptor activation by off-target ligands, thereby increasing sensitivity and selectivity.

To test the feasibility of this approach, we constructed strains with the fluorescent reporter zsGreen1 under the control of the hybrid Gal1-2xtetO promoter in single-copy integration vectors. We also created integration vectors containing tetR with a nuclear localization signal and tetR variants with degradation tags of varying strengths (Lee et al., 2015). Next, we tuned expression of the tetR with low, medium, and high strength constitutive promoters. Our idea was to gauge the dynamic range of the system and potentially mimic the range of induction strengths of receptor activation. After an 18 hour galactosidase induction, we measured the fluorescent output of each of the strains. Without a degradation tag, we observed 121-fold repression when expressing tetR with a high strength promoter (**Figures 6-6 & 6-7**). In contrast, low-level expression of tetR resulted only 20-fold repression in comparison to the control strain lacking tetR expression. The addition of a weak degradation tag (Ubi-M) displayed ranges of 1-73-fold repression when comparing low-high strength expression. The Ubi-Y degradation tag was only able to show low-levels of repression under high expression conditions, and Ubi-R did not show repression under any expression condition, due to the repressor being rapidly degraded. CPR strains, JGXX-3.0, have now been constructed and mock selection with antagonists are currently underway.

**Functional expression of human CB2 and autoactivation by biosynthesis of β-caryophyllene**

We tested the ability of three of our strains to couple to the human CB2 receptor expressed on a high-copy number plasmid. We chose Gα subunits that were predicted to couple to the receptor (Shim et al., 2013). We monitored the dose-response of 2-arachidonyl glycerol, the native endocannabinoid, in strains differing only in the terminal 5 amino acids of the Gα protein. We also built control receptors that contained stop codons in the middle of the gene. We refer to these as 'broken' receptors. Using zsGreen1 as a reporter, each of the three strains was capable of responding to the ligand in a dose-dependent manner only in the presence of a fully functional receptor (**Figure 6-8**). Strains containing broken receptors did not produce a signal at any concentration tested.

A powerful pathway engineering method would involve coupling the biosynthesis of a compound to a CPR circuit using a GPCR. This would allow the directed evolution of the pathway as well as the organism with the addition of more complex CPR circuitry. A literature search revealed that β-caryophyllene, a dietary cannabinoid, was a selective agonist for CB2. This was an attractive molecule because yeast only require one additional enzyme from *Artemisia annua* to convert FPP to β-caryophyllene (Reinsvold et al., 2011). We tested our strains and found that JG04-2.0 and JG05-2.0 both responded to high concentration of β-caryophyllene (**Figure 6-9**).

To produce β-caryophyllene in yeast, the native isoprenoid pathway for the production of farnesyl diphosphate (FPP) was augmented by the heterologous expression of QHSI β-caryophyllene synthase from *A. annua*. To increase yields, the UCP2-1 transcription factor was inserted to upregulate expression of enzymes HMG-CoA synthase (ERG13), mevalonate kinase (ERG12), and phosphomevalonate kinase (ERG8).

In addition, extra copies of the enzymes isopentenyl diphosphate isomerase (IDI1), the truncated HMG-CoA reductase (tHMG1), and farnesyl diphosphate synthase (ERG20) were constructed into expression vectors. Fragmentation mass spectrometry of extract was used to validate production (**Figure 6-10**). Next, we inserted our β-caryophyllene production vectors into a GPCR strain expressing the human CB2. When compared to a strain lacking the pathway, we found that we could activate the GPCR with β-caryophyllene producing strains (**Figure 6-11**). To our knowledge, this is the first autoactivation of a human GPCR with a plant biosynthetic pathway in yeast. We have now inserted the pathway into CPR strains (KOD reporter) and are attempting to evolve the QHSI β-caryophyllene synthase.

**Biosynthesis of Cannabidiol**

The natural biosynthesis of cannabinoids in *Cannabis sativa* can be roughly divided into four distinct metabolic pathways: hexanoate, 2-*C*-methyl-D-erythritol 4-phosphate (MEP), geranyl diphosphate (GPP), and the penultimate cannabinoid production center (**Figure 6-12**). A total of 20 enzymes constitute the complete pathway; intermediates and flux relationships have been identified, providing a framework for alternative biological chassis development. The yeast *Saccharomyces cerevisiae* is an attractive organism in which to port the production of cannabinoids because many of the precursors needed are present, mitigating the need to recreate the pathway wholesale. Existing strains have been engineered to overproduce GPP, which reduces the complexity of the complete pathway from 20 enzymes to 10. Coupled with the ability to add hexanoate exogenously while still obtaining substantial levels of key intermediates, only

5 enzymes are left to optimize to create minimal cannabinoid production strains (Stout et al., 2012).

The genes for hexanoyl-CoA synthetase (CsAEE1), 3,5,7 trioxododecanoyl-CoA synthase (OLS), olivetolic acid cyclase (TKS), geranyl-pyrophosphate-olivetolic acid geranyltransferase (CsPTI), and cannabidiolic acid synthase (CBDA) were codon optimized for expression in *S. cerevisae*. These five enzymes were cloned under the control of endogenous yeast promoters: ScTDH3, ScCCW12, ScPGK1, ScHHF2, and ScTEF11, respectively. Promoters were chosen based on published expression profiles and kinetics (**Table 6-5**). Each enzyme was also given a unique terminator to mitigate the potential ejection of pathway enzymes *via* recombination. All five were assembled along with the Ura3 gene into a single integration vector *via* a restriction-digestion reaction and cloned into E. coli (DH10B). The complete pathway was stably integrated into the ΔUra3 locus of S. cerevisiae BY4741 and a BY4741 ΔPep4 knockout (**Figure 6-13**). To determine if the enzymes were being produced, we performed proteomics and found that each of the cytosolic enzymes was present in great abundance (**Figure 6-14**). We were unable to see spectral counts of the aromatic prenyltransferease, which is an integral membrane protein in the plant. We instead fused a fluorescent reporter and determined that the enzyme was expressed. There were no significant differences observed in protein abundance between BY4741 and the ΔPep4 knockout, leading us to continue with BY4741.

Using this initial strain, we were able to produce a small amount of CBD with supplementation of 1 mM hexanoate in the growth medium after 7 days of growth. We

confirmed the production of CBD by HPLC (**Figure 6-15**). We were also able to show that the bulk of CBD was found within the cell pellet. This is more than likely due to the hydrophobicity of the molecule forcing its aggregation within cell membranes. Control strains, lacking pathway enzymes at the URA3 locus, did not produce CBD.

To determine the bottleneck of our pathway, we performed growth assays with BY4741 cells in the presence of each of the intermediates (**Figure 6-16**). Hexanoic acid and olivetolic acid, the first two molecules in the pathway, were shown to strongly inhibit growth. CBGA and CBD were indistinguishable from one another. However, they produced a modest growth defect. From these data, we concluded that growth would be a suitable surrogate for determining optimal conditions or guiding engineering efforts. Increasing the growth rate would likely correlate to the production of at least CBGA. Due to the abundance of CBDA synthase in the proteomics analyses, we would not expect the conversion of CBGA to CBDA to be problematic unless compartmentalization of enzymes caused partitioning.

Our inability to observe fragments of the aromatic prenyltransferase in our proteomics analyses and the strain's poor growth in hexanoic acid led us to hypothesize that increasing the availability of the aromatic prenyltransferase would lead to robust growth and production of CBD. The aromatic prenyltransferase resides on the outermembrane of chloroplast in the plant. Using the bioinformatics tools TargetP and ChloroP, we predicted that the first 74 amino acids of the integral membrane enzyme were a part of a chloroplast targeting sequence (**Figure 6-17**), which yeast do not contain. The inability of the organism to interpret the plant signal sequence more than likely

triggers degradation, which we observed in the upregulation of the unfolded protein response (UPR) in our proteomics experiments (**Figure 6-18**).

In an effort alleviate the growth defect and increase CBD production, we then rationally designed a number of aromatic prenyltransferases variants. We targeted the inner mitochondrial membrane, the outer mitochondrial membrane, and the plasma membrane using common signal peptides and motifs (**Figure 6-19**). Growth assays in hexanoic acid of these variants did little to distinguish optimal targeting of the aromatic prenyltransferase (**Figure 6-20**).

The two substrates for the aromatic prenyltransferase are olivetolic acid and GPP. Given that CsAEE1, OLS, and the OAC had all been shown to be functional in yeast cells and lysates (Gagne et al., 2012), we concluded that we were more than likely making the toxic molecule olivetolic acid at this point. While yeast are able to synthesize GPP, it is in low abundance and a side product of Erg20, an essential gene in yeast that makes farnesyl pyrophosphate (FPP) as a part of the mevalonate and isoprenoid pathways. Others have shown that the expression of a mutant from of this enzyme (Erg20WW) can increase the amount of GPP available in the cell. With an additional copy of IDI1, a truncated HMG1 (tHMG1), and a mutant transcription factor (UPC2-1), large amounts of GPP can be produced in the cytosol of *S. cerevisiae* (**Figure 6-21**)(Zhao et al., 2016).

We reasoned that including these proteins along with an additional copy of the aromatic prenyltransferase might alleviate the growth defect if GPP was the limiting factor in our growth assays in hexanoic acid. We constructed integration vectors targeting the Leu2 and HO sites within BY4741 with each of the aromatic prenyltransferase

variants being expressed under the control of the pTDH3 promoter (**Figure 6-22**). Preliminary screening of strains harboring the CRC and GPP integration vectors in 10 mM hexanoic acid and 1% acetate revealed noticeable differences in growth (**Figure 6-23**). These strains have subsequently been grown in 1 L batches. We currently await extraction and metabolic profiling of these variants.

The establishment of a robust CBDA-producing strain will enable the production of other natural cannabinoids as well as the capability to further engineer pathways for more exotic, tailor-made molecules. Natural derivatives will result from plugging in key enzymes at pathway branch points. For example, CBGA is a common intermediate in the synthesis of CBCA, CBDA, and THCA. By replacing CBDAS with CBC synthase or THC synthase, CBC or THC can be readily accessed biosynthetically (Figure 2). Another natural branch point occurs upstream of olivetolic acid, which produces the shorter chain divarinic acid derivatives (CBDV, THCV, and CDCV). Additionally, a number of other cannabinoid branch points from *C. sativa* and other cannabinoid producing plants should be amenable to our plug-and-play platform to create a vast array of combinatorial variants. The addition of other hexanoate and GPP derivatives should provide yet another avenue to create novel compounds not otherwise found in nature.

## Conclusions and looking forward

### *Eukaryotic CPR*

We have now adapted CPR to function in the context of the yeast *Saccharomyces cerevisiae*. This greatly expands the range of experiments that can be carried out. By adapting CPR to function in the context of yeast, we can now undertake the optimization

of eukaryotic signal transduction pathways. This was not possible in bacteria, and the extension from transcription and translation to signal transduction pathways was an obvious first step in creating eukaryotic CPR. For example, we have cloned a human GPCR (5HT1A) into yeast and coupled it to the yeast pheromone response signal transduction pathway that ultimately leads to the expression of KOD polymerase. This allows us to optimize or alter the function of this GPCR. Ultimately, similar schemes should be possible with almost any signal transduction protein or small molecule effector that can be cloned into or made in.

***Next-Generation Drug Screening using CPR***

The ability to examine signal transduction pathways in yeast further leads to the ability to identify drugs that can modulate signal transduction pathways. In this instantiation, a drug would be co-expressed with the pathway or added exogenously. Pathways for the production of small molecules can be co-expressed in yeast with a given GPCR or other signal transduction partner. The pathways will lead to the production of drugs, some of which may activate (or inhibit) the GPCR. We have proofed the production of cannabidiol, a 5HT1A receptor agonist and CB1/CB2 receptor antagonist, and β-caryophyllene, a CB2 receptor agonist. By including primers that can amplify the pathways that created the small molecule, emulsion PCR will lead to the selective amplification of those pathways or pathway variants that produce the most functional effector molecule. Using CPR for drug-discovery should now be possible by creating libraries of human receptors and screening drugs in either the agonist or antagonist CPR circuits. This is an especially powerful technique when one considers the ability to screen

for off-target hits in a high-throughput manner by sequencing activated or inactivated receptors.

***Therapeutic Antibody Selections Using CPR***

The ability to screen small molecules is only one possibility. Using CPR, it should also prove possible to select for human antibodies that activate (Agonist CPR) or inactivate (Antagonist CPR) GPCRs. To do so, libraries of antibodies should be co-expressed in GPCR-expressing strains. Rather than tethering the antibody to the outside of the cell, as done in yeast display, antibodies are anchored inside the periplasmic space with a GPI-domain fusion (Flo42). Preliminary experiments with the human Glp1R using a fluorescent reporter suggest that this may be possible. Antibodies that can activate a given GPCR will lead to the production of the DNA polymerase, which leads to amplification of the antibody variants capable of triggering the receptor. The implementation of the inhibitory pathway requires the inverter circuit, in which the repression of a repressor led to KOD polymerase production. This method of antibody discovery is especially important since potentially therapeutic antibodies identified will be identified not only on basis of the strength of interaction (affinity), but also on functionality, such as whether the antibody is an antagonist or agonist of the GPCR. Such functional selections cannot be carried out in almost any other way and would mark the first directed evolution platform capable of producing human therapeutic antibodies.

## Methods & Materials

### GPCR Strain creation and genetic manipulation

Strains BY4741 and BY4741ΔGal2 were obtained from the Marcotte laboratory. Cas9 plasmids were constructed as previously described (Lee et al., 2015). To cure the Cas9 plasmid, we patch plated 15 colonies that were grown on rich media. Human receptor sequences were procured from Uniprot, codon optimized for expression in *S. cerevisiae*, and ordered as GeneBlocks from IDT. Entry vector stocks (YTK001) were made to allow for shuttling between integration, low-copy, and high-copy plasmids. Promoters and terminators were taken from the Yeast Toolkit. A list of GPCR strains is contained in Table 6-4. The copy number of receptor plasmids are indicated where necessary. Integration vectors were digested with NotI. All transformations were done using the EZ Yeast Transformation Kit (Zymo Research). For pathway engineering, vectors were constructed with the same procedures. Plasmid maps and sequences are available upon request.

### Dose-response of human receptors expressed in yeast

For 5HT1A, overnight cultures of receptor transformants of JG05-2.0 and JG05-0.2 were washed and diluted 1:1000 in media containing varying amounts of serotonin. Cells were washed 3 times and resuspended in PBS. Fluorescent (485nm/515nm) and cell density (600 nm) measurements were taken on a Cytation 5. Data were plotted and analyzed using SigmaPlot10. For CB2, strains JG03-,JG04-, and JG05-2.0 were transformed and treated the same. We also included broken receptor (stop codons) in these assays.

**CPR in yeast**

JG05-1.0-5HT1A were grown overnight and diluted into media containing 250 μM serotonin. Control cells were not induced. Cells (3 mL total) were collected after 18 hours of induction and spun in a tabletop centrifuge at 3,000 x g for 5 minutes. The supernatant was discarded and the cell pellet was washed in water and then resuspended in 1 M sorbitol, 25 mM EDTA, 50 mM DTT. Cells were collected again and washed in 1 M sorbitol. After another wash with water, cells were resuspended in 500 μL PCR mastermix (pH 8.0 or pH 8.8).The resuspended cells were placed into a 2 mL tube with a 1mL rubber syringe plunger and 1200 μL of oil mix (876 Tegosoft DEC, 84 Gransurf W9, and 240 μL light mineral oil (Sigma-Aldrich)). The emulsion was created by placing the cell and oil mix on a TissueLyser LT (Qiagen) with a program of 37 Hz for minutes. The emulsified cells were incubated at 37˚C for 15 minutes, followed by 10 minutes at 95˚C, and then thermal-cycled for 35 cycles (Appendix II). Emulsions were broken by spinning the reaction (20,000xg) for 15 minutes, removing the top oil phase, adding 750 μL chloroform, vortexing vigorously, and finally phase separating in a phase lock tube (5Prime). The aqueous phase was cleaned using a Promega PCR purification column. Products of the CPR amplification were visualized using agarose gel electrophoresis.

**Proteomics analysis**

Cells harboring pathways or empty integration vectors were grown at 30˚C for 24-48 hrs in selective media and diluted 1:25 for 6 hours. Cell pellets were collected by

centrifugation and given to the Marcotte laboratory for analysis. Dr. Dan Boutz performed all of the proteomics experiments.

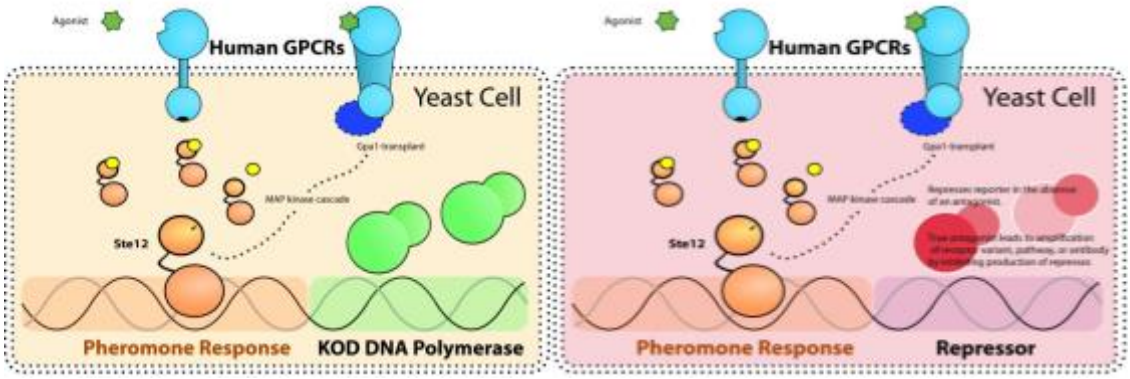
**Metabolite toxicity and growth curves**

BY4741 transformed with a spacer sequence at the URA3 locus were grown for 48 hours and diluted 1:25 in media containing 4 mM of each metabolite (hexanoate, olivetol, CBGA, and CBD). Measurements of cell density (OD600) were taken in 10 min intervals. Each condition was tested in triplicate. For hexanoate growth conditions of aromatic prenyltransferase variants, transformed cells were seeded for 48 hrs and diluted 1:25 in 96-well grow-blocks in media containing 1-5 mM hexanoate. Measurements of cell density were taken in 24 hr intervals using an F500 plate reader. Final screening was performed with 10 mM hexanoate, 1% acetate supplemented synthetic complete media in 50 mL culture volumes.

**Metabolite extraction and analysis**

Supernatant and cell pellets were extracted with 40 mL ethyl acetat, dried under vacuum and resuspended in 250 µL ethanol or methanol. To decarboxylate metabolites, resuspensions were incubated at 65˚C for 24 hrs on a tabletop incubator. 200 µL of the decarboxylated product was then run on a C18 column equilibrated with 25% water, 75% ACN, 0.1% TFA. Metabolites were eluted along a gradient with 100% CAN, 0.1% TFA. Standards were run after samples. We recorded the full spectrum between 0-600 nm and monitored 228 nm for CBD production. Chromatograms were produced and analyzed within the Beckman System Gold 32 Karat software package.

Figure 6-1: Agonist and antagonist CPR circuits




Left, agonist CPR. Ligand induces the production of DNA polymerase, which is then used in emulsion PCR. Right, antagonist CPR circuit. Agonism of the receptor leads to the production of a repressor, which then in turn inhibits the production of a thermostable DNA polymerase for use in emulsion PCR. Only when the receptor is blocked is it possible to produce polymerase.

Table 6-1: G-alpha subunits

| Name | Strain Code | Terminal AA |
|---|---|---|
| Gpa1p | 00 | KIGII |
| Gαq | 01 | EYNLV |
| Gαs | 02 | QYELL |
| Gαo | 03 | GCGLY |
| Gαi2 | 04 | DCGLF |
| Gαi3 | 05 | ECGLY |
| Gαz | 06 | YIGLC |
| Gα12 | 07 | DIMLQ |
| Gα13 | 08 | QLMLQ |
| Gα14 | 09 | EFNLV |
| Gα16 | 10 | EINLL |

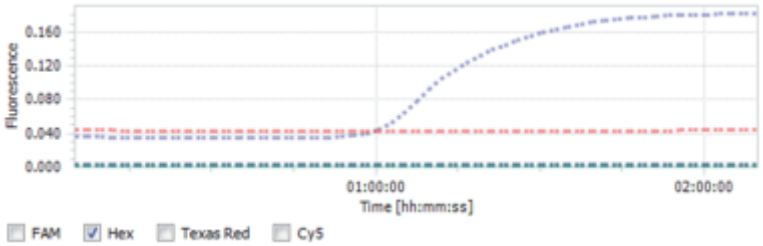
Table 6-2: sgRNA oligos used for genomic knockouts

| Oligo | Sequence |
|---|---|
| EG.sgRNA.Far1.A.F | GACTTTACCAAGTTTGAAGAAAACAG |
| EG.sgRNA.Far1.A.R | AAACCTGTTTTCTTCAAACTTGGTAA |
| EG.sgRNA.Far1.B.F | GACTTTCCACCGAAGAAATTTCTAAG |
| EG.sgRNA.Far1.B.R | AAACCTTAGAAATTTCTTCGGTGGAA |
| EG.sgRNA.Far1.C.F | GACTTTATAGGCTTGGAAAGATTCAG |
| EG.sgRNA.Far1.C.R | AAACCTGAATCTTTCCAAGCCTATAA |
| EG.sgRNA.Sst2.A.F | GACTTTCAGTTCATCAGTAGAGGTGA |
| EG.sgRNA.Sst2.A.R | AAACTCACCTCTACTGATGAACTGAA |
| EG.sgRNA.Sst2.B.F | GACTTTTCAGGAAGCTATTAAAGCAA |
| EG.sgRNA.Sst2.B.R | AAACTTGCTTTAATAGCTTCCTGAAA |
| EG.sgRNA.Sst2.C.F | GACTTTACATAAATAATGTAAGAAGG |
| EG.sgRNA.Sst2.C.R | AAACCCTTCTTACATTATTTATGTAA |
| EG.sgRNA.Ste2.A.F | GACTTTACATAAATAATGTAAGAAGG |
| EG.sgRNA.Ste2.A.R | AAACCCTTCTTACATTATTTATGTAA |
| EG.sgRNA.Ste2.B.F | GACTTTTCAGTGACTTACGCTCTCAC |
| EG.sgRNA.Ste2.B.R | AAACGTGAGAGCGTAAGTCACTGAAA |
| EG.sgRNA.Ste2.C.F | GACTTTTTGCAACTCATCGAAAGTGA |
| EG.sgRNA.Ste2.C.R | AAACTCACTTTCGATGAGTTGCAAAA |

Table 6-3: sgRNA knockout efficiency

| Gene | Guide | KO | % effective sgRNA per gene | total % effective sgRNA |
|---|---|---|---|---|
| Ste2 | A | n | 67% | 67% |
| Ste2 | B | **KO** | | |
| Ste2 | C | **KO** | | |
| Sst2 | A | n | 33% | |
| Sst2 | B | n | | |
| Sst2 | C | **KO** | | |
| Far1 | A | **KO** | 100% | |
| Far1 | B | **KO** | | |
| Far1 | C | **KO** | | |

Figure 6-2: Representative qPCR detection of barcodes

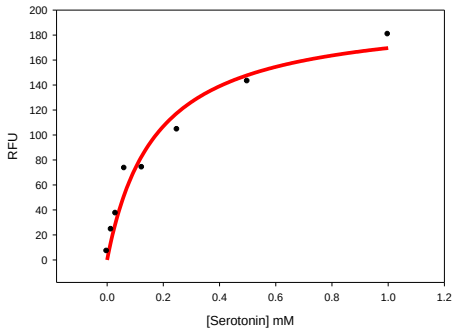



Blue line represents sequence-verified strain containing a disrupted Ste2. Redline is the unaltered BY4741. qPCR could be used to distinguish between the wildtype sequence and a disrupted gene.

Figure 6-3: Inverter circuit reporter design


Inverter circuit reporter is depicted. We inserted the construct into the Leu2 site by supplying an auxotrophic marker. The reporter, zsGreen1, is driven by a modified Gal1 promoter. Two tetO sites are located downstream of the TATA box and prevent transcription when bound by tetR.

Figure 6-4: 5HT1A dose-response to serotonin




Dose-response of JG05-2.0 expressing human 5HT1AR from a 2 micron plasmid. Each point is the average of replicates. EC50 = 171.6 µM.

Table 6-4: GPCR strains constructed in this work

| CODE | Parental | Resistance/Aux | Alterations |
|---|---|---|---|
| JG00-0.0 | BY4741ΔGal2 | G418 | ΔFar1-KO, ΔSte2-KO, ΔSSt2-KO |
| AM00-0.0 | BY4741 | None | ΔFar1-KO, ΔSte2-KO, ΔSSt2-KO |
| JG00-1.0 | JG00-0.0 | G418 | ΔpFig1-KOD |
| JG00-0.1 | JG00-0.0 | G418 | ΔpFus1-KOD |
| JG00-2.0 | JG00-0.0 | G418 | ΔpFig1-zsGreen1 |
| JG00-0.2 | JG00-0.0 | G418 | ΔpFus1-zsGreen1 |
| JG00-3.0 | JG00-0.0 | G418 | ΔpFig1-tetR |
| JG01-0.0 | JG00-0.0 | G418 | ΔGpa1-hGαq |
| JG01-1.0 | JG01-0.0 | G418 | ΔpFig1-KOD |
| JG01-2.0 | JG01-0.0 | G418 | ΔpFig1-zsGreen1 |
| JG01-3.0 | JG01-0.0 | G418 | ΔpFig1-tetR |
| JG02-0.0 | JG00-0.0 | G418 | ΔGpa1-hGαs |
| JG02-1.0 | JG02-0.0 | G418 | ΔpFig1-KOD |
| JG02-2.0 | JG02-0.0 | G418 | ΔpFig1-zsGreen1 |
| JG02-3.0 | JG02-0.0 | G418 | ΔpFig1-tetR |
| JG03-0.0 | JG00-0.0 | G418 | ΔGpa1-hGαi2 |
| JG03-1.0 | JG03-0.0 | G418 | ΔpFig1-KOD |
| JG03-2.0 | JG03-0.0 | G418 | ΔpFig1-zsGreen1 |
| JG03-3.0 | JG03-0.0 | G418 | ΔpFig1-tetR |
| JG04-0.0 | JG00-0.0 | G418 | ΔGpa1-hGαo |
| JG04-1.0 | JG04-0.0 | G418 | ΔpFig1-KOD |
| JG04-2.0 | JG04-0.0 | G418 | ΔpFig1-zsGreen1 |
| JG04-3.0 | JG04-0.0 | G418 | ΔpFig1-tetR |
| JG05-0.0 | JG00-0.0 | G418 | ΔGpa1-hGαi3 |
| JG05-1.0 | JG05-0.0 | G418 | ΔpFig1-KOD |
| JG05-0.1 | JG05-0.0 | G418 | ΔpFus1-KOD |
| JG05-2.0 | JG05-0.0 | G418 | ΔpFig1-zsGreen1 |
| JG05-0.2 | JG05-0.0 | G418 | ΔpFus1-zsGreen1 |
| JG05-2.2 | JG05-2.0 | G418 | ΔpFig1-zsGreen1, ΔpFus1-zsGreen1 |
| JG05-3.0 | JG05-2.0 | G418 | ΔpFig1-tetR |
| JG06-0.0 | JG00-0.0 | G418 | ΔGpa1-hGαz |
| JG06-1.0 | JG06-0.0 | G418 | ΔpFig1-KOD |

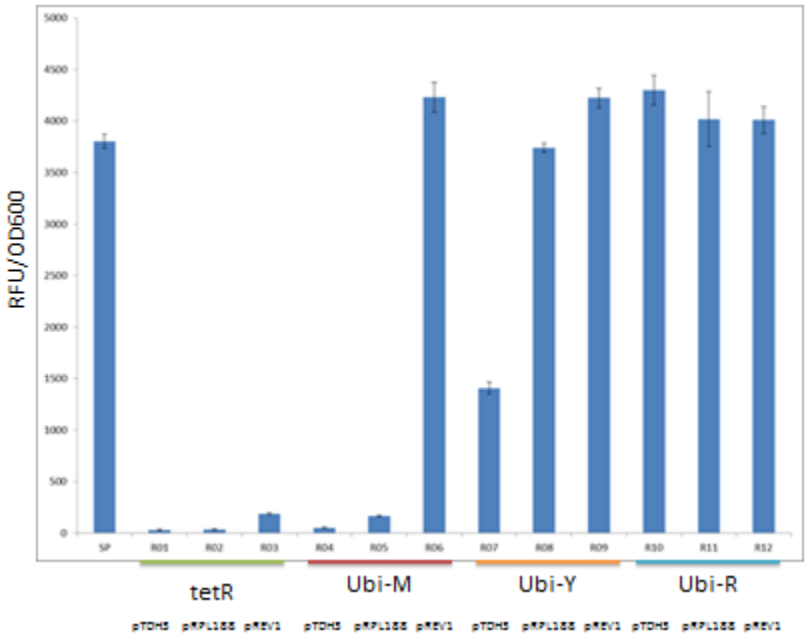
Table 6-4 Cont.

| | | | |
|---|---|---|---|
| JG06-2.0 | JG06-0.0 | G418 | ΔpFig1-zsGreen1 |
| JG06-3.0 | JG06-0.0 | G418 | ΔpFig1-tetR |
| JG07-0.0 | JG00-0.0 | G418 | ΔGpa1-hGα12 |
| JG07-1.0 | JG07-0.0 | G418 | ΔpFig1-KOD |
| JG07-2.0 | JG07-0.0 | G418 | ΔpFig1-zsGreen1 |
| JG07-3.0 | JG07-0.0 | G418 | ΔpFig1-tetR |
| JG08-0.0 | JG00-0.0 | G418 | ΔGpa1-hGα13 |
| JG08-1.0 | JG08-0.0 | G418 | ΔpFig1-KOD |
| JG08-2.0 | JG08-0.0 | G418 | ΔpFig1-zsGreen1 |
| JG08-3.0 | JG08-0.0 | G418 | ΔpFig1-tetR |
| JG09-0.0 | JG00-0.0 | G418 | ΔGpa1-hGα14 |
| JG09-1.0 | JG09-0.0 | G418 | ΔpFig1-KOD |
| JG09-2.0 | JG09-0.0 | G418 | ΔpFig1-zsGreen1 |
| JG10-0.0 | JG00-0.0 | G418 | ΔGpa1-hGα16 |
| JG10-1.0 | JG10-0.0 | G418 | ΔpFig1-KOD |
| JG10-2.0 | JG10-0.0 | G418 | ΔpFig1-zsGreen1 |
| JG10-3.0 | JG10-0.0 | G418 | ΔpFig1-tetR |

Figure 6-5: GPCR CPR proof-of-concept with 5HT1A

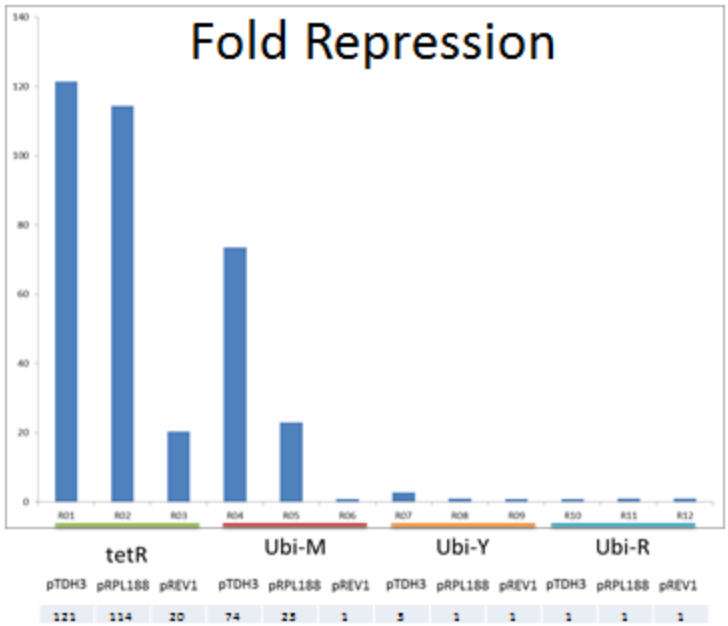
Initial proof-of-concept GPCR-CPR. Human 5HT1A was constitutively expressed from a single-copy integration vector inserted at the HO site. Lane 1 is a 1 kb ladder. Lane 2 is the amplification product of cells treated with 250 μM serotonin. Lane 3 represents the negative control. There is a 20-fold increase in signal with cells treated with the ligand.

Figure 6-6: Inverter circuit proof-of-concept



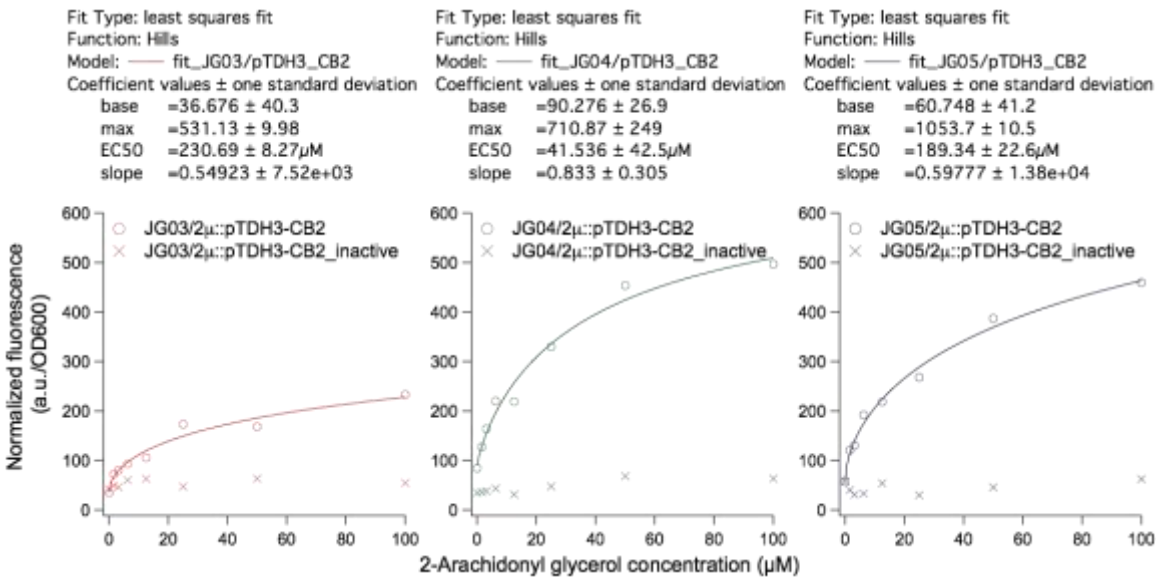

Normalized fluorescent measurements were taken after 18 hrs of galactosidase induction. SP refers to a spacer and does not contain any form of tetR. Ubi-M/Y/R are degradation tags added to the N-terminus of tetR. They are ordered in range of strength. Promoters driving the expression of tetR variants are listed below. They are arranged in order of high to low.

Figure 6-7: Inverter circuit fold repression



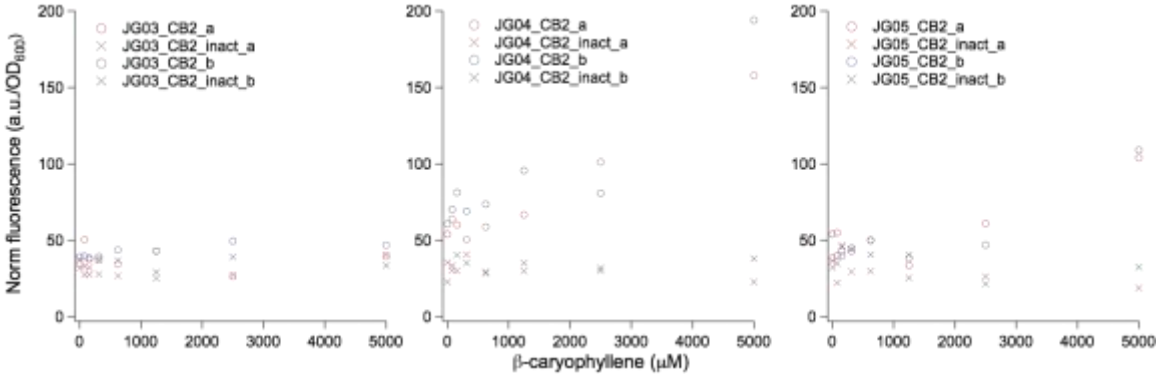

Fold-repression calculated from figure 6-5. Numeric values are listed below. No detectable repression is found when using the fast or high degradation tag on tetR.

Figure 6-8:CB2 dose response with native ligand



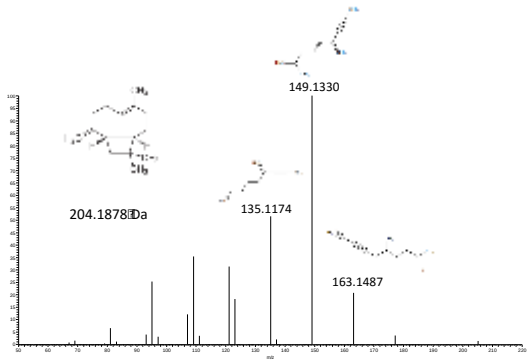

Dose-response of CB2 to 2-arachidonyl glycerol. Each point on the graph represents the average of replicate measurements. Strain JG04-2.0 was selected to be the preferred strain due to the superior EC50.

Figure 6-9**:** CB2 β-caryophyllene dose-response




Dose-response of CB2 to β-caryophyllene. Each point on the graph represents the average of replicate measurements. We later found that our standard was mostly a degradation product, possibly explaining the large amounts of β-caryophyllene needed to activate the receptor.

Figure 6-10: MS ANALYSIS OF β-caryophyllene production



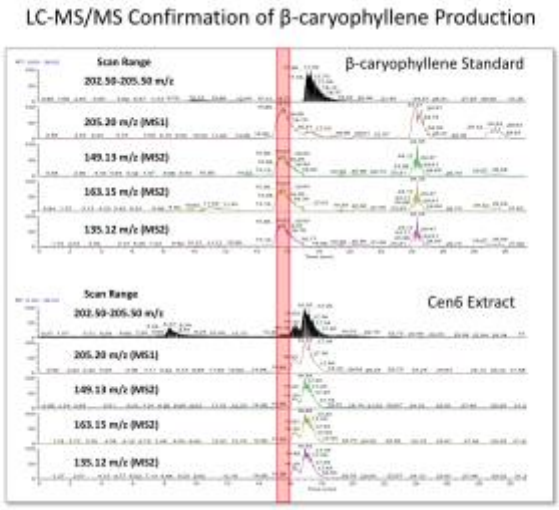
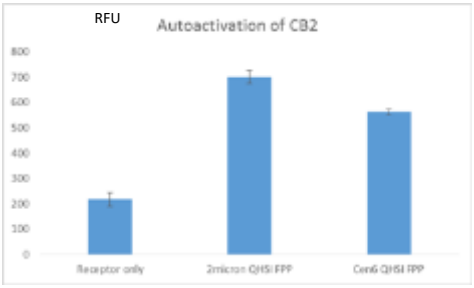
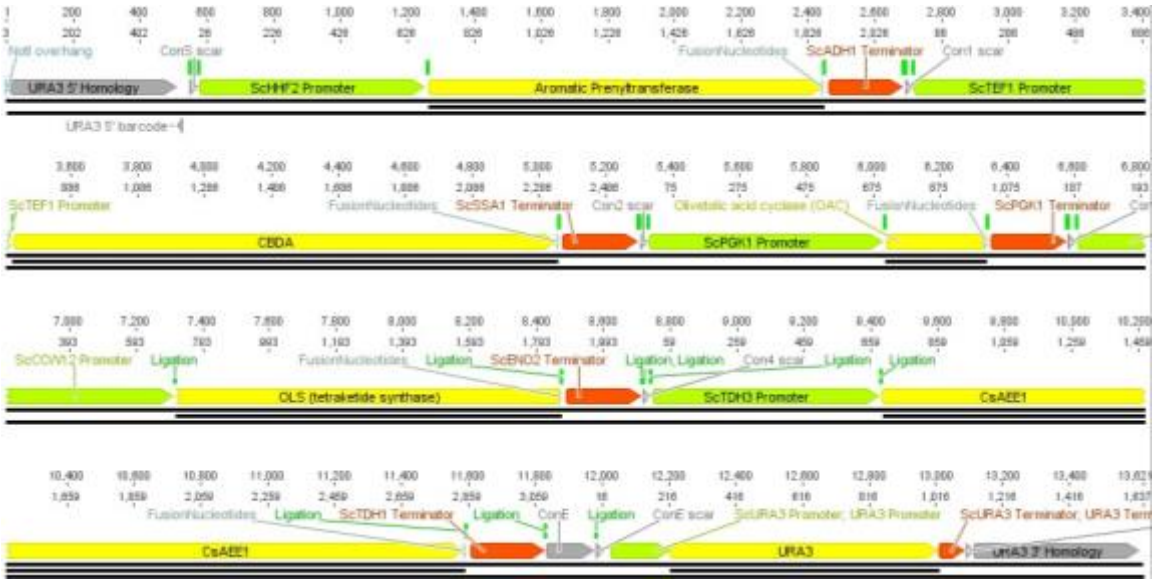



Figure 6-11: Autoactivation of CB2 with β-caryophyllene producing strains



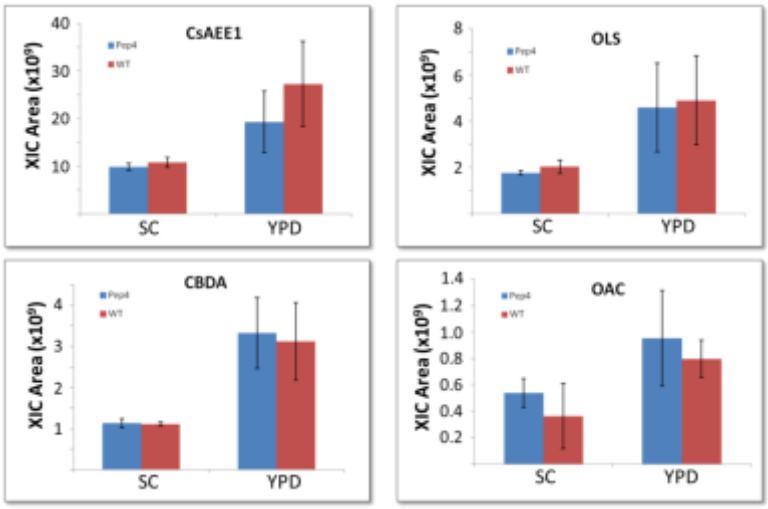

Autoactivation of the CB2 receptor by beta-carophyllene production in the same strain. QHSI was introduced on 2 micron and CEN6 origins and CB2 stimulation was read by ZsGreen reporter. Error bars represent the standard deviation of replicate measurements.

Figure 6-12: CRC integration vector




Cannabinoid reaction center (CRC). The 5 pathway enzymes are shown in yellow. Promoters are depicted in green. Terminators are shown in orange.

Figure 6-13: Proteomics analysis of pathway enzymes

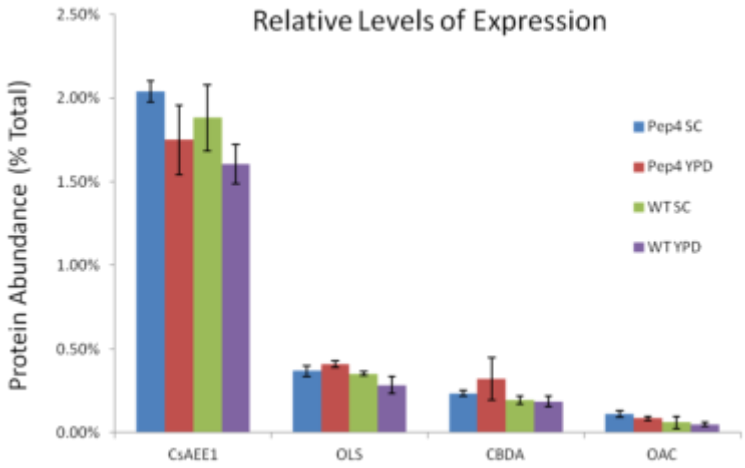



Spectral counts of each of the soluble CBD pathway enzymes are shown. Enzymes were expressed in BY4747 (wt) and ΔPep4 strains. We also measured protein abundances in cells grown in synthetic complete (SC) or rich (YPD) media.

Table 6-5: Published kinetics of CBD enzymes

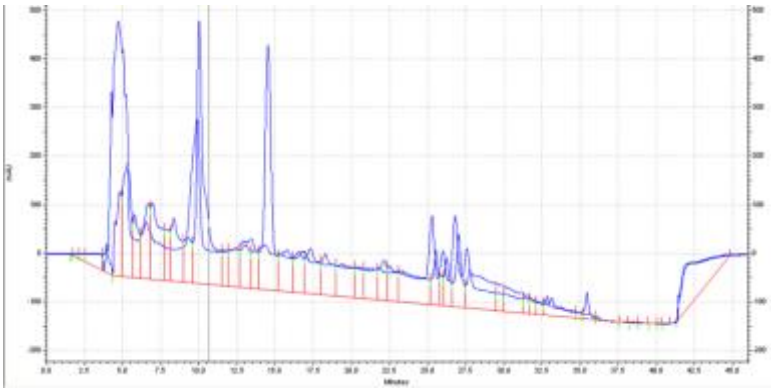
| EC # | Enzyme | Constitutive | kcat | Km | kcat/Km | Substrate |
|---|---|---|---|---|---|---|
| | CsAAE1( O-succinylbenzoate-CoA ligase) Hexanoyl CoA Synthase TKS/OLS | pTDH3 | 2.0 s-1 | 3.7 + 0.7 mM | 540.5 s-1 M-1<br>1013 s-1 M-1 | Hexanoic acid + CoA |
| 2.3.1.206 | 3,5,7-trioxododecanoyl-CoA synthase (OLS) | pCCW12 | 2.96 (min-1) | .0608 (mM) | 811 s-1 M-1 | Hexanoyl-CoA |
| | | | 1.06 (min-1) | .0631 (mM) | 280 s-1 M-1 | Hexanoyl-CoA |
| | | | 3.46 (min-1) | .0569 (mM) | 1013 s-1 M-1 | Hexanoyl-CoA |
| 4.4.1.26 | olivetolic acid cyclase (TKS) | pPGK1 | None available | None available | None available | 3,5,7-trioxododecanoyl-CoA |
| 2.5.1.102 | CsPTI (Aromatic prenyltransferase) geranyl-pyrophosphate-olivetolic acid geranyltransferase | pHHF2 | None available | 2 mM | None available | Olivetolic acid* Mg2++ |
| 1.21.3.8 | cannabidiolic acid synthase (CBDA synthase) | pTEF11 | 0.19 (s-1) | .137 (mM) | 1386 s-1 M-1 | Cannabigerolic acid |
| 1.21.3.7 | tetrahydrocannabinolic acid synthase (THCA synthase) WT | | .201 (s-1) | .134 (mM) | 556 s-1 M-1 | Cannabigerolic acid |

Figure 6-14: Protein abundances by strain



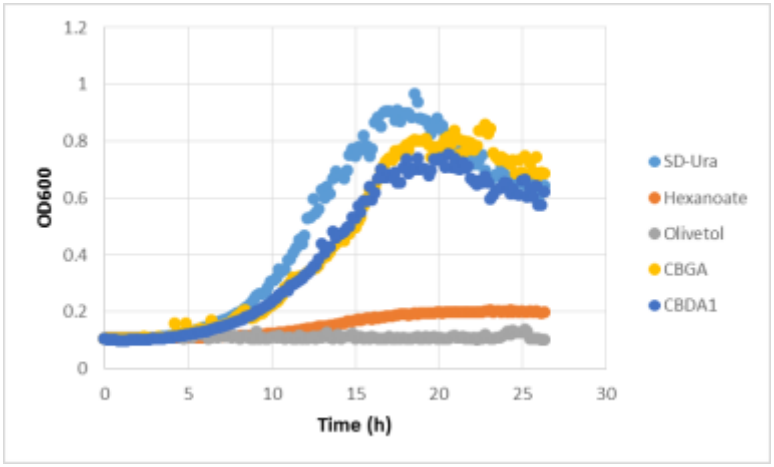

Relative abundance of each of the soluble CBD pathway enzymes are shown. Enzymes were expressed in BY4747 (wt) and ΔPep4 strains in both SC and YPD. Percentages are relative to total protein abundances from their respective strain and growth condition.

Figure 6-15: HPLC of CBD from supernatant and cell pellet extractions



Total ethyl acetate extractions from the supernatant as well as cell pellet are shown. CBD is primarily found in the cell pellet extraction (14.9 min peak).

Figure 6-16: Metabolite toxicity determined by growth assays

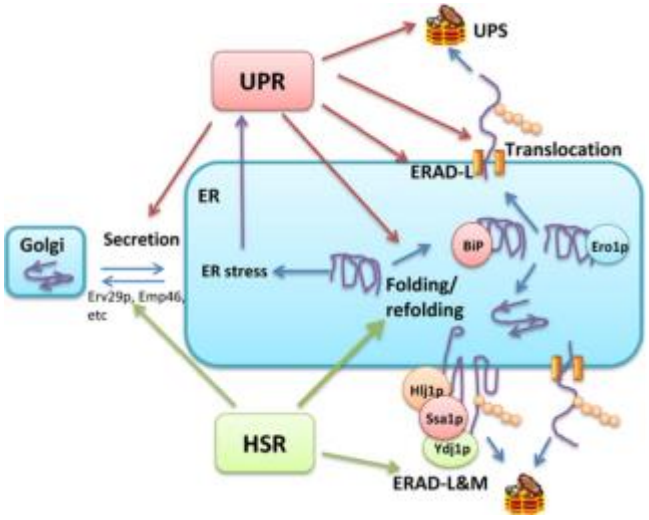



BY4741 with a spacer sequence inserted into the URA3 locus were grown in 4 mM of each intermediate. Hexanoate and olivetol severely inhibit growth. CBGA and CBD were far less inhibitory when compared to growth in synthetic media only (SD-URA).

Figure 6-17: Predicted chloroplast targeting sequence of aromatic prenyltransferase

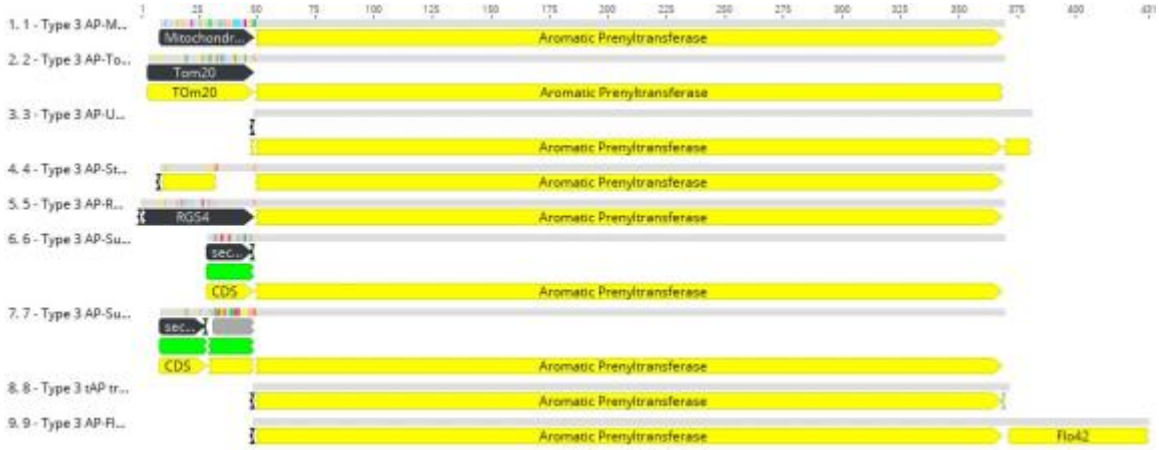

The SignalP and ChloroP-predicted signal sequence is depicted in black. The two common aromatic prenyltransferase active site loops are highlighted in pink.

Figure 6-18: Unfolded protein response observed in cells with pathway



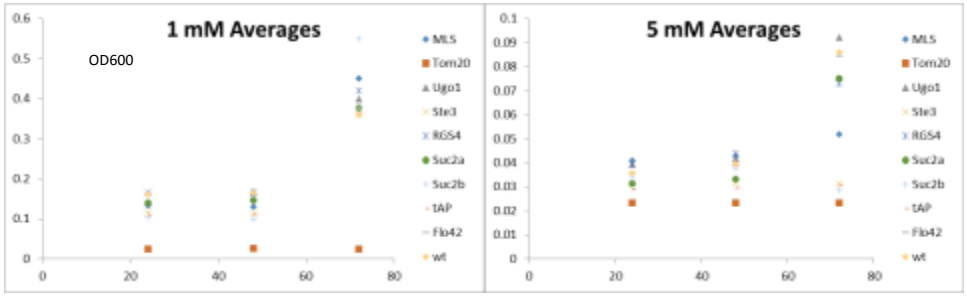

| Gene ID | Z-combined | Ave FC |
|---|---|---|
| KAR2 | 11.69 | 1.78 |
| ERO1 | 6.97 | 2.62 |
| PDI1 | 6.04 | 1.52 |
| MCD4 | 5.81 | 7.72 |
| LHS1 | 5.25 | 2.27 |
| SIL1 | 4.90 | 3.86 |
| YPS1 | 3.51 | 5.13 |
| MPD1 | 3.24 | 3.99 |
| DOG2 | 3.23 | 4.50 |
| SCJ1 | 3.22 | 1.99 |
| NCE103 | 3.11 | 1.51 |
| YMR315W | 2.97 | 1.55 |
| YIL108W | 2.88 | 2.14 |
| EUG1 | 2.68 | 1.83 |

Figure 6-19: Rational design of aromatic prenyltransferase variants



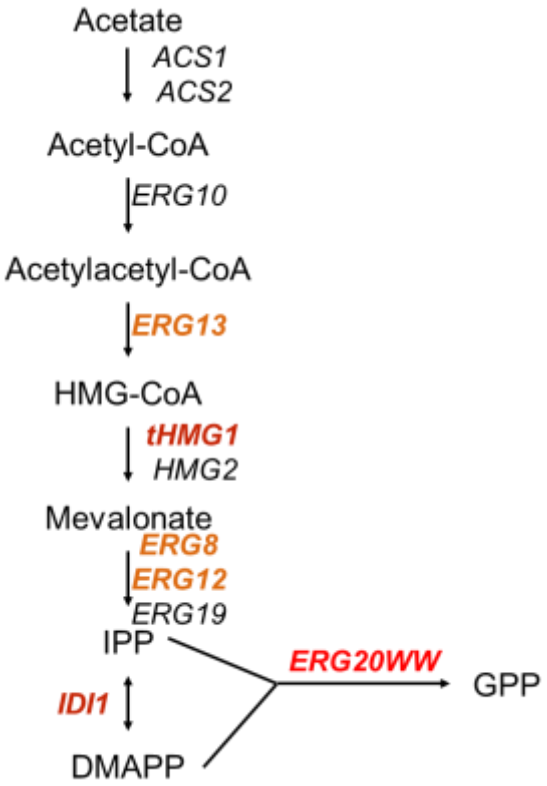

Aromatic prenyltransferase variants are depicted. For each of the designs, the predicted chloroplast targeting sequence was removed. 1-3 represent mitochondrial variants. 4-7 target the plasma membrane. 8 is the wildtype without the signal peptides, and 9 contains a C-terminal domain with the last 42 amino acids of Flo1p. This domain has been shown to anchor to the plasma membrane.

Figure 6-20: Aromatic prenyltransferase variants alone do not confer robust growth in hexanoic acid




Growth of aromatic prenyltransfersae variants in the presence of 1mM (left) and 5 mM (right) hexanoate-supplemented media. Each point represents the average growth of 3 independent colonies. Time points were taken in 24 hr intervals.

Figure 6-21: GPP pathway

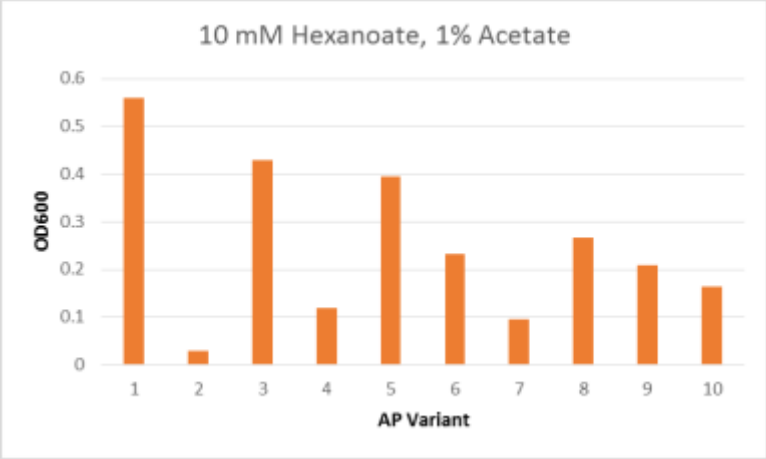



The genes involved in GPP production from acetate in yeast are depicted. Additional enzymes added to CBD strains are depicted in red. Erg20WW is a mutant form of the native Erg20 and has been shown to primarily produce GPP. Enzymes shown in orange are upregulated with the addition of the mutant UPC2-1 transcription factor.

Figure 6-22: GPP integration vector

Integration vector constructed to increase GPP. Pathway enzymes are shown in yellow. Promoters are green, and terminators are denoted in orange.

Figure 6-23: Increasing GPP enables robust growth in the presence of hexanoic acid




Measurements were taken after 3 days of growth in media supplemented with 10 mM hexanoic acid and 1% acetate. The variants are numbered the same as Figure 6-16. Variant 10 is the unaltered wildtype aromatic prenyltransferase.

# Chapter 7: Towards A Robust Ionotropic Activator for Brain-wide Manipulation of Neuronal Function


## ABSTRACT

This chapter outlines methods for the rational design, directed evolution, and characterization of novel neuronal actuators based on the purinergic P2X receptor. The highly original and comprehensive engineering scheme outlined here brings together the various themes explored in previous chapters and will one day yield a new set of transformative tools for functional brain analysis. For nearly a decade, analysis of brain circuitry has relied on methods that allow neuronal activity to be perturbed in an intact brain with cell type-specificity. Genetically-encoded neural actuators have ranged from chimeric G-protein coupled receptors (GPCRs) with orthogonal ligands to light-gated ionotropic channels. While these tools have helped uncover cellular substrates of cognitive and behavioral states, significant limitations remain. Optical fiber implantation is destructive, and illumination is limited by mechanical constraints and the requirement that the target site be identified in advance. GPCRs are often inefficient, display poor temporal control, and often produce long-term functional changes in neurons. The goal of this chapter is to develop the computational and directed evolution methods needed to engineer a neuronal activator that embodies the strongest features of existing approaches, ligand-dependent gating of a channel with high unitary conductance and a negligible desensitization profile. We discuss the computational work, the functional expression of P2X subtypes in yeast, and validation of CPR selection components that will eventually lead to the engineering of orthogonal receptor:ligand pairs. We also discuss the development of a potential drug screening platform using yeast to express P2X receptor subtypes.

## INTRODUCTION

The ability to manipulate defined neuron populations has revolutionized *in vivo* investigations of brain circuitry (Alexander et al., 2009; Boyden et al., 2005; Zemelman et al., 2002). The main requirement for a genetically encoded neuronal actuator is that it normally be absent from the central nervous system, so that only neurons expressing the actuator are sensitized to an orthogonal ligand added in trans. While many actuators that stimulate or silence neurons have been devised, each has its own set of drawbacks, providing a significant opportunity for improvement. Building on lessons learned from modeling and directed evolution, this chapter develops and tests strategies for the engineering of orthogonal receptor:ligand pairs that should one day enable restricted activation of neurons. Based on the P2X nucleotide receptor, this activator will incorporate the strongest features of existing tools, while avoiding many of their limitations.

There are two varieties of actuators: (1) pharmacogenetic, such as metabotropic G-protein coupled receptors (GPCRs), and (2) optogenetic, mostly ligand-gated ion channels. Hence, actuator triggers have been either small molecules or pulses of light. For in vivo use, light delivery is inefficient and requires potentially destructive optical fiber implantation. In this respect, pharmacogenetic tools have an advantage over optogenetic ones: (1) neurons of interest may be distributed throughout the brain, making optical methods impractical; and (2) cell locations may not be known ahead of time. However, while systemically administered small molecule ligands can theoretically reach sensitized neurons throughout the brain, they must cross first cross the blood-brain barrier. In addition, GPCRs are more difficult to regulate, but can be induced for hours; meanwhile, ionotropic channels offer precise temporal control, but are inadequate for long-term

activity management either due to rapid desensitization or because the required ligand is metabolized. Currently, the most widely used actuator is channelrhodopsin2 (Chr2), a light-gated algal cation channel (Nagel et al., 2003). The tradeoff for its very rapid on/off kinetics is poor single channel conductance (estimated at 150 fS) and fast desensitization. These features have limited the use of Chr2 for prolonged or localized stimulation, such as of axonal fibers and terminals.

The functional mapping of neuronal circuits in the brain can be performed anatomically or behaviorally. For anatomical circuit analysis, it is necessary to choose an initial population of neurons within a brain region of interest and sensitize those neurons to activation. Genetically encoded activators that are otherwise absent from the central nervous system can be targeted to a selected subset of neurons using spatially confined viral infection and cell type-specific promoters. When the neurons are subsequently stimulated, their post-synaptic partners can be identified on the basis of elevated activity, using *post hoc* staining for elevated immediate-early gene (IEG) products (Bahrami and Drabløs, 2016), or in transgenic animals where IEG activation is coupled to the synthesis of a fluorescent protein tag. In both instances the pre- and post-synaptic neurons become labeled. Behavioral circuit mapping relies on a change in animal performance (DiBenedictis et al., 2017; Krakauer et al., 2017). Stimulation or silencing of cells in an anatomical domain produces a measurable behavioral outcome. When the manipulation is regional, akin to a lesion, cell types contributing to task performance cannot be readily deduced. When the activity of defined cell types is perturbed and a behavioral phenotype detected, inferences can be made about cell function.

In both cases, experimental success is predicated on the efficiency of cellular perturbations. In direct circuit mapping, post-synaptic neurons will not be detected if the stimulus is too weak. Likewise, a behavioral change may not be apparent when the target

cell population is only partially affected. As a result, pharmacogenetic methods have a slight advantage over optogenetic ones: (1) light guides produce spatially restricted illumination, potentially missing many cells of interest, and can damage neuronal connections—both drawbacks for behavioral analyses of cell function; (2) where the experimental readout is optical or the behavioral assay incorporates a visual component, light stimulation can confound outcomes; and (3) optical methods are often hampered by low conductances and rapid desensitization. Likewise, pharmacogenetic techniques that rely on neuromodulatory second messenger cascades can increase the sensitivity of pre-synaptic neurons to stimulation, but may not reliably activate their post-synaptic partners. Finally, for both pharmacological and optical methods, control over multiple, potentially intermingled cell populations has proven elusive.

## APPROACH

A family of pharmacogenetic ionotropic activators addresses the enumerated limitations listed above. The absence of a light requirement ensures comprehensive stimulation of sensitized neurons irrespective of their location; the gating of a non-specific cation channel produces rapid membrane depolarization followed by sustained cell firing; and the availability of multiple actuators that can be used in parallel provides experimental access to several cell populations.

An activator based on the P2X receptor represents the ideal alternative for robust and selective stimulation of neurons. P2X channels support currents of 10-20 pS (exceeding that of Chr2 by nearly 100 fold). Stimulation is tunable, controlled by ligand concentration. Several members of this receptor family (P2X2 and P2X4) display negligible desensitization. Nucleotide analogs can be administered systemically or fitted

with a photolabile caging group to enable localized activation using visible light (Zemelman et al., 2003). Their stability can also vary, regulating stimulus duration. Since P2X receptors are abundant in the central nervous system, it is necessary to design synthetic variants of P2X, based on published crystal structures, along with matching ligands. These modified P2X receptors will be unresponsive to endogenous nucleotides, but instead be controlled by the orthogonal small molecule ligands. Fundamentally, three main tasks must be accomplished to effectively control neurons with the resulting orthogonal receptor-ligand pairs: (1) validation of our computational methodology in the design of receptors and agonists; (2) optimization of conditions for the directed evolution of receptors; and (3) ensure orthogonality of our engineered channels and small molecule ligands before functionally testing them in an organism.

This chapter outlines the computational work, the functional expression of P2X subtypes in yeast, and validation of selection circuits that will eventually lead to the engineering of orthogonal receptor:ligand pairs. Using structural information of P2X4, candidate mutagenesis sites and designs for complementary ligands are discussed. By extending the directed evolution technique pioneered in our laboratory, Compartmentalized Partnered Replication (CPR), to yeast, we now have a platform to engineer orthogonal purinergic receptor-ligand pairs in parallel with our computational approach. Using alternating rounds of positive and negative selection, we envision libraries that converge on a small number of unique solutions for each new ligand.

The goal of this chapter is to develop orthogonal P2X receptor-ligand pairs using novel computational and directed evolution methods. While computational methods for protein (Miklos et al., 2012; Smith et al., 2014; Tiwari et al., 2012) and small molecule (Rosales-Hernández and Correa-Basurto, 2015; Stanley and De Fabritiis, 2015; Verma and Prabhakar, 2015) engineering have existed for well over a decade, the strategies

typically focus on one component, either a drug-like molecule or the protein itself. For example, medicinal and computational chemists have developed virtual tools for in silico screening of compounds against targets of interest and algorithms for guiding small molecule design based on structure-activity relationships (Doupnik et al., 2015; Giacomini et al., 2015). Quantitative structure activity relationship (QSAR) data are then used to build numerical models for predicting and interpreting experimental results. Refinement by computational fragment-based drug design has also garnered experimental support for use with receptors and enzymes (Böhm et al., 1999; Mueller et al., 2002; Stürzebecher et al., 1995). Finally, techniques to generate molecules with improved drug-like properties have recently emerged (Pierce et al., 2004). This work compounds these methodologies for the development of drug-like molecules to engineered receptors using computational and directed evolution methods (**Figure 7-1**).

In drug development, the refinement of ligand binding through target modification is impossible because targets cannot be altered. In the present case, however, both the ligand and receptor are amenable to engineering. Protein design tools have benefited from better, more robust force fields and scoring methods (Barnoud and Monticelli, 2015; Vanommeslaeghe and MacKerell, 2015). In conjunction with quantum mechanical geometry optimizations (Cho et al., 2005; Fu et al., 2011) and molecular dynamics simulations, protein-by-design strategies are widely displaying their effectiveness. Methodologies laid out below melds protein and ligand design strategies together to create a system to engineer truly orthogonal receptors and tailor-made small molecules.

## RESULTS & DISCUSSION

### Structure-based design approach

An iterative process of computational design was employed to design orthogonal P2X receptor-ligand pairs. Using crystal structures of the zebrafish P2X4 in both bound-open (**Figure 7-2**) and unbound-closed states and homology models of the other receptor subtypes, a fragment-based design approach (Guarnieri, 2015; Ludington, 2015) was utilized to design receptors and corresponding ATP-analogs using a suite of computational tools described below. While the ATP binding motif of P2X family receptors is not otherwise found in nature (**Figures 7-3 & 7-4**), the U-shaped binding orientation is only crudely observed in Type II amino-acyl tRNA synthetases (Hattori and Gouaux, 2012). Additionally, several agonists/antagonists have been identified for each of the receptor subtypes (P2X1-7) (Dal Ben et al., 2015; North, 2002). Together these facts offered a unique opportunity for engineering custom receptors to ATP-analogs and the potential to design novel drug-like molecules to engineered receptors.

#### *Computational design of receptors*

Using the Molecular Operating Environment (MOE09.2014) suite of protein design and medicinal chemistry tools, NAMD Scalable Molecular Dynamics (Phillips et al., 2005), and resources at the Texas Advanced Computing Center, we computationally derived amino acid substitutions that enhance binding to one of a number of ATP-analogs (**Figure 7-5**). As a proof-of-concept and training of our design process, we designed 20 P2X4 receptor variants to bind to 2-Chloro-ATP (2-Cl-ATP) (**Figure 7-6**), 20 receptors that bind to 2'3'-O-(4-benzoyl-benzoyl)- adenosine 5'-triphosphate (BzATP), and an additional 20 receptors that bind adenylyl-imidodiphosphate (AmpPNP), a non-

hydrolysable analog. The sets of designs independently focused on the three structural motifs bound by the receptor (base, sugar and phosphate) in an effort to minimize computational load and potentially allow design modularity. *In silico* selections were scored on binding energy, stability of the mutations in relation to the protein, and the overall potential energy of the agonist bound in the open state using an all atom force field with reaction field treatment of electrostatics. Candidate receptors were refined by rotamer exploration (Lovell et al., 2000), low-mode molecular dynamics (Labute, 2010b), and nanosecond timescale full molecular dynamics simulations in a 6 Å explicit solvent cage. Representative data for an in silica selection against 2-Cl-ATP are shown in **Figure 7-7**.

***In silico* screening of ATP-analogs and in cleft design of drug-like molecules**

Homology models of each of the human and zebrafish P2X receptor subtypes were used in template-based docking studies to ensure the orthogonality of receptor-ligand pairs. To this end, we compiled a database of ATP-analogs that can be used to virtually screen candidate molecules for binding using a template-forced screening methodology, which restricts conformational rearrangements of docked ligands to those more closely matching empirically derived structures. Binding energies of these analogs were calculated for each of the receptor subtypes. Common motifs that resulted in productive binding found across multiple receptor types are eliminated from the design process. Unique compounds or rare functional groups can be identified using this process, and we anticipate the ability to create tailored molecules to abolish binding to wild-type (unaltered) receptors by disrupting hydrogen bonding networks and introducing additional, sterically unfavorable transformations in the cleft. Once candidate receptor-ligand pairs have been identified and validated, we will replace the scaffold of the ATP-

analog with components that are more drug-like in nature (Maass et al., 2007; Stewart et al., 2006). Further refinement of receptor design at this stage should create the first orthogonal P2X receptors and provide molecules that are usable *in vivo* (**Figure 7-8**).

## Directed evolution of orthogonal receptor-ligand pairs

### *Selection scheme overview*

The directed evolution approach to engineer receptor-ligand pairs builds on a recent selection method pioneered in our lab, Compartmentalized Partnered Replication (CPR). In this selection scheme, the function of a genetic circuit drives the production of a DNA polymerase during an *in vivo* stage. After the initial expression of the circuit, individual cells are compartmentalized in emulsions and thermal cycled in the presence of nucleotides and primers specific to the component of interest. Libraries of individual components of the circuit are easily made by site saturation or error-prone PCR; variants that produce the most DNA polymerase will subsequently be more efficiently amplified (**Figure 7-9**). Following multiple rounds of selection, the most active variants predominate, and can be further analyzed by sequencing and assay. The power of CPR as a selection tool derives from the decoupling of host fitness from circuit function and its ability to screen $>10^8$ variants in a single day. The production of thermostable polymerase in vivo is short-lived, negating fitness effects that hamper other selection methodologies, and the PCR amplification of functional genes allows for the ability to quickly and efficiently turn multiple rounds of selection. To date, CPR has been used to engineer: tRNAs, amino-acyl tRNA synthetases, orthogonal T7 RNA polymerase variants (Ellefson et al., 2014), transcription factors, and components of biosynthetic pathways (in preparation).

***Positive Selection Design & Construction***

Positive circuits were engineered in *S. cerevisiae* to generate DNA polymerase upon binding designed agonists from a semi-synthetic promoter, consisting of 4 tandem repeats of a 24-bp calcium-dependent response element (CDRE), a small fragment of the FKS2 promoter, joined to a truncated version of the Cyc1 promoter (Stathopoulos and Cyert, 1997). In the presence of agonist, polymerase is made in response to receptor function, which will in turn allow emulsion-based amplification of functional receptor variants (**Figure 7-10**). Functional testing of the CDRE using a fluorescent reporter showed calcium dependent activation (**Figure 7-11**).

***Negative Selection Design & Construction***

To select against constitutively active receptors, activation by ATP, the native ligand, or other off-target molecules, we designed and built a counterselection circuit where channel opening in the absence of the target molecule results in death of the cell. We designed a URA3-deficient strain of yeast and replaced the polymerase in the construct outlined above with a functional URA3 under the control of a slightly weaker 2xCDRE synthetic promoter. Cells are then be incubated in rich media with the addition of 5- Fluoroorotic acid (5-FOA). Off-target or constitutively active receptors will drive the expression of URA3, which converts 5-FOA to 5-Fluorouracil (5- FU), a highly toxic compound (**Figure 7-12**). Modulation between rounds of positive selection (CPR) and a life/death negative selection will allow for the selection of highly specific receptor variants to target molecules.

***Functional Expression and Testing of P2X Subtypes in S. Cerevisiae***

We started by expressing wt-hP2X4 receptors under the control of TDH3 and Gal promoters in wildtype S. cerevisiae (BY4741) and a ΔCne1p mutant, which has been shown to mitigate endoplasmic reticulum-mediated degradation of heterologously expressed transmembrane proteins (Prinz et al., 2003). By fusing Citrine, a fluorescent reporter, to the carboxy terminus of the receptor, we were able to observe a modest increase in expression using the ΔCne1p mutant (**Figure 7-13**). However, microscopy revealed large differences in localization patterns between the two strains. As predicted, the wild-type BY4741 cells were unable to traffic the receptor out of the ER, presumably explaining the modest decrease in signal during our expression testing.

Our results indicated that the receptor escaped the ER in the ΔCne1p mutant; however, the receptor was largely trapped within vacuoles or pits, leaving very few, if any, at the plasma membrane—likely due to tyrosine-mediated endocytosis. At this point we learned that the carboxy terminus of hP2X4 harbors a well-characterized non-cononical endocytic motif (YXXGθ) (Royle et al., 2002), which is recognized by specific machinery in mammalian systems. Somewhat unsurprisingly, the clatherin-associated receptor recycling system is conserved in yeast (Chapa-y-Lazo et al., 2014; Weinberg and Drubin, 2012). These facts led us to truncate the C-terminus of the receptor by 11 and 14 residues (ΔC11, ΔC14), which has been shown to increase plasma membrane localization in certain mammalian cell types without substantial effects on gating (Royle et al., 2005). Confocal microscopy revealed hP2X4ΔC14 was more efficiently trafficked and largely absent in the vacuoles compared to the wildtype receptor (**Figure 7-14**).

To determine whether or not this truncated channel gated in yeast, we utilized an aequorin-based functional assay. Here aequorin, a bioluminescent protein, emits a photon

upon binding calcium. This is done with the aid of coelenterazine, which must be added exogenously to yeast. The reaction is not catalytic, and the off-rate of calcium is too slow to observe on the timescales utilized in our assays. To optimize conditions for expression of both the receptor and reporter, we built and tested a combination of integration, low-copy, and high-copy vectors. At this point, it was only possible to produce a signal upon addition of ATP, the native ligand, when both the receptor and reporter were expressed on high-copy plasmids (**Figure 7-15**).

While encouraging, the amount of activation was small and unlikely to trigger the CDRE. Visual inspection of the primary sequence revealed that the truncated receptor now contained cannonical C-terminal KKXX ER retention signals (Gaynor et al., 1994). In an effort to increase the amount of receptor at the plasma membrane, we engineered a version with the ER export signal from Kir2.1 (FCYENEV) (Stockklausner et al., 2001). Confocal microscopy verified that the receptor largely escaped intracellular aggregation. Functional testing of this variant expressed from a single-copy integration vector revealed almost a full order of magnitude increase in signal when compared to the truncated version of the receptor (**Figure 7-16**).

Next, we built all 7 P2X subtypes and integrated them into the genome of BY4741ΔCne1p under the control of the GAL1 promoter. After verification of the integrations, we transformed these strains with high-copy plasmids expressing aequorin. Interestingly, many of the receptors were shown to be functional in our assays when expressed from single-copy integration vectors (**Figure 7-16**). Exceptions include P2X3, P2X5, and P2X6. For P2X3, this negative result may be a result of poor trafficking in yeast or desensitization of the receptor. P2X3 desensitizes rapidly, and our assay may not be sensitive enough to detect gating of this receptor. P2X5 and P2X6 were not expected

to produce signals. In humans, these receptors are typically found in heterotrimeric configurations with other subtypes (Lê et al., 1999; North, 2002).

Surprisingly, the ligand-dependent signal generated by P2X2 was several orders of magnitude higher than any of the other subtypes. This is more than likely due to efficient trafficking and a lack of desensitization of the receptor to ATP. Of the 7 subtypes, P2X2 is considered to be 'non-desensitizing' and displays the highest single-channel conductance of the P2X receptor family (North, 2002). Due to the superior signal observed in our aequorin assay, we decided to perform a full kinetic analysis to determine if the receptor expressed in yeast matched data from mammalian hosts. In an ATP dose-response assay, we calculated the $EC_{50}$ to be 6.7±1.5 μM—perfectly matching reported values (**Figure 7-17**) (Lynch et al., 1999). Next, we characterized the P2X2 dose-response to 2-Cl-ATP (**Figure 7-18**) and BzATP (**Figure 7-19**) and found that the receptor was agonized by these compounds with nearly identical $EC_{50}$ values (26.7±3.6 and 29.1±6.0 μM, respectively) and in good agreement with published reports (Coddou et al., 2011; Jarvis and Khakh, 2009). Of particular interest, BzATP has been previously shown to lower the magnitude of response while increasing the time the channel is open (less desensitizing). Our plate-based aequorin assay recapitulated this phenomenon. The amplitude of the BzATP generated signal is 3 orders of magnitude lower than either ATP or 2-Cl-ATP, and the duration that the channel remains open is almost 10-fold longer in our assays. AMP-PNP, a non-hydrolyzable ATP analog, did not gate P2X2. This negative result was expected because AMP-PNP has not been reported to activate P2X2. Suramin is a known P2X2 antagonist. To determine if our assay could detect inhibition, we equilibrated yeast cells expressing P2X2 with varying concentration of suramin. We found that suramin inhibited P2X2 with an $IC_{50}$ of 31.6±1.4 μM (**Figure 7-20**), well within the range of reported values (2-48 μM) (Baqi et al., 2011; Lambrecht et al., 2002).

Finally, a non-functional double-mutant (K81C, K83C) (Roberts et al., 2008) was shown to be completely non-functional in our assay (**Figure 7-21**), giving us confidence that our results were robust and adequately accounted for known P2X2 properties.

To our knowledge, this marks the first direct evidence of a human neural receptor gating in yeast. We believe this plate-based aequorin assay could be used in high-throughput drug discovery platforms. We are particularly interested in how P2X2 is able to gate much more efficiently than the other subtypes. It may prove possible to build chimeric receptors with the N- or C-termini of P2X2 affixed to the extracellular and transmembrane portions of the other receptor subtypes. Doing so may allow more robust signal generation of the other receptors. It may also be useful to construct known 'desensitizing' mutations of the other receptors and test their activity using our yeast-based assay.

***P2X2 selections***

Having built and proofed the individual components of the selection scheme, libraries of P2X2 are now under construction. To initially change the ligand specificity of P2X2 receptors, we saturated positions in the ATP binding pocket and will select for gating to AMP-PNP. Those circuits that can respond most readily to AMP-PNP should be readily identified within a few cycles of emulsion and amplification in the presence of primers specific for the P2X2 receptor. Another avenue to proof the selection methodology outlined above would be to evolve the 'broken' double mutant back to wildtype function using ATP or 2-Cl-ATP.

### *Conclusions*

Here we developed a methodology to engineer robust, orthogonal receptor:ligand pairs. By combining rational design and directed evolution approaches, it should prove possible to engineer both the receptor and ligand. The ability to make complementary modifications in channel and ligand structure will help generate a family of orthogonal receptor-ligand pairs for independent control over multiple cell populations within the brain while eliminating crosstalk with endogenous factors. The strength of this and other pharmacogenetic approaches is that the locations of target neurons need not be known a priori; however, should precise temporal regulation be needed, the ligands could potentially be chemically disabled, enabling brief localized photoactivation. A synthetic purinergic activator (SPArk) will advance functional brain mapping, provide robust control over discrete neuronal populations that represent known neurochemical classes or are selected using pioneering activity-based molecular-genetic methods.

A panoply of tools that will be deployed brain-wide across species to control distinct ensembles of neurons will uncover circuit connectivity and signaling hierarchies. However, just as with existing technologies, much work remains to be done not only to engineer the synthetic receptors, but also to synthesize and screen orthogonal ligands that are well-tolerated, easy to administer, and that readily reach target sites in the brain.

## METHODS & MATERIALS

### *Molecular Modeling of P2X Receptors*

Homology models of human P2X subtypes were constructed using two reference structures (PDB: 4DW1 and 3H9V).The templates were prepared using the Molecular Operating Environment (MOE.09.2015) software package from Chemical Computing Group. The structure was inspected for anomalies and protonated/charged with the

Protonate3D subroutine (310K, pH 7.4, 0.1 M salt). The protonated structure was then lightly tethered to reduce significant deviation from the empirically determined coordinates and minimized using the Amber10:EHT forcefield with R-field treatment of electrostatics to an RMS gradient of 0.1 kcal $mol^{-1}$ $Å^{-1}$. Homology models of the other P2X subtypes were prepared by creating 25 main chain models with 25 sidechain samples at 298K (625 total) within MOE. Intermediates were refined to an RMS gradient of 1 kcal $mol^{-1}$ $Å^{-1}$, scored with the GB/VI methodology, minimized again to an RMS gradient of 0.5 kcal $mol^{-1}$ $Å^{-1}$, and protonated. The final model for each variant was further refined by placing the protein within a 6 Å water sphere and minimizing the solvent enclosed structure to an RMS gradient of 0.001 kcal $mol^{-1}$ $Å^{-1}$. Models were evaluated by calculating Phi-Psi angles and superimposed against the reference structure. Agonist structures were created in cleft, and LigandInteractions subroutine was used to calculate binding energies. ResidueScan was used to identify potential amino acid substitutions for a given analog. Interactions were scored based on stability, affinity, and potential energy of the system.

***Cloning and expression of P2X subtypes in yeast***

Sequences for the wild-type receptors and aequorin were obtained from Uniprot. Amino acid sequences were codon optimized for expression in *S. cerevisiae* and purchased as GeneBlocks from IDT. The Yeast Toolkit was then used for the Golden Gate Assembly of all vectors. Aequorin was expressed under the control of the pPGK2 promoter. P2X variants were expressed under the control of the GAL1 promoter. Fluorescently-tagged (Venus) constructs of receptor variants were also constructed for confocal microscopy and flow cytometry analyses. Flow cytometry was performed by

Michelle Byrom at the Core Facility at the University of Texas at Austin. Plasmid maps and sequences are available upon request.

***Confocal microscopy***

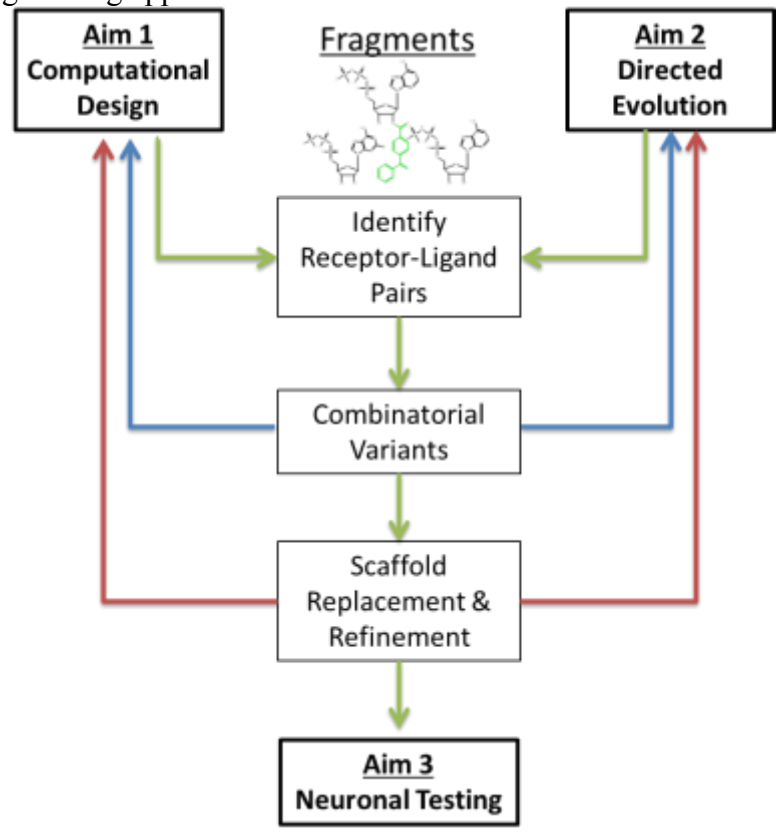
Overnight cultures harboring fluorescently tagged (Venus) receptors were diluted 1:20 in fresh media. After 6 hours of growth, cells were washed 3 times with PBS. A total of 3 mL of cells were concentrated into 100 μL. 5 μL of the concentrated cells were placed on a slide. A coverslip was added and sealed with clear nail polish. Images were taken at the University of Texas at Austin Microscopy center by Dr. Julie Hayes.

***Aequorin assays***

BY4741ΔCne1p cells were transformed with vectors containing P2X variants. Transformants were validated and made competent for a subsequent transformation with a vector containing aequorin under the expression of the constitutive PGK2 promoter. Overnight cells were induced with 2% galactosidase for 20 hours. Following receptor expression, cells were collected by centrifugation and concentrated to 500 μL. The concentrated cells were then incubated with 5 μL of 1 mg/mL coelenterazine for 30 minutes at room temperature. Washed cells were then resuspended in softening buffer (1 M Sorbitol, 25 mM EDTA, 50 mM DTT), followed by washes in 1 M sorbitol and water. Cells were then resuspended in 50 mM MES/Tris pH 6.5 and aliquoted into a 96-well plate. Using an F500, 25 μL of cells were monitored for luminescence for 5 seconds. Upon addition of ligand or buffer (100 μL), cells were monitored for luminescence for varying times. Cell lysis was performed with the addition of 100 μL Tween20. For inhibition, varying concentrations of suramin were incubated with cells. Luminescence

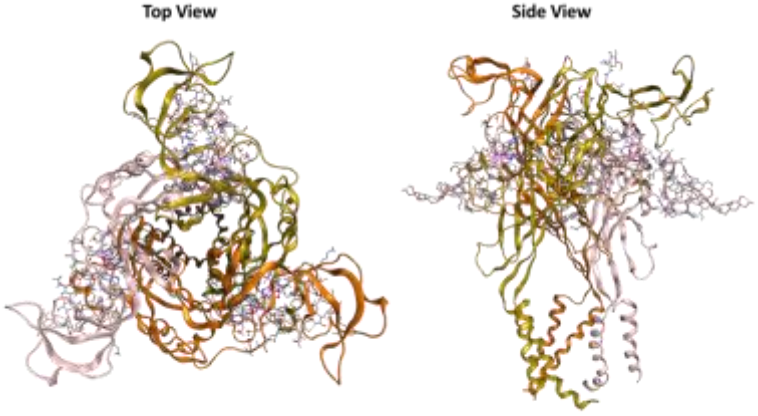
was monitored before and after the addition of 100 μL of 10 μM ATP. Areas under the curve were calculated and plotted against a negative control to determine the IC50.

Figure 7-1: Engineering approach



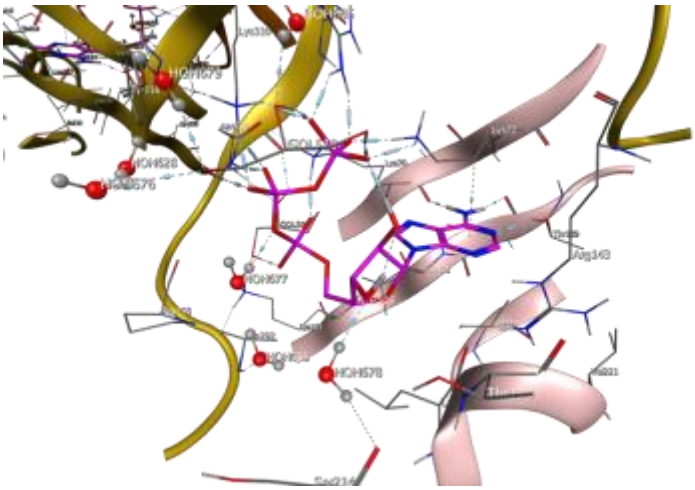

Overview of Orthoganol P2X Receptor-Ligand Engineering. Parallel approaches of rational design and directed evolution are depicted.

Figure 7-2: hP2X4 biomolecule assembly




Biomolecular assembly of the functional P2X4 homotrimer. Individual monomers are depicted in pink, yellow, and orange. Glycosylation sites near the ligand-binding domains are shown using sticks.

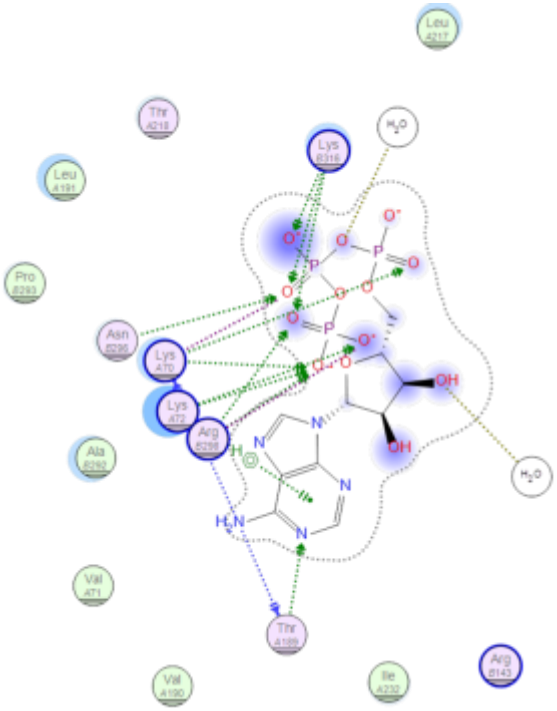
Figure 7-3: P2X4 ATP binding pocket



3D-depiction of a single ATP-binding pocket. Each binding domain rests between two monomer, depicted herein pink and yellow. Ordered water molecules from the crystal structure are shown in ball and stick format.

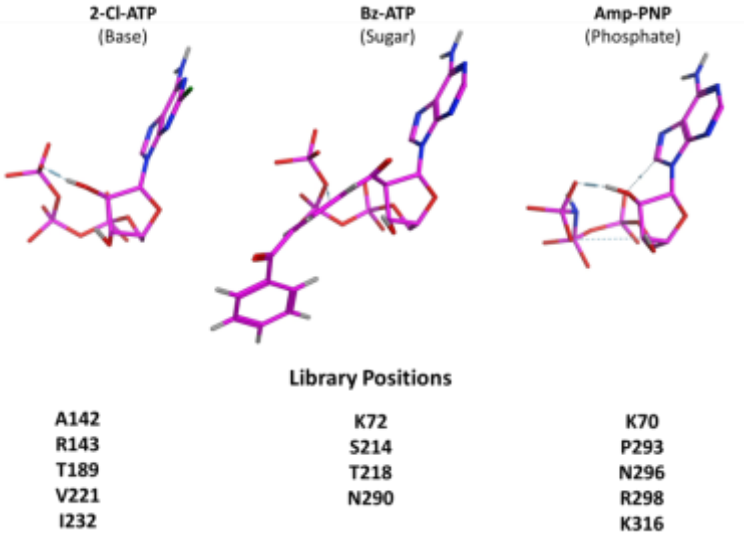
Figure 7-4: Ligand interactions of ATP with P2X4

| Ligand | Receptor | Interaction | Distance | E (kcal/mol) |
|---|---|---|---|---|
| N6 36 | O LYS 70 (A) | H-donor | 2.76 | -2.3 |
| N6 36 | O THR 189 (A) | H-donor | 2.87 | -1.4 |
| O1G 2 | NH2 ARG 298 (A) | H-acceptor | 2.71 | -7.2 |
| O1G 2 | NZ LYS 316 (A) | H-acceptor | 2.69 | -17.8 |
| O2G 3 | NZ LYS 70 (A) | H-acceptor | 2.72 | -20 |
| O2G 3 | NZ LYS 72 (A) | H-acceptor | 2.7 | -20.1 |
| O2G 3 | NH2 ARG 298 (A) | H-acceptor | 2.69 | -7.2 |
| O3G 4 | NZ LYS 72 (A) | H-acceptor | 2.66 | -5.1 |
| O2B 6 | ND2 ASN 296 (A) | H-acceptor | 2.73 | -5.7 |
| O2B 6 | NZ LYS 316 (A) | H-acceptor | 2.82 | -9.3 |
| O1B 7 | NZ LYS 316 (A) | H-acceptor | 2.96 | -9 |
| O1A 10 | NZ LYS 70 (A) | H-acceptor | 2.76 | -15.6 |
| O3A 12 | O HOH 677 (A) | H-acceptor | 2.81 | -1.5 |
| O3' 22 | O HOH 678 (A) | H-acceptor | 2.77 | -1.5 |
| N1 39 | OG1 THR 189 (A) | H-acceptor | 2.76 | -2.9 |
| O1G 2 | NH2 ARG 298 (A) | ionic | 2.71 | -6.7 |
| O1G 2 | NZ LYS 316 (A) | ionic | 2.69 | -6.9 |
| O2G 3 | NZ LYS 70 (A) | ionic | 2.72 | -6.7 |
| O2G 3 | NZ LYS 72 (A) | ionic | 2.7 | -6.8 |
| O2G 3 | NH2 ARG 298 (A) | ionic | 2.69 | -6.9 |
| O3G 4 | NZ LYS 72 (A) | ionic | 2.66 | -7.2 |
| O3G 4 | NH2 ARG 298 (A) | ionic | 3.88 | -0.8 |
| O2B 6 | NZ LYS 70 (A) | ionic | 2.8 | -6 |
| O2B 6 | NZ LYS 316 (A) | ionic | 2.82 | -5.8 |
| O1B 7 | NZ LYS 316 (A) | ionic | 2.96 | -4.7 |
| O1A 10 | NZ LYS 70 (A) | ionic | 2.76 | -6.3 |
| 6-ring | CD LYS 72 (A) | pi-H | 3.56 | -0.7 |



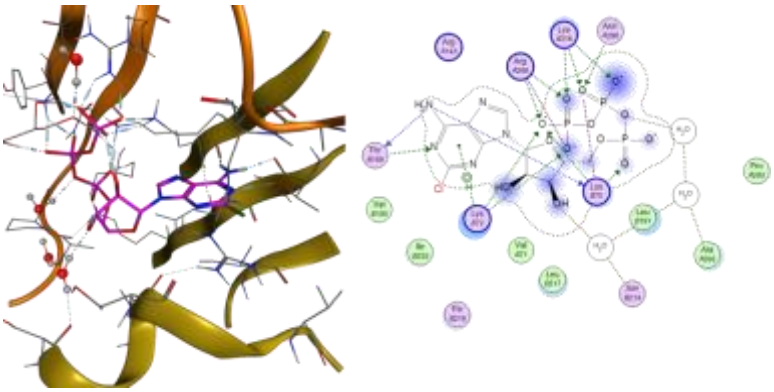
Left, ATP interactions with P2X4 are tabulated. Right, 2D contact map of ATP bound to the wildtype structure. Basic residues are highlighted in blue. Polar contacts are depicted with arrows.

Figure 7-5: Training molecules



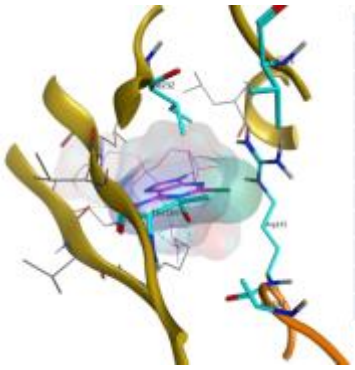

Training molecules and amino acid residues used in the design of P2X4 receptor variants.

Figure 7-6: 2-Cl-ATP binding to P2X4



Left, 3D representation of 2-Cl-ATP in the wild-type P2X4 structure. Right, 2D contact map. Clashing is observed in P2X4.

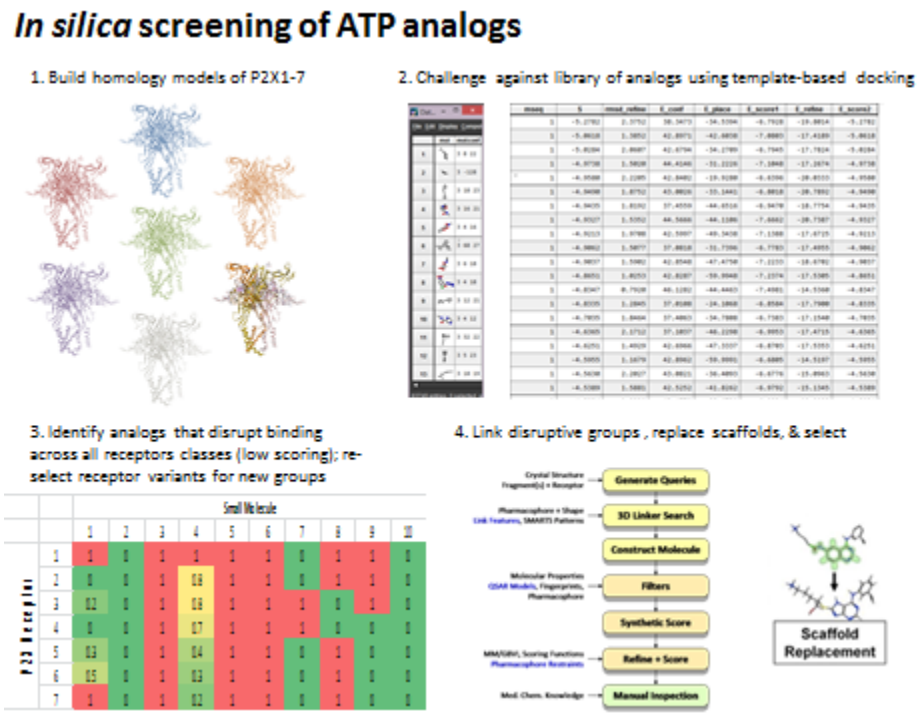
Figure 7-7: In silica selection of P2X4 variants to 2-Cl-ATP



| 2-Chloro-Variant | Affinity | dAffinity | Stability |
|---|---|---|---|
| 1 | -17.0777 | -0.59867 | -74.7449 |
| 2 | -17.0317 | -0.55264 | -77.9894 |
| 3 | -16.9609 | -0.48184 | -75.3053 |
| 4 | -16.932 | -0.45297 | -79.4034 |
| 5 | -16.9021 | -0.42306 | -74.2643 |
| 6 | -16.8768 | -0.39773 | -73.2561 |
| 7 | -16.8547 | -0.37568 | -77.7231 |
| 8 | -16.844 | -0.36494 | -72.8208 |
| 9 | -16.8264 | -0.34731 | -83.6089 |
| 10 | -16.7821 | -0.30302 | -78.637 |

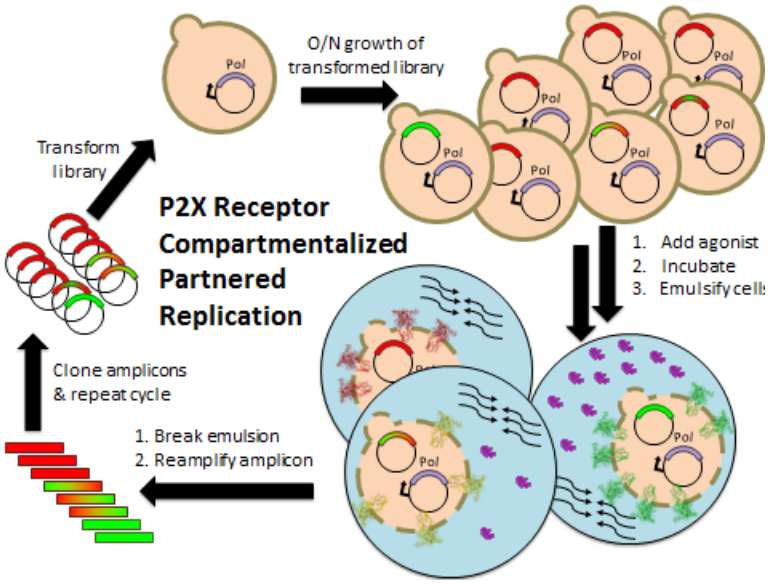
Left, library residues for in silica selection are shown in blue. 2-Cl-ATP is depicted in magenta and with a molecular surface. Right, representative in silica selection data of top 10 computationally derived P2X4 variants.

Figure 7-8: Analog screening methodology



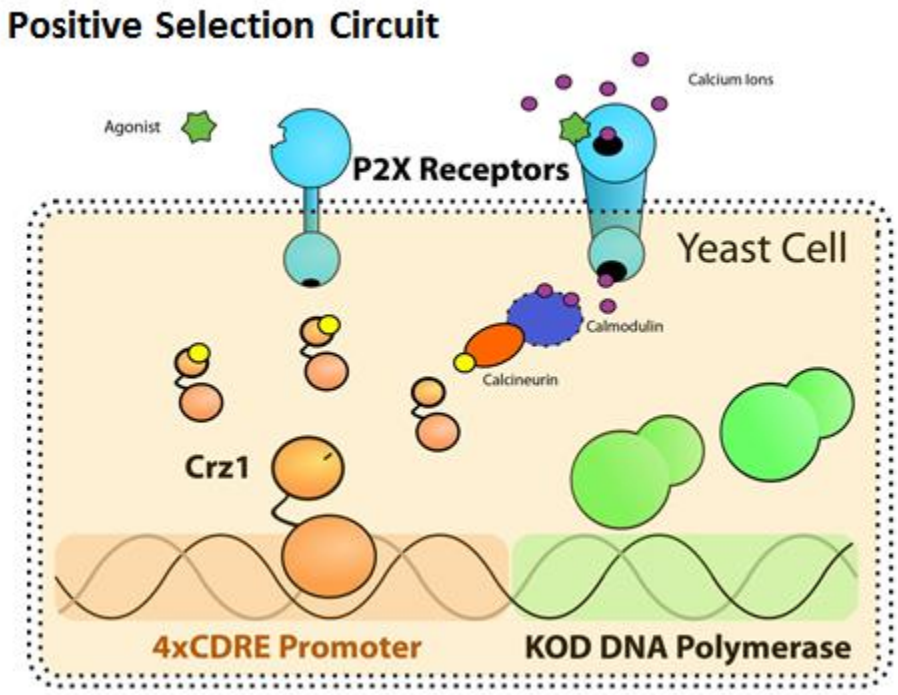

Methodology for selection of orthogonal receptor:ligand pairs. 1. Homology models were constructed for each of the P2X receptor subtypes. 2. Models are screened against a number of small molecules. 3. Small molecules are scored with quantitative structure activity relationships and given scores. 4. Candidate molecules are transformed into druggable molecules through by replacing the scaffold of the selected molecule.

Figure 7-9: yCPR for ion channels



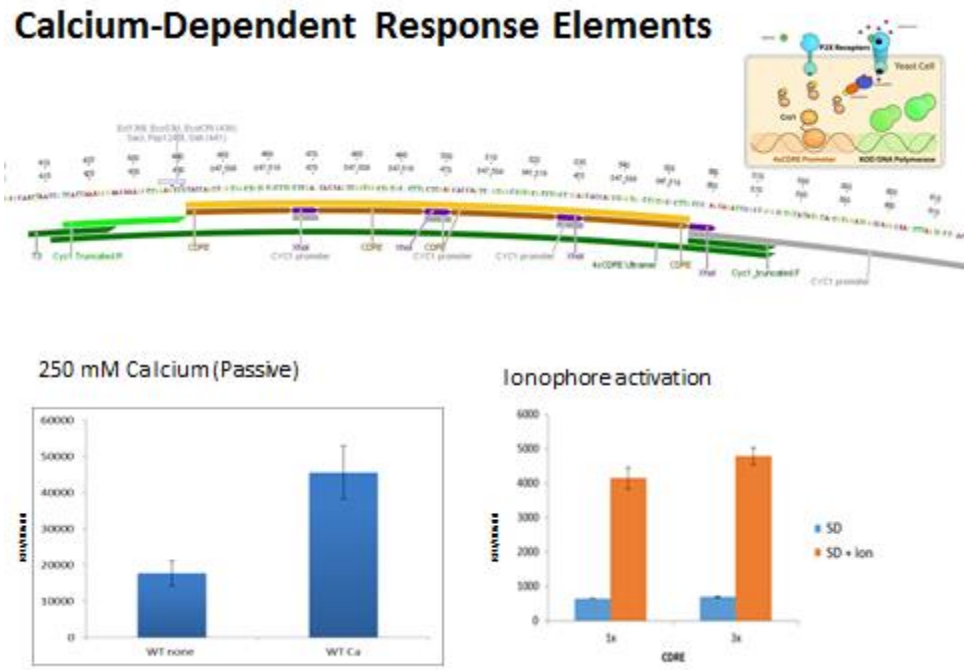

Overview of ionotropic receptor CPR. Libraries of receptors are transformed into yeast and expressed. Variants are challenged with agonist in bulk. Active receptors lead to the production of a thermostable DNA polymerase, used in emulsion PCR. The most active variants produce the most polymerase and thereby resulting in more copies being made during the PCR step. After breaking the emulsion, variants are recloned and subjected to more rounds of selection or assayed individually.

Figure 7-10: Positive selection circuit



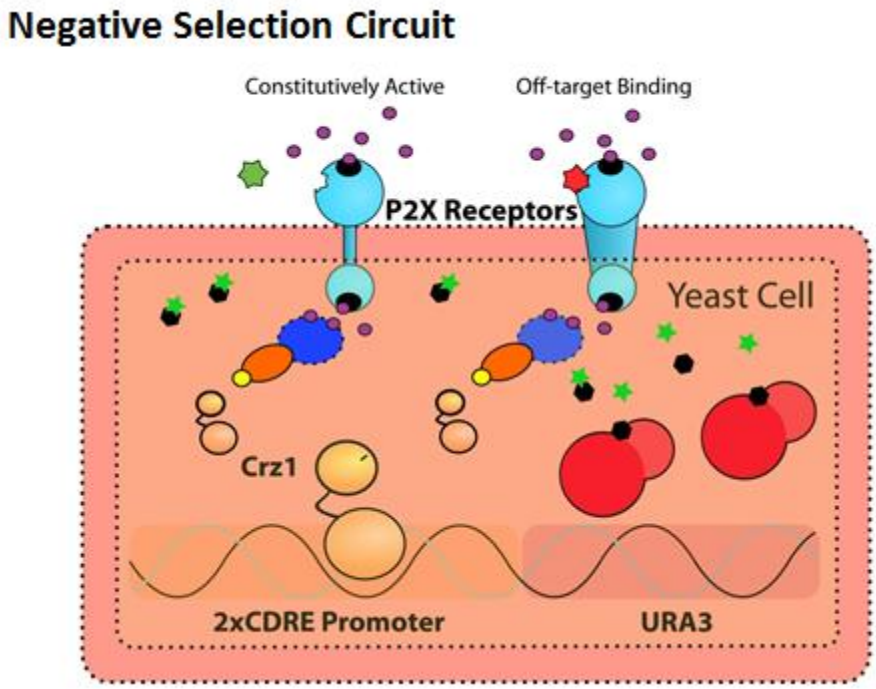

Positive Selection Circuit. DNA polymerase is induced by calcium intake from P2X receptor variants.via the Crz1-dependent pathway. The amount of polymerase produced in each cell is directly influenced by receptor function.

Figure 7-11: Calcium dependent response element characterization



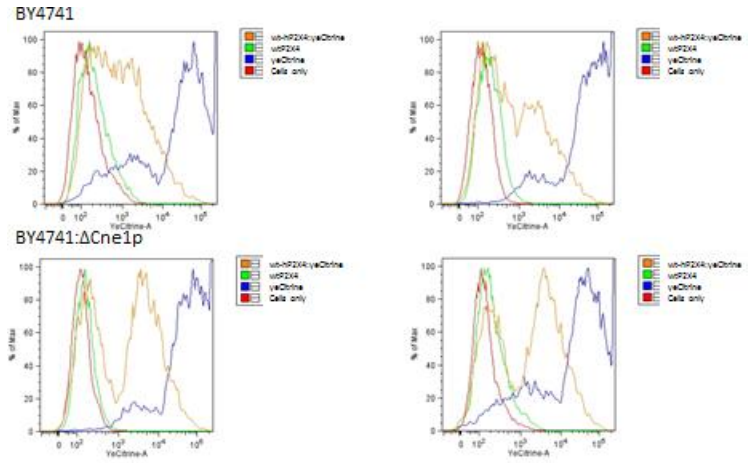

Top, depiction of 4xCDRE-Cyc1 hybrid promoter. Bottom left, passive diffusion of calcium triggers the promoter when compared to a control. Bottom right, the addition of an calcium ionophore triggers expression of a fluorescent reporter under the control of 1x and 3x CDRE hybrid promoters.

Figure 7-12: Negative selection circuit



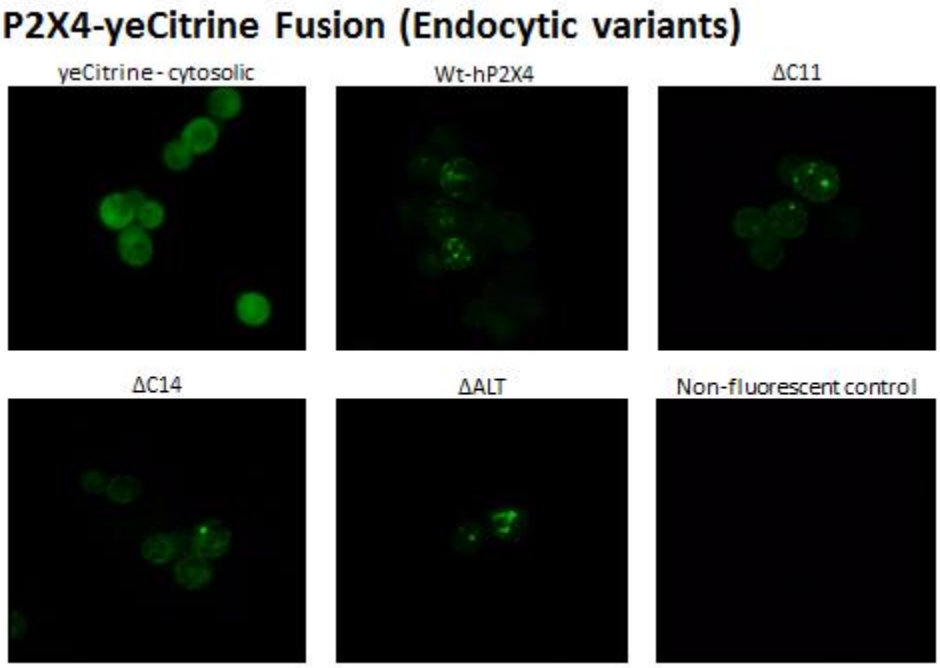

Negative Selection Circuit. URA3 is produced in the presence of constitutively active and off-target binding receptors. URA3 converts 5-FOA to 5-FU and kills the cell.

Figure 7-13: Flow cytometry reveals expression of P2X4 in yeast



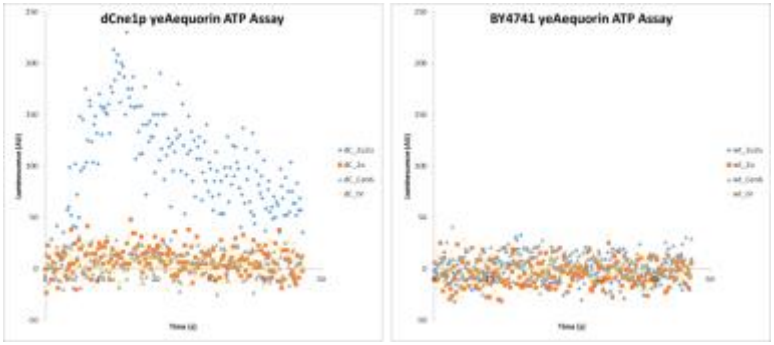

Flow cytometry of low-copy plasmid expression testing of fluorescently –tagged hP2X4 receptors in BY4741 (top) and BY4741ΔCne1p (bottom). See key for traces.

Figure 7-14: Localization of P2X4 in yeast



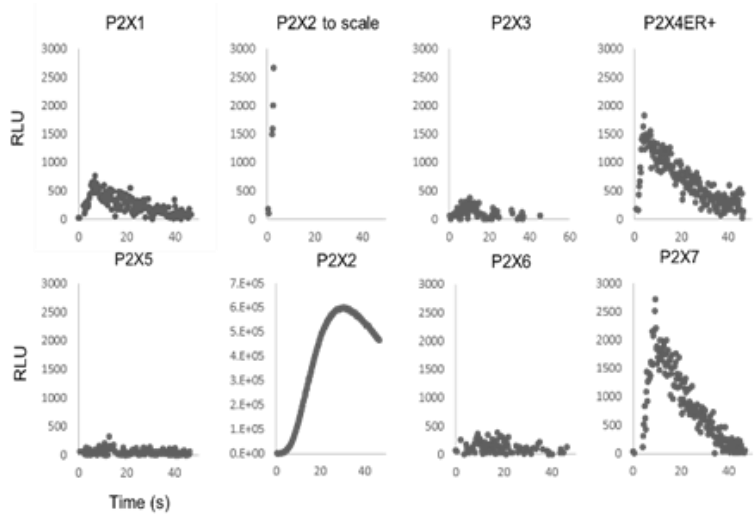

Confocal microscopy of wtP2X4 and endocytic variants. ΔC14 was judged to be largely absent from intracellular vacuoles when compared to the wt receptor.

Figure 7-15: Function assay of P2X4



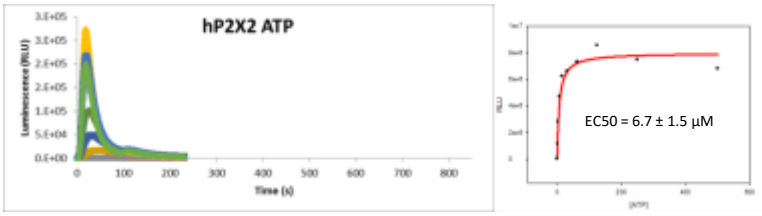

Functional testing of hP2X4ΔC14 using a plate-reader based aequorin assay. Each point represents the average of three independent test. Only when the receptor and aequorin were expressed from 2 micron (high-copy) plasmids could a reproducible signal be observed.

Figure 7-16: Functional assays of all P2X subtypes in yeast



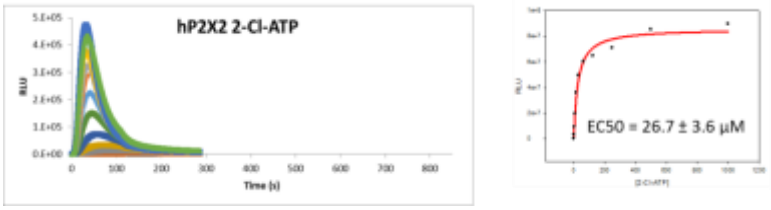

Aequorin assay of all P2X receptor subtypes. P2X4ER+ is the truncated (ΔC14) receptor with the addition of an ER export signal (FCYENEV) to the carboxy terminus. All other receptors are wild-type sequences. Reproducible signals were observed in subtypes 1, 2, 4, and 7.

Figure 7-17: P2X2 dose-response with ATP



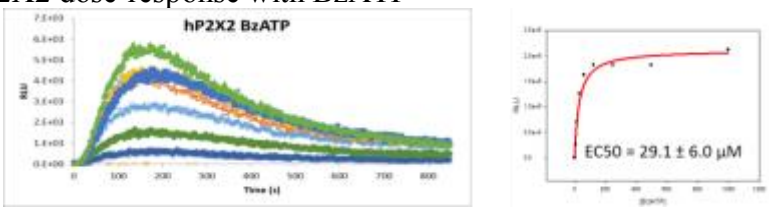

Left, time course of ATP-dependent luminescence of P2X2. Each curve represents a different ATP concentration. Right, area under the curve from the luminescence assay was plotted against concentration and fit.

Figure 7-18: P2X2 does-response with 2-Cl-ATP



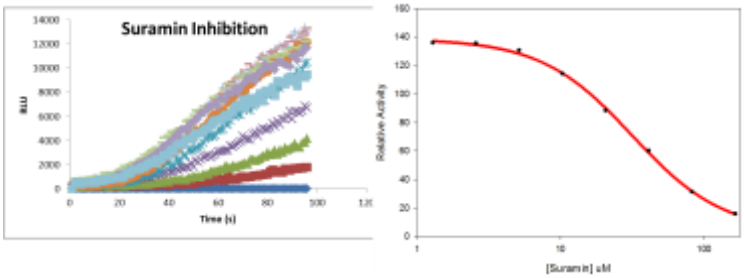

Left, time course of 2-Cl-ATP-dependent luminescence of P2X2. Each curve represents a different 2-Cl-ATP concentration. Right, area under the curve from the luminescence assay was plotted against concentration and fit.

Figure 7-19: P2X2 dose-response with BzATP



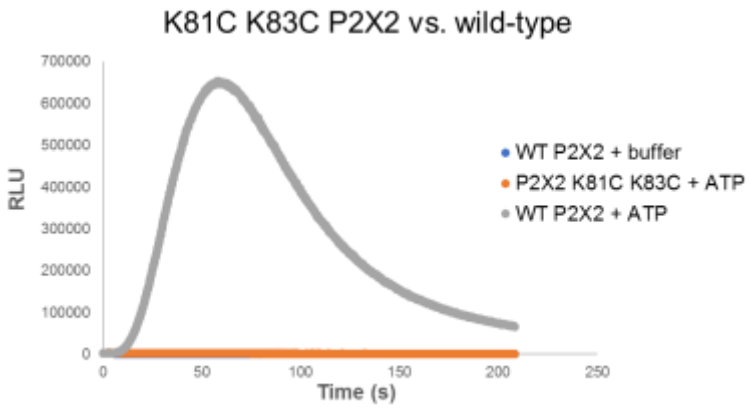

Left, time course of BzATP-dependent luminescence of P2X2. Each curve represents a different BzATP concentration. Right, area under the curve from the luminescence assay was plotted against concentration and fit.

Figure 7-20: P2X2 is inhibited by suramin




Left, luminescence of ATP-induced signaling. Right, inhibition curve generated from timecourse data.

Figure 7-21: Nonfunctional P2X2 mutant does not gate




wtP2X2 response is depicted in gray. Orange represents the response of the broken receptor (K81C, K83C). No response is observed in the double mutant.

# References


Aguayo-Ortiz, R., Méndez-Lucio, O., Romo-Mancillas, A., Castillo, R., Yépez-Mulia, L., Medina-Franco, J.L., and Hernández-Campos, A. (2013a). Molecular basis for benzimidazole resistance from a novel β-tubulin binding site model. J. Mol. Graph. Model. *45*, 26–37.

Aguayo-Ortiz, R., Méndez-Lucio, O., Medina-Franco, J.L., Castillo, R., Yépez-Mulia, L., Hernández-Luis, F., and Hernández-Campos, A. (2013b). Towards the identification of the binding site of benzimidazoles to β-tubulin of Trichinella spiralis: Insights from computational and experimental data. J. Mol. Graph. Model. *41*, 12–19.

Aharoni, A., Amitai, G., Bernath, K., Magdassi, S., and Tawfik, D.S. (2005). High-Throughput Screening of Enzyme Libraries: Thiolactonases Evolved by Fluorescence-Activated Sorting of Single Cells in Emulsion Compartments. Chem. Biol. *12*, 1281–1289.

Akane, A., Matsubara, K., Nakamura, H., Takahashi, S., and Kimura, K. (1994). Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. J. Forensic Sci. *39*, 362–372.

Akanuma, S., Yamagishi, A., Tanaka, N., and Oshima, T. (1999). Further improvement of the thermal stability of a partially stabilized Bacillus subtilis 3-isopropylmalate dehydrogenase variant by random and site-directed mutagenesis. Eur. J. Biochem. *260*, 499–504.

Alexander, G.M., Rogan, S.C., Abbas, A.I., Armbruster, B.N., Pei, Y., Allen, J.A., Nonneman, R.J., Hartmann, J., Moy, S.S., Nicolelis, M.A., et al. (2009). Remote control of neuronal activity in transgenic mice expressing evolved G protein-coupled receptors. Neuron *63*, 27–39.

Al-Soud, W.A., and R\a adström, P. (2001). Purification and characterization of PCR-inhibitory components in blood cells. J. Clin. Microbiol. *39*, 485–493.

Amstutz, P., Pelletier, J.N., Guggisberg, A., Jermutus, L., Cesaro-Tadic, S., Zahnd, C., and Plückthun, A. (2002). In vitro selection for catalytic activity with ribosome display. J. Am. Chem. Soc. *124*, 9396–9403.

Anderson, D.H., and Duckworth, H.W. (1988). In vitro mutagenesis of *Escherichia coli* citrate synthase to clarify the locations of ligand binding sites. J. Biol. Chem. *263*, 2163–2169.

Arul, L., Benita, G., and Balasubramanian, P. (2008). Functional insight for ?-glucuronidase in Escherichia coli and Staphylococcus sp. RLH1. Bioinformation *2*, 339–343.

Asakawa, K., Kume, K., Kanai, M., Goshima, T., Miyahara, K., Dhut, S., Tee, W.W., Hirata, D., and Toda, T. (2006). The V260I Mutation in Fission Yeast α-Tubulin Atb2 Affects Microtubule Dynamics and EB1-Mal3 Localization and Activates the Bub1 Branch of the Spindle Checkpoint. Mol. Biol. Cell *17*, 1421–1435.

Assumpção, A.L.F.V., and da Silva, R.C. (2016). Immuno-PCR in cancer and non-cancer related diseases: a review. Vet. Q. *36*, 63–70.

Baar, C., d'Abbadie, M., Vaisman, A., Arana, M.E., Hofreiter, M., Woodgate, R., Kunkel, T.A., and Holliger, P. (2011). Molecular breeding of polymerases for resistance to environmental inhibitors. Nucleic Acids Res. *39*, e51.

Bahrami, S., and Drabløs, F. (2016). Gene regulation in the immediate-early response process. Adv. Biol. Regul. *62*, 37–49.

Baqi, Y., Hausmann, R., Rosefort, C., Rettinger, J., Schmalzing, G., and Müller, C.E. (2011). Discovery of potent competitive antagonists and positive modulators of the P2X2 receptor. J. Med. Chem. *54*, 817–830.

Barbier, P., Dorléans, A., Devred, F., Sanz, L., Allegro, D., Alfonso, C., Knossow, M., Peyrot, V., and Andreu, J.M. (2010). Stathmin and interfacial microtubule inhibitors recognize a naturally curved conformation of tubulin dimers. J. Biol. Chem. *285*, 31672–31681.

Barnoud, J., and Monticelli, L. (2015). Coarse-grained force fields for molecular simulations. Methods Mol. Biol. Clifton NJ *1215*, 125–149.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Barrick, J.E., and Lenski, R.E. (2013). Genome dynamics during experimental evolution. Nat. Rev. Genet. *14*, 827–839.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

Bhadra, S., Jiang, Y.S., Kumar, M.R., Johnson, R.F., Hensley, L.E., and Ellington, A.D. (2015). Real-time sequence-validated loop-mediated isothermal amplification assays for

detection of Middle East respiratory syndrome coronavirus (MERS-CoV). PloS One *10*, e0123126.

Blount, Z.D., Borland, C.Z., and Lenski, R.E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *105*, 7899–7906.

Blount, Z.D., Barrick, J.E., Davidson, C.J., and Lenski, R.E. (2012). Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature *489*, 513–518.

Bocci, G., Nicolaou, K.C., and Kerbel, R.S. (2002). Protracted low-dose effects on human endothelial cell proliferation and survival in vitro reveal a selective antiangiogenic window for various chemotherapeutic drugs. Cancer Res. *62*, 6938–6943.

Böhm, M., St rzebecher, J., and Klebe, G. (1999). Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. J. Med. Chem. *42*, 458–477.

Bouthors, A.-T., Delettré, J., Mugnier, P., Jarlier, V., and Sougakoff, W. (1999). Site-directed mutagenesis of residues 164, 170, 171, 179, 220, 237 and 242 in PER-1 β-lactamase hydrolysing expanded-spectrum cephalosporins. Protein Eng. Des. Sel. *12*, 313–318.

Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. Nat. Neurosci. *8*, 1263–1268.

Brannigan, J.A., and Wilkinson, A.J. (2002). Protein engineering 20 years on. Nat. Rev. Mol. Cell Biol. *3*, 964–970.

Broll, H. (2010). Polymerase Chain Reaction. In Molecular Biological and Immunological Techniques and Applications for Food Chemists, B. Popping, C. Diaz-Amigo, and K. Hoenicke, eds. (Hoboken, NJ, USA: John Wiley & Sons, Inc.), pp. 41–58.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964.

Brown, J.A., and Suo, Z. (2011). Unlocking the Sugar "Steric Gate" of DNA Polymerases. Biochemistry (Mosc.) *50*, 1135–1142.

Brown, A.J., Dyos, S.L., Whiteway, M.S., White, J.H.M., Watson, M.-A.E.A., Marzioch, M., Clare, J.J., Cousens, D.J., Paddon, C., Plumpton, C., et al. (2000). Functional

coupling of mammalian receptors to the yeast mating pathway using novel yeast/mammalian G protein α-subunit chimeras. Yeast *16*, 11–22.

Cadwell, R.C., and Joyce, G.F. (1994). Mutagenic PCR. PCR Methods Appl. *3*, S136-140.

Carmeliet, P., and Jain, R.K. (2011). Molecular mechanisms and clinical applications of angiogenesis. Nature *473*, 298–307.

Carter, P. (2001). Improving the efficacy of antibody-based cancer therapies. Nat. Rev. Cancer *1*, 118–129.

Cedrone, F., Ménez, A., and Quéméneur, E. (2000). Tailoring new enzyme functions by rational redesign. Curr. Opin. Struct. Biol. *10*, 405–410.

Cha, H.J., Byrom, M., Mead, P.E., Ellington, A.D., Wallingford, J.B., and Marcotte, E.M. (2012). Evolutionarily repurposed networks reveal the well-known antifungal drug thiabendazole to be a novel vascular disrupting agent. PLoS Biol. *10*, e1001379.

Chang, L., Li, J., and Wang, L. (2016). Immuno-PCR: An ultrasensitive immunoassay for biomolecular detection. Anal. Chim. Acta *910*, 12–24.

Chapa-y-Lazo, B., Allwood, E.G., Smaczynska-de Rooij, I.I., Snape, M.L., and Ayscough, K.R. (2014). Yeast Endocytic Adaptor AP-2 Binds the Stress Sensor Mid2 and Functions in Polarized Cell Responses. Traffic Cph. Den. *15*, 546–557.

Chelius, D., Loeb-Hennard, C., Fleischer, S., McIntyre, J.O., Marks, A.R., De, S., Hahn, S., Jehl, M.M., Moeller, J., Philipp, R., et al. (2000). Phosphatidylcholine Activation of Human Heart (R)-3-Hydroxybutyrate Dehydrogenase Mutants Lacking Active Center Sulfhydryls: Site-Directed Mutagenesis of a New Recombinant Fusion Protein. Biochemistry (Mosc.) *39*, 9687–9697.

Cheng, M.M.-C., Cuda, G., Bunimovich, Y.L., Gaspari, M., Heath, J.R., Hill, H.D., Mirkin, C.A., Nijdam, A.J., Terracciano, R., Thundat, T., et al. (2006). Nanotechnologies for biomolecular detection and medical diagnostics. Curr. Opin. Chem. Biol. *10*, 11–19.

Cho, A.E., Guallar, V., Berne, B.J., and Friesner, R. (2005). Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. J. Comput. Chem. *26*, 915–931.

Clark, A.M., and Labute, P. (2007). 2D depiction of protein - Ligand complexes. J. Chem. Inf. Model. *47*, 1933–1944.

Clark, L.C., and Lyons, C. (1962). Electrode systems for continuous monitoring in cardiovascular surgery. Ann. N. Y. Acad. Sci. *102*, 29–45.

Coco, W.M., Levinson, W.E., Crist, M.J., Hektor, H.J., Darzins, A., Pienkos, P.T., Squires, C.H., and Monticello, D.J. (2001). DNA shuffling method for generating highly recombined genes and evolved enzymes. Nat. Biotechnol. *19*, 354–359.

Coddou, C., Yan, Z., Obsil, T., Huidobro-Toro, J.P., and Stojilkovic, S.S. (2011). Activation and Regulation of Purinergic P2X Receptor Channels. Pharmacol. Rev. *63*, 641–683.

Connolly, B.A., Fogg, M.J., Shuttleworth, G., and Wilson, B.T. (2003). Uracil recognition by archaeal family B DNA polymerases. Biochem. Soc. Trans. *31*, 699–702.

Correia, B.E., Bates, J.T., Loomis, R.J., Baneyx, G., Carrico, C., Jardine, J.G., Rupert, P., Correnti, C., Kalyuzhniy, O., Vittal, V., et al. (2014). Proof of principle for epitope-focused vaccine design. Nature *507*, 201–206.

Coward, P., Wada, H.G., Falk, M.S., Chan, S.D., Meng, F., Akil, H., and Conklin, B.R. (1998). Controlling signaling with a specifically designed Gi-coupled receptor. Proc. Natl. Acad. Sci. U. S. A. *95*, 352–357.

Cozens, C., Pinheiro, V.B., Vaisman, A., Woodgate, R., and Holliger, P. (2012). A short adaptive path from DNA to RNA polymerases. Proc. Natl. Acad. Sci. U. S. A. *109*, 8067–8072.

Cunningham, B.C., and Wells, J.A. (1991). Rational design of receptor-specific variants of human growth hormone. Proc. Natl. Acad. Sci. U. S. A. *88*, 3407–3411.

Daar, A.S., Thorsteinsdóttir, H., Martin, D.K., Smith, A.C., Nast, S., and Singer, P.A. (2002). Top ten biotechnologies for improving health in developing countries. Nat. Genet. *32*, 229–232.

Dal Ben, D., Buccioni, M., Lambertucci, C., Marucci, G., Thomas, A., and Volpini, R. (2015). Purinergic P2X receptors: structural models and analysis of ligand-target interaction. Eur. J. Med. Chem. *89*, 561–580.

Dalby, P.A. (2003). Optimising enzyme function by directed evolution. Curr. Opin. Struct. Biol. *13*, 500–505.

Davidse, L.C. (1986). Benzimidazole Fungicides: Mechanism of Action and Biological Impact. Annu. Rev. Phytopathol. *24*, 43–65.

Davidse, L.C., and Flach, W. (1978). Interaction of thiabendazole with fungal tubulin. Biochim. Biophys. Acta *543*, 82–90.

Dawkins,  Richard (2006). The Selfish Gene (Oxford University Press).

Dawson, P.J., Gutteridge, W.E., and Gull, K. (1984). A comparison of the interaction of anthelmintic benzimidazoles with tubulin isolated from mammalian tissue and the parasitic nematode Ascaridia galli. Biochem. Pharmacol. *33*, 1069–1074.

Declerck, N., Machius, M., Wiegand, G., Huber, R., and Gaillardin, C. (2000). Probing structural determinants specifying high thermostability in Bacillus licheniformis α-amylase1. J. Mol. Biol. *301*, 1041–1057.

DeGrado, W.F., Prendergast, F.G., Wolfe, H.R., and Cox, J.A. (1985). The design, synthesis, and characterization of tight-binding inhibitors of calmodulin. J. Cell. Biochem. *29*, 83–93.

Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. Nature *418*, 700–707.

DiBenedictis, B.T., Nussbaum, E.R., Cheung, H.K., and Veenema, A.H. (2017). Quantitative mapping reveals age and sex differences in vasopressin, but not oxytocin, immunoreactivity in the rat social behavior neural network. J. Comp. Neurol.

Dickinson, F.M. (1970). The binding of dihydronicotinamide–adenine dinucleotide and pyridine-3-aldehyde–adenine dinucleotide by yeast alcohol dehydrogenase. Biochem. J. *120*, 821–830.

Doupnik, C.A., Parra, K.C., and Guida, W.C. (2015). A computational design approach for virtual screening of peptide interactions across K(+) channel families. Comput. Struct. Biotechnol. J. *13*, 85–94.

Du, Y., Hughes, R.A., Bhadra, S., Jiang, Y.S., Ellington, A.D., and Li, B. (2015). A Sweet Spot for Molecular Diagnostics: Coupling Isothermal Amplification and Strand Exchange Circuits to Glucometers. Sci. Rep. *5*, 11039.

Du, Y., Pothukuchy, A., Gollihar, J.D., Nourani, A., Li, B., and Ellington, A.D. (2017). Coupling Sensitive Nucleic Acid Amplification with Commercial Pregnancy Test Strips. Angew. Chem. Int. Ed Engl. *56*, 992–996.

Duckworth, H.W., and Tong, E.K. (1976). The binding of reduced nicotinamide adenine dinucleotide to citrate synthase of *Escherichia coli* K12. Biochemistry (Mosc.) *15*, 108–114.

Dunbrack, R.L. (2002). Rotamer libraries in the 21st century. Curr. Opin. Struct. Biol. *12*, 431–440.

Dunbrack, R.L., and Karplus, M. (1993). Backbone-dependent rotamer library for proteins. Application to side-chain prediction. J. Mol. Biol. *230*, 543–574.

Eckhart, L., Bach, J., Ban, J., and Tschachler, E. (2000). Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity. Biochem. Biophys. Res. Commun. *271*, 726–730.

Eisenberg, D., Marcotte, E.M., Xenarios, I., and Yeates, T.O. (2000). Protein function in the post-genomic era. Nature *405*, 823–826.

Ellefson, J.W., and others CPR Reinvigorates Directed Evolution Efforts. Biopolymers *101*.

Ellefson, J.W., Meyer, A.J., Hughes, R.A., Cannon, J.R., Brodbelt, J.S., and Ellington, A.D. (2014). Directed evolution of genetic parts and circuits by compartmentalized partnered replication. Nat. Biotechnol. *32*, 97–101.

Ellefson, J.W., Gollihar, J., Shroff, R., Shivram, H., Iyer, V.R., and Ellington, A.D. (2016). Synthetic evolutionary origin of a proofreading reverse transcriptase. Science *352*, 1590–1593.

Engin, S., Trouillet, V., Franz, C.M., Welle, A., Bruns, M., and Wedlich, D. (2010). Benzylguanine thiol self-assembled monolayers for the immobilization of SNAP-tag proteins on microcontact-printed surface structures. Langmuir *26*, 6097–6101.

Falk, R.T., Rossi, S.C., Fears, T.R., Sepkovic, D.W., Migella, A., Adlercreutz, H., Donaldson, J., Bradlow, H.L., and Ziegler, R.G. (2000). A New ELISA Kit for Measuring Urinary 2-Hydroxyestrone, 16α-Hydroxyestrone, and Their Ratio: Reproducibility, Validity, and Assay Performance after Freeze-Thaw Cycling and Preservation by Boric Acid. Cancer Epidemiol. Biomarkers Prev. *9*, 81–87.

Flores, H., and Ellington, A.D. (2002). Increasing the thermal stability of an oligomeric protein, beta-glucuronidase. J. Mol. Biol. *315*, 325–337.

Fogg, M.J., Pearl, L.H., and Connolly, B.A. (2002). Structural basis for uracil recognition by archaeal family B DNA polymerases. Nat. Struct. Biol. *9*, 922–927.

Folkman, J. (2007). Angiogenesis: an organizing principle for drug discovery? Nat. Rev. Drug Discov. *6*, 273–286.

Forsburg, S.L. (2001). The art and design of genetic screens: yeast. Nat. Rev. Genet. *2*, 659–668.

Freeland, S.J., and Hurst, L.D. (1998). The genetic code is one in a million. J. Mol. Evol. *47*, 238–248.

Fu, Z., Li, X., and Merz, K.M. (2011). Accurate assessment of the strain energy in a protein-bound drug using QM/MM X-ray refinement and converged quantum chemistry. J. Comput. Chem. *32*, 2587–2597.

Gagne, S.J., Stout, J.M., Liu, E., Boubakir, Z., Clark, S.M., and Page, J.E. (2012). Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides. Proc. Natl. Acad. Sci. *109*, 12811–12816.

Gaynor, E.C., te Heesen, S., Graham, T.R., Aebi, M., and Emr, S.D. (1994). Signal-mediated retrieval of a membrane protein from the Golgi to the ER in yeast. J. Cell Biol. *127*, 653–665.

Gebler, J.C., Aebersold, R., and Withers, S.G. (1992). Glu-537, not Glu-461, is the nucleophile in the active site of (lac Z) beta-galactosidase from Escherichia coli. J. Biol. Chem. *267*, 11126–11130.

Geddie, M.L., and Matsumura, I. (2004). Rapid Evolution of β-Glucuronidase Specificity by Saturation Mutagenesis of an Active Site Loop. J. Biol. Chem. *279*, 26462–26468.

Ghadessy, F.J., and Holliger, P. (2007). Compartmentalized self-replication: a novel method for the directed evolution of polymerases and other enzymes. Methods Mol. Biol. Clifton NJ *352*, 237–248.

Ghadessy, F.J., Ong, J.L., and Holliger, P. (2001a). Directed evolution of polymerase function by compartmentalized self-replication. Proc. Natl. Acad. Sci. U. S. A. *98*, 4552–4557.

Ghadessy, F.J., Ong, J.L., and Holliger, P. (2001b). Directed evolution of polymerase function by compartmentalized self-replication. Proc. Natl. Acad. Sci. *98*, 4552–4557.

Giacomini, E., Buonfiglio, R., Masetti, M., Wang, Y., Tseng, G.-N., Roberti, M., and Recanatini, M. (2015). A Ligand-Based Virtual Screening Approach to Identify Small Molecules as hERG Channel Activators. Comb. Chem. High Throughput Screen.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Giljohann, D.A., and Mirkin, C.A. (2009). Drivers of biodiagnostic development. Nature *462*, 461–464.

Gorris, H.H., and Walt, D.R. (2009). Mechanistic Aspects of Horseradish Peroxidase Elucidated through Single-Molecule Studies. J. Am. Chem. Soc. *131*, 6277–6282.

Graaf, M., Boven, E., Scheeren, H.W., Haisma, H.J., and Pinedo, H.M. (2002). Beta-Glucuronidase-Mediated Drug Release. Curr. Pharm. Des. *8*, 1391–1403.

Gu, G.J., Friedman, M., Jost, C., Johnsson, K., Kamali-Moghaddam, M., Plückthun, A., Landegren, U., and Söderberg, O. (2013). Protein tag-mediated conjugation of oligonucleotides to recombinant affinity binders for proximity ligation. New Biotechnol. *30*, 144–152.

Guarnieri, F. (2015). Designing an Orally Available Nontoxic p38 Inhibitor with a Fragment-Based Strategy. Methods Mol. Biol. Clifton NJ *1289*, 211–226.

Guo, J., Ma, X., Cai, Y., Ma, Y., Zhan, Z., Zhou, Y.J., Liu, W., Guan, M., Yang, J., Cui, G., et al. (2016). Cytochrome P450 promiscuity leads to a bifurcating biosynthetic pathway for tanshinones. New Phytol. *210*, 525–534.

Haisma, H.J., Sernee, M.F., Hooijberg, E., Brakenhoff, R.H., vd Meulen-Muileman, I.H., Pinedo, H.M., and Boven, E. (1998). Construction and characterization of a fusion protein of single-chain anti-CD20 antibody and human β-glucuronidase for antibody-directed enzyme prodrug therapy. Blood *92*, 184–190.

Hammerling, M.J., Ellefson, J.W., Boutz, D.R., Marcotte, E.M., Ellington, A.D., and Barrick, J.E. (2014). Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. Nat. Chem. Biol. *10*, 178–180.

Hammerling, M.J., Gollihar, J., Mortensen, C., Alnahhas, R.N., Ellington, A.D., and Barrick, J.E. (2016). Expanded Genetic Codes Create New Mutational Routes to Rifampicin Resistance in *Escherichia coli*. Mol. Biol. Evol. *33*, 2054–2063.

Hammond, L.A., Davidson, K., Lawrence, R., Camden, J.B., Von Hoff, D.D., Weitman, S., and Izbicka, E. (2001). Exploring the mechanisms of action of FB642 at the cellular level. J. Cancer Res. Clin. Oncol. *127*, 301–313.

Harford-Cross, C.F., Carmichael, A.B., Allan, F.K., England, P.A., Rouch, D.A., and Wong, L.-L. (2000). Protein engineering of cytochrome P450cam (CYP101) for the oxidation of polycyclic aromatic hydrocarbons. Protein Eng. Des. Sel. *13*, 121–128.

Harris, J.L., and Craik, C.S. (1998). Engineering enzyme specificity. Curr. Opin. Chem. Biol. *2*, 127–132.

Hattori, M., and Gouaux, E. (2012). Molecular mechanism of ATP binding and ion channel activation in P2X receptors. Nature *485*, 207–212.

He, S.-X., Song, G., Shi, J.-P., Guo, Y.-Q., and Guo, Z.-Y. (2014). Nanoluciferase as a novel quantitative protein fusion tag: Application for overexpression and bioluminescent receptor-binding assays of human leukemia inhibitory factor. Biochimie *106*, 140–148.

Heath, V.L., and Bicknell, R. (2009). Anticancer strategies involving the vasculature. Nat. Rev. Clin. Oncol. *6*, 395–404.

Hellinga, H.W., Caradonna, J.P., and Richards, F.M. (1991). Construction of new ligand binding sites in proteins of known structure. J. Mol. Biol. *222*, 787–803.

Henrissat, B., and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. *316*, 695.

Heuvel, R.H.H. van den, Fraaije, M.W., Ferrer, M., Mattevi, A., and Berkel, W.J.H. van (2000). Inversion of stereospecificity of vanillyl-alcohol oxidase. Proc. Natl. Acad. Sci. *97*, 9455–9460.

Hinnen, P., and Eskens, F. a. L.M. (2007). Vascular disrupting agents in clinical development. Br. J. Cancer *96*, 1159–1165.

Hutchings, C.J., Koglin, M., and Marshall, F.H. (2010). Therapeutic antibodies directed at G protein-coupled receptors. mAbs *2*, 594–606.

IGARASHI, K., OZAWA, T., IKAWA-KITAYAMA, K., HAYASHI, Y., ARAKI, H., ENDO, K., HAGIHARA, H., OZAKI, K., KAWAI, S., and ITO, S. (1999). Thermostabilization by Proline Substitution in an Alkaline, Liquefying α-Amylase from Bacillus sp. Strain KSM-1378. Biosci. Biotechnol. Biochem. *63*, 1535–1540.

Jain, S., Drendel, W.B., Chen, Z., Mathews, F.S., Sly, W.S., and Grubb, J.H. (1996). Structure of human β-glucuronidase reveals candidate lysosomal targeting and active-site motifs. Nat. Struct. Mol. Biol. *3*, 375–381.

Jarvis, M.F., and Khakh, B.S. (2009). ATP-gated P2X cation-channels. Neuropharmacology *56*, 208–215.

Jefferson, R.A. (1993). Plant promoter .alpha.-glucuronidase gene construct.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. *6*, 3901–3907.

Jiang, Y.S., Bhadra, S., Li, B., Wu, Y.R., Milligan, J.N., and Ellington, A.D. (2015). Robust strand exchange reactions for the sequence-specific, real-time detection of nucleic acid amplicons. Anal. Chem. *87*, 3314–3320.

Jo, S., Cheng, X., Lee, J., Kim, S., Park, S.-J., Patel, D.S., Beaven, A.H., Lee, K.I., Rui, H., Park, S., et al. (2017). CHARMM-GUI 10 years for biomolecular modeling and simulation. J. Comput. Chem. *38*, 1114–1124.

Jones, J.B., and DeSantis, G. (1999). Toward Understanding and Tailoring the Specificity of Synthetically Useful Enzymes. Acc. Chem. Res. *32*, 99–107.

Jones, P.T., Dear, P.H., Foote, J., Neuberger, M.S., and Winter, G. (1986). Replacing the complementarity-determining regions in a human antibody with those from a mouse. Nature *321*, 522–525.

Ju, S.-S., Lin, L.-L., Chien, H.R., and Hsu, W.-H. (2000). Substitution of the critical methionine residues in Trigonopsis variabilis D-amino acid oxidase with leucine enhances its resistance to hydrogen peroxide. FEMS Microbiol. Lett. *186*, 215–219.

Jung, M.K., Wilder, I.B., and Oakley, B.R. (1992). Amino acid alterations in thebenA (β-tubulin) gene ofAspergillus nidulans that confer benomyl resistance. Cell Motil. Cytoskeleton *22*, 170–174.

Kampmeier, F., Ribbert, M., Nachreiner, T., Dembski, S., Beaufils, F., Brecht, A., and Barth, S. (2009). Site-Specific, Covalent Labeling of Recombinant Antibody Fragments via Fusion to an Engineered Version of 6-O-Alkylguanine DNA Alkyltransferase. Bioconjug. Chem. *20*, 1010–1015.

Kavallaris, M. (2010). Microtubules and resistance to tubulin-binding agents. Nat. Rev. Cancer *10*, 194–204.

Kenski, D.M., Zhang, C., von Zastrow, M., and Shokat, K.M. (2005). Chemical genetic engineering of G protein-coupled receptor kinase 2. J. Biol. Chem. *280*, 35051–35061.

Keppler, A., Kindermann, M., Gendreizig, S., Pick, H., Vogel, H., and Johnsson, K. (2004). Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro. Methods *32*, 437–444.

Kerbel, R.S. (2008). Tumor angiogenesis. N. Engl. J. Med. *358*, 2039–2049.

Khan, H.A., Ahmad, A., and Mehboob, R. (2015). Nosocomial infections and their control strategies. Asian Pac. J. Trop. Biomed. *5*, 509–514.

Kitabayashi, M., Nishiya, Y., Esaka, M., Itakura, M., and Imanaka, T. (2002). Gene cloning and polymerase chain reaction with proliferating cell nuclear antigen from Thermococcus kodakaraensis KOD1. Biosci. Biotechnol. Biochem. *66*, 2194–2200.

Knight, R.D., Freeland, S.J., and Landweber, L.F. (2001). Rewiring the keyboard: evolvability of the genetic code. Nat. Rev. Genet. *2*, 49–58.

Koenraadt, H. (1992). Characterization of Mutations in the Beta-Tubulin Gene of Benomyl-Resistant Field Strains of *Venturia inaequalis* and Other Plant Pathogenic Fungi. Phytopathology *82*, 1348.

Kotze, A.C., Cowling, K., Bagnall, N.H., Hines, B.M., Ruffell, A.P., Hunt, P.W., and Coleman, G.T. (2012). Relative level of thiabendazole resistance associated with the E198A and F200Y SNPs in larvae of a multi-drug resistant isolate of Haemonchus contortus. Int. J. Parasitol. Drugs Drug Resist. *2*, 92–97.

Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., and Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. Neuron *93*, 480–490.

Kranaster, R., and Marx, A. (2010). Engineered DNA Polymerases in Biotechnology. ChemBioChem *11*, 2077–2084.

Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., and Baker, D. (2003). Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. Science *302*, 1364–1368.

Kumar, S., Tsai, C.-J., and Nussinov, R. (2000). Factors enhancing protein thermostability. Protein Eng. Des. Sel. *13*, 179–191.

Labute, P. (2009). Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. Proteins *75*, 187–205.

Labute, P. (2010a). LowModeMD - Implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. J. Chem. Inf. Model. *50*, 792–800.

Labute, P. (2010b). LowModeMD--implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. J. Chem. Inf. Model. *50*, 792–800.

Ladds, G., Goddard, A., and Davey, J. (2005). Functional analysis of heterologous GPCR signalling pathways in yeast. Trends Biotechnol. *23*, 367–373.

Lambrecht, G., Braun, K., Damer, S., Ganso, M., Hildebrandt, C., Ullmann, H., Kassack, M.U., and Nickel, P. (2002). Structure-activity relationships of suramin and pyridoxal-5'-phosphate derivatives as P2 receptor antagonists. Curr. Pharm. Des. *8*, 2371–2399.

Lan, T., Zhang, J., and Lu, Y. (2016). Transforming the blood glucose meter into a general healthcare meter for in vitro diagnostics in mobile health. Biotechnol. Adv. *34*, 331–341.

Laos, R., Thomson, J.M., and Benner, S.A. (2014). DNA polymerases engineered by directed evolution to incorporate non-standard nucleotides. Front. Microbiol. *5*.

Lassila, J.K., Privett, H.K., Allen, B.D., and Mayo, S.L. (2006). Combinatorial methods for small-molecule placement in computational enzyme design. Proc. Natl. Acad. Sci. U. S. A. *103*, 16710–16715.

Lê, K.T., Boué-Grabot, E., Archambault, V., and Séguéla, P. (1999). Functional and biochemical evidence for heteromeric ATP-gated channels composed of P2X1 and P2X5 subunits. J. Biol. Chem. *274*, 15415–15419.

Leandro-García, L.J., Leskelä, S., Landa, I., Montero-Conde, C., López-Jiménez, E., Letón, R., Cascón, A., Robledo, M., and Rodríguez-Antona, C. (2010). Tumoral and tissue-specific expression of the major human β-tubulin isotypes. Cytoskeleton *67*, 214–223.

Leatherbarrow, R.J., Fersht, A.R., and Winter, G. (1985). Transition-state stabilization in the mechanism of tyrosyl-tRNA synthetase revealed by protein engineering. Proc. Natl. Acad. Sci. U. S. A. *82*, 7840–7844.

Lee, M.E., DeLoache, W.C., Cervantes, B., and Dueber, J.E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synth. Biol. *4*, 975–986.

Lee, M.-H., Pan, S.-M., Ng, T.-W., Chen, P.-S., Wang, L.-Y., and Chung, K.-R. (2011). Mutations of β-tubulin codon 198 or 200 indicate thiabendazole resistance among isolates of Penicillium digitatum collected from citrus in Taiwan. Int. J. Food Microbiol. *150*, 157–163.

Lenski, R.E., and Travisano, M. (1994). Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proc. Natl. Acad. Sci. U. S. A. *91*, 6808–6814.

Lenski, R.E., Rose, M.R., Simpson, S.C., and Tadler, S.C. (1991). Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am Nat *138*, 1315–1341.

Liao, J.-Y., and Li, H. (2014). Target-induced DNAzyme Cleavage Accompanying Bioactive Enzymatic Assembly with Glucometer Readout for Quantitative Monitoring of Lead Ion. Chem. Lett. *43*, 1599–1600.

Lippow, S.M., and Tidor, B. (2007). Progress in computational protein design. Curr. Opin. Biotechnol. *18*, 305.

Liu, Y., Wang, Y., Zhang, Y., Liu, T., Jia, H., Zou, H., Fu, Q., Zhang, Y., Lu, L., Chao, E., et al. (2016). Rational Design of Dual Agonist-Antibody Fusions as Long-acting Therapeutic Hormones. ACS Chem. Biol. *11*, 2991–2995.

Lovell, S.C., Word, J.M., Richardson, J.S., and Richardson, D.C. (2000). The penultimate rotamer library. Proteins *40*, 389–408.

Ludington, J.L. (2015). Virtual fragment preparation for computational fragment-based drug design. Methods Mol. Biol. Clifton NJ *1289*, 31–41.

Lynch, K.J., Touma, E., Niforatos, W., Kage, K.L., Burgard, E.C., van Biesen, T., Kowaluk, E.A., and Jarvis, M.F. (1999). Molecular and functional characterization of human P2X(2) receptors. Mol. Pharmacol. *56*, 1171–1181.

Maass, P., Schulz-Gasch, T., Stahl, M., and Rarey, M. (2007). Recore: a fast and versatile method for scaffold hopping based on small molecule crystal structure conformations. J. Chem. Inf. Model. *47*, 390–399.

MacKenzie, B.A., Striley, C. a. F., Biagini, R.E., Stettler, L.E., and Hines, C.J. (2000). Improved Rapid Analytical Method for the Urinary Determination of 3,5,6 Trichloro-2-Pyridinol, a Metabolite of Chlorpyrifos. Bull. Environ. Contam. Toxicol. *65*, 1–7.

Mahmoudian, L. (2010). The Central Dogma in Molecular Biology. In Unravelling Single Cell Genomics, pp. 15–25.

Maranhao, A.C., and Ellington, A.D. (2017). Evolving Orthogonal Suppressor tRNAs To Incorporate Modified Amino Acids. ACS Synth. Biol. *6*, 108–119.

Marks, R.S., Cullen, D.C., Karube, I., Lowe, C.R., and Weetall, H.H. (2007). Handbook of biosensors and biochips (John Wiley & Sons Chichester, UK:).

Mata, L., Gripon, J.-C., and Mistou, M.-Y. (1999). Deletion of the four C-terminal residues of PepC converts an aminopeptidase into an oligopeptidase. Protein Eng. Des. Sel. *12*, 681–686.

Matsumura, I., and Ellington, A.D. (2001). In vitro Evolution of Beta-glucuronidase into a Beta-galactosidase Proceeds Through Non-specific Intermediates. J. Mol. Biol. *305*, 331–339.

Matsumura, I., Wallingford, J.B., Surana, N.K., Vize, P.D., Ellington, A.D., and others (1999). Directed evolution of the surface chemistry of the reporter enzyme beta-glucuronidase. Nat. Biotechnol. *17*, 696–701.

Maurus, R., Nguyen, N.T., Stokell, D.J., Ayed, A., Hultin, P.G., Duckworth, H.W., and Brayer, G.D. (2003). Insights into the evolution of allosteric properties. The NADH binding site of hexameric type II citrate synthases. Biochemistry (Mosc.) *42*, 5555–5565.

McCafferty, J., Griffiths, A.D., Winter, G., and Chiswell, D.J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. Nature *348*, 552–554.

McGary, K.L., Park, T.J., Woods, J.O., Cha, H.J., Wallingford, J.B., and Marcotte, E.M. (2010). Systematic discovery of nonobvious human disease models through orthologous phenotypes. Proc. Natl. Acad. Sci. *107*, 6544–6549.

McNamara, D.E., Senese, S., Yeates, T.O., and Torres, J.Z. (2015). Structures of potent anticancer compounds bound to tubulin. Protein Sci. Publ. Protein Soc. *24*, 1164–1172.

de Mendoza, A., Sebé-Pedrós, A., and Ruiz-Trillo, I. (2014). The Evolution of the GPCR Signaling System in Eukaryotes: Modularity, Conservation, and the Transition to Metazoan Multicellularity. Genome Biol. Evol. *6*, 606–619.

Meyer, A.J., Ellefson, J.W., and Ellington, A.D. (2014a). Directed Evolution of a Panel of Orthogonal T7 RNA Polymerase Variants for in Vivo or in Vitro Synthetic Circuitry. ACS Synth. Biol.

Meyer, A.J., Ellefson, J., and Ellington, A.D. (2014b). Library Generation by Gene Shuffling. Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al *105*, Unit-15.12.

Miki, B., and McHugh, S. (2004). Selectable marker genes in transgenic plants: applications, alternatives and biosafety. J. Biotechnol. *107*, 193–232.

Miklos, A.E., Kluwe, C., Der, B.S., Pai, S., Sircar, A., Hughes, R.A., Berrondo, M., Xu, J., Codrea, V., Buckley, P.E., et al. (2012). Structure-Based Design of Supercharged, Highly Thermoresistant Antibodies. Chem. Biol. *19*, 449–455.

Molodtsov, V., Nawarathne, I.N., Scharf, N.T., Kirchhoff, P.D., Showalter, H.D.H., Garcia, G. a, and Murakami, K.S. (2013). X-ray crystal structures of the *Escherichia coli* RNA polymerase in complex with benzoxazinorifamycins. J. Med. Chem. *56*, 4758–4763.

Montagnana, M., Caputo, M., Giavarina, D., and Lippi, G. (2009). Overview on self-monitoring of blood glucose. Clin. Chim. Acta Int. J. Clin. Chem. *402*, 7–13.

Morgan, F.J., Birken, S., and Canfield, R.E. (1975). The amino acid sequence of human chorionic gonadotropin. The alpha subunit and beta subunit. J. Biol. Chem. *250*, 5247–5258.

Mouratou, B., Kasper, P., Gehring, H., and Christen, P. (1999). Conversion of Tyrosine Phenol-lyase to Dicarboxylic Amino Acid β-Lyase, an Enzyme Not Found in Nature. J. Biol. Chem. *274*, 1320–1325.

Mueller, M.M., Sperl, S., Stürzebecher, J., Bode, W., and Moroder, L. (2002). (R)-3-Amidinophenylalanine-derived inhibitors of factor Xa with a novel active-site binding mode. Biol. Chem. *383*, 1185–1191.

Mukhopadhyay, T., Sasaki, J., Ramesh, R., and Roth, J.A. (2002). Mebendazole elicits a potent antitumor effect on human cancer cell lines both in vitro and in vivo. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *8*, 2963–2969.

Müller, K.M., Stebel, S.C., Knall, S., Zipf, G., Bernauer, H.S., and Arndt, K.M. (2005). Nucleotide exchange and excision technology (NExT) DNA shuffling: a robust method for DNA fragmentation and directed evolution. Nucleic Acids Res. *33*, e117.

Nagel, G., Szellas, T., Huhn, W., Kateriya, S., Adeishvili, N., Berthold, P., Ollig, D., Hegemann, P., and Bamberg, E. (2003). Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. Proc. Natl. Acad. Sci. U. S. A. *100*, 13940–13945.

Nagy, M. (2007). Automated DNA extraction techniques for forensic analysis. Mol. Forensics 37.

Nakamura, Y., Ishii, J., and Kondo, A. (2015). Applications of yeast-based signaling sensor for characterization of antagonist and analysis of site-directed mutants of the human serotonin 1A receptor. Biotechnol. Bioeng. *112*, 1906–1915.

Ness, J.E., Kim, S., Gottman, A., Pak, R., Krebber, A., Borchert, T.V., Govindarajan, S., Mundorff, E.C., and Minshull, J. (2002). Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. Nat. Biotechnol. *20*, 1251–1255.

Nevozhay, D., Adams, R.M., Murphy, K.F., Josić, K., and Balázsi, G. (2009). Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene expression. Proc. Natl. Acad. Sci. *106*, 5123–5128.

North, R.A. (2002). Molecular Physiology of P2X Receptors. Physiol. Rev. *82*, 1013–1067.

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., and Hase, T. (2000). Loop-mediated isothermal amplification of DNA. Nucleic Acids Res. *28*, E63.

Obritsch, M.D., Fish, D.N., MacLaren, R., and Jung, R. (2005). Nosocomial infections due to multidrug-resistant Pseudomonas aeruginosa: epidemiology and treatment options. Pharmacotherapy *25*, 1353–1364.

Packer, M.S., and Liu, D.R. (2015). Methods for the directed evolution of proteins. Nat. Rev. Genet. *16*, 379–394.

Paluh, J.L., Killilea, A.N., Detrich, H.W., and Downing, K.H. (2004). Meiosis-specific failure of cell cycle progression in fission yeast by mutation of a conserved beta-tubulin residue. Mol. Biol. Cell *15*, 1160–1171.

Pantazes, R.J., Grisewood, M.J., and Maranas, C.D. (2011). Recent advances in computational protein design. Curr. Opin. Struct. Biol. *21*, 467–472.

Pasquier, E., Honore, S., Pourroy, B., Jordan, M.A., Lehmann, M., Briand, C., and Braguer, D. (2005). Antiangiogenic concentrations of paclitaxel induce an increase in microtubule dynamics in endothelial cells but not in cancer cells. Cancer Res. *65*, 2433–2440.

Percy, M.G., and Gründling, A. (2014). Lipoteichoic acid synthesis and function in gram-positive bacteria. Annu. Rev. Microbiol. *68*, 81–100.

Pereira, D.S., Donald, L.J., Hosfield, D.J., and Duckworth, H.W. (1994). Active site mutants of Escherichia coli citrate synthase. Effects of mutations on catalytic and allosteric properties. J. Biol. Chem. *269*, 412–417.

Perugino, G., Vettone, A., Illiano, G., Valenti, A., Ferrara, M.C., Rossi, M., and Ciaramella, M. (2012). Activity and Regulation of Archaeal DNA Alkyltransferase CONSERVED PROTEIN INVOLVED IN REPAIR OF DNA ALKYLATION DAMAGE. J. Biol. Chem. *287*, 4222–4231.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005). Scalable Molecular Dynamics with NAMD. J. Comput. Chem. *26*, 1781–1802.

Pierce, A.C., Rao, G., and Bemis, G.W. (2004). BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, p38, and HIV protease. J. Med. Chem. *47*, 2768–2775.

Pohl, M. (2000). Optimierung von Biokatalysatorenfür technische Prozesse. Chem. Ing. Tech. *72*, 883–885.

Porebski, P.J., Cymborowski, M., Pasenkiewicz-Gierula, M., and Minor, W. (2016). Fitmunk: improving protein structures by accurate, automatic modeling of side-chain conformations. Acta Crystallogr. Sect. Struct. Biol. *72*, 266–280.

Pourroy, B., Honoré, S., Pasquier, E., Bourgarel-Rey, V., Kruczynski, A., Briand, C., and Braguer, D. (2006). Antiangiogenic concentrations of vinflunine increase the interphase microtubule dynamics and decrease the motility of endothelial cells. Cancer Res. *66*, 3256–3263.

Prinz, B., Stahl, U., and Lang, C. (2003). Intracellular transport of a heterologous membrane protein, the human transferrin receptor, in Saccharomyces cerevisiae. Int. Microbiol. Off. J. Span. Soc. Microbiol. *6*, 49–55.

Qiu, J., Xu, J., Yu, J., Bi, C., Chen, C., and Zhou, M. (2011). Localisation of the benzimidazole fungicide binding site of Gibberella zeae β2-tubulin studied by site-directed mutagenesis. Pest Manag. Sci. *67*, 191–198.

Quandt, E.M., Gollihar, J., Blount, Z.D., Ellington, A.D., Georgiou, G., and Barrick, J.E. (2015). Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. eLife *4*.

Ranaivoson, F.M., Gigant, B., Berritt, S., Joullié, M., and Knossow, M. (2012). Structural plasticity of tubulin assembly probed by vinca-domain ligands. Acta Crystallogr. D Biol. Crystallogr. *68*, 927–934.

Reiländer, H., and Weiβ, H.M. (1998). Production of G-protein-coupled receptors in yeast. Curr. Opin. Biotechnol. *9*, 510–517.

Reinsvold, R.E., Jinkerson, R.E., Radakovits, R., Posewitz, M.C., and Basu, C. (2011). The production of the sesquiterpene β-caryophyllene in a transgenic strain of the cyanobacterium Synechocystis. J. Plant Physiol. *168*, 848–852.

Rissin, D.M., and Walt, D.R. (2006). Digital Concentration Readout of Single Enzyme Molecules Using Femtoliter Arrays and Poisson Statistics. Nano Lett. *6*, 520–523.

Rissin, D.M., Gorris, H.H., and Walt, D.R. (2008). Distinct and Long-Lived Activity States of Single Enzyme Molecules. J. Am. Chem. Soc. *130*, 5349–5353.

Roberts, J.A., Digby, H.R., Kara, M., El Ajouz, S., Sutcliffe, M.J., and Evans, R.J. (2008). Cysteine substitution mutagenesis and the effects of methanethiosulfonate reagents at P2X2 and P2X4 receptors support a core common mode of ATP action at P2X receptors. J. Biol. Chem. *283*, 20126–20136.

Rojkova, A.M., Galkin, A.G., Kulakova, L.B., Serov, A.E., Savitsky, P.A., Fedorchuk, V.V., and Tishkov, V.I. (1999). Bacterial formate dehydrogenase. Increasing the enzyme thermal stability by hydrophobization of alpha-helices. FEBS Lett. *445*, 183–188.

Rosales-Hernández, M.C., and Correa-Basurto, J. (2015). The importance of employing computational resources for the automation of drug discovery. Expert Opin. Drug Discov. *10*, 213–219.

Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O., et al. (2008). Kemp elimination catalysts by computational enzyme design. Nature *453*, 190–195.

Rowe, L.A., Geddie, M.L., Alexander, O.B., and Matsumura, I. (2003). A Comparison of Directed Evolution Approaches Using the β-Glucuronidase Model System. J. Mol. Biol. *332*, 851–860.

Royle, S.J., Bobanović, L.K., and Murrell-Lagnado, R.D. (2002). Identification of a Non-canonical Tyrosine-based Endocytic Motif in an Ionotropic Receptor. J. Biol. Chem. *277*, 35378–35385.

Royle, S.J., Qureshi, O.S., Bobanović, L.K., Evans, P.R., Owen, D.J., and Murrell-Lagnado, R.D. (2005). Non-canonical YXXGPhi endocytic motifs: recognition by AP2 and preferential utilization in P2X4 receptors. J. Cell Sci. *118*, 3073–3080.

Ryall, B., Eydallin, G., and Ferenci, T. (2012). Culture history and population heterogeneity as determinants of bacterial adaptation: the adaptomics of a single environmental transition. Microbiol. Mol. Biol. Rev. MMBR *76*, 597–625.

Ryazantsev, D.Y., Voronina, D.V., and Zavriev, S.K. (2016). Immuno-PCR: Achievements and Perspectives. Biochem. Biokhimiia *81*, 1754–1770.

Samish, I., MacDermaid, C.M., Perez-Aguilar, J.M., and Saven, J.G. (2011). Theoretical and Computational Protein Design. Annu. Rev. Phys. Chem. *62*, 129–149.

Scheib, H., Pleiss, J., Stadler, P., Kovac, A., Potthoff, A.P., Haalck, L., Spener, F., Paltauf, F., and Schmid, R.D. (1998). Rational design of Rhizopus oryzae lipase with modified stereoselectivity toward triradylglycerols. Protein Eng. Des. Sel. *11*, 675–682.

Scheib, H., Pleiss, J., Kovac, A., Paltauf, F., and Schmid, R.D. (1999). Stereoselectivity of Mucorales lipases toward triradylglycerols--a simple solution to a complex problem. Protein Sci. Publ. Protein Soc. *8*, 215–221.

Schmidt, L.S., Ghosoph, J.M., Margosan, D.A., and Smilanick, J.L. (2006). Mutation at β-Tubulin Codon 200 Indicated Thiabendazole Resistance in *Penicillium digitatum* Collected from California Citrus Packinghouses. Plant Dis. *90*, 765–770.

Sharma, S.K., and Bagshawe, K.D. (2017). Translating antibody directed enzyme prodrug therapy (ADEPT) and prospects for combination. Expert Opin. Biol. Ther. *17*, 1–13.

Shim, J.-Y., Ahn, K.H., and Kendall, D.A. (2013). Molecular basis of cannabinoid CB1 receptor coupling to the G protein heterotrimer Gαiβγ: identification of key CB1 contacts with the C-terminal helix α5 of Gαi. J. Biol. Chem. *288*, 32449–32465.

Singh, R.K., Lee, J.-K., Selvaraj, C., Singh, R., Li, J., Kim, S.-Y., and Kalia, V.C. (2016). Protein Engineering Approaches in the Post-Genomic Era. Curr. Protein Pept. Sci.

Sirajuddin, M., Rice, L.M., and Vale, R.D. (2014). Regulation of microtubule motors by tubulin isotypes and post-translational modifications. Nat. Cell Biol. *16*, 335–344.

Slusarczyk, H., Felber, S., Kula, M.-R., and Pohl, M. (2000). Stabilization of NAD-dependent formate dehydrogenase from Candida boidinii by site-directed mutagenesis of cysteine residues. Eur. J. Biochem. *267*, 1280–1289.

Sly, W.S., Quinton, B.A., McAlister, W.H., and Rimoin, D.L. (1973). Beta glucuronidase deficiency: Report of clinical, radiologic, and biochemical features of a new mucopolysaccharidosis. J. Pediatr. *82*, 249–257.

Smith, M.D., Zanghellini, A., and Grabs-Röthlisberger, D. (2014). Computational design of novel enzymes without cofactors. Methods Mol. Biol. Clifton NJ *1216*, 197–210.

Smith, M.D., Rao, J.S., Segelken, E., and Cruz, L. (2015). Force-Field Induced Bias in the Structure of Aβ21-30: A Comparison of OPLS, AMBER, CHARMM, and GROMOS Force Fields. J. Chem. Inf. Model. *55*, 2587–2595.

Srere, P.A. (1969). Citrate synthase. Methods Enzymol. *13*, 3–11.

Stanley, N., and De Fabritiis, G. (2015). High throughput molecular dynamics for drug discovery. Silico Pharmacol. *3*, 3.

Stathopoulos, A.M., and Cyert, M.S. (1997). Calcineurin acts through theCRZ1/TCN1-encoded transcription factor to regulate gene expression in yeast. Genes Dev. *11*, 3432–3444.

Stemmer, W.P.C. (1994). Rapid evolution of a protein in vitro by DNA shuffling. Nature *370*, 389–391.

Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M., and Heyneker, H.L. (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene *164*, 49–53.

Stewart, K.D., Shiroda, M., and James, C.A. (2006). Drug Guru: a computer software program for drug design using medicinal chemistry rules. Bioorg. Med. Chem. *14*, 7011–7022.

Stockklausner, C., Ludwig, J., Ruppersberg, J.P., and Klöcker, N. (2001). A sequence motif responsible for ER export and surface expression of Kir2.0 inward rectifier K(+) channels. FEBS Lett. *493*, 129–133.

Stokell, D.J., Donald, L.J., Maurus, R., Nguyen, N.T., Sadler, G., Choudhary, K., Hultin, P.G., Brayer, G.D., and Duckworth, H.W. (2003). Probing the roles of key residues in the unique regulatory NADH binding site of type II citrate synthase of Escherichia coli. J. Biol. Chem. *278*, 35435–35443.

Stout, J.M., Boubakir, Z., Ambrose, S.J., Purves, R.W., and Page, J.E. (2012). The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes. Plant J. Cell Mol. Biol. *71*, 353–365.

Stürzebecher, J., Prasa, D., Wikström, P., and Vieweg, H. (1995). Structure-activity relationships of inhibitors derived from 3-amidinophenylalanine. J. Enzym. Inhib. *9*, 87–99.

Subramaniam, S., and Senes, A. (2012). An energy-based conformer library for side chain optimization: improved prediction and adjustable sampling. Proteins *80*, 2218–2234.

Tack, D.S., Ellefson, J.W., Thyer, R., Wang, B., Gollihar, J., Forster, M.T., and Ellington, A.D. (2016). Addicting diverse bacteria to a noncanonical amino acid. Nat. Chem. Biol. *12*, 138–140.

Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M., and Imanaka, T. (1997). Characterization of DNA polymerase from Pyrococcus sp. strain KOD1 and its application to PCR. Appl. Environ. Microbiol. *63*, 4504–4510.

Tang, L., Wang, X., Ru, B., Sun, H., Huang, J., and Gao, H. (2014). MDC-Analyzer: a novel degenerate primer design tool for the construction of intelligent mutagenesis libraries with contiguous sites. BioTechniques *56*, 301–302, 304, 306–308, passim.

Tautermann, C.S. (2014). GPCR structures in drug design, emerging opportunities with new structures. Bioorg. Med. Chem. Lett. *24*, 4073–4079.

Tawfik, D.S., and Griffiths, A.D. (1998). Man-made cell-like compartments for molecular evolution. Nat. Biotechnol. *16*, 652–656.

Tiwari, M.K., Singh, R., Singh, R.K., Kim, I.-W., and Lee, J.-K. (2012). Computational approaches for rational design of proteins with novel functionalities. Comput. Struct. Biotechnol. J. *2*, e201209002.

Tobin, M.B., Gustafsson, C., and Huisman, G.W. (2000). Directed evolution: the "rational"basis for "irrational"design. Curr. Opin. Struct. Biol. *10*, 421–427.

Tozer, G.M., Kanthou, C., and Baguley, B.C. (2005). Disrupting tumour blood vessels. Nat. Rev. Cancer *5*, 423–435.

Tram, K., Kanda, P., Salena, B.J., Huan, S., and Li, Y. (2014). Translating bacterial detection by DNAzymes into a litmus test. Angew. Chem. Int. Ed Engl. *53*, 12799–12802.

Turanli-Yildiz, B., Alkim, C., and Cakar, Z.P. (2012). Protein Engineering Methods and Applications.

Ueda, T., Masumoto, K., Ishibashi, R., So, T., and Imoto, T. (2000). Remarkable thermal stability of doubly intramolecularly cross-linked hen lysozyme. Protein Eng. Des. Sel. *13*, 193–196.

Ulmer, K.M. (1983). Protein engineering. Science *219*, 666–671.

Urdea, M., Penny, L.A., Olmsted, S.S., Giovanni, M.Y., Kaspar, P., Shepherd, A., Wilson, P., Dahl, C.A., Buchsbaum, S., Moeller, G., et al. (2006). Requirements for high impact diagnostics in the developing world. Nature *444 Suppl 1*, 73–79.

Vanommeslaeghe, K., and MacKerell, A.D. (2015). CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. Biochim. Biophys. Acta *1850*, 861–871.

Vassilatis, D.K., Hohmann, J.G., Zeng, H., Li, F., Ranchalis, J.E., Mortrud, M.T., Brown, A., Rodriguez, S.S., Weller, J.R., Wright, A.C., et al. (2003). The G protein-coupled receptor repertoires of human and mouse. Proc. Natl. Acad. Sci. U. S. A. *100*, 4903–4908.

Verma, S., and Prabhakar, Y.S. (2015). Target Based Drug Design-A Reality in Virtual Sphere. Curr. Med. Chem.

Vilà, C., Savolainen, P., Maldonado, J.E., Amorim, I.R., Rice, J.E., Honeycutt, R.L., Crandall, K.A., Lundeberg, J., and Wayne, R.K. (1997). Multiple and Ancient Origins of the Domestic Dog. Science *276*, 1687–1689.

Vilà, C., Maldonado, J.E., and Wayne, R.K. (1999). Phylogenetic relationships, evolution, and genetic diversity of the domestic dog. J. Hered. *90*, 71–77.

Vo-Dinh, T., and Cullum, B. (2000). Biosensors and biochips: advances in biological and medical diagnostics. Fresenius J. Anal. Chem. *366*, 540–551.

Vögler, O., Barceló, J.M., Ribas, C., and Escribá, P.V. (2008). Membrane interactions of G proteins and other related proteins. Biochim. Biophys. Acta BBA - Biomembr. *1778*, 1640–1652.

Wallace, B.D., Wang, H., Lane, K.T., Scott, J.E., Orans, J., Koo, J.S., Venkatesh, M., Jobin, C., Yeh, L.-A., Mani, S., et al. (2010). Alleviating Cancer Drug Toxicity by Inhibiting a Bacterial Enzyme. Science *330*, 831–835.

Wang, C.C., and Touster, O. (1972a). Studies of catalysis by -glucuronidase. The effect of structure on the rate of hydrolysis of substituted phenyl- -d-glucopyranosiduronic acids. J. Biol. Chem. *247*, 2650–2656.

Wang, C.-C., and Touster, O. (1972b). Studies of Catalysis by β-Glucuronidase ACTIVE SITE. J. Biol. Chem. *247*, 2644–2649.

Wang, A.H., Fujii, S., van Boom, J.H., van der Marel, G.A., van Boeckel, S.A., and Rich, A. (1982). Molecular structure of r(GCG)d(TATACGC): a DNA--RNA hybrid helix joined to double helical DNA. Nature *299*, 601–604.

Wang, B., Lee, C.-H., Johnson, E.L., Kluwe, C.A., Cunningham, J.C., Tanno, H., Crooks, R.M., Georgiou, G., and Ellington, A.D. (2016a). Discovery of high affinity anti-ricin antibodies by B cell receptor sequencing and by yeast display of combinatorial VH:VL libraries from immunized animals. mAbs *8*, 1035–1044.

Wang, L.-P., McKiernan, K.A., Gomes, J., Beauchamp, K.A., Head-Gordon, T., Rice, J.E., Swope, W.C., Martínez, T.J., and Pande, V.S. (2017). Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. J. Phys. Chem. B *121*, 4023–4039.

Wang, Q., Wang, H., Yang, X., Wang, K., Liu, F., Zhao, Q., Liu, P., and Liu, R. (2014). Multiplex detection of nucleic acids using a low cost microfluidic chip and a personal glucose meter at the point-of-care. Chem. Commun. *50*, 3824.

Wang, W., Huang, S., Li, J., Rui, K., Zhang, J.-R., and Zhu, J.-J. (2016b). Coupling a DNA-Based Machine with Glucometer Readouts for Amplified Detection of Telomerase Activity in Cancer Cells. Sci. Rep. *6*, 23504.

Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M., and Vander Horn, P.B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res. *32*, 1197–1207.

Weinberg, J., and Drubin, D.G. (2012). Clathrin-mediated endocytosis in budding yeast. Trends Cell Biol. *22*, 1–13.

Weitzman, P.D. (1966a). Reduced nicotinamide-adenine dinucleotide as an allosteric effector of citrate-synthase activity in *Escherichia coli*. Biochem. J. *101*, 44c–5c.

Weitzman, P.D.J. (1966b). Regulation of citrate synthase activity in *Escherichia coli*. Biochim. Biophys. Acta BBA - Enzymol. Biol. Oxid. *128*, 213–215.

Weitzman, P.D., and Jones, D. (1968). Regulation of citrate synthase and microbial taxonomy. Nature *219*, 270–272.

Wilson, D.S., Keefe, A.D., and Szostak, J.W. (2001). The use of mRNA display to select high-affinity protein-binding peptides. Proc. Natl. Acad. Sci. U. S. A. *98*, 3750–3755.

Wirtz, R.M. (2014). Matrix and method for purifying and/or isolating nucleic acids (Google Patents).

Wolfe, K.A., Breadmore, M.C., Ferrance, J.P., Power, M.E., Conroy, J.F., Norris, P.M., and Landers, J.P. (2002). Toward a microchip-based solid-phase extraction method for isolation of nucleic acids. Electrophoresis *23*, 727–733.

Wong, D.T. (2006). Salivary diagnostics powered by nanotechnologies, proteomics and genomics. J. Am. Dent. Assoc. 1939 *137*, 313–321.

Wong, A.W., He, S., Grubb, J.H., Sly, W.S., and Withers, S.G. (1998). Identification of Glu-540 as the Catalytic Nucleophile of Human β-Glucuronidase Using Electrospray Mass Spectrometry. J. Biol. Chem. *273*, 34057–34062.

Wong, T.S., Tee, K.L., Hauer, B., and Schwaneberg, U. (2004). Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. Nucleic Acids Res. *32*, e26.

Xiang, Y., and Lu, Y. (2011). Using personal glucose meters and functional DNA sensors to quantify a variety of analytical targets. Nat. Chem. *3*, 697–703.

Xiong, A.-S., Peng, R.-H., Zhuang, J., Chen, J.-M., Zhang, B., Zhang, J., and Yao, Q.-H. (2011a). A Thermostable β-Glucuronidase Obtained by Directed Evolution as a Reporter Gene in Transgenic Plants. PLoS ONE *6*, e26773.

Xiong, A.-S., Peng, R.-H., Zhuang, J., Chen, J.-M., Zhang, B., Zhang, J., and Yao, Q.-H. (2011b). A Thermostable β-Glucuronidase Obtained by Directed Evolution as a Reporter Gene in Transgenic Plants. PLoS ONE *6*, e26773.

Xiong, A.-S., Peng, R.-H., Zhuang, J., Davies, J., Zhang, J., and Yao, Q.-H. (2012). Advances in directed molecular evolution of reporter genes. Crit. Rev. Biotechnol. *32*, 133–142.

Yager, P., Domingo, G.J., and Gerdes, J. (2008). Point-of-care diagnostics for global health. Annu. Rev. Biomed. Eng. *10*, 107–144.

Yarden, O., and Katan, T. (1993). Mutations leading to substitutions at amino acids 198 and 200 of beta-tubulin that correlate with benomyl-resistance phenotypes of field strains of Botrytis cinerea. In Phytopathology, (American Phytopathological Society), pp. 1478–1483.

Yoshimoto, N., Tatematsu, K., Iijima, M., Niimi, T., Maturana, A.D., Fujii, I., Kondo, A., Tanizawa, K., and Kuroda, S. 'ichi (2014). High-throughput de novo screening of

receptor agonists with an automated single-cell analysis and isolation system. Sci. Rep. *4*, 4242.

Zalatan, J.G., and Herschlag, D. (2009). The far reaches of enzymology. Nat. Chem. Biol. *5*, 516–520.

Zemelman, B.V., Lee, G.A., Ng, M., and Miesenböck, G. (2002). Selective photostimulation of genetically chARGed neurons. Neuron *33*, 15–22.

Zemelman, B.V., Nesnas, N., Lee, G.A., and Miesenbock, G. (2003). Photochemical gating of heterologous ion channels: remote control over genetically designated populations of neurons. Proc. Natl. Acad. Sci. U. S. A. *100*, 1352–1357.

Zha, D., Eipper, A., and Reetz, M.T. (2003). Assembly of designed oligonucleotides as an efficient method for gene recombination: a new tool in directed evolution. Chembiochem Eur. J. Chem. Biol. *4*, 34–39.

Zhang, R., and Xie, X. (2012). Tools for GPCR drug discovery. Acta Pharmacol. Sin. *33*, 372–384.

Zhao, H., Giver, L., Shao, Z., Affholter, J.A., and Arnold, F.H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat. Biotechnol. *16*, 258–261.

Zhao, J., Bao, X., Li, C., Shen, Y., and Hou, J. (2016). Improving monoterpene geraniol production through geranyl diphosphate synthesis regulation in Saccharomyces cerevisiae. Appl. Microbiol. Biotechnol. *100*, 4561–4571.

Zhu, G.P., Xu, C., Teng, M.K., Tao, L.M., Zhu, X.Y., Wu, C.J., Hang, J., Niu, L.W., and Wang, Y.Z. (1999). Increasing the thermostability of d-xylose isomerase by introduction of a proline into the turn of a random coil. Protein Eng. Des. Sel. *12*, 635–638.

Zhu, X., Xu, H., Lin, R., Yang, G., Lin, Z., and Chen, G. (2014a). Sensitive and portable detection of telomerase activity in HeLa cells using the personal glucose meter. Chem. Commun. Camb. Engl. *50*, 7897–7899.

Zhu, Z., Guan, Z., Jia, S., Lei, Z., Lin, S., Zhang, H., Ma, Y., Tian, Z.-Q., and Yang, C.J. (2014b). Au@Pt nanoparticle encapsulated target-responsive hydrogel with volumetric bar-chart chip readout for quantitative point-of-care testing. Angew. Chem. Int. Ed Engl. *53*, 12503–12507.