Combining Rational and Evolutionary Approaches to Optimize Enzyme Activity in *Saccharomyces cerevisiae*

Thesis by
Joshua Kieran Michener

In Partial Fulfillment of the Requirements for the Degree
of
Doctor of Philosophy

California Institute of Technology
Pasadena, California
2012
(Defended May 7, 2012)

Acknowledgements

I was very fortunate during my graduate career to have been given the freedom to wend my way down a number of different paths, and I am grateful to my advisors for their support along the way. Professor Christina Smolke has been a constant source of advice and encouragement, and I am inspired by her dedication and insight. I always knew that I could count on Professor Frances Arnold to give me her unvarnished opinion, and I credit her for training me to critically analyze my research. Finally, I enjoyed working with Professor Jens Nielsen and am grateful that he gave me the opportunity to learn from him and his lab.

I would also like to thank my committee members, Professors Richard Murray and Jared Leadbetter. I have enjoyed my opportunities to talk with and learn from Richard, and a comment of Jared's eventually led to my time in the Nielsen lab.

During my graduate career, I have had the opportunity to work with a number of amazingly talented people who have greatly influenced me. Specifically, I would like to thank the entire Smolke Laboratory Metabolic Engineering subgroup, for shared commiseration and troubleshooting; Joe Liang, Katie Galloway, and Dr. Chase Beisel, for an entertaining and stimulating laboratory environment; Andrew Sawayama and Mike Chen, for an introduction to directed evolution; and the Caltech Biocontrols group, particularly Fiona Chandra and Mary Dunlop, for consistently insightful and enjoyable discussions.

Finally, I would like to thank Lloyd, Gwen, Becca, and Josie for their love and support.

Abstract

Metabolic engineering has become an increasingly important tool for the production of bulk and fine chemicals. New biosynthetic pathways can be built in a tractable production host using enzymes from a wide variety of organisms. However, these enzymes did not evolve to function in their new host, and as a result their activity may be unacceptably low. Additionally, the host has not adapted to support this new pathway, and its response to any new stresses imposed by the pathway may further limit productivity. I describe two methods for optimizing the host-enzyme interface, using an evolutionary approach to adapt an enzyme to its new host and a rational approach to modify the host in response. Using a synthetic RNA switch to screen for improvements in enzymatic activity *in vivo*, I increased the activity of a model enzyme more than 30-fold. I then used a systems-level analysis of the host to identify a stress, heme depletion, that the enzyme placed on its host. Alleviating that stress increased the activity of an optimized enzyme by a further 2.3-fold. These results highlight the advantages of combining systems and synthetic biology during the construction of a metabolic pathway. I also consider options for extending the uses of synthetic RNA switches both earlier and later in the pathway development process. An RNA switch could first be used in a functional screen for enzyme discovery and then be used to adapt the newly discovered enzyme to its production host. Finally, a variant of that switch could be used to dynamically regulate a biosynthetic pathway and improve the pathway reliability.

Table of Contents

Acknowledgements iii

Abstract iv

Table of Contents v

List of Figures and Tables vii

1 Introduction 1

1.1 Open challenges in metabolic engineering 1

1.1.1 Challenges in the predictable design of biological systems 2

1.1.2 Challenges in the application of evolutionary methods 5

1.1.3 Challenges in the discovery of uncharacterized enzymes 6

1.2 Tools from systems and synthetic biology 7

1.2.1 Characterization of biological systems 7

1.2.2 Managing biological complexity 9

2.2.3 Small-molecule screens and selections 11

2.2.4 New techniques for enzyme discovery 13

1.3 Thesis overview 14

2 High-Throughput Enzyme Evolution Using a Synthetic RNA Switch 17

2.1 Introduction 17

2.2 Results

2.2.1 Development of the screening system 22

2.2.2 Identification of a starting enzyme 24

2.2.3 Plate-based enzyme screening 26

2.2.4 Screening enzymes in single cells 29

2.2.5 Characterization of improved mutants 33

2.2.6 Comparison to *in vitro* evolution 36

2.2.7 Increasing thermostability also increases mutational tolerance 37

2.3 Discussion 39

2.4 Methods

2.4.1 General molecular biology techniques 43

2.4.2 *In vitro* enzyme screening 44

2.4.3 Construction of the *S. cerevisiae* enzyme expression vectors 44

2.4.4 Product biosensor yeast strain construction 47

2.4.5 Random mutagenesis 47

2.4.6 Library shuffling 48

2.4.7 Growth conditions for liquid culture *S. cerevisiae* assays 49

2.4.8 Flow-cytometry-based library screening 49

2.4.9 FACS-based library screening 50

2.4.10 HPLC methods 51

2.4.11 Thermostability assays 52

2.4.12 Determination of apparent kinetic constants 52

2.4.13 Western blots 53

2.4.14 Enzyme stabilization 53

2.5 Tables 54

2.6 Acknowledgements 60

3 Use of a Synthetic RNA Switch for Enzyme Discovery 61

3.1 Introduction 61

3.2 Results

3.2.1 cDNA library construction 64

3.2.2 *In vivo* cDNA screening 65

3.3 Discussion 68

3.4 Methods

3.4.1 cDNA synthesis 72

3.4.2 Subtractive hybridization 73

3.4.3 Growth conditions for liquid culture *S. cerevisiae* assays 74

3.4.4 Flow-cytometry-based library screening 74

3.4.5 FACS-based library screening 74

3.4.6 HPLC methods 75

3.5 Acknowledgements 76

4 Identifying and Alleviating Stress from Monooxygenase Overexpression 77

4.1 Introduction 78

4.2 Results

4.2.1 Improvements at low copy do not translate to high copy 82
4.2.2 Microarrays identify heme depletion as the major cellular stress 83
4.2.3 Increased heme biosynthesis raises the intracellular heme level 84
4.2.4 Heme limits the activity of cytosolic hemoproteins 86
4.2.5 Heme depletion limits total enzyme expression 90
4.3 Discussion 91
4.4 Methods
4.4.1 Strains and cultivation 95
4.4.2 Metabolite analysis 95
4.4.3 DNA microarray experiments 96
4.4.4 Proteasomal activity measurements 97
4.4.5 Heme measurements 97
4.4.6 Quantitative Western blots 98
4.4.7 Catalase activity assays 99
4.4.8 qRT-PCR measurements 100
4.5 Tables 102
4.5 Acknowledgements 102
5 Design of *In Vivo* Feedback Controllers Using Synthetic RNA Switches 103
5.1 Introduction 103
5.2 Results
5.2.1 Minimizing disruptions from input perturbations 108
5.2.2 Minimizing retroactivity in a biosynthetic pathway 110
5.2.3 Experimental characterization 113
5.3 Discussion 117
5.4 Methods
5.4.1 Variable enzyme expression 120
5.4.2 RNA switch characterization 120
5.4.3 Intracellular metabolite assays 121
5.5 Modeling
5.5.1 Modeling an input disturbance 122
5.5.2 Modeling a variable load 123
5.5.3 Modeling a branch point 124

5.6 Acknowledgements ... 125

6 Conclusions ... 127

6.1 Applications of RNA switches for metabolic engineering ... 127

6.2 Constructing new RNA switches for metabolic engineering ... 128

6.3 Analysis of heterologous pathways ... 129

6.4 Tools and techniques to accommodate metabolic pathways ... 131

7 References ... 133

List of Figures and Tables

Figure **Page**

Figure 1.1 Feedback regulation allows a pathway to resist disturbances. 10

Figure 1.2 Small-molecule screens and selections allow high throughput screening *in vivo.* 11

Figure 1.3 Functional screens complement computational methods for enzyme discovery. 14

Figure 2.1 Applications of high-throughput *in vivo* biosensors for metabolic pathway engineering 19

Figure 2.2 Optimization of the RNA switch for use in enzyme screening. 23

Figure 2.3 Caffeine oxidation in *S. cerevisiae* using human cytochrome P450s 25

Figure 2.4 Workflow associated with a RNA switch-based screen for enzyme activity 26

Figure 2.5 One-color fluorescence response histograms 27

Figure 2.6 Screening by fluorescence with an RNA switch-based biosensor in 96-well plates enriches cells containing highly active enzymes. 28

Figure 2.7 Screening enzyme libraries by FACS allows significantly higher throughput. 30

Figure 2.8 Changing the plasmid copy number does not significantly affect enzymatic activity. 31

Figure 2.9 Screening by fluorescence repeatedly identifies improved enzyme variants. 32

Figure 2.10 Enzyme characterization suggests that the screening strategy selected for improved catalysis 34

Figure 2.11 Summary of *in vitro* enzyme evolution 37

Figure 2.12 Increasing enzyme thermostability improves mutational tolerance *in vivo.* 38

Figure 2.13 Plasmid maps 46

Table 2.1 Primers used in this chapter 55

Table 2.2 Plasmids and strains constructed in this chapter 57

Table 2.3 Mutations in the BM3 enzyme variants generated in this chapter 59

Table 2.4 Summary of functional characterization data for enzyme variants 60

Figure 3.1 Overview of the enzyme discovery process 64

Figure 3.2 cDNA library construction.. 65

Figure 3.3 Fluorescence histograms of the cDNA library... 66

Figure 3.4 Construction and screening of a cDNA library after subtractive hybridization 67

Figure 3.5 Effect of switch optimization on metabolite detection 70

Figure 4.1 Comparison of theophylline accumulation for high- and low-copy enzyme expression 82

Figure 4.2 Microarray analysis identifies heme depletion as the major cellular stress. 84

Figure 4.3 Heme synthesis and usage must both be elevated to produce large changes in heme content. 86

Figure 4.4 Increasing the total cellular heme level leads to an increase in total enzymatic activity at the cost of slower growth. 88

Figure 4.5 Integrated heme overexpression constructs accumulate highest observed theophylline titers. 89

Figure 4.6 Heme overexpression increases the activity of soluble, but not membrane-associated, hemoproteins. 90

Figure 4.7 Heme depletion leads to increased proteasomal activity and low enzyme expression. 91

Table 4.1 Heme overexpression data 102

Figure 5.1 Dynamic regulation of the lycopene biosynthetic pathway 106

Figure 5.2 Feedback control can reduce the effects of a disturbance to the pathway input... 109

Figure 5.3 A proportional feedback controller can reduce the retroactivity of a downstream reaction... 111

Figure 5.4 Feedback control can accommodate competition over a common substrate. 112

Figure 5.5 Enzymatic activity shows a sublinear dependence on transcription rate. 114

Figure 5.6 Activation of a theophylline-dependent OFF switch 115

Figure 5.7 Measurements of internal metabolite concentrations 116

Figure 5.8 Transfer function of the sensor limits controller performance 119

Figure 6.1 A synthetic RNA switch can be used to identify, optimize, and regulate a biosynthetic enzyme.. 127

1 Introduction

1.1 Open challenges in metabolic engineering

Biological systems are amazingly adept at performing complex synthetic chemistry (Mizutani & Ohta, 2010; Walsh, 2008). Many of our dyes, fragrances, and pharmaceuticals are derived from natural products. However, producing such molecules at the necessary volume and cost can be difficult. Many useful compounds are produced at low concentration in their native host, necessitating the expense of growing large amounts of biomass followed by extensive purification of the desired compound. As an alternative to traditional production methods, researchers can move enzymes and pathways from their native hosts into new, more-tractable production organisms such as *Escherichia coli* or *Saccharomyces cerevisiae*. Freed from the constraints of the native context, we can optimize these engineered organisms to produce the precise compound desired, with high purity and yield (Keasling, 2010). However, the construction of a new metabolic pathway still requires the investment of an enormous amount of time and money.

There are several challenges preventing the efficient construction and optimization of biosynthetic pathways. First, the necessary enzymes must be identified. While improvements in sequencing technology have greatly simplified this process, the identification of a single enzyme can still require significant effort. Next, the pathway must be constructed and optimized. Ideally, this process would largely be a design challenge, and researchers would be able to predictably combine enzymes into a pathway that would behave the same way in a cell as it did in a computer. Unfortunately, our current abilities are far from this ideal. We lack the tools and understanding necessary for forward design of biological systems. In the absence of reliable design tools, we can instead turn to evolutionary methods for pathway optimization. Nature has certainly proven that evolution can be an enormously powerful optimization tool. Unfortunately, here too our abilities are limited, largely by the paucity of general methods for quickly measuring pathway performance.

However, while our current abilities are limited, new tools are rapidly being developed to overcome these challenges. In moving from a reductionist view of biology to a more-holistic perspective, recent advances in systems biology are improving our ability to measure and model biological processes. Similarly, synthetic biology is providing new tools to construct and control biological systems. Our challenge is to apply these new capabilities to metabolic engineering and thereby improve our ability to rapidly construct efficient metabolic pathways in microbes.

1.1.1 Challenges in the predictable design of biological systems

Our ability to modify organisms has grown enormously in recent years. Genomes can be constructed *de novo* (Gibson et al., 2010) or modified on a genome-wide scale (Wang et al., 2009; Warner et al., 2010). New tools are rapidly being developed to aid in the construction of genes (Gibson, 2011) and pathways (Shao & Zhao, 2009; Wingler & Cornish, 2011). Even the cost of direct gene synthesis is decreasing at an exponential rate (Carlson, 2009). Unfortunately, our ability to design biological systems has not kept pace. For example, the first fully synthetic genome was copied almost verbatim from a natural organism with only minor modifications. Even when we reduce the problem down to predicting the relationship between the sequence and function (protein expression level) of a simple genetic element such as a ribosome binding site, the best available computational design tool provides a 47% chance that the actual expression is within twofold of the desired level (Salis et al., 2009). Moving to more-complicated systems, such as protein design, researchers have been able to computationally design functional enzymes (Jiang et al., 2008; Rothlisberger et al., 2008; Siegel et al., 2010). However, the initial designs have only been marginally active, and enzyme optimization required directed evolution (Khersonsky et al., 2010). In particular, ligand binding seems to be a challenging problem to model computationally, as the K_M for published examples of designed enzymes are typically in the range of several hundred μM (Schreier et al., 2009). Designing a multicomponent biological system (Tabor et al., 2009) or trying to predict the interactions between an engineered system and the host organism (Blazeck & Alper, 2010) is more challenging still. Such models are often able to explain observed behavior but are of limited predictive utility.

There are a number of reasons that modeling multicomponent biological systems is difficult. A model is only as good as the underlying data and assumptions. In many cases, even when the biology has been well studied, one unknown interaction can invalidate a carefully constructed model. For example, the regulatory system controlling sugar consumption in yeast has been studied for decades. However, attempts to predict how yeast would respond to a sinusoidally varying glucose concentration were inaccurate (Bennett et al., 2008). Researchers eventually discovered previously unknown interactions between components of the system — a worthwhile result in itself, but one that illustrates the difficulty of constructing a predictive model. Similarly, when researchers attempted to use a model to inform their design of a genetic oscillator, they fortuitously found that the oscillator functioned under conditions that the model predicted would fail (Stricker et al., 2008). Ultimately, this additional robustness was explained by an unexpected coupling of two components that were both degraded by the same proteasome (Cookson et al., 2011). As before, the resulting knowledge was interesting, but the model was inaccurate until the new biology was carefully analyzed.

In other cases, we lack sufficient knowledge to even begin to model the desired system. Building a new metabolic pathway might involve the expression of multiple heterologous enzymes. Characterizing the performance and interactions of each enzyme in the new context requires an enormous amount of work before model building can even begin. The characterization process is not straightforward, and it is unclear exactly what information is needed in order to construct informative models. Once a set of enzymes is characterized in a particular host, the enzymes could perhaps be reused in a predictable fashion, for example, when combining two heterologous pathways in a single host. However, the current lack of consistent characterization, in addition to the uncertainty surrounding what exactly would constitute sufficient characterization, prevents the use of forward design in these situations.

Even when potential pitfalls have been identified, integrating them into a design can be difficult. For example, high plasmid copy numbers have long been recognized as a potentially deleterious load (Jones et al., 2000) and the host can be modified to reduce the effects of this load (Flores et al., 2004). Still, small changes to the plasmid load in a cell can have an unexpectedly large effect on the output of an optimized pathway (Ajikumar et al., 2010). Understanding that there *might be* a problem is mainly useful retrospectively, to explain why a process failed. Instead, we need consistent methods for measuring the ways in which heterologous pathways interact with their host and an explicit understanding of the ways in which these interactions may affect productivity. In the previous example, forward design would require a quantitative measure of the burden due to carrying the various plasmids, as well as a model explaining how changes to that burden would change the concentration of the final product. While the individual pieces of information are available, the lack of consistent characterization standards means that combining them into a cohesive design model is still challenging.

1.1.2 Challenges in the application of evolutionary methods

Lacking the ability to reliably design biological systems that perform to specification, researchers commonly resort to constructing many variants of the desired system, followed by screening to identify the best version. However, there are few available screens that can match the scale of the recent techniques for pathway construction and modification (Dietrich et al., 2010). As a result, researchers are limited to processes that produce obvious phenotypes such as color (Wang et al., 2009) or can easily be coupled to growth (Warner et al., 2010).

One alternative is to simplify the screening process by moving from whole cells to cell lysate. Without the complication of transport limitations or the need to consider toxicity, many more screening techniques become feasible (Arnold & Georgiou, 2003). Unfortunately, while these *in vitro* techniques solve some problems, they also introduce others. Many mutations that improve activity *in vitro* will not be beneficial *in vivo* (Fasan et al., 2007). While false positives can be identified by rescreening *in vivo*, false negatives due to mutations that improve activity *in vivo* but are not beneficial *in vitro* will go undetected entirely. It would be preferable, therefore, to conduct the screening *in vivo*, in as close an approximation to production conditions as possible.

Another option is to use a surrogate substrate that has been modified chemically to introduce a screenable phenotype. Dye molecules (Aharoni et al., 2006; Yang et al., 2010) or chemical groups with robust binding partners (Baker et al., 2002; Peralta-Yahya et al., 2008) can allow the use of simple screens for reactions that would otherwise be very difficult to interrogate. However, these screening systems are often limited by the specific substrate or reaction chemistry or by the need to transport the surrogate into living cells. As a result, these techniques are only applicable to a subset of interesting metabolic pathways and reactions. More generally, the use of a surrogate substrate raises the possibility of finding variants that improve their activity towards the surrogate but not towards the authentic substrate (Aharoni et al., 2006). In addition to screening under conditions that closely approximate production conditions, screening should be performed with authentic substrates whenever possible. New tools will be necessary to allow such screens and selections.

1.1.3 Challenges in discovering uncharacterized enzymes

As sequencing costs decrease and sequenced genomes proliferate, both from isolated strains (Song et al., 2010) and complex uncultured mixtures (Warnecke et al., 2007), the challenge of enzyme discovery has switched from having too little data to having too much. Metagenomic sequencing is a powerful technique, but it can easily produce far more potential sequences than can be individually characterized. For example, a metagenomic library from the termite hindgut produced 700 putative carbohydrate-active enzymes (Warnecke et al., 2007), and a separate library from the cow rumen produced another 28,000 (Hess et al., 2011). Similarly, while plant genomes and transcriptomes are being sequenced at an increasingly rapid pace, the sheer number and diversity of genes involved in secondary metabolism can confound traditional discovery methods. A single plant genome can contain more than 100 terpene synthases (Chen et al., 2011a) and ~ 1% of the genome can consist of P450 monooxygenases (Mizutani & Ohta, 2010). Dealing with the deluge of data will require a combination of computational techniques and new functional screens to reduce the scale down to a level suitable for traditional characterization techniques.

1.2 Tools from systems and synthetic biology

I have described above a number of areas in which current metabolic engineering efforts face serious challenges. However, recent progress in systems and synthetic biology has begun to overcome these obstacles. While our current understanding of natural systems might be insufficient to permit forward modeling of new metabolic pathways, new techniques for characterizing biological systems are narrowing this knowledge gap, and new tools for managing biological complexity are allowing us to produce increasingly reliable designs by reducing or accommodating unknown and undesired interactions. New small-molecule screens and selections are enabling the high-throughput detection of authentic compounds *in vivo* using scalable screening platforms that can quickly be modified to recognize new targets. Finally, these screening tools, in combination with advances in sequencing and synthesis of DNA, are transforming the ways in which we discover new enzymes from nature.

1.2.1 Characterization of biological systems

When transplanted into a new organism, many heterologous enzymes function poorly or not at all. Understanding the reasons for this poor performance is the first step to rationally designing new metabolic pathways. However, in any given pathway there are many potential reasons, and identifying the most significant of these can be difficult.

Some of these reasons are common among many heterologous proteins, and these are the areas in which we have made the most progress. If, for example, a protein requires accessory factors for proper folding and localization, a new environment might lead to misfolding and low expression. In some cases, the protein folding machinery is sufficiently well understood as to allow rational modification of the host to accommodate the demand for additional chaperones (Shusta et al., 1998; Tokuriki & Tawfik, 2009). In other cases, the interactions between a heterologous protein and its host can be very complex, and we are only beginning to catalog all of the interactions (Geiler-Samerotte et al., 2011). This type of catalog is a necessary first step to understanding how a cell is trying to cope with expression of a new enzyme and ultimately to assisting the cell in this process.

In other cases, the interactions between an enzyme or pathway and its new host can be unique to that system. In these cases, analysis tools are critical for understanding the mechanisms by which the pathway places stress on its host. Global measurements of RNA or protein levels can allow an unbiased appraisal of the various limitations that a host places on a heterologous pathway. Global analyses are often necessary simply to understand what a pathway is doing to its host. Seemingly minor perturbations can have effects on a wide range of host processes (Lee et al., 2009) or on seemingly unconnected pathways (Kizer et al., 2008). Proteomics can be used in a similar fashion to identify proteins whose expression is changed due to the introduction of the heterologous pathway, presumably as the host tries to cope with the new stress (Xia et al., 2010). Additional expression of these genes may increase productivity. Finally, even when the broad outlines of the problem are known, such as a heterologous pathway inducing transcriptional feedback inhibition of an endogenous pathway, the specific targets of the feedback might be unclear. Researchers can use transcriptome analyses to identify these targets and then overexpress the relevant genes to increase the pathway output (Choi et al., 2003; Park et al., 2007).

Ultimately, we hope that as we gather more examples of deleterious interactions, patterns will start to emerge. Perhaps some of these seemingly unique stresses will turn out to be common, and we can describe standard protocols for characterizing them and their effect on the host (Canton et al., 2008).

1.2.2 Managing biological complexity

Rather than trying to accurately understand and model the complex interactions inherent in a biological system, an alternative approach is to minimize those interactions and thereby make the system more predictable. For example, synthetic systems might combine orthogonal mechanisms for transcription (An & Chin, 2009) and translation (Rackham & Chin, 2005) with an orthogonal genetic code (Neumann et al., 2010) to minimize interference with the analogous host processes. Additionally, enzymes could be localized to scaffolds (Dueber et al., 2009) or protein microcompartments (Bonacci et al., 2011; Fan et al., 2010), improving pathway productivity while preventing intermediates from diffusing away and interacting with the host. Protein scaffolds have been shown to reduce the enzyme loading required for a given level of pathway output, minimizing the burden on the cell (Dueber et al., 2009). In a similar vein, eukaryotes such as yeast could be modified with designer organelles to sequester heterologous pathways. By reducing the interactions between components of a biological system, these techniques would allow us to more fully understand and model the system.

As we gain a better understanding of the different mechanisms through which unexpected interactions arise (Ventura et al., 2010), we can modify our designs to minimize these effects (Del Vecchio et al., 2008). In metabolic pathways, retroactivity typically manifests itself though allosteric feedback control. This type of retroactivity can be eliminated through the use of feedback-insensitive enzyme variants or promoters (Lee et al., 2007). In other cases, retroactivity might be unavoidable. For example, a metabolic pathway in which a single compound is the substrate for multiple reactions (Nakagawa et al., 2011) would also demonstrate retroactivity, since modifications to the rate of one reaction would change the concentration of the joint substrate and therefore the rates of the other reactions. Similarly, appending a new reaction can cause a previously optimized pathway to fail, due to interference from a new plasmid (Ajikumar et al., 2010) or a new enzyme that competes for cofactors. While cofactor competition could, in theory, be avoided by reengineering one enzyme to use an alternate cofactor (Bastian et al., 2011) or by discovery of an alternative, non-cofactor-dependent enzyme (Gonzalez-Pajuelo et al., 2005), substrate competition is unavoidable. In these cases, minimizing retroactivity requires the introduction of feedback controllers (Figure 1.1).

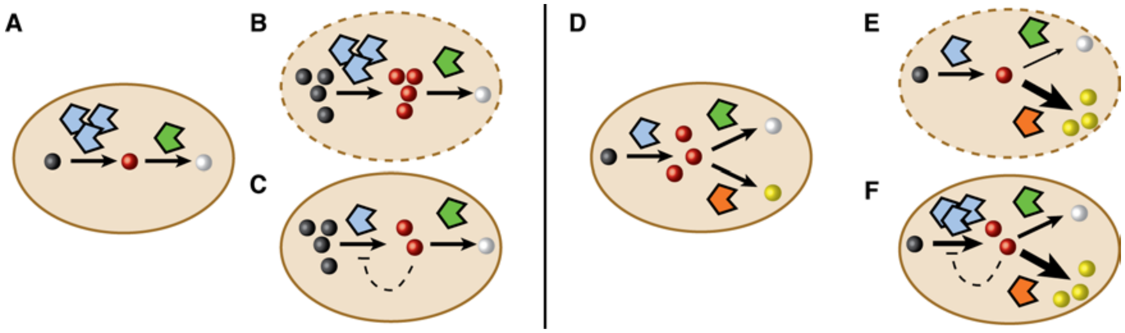

Figure 1.1 Feedback regulation allows a pathway to resist disturbances. (**A**) An engineered metabolic pathway produces a toxic intermediate (red). (**B**) An increase in the concentration of the substrate (black) can lead to accumulation of the intermediate, harming the host cell. (**C**) Feedback regulation allows the pathway to respond to this disturbance by reducing the expression of the upstream enzyme and keeping the concentration of the intermediate acceptably low. (**D**) Alternately, a pathway might contain a branch point, such as an intermediate used both for the desired product (yellow) and a necessary compound in the host metabolism (white). (**E**) Diversion of too much flux to the engineered pathway might deplete the cell of a necessary metabolite. (**F**) Feedback regulation would initially keep the upstream enzyme expression low. As the common intermediate (red) was depleted, the relief of that repression would lead to an increase in expression, thereby rebalancing the pathway.

Feedback regulation is a ubiquitous feature both in natural systems (Winkler et al., 2004) and in other engineering disciplines, but is rarely added to heterologous metabolic pathways (Farmer & Liao, 2000). Including feedback regulation in biological designs could improve system performance (Dunlop et al., 2010) and reduce retroactivity. Researchers could intentionally build in excess capacity to synthesize cofactors or substrates and then adjust the utilization of this spare capacity in response to changing demands. Introduction of a new enzyme that competes for a substrate would lead to increased synthesis of the substrate rather than undesirable retroactivity. Additionally, engineered microbes are expanding into less-predictable environments, such as open pond cultivation of engineered algae for biofuels (Scott et al., 2010) or the production of anticancer compounds inside tumors (Anderson et al., 2006). As a result, the ability to respond to environmental variation will become increasingly important.

1.2.3 Small-molecule screens and selections

An ideal screening system would allow high throughput screening *in vivo* using a biosensor that is specific to the authentic compound desired (Figure 1.2). This ideal system would be easily reconfigured to recognize a new compound, independent of the specific functionality of the compound or the reaction chemistry of the associated pathway. While no current screen meets all of these goals, significant progress has been made (Michener et al., 2012).

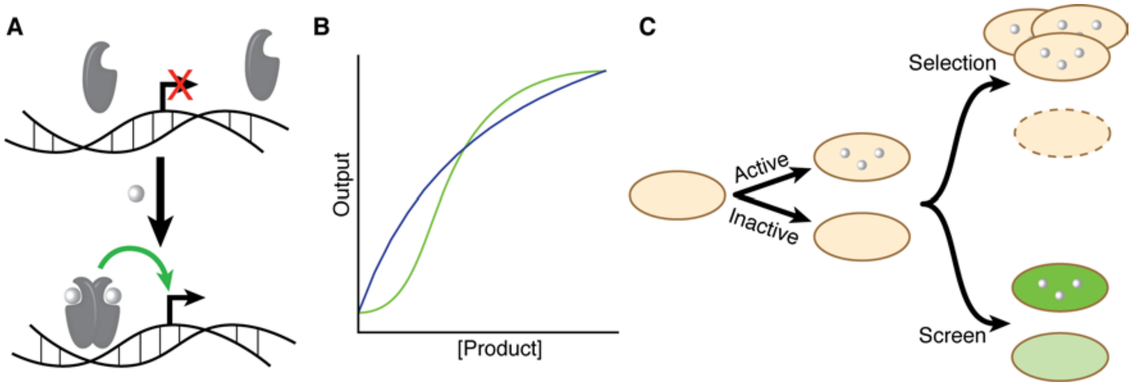

Figure 1.2 Small-molecule screens and selections allow high-throughput screening *in vivo*. (**A**) In a simple sensor system, a promoter is initially inactive. Binding of a small molecule (white) to a transcriptional activator (grey) leads to gene expression from the associated promoter. (**B**) Depending on the details of the sensor and reporter, the screen could produce different transfer curves, such as a graded response (blue) or a cooperative, threshold response (green). (**C**) These sensors can be used to implement screens or selections. Cells containing highly active pathways will produce large amounts of the target molecule (white), while those with inactive pathways will not. In a selection, the cells with active pathways will grow more than cells without. In a screen, they will express more of an easily measured marker, such as GFP.

There are several possible platforms that could fulfill these requirements. Protein transcription factors naturally recognize a diverse range of small molecules. Expression of a screenable or selectable marker from the associated promoter would then allow high-throughput assays *in vivo* (Mohn et al., 2006; Tang & Cirino, 2011; van Sint Fiet et al., 2006). However, there are many small molecules for which no specific transcription factor has been identified. In such a situation, development of a new screen would require modifying a transcription factor to recognize a new ligand, which can be a challenging task (Collins et al., 2005; Collins et al., 2006; Tang & Cirino, 2010). Similarly, proteins can be modified to allow allosteric control of fluorescence (Fehr et al., 2002) or enzymatic activity (Edwards et al., 2008; Guntas et al., 2005; Guntas & Ostermeier, 2004). Fluorescence is a directly screenable phenotype and the new allosteric enzyme, if it provides a phenotype such as antibiotic resistance, can readily be screened or selected. However, these sensors are artificial constructs combining unrelated domains for ligand binding and enzymatic activity. At best, development of a new sensor requires the time consuming integration of a new ligand-binding domain. If no natural binding domain is available, an existing domain must be modified while still maintaining the allosteric linkage between ligand binding and enzymatic activity — certainly a difficult task.

RNA-based biosensors could also be used to sense and respond to small molecules. Synthetic RNA switches can regulate gene expression through a variety of mechanisms, including controlling transcription (Buskirk et al., 2004), mRNA stability (Win & Smolke, 2007), and translation (Desai & Gallivan, 2004). Ligand-binding domains can be selected *in vitro* from random RNA pools (Jenison et al., 1994) and integrated into existing switch platforms (Win & Smolke, 2007). However, the chemical functionality of RNA is quite limited, given the similarities between the four available bases, and as a result there may be entire classes of molecules to which no RNA binding domain can be selected. Still, these new techniques are moving us closer to a time when new high-throughput *in vivo* screens could rapidly be constructed for a wide range of small-molecule targets.

1.2.4 New techniques for enzyme discovery

New sequencing techniques are producing an explosion in the available sequence information. As a result, computational techniques are becoming increasingly important tools for enzyme discovery. In some cases, such as polyketide and nonribosomal polypeptide synthases, conserved sequence signatures allow automatic identification of enzymes by genome mining (Kersten et al., 2011), and the consistent sequence-function relationship allows automated prediction of pathway assembly (Yadav et al., 2009). For other enzymes, comparing genomes of species known to perform a given reaction can allow identification of the associated enzymes (Balskus & Walsh, 2010). If all the species known to perform the desired reaction are closely related, *in silico* subtraction of a related but non-producing strain, either a mutant (Hagel & Facchini, 2010) or a close evolutionary relative (Schirmer et al., 2010), can significantly narrow the list of potential targets. As DNA synthesis costs continue to drop, simply synthesizing and testing all available enzyme homologs can be an elegantly brute force solution to identifying the best enzyme variant for metabolic engineering (Bayer et al., 2009). Sequence information from formerly intractable organisms such as plants (Facchini et al., 2012; Fridman & Pichersky, 2005) is becoming increasingly available, further expanding the reach of computational techniques for enzyme discovery.

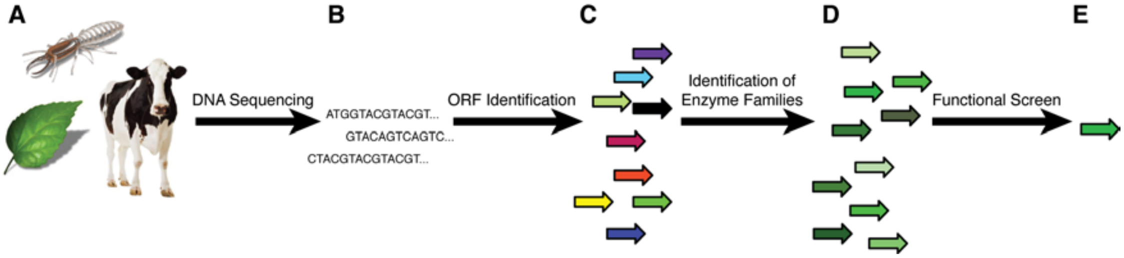

Figure 1.3 Functional screens complement computational methods for enzyme discovery. (**A**) Enzymes can be identified from a wide range of organisms (**B**) using DNA sequencing followed by (**C**) computational predictions of open reading frames to produce a pool of candidate enzymes. (**D**) Bioinformatics techniques, such as homology searches to known enzymes, can narrow down the pool of potential enzymes. (**E**) However, even focusing on a single enzyme family may produce too many enzymes to test individually. In these cases, a functional screen is necessary to identify the single best enzyme variant for the desired conditions.

In many cases, however, enzyme discovery based on sequence homology is only the first step in identifying a promising enzyme for metabolic engineering. Narrowing the list of candidate enzymes to a single, highly homologous population may still produce too many candidates to exhaustively characterize (Hess et al., 2011). In cases such as these, functional screens become necessary to narrow the pool down to a scale that can be individually screened (Taupp et al., 2011). The tools described above for enzyme evolution can also readily be applied to screening metagenomic libraries, allowing the identification of enzymes that produce a specific product. Those same screening systems could then be used to optimize the enzyme to its new production host and to construct feedback control systems that allow predictable integration into an engineered pathway.

1.3 Thesis overview

In this thesis, I demonstrate the value of combining approaches from systems and synthetic biology to advance the metabolic engineering of *S. cerevisiae*. In Chapter 2, I consider a case where the problem, low activity from a specific enzyme, is clear but we lack the understanding to rationally solve the problem. While great strides have been made in computational enzyme design, enzyme optimization is beyond the scope of current techniques. Instead, in these situations we turn to evolutionary approaches, since they require significantly less knowledge of the underlying biology. The challenge in using evolution to optimize a biological system lies in rarity of improvements produced through random mutation, at frequencies of 10^{-3} to less than 10^{-6}. Therefore, the screening system used to identify improved variants is of critical importance. Currently, we have few general techniques for *in vivo* enzyme evolution. I describe the development of a new high-throughput screen for enzyme evolution *in vivo* using a synthetic RNA switch. I apply this novel screen to the evolution of a model protein, a P450 monooxygenase, and ultimately produce a 33-fold improvement in the enzymatic activity and a 22-fold increase in the product selectivity. Finally, I compare my efforts to evolve an enzyme *in vivo* to a parallel evolutionary trajectory *in vitro*, highlighting the difficulties involved in screening *in vitro* when one desires activity *in vivo*. I also use the *in vitro* results to experimentally demonstrate the connection between an enzyme's thermostability and mutational tolerance.

Having demonstrated that the synthetic RNA switch could identify enzymes with caffeine demethylase activity *in vivo*, I next asked whether I could use that same screening technique to identify natural caffeine demethylases. In Chapter 3, I describe my efforts to construct and screen cDNA libraries from *Coffea dewevrei*, a species that produces naturally low-caffeine coffee beans. Previous research has demonstrated that the low caffeine content is due to rapid enzymatic demethylation of caffeine to theophylline, but the enzyme responsible for this transformation has not been identified. I successfully constructed cDNA libraries based on total RNA extracted from leaves of *C. dewevrei*, including a cDNA library in which I used subtractive hybridization with *C. arabica* RNA to enrich for differentially expressed cDNAs. Unfortunately, screening these libraries in *S. cerevisiae* did not identify any caffeine demethylases, and I discuss possible explanations for this negative result.

While evolutionary methods, like those described in Chapter 2, are very powerful, there are some situations in which such methods are not readily applicable. In Chapter 4, I move to a case where the bottleneck in the pathway is less obvious, namely the deleterious interactions between an engineered pathway and the cell in which that pathway resides. To apply evolutionary methods to such a situation, we must mutate every possible target, drastically reducing the frequency of beneficial mutations since most mutations would hit the wrong targets. Instead, the challenge is to narrow down the list of potential targets, at which point we can either modify the pathway to limit the stress it places on the host or engineer the host to accommodate that stress. I demonstrate that a comparative systems-level analysis of multiple pathway variants can help in identification of the major stresses that the pathway places on its host. For my model system, the same cytochrome P450 expressed in *S. cerevisiae*, I show that heme levels limit enzyme expression and, as a result, total enzyme activity. Overexpression of key genes in the heme biosynthetic pathway, in addition to feeding a heme precursor, increases the level of total heme by more than tenfold and the enzymatic activity by 2.3-fold.

Finally, in Chapter 5, I consider further applications of synthetic RNA switches, focusing on their use for controlling metabolic pathways. I describe the design of synthetic feedback control systems to reduce the change in product concentration resulting from variations in the concentration of the substrate. I construct computational models of these controllers and explain the necessary design parameters for components used in such a controller. I also discuss possible methods for experimental validation of a feedback controller as well as applications of such a controller to provide greater composability in engineered metabolic pathways.

2 High-Throughput Enzyme Engineering Using a Synthetic RNA Switch

Portions of this chapter are adapted with permission from Michener JK and Smolke CD (2012) High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. *Metab Eng*. Jul;14(4):306-16.

Metabolic engineering can produce a wide range of bulk and fine chemicals using renewable resources. These approaches frequently require high levels of activity from multiple heterologous enzymes. Directed evolution techniques have been used to improve the activity of a wide range of enzymes but can be difficult to apply when the enzyme is used in whole cells. To address this limitation, I developed generalizable *in vivo* biosensors using engineered RNA switches to link metabolite concentrations and GFP expression levels in living cells. Using such a sensor, I quantitatively screened large enzyme libraries in high throughput based on fluorescence, either in clonal cultures or in single cells by fluorescence-activated cell sorting (FACS). By iteratively screening libraries of a caffeine demethylase, I identified beneficial mutations that ultimately increased the enzyme activity *in vivo* by 33-fold and the product selectivity by 22-fold. As aptamer selection strategies allow RNA switches to be readily adapted to recognize new small molecules, these RNA-based screening techniques are applicable to a broad range of enzymes and metabolic pathways.

2.1 Introduction

Recent advances in metabolic engineering have involved the construction of multi-step enzymatic pathways to synthesize complex molecules, such as isoprenoids, benzylisoquinoline alkaloids, and steroids, from simple precursors (Hawkins & Smolke, 2008; Ro et al., 2006; Szczebara et al., 2003). The enzymes responsible for these reactions can be taken from a variety of sources and combined into a single production host, and as a result they often require modification before they function well in the new environment (Chang et al., 2007). Additionally, many natural biosynthetic pathways have uncharacterized reactions for which alternative enzymes must be identified (Yim et al., 2011) or engineered (Bastian et al., 2011) to reconstruct these pathways in synthetic hosts. Finally, once the pathway is constructed, enzyme activities must be balanced to optimize pathway productivity and yield (Ajikumar et al., 2010). Typically, each of these optimization steps involves the construction of many pathway variants followed by the identification of the best resulting pathway. Therefore, optimization requires the ability to measure the production of the desired metabolite at high throughput using an appropriate screen or selection (Dietrich et al., 2010).

When measuring the productivity of a reaction or pathway, an ideal screening system would (a) function *in vivo*, so that screening is performed under the same conditions as production; (b) allow high-throughput analysis, enabling the characterization of large libraries of variants; (c) be scalable, or readily adapted to recognize new small molecules and discriminate between structurally similar compounds; (d) measure the specific reaction desired, without being limited to surrogate substrates or specific reaction chemistries; and (e) be parallelizable, and thus capable of simultaneously measuring multiple metabolites (Figure 2.1A). No current small-molecule screening system meets all of these requirements.

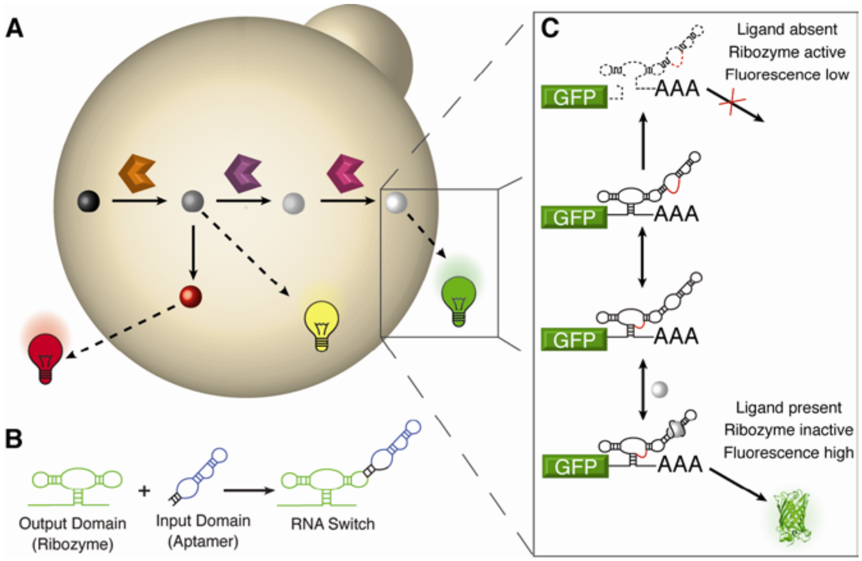

Figure 2.1 Applications of high-throughput *in vivo* biosensors for metabolic pathway engineering. (**A**) The ideal biosensor platform would convert the concentration of the desired metabolite (not a surrogate) into an easily measured signal, be readily modified to detect new metabolites, and function in parallel to allow simultaneous measurement of multiple points in a metabolic pathway. Circles: metabolites; hexagons: enzymes; light bulbs: biosensor signals. (**B**) Synthetic RNA switches are built through the modular assembly of an input domain and an output domain. When properly folded, the input domain (encoded in an RNA aptamer, blue) binds the desired small molecule. If the output domain (encoded in a ribozyme, green) folds correctly, it cleaves itself. In an ON switch, the two domains are connected in such a way that only one domain can properly fold at any given time. (**C**) Engineered RNA switches act as programmable *in vivo* biosensors for desired metabolites. The RNA switch is placed in the 3' untranslated region of a fluorescent reporter gene. If the output domain folds and cleaves, it removes the poly-A tail of the associated mRNA, leading to rapid degradation and low gene expression. Addition of the small molecule ligand favors the conformation where the input domain is properly folded and the output domain is misfolded. Therefore, increasing concentrations of ligand lead to lower cleavage rates and higher gene expression.

Many enzymes can be assayed and evolved *in vitro*, where the reaction conditions may be more carefully controlled (Arnold & Georgiou, 2003). However, mutations that improve activity *in vitro* may be neutral or deleterious *in vivo* (Fasan et al., 2007), and mutations that improve activity *in vivo* may occur through mechanisms that are absent *in vitro* (Bulter et al., 2003). While precise analytical techniques, such as liquid or gas chromatography coupled to mass spectrometry, are generally available to measure any desired small molecule, their slow speed limits the throughput of any resulting screen (Leonard et al., 2010). In some cases, the desired compound either produces (Wang et al., 2009) or can be linked to (Santos & Stephanopoulos, 2008) phenotypes, such as color, that are rapidly identifiable. These compounds can easily be screened in high throughput, but applications are limited by the scarcity of such phenotypes among relevant compounds. Similarly, when the desired compound is required for cell growth, selections can be powerful tools for improving pathway yield (Pfleger et al., 2007). However, molecules linked to cell growth tend to be endogenous, so these auxotrophic selections are useful primarily for increasing substrate availability, rather than optimizing a heterologous pathway. In general, selection strategies allow larger libraries (effectively limited only by transformation efficiency (Peralta-Yahya et al., 2008)), but produce a threshold rather than graded response (Desai & Gallivan, 2004). Additionally, responding to multiple signals using a selection requires a genetic logic gate (Anderson et al., 2006) to integrate multiple signals into a single response (growth), while in a screen the researcher has greater flexibility to independently adjust the screening threshold for each signal. Other screening systems have been developed using transcription factors that respond to the desired product (Mustafi et al., 2012; Tang & Cirino, 2011; van Sint Fiet et al., 2006). While these assays can precisely report the concentration of the desired compound, their reuse for detection of even a slightly modified compound requires significant understanding and engineering of the associated biosensor (Tang & Cirino, 2011). The generation of a protein-based biosensor *de novo* is still challenging (Schreier et al., 2009). Finally, chemical complementation, a modified yeast three hybrid assay, has been used to screen enzyme libraries for a variety of different chemistries (Baker et al., 2002; Lin et al., 2004; Peralta-Yahya et al., 2008). Unfortunately, the assay requires the use of extensively modified surrogate substrates, limiting the types of reactions that can be screened with this approach.

Advances in synthetic biology have led to the design of modular, programmable, RNA-based control elements, or RNA switches (Figure 2.1B) (Win et al., 2009). RNA switches generally link an input domain (an RNA aptamer) to an output domain (an RNA gene-regulatory component), resulting in a control element that regulates gene expression in response to binding of a ligand, such as a protein or small molecule (Figure 2.1C). Synthetic RNA switches responsive to exogenous small molecules have been constructed using a variety of output domains in a diverse range of hosts (Buskirk et al., 2004; Soukup & Breaker, 1999; Suess et al., 2004; Topp et al., 2010). When the input and output domains of an RNA switch are distinct, new input domains can be selected *de novo* (Jenison et al., 1994) and then readily integrated into existing switch platforms (Win & Smolke, 2007). As such, an *in vivo* screening system using RNA switches can be readily reconfigured to respond to new metabolites, providing a platform for the development of scalable, high-throughput, *in vivo* biosensors for metabolic and enzyme engineering (Desai & Gallivan, 2004). However, previous efforts using RNA switches in high-throughput screens have focused only on evolving improved switches (Fowler et al., 2008; Lynch & Gallivan, 2009). These screens can use saturating concentrations of an exogenous ligand and thereby take advantage of the entire dynamic range of the switch. In order to use RNA switches as a platform for screening enzyme libraries, the switches must accurately and precisely discriminate between small differences in the concentrations of heterologous metabolites.

I have developed a generalizable *in vivo* screening strategy for product accumulation using engineered RNA switches as the key biosensor components. These novel biosensors link the concentration of a product metabolite to GFP expression levels in living cells. I use an RNA-based biosensor to quantitatively screen large enzyme libraries in high throughput based on fluorescence, either in clonal culture by flow cytometry or in single cells by fluorescence-activated cell sorting (FACS). I demonstrate that the RNA-based biosensor has sufficient precision to distinguish small changes in fluorescence and therefore identify relatively small improvements in activity. Additionally, the biosensor can be coupled to

FACS to allow screening of large enzyme libraries (~ 10^6). By iteratively applying this screen to libraries of a caffeine demethylase enzyme in yeast, I identified a series of beneficial mutations that ultimately increased the enzyme activity *in vivo* by 33-fold and the product selectivity by 22-fold. My work demonstrates that modular RNA switches provide a flexible screening platform for metabolic and enzyme engineering.

2.2 Results

To demonstrate the RNA-based *in vivo* screening system, I studied the production of the purine alkaloid theophylline in *Saccharomyces cerevisiae* through the enzymatic demethylation of caffeine. Setting up a new screen required that I develop two components: (1) an appropriate biosensor that could precisely report the concentration of the product, theophylline, without interference from the substrate, caffeine; and (2) an enzyme that regiospecifically produces theophylline from caffeine *in vivo*.

2.2.1 Development of the screening system

Previous work in the Smolke Laboratory led to the construction of a theophylline-responsive RNA switch to control GFP expression in *S. cerevisiae*, providing the basis for an *in vivo* screen (Win & Smolke, 2007). When fed increasing amounts of theophylline, ranging from 10 μM to 5 mM, the switch produces a graded increase in fluorescence. The input domain of this RNA switch is an aptamer that binds theophylline ~ 10,000-fold more tightly than it binds caffeine, making it an ideal biosensor for this screening strategy (Jenison et al., 1994).

However, the original switch was too noisy to be used as a screen for enzyme activity (Figure 2.2A). First, the biosensor construct was expressed from an inducible promoter, the

Gal1-10 promoter, that produced a bimodal induction profile. Even in the presence of saturating concentrations of the inducer, a sizable population of cells remained in the uninduced state. These cells would be false negatives in a positive screen and false positives in a negative screen, sufficient in either case to disrupt the screen. Second, the biosensor's ratio of signal-to-noise was very low. The change in fluorescence due to the presence of the metabolite was significantly smaller than the variation in expression due to expression noise. As a result, screening in single cells would be extremely difficult, and even measuring population mean fluorescence would show significant culture-to-culture variability (Figure 2.2B). Finally, the biosensor was expressed from a centromeric plasmid. In such a situation, a simple evolutionary solution to increase the mean fluorescence is to increase the plasmid copy number by disrupting the centromere (Hill & Bloom, 1987). Each of these problems needed to be addressed before the biosensor could be used for enzyme screening.

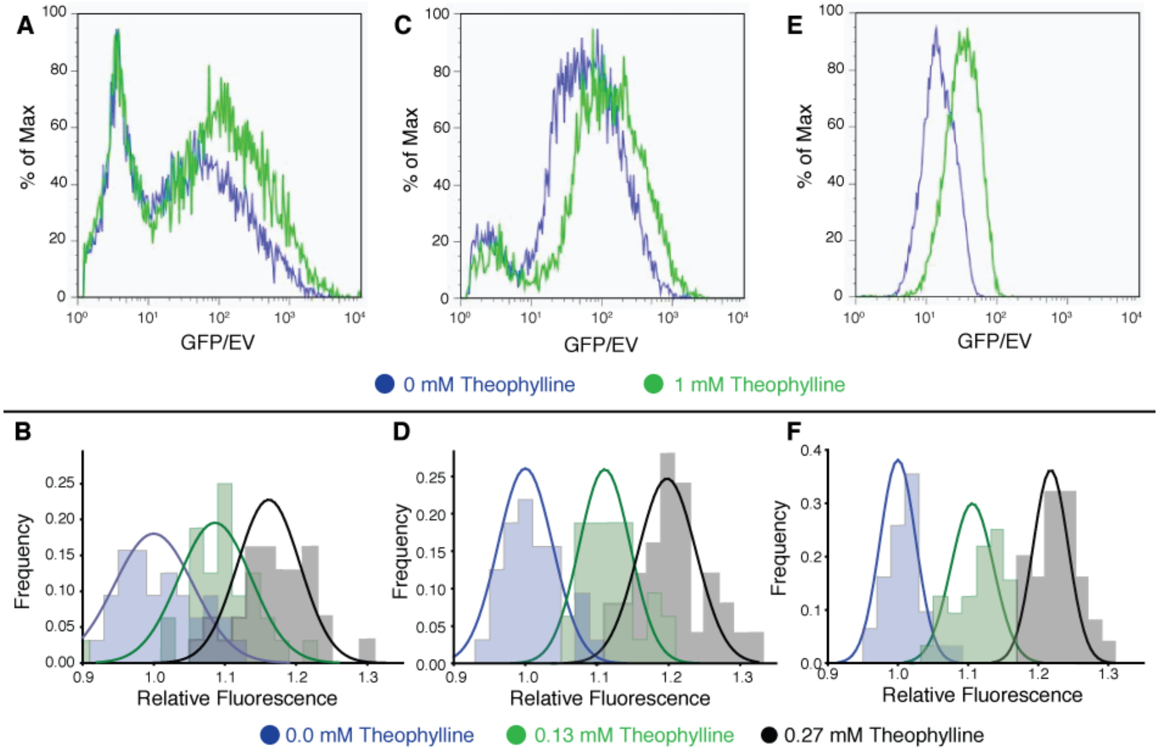

Figure 2.2 Optimization of the RNA switch for use in enzyme screening. (**A**) Single-color fluorescence histogram of yeast expressing the original RNA switch from an inducible GAL1-10 promoter on a centromeric plasmid, grown under the indicated theophylline concentration. The GFP fluorescence of each cell is normalized by the electronic volume (EV). The construct exhibits substantial noise in expression that makes the construct unsuitable as a biosensor for enzyme screening. (**B**) Decreasing the cell-to-cell variation also decreases the culture-to-culture variation. When screening by flow cytometry, screening efficiency is determined by the coefficient of variation (CV) between replicate cultures. For each of the screening constructs, replicate cultures (n=32) were grown in the presence of varying amounts of theophylline and the geometric mean fluorescence of each culture was determined by flow cytometry. The relative fluorescence is the ratio of the geometric mean fluorescence for a single culture relative to the average geometric mean for all 32 uninduced cultures. For the original screening construct, the CV is 5.0%. (**C+D**) Modification of the biosensor to use a constitutive TEF1 promoter eliminated the uninduced population and narrowed the distribution of GFP expression, reducing the CV to 3.8%. (**E+F**) An RNA switch-based biosensor that can readily distinguish between small changes in fluorescence was developed by integrating the biosensor construct into the genome to further reduce variability in expression, ultimately lowering the CV to 2.7%.

To eliminate the uninduced population, I replaced the GAL1-10 promoter with the strong, constitutive TEF1 promoter (Figure 2.2C–D). This change had the additional effect of moderately reducing both the cell-to-cell and culture-to-culture variability in expression, by approximately one third and one quarter, respectively. Next, the biosensor construct was integrated into the yeast chromosome in the lys2 locus to yield strain CSY492 (Figure 2.2E–

F). The integration further reduced the cell-to-cell variability in expression by approximately twofold, presumably by avoiding variability in plasmid copy number, in addition to reducing the culture-to-culture variability by a further 30% and eliminating the potential for spontaneous loss of plasmid copy number control. The resulting biosensor strain shows clear separation between populations with and without theophylline and no response to caffeine (Figure 2.5A).

2.2.2 Identification of a starting enzyme

With a theophylline-responsive screening system in hand, the next challenge was to identify an enzyme capable of demethylating caffeine to theophylline. This demethylation reaction occurs in plants, but the enzyme has not been cloned (Mazzafera, 2004). Similarly, human liver P450 monooxygenases can catalyze the desired demethylation, but with poor product selectivity (Tassaneeyakul et al., 1994). I expressed CYP2D6 in *S. cerevisiae*, confirming both the presence of low levels of caffeine demethylase activity as well as the expected product promiscuity (Figure 2.3). In general, human liver P450s are difficult engineering substrates, as evolution has selected for both substrate and product promiscuity.

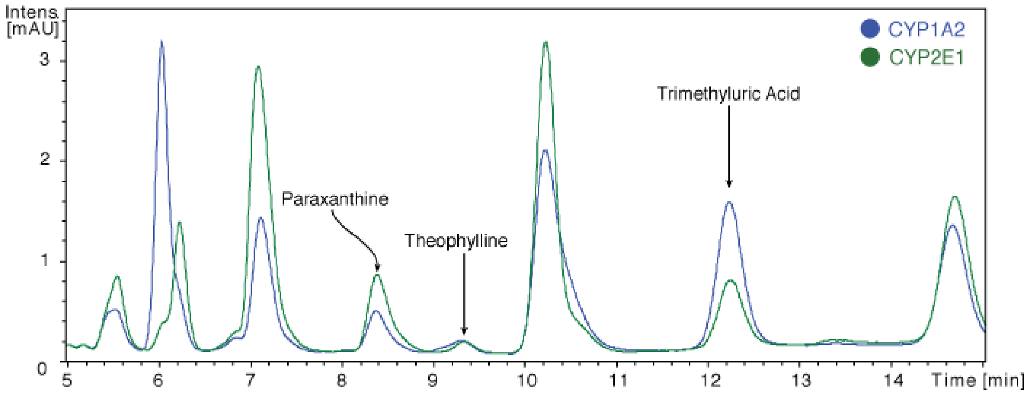

Figure 2.3 Caffeine oxidation in *S. cerevisiae* using human cytochrome P450s. The enzymes showed the expected caffeine oxidation (to trimethyluric acid) and demethylation (to paraxanthine and theophylline). However, theophylline production was very low, and the enzymes showed high product promiscuity.

Lacking the native plant enzyme and preferring to avoid working with CYP2D6, I engineered an alternative reaction using the bacterial P450 monooxygenase CYP102A1 from *Bacillus megaterium* (Narhi & Fulco, 1986). This P450, known as BM3, is soluble, highly active, and, though its natural substrates are long chain fatty acids, can be readily evolved to accept a variety of novel substrates (Dietrich et al., 2009; Fasan et al., 2007; Kille et al., 2011; Lewis et al., 2009; Rentmeister et al., 2009). Since wild-type BM3 does not show activity on caffeine, I instead screened a collection of existing BM3 variants *in vitro* by HPLC to identify enzymes capable of regiospecifically demethylating caffeine to theophylline. The most active enzymes were then expressed in *S. cerevisiae* and assayed by HPLC for activity *in vivo*. When fed 1 mM caffeine, the cells produced low levels of theophylline in the media, as well as minor amounts of the side product paraxanthine. The best of these BM3 mutants, termed caffeine demethylase 1 or CDM1, was then yeast codon optimized to give the enzyme yCDM1.

2.2.3 Plate-based enzyme screening

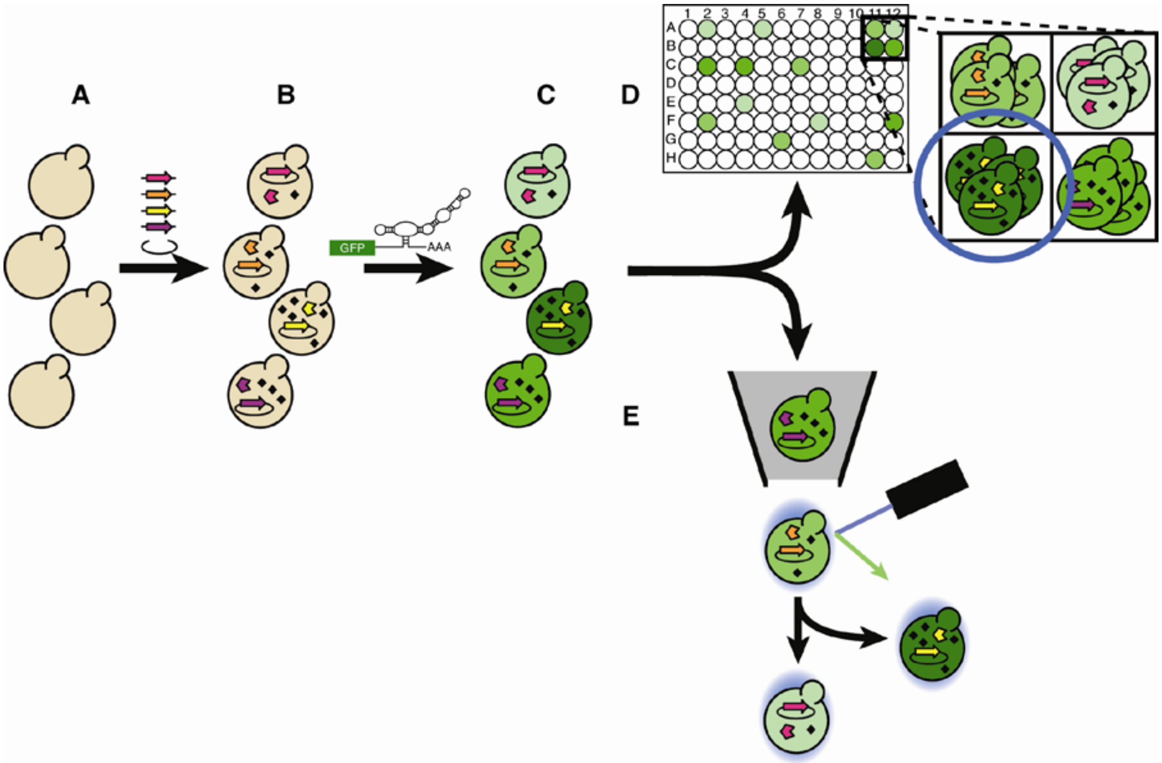

Figure 2.4 Workflow associated with a RNA switch-based screen for enzyme activity. (**A**) An enzyme library is constructed in yeast using either error-prone PCR or DNA shuffling strategies. The resulting library is transformed into yeast cells using a high-efficiency gap repair method. (**B**) The substrate molecule is added to the culture, and the enzymes convert it to product (black diamonds). Cells harboring more highly active enzymes will make more of the product molecule. (**C**) The product interacts with the RNA switch. Cells that make more product have a greater percentage of switches in the conformation where the output domain is misfolded and thus exhibit higher GFP fluorescence. (**D**) Cells can be screened by flow cytometry in clonal culture in 96-well plates. This screening method effectively limits library sizes to ~ 10^3, but can distinguish small changes in enzymatic activity. (**E**) Alternately, mixed cultures can be sorted by fluorescence-activated cell sorting (FACS). Using FACS, libraries of ~ 10^6–10^7 can be screened in a matter of hours, though with less precision than in 96-well plates.

While yCDM1 produced theophylline, the enzymatic activity was low, making this a good model for a reaction in an engineered pathway that depends on an alternative enzyme reacting with a nonnative substrate. Therefore, I investigated whether the RNA-based biosensor would allow me to screen enzyme libraries *in vivo* by fluorescence and optimize the activity of the caffeine demethylase. Screening by fluorescence can be accomplished either in 96-well plates or by FACS (Figure 2.4D-E). Screening in well plates is more precise, since many cells can be analyzed to determine the mean fluorescence of the population. When I expressed yCDM1 in strain CSY492, a small but consistent increase in fluorescence was detected on addition of caffeine, indicating that the caffeine was demethylated to theophylline, which then activated the RNA switch and increased GFP expression (Figure 2.4A-C and Figure 2.5B). Elimination of the catalytic activity of yCDM1 (Neeli et al., 2005) removed the fluorescence response to caffeine, demonstrating that the biosensor was specifically detecting the enzymatic production of theophylline (Figure 2.5B).

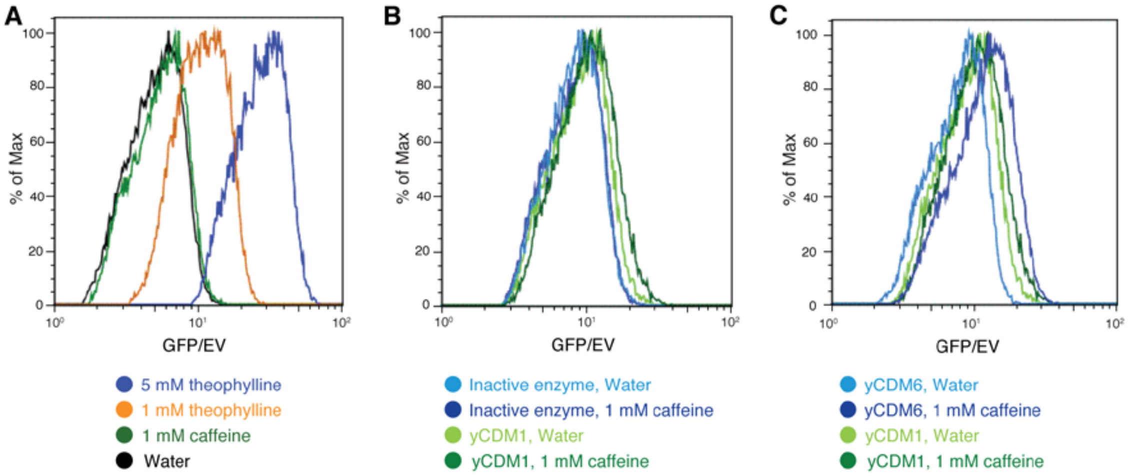

Figure 2.5 One-color fluorescence response histograms. (**A**) Fluorescence histogram of CSY492 grown in the presence of theophylline, caffeine, or water. The screening strain has a graded response to theophylline and no response to caffeine. (**B**) Fluorescence histograms of CSY492 containing active (green) and inactive (blue) versions of yCDM1. The fluorescence increases when caffeine is added to cells containing the active enzyme but is unchanged when caffeine is added to cells with the inactive enzyme. (**C**) Fluorescence histograms of CSY492 containing yCDM1 (green) and yCDM6 (blue). The more-active enzyme produces a larger increase in fluorescence upon addition of caffeine.

Because yCDM1 produced a very small change in fluorescence, I performed my initial screening in 96-well plates using flow cytometry. Measuring fluorescence by flow cytometry is a relatively slow process, potentially limiting the throughput of the resulting screen. In order to maximize the throughput, I streamlined the assay conditions by determining the optimal combination of cell density, flow rate, and acquisition time. Dense cell suspensions allow faster analysis but also increase the probability of clogs. Similarly, faster flow rates increase the analysis speed, but also raise the likelihood of analyzing aggregates. Finally, I had to balance the need for accuracy, which required long acquisition times for precise population-level statistics, with the need for throughput. This optimization brought the assay time down to ~ 40 seconds per sample, compared to 7 minutes per sample for analysis by HPLC. Using these conditions, my library sizes were limited by the necessity of growing cells in 96-well plates rather than my ability to assay the cultures on the flow cytometer.

Next, I generated a library of yCDM1 mutants by error-prone PCR and transformed them by gap repair into a high-copy expression plasmid in CSY492. Individual clones were selected from this library, grown in the presence of caffeine, and assayed for mean fluorescence and for theophylline production by HPLC analysis. A correlation was shown between the fluorescence change and theophylline levels (Figure 2.6A, inset), suggesting that screening by fluorescence would enrich the population for active enzymes. Therefore, I generated a new collection of ~ 800 yCDM1 mutants and measured the fluorescence of each mutant in the presence of caffeine. The brightest ~ 25% of the screened colonies were then assayed for theophylline production by HPLC analysis. The screened population showed a significant enrichment of active mutants relative to the unscreened library (Figure 2.6B). The enzyme variants with the highest activity were recloned to confirm activity, and the best validated mutants increased theophylline production by 50–80% relative to yCDM1.

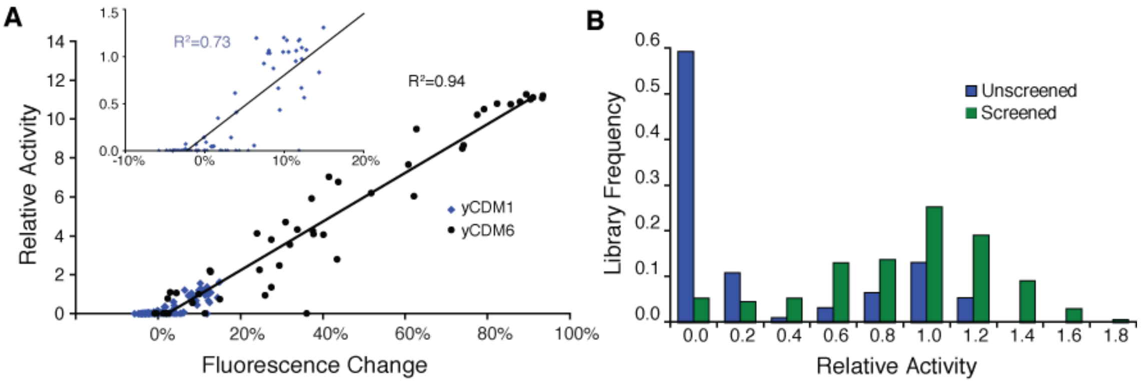

Figure 2.6 Screening by fluorescence with an RNA-switch-based biosensor in 96-well plates enriches cells containing highly active enzymes. (**A**) Theophylline accumulation correlates to population mean fluorescence. Each point represents a single clonal culture from a population of random enzyme mutants, using either yCDM1 (blue) or yCDM6 (black) as the template. Inset shows only the data for the initial enzyme yCDM1. As the theophylline production increases, the fluorescence measurements become more discriminating. (**B**) Screening clonal cultures for high fluorescence levels allows enrichment of active enzymes. A library of random mutants of yCDM1 was constructed in yeast. The distribution of product accumulation (relative to yCDM1) was measured before (blue, 92 clones) and after (green, 208 clones) screening by fluorescence. Screening clonal cultures by fluorescence allows the identification of several enzymes with 50–80% improvements in accumulation.

The improved mutants identified in the first screen encompassed a total of twelve amino acid mutations, which I randomly recombined to produce a new library and screened using the same criteria. I assayed ~ 1600 variants by fluorescence, and ~ 15% exhibiting the highest fluorescence levels were analyzed by HPLC. The best of the validated enzyme variants, a double mutant S72F/A603T identified as yCDM3, showed a 3.4-fold increase in theophylline production relative to yCDM1. The mutant yCDM3 was subjected to another round of random mutagenesis by error-prone PCR and shuffling of the resulting mutations, screening each library using the plate-based assay described above. This process yielded mutant yCDM5 with additional mutations Q27H, R47S, and F72I and an increase in theophylline production of 2.5-fold relative to yCDM3. However, attempts to identify further improved mutants of yCDM5 using flow cytometry-based screening in 96-well plates were unsuccessful, despite screening more than 5000 colonies.

2.2.4 Screening enzymes in single cells

While screening by flow cytometry is significantly faster than screening directly by HPLC analysis, the throughput is still limited by the need to culture isolated colonies in 96-well plates. Using such plate-based screening strategies, library sizes of ~ 10^3 are feasible, but even ~ 10^4 would be unwieldy. By moving to single-cell analysis and recovery with fluorescence-activated cell sorting (FACS), libraries of ~ 10^6–10^7 may be screened by fluorescence in a matter of hours (Chen et al., 2011b; Yang et al., 2010). However, single-cell measurements by FACS exhibit more variation than the population mean values collected through flow cytometry of clonal cultures. To minimize this variation, I used a modified screening system in which a constitutively expressed mCherry protein was used to normalize for extrinsic noise in gene expression (Liang JC, Chang AL, Kennedy AB, and CDS, in submission). This dual color screening system was integrated into the yeast chromosome to further reduce expression noise, yielding strain CSY820 (Figure 2.7A).

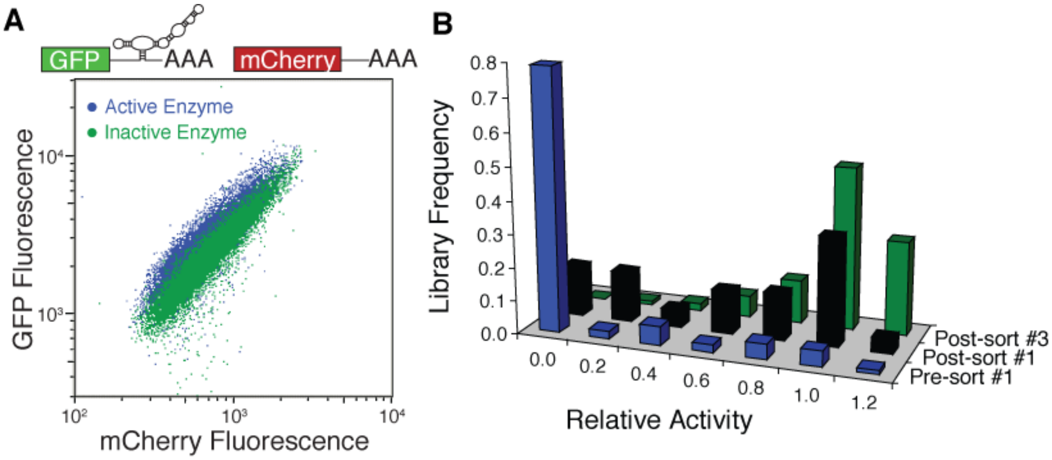

Figure 2.7 Screening enzyme libraries by FACS allows significantly higher throughput. (**A**) A dual color biosensor allows screening by FACS. A second unregulated fluorescent reporter (mCherry) is used to normalize for the extrinsic expression noise, providing better resolution between cell populations exhibiting small changes in fluorescence. Fluorescence dot plots are shown for yeast harboring the two-color construct and expressing either active or inactive enzyme grown in 1 mM caffeine. (**B**) Screening single cells by FACS enriches for active enzymes. A library of random mutants of yCDM6 was constructed in yeast. The distribution of theophylline accumulation by members of the library were measured before sorting (blue), after a single positive sort (black), and after a total of three sorts (positive/negative/positive; green). Sorting virtually eliminated the inactive enzymes from the population and enriched for enzymes that are at least as productive as the parent.

I constructed a library of yCDM5 mutants by error-prone PCR and transformed them into the dual color screening strain CSY820, producing ~ 10^6 transformants. This library was sorted three times: first, in the presence of caffeine and selecting for the brightest 5% of cells by mCherry-normalized GFP fluorescence; second, in the absence of caffeine, selecting for the ~ 70% of cells that showed background levels of fluorescence; and third, in the presence of caffeine, selecting again for the brightest 0.5% of cells by normalized GFP fluorescence. Sorting by fluorescence reduced the library size from ~ 10^6 down to ~ 10^3. The resulting cells were isolated on agar plates and then screened by fluorescence in 96-well plates to further narrow the library. Screening ~ 750 colonies by flow cytometry and the top ~ 250 by HPLC yielded mutant yCDM6, with a single additional mutation E435G and a 1.6-fold increase in theophylline production relative to yCDM5.

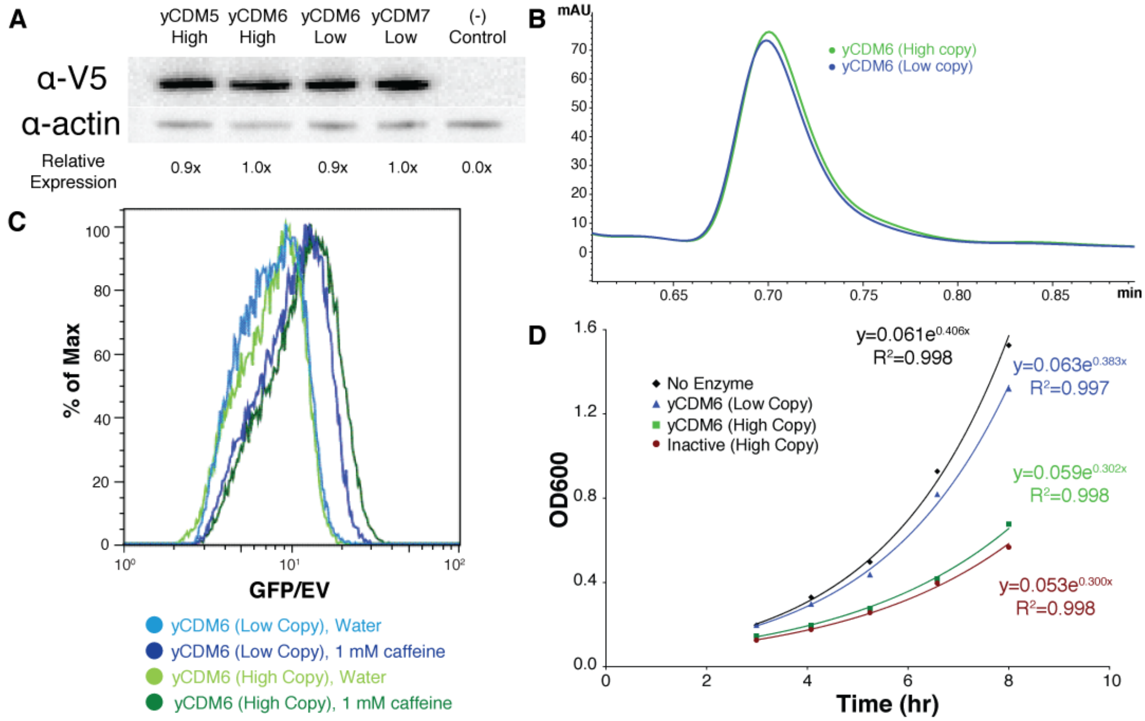

Figure 2.8 Changing the plasmid copy number does not significantly affect enzymatic activity. (**A**) A Western blot demonstrates that total enzyme expression does not change as the DNA copy number is decreased. The enzyme contains a C-terminal V5 epitope, and the anti-actin antibody is used as a loading control. The relative expression values are calculated as the ratio of anti-V5 intensity/anti-actin intensity, normalized to yCDM6-High. The blot shown is representative of three independent experiments. (**B**) Total theophylline production differs by less than 10% between yCDM6-High (green) and yCDM6-Low (blue). Theophylline elutes at 0.70 minutes. (**C**) Fluorescence response histograms for yCDM6-High (green) and yCDM6-Low (blue). Despite the similar levels of theophylline production, the low-copy expression system shows a smaller change in fluorescence, presumably indicating lower theophylline per cell. (**D**) Growth curves for cells containing empty plasmid (black), yCDM6-High (green), yCDM6-Low (blue), or yCDM1 A264H (red). The curves are an exponential fit to the data. High-copy expression of yCDM6 causes a significant decrease in growth rate. Lowering the plasmid copy number relieves ~ 80% of the growth inhibition. While the cells with the low expression system may produce less theophylline per cell, they grow faster and therefore have more time to make theophylline, resulting in similar total production.

While the FACS-based screen was successful, the sorts did not provide as strong an enrichment for active enzymes as I had expected. Upon further examination of the FACS protocol, I discovered that a subset of cells with inactive enzymes were growing significantly faster than those with active enzymes (Figure 2.8D). As a result, cells with inactive enzymes outcompeted the cells with active enzymes during the growth phases between sorting runs, thus limiting the enrichment from the FACS-based screen. This effect was not dependent on catalysis, as a catalytically inactive mutant still showed growth inhibition (Figure 2.8D). To relieve the stress associated with enzyme overexpression, yCDM6 was moved from a high-copy 2μ plasmid to a centromeric plasmid. Despite the expected change in enzyme DNA copy number (~ 10x), the total enzyme expression did not change significantly (Figure 2.8A) and the supernatant theophylline concentration decreased by less than 10% (Figure 2.8B). However, lowering the enzyme expression increased the growth rate of cells with active enzyme near that of cells with an empty plasmid (Figure 2.8D). Using this new expression system, sorting by FACS was shown to enrich a library for active enzymes (Figure 2.7B). Two more rounds of random mutagenesis by error-prone PCR and FACS-based screening led to the identification of mutant yCDM8, with 1.7-fold higher theophylline production and additional A87S/I174V mutations relative to yCDM6. Overall, the theophylline production increased 23.2±2.5-fold relative to yCDM1 (Figure 2.9). I also calculated the ratio of apparent v_{max} to apparent K_M for each enzyme (Figure 2.9A). This ratio increased by 33±4-fold relative to yCDM1 (Figure 2.9B).

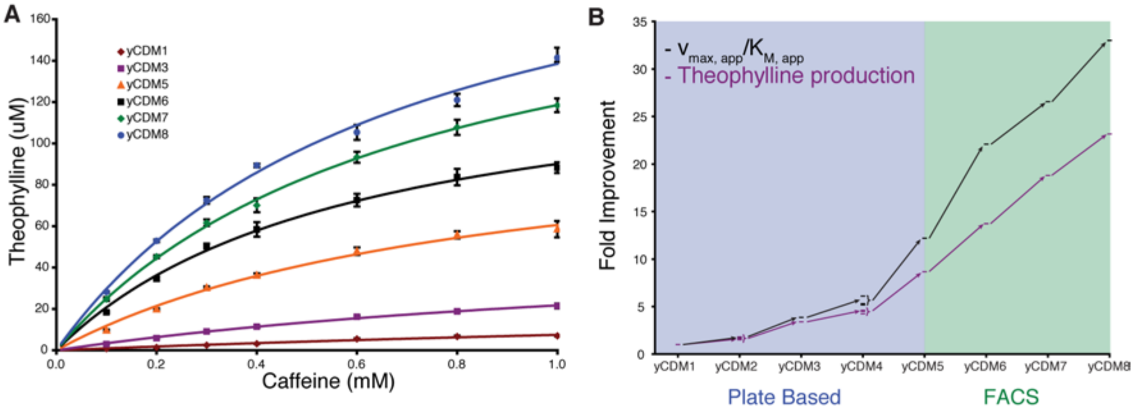

Figure 2.9 Screening by fluorescence repeatedly identifies improved enzyme variants. (**A**) Theophylline production as a function of substrate concentration for the evolved enzyme variants. Cells containing each enzyme were grown in the presence of a range of caffeine concentrations. The resulting data were fit to a Michaelis-Menten equation to determine the apparent K_M and apparent v_{max}. The error bars show ± one standard deviation, calculated from three biological replicates. (**B**) Summary of improvements over the course of the evolution process. The data shown are the average of three biological replicates, with standard deviations listed in Table 2.4.

2.2.5 Characterization of improved mutants

I next sought to understand the mechanism leading to the increased activity of the evolved enzymes. The RNA switch-based screening strategy identifies improved enzymes based on the final product concentration. In general, there are two paths through which product titers could improve: either by increasing enzyme expression or by increasing the specific activity of the enzyme. Western blot analysis indicates that total enzyme expression does not increase significantly over the course of the evolution (Figure 2.10A), although I cannot rule out an increase in the fraction of active, properly folded enzyme. However, the apparent K_M decreases from 1.5±0.1 mM to 0.69±0.04 mM, and the product selectivity (the ratio of the desired product, theophylline, to the undesired side product, paraxanthine) increases from 10.3±0.5 to 230±20 (Figure 2.10B). There was no selective pressure for product selectivity applied over the course of the evolution; simply selecting for theophylline production appears to have imposed a selection against paraxanthine. The apparent K_M stabilizes at ~ 700 μM, approaching the K_M of 290 μM for the wild-type BM3 with one of its native substrates, lauric acid (Noble et al., 1999). Since the assays were conducted while feeding 1 mM caffeine, a mutant with lower apparent K_M would not show significant enrichment, as the parent is already attaining 60% of v_{max}. Both of these measures, the apparent K_M and product selectivity, suggest that the specific activity of the enzyme is increasing.

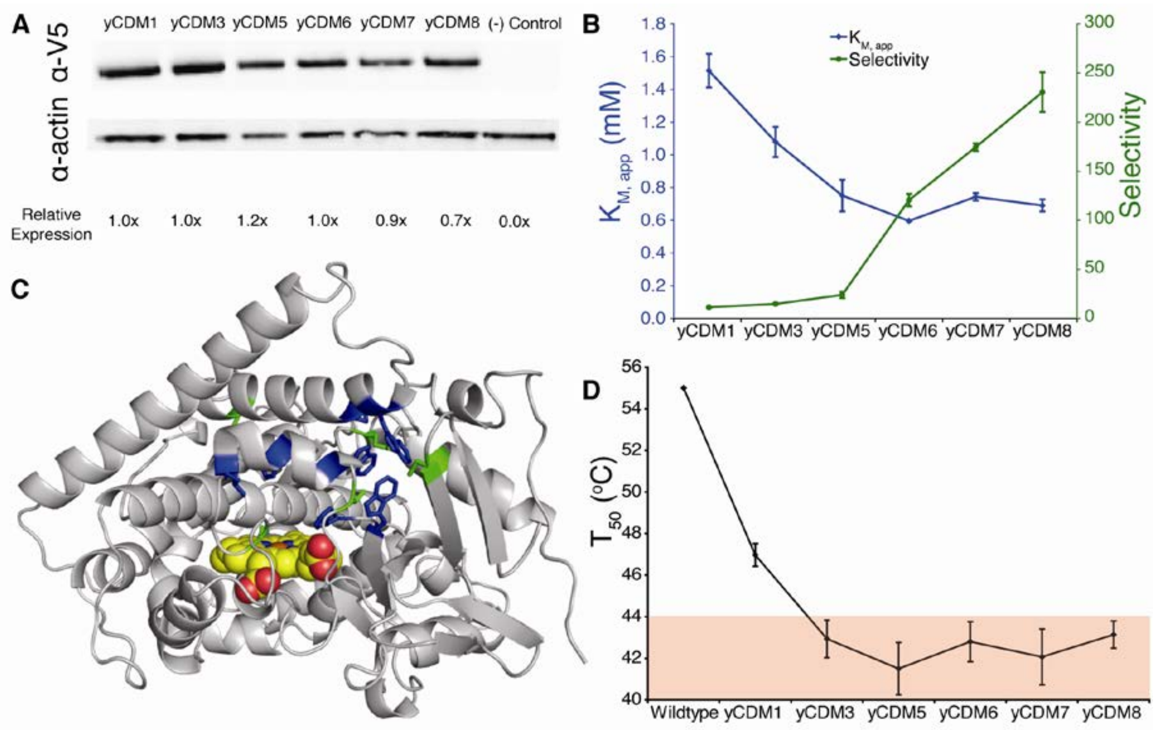

Figure 2.10 Enzyme characterization suggests that the screening strategy selected for improved catalysis. (**A**) A Western blot shows no significant increase in total enzyme expression over the course of the evolution process. The enzyme contains a C-terminal V5 epitope, and the anti-actin antibody is used as a loading control. The relative expression is reported as the ratio of anti-V5 intensity to anti-actin intensity, normalized to yCDM1. The blot shown is representative of three independent experiments. (**B**) As the theophylline production increases, the apparent K_M decreases and the product selectivity (the ratio of the product theophylline to the side product paraxanthine) increases. The lines shown are a guide for the eye, and error bars show ± one standard deviation calculated from three independent experiments. (**C**) A homology model of yCDM1, indicating the initial mutations to yCDM1 (blue) and the new mutations in yCDM8 (green). Catalysis occurs at the upper face of the central heme group, and the substrate channel begins in roughly the upper-right corner. The mutations cluster around the active site and substrate channel. (**D**) Enzyme thermostability decreases as the enzymatic activity increases. The T_{50} measures the temperature at which a 10 minute incubation inactivates 50% of the enzyme population *in vitro*. Previous studies have shown that when the thermostability drops below ~ 44 ^{o}C (indicated by the red shading), further evolution becomes difficult. The T_{50} for wild-type BM3 is shown for comparison from previously published data (Fasan et al., 2007). The lines are a guide for the eye, and error bars show ± one standard deviation calculated from three independent experiments.

I next examined the location of the mutations to determine whether they were consistent with an increase in catalytic efficiency. The heme domain mutations generally cluster around the substrate channel and enzyme active site (Figure 2.10C), as I would expect for mutations that improve catalysis on a new substrate. For example, R47 has been implicated in binding to the carboxylate of the native fatty acid substrate, a role that is no longer necessary with caffeine as the substrate (Graham-Lorence et al., 1997). Similarly, F87 sits at the other end of the substrate channel and controls the positioning of the substrate near the heme (Graham-Lorence et al., 1997). This residue was mutated twice: to alanine in yCDM1 and to serine in yCDM8. S72 forms a portion of the binding pocket surrounding the heme. The initial mutations in CDM1 (Table 2.3) radically restructured the binding pocket, and the series of mutations to S72, first to phenylalanine and then to isoleucine, represent further modifications to optimize the pocket. The location of the mutations, in combination with the increased affinity and selectivity of the enzyme, strongly suggest that I have selected for improved catalysis rather than increased expression in *S. cerevisiae*.

The final caffeine demethylase, yCDM8, has a total of 14 mutations from wild-type, many of which are nonconservative. In general, most mutations are destabilizing (Bloom et al., 2006) and, as expected, the demethylase stability decreases during its evolution (Figure 2.10D). Previous studies have suggested that enzyme stability must stay above a threshold value in order to properly fold (Bloom et al., 2006). When the enzyme stability drops too low, further evolution becomes extremely difficult since mutations must increase activity without decreasing stability. BM3 unfolds irreversibly, so standard measures of protein folding energy are inappropriate. Instead, we can quantify the protein stability by the temperature at which a 10 minute incubation causes 50% of the enzyme to unfold, known as the T_{50}. For BM3, the threshold at which further evolution becomes difficult is estimated at a T_{50} of ~ 44 °C (Fasan et al., 2007). Starting with yCDM3, the thermostability of the evolved enzymes is sufficiently low as to restrict the accessible mutations. As long as the library sizes are limited by screening clonal populations in 96-well plates, low thermostability would require enzyme stabilization before further beneficial mutations are accessible (Fasan et al., 2007). I believe that low thermostability explains my inability to find improved mutants of yCDM5 when screening in 96-well plates. However, since screening by FACS allows much larger libraries to be investigated, increasingly rare beneficial mutations can still be identified even for an unstable enzyme. In support of this hypothesis, all of the mutations identified by the FACS-based screening strategy increase activity without decreasing stability.

2.2.6 Comparison to *in vitro* evolution

It was not initially evident that yCDM1 would produce sufficient theophylline to active the RNA switch *in vivo*. Therefore, in addition to screening variants of the caffeine demethylase in *S. cerevisiae*, I also attempted to evolve CDM1 *in vitro*, in *E. coli* cell lysate. While I ultimately wanted to increase the enzymatic activity in *S. cerevisiae*, established protocols were available for evolving the demethylase *in vitro*. I screened ~ 1,600 variants of CDM1 *in vitro*, identifying improved enzymes based on increased total turnover quantified by measuring the formaldehyde produced as a by-product of demethylation. I rescreened the top 35 hits *in vitro* in replicate, then cloned the best eight into a yeast expression vector to determine their activity *in vivo*. The best variant, a triple mutant K202Q/F331S/P346H denoted CDM2b, showed roughly a twofold increase in activity *in vitro* and a 1.3-fold increase in activity *in vivo* (Figure 2.11B). It is important to note that this mutant was identified as the best variant *in vivo*; there were enzyme variants that showed larger improvements *in vitro*, but these variants were less active *in vivo* than CDM2b (Figure 2.11A). Screening ~ 2,000 variants of CDM2b *in vitro*, rescreening 52 *in vitro*, and testing 7 *in vivo* identified a single additional beneficial mutation, N283I. The quadruple mutant, named CDM3b, showed an overall increase of 2.6-fold *in vitro* and 1.8-fold *in vivo* (Figure 2.11).

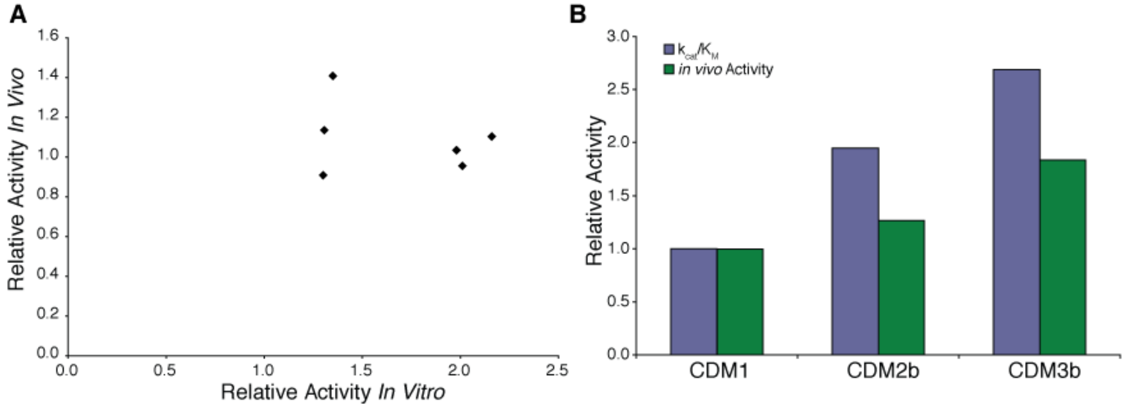

Figure 2.11 Summary of *in vitro* enzyme evolution. (**A**) Comparison of activity *in vitro* and *in vivo* for mutants of CDM2b. Several enzymes that were more active than parent *in vitro* were not improved *in vivo*. (**B**) Caffeine demethylases were screened in *E. coli* cell lysate, then cloned into a yeast expression vector to assay activity in live *S. cerevisiae*. As expected, the activity increased faster *in vitro* than *in vivo*.

These results illustrate a common difficulty in directed evolution, where a round of evolution improves the property that is screened for, rather than the property that is desired. I screened for activity *in vitro* but wanted activity *in vivo*. There is a correlation between those two properties, demonstrated by the presence of improved variants among the ~ 8 that I tested *in vivo*, but the correlation is poor. Many of the mutants that were improved *in vitro* were similar or worse than parent *in vivo*, and I undoubtedly missed many potential mutations that increased the enzyme activity *in vivo* but not *in vitro* (Figure 2.11A). The amino acid mutations that I identified *in vitro* do not overlap with the mutations identified by screening *in vivo*, consistent with the expectation that these two screening methods would select for traits that only partially overlap.

2.2.7 Increasing thermostability also increases mutational tolerance

While running further controls, I identified a serendipitous mutant that added two additional mutations, D80G/S332P, to CDM3b to yield CDM4b and a further 1.5-fold increase in activity *in vivo*. However, I was unable to find further mutations in this lineage that increased activity *in vivo*, despite screening ~ 4,000 variants among several different libraries. I hypothesized that low thermostability was limiting the available mutations. Consistent with this hypothesis, measurements of the enzyme thermostability showed that CDM4b had a T_{50} of ~ 42.5 °C (Figure 2.12A), low enough to limit the utility of screening colonies in well plates. I used site-directed mutagenesis to introduce a collection of mutations that had been shown to be thermostabilizing in a different lineage of BM3 mutants (Fasan et al., 2007). A combination of two additional mutations, adding I366V/E442K to produce CDM5b, increased the T_{50} from 42.5 °C to 47.0 °C, roughly the same T_{50} as CDM1 (Figure 2.11A).

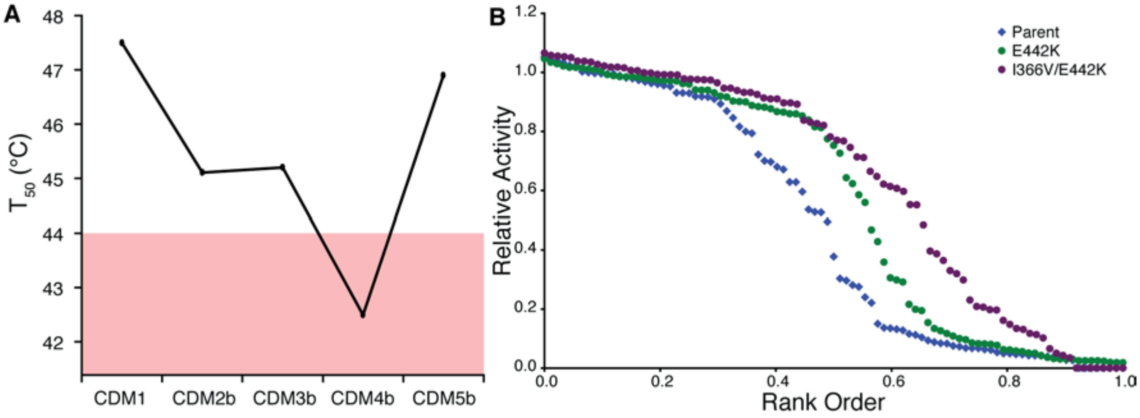

Figure 2.12 Increasing enzyme thermostability improves mutational tolerance *in vivo*. (**A**) T_{50} measurements for *in vitro* BM3 lineage. Enzyme evolution reduces the thermostability, but the rational introduction of I366V/E442K restores the initial stability. (**B**) Increasing thermostability increases mutational tolerance. A single mutagenic PCR was performed on residues 86-326 and transformed into three different vector backbones: CDM4b, CDM4b+E442K and CDM4b+E442K/I366V. 90 clones from each library were tested for activity relative to the unmutated backbone. Each point represents a single variant from the resulting library, rank ordered by relative activity. The stabilized enzymes can better tolerate mutations, reflected in the rightward shift of the distribution.

I next tested whether the increase in enzyme thermostability would also increase the mutational tolerance. I performed a single mutagenic PCR on residues 86-326 and transformed the resulting library into *S. cerevisiae in vivo* using gap repair to replace residues 86-326 of CDM4b, CDM4b+E442K, and CDM4b+I366V/E442K. I then measured the enzymatic activity of 90 members of each library, normalizing by the activity of the appropriate unmutated parent (Figure 2.12B). The distribution of mutations in each population should be the same, since they all are derived from a single error-prone PCR. Differences in the resulting distribution of activities reflects the ability of the remaining portion of the enzyme to tolerate those mutations. The stabilized enzymes can better tolerate destabilizing mutations, and as a result the libraries built from these enzymes show more folded, active clones. Unfortunately, the I366V/E442K mutations did not stabilize the enzymes from the BM3 lineage that had been evolved exclusively *in vivo*, indicating that the stabilizing mutations compensated for disruptions due to the specific set of mutations accumulated *in vitro*.

Finally, when I introduced further mutations to CDM5b, screening either *in vitro* or *in vivo* using well plates, I was unable to identify further improvements *in vivo*. I found mutants that showed improved activity *in vitro*, but despite the increased mutational tolerance of CDM5b none showed improved activity *in vivo*. Due to the success of the lineage screened exclusively *in vivo*, I chose to pursue that lineage instead.

2.3 Discussion

I have developed a novel high-throughput screen for enzyme activity *in vivo*. The core element of the *in vivo* screen is a synthetic RNA switch that connects the concentration of an enzymatic product to an easily screenable phenotype such as GFP fluorescence. Flow cytometry can then be used to distinguish small changes in fluorescence and therefore identify relatively small changes in activity. FACS allows screening of significantly larger libraries, ~ 10^7 rather than ~ 10^3, though with correspondingly lower accuracy. Using the two methods in combination allows efficient identification of improved mutants from a large library of enzyme variants. When this screening system was applied to optimize a novel caffeine demethylase, seven rounds of screening increased the enzyme activity *in vivo* by 33-

fold and the product selectivity by 22-fold.

It is important to note that, in contrast to many other screening platforms, this RNA-based screening system directly identified enzyme variants showing improved activity on the desired substrate and under the desired conditions. Since the RNA switch is highly specific for theophylline, I did not identify any false positive screening hits due to promiscuous production of the side product paraxanthine. Similarly, as the screening conditions are the same as the production conditions, potential screening hits need only be validated for consistency rather than tested with authentic substrates or new assay conditions. Correspondingly, the validation rate was extremely high; virtually all of the initial screening hits were at least as active as the parent enzyme, and the top enzyme(s) from the screening were typically the most improved variants after validation.

Additionally, by performing the enzyme screening *in vivo*, I impose a selective pressure against deleterious interactions between the enzyme and the host cell. This selective pressure may explain the selection for improved catalysis rather than increased expression. Expression of the enzyme from a high-copy plasmid placed significant stress on the host cell, reflected in a lower growth rate. Any further increase in enzyme expression would be counteracted by an increase in the cellular stress, producing little or no improvement in the product concentration. If screening had not been performed under the same conditions as production, this additional selective pressure to keep enzyme expression low might have been lost, leading to the identification of enzyme variants that were improved under the screening conditions but not under the production conditions. Similarly, evolving the enzyme *in vivo* was significantly more efficient than evolving it *in vitro*. While the enzyme's activity *in vitro* was correlated to its activity *in vivo*, the correlation was weak. I identified many false positives, where a mutation was beneficial *in vitro* but neutral or deleterious *in vivo*, and undoubtedly missed many false negatives, mutations that were beneficial *in vivo* but neutral or deleterious *in vitro*. Finally, the *in vitro* evolution plateaued after a ~ 4 fold increase in activity, compared to 33-fold *in vivo*. While I cannot draw any conclusions from a single evolutionary trajectory in each situation, I have seen no benefit to screening *in vitro* when an *in vivo* screen is available.

The *in vivo* biosensor is sensitive to a roughly 500-fold range of exogenous input concentrations, from ~ 10 μM to ~ 5 mM, though the output is no longer linear above ~ 1 mM. However, if the parent enzyme produces metabolite concentrations outside this range, either above or below, the sensor will not be able to identify improved variants. If the product concentration is too high, further screening will require reduced enzyme levels (Neuenschwander et al., 2007), lower substrate concentrations, or a modified RNA switch with a different affinity for the ligand (Zimmermann et al., 2000). Similarly, low levels of enzymatic activity require high substrate concentrations and precise screening of smaller libraries with a flow cytometer. For example, in order to maximize the signal from the biosensor, I used a substrate concentration that is close to the upper limit at which toxicity effects are observed for yeast. However, as long as the initial enzyme activity provides a small signal from the biosensor, the evolution process can provide a beneficial cycle whereby increased enzymatic activity makes the screening process more powerful and better screens then allow the identification of mutants with further increases in activity. The choice of screening conditions will determine the final enzyme properties, such as the apparent K_M and apparent v_{max}. For instance, screening at a lower substrate concentration will apply additional selective pressure to minimize the apparent K_M. By varying the selection conditions, either by choice or necessity, researchers can tune the properties of the resulting enzymes.

Modular RNA switches provide a flexible and generalizable screening platform for performing directed evolution *in vivo* and, in the future, will be used to develop advanced metabolic pathway optimization strategies. Since the synthetic RNA switches are modular, aptamers to desired metabolites can be selected *de novo* and integrated into existing switch platforms. For example, the theophylline aptamer used in this work was replaced with an aptamer to the polyketide tetracycline (Berens et al., 2001) to produce a tetracycline-responsive switch (Win & Smolke, 2007). While there are certain classes of molecules, such as short chain alkanes, that are unlikely to participate in specific binding interactions with RNA, aptamers have been selected against a wide range of metabolites and cofactors (Berens et al., 2001; Lorsch & Szostak, 1994; Manniròni et al., 1997; Sinha et al., 2010; Werstuck & Green, 1998; Win et al., 2006; Zimmermann et al., 2000), suggesting that there are many biosynthetic pathways to which this strategy may be applied. Additionally, since aptamers can be selected to discriminate between closely related molecules, multiple RNA switches, each responding to a different metabolite and controlling a different fluorescent reporter, could enable screening of several points along an engineered pathway or simultaneous screening for a desired product and against an undesirable side product.

However, these future applications may be more effective if they incorporate biosensors with a higher signal-to-noise, either using RNA switches with a greater fold change in output or using a nonlinear amplifier (Karig & Weiss, 2005) to increase the output of an existing RNA switch. In addition, the broad application of RNA switches to metabolic pathways will benefit from improved methods for rapidly selecting aptamers to new small molecules of interest, particularly aptamers that can distinguish between families of metabolites that differ by small functional groups. Finally, once an RNA switch is constructed and used to increase the activity of a metabolic pathway, researchers can take inspiration from natural riboswitches (Winkler et al., 2004) and adapt that same engineered RNA switch to dynamically regulate flux through the pathway in response to changing metabolite and cofactor levels by coupling the switch to the control of targeted pathway enzymes. Such dynamic control strategies can be used to design sophisticated synthetic metabolic networks that use cellular resources more efficiently (Chubukov et al., 2012; Zaslaver et al., 2004), minimize accumulation of toxic intermediates (Farmer & Liao, 2000), and improve the reliability of the engineered pathway (Bennett et al., 2008).

2.4 Methods

2.4.1 General molecular biology techniques

Restriction enzymes, T4 DNA ligase, and other cloning enzymes were obtained from New England Biolabs (Ipswich, MA). PfuUltraII (Agilent Technologies, Santa Clara, CA) was used for high-fidelity PCR amplification. Oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA) and the Stanford Protein and Nucleic Acid Facility (Stanford, CA). Standard molecular biology techniques were used for DNA manipulation (Sambrook & Russell, 2001). Ligation products were transformed into electrocompetent DH10B (Invitrogen, Carlsbad, CA; F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 deoR recA1 endA1 araD139 Δ (ara, leu)7697 galU galK λ-rpsL nupG) using a Gene Pulser Xcell System (Bio-Rad, Hercules, CA). Individual plasmids were transformed into yeast using standard lithium-acetate methods (Gietz & Woods, 2002). Escherichia coli were grown in LB media (BD, Franklin Lakes, NJ) with 100 μg/mL ampicillin (EMD Chemicals, Gibbstown, NJ). Yeast were grown in YPD or appropriate dropout media (Clontech, Mountain View, CA) with 2% glucose. Plasmids were prepped from overnight cultures of *E. coli* and *S. cerevisiae* using Spin Columns (Epoch Biolabs, Missouri City, TX) and Zymoprep Yeast Plasmid Miniprep II kit from Zymo Research (Irvine, CA), respectively, according to the manufacturers' instructions. Sequencing was performed by Laragen Inc. (Los Angeles, CA) and Elim Biopharmaceuticals (Hayward, CA). Caffeine, theophylline, and paraxanthine were obtained from Sigma-Aldrich (St. Louis, MO).

2.4.2 *In vitro* enzyme screening

In vitro enzyme screening in *E. coli* cell lysates was performed as described previously (Peters et al., 2003). I began with a panel of 95 enzymes that had been semi-rationally designed for hydroxylation of short chain alkanes (Chen et al., 2012). The original enzyme library had been tested for activity on caffeine during the initial screening. The subset of 95 enzymes used in this work had shown measurable caffeine demethylase activity *in vitro*. Cleared lysates were incubated with 10 mM caffeine and 2 mM NADPH (Gevo, Englewood, CO). The enzymatic products were assayed using HPLC to confirm that demethylation occurred at the 7-position to produce theophylline. Top hits were cloned into the high-copy 2μ *S. cerevisiae* expression vector for *in vivo* characterization of activity.

2.4.3 Construction of the *S. cerevisiae* enzyme expression vectors

The original BM3 variants were present on the *E. coli* expression vector pCWori. The coding regions were amplified by PCR using primers BM3-FromWori-FWD and BM3-FromWori-REV. The BM3 mutants were cloned between the EcoRI and NotI sites of a 2μ shuttle plasmid containing a URA marker (Hawkins & Smolke, 2008). Each gene was placed between a strong mutant TEF promoter (mutant #6) (Nevoigt et al., 2006) and a CYC1 terminator. Enzyme yCDM1 was generated by synthesis of a yeast codon optimized version of the CDM1 enzyme (GENEART, Regensburg, Germany). A V5 epitope tag and an additional six histidines were added to the C terminus of the synthesized enzyme for blotting and purification experiments. The synthesized enzyme was then cloned between the same EcoRI and NotI sites as CDM1. Enzyme yCDM6 was amplified from the 2μ shuttle plasmid using primers TEF-FWD and yCDM-CEN-REV and cloned between the EcoRI and AvrII sites in the centromeric (CEN6/ARSH4) shuttle plasmid pCS1585 (Liang JC, Chang AL, Kennedy AB, and CDS, in submission). During the recloning process, the CYC1 terminator was replaced with the ADH1 terminator. Plasmid maps for representative constructs used in this study are provided in Figure 2.13.

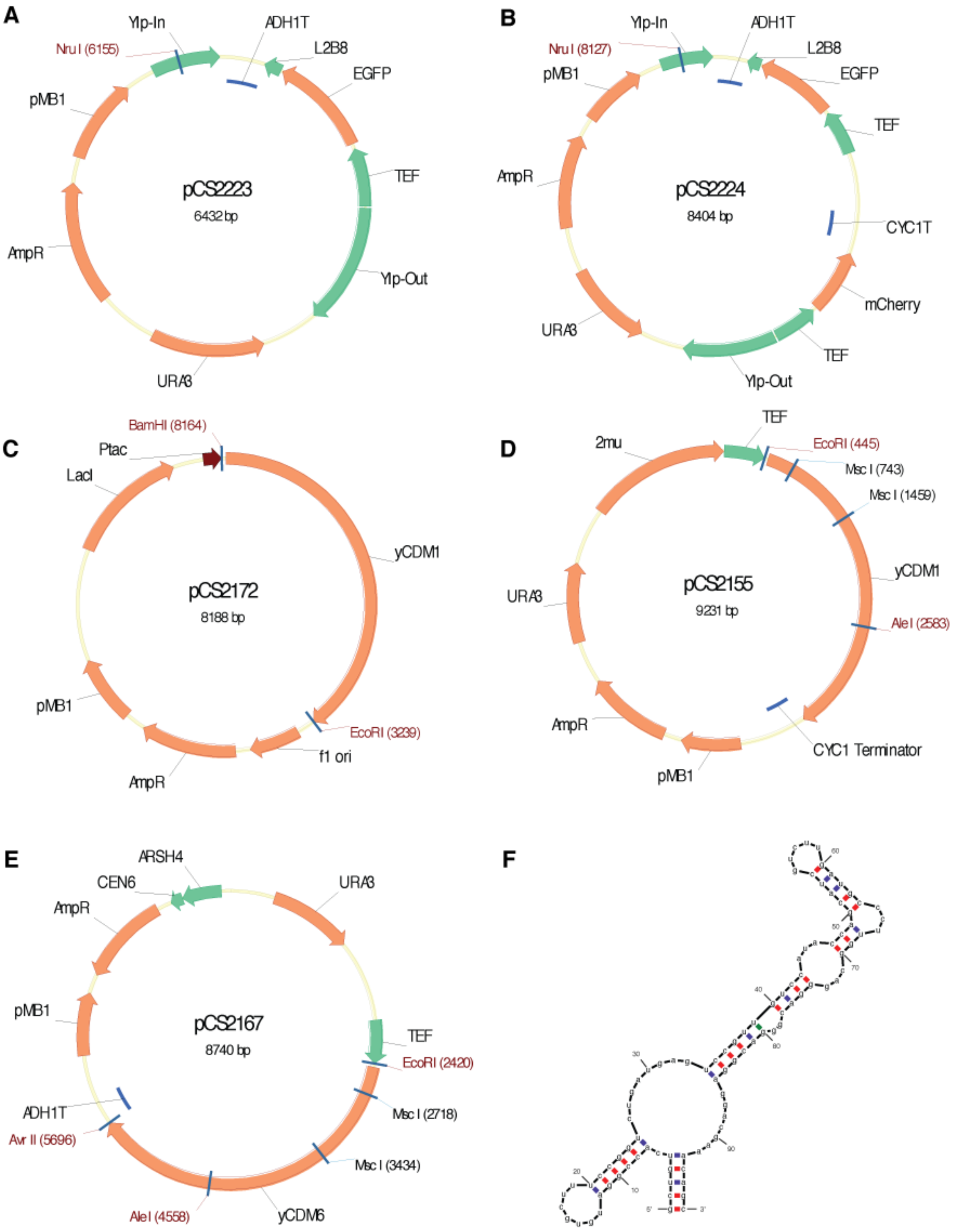

Figure 2.13 Plasmid maps. (**A**) pCS2223, the single-color integration vector used to construct CSY492. (**B**) pCS2224, the dual color integration vector used to construct CSY820. (**C**) pCS2172, the *E. coli* expression vector with yCDM1, used for T_{50} measurements. (**D**) pCS2155, the high-copy yeast expression vector with yCDM1. (**E**) pCS2167, the low-copy yeast expression vector with yCDM6. (**F**) The sequence and predicted structure of L2B8 (adapted from Win and Smolke, 2007)

2.4.4 Product biosensor yeast strain construction

The theophylline-sensing yeast strain CSY492 was constructed from W303α (MATα leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15). Previous work used a theophylline-responsive ON switch L2b8 behind yEGFP under the control of a GAL1/10 promoter in a centromeric plasmid (Win & Smolke, 2007). I first cloned the constitutive TEF1 promoter between EcoRI and SacI, replacing the GAL1/10 promoter. This expression construct, PTEF1-GFP-L2B8-ADH1T, was then cloned between the SacI and KpnI sites of the pIS385 yeast disintegrator vector (Sadowski et al., 2007). The resulting plasmid, pCS2223, was used to integrate the RNA switch-GFP construct into the lys2 locus of W303 and then remove the URA-selectable marker as previously described (Sadowski et al., 2007), producing strain CSY492. Alternately, I used a construct containing pTEF-mCherry-CYC1T upstream of the RNA switch-GFP expression cassette, pCS1748 (Liang JC, Chang AL, Kennedy AB, and CDS, in submission). This dual fluorescence construct was cloned into the disintegrator plasmid, resulting in pCS2224, and integrated into the W303 chromosome, using the same methods as before, producing strain CSY820.

2.4.5 Random mutagenesis

Briefly, error-prone PCR (epPCR) was performed using the GeneMorphII kit (Agilent Technologies). Libraries that were screened *in vivo* used primers yMutF and yMutR, while those screened *in vitro* were constructed using Heme MutF and Heme MutR. Typically, four epPCR reactions were performed for each library to produce a range of mutation rates.

For the *in vivo* screening, epPCR products were treated with DpnI, purified, and reamplified in 2x100 μL reactions per template using Pfu. For each epPCR, 1 μg of recipient vector was linearized with EcoRI, MscI, and AleI. Insert and vector were combined, phenol/chloroform extracted, ethanol precipitated, and then electroporated into CSY492 (Chao et al., 2006). Each transformation was plated at a range of dilutions to determine the transformation efficiency.

For FACS screening, the initial epPCR products were split. One fraction was amplified by PCR using Pfu, transformed by gap repair into CSY492, and 46 clones per epPCR were screened in 96-well plates using HPLC analysis. The epPCR product that produced the best distribution of activities, aiming for ~ 20% active enzymes, was then reamplified into 8x100 µL PCR reactions using Pfu and transformed by gap repair into CSY820. The remaining portion of the selected epPCR was then reamplified in 8x100 µL PCR reactions and transformed into CSY820, using the same method as described earlier, to construct the final sorting library.

For the *in vitro* screening, the epPCR reactions and pCWori-CDM1 plasmid were digested overnight with BamHI and SacI. The plasmid was treated with calf intestinal phosphatase and gel purified. The digested PCR products were cleaned up on a PCR column (Qiagen). The resulting insert and vector were ligated together and transformed into ElectroMax electrocompetent *E. coli* (Invitrogen). The resulting libraries were screened as described previously.

2.4.6 Library shuffling

Mutations from four round 2 variants (Supplementary Table 3) were shuffled to produce the round 3 library. These mutants encompassed a total of 12 mutations. Primers were designed that bound in the regions separating each adjacent mutation (Primers 30F through 663F, Supplementary Table 2). These primers, as well as their reverse complements (Primers 30R through 663R) were synthesized. For each mutation, the primers bracketing that mutation (for example, 30F and 60R) were used to amplify either the mutant template or parent. The resulting PCRs were cleaned up and mixed at a 2:1 ratio of parent to mutant fragment. The fragments were then stitched together by overlap PCR, amplified using yMutF and yMutR, and transformed into yeast as described above. In addition, mutations from four round 4 variants, comprising 9 mutations (Table 2.3), were shuffled together using a similar strategy to produce the library from which yCDM5 was identified.

2.4.7 Growth conditions for liquid culture *S. cerevisiae* assays

Yeast cultures were grown in 96-well plates (BD Falcon) using AeraSeal film (Excel Scientific, Victorville, CA) to allow for thorough aeration. Colonies were picked from agar plates with toothpicks, inoculated into 400 μL of SD-Ura media, and grown for 24 hours in a Kuhner LT-X plate shaker at 30 ºC, 480 RPM, a 1.24 cm orbital diameter, and 80% humidity. The cultures were subsequently backdiluted 100x into fresh SD-Ura with or without 1 mM caffeine and regrown for an additional 24 hours.

2.4.8 Flow-cytometry-based library screening

At the beginning of each screen, four libraries were constructed and tested to identify the library with the appropriate mutation rate. In these initial tests, 46 clones per transformation were grown as described previously and assayed for fluorescence and theophylline production. Based on the apparent mutation rates evident from the activity distribution of the populations one transformation was chosen for further screening, typically aiming for 50% of the population to retain significant activity.

When screening enzyme libraries, transformants were grown in 96-well plates and assayed for fluorescence. Yeast cultures harboring the enzyme expression vectors and integrated metabolite-sensing device were diluted 4x into water and assayed for fluorescence in 96-well plates on a Beckman Quanta flow cytometer. Cells were excited at 488 nm and GFP fluorescence was measured at 525 nm. Samples were gated first by electronic volume and side scatter to capture the cell population and then by fluorescence to remove the outliers with significantly low fluorescence. Approximately 8,000 cells were analyzed for each culture. The geometric mean fluorescence, normalized by the electronic volume, was then compared to the parental control. The brightest clones were selected for further screening by HPLC analysis. After a library was screened by both methods, the clones with the highest theophylline production (typically 10–35) were rescreened in triplicate by fluorescence, both with and without caffeine, and assayed for theophylline production. Consistent hits were sequenced and recloned by gap repair into fresh vector backbone to confirm that changes in activity were enzyme dependent. The top hit(s) after recloning were chosen as the template for the next round of evolution or shuffling as appropriate.

2.4.9 FACS-based library screening

Immediately following the gap repair transformation, the FACS library was diluted into 250 mL of SD-Ura. Dilutions of the transformation culture were plated on SD-Ura plates to determine the library size. The library culture was diluted into fresh SD-Ura every 24 hours. The initial dilution factor was tenfold, increasing to 30-fold and then 100-fold for the final presort dilution 24 hours prior to sorting. In a positive sort, caffeine was added to the growing culture to a final concentration of 1 mM during the final back-dilution. In a negative sort, the culture was grown in the absence of caffeine. Immediately prior to sorting, cells were centrifuged at 6,000 g for 5 minutes at 4 °C. The supernatant was discarded, the cells were resuspended in 1x phosphate buffered saline + 1% bovine serum albumin

(Fraction V, EMD Chemicals), stained with DAPI (Invitrogen), and filtered through a 40 μm cell strainer (BD Falcon). The cells were sorted using a BD Aria II sorter at the Stanford Shared FACS Facility. GFP was excited at 488 nm and captured by a 505 nm beam splitter and a 525/20 nm bandpass filter. mCherry was excited at 532 nm and captured by a 600 nm beam splitter and a 610/20 bandpass filter. DAPI was excited at 355 nm and captured by a 450/50 bandpass filter. Viable cells were isolated by gating on forward vs. side scatter, followed by a viability gate for DAPI negative cells. The cells resulting from one sort were grown overnight, then back-diluted 100-fold into the appropriate media (with or without substrate) and grown for 24 hours before the next sort.

2.4.10 HPLC methods

During screening, metabolite analysis was performed on an XDB-C18 2.1 x 50 mm, 3.5 μm column (Agilent Technologies). I injected 5 μL of sample onto the column. The mobile phase was 0.35 mL/min of 15% methanol/85% water with 0.1% acetic acid. Theophylline eluted at 1.65 minutes and was detected by UV absorbance at 274 nm. Culture supernatant from wild-type W303 showed no detectable peak. I switched to a Poroshell 120 SB-C18 2.1 x 50 mm, 2.7 μm column (Agilent) for the final two rounds of screening and the enzyme characterization, since the Poroshell column allowed shorter HPLC runs. The mobile phase was 0.50 mL/min of 20% methanol/80% water with 0.1% acetic acid. Using the Poroshell column, theophylline eluted at 0.70 minutes. For each sample, 3 μL was injected onto the columns. The identity of the theophylline peak was confirmed with each assay by the use of an authentic standard (Sigma-Aldrich), and the concentration of theophylline in each sample was determined by comparison to a series of reference standards.

2.4.11 Thermostability assays

Mutant CDM enzymes were moved from the *S. cerevisiae* expression vectors into pCWori for protein expression in *E. coli*. Enzymes were PCR amplified from the appropriate yeast vectors using primers yCDM-ToWori-FWD and yCDM-ToWori-REV and cloned between the BamHI and EcoRI sites of pCWori. Lysates of *E. coli* cells expressing the mutant CDM enzymes were prepared in the same fashion as for *in vitro* screening. Cleared lysate was incubated for 10 minutes at temperatures ranging from 30 °C to 55 °C and then cooled on ice. Reactions were set up with 140 μL of heat-treated lysate, 20 μL of 20 mM NADPH (Sigma), and 40 μL of 25 mM caffeine, incubated for 2 hours at room temperature, and centrifuged at 16,000 g for 10 minutes at 4 °C. The reactions were then assayed by HPLC for residual enzymatic activity. T_{50} values were calculated as the temperature at which 50% of the residual theophylline production remained following the 10 minute heat inactivation. Three technical replicates were conducted at each temperature for each enzyme. Each inactivation curve was fit to an Arrhenius model, and the fit was used to calculate the T_{50}. This entire process, beginning with fresh lysate, was performed three times. The reported T_{50} values are the average of the three independent measurements.

2.4.12 Determination of apparent kinetic constants

To determine the apparent K_M of the enzymes, each enzyme was grown overnight in 5 mL of SD-Ura. They were then backdiluted 100-fold into 96-well plates containing 400 μL of fresh SD-Ura with a range of caffeine concentrations (0.1, 0.2, 0.3, 0.4, 0.6, 0.8, and 1.0 mM) and grown for 24 hours. Three biological replicates were performed at each substrate concentration. A Michaelis-Menten curve was fit to the data using MATLAB (MathWorks, Natick, MA) to determine the apparent K_M and apparent v_{max}. This process, beginning with fresh overnight cultures, was performed three times and the reported kinetic constants are the average of the three independent measurements.

2.4.13 Western blots

Yeast strains harboring the appropriate enzyme expression constructs were grown overnight in 5 mL of SD-Ura. Protein extraction was carried out using 0.1 M NaOH (Kushnirov, 2000) followed by lysis in protein loading buffer (Invitrogen). Samples and ladder (New England Biolabs P7711S) were resolved on 4–12% Bis-Tris SDS-PAGE gels in 1x MOPS (Invitrogen). Protein was transferred to a nitrocellulose membrane using semidry transfer (Bio-Rad) in 2x NuPAGE transfer buffer (Invitrogen) + 10% MeOH. After transfer, the membrane was cut in half at ~ 55 kDa. Both membrane halves were blocked in 5% BSA for 1 hour. The membrane with higher-molecular-weight proteins was blotted with an anti-V5 HRP antibody according to the manufacturer's instructions (Invitrogen). The membrane with lower-molecular-weight proteins was blotted with a mouse anti-actin antibody (Abcam 8224, Cambridge, UK) and a rabbit anti-mouse HRP (Abcam 6728) according to the manufacturer's instructions. Both HRP antibodies were detected by chemiluminescence, following the manufacturer's instructions, (Pierce, Rockford, IL) using a Chemi-Doc XRS imager (Bio-Rad). Blots were analyzed using the QuantityOne analysis software (Bio-Rad).

2.4.14 Enzyme stabilization

Site directed mutagenesis was used to introduce potentially stabilizing point mutations into CDM4b. Three mutations previously shown to be stabilizing (L52I, I366V, E442K) and one reversion (H346P) were introduced singly and in combination using the Quikchange mutagenesis method with PfuUltra (Agilent) and the appropriate primers listed below. Mutagenesis was confirmed using the respective screening primer, designed to have the 3' base match the newly mutated residue. Each PCR screening reaction was optimized to ensure that unmutated DNA gave poor or no amplification while the correct mutated residue was more strongly amplified.

Once the mutants were constructed, error-prone PCR was conducted on a single CDM4b template using primers Stability TestF and Stability TestR to mutate residues 86-326. The resulting PCR product was then transformed into each mutant backbone in *S. cerevisiae* using gap repair as described previously. 90 members were picked from each library and compared to the respective unmutated backbone (e.g., the CDM4b+E442K library was compared to the CDM4b+E442K parent). The distribution of activities in the resulting library showed the extent to which the mutations increased the enzyme's mutational tolerance *in vivo*.

2.5 Tables

Table 2.1 Primers used in this chapter

Primer Name	**Primer Sequence**
BM3-FromWori-FWD	5’-TATAGAATTCGATATCAAGCTTGGAGATCTAAAAGAAAACAATGACAATTAAAGAAATGCCTCAG-3’
BM3-FromWori-REV	5’-CTATGCGGCCGCTCACCCAGCCCACACGTCTTTTG-3’
TEF-FWD	5’-ACTTCTTGCTCATTAGAAAGAAAGC-3’
yCDM-CEN-REV	5’-TATACCTAGGCTTCAATGGTGGTGGTGATGG-3’
yCDM-ToWori-FWD	5’-AATTGGATCCATCGATGCTTAGGAGGTCATATGTCTATCAAAGAAATGCCAC-3’
yCDM-ToWori-REV	5’-TAATGAATTCTCAATGGTGGTGGTGATGGTG-3’
yMutF	5’-TCTTGCTCATTAGAAAGAAAGCATAGCAATCTAATCTAAGTTTTAATTAC-3’
yMutR	5’-AATCTAGCAGTAACTCTGTTGACGATACCTTCGTAGTTTCTTGGAATAAC-3’
30R	5’-CTTAAAGATTTCACCCAATTCGTCAGCAATTTTCATC-3’
30F	5’-GATGAAAATTGCTGACGAATTGGGTGAAATCTTTAAG-3’
60R	5’-CAAGTTCTTGTCGAATCTAGATTCATCACAAGCTTCC-3’
60F	5’-GGAAGCTTGTGATGAATCTAGATTCGACAAGAACTTG-3’
61R	5’-CAAGTTCTTGTCGAATCTAGATTCATCACAAGC-3’
61F	5’-GCTTGTGATGAATCTAGATTCGACAAGAACTTG-3’
85R	5’-CAGTTCTTTTCGTGGGTCCAGGAAGTGGCCAAACCG-3’
85F	5’-CGGTTTGGCCACTTCCTGGACCCACGAAAAGAACTG-3’
216R	5’-GGAGGCCTTTCTGTCAGCGATGATCTTGTCAAC-3’
216F	5’-GTTGACAAGATCATCGCTGACAGAAAGGCCTCC-3’
329R	5’-GTGTCTTCCTTAGCGTACAAAGAGAACCATGG-3’
329F	5’-CCATGGTTCTCTTTGTACGCTAAGGAAGACAC-3’
341R	5’-CACCCTTTTCCAATGGGTATTCACCACCCAAG-3’
341F	5’-CTTGGGTGGTGAATACCCATTGGAAAAGGGTG-3’
396R	5’-GCGAATTGTTGACCGATACAGGCTCTTTGACCG-3’
396F	5’-CGGTCAAAGAGCCTGTATCGGTCAACAATTCGC-3’
481R	5’-CCATGTTAGAACCGTACAAAACCAACAATG-3’
481F	5’-CATTGTTGGTTTTGTACGGTTCTAACATGG-3’
535R	5’-GCGTTATCTGCTGGATGACCGTTGTAGG-3’
535F	5’-CCTACAACGGTCATCCAGCAGATAACGC-3’
570R	5’-CCCAGTTTTTATCACCA-3’
570F	5’-TGGTGATAAAAACTGGG-3’
586R	5’-CACCCTTAGCAGCCAAAGTTTCGTC-3’
586F	5’-GACGAAACTTTGGCTGCTAAGGGTG-3’
663R	5’-CAATTCCTTGGAGGCAACGACGTTGGTAGAG-3’
663F	5’-CTCTACCAACGTCGTTGCCTCCAAGGAATTG-3’
Heme MutF	5’- CACAGGAAACAGGATCCATCGTGCTTAGG-3’
Heme MutR	5’-CTAGGTGAAGGAATACCGCCAAGCGGA-3’
Stability TestF	5’-AAGAACTTGTCTCAATGGTTGAAGTTCATTAGAGGTTTCTTGGGTGACGG-3’
Stability TestR	5’-CCCAAGACAGTGTCTTCCTTAGCGTACAATGGGGACCATGGGAAAGTTGG-3’
L52I SDM FWD	5’-ccaggtagagttaccagatacAtCtcttcccaaagattgattaagg-3’
L52I SDM REV	5’-CCTTAATCAATCTTTGGGAAGAGATGTATCTGGTAAC

	TCTACCTGG-3’
H346 SDM FWD	5’-ctgtcttgggtggtgaataccCAttggaaaagggtgacgaattg-3’
H346P SDM REV	5’-CAATTCGTCACCCTTTTCCAATGGGTATTCACCACCCAAGACAG-3’
I366V SDM FWD	5’-cacaattgcacagagacaaaaccGtctggggtgacgatgttg-3’
I366V SDM REV	5’-CAACATCGTCACCCCAGACGGTTTTGTCTCTGTGCAATTGTG-3’
E442K SDM FWD	5’ggaaaccttgaccttaaaaccaAaGggtttcgttgtcaag-3’
E442K SDM REV	5’-CTTGACAACGAAACCCTTTGGTTTTAAGGTCAAGGTTTCC-3’
L52I cPCR FWD	5’ccaggtagagttaccagatacAtC-3’
H346P cPCR FWD	5’ggtggtgaataccCA-3’
I366V cPCR FWD	5’-cacagagacaaaaccG-3’
E442K cPCR FWD	5’-cttgaccttaaaaccaAaG-3’

Table 2.2 Plasmids and strains constructed in this chapter

Strain	Genotype
W303	MATα *leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15*
CSY492	W303 *lys2::*P_{TEF}*-GFP-L2Bulge8-ADH1T*
CSY820	W303 *lys2::* P_{TEF}*-mCherry-CYC1T-*P_{TEF}*-GFP-L2Bulge8-ADH1T*
pCS2155	2μ URA P_{TEF}-yCDM1
pCS2156	2μ URA P_{TEF}-yCDM2a
pCS2157	2μ URA P_{TEF}-yCDM2b
pCS2158	2μ URA P_{TEF}-yCDM2c
pCS2159	2μ URA P_{TEF}-yCDM2d
pCS2160	2μ URA P_{TEF}-yCDM3
pCS2161	2μ URA P_{TEF}-yCDM4a
pCS2162	2μ URA P_{TEF}-yCDM4b
pCS2163	2μ URA P_{TEF}-yCDM4c
pCS2164	2μ URA P_{TEF}-yCDM4d
pCS2165	2μ URA P_{TEF}-yCDM5
pCS2166	2μ URA P_{TEF}-yCDM6
pCS2167	Centromeric URA P_{TEF}-yCDM6
pCS2168	Centromeric URA P_{TEF}-yCDM7
pCS2169	Centromeric URA P_{TEF}-yCDM8
pCS2170	2μ URA P_{TEF}-yCDM1 (A264H)
pCS2172	pCWori + yCDM1
pCS2173	pCWori + yCDM3
pCS2174	pCWori + yCDM5
pCS2175	pCWori + yCDM6
pCS2176	pCWori + yCDM7
pCS2177	pCWori + yCDM8
pCS2223	pIS385 + P_{TEF}-GFP-L2Bulge8-ADH1T
pCS2224	pIS385 + P_{TEF}-mCherry-CYC1T-P_{TEF}-GFP-L2Bulge8-ADH1T

CSY821	CSY492+pCS2155
CSY822	CSY492+pCS2160
CSY823	CSY492+pCS2165
CSY824	CSY492+pCS2166
CSY825	CSY492+pCS2167
CSY826	CSY492+pCS2168
CSY827	CSY492+pCS2169
CSY828	CSY492+pCS2170
CSY829	CSY492+pCS2171
CSY830	CSY492+pCS4 (empty centromeric plasmid)
CSY831	CSY492+pCS31 (empty 2µ plasmid)

Table 2.3 Mutations in the BM3 enzyme variants generated in this chapter

Enzyme	**Mutations**
yCDM1	A74W, V78I, A82L, F87A, M185V, L188W, A328F, A330W
yCDM2a	yCDM1 + I58T, P461L, A575V
yCDM2b	yCDM1 + N522S, C569Y
yCDM2c	yCDM1 + T22R, D194N, Q387R, A603T
yCDM2d	yCDM1 + S72F, P301L, G457D
yCDM3	yCDM1 + S72F, A603T
yCDM4a	yCDM3 + M354L, T576R, Q673K
yCDM4b	yCDM3 + F72I, T339I
yCDM4c	yCDM3 + R47S
yCDM4d	yCDM3 + Q27H, G660D
yCDM5	yCDM3 + Q27H, R47S, F72I
yCDM6	yCDM5 + E435G
yCDM7	yCDM6 + I174V
yCDM8	yCDM7+A87S
CDM2b	CDM1 + K202Q, F331S, P346H
CDM3b	CDM2b + N283I
CDM4b	CDM3b + D80G, S332P
CDM5b	CDM4b + I366V, E442K

Table 2.4 Summary of functional characterization data for enzyme variants

Enzyme	Relative $v_{max,\,app}/K_{M,\,app}$	$K_{M,\,app}$ (mM)	T_{50} (°C)	Selectivity
yCDM1	1.0	1.5 ±0.1	47.0 ± 0.6	10.3 ± 0.5
yCDM3	3.9 ±0.4	1.1 ±0.1	42.9 ± 0.9	14 ± 1
yCDM5	12.2 ±1.0	0.75 ±0.10	41.5 ± 1.3	23 ± 3
yCDM6	22.1 ±1.6	0.59 ±0.01	42.8 ± 1.0	100 ±25
yCDM7	26.6 ±3.1	0.74 ±0.02	42.1 ± 1.3	175 ± 4
yCDM8	33.0 ±4.2	0.69 ±0.04	43.1 ± 0.7	230 ±20

2.6 Acknowledgements

I would like to thank Mike Chen for building and characterizing the original library from which the CDM1 demethylase was isolated and for assistance in troubleshooting the initial stages of the project; Frances Arnold for serving as an invaluable resource on the directed evolution process; Andrew Sawayama for providing an introduction to directed evolution techniques; Kristy Hawkins for assisting with the design of the yeast expression system; Joe Liang for aiding in the FACS screening; Derek Macklin for cloning and characterizing several of the enzymes in new expression plasmids; and Chris Snow for providing the homology model of CDM1. This work was supported by the National Science Foundation (CBET-0917638; fellowship to JKM), the Institute for Collaborative Biotechnologies (U.S. Army Research Office DAAD19-03-D-0004), and the Alfred P. Sloan Foundation (fellowship to CDS).

3 Use of a Synthetic RNA Switch for Enzyme Discovery

Many biosynthetic pathways, including those of medically relevant small molecules, contain reactions for which no enzyme has yet been identified. Even after an enzyme has been identified to catalyze a particular reaction, further enzyme discovery might produce a new enzyme with better properties for use in a heterologous pathway. Advances in DNA sequencing and synthesis mean that it will become increasingly easy to identify and express candidate enzymes. However, the sheer volume of the resulting candidates will overwhelm our ability to characterize all of the potentially relevant enzymes. New functional screens will be necessary to sort through the list of candidate enzymes to identify the best variant. In this chapter, I describe the use of synthetic RNA switches to perform functional screening of a plant cDNA library. The plant, *Coffea dewevrei*, is known to enzymatically demethylate caffeine to theophylline. I used a theophylline-responsive RNA switch to screen for members of a *C. dewevrei* cDNA library that were capable of producing theophylline when heterologously expressed in *S. cerevisiae* and fed the substrate, caffeine. Unfortunately, I was unable to identify any candidate enzymes, and I discuss possible reasons for this lack of success.

3.1 Introduction

Caffeine is a classic example of a plant-derived natural product and has been in common use, mainly in the form of coffee and tea, for centuries. The typical commercial coffee species, *Coffea arabica*, has been bred for its caffeine production and can accumulate caffeine in its mature fruit to ~ 1% by dry weight (Ashihara et al., 1996). However, decaffeinated coffee has claimed a significant fraction of the coffee market, and as a result there is now interest in developing naturally decaffeinated species of coffee (Ogita et al., 2003). In addition to reducing caffeine content by decreasing the rate of caffeine synthesis, the same result could be achieved by increasing the rate of caffeine catabolism. However, in *C. arabica* the rate of caffeine catabolism is very low (Ashihara et al., 1996). Several coffee species, such as *Coffea dewevrei* (Mazzafera et al., 1994) and *Coffea eugenioides* (Ashihara & Crozier, 1999), have naturally low caffeine concentrations in the mature fruit due to rapid catabolism of caffeine, with demethylation of caffeine to theophylline thought to be the rate-limiting step. While caffeine degradation can readily be shown to occur in these species, the enzyme responsible for the catabolism has not been identified. If the relevant enzyme could be identified, overexpression of the associated gene in *C. arabica* could further lower the caffeine concentration in transgenic plants.

Unfortunately, sequencing-based approaches for enzyme discovery from plants can be difficult. First, plant genome sequencing can be difficult, owing to the size and repetitive structure of many plant genomes (Feuillet et al., 2011). Instead, researchers often must turn to lower-quality EST libraries (Facchini et al., 2011; Mondego et al., 2011). Second, plant genomes often contain multiple members of a given enzyme family, complicating attempts to screen by sequence homology. For example, approximately 2% of the coding sequences from an EST library of *C. arabica* consist of members of the P450 superfamily (Mondego et al., 2011). Finally, the biochemical evidence is often inconclusive. The available evidence suggests that caffeine is demethylated in *Coffea* species by a P450 monooxygenase (Mazzafera, 2004). However, other studies using selective enzyme inhibitors point towards a flavin monooxygenase as the relevant enzyme (Paulo Mazzafera, personal communication). Without a precise identification of the gene family, the search must be broadened to other enzyme classes, reducing the utility of the sequence-based approach. When sequence-based approaches are ineffective, functional screening of cDNA libraries can be used for enzyme discovery (Uchiyama & Miyazaki, 2009).

We have previously demonstrated our ability to screen large enzyme libraries in *S. cerevisiae* to identify enzyme variants capable of demethylating caffeine to theophylline. The same screening platform can be used for functional screening of a cDNA library, simply by replacing the library of enzyme mutants with a cDNA pool (Figure 3.1). *S. cerevisiae* is a good host organism for heterologous screening of plant cDNA libraries, as it has the appropriate membranes for proper expression of membrane-bound proteins. Total RNA samples from the leaves of *C. arabica* and *C. dewevrei* were provided by researchers at the Campinas State University in Brazil. cDNA library construction went smoothly, and sequencing results confirmed the presence of full-length *C. dewevrei* cDNAs. However, screening three separate cDNA libraries failed to identify a caffeine demethylase. There are several possible reasons for this lack of success, including the difficulty of functional expression of plant P450s, the limitations due to scaling problems involved in constructing and screening cDNA libraries, and the sensitivity of the RNA switch used in the screen.

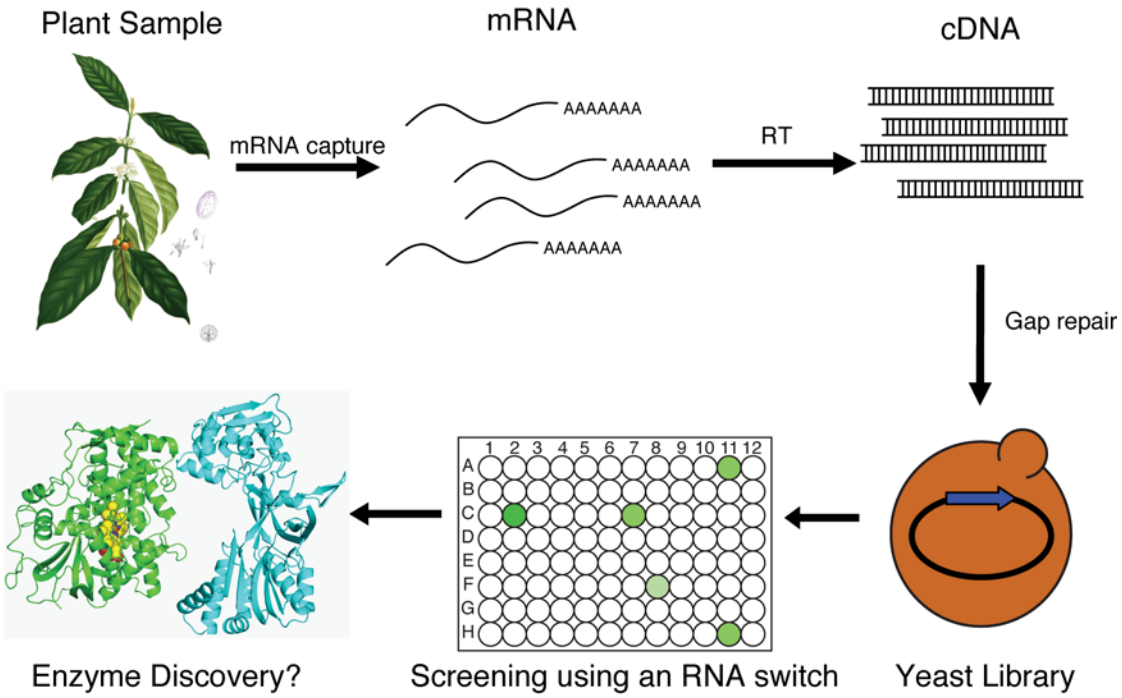

Figure 3.1 Overview of the enzyme discovery process. mRNA is extracted from plant samples, reverse transcribed into cDNA, and transformed by gap repair into *S. cerevisiae*. As described in Chapter 2, the resulting cDNA library can be functionally screened for caffeine demethylase activity using a theophylline-responsive RNA switch.

3.2 Results

3.2.1 cDNA library construction

Total RNA from *C. arabica* and *C. dewevrei* were provided by Professor Paulo Mazzafera at the Campinas State University, Brazil. Poly-A$^+$ mRNA was purified from the *C. dewevrei* sample using oligo-dT magnetic beads. The poly-A$^+$ mRNA was then reverse transcribed using a template switching technique to enrich for full-length cDNAs (Zhu et al., 2001b). The forward and reverse primers included homology to the yeast expression plasmid to facilitate cloning by gap repair. The first strand cDNA was amplified using PCR (Figure 3.2). The PCR products were transformed into a yeast expression vector in *S. cerevisiae*, producing ~ 10^4 clones.

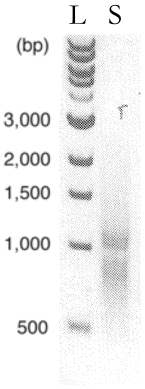

Figure 3.2 cDNA library construction. Poly-A+ RNA was purified from *C. dewevrei* total RNA. A template-switching reverse transcriptase was used to selectively amplify full-length mRNA templates. The resulting first-strand cDNA was amplified using PCR to add homology regions that allowed gap repair library construction in *S. cerevisiae*. The distinct band at ~ 1 kb likely corresponds to the cDNA of a highly expressed chitinase. L = DNA ladder. S = cDNA sample

Twelve members of the cDNA library were randomly selected and sequenced. Of the twelve, four showed neither an open reading frame nor a poly-A tail. Two showed significant homology to a *C. arabica* chitinase known to be highly expressed as a fungal defense mechanism (Guerra-Guimarães et al., 2009). The remaining six sequences were full-length cDNAs that showed homology to plant genes, frequently with *Vitis vinifera* as the nearest homolog (Mondego et al., 2011). The cDNAs consisted of two separate ribosomal proteins, a dehydrogenase, a histone, a transcription factor, and a lipid transferase. The sequencing results confirmed that the library contained full-length cDNAs from *C. dewevrei*.

3.2.1 *In vivo* cDNA screening

After optimization of the cDNA construction process, a new library of 10^5 clones was transformed into the screening strain CSY820. As described in section 2.2.4, this strain expresses GFP under the control of a theophylline-dependent RNA switch as well as constitutively expressing mCherry. Theophylline production can be detected *in vivo* at the single-cell level using FACS to identify cells with an increase in the ratio of GFP to mCherry fluorescence. 2×10^6 cells were screened by FACS in the presence of 1 mM caffeine, selecting for the top 1% by normalized fluorescence. The resulting population was grown in the absence of caffeine and sorted for the ~ 20% of the population with background levels of fluorescence. These cells were split and grown in both the presence and absence of 1 mM caffeine. Comparing the resulting fluorescence distributions, the population grown in 1 mM caffeine showed a slight increase in normalized GFP fluorescence (Figure 3.3). The culture grown in 1 mM caffeine was sorted again to collect the top 1% by normalized fluorescence. Cells were isolated on agar plates, picked into liquid culture in a 96-well plate, and assayed for theophylline production in the presence of 1 mM caffeine. None of the 92 cultures showed any theophylline production.

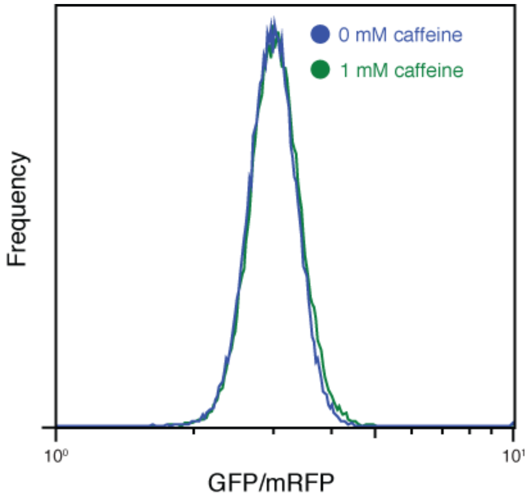

Figure 3.3 Fluorescence histograms of the cDNA library. After two rounds of sorting (one positive, one negative), the resulting library was grown in the presence (green) and absence (blue) of caffeine. The culture grown with caffeine showed a slight increase in fluorescence, possibly indicating the presence of a caffeine demethylase in the cDNA library.

Expecting that the caffeine demethylase mRNA might constitute a small fraction of the total mRNA in *C. dewevrei*, I constructed a new cDNA library using on-bead subtractive hybridization to enrich for mRNAs overexpressed in the high-theophylline *C. dewevrei* relative to the low-theophylline *C. arabica* (Figure 3.4A). After transformation into *S. cerevisiae*, the resulting library of ~ 10^5 clones was screened as before. Again, the final sorting library showed an increase in fluorescence when grown in the presence of caffeine (Figure 3.4B).

368 colonies from the unsorted library and 1104 from the sorted library were assayed for fluorescence in 96-well plates using flow cytometry. Any clone that showed an increase in fluorescence relative to the negative control was measured for theophylline production by liquid chromatography. None of the assayed clones produced theophylline.

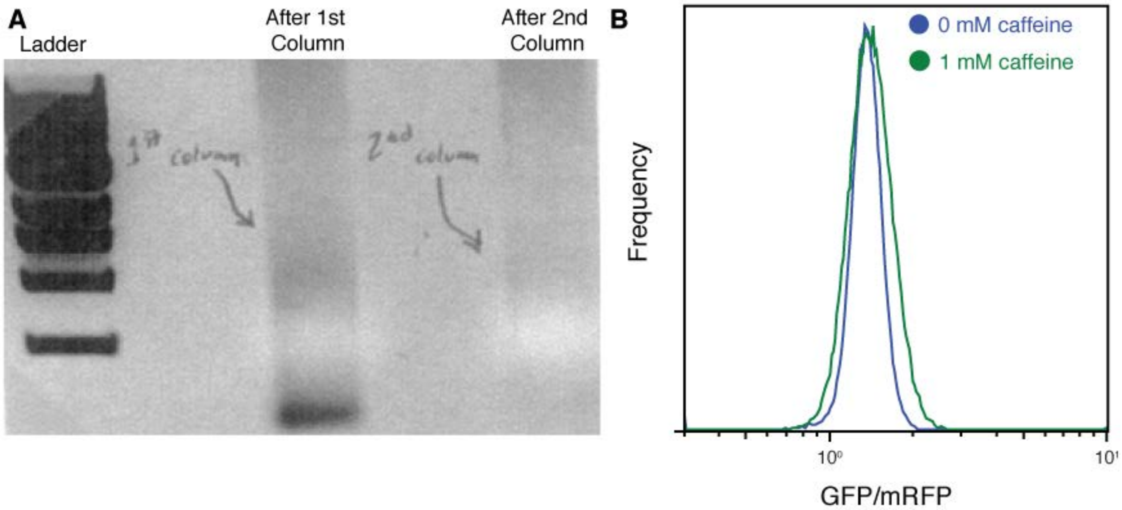

Figure 3.4 Construction and screening of a cDNA library after subtractive hybridization. (**A**) After subtractive hybridization, the cDNA library has a large population of short (< 500 bp) fragments. A size-exclusion column is used to selectively eliminate these fragments while retaining the longer (full-length) fragments. (**B**) After two rounds of sorting (one positive, one negative), the resulting library was grown in the presence (green) or absence (blue) of caffeine. The culture grown in caffeine showed an increase in fluorescence, suggesting the presence of theophylline-producing enzymes. After a further positive sort, the cells were isolated on agar plates and screened in clonal culture.

Hypothesizing that the caffeine demethylase might be a cytochrome P450, I tested a version of CSY820 with an integrated copy of the *Arabidopsis thaliana* cytochrome P450 reductase. I synthesized fresh cDNA, again using subtractive hybridization to enrich for mRNAs that were overexpressed relative to *C. arabica*, and produced a library of ~ 10^5 clones in *S. cerevisiae*. The library was sorted as described previously. Approximately 1,400 of the sorted clones were rescreened by fluorescence in 96-well plates. As before, any clone showing an increase in fluorescence was assayed for theophylline production by liquid chromatography, but no theophylline production was observed.

3.3 Discussion

Despite carefully screening three separate cDNA libraries, I was unsuccessful in identifying a caffeine demethylase. There are several potential explanations for this lack of success. The obvious explanation is that the enzyme simply was not present in the libraries that I screened. Sequencing results suggested that a majority of the cDNAs in the unsubstracted library were full length, though similar sequencing was not performed on the subtracted cDNA library. However, particularly large or small cDNAs might have been lost during the library construction process. Additionally, the PCR amplification process introduces a bias into the resulting library. If the caffeine demethylase was poorly amplified, it might be absent from the resulting library. Finally, the library sizes of 10^5 might have been insufficient to cover the *C. dewevrei* cDNA population. I used subtractive hybridization in an attempt to reduce the cDNA library diversity to a level suitable for screening. However, there are drawbacks to the use of subtractive hybridization. Perhaps mRNA expression of the caffeine demethylase is similar in *C. arabica* and *C. dewevrei*, and the difference in caffeine demethylase activity is due to post-transcriptional control or specific enzyme activity. In that case, subtractive hybridization would simply remove the demethylase mRNAs and screening would be unsuccessful.

An alternate possibility is that the desired enzymatic activity is not actually present in *C. dewevrei*. While the available evidence suggests that the primary pathway for caffeine catabolism in *Coffea* proceeds through theophylline (Suzuki & Waller, 1984), the evidence is not conclusive. Plant secondary metabolism is perhaps better viewed as a network than a set of discrete pathways, making analysis difficult. If, for example, the primary catabolic pathway is to demethylate caffeine to theobromine, no cDNA would be capable of theophylline production.

Finally, there is a strong suspicion that the first demethylation is catalyzed by a P450 monooxygenase (Mazzafera, 2004). Heterologous expression of P450 monooxygenases from plants is particularly difficult, as these enzymes are typically membrane associated and require the presence of an additional reductase domain (Mizutani & Ohta, 2010). While *S. cerevisiae* is a preferred host for heterologous plant P450 expression (Urban et al., 1994), functional expression can still be difficult. The P450 may not express well or may not efficiently receive electrons from the native yeast reductase partner. While I screened the cDNA library in a strain that overexpresses the *A. thaliana* P450 reductase, the caffeine demethylase might be specific for its native *C. dewevrei* reductase. The detection limit for the RNA switch that I used in this work is ~ 10 µM in well plates and ~ 100 µM for FACS. The target enzyme might be both present and active, but simply not sufficiently active to be detected in my screen.

Despite my inability to isolate a caffeine demethylase, I have successfully developed procedures for the functional screening of cDNA libraries in *S. cerevisiae* using a synthetic RNA switch. Aside from the theophylline aptamer used in the RNA switch, nothing in this method is specific for this organism or reaction, so I expect that the same procedure could be used for enzyme discovery of other uncharacterized enzymatic activities. There are several potential changes to my procedure that might improve the future chances of success. The most important would be to target enzymes with a high probability of functional expression in *S. cerevisiae*. Yeast are the best available host for heterologous P450 expression, but even still many P450s fail to show significant activity.

Additionally, there is a trade-off between the sensitivity of screening in well plates and the speed of screening in single cells. Plate-based screening can detect enzymes with low activity, but requires small libraries and therefore an enrichment technique like subtractive hybridization. Unfortunately, subtractive hybridization might erroneously eliminate the target enzyme from the library. Avoiding subtractive hybridization would require screening larger libraries in order to identify rare cDNAs without prior enrichment. FACS-based screening can easily screen libraries of 10^7 and could be extended to libraries of 10^8 if needed. My library sizes were typically in the range of 10^5, so the library construction process would need to be optimized in order to cover the full range of cDNAs. However, the use of a FACS-based screen requires either that the enzyme display a higher level of heterologous activity or the use of more-sensitive switches. Switch optimization could greatly improve the efficacy of a function screen (Figure 3.5). Decreasing the EC_{50} of the switch would directly increase the sensitivity of the screen. Similarly, the dynamic range of the switch determines the relationship between the EC_{50} and the minimum detectable concentration. A switch with a larger dynamic range can detect lower concentrations, relative to the switch EC_{50}, than a switch with a smaller dynamic range. Combining a more-tractable target, a more-sensitive switch, and an optimized library construction protocol would greatly increase the chances of success.

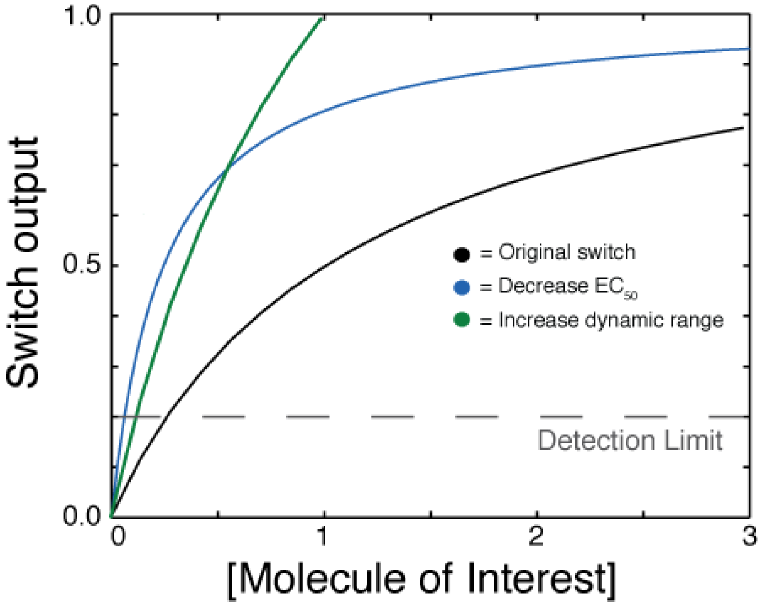

Figure 3.5 A schematic of the effect of switch optimization on metabolite detection. The switch output is plotted as a function of the ligand concentration (given in arbitrary units). A given screen has a detection limit, below which the noise in the sensor prevents differentiation of active and inactive enzymes. The characteristics of the switch will determine the concentration at which this detection limit is crossed. Starting from an initial switch (black), decreasing the EC$_{50}$ (blue) or increasing the dynamic range (green) would produce a switch that crosses the detection limit at a lower concentration of the target molecule. Therefore, these improved switches would be capable of detecting enzymes with lower levels of activity.

The main limitation of this technique is the requirement for an RNA sensor to the desired enzymatic product. Existing RNA selection strategies require significant quantities of the target molecule, and many metabolic intermediates are not commercially available. In some cases, the desired intermediate may be available. In other cases, this strategy might be more easily applied to identify replacement enzymes in a pathway that shows limited, but nonzero, activity. For example, a promiscuous and marginally active enzyme, like CYP2D6 (Hawkins & Smolke, 2008), might be used to construct an initial pathway and produce enough of the target molecule for sensor construction. The sensor could then be used to screen cDNA libraries to identify enzymes from the native pathway that show higher activity and selectivity.

Functional screens for enzyme activity remain an important avenue for further research. Despite recent improvements in computational prediction of enzyme activity and specificity (Lukk et al., 2012), accurately predicting the substrate of an arbitrary enzyme is an uncertain proposal requiring significant effort. While my attempt to use a functional screen to identify a caffeine demethylase from a low-caffeine strain of *Coffea* was unsuccessful, I believe that the fundamental process was valid and could be successful on a more-tractable target.

3.4 Methods

3.4.1 cDNA synthesis

Total RNA extracted from *C. dewevrei* and *C. arabica* leaves was shipped from Brazil precipitated in ethanol at room temperature. Agarose gel electrophoresis confirmed that the samples contained clean rRNA bands. Poly-A$^+$ mRNA was purified using a Dynabeads mRNA Direct kit (Invitrogen) according to the manufacturer's directions. The beads were washed twice to remove rRNA contamination and eluted into 10 µL of the appropriate buffer. For first strand synthesis, 3 µL of RNA were combined with 1 µL of each 12 µM primer (cDNA-FWD 5'- tgctcattagaaagaaagcatagcaatctaatctaagttttaattac rG rG rG-3' and cDNA-REV 5'- AAAATCATAAATCATAAGAAATTCGCTTATTTAGAAGTGGT$_{31}$VN -3'). The mixture was heated to 72 °C for 2 minutes then cooled on ice for 2 minutes. Next 2 µL of first strand buffer, 1 µL of 20 mM DTT, 1 µL of 10 mM dNTPs, and 1 µL of SMARTScribe (Clontech, Mountain View, CA) were added to the reaction mixture. The reaction mixture was heated to 42 °C for 1 hour, then cooled on ice. 1 µL of 25 mM NaOH was added to the mixture, and the mixture was incubated at 68 °C for 30 minutes to degrade the RNA. For second strand synthesis, 71 µL of water, 10 µL 10x PfuUltraII buffer, 2 µL 10 mM dNTPs, 2 µL each 10 mM primer, and 2 µL PfuUltraII (Agilent) were added to the mixture. The reaction was amplified by 25 cycles of PCR with an annealing temperature of

56 °C and a 10 minute extension time. The resulting cDNA library was cleaned up on a commercial column (Qiagen) and then used as a template for 8x100 μL PCRs, run as before except substituting Pfu for PfuUltraII. The resulting ~ 40 μg of cDNA was then ready for gap repair transformation into *S. cerevisiae*.

Gap repair transformations were performed using the method of Chao et al. (Chao et al., 2006). Briefly, 6 μg of the destination vector, a centromeric yeast shuttle vector with a uracil marker named pCS1585, was linearized with EcoRI and AvrII overnight. The cDNA and destination vector were extracted with phenol/chloroform and coprecipitated with ethanol. The resulting DNA was then electroporated into the screening strain CSY820.

3.4.2 Subtractive hybridization

To perform subtractive hybridization, *C. arabica* total RNA was purified using Dynabeads. The resulting poly-A$^+$ mRNA was left hybridized to the beads and resuspended in 10 μL 10 mM Tris, pH 7.5. On-bead first strand synthesis was performed by adding 4 μL of first strand buffer, 2 μL of 20 mM DTT, 2 μL of 10 mM dNTPs, and 2 μL of SMARTScribe and incubating at 42 °C for 1 hour. The mRNA was eluted of the beads, the beads were washed with Buffer 2 (NEB), and resuspended in 50 μL Buffer 2, 1x BSA, with 1 μL 10 mM dNTPs and 1 μL T4 polymerase (NEB).

Next, *C. dewevrei* total RNA was hybridized to the beads. After hybridization, the supernatant was removed and saved. The beads were washed once with Buffer B (Invitrogen), then eluted with 200 μL 10 mM Tris, pH 7.5. The beads were regenerated according to the manufacturer's instructions, and the entire cycle was repeated for a total of three times. The supernatant resulting from the third annealing was then annealed to fresh poly-T beads, and first and second strand synthesis were performed as described previously.

After second strand synthesis, short cDNA fragments (< 500 bp) were removed using a ChromaSpin TE+400 size exclusion column (Clontech) according to the manufacturer's instructions. After size exclusion, the resulting cDNA was reamplified by PCR to the appropriate volume for gap repair library construction.

3.4.3 Growth conditions for liquid culture *S. cerevisiae* assays

Yeast cultures were grown in 96-well plates (BD Falcon) using AeraSeal film (Excel Scientific, Victorville, CA) to allow for thorough aeration. Colonies were picked from agar plates with toothpicks, inoculated into 400 μL of SD-Ura media, and grown for 24 hours in a Kuhner LT-X plate shaker at 30 ºC, 480 RPM, a 1.24 cm orbital diameter, and 80% humidity. The cultures were subsequently backdiluted 100x into fresh SD-Ura with or without 1 mM caffeine and regrown for an additional 24 hours.

3.4.4 Flow-cytometry-based library screening

When screening enzyme libraries, transformants were grown in 96-well plates and assayed for fluorescence. Yeast cultures harboring the enzyme expression vectors and integrated metabolite-sensing device were diluted 4x into water and assayed for fluorescence in 96-well plates on a Beckman Quanta flow cytometer. Cells were excited at 488 nm and GFP fluorescence was measured at 525 nm. Samples were gated first by electronic volume and side scatter to capture the cell population and then by fluorescence to remove the outliers with significantly low fluorescence. Approximately 8,000 cells were analyzed for each culture. The geometric mean fluorescence, normalized by the electronic volume, was then compared to the parental control. The brightest clones were selected for further screening by HPLC analysis.

3.4.5 FACS-based library screening

Immediately following the gap repair transformation, the FACS library was diluted into 250 mL of SD-Ura. Dilutions of the transformation culture were plated on SD-Ura plates to determine the library size. The library culture was diluted into fresh SD-Ura every

24 hours. The initial dilution factor was 10-fold, increasing to 30-fold and then 100-fold for the final presort dilution 24 hours prior to sorting. In a positive sort, caffeine was added to the growing culture to a final concentration of 1 mM during the final back-dilution. In a negative sort, the culture was grown in the absence of caffeine. Immediately prior to sorting, cells were centrifuged at 6,000 g for 5 minutes at 4 °C. The supernatant was discarded; the cells were resuspended in 1x phosphate-buffered saline + 1% bovine serum albumin (Fraction V, EMD Chemicals), stained with DAPI (Invitrogen), and filtered through a 40 μm cell strainer (BD Falcon). The cells were sorted using a BD Aria II sorter at the Stanford Shared FACS Facility. GFP was excited at 488 nm and captured by a 505 nm beam splitter and a 525/20 nm bandpass filter. mCherry was excited at 532 nm and captured by a 600 nm beam splitter and a 610/20 bandpass filter. DAPI was excited at 355 nm and captured by a 450/50 bandpass filter. Viable cells were isolated by gating on forward vs. side scatter, followed by a viability gate for DAPI negative cells. The cells resulting from one sort were grown overnight, then back-diluted 100-fold into the appropriate media (with or without substrate) and grown for 24 hours before the next sort.

3.4.6 HPLC methods

During screening, metabolite analysis was performed on an XDB-C18 2.1 x 50 mm, 3.5 μm column (Agilent Technologies). I injected 5 μL of sample onto the column. The mobile phase was 0.35 mL/min of 15% methanol/85% water with 0.1% acetic acid. Theophylline eluted at 1.65 minutes and was detected by UV absorbance at 274 nm. Culture supernatant from wild-type W303 showed no detectable peak. The identity of the theophylline peak was confirmed with each assay by the use of an authentic standard (Sigma-Aldrich).

3.5 Acknowledgements

I would like to thank Igor Cesarino and Paulo Mazzafera for providing the *C. arabica* and *C. dewevrei* total RNA. Isis Trenchard integrated the *A. thaliana* cytochrome P450 reductase into CSY820. Joe Liang assisted with the FACS-based library screening. This work was supported by the National Science Foundation (CBET-0917638; fellowship to JKM), the Institute for Collaborative Biotechnologies (U.S. Army Research Office DAAD19-03-D-0004), and the Alfred P. Sloan Foundation (fellowship to CDS).

4 Identifying and Alleviating Stress from Monooxygenase Overexpression

Portions of this chapter are adapted with permission from Michener JK, Nielsen J, and Smolke CD (2012) Identification and treatment of heme depletion attributed to overexpression of a lineage of evolved P450 monooxygenases. *Proc Natl Acad Sci USA*. doi: 10.1073/pnas.1212287109

Recent advances in metabolic engineering have demonstrated that novel biosynthetic pathways in microbes can provide a viable alternative to chemical synthesis for the production of both bulk and fine chemicals. The introduction of a new biosynthetic pathway typically requires the expression of multiple heterologous enzymes in the production host, which can place severe stress on the host cell. The host has not evolved to deal with the specific stresses of the engineered pathway, and the cell's response to the new stress may limit pathway productivity. Unfortunately, analysis and treatment of the host stress response can be difficult, as there are many sources of stress that may interact in complex fashions. I used global transcript measurements to identify heme depletion as the major source of stress resulting from overexpression of a lineage of evolved heterologous P450 monooxygenases in *Saccharomyces cerevisiae*. Heme depletion leads to low enzyme expression due to increased protein degradation. I further demonstrate that this stress decreases during rounds of evolution when the enzyme is highly expressed and increases during rounds when the expression is low. Overexpression of a rate limiting enzyme in the heme biosynthetic pathway alleviates this stress, increasing the enzymatic activity of the P450 by 2.3-fold. Heme overexpression can also increase the expression of a cytosolic heme-containing catalase but not a membrane-bound P450, implying that other factors may limit expression of some hemoproteins. This work demonstrates the utility of combining systems and synthetic biology to analyze and optimize heterologous biosynthetic pathways in microorganisms.

4.1 Introduction

The burgeoning field of metabolic engineering offers the promise of efficient, controlled, scalable production of both fine and bulk chemicals (Atsumi et al., 2008; Hawkins & Smolke, 2008; Ro et al., 2006; Yim et al., 2011). This approach is particularly useful when synthesizing complex molecules with defined stereochemistry, as such molecules are difficult to synthesize chemically. For example, plant secondary metabolites are a rich source of pharmaceuticals, such as the antimalarial isoprenoid artemisinin and the analgesic benzylisoquinoline alkaloid morphine. However, there are significant limitations to the industrial production of these compounds. The total chemical synthesis of complex metabolites is prohibitively costly (Rice, 1980) and the extraction from plants can be unpredictable, as it depends on complicated environmental and human factors (Hale et al., 2007). Additionally, evolution has optimized plants for the production of a final product; if the desired product is an intermediate, accumulation of sufficient quantities may be infeasible.

Microbial production of plant secondary metabolites offers a powerful alternative to traditional extraction methods (Ro et al., 2006). Microbes can produce specific molecules in a cost-effective manner, and the pathway can be tailored to produce a wide range of natural (Hawkins & Smolke, 2008) and nonnatural (Runguphan et al., 2010) compounds. However, these biosynthetic pathways can be very complex and therefore require the simultaneous expression of many heterologous enzymes in the production host. For example, the complete synthesis of morphine from tyrosine requires a total of 14 separate enzymatic reactions (Liscombe & Facchini, 2008). These enzymes can potentially interact with each other and with the host cell in deleterious ways (Ro et al., 2008), and the longer the biosynthetic pathway the more opportunities arise for such interactions. Understanding and alleviating these harmful interactions can significantly improve the pathway productivity and yield (Lee et al., 2007; Park et al., 2007).

However, there are many different potential stresses resulting from heterologous pathway expression. These stresses range from predictable, including the common stresses of heterologous protein production (Goff & Goldberg, 1985) and by-product toxicity (Zhu et al., 2002; Zhu et al., 2001a), to the novel, such as the overexpression of spider silk in *Escherichia coli* leading to the specific depletion of glycyl-tRNA (Xia et al., 2010). Individually analyzing each source of stress would be a lengthy process requiring an exhaustive and accurate list of potential stresses. As an alternative to such bottom-up approaches, we can instead use global analyses to identify the host cell's response to the induced stress looking, for example, at changes in transcript (Kizer et al., 2008) or protein levels (Han et al., 2001). Top-down approaches rely on the crucial assumptions that (a) the heterologous stress will trigger an endogenous response and (b) the endogenous response will produce a change in protein or RNA levels. These assumptions are generally valid, but as a result it is often easier to demonstrate that a particular stress is present than to conclusively rule out a potential source of stress.

Once the stresses are identified, the next step is to increase pathway productivity by treating the stresses, either by eliminating the source of the stress or by augmenting the cell's ability to respond to the stress. For example, after glycyl-tRNA depletion was identified as a cause of low expression of spider silk in *E. coli*, researchers modified the host to accommodate the demand for glycyl-tRNA by overexpressing both the tRNAGly and the glycine biosynthetic pathway (Xia et al., 2010). Alternatively, when the by-product glycerol-3-phosphate was shown to inhibit the production of 1,3-propanediol, a glycerol kinase knockout prevented by-product formation and increased the 1,3-propanediol yield (Zhu et al., 2002).

In this chapter, I considered the stress due to heterologous overexpression of a P450 monooxygenase in *S. cerevisiae*. P450s are an important class of enzymes in plant secondary metabolism, participating in the biosynthesis of metabolites ranging from alkaloids and terpenoids to hormones and lipids (Mizutani & Ohta, 2010). There are many potential sources of stress resulting from heterologous overexpression of a P450 in yeast: the enzyme binds heme as a cofactor, sequestering it from other endogenous enzymes. The monooxygenase consumes NADPH and may produce toxic reaction by-products such as formaldehyde. The enzyme may be uncoupled, so many of the electrons taken from NADPH are not transferred to the substrate but instead produce reactive oxygen species (ROS) (Fasan et al., 2008). Additionally, the relaxed substrate selectivity of the monooxygenase may allow it to oxidize endogenous compounds, consuming important metabolites and producing potentially toxic side products.

In order to narrow down this list of potential stresses, I first used DNA microarrays to identify the major stresses involved. However, global analysis methods often identify many different cellular responses, and selecting the stresses that are the best targets for treatment can be difficult, requiring either a lengthy search of the potential targets (Lee et al., 2007) or intuition about the likely targets (Choi et al., 2003). Instead, I took inspiration from inverse metabolic engineering, where researchers frequently compare multiple strains with varying levels of productivity to identify the components responsible for the observed variation in productivity (Askenazi et al., 2003; Bro et al., 2005). The monooxygenase used in this study had been evolved *in vivo* to improve the enzyme's ability to demethylate caffeine to theophylline. As a result, I had an entire lineage of enzyme mutants with a range of different activities. Enzyme overexpression places a significant stress on the cell, reflected in a reduced growth rate, and that stress produces a selective pressure during the evolutionary process. As a result of this selective pressure, I expected the stress to change as the enzyme is evolved. Additionally, I varied the expression level of the mutant enzymes, further modulating the cellular stress. By tracking the changes in stress as the enzyme evolves and the expression level is changed, I can better identify the relevant stresses.

My global analysis indicated that enzyme overexpression starves the cells for heme, leading to a significant change in the host physiology. The lack of heme also causes the P450 to misfold, reducing the expression of active enzyme and consuming cellular resources to degrade the misfolded protein. Overexpression of three rate limiting steps in heme biosynthesis, in addition to feeding iron and the heme precursor δ-aminolevulinic acid, increased the heme level by up to 90-fold and the product concentration by 2.3-fold. Additionally, while enzyme overexpression led to an increase in proteasomal activity, concomitant overexpression of heme biosynthesis reduced proteasomal activity to background levels. These results demonstrate that systems and synthetic biology can be successfully combined to analyze the stress due to heterologous enzyme expression, identify the most significant sources of stress, and alleviate those sources of stress to ultimately increase pathway productivity.

4.2 Results

The caffeine demethylase that I evolved in Chapter 2 placed a significant stress on the cell, reflected in a decreased growth rate. When I investigated the activity of several enzyme variants under conditions different from those in which the variants were selected, I found evidence that the stress produced by the enzyme was changing as the enzyme was evolved, and the conditions of the evolution determined the selective pressure on the enzyme-dependent stress.

4.2.1 Improvements at low copy do not translate to high copy

The original caffeine demethylase underwent five rounds of evolution while expressed from a high-copy plasmid, producing enzyme yCDM6, followed by two further rounds of evolution on a low-copy plasmid to give yCDM8. yCDM6 exhibited similar levels of activity when expressed from high- and low-copy plasmids. However, when I expressed the entire lineage of mutant enzymes from both high- and low-copy plasmids, I found that yCDM7 and yCDM8, the enzymes that were evolved at low copy, were only marginally more active than yCDM6 expressed from a high-copy plasmid (Figure 4.1). In contrast, the enzymes that were evolved on high-copy plasmids, yCDM2 through yCDM6, were equally active at high- and low-copy conditions. I hypothesized that the later enzymes, yCDM7 and yCDM8, were placing a stress on the cell that was tolerable when the enzyme was expressed at low levels but deleterious when the enzyme expression was increased. However, there were too many potential sources of stress to exhaustively verify each one.

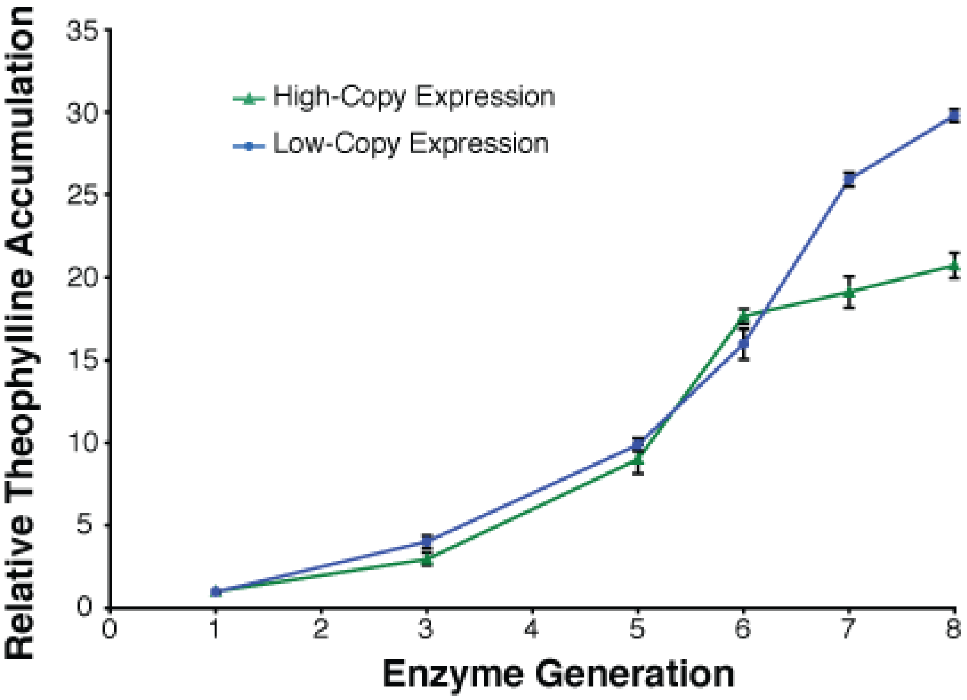

Figure 4.1 Comparison of theophylline accumulation for high- and low-copy enzyme expression. The early enzymes, through yCDM6, were identified based on activity when expressed from a high-copy plasmid in yeast. The last two enzymes, yCDM7 and yCDM8, were evolved on a low-copy plasmid backbone. Cells containing each of the enzymes developed in Chapter 2 were grown in the presence of caffeine. Theophylline accumulation was assayed after 24 hours. The data are all normalized to yCDM1 at high copy (relative accumulation = 1.0). The error bars show ±1 standard deviation, calculated from three biological replicates, and the lines are a guide for the eye.

4.2.2 Microarrays identify heme depletion as the major cellular stress

Rather than individually testing each possible stress, I instead used DNA microarrays to identify the source of this stress by analyzing the global transcriptional response to monooxygenase overexpression. I selected a total of eight strains for analysis: (1–6) yCDM1, yCDM6, and yCDM8, each expressed from high- and low-copy plasmids; (7) the catalytically inactive variant yCDM1-A264H (Neeli et al., 2005) expressed from a high-copy plasmid; and (8) an empty high-copy plasmid. Each strain was assayed in triplicate. I then used principal component analysis (PCA) to identify common patterns in expression across the different samples.

Critically, in my application of PCA I treated the samples as variables and the genes as observations (Raychaudhuri et al., 2000), while previous metabolic engineering analyses have done the reverse. Using PCA in this fashion allowed me to identify profiles of gene expression across the samples (expression is high in sample A, low in sample B, etc.) that explain the observed variation in gene expression. By looking at the genes that are most strongly associated with a given pattern of gene expression, I could identify the specific stresses that elicited the pattern. Ultimately, then, this pattern can be interpreted as a measure of the magnitude of the associated stress.

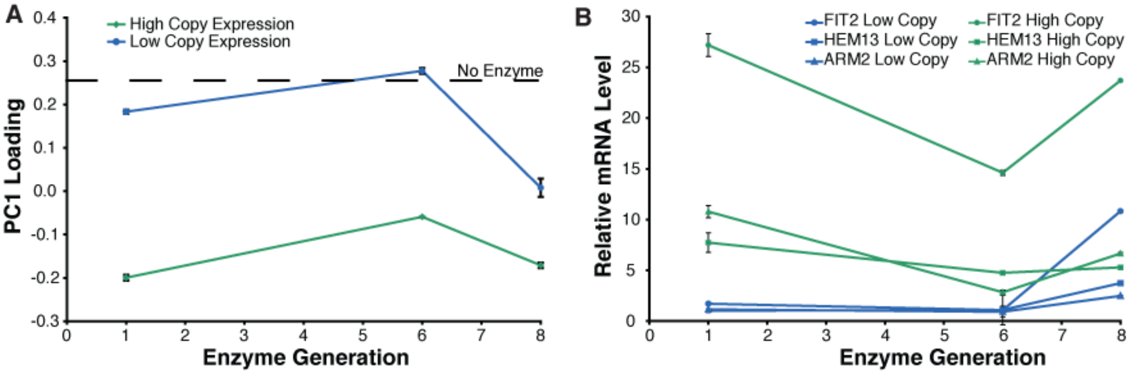

Figure 4.2 Microarray analysis identifies heme depletion as the major cellular stress. (**A**) The loadings for the first principal component are plotted for each of the six enzymes assayed. The dotted line denotes the loading for the no enzyme control. The error bars show ±1 standard deviation, calculated from three biological replicates. (**B**) mRNA levels, relative to the no enzyme control, are shown for three iron- and heme-regulated genes (FIT2, HEM13, and ARN2). The three genes are repressed by heme and/or iron, so the pattern in the mRNA levels is the inverse of the pattern in the loading. The error bars show ±1 standard deviation, calculated from three biological replicates.

The loadings for the first principal component, shown in Figure 4.2A, identify a pattern of expression that explains ~ 53% of the variability between samples. Next, the genes are given scores that indicate how well the observed expression matches this pattern of expression. I analyzed the genes with the highest magnitude scores to identify transcription factors that were likely involved in coordinating the response (Oliveira et al., 2008). The transcription factors identified in this analysis included Rcs1, which responds to iron starvation, as well as Hap1 and Rox1, which regulate genes in a heme-dependent manner. Indeed, a closer look at several Rcs1-, Rox1-, and Hap1-dependent genes shows a similar pattern to the loadings of the first principal component (Figure 4.2B), suggesting that the loadings may be interpreted as a measure of heme and iron levels, with lower values of the loading correlating to lower intracellular levels of heme.

4.2.3 Increased heme biosynthesis raises the intracellular heme level

Having implicated heme limitation as the major source of cellular stress, I next sought to restore the intracellular heme concentration to its native level. In bacteria, the first committed step in heme biosynthesis, producing δ-aminolevulinic acid (ALA), is limiting and feeding additional ALA can increase the expression of functional hemoproteins (Kraus & Kery, 1997). Previous efforts to expand the pool of heme available for P450 expression in yeast have produced modest increases in enzymatic activity by feeding iron and ALA (Jiang & Morgan, 2004). However, ALA synthesis is not the rate limiting step in heme biosynthesis (Hoffman et al., 2003). Therefore, I simultaneously overexpressed three rate limiting enzymes - HEM2, HEM3, and HEM12 - in addition to feeding iron(II) and ALA. I compared total cellular levels of heme and the heme intermediate porphyrins in cells with different levels of heme overexpression and heme usage. This assay that I used does not distinguish between free and protein-bound heme, so free heme levels must be inferred from observing the response of genes that are regulated in a heme-dependent manner.

For a given level of heme usage, increasing the capacity of the heme biosynthetic pathway led to an increase in the total heme concentration. However, I only observed large increases in total heme content when heme biosynthesis and usage were both elevated (Figure 4.3A). As expected, increasing the capacity of the initial stages of heme biosynthesis led to accumulation of biosynthetic intermediates (Figure 4.3B). However, strong overexpression of a heme-containing enzyme led to lower porphyrin levels, suggesting that the cell is actively controlling the conversion of porphyrins to heme, presumably by transcriptional regulation of HEM13 in response to increasing levels of free heme. To test this theory, I measured HEM13 mRNA levels by qRT-PCR. HEM13 expression is repressed by free heme, so high levels of HEM13 mRNA correspond to low levels of free heme. HEM13 expression increased as the P450 expression did, and decreased below background levels when both the P450 and HEM3 were highly expressed (Figure 4.3C). These results are consistent with my hypothesis that P450 overexpression starves the cell of free heme but overexpression of the heme biosynthetic pathway can restore the free-heme level.

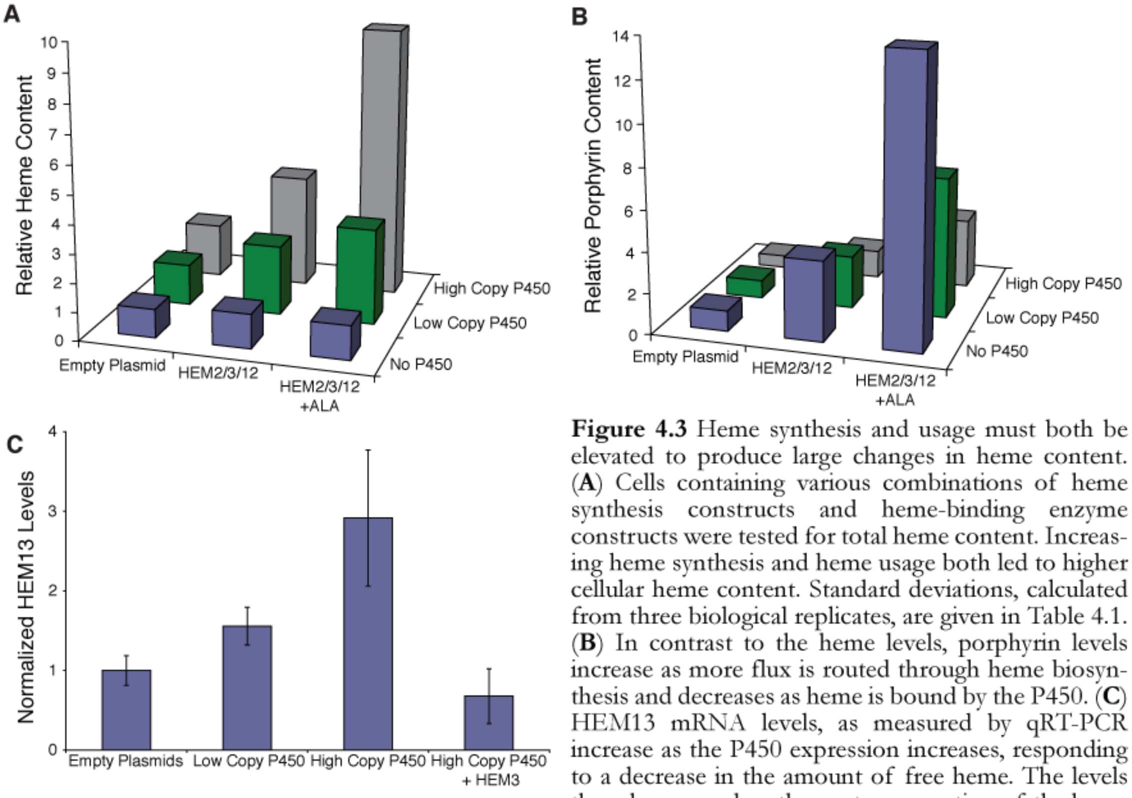

Figure 4.3 Heme synthesis and usage must both be elevated to produce large changes in heme content. (**A**) Cells containing various combinations of heme synthesis constructs and heme-binding enzyme constructs were tested for total heme content. Increasing heme synthesis and heme usage both led to higher cellular heme content. Standard deviations, calculated from three biological replicates, are given in Table 4.1. (**B**) In contrast to the heme levels, porphyrin levels increase as more flux is routed through heme biosynthesis and decreases as heme is bound by the P450. (**C**) HEM13 mRNA levels, as measured by qRT-PCR increase as the P450 expression increases, responding to a decrease in the amount of free heme. The levels then decrease when the upstream portion of the heme biosynthetic pathway is overexpressed, demonstrating that the free-heme levels have recovered and the cell is now restricting heme biosynthesis at HEM13 and accumulating porphyrins. The error bars show ±1 standard deviation, calculated from three biological replicates.

4.2.4 Heme limits the activity of cytosolic hemoproteins

Having shown that increasing the cells' capacity to produce heme leads to an increase in the total heme level, I next considered the effect this extra heme has on theophylline accumulation. The connection is not obvious: while the evidence suggests that heme is limiting in cells that overexpress CYP102A1, total product accumulation depends on a large number of factors. Heme is costly to produce, and free heme is toxic to the cell. The reward of alleviating the heme stress may be outweighed by the burden of producing the necessary heme.

When I increased the cells' capacity to produce heme, I observed an increase in the amount of theophylline produced, averaging a 2.3-fold improvement at the optimal heme level (Figure 4.4A). Unfortunately, the experiments showed significant day-to-day and culture-to-culture variability, likely a result of varying copy numbers of the two plasmids. There is a selective pressure to maintain both plasmids at some nonzero copy number, but the copy number of each plasmid is likely to vary across cells and across time. Replicate cultures may have different relative copy numbers and therefore different phenotypes.

Additionally, while the cells overexpressing heme biosynthesis produced more theophylline relative to the empty plasmid control, the absolute theophylline production with two plasmids (high-copy CDM plasmid and high-copy empty plasmid) was lower than the single-plasmid control (high-copy CDM plasmid, Figure 4.1), likely due to the burden of carrying an additional plasmid (Figure 4.4B and Figure 2.8D). I discovered that cells expressing both the enzyme and the heme overexpression constructs grew more slowly than cells with the enzyme and an empty plasmid (Figure 4.4B). However, the burden appears to be associated with the expression of the heme biosynthetic enzymes rather than the production of heme. The growth rate was similar for cells overexpressing HEM3, overexpressing HEM2/3/12, or overexpressing HEM2/3/12 and fed ALA. Therefore, I can conclude that caffeine demethylase activity is heme limited, the cost of overexpressing heme biosynthetic genes is significant, and the benefits of producing more heme are greater than the cost.

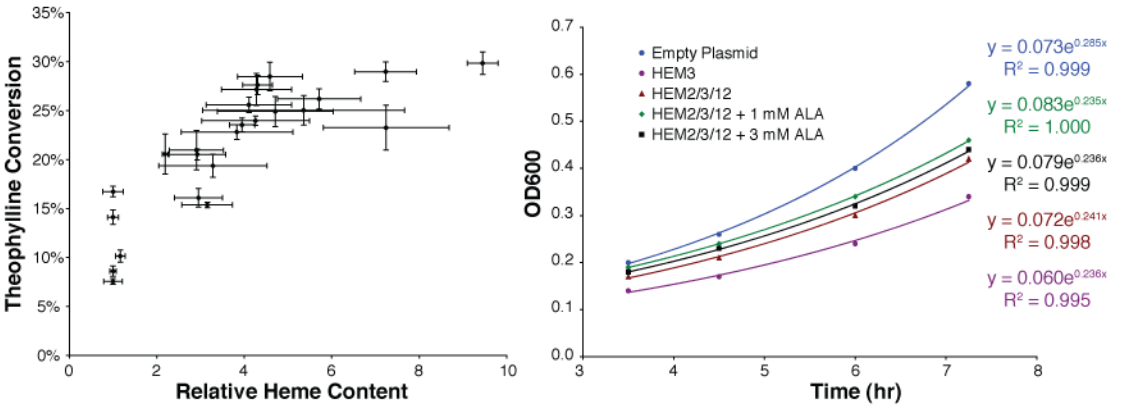

Figure 4.4 Increasing the total cellular heme level leads to an increase in total enzymatic activity at the cost of slower growth. (**A**) Cells expressing yCDM8 from a high-copy plasmid were cotransformed with high-copy plasmids overexpressing HEM3 or HEM2/3/12. Cultures were grown in the presence of caffeine and varying amounts of iron and ALA. After 48 hours, the theophylline concentration was measured in the supernatant and the cell pellets were assayed for heme content. Each point represents the average of three biological replicates, and error bars show ±1 standard deviation. The data shown are concatenated from four separate experiments to account for day-to-day variability. (**B**) Growth curves were measured for cells expressing yCDM8 from a high-copy plasmid in combination with heme overexpression. The heme overexpressing cultures grow more slowly than cultures that carry a similar plasmid but do not overexpress heme. The curves are an exponential fit to the data.

Given the stress associated with heme overexpression, I next sought to tune the overexpression constructs to minimize this stress. I integrated the heme overexpression constructs into the yeast genome, eliminating the need to maintain a high-copy plasmid while also presumably lowering the expression of the genes. Unexpectedly, I found that cells with the integrated copies of the overexpression constructs produced more heme than cells with those same constructs on high-copy plasmids. The strain with HEM2, HEM3, and HEM12 integrated simultaneously grew very slowly (Figure 4.5B) and produced enormous quantities of heme (> 90-fold more than the control, data not shown). However, the strain with only HEM3 integrated showed no growth defect (Figure 4.5B) and produced the highest theophylline titers yet seen, reaching 42% conversion (Figure 4.5A). Notably, HEM3 overexpression had no effect on theophylline production, either positive or negative, when yCDM8 was expressed from a low-copy plasmid, demonstrating that integrated HEM3 does not have an inherent deleterious effect and that the benefits of heme overexpression are dependent on the simultaneous overexpression of the hemoprotein yCDM8.

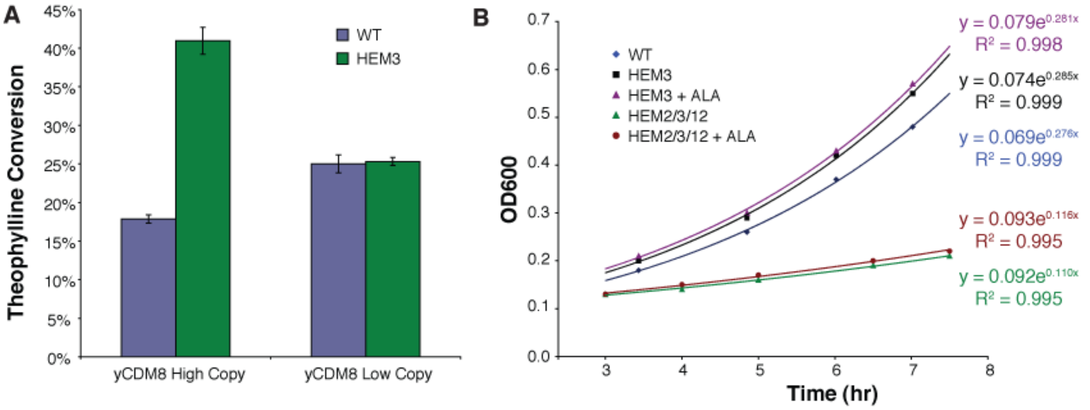

Figure 4.5 Integrated HEM3 overexpression improves theophylline production from highly expressed P450. (**A**) Cells expressing yCDM8 from either a high- or low-copy plasmid were transformed into strains with and without an integrated HEM3 expression construct. Adding ALA to either the WT or HEM3 strains did not affect activity. After 48 hours, the theophylline concentration was measured in the supernatant. The error bars show ±1 standard deviation calculated from three biological replicates. (**B**) Growth curves were measured for cells expressing yCDM8 from a high-copy plasmid in combination with integrated heme overexpression. Integrating HEM2/3/12 led to very slow growth. The curves are an exponential fit to the data.

Finally, I asked whether the expression of other hemoproteins was limited by heme biosynthesis. I tested two representative enzymes: a membrane-associated P450, CYP2D6 (Hawkins & Smolke, 2008), and a cytosolic catalase, CTT1. In contrast to the soluble CYP102A1, CYP2D6 was not heme limited and showed a slight decrease in activity with increasing concentrations of heme (Figure 4.6A). I expect that CYP2D6 expression is limited instead by the ability of the cell to accommodate large quantities of functional membrane proteins (Schunck et al., 1991). However, the soluble catalase CTT1 exhibited nearly twice the activity when both heme and enzyme are overexpressed (Figure 4.6B). For comparison, previous work overexpressing HEM2 alone led to a 20–40% increase in activity (Mattoon & Bajszar, 1998).

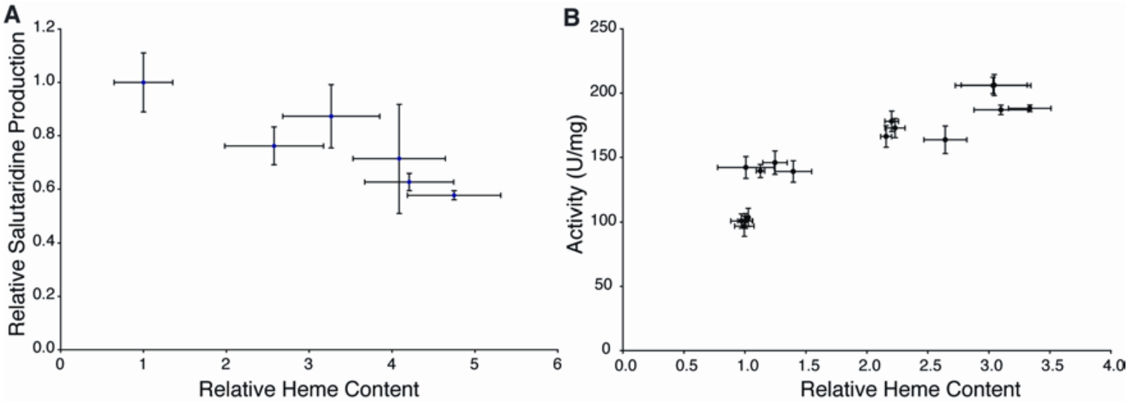

Figure 4.6 Heme overexpression increases the activity of soluble, but not membrane-associated, hemoproteins. (**A**) Cells expressing membrane-associated CYP2D6 from a high-copy plasmid were cotransformed with high-copy heme overexpression plasmids. Cultures were grown in the presence of norlaudanosline and varying amounts of iron and ALA. After 48 hours, the salutaridine concentration was measured in the supernatant and the cell pellets were assayed for heme content. Each point represents the average of three biological replicates with error bars showing ±1 standard deviation. (**B**) Cells expressing CTT1 from a high-copy plasmid were cotransformed with heme overexpression plasmids. Cultures were grown to midexponential phase (OD ~ 0.4) and lysed with glass beads. Catalase activity was determined by monitoring the degradation of H_2O_2 *in vitro*. Heme content was measured as described previously. Each point represents a single biological replicate and the error bars show ±1 standard deviation calculated from technical replicates.

These results demonstrate that heme overexpression is not a general solution to increase the functional expression of a hemoprotein. Cytosolic hemoproteins may be heme limited, but membrane hemoproteins are likely to be limited by other factors. The complicated interactions underlying an apparently simple process of cofactor production highlights the importance of my microarray analysis. Lacking the systems-level analysis, each potential source of stress would need a separate method of characterization and rules predicting when the stress was likely to occur. In contrast, the microarray analysis allowed me to directly identify a limiting stress.

4.2.5 Heme depletion limits total enzyme expression

I previously demonstrated in Chapter 2 that total CDM expression was roughly constant, irrespective of the enzyme generation or the plasmid copy number. I hypothesized that most of the additional proteins produced from the high-copy plasmid were misfolding, leading to increased rates of protein degradation and equivalent steady-state protein levels.

To test this hypothesis, I measured the proteasomal activity of cells expressing high or low levels of the final enzyme yCDM8 (Figure 4.7A). Cultures that express yCDM8 from a high copy plasmid showed increased levels of proteasomal activity, in contrast to cells expressing the same enzyme from a low-copy plasmid. However, when we increased the heme levels, the proteasomal activity decreased to background levels (Figure 4.7B) and the total CDM expression increased (Figure 4.7C).

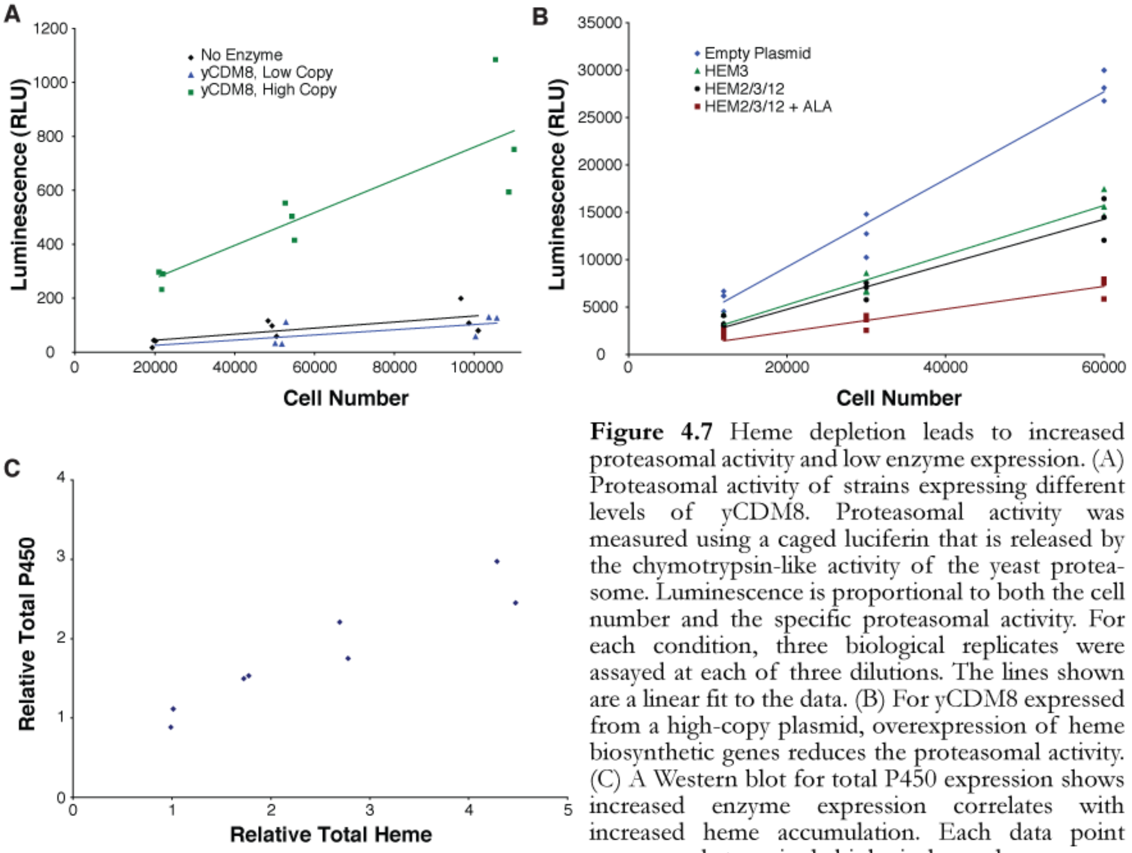

Figure 4.7 Heme depletion leads to increased proteasomal activity and low enzyme expression. (A) Proteasomal activity of strains expressing different levels of yCDM8. Proteasomal activity was measured using a caged luciferin that is released by the chymotrypsin-like activity of the yeast proteasome. Luminescence is proportional to both the cell number and the specific proteasomal activity. For each condition, three biological replicates were assayed at each of three dilutions. The lines shown are a linear fit to the data. (B) For yCDM8 expressed from a high-copy plasmid, overexpression of heme biosynthetic genes reduces the proteasomal activity. (C) A Western blot for total P450 expression shows increased enzyme expression correlates with increased heme accumulation. Each data point corresponds to a single biological sample.

4.3 Discussion

In the past, systems biology has generally been applied to metabolic engineering in two distinct fashions. When used for forward metabolic engineering, researchers typically consider a single strain, the best producer, and ask how that strain differs from a control. The difficulty of a using a global analysis is the sheer volume of data that can be obtained. In one of my engineered strains, for example, the transcription of 953 genes was significantly different than the control ($p < 0.001$), and greater than 20% of the genes measured showed statistically significant variation in at least one of the seven experimental strains. In inverse metabolic engineering, researchers deal with this problem by constructing many variant strains and then looking for consistent patterns between the variants. However, this approach has not previously been used for forward engineering. In this chapter, I successfully applied the techniques of inverse metabolic engineering to a forward engineering problem, quickly narrowing my focus to those genes that show consistent patterns of expression across the variants tested and then to the transcription factors that produce the coordinated response.

Based on my global transcript analysis, I propose that the overexpression of a heme-containing monooxygenase depletes the intracellular pool of heme and the resulting lack of heme places a stress on the cell that limits the total enzymatic activity. High-copy expression of the enzyme sequesters more heme and therefore produces a greater stress than low-copy expression. Evolution of the enzyme on a high-copy plasmid reduced the stress, suggesting that the stress imposed a selective pressure to minimize the deleterious effects. However, decreasing the enzyme copy number removed both the stress of heme depletion and the selective pressure to minimize that depletion. Further evolution of the enzyme on a low-copy plasmid led to an increase in enzymatic activity when expressed at low-copy, but also to greater heme usage by the monooxygenase. As a result, the enzymes that show improved activity under low-copy expression conditions do not show corresponding increases in activity when expressed from a high-copy plasmid.

Heme depletion has several effects on the cell. First, lack of heme disrupts the host metabolism. Without heme, the cells show gene expression profiles consistent with anaerobic growth, despite the presence of sufficient oxygen. For example, transcription of heme-dependent genes such as mitochondrial cytochromes is significantly reduced. Second, low heme levels limit the production of active P450. When heme levels are low, additional nascent peptides are unable to properly bind heme and, as a result, misfold. The misfolded proteins are then degraded by the proteasome. Disruption of aerobiosis, in addition to the additional burden of recycling misfolded protein (Geiler-Samerotte et al., 2011), would tend to reduce the host's growth rate and correspondingly reduce theophylline production. Additionally, heme-dependent protein misfolding limits the expression of active, properly-folded P450, further reducing total activity. In my system, increasing the heme supply, by overexpressing three rate limiting enzymes and feeding additional substrates, alleviated these stresses and increased the total enzymatic activity by 2.3-fold.

I have demonstrated that CYP102A1 is not the only heme limited enzyme in yeast, but also that not all hemoproteins are heme limited. In situations where activity from a soluble hemoprotein is limiting, a similar heme overexpression strategy may be worthwhile. If optimal activity is required, the heme overexpression constructs could be streamlined and optimized by tuning the expression level of the HEM genes. Similarly, while feeding ALA is a straightforward method for increasing precursor availability for the heme biosynthetic pathway, the cost would likely be prohibitive on an industrial scale. Accordingly, overexpression of the ALA biosynthetic pathway (Kang et al., 2011) might be a preferable solution. A similar strategy might also be useful for other enzyme classes, such as S-adenosyl methionine (Okamoto et al., 2003) or the phosphopantetheinyl group of acyl carrier proteins (Siewers et al., 2009), where a focus on cofactor availability and loading might identify novel factors limiting productivity.

Currently, this type of systems analysis is limited by the assumption that cellular stresses will be reflected in the transcript or protein levels. When we see changes in transcript levels, we can generally trust that they result from a disturbance to the host. However, constant transcript levels can mean either that the corresponding stress is absent or that the stress does not produce a change in transcript levels. If we cannot distinguish between these two situations, we run the risk of missing important data. For example, my global analysis did not identify protein degradation as a major cellular stress even though I later found that the proteasomal activity was significantly higher in some strains. This misidentification is likely due to posttranslational regulation of the proteasome (Mason et al., 1996; Zhang et al., 2003). Had I not hypothesized that protein turnover was increasing and specifically measured the proteasomal activity, this would have remained a false negative in my global analysis. Conversely, the demethylase used in this work produces formaldehyde as a by-product, and therefore I identified formaldehyde toxicity as one potential source of stress. The gene responsible for formaldehyde detoxification, Sfa1, is not differentially transcribed in the strains that I analyzed. I could only rule out formaldehyde toxicity by verifying in the literature that Sfa1 would be induced were formaldehyde present at high concentrations (Wehner et al., 1993). Global analyses would be more informative if we had a better method of screening through this negative data to separate situations where a lack of response is meaningful, such as Sfa1, from those where measuring transcript or protein levels does not inform us about the underlying stress, such as the proteasome.

I anticipate that the approaches described in this chapter will be generally useful in the optimization of heterologous metabolic pathways. In addition to my specific solution, where heme overexpression leads to increased expression of a P450 monooxygenase, I believe that a systems analysis of multiple variants of a heterologous pathway will generally provide additional insight into the factors limiting pathway productivity and therefore enable further pathway optimization.

4.4 Methods

4.4.1 Strains and cultivation

The strains used in this work are derivatives of W303α (MATα leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15). Transcript analysis was conducted in CSY492 (W303 lys2::PTEF-GFP-L2Bulge8-ADH1T). Cultures were grown in shake flasks at 30 °C and 200 RPM in the appropriate synthetic dropout media (Formedium, Hunstanton, UK) supplemented with 2% glucose, an additional 10 mg/L of adenine (Sigma-Aldrich, St. Louis, MO), and 1 mM caffeine (Sigma-Aldrich). Heme overexpression strains were fed varying amounts of iron (II) citrate and δ-aminolevulinic acid (Sigma-Aldrich). HEM overexpression plasmids were provided by L. Liu, J.L.M Ruiz, and J. Nielsen. HEM overexpression constructs were integrated into the lys2 locus of W303 (Sadowski et al., 2007). Assays for salutaridine production were fed 4 mM norlaudanosoline (Santa Cruz Biotech, Santa Cruz, CA).

4.4.2 Metabolite analysis

Supernatant theophylline production was assayed on an Agilent 1200 series liquid chromatograph using a Poroshell 120 SB-C18 2.1 x 50 mm, 2.7 μm column (Agilent). The mobile phase was 0.50 mL/min of 20% methanol/80% water with 0.1% acetic acid. Theophylline eluted at 0.70 minutes and was detected at 274 nm. For each sample, 3 μL was injected onto the columns. The identity of the theophylline peak was confirmed with each assay by the use of an authentic standard (Sigma-Aldrich), and the concentration of theophylline in each sample was determined by comparison to a series of reference standards.

Supernatant salutaridine production was assayed on an Agilent 1200 series liquid chromatograph using a Zorbax SB-Aq 3.0 x 50 mm, 1.8 μm column (Agilent). The mobile phase was 0.60 mL/min of a mixture of water (Buffer A) and methanol (Buffer B), both with 0.1% acetic acid. The mobile phase started at 100% A for 1 minute, followed by a gradient to 75% A/25% B over three minutes, then held at 75% A/25% B for three minutes. After a total of seven minutes, there was a further gradient to 100% B over one minute, then held at 100% B for four minutes. Finally, the mobile phase was switched back to 100% A and reequilibrated for 6 minutes. Salutaridine was detected using an Agilent 6320 Ion Trap Mass Spectrometer, measuring the 265 m/z fragment of the 328 m/z ion.

4.4.3 DNA microarray experiments

Each strain for microarray analysis was grown overnight in appropriate dropout media. Each culture was diluted to OD 0.05 in 30 mL of fresh media, with four biological replicates per strain. When the cultures reached OD 0.3–0.4 they were quenched by decanting into a 50 mL centrifuge tube filled with ice. The cultures were centrifuged for 3 minutes at 4 °C and 5000 RCF, washed with 1 mL of water, transferred to a 1.5 mL centrifuge tube and centrifuged for 2 minutes at 4 °C and 8000 RCF. The resulting cell pellet was frozen in liquid nitrogen and stored at -80 °C in preparation for analysis.

For each strain, three cell pellets representing three biological replicates were lysed using the RNeasy kit (Qiagen, Valencia, CA) following the manufacturer's instructions. cDNA synthesis followed by aRNA synthesis and fragmentation were performed using the 3' IVT Express kit (Affymetrix, Santa Clara, CA) following the manufacturer's instructions.

aRNA synthesis and fragmentation were monitored using an Agilent 2100 Bioanalyzer and RNA 6000 Nano chips (Agilent Technologies, Santa Clara, CA). Fragmented aRNA was hybridized to Yeast Genome 2.0 DNA chips and scanned using a GeneChip 3000 7G Scanner (Affymetrix), according to the manufacturer's instructions.

Microarray data were analyzed using the BioConductor suite in R. Principal components analysis, treating the samples as the variables and gene expression data as observations (Raychaudhuri et al., 2000), was used to identify genes with consistent patterns of expression between the different strains. Prior to PCA, the microarray data were normalized to correct for the steady-state expression level (Holter et al., 2000). The ~ 400 genes with the highest magnitude scores for PC1 were used as input to Reporter Features (Oliveira et al., 2008) to identify transcription factors whose targets were overrepresented.

4.4.4 Proteasomal activity measurements

Proteasomal activity was measured using the Proteasome-Glo kit (Promega, Madison, WI), according to the manufacturer's directions. Cultures were grown to mid-log phase (OD 0.2–0.4) then diluted in fresh media to OD 0.1, 0.05, and 0.02 (corresponding to ~ 100,000 to 20,000 cells per 100 μL, respectively). 100 μL of the resulting cell suspension was mixed with 100 μL of the assay reagent, prepared according to the manufacturer's directions. After a 10 minute incubation at room temperature, the luminescence was measured on a Wallac 1420 Victor3 microplate reader (PerkinElmer, Waltham, MA).

4.4.5 Heme measurements

The intracellular heme concentration was measured using a derivative of a previously described protocol (Sassa, 1976). 5 mL samples of mid-log cultures (OD ~ 0.4) were centrifuged at 4 °C and 5000 g for 5 minutes. The pellet was washed with water, transferred to a centrifuge tube, and centrifuged again at 4 °C and 8000 g for 5 minutes. The pellet was then resuspended in 500 μL of 20 mM oxalic acid (Sigma-Aldrich) and stored at 4 °C in the dark for 16 hours. After the acid extraction, 500 μL of 2 M oxalic acid was added to each tube. 500 μL of the resulting suspension was transferred to a new centrifuge tube. The original centrifuge tube was heated to 95 °C for 30 minutes, removing the iron from non-fluorescent heme and producing a fluorescence porphyrin ring. 200 μL of each sample (heated and unheated) were measured in a microplate reader (Tecan Safire, Männedorf, Switzerland), exciting at 400 nm and measuring emission at 620 nm. A standard curve was constructed using variable concentrations of hemin (Sigma-Aldrich).

4.4.6 Quantitative Western blots

Yeast strains harboring the appropriate enzyme expression constructs were grown overnight in 5 mL of SD-Ura. Protein extraction was carried out using 0.1 M NaOH (Kushnirov, 2000) followed by lysis in protein loading buffer (Invitrogen). Samples and ladder (New England Biolabs P7711S) were resolved on 4–12% Bis-Tris SDS-PAGE gels in 1x MOPS (Invitrogen). Protein was transferred to a nitrocellulose membrane using semidry transfer (Bio-Rad) in 2x NuPAGE transfer buffer (Invitrogen) + 10% MeOH. After transfer, the membrane was cut in half at ~ 55 kDa. Both membrane halves were blocked in 5% BSA for 1 hour. The membrane with higher-molecular-weight proteins was blotted with an anti-V5 HRP antibody according to the manufacturer's instructions (Invitrogen). The membrane with lower-molecular-weight proteins was blotted with a mouse anti-actin antibody (Abcam 8224, Cambridge, UK) and a rabbit anti-mouse HRP (Abcam 6728) according to the manufacturer's instructions. Both HRP antibodies were detected by chemiluminescence, following the manufacturer's instructions, (Pierce, Rockford, IL) using a Chemi-Doc XRS imager (Bio-Rad). Blots were analyzed using the QuantityOne analysis software (Bio-Rad).

4.4.7 Catalase activity assays

Cells containing the catalase overexpression construct as well as various heme overexpression constructs were grown to saturation overnight. The cultures were diluted 20x into 50 mL of fresh dropout media and regrown to mid-log (OD ~ 0.4). 5 mL samples were taken to measure heme content as described previously. The remaining culture volume was centrifuged at 4 °C and 6000 g for 5 minutes, washed once with 1 mL of resuspension buffer (0.1 M potassium phosphate, 0.5 mM EDTA), and centrifuged again. The pellet was then resuspended in 1 mL of resuspension buffer plus protease inhibitor (HALT, Pierce) and transferred to a tube containing 500 mg of acid washed glass beads (Sigma). The samples were lysed by 5 cycles of 1 minute vortexing followed by 1 minute on ice. After vortexing, the crude lysate was centrifuged at 4 °C and 16,000 g for 5 minutes, and the supernatant was transferred to a new tube. The total protein concentration was measured using a Bradford reagent (Bio-Rad) according to the manufacturer's instructions, using 160 μL of sample dilutions and 40 μL in a microwell plate. Absorbance was assayed using a Tecan Safire microplate reader. Sample values were compared to a standard curve was constructed using BSA to determine the total protein concentration.

Next, the catalase activity of each sample was measured. Samples were diluted to ~ 10 μg/mL. 40 μL of protein was mixed with 160 μL of 250 uM H_2O_2. Aliquots were taken at 30, 60, and 120 seconds and quenched in 200 μL of Peroxide Assay Reagent (Pierce), according to the manufacturer's instructions. Absorbance was measured on a Tecan Safire microplate reader. Residual peroxide was calculated by comparison to a standard curve of H_2O_2 dilutions. One unit of catalase activity was calculated as the amount of active protein necessary to degrade 1 mM of H_2O_2 in 1 minute.

4.4.8 qRT-PCR measurements

Cells containing various combinations of heme and P450 overexpression plasmids were grown to saturation overnight. They were then diluted in triplicate 30 mL cultures and regrown to midexponential phase (OD_{600} ~ 0.5). 10 mL of each culture was centrifuged for 5 minutes at 6,000 g and 4 ^{o}C, washed with 1 mL of water, and repelleted for 5 minutes at 8,000 g and 4 ^{o}C. The supernatant was removed and the cells were frozen in liquid nitrogen and stored at -80 ^{o}C.

The cell pellets were resuspended in 500 µL buffer AE (50 mM NaOAc, 10 mM EDTA) with 1.5% SDS. 500 µL of acid phenol was added to each suspension and the mixture was heated at 65 ^{o}C for 10 minutes with regular vortexing. The tubes were cooled on ice for 5 minutes, then centrifuged for 12 minutes at 10,000 g and 4 ^{o}C. The supernatant was transferred to a new tube and mixed with an equal volume of chloroform. The tubes were again centrifuged, and the supernatant transferred to a new tube. Nucleic acids were precipitated with 1/10th volume NaOAc and 2 volumes 100% ethanol. Tubes were stored at -20 ^{o}C for 30 minutes, the centrifuged for 30 minutes at 16,000 g and 4 ^{o}C. The supernatant was removed, the pellets were washed with 500 µL 70% ethanol, and the tubes centrifuged again for 20 minutes at 16,000 g and 4 ^{o}C. The supernatant was removed and the pellets allowed to air dry. The pellets were then resuspended in 20 µL of water. 2 µL of DNAseI buffer and 1 µL of DNAseI (NEB) were added to each tube. The tubes were incubated at 37 ^{o}C for 10 minutes. Next, EDTA was added to a final concentration of 5 mM and the tubes incubated at 75 ^{o}C for 10 minutes. Finally, the ethanol precipitation procedure was repeated to remove the EDTA.

The RNA was quantified using a Nano-Drop spectrophotometer. Total RNA was reverse transcribed using SuperScript III (Invitrogen) and gene specific primers for HEM13 and ACT1, according to the manufacturer's instructions. Approximately 1.5 μg of total RNA was loaded into each RT reaction. Following reverse transcription, qPCR was performed according to the manufacturer's instructions using the iQ SYBR Green Supermix (Bio-Rad) and 3 μL of cDNA in a 20 μL reaction. The qPCR reactions were monitored on a Bio-Rad iCycler. For each biological replicate, three technical replicates were performed for each of the gene specific primers. A dilution series was conducted for one sample, using both primer pairs, to measure the cycle efficiency. For each biological replicate, the technical replicates were averaged and the measured HEM13 level was normalized to the ACT1 level. The normalized expression was then averaged for the three biological replicates.

4.5 Tables

Table 4.1 Heme overexpression data

Enzyme Expression	Heme Overexpression	Relative Total Heme Level	Relative Total Porphyrin Level
No Enzyme	Empty Plasmid	1.00 ±0.14	1.00 ±0.14
No Enzyme	HEM2/3/12	1.20 ±0.20	3.91 ±0.52
No Enzyme	HEM2/3/12 + ALA	1.19 ±0.33	13.81 ±1.09
Low-Copy P450	Empty Plasmid	1.44 ±0.14	0.89 ±0.04
Low-Copy P450	HEM2/3/12	2.44 ±0.14	2.58 ±0.26
Low-Copy P450	HEM2/3/12 + ALA	3.35 ±0.13	6.87 ±0.15
High-Copy P450	Empty Plasmid	1.90 ±0.11	0.62 ±0.04
High-Copy P450	HEM2/3/12	3.97 ±0.17	1.37 ±0.17
High-Copy P450	HEM2/3/12 + ALA	9.54 ±0.94	3.41 ±0.71

4.6 Acknowledgements

I would like to thank R. Kumar for assistance with the microarray experiments. JLM Ruiz and L. Liu provided the HEM overexpression plasmids. J. Nielsen and V. Siewers provided invaluable advice in designing and analyzing the microarray experiments. S. Galanie assisted in the construction and characterization of the CYP2D6 strains. This work was supported by the National Science Foundation (USA) and Vetenskapsrådet (Sweden) through a Nordic Research Fellowship to JKM.

5 Design of *In Vivo* Feedback Controllers Using Synthetic RNA Switches

Portions of this chapter are adapted with permission from Michener JK, Thodey K, Liang JC, Smolke CD (2011) Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metab Eng.* May;14(3):212-22.

Cells are filled with biosensors, molecular systems that measure the state of the cell and respond by regulating host processes. In much the same way that an engineer would monitor a chemical reactor, the cell uses these sensors to monitor changing intracellular environments and produce consistent behavior despite the variation. While natural systems clearly derive benefit from pathway regulation, past research efforts in engineering cellular metabolism has focused on introducing new pathways and removing existing regulation, and researchers have rarely used genetically encoded biosensors as tools for optimizing and regulating heterologous pathways. In this chapter, I describe several ways in which biosensors could be used to introduce feedback control into metabolic pathways, providing dynamic control of metabolism to increase pathway efficiency and reliability.

5.1 Introduction

Natural metabolic pathways have evolved intricate regulatory networks to allow cells to respond to changing conditions (Bennett et al., 2008; Zaslaver et al., 2004) with a minimum of wasted energy (Chubukov et al., 2012). In contrast, engineered pathways rarely introduce any regulation beyond the use of an inducible promoter. More commonly, engineering efforts focus on removing regulation (Lutke-Eversloh & Stephanopoulos, 2007) rather than adding it anew. However, there are important reasons that cells use dynamic regulation, ranging from responding to variable environments to coping with stochastic variation in transcription and translation (Elowitz et al., 2002). However, efforts to replicate these types of controllers have been slow, due in large part to the difficulty of engineering a sensor that can recognize the desired pathway intermediate (Farmer & Liao, 2000; Zhang et al., 2012). RNA sensors have the potential to simplify the construction of dynamic controllers that sense and respond to novel metabolites.

In Chapter 2, I describe the optimization of an enzymatic reaction using a sensor, a theophylline-dependent RNA switch, to control expression of a reporter, GFP. Once a sensor has been used in this fashion for pathway optimization, it can then be linked to an actuator to dynamically regulate a metabolic pathway. In this context, an actuator refers to a molecule that affects the pathway being regulated. The combination of sensor and actuator forms a controller. Controllers may be divided into two broad categories, open loop and closed loop, based on the ligand being sensed. An open loop controller responds to a ligand that is distinct from the pathway being controlled; examples of open loop control include inducible promoters, native promoters that demonstrate a desired temporal response (Scalcinati et al., 2012), and quorum-sensing systems (Tsao et al., 2010). In contrast, closed loop controllers directly measure the current status of the pathway and respond accordingly (Zhang et al., 2012).

For a closed loop controller, the details of the linkage between sensor and actuator define the control law for the circuit. In a simple negative feedback loop, the sensor might respond to the product of an enzymatic reaction and directly regulate the expression of the associated enzyme; the enzyme serves as the actuator, and the feedback is roughly proportional. In a more-complicated controller, an increase in pathway output would lead to increased expression of a repressor that lowers the pathway expression (Ang et al., 2010). This system can be switched from proportional to integral control if the repressor is made to degrade at a constant rate. A truly constant decay rate is not possible, but pseudozeroth-order decay can be achieved through saturated enzymatic proteolysis with much greater magnitude than that of first-order dilution (Grilly et al., 2007). Proportional negative feedback can diminish the effects of disturbances to a pathway, but integral control, particularly in conjunction with proportional control, can completely reject disturbances. However, sensors linked to actuators require a great deal of design flexibility to ensure proper operation. In contrast to electronic controllers, the set point and control law of a metabolic controller are fixed by biochemical parameters of the component elements. If the metabolic engineer lacks sufficient tools to manipulate these parameters, the controller is unlikely to work reliably.

The primary advantage of using RNA switches to construct controllers is their modularity, in both sensing and actuation. As described previously, new ligand binding domains can be selected *in vitro* and integrated into existing RNA switch platforms. Similarly, taking an existing switch of the type used in Chapters 2 and 3 and using it to regulate a new gene is simply a matter of cloning the switch into the 3' UTR of the gene to be regulated. The characteristics of an RNA switch can also be quantitatively tuned (Liang JC, Chang AL, Kennedy AB, and CDS, in submission), aiding in the construction of controllers. There are two main weaknesses associated with the use of RNA switches in feedback controllers. First, the currently available switches have relatively small dynamic ranges, and a feedback controller requires a much larger range than does enzyme screening. Second, the RNA switch operates posttranscriptionally, so actuation requires a relatively slow translation step. If the input is changing on a faster time-scale than the switch can handle, the controller will not function properly. Protein transcription factors can be used in place of RNA switches, though with similar weaknesses. In fact, a transcription factor will show an even slower response, as actuation requires both transcription and translation. Future options for post-translation actuation could control enzyme activity (Guntas & Ostermeier, 2004), localization

(Czlapinski et al., 2008), or degradation (Davis et al., 2009). Controllers built using these actuators could respond on much faster time-scales. However, engineering a new post-translational actuator would be more challenging than constructing a transcriptional or post-transcriptional actuator.

In a classic example of engineering closed loop metabolic control, a transcription factor-based sensor has been used to detect excess flux through the glycolytic pathway (Farmer & Liao, 2000). The sensor, based on a natural *E. coli* promoter, linked increases in acetyl phosphate concentration to increases in transcription from its cognate promoter, *glnAp2*. This promoter was used to express genes that divert the glycolytic flux away from acetyl phosphate to an engineered lycopene biosynthetic pathway, producing a closed loop control system designed to maintain acetyl phosphate levels at a set value (Figure 5.1). Two genes in the lycopene pathway were placed under the control of either a strong constitutive promoter or the acetyl phosphate-responsive *glnAp2* promoter. Expression from the strong constitutive promoter led to growth arrest and low production of lycopene. When the genes were expressed in an acetyl phosphate-dependent manner, the cells grew normally and produced high titers of lycopene.

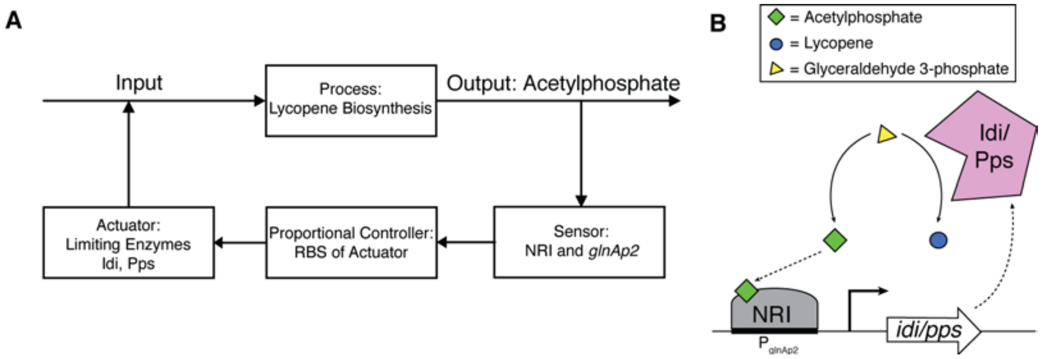

Figure 5.1 Dynamic regulation of the lycopene biosynthetic pathway. (**A**) The controller senses acetylphosphate, a signal that glycolytic intermediates are building up, and regulates *idi* and *pps,* the genes coding for rate limiting enzymes in lycopene biosynthesis. (**B**) Glyceraldehyde-3-phosphate can either be metabolized through glycolysis or converted into lycopene. An increase in acetylphosphate causes an increase in Idi and Pps, diverting flux from glycolysis into lycopene biosynthesis.

In this example, the controller functioned similarly to a stationary phase promoter and allowed the cells to switch on the pathway only after the cell density had reached a critical level. However, this controller also demonstrates several disadvantages of relying on natural transcription factors. The system can only respond to acetyl phosphate and shifting to regulate the pathway based on another metabolite would be difficult. Acetyl phosphate is quite close to central metabolism, and is embedded in an existing regulatory network. Analysis of the controller is therefore quite difficult, as it may interact with the endogenous network in unexpected ways. Additionally, the control law is difficult to modify, since the set point of the controller is largely fixed by the relationship between acetyl phosphate concentration and transcription factor activation. Varying regulatory elements, such as the RBS (Salis et al., 2009), may allow some tuning of the control response by changing the strength of the linkage between sensor and actuator, but much of the system response is fixed by the choice of components.

Due to the properties of its components, this type of controller is difficult to modify and analyze. Simple changes to the sensor and pathway, such as using an RNA switch and a pathway with an exogenous substrate, can make the controller much more tractable. The tunability of RNA switches would allow targeted modifications to the controller, and the use of an exogenous substrate minimizes the interactions between the system and its host. Built in this fashion, negative feedback controllers can be used in many different situations. In this chapter, I consider two such applications: the use of feedback control to accommodate disturbances to the concentration of the substrate of an enzymatic reaction, and its use to reduce retroactivity when connecting enzymes into pathways. I construct computational models of these controllers and explain the necessary design parameters for components used in such a controller. Finally, I discuss experimental characterization of potential components for use in a feedback controller.

5.2 Results

5.2.1 Minimizing disturbances from input perturbations

Substrate-limited enzymes can allow changes in the concentration of an initial reactant to propagate through a metabolic pathway. These changes could have deleterious consequences, including starving the cell of a necessary metabolite or overproducing a toxic intermediate. In a substrate-limited reaction, an increase in the substrate concentration would lead to an increase in the product concentration. A feedback controller could effectively convert this substrate-limited reaction into a substrate-independent step by sensing the increase in product concentration and decreasing the enzyme expression to bring the concentration back to the basal level. The simplest type of controller would use proportional feedback, where the enzyme expression was regulated in response to the current product concentration (Figure 5.2A). A more-complicated integral controller adds a repressor protein to effectively integrate the product concentration over time (Figure 5.2B). A true integral controller would require that this repressor protein degrade at a constant rate, so that the repressor concentration would be the integral of the difference between the product-dependent synthesis rate and the constant degradation rate. However, a constant degradation rate is not possible in a growing cell due to first-order dilution. The best that we can achieve is to degrade the repressor enzymatically using a saturated protease. The protease will introduce an effectively constant decay term, and if the magnitude of this term is significantly larger than the first-order dilution, the system will behave as an integral controller.

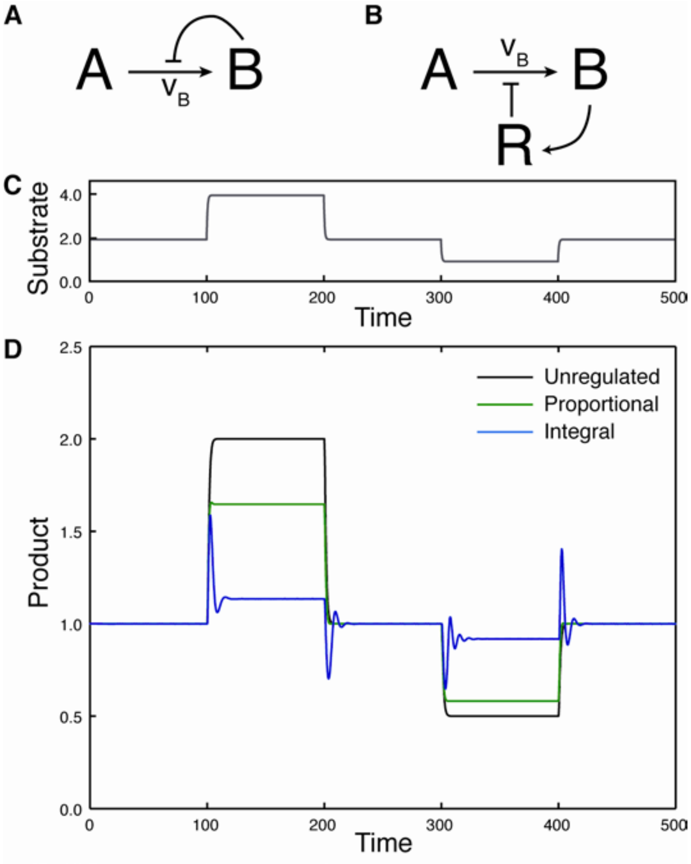

Figure 5.2 Feedback control can reduce the effects of a disturbance to the pathway input. (**A**) An enzyme is used to convert a substrate, A, into a product, B. Proportional feedback control would regulate enzyme expression in response to the current product concentration. (**B**) An integral feedback controller uses a repressor protein to integrate the product concentration. Synthesis of the repressor is induced by the product, and the repressor then reduces expression of the enzyme. (**C+D**) Feedback controllers can reduce the effects of a change in substrate concentration. The unregulated module (black) has a linear dependence on the substrate concentration. Proportional regulation can reduce this effect but cannot eliminate it entirely. An integral controller could completely reject this type of disturbance. However, true integral control requires the repressor protein be degraded at a constant rate, which is not possible in the presence of dilution due to growth. The presence of a small-magnitude first-order decay term prevents perfect adaptation (blue).

I have modeled the effects of introducing these two controllers into a substrate-limited enzymatic reaction that is challenged by fluctuations in the substrate concentration (Figure 5.2C). The cells are grown in continuous culture with varying concentrations of substrate added to the media feed. In the absence of a control system, changes in the substrate concentration propagate directly to the product concentration (Figure 5.2D). For example, doubling the substrate concentration will double the product concentration.

Proportional regulation can reduce, but not eliminate, these disturbances. With a proportional controller, a twofold change in the substrate concentration produces a 1.6-fold change in the product concentration. The integral controller further reduces the disturbance, but only at the cost of significant transient fluctuations as the system adapts to the new equilibrium. Additionally, the presence of first-order degradation due to dilution prevents the controller from behaving as a true integral controller. As a result, the system is not able to perfectly adapt to a changing substrate concentration, and there is still a steady-state disturbance. Neither of these controllers, proportional or integral, is perfect, but either controller would reduce the effect as an input change propagates through a metabolic pathway.

5.2.2 Minimizing retroactivity in a biosynthetic pathway

In addition to using controllers to prevent disturbances from propagating along a pathway, they can also be used to accommodate disturbances that propagate up a pathway, a situation described as retroactivity. Metabolic pathways can demonstrate retroactivity when two reactions in a pathway compete for the same substrate or cofactor. Changing the specifics of the downstream reaction, such as its cofactor utilization, can propagate up a pathway to affect the enzyme that shares that cofactor. In the case of cofactor-dependent retroactivity, the host cell will generally already use control systems to sense cofactor depletion and react accordingly. While this adaptation might not be sufficient to counteract the new load, as in the case of heme biosynthesis (described in Chapter 4), solutions typically take the form of augmenting the native response. Minimizing the retroactivity from competition over a mutual substrate, however, will often require the introduction of an entirely new controller. This controller must sense the concentration of the mutual substrate and act to keep that substrate concentration constant despite varying rates of consumption.

I first considered a simple example where a feedback controller was used to maintain the concentration of an intermediate, B, as the rate of consumption was increased (Figure 5.3A). In my model, the parameter ‘m’ denotes the activation ratio of the sensor, roughly equivalent to the gain of the proportional controller. If m = 1, there is no feedback and the enzyme that converts A into B is working at its maximum reaction rate. When m = 10, the enzyme expression can vary by a factor of 10. I used realistic parameters for the RNA switch and modified the basal expression level of the regulated enzyme variants to ensure that, in the absence of a downstream reaction, the various controllers produced the same output (Figure 5.3B). The controller is able to reduce, but not eliminate, the variation in the concentration of the intermediate, B (Figure 5.3C). As a result, the pathway would show less retroactivity and an enzyme that competes for B would produce more-consistent results.

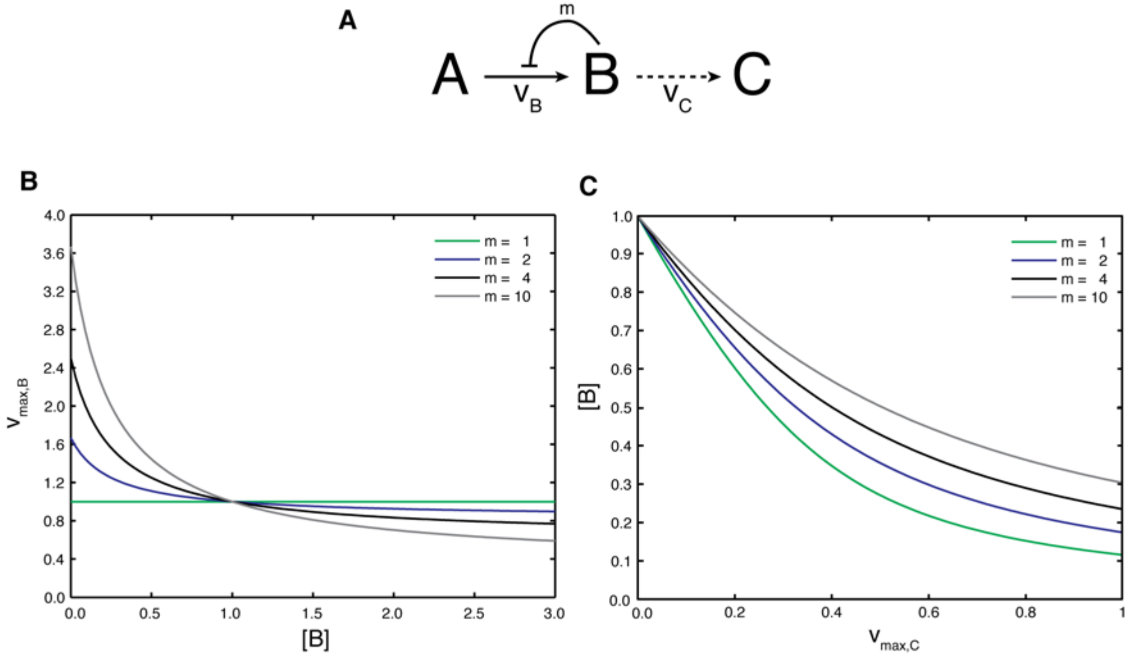

Figure 5.3 A proportional feedback controller can reduce the retroactivity of a downstream reaction. (**A**) Diagram of the model system. A substrate, A, is converted to an intermediate B. The consumption rate of B is varied. A feedback controller attempts to maintain the concentration of B at a fixed value despite this variable consumption. (**B**) In an unregulated pathway, a small increase in the rate of conversion of B to C can have a large effect on the concentration of free B. A feedback controller can respond by increasing the rate at which B is synthesized. As a result, the concentration of B is less dependent on the rate of the downstream reaction.

The concentration of B is reported relative to its basal rate in the absence of any downstream reaction (where loss of B occurs only through dilution). The downstream reaction velocity is normalized by the basal dilution rate; a reaction velocity of 1 means that the rate of B→C is equal to the basal dilution rate of B.

Next, I expanded my simple model to a more-realistic pathway (Figure 5.4A), in which a common substrate, B, is used to make two intermediates, C and D. These two intermediates then condense into the product, E. This pathway structure occurs in the biosynthesis of norcoclaurine (Nakagawa et al., 2011). In this type of pathway, optimizing one branch of the pathway, perhaps by evolving the enzyme that makes D to improve its activity, could lead to lower production of the final product E by reducing the concentration of the common substrate B and therefore the rate of reaction in the opposite branch making C. A feedback controller can sense the increased consumption of B and respond by increasing the rate of synthesis of B.

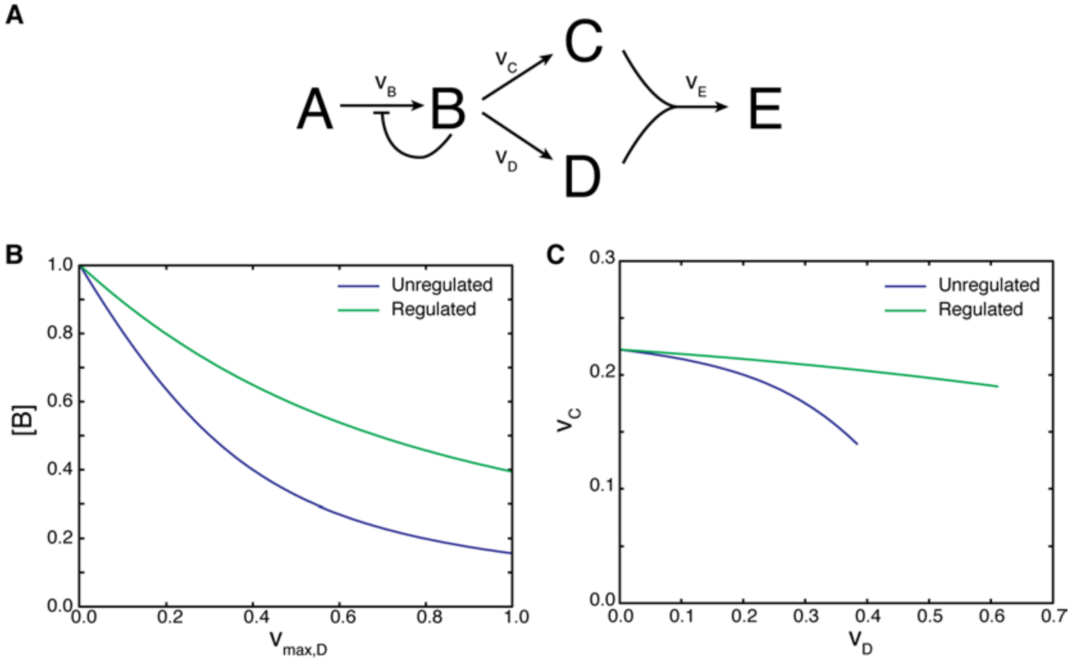

Figure 5.4 Feedback control can accommodate competition over a common substrate. (**A**) Diagram of the model pathway. A single intermediate is used to produce two different compounds that condense into the final product. (**B**) As one branch of the pathway is optimized, the maximum rate of that reaction will increase. In the absence of regulation, the concentration of the common substrate, B, will decrease. (**C**) As a result, the rate of the competing reaction will decrease and quickly become limiting. Since the overall flux is limited by the slower of the two reactions, an improvement to one branch can lower the concentration of the final product. However, a feedback controller can respond to lessen the decrease in the concentration of B. As a result, the competing reaction is not affected by the optimization. The optimization may not increase pathway output, but it will not cause a large decrease in output.

I considered a feedback controller with a 10-fold dynamic range. As in the simpler model, the feedback controller was able to lessen the change in the concentration of B as the maximum reaction velocity converting B to D was increased (Figure 5.4B). As a result, the rate of the competing reaction, v_c, was nearly constant despite the changes in $v_{max,D}$. The retroactivity has not been eliminated completely, and we would not expect a proportional controller to be able to do so. However, the behavior of the enzymatic module that converts A to B is now much less sensitive to the characteristics of the downstream processes that might be connected to it, allowing more-predictable composition of disparate enzymes.

5.2.3 Experimental characterization

I planned to build a negative feedback controller using a theophylline-dependent RNA OFF switch to control expression of the caffeine demethylase that I evolved in Chapter 2. However, the characteristics of the system components, both the switch and the enzyme, were ill suited to this application. As I describe in Chapter 4, the enzymatic activity of the engineered demethylase is not limited by the transcription rate. Decreasing the plasmid copy number by a factor of ~ 10 made little difference to the amount of theophylline produced. I decided to investigate this relationship further, to see whether a change in mRNA stability, such as that produced by an RNA switch, would lead to a change in enzymatic activity.

I built a series of constructs to express the caffeine demethylase from high- and low-copy plasmids behind a series of promoter variants whose transcription rates varied by ~ 12 fold (Nevoigt et al., 2006). As expected, changing the promoter strength on a high-copy plasmid had little effect (Figure 5.5). However, combining a weak promoter with a low-copy plasmid reduced the total enzyme activity. Unfortunately, at low expression the enzymatic activity shows a logarithmic relationship with the expression level. As a result, an RNA switch that can produce a twofold change in the expression level might only produce a 1.5-fold change in enzymatic activity. This logarithmic relationship effectively reduces the dynamic range of the RNA switch used for control.

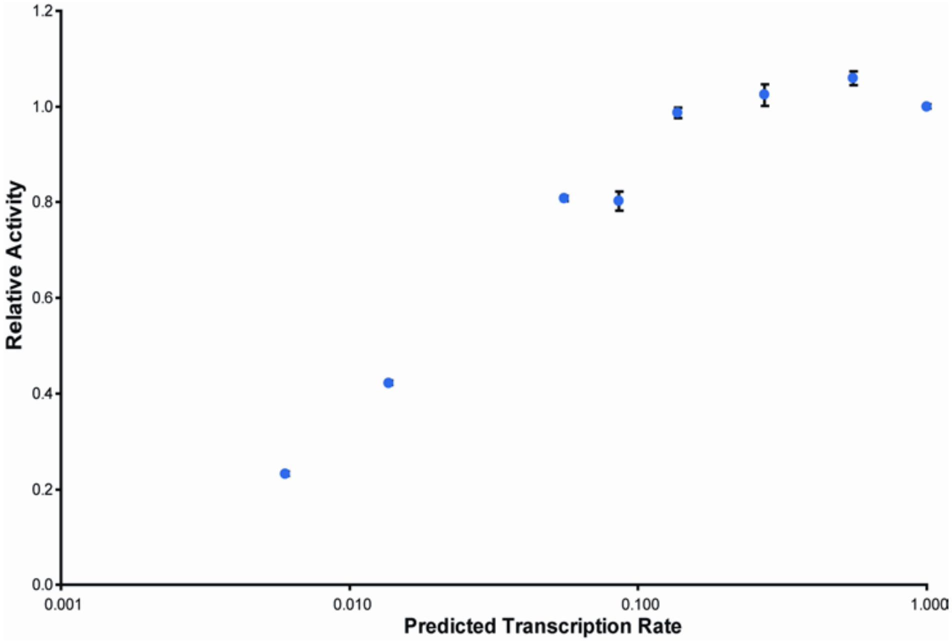

Figure 5.5 Enzymatic activity shows a sublinear dependence on transcription rate. A caffeine demethylase was expressed from a variety of constructs, including high- and low-copy plasmids as well as a range of promoter strengths. The predicted transcription rate is based on similar results for GFP, where fluorescence is presumed to be a linear function of the transcription rate. The enzymatic activity, measured as the final theophylline concentration, was independent of concentration over a roughly tenfold change in transcription rate. At lower expression levels, changes in transcription resulted in changes in activity, but the relationship was logarithmic.

Further complicating the construction of a feedback controller, the existing OFF switches do not show a large dynamic range. I tested a previously described OFF switch (Win & Smolke, 2007) and constructed two mutated switches that should show constitutive expression, either at the high or low end of the OFF switch dynamic range. When GFP was expressed under the control of the OFF switch, I saw a ~ 1.5-fold change in fluorescence between the uninduced and fully induced switch (Figure 5.6), which was consistent with the controls. However, the dynamic range of the OFF switch is less than ideal for use in a feedback controller (Figure 5.2B).

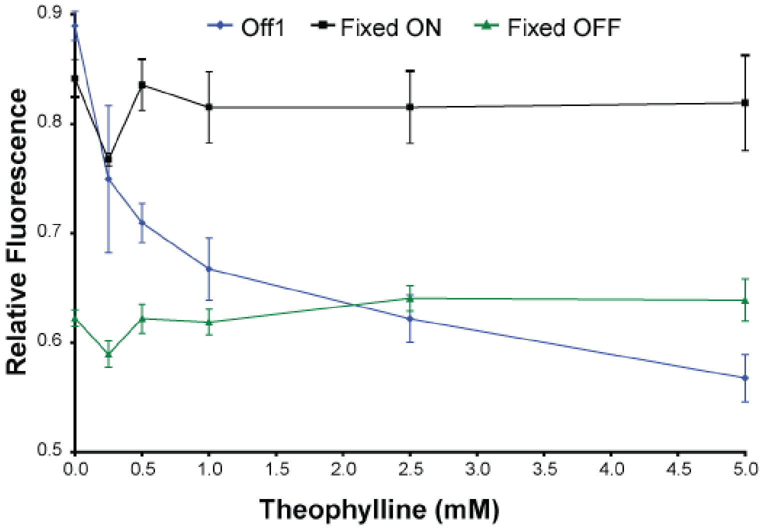

Figure 5.6 Activation of a theophylline-dependent OFF switch. GFP was expressed from a constitutive promoter with the OFF switch placed in the 3' UTR. Samples were grown in the presence of varying amounts of theophylline and assayed for geometric mean fluorescence using a flow cytometer. Two controls (shown in black and green) have mutations in the RNA switch that lock the switch into the ON and OFF conformations.

To aid in my modeling work, I also sought to measure the internal concentrations of the relevant metabolites. Previous results have assumed that feeding theophylline to cells, either yeast or bacteria, produces an internal concentration that is significantly lower than the concentration in the culture media but the justification for such assumptions is either inferred (Chen & Ellington, 2009) or based on a misinterpretation of experimental data (Koch, 1956). Measurements of the intracellular methylxanthine concentration are technically very difficult: since the intracellular concentration is so low, any carry-over of supernatant will drastically skew the measured concentration. To minimize theophylline export during the wash steps, I immobilized the cells on a membrane filter and rinsed them with 3 volumes of PBS in < 1 minute (Wittmann et al., 2004). No theophylline was observed when the cells were simply washed by centrifugation, demonstrating that theophylline was exported from the cell on a time-scale shorter than that required for centrifugation. The cells attached to the membrane filter were then lysed in boiling buffered ethanol (Gonzalez et al., 1997), concentrated to 50 μL, and measured by LC-MS.

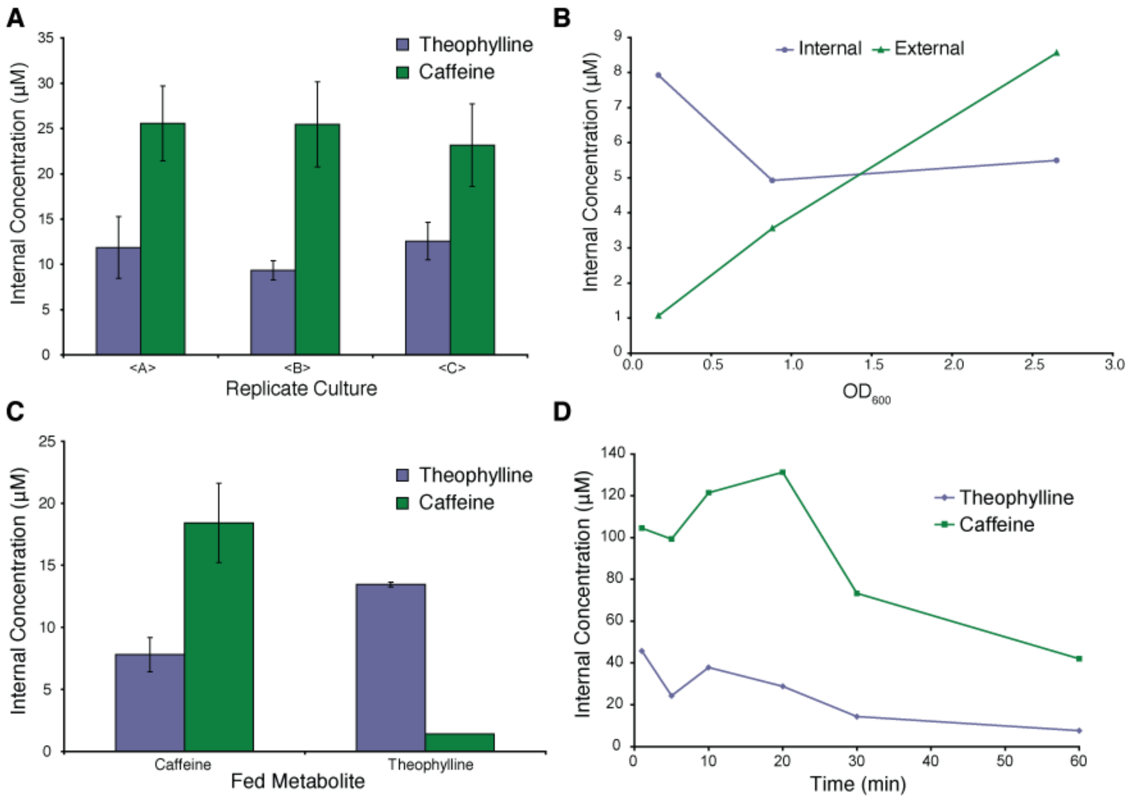

Figure 5.7 Measurements of internal metabolite concentrations. (**A**) When fed 1 mM theophylline and caffeine, internal concentrations were ~ 40-fold lower for caffeine (green) and ~ 80-fold lower for theophylline (blue). Technical and biological replicates showed similar variance, ~ 15–20%. (**B**) As expected, cultures at a higher OD accumulate more theophylline in the supernatant. However, the internal concentrations are similar, demonstrating that the measured internal concentration is not simply carry-over from the supernatant. (**C**) When fed 1 mM of caffeine, a variant of the caffeine demethylase sees ~ 18 µM caffeine and produces ~ 8 µM theophylline. When fed 1 mM theophylline, the same cell sees ~ 13 µM theophylline intracellularly. These results are consistent with comparative measurements of fluorescence when feeding or producing theophylline. (**D**) A timecourse of the intracellular metabolite accumulation suggests passive import and active export of caffeine and theophylline. 1 mM of each metabolite was added to the culture at t=0 and samples were taken at the indicated timepoints.

I consistently observed a ~ 40-fold drop in caffeine concentration and an ~ 80-fold drop in theophylline concentration across the yeast cell membrane (Figure 5.7A). When theophylline was produced intracellularly, I observed similar internal theophylline concentrations in cultures of variable density, despite different levels of extracellular theophylline accumulation (Figure 5.7B). The relative measurements of the internal concentration of cells fed or producing theophylline were consistent with fluorescence measurements of the respective cultures (Figure 5.7C, Figure 2.5). When I spiked a culture with 1 mM caffeine and theophylline and followed the intracellular concentration over time, I saw a large, rapid increase in the internal concentration followed by a slow decay back to the steady-state concentration (Figure 5.7D). Previous research has demonstrated that transporter mutants can show increased caffeine sensitivity (Parsons et al., 2004). Several different pieces of data — my metabolite measurements, the connection between caffeine sensitivity and export protein knockouts, the similarity between the internal concentrations of cells fed or producing theophylline despite significantly different external concentrations, and my observation that slow washes led to negligible theophylline recovery — all support the hypothesis that caffeine and theophylline are actively exported from the cell, and as a result the internal concentration is significantly lower.

5.3 Discussion

I made several simplifying assumptions in my first computational model, describing the use of controllers to respond to fluctuations in the input. First, I linearized several Hill functions, consistent with a situation in which both the enzyme converting substrate to product and the sensor controlling the product-dependent synthesis of the repressor protein are substrate limited. These assumptions made the model more tractable, but as a result the model only represents a subset of possible systems. Additionally, the model does not include the time delays involved in transcription and translation. These delays would slow the response of the regulated systems and possibly destabilize the integral controller.

Integral control requires an approximately constant decay rate. To achieve such a rate, the repressor must be degraded by a saturated protease at a rate much greater in magnitude than that due to first-order dilution. As a result, this controller would be extremely wasteful, producing and degrading large numbers of repressor proteins, and attempts to improve the integral controller would also increase the load. As I demonstrated in Chapter 4, when a pathway places a burden on its host, the host's response to that burden can affect the pathway. A better integral controller would use less-costly posttranslational modifications such as phosphorylation or methylation as a control variable in the place of protein synthesis and degradation (Yi et al., 2000).

A proportional feedback controller that uses a sensor whose transfer function is a non-cooperative repressed Hill function, such as the RNA switch that I characterized for this work (Figure 5.6), will be limited by the shape of that transfer function, even at the limit of sensors with a large dynamic range. As the dynamic range of the controller increases, the sensor output approaches a simple hyperbola. Even assuming that the researcher has complete flexibility to tune the sensitivity of the sensor and the basal expression of the enzyme, certain trade-offs will still be unavoidable. The expression level will be fixed by the requirement that the controller produce a certain behavior in the absence of any load (Figure 5.3B). Tuning the sensitivity will move the basal position along the hyperbola (Figure 5.8). If the basal substrate concentration is low compared to the binding constant of the sensor, the controller will produce a strong response to any disturbance. However, the controller will only be able to respond to small disturbances before saturating the sensor. If the basal substrate is high relative to the sensor binding constant, nearly the entire dynamic range of the sensor will be available, and the controller will be able to produce large changes in output. Unfortunately, the sensitivity will be low, and the concentration of the controlled substrate will have to change dramatically to produce a large change in output. Sensors with cooperative transfer functions would produce more-effective controllers by producing a large change in output with higher sensitivity.

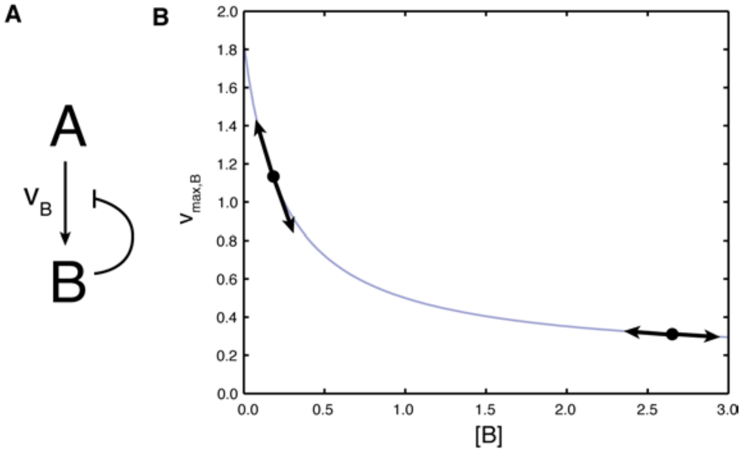

Figure 5.8 Transfer function of the sensor limits controller performance. (**A**) An enzyme converts substrate A to product B. Expression of the enzyme is regulated by the concentration of B, such that increases in the concentration of B lead to lower expression of the enzyme that produces B. (**B**) Transfer function of a non-cooperative RNA switch. A larger dynamic range will lower the enzyme expression at large [B] but does not change the basic shape of the transfer curve. Tuning the affinity of the switch binding domain allows the researcher to choose where on the curve to set the basal expression level. If the basal state is significantly below the EC_{50} of the switch, the controller will display high sensitivity but can use only a small fraction of the potential dynamic range of the switch. If the basal state is well above the EC_{50}, the controller will be able to use the entire dynamic range of the switch, but will require a large change in [B] to produce a small change in $v_{max,B}$.

Due to the logarithmic scaling between predicted mRNA expression and enzymatic activity, the caffeine demethylase is not an ideal enzyme for use in a feedback controller. The existing OFF switches have relatively small dynamic ranges as well as hyperbolic transfer functions. In combination, the characteristics of these components severely limit the effectiveness of any feedback controller built from them. My computational results demonstrate that a moderate dynamic range (2–4-fold) is necessary for measurable controller performance, so the dynamic range of the OFF switch will need to be increased in order to build an effective controller. A further increase in the switch range will be necessary if the current demethylase is used in the controller, perhaps by adapting the screening strategy used for ON switches (Liang JC, Chang AL, Kennedy AB, and CDS, in submission) to screen for improved OFF switches. Alternately, a different enzyme with a linear rather than logarithmic relationship would allow the use of a switch with a smaller activation ratio.

Despite the difficulties that I faced in implementing my controller designs, I believe that efforts to build and characterize dynamic controllers will play an increasingly important role in constructing predictable metabolic pathways. While constructing novel controllers, we will also need to develop a deeper understanding of exactly how controllers, both native and engineered, work to increase the pathway productivity and reliability. Some controllers may function by reducing variability, either between cells or over time, while others mainly provide benefits by programming the bulk temporal expression of genes in the pathway. Differentiating between subtle differences such as these will aid in the forward design of future controllers for novel situations.

5.4 Methods

5.4.1 Variable enzyme expression

The enzyme yCDM4b was initially expressed from a high-copy plasmid with a strong TEF promoter, as described in Chapter 2. First, the enzyme was cloned into a low-copy centromeric plasmid, following the same protocol as in Chapter 2. Next, the promoter in each plasmid was replaced by a series of TEF promoter mutants (Nevoigt et al., 2006). In total, I tested TEF mutants 4 (65% native activity), 3 (32% native activity), and 7 (16% native activity) in the high-copy plasmid and mutants 4, 7, and 2 (7% native activity) in the low-copy plasmid. Cultures containing these enzyme expression constructs were grown in the presence of caffeine as described previously and then assayed for theophylline production by HPLC.

5.4.2 RNA switch characterization

A previously described OFF switch (5’-AAACAAACAAAGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGTTGCTGAtACCAGCATCGTCTTGATGCCct

TGGCAGCAGTGGACGAGGACGAAACAGCAAAAAGAAAAATAAAAA-3') was constructed in two overlapping DNA oligos and cloned between the XhoI and AvrII sites of pCS1585. Two variants, indicated by the lowercase letters in the sequence above, were also used as controls. The double mutant C89T/T90A fixes the switch in the ON confirmation, where expression is high. A triple mutant, T58G/C89T/T90A locks the switch OFF. These variants were constructed in a similar fashion using overlapping oligos. Cells containing these plasmids were grown in the presence of variable amounts of theophylline and assayed for fluorescence by flow cytometry. The yeast cultures were diluted 4x into water and assayed for fluorescence in 96-well plates on a Beckman Quanta flow cytometer. Cells were excited at 488 nm and GFP fluorescence was measured at 525 nm. Samples were gated first by electronic volume and side scatter to capture the cell population and then by fluorescence to remove the outliers with significantly low fluorescence. Approximately 8,000 cells were analyzed for each culture. The geometric mean fluorescence, normalized by the electronic volume, was then compared to cells containing an unregulated copy of GFP.

5.4.3 Intracellular metabolite assays

Yeast cultures were grown in the appropriate culture medium to the desired density, approximately 20 OD*mL of cells per extraction. Fewer cells resulted in a lower signal, and more cells clogged the filter paper and slowed the wash steps. The cells were then applied to a vacuum filtration unit containing a cellulose nitrate filter with a 25 mm diameter and 0.45 μm pore size. The filter disks were washed with 90 mL of phosphate-buffered saline. The filter was then removed from the holder, transferred to a clean tube, and vortexed with 500 μL of 80 °C buffered ethanol (75% ethanol, 10 mM HEPES pH 7.0) to wash the cells off the filter. The resulting cell suspension was incubated at 80 °C for 5 minutes, then centrifuged at 10,000 g for 10 minutes. The supernatant was transferred to a fresh tube and concentrated to < 50 μL, centrifuged again to remove particulates, and adjusted to 50 μL before being analyzed by LC-MS. Based on previous data (Hans et al., 2001), I estimated that the intracellular volume of our standard yeast strain is 1.3 μL/(OD•mL). By calculating the internal volume of each sample, I can estimate the dilution factor and therefore the internal concentration.

5.4.4 HPLC and LC-MS Characterization

HPLC samples were analyzed on a Poroshell 120 SB-C18 2.1 x 50 mm, 2.7 μm column (Agilent). The mobile phase was 0.50 mL/min of 20% methanol/80% water with 0.1% acetic acid. Using the Poroshell column, 3 μL of each sample was injected onto the column and theophylline eluted at 0.70 minutes.

LC-MS samples were analyzed on an XDB-C18 2.1 x 50 mm, 3.5 μm column (Agilent Technologies). 5 μL of each sample was injected onto the column. The mobile phase was 0.35 mL/min of 15% methanol/85% water with 0.1% acetic acid. Theophylline eluted at 1.65 minutes and was detected by mass spectrometry (Agilent 6320 Ion Trap) as a peak with m/z of 181.

5.5 Modeling

5.5.1 Modeling an Input Disturbance

In this model, an enzyme E converts a substrate A into a product B. The substrate is fed at a constant rate, F_c, and the cells grow and dilute at a rate k_d. The enzyme is in a linear regime, such that the rate of enzymatic conversion of A to B is linearly proportional to A. In the unregulated case, the enzyme is produced at a constant rate. With proportional regulation, the rate of enzyme synthesis is dependent on the concentration of the product, B, and the inhibition constant of the RNA switch, K_i. For integral regulation, a new component, a repressor, is introduced. The enzyme synthesis rate is dependent on the concentration of the repressor and its inhibition constant, K_R. Repressor production follows a noncooperative Hill function, dependent on the concentration of B. The repressor is degraded both by dilution and by a protease. The repressor concentration, R, is much larger than the binding affinity of the protease, K_{deg}, so proteolysis effectively occurs at a constant rate k_R. Reduced to equations, this becomes:

$$\dot{A} = F_C - k_B EA - k_d A$$
$$\dot{B} = k_B EA - k_d B$$
$$\dot{E}_U = S_{E,U} - k_d E_U$$
$$\dot{E}_P = \frac{S_{E,P}}{1 + B / K_i} - k_d E_P$$
$$\dot{E}_I = \frac{S_{E,I}}{1 + R / K_R} - k_d E_I$$
$$\dot{R} = \frac{S_R B}{B + K_D} - \frac{k_R R}{R + K_{\deg}} - k_d R$$

To simulate a disturbance in the input concentration, F_C was varied as a function of time. The parameters used were: $F_C = [1.5, 3, 6]$, $k_B = 1$, $k_d = 1$, $S_{E,U} = 0.5$, $S_{E,P} = 1$, $S_{E,I} = 4$, $K_i = 1$, $K_R = 0.2$, $S_R = 1$, $K_D = 0.1$, $k_R = 10$, $K_{deg} = 0.001$. The differential equations were solved in MATLAB using ode23s.

5.5.2 Modeling a Variable Load

In this model, an enzyme E converts a substrate A into an intermediate B. The intermediate B then reacts, at a variable rate, to produce the final product C. The steady-state concentration of enzyme is dependent on the concentration of B and the dynamic range of the sensor, m. The sensor can switch from full expression to a fraction 1/m of the maximal expression. Thus, at m = 2 the sensor range is 100% to 50% and at m = 10 the range is 100% to 10%. Within that switching range, the transfer curve is a repressed Michaelis-Menten equation with inhibition constant K_B. However, the enzyme expression must be normalized to ensure that the steady-state concentration of B in the absence of any load, termed B_f, is the same for each switch. Therefore, the enzyme expression is normalized to ensure that $E(B_f) = 1$.

$$\dot{B} = \frac{k_B E A}{A + K_A} - \frac{v_{\max,C} B}{B + K_B} - k_d B$$

$$B_f = \frac{k_B}{k_d} \frac{A}{A + K_A}$$

$$E = \frac{\frac{1}{m} + \left(\frac{m-1}{m}\right)\left(\frac{1}{1 + B/K_I}\right)}{\frac{1}{m} + \left(\frac{m-1}{m}\right)\left(\frac{1}{1 + B_f/K_I}\right)}$$

Or, at steady state:

$$\dot{B} = k_d B_f E - \frac{v_{\max,C} B}{B + K_B} - k_d B = 0$$

The parameters used were: $B_f = 1$, $k_d = 0.36$, $K_B = 0.25$, $K_I = 0.25$. In Figure 5.3A, I plot E as a function of B for various values of m. In Figure 5.3B, I used the fzero function in MATLAB to numerically solve for the steady-state concentration of B as $v_{\max,C}$ was varied from 0 to 1.

5.5.3 Modeling a Branch Point

In this final model, the previous example of a variable load was extended to two competing reactions for the intermediate B. B can be converted to C with rate constants $v_{max,C}$ and K_C or to D with rate constants $v_{max,\ D}$ and K_D. The rate of the first reaction, $v_{max,C}$, is kept constant while the rate of the second reaction, $v_{max,D}$, is varied. In comparison to the previous model, the normalization factor for E must be adjusted to account for the constant conversion of B into C. Otherwise, the modeling is very similar. Reduced to equations, this becomes:

$$\dot{B} = \frac{k_B EA}{A + K_A} - \frac{v_{\max,C} B}{B + K_C} - \frac{v_{\max,D} B}{B + K_D} - k_d B$$

$$B_f = \frac{k_B}{k_d} \frac{A}{A + K_A}$$

$$E = \frac{\frac{1}{m} + \left(\frac{m-1}{m}\right)\left(\frac{1}{1 + B / K_I}\right)}{\frac{1}{m} + \left(\frac{m-1}{m}\right)\left(\frac{1}{1 + B_f / K_I}\right)} \left(\frac{1}{1 + \frac{v_{\max,C}}{d(B_f + K_C)}}\right)$$

Or, at steady state:

$$\dot{B} = k_d B_f E - \frac{v_{\max,C} B}{B + K_B} - \frac{v_{\max,D} B}{B + K_D} - k_d B = 0$$

The parameters used were: $B_f = 1$, $k_d = 0.36$, $K_C = 0.25$, $K_D = 0.125$, $K_I = 0.25$, $v_{max,\ D} = 0.25$. In Figure 5.4B, I solved this equation numerically for B using fzero in MATLAB as $v_{max,D}$ was varied between 0 and 1. In the unregulated case, m = 1, and in the regulated case m = 10. In Figure 5.4C, I used the resulting concentration of B to calculate the actual reaction rates, v_C and v_D, over the same range of $v_{max,D}$.

5.6 Acknowledgements

I would like to thank Richard Murray, Domitilla Del Vecchio, and the Caltech Biocontrols group for useful discussions on this chapter. This work was supported by the National Science Foundation (CBET-0917638; fellowship to JKM), the Institute for Collaborative Biotechnologies (U.S. Army Research Office DAAD19-03-D-0004), and the Alfred P. Sloan Foundation (fellowship to CDS).

6 Conclusions and Future Prospects

6.1 Applications of RNA switches for metabolic engineering

In this thesis, I have discussed three ways in which synthetic RNA switches could advance efforts in metabolic engineering. RNA switches can be used to identify new enzymes from cDNA pools, to evolve enzymes and increase their activity in a heterologous host, and to regulate enzymes and improve their predictability. These applications can form a continuous process, where a single ligand-binding domain can be integrated into a variety of switch platforms and used sequentially to identify, optimize, and regulate an enzyme.

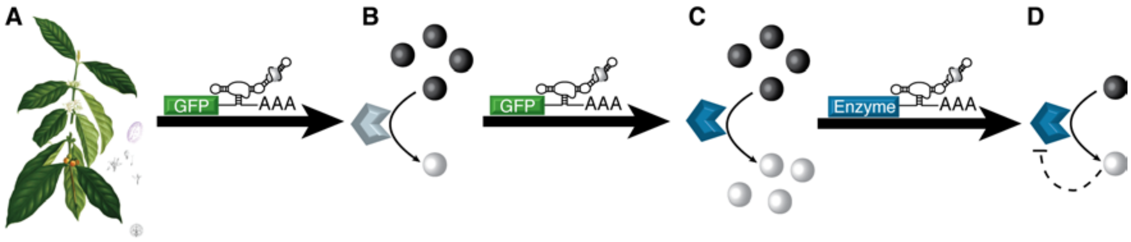

Figure 6.1 A synthetic RNA switch can be used to identify, optimize, and regulate a biosynthetic enzyme. (**A**) Starting from an uncharacterized source, such as plant cDNA or metagenomic DNA, a functional screen using an RNA switch can (**B**) identify an enzyme capable of producing the target molecule. (**C**) The same switch can then be used to optimize the enzyme activity in a heterologous host. (**D**) Finally, a variant of the RNA switch can be used to regulate the expression of the enzyme, allowing the construction of a synthetic feedback controller.

I have demonstrated that synthetic RNA switches can be used as biosensors for *in vivo* enzyme evolution. I was able to perform seven iterative rounds of directed evolution, screening for increases in enzymatic activity using an RNA switch and ultimately increasing the activity of the enzyme by more than 30-fold. Assaying mutants in clonal culture allowed me to discriminate small changes in the switch output, while screening at the single-cell level allowed me to screen libraries of more than a million members. Given the modular nature of the RNA switch that I used (Win & Smolke, 2007), I expect that these techniques will be generally applicable to a wide range of enzymes and pathways. Other RNA switches, such as those developed for *E. coli*, could allow the use of similar strategies in other host organisms (Topp et al., 2010).

I also developed a process for using an RNA switch to perform functional screening of a cDNA library. While my screening attempts were unsuccessful, I believe that the limitation lay in the choice of a target that may be difficult to functionally express in a heterologous host. I expect that the same process could successfully be used to identify a more-tractable target, such as a cytosolic enzyme. Additional improvements to the library construction process and the RNA sensor would also increase the chances of success.

Finally, I have described several ways in which RNA switches could be used to build synthetic regulatory controllers for metabolic pathways. These controllers would allow researchers to engineer more-predictable pathway output in the presence of various disturbances, such as changes to the pathway input or the addition of a competing reaction. Unfortunately, the components currently available for use in a controller are not well suited to the task. New components, including both new switches and new model pathways, will be necessary before these applications can be tested experimentally.

6.2 Constructing new RNA switches for metabolic engineering

The main barrier to the broad application of RNA switches lies in the limited number of available ligand-binding domains that recognize metabolites of interest. While the SELEX technique for selecting new binding domains *de novo* is well established, the process is still quite tedious. Additionally, traditional SELEX has several inherent disadvantages. First, the target molecule must be conjugated to a chromatographic resin. This coupling process is dependent on the presence of reactive groups in the target molecule, and there are some molecules that are simply incompatible with the available coupling chemistries. Additionally, the coupling process removes from functional group that could potentially participate in binding. Second, the selections take place on the surface of the resin, with the target molecule fixed in a specific orientation. Fixing the orientation of the target limits the potential binding modes, possibly forcing the target into a conformation that is difficult to bind. Limiting the conformational entropy of the target may also produce false positives in the selection process, such as binding domains that can only recognize the target when it is attached to the resin. The presence of the resin also requires tedious counterselections to avoid selecting binding domains that recognize the scaffold and not the target. Finally, the applications described above would be more effective if they could draw on a pool of binding domains with identical specificity but varying affinity. Screening for novel enzymes would benefit from an RNA switch with high affinity, while a feedback controller regulating an optimized enzyme would be more effective using a lower-affinity switch. New methods for rapidly selecting and tuning binding domains in solution would be very valuable (Liang, 2012).

A second barrier concerns the limited dynamic range of existing synthetic RNA switches. Screening enzyme variants in single cells requires a large signal-to-noise ratio in order to eliminate inactive enzyme mutants. As a result, any switch has a minimum sensitivity threshold. An enzyme must be sufficiently active so as to cross this threshold before the switch can be used for enzyme discovery and optimization. All else being equal, a switch with a larger dynamic range has a lower sensitivity threshold. Additionally, a small dynamic range brings with it the risk of saturating the switch when evolving or controlling highly active enzymes. The smaller the dynamic range of the switch, the more important our ability to tune the switch sensitivity becomes, and vice versa. Larger dynamic ranges would hide a number of other weaknesses in switch construction.

Finally, when switches are used as sensors in metabolic controllers, the shape of the transfer curve becomes particularly important. In these applications, switches that display cooperative binding would be particularly valuable, as they would allow a controller to respond with high sensitivity over a large dynamic range. Our ability to tune the switch parameters and the linkage between the switch and enzyme will also become more important. If the switch transfer curve is ideal but the binding affinity is poorly chosen, the controller will not function effectively. Alternately, if the switch properties, such as binding affinity and transfer curve, are not matched to the enzyme expression and activity, the basal enzyme activity will be suboptimal. Useful metabolic controllers will require the ability to precisely specify characteristics of the sensor its linkage to the actuator.

6.3 Analysis of heterologous pathways

Many biosynthetic pathways are highly productive in their native host but fail to function effectively when transferred to a new host. In Chapter 2, I described methods by which a pathway can be evolved to adapt it to its new host. However, in many situations it may also be necessary to adapt a host to its new pathway. Therefore, in Chapter 4 I demonstrated one technique for accommodating the specific stress of a heterologous pathway, using measurements of mRNA levels to identify the stress and targeted overexpression of native genes to alleviate that stress.

The first step in this process, identification of the stress or stresses, can be very challenging. First, a typical heterologous metabolic pathway will induce many changes in its host, and identifying a small set of significant stresses among all the data can be difficult. I have shown that analysis of multiple pathway variants can be helpful in this process, but a deeper knowledge of the host stress response and better models and bioinformatics tools would greatly aid in this effort. Second, a given measurement technique, such as my choice to measure mRNA levels, will miss stresses that provoke responses through other mechanisms, such as changes in translational or posttranslational control. As before, a better understanding of the underlying biology will allow us to focus on the specific biochemical players that are likely to be important. Similarly, improved methods for rapidly characterizing a cell on multiple levels (mRNA, protein, metabolites, etc.) and integrating and analyzing the large resulting data sets will be very useful (Moxley et al., 2009). However, our ultimate goal is to increase the pathway productivity, and identifying the stress is only the first step in that process. Once a stress is identified, a metabolic engineer must then treat that stress in order to have a real impact.

6.4 Tools and techniques to accommodate metabolic pathways

I was fortunate to have identified a stress, heme depletion, that is amenable to a rational, targeted approach. Overexpressing three native genes and feeding the precursor led to a large increase in the heme level and therefore pathway output. Had the source of stress proven to be more complicated, such as reactive oxygen (Ro et al., 2008) or protein folding and expression (Wiedmann et al., 1993), I would have had few targets for rationally modifying the host. The host systems responsible for detoxifying reactive oxygen species (Morano et al., 2011) or ensuring proper folding of marginally stable proteins (Geiler-Samerotte et al., 2011) are very complex, and no simple modification will increase the stress tolerance without significantly altering the basal functions of the host (Zakrzewska et al., 2011).

In cases such as these, researchers must turn to more-complicated strategies, either to limit the interactions between the pathway and its host or to identify unexpected host modifications that would better accommodate the pathway. Encapsulating a pathway in an organelle (Farhi et al., 2011) or protein microcompartment (Sampson & Bobik, 2008) could limit the stress by physically separating the heterologous pathway from the rest of the cell. Alternately, the host could be modified on a genome-wide scale to identify (Warner et al., 2010) and then optimize (Wang et al., 2009) novel modulators of complex traits. Finally, experimental coevolution of a pathway and its new host could simultaneously optimize both components (Chou et al., 2011). The development of techniques such as these will be critical for future efforts to optimize the interactions between a heterologous pathway and its new production host.

Metabolic engineering has the potential to play an important role in the transition away from an economy based on increasingly limited petrochemicals. Efficient biosyntheses of chemicals and fuels would allow the development of a sustainable industry built around the production of a wide array of valuable chemicals from feedstocks such as agricultural and municipal waste. Given the decades of optimization in existing petrochemical processes, a bioconversion must be extraordinarily efficient in order to be economically viable. Similarly, development costs must be minimized if biobased chemicals are to replace all of the niche chemicals currently produced from petroleum. New tools from systems and synthetic biology, such as those described in this thesis, will be necessary to enable the rapid construction and optimization of efficient metabolic pathways for chemical synthesis.

7 References

Aharoni A, Thieme K, Chiu CP, Buchini S, Lairson LL, Chen H, Strynadka NC, Wakarchuk WW, Withers SG (2006) High-throughput screening methodology for the directed evolution of glycosyltransferases. *Nat Methods* **3:** 609–614

Ajikumar PK, Xiao WH, Tyo KE, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G (2010) Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* **330:** 70–74

An W, Chin JW (2009) Synthesis of orthogonal transcription-translation networks. *Proc Natl Acad Sci U S A* **106:** 8477–8482

Anderson JC, Clarke EJ, Arkin AP, Voigt CA (2006) Environmentally controlled invasion of cancer cells by engineered bacteria. *Journal of Molecular Biology* **355:** 619–627

Ang J, Bagh S, Ingalls BP, McMillen DR (2010) Considerations for using integral feedback control to construct a perfectly adapting synthetic gene network. *J Theor Biol* **266:** 723–738

Arnold FH, Georgiou G (2003) *Directed enzyme evolution: screening and selection methods*, Totowa, N.J.: Humana Press.

Ashihara H, Crozier A (1999) Biosynthesis and Catabolism of Caffeine in Low-Caffeine-Containing Species of Coffea. *Journal of Agricultural and Food Chemistry* **47:** 3425–3431

Ashihara H, Monteiro AM, Moritz T, Gillies FM, Crozier A (1996) Catabolism of caffeine and related purine alkaloids in leaves of *Coffea arabica. Planta* **198:** 334–339

Askenazi M, Driggers EM, Holtzman DA, Norman TC, Iverson S, Zimmer DP, Boers M-E, Blomquist PR, Martinez EJ, Monreal AW, Feibelman TP, Mayorga ME, Maxon ME, Sykes K, Tobin JV, Cordero E, Salama SR, Trueheart J, Royer JC, Madden KT (2003) Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. *Nat Biotech* **21:** 150–156

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

Baker K, Bleczinski C, Lin H, Salazar-Jimenez G, Sengupta D, Krane S, Cornish VW (2002) Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proc Natl Acad Sci U S A* **99:** 16537–16542

Balskus EP, Walsh CT (2010) The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. *Science* **329:** 1653–1656

Bastian S, Liu X, Meyerowitz JT, Snow CD, Chen MM, Arnold FH (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli. *Metab Eng* **13:** 345–352

Bayer TS, Widmaier DM, Temme K, Mirsky EA, Santi DV, Voigt CA (2009) Synthesis of methyl halides from biomass using engineered microbes. *J Am Chem Soc* **131:** 6508–6515

Bennett MR, Pang WL, Ostroff NA, Baumgartner BL, Nayak S, Tsimring LS, Hasty J (2008) Metabolic gene regulation in a dynamically changing environment. *Nature* **454:** 1119–1122

Berens C, Thain A, Schroeder R (2001) A tetracycline-binding RNA aptamer. *Bioorganic & Medicinal Chemistry* **9:** 2549–2556

Blazeck J, Alper H (2010) Systems metabolic engineering: genome-scale models and beyond. *Biotechnol J* **5:** 647–659

Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* **103:** 5869–5874

Bonacci W, Teng PK, Afonso B, Niederholtmeyer H, Grob P, Silver PA, Savage DF (2012) Modularity of a carbon-fixing protein organelle. *Proc Natl Acad Sci U S A* **109:** 478–483

Bro C, Knudsen S, Regenberg B, Olsson L, Nielsen J (2005) Improvement of galactose uptake in Saccharomyces cerevisiae through overexpression of phosphoglucomutase: example of transcript analysis as a tool in inverse metabolic engineering. *Appl Environ Microbiol* **71:** 6465–6472

Bulter T, Alcalde M, Sieber V, Meinhold P, Schlachtbauer C, Arnold FH (2003) Functional expression of a fungal laccase in Saccharomyces cerevisiae by directed evolution. *Appl Environ Microbiol* **69:** 987–995

Buskirk AR, Landrigan A, Liu DR (2004) Engineering a ligand-dependent RNA transcriptional activator. *Chem Biol* **11:** 1157–1163

Canton B, Labno A, Endy D (2008) Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol* **26:** 787–793

Carlson R (2009) The changing economics of DNA synthesis. *Nat Biotechnol* **27:** 1091–1094

Chang MC, Eachus RA, Trieu W, Ro DK, Keasling JD (2007) Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. *Nat Chem Biol* **3:** 274–277

Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD (2006) Isolating and engineering human antibodies using yeast surface display. *Nat Protoc* **1:** 755–768

Chen F, Tholl D, Bohlmann J, Pichersky E (2011a) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal* **66:** 212–229

Chen I, Dorr BM, Liu DR (2011b) A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc Natl Acad Sci U S A* **108:** 11399–11404

Chen MM, Snow CD, Vizcarra CL, Mayo SL, Arnold FH (2012) Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. *Protein Eng Des Sel* **25:** 171–178

Chen X, Ellington AD (2009) Design principles for ligand-sensing, conformation-switching ribozymes. *PLoS Comput Biol* **5:** e1000620

Choi JH, Lee SJ, Lee SY (2003) Enhanced production of insulin-like growth factor I fusion protein in Escherichia coli by coexpression of the down-regulated genes identified by transcriptome profiling. *Appl Environ Microbiol* **69:** 4737–4742

Chou HH, Chiu HC, Delaney NF, Segre D, Marx CJ (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332:** 1190–1192

Chubukov V, Zuleta IA, Li H (2012) Regulatory architecture determines optimal regulation of gene expression in metabolic pathways. *Proc Natl Acad Sci U S A* **109:** 5127–5132

Collins CH, Arnold FH, Leadbetter JR (2005) Directed evolution of Vibrio fischeri LuxR for increased sensitivity to a broad spectrum of acyl-homoserine lactones. *Mol Microbiol* **55:** 712–723

Collins CH, Leadbetter JR, Arnold FH (2006) Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat Biotechnol* **24:** 708–712

Cookson NA, Mather WH, Danino T, Mondragon-Palomino O, Williams RJ, Tsimring LS, Hasty J (2011) Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Mol Syst Biol* **7:** 561

Czlapinski JL, Schelle MW, Miller LW, Laughlin ST, Kohler JJ, Cornish VW, Bertozzi CR (2008) Conditional glycosylation in eukaryotic cells using a biocompatible chemical inducer of dimerization. *J Am Chem Soc* **130:** 13186–13187

Davis JH, Baker TA, Sauer RT (2009) Engineering synthetic adaptors and substrates for controlled ClpXP degradation. *J Biol Chem* **284:** 21848–21855

Del Vecchio D, Ninfa AJ, Sontag ED (2008) Modular cell biology: retroactivity and insulation. *Mol Syst Biol* **4:** 161

Desai SK, Gallivan JP (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J Am Chem Soc* **126:** 13247–13254

Dietrich JA, McKee AE, Keasling JD (2010) High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu Rev Biochem* **79:** 563–590

Dietrich JA, Yoshikuni Y, Fisher KJ, Woolard FX, Ockey D, McPhee DJ, Renninger NS, Chang MC, Baker D, Keasling JD (2009) A novel semi-biosynthetic route for artemisinin production using engineered substrate-promiscuous P450(BM3). *ACS Chem Biol* **4:** 261–267

Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KL, Keasling JD (2009) Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* **27:** 753–759

Dunlop MJ, Keasling JD, Mukhopadhyay A (2010) A model for improving microbial biofuel production using a synthetic feedback loop. *Syst Synth Biol* **4:** 95–104

Edwards WR, Busse K, Allemann RK, Jones DD (2008) Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. *Nucleic Acids Res* **36:** e78

Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* **297:** 1183–1186

Facchini PJ, Bohlmann J, Covello PS, De Luca V, Mahadevan R, Page JE, Ro DK, Sensen CW, Storms R, Martin VJ (2012) Synthetic biosystems for the production of high-value plant metabolites. *Trends Biotechnol* **30:** 127–131

Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, Jefferson RE, Yeates TO, Bobik TA (2010) Short N-terminal sequences package proteins into bacterial microcompartments. *Proc Natl Acad Sci U S A* **107:** 7509–7514

Farhi M, Marhevka E, Masci T, Marcos E, Eyal Y, Ovadis M, Abeliovich H, Vainstein A (2011) Harnessing yeast subcellular compartments for the production of plant terpenoids. *Metabolic Engineering* **13:** 474–481

Farmer WR, Liao JC (2000) Improving lycopene production in Escherichia coli by engineering metabolic control. *Nat Biotech* **18:** 533–537

Fasan R, Chen MM, Crook NC, Arnold FH (2007) Engineered alkane-hydroxylating cytochrome P450(BM3) exhibiting nativelike catalytic properties. *Angew Chem Int Ed Engl* **46:** 8414–8418

Fasan R, Meharenna YT, Snow CD, Poulos TL, Arnold FH (2008) Evolutionary history of a specialized p450 propane monooxygenase. *Journal of Molecular Biology* **383:** 1069–1080

Fehr M, Frommer WB, Lalonde S (2002) Visualization of maltose uptake in living yeast cells by fluorescent nanosensors. *Proc Natl Acad Sci U S A* **99:** 9846–9851

Feuillet C, Leach JE, Rogers J, Schnable PS, Eversole K (2011) Crop genome sequencing: lessons and rationales. *Trends in Plant Science* **16:** 77–88

Flores S, de Anda-Herrera R, Gosset G, Bolivar FG (2004) Growth-rate recovery of Escherichia coli cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway. *Biotechnology and Bioengineering* **87:** 485–494

Fowler CC, Brown ED, Li Y (2008) A FACS-Based Approach to Engineering Artificial Riboswitches. *Chembiochem* **9:** 1906–1911

Fridman E, Pichersky E (2005) Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. *Current Opinion in Plant Biology* **8:** 242–248

Geiler-Samerotte KA, Dion MF, Budnik BA, Wang SM, Hartl DL, Drummond DA (2011) Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proc Natl Acad Sci U S A* **108:** 680–685

Gibson DG (2011) Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol* **498:** 349–361

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329:** 52–56

Gietz RD, Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350:** 87–96

Goff SA, Goldberg AL (1985) Production of abnormal proteins in E. coli stimulates transcription of lon and other heat shock genes. *Cell* **41:** 587–595

Gonzalez-Pajuelo M, Meynial-Salles I, Mendes F, Andrade JC, Vasconcelos I, Soucaille P (2005) Metabolic engineering of Clostridium acetobutylicum for the industrial production of 1,3-propanediol from glycerol. *Metab Eng* **7:** 329–336

Gonzalez B, François J, Renaud M (1997) A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol. *Yeast* **13:** 1347–1355

Graham-Lorence S, Truan G, Peterson JA, Falck JR, Wei S, Helvig C, Capdevila JH (1997) An active site substitution, F87V, converts cytochrome P450 BM-3 into a regio- and stereoselective (14S,15R)-arachidonic acid epoxygenase. *J Biol Chem* **272:** 1127–1135

Grilly C, Stricker J, Pang WL, Bennett MR, Hasty J (2007) A synthetic gene network for tuning protein degradation in Saccharomyces cerevisiae. *Mol Syst Biol* **3:** 127

Guerra-Guimarães L, Silva M, Struck C, Loureiro A, Nicole M, Rodrigues C, Ricardo C (2009) Chitinases of *Coffea arabica* genotypes resistant to orange rust *Hemileia vastatrix*. *Biologia Plantarum* **53:** 702–706

Guntas G, Mansell TJ, Kim JR, Ostermeier M (2005) Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc Natl Acad Sci U S A* **102:** 11224–11229

Guntas G, Ostermeier M (2004) Creation of an allosteric enzyme by domain insertion. *Journal of Molecular Biology* **336:** 263–273

Hagel JM, Facchini PJ (2010) Dioxygenases catalyze the O-demethylation steps of morphine biosynthesis in opium poppy. *Nat Chem Biol* **6:** 273–275

Hale V, Keasling JD, Renninger N, Diagana TT (2007) Microbially derived artemisinin: a biotechnology solution to the global problem of access to affordable antimalarial drugs. *The American Journal of Tropical Medicine and Hygiene* **77:** 198–202

Han MJ, Yoon SS, Lee SY (2001) Proteome analysis of metabolically engineered Escherichia coli producing Poly(3-hydroxybutyrate). *J Bacteriol* **183:** 301–308

Hans MA, Heinzle E, Wittmann C (2001) Quantification of intracellular amino acids in batch cultures of Saccharomyces cerevisiae. *Appl Microbiol Biotechnol* **56:** 776–779

Hawkins KM, Smolke CD (2008) Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat Chem Biol* **4:** 564–573

Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331:** 463–467

Hill A, Bloom K (1987) Genetic manipulation of centromere function. *Molecular and Cellular Biology* **7:** 2397–2405

Hoffman M, Gora M, Rytka J (2003) Identification of rate-limiting steps in yeast heme biosynthesis. *Biochemical and Biophysical Research Communications* **310:** 1247–1253

Holter NS, Mitra M, Maritan A, Cieplak M, Banavar JR, Fedoroff NV (2000) Fundamental patterns underlying gene expression profiles: Simplicity from complexity. *Proc Natl Acad Sci U S A* **97:** 8409–8414

Jenison RD, Gill SC, Pardi A, Polisky B (1994) High-resolution molecular discrimination by RNA. *Science* **263:** 1425–1429

Jiang H, Morgan JA (2004) Optimization of an in vivo plant P450 monooxygenase system in Saccharomyces cerevisiae. *Biotechnology and bioengineering* **85:** 130–137

Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, 3rd, Hilvert D, Houk KN, Stoddard BL, Baker D (2008) De novo computational design of retro-aldol enzymes. *Science* **319:** 1387–1391

Jones KL, Kim SW, Keasling JD (2000) Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab Eng* **2:** 328–338

Kang Z, Wang Y, Gu PF, Wang Q, Qi QS (2011) Engineering Escherichia coli for efficient production of 5-aminolevulinic acid from glucose. *Metab Eng* **13:** 492–498

Karig D, Weiss R (2005) Signal-amplifying genetic circuit enables in vivo observation of weak promoter activation in the Rhl quorum sensing system. *Biotechnology and Bioengineering* **89:** 709–718

Keasling JD (2010) Manufacturing molecules through metabolic engineering. *Science* **330:** 1355–1358

Kersten RD, Yang YL, Xu Y, Cimermancic P, Nam SJ, Fenical W, Fischbach MA, Moore BS, Dorrestein PC (2011) A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat Chem Biol* **7:** 794–802

Khersonsky O, Rothlisberger D, Dym O, Albeck S, Jackson CJ, Baker D, Tawfik DS (2010) Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *Journal of Molecular Biology* **396:** 1025–1042

Kille S, Zilly FE, Acevedo JP, Reetz MT (2011) Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nature Chemistry* **3:** 738–743

Kizer L, Pitera DJ, Pfleger BF, Keasling JD (2008) Application of functional genomics to pathway optimization for increased isoprenoid production. *Appl Environ Microbiol* **74:** 3229–3241

Koch AL (1956) The metabolism of methylpurines by Escherichia coli. I. Tracer studies. *J Biol Chem* **219:** 181–188

Kraus JP, Kery V (1997) Method of increasing the yield and heme saturation of cystathione β-synthase. *United States Patent*, Number 5,635,375.

Kushnirov VV (2000) Rapid and reliable protein extraction from yeast. *Yeast* **16:** 857-860

Lee KH, Park JH, Kim TY, Kim HU, Lee SY (2007) Systems metabolic engineering of Escherichia coli for L-threonine production. *Mol Syst Biol* **3:** 149

Lee SJ, Trostel A, Le P, Harinarayanan R, Fitzgerald PC, Adhya S (2009) Cellular stress created by intermediary metabolite imbalances. *Proc Natl Acad Sci U S A* **106:** 19515–19520

Leonard E, Ajikumar PK, Thayer K, Xiao WH, Mo JD, Tidor B, Stephanopoulos G, Prather KL (2010) Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc Natl Acad Sci U S A* **107:** 13654–13659

Lewis JC, Bastian S, Bennett CS, Fu Y, Mitsuda Y, Chen MM, Greenberg WA, Wong CH, Arnold FH (2009) Chemoenzymatic elaboration of monosaccharides using engineered cytochrome P450BM3 demethylases. *Proc Natl Acad Sci U S A* **106:** 16550–16555

Liang JC (2012) *High-Throughput Strategies for the Scalable Generation of RNA Component Functions*. PhD Thesis, California Institute of Technology

Lin H, Tao H, Cornish VW (2004) Directed evolution of a glycosynthase via chemical complementation. *J Am Chem Soc* **126:** 15051–15059

Liscombe DK, Facchini PJ (2008) Evolutionary and cellular webs in benzylisoquinoline alkaloid biosynthesis. *Curr Opin Biotechnol* **19:** 173–180

Lorsch JR, Szostak JW (1994) In vitro selection of RNA aptamers specific for cyanocobalamin. *Biochemistry* **33:** 973–982

Lukk T, Sakai A, Kalyanaraman C, Brown SD, Imker HJ, Song L, Fedorov AA, Fedorov EV, Toro R, Hillerich B, Seidel R, Patskovsky Y, Vetting MW, Nair SK, Babbitt PC, Almo SC, Gerlt JA, Jacobson MP (2012) Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily. *Proc Natl Acad Sci U S A* **109:** 4122–4127

Lutke-Eversloh T, Stephanopoulos G (2007) L-tyrosine production by deregulated strains of Escherichia coli. *Appl Microbiol Biotechnol* **75:** 103–110

Lynch SA, Gallivan JP (2009) A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res* **37:** 184–192

Mannironi C, Di Nardo A, Fruscoloni P, Tocchini-Valentini GP (1997) In vitro selection of dopamine RNA ligands. *Biochemistry* **36:** 9726–9734

Mason GG, Hendil KB, Rivett AJ (1996) Phosphorylation of proteasomes in mammalian cells. Identification of two phosphorylated subunits and the effect of phosphorylation on activity. *Eur J Biochem* **238:** 453–462

Mattoon JR, Bajszar G (1998) Method for enhancing the production of hemoproteins. *United States Patent*, Number 5,824,511.

Mazzafera P (2004) Catabolism of caffeine in plants and microorganisms. *Frontiers in Bioscience: a Journal and Virtual Library* **9:** 1348–1359

Mazzafera P, Crozier A, Sandberg G (1994) Studies on the Metabolic Control of Caffeine Turnover in Developing Endosperms and Leaves of Coffea arabica and Coffea dewevrei. *Journal of Agricultural and Food Chemistry* **42:** 1423–1427

Michener JK, Thodey K, Liang JC, Smolke CD (2012) Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metab Eng* **14:** 212–222

Mizutani M, Ohta D (2010) Diversification of P450 genes during land plant evolution. *Annu Rev Plant Biol* **61:** 291–315

Mohn WW, Garmendia J, Galvao TC, de Lorenzo V (2006) Surveying biotransformations with a la carte genetic traps: translating dehydrochlorination of lindane (gamma-hexachlorocyclohexane) into lacZ-based phenotypes. *Environ Microbiol* **8:** 546–555

Mondego JM, Vidal RO, Carazzolle MF, Tokuda EK, Parizzi LP, Costa GG, Pereira LF, Andrade AC, Colombo CA, Vieira LG, Pereira GA (2011) An EST-based analysis identifies new genes and reveals distinctive gene expression features of Coffea arabica and Coffea canephora. *BMC Plant Biol* **11:** 30

Morano KA, Grant CM, Moye-Rowley WS (2012) The Response to Heat Shock and Oxidative Stress in Saccharomyces cerevisiae. *Genetics* **190:** 1157–1195

Moxley JF, Jewett MC, Antoniewicz MR, Villas-Boas SG, Alper H, Wheeler RT, Tong L, Hinnebusch AG, Ideker T, Nielsen J, Stephanopoulos G (2009) Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proc Natl Acad Sci U S A* **106:** 6477–6482

Mustafi N, Grünberger A, Kohlheyer D, Bott M, Frunzke J (2012) The development and application of a single-cell biosensor for the detection of l-methionine and branched-chain amino acids. *Metabolic Engineering*

Nakagawa A, Minami H, Kim JS, Koyanagi T, Katayama T, Sato F, Kumagai H (2011) A bacterial platform for fermentative production of plant alkaloids. *Nat Commun* **2:** 326

Narhi LO, Fulco AJ (1986) Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in Bacillus megaterium. *J Biol Chem* **261:** 7160–7169

Neeli R, Girvan HM, Lawrence A, Warren MJ, Leys D, Scrutton NS, Munro AW (2005) The dimeric form of flavocytochrome P450 BM3 is catalytically functional as a fatty acid hydroxylase. *FEBS Lett* **579:** 5582–5588

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

Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464:** 441–444

Nevoigt E, Kohnke J, Fischer CR, Alper H, Stahl U, Stephanopoulos G (2006) Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. *Appl Environ Microbiol* **72:** 5266–5273

Noble MA, Miles CS, Chapman SK, Lysek DA, MacKay AC, Reid GA, Hanzlik RP, Munro AW (1999) Roles of key active-site residues in flavocytochrome P450 BM3. *Biochemical Journal* **339 (2):** 371–379

Ogita S, Uefuji H, Yamaguchi Y, Koizumi N, Sano H (2003) RNA interference: Producing decaffeinated coffee plants. *Nature* **423:** 823–823

Okamoto S, Lezhava A, Hosaka T, Okamoto-Hosoya Y, Ochi K (2003) Enhanced expression of S-adenosylmethionine synthetase causes overproduction of actinorhodin in Streptomyces coelicolor A3(2). *J Bacteriol* **185:** 601–609

Oliveira AP, Patil KR, Nielsen J (2008) Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Systems Biology* **2:** 17

Park JH, Lee KH, Kim TY, Lee SY (2007) Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc Natl Acad Sci U S A* **104:** 7797–7802

Parsons AB, Brost RL, Ding H, Li Z, Zhang C, Sheikh B, Brown GW, Kane PM, Hughes TR, Boone C (2004) Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat Biotechnol* **22:** 62–69

Peralta-Yahya P, Carter BT, Lin H, Tao H, Cornish VW (2008) High-throughput selection for cellulase catalysts using chemical complementation. *J Am Chem Soc* **130:** 17446–17452

Peters MW, Meinhold P, Glieder A, Arnold FH (2003) Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. *J Am Chem Soc* **125:** 13442–13450

Pfleger BF, Pitera DJ, Newman JD, Martin VJ, Keasling JD (2007) Microbial sensors for small molecules: development of a mevalonate biosensor. *Metab Eng* **9:** 30–38

Rackham O, Chin JW (2005) A network of orthogonal ribosome x mRNA pairs. *Nat Chem Biol* **1:** 159–166

Raychaudhuri S, Stuart JM, Altman RB (2000) Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing***:** 455–466

Rentmeister A, Arnold FH, Fasan R (2009) Chemo-enzymatic fluorination of unactivated organic compounds. *Nat Chem Biol* **5:** 26–28

Rice KC (1980) Synthetic opium alkaloids and derivatives. A short total synthesis of (.+-.)-dihydrothebainone, (.+-.)-dihydrocodeinone, and (.+-.)-nordihydrocodeinone as an approach to a practical synthesis of morphine, codeine, and congeners. *The Journal of Organic Chemistry* **45:** 3135–3137

Ro DK, Ouellet M, Paradise EM, Burd H, Eng D, Paddon CJ, Newman JD, Keasling JD (2008) Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid. *BMC Biotechnol* **8:** 83

Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MC, Withers ST, Shiba Y, Sarpong R, Keasling JD (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440:** 940–943

Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D (2008) Kemp elimination catalysts by computational enzyme design. *Nature* **453:** 190–195

Runguphan W, Qu X, O/'Connor SE (2010) Integrating carbon-halogen bond formation into medicinal plant metabolism. *Nature* **468:** 461–464

Sadowski I, Su TC, Parent J (2007) Disintegrator vectors for single-copy yeast chromosomal integration. *Yeast* **24:** 447–455

Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27:** 946–950

Sambrook J, Russell D (eds) (2001) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Lab Press

Sampson EM, Bobik TA (2008) Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J Bacteriol* **190:** 2966–2971

Santos CN, Stephanopoulos G (2008) Melanin-based high-throughput screen for L-tyrosine production in Escherichia coli. *Appl Environ Microbiol* **74:** 1190–1197

Sassa S (1976) Sequential induction of heme pathway enzymes during erythroid differentiation of mouse Friend leukemia virus-infected cells. *The Journal of Experimental Medicine* **143:** 305–315

Scalcinati G, Knuf C, Partow S, Chen Y, Maury J, Schalk M, Daviet L, Nielsen J, Siewers V (2012) Dynamic control of gene expression in Saccharomyces cerevisiae engineered for the production of plant sesquitepene alpha-santalene in a fed-batch mode. *Metab Eng* **14:** 91–103

Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB (2010) Microbial biosynthesis of alkanes. *Science* **329:** 559–562

Schreier B, Stumpp C, Wiesner S, Hocker B (2009) Computational design of ligand binding is not a solved problem. *Proc Natl Acad Sci U S A* **106:** 18491-18496

Schunck WH, Vogel F, Gross B, Kargel E, Mauersberger S, Kopke K, Gengnagel C, Muller HG (1991) Comparison of two cytochromes P-450 from Candida maltosa: primary structures, substrate specificities and effects of their expression in Saccharomyces cerevisiae on the proliferation of the endoplasmic reticulum. *European Journal of Cell Biology* **55:** 336–345

Scott SA, Davey MP, Dennis JS, Horst I, Howe CJ, Lea-Smith DJ, Smith AG (2010) Biodiesel from algae: challenges and prospects. *Curr Opin Biotechnol* **21:** 277–286

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

Shusta EV, Raines RT, Pluckthun A, Wittrup KD (1998) Increasing the secretory capacity of Saccharomyces cerevisiae for production of single-chain antibody fragments. *Nat Biotechnol* **16:** 773–777

Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, Houk KN, Michael FE, Baker D (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329:** 309–313

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

Sinha J, Reyes SJ, Gallivan JP (2010) Reprogramming bacteria to seek and destroy an herbicide. *Nat Chem Biol* **6:** 464–470

Song JY, Jeong H, Yu DS, Fischbach MA, Park HS, Kim JJ, Seo JS, Jensen SE, Oh TK, Lee KJ, Kim JF (2010) Draft genome sequence of Streptomyces clavuligerus NRRL 3585, a producer of diverse secondary metabolites. *J Bacteriol* **192:** 6317–6318

Soukup GA, Breaker RR (1999) Engineering precision RNA molecular switches. *Proc Natl Acad Sci U S A* **96:** 3584–3589

Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J (2008) A fast, robust and tunable synthetic gene oscillator. *Nature* **456:** 516–519

Suess B, Fink B, Berens C, Stentz R, Hillen W (2004) A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res* **32:** 1610–1614

Suzuki T, Waller GR (1984) Biodegradation of caffeine: Formation of theophylline and theobromine from caffeine in mature Coffea arabica fruits. *Journal of the Science of Food and Agriculture* **35:** 66–70

Szczebara FM, Chandelier C, Villeret C, Masurel A, Bourot S, Duport C, Blanchard S, Groisillier A, Testet E, Costaglioli P, Cauet G, Degryse E, Balbuena D, Winter J, Achstetter T, Spagnoli R, Pompon D, Dumas B (2003) Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nat Biotechnol* **21:** 143–149

Tabor JJ, Salis HM, Simpson ZB, Chevalier AA, Levskaya A, Marcotte EM, Voigt CA, Ellington AD (2009) A synthetic genetic edge detection program. *Cell* **137:** 1272–1281

Tang SY, Cirino PC (2010) Elucidating residue roles in engineered variants of AraC regulatory protein. *Protein Sci* **19:** 291–298

Tang SY, Cirino PC (2011) Design and application of a mevalonate-responsive regulatory protein. *Angew Chem Int Ed Engl* **50:** 1084–1086

Tassaneeyakul W, Birkett DJ, McManus ME, Veronese ME, Andersson T, Tukey RH, Miners JO (1994) Caffeine metabolism by human hepatic cytochromes P450: contributions of 1A2, 2E1 and 3A isoforms. *Biochem Pharmacol* **47:** 1767–1776

Taupp M, Mewis K, Hallam SJ (2011) The art and design of functional metagenomic screens. *Current Opinion in Biotechnology* **22:** 465–472

Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459:** 668–673

Topp S, Reynoso CM, Seeliger JC, Goldlust IS, Desai SK, Murat D, Shen A, Puri AW, Komeili A, Bertozzi CR, Scott JR, Gallivan JP (2010) Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl Environ Microbiol* **76:** 7881–7884

Tsao CY, Hooshangi S, Wu HC, Valdes JJ, Bentley WE (2010) Autonomous induction of recombinant proteins by minimally rewiring native quorum sensing regulon of E. coli. *Metab Eng* **12:** 291–297

Uchiyama T, Miyazaki K (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* **20:** 616–622

Urban P, Werck-Reichhart D, Teutsch HG, Durst F, Regnier S, Kazmaier M, Pompon D (1994) Characterization of recombinant plant cinnamate 4-hydroxylase produced in yeast. Kinetic and spectral properties of the major plant P450 of the phenylpropanoid pathway. *Eur J Biochem* **222:** 843–850

van Sint Fiet S, van Beilen JB, Witholt B (2006) Selection of biocatalysts for chemical synthesis. *Proc Natl Acad Sci U S A* **103:** 1693–1698

Ventura AC, Jiang P, Van Wassenhove L, Del Vecchio D, Merajver SD, Ninfa AJ (2010) Signaling properties of a covalent modification cycle are altered by a downstream target. *Proc Natl Acad Sci U S A* **107:** 10032–10037

Walsh CT (2008) The chemical versatility of natural-product assembly lines. *Acc Chem Res* **41:** 4–10

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460:** 894–898

Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E et al (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450:** 560–565

Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LB, Gill RT (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol* **28:** 856–862

Wehner EP, Rao E, Brendel M (1993) Molecular structure and genetic regulation of SFA, a gene responsible for resistance to formaldehyde in Saccharomyces cerevisiae, and characterization of its protein product. *Molecular & General Genetics (MGG)* **237:** 351–358

Werstuck G, Green MR (1998) Controlling gene expression in living cells through small molecule-RNA interactions. *Science* **282:** 296–298

Wiedmann B, Silver P, Schunck WH, Wiedmann M (1993) Overexpression of the ER-membrane protein P-450 CYP52A3 mimics sec mutant characteristics in Saccharomyces cerevisiae. *Biochim Biophys Acta* **1153:** 267–276

Win MN, Klein JS, Smolke CD (2006) Codeine-binding RNA aptamers and rapid determination of their binding constants using a direct coupling surface plasmon resonance assay. *Nucleic Acids Res* **34:** 5670–5682

Win MN, Liang JC, Smolke CD (2009) Frameworks for programming biological function through RNA parts and devices. *Chem Biol* **16:** 298–310

Win MN, Smolke CD (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A* **104:** 14283–14288

Wingler LM, Cornish VW (2011) Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. *Proc Natl Acad Sci U S A* **108:** 15135–15140

Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428:** 281–286

Wittmann C, Kromer JO, Kiefer P, Binz T, Heinzle E (2004) Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria. *Anal Biochem* **327:** 135–139

Xia XX, Qian ZG, Ki CS, Park YH, Kaplan DL, Lee SY (2010) Native-sized recombinant spider silk protein produced in metabolically engineered Escherichia coli results in a strong fiber. *Proc Natl Acad Sci U S A* **107:** 14059–14063

Yadav G, Gokhale RS, Mohanty D (2009) Towards prediction of metabolic products of polyketide synthases: an in silico analysis. *PLoS Comput Biol* **5:** e1000351

Yang G, Rich JR, Gilbert M, Wakarchuk WW, Feng Y, Withers SG (2010) Fluorescence activated cell sorting as a general ultra-high-throughput screening method for directed evolution of glycosyltransferases. *J Am Chem Soc* **132:** 10570–10577

Yi T-M, Huang Y, Simon MI, Doyle J (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci U S A* **97:** 4649–4653

Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, Estadilla J, Teisan S, Schreyer HB, Andrae S, Yang TH, Lee SY, Burk MJ, Van Dien S (2011) Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat Chem Biol* **7:** 445–452

Zakrzewska A, van Eikenhorst G, Burggraaff JE, Vis DJ, Hoefsloot H, Delneri D, Oliver SG, Brul S, Smits GJ (2011) Genome-wide analysis of yeast stress survival and tolerance acquisition to analyze the central trade-off between growth rate and cellular robustness. *Mol Biol Cell* **22:** 4435–4446

Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, Alon U (2004) Just-in-time transcription program in metabolic pathways. *Nature Genetics* **36:** 486–491

Zhang F, Carothers JM, Keasling JD (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat Biotechnol* **30:** 354–359

Zhang F, Su K, Yang X, Bowe DB, Paterson AJ, Kudlow JE (2003) O-GlcNAc modification is an endogenous inhibitor of the proteasome. *Cell* **115:** 715–725

Zhu MM, Lawman PD, Cameron DC (2002) Improving 1,3-propanediol production from glycerol in a metabolically engineered Escherichia coli by reducing accumulation of sn-glycerol-3-phosphate. *Biotechnology Progress* **18:** 694–699

Zhu MM, Skraly FA, Cameron DC (2001a) Accumulation of methylglyoxal in anaerobically grown Escherichia coli and its detoxification by expression of the Pseudomonas putida glyoxalase I gene. *Metab Eng* **3:** 218–225

Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD (2001b) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* **30:** 892–897

Zimmermann GR, Wick CL, Shields TP, Jenison RD, Pardi A (2000) Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *RNA* **6:** 659–667