Expanding Biological Engineering from Single Enzymes

to Cellular Pathways

Nili Ostrov

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

ABSTRACT

Expanding Biological Engineering from Single Enzymes to Cellular Pathways

Nili Ostrov

The emerging field of synthetic biology evolved from biological research much the same way synthetic chemistry evolved from chemical research; with accumulated knowledge of the structure of single genes and proteins and the methodologies to manipulate them, researchers turn to forward engineer complex biological systems to effectively manipulate living systems. Much like in the case of enzyme engineering, a rationally designed biological network is currently beyond our reach, and we turn to directed evolution to circumvent this gap in knowledge. Yet the unique nature of live biological networks uncovered new challenges previously unmet by single-gene molecular technologies, and extrapolation of current technologies to the manipulation of multi-component has proven laborious and inefficient.

To establish engineering technologies for living cells, novel directed evolution techniques are sought for that are compatible with simultaneous manipulation of multiple biological components *in vivo*. In this work, we explore techniques for library DNA mutagenesis in the context of single and multiple genes. **Chapter 1** provides an overview of the challenges in expanding current in vivo directed evolution methods from single enzymes, to the design pathways and cells. **Chapter 2** describes the design and characterization of an assay for combinatorial directed evolution of a single metabolic enzyme. In **Chapter 3** we present the utilization of our DNA assembly system, Reiterative Recombination, for attenuation of metabolic pathways. We use a library of promoters to combinatorially vary

the expression of genes in the heterologous lycopene biosynthetic pathway in S. cerevisiae. Finally, **Chapter 4** explores the calibration of the dynamic range of genetic selection, using metabolic enzyme activity as probe for cell survival.

Table of Contents

iii

List of Figures

List of Tables

List of Abbreviations

5-FOA	5-fluoroorotic acid
A	deoxyadenosine
Amp^R	ampicillin resistance
A_x	absorbance at x nm
bp	base pair
°C	degrees Celsius
C	deoxycytoside
DMF	dimethyl formamide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DSB	double-strand DNA break
E. coli	*Escherichia coli*
et al.	*et alia*
FACS	fluorescence-activated cell sorting
FPP	farnesyl pyrophosphate
g	gram
G	deoxyguanosine
G418	geneticin
gal	galactose
gDNA	genomic DNA
h	hour
IPP	isopentenyl pyrophosphate

IPTG	isopropyl β-D-thiogalactopyranoside
Kan^R	kanamycin resistance
kb	kilobase
k_{cat}	catalytic turnover number
K_m	Michaelis constant
L	liter
M	moles per liter
MAGE	multiplex automated genome engineering
Mb	megabase
MCS	multiple cloning site
mg	milligram
min	minute
mL	milliliter
nm	nanometer
nM	nanomoles per liter
NRPS	nonribosomal peptide synthase
OD_x	optical density at x nm
ONPG	ortho-nitrophenyl-β-galactoside
ori	origin of replication
PACE	phage-assisted continuous evolution
PCR	polymerase chain reaction
PKS	polyketide synthase
pM	picomoles per liter

raf	raffinose
RNA	ribonucleic acid
rpm	revolutions per minute
s	second
S. cerevisiae	*Saccharomyces cerevisiae*
SC	synthetic complete
SDS	sodium dodecyl sulfate
t	time
T	deoxythymidine
UV	ultraviolet
YPD	yeast peptone dextrose media
μg	microgram
μL	microliter
μM	micromoles per liter

Acknowledgments

First and foremost, I would like to thank my supervisor, Professor Virginia Cornish. Her passion in rigorously pursuing highly interdisciplinary science, while thinking critically about the smallest details and never forgetting the big picture, was an inspiration to me. The thought and attention put into every aspect of her lab, from fostering a dynamic scientific environment to tending to the well-being of her group members, have made my graduate studies a fascinating learning experience and an everyday pleasure. I'm particularly thankful for her unhindered support and encouragement of my exploratory scientific adventures. Pushing for excellence has never been so much fun as in the Cornish lab.

I also extend my thanks to my graduate committee members, who have all been a great source of scientific knowledge and support throughout my time at Columbia. I wish to express my gratitude to Professor Julio Fernandez, whose door was always open for consultation and guidance, and has constantly encouraged me to keep reaching for the stars. Professor Ruben Gonzalez has been abundantly helpful and offered insights and assistance in both scientific and graduate life. I am also grateful to Professor Gordana Vunjak-Novakovic and Professor Mattheos Koffas for their willingness to serve on my thesis committee. Even more so, I am grateful for the opportunity they have given me to share their research in exciting collaborative projects, putting time and effort to bridge the gap between yeast genetics, stem cells, and metabolic engineering.

Special thanks is extended to Professor Alastair Ager, Professor James Eliades and Professor Leslie Roberts of Columbia Department of Public Health. Without their insightful advice on how things really work outside of lab, the quest to develop a biosensor would not have been possible.

I have been extremely lucky to work with an exceptional group of people in my five years in the Cornish group. They have been an endless source of scientific enthusiasm and challenge, friendly support, and at times of failed experiments - much needed entertainment. My deepest gratitude is due to each one of them, and especially to the members of the directed evolution group. In particular, my partner in crime and dear friend

To Omer, who has shared the great pleasures and challenges of this exciting period with me, for making sure I walk the line.

Chapter 1

Harnessing *in vivo* Directed Evolution for Enzyme and

Cell Design

1   Chapter outlook

Throughout history, human societies have continuously sought to harness their natural environment seeking better food, health, and shelter. The cultivation of crops and breeding of domesticated animals are perhaps the earliest attempts to beneficially shape the biological world. It was not until the late 20th century, however, that the powerful tools of recombinant DNA technology enabled highly precise manipulation of single genes, revolutionizing the way we utilize useful biological systems. Whereas our limited understanding of molecular mechanisms hampers true tailoring of biological behavior, directed evolution techniques combine rational design with molecular evolution to conveniently circumvent this difficulty. Traditionally targeting single genes by way of *in vitro* DNA mutagenesis, directed evolution is currently widely used for the development of enzymatic biocatalysts. However, as advances in synthetic biology improve our cell engineering capabilities, increasing research efforts are aimed at generating man-made networks of molecules that function *in vivo* for production of low-cost drugs, chemicals and energy. Technologies are thus needed to expand directed evolution from single to multiple biological components and to be carried out directly in the cell. Here we review current methods for *in vivo* directed evolution in light of the new challenges posed by the emerging fields of pathway and whole cell engineering.

1.1   *In vivo* **directed evolution**

The field of synthetic biology aims to utilize natural building blocks to make new, useful biological tools. Yet our limited understanding of natural molecular systems makes rational design of even the simplest biological machine largely unattainable and the dynamic nature and complexity of living organisms greatly challenges the well-defined, predictable principles of traditional engineering disciplines. Paradoxically, However, it is the same cellular properties that provide us with the most powerful tools for their engineering.[1]

Laboratory directed evolution is now routinely used to develop improved catalysts for pharmaceutical and industrial purposes. It implements a Darwinian optimization cycle of DNA mutagenesis and selection by which the fittest variants with a desired phenotype are selected from a collection ('library') of mutants (Figure 1-1). It is hard to underestimate the contribution of recombinant DNA technology[2] to the improvement of directed evolution techniques, enabling controlled manipulation of DNA to readily produce extensive genetic variation in the laboratory. *In vitro* molecular manipulation was fundamental for the development of directed evolution, as demonstrated by numerous successfully evolved biocatalysts.[3] However, many of the methodologies traditionally used for
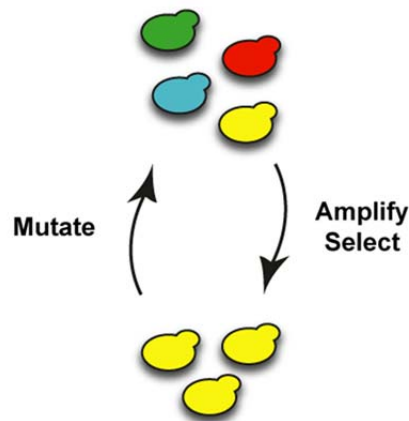
Figure 1-1. Laboratory directed evolution. Mutagenesis is applied to increase genetic variability. The collection of mutants is then amplified and selected for a desired property.

evolution of single molecules prove laborious and inefficient once challenged by the complexity of multiple-component networks. As synthetic biology now pushes the boundaries of biological engineering, traditional methodologies for directed evolution face new challenged.

In this chapter we introduce *in vivo* approaches for directed evolution, starting from engineering of single biocatalysts. We then discuss the challenges involved in engineering multi-component biological systems such as metabolic pathways and entire cellular networks, in light of currently available methodologies. Finally, we present new approaches and enabling technologies put forward to address large scale biological design. We limit ourselves to the work done in *E. coli* and *S. cerevisiae*, the most commonly used laboratory host thus far.

1.2   Technologies for *in vivo* directed enzyme evolution

Directed evolution has traditionally been an enzyme-oriented research field.[4,5] It provides an efficient technological route to an otherwise unattainable challenge – altering enzyme functionality without having detailed understanding of its structure-function relationship. The overwhelming variety of highly efficient biological catalysts make them an extremely attractive technological tool, yet natural enzymes are rarely suitable for industrial purposes without adjustments in stability, specificity and efficiency.[3] Without fundamental understanding of the underlying design rules, recombinant DNA techniques alone could not have been used to rationally build a tailor-made enzyme active site. However, in the context of laboratory evolution, the capability to efficiently mutate and select enzyme functionality allows us to search much beyond the limits of our

understanding of biological systems, and explore not only alteration of existing catalysts but also the evolution of novel non-natural functions.[6]

Considering the ease of manipulation of single-gene fragments by PCR, it is not surprising that *in vitro* DNA mutagenesis is now routinely used to create libraries of protein and nucleic acid molecules for directed enzyme evolution.[7] *In vitro* mutagenesis techniques for library generation have been developed extensively, and robust methods for both random library mutagenesis, such as error prone PCR and DNA shuffling, and targeted library mutagenesis, such as cassette mutagenesis, are available today.[8,9] The most popular *in vitro* library methods have been thoroughly analyzed for their mutagenesis rate, mutational bias, dependence on DNA sequence and length and reproducibility.[2,10] This comprehensive development and characterization of *in vitro* library mutagenesis methods have made these approaches generally accessible.

By contrast, *in vivo* library mutagenesis approaches are less developed and reported optimization and characterization of these newer mutagenesis methods is limited. With the aim of minimizing experimental effort and maximize catalytic yield, researchers turn back to natural systems for new evolution approaches. Below we discuss current methods for *in vivo* mutagenesis and selection of biocatalysts.

1.2.1 Methods for DNA mutagenesis

Random mutagenesis. Though we are now able to control DNA mutagenesis at single nucleotide precision, *in vitro* random mutagenesis remains one of the most popular approaches for enzyme evolution, simply because it does not require any prior knowledge of protein structure or mechanism.[11,12] Exhaustive random mutagenesis finds further

support in successful evolution experiments where beneficial mutations were found well outside of the protein active site.[13,14]

While it is widely performed *in vitro* using PCR,[3,14-18] random mutagenesis can also be carried out *in vivo* in 'mutator' strains in which one or more of the DNA repair components are compromised.[19] Accumulation of mutations during DNA replication allow very large library size independent of DNA transformation efficiency, as demonstrated for both single enzyme[20,21] and genomes.[22] Nonetheless, repair-deficient strains inevitably randomize the host chromosomal DNA at a higher rate, limiting the potential for carrying out multiple rounds of mutagenesis and selection without removing the target gene of interest from the host strain. Furthermore, the mutagenesis cannot be targeted to residues of particular interest.[2]

While toxic to the host cell in its current inception, an error-prone DNA polymerase engineered by Loeb and co-workers offers a clever strategy for carrying out *in vivo* mutagenesis without chromosomal modification.[23] Polymerase I, specifically active in the replication of endogenous ColEI plasmids, has been mutated to have 80,000-fold increased mutagenesis over the natural enzyme. It was then used to improve beta-lactamase antibiotic resistance, increasing it 150-fold.

Significantly, Tsien and co-workers outperformed traditional *in vitro* mutagenesis techniques performing multiple rounds of random *in vivo* mutagenesis by somatic hypermutation. Taking advantage of the fact that the rate of somatic hypermutation, a mechanism unique to B cell lines, is 10^6 times higher than that in the rest of the genome,[5] they demonstrated substantial improvement of a red fluorescent protein variant.[24]

Borrowing a natural mutagenesis approach, mutation accumulation by neutral drift has been used to allow comprehensive examination of sequence space for superior mutants.[25] In this method, mutagenesis occurs while the original enzyme functionality is retained. The accumulation of non-deleterious mutations facilitates diversity that is then tested for new properties.[26] For example, Tawfik and colleagues showed serum paraoxonase and P450 enzymes[26] to evolve under neutral drift exhibiting significant changes in promiscuous activities and enzyme specificity.[27-29]

Targeted Mutagenesis. As we cannot realistically test all possible random sequence, targeted mutagenesis is necessary to enrich the variant library with sequences that are most likely to produce a functional protein.[3] Furthermore, Studies of protein tolerance to mutagenesis show each random mutation is more likely to damage then enhance catalytic activity, and typically inactivates many of the remaining active protein variants[30] and disrupts synergistic beneficial mutations.[31] In addition, it was found that most enzymes inherently carry additional, weaker promiscuous activities that require only a small number of mutations to enhance.[32,33]

'S*mart'* focused libraries are increasingly being used[34] to identify likely sites for evolutionary improvement[35,36] permitting faster screening of smaller, high quality libraries. The recent evolution of transaminase for synthesis of the diabetic drug stagliptin[37] using homology modeling, site directed mutagenesis and DNA recombination by shuffling demonstrate that combination of methods is often advantageous for successful enzymes evolution.

Of the methods available for *in vivo* targeted mutagenesis, Homologous Recombination (HR) is central to diversification of natural systems. Once identified in

yeast, DNA recombination quickly became a useful tool for genetic manipulation[38,39] [40] and centered much of eukaryotic genetic studies on S. cerevisiae as a model system.[10,41] The highly efficient HR machinery of *S. cerevisiae* is now routinely used to insert and knockout genes for strain and plasmid construction (Figure 1-2).[10,39,42-45] Significantly, recombination-based mutagenesis methods are also found to be more likely to retain synergistic mutations.[31]
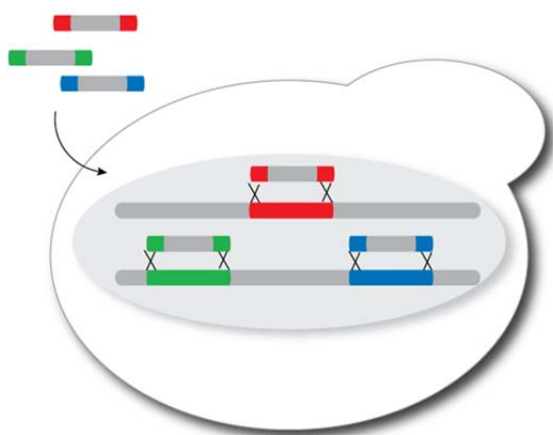
Figure 1-2. Homologous recombination for targeted DNA mutagenesis. The highly efficient homologous recombination mechanism in S. cerevisiae can be used to target desired sequenced. Homology sequences on either end of a DNA cassette guide its integration at the respective locus.

A handful of reports exploit HR for library mutagenesis of enzymes. Already in 1995 Sherman and co-workers used short DNA oligonucleotides to randomized a single position in the chromosomal CYC1 gene in yeast.[46] Novo Nordisk and others reported the use of HR in yeast to shuffle beneficial mutations obtained from previous rounds of directed evolution.[47-49] Resnick and co-workers have constructed *Dellito perfetto*, an *in vivo* genetic method for site directed mutagenesis in yeast.[50,51] Remarkably, they specifically enhanced the efficiency of recombination by induction of DSB upon transformation of oligonucleotides,[52] reaching recombination efficiency >10% using

ssDNA to replaces a counter-selectable marker. Furthermore, Wittrup and co-workers applied *in vivo* loop shuffling to engineer a fibronectin variant with picomolar affinity to lysozyme.[53]

Although naturally less efficient relative to *S. cerevisiae*, recombination-based targeted mutagenesis was successfully demonstrated in *E. coli*[54-60] and mammalian systems.[45,61-65] The Court laboratory has shown that recombination machinery from bacteriophage λ can be used to support efficient recombination in *E. coli* in a technique they call recombineering,[66,67] and others have used recET proteins in *E. coli* to demonstrate recombination between linear and circular DNA.[68] This technology was further utilized by Church and colleagues for high-throughput targeted mutagenesis using short DNA nucleotides (see below).

HR holds the possibility for a simple and powerful technique for *in vivo* directed evolution, offering targeted, well controlled and highly efficient library mutagenesis. Our laboratory have previously demonstrated the use of *S. cerevisiae* recombination mechanism for directed evolution of plasmid-borne and chromosomal targets for the engineering of enzymes[69] and pathways.[70] Ongoing efforts are made to develop a heritable recombination system, where the induction of mutagenesis is performed completely *in vivo*. By induction of homing endonucleases, a mutagenic DNA cassette is liberated from its carrier plasmid and guided by homologous sequences to recombine with its target gene. The process can be iterated multiple times, and beneficial mutations can be further accumulated by mating strains carrying different mutagenic libraries. This strategy should be extendable to chromosomal modifications in subsequent versions of the technology. Looking forward, there is significant potential for the development of diverse approaches to *in vivo* DNA library mutagenesis.

1.2.2   Genetic Selection methods

The numbers game. The complexity of mutagenic libraries quickly increases with the number of targeted sites. For example, while one random amino acid exchange in a 300-aa protein would provide 5,700 different mutants, and with an average of two simultaneous exchanges, the number of mutants increases to 16 million.[36,71] These numbers suggest that the method of isolation step of a desired functional variant from a large pool of inactive ones may very well be the bottleneck for a successful evolution experiment, pressing the need for efficient, high throughput isolation schemes.

Functional mutants can be isolated by way of screening individual variants for the desired activity, or alternatively by creating the conditions to eliminate all irrelevant variants so that only the desired phenotype is observed (Figure 1-3). While *in vitro* screening technologies have greatly improved in recent years, for example by the availability of fluorescence activated cell sorting (FACS),[72] and these systems eliminate cellular limitations on the attainable library size such as transformation efficiency or toxicity of proteins, screening becomes more difficult as the variant library size increases.[5,11] It is when very large variant libraries are tested that selection schemes become highly advantageous.[73]
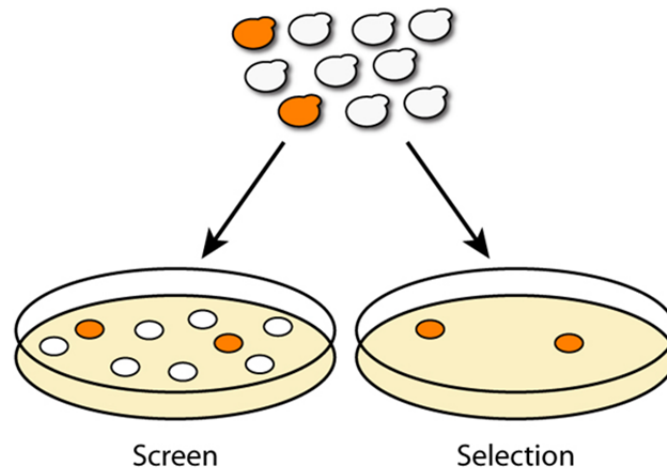
Figure 1-3. Selection vs. screen for isolation of a desired phenotype. Screening is performed by growing all genetic variants under non selective conditions. The desired phenotype is detected by screening all colonies for visible phenotype. Alternatively, when subjecting all variants to selective growth conditions, only colonies expressing the desired selectable trait are isolated.

The efficiency of a selection is a function of its throughput, sensitivity, precision and dynamic range.[33] The most efficient *in vivo* selection scheme is the one provided by cell growth by genetic complementation. In genetic selections, the desired protein functionality is connected to cell growth by some detectable phenotype. The coupling of functionality to cell growth allows measuring cell fitness through competition, where millions of DNA variants are individually expressed in each cell, and active enzymes can readily be isolated by a simple growth assay using a defined, controlled selective pressure.[12,74] Furthermore, identifying an efficient variant involves not only detection of protein activity, but also the unique DNA sequence encoding it. Cell-based *in vivo* assays greatly simplify this process by co-compartmentalization of both protein and DNA encoding it.[6]

However, genetic assay are inherently limited to selectable cellular functions, and although they offer high throughput (~10^8 cells), they often display limited sensitivity.[33] In fact, many desired biological functions are difficult to select for, and high throughput screens and selections are often developed for each unique application.[33] Our lab has reported an *in vivo* method detecting functional enzymes by coupling catalytic activity to expression of selectable reporter gene.[75,76] However, as increasingly demanding functions are desired by directed evolution, there is much need for new methods for the isolation of complex catalytic and cellular functions.

1.3   Engineering multi-component biological systems

Biological engineering is broadly applicable and is giving rise to promising research in cost effective production of foods, foods, fuels and pharmaceuticals in microbial 'cell factories',[77,78] diagnostics development and drug delivery.[79] Recent progress in molecular and systems biology has greatly expanded the toolkit available for biologists, chemists and engineers to selectively refine not only single genes and proteins, but entire living cells. In fact, we are witnessing a shift away from individual enzymatic reactions and towards systems of interacting biochemical reactions,[80] pushing the boundaries of recombinant DNA technology to engineer new and useful biological systems.[1]

With accumulated knowledge of the structure of single genes and proteins and the methodologies to manipulate them, researchers turn to engineer complex biological systems to effectively manipulate living systems.[81,82] Synthetic biology is therefore distinct from traditional molecular biology or genetics in that it is an engineering-driven effort,

seeking to make the design and construction of complex biological systems easier, faster, cheaper and more efficient.[1]

Much like in the case of enzyme engineering, we turn to directed evolution to circumvent our gap in knowledge of biological behavior. Yet the unique nature of live biological networks uncovered new challenges for which the extrapolation of current technologies to multi-component systems has proven not only extremely laborious but also inefficient.

Thus, to accomplish an engineering technology based on living cells[79], novel systems-level techniques are sought for that are compatible with large complex biological systems[83] (Figure 1-4). Below, we consider key challenges of systems level engineering, and discussion emergent techniques for manipulation of pathways and entire cells.

1.3.1   Challenges of scaling up directed evolution methods

The tools used for system-level engineering are often the same ones developed for single genes, addressing the entire biological network instead of its constituent parts. Yet reductionist methods prove laborious and inefficient and require large-scale and high throughput adaptations. Here, we touch upon the main challenges in directed evolution of multi-component systems.

Predictability of biological behavior. Biological systems display a degree of modularity that we only partly understand. The inherent sensitivity of biological systems to their environment and their natural tendency to mutate and evolve as the environment changes is a significant hurdle in the engineering of predictable biological systems, while it is also its greatest asset.[1] In order to systematically engineer cellular properties, we will need a better understanding of the principles governing cellular architecture and regulation.

We further need to systematically standardize biological functions and building blocks, an effort pioneered by the BioBrick repository for biological parts.[79]

Large Scale DNA manipulation. Perhaps the most obvious technical challenge one faces for system engineering is the scale of DNA to be manipulated. The length of even the simplest designed network may be several thousand base pairs long, and the smallest known genome is several million bases.[84] This scale of DNA manipulation exceeds the currently available vectors and the limits of DNA elongation by PCR and chemical DNA synthesis methods.[85] Hence, it is often necessary to shift from engineering of small mobile genetic elements, such as transposons, DNA cassette and plasmids, to engineering of chromosomes, a much more challenging undertaking.
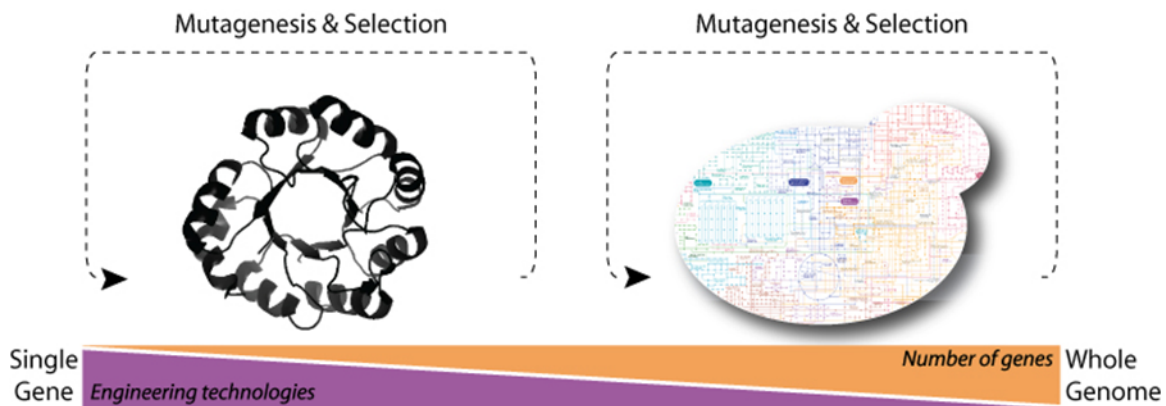
Figure 1-4. Expanding single enzyme engineering to multiple-component biological systems. As the number of genes and biological complexity increases, less technological tools are available for efficient manipulation of DNA and directed evolution.

Mutagenesis efficiency. The complexity and size of genetic libraries exponentially increases with the addition of each new target.[36,71] as targeted mutagenesis methods

become highly inefficient in the context of multiple targets,[69,70] simpler, more efficient mutagenesis methods are required for rapid and efficient mutagenesis.

Computational assisted design. metabolic engineering and synthetic biology both build on quantitative expansion of early studies in metabolic flux and genetic components, respectively.[80] Computational models for both metabolic flux analysis (MFA),[86,87] and synthetic genetic circuits[88] are being developed to allow researchers to embark on ambitious engineering of living cellular networks.

Integration with live host. While the expression of single enzymes is now commonly done in molecular biology research, the engineering of multiple heterologous enzymes in the same host often significantly effects host homeostasis and requires extensive optimization.[80] As all metabolites are derived from a limited set of 12 precursor originating from central carbon metabolism,[89] the interruption of metabolic flux is highly likely. One has to consider not only the metabolites feeding into and out of a heterologous genetic circuit, but also possible disruption of energy and information flux of the entire host cell. Another consideration could be toxicity of heterologous metabolic products, especially for biofuel and drug production.[86].

In vivo design by genetic engineering is therefore less limited by our ability to produce DNA fragments *in vitro*, and more by the methods available for DNA assembly and functional optimization in a living host. The variety of natural parts and the possibility of evolution of these parts bring great promise to the field. Next, we discuss recent advances in pathway and whole cell engineering.

1.3.2 Emerging Technologies for pathway and whole cell engineering

Two major synergistic fields drive the development of new methods for engineering of multi-component systems.[79,93] Metabolic engineering aims at the directed improvement of product formation through the modification of specific biochemical reactions,[80,90] and synthetic biology attempts to construct non-natural devices using biological building blocks.[85,91,92] The two fields interconnect in the investigation of metabolic flux, network regulation and high-throughput DNA manipulation.[92]

For example, advances in metabolic engineering for industrial microbiology provided successful examples for production of commodity chemicals,[90] biofuels[93-97] and drugs,[98-116117] with significant recent breakthroughs include expression of drug precursors for the anticancer drug Taxol[118] and antimalarial drug artemisinin.[119] However, though many studies have demonstrated the feasibility of metabolic engineering, only few achieve the yields, rates or titers required for practical production process.[90]

Synthetic biology applications include the formation of artificial DNA replicating system using unnatural nucleotides,[120] the incorporation of unnatural amino acids into proteins using an orthogonal translation system[121] and the introduction of hydrogen producing machineries[122] into heterologous hosts.

DNA manipulation. New methods for DNA manipulation and assembly play key component in advances for multi-component biological systems. High speed, low cost DNA synthesis and sequencing methods with very low error rate[123] are becoming more and more available,[85,124] and the capacity and cost of synthesis, manipulation and analysis of DNA has exponentially increased.[85] While *in vitro* biochemical enabled the leap from oligos to genes synthesis, *in vivo* processing made possible the leap from genes to

genomes, as large scale DNA manipulation relies heavily on *in vivo* recombination techniques.

Several laboratories have demonstrated DNA assembly directly in cells[125,126] using homologous recombination. Gibson et al[127-129] constructed a synthetic microbial genome by assembly of synthetic DNA cassettes. Church and colleagues[130] developed multiplex automated genome engineering (MAGE), an *in vivo* platform to combinatorially mutagenize the ribosome binding sites (RBS) using recombination of short oligonucleotides in *E. coli*. They bring genomic recombination efficiency up to ~30% to delete, replace or insert short (<30bp) DNA fragments. In a recently described technique, Conjugative Assembly Genome Engineering (CAGE), bacterial conjugation is applied to combine large contiguous regions of *E. coli* chromosomes,[131] to site-specifically replace all 314 TAG stop codons with synonymous TAA codons in parallel across the 32 *E. coli* strains later combined.

Pioneering work by Itaya et al. uses homologous recombination in conjunction with elegant marker recycling strategies to integrate constructs 16 kb-3.5 Mb in size into the *Bacillus subtilis* genome,[132,133] while Posfai and coworkers deleted large parts of *E. coli* genome to form a compact genome, eliminating unstable elements.[134] In a remarkable genome design effort, Boeke and co-workers have constructed a partially synthetic eukaryotic chromosome, maintaining near wild-type phenotype while carrying an inducible genetic scrambling system to facilitate future studies.[135]

Our lab has recently demonstrated Reiterative Recombination, a highly efficient technique for assembly of DNA into the yeast chromosome.[70] The method relies on alternating pairs of endonucleases and selective markers, harnessing DSB-induced

recombination to sequentially incorporate DNA directly into the chromosome. Importantly, this system was further shown to be compatible with the construction of up to 10^4 variant libraries of metabolic pathway.

Regulation of gene expression. The direct regulation of gene transcription has been a key tool for biological engineering since the early days of molecular research.[136] Traditionally, regulation of gene expression was performed by either complete deletion or overexpression. However, it is rarely the case in the cell, where metabolic steady state is constantly fine-tuned by balancing transcription, translation and degradation levels. It is desired to incorporate such dynamic regulation of gene expression to successfully design biological networks.

Considering the extensive use of promoters to regulate gene transcription,[136] it is surprising that relatively few studies attempted to extensively compare promoter strength.[137] Systematic characterization and standardization of promoter properties has only recently taken place.[138,139] Significantly, several groups have constructed well characterized promoter libraries in the common hosts E.coli [140-142], *S. cerevisiae* [143 144], P. pastoris[145] and mammalian systems[146], providing precise quantitative control of *in vivo* gene expression.

Utilizing a targeted mutagenesis approach, Jensen and coworkers demonstrated the functionality of synthetic promoter libraries by randomizing the -35 to -10 upstream 'spacer' region of prokaryotic [140,147], yeast[144] and mammalian genes [146]. Using random mutagenesis, Stephanopoulos and colleges used constitutive promoters to construct and rigorously characterize libraries of promoter variants with broad range activity to be use in *E. coli* and *S. cerevisiae*. Both libraries demonstrated reproducible, homogenous, and

linear relationship between promoter strength and reporter activation, and further used to modulate gene expression for E.coli lycopene production[141], and S. cerecisiae glycerol yield.[143] Sauer and coworkers[142] recently designed an insulated bacterial promoter library spanning two orders of magnitude by adding sequences extending beyond the core-polymerase binding region in both the 5'- and 3'- directions, to reduce stimulatory and repressive effects of adjacent genetic elements.

Synthetic RNA regulation provides another effective tool for modulating gene expression in response to changes in cellular metabolism.[92] A variety of these riboswitch regulators have been used to interact with mRNA in response to small molecules[148,149] or silence the expression of target genes.[150,151]

A method termed Global Transcription Machinery Engineering (GTME) enables broad perturbation of the entire cellular transcriptome by alteration of global regulator proteins.[152] It was utilized for engineering of sigma factors in *E. coli* or TFIII regulator in *S. cerevisiae*, eliciting multi-gene response to enhance ethanol tolerance and lycopene yield, respectively.[153] The use of gene array technology enables investigation of whole cell changes to identify such regulators.[4]

Engineering synthetic cellular pathways. Several examples exist where directed evolution have been implemented to engineer multiple-component pathways and novel genetic circuits. The shuffling of metabolic pathways was demonstrated for making chimeric nonribosomal peptides and polyketide analogs[154-156] and further improving their functionality by directed evolution.[157] Novel cellular behavior has been obtained using artificial "genetic oscillators",[81] synthetic genetic circuits designed to elicit programmable response to predefined stimuli.[158-162]

For example, synthetic auto regulatory feedback was used to detect 'cellular memory' in yeast[163,164] in response to DNA damage.[162,163] Arnold and coworkers have engineering microbial consortium communication signals, such that gene expression response is generated in a cell density-dependent fashion. Response is elicited by quorum sensing molecule,[165] and was shown to sustain also upon biofilm formation.[166,167]

Collins and coworkers utilized in silico modeling to design artificial gene networks guiding the production of a 'timer' controlling S. cereviaie sedimentation phenotype.[168]. Recently, the construction of an 'olfactory yeast' strain was reported, engineering S. cerevisiae with mammalian olfactory signaling pathway to detect the presence of DNT explosive mimic.[169] Furthermore, spatial organization of bacterial metabolism was addressed by Delebecque et al.,[122] who rationally designed multidimensional RNA structures as scaffolds for controlled *in vivo* organization of a hydrogen-producing pathway.

1.4 Conclusions

New technologies are emerging for faster, more efficient and cost effective engineering of complex biological systems. To enable breakthroughs in our cell engineering capabilities, the efficiency of existing molecular tools must be challenged and creatively adapted to be compatible with complex cellular networks. Modern directed evolution approaches, harnessing cellular processes to explore the interplay of random and rational design, are expected to provide ever more powerful tools for biological engineering.

1.5   References

1   Arkin, A. P. & Fletcher, D. A. Fast, cheap and somewhat in control. *Genome Biol* **7**, 114, (2006).

2   Botstein, D. & Shortle, D. Strategies and applications of in vitro mutagenesis. *Science* **229**, 1193-1201, (1985).

3   Arnold, F. H. Combinatorial and computational challenges for biocatalyst design. *Nature* **409**, 253-257, (2001).

4   Chatterjee, R. & Yuan, L. Directed evolution of metabolic pathways. *Trends Biotechnol.* **24**, 28-38, (2006).

5   Blagodatski, A. & Katanaev, V. L. Technologies of directed protein evolution in vivo. *Cell. Mol. Life Sci.* **68**, 1207-1214, (2011).

6   Lin, H. & Cornish, V. W. Screening and selection methods for large-scale analysis of protein function. *Angew. Chem. Int. Ed. Engl.* **41**, 4402-4425, (2002).

7   Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* **15**, 553-557, (1997).

8   Smith, M. In vitro mutagenesis. *Annu. Rev. Genet.* **19**, 423-462, (1985).

9   Bloom, J. D. *et al.* Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* **15**, 447-452, (2005).

10   Klinner, U. & Schafer, B. Genetic aspects of targeted insertion mutagenesis in yeasts. *FEMS Microbiol. Rev.* **28**, 201-223, (2004).

11   Bloom, J. D. *et al.* Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* **15**, 447-452, (2005).

12   Sen, S., Venkata Dasu, V. & Mandal, B. Developments in directed evolution for improving enzyme functions. *Appl. Biochem. Biotechnol.* **143**, 212-223, (2007).

13   May, O., Nguyen, P. T. & Arnold, F. H. Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. *Nat. Biotechnol.* **18**, 317-320, (2000).

14   Oue, S., Okamoto, A., Yano, T. & Kagamiyama, H. Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues. *J. Biol. Chem.* **274**, 2344-2349, (1999).

15   Hermes, J. D., Blacklow, S. C. & Knowles, J. R. Searching sequence space by definably random mutagenesis: improving the catalytic potency of an enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 696-700, (1990).

16   Giver, L., Gershenson, A., Freskgard, P. O. & Arnold, F. H. Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 12809-12813, (1998).

17   Yano, T., Oue, S. & Kagamiyama, H. Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5511-5515, (1998).

18   Leung, D. W., Chen, E. & Goeddel, D. V. A Method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique (Philadelphia)* **1**, 11-15, (1989).

19   Greener, A., Callahan, M. & Jerpseth, B. An efficient random mutagenesis technique using an E. coli mutator strain. *Methods Mol. Biol.* **57**, 375-385, (1996).

20   Bornscheuer, U. T., Altenbuchner, J. & Meyer, H. H. Directed evolution of an esterase for the stereoselective resolution of a key intermediate in the synthesis of epothilones. *Biotechnol. Bioeng.* **58**, 554-559, (1998).

21   Callanan, M. J., Russell, W. M. & Klaenhammer, T. R. Modification of Lactobacillus beta-glucuronidase activity by random mutagenesis. *Gene* **389**, 122-127, (2007).

22   Lu, X., Hirata, H., Yamaji, Y., Ugaki, M. & Namba, S. Random mutagenesis in a plant viral genome using a DNA repair-deficient mutator Escherichia coli strain. *J. Virol. Methods* **94**, 37-43, (2001).

23   Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proc Natl Acad Sci U S A* **100**, 9727-9732, (2003).

24   Wang, L., Jackson, W. C., Steinbach, P. A. & Tsien, R. Y. Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16745-16749, (2004).

25   Bergquist, P. L., Reeves, R. A. & Gibbs, M. D. Degenerate oligonucleotide gene shuffling (DOGS) and random drift mutagenesis (RNDM): two complementary techniques for enzyme evolution. *Biomol Eng* **22**, 63-72, (2005).

26   Bloom, J. D., Romero, P. A., Lu, Z. & Arnold, F. H. Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol Direct* **2**, 17, (2007).

27   Tokuriki, N. & Tawfik, D. S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668-673, (2009).

28   Amitai, G., Gupta, R. D. & Tawfik, D. S. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J* **1**, 67-78, (2007).

29   Gupta, R. D. & Tawfik, D. S. Directed enzyme evolution via small and effective neutral drift libraries. *Nat. Methods* **5**, 939-942, (2008).

30   Guo, H. H., Choe, J. & Loeb, L. A. Protein tolerance to random amino acid change. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9205-9210, (2004).

31   Dalby, P. A. Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol.* **21**, 473-480, (2011).

32   Aharoni, A. *et al.* The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37**, 73-76, (2005).

33   Rowe, L. A., Geddie, M. L., Alexander, O. B. & Matsumura, I. A comparison of directed evolution approaches using the beta-glucuronidase model system. *J. Mol. Biol.* **332**, 851-860, (2003).

34   Mena, M. A., Treynor, T. P., Mayo, S. L. & Daugherty, P. S. Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. *Nat. Biotechnol.* **24**, 1569-1571, (2006).

35   Voigt, C. A., Mayo, S. L., Arnold, F. H. & Wang, Z. G. Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3778-3783, (2001).

36   Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc* **2**, 891-903, (2007).

37   Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305-309, (2010).

38   Rothstein, R. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol.* **194**, 281-301, (1991).

39   Raymond, C. K., Pownder, T. A. & Sexson, S. L. General method for plasmid construction using homologous recombination. *Biotechniques* **26**, 134-138, 140-131, (1999).

40   Wells, J. A., Vasser, M. & Powers, D. B. Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene* **34**, 315-323, (1985).

41   Hinnen, A., Hicks, J. B. & Fink, G. R. Transformation of yeast. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 1929-1933, (1978).

42   Rothstein, R. J. One-step gene disruption in yeast. *Methods Enzymol.* **101**, 202-211, (1983).

43   Kunes, S., Ma, H., Overbye, K., Fox, M. S. & Botstein, D. Fine structure recombinational analysis of cloned genes using yeast transformation. *Genetics* **115**, 73-81, (1987).

44   Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. *Nucleic Acids Res* **21**, 3329-3330, (1993).

45   Igoucheva, O., Alexeev, V. & Yoon, K. Oligonucleotide-directed mutagenesis and targeted gene correction: a mechanistic point of view. *Curr Mol Med* **4**, 445-463, (2004).

46   Linske-O'Connell, L. I., Sherman, F. & McLendon, G. Stabilizing amino acid replacements at position 52 in yeast iso-1-cytochrome c: in vivo and in vitro effects. *Biochemistry* **34**, 7094-7102, (1995).

47   Cherry, J. R. *et al.* Directed evolution of a fungal peroxidase. *Nat. Biotechnol.* **17**, 379-384, (1999).

48   Abecassis, V., Pompon, D. & Truan, G. High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *Nucleic Acids Res* **28**, E88, (2000).

49   Swers, J. S., Kellogg, B. A. & Wittrup, K. D. Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display. *Nucleic Acids Res* **32**, e36, (2004).

50   Storici, F., Lewis, L. K. & Resnick, M. A. In vivo site-directed mutagenesis using oligonucleotides. *Nat. Biotechnol.* **19**, 773-776, (2001).

51   Storici, F. & Resnick, M. A. The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.* **409**, 329-345, (2006).

52   Storici, F., Durham, C. L., Gordenin, D. A. & Resnick, M. A. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14994-14999, (2003).

53   Hackel, B. J., Kapila, A. & Wittrup, K. D. Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *J. Mol. Biol.* **381**, 1238-1252, (2008).

54   Borrego, B., Wienecke, A. & Schwienhorst, A. Combinatorial libraries by cassette mutagenesis. *Nucleic Acids Res* **23**, 1834-1835, (1995).

55   Zhang, Y., Buchholz, F., Muyrers, J. P. & Stewart, A. F. A new logic for DNA engineering using recombination in Escherichia coli. *Nat. Genet.* **20**, 123-128, (1998).

56   Ellis, H. M., Yu, D., DiTizio, T. & Court, D. L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6742-6746, (2001).

57   Court, D. L., Sawitzke, J. A. & Thomason, L. C. Genetic engineering using homologous recombination. *Annu. Rev. Genet.* **36**, 361-388, (2002).

58   Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* **4**, 206-223, (2009).

59   Datta, S., Costantino, N., Zhou, X. & Court, D. L. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1626-1631, (2008).

60   Costantino, N. & Court, D. L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15748-15753, (2003).

61   Donoho, G., Jasin, M. & Berg, P. Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. *Mol. Cell. Biol.* **18**, 4070-4078, (1998).

62   Vasquez, K. M., Marburger, K., Intody, Z. & Wilson, J. H. Manipulating the mammalian genome by homologous recombination. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8403-8410, (2001).

63   Agarwal, S., Gamper, H. B. & Kmiec, E. B. Nucleotide replacement at two sites can be directed by modified single-stranded oligonucleotides in vitro and in vivo. *Biomol Eng* **20**, 7-20, (2003).

64   Parekh-Olmedo, H. & Kmiec, E. B. Progress and prospects: targeted gene alteration (TGA). *Gene Ther.* **14**, 1675-1680, (2007).

65   Wu, X. S. *et al.* Increased efficiency of oligonucleotide-mediated gene repair through slowing replication fork progression. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2508-2513, (2005).

66   Costantino, N. & Court, D. L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15748-15753, (2003).

67   Sawitzke, J. A. *et al.* Recombineering: in vivo genetic engineering in E. coli, S. enterica, and beyond. *Methods Enzymol.* **421**, 171-199, (2007).

68   Zhang, Y., Buchholz, F., Muyrers, J. P. & Stewart, A. F. A new logic for DNA engineering using recombination in Escherichia coli. *Nat. Genet.* **20**, 123-128, (1998).

69   Pirakitikulr, N., Ostrov, N., Peralta-Yahya, P. & Cornish, V. W. PCRless library mutagenesis via oligonucleotide recombination in yeast. *Protein Sci.* **19**, 2336-2346, (2010).

70   Wingler, L. M. & Cornish, V. W. Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15135-15140, (2011).

71   Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the numbers problem in directed evolution. *ChemBioChem* **9**, 1797-1804, (2008).

72   Bershtein, S. & Tawfik, D. S. Advances in laboratory evolution of enzymes. *Curr. Opin. Chem. Biol.* **12**, 151-158, (2008).

73   Fastrez, J. In vivo versus in vitro screening or selection for catalytic activity in enzymes and abzymes. *Mol. Biotechnol.* **7**, 37-55, (1997).

74   Woycechowsky, K. J. & Hilvert, D. Deciphering enzymes. Genetic selection as a probe of structure and mechanism. *Eur. J. Biochem.* **271**, 1630-1637, (2004).

75   Baker, K. *et al.* Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16537-16542, (2002).

76   Lin, H. & Cornish, V. W. Screening and selection methods for large-scale analysis of protein function. *Angew. Chem. Int. Ed. Engl.* **41**, 4402-4425, (2002).

77   Boyle, P. M. & Silver, P. A. Parts plus pipes: Synthetic biology approaches to metabolic engineering. *Metab Eng*, (2011).

78   Nielsen, J. & Keasling, J. D. Synergies between synthetic biology and metabolic engineering. *Nat. Biotechnol.* **29**, 693-695, (2011).

79   Endy, D. Foundations for engineering biology. *Nature* **438**, 449-453, (2005).

80   Stephanopoulos, G. N., Aristidou, A. A. & Nielsen, J. *Metabolic Engineering Principles and Methodologies.* (Elsevier, 1998).

81   Haseltine, E. L. & Arnold, F. H. Synthetic gene circuits: design with directed evolution. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 1-19, (2007).

82   Evanko, D. Engineering life. *Nat Methods* **2**, 640, (2005).

83   Conrad, T. M., Lewis, N. E. & Palsson, B. O. Microbial laboratory evolution in the era of genome-scale science. *Mol Syst Biol* **7**, 509, (2011).

84   Baker, D. *et al.* Engineering life: building a fab for biology. *Sci. Am.* **294**, 44-51, (2006).

85   Carr, P. A. & Church, G. M. Genome engineering. *Nat. Biotechnol.* **27**, 1151-1162, (2009).

86   Koffas, M., Roberge, C., Lee, K. & Stephanopoulos, G. Metabolic engineering. *Annu Rev Biomed Eng* **1**, 535-557, (1999).

87   Armstrong, A. J. *et al.* A pharmacodynamic study of rapamycin in men with intermediate- to high-risk localized prostate cancer. *Clin. Cancer Res.* **16**, 3057-3066, (2010).

88   Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410-422, (2009).

89   Nielsen, J. It is all about metabolic fluxes. *J. Bacteriol.* **185**, 7031-7035, (2003).

90   Bailey, J. E. Toward a science of metabolic engineering. *Science* **252**, 1668-1675, (1991).

91   Krivoruchko, A., Siewers, V. & Nielsen, J. Opportunities for yeast metabolic engineering: Lessons from synthetic biology. *Biotechnol J* **6**, 262-276, (2011).

92   Boyle, P. M. & Silver, P. A. Parts plus pipes: Synthetic biology approaches to metabolic engineering. *Metab Eng* **14**, 223-232, (2012).

93   Alper, H. & Stephanopoulos, G. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat Rev Microbiol* **7**, 715-723, (2009).

94   Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86-89, (2008).

95   Steen, E. J. *et al.* Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**, 559-562, (2010).

96   Bro, C., Regenberg, B., Forster, J. & Nielsen, J. In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production. *Metab Eng* **8**, 102-111, (2006).

97   Beaudoin, F. *et al.* Heterologous reconstitution in yeast of the polyunsaturated fatty acid biosynthetic pathway. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6421-6426, (2000).

98   Ma, S. M. *et al.* Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **326**, 589-592, (2009).

99   Watanabe, K., Rude, M. A., Walsh, C. T. & Khosla, C. Engineered biosynthesis of an ansamycin polyketide precursor in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9774-9778, (2003).

100   Malpartida, F., Niemi, J., Navarrete, R. & Hopwood, D. A. Cloning and expression in a heterologous host of the complete set of genes for biosynthesis of the Streptomyces coelicolor antibiotic undecylprodigiosin. *Gene* **93**, 91-99, (1990).

101   Malpartida, F. & Hopwood, D. A. Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature* **309**, 462-464, (1984).

102   Smith, D. J., Burnham, M. K., Edwards, J., Earl, A. J. & Turner, G. Cloning and heterologous expression of the penicillin biosynthetic gene cluster from penicillum chrysogenum. *Biotechnology (N Y)* **8**, 39-41, (1990).

103   Jacobsen, J. R. & Khosla, C. New directions in metabolic engineering. *Curr. Opin. Chem. Biol.* **2**, 133-137, (1998).

104   Hopwood, D. A. *et al.* Production of 'hybrid' antibiotics by genetic engineering. *Nature* **314**, 642-644, (1985).

105   Thykaer, J. & Nielsen, J. Metabolic engineering of beta-lactam production. *Metab Eng* **5**, 56-69, (2003).

106   Isogai, T. *et al.* Construction of a 7-aminocephalosporanic acid (7ACA) biosynthetic operon and direct production of 7ACA in Acremonium chrysogenum. *Biotechnology (N Y)* **9**, 188-191, (1991).

107   Siewers, V., Chen, X., Huang, L., Zhang, J. & Nielsen, J. Heterologous production of non-ribosomal peptide LLD-ACV in Saccharomyces cerevisiae. *Metab Eng* **11**, 391-397, (2009).

108   Jacobsen, J. R., Hutchinson, C. R., Cane, D. E. & Khosla, C. Precursor-directed biosynthesis of erythromycin analogs by an engineered polyketide synthase. *Science* **277**, 367-369, (1997).

109   Kao, C. M., Katz, L. & Khosla, C. Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* **265**, 509-512, (1994).

110   Mutka, S. C., Bondi, S. M., Carney, J. R., Da Silva, N. A. & Kealey, J. T. Metabolic pathway engineering for complex polyketide biosynthesis in Saccharomyces cerevisiae. *FEMS Yeast Res* **6**, 40-47, (2006).

111   McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1846-1851, (1999).

112   Xue, Q., Ashley, G., Hutchinson, C. R. & Santi, D. V. A multiplasmid approach to preparing large libraries of polyketides. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11740-11745, (1999).

113   Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171-1176, (2005).

114   Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* **291**, 1790-1792, (2001).

115   Minami, H. *et al.* Microbial production of plant benzylisoquinoline alkaloids. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7393-7398, (2008).

116   Chappell, J. Production platforms for the molecular pharming of alkaloid diversity. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7897-7898, (2008).

117   Hawkins, K. M. & Smolke, C. D. Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat. Chem. Biol.* **4**, 564-573, (2008).

118   Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* **330**, 70-74, (2010).

119   Westfall, P. J. *et al.* Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E111-118, (2012).

120   Piccirilli, J. A., Krauch, T., Moroney, S. E. & Benner, S. A. Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* **343**, 33-37, (1990).

121   Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498-500, (2001).

122   Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470-474, (2011).

123   Fuller, C. W. *et al.* The challenges of sequencing by synthesis. *Nat. Biotechnol.* **27**, 1013-1023, (2009).

124   Church, G. M. Genomes for all. *Sci. Am.* **294**, 46-54, (2006).

125   Shao, Z., Luo, Y. & Zhao, H. Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. *Mol Biosyst* **7**, 1056-1059, (2011).

126   Shao, Z. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res* **37**, e16, (2009).

127   Gibson, D. G. *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20404-20409, (2008).

128   Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319**, 1215-1220, (2008).

129   Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343-345, (2009).

130   Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898, (2009).

131   Isaacs, F. J. *et al.* Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348-353, (2011).

132   Itaya, M., Fujita, K., Kuroki, A. & Tsuge, K. Bottom-up genome assembly using the Bacillus subtilis genome vector. *Nat. Methods* **5**, 41-43, (2008).

133   Itaya, M., Tsuge, K., Koizumi, M. & Fujita, K. Combining two genomes in one cell: Stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15971-15976, (2005).

134   Posfai, G. *et al.* Emergent properties of reduced-genome Escherichia coli. *Science* **312**, 1044-1046, (2006).

135   Dymond, J. S. *et al.* Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471-476, (2011).

136   Reznikoff, W. S., Miller, J. H., Scaife, J. G. & Beckwith, J. R. A mechanism for repressor action. *J. Mol. Biol.* **43**, 201-213, (1969).

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

138   Sun, J. *et al.* Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in Saccharomyces cerevisiae. *Biotechnol. Bioeng.*, (2012).

139   Blount, B. A., Weenink, T., Vasylechko, S. & Ellis, T. Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS One* **7**, e33279, (2012).

140   Jensen, P. R. & Hammer, K. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl. Environ. Microbiol.* **64**, 82-87, (1998).

141   Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12678-12683, (2005).

142   Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res* **39**, 1131-1141, (2011).

143   Nevoigt, E. *et al.* Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **72**, 5266-5273, (2006).

144   Jeppsson, M., Johansson, B., Jensen, P. R., Hahn-Hagerdal, B. & Gorwa-Grauslund, M. F. The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant Saccharomyces cerevisiae strains. *Yeast* **20**, 1263-1272, (2003).

145   Hartner, F. S. *et al.* Promoter library designed for fine-tuned gene expression in Pichia pastoris. *Nucleic Acids Res* **36**, e76, (2008).

146   Tornoe, J., Kusk, P., Johansen, T. E. & Jensen, P. R. Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. *Gene* **297**, 21-32, (2002).

147   Jensen, P. R. & Hammer, K. Artificial promoters for metabolic optimization. *Biotechnol. Bioeng.* **58**, 191-195, (1998).

148   Bayer, T. S. & Smolke, C. D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat. Biotechnol.* **23**, 337-343, (2005).

149   Win, M. N. & Smolke, C. D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456-460, (2008).

150   Callura, J. M., Dwyer, D. J., Isaacs, F. J., Cantor, C. R. & Collins, J. J. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15898-15903, (2010).

151   Isaacs, F. J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* **22**, 841-847, (2004).

152   Alper, H. & Stephanopoulos, G. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng* **9**, 258-267, (2007).

153   Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565-1568, (2006).

154   Kittendorf, J. D. & Sherman, D. H. Developing tools for engineering hybrid polyketide synthetic pathways. *Curr. Opin. Biotechnol.* **17**, 597-605, (2006).

155   McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. Engineered biosynthesis of novel polyketides. *Science* **262**, 1546-1550, (1993).

156   Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171-1176, (2005).

157   Fischbach, M. A., Lai, J. R., Roche, E. D., Walsh, C. T. & Liu, D. R. Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11951-11956, (2007).

158   Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* **469**, 212-215, (2011).

159   Kobayashi, H. *et al.* Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8414-8419, (2004).

160   Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338, (2000).

161   Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. *Cell* **113**, 597-607, (2003).

162   Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516-519, (2008).

163   Ajo-Franklin, C. M. *et al.* Rational design of memory in eukaryotic cells. *Genes Dev.* **21**, 2271-2276, (2007).

164   Burrill, D. R. & Silver, P. A. Synthetic circuit identifies subpopulations with sustained memory of DNA damage. *Genes Dev.* **25**, 434-439, (2011).

165   Haseltine, E. L. & Arnold, F. H. Implications of rewiring bacterial quorum sensing. *Appl. Environ. Microbiol.* **74**, 437-445, (2008).

166   Brenner, K., Karig, D. K., Weiss, R. & Arnold, F. H. Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 17300-17304, (2007).

167   Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16587-16591, (2002).

168   Ellis, T., Wang, X. & Collins, J. J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* **27**, 465-471, (2009).

169   Radhika, V. *et al.* Chemical sensing of DNT by engineered olfactory yeast strain. *Nat Chem Biol* **3**, 325-330, (2007).

Chapter 2

Library Mutagenesis of PRA-Isomerase via Oligonucleotide

Recombination

2.1   Chapter outlook

The directed evolution of biomolecules with new functions is largely performed *in vitro*, with PCR mutagenesis followed by high-throughput assays for desired activities. As synthetic biology creates impetus for generating biomolecules that function in living cells, new technologies are needed for performing mutagenesis and selection for directed evolution *in vivo*. Homologous recombination, routinely exploited for targeted gene alteration, is an attractive tool for *in vivo* library mutagenesis, yet surprisingly is not routinely used for this purpose. Here, we report the design and characterization of a yeast-based system for library mutagenesis of protein loops via oligonucleotide recombination. In this system, a linear vector is co-transformed with single-stranded mutagenic oligonucleotides. Using repair of nonsense codons engineered in three different active-site loops in the selectable marker *TRP1*, we first optimized the recombination efficiency at single and multiple protein loops. Then, the utility of this system for directed evolution was explicitly tested by selecting functional variants from a mock library of $1:10^6$ wild-type:nonsense codons. Sequencing showed that oligonucleotide recombination readily covered this large library, mutating not only the target codon but also encoding silent mutations on either side of the library cassette. Together these results establish oligonucleotide recombination as a simple and powerful library mutagenesis technique and advance efforts to engineer the cell for fully *in vivo* directed evolution.

2.2   Introduction

As advances in synthetic biology improve our cell engineering capabilities and increasing research efforts are aimed at generating networks of molecules that function *in vivo*, technologies are needed to allow directed evolution to be carried out directly in the cell. Homologous recombination (HR) holds the possibility for a simple and powerful library mutagenesis technique.

The highly efficient HR machinery of *S. cerevisiae* is now routinely used, both with single-stand (ss)[1-3] and double-stranded (ds)[41,48-50] DNA, to insert and knockout genes for strain and plasmid construction.[3,6-13] Classical studies in yeast genetics have laid the foundations not only for efficient delivery of foreign DNA[1,14,15] but also for informed design of plasmids[6,16] and oligonucleotides[2,8,17,18] as well as the elucidation of the mechanism underlying homologous recombination.[19,20,24]

For example, Sherman and others explored synthetic oligonucleotide recombination into the yeast chromosome by way of DNA transformation, and found that the length of the oligonucleotide and homology region, as well as its GC content greatly contributed to the efficiency of integration.[21-26] Interestingly, it was found that the frequency of gene targeting depends on the number of target copies[27] and the targeted DNA strand.[28] It was further shown that the efficiency of transformation was enhanced by using a large DNA 'carrier' molecule.[29] Gene alteration by oligonucleotide recombination has also been reported in *E. coli*[30-32] and mammalian cells,[10,33-36] where various methods are used to enhance the efficiency of HR which is naturally significantly low.[37]

Yet this prior work predominantly focused on the recombination of individual cassettes targeting single loci,[38-40] whereas the efficiency of simultaneously mutating, for

example, multiple loops in a protein or multiple distant genes is yet to be explored. Furthermore, while several recent studies demonstrate the potential of HR for the formation of libraries,[41,42] it is yet to become a mainstay technology for library generation and directed evolution. Working towards truly *in vivo* targeted mutagenesis requires that the cell carries the potential library and shuffles upon induction. Reported optimization and characterization of newer *in vivo* mutagenesis methods is limited. Looking forward, there is significant potential for the development of diverse approaches to *in vivo* DNA library mutagenesis.

Here, as a first step towards the long-term goal of developing HR as a robust technology for *in vivo* library mutagenesis, we adapt HR techniques routinely used for targeted gene alteration for cassette library mutagenesis of multiple protein loops. Inspiration for this notion comes both from the longstanding use of bacterial mutator strains to generate libraries of random DNA mutations[43] and the widespread use of HR for making gene deletions.[11] We designed a recombination-based system specifically to meet the needs of protein directed evolution and challenged this system to search large libraries of mutations.

The mutagenesis is affected by simple co-transformation of a linearized vector carrying the target gene and ss oligonucleotide(s) encoding the library mutations under standard electroporation conditions. Notably, with optimization of DNA transformation efficiency, it is currently possible to achieve 10^7-10^9 transformants which is sufficient for library experiments.[44,45] The selectable TRP1 gene was used as the model system for the development of this technology. Three different active site loops were engineered with nonsense codons, and the efficiency of repair by oligonucleotide recombination was

studied. First, the efficiency of oligonucleotide recombination was optimized at a single loop. Then, these optimized conditions were used to evaluate the feasibility of simultaneously mutating two or three loops. Finally, the utility of oligonucleotide recombination for directed evolution was challenged by carrying out a mock library selection. This library-oriented approach expands the use of HR for DNA manipulation in yeast, offering a straightforward method for mutagenesis *in vivo* that could ultimately replace current PCR-based *in vitro* methods.

2.3   Results

2.3.1   Trp1 model system

Repair of nonsense codons in the classic yeast selection marker TRP1 presented a convenient model to develop our oligonucleotide recombination system. The gene product N-(5'-phosphoribosyl)-anthranilate isomerase (PRAI) is part of a large family of TIM-barrel enzymes, a common fold shared by approximately 10% of all known enzyme.[46] Conveniently, the active-site in TIM-barrel enzymes is positioned on loops extending between α-helices and β-sheets (Figure 2-1), making it particularly well placed for oligonucleotide recombination. The enzyme yPRAI catalyzes an essential step in the tryptophan biosynthesis pathway of *S. cerevisiae* and thus offers a simple selection for growth in the absence of tryptophan.[47] Nevertheless, oligonucleotide recombination as characterized in this work is not limited to mutagenesis of protein loops, but rather is a general technology readily applicable for library mutagenesis of any multiple-component system.
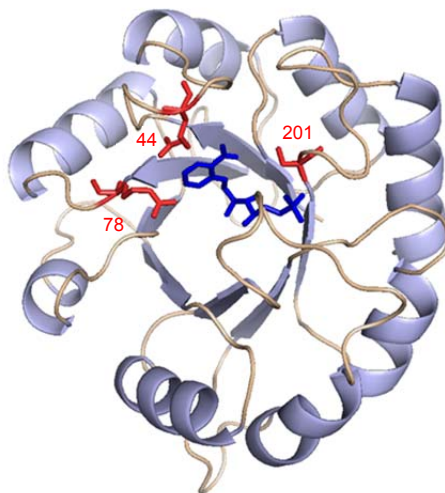
Figure 2-1. Residues selected for cassette mutagenesis of N-(5'-phosphoribosyl)-anthranilate isomerase (yPRAI). Homology model was constructed using PRAI from *Thermatoga maritima* (PDB: 1LBM[4], PDB: 1NSJ[19]) on the Swiss-Model server. Here, the product analogue (reduced 1-(o-carboxyphenylamino)-1-deoxyribulose 5-phosphate (rCDRP), dark blue) and residues 43-45 were taken from a homology model with PDB: 1NSJ.[19] Residues Arg44, Arg78 and Ser201 (red) in loops 2, 3 and 8, respectively, were replaced with stop codons to knockout protein function .

We identified three target residues, Arg44, Arg78 and Ser201, in yPRAI for engineering of ochre (TAA) nonsense codons (Figure 2-1). Although the crystal structure for *S. cerevisiae* PRAI is not available, these residues were selected based on inspection of a homology model constructed using the *Thermotoga Maritima* PRAI using Swiss Model.[48] The three target residues Arg44, Arg78 and Ser201 lie in catalytic loops 2, 3 and 8 in yPRAI, respectively. The residues vary in their distance from the 3' end of Trp1 gene, where a double strand break (DSB) is introduced, to control for the dependence of HR efficiency on proximity to the DSB. Additionally, they also vary in their distance from one

another, to explicitly test whether simultaneous mutagenesis of two loops is distance dependent. The reversion rate of plasmids carrying yPRAI with single, double or triple ochre codons was tested and found to be $<3 \times 10^6$.

As shown in Figure 2-2, the yPRAI gene was encoded on a 2μ (high copy number) vector carrying a URA3 marker, such that the linearized vector and fixing oligonucleotides are co-transformed into yeast. Cell survival on media lacking uracil depends on circularization of vector. Cell survival on media lacking tryptophan depends on repair of stop codons by oligonucleotide cassettes. The efficiency of oligonucleotide recombination could be scored simply as the ratio of colonies that survive on tryptophan and uracil deficient plates versus uracil deficient plates. We consider the recombination efficiency to be a working, rather than a theoretical, definition, because formally the vector must circularize in order to survive in uracil deficient media. Interestingly, in control experiments where the number of viable URA+ transformants was compared, transformation of linearized vector resulted in an order of magnitude less transformants than circular vector. However, no difference in the number of transformants was observed when linear vector was co-transformed with oligonucleotides (data not shown). To minimize false positives, all experiments were carried out in yeast strain NP2273, which has a complete deletion of the TRP1 gene.[49]

Initial quantification of oligonucleotide recombination efficiency was carried out by simply targeting an ochre codon. The mutagenic oligonucleotides used for Ser201 were then designed with additional silent mutations. The two additional silent mutations were added, one upstream and the other downstream of the mutated target codon, so that three codons were effectively mutagenized by each oligonucleotide. Incorporation of these

additional silent mutations both provides record that the recombination event occurred and, in combination with the ochre codon mutations, adds a unique restriction site. Importantly, control experiments using a single oligonucleotide with or without silent mutations showed the efficiency of recombination to be very similar. Hence, successful repair of the nonsense Ser201 codon in *trp1* by HR with a fixing oligonucleotide generated functional yPRAI with the expected silent mutations and led to cell survival on media lacking tryptophan.
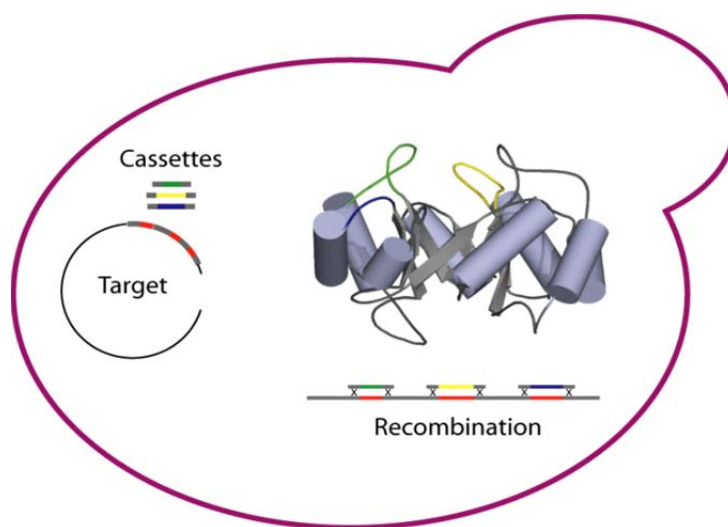
Figure 2-2. Oligonucleotide recombination via yeast HR. Oligonucleotide recombination provides a general method for generating targeted libraries of DNA mutants *in vivo*. A linearized vector expressing the target gene (gray) and linear ssDNA oligonucleotides (green, yellow, blue) are co-transformed into yeast. HR between the target gene and oligonucleotides yields libraries of mutated target gene that can be detected by growth selection.

2.3.2   Design of oligonucleotide recombination system for library mutagenesis

While there are now standard protocols for performing gene knockouts using HR, these protocols cannot be simply translated to library mutagenesis. For library mutagenesis

it is important to be able to mutate multiple regions simultaneously, to have a simple protocol that can be implemented readily and rapidly during the iterative steps of mutation and assay, and to have high efficiency HR to cover large libraries of mutations. In fact, the handful of papers that have exploited HR for library mutagenesis have used different strategies for recombination.[24,27,50,51] Thus, we designed an oligonucleotide recombination system explicitly for the needs of directed evolution with these criteria in mind.

As illustrated in Figure 2-2, the vector carrying the target gene was linearized by restriction enzymes, causing a double strand break (DSB). The DSB was introduced just downstream of the 3' terminus of the target gene TRP1 on a high copy plasmid to achieve high efficiency HR[52] in order to enable the coverage of large libraries. Since the cut site is located outside of the coding gene, as opposed to internally as in most HR technologies, it allows for the possibility of mutating multiple regions simultaneously.

It is important to distinguish the two mutagenesis schemes addressed in this work: (1) mutagenesis of multiple codons within a single loop using a single oligonucleotide, and (2) mutagenesis of multiple loops simultaneously using multiple oligonucleotides. Assuming typical protein loop sizes and homology regions, the ss oligonucleotide can simply be synthesized, making the technique very straightforward to implement.

The mutagenic ss oligonucleotides were designed to have sufficient homology for high efficiency HR and to be short enough for commercial synthesis. Specifically, each oligonucleotide was designed with 30 bp homology on both sides of the codon to be mutagenized. Published studies have established that 30 bp is the minimal homology required for high-efficiency HR.[22] Conveniently, 30 bp is shorter than a typical protein β-strand or α-helix,[46] allowing multiple mutagenic oligonucleotides to be used

simultaneously. Assuming a typical protein loop size of 10 amino acids, or 30 bp, the total oligonucleotide size (with homology) would be 90 bp. However, further investigation is required to optimize the number of residues to be mutagenized in each protein loop. Nevertheless, sequences of this length scale can be directly made by solid-phase synthesis,[53] eliminating the need for further enzymatic manipulation of the mutagenic oligonucleotide.

To optimize the likelihood of high efficiency recombination, particularly at multiple loops, the transformation was carried out at high concentrations of plasmid and oligonucleotides. While standard plasmid transformation protocols are expected to yield on average a single circular plasmid per yeast cell, it is now established that oligonucleotide transformation results in multiple oligonucleotides per cell. This is supported by multiple studies of gene targeting[27] and further demonstrated in a recent report by Venter and colleagues, showing the average yeast cell likely takes up substantially more than 25 (DNA) pieces in a single transformation experiment.[54] Furthermore, early experiments demonstrated that co-transformation of large DNA fragments enhances oligonucleotide transformation efficiency.[29] Thus, our working model is that each cell receives the plasmid and multiple copies of mutagenic oligonucleotide(s).

2.3.3 Oligonucleotide recombination at a single loop

First we examined a broad range of conditions to optimize the efficiency of oligonucleotide recombination targeting a single loop. Specifically, we sought conditions that result in the greatest number of recombinants possible while maintaining a large number of overall transformants. Based on the *S. cerevisiae* HR machinery literature, we considered three main variables: the nature of the vector DNA (circular vs. linear),[16] the

nature of the oligonucleotide DNA (ss vs. ds),[12,50] and the ratio of vector to oligonucleotide.[29] We used a sense strand oligonucleotide with forward orientation for all single strand oligonucleotide experiments, unless otherwise indicated. All experiments were performed at least in triplicate, and data were only included from experiments in which at least 10^6 transformants were achieved.

The recombination efficiency was initially optimized based on repair of the nonsense codon in the *trp1-Arg44** gene encoded on vector pNP2279 by the oligonucleotide Arg44Fix. To test whether co-transformation of circular or linearized vector yielded the largest number of recombinants, vector pNP2279 was linearized by digesting 30 bp downstream of the trp1 gene using the ClaI restriction enzyme. As Figure 2-3 shows, oligonucleotide recombination using linearized vector gave 50-fold higher recombination efficiencies than that using circular vector.

Next, recombination efficiencies via co-transformation of varying ratios of linearized vector and either ss or ds ARG44Fix oligonucleotides were compared. Figure 2-3 presents the recombination efficiencies using 1:10, 1:100, 1:1000 and 1:10000 molar ratio of linearized vector to ss or ds oligonucleotide. A 1:1000 molar ratio of linearized vector to ss oligonucleotide yielded the highest recombination efficiency, $4.5 \pm 0.5\%$. The efficiency at the 1:1000 molar ratio was 190-times that at the 1:10 molar ratio. The efficiency of ss oligonucleotide at the 1:1000 molar ratio was 4 times that of ds oligonucleotide at the same ratio.
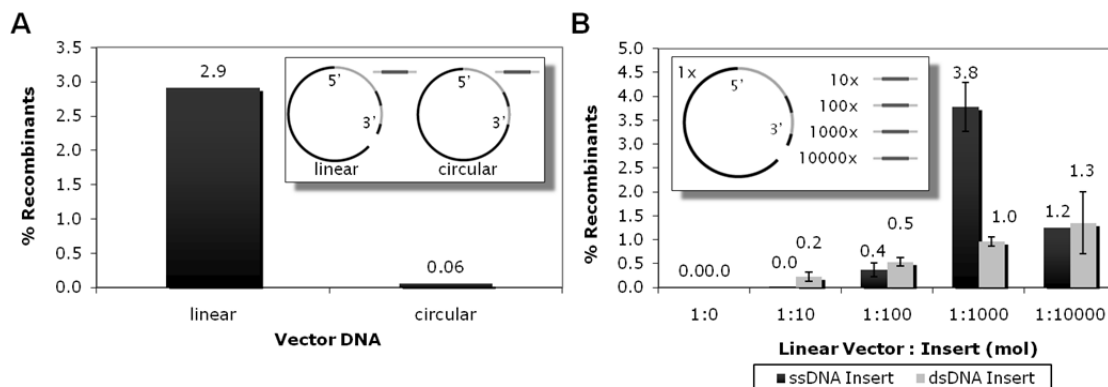
Figure 2-3. Optimization of experimental parameters for high rates of recombination. (A) Linear or circular *trp1-Arg44** vector (1 μg) is co-transformed with 5 μg of the ss oligonucleotide ARG44Fix (1:1000 mol vector: oligonucleotide) into the Δ*trp1 S. cerevisiae* strain NP2273. Co-transformation of linear vector and oligonucleotide yields the greatest percentage of recombinants. (B) Linear *trp1-Arg44** vector (1 μg) and varying amounts of ss and ds oligonucleotides ARG44Fix and ARG44Fix_ds, respectively, are co-transformed and recombinant colonies are scored by plating on SC (Ura−) and SC (Ura−Trp−) selective plates. Co-transformation of 1:1000 molar ratio vector: oligonucleotide yields the greatest percentage of recombinants. The data shown are the mean ± the standard error of at least three separate experiments.

To ensure that this measured efficiency for single-site recombination is general, the same optimized conditions were then used to repair nonsense codons in the trp1-Arg78* and trp1-Ser201* genes, encoded on vectors pNP2284 and pNP2282, respectively, by the oligonucleotides Arg78Fix and Ser201Fix (Figure 2-4). The recombination efficiencies determined at positions 78 and 201 were 1.4 ± 0.4% and 9.0 ± 0.4%, respectively. Thus, the average efficiency of oligonucleotide recombination at a single loop was 5 ± 2%, consistent with published efficiencies of HR at a DSB.[22,55]

Interestingly, the measured efficiency was slightly higher at position 201, which is closer to the DSB, although there is not a correlation between the distance from the DSB and the efficiency of recombination. The small differences in measured efficiencies alternatively may arise from the permissiveness of the individual residue to amino acid or codon substitutions. Indeed in the mock selection experiment *vide infra*, we recovered not only the encoded TCC Ser codon, but also non-encoded codons (see below).

Sequence analysis of recombinant colonies repaired at position Ser201 revealed that all colonies carried the oligonucleotide-encoded fixing codon (TCC) (Figure 2-5). Interestingly, 80% of tested colonies (16/20) were found to carry both downstream and upstream silent mutations, whereas 20% carried only the upstream silent mutation. This could be caused by either partial incorporation of the oligonucleotide or, alternatively, suggests a role for mismatch repair mechanism in oligonucleotide recombination. Further investigation is required to determine the underlying recombination mechanism.

2.3.4   Oligonucleotide recombination at multiple loops

With optimized conditions for oligonucleotide recombination at a single loop, we proceeded to test the efficiency of simultaneous mutagenesis at two and three loops. These experiments were carried out essentially as for the single-site recombination, except that equal molar quantities of the appropriate combinations of two or three of the oligonucleotides Arg44Fix, Arg78Fix, or Ser201Fix were co-transformed with the linearized plasmids pNP2283, pNP2280 or pNP2281. Simultaneous mutagenesis at positions Arg44 and Arg78 had an efficiency of 0.11 ± 0.03%, and simultaneous mutagenesis of positions Arg44 and Ser201 had an efficiency of 0.32 ± 0.15% (Figure 2-4). Oligonucleotide recombination at two loops simultaneously therefore had an average

efficiency of 0.2 ± 0.1%. The efficiency may be slightly higher when the two loops are a greater distance from one another, but the difference is at most slight.
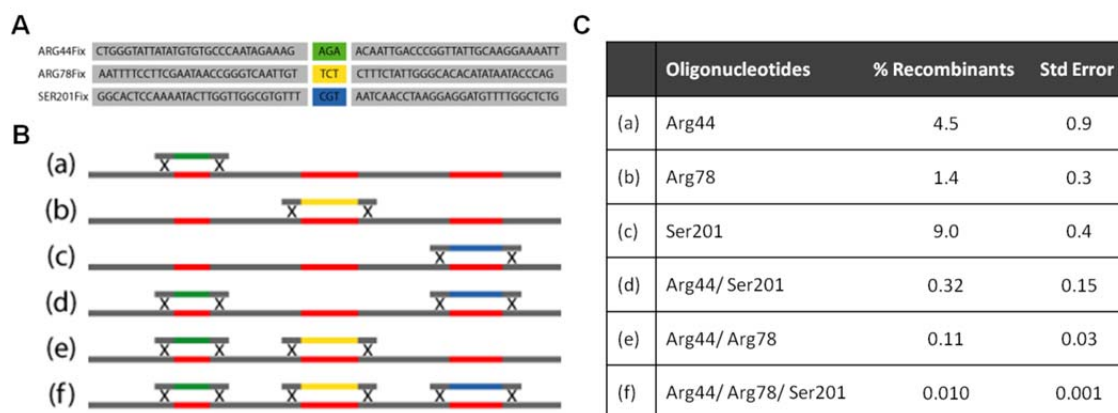
Figure 2-4. Efficiency of oligonucleotide recombination at multiple loops. Linear vectors carrying *trp1-Arg44**, *trp1-Arg78**, *trp1-Ser201**, *trp1-Arg44*Arg78**, *trp1-Arg44*Ser201**, and *trp1-Arg44*Arg78*Ser201** were co-transformed into the Δ*trp1 S. cerevisiae* strain NP2273 with the appropriate oligonucleotide (see Table 2-3). (A) Oligonucleotide used in this study to fix nonsense mutations in residues Arg44, Arg78, and Ser201 in yPRAI. Sequences of target codons are highlighted. (B–C) Simultaneous mutagenesis of multiple loops exhibits multiplicative efficiency. (B) Schematic representation of one (a–c), two (d–e), or three (f) oligonucleotide mutagenesis. (C) The data shown are mean ± standard error for percent recombination of at least three separate experiments.

Figure 2-5. Single loop oligonucleotide mutagenesis at position Ser201: sequencing analysis of recombinant colonies. The DNA and protein sequence of the trp1-Ser201* vector are indicated above. The mutated Ser201 codon is framed. Arrows indicate location of silent mutation.

Next, we measured the efficiency of simultaneous mutagenesis at all three positions, Arg44, Arg78 and Ser201; it was $0.010 \pm 0.001\%$ (Figure 2-4). Thus, simultaneous oligonucleotide recombination at two loops was 25-fold less efficient than that at a single loop, and at three loops was 20-fold less efficient than that at two loops. Notably, the efficiencies at multiple loops were nearly multiplicative.

To increase the efficiency of simultaneous recombination at multiple loops, we tried two classic methods in yeast genetics.[3] First, an overlapping oligomer was used to link two mutagenic oligonucleotides in an attempt to improve the efficiency of multi-loop mutagenesis. In our system, such an approach did not lead to a significant increase in the number of recombinants (data not shown). Next, we aimed at improving the efficiency of vector re-circularization by co-transformation with an oligonucleotide that overlapped the DSB in the vector (Table 2-3, Fig. 2-6). Again, inclusion of this additional oligonucleotide did not improve the recombination efficiency. Thus, in its current form, our protocol for oligonucleotide recombination leads to drops in efficiency as additional loops are mutagenized.
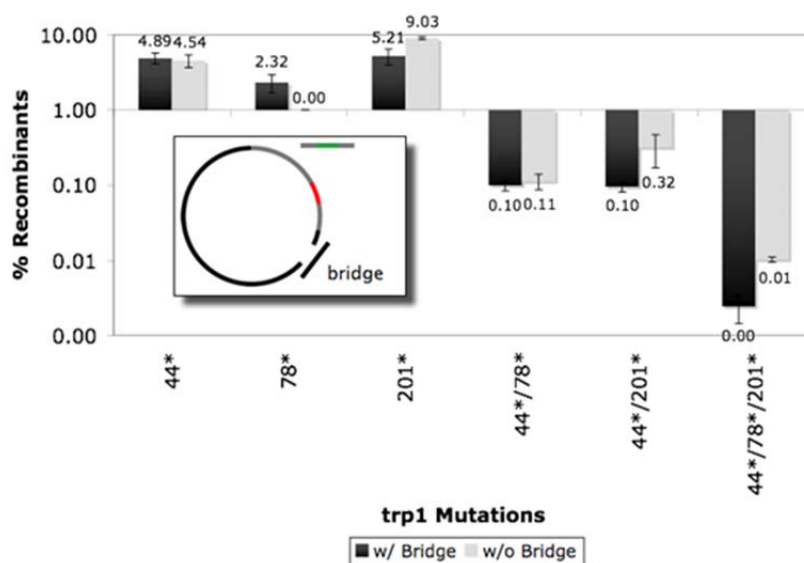
Figure 2-6. Co-transformation of linker Oligonucleotide. Linear vectors carrying trp1-R44*, trp1-R78*, trp1-S201*, trp1-R44*R78*, trp1-R44*S201*, and trp1-R44*R78*S201* are co-transformed into the Δtrp1 *S. cerevisiae* strain NP2273 with the appropriate oligonucleotides and with or without a 60 bp oligomer bridge (see Table 2-3). Co-transformation with bridge showed no significant improvement in recombination efficiency. The data shown are the mean ± the standard error of at least three separate experiments

2.3.5 Mock selection via oligonucleotide recombination

While the high efficiency of single-site recombination with our system suggests that it should allow for the generation of large libraries *in vivo*, and since HR is not routinely used for directed evolution, we explicitly challenged our system in a mock selection experiment. Given that 10^6 transformants can be readily obtained in *S. cerevisiae*, the mock selection experiment was designed to test the feasibility of enriching a functional TRP1 gene obtained by oligonucleotide recombination from a pool of oligonucleotides encoding a mock library of 1 active to 10^6 inactive trp1 variants.

Specifically, linearized vector pNP2282, encoding the trp1-Ser201* gene with an ochre codon at position 201, was co-transformed with a $1:10^6$ mixture of oligonucleotide Ser201LibraryFix: Ser201LibraryOpal (Figure 2-7). Both the Ser201LibraryFix and Ser201LibraryOpal oligonucleotides were designed such that they not only introduced a unique restriction site, but also encoded unique silent mutations on either side of the 201 codon to mark the recombination event. Use of the opal codon in the mock library allows it to be readily distinguished from the vector ochre codon. The mock selection was carried out under the same optimized conditions used to measure the single-loop recombination efficiency. Thus, 10^8 cells were co-transformed with 3.5×10^{10} linear vector molecules and 3.6×10^{13} total oligonucleotides. Transformants were then selected on SC (Ura-) plates to determine the library size, and recombinants were scored on SC (Ura-Trp-) plates to select for functional TRP1.

Encouragingly, with 2.6×10^6 successful transformants on SC (Ura-) plates, 8328 colonies survived on SC (Ura$^-$Trp$^-$) plates, demonstrating that a library of $1:10^6$ was fully covered with this protocol. The mock selection from a library of $1:10^6$ was repeated three times. While the number of TRP$^+$ colonies varied in each experiment, the library was successfully covered each time.

We speculate that the library size that can be covered is greater than that predicted by the recombination efficiency because there is a vast excess of oligonucleotide molecules to number of linearized vector molecules, and the vector molecules again are in excess of the number of transformed cells ($\sim 3 \times 10^5$ oligonucleotides and ~ 300 plasmids per cell). However, this rationale has not yet been explicitly tested. The fact that the library size that can be covered cannot be simply calculated from the recombination efficiency

demonstrates the importance of directly testing HR in the context of library construction. Furthermore, since the library of 10^6 was covered in excess, it may be possible to cover even larger libraries not only for single-loop but also multiple-loop recombination.

Eighteen individual colonies were first subjected to further selection in liquid media to cure the strain of what is presumed to be non-recombinant or recombinant non-viable pNP2282 vector and then analyzed by sequencing. Of the 18 colonies, 13 were successfully cured of non-recombinant pNP2282 and hence could be assigned to a *TRP1* gene and a 201 codon. The remaining 5 colonies that were not successfully cured were not analyzed further. As shown in Figure 2-7, all 13 recombinant colonies encoded functional *TRP1* using the TCC Ser codon encoded by the oligonucleotide Ser201LibraryFix and its silent mutations.

Because of inherent variability in the mock selection experiment, we repeated the experiment multiple times. It should be noted that in some of these mock selections, we additionally observed a handful of viable recombinant carrying non encoded codons such as TGT (Cys) and TCA (Ser) at position 201. While the mechanism by which non-encoded codons arise is not known at this point, silent mutation analysis suggests they originate from alteration of the SER201LibraryOpal oligonucleotide. Furthermore, chimeric variants were previously suggested to arise from multiple crossover events between library oligonucleotides during *in vivo* recombination.[25,51] Alternatively, it was suggested that mismatch repair mechanism may play a role in oligonucleotide recombination.[34] Importantly, in addition to potential biochemical significance of these residues that was not addressed in this study, from the vantage point of library creation these non-coded codons may be viewed as an additional source of mutation.

A

Target sequence	TAGATTAAATGGCGTTATTGGTGTTGATGTA **TAA** GGAGGTGTGGAGACAAATGGTGTAAAAGAC
Fixing oligo (SER201LibraryFix)	TAGATTAAATGGCGTTATTGGTGTTGATGT**G TCC** GGAGG**C**GTGGAGACAAATGGTGTAAAAGAC
Nonsense oligo (SER201LibraryOpal)	TAGATTAAATGGCGTTATTGGTGTTGATGTC **TGA** GGAGG**A**GTGGAGACAAATGGTGTAAAAGAC

B
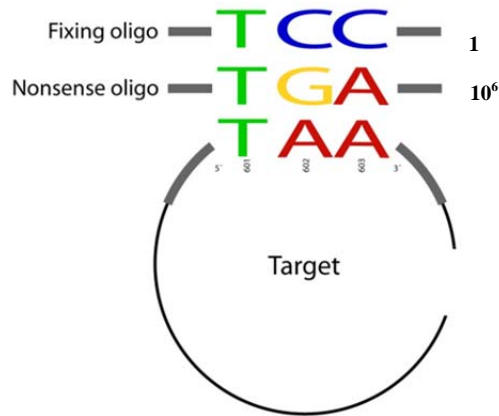
C
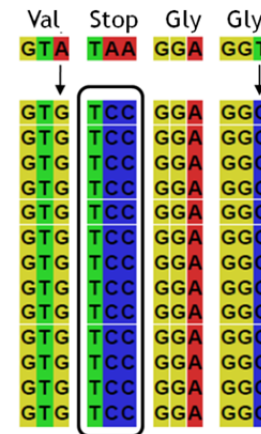
Figure 2-7. Mock library mutagenesis of *trp1-Ser201** vector. (A) oligonucleotide cassette use for mock library mutagenesis. Target codon is marked in red. silent mutations are marked blue. (B) Fixing codon was enriched from a mock library of 10^6 inactive variants in a single step. Vector NP2282, carrying a *trp1-Ser201** allele, is targeted with a $1:10^6$ mix of fixing to nonsense ssDNA oligonucleotides (oligonucleotides SER201LibraryFix and SER201Library*Opal,* respectively). (C) Eighteen viable colonies were analyzed by sequencing (see text), and 13 high quality recombinant sequences were aligned using the clustalW server.[5] The vector sequence is shown at the top. All sequenced colonies carried the fixing codon (TCC) at position 201 (framed) as well as the encoded silent mutations upstream (A>G) and downstream (T>C) (highlighted). Library results suggest that the library size is fully covered and therefore may allow for larger library experiments.

2.4   Discussion

Together these results establish that oligonucleotide recombination can be used to construct large DNA libraries entirely *in vivo*. For library generation, as opposed to

targeted gene alteration, the efficiency of recombination is critical because it directly limits the number of variants that can be tested. We were able to optimize the recombination efficiency at a single loop to ~5%. These optimized conditions allowed a wt codon to be enriched from a mock library of 10^6 inactive variants (a typical library size for directed evolution) in a single step. Furthermore, the method is particularly straightforward to implement. All that is required is co-transformation of commercial ss oligonucleotides and linearized vector using a standard electroporation protocol. Thus, oligonucleotide recombination is competitive with, and for *in vivo* selections easier to implement, than *in vitro* PCR library mutagenesis techniques.

The efficiency of oligonucleotide recombination at a single loop reported here is high and consistent with that reported in the recombination field with use of either a DSB or viral machinery.[31,55] The significant enhancement in recombination efficiency using a linearized vector is in agreement with previous studies demonstrating that DSB induction significantly enhances oligonucleotide recombination in yeast from ~0.03% [10] up to 20%.[55,56]

To our knowledge this study is the first attempt to mutagenize multiple loops simultaneously by co-transformation of two or more oligonucleotides. Kmiec and others demonstrated that multiple codons could be mutagenized simultaneously using a single oligonucleotide.[57,58] In MAGE, automation allowed for accumulation of multiple mutations by repetitive transformation.[59] Significantly, for directed evolution it is advantageous to be able to mutate multiple positions at the same time.

Single-loop recombination was highly efficient, averaging 5%, or 4.0×10^5 recombinants. Multiple loops could be simultaneously mutagenized, although the

efficiencies dropped to 0.2%, or $6.0{\times}10^3$ recombinants, for two loops and 0.01% efficiency, or $1.5{\times}10^2$ recombinants, for three loops.

We find that simultaneous transformation of two or three targeting oligonucleotides gives multiplicative recombination efficiencies. Compared to 5% average efficiency for single oligonucleotide recombination, the efficiencies for mutation using two and three oligonucleotides simultaneously were 0.2% and 0.01%, respectively. The multiplicative efficiency is consistent with current fundamental understanding of crossover events, which are independent events whose frequency is proportional to the distance between potentially homologous regions.[60] Alternatively, the multiplicative efficiency could be explained simply as the probability of the two events occurring at the same time.

We tried to overcome the multiplicative effect of simultaneous oligonucleotide recombination, first by linking two oligonucleotides together with a third oligomer to minimize the number of necessary crossovers, and next by using an oligomer to close the linearized vector. Neither of these strategies increased the recombination efficiency significantly. However, it is possible that moving the vector cut site into the target gene will allow for increased recombination at multiple locations. Finally, if the multiplicative effect arises from probabilities, it is possible that for targets farther apart, such as different chromosomes, this effect could be eliminated.

Therefore, oligonucleotide recombination should allow for the construction of large libraries at a single loop or moderate size libraries at two loops, but it is not yet sufficiently efficient to simultaneously mutate more than two loops. Assuming 10^7 transformants, an efficiency of 5% at a single loop predicts complete coverage of a library of $2{\times}10^5$. Interestingly, our actual coverage in the mock selection experiment was even larger than

predicted by this estimation. We obtained ~8000 colonies from 2.6×10^6 unique transformants. This greater coverage may be attributed to the vast excess of oligonucleotides and plasmid molecules compared with the number of transformed cells, as was previously suggested by Truan and colleagues.[61] Thus, we may be able to generate libraries even higher than our recombination efficiencies would predict. Notably, this result demonstrates the significance of library-oriented experimental setups for testing novel mutagenesis techniques. The efficiency of nucleotide recombination at two loops of 0.2% predicts library coverage of 10^4. Therefore, multiple oligonucleotide recombinations could be attractive for replacing iterative mutagenesis approaches where smaller, structure-based libraries are designed. For example, it has been shown that for directed evolution of enantioselective enzymes, simultaneous randomization is far more efficient than consecutive rounds of error-prone-PCR.[62]

The recombination system presented here is engineered, and its mechanism is undefined at this point. The mutagenesis may occur via a combination of DSB repair and single-strand annealing (SSA) or during DNA replication.[30,63,64] Currently, we are investigating the mechanism using knockouts of yeast recombination machinery with the goal of improving the recombination efficiency, particularly at multiple loops. For example, deletion of RAD51 could be used to establish that the mechanism involved is DSB repair,[63] while RAD59 deletion may indicate that SSA is taking place.[65,66] Furthermore, as recent studies in both bacteria and yeast suggest a strand bias exists that affects the efficiency of oligonucleotide mutagenesis,[34,50,67-72] it will be interesting to see if such bias can be exploited to increase recombination efficiency at multiple loops.

While used here for the mutagenesis of loops in the active site of an enzyme, the oligonucleotide recombination system reported here should be broadly useful for library mutagenesis not only of individual proteins, but also of other biomolecules such as RNA, multi-component systems such as metabolic pathways, and regions other than loops such as gene promoters. For example, this mutagenesis strategy could be used to randomize promoter strengths of multiple genes in a biosynthetic pathway to maximize production of a natural product in a heterologous host. Alternatively, the technology could be used to randomize the strength of interactions among multiple ribozymes in an engineered, artificial circuit. As synthetic biologists seek to engineer complex systems with increasing numbers of components, the need for directed evolution tools that allow large numbers of variations to be tested will only increase.

Recombination is now widely employed as a tool for targeted gene alteration, yet surprisingly it is not yet routinely used to generate large libraries of DNA mutants. Oligonucleotide recombination has not been characterized and optimized specifically for library generation. Problems of how to mutagenize multiple positions simultaneously as well as the mechanism underlying chimeric codon formation have not been addressed.

In vivo recombination is faster, cheaper, and more straightforward to implement than current *in vitro* PCR-based mutagenesis techniques, although the technology is less well developed for library mutagenesis at this time. To meet synthetic biology's goal of cell engineering, the cell's own synthetic machineries will need to be co-opted to construct new building blocks and pathways directly in living cells.

2.5   Experimental methods

General materials and methods. Standard methods for molecular biology in *Saccharomyces cerevisiae* and *Escherichia coli* were used.[73,74] *S. cerevisiae* strains were grown at 30 °C in media containing 2% glucose unless otherwise noted. Restriction enzymes and Vent DNA polymerase were purchased from New England Biolabs. Vent polymerase was used for all PCR reactions except yeast or *E. coli* colony PCR unless otherwise noted. For yeast colony PCR, cells were prepared according to a reported protocol (http://labs.fhcrc.org/hahn/Methods/mol_bio_meth/pcr_yeast_colony.html), and amplifications were performed with GoTaq DNA polymerase (Promega). The dNTPs used for PCR were purchased from GE Healthcare Life Sciences DNA cassette oligonucleotides for each mutation along with complementary strands were purchased from Invitrogen or Integrated DNA Technologies. DNA sequencing was performed by Genewiz or in the lab of Dr. Jingyue Ju at the Columbia Genome Center. Pellet Paint was purchased from Novagen. Plasmid DNA was purified using QIAprep miniprep kits (Qiagen); for yeast minipreps, cells were vortexed with acid-washed glass beads (Sigma) for five minutes before cell lysis. PCR products were purified with agarose gel electrophoresis and QIAquick spin columns purchased from Qiagen. Yeast genomic DNA was purified using a YeaStar Genomic DNA Kit (Zymo Research). DNA concentrations were determined by absorption at 260 nm, and absorbance measurements were taken on a Molecular Devices SpectraMax Plus 384 instrument. DNA Gels were run at 120-140 V for 45 min using a Bio-Rad PowerPac 100. Cells were incubated in a New Brunswick Scientific Series 25 Incubator Shaker. All aqueous solutions were made with distilled water prepared from a Milli-Q Water System. For PCR, a MJ Research PTC-200 Pellier Thermal Cycler was

employed. Transformation of *E. coli* was carried out by electroporation using a Bio-Rad *E. coli* Pulser. Yeast electroporation was carried out using a Bio-Rad Gene Pulser Xcell and a previously reported protocol.[75]

Plasmid Construction. Standard protocols for molecular biology and yeast genetics were used.[76,77] The materials and primers used in this study are listed in Tables 2-1 to 2-3. The *TRP1* gene was subcloned into pHT2150 under control of the *MET25* promoter using primers VWC2031 and VWC2032 to generate plasmid pNP2278. Plasmids for expressing *trp1* mutants were created using overlap extension PCR. Two modified strands incorporating the *trp1-Arg44** mutation were made from p424MET25 using the primers VWC2031 and VWC2036, and VWC2032 and VWC2035. Following fusion, this fragment was amplified using VWC2034 and VWC2033, digested with SfiI and inserted into the multiple cloning site of the pHT2150 vector to generate pNP2279. These steps were repeated on pNP2279 using primers VWC2031 and VWC2038, and VWC2032 and VWC2037 to generate pNP2280, *trp1-Arg44*-Ser201** mutations. These steps were then repeated on pNP2280 but instead using primers VWC2031 and VWC2040, and VWC2032 and VWC2039 to generate pNP2281, *trp1-Arg44*Arg78*Ser201** mutations. The remaining three plasmids pNP2282, *trp1-Ser201** mutation, pNP2283, *trp1-Arg44*Arg78** mutations, and pNP2284, *trp1-Arg78** mutation, were created using restriction enzymes and ligation.

pNP2278, pNP2280 and pNP2281 were digested with BstXI and SfiI to generate fragments 303 bp and 403 bp in length. The 303 bp fragment from pNP2278, carrying no mutations, and the 403 bp fragment from pNP2280, carrying the *trp1-Ser201** mutation, were ligated to make pNP2282. The 303 bp fragment from pNP2281, carrying the *trp1-*

Arg44*Arg78* mutations, and the 403 bp fragment from pNP2278, carrying no mutations were ligated to make pNP2283. Finally, pNP2278 and pNP2283 were digested with MfeI and SfiI to generate fragments 145 bp and 561 bp in length. The 145 bp fragment from pNP2278, carrying no mutations, and the 561 bp fragment from pNP2281, carrying the *trp1-Arg78** mutation, were ligated to make pNP2284. Linearized plasmids were prepared by digesting with SalI or ClaI, which cut once at the 3' of *TRP1* gene, within the multiple cloning site.

Yeast Transformation. We used a high efficiency yeast electroporation protocol with slight modifications.[45] A 10 mL culture of NP2273 was inoculated in fresh YPD and incubated at 30ºC overnight (250 mL). The overnight culture was used to inoculate 100 mL of 30°C YPD to reach an $OD_{600} = 0.1$. This culture was grown to an $OD_{600} = 0.6 - 0.8$ (~ 5 hr). Then 1 mL filter-sterilized 1,4-dithiothreitol (DTT) solution (1M Tris, pH 8.0, 2.5 M DTT) was then added to the 100 mL of YPD. The cells were grown at 30ºC for another 20 min. From this point forward, the cells were kept on ice. The cells were harvested in 50 mL Falcon tubes at 2000 rpm at 4ºC for 5 min. The supernatant was discarded and the cells were washed with 25 mL of E-buffer (10 mM Tris, pH 7.5, 270 mM sucrose, 1 mM $MgCl_2$) by pipetting up and down. The cells were reharvested and resuspended in 1 mL of E-buffer and transferred to an autoclaved 1.5 mL microcentrifuge tube and spun down again in a tabletop centrifuge at 12000 rpm for 10-20 s. The supernatant was discarded and resuspended using 50 µL of E-buffer.

The cells were distributed into 4 tubes and added with vector and either ss or ds DNA oligonucleotides, and incubated on ice for 10 min (approximately 100 µL cells per tube). The gene pulser was set to 540 V, 25 µF, 0 Ω, and the cells were transferred to 2

mm electroporation cuvettes. Immediately after electroporation the cells were recovered with 1 mL of 30ºC YPD. The cells were then incubated at 30ºC for 1 hr before plating on SC(Ura⁻) and SC(Ura⁻Trp⁻), which gave transformants and recombinants, respectively. Prior to analysis, recombinant colonies were further grown in SC (Ura⁻Trp⁻) liquid selective media for 4 days with multiple seedings using fresh selective media. The cells were then plated on SC (Ura⁻Trp⁻) selective plates and the resulted colonies were sequenced.

Mock library experiment. To generate a mock library, fixing and nonsense oligonucleotides (Ser201LibraryFix and Ser201LibraryOpal, respectively) were mixed to a final molar ratio of 1 to 10^6 Ser201LibraryFix to Ser201LibraryOpal and co-transformed with a linear trp1*-Ser201 vector by electroporation. The same amount of DNA oligonucleotides was used as in previous experiments, as well as the same optimized molar ratio of vector to oligonucleotides (1:1000). Each experiment yielded a transformation efficiency of an order of magnitude 10^5/µg of oligonucleotide DNA. At least 10^6 transformants were achieved in each experiment, using up to 10 electroporation samples.

Once electroporated, samples were rescued In YPD for one hour at 30˚C and then a 1000-fold dilution was plated on SC (Ura-) plates to score for transformation efficiency and in 3-fold dilution was plated on SC(Ura-Trp-) plates to score for recombination efficiency. 2.6 x10^3 total colonies were counted on SC (Ura-) plates, whereas 2776 colonies were counted on dual selection SC (Ura-Trp-) plates. Although fully fixed viable colonies were always recovered, variability in recombination efficiency was observed between experiments, with the number of colonies varying within an order of magnitude. The experiment was further performed independently by another student in the lab, Laura

Wingler, with similar results. No colonies were observed when linear vector was transformed without oligonucleotides.

Seeding and sequencing of recombinant colonies. In order to isolate the recombinant plasmids carrying functional TRP1 that allows for cell survival, it is necessary to cure the cells of plasmids that did not undergo recombination. For this purpose, we used a continuous selection process in liquid selective media. 20-30 colonies were analyzed for each experiment. Each recombinant colony was further selected in SC (Ura-Trp-) liquid media for 4 days at 30ºC to cure the cell of additional trp1*-201 vectors. To avoid depletion of nutrients and maintain the stringency of selection, cells were diluted into fresh selective media each day, a process we refer to as 'seeding'. After 4 'seedings' (day 5) each sample (i.e a single recombinant colony seeded in liquid selective media) was plated on selective SC (Ura-Trp-) plates.

Colony PCR was performed on a single colony from each plate using the following protocol: the colony was resuspended in 50 μL dH$_2$O, boiled in 100 ˚C bath for 1 minute and then vortexed for 1 minute. The supernatant was subjected to PCR amplification of TRP1 gene. Colony PCR protocol: 77 μL dH$_2$O, 2 μL colony supernatant, 20 μL 5x GoTaq Buffer, 0.2 μL dNTPs, 0.2 μL primer VWC1051, 0.2 μL primer VWC1052 and 0.2 μL GoTaq Polymerase to a total of 100 μL reaction. The PCR reaction product was verified by gel electrophoresis for the correct size (1000 bp), purified using a Qiagen PCR purification kit and sequenced using the reverse primer VWC1051. Only high quality sequences were used for analysis. We compared sequencing results of the single seeded colonies to retransformed single colonies and found the seeding technique to be highly efficient for isolation of recombinant plasmids. (Data not shown)

Double stranded oligonucleotide preparation. dsDNA was prepared by annealing complementary ss oligonucleotides using the following annealing protocol. Complementary single strand oligonucleotide, dissolved in annealing buffer (500 mM NaCl, 10 mM EDTA, 10 mM Tris pH 7.5-8.0), were mixed (0.5 µg of each) and subjected to gradual cooling using a PCR machine: the sample was first heated to 95˚ C for 5 minutes, and then cooled at 0.1 ˚C/sec to 46 ˚C, held constant for 30 minutes and cooled again to 8˚C at the same rate.

Trp1* allele reversion rate. Reversion rate of target plasmids carrying trp1* allele with one (trp1*-201), two (trp1*-44-201) or three (trp1*44-78-201) ochre mutations was tested by plating $3x10^6$ plasmid-carrying cells on double selective media SC (Ura-Trp-). No colonies were observed after 5 days.

2.6   Strains, plasmids, and oligonucleotides

Table 2-1. Strains used in this study

Strain	Origin	Genes
NP2273	ATCC4017202	MATa his3 Δ1 leu2Δ0 met15Δ0 ura3Δ0 ΔTRP1

Table 2-2. Plasmids used in this study

Vector	Mutation	Genes	Strain
pHT2150	H. Tao, P. Peralta-Yahya, J. Decatur, V.W. Cornish. Characterization of a new glycosynthase cloned using chemical complementation. ChemBiochem 2008, 9(5).		
p424MET25			ATCC 87321
pNP2278		URA3, TRP1	NP2278
pNP2279	trp1-R44*	URA3, trp1	NP2279
pNP2284	trp1-R78*	URA3, trp1	NP2284
pNP2282	trp1-S201*	URA3, trp1	NP2282
pNP2283	trp1-R44*R78*	URA3, trp1	NP2283
pNP2280	trp1-R44*S201*	URA3, trp1	NP2280
pNP2281	trp1-R44*R78*S201*	URA3, trp1	NP2281

Table 2-3. Oligonucleotides used in this study. Point mutations are written in lowercase.

Name		Sequence (5' to 3')
ARG44Fix	VWC2041	CTGGGTATTATATGTGTGCCCAATAGAAAGagaACAATTGACCCGGTTATTGCAAGGAAAATT
ARG44Fix_ds	VWC2042	AATTTTCCTTCGAATAACCGGGTCAATTGTtctCTTTCTATTGGGCACACATATAATACCCAG
ARG78Fix	VWC2043	GGCACTCCAAAATACTTGGTTGGCGTGTTTcgtAATCAACCTAAGGAGGATGTTTTGGCTCTG
SER201Fix	VWC2044	AGATTAAATGGCGTTATTGGTGTTGATGTAagcGGAGGTGTGGAGACAAATGGTGTAAAAGAC
SER201LibraryFix	VWC2218	TAGATTAAATGGCGTTATTGGTGTTGATGTGtccGGAGGCGTGGAGACAAATGGTGTAAAAGAC
SER201LibraryOpal	VWC2219	TAGATTAAATGGCGTTATTGGTGTTGATGTGtgaGGAGGCGTGGAGACAAATGGTGTAAAAGAC

Table 2-4. Oligonucleotides used to construct trp1 variants.

Name	Sequence (5' to 3')
VWC2031	GCATACGTCGGCCCCCGGGGCCATGTCTGTTATTAATTTCACA
VWC2032	GCATACGTCGGCCGGCAGGGCCTTAATGGTGAT
VWC2033	GGCCGGCAGGGCCTTA
VWC2034	GGCCCCCGGGGCCAT
VWC2035	ATATCTGTGCCCAATTAAAAGAGAACAATTGACCCGG
VWC2036	GTCAATTGTTCTCTTTTAATTGGGCACACATATAATAC
VWC2037	ATTGGTGTTGATGTATAAGGAGGTGTGGAGACAAAT
VWC2038	TGTCTCCACACCTCCTTATACATCAACACCAATAACGC
VWC2039	TTGGTTGGCGTGTTTTAAAATCAACCTAAGGAGGATGT
VWC2040	CTCCTTAGGTTGATTTTAAAACACGCCAACCAAGTATT

2.7 References

1   Schiestl, R. H. & Gietz, R. D. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr. Genet.* **16**, 339-346, (1989).

2   Simon, J. R. & Moore, P. D. Homologous recombination between single-stranded DNA and chromosomal genes in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **7**, 2329-2334, (1987).

3   Klinner, U. & Schafer, B. Genetic aspects of targeted insertion mutagenesis in yeasts. *FEMS Microbiol. Rev.* **28**, 201-223, (2004).

4   Henn-Sax, M. *et al.* Two (betaalpha)(8)-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry* **41**, 12032-12042, (2002).

5   Chenna, R. *et al.* Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**, 3497-3500, (2003).

6   Rothstein, R. J. One-step gene disruption in yeast. *Methods Enzymol.* **101**, 202-211, (1983).

7   Kunes, S., Ma, H., Overbye, K., Fox, M. S. & Botstein, D. Fine structure recombinational analysis of cloned genes using yeast transformation. *Genetics* **115**, 73-81, (1987).

8   Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. *Nucleic Acids Res* **21**, 3329-3330, (1993).

9   Raymond, C. K., Pownder, T. A. & Sexson, S. L. General method for plasmid construction using homologous recombination. *Biotechniques* **26**, 134-138, 140-131, (1999).

10   Igoucheva, O., Alexeev, V. & Yoon, K. Oligonucleotide-directed mutagenesis and targeted gene correction: a mechanistic point of view. *Curr Mol Med* **4**, 445-463, (2004).

11   Rothstein, R. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol.* **194**, 281-301, (1991).

12   Wendland, J. PCR-based methods facilitate targeted gene manipulations and cloning procedures. *Curr. Genet.* **44**, 115-123, (2003).

13   Hegemann, J. H., Gldener, U. & Kohler, G. J. Gene disruption in the budding yeast Saccharomyces cerevisiae. *Methods Mol. Biol.* **313**, 129-144, (2006).

14   Hinnen, A., Hicks, J. B. & Fink, G. R. Transformation of yeast. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 1929-1933, (1978).

15   Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C. M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155-159, (2010).

16   Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 6354-6358, (1981).

17   Moerschell, R. P., Tsunasawa, S. & Sherman, F. Transformation of yeast with synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 524-528, (1988).

18   Scherer, S. & Davis, R. W. Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 4951-4955, (1979).

19   Langston, L. D. & Symington, L. S. Gene targeting in yeast is initiated by two independent strand invasions. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15392-15397, (2004).

20   Storici, F., Snipe, J. R., Chan, G. K., Gordenin, D. A. & Resnick, M. A. Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. *Mol. Cell. Biol.* **26**, 7645-7657, (2006).

21   Gray, M. & Honigberg, S. M. Effect of chromosomal locus, GC content and length of homology on PCR-mediated targeted gene replacement in Saccharomyces. *Nucleic Acids Res* **29**, 5156-5162, (2001).

22   Hua, S. B., Qiu, M., Chan, E., Zhu, L. & Luo, Y. Minimum length of sequence homology required for in vivo cloning by homologous recombination in yeast. *Plasmid* **38**, 91-96, (1997).

23   Manivasakam, P., Weber, S. C., McElver, J. & Schiestl, R. H. Micro-homology mediated PCR targeting in Saccharomyces cerevisiae. *Nucleic Acids Res* **23**, 2799-2800, (1995).

24   Negritto, M. T., Wu, X., Kuo, T., Chu, S. & Bailis, A. M. Influence of DNA sequence identity on efficiency of targeted gene replacement. *Mol. Cell. Biol.* **17**, 278-286, (1997).

25   Mezard, C., Pompon, D. & Nicolas, A. Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. *Cell* **70**, 659-670, (1992).

26   Liu, L., Rice, M. C. & Kmiec, E. B. In vivo gene repair of point and frameshift mutations directed by chimeric RNA/DNA oligonucleotides and modified single-stranded oligonucleotides. *Nucleic Acids Res* **29**, 4238-4250, (2001).

27   Wilson, J. H., Leung, W. Y., Bosco, G., Dieu, D. & Haber, J. E. The frequency of gene targeting in yeast depends on the number of target copies. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 177-181, (1994).

28   Yamamoto, T., Moerschell, R. P., Wakem, L. P., Komar-Panicucci, S. & Sherman, F. Strand-specificity in the transformation of yeast with synthetic oligonucleotides. *Genetics* **131**, 811-819, (1992).

29   Yamamoto, T., Moerschell, R. P., Wakem, L. P., Ferguson, D. & Sherman, F. Parameters affecting the frequencies of transformation and co-transformation with synthetic oligonucleotides in yeast. *Yeast* **8**, 935-948, (1992).

30   Court, D. L., Sawitzke, J. A. & Thomason, L. C. Genetic engineering using homologous recombination. *Annu. Rev. Genet.* **36**, 361-388, (2002).

31   Costantino, N. & Court, D. L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15748-15753, (2003).

32   Ellis, H. M., Yu, D., DiTizio, T. & Court, D. L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6742-6746, (2001).

33   Vasquez, K. M., Marburger, K., Intody, Z. & Wilson, J. H. Manipulating the mammalian genome by homologous recombination. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8403-8410, (2001).

34   Parekh-Olmedo, H. & Kmiec, E. B. Progress and prospects: targeted gene alteration (TGA). *Gene Ther.* **14**, 1675-1680, (2007).

35   Cole-Strauss, A. *et al.* Correction of the mutation responsible for sickle cell anemia by an RNA-DNA oligonucleotide. *Science* **273**, 1386-1389, (1996).

36   Donoho, G., Jasin, M. & Berg, P. Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. *Mol. Cell. Biol.* **18**, 4070-4078, (1998).

37   Vasquez, K. M., Marburger, K., Intody, Z. & Wilson, J. H. Manipulating the mammalian genome by homologous recombination. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8403-8410, (2001).

38   Wells, J. A., Vasser, M. & Powers, D. B. Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene* **34**, 315-323, (1985).

39   Linske-O'Connell, L. I., Sherman, F. & McLendon, G. Stabilizing amino acid replacements at position 52 in yeast iso-1-cytochrome c: in vivo and in vitro effects. *Biochemistry* **34**, 7094-7102, (1995).

40   Storici, F., Lewis, L. K. & Resnick, M. A. In vivo site-directed mutagenesis using oligonucleotides. *Nat. Biotechnol.* **19**, 773-776, (2001).

41   Swers, J. S., Kellogg, B. A. & Wittrup, K. D. Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display. *Nucleic Acids Res.* **32**, e36, (2004).

42   Wang, L., Jackson, W. C., Steinbach, P. A. & Tsien, R. Y. Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16745-16749, (2004).

43   Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9727-9732, (2003).

44   Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C. M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155-159, (2010).

45   Wittrup, K. D. Wittrup: Yeast_Transformation. (2007). <http://openwetware.org/wiki/Wittrup:_Yeast_Transformation>.

46   Branden, C., Tooze, J. *Introduction to Protein Structure*. 2nd edn, (Garland publishing Inc., 1999).

47   Hommel, U., Eberhard, M. & Kirschner, K. Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction. *Biochemistry* **34**, 5429-5439, (1995).

48   Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201, (2006).

49   Winzeler, E. A. *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**, 901-906, (1999).

50   Liu, L. *et al.* Strand bias in targeted gene repair is influenced by transcriptional activity. *Mol. Cell. Biol.* **22**, 3852-3863, (2002).

51   Swers, J. S., Kellogg, B. A. & Wittrup, K. D. Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display. *Nucleic Acids Res* **32**, e36, (2004).

52   Storici, F., Durham, C. L., Gordenin, D. A. & Resnick, M. A. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14994-14999, (2003).

53   May, M. Engineering a new business. *Nat. Biotechnol.* **27**, 1112-1120, (2009).

54   Gibson, D. G. *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20404-20409, (2008).

55   Storici, F. & Resnick, M. A. The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.* **409**, 329-345, (2006).

56   Smith, G. R. Homologous recombination near and far from DNA breaks: alternative roles and contrasting views. *Annu. Rev. Genet.* **35**, 243-274, (2001).

57   Alcalde, M., Zumarraga, M., Polaina, J., Ballesteros, A. & Plou, F. J. Combinatorial saturation mutagenesis by in vivo overlap extension for the engineering of fungal laccases. *Comb Chem High Throughput Screen* **9**, 719-727, (2006).

58   Agarwal, S., Gamper, H. B. & Kmiec, E. B. Nucleotide replacement at two sites can be directed by modified single-stranded oligonucleotides in vitro and in vivo. *Biomol Eng* **20**, 7-20, (2003).

59   Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898, (2009).

60   Hartwell, L. *Genetics : from genes to genomes*. 2nd edn, (McGraw-Hill Higher Education, 2004).

61   Abecassis, V., Pompon, D. & Truan, G. High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *Nucleic Acids Res* **28**, E88, (2000).

62   Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the numbers problem in directed evolution. *Chembiochem* **9**, 1797-1804, (2008).

63   Bartsch, S., Kang, L. E. & Symington, L. S. RAD51 is required for the repair of plasmid double-stranded DNA gaps from either plasmid or chromosomal templates. *Mol. Cell. Biol.* **20**, 1194-1205, (2000).

64   Agmon, N., Pur, S., Liefshitz, B. & Kupiec, M. Analysis of repair mechanism choice during homologous recombination. *Nucleic Acids Res* **37**, 5081-5092, (2009).

65   Davis, A. P. & Symington, L. S. The yeast recombinational repair protein Rad59 interacts with Rad52 and stimulates single-strand annealing. *Genetics* **159**, 515-525, (2001).

66   Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. *Annu. Rev. Genet.* **38**, 233-271, (2004).

67   Kow, Y. W., Bao, G., Reeves, J. W., Jinks-Robertson, S. & Crouse, G. F. Oligonucleotide transformation of yeast reveals mismatch repair complexes to be differentially active on DNA replication strands. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11352-11357, (2007).

68   Rothstein, R., Michel, B. & Gangloff, S. Replication fork pausing and recombination or "gimme a break". *Genes Dev.* **14**, 1-10, (2000).

69   Huen, M. S. *et al.* The involvement of replication in single stranded oligonucleotide-mediated gene repair. *Nucleic Acids Res* **34**, 6183-6194, (2006).

70   Huen, M. S., Lu, L. Y., Liu, D. P. & Huang, J. D. Active transcription promotes single-stranded oligonucleotide mediated gene repair. *Biochem. Biophys. Res. Commun.* **353**, 33-39, (2007).

71   Gottipati, P. & Helleday, T. Transcription-associated recombination in eukaryotes: link between transcription, replication and recombination. *Mutagenesis* **24**, 203-210, (2009).

72   Smith, D. J. & Whitehouse, I. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* **483**, 434-438, (2012).

73   Ausubel, F. *et al. Current Protocols in Molecular Biology*. (Wiley, 1995).

74   Adams, A., Gottschling, D., Kaiser, C. & Stearns, T.   (Cold Spring Harbor Laboratory Press Plainview, NY 1998).

75   Pirakitikulr, N., Ostrov, N., Peralta-Yahya, P. & Cornish, V. W. PCRless library mutagenesis via oligonucleotide recombination in yeast. *Protein Sci.* **19**, 2336-2346, (2010).

76   Ausubel, F. M. *Current protocols in molecular biology*, <http://www.columbia.edu/cgi-bin/cul/resolve?clio5406135> (2001).

77   Adams, A. *Methods in yeast genetics : a Cold Spring Harbor Laboratory course manual*. 1997 edn, (Cold Spring Harbor Laboratory Press, 1998).

Chapter 3

Combinatorial Pathway Engineering via Reiterative Recombination

3.1 Chapter outlook

Metabolic engineering is a rapidly growing research field with potential to profoundly impact the production of drugs, biofuels and commodity chemicals. Valuable complex natural products, unattainable or too complex for cost effective production by chemical synthesis, have now been successfully obtained from biosynthesis in microbial hosts. Yet current methods for gene deletion and insertion prove laborious and inefficient when applied to engineering of entire metabolic pathways. Furthermore, the tools available for directed evolution of single enzymes cannot be extrapolated for construction of combinatorial libraries required for optimization of metabolite yield. Our lab previously developed Reiterative Recombination, a powerful DNA assembly method which enables the infinite integration of DNA fragments directly into the yeast chromosome with high efficiency and in a user friendly fashion. In this work, we challenged this novel technology for combinatorial attenuation of gene transcription to regulate the flux through a lycopene biosynthetic pathway. Using three cycles of assembly, five promoters were shuffled to vary the expression of each gene, constructing a combinatorial pathway library. We discuss Reiterative Recombination as a generic technique for engineering of multi-component pathways.

3.2 Introduction

Recent breakthroughs in the engineering of microbial hosts for the production of valuable natural product drugs[2-4] suggest prospects for routine production of desired natural products in heterologous system.[5-7] However, expression of complex molecules in microbial hosts is often characterized by low product yield, product toxicity, genome instability, and compromised cell growth.[7,8] Once installed, optimization of product yield and expression of product analogues add further challenges.[10-15] Much like for protein engineering, directed evolution offers to overcome some of these challenges, provided single-enzyme methodologies be adapted to multi-component, *in vivo* metabolic networks.

For example, while plasmids are widely used for protein expression in microbial hosts, it is now widely accepted that cloning of heterologous pathways necessitates genomic integration in order to avoid the metabolic burden caused by multiple plasmids.[16] Yet integration of multiple genes is not only cumbersome and time consuming, but also restricted by the limited availability of selective markers in both *E. coli* and *S. cerevisiae*. In addition, the iterative process of mutagenesis and selection easily performed for evolution of single enzymes by isolation and retransformation of plasmids becomes substantially more complex in the context of multiple chromosomal components. Finally, the tools available for characterization of enzyme variants are not sufficient for characterization of complex cellular phenotype resulting from perturbation of metabolic networks.

Reiterative Recombination allows *in vivo* construction of optimized, multi-gene pathways.[9] It overcomes the critical shortcoming of existing *in vivo* DNA assembly methods - their very low efficiencies - by coupling DNA recombination to the repair of

double-strand DNA breaks (DSB's). Our lab previously demonstrated the use of Reiterative Recombination for assembly of gene pathways and confirmed the method is suitable for construction of up to 10^4 member libraries.

Here, we demonstrate Reiterative Recombination to be a highly efficient, technically straightforward technique for assembly and combinatorial shuffling of metabolic pathways, using the simple three gene lycopene pathway as a model system. We utilized libraries of widely used yeast promoters to assemble multiple libraries in the same host strain, combinatorially shuffling promoters for each of the pathway genes. We then screened the resulting colonies by eye for expression of the red pigment lycopene. Characterization of this initial library provides insight into the flux through the isoprenoid pathway and a basis for further development of this promising technology for metabolic engineering.

3.2.1 DNA assembly via Reiterative Recombination

Reiterative Recombination serves as an efficient DNA assembly system in *S. cerevisiae*. At each assembly cycle, the DNA fragment of interest is inserted downstream of the previous assembled fragment by recombination between a 'donor' module (vector) and an 'acceptor' module (genome), as depicted in Figure 3-1. The recombination event results in elongation of the construct of interest and simultaneously replaces a selectable marker, allowing selection of cells with the desired construct.

In order to guide the assembly process, the DNA fragment to be assembled is first added with short homology sequences on both ends. The fragment is then introduced into the acceptor strain by co-transformation with a generic linear donor plasmid. Overlapping sequences allow the plasmid to circularize with the fragment upon transformation (a

process termed 'gap repair'). Next, double strand break is induced at a designated locus in the acceptor strain chromosome to trigger recombination between the donor plasmid and acceptor chromosome at a unique designated locus. Building on an established technique for targeted gene disruption,[17,18] Reiterative Recombination utilizes homing endonucleases with large unique recognition sequences (~20bp) to introduce DSBs at designated sites[9] stimulating homologous recombination between a donor and acceptor modules. DSB induction is facilitated by a homing endonuclease that is encoded by the donor plasmid. Importantly, the enzyme is tightly controlled under a galactose promoter which is activated only when cells are transferred into galactose containing media. Upon cleavage, the donor plasmid provides homology to repair the acceptor chromosome DSB guided by short regions of overlap. The fragment carried by the donor plasmid is added to the acceptor's growing construct and simultaneously replaces the acceptor module's endonuclease cleavage site and selectable marker.

Selection of successful recombination events is enabled since after fragment integration only the acceptor module carries a promoter to actively transcribe the adjacent selectable marker. Pairs of orthogonal selective markers and endonucleases are used for repeated iterations of the elongation cycle. All assembly occurs at the HO gene locus in a *S. cerevisiae* strain with full deletions of all markers used in our system.[19]

Figure 3-1. DNA assembly by Reiterative Recombination.[9] General scheme of Reiterative Recombination, showing two rounds of elongation. Each round of elongation is achieved by recombination between an "acceptor module" (in the linear chromosome) and a "donor module" (in the circular plasmid). The two modules contain orthogonal homing endonuclease cleavage sites (triangles) adjacent to different selectable markers (purple and green). Both markers are downstream of a homology region (gray), but only the acceptor module contains a promoter (white 'prom') driving marker expression. Endonuclease cleavage of the acceptor module stimulates recombination, joining the fragments being assembled (orange) and replacing the acceptor module's endonuclease site and expressed selectable marker with those of the donor module. Repeating this procedure with a donor module of the opposite polarity returns the acceptor module to its original state, allowing the assembly to be elongated indefinitely.

3.2.2   Terpenoids pathway as model system

Terpenoids are a particularly prominent class of metabolic engineering targets. Though they are all derived from the universal precursor isopentenyl pyrophosphate (IPP), the >55,000 terpenoids isolated to date exhibit vast structural and functional diversity.[20] This important class of functionally modified terpene secondary metabolites is present in all living organisms, and display anticancer, antiparasitic, antimicrobial, antiallergenic and anti-inflammatory properties. Terpenoids are further used as industrial products such as flavors, fragrances, insecticides or food colorants.[20] Many terpenoids, such as the chemotherapeutics taxol (paclitaxel), vinblastine and the antimalarial artemisinin are frontline therapeutics but cannot be sustainably supplied to the commercial market from their natural sources.[3,21-23] Thus, recombinant biosynthesis of terpenoid in microbial host cells could replace the massive harvesting currently required to enable their large scale utilization.

In recent years, considerable improvement has been achieved in microbial terpenoid yield.[20] Particularly, efforts have been made to improve the availability of isoprenoid precursor molecules in microbial hosts, altering either the classical mevalonate (MVA) biosynthetic pathway in eukaryotes (Figure 3-2), or the recently discovered non-mevalonate (DXP) pathway in bacteria.[20,21,24-26]

Carotenoids such as lycopene and beta-carotene are neutraceutical terpenoids with remarkable antioxidant activity. The broad range of carotenoid colors conveniently allow for a visible screen, making them the focus of many metabolic engineering studies[27-30] and particularly combinatorial biosynthesis.[31] Specifically, lycopene can be monitored at 470 nm wavelength or extracted into organic solvent to provide a sensitive assay, reaching

submilligram per liter differences in yield, and has therefore been the major carotenoid used for production in microbial hosts.[32]

We initiated collaboration with the labs of Greg Stephanopoulos (MIT) and Mattheos Koffas at Rensselaer Polytechnic Institute (RPI) in aim to develop new methodologies for the construction of *Saccharomyces cerevisiae* phenotypes for the efficient synthesis of the isoprenoid precursors, IPP, GPP, FPP and GGPP. Taking advantage of its facile colorimetric screen and relatively short pathway, we chose lycopene as a proof of principle system to demonstrate combinatorial attenuation of gene expression using Reiterative Recombination.

3.3   Results

3.3.1   Efficiency of genomic cassette mutagenesis in *S. cerevisiae*

Though highly efficient for DNA exchange, the process of homologous recombination is also tightly regulated in *S. cerevisiae* and maintained at low level during most phases of cell cycle to ensure genome integrity.[33] Thus, while spontaneous recombination of foreign DNA is sufficient for single gene insertions and deletions, it has been a significant barrier for construction of large libraries directly in the yeast genome. Building on our previous studies of plasmid-based cassette mutagenesis, we first quantified the efficiency of cassette mutagenesis on the yeast chromosome by way of DNA transformation.
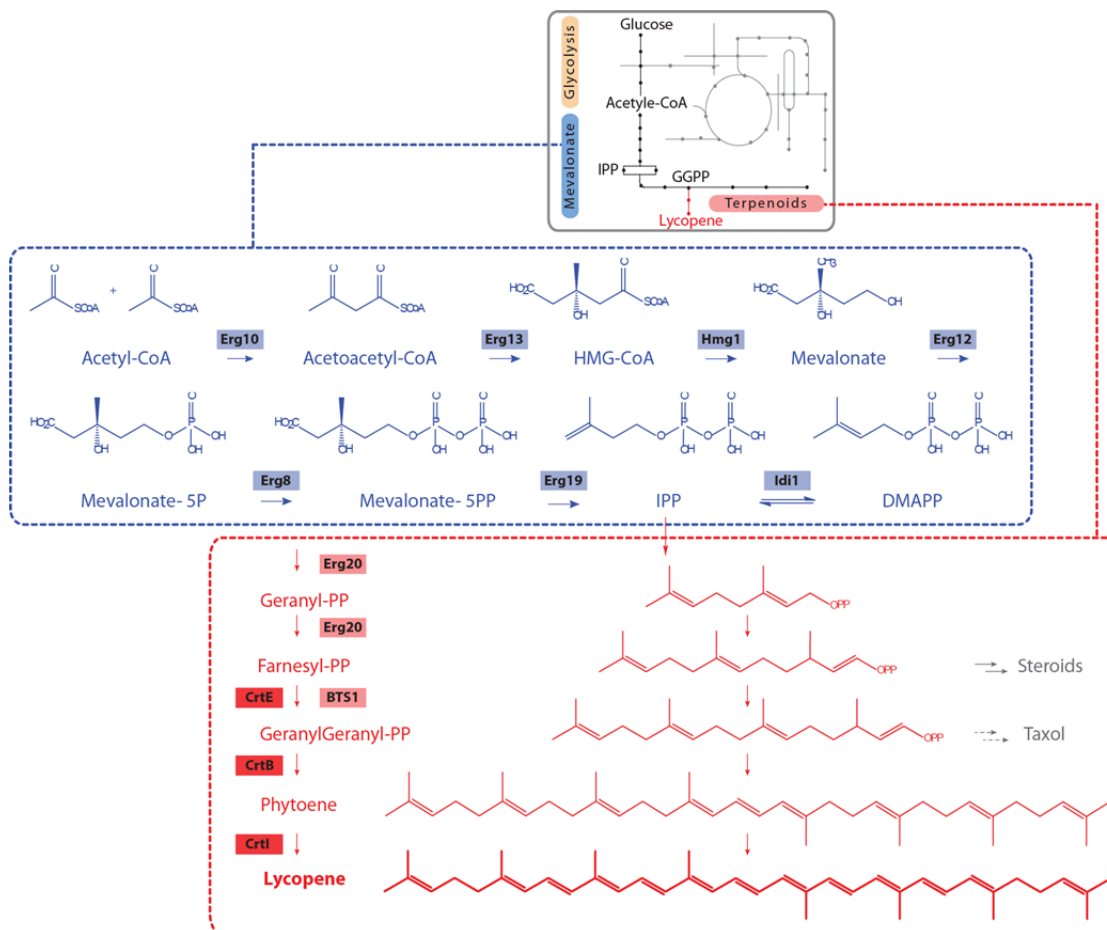
Figure 3-2. Lycopene biosynthetic pathway. The mevalonate pathway (blue) provides IPP and DMAPP which are the universal precursors for biosynthesis of terpenoids including, among others, phytoene and lycopene. Farnesyl-PP (FPP) is the natural precursor for steroid production in yeast, whereas GeranylGeranyl-PP (GGPP) can be used as precursor for heterologous expression of the anticancer drug Taxol. Heterologous genes that are used in this work to produce lycopene in yeast are marked in red (CrtE, CrtB and CrtI).

We therefore measured the efficiency of repairing point mutations and large deletions in the chromosomal target TRP1 by direct transformation of dsDNA. TRP1 gene is widely used as auxotrophic marker in *S. cerevisiae* and previously served as the model system for our plasmid-based cassette mutagenesis. Notably, we avoid the induction of DSB or over expression of proteins known to enhance recombination in yeast in order to establish baseline for the efficiency of chromosomal repair.

To examine the efficiency of repairing a point mutation we used strain VC2445 (a variant of the common laboratory auxotrophic strain W303, kindly provided by R. Rothstein) carrying an intact recombination mechanism and a single mutation at TRP1 nucleotide 247 which changes codon GAG (glutamic acid) to codon TAG (amber stop). To explore repair of deletion mutations we used the common laboratory strain FY251 carrying 600 bp deletion spanning (-100) to (+502) of the TRP1 gene. Control experiments confirmed both strains do not grow in the absence of tryptophan (data not shown).

Two fixing DNA cassettes were made by PCR using an intact TRP1 gene. For repair of point mutations, a 300 bp cassette was produced, and for repair of deletion mutants, 800 bp cassette spanning (-205) to (+620) of the TRP1 reading frame was produced. Each cassette was introduced into yeast by transformation. $\sim 10^7$ cells / ml survived transformation by electroporation. The efficiency of gene repair by cassette mutagenesis was calculated by comparing the number of cells growing on media lacking tryptophan (SC(T-)), where successful recombination occurred, and the number of overall surviving cells after transformation (SC, non-selective media).

Results are shown in Figure 3-3. Overall, the efficiency of dsDNA cassette recombination into the chromosome ranged 0.001 to 0.1 percent, significantly lower than

the previously observed efficiency of plasmid-based cassette mutagenesis.[34] While increasing the amount of DNA cassette slightly improved the repair efficiency of point mutation, the repair of deletion mutation within the range tested was not significantly changed. Increasing the homology sequence on both sides of the target mutation greatly enhances recombination efficiency, as demonstrated for targeting TRP1 point mutation using the longer 800 bp repair cassette.

Since transformation efficiency (~10^7 cells) is not limiting in this experiment and typically results in multiple copies of the fixing cassette per cell,[35,36] the observed spontaneous recombination efficiency of ~0.01-0.05% is very low. It is, however, on par with previous cassette mutagenesis literature in the absence of DSB.[37,38] dsDNA cassette mutagenesis thus proves insufficient to support large library size (>10^4) required for metabolic engineering.

3.3.2 Adapting Reiterative Recombination for library assembly

To overcome the limited efficiency observed by cassette mutagenesis, we adapt the highly efficient DNA assembly system Reiterative recombination for integration of pathway libraries directly in the yeast chromosome. Specifically, we use iterative cycles of targeted chromosomal recombination, originally designed for sequential assembly of single DNA fragments, to simultaneously assemble collections of DNA sequence variants. Slightly modifying the protocol described for single gene assembly (Figure 3-1),[9] we simply co-transformed libraries of variants with a linear donor plasmid into the acceptor strain (Figure 3-4). Libraries are accumulated in the same host by repeating the assembly cycle several times. The result is a collection of cells each carrying one of many possible gene combinations.
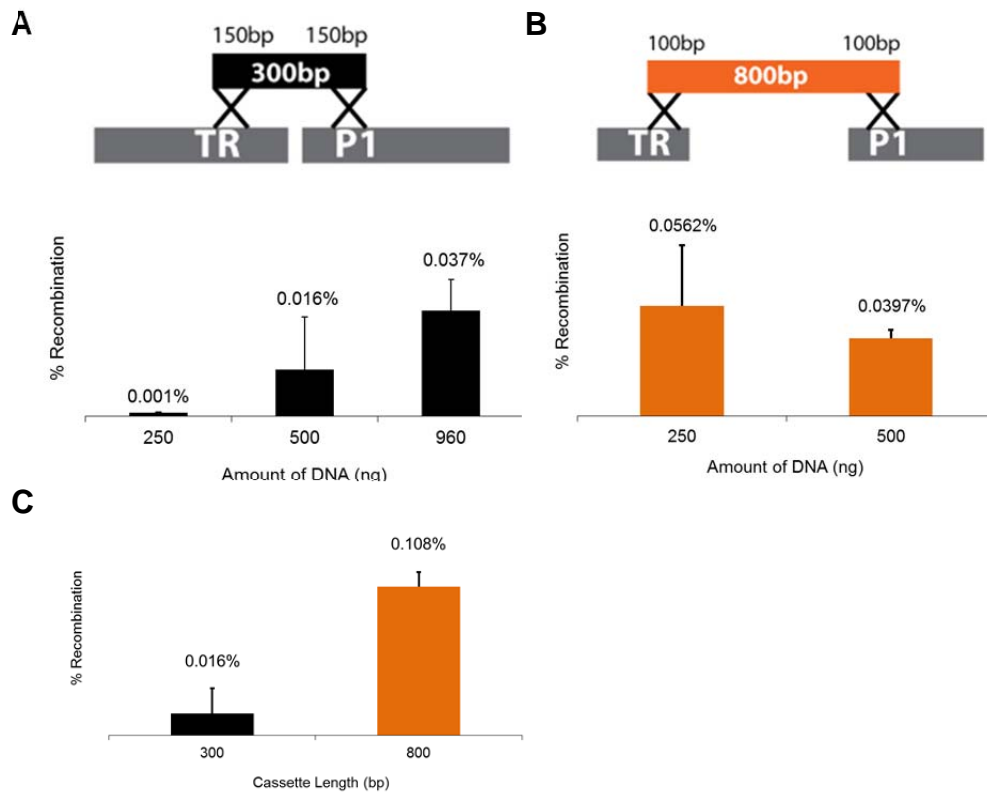
Figure 3-3. Efficiency of cassette mutagenesis in chromosomal TRP1 gene. Chromosomal mutations in the gene TRP1 were targeted for repair by transformation of dsDNA. The efficiency was calculated by comparing the number of cells able to grow on selective media lacking tryptophan and the number of total cells after transformation (no tryptophan selection). (A) Recombination efficiency for repair of point mutation using increasing amount of 300 bp dsDNA cassette. (B) Recombination efficiency for repair of deletion mutation using increasing amount of 800 bp cassette. (C) Effect of increased homology sequence (cassette length) on the efficiency of point mutation repair. 500 ng of either 300 bp or 800 bp dsDNA cassette was transformed into strain VWC2445. All experiments were performed in triplicates.

Like in the case of non-library DNA assembly, sequence homology guides the integration of fragments into the donor vector and acceptor strain. PCR is thus used to add homology on both ends of each of the library members. Thus, all library members are equally likely to undergo recombination with the donor plasmid and become integrated into the chromosome. However, since it is known each cell receives a random collection of library variants upon transformation and only single copy of the donor (centromeric) donor plasmid that guides its integration, each cycle eventually gives rise to a library of host cells, each carrying one integrated library variant at the acceptor site. Adding an additional library into an existing pool of variant cells increases the complexity of the library at each cycle.

3.3.3 Single gene attenuation using endogenous promoter library

We chose to demonstrate the utility of Reiterative Recombination for library construction by attenuation of gene expression using promoter libraries. Promoter elements for the regulation of transcription are widely used in all fields of molecular biology, and are a central tool for the optimization of pathways' flux in metabolic engineering strains.[2,39-42]

For attenuation of gene expression we used a promoter library consisting of five endogenous constitutive yeast promoters. The library members, CYC1, KEX2, GPD1, PGK1 and TEF1 are all widely used *S. cerevisiae* promoters previously characterized as having a wide range of promoter activity.[43] We used the same protocol as described for assembly of single DNA fragment with the following modifications. First, PCR was used to add all library promoters with common arbitrary 20 bp sequences on both ends to

facilitate analysis of recombinants (Figure 3-5). Additional PCR was used as previously described[9] to add homology to both the donor plasmid and the gene downstream of the promoter library. The pool of PCR products carrying all library variants with identical 5' and 3' sequences was transformed into the acceptor strain with the respective donor plasmid for each cycle.[9]
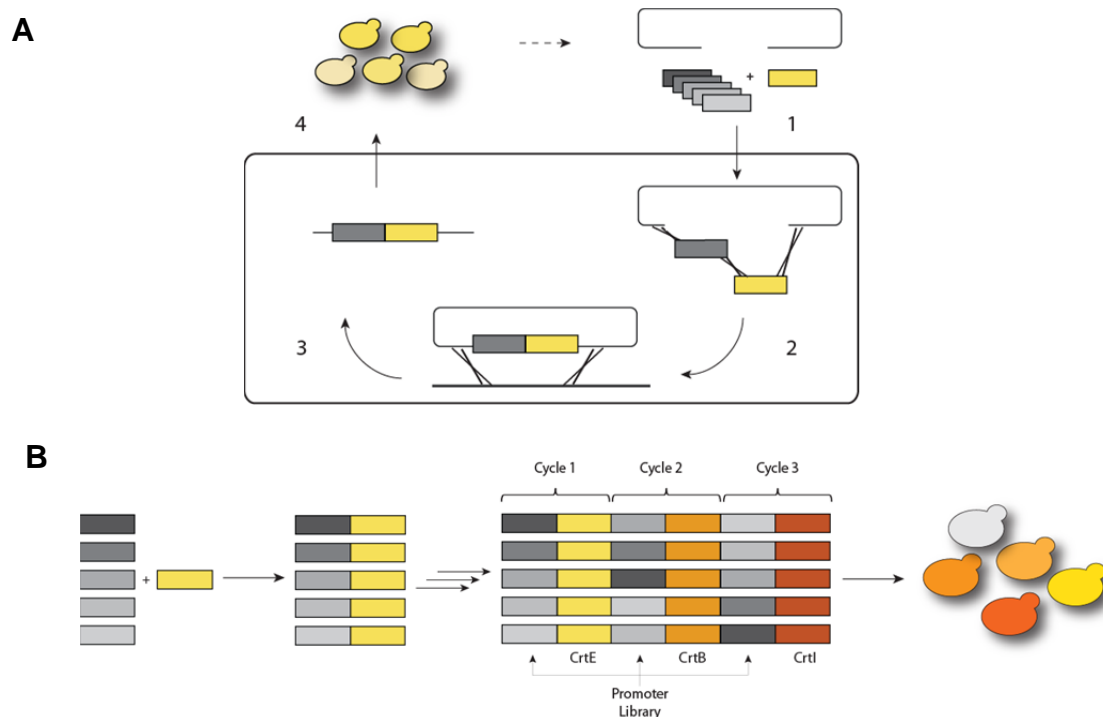
Figure 3-4. Adapting Reiterative Recombination for assembly of DNA libraries. (A) Schematic single cycle assembly of a DNA library. All fragments for assembly, including library of promoter variants (gray) and the respective Crt genes (yellow) are co-transformed with the donor plasmid into the acceptor strain (1). A single variant is then incorporated into the donor (2) and then the acceptor module (3), resulting in a pool of recombinant cells (4). (B) Iterative assembly of combinatorial libraries. Multiple gene libraries can be repeatedly assembled to form complex combinatorial pathways. Using a library of promoters (gray), the expression of multiple enzymes is altered to explore synergistic effects on lycopene yield.
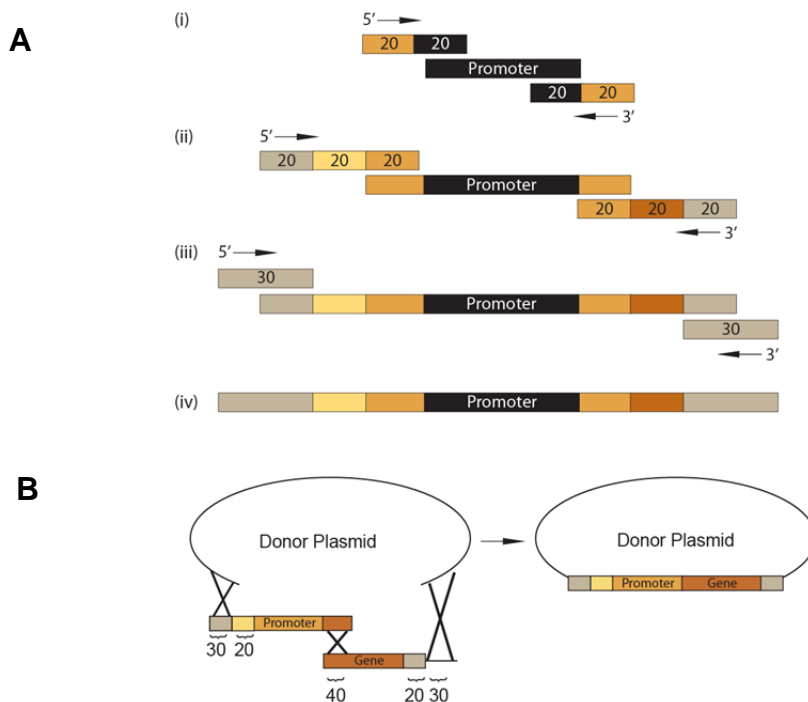
Figure 3-5: Preparation of DNA for library assembly via Reiterative Recombination. (A) PCR overview, demonstrated here for a promoter library. Each fragment was added with homology on both ends to enable recombination. (i) Common arbitrary sequences are added on both ends of each promoter. (ii) Homology is added to preceding assembly fragments (yellow), the respective Crt gene (brown) and donor plasmid (gray). (iii) Homology to donor plasmid is extended to make the final fragment (iv). (B) Upon transformation into yeast, spontaneous recombination occurs between overlapping sequences to circularize the final donor plasmid.

To test the adapted Reiterative recombination protocol, we first transformed the promoter library with a LacZ reporter gene. The assembly of LacZ with five promoters simultaneously was used to both validate the library protocol and calibrate the dynamic range of each promoter activity as a baseline for attenuation of lycopene pathway genes. A single Reiterative Recombination library assembly cycle was used to co-transform LacZ

with all five members of the library into an acceptor strain LMW2591. Induction of recombination was performed by galactose media, cells were cured of excess donor plasmid and selected for successful recombination events as previously described (Figure 3-1, 3-4).[9] Several recombinant colonies were picked, tested for LacZ activity by liquid assay and further analyzed by PCR of genomic DNA to confirm correct integration of both promoters and LacZ gene.

As shown in Figure 3-6, we were able to recover a variety of colonies displaying variable levels of LacZ activity. Sequencing confirmed all five library members were represented in the final pool of recombinant colonies. We then characterized gene expression level of each library promoter, and observed a dynamic range spanning 2 orders of magnitude in promoter activation. Interestingly, sequencing suggests the CYC1 weak promoter to be slightly more abundant among recombinant colonies than other library promoters (data not shown).

Next, we similarly characterized the attenuation of the first gene in the lycopene pathway, CrtE, with the same five-member promoter library. We used a codon optimized CrtE gene, encoding Geranylgeranyl pyrophosphate (GGPP) synthetase from *Erwinia herbicola* (kindly provided by Greg Stephanopoulos). CrtE performs the same function as the endogenous *S. cerevisiae* GGPP synthetase protein (BTS1 gene) and is expected to increase the cellular pool of GGPP (Figure 3-2). The ADH transcription terminator sequence was added in all experiments.
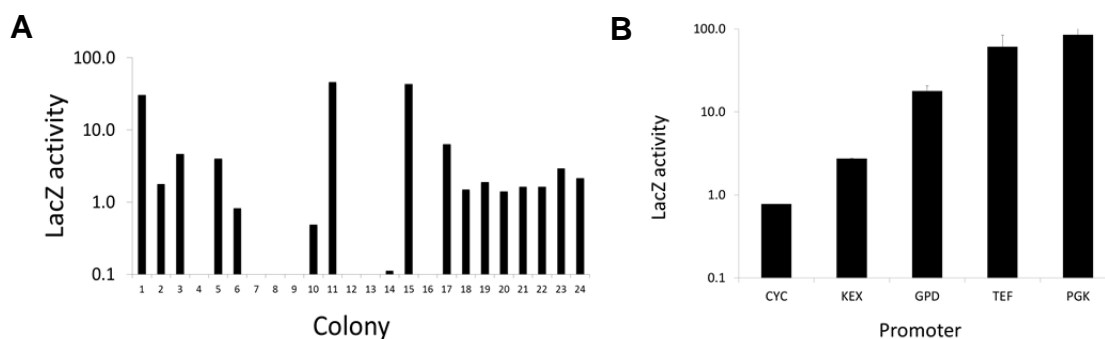
Figure 3-6. Attenuation of LacZ expression using constitutive promoter library. (A) A mix of five promoters was co transformed with the LacZ gene PCR product and integrated into the chromosome via Reiterative Recombination. The levels of LacZ activity were tested in 24 randomly selected recombinants by liquid assay. (B) Recombinant Colonies carrying each of the library promoters were individually used to characterize the expression level of each library member. The experiment was performed in triplicates.

We transformed CrtE into the acceptor strain with one, two, three or five of the library promoters in a single cycle. To control for bias towards either library members, equimolar amount of all library members were used at each transformation, keeping the total amount of DNA as well as induction and curing conditions constant. The distribution of promoters observed in one cycle of Reiterative Recombination assembly is presented in Figure 3-7.

Except for single promoter experiments, where all recombinants carry the expected promoter, the results suggest over-representation of the weak promoter CYC1 among CrtE library recombinants. Interestingly, the same trend was also suggested by sequencing of the promoter library used for LacZ (data not shown). It is unlikely that such bias originates from differences in recombination efficiency among library variants, since all members

carry the same 3' and 5' homology sequences and thus have equal probability for recombination with both donor and acceptor modules.

The bias can be explained, however, by competitive advantage resulting from integrating of CYC1 over other stronger library promoters. During the assembly process, variants are grown in liquid selective media. Fast growing variants might overtake the culture during liquid selection growth, while the number of slow growing variants becomes exponentially scarce. For example, strong expression of heterologous genes which do not contribute to cell survival, could cause a metabolic burden and lead to slower cell growth. Cells carrying strong promoters might thus be selected against during the curing and recombinant growth selection assay while variants carrying weak promoters grow faster and take over the population. Nevertheless, it should be noted that the other weak library promoter, Kex2, was not found to be over-represented among recombinant colonies. Extensive investigation of promoter bias at the single gene level is required to draw further conclusion. Such inherent selective bias might eliminate many variants at early cycles of assembly, decreasing combinatorial complexity of the final pathway library.
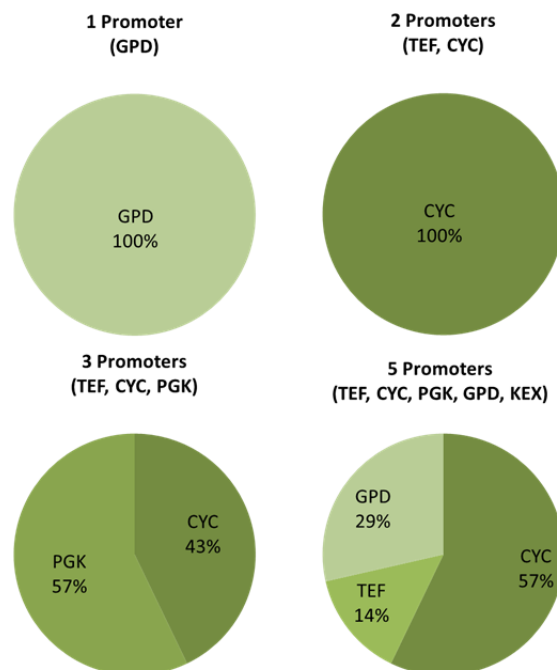
Figure 3-7. Distribution of library promoters observed for single assembly cycle. (A) CrtE gene was assembled with one, two, three or five promoters in a single Reiterative Recombination assembly cycle. Eight recombinant colonies were randomly picked, genomic DNA was extracted and promoters upstream of CrtE were identified by sequencing. The distribution of promoters among the sequenced colonies is presented. The experiment was performed three times with similar results.

3.3.4 Combinatorial attenuation of the lycopene pathway

Next, we proceeded to add the genes CrtB and CrtI to achieve expression of lycopene. Significantly, each of the Crt genes are added with the same five member promoter library, using the same assembly process described for CrtE. Specifically, ~1400 recombinant cells carrying CrtE with promoter library from the first cycle were used as the acceptor strain for transformation of the second cycle DNA, consisting of CrtB, five-member promoter library, and the respective donor plasmid for cycle 2. Induction and

curing steps were performed as previously described. For the third cycle of library assembly, CrtI was transformed with the same five-member promoter library into an acceptor strain comprised of cycle 2 recombinants. Finally, cycle 3 recombinants carrying all three lycopene biosynthetic pathway genes were plated and visibly screened for the expression of red color as a result of lycopene expression (Figure 3-8)

Notably, using the collection of recombinant colonies to initiate each cycle, we might be carrying on a certain number of non-cured colonies to the next cycle (i.e cells carrying excess donor plasmids that escaped the curing phase and therefore contain additional copies of library promoters). We thus added a selective marker TRP1 with the last pathway gene CrtI, to help 'clean out' the last library cycle and further verify the selection of only these recombinants carrying both integrated markers. For each cycle, we observe ~1400 recombinant colonies (i.e galactose-induced, diluted for screening on selective plates from 1 ml) that were used to inoculate the next cycle of library assembly.
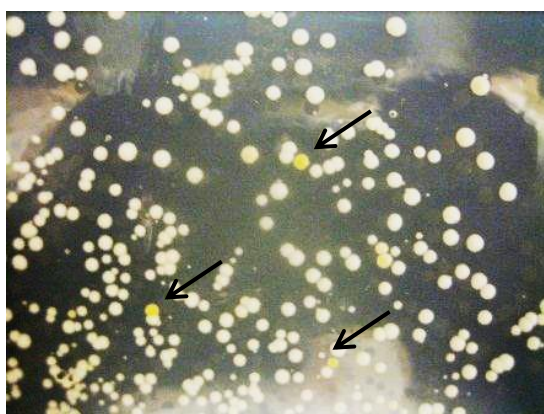
Figure 3-8. Recombinant colonies carrying a library of lycopene pathways. Three genes of the lycopene pathway were assembled, each with a library of constitutive promoters. The combinatorial library was plated on selective SC(LT-/FOA) media and incubated 3 days at 30 °C. 30 colonies were randomly selected for further analysis.

Interestingly, while the occurrence of white and orange colonies indicates that the multiple-gene attenuation library contains a range of lycopene expression levels, the majority of colonies observed were white with several easily detected orange colonies (Figure 3-8). The size of the library demonstrated in this work is relatively small, consisting of three cycles each with five-member library totaling 5^3=125 possible variants. Thus, to ensure the variability of the library is fully covered, the number of recombinants screened should exceed 3 times the library size. The number of colonies observed at each cycle after transformation (10^7 cells/ ml) and recombination ($10^{4-}10^5$ cells/ ml) ensures the library is fully covered. Notably, our analysis of recombinant colonies is aimed at demonstrating the feasibility and diversity of the library achieved by Reiterative Recombination technology, rather than isolating the highest yielding colony.

Thirty colonies of the final cycle (cycle 3) were chosen for further analysis. To gain insight into the distribution of promoters as well as the effect of expression level on cell growth and lycopene yield, we characterized genomic and cellular phenotype. First, PCR and sequencing of genomic DNA was performed to identify which promoter is driving the expression of lycopene biosynthetic genes CrtE, CrtB and CrtI. Lycopene content was quantified by extraction into organic solvent. Finally, growth assay was performed to assess the degree of metabolic burden or other changes affecting cell survival. The results are summarized in Figure 3-9.

Several trends were observed. First, we found the promoter upstream of the first pathway gene, the GGPP synthase CrtE, to be a strong promoter for the majority of recombinant colonies. The second enzyme CrtB, which converts GGPP to phytoene, carried strong promoters (GPD) at colonies with low lycopene production, and weak

promoter (CYC1) at colonies expressing high level of lycopene. Interestingly, the promoter upstream of the third biosynthetic gene CrtI, encoding lycopene synthase, shows the opposite trend; weak promoters are observed for colonies with low lycopene expression whereas strong promoters are observed for colonies with high lycopene expression (Figure 3-9).

CrtE is a GGPP synthase that duplicates the functionality of the endogenous *S. cerevisiae* GGPP synthetase gene BST1 (Figure 3-2). It thus increases the pool of GGPP for the production of lycopene, but also as a precursor for multiple other cellular pathways. It is thus not surprising to find all colonies to carry strong CrtE promoters.

CrtB and CrtI both encodes heterologous functional gene diverting *S. cerevisiae* metabolite flux towards the production of lycopene (Figure 3-2). It is interesting, thus, that the distribution of promoters driving CrtB and CrtI shows opposite trends. Clearly, strong GGPP synthase (CrtE) and lycopene synthase (CrtI) expression is sufficient for the accumulation of lycopene, even with low CrtB expression, Suggesting these enzymes are the bottleneck for flux thorough the pathway. It is possible that other factors not considered here such as cell growth or regulation of the precursor availability, also contribute to the regulation of lycopene production and therefore have an effect on the final composition of recombinants.

Overall, library characterization provided interesting insights into the flux through the lycopene pathway that is in agreement with previous published work. Extensive characterization of larger number of variants from each cycle is required in order to unequivocally determine the most efficient scheme for promoter attenuation supporting high lycopene yield.

It was previously demonstrated that high level expression of heterologous products could result in metabolic burden.[16,44] To test for metabolic burden caused by the lycopene pathway library, we determined the doubling time for each variant and compared it with the doubling time of the parent Reiterative Recombination strain (carrying no lycopene expression) (Figure 3-9). As shown in Figure 3-9, we found no correlation between doubling time and lycopene expression levels for the thirty colonies characterized here. It is possible that the levels of lycopene in the selected variants were too low to induce metabolic burden. Alternatively, it is possible that the process of library assembly by Reiterative Recombination inherently selects for fast growing variants with low metabolic burden.

3.3.5   Synthetic promoter library

Several synthetic promoter libraries have been constructed for metabolic engineering applications.[1,45-47] We used a synthetic promoter library constructed by the Stephanopoulos lab that consists of 11 variants of the constitutive Translation elongation factor 1 (TEF1) promoter,[1] spanning 8 to 120% of the unmutated TEF1 activity. Importantly, each variant carries between 4 and 71 point mutations of the wild type promoter, retaining very high degree of homology among library members (Figure 3-10). This library was specifically chosen to challenge our recombination-dependent assembly process, testing the effect of highly homologous variants on the effectiveness of donor-acceptor DNA exchange.

A

B

Colony	CrtE	CrtB	CrtI	Doubling time
5	GPD	GPD	CYC	4.0
2	GPD	GPD	CYC	4.3
3	GPD	GPD	CYC	4.4
4	GPD	GPD	CYC	3.8
7	TEF	GPD	CYC	4.3
1	GPD	GPD	CYC	2.5
8	CYC	GPD	CYC	3.9
13	GPD	GPD	CYC	3.8
12	TEF	CYC	CYC	3.4
6	GPD	GPD	CYC	9.4
27	GPD	CYC	CYC	4.0
24	CYC	KEX	PGK	7.5
23	GPD	CYC	CYC	4.7
22	GPD	CYC	PGK	3.7
21	GPD	CYC	CYC	4.8
25	TEF	CYC	PGK	4.5
28	GPD	CYC	PGK	2.9
30	GPD	CYC	CYC	4.2
20	GPD	CYC	PGK	3.2
18	GPD	CYC	PGK	4.0
11	GPD	CYC	PGK	3.7
LMW2681	GPD	PGK	TEF	2.9

Lycopene

C

Three synthetic promoter variants were chosen to evaluate the utility of the library for the attenuate the lycopene pathway by Reiterative Recombination. The weak, medium and strong promoters TEF2, TEF4 and TEF6, respectively, were added with appropriate homology on each end using PCR, as previously described (Figure 3-5).

Figure 3-10. Sequence alignment of synthetic TEF1 promoters used for pathway attenuation. Variants TEF2, TEF4 and TEF6 exhibit 0.07, 0.65 and 1.17 of native un-mutated TEF1 promoter activity, respectively, as measured by reported gene expression by Nevoigt et al.[1] Interestingly, the variant with the strongest activity, TEF6, carries only a single nucleotide deletion. The promoters were transformed with each of the CRT genes to attenuate lycopene production.

As for the previous promoter library, attenuation of the lycopene pathway gene CrtE was performed by co-transformation of linear donor plasmid with equimolar amounts of all three promoters, the fragment encoding CrtE and the tADH transcription terminator. The resulting recombinant colonies were analyzed for chromosomal integration by PCR restriction analysis and sequencing. As previously described, cycle 1 (CrtE) library recombinants were used as an acceptor strain for the following two cycles of library

assembly for CrtB and CrtI. As before, cycle 3 was added with an additional selective marker TRP1 and screened by eye for expression of lycopene.

Sequencing characterization of only 10 colonies of the first cycle of library assembly (CrtE) suggested all three promoters to be represented with no bias in promoter representation as previously observed for the natural five-member promoter library (data not shown). This could be due to the difference in the range of promoter activity between the synthetic and natural promoter libraries, since the two libraries have been calibrated separately and were not directly compared; the current synthetic library consists of 3 promoters ranging 0.07 to 1.17 of natural strong promoter TEF1 as measured by GFP expression,[1] whereas the previous five-member natural library ranges ~100-fold as measured by LacZ expression above.

Unlike previous library assembly experiment, no orange colonies were detected in recombinants resulting of three-cycle assembly using TEF1 synthetic promoters. Although the correct selective marker exchange occurred as expected in all three cycles (as confirmed by selective conditions), we could not detect the genes of the second library (CrtB) in the final cycle 3 recombinants, which explains the lack of orange lycopene-expressing colonies. This may be a result of spontaneous 'pop-out' mechanism triggered by multiple repeats of the almost identical TEF1 promoters at the acceptor locus, as frequently occurs in *S. cerevisiae* to eliminate sequences flanked by overlapping repeats.[48] Thus, the final recombinants, though carrying the correct selective markers as expected by Reiterative Recombination assembly, could result from a different mechanism that does not support the assembly of lycopene genes. We did not pursue this library further.

3.3.6 Combinatorial gene library guided by Flux Balance Analysis

To demonstrate the utilization of our assembly method beyond promoter libraries, we explored the assembly of entire gene libraries. Previously, it has been shown that changes in expression of host genes involved in the isoprenoid or competing pathways can help alleviate terpenoid yield by increasing precursor availability.[3,27,49,50] For example, HMG-CoA reductase (enzyme HMG1), which converts HMG-CoA to mevalonate (Figure 3-2), was found to have a significant effect on isoprenoid pathway yield.[3,50] Thus, we used Reiterative Recombination to simultaneously examine the effect of several *S. cerevisiae* genes on lycopene yield.

In order to identify host gene likely to affect the biosynthesis of lycopene, we turned to computational modeling via Flux Balance Analysis. In collaboration with Mattheos Koffas laboratory at Rensselaer Polytechnic Institute (RPI), *S. cerevisiae* metabolic network was modeled to predict endogenous host genes whose deletion, over expression or attenuation is expected to significantly enhance lycopene yield. Of the genes identified by the algorithm, we chose the following five member library: ALD6, MVD1 and ERG10 were selected of the genes suggested for over expression. ALD6 is a cytosolic aldehyde dehydrogenase required for conversion of acetaldehyde to acetate. MVD1 (ERG19) is mevalonate pyrophosphate decarboxylasethe. Acetoacetyl-CoA thiolase ERG10 is involved in the first step in mevalonate biosynthesis, forming acetoacetyl-CoA. All three genes are expected to alleviate precursor availability for lycopene biosynthesis. The genes ARO2, involved in the biosynthesis of chorismate, and HXK1, catalyzing phosphorylation of glucose during glucose metabolism, were suggested as deletion targets and were chosen as negative controls.

We chose to shuffle these genes with the promoters previously used for attenuation of the pathway, namely the five-member constitutive promoter library (Figure 3-6). Each gene could pair with any of the five promoters giving rise to 25 possible combinations. Each library gene was added with arbitrary common 20 bp sequence tags on both 5' and 3' ends to facilitate recombinant analysis, as well as the necessary homology for assembly and to a common terminator sequence. Equimolar amounts of gene and promoter libraries were co-transformed into an acceptor strain carrying all lycopene biosynthetic genes under constitutive strong promoters (pGPD-CrtE, pPGK-CrtB, pTEF-CrtI, visibly orange in color). Recombinant colonies were plated on selective media and screened by eye for change in color after 3-5 days incubation.

We could not visibly detect significant changes in phenotype between the starting acceptor strain expressing lycopene and the resulting library recombinants carrying an additional gene. Nevertheless, growth rate and lycopene yield were measured for 30 colonies picked at random. As seen in Figure 3-11, though all recombinant colonies displayed higher lycopene yield than the previously constructed CrtEBI promoter attenuation library, we observed very small changes among variants.

Several explanations could account for the lack of change in lycopene yield. First, it is possible that the genes chosen for library mutagenesis have no significant effect on lycopene yield under the experimental conditions used. Alternatively, the competitive pressure applied by liquid selection in Reiterative Recombination could have resulted in selection of only specific promoter/gene pairs in all colonies. The library was not characterized further.

A

B
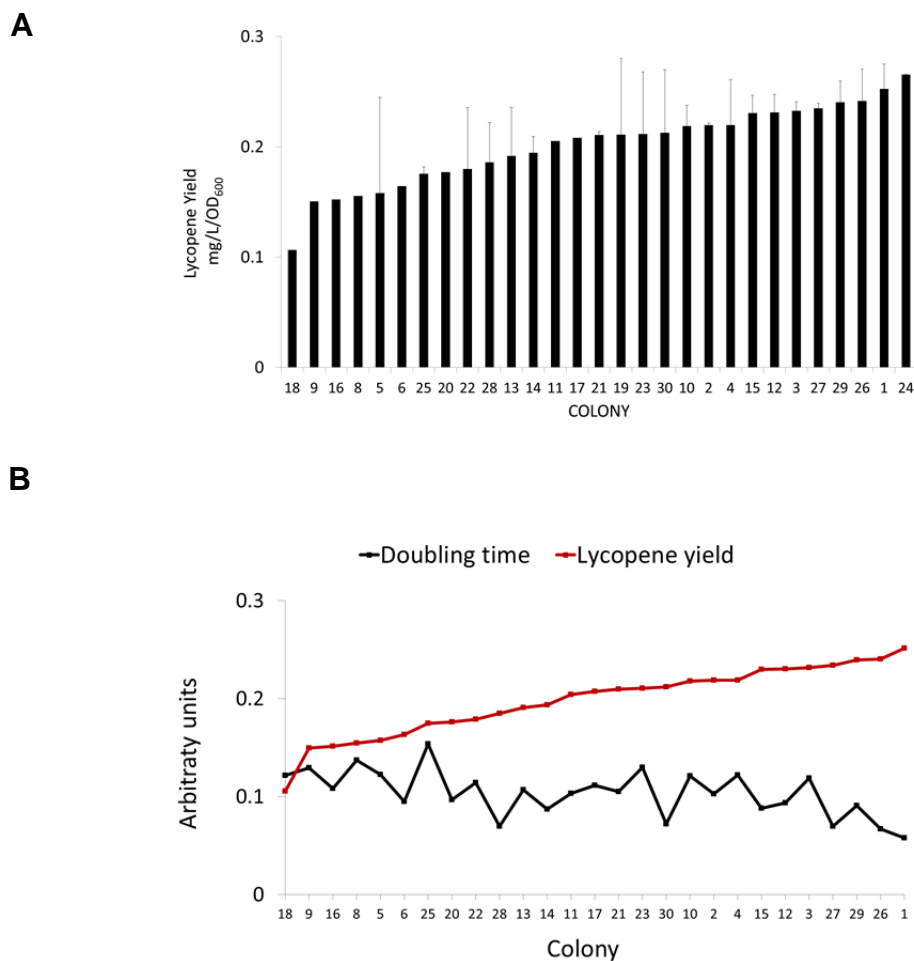
Figure 3-11. Characterization of gene library recombinants. (A) Lycopene yield of thirty recombinant colonies of the fourth cycle of library assembly, adding additional copy of host genes to lycopene producing strains. Yield was measured by extraction of lycopene using acetone. The results are an average of three independent experiments. (C) Plotting of both lycopene yield and growth rate for thirty recombinants (both parameters normalized to arbitrary units for comparison).

3.3.7 Exploring lycopene screening methods

The colorimetric screen of lycopene expression has a limited dynamic range in *S. cerevisiae* compared to *E. coli*, making it difficult to distinguish different levels of lycopene production by visual screen of colonies. Furthermore, the extraction of lycopene from individual colonies for characterization of lycopene yield is a laborious time consuming task. Thus, In search of alternative lycopene-dependent *in vivo* phenotypes with increased sensitivity and dynamic range, we tested lycopene absorbance in liquid cultures as well as the effect of the steroid inhibitor Fluconazole as potential alternative methods to monitor and select for lycopene production.

Liquid growth selection we first tested whether the accumulation of lycopene, characterize by visible change in colony color on plates, can be efficiently adapted to a more sensitive liquid detection assay. Lycopene absorbance has characteristic peaks at 360nm, 443nm, 471nm and 502nm. This absorbance signature is clearly detectable in liquid lycopene-expressing cultures (Figure 3-12). We thus tested the feasibility of measuring increase in lycopene yield using liquid culture absorbance at 480nm.

The following strains were compared: strain LMW2591 is a Reiterative Recombination acceptor strain carrying no lycopene pathway genes which served as a negative control. Strains LMW2681 and LMW2671 express lycopene using CrtEBI genes on the chromosome, with or without an additional copy of HMG1 host genes, respectively. Strain DR117 expresses lycopene using CrtEBI genes on a low copy plasmid. All strains were grown under the appropriate selective conditions to maintain the pathway genes. OD_{600} was measured as indication of cell growth and OD_{480} was measured as indication of changes in lycopene content. The results are shown in Figure 3-12.

Although growth rate was similar for all strains, we found a significant difference in absorbance between lycopene producing strains versus control. Interestingly, we see an increase in the ratio OD_{480}/OD_{600} over time, with expression from low copy plasmid slightly lower than chromosomal expression. This might be explained by variation in the accessibility of DNA or transcription regulation at different loci. Lycopene content seems to increase as the cells proceed into stationary phase. Overall, while we can clearly detect difference in absorbance between control and lycopene producing strains, these differences are relatively small (difference of ~0.1 in ratio $OD_{480/600}$). These results suggest a liquid assay can be used to study the accumulation of lycopene and its effect on cellular phenotype, but might not be sufficiently sensitive to identify small differences in lycopene yield between variants in a directed evolution experiment.

Fluconazole inhibition. Lycopene is made in *S. cerevisiae* from endogenous FPP molecules. However, FPP also serves as precursor for the endogenous biosynthesis of essential metabolites such as farnesyl and steroids (Figure 3-13). We therefore hypothesized that expression of lycopene might affect the flux towards the steroid pathway, resulting in alteration of cellular resistance to steroid inhibitors. Differences in resistance phenotype could be utilized as a growth selection to isolate lycopene producing variants out of large metabolic libraries. The commonly used antifungal drug Floconazole, inhibiting the conversion of inositol to ergosterol in the biosynthesis of steroids, was used to test this hypothesis.[51]
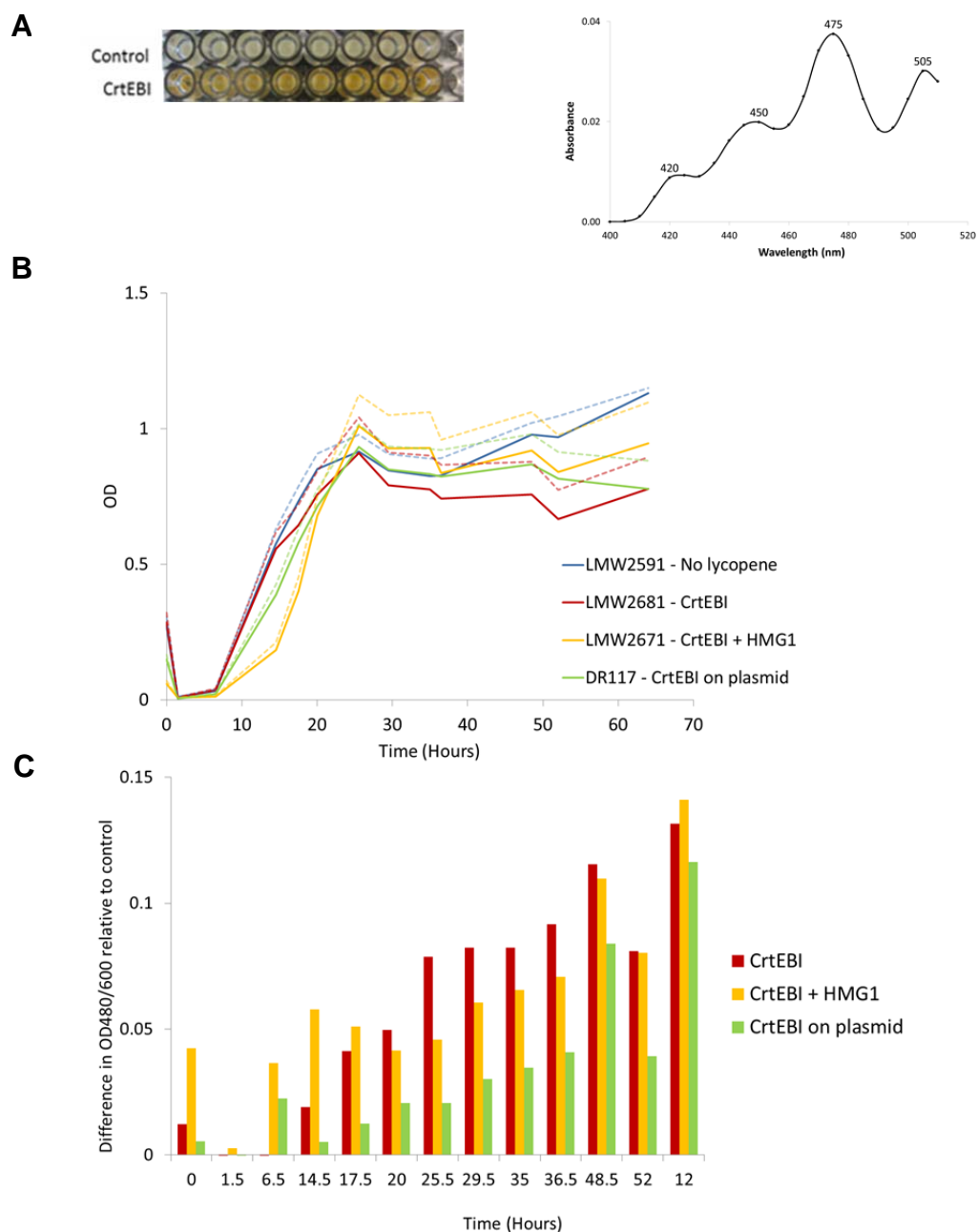
Figure 3-12. Liquid screen assay for lycopene expression. (A) Phenotype and absorbance spectrum of lycopene expressing *S. cerevisiae* in liquid culture. (B) Growth assay monitoring both cell density (OD_{600} -solid lines) and lycopene absorbance (OD_{480} -dashed lines) for strains LMW2591 LMW2681, DR117 and LMW2671 in selective media. (C) The difference in absorbance ratio OD_{480}/OD_{600} was calculated, normalized to control strain LMW2591 and plotted against growth duration.

An 'inhibition zone assay' (Halo assay) was performed to test for strain sensitivity to fluconazole using lycopene producing strain (LMW2681, carrying gene CrtEBI on the chromosome) and control strain carrying no Crt genes (LMW2591). To further evaluate the sensitivity of this assay, we also tested several variants of the promoter CrtEBI library described above which exhibit low lycopene yield. As expected, we found lycopene expression to elicit a significant difference in the resistance of *S. cerevisiae* to fluconazole. Surprisingly, however, lycopene did not increase the sensitivity of the strain but rather increased its resistance to the inhibitor, as indicated by a decrease in inhibition zone (Figure 3-13). All library variants exhibited similar inhibition to that of the control strain lacking lycopene, regardless of their measured lycopene yield (data not shown).

We thus conclude that the increased resistance of lycopene producing cells to the steroid inhibitor fluconazole can be potentially utilized to increase the selective pressure in directed evolution experiments. For example, fluconazole added to the growth media might provide a selective advantage to high yielding cells, facilitating their isolation in competitive liquid selection. Yet the sensitivity of this inhibition assay, at least under current conditions, is not sufficient to distinguish between variants with low lycopene yield.
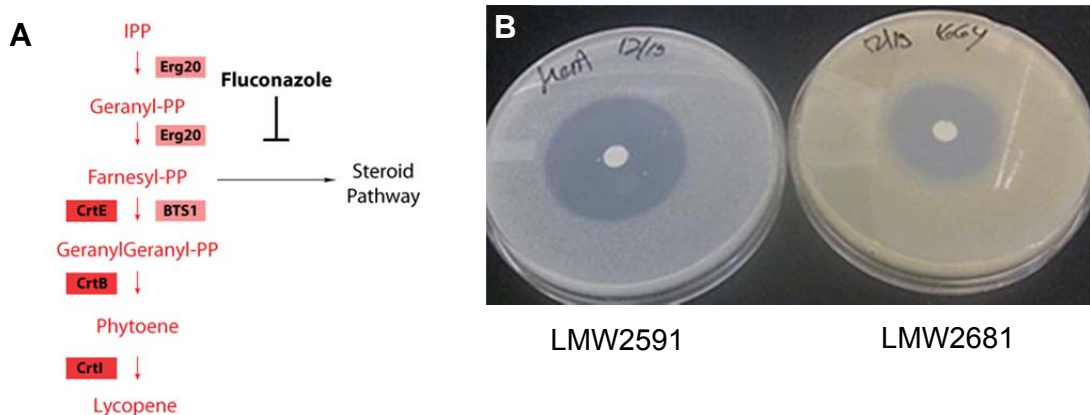
Figure 3-13. Effect of lycopene expression on fluconazole resistance. (A) Schematic of Fluconazole metabolic inhibition. (B) Fluconazole inhibition assay. 300μg/ml Fluconazole was applied onto sterile paper disc and placed on culture plates. The inhibition zone of lycopene expressing *S. cerevisiae* (LMW2681) and control (LMW2591) was monitored after 24 hours incubation at 30ºC.

3.4 Discussion

Despite great advances in industrial microbiology, genetics and computational tools, efficient engineering of metabolic pathways is still a complex and laborious task. Recent breakthroughs in the field[2-4] demonstrate the enormous potential of microbial host engineering for the production of pharmaceuticals, commodity chemicals and biofuels, but also emphasize the laborious nature of engineering each new target molecule.

Much like in the design of new catalysts, directed evolution has the potential to circumvent our lack of comprehensive understanding of biological networks, providing powerful tools for their manipulation. However, while enzyme evolution is now common practice for engineering of new catalysts, the highly efficient methodologies optimized for

single catalysts cannot be simply extrapolated for metabolic networks. Classic molecular biology tools for plasmid construction, mutagenesis, optimization and analysis for cloning of single genes prove laborious and inefficient when applied to multiple genes, and specifically in the context of large combinatorial libraries.

For example, we have previously shown DNA cassette mutagenesis to be compatible with library mutagenesis by transformation when targeting a high copy plasmid. However, the low efficiency of this commonly used method for chromosomal mutagenesis prevents us from using it for metabolic engineering. The same problem was recently addressed, for example, by enhancing recombination efficiency using viral proteins combined with high throughput automated mutagenesis in MAGE.[30,41]

Here, we show that Reiterative Recombination technique can be utilized for directed evolution of metabolic pathways. The key property allowing the translation of this system to library mutagenesis is its significantly high recombination efficiency in the yeast chromosome (up to 10%), which stems from the coupling of DNA recombination with timely induction of DSB at the target site. This unique assembly system provides access to an optimization process that is at the heart of metabolic engineering, enabling both the assembly of heterologous genes and the attenuation of expression simultaneously. Attenuation of gene expression, as opposed to overexpression alone, opens access to metabolic landscapes previously unattained by traditional methods. Thus, Reiterative Recombination offers a user friendly, efficient technology for metabolic engineering that is available to non-experts yet is compatible with the complexity of metabolic design.

To adapt the method from single fragments to library assembly, by co-transformation of multiple DNA sequence variants at each cycle. Upon chromosomal

integration, a single library variant is integrated at the acceptor loci on the chromosome. Iteration of this assembly cycle enables accumulation of libraries to increase combinatorial complexity.

We demonstrate the construction of a small library for the attenuation of lycopene expression. A body of literature on the engineering of microbial systems for expression of lycopene provided us with a wealth of information to test our system.[27,28,31,50,52] We tested two promoter libraries, a five-member natural promoter library and a three- member synthetic promoter library constructed by the Stephanopoulos laboratory. Both libraries contain constitutive promoters and display a wide dynamic range of activity.

While the synthetic library was successfully used for single cycle library assembly, we observed DNA instability when carried to the second (CrtB) and third (CrtI) libraries. We hypothesize the high similarity among library variants, all mutants of the same promoter TEF1, contributed to increase in undesired recombination among promoters as more libraries were added. This observation is supported by sequencing data suggesting recombination also between individual TEF1 promoters to make hybrid variants. Similar mixing of promoters has not been observed for the other, natural promoter library.

Since many natural pathways, such as polyketide biosynthetic pathways, carry repeated and duplicated sequences, we find it extremely important to address the issue of highly homologous sequences for Reiterative Recombination assembly.

Alternatively, five commonly used constitutive yeast promoters carrying very little sequence homology were used to attenuate the lycopene biosynthetic genes CrtE, CrtB and CrtI. We successfully carried out 3 cycles of assembly, sequentially integrating each of the

lycopene genes with its own library of promoters. This small library of 125 combinations was well covered by the number of recombinant colonies at each cycle.

Sequence analysis of a single library assembly cycle suggests possible bias towards the selection of weak library promoters. This could be due to selection against metabolic burden caused by overexpression of a heterologous gene that does not contribute to cell survival. Growth rate could be a significant factor contributing to the enrichment of specific variants. We continue to investigate this issue by expanding our characterization effort to a larger number of variants and to previously unstudied intermediates of the assembly process. Specifically, decreasing the stringency of competitive growth selection by exchanging liquid culture with solid media could increase the survival of lycopene producing recombinants. We further plan to compare growth rate of individual strains carrying different promoter activities to test this concern.

A wide range of lycopene production levels was observed in the final three-library recombinant pool, spanning ~10-fold in lycopene content. Notably, while the vast majority of colonies exhibit very low lycopene yield, single orange colonies were clearly visible. We characterized 30 recombinant colonies by DNA sequencing, lycopene extraction and growth rate analysis. Overall, we did not find a correlation between lycopene content and growth rate, indicating no metabolic burden is caused by low level lycopene expression.

Several interesting trends were observed for the attenuation of lycopene biosynthetic genes by the five-member promoter library. First, we note no colony was observed where all Crt genes are driven by strong promoters. Exhaustive characterization of a larger number of colonies is needed to determine if indeed such combination was not produced by the library at hand. Regardless of lycopene yield, the vast majority of

colonies tested were found to carry a strong promoter upstream of the CrtE gene. This is not surprising since this additional copy of GGPP synthase is expected to enhance the availability of isoprenoid precursors for many cellular pathways.

An interesting relationship was observed between the promoters found upstream of the heterologous genes CrtB and CrtI. While high levels of lycopene seem to require a strong promoter upstream of CrtI, these colonies also all carry a weak promoter upstream of CrtB. First, this finding suggests CrtI expression is sufficient for high level of lycopene production, even at low CrtB transcription, given the availability of GGPP precursor. The flux though the lycopene pathway is thus determined predominantly by the activity of lycopene synthase (CrtI). (Figure 3-9).

Expanded the library assembly beyond promoters, five genes of the host S. cerevisiae were selected by flux balance analysis modeling and shuffled into a strain producing lycopene with the same library of promoters previously used. Initial characterization of the library suggests no significant changes in lycopene yield resulted. This might be attributed to low contribution of the selected genes to lycopene yield, inherent selection against increase of yield due to metabolic burden, or may also result from additional mutations in either CRT or library genes. We did not characterize this library by sequencing, and thus further analysis is required to assess the effect of gene shuffling.

Finally, in light of the low sensitivity of visible colony screen for lycopene, we attempted to find a cellular phenotype that can be used for high throughput identification of lycopene production in yeast. Both liquid colorimetric screen and the steroid inhibitor fluconazole were tested and found to provide significant detectable signal correlating with

lycopene production. While suggesting a potential enhancement in the isolation of high yielding strains, both methods were shown to be insufficiently sensitive to distinguish low level of lycopene, and therefore cannot be used for metabolic libraries such as the one presented above.

Overall, these results demonstrate Reiterative Recombination to be a powerful method for combinatorial assembly of metabolic libraries in yeast. Each cycle for assembly of each library was constructed in one week, using a technically simple protocol and basic yeast genetic techniques. Further investigation of intermediate recombinants (i.e. cycle 1 and 2) is required to support inherent selective pressure under current conditions.

The selection for cell growth, rather than specifically selecting for a desired metabolic product, is a major recognized bottleneck in the field of metabolic engineering. Our library construction method allows assembly of any desired pathway, independent of selectable product. However, the selection for auxotrophic markers inherently implies that changes in growth rate can indirectly affect the library outcome. Altering the assembly conditions as well as testing other metabolic pathways using the same protocol can provide insight into the generality of our selection system and its limitations. For example, changing the duration or stringency of growth selection could provide further insights into the underlying evolutionary pressure, and attenuation of other pathways with the same promoter library could confirm the trends observed in initial experiments.

Though directed evolution technologies clearly have the potential to revolutionize the way metabolic engineering is currently performed, our results also shed light on some of the challenges that will accompany this approach. Extensive investigation of these

bottlenecks can help us get a head start on designing the next generation of transformative technologies to address these challenges.

In addition, the results presented here provide the basis for investigation of the flux through the terpenoid pathway. It is possible that the combination of promoters observed in this initial library is a result of additional factors not considered here, such as feedback inhibition or unknown effects of phytoene and lycopene in yeast. It is also interesting to see if similar patterns emerge when constitutive promoter library is replaced with inducible promoters that are activated only after the entire pathway is assembled. Another intriguing option is to change the order in which the genes are assembled to give the same final pathway library, or simultaneously assemble multiple genes in a single cycle.

Building from this work, libraries can be constructed using not only promoters, but also enzymes, receptors, transcription factors and mRNA. Reiterative Recombination can be further used to construct new signaling pathways or encode synthetic circuits, harnessing yeast as a scaffold not only for the production of metabolites but also for developing new useful devices from biological components.

3.5 Experimental methods

General materials and methods. General materials and methods were as in **Chapter 2**.

Reiterative Recombination was performed as previously described[9] with the following adjustments: Equimolar amounts of all library members (PCR product) were co-transformed with the respective gene and donor plasmid. After each library was cured, the entire plate (rather than single colonies) was scraped and used to inoculate the next library cycle.

Lycopene extraction protocol (Stephanopoulos lab/MIT). *S. cerevisiae* cultures were grown in 30C shaker for 48 hours in selective media. Before extraction, the culture OD_{600} was measured for normalization. 2 ml of culture were then centrifuged for 5 minutes at 14,000 rpm and the supernatant was removed. 200 µl glass beads (425-600 µm – Sigma product# G8772) and 1 ml of acetone (containing 100 mg/L BHT – Sigma product# B1378) were added to the cell pellet. The tube is kept on ice. a FastPrep vortex machine was used to break the cells by 3 cycles of 45 seconds each, at maximum speed (6.5). Cells were kept on ice in between cycles. The resulting lysate is carefully filtered using a 1 ml syringe and a 13 mm 0.2 µm PTFE syringe filter (VWR North American Cat.# 28145-491). OD_{474} is measured for each sample in a quartz cuvette, using acetone containing 100 mg/L BHT as reference sample. Lycopene concentration (mg/L) is determined from a standard curve (provided by Stephanopoulos lab at MIT) and divided by OD_{600} in order to normalize all samples to the number of cells.

LacZ ONPG assay. 200ul cultures of all tested strains and LWM2591 acceptor strain control were inoculated in SC selective media in round bottom 96 well plate and grown at 30ºC overnight . The next day, the cells were harvested by centrifugation and the pellets were resuspended in 100 µL of distilled water and transferred to a flat-bottomed 96-well plate to measure the absorbance at 600 nm. Then, the cultures were transferred back to U-bottomed 96-well plate, centrifuged for 5 minutes at 2000 rpm, the supernatant was removed, and the pellets were resuspended in 100 µL of the Y-Per Protein Extraction Reagent and lysed for 30 minutes. To measure the absolute β-galactosidase activity of the cells, the lysate was incubated with 8.5 µL of 10 mg/mL ONPG for about 10 min at 37°C until the positive control turned a yellow color. After the incubation, the reaction was

stopped by adding 110 μL of 1M Na_2CO_3. The lysates were centrifuged for 5 min at 2000 rpm and the supernatants were transferred to a flat-bottomed 96-well plate. The absorbance of the solution at 420 and 550 nm was measured using the HTS 7000 Plus BioAssay Plate Reader. The β-galactosidase units were calculated using the following equation: β-galactosidase = 1000 * (A_{420}/[A_{600} * time in minutes * volume assayed in mL]).

Preparation of dsDNA cassette for TRP1 mutagenesis: Two fixing DNA cassettes were made by PCR using an intact TRP1 gene. For repair of point mutations, a 300bp cassette was produced using primers NO32/NO33 and template vector pVC2278. For repair of deletion mutants, 800bp cassette spanning (-205) to (+620) of the TRP1 reading frame was produced using primers NO36/NO37 on template strain VC2271 (genomic DNA prep).

Preparation of library fragments. *Synthetic promoter library. F*ragments were were prepared by PCR as previously described for Reiterative recombination[9] using the following primers: cycle 1: TEF2, 4, 6: NO144-145, NO146-147, NO148-149, respectively. CrtE-tADH: LW449 with either NO150/151/152 for TEF2,4,6, respectively. Cycle 2 primers: TEF2,4,6: LW375 with NO156/ 155/ 154 (respectively), CrtB: LW374-NO153. For cycle 3: TEF2, 4, 6: NO160-161, NO162-163, NO164-165. CrtI-tACT-TRP1: LW461-456. Restriction analysis of library promoters: Spe1 digests TEF2 to make 129 bp and 273 bp. HindIII digests TEF4 to make 314 bp and 87 bp. HincII digests TEF6 to make 129 bp and 273 bp. for restriction analysis, individual colonies genomic DNA was purified and served as template for PCR u LW317-LW447 to give 1320bp PCR product. The

product was then digested with either enzymes (15 µl PCR product, 2 µl SpeI or HindIII or HincII, 2 µl NEB buffer 2, 2 µl H₂O), and incubated 4 hours at 37 ºC in a PCR block.

Natural promoter library was prepared using the same protocol, with the following adjustments: Cycle 1: Promoters: NO285-271. CrtE: NO272-LW447. tADH: LW448-449. Cycle 2: Promoters: NO286-303. CrtB: NO304-305. Cycle 3: promoters: NO267-268. CrtI: NO-269-229. tACT: NO221-LW459. Trp1: LW460-461. LacZ promoter library primers: Promoter library: LW584-NO252. LacZ gene: NO253-254 to add Flag tag, then NO253-255. tADH: NO256-296. Plasmid pSC201 (provided by Stephanopoulos lab) was used as template for PCR of CRT genes. The following PCR conditions were used: 95 ºC - 1 min, 95 ºC - 45 sec, 52 ºC – 1 min, 72c – 2 min, 72 ºC – 15 min. Repated steps 2-4 x25 times. Digestion products were analyzed on 2% agarose gel.

Gene library: the following arbitrary sequences were added to each library member gene 5' sequence: CTTTAATTCTAGCAAGTAAt. 3' sequence: TATGCATATGGTTCACAGGA. Generation of random Sequences was made by random DNA sequences generator (http://www.faculty.ucr.edu/~mmaduro/random.htm). PCR was performed using the following primers: ALD6: NO275-276, HXK1: NO277-278. ARO2: NO279-280. ERG10: NO281-282. MVD1: NO283-284. The following primers are used to add homology to all promoters 1-5 carry the same common ends: all PCR with primers NO291-292. Five host genes carrying the same common ends: all PCR with primers NO293-294. tADH: NO295-296.

Transformation of promoter library. Equimolar amounts of purified PCR products of library members were co-transformed with the gene of interest, a terminator sequence, and the relevant donor plasmid for each cycle. All fragments contain the

respective homology (30-40 bp) to allow their assembly into the vector by gap repair. The total amount of DNA was 500-1000ng. The resulting transformants were plated on the appropriate media. Induction and curing were performed as described [9]. After the final cured recombinants were plated, the entire plate was scraped using 1ml water to start the next cycle of library transformation.

Growth assay. Liquid growth assay was performed in 96 well plates using the appropriate selective media for the respective round of Reiterative Recombination assembly.[9] Promoter library variants (cycle 3) were growth in SC(LT-) media, whereas gene library variants (cycle 4) were grown in SC(HT-) media. 200 μL media was inoculated to initial OD_{600} of 0.05, and growth was monitored at OD_{600} or at OD_{480} for lycopene absorbance. Plate growth assay for lycopene promoter library was performed by diluting an overnight culture to OD_{600} of 0.1, 0.01 and 0.001, and plating 2 μL of each variant on selective SC (LT-) and non-selective SC(L-) agar plates.

Fluconazole Halo Assay. (Adapted from protocol[53] (http://www.med.unc.edu/~hdohlman/haloassay.html). A starter culture was grown at 30º C with shaking overnight, and 4 ml was mixed with sterilized solution of 0.5 % agar (autoclaved and cooled below 50ºC). The culture was then poured over the relevant selective plates - (SC (TL-) for LMW2681 or SC(H-) for LMW2591. Fluconazole was dissolved in sterilized water to final concentration 10 mg/ml. 300 μg fluconazole was applied on sterile filter paper disc, dried, and placed at the middle of the cultured plate using sterile technique. Plates were incubated at 30º C and inhibition zone was measured after 24 hours.

3.6 Strains, plasmids, and oligonucleotides

Table 3-1 Strains used in this study

Strain	Genes
FY251	MATa, ura3–52, his3Δ200, trp1Δ63, leu2Δ
VWC2245	W303 with RAD5+ (provided by Rothstein's lab)
LMW2591	Reiterative Recombination parental acceptor strain BY4733 *MATa-inc*
LMW2671	Round 4 recombinant from lycopene pathway
LMW2681	Reiterative Recombination parental acceptor strain BY4733 *MATa-inc,* Round 3 recombinants from lycopene library
DR117	FY251 carrying CEN plasmid with CrtEBI.

Table 3-2. Plasmids used in this study

Vector	Genes
pSC201	*URA3* integration plasmid with overexpressed *crtE*, *crtB*, and *crtI (*P. Ajikumar and G. Stephanopoulos)
pVWC2278	*pRS426 carrying wt S.Cerevisiae TRP1*

Table 3-3 Oligonucleotides used in this study

Name	Sequence
NO32	GATGCTGACTTGCTGGGT
NO33	AGCTGCACTGAGTAGTATG
NO36	TACTGTTGACATTGCGAAGA
NO37	TTAGTTTTGCTGGCCGCAT
NO38	TGTGTACTTTGCAGTTATGAC
NO39	CTTTTACACCATTTGTCTCC
NO144	AaaattgtgcctttggacttaaaatggcgtACGGCTCTAAAGTGCTTCG
NO145	GCTTTCGAACCAGAAACCATAAACTTGGATTAGATTGCTATG
NO146	aaaattgtgcctttggacttaaaatggcgt ATAGCTTCAAAACGCTTCTAC
NO147	GCTTTCGAACCAGAAACCATGAACTTAGATTAGATTGCTATG
NO148	aaaattgtgcctttggacttaaaatggcgtATAGCTTCAGAATGTTTCTACT
NO149	GCTTTCGAACCAGAAACCATAAACTTAGATTAGATTGCTATG
NO150	TAGCAATCTAATCCAAGTTTATGGTTTCTGGTTCGAAAG
NO151	TAGCAATCTAATCTAAGTTCATGGTTTCTGGTTCGAAAG
NO152	TAGCAATCTAATCTAAGTTTATGGTTTCTGGTTCGAAAG

NO153	TAGCAATCTAATCTAAGTTTATGAGTCAACCACCTTTGTT
NO154	AACAAAGGTGGTTGACTCATAAACTTAGATTAGATTGCTA
NO155	AACAAAGGTGGTTGACTCATGAACTTAGATTAGATTGCTA
NO156	AACAAAGGTGGTTGACTCATAAACTTGGATTAGATTGCTA
NO157	CTGTTGCGGAAAGCTGAAATAGAAACATTTTGAAGCTATATAGCTTCAGAATGTTTCTA
NO158	CTGTTGCGGAAAGCTGAAATAGAAACATTTTGAAGCTATATAGCTTCAAAACGCTTCT
NO159	CTGTTGCGGAAAGCTGAAATAGAAACATTTTGAAGCTATACGGCTCTAAAGTGCTTCGG
NO160	GgacgctcgaaggctttTCTTCGAAAAATTCGCGTCTACGGCTCTAAAGTGCTTCG
NO161	GTTTTCTTCATTTTGTAATTAAAACTTGGATTAGATTGCTATG
NO162	GgacgctcgaaggctttTCTTCGAAAAATTCGCGTCTATAGCTTCAAAACGCTTCTAC
NO163	GTTTTCTTCATTTTGTAATTAGAACTTAGATTAGATTGCTATG
NO164	Ggacgctcgaaggcttt TCTTCGAAAAATTCGCGTCTATAGCTTCAGAATGTTTCTACT
NO165	V GTTTTCTTCATTTTGTAATTAAAACTTAGATTAGATTGCTATG
NO275	CTTTAATTCTAGCAAGTAATATGACTAAGCTACACTTTGA
NO276	TCCTGTGAACCATATGCATATTACAACTTAATTCTGACAGC
NO277	CTTTAATTCTAGCAAGTAATATGGTTCATTTAGGTCCAAA
NO278	TCCTGTGAACCATATGCATATTAAGCGCCAATGATACCAA
NO279	CTTTAATTCTAGCAAGTAATATGTCAACGTTTGGGAAACT
NO280	TCCTGTGAACCATATGCATATTAATGAACCACGGATCTG
NO281	CTTTAATTCTAGCAAGTAATATGTCTCAGAACGTTTACAT
NO282	TCCTGTGAACCATATGCATATCATATCTTTTCAATGACAATA
NO283	CTTTAATTCTAGCAAGTAATATGACCGTTTACACAGCAT
NO284	TCCTGTGAACCATATGCATATTATTCCTTTGGTAGACCAG

3.7   References

1   Nevoigt, E. *et al.* Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **72**, 5266-5273, (2006).

2   Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* **330**, 70-74, (2010).

3   Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943, (2006).

4   Hawkins, K. M. & Smolke, C. D. Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat. Chem. Biol.* **4**, 564-573, (2008).

5   Wenzel, S. C. & Muller, R. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr. Opin. Biotechnol.* **16**, 594-606, (2005).

6   Galm, U. & Shen, B. Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis. *Expert Opin. Drug Discovery* **1**, 409-437, (2004).

7   Olano, C., Lombo, F., Mendez, C. & Salas, J. A. Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. *Metab Eng* **10**, 281-292, (2008).

8   Mutka, S. C., Bondi, S. M., Carney, J. R., Da Silva, N. A. & Kealey, J. T. Metabolic pathway engineering for complex polyketide biosynthesis in Saccharomyces cerevisiae. *FEMS Yeast Res* **6**, 40-47, (2006).

9   Wingler, L. M. & Cornish, V. W. Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15135-15140, (2011).

10   Kittendorf, J. D. & Sherman, D. H. Developing tools for engineering hybrid polyketide synthetic pathways. *Curr. Opin. Biotechnol.* **17**, 597-605, (2006).

11   Zhou, H., Xie, X. & Tang, Y. Engineering natural products using combinatorial biosynthesis and biocatalysis. *Curr. Opin. Biotechnol.* **19**, 590-596, (2008).

12   Runguphan, W., Qu, X. & O'Connor, S. E. Integrating carbon-halogen bond formation into medicinal plant metabolism. *Nature* **468**, 461-464, (2010).

13   Eustaquio, A. S., O'Hagan, D. & Moore, B. S. Engineering fluorometabolite production: fluorinase expression in Salinispora tropica Yields Fluorosalinosporamide. *J. Nat. Prod.* **73**, 378-382, (2010).

14   Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171-1176, (2005).

15   Xue, Q., Ashley, G., Hutchinson, C. R. & Santi, D. V. A multiplasmid approach to preparing large libraries of polyketides. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11740-11745, (1999).

16   Koffas, M., Roberge, C., Lee, K. & Stephanopoulos, G. Metabolic engineering. *Annu Rev Biomed Eng* **1**, 535-557, (1999).

17   Lartigue, C. *et al.* Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* **325**, 1693-1696, (2009).

18   Storici, F., Durham, C. L., Gordenin, D. A. & Resnick, M. A. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14994-14999, (2003).

19   Brachmann, C. B. *et al.* Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115-132, (1998).

20   Ajikumar, P. K. *et al.* Terpenoids: Opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharmaceutics* **5**, 167-190, (2008).

21   Chang, M. C., Eachus, R. A., Trieu, W., Ro, D. K. & Keasling, J. D. Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. *Nat. Chem. Biol.* **3**, 274-277, (2007).

22   Chang, M. C. & Keasling, J. D. Production of isoprenoid pharmaceuticals by engineered microbes. *Nat. Chem. Biol.* **2**, 674-681, (2006).

23   Roberts, S. C. Production and engineering of terpenoids in plant cell culture. *Nat. Chem. Biol.* **3**, 387-395, (2007).

24   Mijts, B. N. & Schmidt-Dannert, C. Engineering of secondary metabolite pathways. *Curr. Opin. Biotechnol.* **14**, 597-602, (2003).

25   Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat. Biotechnol.* **21**, 796-802, (2003).

26   Roberts, S. C. Production and engineering of terpenoids in plant cell culture. *Nat. Chem. Biol.* **3**, 387-395, (2007).

27   Farmer, W. R. & Liao, J. C. Improving lycopene production in Escherichia coli by engineering metabolic control. *Nat. Biotechnol.* **18**, 533-537, (2000).

28   Lee, P. C., Momen, A. Z. R., Mijts, B. N. & Schmidt-Dannert, C. Biosynthesis of structurally novel carotenoids in Escherichia coli. *Chem. Biol.* **10**, 453-462, (2003).

29   Sandmann, G., Albrecht, M., Schnurr, G., Knorzer, O. & Boger, P. The biotechnological potential and design of novel carotenoids by gene combination in Escherichia coli. *Trends Biotechnol.* **17**, 233-237, (1999).

30   Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898, (2009).

31   Schmidt-Dannert, C., Umeno, D. & Arnold, F. H. Molecular breeding of carotenoid biosynthetic pathways. *Nat. Biotechnol.* **18**, 750-753, (2000).

32   Dietrich, J. A., McKee, A. E. & Keasling, J. D. High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu. Rev. Biochem.* **79**, 563-590, (2010).

33   Igoucheva, O., Alexeev, V. & Yoon, K. Oligonucleotide-directed mutagenesis and targeted gene correction: a mechanistic point of view. *Curr Mol Med* **4**, 445-463, (2004).

34   Pirakitikulr, N., Ostrov, N., Peralta-Yahya, P. & Cornish, V. W. PCRless library mutagenesis via oligonucleotide recombination in yeast. *Protein Sci.* **19**, 2336-2346, (2010).

35   Gibson, D. G. *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20404-20409, (2008).

36   Pirakitikulr, N., Ostrov, N., Peralta-Yahya, P. & Cornish, V. W. PCRless library mutagenesis via oligonucleotide recombination in yeast. *Protein Sci.* **19**, 2336-2346, (2010).

37   Storici, F. & Resnick, M. A. The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.* **409**, 329-345, (2006).

38   Yamamoto, T., Moerschell, R. P., Wakem, L. P., Ferguson, D. & Sherman, F. Parameters affecting the frequencies of transformation and co-transformation with synthetic oligonucleotides in yeast. *Yeast* **8**, 935-948, (1992).

39   Sun, J. *et al.* Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in Saccharomyces cerevisiae. *Biotechnol. Bioeng.*, (2012).

40   Tyo, K. E., Alper, H. S. & Stephanopoulos, G. N. Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol.* **25**, 132-137, (2007).

41   Wang, H. H. *et al.* Genome-scale promoter engineering by coselection MAGE. *Nat Methods*, (2012).

42   Westfall, P. J. *et al.* Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E111-118, (2012).

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

44   Stephanopoulos, G. N., Aristidou, A. A. & Nielsen, J. *Metabolic Engineering Principles and Methodologies*. (Elsevier, 1998).

45   Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12678-12683, (2005).

46   Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res* **39**, 1131-1141, (2011).

47   Hammer, K., Mijakovic, I. & Jensen, P. R. Synthetic promoter libraries--tuning of gene expression. *Trends Biotechnol.* **24**, 53-55, (2006).

48   Rothstein, R. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol.* **194**, 281-301, (1991).

49   Alper, H., Miyaoku, K. & Stephanopoulos, G. Characterization of lycopene-overproducing E. coli strains in high cell density fermentations. *Appl. Microbiol. Biotechnol.* **72**, 968-974, (2006).

50   Alper, H., Miyaoku, K. & Stephanopoulos, G. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat. Biotechnol.* **23**, 612-616, (2005).

51   Kontoyiannis, D. P. Modulation of fluconazole sensitivity by the interaction of mitochondria and erg3p in Saccharomyces cerevisiae. *J. Antimicrob. Chemother.* **46**, 191-197, (2000).

52   Yamano, S., Ishii, T., Nakagawa, M., Ikenaga, H. & Misawa, N. Metabolic engineering for production of beta-carotene and lycopene in Saccharomyces cerevisiae. *Biosci. Biotechnol. Biochem.* **58**, 1112-1114, (1994).

53   Hoffman, G. A., Garrison, T. R. & Dohlman, H. G. Analysis of RGS proteins in Saccharomyces cerevisiae. *Methods Enzymol.* **344**, 617-631, (2002).

Chapter 4

Exploring the Dynamic Range of Genetic Selection

Using PRA-Isomerase Catalysis

4.1 Chapter outlook

A key component of a successful directed evolution experiment is the efficiency of its selection scheme. Genetic selection is perhaps the most powerful method for the isolation of a desired phenotype from a large population of cells, eliminating the growth of all undesired phenotypes under defined selective conditions. In *S. cerevisiae*, essential metabolic enzymes traditionally serve as genetic selective markers for plasmid and strain construction. However, in order to use growth selection for the identification of highly active enzyme variants, cell growth should be coupled with the level of enzyme activity. Here we examined the dynamic range of genetic selection by the common yeast auxotrophic marker TRP1, encoding the third enzyme in *S. cerevisiae* tryptophan pathway, phophoribosyl anthranilate isomerase (PRAI). We use mutants of PRAI varying in catalytic activity to study the effect of PRAI activity on the rate of cell growth. The role of metabolic enzymes in maintaining metabolic robustness and their utilization for broad range genetic selection is discussed.

4.2　Introduction

A broad variety of methodologies has been applied to generate improved enzymes for therapeutic and industrial applications by Directed evolution. A typical laboratory evolution experiment consists of iterations of mutagenesis and selection of desired functionality; [1] hence experimental evolution is only as powerful as its mutagenesis and selection efficiency allow for.

In a classic experiment, Beadle and Tatum pointed out that mutants can be screened based on enzymatic activity.[2] Selection schemes, unlike screens, do not require an active search for desired properties, but rather set the conditions to prevent the survival of undesired mutants, assessing up to 10^{10} clones in a single experiment in *E. coli* cells.[3] Specifically, it is in the realm of very large combinatorial libraries that selection gains critical importance.

Genetic selection is perhaps the most powerful technique currently available for analyzing large libraries in a fast, high-throughput manner. It involves the complementation of a missing essential function for cell growth and requires the coupling of enzyme activity to cell survival. Enzyme activity represents very stringent selection criteria, as even the smallest change in catalytic residues or structural integrity could cause complete inactivation.[4] Importantly, to distinguish highly active enzymes from less active ones, a genetic selection is needed that is sensitive over a wide dynamic range or enzyme activity.[5] Indeed, many directed evolution examples exist where increased stringency of genetic selection is utilized to detect improved enzyme catalysis, include the evolution of triosephosphate isomerase (TIM),[6] β-glucuronidase,[7] β-lactamase antibiotic resistance genes,[8] HIV thymidine kinase[9] and catalytic antibodies.[10]

Complementation of biosynthetic enzymes is traditionally used as selection assay in *S. cerevisiae*.[11] Yet central metabolism enzymes are known to be heavily regulated, making it difficult to predict the range of functional complementation of single enzymes of cell growth.[12] Despite several attempts to optimize the selection stringency of metabolic genes,[13] the dynamic range of such systems is still poorly characterized. To lay the foundations for using Trp1 genetic selection as a model system for developing novel *in vivo* evolution methodologies, we systematically characterized the dynamic range of this widely used selection scheme.
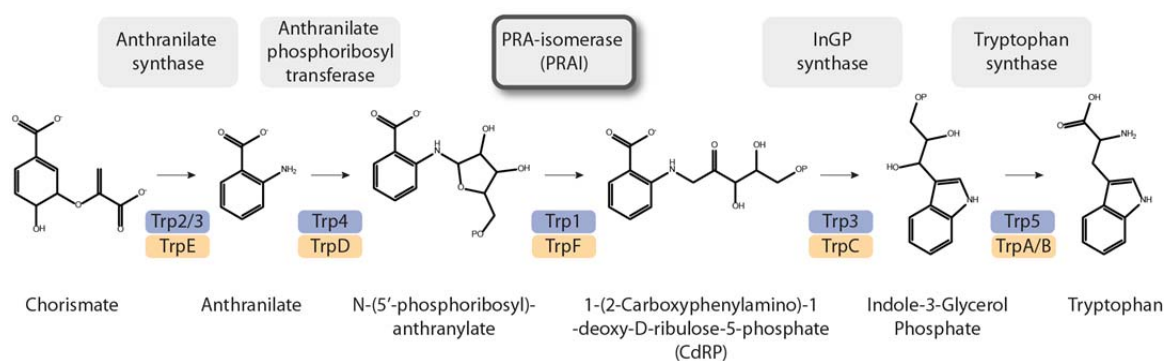
Figure 4-1. Tryptophan biosynthetic pathway. Enzymes (gray) and intermediates are indicated. PRA-isomerase (PRAI) is highlighted. S. cerevisiae genes indicated in purple, bacterial genes indicated in pink.

4.2.1 PRA- Isomerase as a model enzyme for genetic selection

Tryptophan synthesis from chorismic acid is a key metabolic pathway comprising five enzymes in *S. cerevisiae*.[14] Phosphoribosyl anthranilate isomerase (PRAI) is the third biosynthetic enzyme (Figure 4-1), catalyzing an Amadori rearrangement which is the irreversible isomerization of an aminoaldose to an aminoketose.[15] PRAI has been

classically used as an auxotrophic selection marker in yeast genetics, as cells lacking a functional enzyme cannot survive without exogenous tryptophan.[16,17] It belongs to the family of triose phosphate isomerase (TIM) $(\beta\alpha)_8$-barrel proteins (figure 4-2), the most frequently encountered fold among single domain enzymes which is adopted by 10% of all known protein structures.[18]

Extensive characterization of TIM $(\beta\alpha)_8$-barrel enzymes now support the notion of their divergent evolution from a common ancestor enzyme.[19-21] Considering the ubiquity of TIM barrel enzymes in biosynthetic pathways of central metabolism and their catalytic versatility,[22-24] it is perhaps not surprising that many TIM barrel enzymes display functional promiscuity[25-27] and that their catalytic activity can be interconverted via directed evolution.[28-33] In particular, PRAI has been the focus of much research in directed enzyme evolution,[28-30,34,35] providing surprisingly broad insights into the evolution of different metabolic pathways.[24,25,32]

Remarkably, it was shown that a single amino acid substitution is sufficient to elicit PRA isomerase activity in another TIM barrel enzyme, ProFAR isomerase, which performs an Amadori rearrangement in the histidine biosynthetic pathways,[24,34] although the two enzymes share only 10% sequence similarity. The ability of a single amino acid exchange to drastically affect enzyme activity implies a close correlation between structure and function.

To pursue our long-term aim of utilizing PRAI as a model system in the context of enzyme engineering, we sought to calibrate the dynamic range of the tryptophan genetic selection in *S. cerevisiae*. By mutating the active site of PRAI we constructed variants

ranging in catalytic efficiency, and used them to systematically probe the relationship between enzyme catalysis and cell survival.
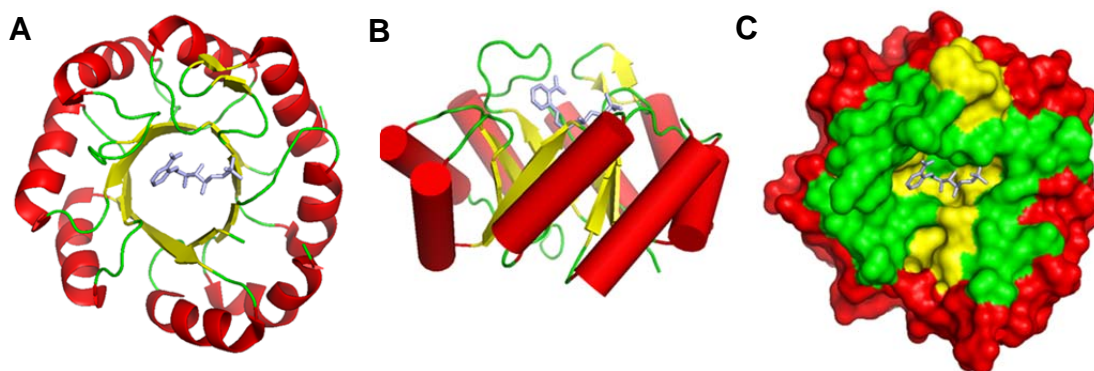
Figure 4-2. PRAI structure. Phosphoribosyl anthranilate isomerase (PRAI) of *Thermotoga Maritima* in complex with reduced product (rCdRP). α-helixes are shown in red, β sheets shown in yellow, and flexible loops shown in green. (PDB: 1LBM). (A) Top view. The active site is located at the central barrel and flexible loops, stabilized by surrounding helices. (B) Side view. The C-terminal face of the barrel β-sheets typically carries active site residues, while stability-determining residues are located at the N-terminal face. (C) Enzyme Surface. The substrate binding pocket is clearly visible.

4.3 Results

4.3.1 Mutagenesis of selected active site residues

To calibrate the tryptophan genetic selection, we generated a series of active site mutations of PRAI enzyme using site-directed mutagenesis, with a goal of achieving a range of catalytic activities. While the crystal structure of *S. cerevisiae* yPRAI is not yet available, its remarkably stable homolog from the hyperthermophile bacteria *T. maritima*,

encoded by the gene *t*TrpF, has been extensively studied,[18,22,24,36-38] shown to complement PRAI deficiency[18] and further investigated by directed evolution experiments.[39]

In order to construct mutants that vary in catalytic activity, we examined the active site for candidate residues for mutagenesis. The residues to be mutated were chosen both for their potential significance for substrate interaction at the active site, and the degree of their evolutionary conservation among TIM barrel enzymes. The degree to which an amino acid position is evolutionarily conserved is strongly dependent on its structural and functional importance. We used CONSURF,[40] a web-based tool to calculate a conservation score for each residue based on phylogenetic relations among sequence homologues (figure 4-3). Notably, selection of residues is also facilitated by the fact that in TIM barrel enzymes the stability of the fold is mostly maintained by the N-terminal face of the beta-strands, while the active site is located on the opposite face of the barrel.[41] Thus, varying residues in the active-site forming loops could alter catalytic function without affecting enzyme stability.[42]
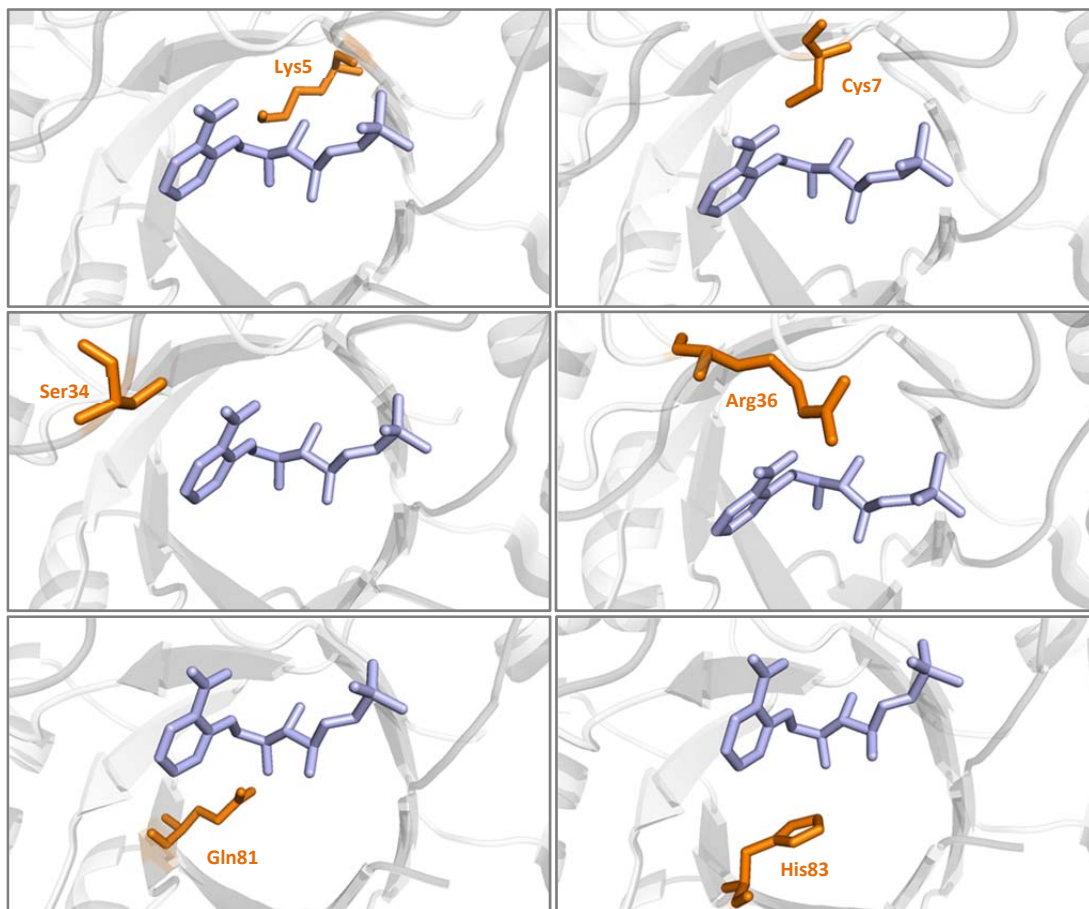
Six highly conserved PRAI active site residues were chosen for mutagenesis (Figure 4-4). Lysine at position 5 was replaced by another polar basic residue, arginine, which has a higher pKa, testing its effect on the acid-base reaction mechanism. Cysteine at position 7 was replaced with serine to test the effect of lower acidity on the Amadori rearrangement. The serine and arginine at positions 34 and 36, respectively, both form hydrogen bonds with the carboxyl groups of CdRP. Changing Serine at position 34, a primary alcohol, to a threonine, a secondary alcohol, may weaken hydrogen bond interaction. The same effect is expected for mutagenesis of Arg36 to lysine which has lower pKa. Similarly, glutamine at position 81 and histidine at position 83 both form

hydrogen bonds with substrate hydroxyl group. Replacing glutamine at position 81 with a charged histidine residue would most likely strengthen the hydrogen bond with the substrate, whereas replacing histidine at position 83 with uncharged asparagine would mostly likely lower the strength of hydrogen bonding. In addition, all residues but Cys7 were further mutagenized to alanine (as C7A mutation was previously shown to be inviable).[24] Mutagenesis was performed by either fusion PCR or site directed mutagenesis. Mutants were then purified for *in vitro* kinetic analysis and further tested for efficiency of complementation of *S. cerevisiae* PRAI functionality.
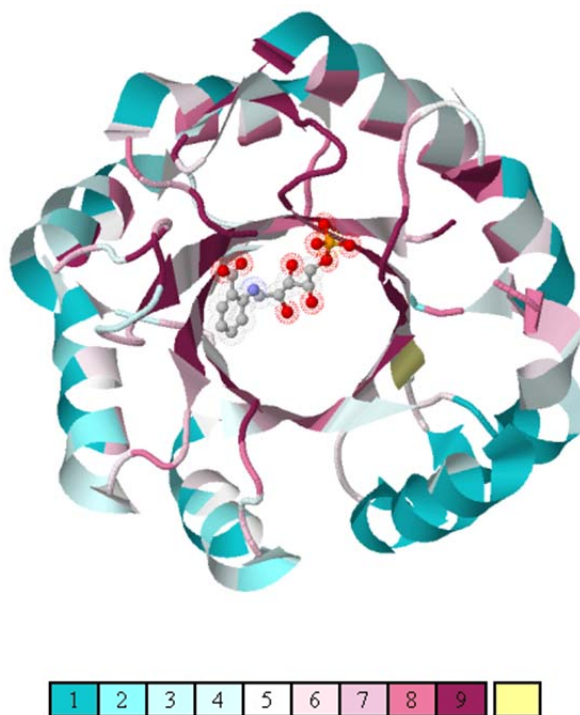
Figure 4-3. Position-specific conservation analysis for *T.Maritima* PRAI. Using CONSERF, Multiple sequence alignment of homologous structures is used to compute position-specific conservation score ranging 1 (not conserved, blue) to 9 (conserved, purple). Yellow color indicates insufficient data for analysis. The high degree of conservation is observed around rCdRP product, indicating active site residues. Peripheral α-helixes provide structural stability and displaying lower degree of evolutionary conservation. Highly conserved residues surrounding the active site were chosen for analysis. (PDB file: 1LBM)

Figure 4-4. PRAI residues chosen for mutagenesis. Structural and conservation analysis suggests residues to be evolutionary conserved and in close proximity to the active site. rCdRP product is shown in blue, candidate residues for mutagenesis are shown in orange.

**4.3.2 Purification and *in vitro* characterization of tTrpF variants**

All PRAI variants were overexpressed and purified in *E. coli* to conduct structural and kinetic analysis *in vitro*. To avoid contamination by endogenous *E. coli* tryptophan biosynthetic enzymes, we used an expression strain (KK8) which lacks the entire tryptophan biosynthetic pathway. Wild type *T. Maritima* TrpF and all variants were cloned into vector pVWC2323 and expression was induced using IPTG. To facilitate protein purification, cell extract was boiled at 80ºC to eliminate the vast majority of *E. coli* proteins, taking advantage of the extreme heat stability of T. Maritima proteins.[43]

Finally, the boiled cell extract was further purified by anion exchange column. All mutants were successfully expressed and purified from *E. coli* with the exception of R36A that could not be purified by anion exchange column (Figure 4-5). Notably, we did not use affinity purification for tTrpF variants to avoid any potential effects of a histidine tag on its structure or catalytic efficiency. However, the wt tTrpF was later successfully purified by nickel affinity using an N-terminal six histidine tag (data not shown).

4.3.3 Kinetic assay

A coupled enzyme assay[15] was used to measure PRAI activity for all PRAI variants (Figure 4-6). In this assay, the measured PRAI activity is coupled to the activity f the preceding and the following biosynthetic enzymes in the tryptophan biosynthetic pathway to provide the substrate for PRAI catalysis and eliminate product inhibition, respectively. The reaction is initiated by conversion of the substrates PRPP and anthranilate into PRA and diphsphate by the enzyme Anthranilate-phosphoribosyl transferase. Then, the conversion of PRA to CdRP is measured by adding PRAI to the reaction mix and monitoring the rate of consumption of PRA. Finally, the reaction also contains the third

enzyme Indoleglycerol-phosphate synthase to avoid accumulation of the PRAI product, CdRP. Excitation and emission wavelength of 310 nm and 400 nm were used, respectively. The relative fluorescent quantum yields of anthranilate, PRA, and CdRP at 400 nm are 1: 0.3: 0.06, respectively,[44] providing convenient assay for separate detection of each reaction step by following decay in substrate fluorescence.
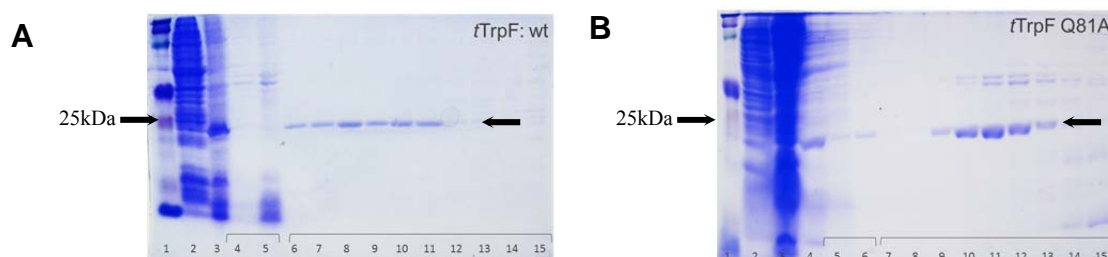
Figure 4-5. *t*PRAI purification. Following 5 hours Induction of PRAI expression in *E. coli*, cell extract were boiled to eliminate host proteins and the supernatant was further purified by anion exchange column. *t*PRAI monomer is expected at 23 kDa. (A) SDS PAGE analysis for purification of wild type *t*PRAI by anion exchange column. (1: protein ladder, 2: boiled extract, pellet. 3: boiled extract, supernatant. 4-5: flow through fractions. 6-15: protein elution fractions). (B) SDS PAGE for purification of variant Q81A purified by anion exchange column. (1: protein ladder. 2: cell exract. 3: boiled extract, pellet. 4: boiled extract, supernatant. 5-6: flow through fractions. 7-15: protein elution fractions).

Specifically, for the first reaction, we observe the decay in anthranilate fluorescence induced by the catalytic activity of Anthranilate-phosphoribosyl transferase (purified from *S. cerevisiae*, $K_{cat}/K_m = 18.1 \times 10^5 \ M^{-1}S^{-1}$). Thirty-fold excess of PRPP substrate is added to ensure anthranilate is exhausted from the reaction mixture. Next, either natural or mutant PRAI is added, resulting in further decay of fluorescence as PRA

is converted to CdRP. Finally, the CdRP product is converted to Indole-3-glycerol phosphate by the IGPS enzyme from *T. Maritima* ($K_{cat}/K_m = 3.8 \times 10^6$ M^{-1}s^{-1}).
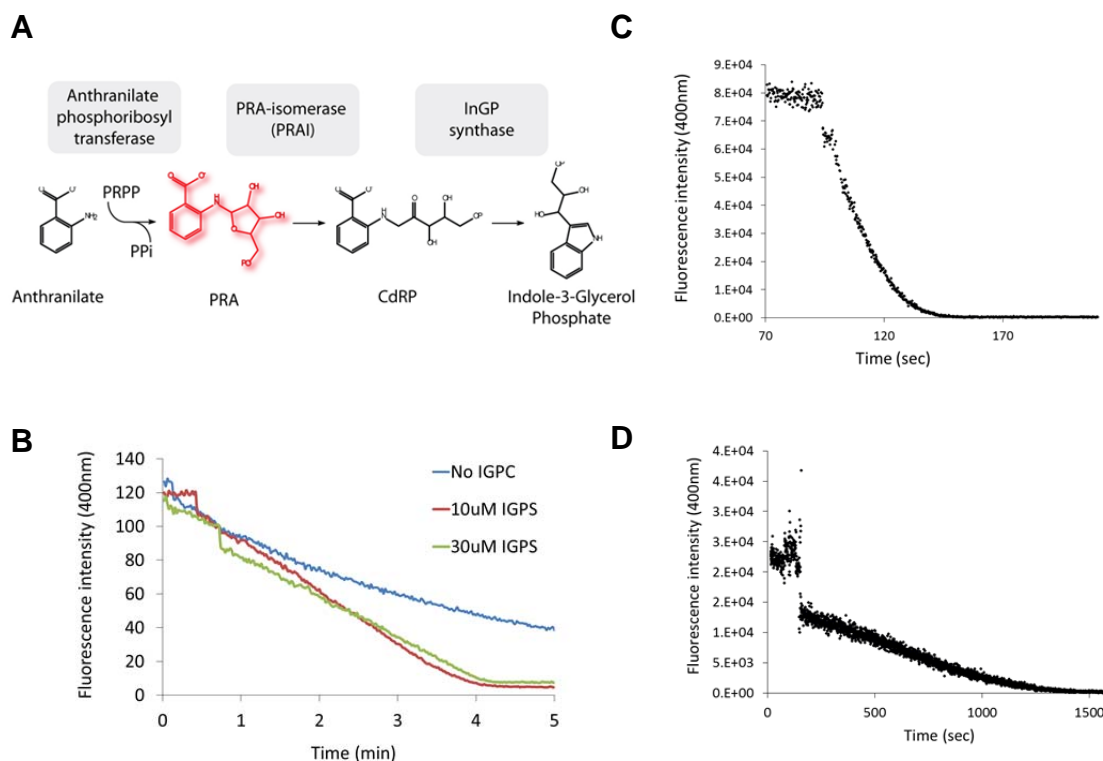
Figure 4-6. *In vitro* coupled kinetic assay for PRAI activity. All measurements are performed at room temperature by monitoring the decay in substrate fluorescence at 400 nm, using access of PRPP and IGPS. (A) Schematic of overall coupled kinetic assay. PRA, the substrate monitored to measure PRAI activity, is indicated in red. (B) Product inhibition of PRAI was observed by increasing the amount of enzyme IGPS (C-D) Decay of PRA substrate fluorescence as measured at 400 nm, shown for (C) variants Q81A (0.5 μM) and (D) K5A (5 μM).

Kinetic parameters were extracted from fluorescent decay data using 'progress curve analysis' method enabling kinetic parameters to be extracted from a single progress curve.[45,46] Briefly, the substrate concentration at each data point can be back calculated to

extract the enzyme Km that best fit to a defined first order decay model curve, assuming the enzyme concentration and initial substrate concentration are known. As previously reported,[15] in the absences of IGPS enzyme we observed product inhibition of PRAI activity (Figure 4-6). Therefore, all experiments were performed with excess of IGPS to eliminate product inhibition.

K_{cat}/K_m values were obtained for all variants (Figure 4-7) with the exception of variants K5R and C7S which displayed very slow kinetic activity that did not allow interpretation by progress curve analysis. Overall, we found a wide range of kinetic activity, with K_{cat}/K_m values spanning $8x10^2$ to $1.5x10^5$. Variant Q81H displayed exceptional curve shape with abrupt changes in initial velocity after reaction initiation and therefore could not be assigned kinetic parameters by progress curve analysis. We did not pursue further analysis of this variant. Importantly, the parameters measured for un-mutated tPRAI, $K_{cat}/K_m=5x10^6$ $M^{-1}s^{-1}$, was consistent with previously published data ($K_{cat}/K_m= 13.3x10^6 M^{-1}s^{-1}$),[43] validating the integrity of this assay.
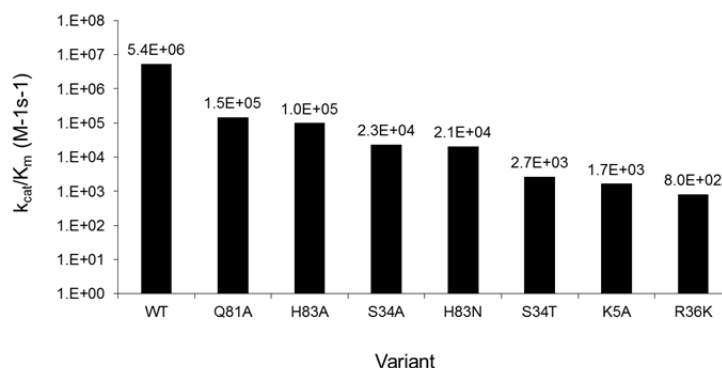

Figure 4-7. K_{cat}/K_m values for *t*PRAI variants. Kinetic parameters were extracted for each wt PRAI and variant using progress curve analysis.

The low catalytic efficiency observed for variants K5R and C7S is not surprising, as both residues are in high proximity to the active site and position 7 was previously suggested to take part in reaction mechanism.[24] In addition, variant K5A showed significantly reduced kinetic activity compared to wt PRAI. A large decrease in catalytic activity was also observed when changing arginine at position 36 to lysine. Variants carrying a mutant serine residue at position 34 show reduced catalytic activity when replaced by alanine (K_{cat}/K_m= 2.3x10^4 $M^{-1}s^{-1}$) and further decreases when replaced with threonine (K_{cat}/K_m= 2.7x10^3 $M^{-1}s^{-1}$). Replacing glycine at position 81 to alanine seem to have a marginal effect on enzyme activity, with K_{cat}/K_m= 1.5x10^5 $M^{-1}s^{-1}$. Histidine residue at position 83 also confers somewhat reduced catalytic activity when replaced by either alanine (K_{cat}/K_m= 1x10^5 $M^{-1}s^{-1}$) which is further reduced for mutagenesis with asparagine (K_{cat}/K_m= 2.1x10^4 $M^{-1}s^{-1)}$.

4.3.4 Circular Dichroism

Secondary structure of wt PRAI and variants was analyzed using circular dichroism spectroscopy (Figure 4-8). While wild type *t*PRAI displayed an expected secondary structure signal, with alpha helix peaks at 208nm and 220nm in addition to 215nm beta sheet signal, the variants all display deviation from wild type that could suggest loss of alpha helical structures. Interestingly, mutant C7S with significantly low catalytic activity displayed a different CD signal then the other variants tested, all with higher catalytic activity. The change in structure was not expected for mutagenesis of the selected residues, based on previous TIM barrel structure analysis.[47] The difference in CD signal could result from perturbation of fold or stability in PRAI variants, possibly altering their resistance to the purification process (involving boiling), or alternatively their stability in solution over

time. However, further characterization of enzyme structure is required to determine if the alteration of enzyme structure could explain variation observed in catalytic activity.
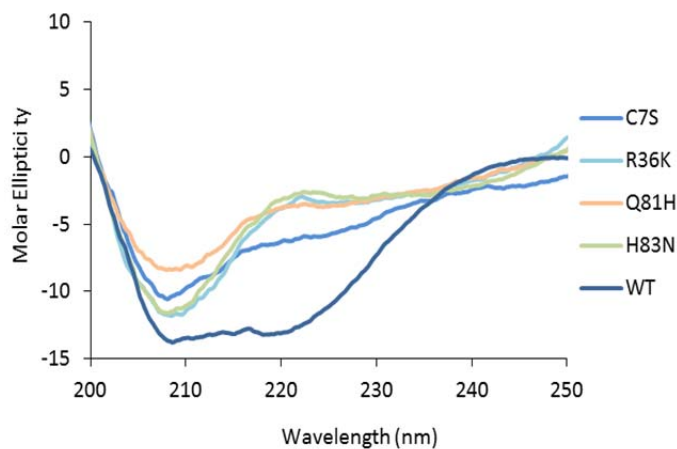
Figure 4-8. Far UV circular dichroism of wt *t*PRAI and selected variants. All measurements were done in room temperature, using 800 μM enzymes in phosphate buffer. Each graph represents an average of 3 scans.

4.3.5 *In vivo* **characterization of tTrpF variants**

In order to study the effect of tTrpF variants on yeast cell growth, variants were further tested for complementation of PRAI functionality. *S. cerevisiae* lacking the entire TRP1 gene was transformed with a plasmid pVWC2148 carrying wild type or variant T. Maritimia TrpF gene, and growth was monitored in the absence of tryptophan, both in solid and liquid media (Figure 4-9). Doubling time for each variant was extracted.

To test whether the complementation assay was affected by the concentration of variant PRAI protein in the cell, we also performed growth assays under lower level of

protein expression (Figure 4-9). Specifically, we used an inducible promoter to lower the level of transcription by adding methionine to the growth media. While all variants show slower growth rate under these conditions, the results were consistent with those observed under the standard high transcription level. we thus conclude protein concentration does not affect the differences in fitness observed between PRAI variants.
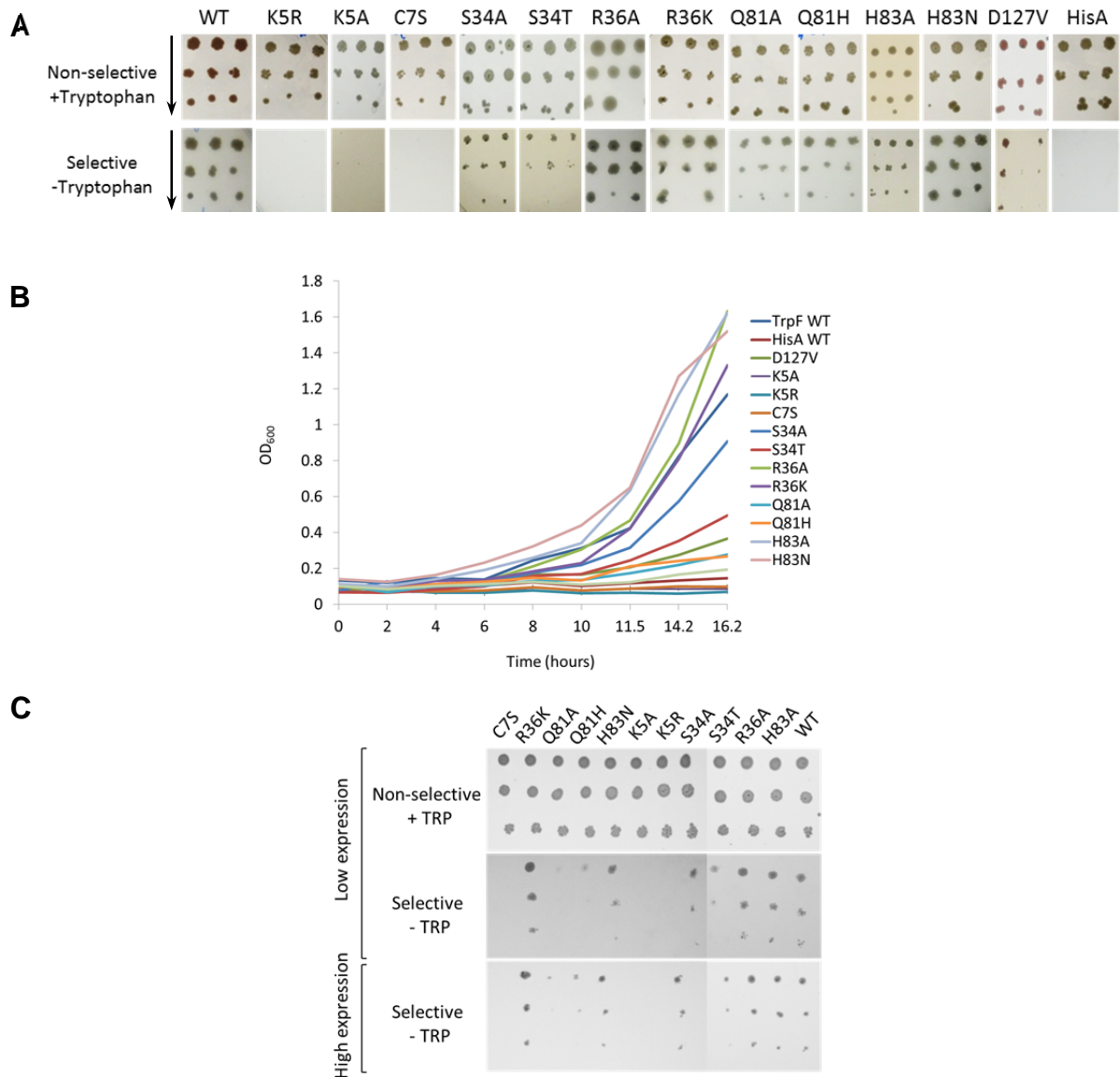
Figure 4-9. *In vivo* characterization of PRAI variants. Variants were expressed in S. cerevisiae lacking yTrp1 gene and cell growth was monitored in both solid and liquid selective media lacking tryptophan. (A) Plate growth assay. Colonies were plated on selective plates (arrow indicates increasing dilution) and colony growth was monitored after 3 days incubation at 30ºC. (B) Liquid growth assay in selective media. Cell density was monitored at OD_{600} at 30ºC shaker. (C) Plate growth assay under reduced expression levels. Variant transcription is regulated by a methionine-induced promoter, so that increasing methionine in the growth media reduces the level of protein expression. Upper panel dispays growth in non-selective (tryptophan rich) media. Middle and bottom panel display growth in the absence of tryptophan using high (middle panel) or low (bottom panel) expression of PRAI enzyme.

Finally, to further verify the difference in fitness between *t*PRAI variants under selective conditions, a competition assay was used where two variants are mixes at 1:100 cell ratios and incubated in liquid selective media lacking tryptophan. The faster growing variant is expected to take over the population as it undergoes more rounds of division. The resulting culture was analyzed by PCR restriction analysis of the entire pool of plasmid DNA using variant-specific restriction enzymes (Figure 4-10). wt *t*PRAI was observed to grow faster in selective media then both Q81H and R36K variants, although all demonstrated similar fitness by liquid growth analysis. We show variant R36K to grow faster than Q81H, and variant H83A to grow faster than variant Q81A. The competition assay is thus in agreement with initial characterization of variant fitness by growth assay.
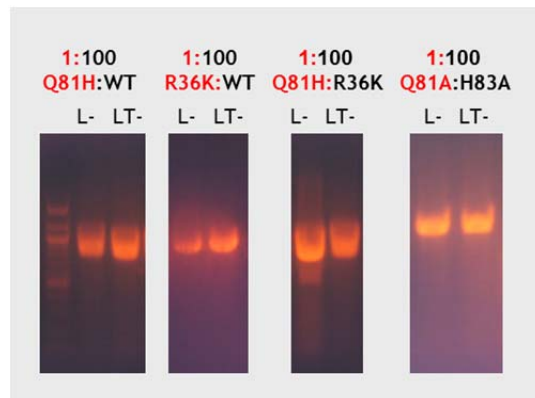
Figure 4-10. Competition growth assay between tPRAI variants. Variant fitness is tested by mixing yeast cultures bearing different variants in a 1:100 ratio. cultures are growth in selective SC(LT-) and non-selective SC(L-) media from OD_{600}=0.1 to OD_{600}=1. Plasmid DNA is extracted from the entire culture and analyzed by PCR and restriction analysis with an enzyme specific for the variant marked in red. the occurrence of same band both lanes indicates the DNA has not been digested, meaning the red variant is absence of the culture after selection.

Figure 4-11 summarized our finding for both *in vivo* and *in vitro* characterization of *t*PRAI variants. Overall, we did not find a correlation between PRAI catalytic activity and the respective cell growth rate, as indicated by doubling time and lag time (Figure4-11). The wide range of catalytic activities observed for the collection of variants was not reflected in cell growth differences under *in vivo* complementation assays. The lack of correlation could not be explained by the effect of protein concentration, since we tested concentration-dependence using methionine regulated promoter and observed the same differences in growth rate.

A

Variant	Kcat/Km ($M^{-1}s^{-1}$)	DT (H)
WT	5.37E+06	2.5
R36K	8.03E+02	2.5
S34A	2.32E+04	3.2
H83N	2.05E+04	3.4
S34T	2.68E+03	3.6
H83A	1.00E+05	3.8
Q81A	1.48E+05	4.1

B
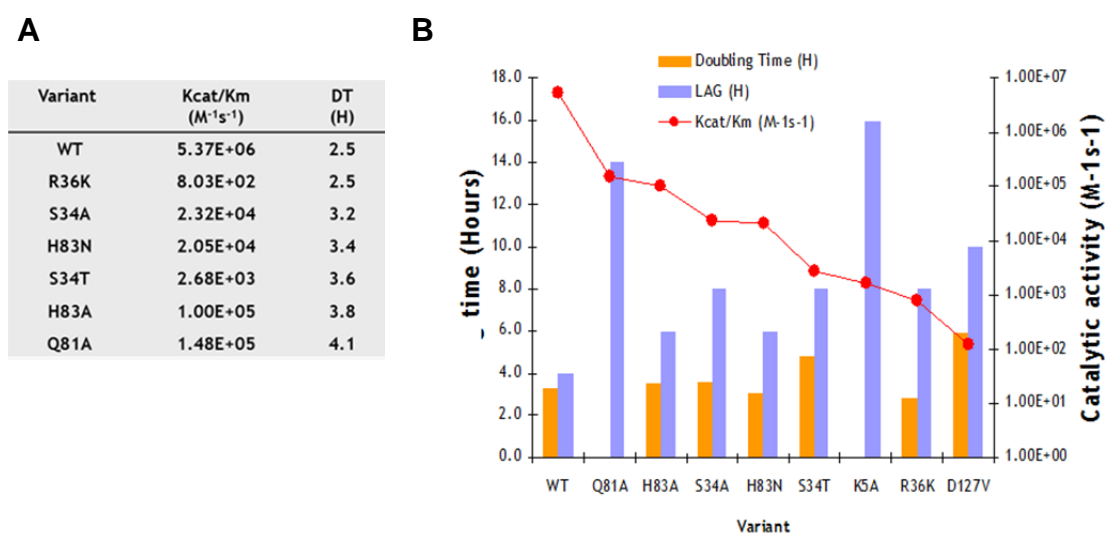
Figure 4-11. Comparing *in vivo* and *in vitro* properties of tPRAI variants. (A) Doubling time and kinetic parameters were extracted for each variant, as described above. (B) Graphic display of relationship between growth and catalytic functionality. K_{cat}/K_m is presented in red (right axis). Strain doubling time (DT) is presented in orange bars and lag time (time to reach OD_{600}=0.2) is presented in blue bars (left axis).

4.4   Discussion

Pioneering work by Jeremy Knowles,[6,48-50] Barry Hall[51] and others established genetic complementation as a powerful method for studying enzyme catalysis specifically in the context of directed enzyme evolution. Here, we examined the dynamic range of a widely used tryptophan selection marker in *S. cerevisiae* to evaluate its potential as a model selection system for development of directed evolution techniques. Specifically, mapping enzyme activity in terms of *in vivo* selection assays could provide us with a straightforward, calibrated assay for detection and isolation active mutants from a large library of variants simply by monitoring cell growth.

The extensively studied homologues PRAI enzyme of the thermophile bacteria *T. maritima* was chosen for its complementation of *S. cerevisiae* TRP1 deficiency, having been shown to be a highly stable protein and was previously used as scaffold for directed evolution. It is a highly efficient enzyme ($K_{cat}/K_m = 13.3 \times 10^6 \ M^{-1}s^{-1}$) and thus we expected active site mutagenesis to predominantly reduce its catalytic activity. We successfully generated a series of active site mutants and used them to test the correlation of enzyme activity with its complementation efficiency in a strain lacking endogenous PRAI enzyme. All variants were purified and characterized both *in vivo* and *in vitro*.

A coupled kinetic assay was used to extract kinetic parameters for wt and active variants. Significantly, kinetic parameters for tPRAI were in agreement with previously published results.[43] As expected, it was found that all variants have reduced catalytic activity compared to wild type tPRAI. Yet we were able to generate a wide range of catalytic activities by targeting residues surrounding the substrate binding site, ranging from 8×10^2 to 1.5×10^5 K_{cat}/K_m.   Three of the variants, namely K5R, K5A and C7S

demonstrated very slow kinetics and could not be analyzed by progress curve analysis. In fact, variant C7A has been previously found to be kinetically inactive and does not grow on selective media.[24] This is in agreement with the hypothesis that this is a catalytic residue acting as a base in the suggested acid-base catalysis mechanism.[24]

Mutagenesis of residues Q81 and H83 had the least effect on enzyme catalysis. For both residues the conservative exchange (Q81H, H83N) causes more significant decrease in enzyme activity then alanine. These results suggest that residues 81 and 83 have a minor effect on active site functionality. Notably, variant Q81H displays exceptional kinetics, with changing velocity as substrate concentration decreases. It can be hypothesized that the exchange of glutamine with histidine might strengthens the to the reaction product which results in stalling product release and possibly effecting enzyme turnover rate. This hypothesis has not been investigated further. Replacement of residues S34 and R36 results in significant decrease in enzyme activity, both for alanine and conservative amino acid exchange, suggesting a possible role in substrate binding.

We used Far UV circular dichroism analysis to evaluate the structural integrity of PRAI variants. Since all the mutations were selected within the active site of the enzyme, rather than on the stabilizing N-terminal face of the beta-barrel, we did not expect the selected resides to alter stability. Surprisingly, significant differences were observed in secondary structure for PRAI variants. We observed significant decreases in the characteristic alpha helix indicator peak at 220nm. Changes in folding efficiency, interruption of dimerization or increased susceptibility to destabilization upon boiling could explain such differences, as boiling is used in the purification process. The alteration of secondary structure might be a direct result of the mutagenized residue, in which case

we have probed the active site for stability-related residues. Nevertheless, the possible loss of structure did not completely eliminate enzyme activity, as observed by *in vivo* and *in vitro* assays.

Finally, we used an *in vivo* complementation assay to study the effect of enzyme catalysis on cell survival. Variants were expressed from high copy plasmids under a strong promoter in *S. cerevisiae* lacking the PRAI enzyme. We found good agreement between solid and liquid media results, with mutants at residue 36 and 83 displaying fast wild type-like growth, mutants of residue 34 and 81 displaying long lag time followed by rapid growth, and mutants of residues 5 and 7 showing very little growth under selective conditions. We further assessed variant fitness using competition assay, supporting relative fitness results as extracted from liquid and solid growth assay.

Cell growth phenotype in genetic complementation assay depends on the total amount of active enzyme, which is determined by both specific enzyme activity and enzyme concentration in the cell.[4] Thus, growth rate differences may provide information regarding changes in protein production, stability or catalytic activity, so that if a catalysts is active but in low amount, it might not support cell growth. We thus repeated the plate growth assay under lower expression of protein, so we can identify differences between strong variants by eliminating their overexpression. We found no change in relative growth between the variant, indicating enzyme concentration is no the determining factor for cell survival.

Overall, no correlation was observed between cell growth and the respective catalytic efficiency of PRAI enzyme. While kinetic parameters of wild type and variants span several orders of magnitude, no such differences were observed for cell growth. It is

possible that in order to draw meaningful conclusions, a larger number of variants must be tested. Nevertheless, these preliminary results raise questions regarding the correlation between enzyme mutagenesis and selection in directed evolution experiments using TRP1.

Early research in the field of metabolic flux previously observed *in vitro* enzyme activity to be a poor indicator for *in vivo* fitness,[52-54] noting the lack of linear correlation between gene dosage and phenotypic expression. Moreover, it was observed that inherent buffering of flux makes it relatively insensitive to small variations in activity. Hartl described the saturation curve for the dependence of fitness on flux, noting the "diminution of selective effect" in which the increment of fitness for a given increase in enzyme activity can be calculated by $[1/ (activity)^2]$. For example, if activity change is 100 fold, the change in fitness is only 10^{-4}. Thus, changes in single enzyme activity would be virtually too small to effect the overall fitness parameters.[55,56]

On the other hand, a handful of directed evolution experiments harness this and similar genetic selection as a basis for isolation of mutants with varying catalytic activity. For example, Hilvert and colleagues have used similar tryptophan genetic selection to evolve *S. cerevisiae* chorismate mutase, a key enzyme in the biosynthesis of aromatic amino acids phenylalanine and tyrosine.[4] In this case, enzyme mutagenesis widely varied cell growth rates, with a highly active variant significantly increasing cell growth rate.[57]

Although TRP1 is extensively used as selection marker in *S. cerevisiae* and further utilized to study molecular evolution of TIM barrel of the tryptophan and other pathways, systematic characterization of the dynamic range of this selection scheme as yet to be published.

We believe it is highly interesting to revisit this problem in the context of enzyme directed evolution using state of the art molecular biochemistry and systems biology tools. It is possible that in the case of highly valuable and highly energy-consuming metabolite such as tryptophan,[14,16] the complex regulation of the pathway is masking changes in single enzyme activity. Thus, the coupling between pathway enzymes via shared substrates/products buffers the effects of changing single catalytic function.

Further experiments using lower PRAI enzyme catalysis then the one achieved in this work should be tested to probe the range of the TRP1 genetic selection. Similar studies using the less active S. cerevisiae PRAI enzyme, rather than T. maritime PRAI, might be used for this purpose to eliminate differences in enzyme stability or oligomerization that might have affected this study.

Overall, our work highlights the significance of rigorously calibrating each selection scheme used for directed evolution experiments. Metabolic enzymes are widely utilized as markers for genetic selection in yeast for basic molecular biology as well as directed evolution applications. However, while selectable markers are a powerful tool for identification of functional enzymes, they might not provide the sensitivity and dynamic range required to efficiently detect a range of catalytic activities in the context of enzyme directed evolution.

The lack of selection schemes for desired natural products and metabolites is a recognized bottleneck in the field of enzyme and metabolic engineering. As the complexity of molecules produced by microbial hosts in increasing, new methods are required to allow generic selection for metabolites regardless of their cellular function.

4.5   Experimental methods

General materials and methods. General materials and methods were as in **Chapter 2**.

Mutagenesis of *T. Maritima* TrpF. Fusion PCR was used to make point mutations in the tTrpF gene using the primers specified in Table 4-3. Two fragments of TrpF were constructed using separate PCR reactions with mutation-containing primers. Then, the two fragments were fused using the two outmost primers of either PCR to produce the full length TrpF containing a point mutation. The template plasmid for fusion PCR was plasmid pVWC2148 containing wild type TrpF. Each point mutation also inserted a new restriction site allowing the verification of mutagenesis. The mutated fragments were subcloned back into plasmid pCWV2148, replacing the wt gene, and additionally subcloned into plasmid pVC2323 for *E. coli*expression.

Purification of T. Maritima tPRAI and mutants. Plasmid pVWC2323 bearing wild type or mutant tTrpF was transformed into *E. coli*KK8 cells purified using the following protocol. [43]   Transformed cells were grown overnight at 37ºC in LB media supplemented with 0.1 mg/ml ampicillin and 0.025 mg/ml kanamycin. This culture was used to inoculate 4 L of the same medium. Shake at 37ºC until OD_{600} reached 0.5 - 0.6. Protein overexpression was induced by adding IPTG to final concentration of 1 mM and allowed to grow for 5 hours. Cells were harvested at 6000 rpm for 30 minutes, washed with 50 mM potassium phosphate buffer pH 7.8 added with 300 mM NaCl and harvested again, resulting in ~2.5 g/L of cells (wet weight). Cell were resuspend at in 100 mM potassium phosphate pH 7.8, 2 mM EDTA, 1 mM DTT, 0.3 mM PMSF and broken by 2 cycles of french press and centrifuged at 12000 rpm for 60 minutes at 4ºC. TrpF was purified from the soluble fraction only. Supernatant was incubated at 80 ºC for 10 minutes

in 1.5 ml eppendorf vials.to remove *E. coli*host proteins and centrifuged again at 12,000 rpm for 20min at 4 ℃. Pellet was discarded, and supernatant was ialyze against 10 mM potassium phosphate pH 7.5, 2 mM EDTA, 0.4 mM DTT. A 5 ml DEAE column was equilibrated with equilibration buffer (10 mM potassium phosphate pH 7.5, 2 mM EDTA, 0.4 mM DTT), and the supernatant was loaded, wash with 5 volumes of equilibration buffer, and eluted using a linear gradient of 20 to 180 mM potassium phosphate buffer. WT tTrpF elutes approximately at 70 mM potassium phosphate (as measured by SDS-PAGE). Fractions containing TrpF were analyzed on SDS-PAGE, pooled and concentrated using Amicon centriprep (10,000 molecular cutoff).

Growth analysis in *S. cerevisiae*. Strain VWC2273 was transformed with vector pVWC2148 carrying wt or mutant tTrpF. For liquid growth assay, an overnight culture was started from a fresh patch in 200 μl SC (L⁻) media. The next day, the culture was diluted to $OD_{600} = 0.1$ and in liquid SC(LT⁻) media. OD_{600} was monitored every 2-3 hours. For solid growth assay, the same overnight culture was further used to plate all variants on SC (LT-) and SC(L-) solid plates, using 2 μl of cultures diluted to $OD_{600} = 0.1$ or 0.01 or 0.001. All measurements were done using three colonies of each variant, in triplicates. The OD was plotted in excel, the slope of the exponential growth phase (between OD=0.3 and OD=0.6) was calculated, and the doubling time was extracted using the formula [Ln2/ (exponential growth slope)].

Growth analysis under lower expression levels. All variants expressed under the Met15 inducible promoter were down regulated by addition of 134 μM methionine to the growth media. colony growth on plates was monitored after 3 days incubation in 30℃.

Competition selection assay. Cultures of two candidate variants for competition were mixed at 1:100 ratio ratios, as measured by OD_{600}. cells were mixed in 1ml volume using selective SC(LT-) or noon selective SC(L-) media. All experiments started at OD_{600}=0.1 and grown in 30ºC shaker until OD_{600}=0.1. Then, the competition culture was used to inoculate miniprep cultures in non-selective media and the DNA extracted and subjected to PCR reaction with primers VWC1051 and VWC1052 to amplify the TrpF gene. PCR samples are then digested using specific restriction enzyme for one variant.

Kinetic analysis. The measurement was performed at room temperature at 100 μl final volume, and repeated three times. The assay is performed in Tris buffer (50mM Tris 4mM EDTA 4mM MgCl, 2mM DTT). Calibration curve was constructed for PRA and anthranilate prior to measurement, and used as reference for analysis (the slope of the curve is used as extinction coefficient for progress curve analysis). Cuvette is highly recommended over 96 well plates. All pipettes were calibrated prior to use, and reaction was stirred after addition of enzyme. The following amount of enzymes and reagents were used in each reaction mix for total volume 1 ml: 100 μM Anthranilate, 3 mM PRPP, 100 μM Anthranilate phosphoribosyl transferase enzyme (TrpD), 10 μM IGPS enzyme (TrpC). $MgCl_2$, EDTA and DTT were added to final concentrations 1.7 mM, 1.7 mM, 850 μM, respectively. PRAI concentration ranged 0.5 – 10 μM for wt and variants.

Circular dichroism. Far UV CD spectra was measured at 800 μM wt or variant in 10mM KH_2PO_4 buffer, All samples were kept at 4 ºC and filtered in .45 μm filter prior to scan. Far UV signal was measured using 0.1 cm cuvette, spanning 190 nm -250 nm, every 1 nm, 100 nm/min, 1 second response tieme and 1 μm bandwidth, averaging 3 scans.

4.6 Strains, plasmids, and oligonucleotides

Table 4-1 Strains used in this study

Strain	Genes
NP2273	ATCC4017202 - MATa his3 Δ1 leu2Δ0 met15Δ0 ura3Δ0 ΔTRP1 (*S. cerevisiae*)

Table 4-2. Plasmids used in this study

Mutation	*S. cerevisiae* (pVWC2148)	*E. coli* -TG1 (pVWC2148)	*E. coli* -KK8 (pVWC2323)
Wild type tTrpF		VWC2258	VWC2422
*t*TrpF-K5A	VWC2397	VWC2365	VWC2375
*t*TrpF-K5R	VWC2398	VWC2361	VWC2376
*t*TrpF-C7S	VWC2370	VWC2349	VWC2377
*t*TrpF-S34A	VWC2399	VWC2362	VWC2378
*t*TrpF-S34T	VWC2400	VWC2360	VWC2379
*t*TrpF-R36A	VWC2401	VWC2359	VWC2405
*t*TrpF-R36K	VWC2371	VWC2366	VWC2380
*t*TrpF-Q81A	VWC2372	VWC2367	VWC2381
*t*TrpF-Q81H	VWC2373	VWC2368	VWC2382
*t*TrpF-H83A	VWC2402	VWC2350	VWC2383
*t*TrpF-H83N	VWC2374	VWC2369	VWC2422

Table 4-3 Oligonucleotides used in this study

Name	Mutation		Sequence	
MC111	K5A	AAA to Gcc		5`GCAGGGCCGGCAGGGCCATGGTCAGAGTG**GCC**ATCTGCGGGATAACGAACC
MC112	K5R	AAA to CGG		5`GCAGGGCCGGCAGGGCCATGGTCAGAGTG**CGG**ATCTGCGGGATAACGAACC
MC113	C7S	TGC to TCA		5`GCAGGGCCGGCAGGGCCATGGTCAGAGTG**CGG**ATCTGCGGGATAACGAACC
MC114	S34A	AGC to GCT	StyI	5` TTCTATC<u>CCAAG</u>gctAAAAGGTACATCTCTCCTGA
MC115	S34A			3` TACCTTTTag<u>cCTTGG</u>GATAGAAAACGAATCCCA
MC116	S34T	AGC to ACT		5` TTCTATCCTAAGactAAAAGGTACATCTCTCCTGA
MC117	S34T			3` GATGTACCTTTTagtCTTAGGATAGAAAACGAATC
MC118	R36A	AGG to GCT	EarI	5` CC<u>GAAGAG</u>CAAAgctTACATCTCTCCTGAGGATG
MC119	R36A			3` TAacgTTTG<u>CTCTTC</u>GGATAGAAAACGAATCCCA
MC103	R36K	AGG to AAA	EarI	5` CC<u>GAAGAG</u>CAAAaaaTACATCTCTCCTGAGGATG
MC104	R36K			3` TAtttTTTG<u>CTCTTC</u>GGATAGAAAACGAATCCCA
MC120	Q81A	CAG to GCT	BsmI	5` CT<u>GAATGCG</u>GTCgctCTACATGGAGAAGAACCGA
MC121	Q81A			3` TGTAGagcGAC<u>CGCATTC</u>AGCTGAACATAAGAGG
MC105	Q81H	CAG to CAT	BsmI	5` CT<u>GAATGCG</u>GTCcatCTACATGGAGAAGAACCGA
MC106	Q81H			3` TGTAGatgGAC<u>CGCATTC</u>AGCTGAACATAAGAGG
MC122	H83A	CAT to GCT	TseI	5` GCGGT<u>GCAGC</u>TAgctGGAGAAGAACCGATCGAA
MC123	H83A			3` TCTCCagcTAGCTGCACCGCGTTCAGCTGAA
MC107	H83N	CAT to AAT	TseI	5` GCGGT<u>GCAGC</u>TAaatGGAGAAGAACCGATCGAA
MC108	H83N			3` TCTCCattTA<u>GCTGC</u>ACCGCGTTCAGCTGAA

4.7 References

1   Arnold, F. H. & Volkov, A. A. Directed evolution of biocatalysts. *Curr Opin Chem Biol* **3**, 54-59, (1999).

2   Beadle, G. W. & Tatum, E. L. Genetic Control of Biochemical Reactions in Neurospora. *Proc. Natl. Acad. Sci. U. S. A.* **27**, 499-506, (1941).

3   Taylor, S. V., Kast, P. & Hilvert, D. Investigating and Engineering Enzymes by Genetic Selection. *Angew Chem Int Ed Engl* **40**, 3310-3335, (2001).

4   Taylor, S. V., Kast, P. & Hilvert, D. Investigating and Engineering Enzymes by Genetic Selection. *Angew. Chem. Int. Ed. Engl.* **40**, 3310-3335, (2001).

5   Aharoni, A., Griffiths, A. D. & Tawfik, D. S. High-throughput screens and selections of enzyme-encoding genes. *Curr. Opin. Chem. Biol.* **9**, 210-216, (2005).

6   Hermes, J. D., Blacklow, S. C. & Knowles, J. R. Searching sequence space by definably random mutagenesis: improving the catalytic potency of an enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 696-700, (1990).

7   Rowe, L. A., Geddie, M. L., Alexander, O. B. & Matsumura, I. A comparison of directed evolution approaches using the beta-glucuronidase model system. *J. Mol. Biol.* **332**, 851-860, (2003).

8   Hall, A. & Knowles, J. R. Directed selective pressure on a beta-lactamase to analyse molecular changes involved in development of enzyme function. *Nature* **264**, 803-804, (1976).

9   Munir, K. M., French, D. C. & Loeb, L. A. Thymidine kinase mutants obtained by random sequence selection. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 4012-4016, (1993).

10   Smiley, J. A. & Benkovic, S. J. Selection of catalytic antibodies for a biosynthetic reaction from a combinatorial cDNA library by complementation of an auxotrophic Escherichia coli: antibodies for orotate decarboxylation. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 8319-8323, (1994).

11   Amberg, D. C., Burke, D., Strathern, J. N. & Cold Spring Harbor Laboratory. *Methods in yeast genetics : a Cold Spring Harbor Laboratory course manual*. 2005 edn, (Cold Spring Harbor Laboratory Press, 2005).

12   Fastrez, J. In vivo versus in vitro screening or selection for catalytic activity in enzymes and abzymes. *Mol. Biotechnol.* **7**, 37-55, (1997).

13   Neuenschwander, M., Butz, M., Heintz, C., Kast, P. & Hilvert, D. A simple selection strategy for evolving highly efficient enzymes. *Nat Biotechnol* **25**, 1145-1147, (2007).

14   Miozzari, G., Niederberger, P. & Hutter, R. Tryptophan biosynthesis in Saccharomyces cerevisiae: control of the flux through the pathway. *J. Bacteriol.* **134**, 48-59, (1978).

15   Hommel, U., Eberhard, M. & Kirschner, K. Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction. *Biochemistry* **34**, 5429-5439, (1995).

16   Braus, G. H. *et al.* The role of the TRP1 gene in yeast tryptophan biosynthesis. *J. Biol. Chem.* **263**, 7868-7875, (1988).

17   Gonzalez, A., Larroy, C., Biosca, J. A. & Arino, J. Use of the TRP1 auxotrophic marker for gene disruption and phenotypic analysis in yeast: a note of warning. *FEMS Yeast Res* **8**, 2-5, (2008).

18   Sterner, R. *et al.* Phosphoribosyl anthranilate isomerase from Thermotoga maritima is an extremely stable and active homodimer. *Protein Sci* **5**, 2000-2008, (1996).

19   Copley, R. R. & Bork, P. Homology among (betaalpha)(8) barrels: implications for the evolution of metabolic pathways. *J. Mol. Biol.* **303**, 627-641, (2000).

20   Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. & Wilmanns, M. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science* **289**, 1546-1550, (2000).

21   Hocker, B. Directed evolution of (betaalpha)(8)-barrel enzymes. *Biomol Eng* **22**, 31-38, (2005).

22   Sterner, R. *et al.* (Beta alpha)8-barrel proteins of tryptophan biosynthesis in the hyperthermophile Thermotoga maritima. *EMBO J.* **14**, 4395-4402, (1995).

23   Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. & Jansonius, J. N. Structural conservation in parallel beta/alpha-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry* **30**, 9161-9169, (1991).

24   Henn-Sax, M. *et al.* Two (betaalpha)(8)-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry* **41**, 12032-12042, (2002).

25   Barona-Gomez, F. & Hodgson, D. A. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep* **4**, 296-300, (2003).

26   Wright, H. *et al.* The structure/function relationship of a dual-substrate (betaalpha)8-isomerase. *Biochem. Biophys. Res. Commun.* **365**, 16-21, (2008).

27   Kuper, J., Doenges, C. & Wilmanns, M. Two-fold repeated (betaalpha)4 half-barrels may provide a molecular tool for dual substrate specificity. *EMBO Rep* **6**, 134-139, (2005).

28   Leopoldseder, S., Claren, J., Jurgens, C. & Sterner, R. Interconverting the catalytic activities of (betaalpha)(8)-barrel enzymes from different metabolic pathways: sequence requirements and molecular analysis. *J. Mol. Biol.* **337**, 871-879, (2004).

29   Saab-Rincon, G., Mancera, E., Montero-Moran, G., Sanchez, F. & Soberon, X. Generation of variability by in vivo recombination of halves of a (beta/alpha)8 barrel protein. *Biomol Eng* **22**, 113-120, (2005).

30   Ochoa-Leyva, A. *et al.* Protein design through systematic catalytic loop exchange in the (beta/alpha)8 fold. *J. Mol. Biol.* **387**, 949-964, (2009).

31   Claren, J., Malisi, C., Hocker, B. & Sterner, R. Establishing wild-type levels of catalytic activity on natural and artificial (beta alpha)8-barrel protein scaffolds. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3704-3709, (2009).

32   Patrick, W. M. & Matsumura, I. A study in molecular contingency: glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J. Mol. Biol.* **377**, 323-336, (2008).

33   Hocker, B., Claren, J. & Sterner, R. Mimicking enzyme evolution by generating new (betaalpha)8-barrels from (betaalpha)4-half-barrels. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16448-16453, (2004).

34   Jurgens, C. *et al.* Directed evolution of a (beta alpha)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9925-9930, (2000).

35   Tafelmeyer, P., Johnsson, N. & Johnsson, K. Transforming a (beta/alpha)8--barrel enzyme into a split-protein sensor through directed evolution. *Chem. Biol.* **11**, 681-689, (2004).

36   Sterner, R. *et al.* (Beta alpha)8-barrel proteins of tryptophan biosynthesis in the hyperthermophile Thermotoga maritima. *Embo J* **14**, 4395-4402, (1995).

37   Hennig, M., Sterner, R., Kirschner, K. & Jansonius, J. N. Crystal structure at 2.0 A resolution of phosphoribosyl anthranilate isomerase from the hyperthermophile Thermotoga maritima: possible determinants of protein stability. *Biochemistry* **36**, 6009-6016, (1997).

38   Thoma, R., Hennig, M., Sterner, R. & Kirschner, K. Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from Thermotoga maritima. *Structure* **8**, 265-276, (2000).

39   Jurgens, C. *et al.* Directed evolution of a (beta alpha)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci U S A* **97**, 9925-9930, (2000).

40   Glaser, F. *et al.* ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**, 163-164, (2003).

41   Urfer, R. & Kirschner, K. The importance of surface loops for stabilizing an eightfold beta alpha barrel protein. *Protein Sci.* **1**, 31-45, (1992).

42   Patrick, W. M. & Blackburn, J. M. In vitro selection and characterization of a stable subdomain of phosphoribosylanthranilate isomerase. *FEBS J.* **272**, 3684-3697, (2005).

43   Sterner, R. *et al.* Phosphoribosyl anthranilate isomerase from Thermotoga maritima is an extremely stable and active homodimer. *Protein Sci.* **5**, 2000-2008, (1996).

44   Kirschner, K., Szadkowski, H., Jardetzky, T., Hager, V. Phosphoribosylanthranilate Isomerase-Indoleglycerol-Phosphate Synthase from Escherichia coli. *Methods Enzymol.* **142**, (1987).

45   Lefurgy, S. T. *et al.* Probing ligand-binding pockets of the mevalonate pathway enzymes from Streptococcus pneumoniae. *J. Biol. Chem.* **285**, 20654-20663, (2010).

46   Tang, Q. & Leyh, T. S. Precise, facile initial rate measurements. *J Phys Chem B* **114**, 16131-16136, (2010).

47   Henn-Sax, M., Hocker, B., Wilmanns, M. & Sterner, R. Divergent evolution of (betaalpha)8-barrel enzymes. *Biol. Chem.* **382**, 1315-1320, (2001).

48   Knowles, J. A. & Albery, W. J. Perfection in Enzyme Catalysis: The Energetics of Triosephosphate Isomerase. *Accounts of Chemical Research* **10**, 105-111, (1977).

49   Albery, W. J. & Knowles, J. R. Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* **15**, 5631-5640, (1976).

50   Burbaum, J. J., Raines, R. T., Albery, W. J. & Knowles, J. R. Evolutionary optimization of the catalytic effectiveness of an enzyme. *Biochemistry* **28**, 9293-9305, (1989).

51   Hall, B. G. & Clarke, N. D. Regulation of newly evolved enzymes. III Evolution of the ebg repressor during selection for enhanced lactase activity. *Genetics* **85**, 193-201, (1977).

52   Hartl, D. L., Dykhuizen, D. E. & Dean, A. M. Limits of adaptation: the evolution of selective neutrality. *Genetics* **111**, 655-674, (1985).

53   Middleton, R. J. & Kacser, H. Enzyme variation, metabolic flux and fitness: alcohol dehydrogenase in Drosophila melanogaster. *Genetics* **105**, 633-650, (1983).

54   Kacser, H. & Burns, J. A. The molecular basis of dominance. *Genetics* **97**, 639-666, (1981).

55   Hartl, D. L., Dykhuizen, D. E. & Dean, A. M. Limits of adaptation: the evolution of selective neutrality. *Genetics* **111**, 655-674, (1985).

56   Dean, A. M., Dykhuizen, D. E. & Hartl, D. L. Fitness as a function of beta-galactosidase activity in Escherichia coli. *Genet. Res.* **48**, 1-8, (1986).

57   Vamvaca, K., Butz, M., Walter, K. U., Taylor, S. V. & Hilvert, D. Simultaneous optimization of enzyme activity and quaternary structure by directed evolution. *Protein Sci.* **14**, 2103-2114, (2005).