Molecular Bioengineering: From Protein Stability to Population Suicide

by

Philippe Robert Marguet

Department of Biochemistry
Duke University

Date:_______________________
Approved:

___________________________
Homme Hellinga, Co-Supervisor

___________________________
Lingchong You, Co-Supervisor

___________________________
Kenneth Kreuzer

___________________________
Pei Zhou

___________________________
David Beratan

___________________________
Jingdong Tian

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2010
ABSTRACT

Molecular Bioengineering: From Protein Stability to Population Suicide

by

Philippe Robert Marguet

Department of Biochemistry
Duke University

Date: _________________________

Approved:

__________________________________
Homme Hellinga, Co-Supervisor

__________________________________
Lingchong You, Co-Supervisor

__________________________________
Kenneth Kreuzer

__________________________________
Pei Zhou

__________________________________
David Beratan

__________________________________
Jingdong Tian

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2010
Abstract

Driven by the development of new technologies and an ever expanding knowledge base of molecular and cellular function, Biology is rapidly gaining the potential to develop into a veritable engineering discipline – the so-called ‘era of synthetic biology’ is upon us. Designing biological systems is advantageous because the engineer can leverage existing capacity for self-replication, elaborate chemistry, and dynamic information processing. On the other hand these functions are complex, highly intertwined, and in most cases, remain incompletely understood. Brazenly designing within these systems, despite large gaps in understanding, engenders understanding because the design process itself highlights gaps and discredits false assumptions.

Here we cover results from design projects that span several scales of complexity. First we describe the adaptation and experimental validation of protein functional assays on minute amounts of material. This work enables the application of cell-free protein expression tools in a high-throughput protein engineering pipeline, dramatically increasing turnaround time and reducing costs. The parts production pipeline can provide new building blocks for synthetic biology efforts with unprecedented speed. Tools to streamline the transition from the \textit{in vitro} pipeline to conventional cloning were also developed. Next we detail an effort to expand the scope of a cysteine reactivity assay for generating information-rich datasets on protein stability and unfolding kinetics. We go on to demonstrate how the degree of site-specific local
unfolding can also be determined by this method. This knowledge will be critical to understanding how proteins behave in the cellular context, particularly with regards to covalent modification reactions. Finally, we present results from an effort to engineer bacterial cell suicide in a population-dependent manner, and show how an underappreciated facet of plasmid physiology can produce complex oscillatory dynamics. This work is a prime example of engineering towards understanding.
Dedication

To Amelia, for always being by my side
# Table of Contents

Abstract .................................................................................................................................................. iv

List of Tables .......................................................................................................................................... xi

List of Figures ......................................................................................................................................... xii

List of Abbreviations .......................................................................................................................... xv

Acknowledgements ............................................................................................................................... xvii

1. Background ........................................................................................................................................ 1

   1.1 Biology by Design .................................................................................................................. 1

   1.2 Foundational Technologies ................................................................................................. 6

      1.2.1 DNA Synthesis ........................................................................................................ 8

      1.2.2 Design and Optimization of Parts ........................................................................... 10

      1.2.3 Modeling-Guided Circuit Engineering .................................................................. 16

      1.2.4 Culturing and Monitoring Technologies ............................................................... 23

   1.3 Outlook .................................................................................................................................... 29

2. A Protein Production and Characterization Pipeline ...................................................................... 35

   2.1 Introduction .............................................................................................................................. 35

      2.1.1 In House Gene Synthesis and Assembly ............................................................... 35

      2.1.2 *in vitro* Transcription and Translation (TnT) ....................................................... 36

   2.2 Results ..................................................................................................................................... 38

      2.2.1 Covalently Coupled Conformationally Sensitive Fluorophores ......................... 38

      2.2.2 Miniaturized Thermal Shift Assay for Ligand Binding ......................................... 46

      2.2.3 SYPRO Orange Assay for Screening Cloned Genes ............................................ 57
2.3 Discussion ........................................................................................................................ 65
2.4 Materials and Methods .................................................................................................. 67
  2.4.1 tmGBP Expression and Fluorophore Conjugation ............................................. 67
  2.4.2 tmGBP Ligand Titrations ...................................................................................... 68
  2.4.3 SYPRO Thermal Shift Assays .............................................................................. 68
  2.4.3 Gel Based Screening ............................................................................................ 69
3. Quantitative Cysteine Reactivity using ABD ...................................................................... 70
  3.1 Introduction .................................................................................................................. 70
  3.2 Results .......................................................................................................................... 73
    3.2.1 Measuring Protein Stability by QCR. ................................................................. 73
    3.2.2 The Mechanism of Cysteine Labeling. ............................................................... 75
    3.2.3 Determinants of \( k_{int} \) ...................................................................................... 78
    3.2.4 Measuring Protein Stability Using fQCR. ......................................................... 81
    3.2.5 EX2 Conditions ................................................................................................... 85
    3.2.6 Estimation of Folding Rate Constants ............................................................... 86
    3.2.7 Experimental fQCR Results. .............................................................................. 87
    3.2.8 \( T_m \) Values ...................................................................................................... 91
    3.2.9 Assessing Protein Function by fQCR. ................................................................. 92
  3.3 Discussion ...................................................................................................................... 95
  3.4 Materials and Methods .............................................................................................. 97
    3.4.1 Protein Engineering and Purification ................................................................. 97
    3.4.2 Preparation of ABD Stock Solutions ................................................................. 98
3.4.3 Reaction of ABD with Glutathione................................................................. 99
3.4.4 The fQCR Experiment....................................................................................... 100
3.4.5 Thermal Unfolding Monitored by Circular Dichroism................................. 101

4. Site Specificity and Local Unfolding in QCR ....................................................... 103
4.1 Introduction............................................................................................................. 103
4.2 Results .................................................................................................................... 107
5.2.1 Planning a Cysteine Scan in E. Coli Ribose Binding Protein ...................... 107
5.2.2 Site Dependent Properties of Cysteine Variants........................................... 110
5.2.3 QCR and Local Unfolding................................................................................. 117
4.3 Discussion.............................................................................................................. 123
4.4 Materials and Methods....................................................................................... 125
4.4.1 Variant Generation and Protein Production.................................................... 125
4.4.2 Characterization Of Site Dependent Properties.............................................. 125
4.4.3 fQCR for Local Unfolding............................................................................... 125

5. Oscillations by Minimal Bacterial Suicide Circuits Reveal Hidden Facets of Host-
Circuit Physiology ........................................................................................................ 127
5.1 Introduction.......................................................................................................... 127
5.2 Results .................................................................................................................. 130
5.2.1 ePop Causes Oscillations In Cell Density ....................................................... 130
5.2.2 Oscillations do not Depend on luxRI Quorum Sensing or Promoter-Level
Regulation .................................................................................................................... 136
5.2.3 Plasmid Amplification Can Account for Negative Feedback....................... 143
5.2.4 ColE1 Replication ............................................................................................ 144
5.2.5 Construction of a Mathematic Model that Captures ePop Dynamics .......... 147
5.2.6 Perturbations that Modulate Copy Number Impact Oscillations .............. 150
5.3 Discussion ........................................................................................................ 154
5.4 Materials and Methods .................................................................................... 159
  5.4.1 Strains, Growth Conditions, Chemicals and Media ........................................ 159
  5.4.2 Liquid Culture Monitoring ........................................................................... 159
  5.4.3 Plasmids ......................................................................................................... 160
  5.4.4 Plasmid Content Determination .................................................................... 160
  5.4.5 Solid Phase Monitoring ................................................................................ 161
References .............................................................................................................. 163
Biography .............................................................................................................. 184
List of Tables

Table 1: Fit parameters for glucose titrations of a tmGBP alanine scan variants...............42
Table 2: Experimentally derived fQCR parameters for SN and ecRBP ................................90
Table 3: $K_d$ of SN and ecRBP as a function of temperature ...........................................94
Table 4: Site-dependent properties of ecRBP Cysteine variants.......................................116
List of Figures

Figure 1: Interplay between engineering tools with a biological hierarchy. ......................... 7
Figure 2: The typical process for engineering gene circuits.................................................. 18
Figure 3: Complexity and uncertainty in biological circuit design........................................ 22
Figure 4: A microfluidic chip with six parallel microchemostat reactors............................... 27
Figure 5: Schematic of a tmGBP biosensor............................................................................. 39
Figure 6: Impact of BSA on tmGBP titrations........................................................................ 40
Figure 7: Rapid response protein engineering...................................................................... 41
Figure 8: Normalized titrations of tmGBP alanine variants.................................................. 43
Figure 9: tmGBP binding site color coded by mutational impact.......................................... 45
Figure 10: Linked equilibrium for ligand binding and protein unfolding............................ 48
Figure 11: Representative SYPRO melts of ecRBP variants.................................................. 51
Figure 12: Thermal shifts for each ecRBP alanine variant at two ribose concentrations........ 52
Figure 13: SYPRO thermal shift assay vs. affinities reported in the literature....................... 53
Figure 14: Stability shifts produced by ribose analogs......................................................... 55
Figure 15: Flow chart for streamlined cloning, sequence verification, and expression........... 60
Figure 16: SDS-PAGE screening of full length in-frame genes............................................ 61
Figure 17: Discriminating probable mutants on the basis of SYPRO profiles....................... 63
Figure 18: pH and temperature dependence of the reaction of ABD with L-GSH................ 80
Figure 19: The relationships between fractional labeling and protein stability.................... 84
Figure 20: Simulation of EX2 behavior.................................................................................. 86
Figure 21: fQCR results for SN and ecRBP.......................................................................... 88
Figure 22: Thermal unfolding of SN and ecRBP.................................................................92
Figure 23: Ligand binding by fQCR. ....................................................................................94
Figure 24: Simulated global and local contributions to apparent free energy of unfolding. ..............................................................................................................................106
Figure 25: Ca of sites selected for a cysteine scan of ecRBP color coded by side chain. 108
Figure 26: Some ecRBP cysteine mutants form covalent dimers under mild conditions. ..............................................................................................................................109
Figure 27: Tm of cysteine variants determined by SYPRO thermal melts. .........................110
Figure 28: Examples of temperature-dependent transitions in ABD labeled proteins. ..112
Figure 29: Wavelength and maximal fluorescence distribution across ecRBP cysteine variants. ..............................................................................................................................115
Figure 30: Raw QCR data, ΔG conversion, and the EX2 test for L265C. .........................119
Figure 31: Three examples of local unfolding in ecRBP measured by QCR.................120
Figure 32: ePop dynamics in liquid culture. .....................................................................131
Figure 33: Optical density measurements are consistent with cell density.................132
Figure 34: ePop plasmid is required for oscillations.......................................................133
Figure 35: Solid phase growth of BW25113 carrying ePop. ...........................................134
Figure 36: ePop oscillates in defined media as well as complex media. ....................135
Figure 37: Effect of culturing temperature on ePop oscillations. ................................135
Figure 38: ePop based circuits treated with AHL..........................................................137
Figure 39: Possible sources of negative feedback. .......................................................139
Figure 40: Effect of DPD on oscillations.......................................................................140
Figure 41: Feedback is not at the promoter level.........................................................142
Figure 42: Plasmid level as a function of growth phase and media composition. .......144
Figure 43: Stringent control, plasmid replication, and a possible mechanistic link. 146
Figure 44: A simplified model for ePop function. 147
Figure 45: Simplified model equations and parameters. 149
Figure 46: Impact of $\delta_1$ value on oscillations. 150
Figure 47: Model predictions and experimental responses to system perturbations. 152
Figure 48: Impact of chloramphenicol on plasmid amplification in pNewTet.E. 153
Figure 49: Gate matching and unexpected feedbacks. 156
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABD</td>
<td>4-(aminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole</td>
</tr>
<tr>
<td>AHL</td>
<td>acyl- homoserine lactone</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>cAMP</td>
<td>cyclic adenosine monophosphate</td>
</tr>
<tr>
<td>CRP</td>
<td>cAMP receptor protein</td>
</tr>
<tr>
<td>ΔGU</td>
<td>Gibbs free energy of protein unfolding</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DTNB</td>
<td>5,5¢-dithiobis(2-nitrobenzoate)</td>
</tr>
<tr>
<td>ecRBP</td>
<td><em>Escherichia coli</em> ribose binding protein</td>
</tr>
<tr>
<td>EX2</td>
<td>exchange conditions that favor a single kinetic phase</td>
</tr>
<tr>
<td>FACS</td>
<td>fluorescence activated cell sorter</td>
</tr>
<tr>
<td>fQCR</td>
<td>fast determination of Quantitative Cysteine Reactivity</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>GUWO</td>
<td>global unfolding window of observation</td>
</tr>
<tr>
<td>HDX</td>
<td>hydrogen-deuterium exchange</td>
</tr>
<tr>
<td>IAM</td>
<td>iodoacetamide</td>
</tr>
<tr>
<td>IANBD amide</td>
<td>N,N’-dimethyl-N-(iodoacetyl)-N’-(7-nitrobenz-2-oxa-1,3-diazol-4-y1)ethylenediamine</td>
</tr>
<tr>
<td>iGEM</td>
<td>international Genetically Engineered Machine competition</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>Kd</td>
<td>dissociation constant</td>
</tr>
<tr>
<td>L-GSH</td>
<td>reduced glutathione</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-(N-morpholino)propanesulfonic acid</td>
</tr>
<tr>
<td>MSL</td>
<td>multilayer soft lithography</td>
</tr>
<tr>
<td>MTS</td>
<td>methanethiosulfonate</td>
</tr>
<tr>
<td>Acronym</td>
<td>Term</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>ODE</td>
<td>ordinary differential equations</td>
</tr>
<tr>
<td>oligo</td>
<td>oligonucleotide</td>
</tr>
<tr>
<td>ppGpp</td>
<td>guanosine pentaphosphate</td>
</tr>
<tr>
<td>QCR</td>
<td>quantitative cysteine reactivity</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SN</td>
<td><em>Staphylococcal</em> nuclease</td>
</tr>
<tr>
<td>TBS</td>
<td>tris(hydroxymethyl)aminomethane buffered saline</td>
</tr>
<tr>
<td>TCEP</td>
<td><em>tris</em>(2-carboxyethyl)phosphine</td>
</tr>
<tr>
<td>$T_m$</td>
<td>thermal midpoint of denaturation</td>
</tr>
<tr>
<td>tmGBP</td>
<td><em>Thermotoga maritima</em> glucose binding protein</td>
</tr>
<tr>
<td>TnT</td>
<td>in vitro Transcription and Translation</td>
</tr>
<tr>
<td>TNT</td>
<td>trinitrotoluene</td>
</tr>
</tbody>
</table>
Acknowledgements

I thank my advisors, Dr. Hellinga and Dr. You, for their guidance, support, and critiques that have helped me become a better scientist. I acknowledge the time spent by and feedback from my other committee members, Dr. Beratan, Dr. Tian, Dr. Zhou, and Dr. Kreuzer. I am deeply indebted to the current and former members of the Hellinga and You labs for their invaluable training, assistance, suggestions, and camaraderie. I would not be where I am today without the love and encouragement of my family, particularly my parents who would always remind me of the high points, and my brother Stephan who could commiserate about the low points. Most of all, I will be forever grateful to my wife Amelia who suffered the hardships of graduate school alongside me and enabled me to pursue my passion with her love and support.
1. Background

1.1 Biology by Design

Biological research is experiencing an increasing focus on the application of knowledge rather than on its generation. Thanks to increased understanding of cellular systems and technological advances, biologists are more frequently asking not only “how can I understand the structure and behavior of this biological system?” but also “how can I apply that knowledge to generate novel functions in different biological systems or in other contexts?” Active pursuit of the latter has nurtured the emergence of synthetic biology.

Imagine you have been charged with building a robot capable of complex and autonomous operations in a dynamic environment. What are the most valuable characteristics to build into such a machine? To perform work, energy will be needed -- renewable energy extracted from the environment is ideal. To respond with meaningful behavior, information gathering and possessing capabilities will be required. For coordinated operations, communication with others is essential. To maintain long term function, a self-contained repair or reproduction system will be necessary. For many applications, it would need to be miniscule. To achieve economic feasibility, production costs will have to be low. While all these requirements are significant hurdles to the robotics engineer on a budget, they are feats that life has accomplished time and time again.
Consider one of the simplest forms of life, bacteria. Bacteria, often only micrometers in length, are capable of many of the above requirements including, entering miniscule environments, surviving on local nutrients, and responding to fluctuations in their environment with adaptive behavior (such as chemotaxis [1], altered nutrient utilization [2], and temperature-dependent gene expression [3]). Many bacterial species communicate in order to produce coordinated behavior [4] and with doubling times as fast as 20 minutes, their reproduction capacity is remarkable.

In fact, an engineer building a device on a bacterial “chassis” would only need to build one functioning prototype, culture overnight in low-cost media, and return the next morning to obtain trillions of virtually identical copies. In a sense, this is like programming a minuscule but complex computer that can also reproduce. As appealing as this concept may seem, several fundamental questions arise: what functions are we capable of programming into a living organism? To what extent will these functions be performed predictably and robustly? What is the best way to implement a pre-defined design goal, and what challenges and opportunities may arise? These are some of the questions that the burgeoning field of synthetic biology is beginning to address.

Over the past few years, synthetic biologists have generated remarkable systems including: an expanded genetic code in *Escherichia coli* [5], various logic gates [6,7], rewired yeast mating and osmolarity response circuitry [8], bistable switches in bacteria [9,10], yeast [11], and mammalian cells [12], photographic bacteria [13], genetic and
metabolic oscillators [14,15,16], artificial communication in bacteria [17] and yeast [18], and many other interesting and useful systems.

Although there is debate about the scope and boundaries of the field, some advocates supply that “synthetic biology” is:

A) the design and construction of new biological parts, devices, and systems, and

B) the re-design of existing, natural biological systems for useful purposes.” [19]

It is worth examining this definition more closely. Inherent in part A are engineering principles -- the notions of abstraction and hierarchy. One level of abstraction consists of biological components with simple albeit well-defined functions, operating under defined conditions, i.e. parts. At a higher level of abstraction, parts can be combined to form devices. Similarly, devices come together to form systems on a third level of abstraction. The basic premise is that an individual researcher can work at one of these levels without necessarily requiring details about the precise mechanics of operation on another level [20].

Part B states that biology is being redesigned for “useful purposes”. What purposes you might wonder? The first purpose may be obvious, and it is the practical application of biologically modified organisms in human life. Although our ancestors did not possess the advanced genetic tools available today, the litany of domesticated species including fermentation yeasts, crop grains, and silkworms is a testament to the vast utility of modified living organisms to humans. However, modification of living
organisms by traditional means, i.e. artificial selection, is an incremental and slow process with limited pay-offs during an individual’s lifetime. For example, it has taken approximately 15,000 years of domestication by selective breeding to turn wolves into present day dogs [21], a process which grouped desirable genes in particular breeds. Improvements in DNA synthesis and genomic engineering methods have enabled the introduction of genetic changes in relatively short time frames. Such technologies will engender the practical application of modified biological systems to new areas such as therapeutics, renewable energies and others. The practical applications of modified biological systems represent the first useful purpose behind redesign.

Of course, even possessing large scale DNA technology capable of making the changes needed to produce, for example, a guide dog from a wolf would not be enough. The necessary DNA changes have to be known in advance in order to be made. This is far from the reality - especially for a complex organism like the dog. Comparative genomics can elucidate the differences between the organisms, but does not yield the full understanding needed to prospectively say “If I want to program guide animal functions into organism X, here are the changes I will make and this is how those changes work.” In the venerable words of physicist Richard Feynman, “what I cannot create, I do not understand” [22]. The laws of physics and chemistry apply to living systems just as they apply to non-living things, such as mechanical engines. Yet designing and constructing even simple biological systems remain major challenges
whereas mechanical engines can be predictably engineered. Feynman would conclude that there must exist fundamental gaps in our understanding of how biological systems operate. Synthetic biology is exploring these gaps in understanding by attempting to build and apply such systems.

Scientific experiments are run under specific conditions in hope that the conclusions drawn will be applicable in a broader context. The creation of biological systems by utilizing currently accepted (or debated) principles will test the limitations and applicability of those principles. Likewise, implementation of existing genes, proteins, and pathways in non-native settings can help elucidate their functions and reveal unknown requirements for their operation. Synthetic biologists therefore aim not only to produce interesting and useful designs, but to simultaneously develop a greater understanding of biological components and design principles in general [23]. Therefore the second, and equally important, purpose of synthetic biology is to gain the biological insight that arises from testing our knowledge during the design and implementation process.
1.2 Foundational Technologies

Just as the development of the microscope made possible the discovery of cells [24], new technologies are providing the critical foundation needed for synthetic biology. Here we discuss four major advances that have produced enabling tools for experimentation and analysis in this regard. They are: DNA synthesis, parts and devices design and optimization, systems modeling, and observational capabilities. Figure 1 illustrates an overview of where these technologies interact with synthetic biology.
Figure 1: Interplay between engineering tools with a biological hierarchy. In order to simplify biological design, it is valuable to apply an abstraction hierarchy. DNA codes for "parts" that interact with each other to form circuits. The totality of all circuits and structures forms a single cell, which interacts with its neighbors and environment to form a population. At each level, technologies have been developed to assist and enable design. Shown here are the major advances that significantly reduce longstanding design, analysis and production barriers. Together, these technologies are helping to make integrated biological design a reality.
1.2.1 DNA Synthesis

At the core of every living thing, dictating that organism’s characteristics and behavior is a string of nucleotide bases - its DNA. To reprogram an organism, that DNA needs to be altered or supplemented. Until recently, DNA manipulations were almost exclusively done in a ‘copy, cut, and paste’ manner using polymerases, restriction endonucleases, and ligases respectively. While this enzymatic approach has produced a wealth of scientific advances, implementing a complicated biological design by these means is the literary equivalent of writing a paper using a photocopier, scissors, and a stick of glue. Recently, however, biologists have received their metaphorical “typewriter”. Tian and colleagues developed a large scale DNA synthesis method by utilizing parallel oligonucleotide synthesis on a programmable microfluidics chip, followed by PCR amplification [25]. In order to reduce the error rate to 1 in 1394 bases, the authors hybridized their “construction oligos” against complimentary “selection oligos” and washed away mismatches. Construction oligos were assembled into larger genes using polymerase assembly multiplexing, an overlap PCR based method. Utilizing their chip-based technology, the authors simultaneously synthesized and optimized all 21 genes encoding the 30S ribosomal subunit from *E. coli*.

In a different study, Jacobson and co-workers develop a method that can further improve the removal of error-containing DNA fragments [26]. Using this method, which exploits the gel mobility shift apparent when MutS binds a mismatched double stranded
DNA, they were able to obtain an error rate as low as 1 in 10,000 (the average length of a prokaryotic gene is 924 base pairs, while that of a eukaryote gene is 1346 [27]). By applying these and other technologies, commercial companies are now able to offer large-scale (multi-kilobase and up) DNA synthesis for under $0.50 per base with a 2-3 week turnaround time. Whereas these prices do not make a 10 kb construct inexpensive for most researchers, they imply that commercial synthesis has begun to rival the equipment, materials, labor, and validation costs incurred by traditional cloning and construction means for select applications. It may be possible to further reduce these costs by 10-100 fold, i.e. 1 to 10 cents per base, within the next decade by fully automating and streamlining new high-throughput techniques (Jingdong Tian, personal communication, 2006).

Total DNA synthesis can be used to alter or improve the sequences being built. In traditional cloning, targeted mutational changes are made only to small regions (~20 bases) at once. Furthermore, each region altered imposes additional experimental steps. DNA synthesis methods, however, can synthesize an altered sequence with no more effort than that necessary to synthesize a wild type sequence of the same length. For example, a protein coding sequence can be matched with regard to codon usage in the host organism where it will be expressed. In this case the sequence of amino acids in a protein is left unaltered by the modification, but translation efficiency can be improved by utilizing codons whose cognate tRNAs are more abundant. Similarly, the sequence...
can be altered to remove or create mRNA secondary structures without changing the resulting amino acid sequence. Furthermore, a gene whose sequence is known, but whose DNA is hard to obtain, can be easily synthesized.

1.2.2 Design and Optimization of Parts

One level of abstraction from the DNA synthesis and manipulation, is parts production, which can be accomplished either through rational design or directed evolution. Recently, improved algorithms and processor power have allowed computational design efforts to achieve new milestones in reprogramming the function of many well-characterized natural proteins. In a series of studies integrating both computation and experiments, the Hellinga lab succeeded in introducing an allosteric control switch into the proton-ATP pump (in collaboration with the Montemango group [28]); and retooled sugar-sensing receptors to bind novel ligands such as lactate, trinitrotoluene (TNT), and serotonin. Significantly, they even demonstrate that designed parts are active in vivo and can be used to produce more complex systems. The TNT receptor and a designed Zn receptor were shown to induce gene expression in response to exogenous ligands when implemented in some of the earlier reported examples of synthetic signaling pathways [29,30].

Computational design has also found applications beyond altering specificity or enzymatic function. For example, the Baker lab has designed a new protein that folds to form a novel structure – matching their modeling predictions [31]. They also apply their
computational methods to increase the thermostability of an enzyme by identifying key mutations. When some mutations were applied in concert the result was a 30-fold increase in half-life at 50°C [32]. The examples here illustrate altered specificity, novel functions and structures, improved stability, and introduction of allosteric control. They highlight some of the contributions that computational protein design has made for parts generation and improvement. While we have only drawn examples from two research groups, computational protein design is a vast and growing field, with important contributions made by numerous other labs [33].

Applying rational design to parts alteration or creation is advantageous in that it can both generate products with defined function, and also produce biological insights into how the designed function comes about. However, rational design requires prior structural knowledge of the part, which is frequently unavailable. Directed evolution is an alternative method that can effectively address this limitation by allowing parts engineering without design. In essence, directed evolution begins with the generation of a library containing many different DNA molecules, often by error-prone DNA replication, DNA shuffling, or combinatorial synthesis. The library is then subjected to high-throughput screening or selection methods that maintain a link between genotype and phenotype in order to enrich for molecules that produce the desired function. The process is then iterated to approach a desired endpoint [34,35,36]. A recent example of parts creation by directed evolution is expansion and alteration of LuxR specificity for
acyl-homoserine lactone ligands [37,38]. LuxR is a transcriptional activator from the marine bacteria *Vibrio fischeri*, and is naturally responsive to the signaling molecule 3OC6HSL. Collins *et al.* first employed a screening scheme to identify mutations that broadened the binding specificity of LuxR to other small molecules in the same class as 3OC6HSL [37]. They then used a dual-selection method [39] to redirect LuxR specificity to one of those molecules, C10HSL [38]. The result was a new protein that responds to the second chemical but no longer to the first. These parts may be particularly beneficial to designers desiring multiple channels of simultaneous communication between cells.

Directed evolution can also be applied at other levels of biological hierarchy, for example to evolve entire gene circuits [40].

Rational design and directed evolution should not be viewed as opposing methods, but as alternate ways to produce and optimize parts, each with their own unique strengths and weaknesses. Directed evolution requires a high-throughput way to screen or select for desired function and requires that functional mutants exist in the sequence space sampled. This second constraint becomes less likely as the desired function diverges further from the initial function. On the other hand, while rational design strategies can make multiple changes or large scale alterations that incorporate scientific knowledge, these strategies are rarely precise enough to finely tune system behavior. Furthermore, it is difficult to know if additional optimization is possible when employing rational design. For these reasons, both methods can and should be used in
conjunction and will hopefully continue to be applied in unison during the years to come.

Recent years have witnessed increasing interest in using parts based on RNA for intricate control of gene expression [41,42]. One particular line of research has been largely inspired from metabolite-controlled riboswitches prevalent in nature [43,44]. RNA switches are advantageous in their fast response, broad applicability, and chemical nature. RNA switches contain a ligand binding region, or aptamer domain, that controls the function of an effector domain through binding induced conformational changes. Strategies for evolution of RNA aptamers and functional RNAs developed early on [45,46,47] due to the fact that the same molecule plays both functional and information-encoding roles (i.e. the genotype-phenotype link required for directed evolution schemes is intrinsic to the molecule). This allows the generation of a library directly from the products of a competitive screen in the previous round. Furthermore, the entire selection, amplification, and iteration procedure can be economically accomplished in vitro. The chemical nature of RNA, with 4 bases possible at each position, means that a higher percentage of available sample space can be covered while evolving an RNA molecule than a protein of similar length (20 amino acid possibilities per position). Additionally, the interactions within an RNA molecule are largely driven by complementary base pairing. As a result, relatively accurate methods for secondary structure prediction of RNA have been developed and are widely used [48,49].
Secondary structure information is valuable because it can allow a researcher to make rationally guided changes.

In recent work from the Smolke group [50], switches were developed that exposed an anti-sense stem sequence upon binding a ligand, producing a riboregulator. Ligands, such as theophylline, controlled switches that turned on gene expression, as well as switches that turned off gene expression. These switches were shown to be tunable by making simple changes to the RNA sequence guided by thermodynamic properties. Multiple switches functioned independently in yeast even when binding similar molecules. Switches such as these may be useful in sensing cellular conditions and could also act as feedback mechanisms for tuning metabolic pathways in response to the depletion or accumulation of reactants, intermediates, or products. Gallivan and colleagues demonstrate a synthetic RNA switch that is functional in prokaryotes and can be applied in screening or selection schemes that tie in vivo levels of small molecules to a reporter gene or cell survival, respectively [51]. In this manner, one could screen enzyme libraries for a desired catalytic function. Inversely, if the small chemical is supplied, then a library of riboswitches could be screened for binders that alter gene expression. Suess and colleagues, who first describe a rationally designed in vivo RNA switch, implement it in such a way that it functions as a logic gate with another ligand, xylose [52]. Perhaps the best known form of gene regulation by RNA, however, is the role of interfering RNA [53]. Yokoboyashi and colleagues show that it is possible to modulate shRNA activity
through the action of a small chemical by fusing the shRNA to an aptamer that responds to the chemical [54].

Synthetic riboregulators need not be ligand controlled. Collins and coworkers demonstrate a general method to introduce RNA mediated post-transcriptional regulation into prokaryotic genes [55]. They introduced a short sequence between the promoter and ribosome binding site that when translated into mRNA folds into a hairpin with the adjacent ribosome binding site, sequestering the site and preventing translation. Translation can be restored by expressing a trans-acting RNA that binds the hairpin and forms a more stable structure, which frees the ribosome binding site.

These examples demonstrate that the cellular engineer of the future will not be restricted to the catalogue of known biological parts, but will also have the tools needed to supplement natural parts with custom-made parts for specific applications.
1.2.3 Modeling-Guided Circuit Engineering

The engineering process usually involves multiple cycles of design, optimization, and revision (Box 1 and Figure 2). This is particularly apparent in the process of constructing gene circuits. As the number of interacting parts and reactions increases, it becomes more difficult to intuitively predict circuit behavior. Towards these ends, mathematical modeling is a useful design tool, in particular for systems with complex dynamics such as bistability and oscillations. The importance of mathematical modeling has been increasingly appreciated, as evidenced by its extensive application in systems biology as a way to decipher “design principles” of natural biological systems [56,57,58,59]. In comparison, the utility of modeling in synthetic biology seems even more dominant [60,61].
Box 1: A recipe for engineering gene circuits (also see Figure 2)

DESIGN
1) **Determine the design goal**
   For the purposes of this tutorial we will attempt to construct a population of cells that restricts its cell density below that imposed by nutrient limitations [62]. The implementation discussed below is a revised version of the circuit[63].

2) **Pick suitable host organisms/strains**
   Key characteristics to consider here are: ease of genetic manipulation, growth rate, survivability under the desired conditions, and endogenous machinery you wish to exploit. *E. coli* could be used for this application.

3) **Identify necessary “parts”**
   Available places to draw from include the literature, genome sequences, colleagues, and the MIT registry [64]. Recall that: a) Parts need not be from the host organism. While native parts are likely to function properly, they can lead to crosstalk with endogenous systems. b) Parts need not exist; they can be developed by rational design or directed evolution. c) The better characterized the parts, the easier your job will be. d) It is advantageous to include parts as reporters. In this tutorial we will pick the quorum sensing genes *luxR* and *luxI*, as well as the toxin gene (*ccdB*) from F plasmid segregation.

MODELING
4) **Build a mathematical model**
   Start with the simplest model that can capture the circuit dynamics (for example a simplifying assumption might be to assume a protein’s production rate depends on a transcription factor rather than explicitly modeling mRNA production, translation, and decay).

5) **Explore circuit dynamics in silico**
   Address questions like: Can the network architecture give you the function you want? What parameters are most critical for success? How do circuit dynamics change with parameters?

IMPLEMENTATION, TESTING, AND DEBUGGING
6) **Determine the DNA implementation of your circuit.**
   In our case we will implement our circuit on a plasmid and need decide on copy number, what promoters, RBSs, transcription terminators, and perhaps degron tags to use. Another choice at this time is to decide if any components need to be expressed together on a polycistronic RNA. In this example, the circuit is implemented in a medium copy number plasmid (p15a origin) which the *luxR* and *luxI* gene are co-expressed by a P<sub>lac/ara</sub> promoter. The *ccdB* gene is controlled by a P<sub>luxI</sub> promoter. Kanamycin resistance is used as a selection marker.

7) **If modeling indicates that a particular parameter is critical, build multiple versions**
   It is rare for all parameters to be perfectly balanced on the first experimental implementation. Designing multiple circuits at once to sample a critical parameter space can increase the chance for initial success. It may also yield interesting information about whether that particular parameter is truly critical.

8) **Test your circuit and decide whether to retest, revise, or redesign**
   If it works as predicted you can continue to fully characterize it. If not, can you fit your model to explain the behavior that is observed? What parameters may need altering to generate the desired function? At this point you can: (1) redesign the circuit to address critical parameter changes, and perhaps “fine tune” the circuit function by directed evolution; or (2) test the circuit in other strains or growth conditions.

A working design usually requires multiple rounds of iteration of steps listed above, which is often the most time consuming portion of biological design.
Figure 2: The typical process for engineering gene circuits (this figure accompanies Box 1).

Various mathematical formulations can be used to model gene circuits. At the population level, gene circuits can be modeled using ordinary differential equations (ODEs). In an ODE formulation, the dynamics of the interactions within the circuit are deterministic. That is, given the same initial condition and numerical configurations, different rounds of simulations will lead to exactly the same results. In other words, the ODE formulation ignores the randomness intrinsic to cellular processes and is convenient for circuit designs that are thought to be less affected by noise or when the impact of noise is irrelevant. For instance, ODE models have been used to guide
experimental efforts to program population dynamics in the temporal domain [62,63] or
the spatial domain [65,66]. Importantly, an ODE model facilitates further sophisticated
analyses, such as sensitivity analysis and bifurcation analysis. Such analyses are useful
to determine how quantitative or qualitative circuit behavior will be impacted by
changes in circuit parameters; this has been almost a standard practice in engineering of
most gene circuits accomplished so far (Box1). For instance, in designing a bistable
toggle switch, bifurcation analysis was used to explore how qualitative features of the
circuit may depend on reaction parameters [9]. Results of the analysis were used to
guide choice of genetic components (genes, promoters, and ribosome binding sites) and
growth conditions to favor a successful implementation of designed circuit function.

In a single cell, however, a gene circuit’s dynamics often involve small numbers
of interacting molecules. Such small numbers will result in highly noisy dynamics even
for expression of a single gene [67,68]. For many gene circuits the impact of such cellular
noise may be critical and needs to be considered. This can be done using stochastic
models [69]. Different rounds of simulation using a stochastic model will lead to
different results each time, which presumably reflect aspects of noisy dynamics inside a
cell. For synthetic biology applications, the key of such analysis is not necessarily to
accurately predict the exact noise level at each time point. This is not possible even for
the simplest circuits due to the “extrinsic” noise component for each circuit [68]. Rather,
it is a way to determine to what extent the designed function can be maintained and,
given a certain level of uncertainty or randomness, to what extent additional layers of control can minimize or exploit such variations. For instance, a number of computational studies have been conducted to analyze the potential of cell-cell communication to synchronize intrinsically noisy and unreliable oscillators [70,71].

Mathematical models, either stochastic or deterministic, can be digitally “evolved” in silico to generate optimal circuit designs that satisfy a particular objective. Francois and Hakim used genetic algorithms to design gene regulatory networks that exhibited hysteresis or oscillations [72]. Initially, a pool of gene circuits was constructed from basic reactions representing activation, repression, and post-translational modification. These circuits were subsequently evolved using numerical simulations to obtain a desired output by repeated rounds of digital “mutations” and functional “screening”. Several unique designs were generated that satisfy each design goal. These designs could serve as alternatives to consider, model, or test during the circuit engineering cycle.

One of the most exciting aspects of synthetic biology is the multiple avenues being used to address questions. While some researchers may only apply a particular method for a given application, the domain as a whole will benefit from the use of these complementary approaches. For example, a simple linear cascade can be implemented using transcriptional regulation or reversible protein modification, both of which are prevalent in nature. Implementation by transcriptional control is appealing because it is
generally easier to stitch multiple DNA elements together. However, multi-component transcriptional cascades can introduce a significant time delay, as shown by Hooshangi et al. [73]. In this work, a one-stage cascade reached its half-maximal activation in minutes, whereas a 3-stage cascade took several hours. Rosenfeld and Alon found that long transcriptional cascades are rare in the sensory systems of relatively short lived E. coli and S. cerevisiae [74]. Protein-modification based circuits can offer much faster temporal response [75]. As the field matures, it is probable that synthetic circuits, like nature, will integrate both DNA and protein regulatory logic in their design. The combination will exploit advantages of each method while mitigating their weaknesses. These choices will require mathematical modeling to ensure that the circuits can perform on the desired time-scale for a particular operation.

In most attempts to engineer gene circuits, mathematical models are often purposefully simplified to capture the qualitative behavior of the underlying systems. Simplification is beneficial partially due to limited quantitative characterization of circuit elements, one limitation that the BioBricks project aims to address [64], and partially because simpler models may better reveal key design constraints. The caveat, however, is that a simplified model may fail to capture richer dynamics intrinsic to a circuit. When engineering a population controller, we built a highly simplified kinetic model to capture the essence of the circuit dynamics including cell growth, signal accumulation, killer protein accumulation, and subsequent cell killing. The model predicts that the
system will always lead to a stable regulated state, and this prediction was supported by observations made in batch cultures [62]. Yet, later, we observed sustained oscillations when cells expressing the circuit were grown in a microchemostat [63]. One way to reconcile the experimental and modeling results was to introduce an extra step of regulation in our model, which indeed resulted in sustained oscillations for biologically feasible parameters. We note that still more layers of regulation are involved, further complicating modeling analysis (Figure 3).

Figure 3: Complexity and uncertainty in biological circuit design. Although we can build and model the circuit from Box1, it is remarkably difficult to capture even all the known interactions (let alone the unknown interactions). In our model, we have a single killing term which sets the rate of cell death proportional to the product of the killing rate constant, CcdB level, and cell number. In reality, the situation is far more muddled. CcdB operates on DNA gyrase in a manner whose mechanistic details are still open to debate. The downstream effects of CcdB are plural and inter-related, and each of these involves many components. For example the SOS response involves over a dozen players. Attempting to incorporate all the partially understood downstream effects would complicate the model with no guarantee of improving its accuracy. Nevertheless, by omitting them, we make the implicit assumption that they do not affect system dynamics.
1.2.4 Culturing and Monitoring Technologies

To determine if a synthetic circuit works as designed, one must be able to test it and observe its dynamics. These tasks have benefited from the rapid development of improved culturing and observational technologies. An ideal method for monitoring cellular dynamics over time should be easy to perform and should not significantly affect the properties being measured. One step towards this ideal has been the engineering of fluorescent proteins variants [76]. These proteins are genetically encoded and mature to functionality without requiring co-factors. Each variant fluoresces with a specific visible wavelength upon excitation, allowing multiple variants to be discerned in one cell.

Fluorescent proteins can directly report on protein levels when present in translational fusions or indirectly report when present in transcriptional fusions. A translational fusion is made by inserting a fluorescent protein into the reading frame of the target protein resulting in translation of the fluorescent protein and target protein as one molecule. That is, one can tag a target protein with a fluorescent tail. In many cases, this does not significantly affect the function of either a target or a fluorescent protein. A transcriptional fusion is made by co-expressing a fluorescent protein and a target protein by placing each behind the same promoter. While this strategy reports on promoter activity, a key determinant of intracellular levels, it fails to capture any post-transcriptional or post-translational regulation, such as the action of regulatory RNAs or
proteases. With both transcriptional and translational fusions, fluorescence measurements are noninvasive to live cells, and the process can be automated for long term measurements. Fluorescent proteins therefore represent an elegant solution for monitoring *in vivo* protein levels. Caution must be exercised with translational fusions, however, because even if the fluorescent tag does not alter the target protein’s function *per se*, it may significantly impact its localization. Although many of such cases are unreported, the literature is spotted with examples of mis-localized or mis-transported fluorescent fusion proteins [77,78]. This is an important issue not only for studies that explore protein trafficking, but for any system where altered localization will affect function.

A particularly appealing application of fluorescent proteins is to monitor single cell dynamics in real time through optical microscopy. Single cell measurements are critical for revealing heterogeneity in gene expression or differences in other phenotypic traits between cells that are often masked in population-level measurements. In one of the earliest synthetic circuits published, Elowitz and Leibler built a circuit capable of producing oscillations in gene expression, but it was only through the microscopic tracking of individual lineages of bacteria that the oscillations became truly apparent [14]. Similar techniques were used to characterize other oscillators implemented later [15,16]. Recently, single cell measurements have become the workhorse for a series of
elegant experimental studies aimed at deciphering that origin and characteristics of cellular noise [67,68,73,79,80,81,82,83,84,85].

Remarkably, measurement capabilities are continuing to improve in resolution as tools to track single molecules in vivo have also been developed. Building on previous mRNA visualization techniques [86], it is now possible to track individual mRNAs in vivo by using multiple fluorescent mRNA binding proteins [87,88,89]. Yu et al. show that it is even possible to detect a single fast maturing fluorescent protein by targeting it to the membrane [90]. These detection methods improve researchers’ abilities to quantify the abundance and localization of cellular components. Researchers can then determine when and where the experimental system deviates from their expectations, improving their ability to test and troubleshoot designs.

It is a rare and joyous occasion when a synthetic genetic circuit actually works as expected the first time. The laborious and time consuming process of characterizing and debugging biological programs will become more significant as the circuits increase in complexity. This process is, by and large, the rate-limiting step for engineering gene circuits that program sophisticated dynamic behavior (See Box 1). An important advance in this area is the miniaturization of characterization processes through microfluidics -- the science and technology of systems that manipulate small amounts of fluids (10^-9 to 10^-18 liters), using micro-sized channels [91,92]. Microfluidic metering enables ultra-low consumption of biological samples and reagents, allowing high-throughput research at
low cost with short analysis time. Microfluidic miniaturization also facilitates automation and integration of complex chemical or biological procedures into a single process that is faster, more precise, and more reproducible than its manual counterparts. Pioneered by the Quake lab, the development of actutable pneumatic valves through Multilayer Soft Lithography (MSL) has facilitated the design of complicated devices equipped with pumps, fluidic isolation, and mixers [93].

As a proof-of-concept for synthetic biology application, Balagadde and colleagues devised and implemented a miniaturized 16 nL bioreactor, called a microchemostat, that enables automated culturing and monitoring of small populations (10^2 to 10^4) of bacteria for hundreds of hours with single cell resolution [63]. By reducing the reactor volume by a factor of 10^5 as compared with traditional chemostats, microchemostat populations undergo proportionately fewer divisions per hour, which suppresses the total mutation rate of the population. This in turn effectively insulates the micro-cultures from rapid evolution, prolonging monitoring of genetically homogeneous populations. The microchemostat system is automated by custom software which controls periodic media dilution, culture mixing, image acquisition, and image analysis. Its unique design also allows multiple experiments to be run in parallel on the same chip (Figure 4).
Figure 4: A microfluidic chip with six parallel microchemostat reactors, used to study the growth of microbial populations. The coin is 18 mm in diameter.

In another microfluidics application, Thorsen and colleagues created a “comparator” capable of screening individual cells for desired functionality in a high-throughput manner. In this device, two reagents can be separately loaded into 256 pairs of subnanoliter reaction chambers. Adjacent chambers are united allowing the reagents to mix and react. The products of each reaction can then be selectively recovered. This system was used to perform a high-throughput detection of single bacterial cells expressing recombinant cytochrome c peroxidase [94].

Fu and colleagues fabricated a microfluidic fluorescence activated cell sorter (FACS) to sort live fluorescent E. coli cells. Compared with conventional FACS machines,
the microfluidic device allows for more sensitive optical detection of bacterial cells as well as DNA strands, and is also capable of “reverse” sorting. Reverse sorting is a procedure where cells are scanned at a high flow rate, until a fluorescent cell is detected. Flow is then stopped and reversed, allowing the cell to be measured a second time and diverted into a collection tube. Reverse sorting is particularly useful for isolating rare cells or making multiple measurements on a single cell [95].

The aforementioned microfluidic devices can be used in stand-alone applications or as part of an integrated system. They are also disposable, which eliminates any cross-contamination in between runs. These and many other microfluidic systems [96,97,98] being actively developed will become important tools for synthetic biologists [99].
### 1.3 Outlook

It has been suggested that many of the difficulties in the production and optimization of biological circuits are due to improper and incomplete description of parts [20]. These limitations are twofold: first, functional characteristics are often unknown for many parts; second, even if they are known, they are rarely described using standardized measures and are often buried in the literature. Towards addressing these limitations, the BioBricks Foundation has established a “registry of standard biological parts” [64]. The registry categorizes parts, devices, and systems. Ultimately the registry strives to provide information on not only sequence but also functional characteristics, and make information available through a central portal. Many of these parts have been cloned into plasmids that enable easy assembly. The plasmids are made available to students participating in the international Genetically Engineered Machine competition (iGEM). Members of the BioBricks Foundation hope that the registry will decrease the time and research costs needed to design and implement gene circuits. Such efforts are analogous in spirit to ongoing attempts to standardize mathematical models [100] and formats for microarray data [101]. The limits in achieving parts standardization for *E. coli* and other organisms remain to be seen.

Even with a repository of information about standardized parts, a major challenge to applying this information will be developing strategies to deal with context dependence [102,103]. For example, synthetic gene circuits often exhibit varying
behavior in different cell strains. In some cases this can be easy to rationalize by the presence or absence of a particular gene, or a documented difference in growth rate. In other cases, causes of variability are much more difficult to ascribe due to many hidden interactions between the designed circuit and a far-from-elucidated host circuitry.

To address this issue, one may imagine selecting a standard cell strain, in which standard parts under standard conditions are to be quantified. A starting point for such a standard strain may be on its way. The Blattner group has recently engineered a series of multiple deletion strains (MDS), that have up to 15% of their parental MG1655 genome removed but maintain similar growth rates on minimal media [104]. Deletions were guided by comparative genomics with related strains. Removing “unnecessary” portions of the genome can presumably reduce the number of hidden interactions. Notably, the deletions cleaned the cells of mobile DNA elements called insertion sequences (IS) that might reduce the genetic stability of a circuit by inserting themselves into and disrupting a DNA sequence unpredictably. Interestingly, the MDS strains produced some unanticipated benefits, including higher electroporation efficiencies than their parent strain and the ability to propagate some plasmids that the parent strain could not.

In an alternate approach, researchers at the Venter Institute have used \textit{M. genitalium} as a starting point in their attempts to determine a minimal gene set by systematically mutating every gene [105]. \textit{M. genitalium} has the smallest known genome.
of organisms capable of growth in the absence of other species. They conclude that in a laboratory setting only 382 of the strain’s 482 genes are essential. In principle, a minimal strain could serve as a bare bones platform upon which desired functionality can be added. Such a small number of genes might allow a greater percentage of the cell’s molecular interactions and metabolic processes to be understood, making the strain more predictable and desirable as a starting point. However, of the 382 essential genes determined, 110 are annotated as hypothetical proteins or as proteins of unknown function, indicating that a truly complete cellular model, even for this simplest of cells, cannot yet be produced.

Despite characterizing parts in a standard strain under defined conditions, individual parts may impact the physiology of the host strain differently, for instance by placing varying burdens on the host translation machinery. For this reason, one may wish to minimize such interactions by creating privileged sets of machinery. For instance, Rackham and Chin [106] describe the formation of orthogonal ribosome–mRNA pairs that could be used to keep a synthetic system and host more isolated. Using a dual positive-negative selection scheme, they isolated mRNAs with modified Shine-Dalgarno regions not recognized by endogenous ribosomes, but instead recognized by alternative ribosomes. Translation by orthogonal pairs should be unaffected by endogenous ribosomes and there should be no competition for ribosomes between orthogonal mRNAs and traditional mRNAs. In principle, multiple ribosome
types can be implemented for specified function, just as cells already possess multiple DNA or RNA polymerase types which play specialized roles.

Previous and current progress promises an ever growing infrastructure that will no doubt tremendously benefit future synthetic biology research, fundamental and applied alike. Concerning standardization, however, two critical questions remain to be addressed by the community. First, given the amount of cell physiology (even for highly characterized organisms such as *E. coli*) that is still poorly understood, to what extent can we standardize parts or systems with confidence? Second, how much standardization can we afford and still hope to create useful systems that can work in complex environments such as in a cancer or a polluted environment?

There is little difficulty in unambiguously defining the DNA sequences that code for parts, be they proteins or RNAs. The true challenge lies at the functional levels. Parts will impact and be impacted by cell physiology, which also changes in response to environmental conditions. In addition, parts tested in isolation may unpredictably impact each other’s functions when combined. For example, connecting one part’s DNA with another part’s DNA may introduce unintended regulation by introducing enough flexibility in DNA to allow DNA looping. For these reasons one can rarely have complete confidence in the part’s function even if used it in a standard strain characterized under a standard condition. Many such interactions are still poorly understood, complicating the use of standard parts. Yet, it is precisely this complexity
that makes engineering biology challenging and interesting. Decoding this complexity is at least one important application of synthetic gene circuits. Without a much deeper understanding of cellular functions at all levels, it is difficult to even define standards meaningfully.

From a practical standpoint, too much standardization may remove flexibility in engineering useful systems. It would be illogical to rely only on standard strains that lack desirable properties for a particular application. Consider thermophilic bacteria, capable of life at temperatures as high as 113°C [107]. The ability to thrive at elevated temperatures may be a useful property for synthetic organisms involved in chemical processing, because higher temperatures speed kinetic rates. Given the difficulty in thermo-stabilizing even a single protein, however, it is unlikely this quality can be engineered into a standard strain. For many applications the researcher is left with no appealing options except to use non-standard strains. No single strain or growth condition can ever cover all potential synthetic biology applications.

If we remain dedicated to standardization, gathering standardized information for a set of potentially useful parts, in a set of useful strains, under a set of relevant conditions becomes a combinatorial nightmare. The inevitable result is that standards will only be available for a limited number of strains and conditions. Although some information is preferable to none, a potential danger is undue weight will be placed on the limited information available and the tempting assumption that a part’s behavior
will not vary significantly from the context in which it was described will be made. In this situation, “significant” is considered to be variation that exceeds the acceptable tolerance limits of a part in its new device. Accepting standardized information at face value, without acknowledging its limitations, will lead one to design many systems doomed to fail. However, being aware of the limitations allows one to use standard information without depending on it, to be guided by the information while simultaneously embracing strategies like combinatorial design [108] and directed evolution [40] of circuits -- strategies that would be unnecessary in a fully standardized and predictable world.
2. A Protein Production and Characterization Pipeline

2.1 Introduction

2.1.1 In House Gene Synthesis and Assembly

As covered in the introduction, DNA synthesis is a critical enabling technology for bioengineering efforts. Although synthetic DNA is available at ever decreasing costs and increasing lengths [109], commercial partners generally cannot offer the control, flexibility and turnaround time achievable in-house. For this reason previous efforts in the Hellinga lab focused on the development of on-site gene synthesis methods, encompassing everything from in silico sequence planning through automated gene assembly and re-amplification [110]. Briefly, the desired mutations are input into custom software that designs long DNA oligonucleotides (~80-90 bp) and generates instructions for a liquid-handling robot to assemble them into a gene using a robust PCR-based assembly scheme. Oligonucleotides are synthesized in-house on two 192-well synthesizers, which reduces costs and ensures quality. Automated robotic assembly occurs in two stages. First ~300-500 bp fragments are made using an inside-out nucleation scheme; second, the resulting PCR reaction fragments are assembled into full length genes using a splice overlap extension method. With this technique, genes up to 2.5 kb have been synthesized. Flexibility in isocoding sequences allows genes to be optimized for codon usage, secondary structure, and mRNA GC content. Optimization is performed in accordance with the organism for which the DNA is intended and
ensures the potential for high expression [111]. Full protocols for gene synthesis can be found online. [112] Outputs from this stage of the pipeline are full length linear genes complete with promoters and transcriptional terminators.

2.1.2 in vitro Transcription and Translation (TnT)

The development of tools for gene cloning in the 1970s ushered in a new era in the study of protein function and enabled the development of biologic therapeutics. Nevertheless, traditional recombinant DNA tools are non-ideal when applied to a high-throughput protein engineering effort. Consider the paths to protein characterization starting from synthetically produced linear genes generated as described in the previous section. Traditionally, assessing the function of the protein variant requires a great many steps: digestion, DNA purification, ligation, transformation, colony picking, culturing, DNA isolation, archiving, DNA sequencing, sequence analysis, transformation to an expression strain, colony picking, culturing, induction, protein purification, and functional testing. Even under ideal circumstances this process requires over a week of time and upwards of $30/variant.

By contrast, given the same linear synthetic DNA, an approach featuring \textit{in vitro} expression requires few steps: set-up, incubation, purification and assaying. This approach is correspondingly fast and inexpensive - protein function can be addressed within 12 hours and for less than $3/variant. \textit{E. coli} lysate and master mix recipes and protocols were adapted and modified from those of the Swartz lab [113]. Briefly, in the
final step of gene synthesis, DNA is re-amplified using end primers that are 5’ biotinylated, a modification thought to protect the linear construct against exonucleases present in the lysate. Linear DNA fragments are mixed with the cell lysate and a master mix solution that contains various amino acids, dNTPs, and energy sources. The mixture is typically incubated at 30°C with shaking for 4-10 hours before purification. Purification is accomplished using affinity tags resin and is compatible with a 96-well format. Full protocols for in vitro transcription and translation (TnT) reactions can be found online [112].

The major limitation of in vitro expression is that minute amounts of protein, on the order of a microgram, are produced relative to traditional cloning. Sensitive and small-scale functional assays are therefore critical in order to take advantage of cell-free expression in a protein engineering pipeline. In this chapter I present the results of two assays adapted to TnT scale work, a conformationally sensitive fluorophore assay, and a thermal shift assay. In the final portion of the chapter, the thermal shift assay is exploited as a tool for screening full length in-frame clones, an important step in extending the pipeline beyond high-throughput screening to engineered protein validation.
2.2 Results

2.2.1 Covalently Coupled Conformationally Sensitive Fluorophores

Ligand binding can induce detectable conformational rearrangements in protein structure. Such movements frequently impact local structure in regions distinct from the binding site. Allosteric protein regulation is a prime example of harnessing structural rearrangements that are coupled with binding, thereby controlling function in a region away from the binding site. The same principle can be exploited in engineered proteins to produce biosensors that monitor binding. In one strategy, an environmentally-sensitive fluorophore is conjugated to a region of the protein where binding-induced changes in local structural cause an increase or decrease in fluorescence. Fluorescent signal therefore reports on ligand occupancy. Ideally fluorophore conjugation does not perturb affinity, but some loss of affinity is generally observed – on rare occasion, affinity increases. The periplasmic binding protein family, which features large hinge-bending motions between two sub-domains upon ligand binding, has been shown to be very amenable to this general strategy [114]. Although it is difficult to predict which combinations of site and fluorophore will result in large changes upon binding, a signaling variant with a 790% signal change has been produced [115]. This particular PBP variant, a cysteine introduced at position 131 in glucose binding protein from *Thermotoga maritima* (tmGBP) labeled with N,N'-dimethyl-N-(iodoacetyl)-N'-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)ethylenediamine (IANBD amide), was selected for
validation of the binding assay at the TnT level (Figure 5). Specifically, we sought to
determine whether the impact of alanine mutations in the binding site could be rapidly
and precisely determined using synthetic gene constructs expressed in TnT systems.

Figure 5: Schematic of a tmGBP biosensor. The crystal structures of wild type
tmGBP in the open form (top left, PDB: 3C6Q) and closed form (top right, PDB:
2H3H) show the structural re-arrangement upon ligand binding (red, glucose). The
location of the S131C mutation where the IANBD fluorophore is conjugated is shown
as space filled atoms. It is modeled dim in the apo state but bright in the closed state,
representing the large signal increase upon binding. Some of the binding site
residues selected for mutation to alanine are shown in the close-up view of the
binding pocket (bottom).

Fourteen side chains that comprise the ligand binding site were selected by
visual inspection on the basis of proximity to ribose in the crystal structure. Side chains
were individually mutated to alanine, gene variants constructed from synthetic
oligonucleotides, and protein expressed in vitro. Standard affinity purification protocols were modified to include an “on-bead” labeling step such that subsequent rinses served the dual purpose of removing contaminant protein and eliminating free, unconjugated, fluorophore. Ligand titrations were performed in 384 well plates in the presence of 0.1 mg/mL bovine serum albumin (BSA) because it was determined that without BSA, signal degraded if protein concentration was low (Figure 6). We hypothesize this effect is the result of nonspecific plastic binding by the protein that decreases the concentration of soluble protein in the center of well. Furthermore, the presence of BSA appeared to decrease noise at high signal, likely associated with the normalization of well menisci that is observed when BSA is added. BSA had no impact on apparent affinity.

Figure 6: Impact of BSA on tmGBP titrations. At modest protein concentrations (1.5μM, left) only slight signal degradation is observed over the course of three hours. However, BSA appeared to reduce noise at high fluorescence. At low protein concentration (15nM, right), BSA prevents the loss of signal over three hours at room temperature.

As a proof of principle to demonstrate the potential for “rapid response” protein engineering, gene synthesis was initiated on a Monday morning and subsequent steps
performed with as little downtime as possible. At this maximal speed, full results were obtained on Thursday afternoon and fully processed by Friday (Figure 7). Combined with the appropriate computational protein design tools, the 5-day turnaround achieved could ultimately be deployed to build biosensors against new chemical threats (nerve agents, etc.) with unprecedented speed.

**Figure 7: Rapid response protein engineering.** Actual timeline for the gene synthesis, expression, and testing of the first replicate of a tmGBP alanine scan. Although only 14 variants were tested in this case, scale up to 96 variants would require only one additional day at the stage of oligo synthesis.

The second half of the pipeline, TnT expression through data analysis, was repeated twice to determine the reproducibility of the binding parameters across three independent experiments. Representative ligand titrations (Figure 8) demonstrate that the majority of variants retain the ability to signal, and that most have reduced affinity. Full fit details for each variant can be found in Table 1. The variants were classified into three broad categories of signal quality. The first class, which contains wild type and six alanine mutants, is characterized by large fluorescence intensity changes (70% of wild type change or more), good fits to the Hill binding equation ($R^2>0.97$), reasonable cooperativity values ($n = 0.78-1.07$), and excellent reproducibility between replicates (Figure 8, top). The second class, which contains five alanine mutants, yields
reproducible data across replicates but fails in one or more of the mentioned categories (Figure 8, middle). The final class of variants suffers poor reproducibility, and would not be scored as signalers in a screening assay (Figure 8, bottom).

Table 1: Fit parameters for glucose titrations of a tmGBP alanine scan variants

<table>
<thead>
<tr>
<th></th>
<th>$K_d$*</th>
<th>Hill Coefficient*</th>
<th>R-square*</th>
<th>% wt signal*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt</td>
<td>21 (4) µM</td>
<td>1.06 (.13)</td>
<td>0.98 (.01)</td>
<td>100 (0)</td>
</tr>
<tr>
<td>K9A</td>
<td>37 (6) µM</td>
<td>1.07 (.15)</td>
<td>0.99 (.01)</td>
<td>77 (5)</td>
</tr>
<tr>
<td>S10A</td>
<td>110 (20) µM</td>
<td>1.04 (.12)</td>
<td>0.99 (.00)</td>
<td>76 (5)</td>
</tr>
<tr>
<td>H12A</td>
<td>470 (70) µM</td>
<td>0.90 (.13)</td>
<td>0.99 (.01)</td>
<td>72 (9)</td>
</tr>
<tr>
<td>F215A</td>
<td>1.8 (0.4) mM</td>
<td>0.82 (.14)</td>
<td>0.98 (.01)</td>
<td>78 (6)</td>
</tr>
<tr>
<td>Q236A</td>
<td>29 (14) mM</td>
<td>0.91 (.21)</td>
<td>0.99 (.01)</td>
<td>71 (13)</td>
</tr>
<tr>
<td>Y192A</td>
<td>5.6 (0.7) µM</td>
<td>0.78 (.14)</td>
<td>0.97 (.02)</td>
<td>77 (7)</td>
</tr>
<tr>
<td>Y14A</td>
<td>NR†</td>
<td>0.68 (0.71)</td>
<td>0.91 (.03)</td>
<td>51 (6)</td>
</tr>
<tr>
<td>W15A</td>
<td>3.6 (0.9) mM</td>
<td>0.46 (0.14)</td>
<td>0.98 (.02)</td>
<td>71 (3)</td>
</tr>
<tr>
<td>N137A</td>
<td>NR†</td>
<td>0.16 (0.05)</td>
<td>0.93 (.01)</td>
<td>45 (9)</td>
</tr>
<tr>
<td>R141A</td>
<td>21 (5) µM</td>
<td>0.80 (0.29)</td>
<td>0.90 (.04)</td>
<td>40 (9)</td>
</tr>
<tr>
<td>E164A</td>
<td>8 (1) µM</td>
<td>0.92 (0.26)</td>
<td>0.93 (.01)</td>
<td>47 (4)</td>
</tr>
<tr>
<td>D89A</td>
<td>NR†</td>
<td>45.14 (39.8)</td>
<td>0.01 (.04)</td>
<td>27 (11)</td>
</tr>
<tr>
<td>Y190A</td>
<td>NR‡</td>
<td>10.2 (13.0)</td>
<td>0.21 (.72)</td>
<td>39 (24)</td>
</tr>
<tr>
<td>D216A</td>
<td>NR‡</td>
<td>0.79 (0.59)</td>
<td>0.58 (.11)</td>
<td>30 (10)</td>
</tr>
</tbody>
</table>

* Numbers in parenthesis are the standard deviation from three independent expression and titration experiments.
†Not reported. Although reproducibility across experiments was good, these titrations do not saturate fluorescence and thus there is large uncertainty in the $K_d$ determined in the fit.
‡ Not reported. These titrations show poor reproducibility. $K_d$ values from fits are unlikely to reflect true affinity.
Figure 8: Normalized titrations of tmGBP alanine variants. All titrations are from the second set of three replicate experiments. Traces from high signal quality variants (top) and intermediate quality variants (middle) are representative. Those from un-interpretable variants* (bottom) are representative of the noise levels observed but the fit curves differed greatly across replicates. Normalization of raw signal was performed on minimum and maximum data points, and thus fit curves do not perfectly converge to one or zero.
Mapping the impact of each mutation on affinity to the crystal structure (Figure 9) reveals that the most detrimental mutations are found on the hinge side of the tmGBP pocket. The notable exception to this observation is R141A. Mutating R141 might be expected to cause significant loss of affinity because the residue is positioned to make hydrogen bonds to two glucose hydroxyls. However these two hydroxyls are also within hydrogen bonding distance of two other side chains each and therefore hydrogen bonding could remain satisfied even in the absence of R141. The general spatial pattern in loss of affinity is consistent with an increased importance of hinge residues in mediating the open to closed transition [116,117,118,119]. Alternatively, it could be that conjugation of IANBD at S131C perturbs protein ligand contacts on the far side of the pocket and that a portion of the energetic contributions made by the original residue are thereby lost. This hypothesis is supported by the fact that although IANBD-labeled tmGBP showed little difference in affinity for glucose or xylose, the affinity of the wild type protein for xylose has been reported at 13.5nM [120] – an affinity three orders of magnitude tighter than observed for the IANBD-labeled form. Testing these two hypotheses, however, is beyond the scope of this investigation which focused on validating the screening method for TnT level use.

In the protein design context, TnT-level hits from a panel of designed proteins would be cloned conventionally and expressed for more rigorous analysis and validation in subsequent steps. Therefore, at a minimum, a screening assay need only
separate binders above a certain affinity from non-binders. In this regards the assay described here drastically outperformed the requirements by enabling reproducible calculations of $K_d$ spanning several orders of magnitude.

![Figure 9: tmGBP binding site color coded by mutational impact. Residues that when mutated to alanine result in signaling loss or affinities greater than 20 mM are shown in blue, between 1-5 mM are shown in green, 100-500 μM are shown in orange, and 5-40 μM in red. The position of IANBD conjugation is shown space filled. The strands that comprise the hinge can be seen on the right side of the figure.]

In addition to data quality, speed, and ease; screening with coupled fluorophores has the advantage that if a successfully redesigned receptor is identified, a biosensor is simultaneously obtained. However, because mutations in the active site may lead to loss of signaling without loss of affinity, the possibility for false negatives cannot be ignored. For example, from only the data presented here, it is impossible to confidently conclude whether loss of binding was responsible for loss of signaling in D89A, Y190A, and
D216A, or whether the mutations simply perturbed the local structural rearrangements responsible for signaling. This concern is supported by the observation that each and every one of the 14 variants produced smaller signal changes than wild type, despite mutating only a single residue. There is no guarantee that larger scale binding pocket redesign will retain any signaling ability at all.

In part for this reason, more generalizable assays that simultaneously satisfied speed and sensitivity requirements were sought. One general strategy is to exploit the fundamental linkage relationship between ligand binding and protein stability. Two methods that rely on this principle were investigated. The SYPRO Orange™ thermal shift assay is covered in the remainder of this chapter and a variant of the Quantitative Cysteine Reactivity (QCR) assay is covered in Chapter 3. These assays are generalizable in the sense that they do not rely on prior knowledge, such as which particular cysteine/fluorophore pairs enables signaling, and that virtually all protein:ligand interactions produce a stability shift.

### 2.2.2 Miniaturized Thermal Shift Assay for Ligand Binding

Proteins are thermodynamically stabilized in the presence of their cognate ligands. To understand this thermodynamic linkage phenomenon, consider the fundamental requirement for ligand binding, that the free energy of protein:ligand complex be lower than that of the unbound protein plus free, solvated, ligand.
Formation of the low-energy complex relies in large part on the ordered placement and orientation of active site residues that form energetically favorable interactions with the ligand (e.g. hydrogen bonds, electrostatics, and van der Walls). Geometric arrangement of active site residues in turn relies on the structural conformations found in the folded ensemble, which are represented in part by crystal or NMR structures. Unfolded proteins lose the organization needed to form many simultaneous, energetically favorable, interactions with ligand. For this reason ligands that bind the folded protein generally do not bind the unfolded form or, if they do, bind it with reduced affinity. Preferential affinity for the folded state is the key requirement for stabilization. To explain how the effect is exerted, we turn to the linked equilibrium in the system.

Like all thermodynamic processes the transition between the folded and unfolded states is in constant dynamic equilibrium. Adding a cognate ligand, which binds the folded state, depletes the concentration of the apo folded form and draws the unfolding equilibrium towards the folded ensemble by Le Chatelier’s principle (Figure 10). Consequently a greater fraction of the protein population is found in the folded state (bound + free) and thermodynamic stabilization occurs concomitantly with the addition of ligand. Arising from this mechanism is one of the most non-intuitive and unappreciated aspects of ligand-induced protein stabilization – stabilization does not saturate when fractional ligand occupancy saturates. Increasing ligand concentration well beyond the $K_d$ will continue to further stabilize protein because it continues to exert
an effect on the linked equilibrium. The only way stabilization ceases to increase is if the $K_d$ for the unfolded state is also surpassed (a situation which can occur in practice but seldom does). Under these circumstances additional ligand exerts equal influence on each side of the unfolding equilibrium and so no net stability change is observed. Generally then, a greater degree of stabilization induced by a given concentration is reflective of a tighter affinity in the binding interaction.

![Diagram of equilibrium]

**Figure 10:** Linked equilibrium for ligand binding and protein unfolding. The unfolded but bound state (U:L) and the reactions to interconvert from it are shown in grey because in this simplified example the unfolded form has negligible affinity for ligand. Adding ligand draws the equilibrium to the lower left native bound state (N:L) and thus away from the unfolded state (U). The change in the proportion of native states (N & N:L) to the unfolded state under given conditions produces the net stabilization effect.

Experimentally, using stability to investigate binding requires an observable that changes upon unfolding and a means with which to induce unfolding. Typical observables include changes in the absorbance of circularly polarized light, fluorescence (extrinsic dyes or intrinsic residues), or the aggregation rate. Denaturation is typically induced with temperature or chemical agents, and the extent needed to reach the
unfolding midpoint serves as a quantifiable measure of stability. For high-throughput screening, temperature has the advantage of being able to monitor a single well that can be continuously scanned and finely sampled. In contrast, chemical denaturants require multiple wells at different denaturant concentrations or complex liquid handling titrations that challenge parallelization.

In 2001 Pantoliano et al. described an implementation of the stability shift assay adapted for 384-well format that used temperature as the denaturant and fluorescence of extrinsic dyes as the observable [121]. Dyes were selected for quantum yields that increased upon partitioning to the low-dielectric interior of the molten globule, present upon unfolding. Although the initial application was for high-throughput drug screening, since then this assay has adapted to identifying gene function [122], determining crystallization conditions [123,124], evaluating SNP impact [125], and protein engineering [126]. Amongst the dyes used for the thermal shift assay, SYPRO Orange™ has emerged as a favorite due to its commercial availability, low cost, high sensitivity, and broad applicability to different proteins. We sought to test whether the SYPRO Orange thermal shift assay satisfied the criteria for inclusion in the protein engineering pipeline. Towards these ends a receptor mutagenesis and in vitro expression study was performed and screened with SYPRO.

E. coli Ribose binding protein (ecRBP) was selected as the model and alanine mutants selected for comparison with a published mutagenesis study by the Dattelbaum
laboratory. The Dattelbaum study used a conjugated fluorophore technique similar to that described in the previous section with tmGBP. In this ecRBP study, mutations were made in two different backgrounds that differed from wild type in that they contained a single buried cysteine at one of two positions (L62 or A188) for further cross validation with QCR. Due to the limitations of available instrumentation, only three ligand concentrations were selected for screening: apo, 100µM ribose, and 1mM ribose. Proteins were expressed \textit{in vitro}, purified using affinity resin, and eluted with FLAG peptide. SYPRO fluorescence was read in the presence of protein while heating samples from 28°C to 75°C (Figure 11). \( \Delta T_m \) values were calculated for each ligand condition by comparing the temperatures at which the maximal rate of fluorescence increase occurs (Figure 12). Across all mutations, variants in the A188C background were less stabilized by ribose than those from the L62C background indicating either a loss of affinity by the 188C mutation itself or a change in enthalpy. Enthalpy of unfolding determines the magnitude change in \( T_m \) for a given \( \Delta\Delta G \).
Figure 11: Representative SYPRO melts of ecRBP variants. Raw fluorescence is plotted in the left column and the normalized derivative is plotted in the right column. Maximal rate of change was used to approximate the $T_m$. The 188C background produces large thermal shifts in the presence of ribose (top). The Q235A mutant is less stabilized by ribose indicating reduced affinity (middle). The D89A shows no stabilization indicating loss of affinity (bottom).
Figure 12: Thermal shifts for each ecRBP alanine variant at two ribose concentrations. Note that when a thermal shift is observed, 1mM ligand always produces a greater thermal shift than 0.1 mM Ribose, on the order of an additional 5°C. Furthermore, the L62C background is always shifted more than the 188C background.

Thermal shift results were compared with affinities determined by the Dattelbaum group. Comparison of the two methods should be done in consideration of the fact that the proteins themselves are not identical, namely they carry different cysteine mutations and there is a conjugated fluorophore in Dattelbaum case. Nevertheless the results of the two trials are quite consistent with one another (Figure 13). For simplicity of comparison, the four $T_m$ values for each variant shown in Figure 12 were averaged. Of the five mutations that cause total loss of affinity in the Dattelbaum study (F16A, D89A, F164A, F214A, and D215A), four show no thermal shift and the fifth (F214A) has the smallest thermal shift detected. Similarly, the only mutant that matched
wild type affinity in the Dattelbaum study, S103A (3.4 vs. 3.9 μM) also showed the greatest thermal shift in our study (average of 12.3 vs. 12.4 °C). Rank order of intermediate affinity mutants does not perfectly match rank order of stabilization, but is not unreasonable given the experimental error and different backgrounds. F15A does show significant disagreement between the assays, but whether this disagreement arises purely from the different backgrounds or is the result of some error is unknown.

Figure 13: SYPRO thermal shift assay vs. affinities reported in the literature. Variants for which no binding was reported in the Dattelbaum study were plotted at an arbitrarily high value of 5mM (●). F15A, for which the methods produce conflicting results, is shown in red (●). All other variants are shown cross hatched (X).

We concluded that the SYPRO assay met the criteria in inclusion in the protein engineering pipeline. In comparison to a simple binding assay, the thermal shift assay
has the added advantage of also determining stability of the unliganded variants. This is critical in the protein (re)design realm as the majority of mutations made to wild type proteins are destabilizing, and stability is a critical factor contributing to long term function. Instrumentation that allowed greater resolution, higher accuracy, faster reads, and increased throughput was obtained in the form of a Roche LightCyler RT-PCR machine. The power of this instrument was demonstrated in a study that sought to explore the binding specificity of ribose analogs.

Ribose analogs were selected for their similarity to ribose regarding functional groups and stereochemistry, as well as commercial availability (Figure 14, top). Thermal shifts were recorded across many ligand concentrations to produce high density data sets (Figure 14, bottom). In principle, these higher resolution data sets can be fit to solve for binding affinity and enthalpy [127]. However, it was later determined that the slow denaturation kinetics of ecRBP results in an overestimated $T_m$ that does not represent the equilibrium $T_m$. Nevertheless, the results are still valuable from a comparative standpoint, and beautifully demonstrate how stability does not saturate with increasing ligand.
Figure 14: Stability shifts produced by ribose analogs. The majority of the analogs alter the functional group or stereochemistry at only a single ring position (red numbers) and are shown in columns arranged accordingly (top). Where there is more than one change, the change to the ring hydroxyl is used for classification. Data points represent the average of quadruplicate reads and error bars are hidden because standard deviations were on the order of the marker size.
Of the analogs, the amine substituted sugar β-D-ribosyranosyl amine appeared to bind ecRBP with the highest, almost wildtype, affinity. Importantly, although ribose can interconvert between anomic forms as well as between pyranose and furanose rings, it is the β-D-pyranose form that is most energetically favored in solution [128] as well as the form found in the crystal structure [129]. Therefore β-D-Ribopyanosyl amine features the appropriate ring structure and substitutes only an amine for a hydroxyl which is capable of forming similar hydrogen bonding. D-Allose, the ligand with the next greatest affinity adds a CH₂OH group, but this additional group is directed towards the solvent side of the pocket. Methyl beta-D-ribopyranoside, and L-Lyxose also show binding, albeit with noticeably decreased affinity. All the sugars appeared to induce mild stabilization at ~100mM. Given the high concentration and ubiquity of the effect we suspect it is the result of non-specific interactions.

Although outside the scope of this study, ecRBP and the collection of moderate affinity sugars represent a powerful model system for testing incremental computational design hypotheses, as well as for probing the adaptive plasticity of the binding pocket. Unlike radical redesign against ligands for which the starting scaffolds show no affinity, ligands with pre-existing, detectable, affinity have that advantage that negative results become informative. That is to say, design variants that show a loss of affinity represent an observable departure from a pre-existing low energy solution. This is in stark contrast to radical redesign where the inability to produce binders reflects an un-interpretable
move in energetic space above the threshold of detection. Furthermore, the modest changes to ligand structure would restrict design to a particular region of the binding pocket and limit the number of residue changes. Combined with the high-throughput capability of producing and testing variants described in this chapter a greater fraction of design space can be covered.

2.2.3 SYPRO Orange Assay for Screening Cloned Genes

The addition of small scale functional assays to the gene synthesis and cell-free expression technologies covered earlier formally enables the full deployment of the high-throughput protein engineering pipeline. Despite their important role, these assays are not meant to be the final step in design characterization. On the contrary, each represents only a launching point, but one that enables rapidly sorting positive hits from a larger pool of nonfunctional designs. The fact that these assays also provide accurate estimations of affinity does not abrogate the need for confirmation and more rigorous study. Further characterization would likely turn to established and orthogonal biophysical assays. In the case of receptor design, assays could include isothermal titration calorimetry and equilibrium dialysis with radioligand. However in addition to being lower throughput, these typically require increased amounts of material, which can only be provided by cloning.
Cloning synthetic genes can be problematic. During the oligonucleotide production phase, \( n-1 \) deletions result from imperfect efficiency at each round of extension. Missing bases in oligos subsequently introduce frameshift mutations in the final gene product. Mathematically, even with an extremely high coupling efficiency, lengthy oligos will contain a meaningful fraction of errors. For example 99.8% efficiency at each coupling step in an 88-mer synthesis produces a population of oligos, a full 16% of which are incorrect. In principle, at this error rate, assembling the eighteen oligos of this length required to make the ecRBP expression cassette predicts only a 4% likelihood of error free product. This calculation assumes that both correct and incorrect oligo are equivalently likely to be incorporated. In reality, intrinsic proofreading that occurs because of selection against mutations in annealing regions increases the final accuracy rate. Nevertheless, it is not surprising for cloned product to contain a high frequency of plasmids with at least one error, and high error rates been repeatedly observed by Hellinga lab members (although rates are also variable across syntheses and genes). A high error rate requires that an increased number of colonies be picked and sequenced to ensure the correct clone is obtained, thus raising costs and causing delays.

Given the degree of error in the linear DNA pool, it is perhaps surprising that consistent and reasonable functional results are possible from TnT expression, as demonstrated in the previous two sections. The paradox is resolved by the nature of the \( n-1 \) mutation and the purification protocol. \( n-1 \) mutations cause frameshifts which result
in early stop codons or out of frame purification tags at the C terminus. Consequently, purification serves as a filter against mutants, with the obvious exceptions that neither point mutants nor insertions and deletions in multiples of three are prevented. This filter is almost certainly at play when C-terminal tags are used to purify TnT-expressed protein made directly from synthesized linear DNA. We sought to exploit this feature for the purpose of using micro-purifications of clones to rapidly and economically determine those clones not worth sequencing. Expression was verified using either gel visualization or a SYPRO Orange melt. To streamline the process and prevent the need for multiple transformations, a dual cloning and T7 expression stain (KRX, Promega) was employed according to the protocol in Figure 15.
Figure 15: Flow chart for streamlined cloning, sequence verification, and expression. Cultures are split three ways at the end of Day 1 corresponding to growth blocks for in-frame screening, minipreps, and long term glycerol storage. To prevent wasted effort and costs, only variants that pass expression screening proceed to miniprep and sequencing, and only those that are sequence validated go on to large scale expression. Note that the rapidity of SYRPO screening means minipreps are delayed less than a day.

As a test set, we used 27 ribose binding protein variants designed to contain combinations of consensus mutations from a group of eight highly homologous ribose
binding proteins found in the genomes of thermophilic organisms. A first round of gel visualization-based screening was performed on three clones each of the 27 variants for a total of 81 samples. Insert check PCR eliminated 19 clones from consideration as empty vectors. The remaining 62 were purified on a small scale and run on polyacrylamide gels (Figure 16). Of these, 38 showed production of in-frame protein, and a maximum of two clones per variant submitted to sequencing for a total of 31 samples. Sequencing established that 14 samples were correct representing 11 of the 27 desired variants.

![Figure 16: SDS-PAGE screening of full length in-frame genes. Clones that expressed protein with an in-frame HIS tag were easily distinguishable from those that did not, and have their labels boxed. No bands were observed outside the regions shown on any of the three gels shown here.](image)

---

1 Inspired by the amazing degree of structural homology between ecRBP and ribose binding protein from *Thermoanaerobacter tengcongensis* (tteRBP), we hypothesized we could use this sequence information to build PBFs of intermediate stability between ecRBP and tteRBP. These might have the advantages of increased tolerance to potentially destabilizing mutations introduced during design. However, they would not rely on destabilizing mutations to bring the $T_m$ of tteRBP below 100°C, and into the range observable by SYPRO. The exact nature of these variants is not critical to their role in testing expression screening of in-frame clones, and for that reason is not presented in detail.
A second round of screening, this time applying a SYPRO method, was performed on the remaining 16 variants. At the most basic level, like gel bands, SYPRO melts can determine the presence or absence of protein by the presence or absence of signal. However, SYPRO melts are more information rich than gel bands because they can provide stability and ligand binding information. Furthermore, this information is potentially valuable from the standpoint of predicting which clones are mutants and not worth sequencing – the goal of this experiment. First, consider that random mutations are predominantly destabilizing and detrimental towards function. Therefore, when prioritizing between clones of a given variant, the more stable, higher affinity variants are more likely to be correct and should be sequenced first. Second, random mutations are expected to exert dissimilar effects, whereas all clones with the correct sequence will show the same characteristics. Therefore, if amongst a set of SYPRO melts for a given variant, the profiles of several clones are identical but there are individual outliers, it is less likely that the correct sequence be found amongst the outliers.

For the remaining 16 variants, six colonies apiece were picked to fill a 96 well block. Of these, 40 showed in-frame protein expression by SYPRO signal, a portion of which were further eliminated on the basis of the stability and binding criteria previously mentioned (Figure 17). Of 32 clones sequenced, 19 coded for the correct sequence, and 14 of the 16 remaining variants were obtained. The final two variants were subjected to a third round of screening, this time 32 colonies each were tested by
SYPRO. In one case 6 of 32 colonies showed expression and in the other case 15 of 32 did.

Figure 17: Discriminating probable mutants on the basis of SYPRO profiles. Shown here are four examples from the 16 variants screened. Each clone is shown with four melt traces, two in the absence of ligand (light shades) and two in the presence of 1mM ribose (dark shades of the same color). A) All but the red and green clones showed only baseline signal, indicating none of these clones expressed in-frame protein. B) Red, blue, and purple traces show a consistent profile, indicating the orange traces are unlikely to be from a correct clone. C) The blue and green clones respond to ligand, but the purple does not, indicating it is less likely to be from a correct clone. D) The orange colony is greater stability and has a lower baseline than the blue, making it more likely to be wild type. In this case both colonies were sequenced and the orange clone was correct while the blue clone did, in fact, contain an unexpected M194I mutation.
Compared to gel visualization, SYPRO is the faster of the two methods. It can be completed within 30 minutes of purification vs. several hours for loading, running, staining, and imaging a gel. SYPRO is also the cheaper of the two methods. Consumables costs for SYPRO total ~$0.50/sample, whereas using precast polyacrylamide gels, ladder, and stain raises costs to ~$1.00/sample. Both are still more economical than minipreps, which cost over $2/sample. The majority of the savings are realized by preventing unnecessary sequencing, which runs $5/read and thus $10/clone for genes requiring forwards and reverse reads. In this particular experiment, of the errors discovered, over 75% were found at the screening stage rather than at the sequencing stage. Coupled with the low rate of correct sequences in cloned product, eliminating most of the mutants prior to minipreps and sequencing represents a substantial savings and reduces the demand on sequencing infrastructure.
2.3 Discussion

Combining the rapid gene design, assembly, and cell-free testing shown in Figure 7 with the cloning, screening, sequencing and expression protocol shown in Figure 15 results in a ten-day process. If we allow a couple days upstream for *in silico* calculation and another couple downstream for orthogonal functional assays, then the total time from beginning design though completing conventional-scale validation totals approximately two weeks. Notably, this timeframe meets the “within weeks” target set by DARPA’s Biological Warfare Defense unit for the development of sensors against unanticipated threats [130]. Of course, success demands that binders be present within the set of top ranking computational designs selected for experimental testing. But, computation need not be perfect. Within the schema presented, it is entirely reasonable for a single individual to build and screen 96 designs and focus on cloning the top 5-10 hits in this two-week timeframe. From a basic research standpoint, the accelerated timeline and increased throughput meet a critical need. They provides a means to close the historically large time gap between computational design hypotheses and experimental testing while providing a more statistically significant number of data points from which to draw conclusions. This is particularly true in the case of design against ligands with weak pre-existing affinity where every design, binder or not, is informative.
In addition to the assays documented in this chapter, other Hellinga lab members have, in parallel, developed a ligand binding assay capable of exquisite sensitivity called quantitative cysteine reactivity (QCR) [131]. While the original implementation suffered from expensive reagents and limited throughput, the potential for high precision measurements of protein thermodynamics and kinetics was apparent. I became involved in an effort to continue QCR development to decrease costs and increase throughput. The results of that collaboration with Dr. Daniel Isom are a joint effort and described in the following chapter.
2.4 Materials and Methods

2.4.1 tmGBP Expression and Fluorophore Conjugation

Synthetic linear DNA for each variant was re-amplified using biotinylated end primers and KOD polymerase (EMD, #71086). TnT reactions were prepared according to standard lab protocols [112] on a 350 µL scale with the addition of 8µg DNA. Reactions were incubated in 2mL eppendorf tubes for 8 hours at 30°C and 350 rpm, after which they were stored for several hours at 4°C until purification. Initial centrifugation (10 minutes, 13,000g) separated out insoluble protein. TnT supernatant containing soluble protein was added to 500µL of a 50:50 mixture of binding buffer (15 mM imidazole, 20 mM MOPS, 500 mM NaCl, pH 7.5) and TBS (20 mM Tris, 136 mM NaCl, pH 7.4), to which a 50µL bed volume of washed EZView HIS beads (Sigma, E3528) was added. After one hour binding incubation at 4°C, samples were spun at 8.2 G, supernatant disposed of, and labeling initiated by re-suspending in 1mL 50:50 bind:TBS to which IANBD-amide was added at a final concentration of 0.5mM. IANBD-amide stock solutions had been previously prepared at 100mM in DMSO and stored at -20°C. Labeling proceeded for 30 minutes at 37°C, after which five 1mL washes of 50:50 bind:TBS were used to further purify the protein and remove unreacted dye. Protein was eluted with 200µL Elution buffer (400 mM imidazole, 20 mM MOPS, 500 mM NaCl, pH 7.5) and stored on ice.
2.4.2 tmGBP Ligand Titrations

Titrations were performed in a 384 well flat bottom microplate (Corning, #3821). In each well 10 µL eluted protein was mixed with 30 µL of glucose solution in TBS, and 16 concentrations ranging from nM to mM were tested. Previous control experiments established that the presence of residual imidazole did not impact affinity. BSA (NEB) was supplemented at a concentration of 0.1 mg/mL. Fluorescence was read at 30°C in a FluoDia T70 (Photon Technology International) using a 475nm excitation filter and a 530nm emission filter. Data were fit in Origin software using a Hill cooperative binding equation where \( \theta \) is fractional occupancy, \( F_o \) is background fluorescence, \( F_{\text{max}} \) is maximal fluorescence, \([L]\) is glucose concentration, and \( K_d \) is the dissociation constant.

\[
\theta = F_o + (F_{\text{max}} - F_o) \left( \frac{[L]^n}{K_d + [L]^n} \right)
\]  

[1]

2.4.3 SYPRO Thermal Shift Assays

Alanine scan SYPRO assays were performed in the FluoDia T70 (Photon Technology International), whereas ribose analog and cloned gene screening SYPRO assays were performed in a LightCycler 480 II (Roche). Excitation and emission filters were 465/590 and 465/580 for the FluoDia and LightCycler respectively. Heating rate in the FluoDia was 1°C/min whereas in the LightCycler it was increased to 3.6°C/min. In both cases \( T_m \) was called as the maximal value of the smoothed derivative. SYPRO orange dye was obtained from Invitrogen (S-6650) and used at a final concentration of
20X§. Well volumes were 100µL in the FluoDia and covered with PDMS oil (Sigma, #378321), and 20 µL in the LightCycler and covered with optically clear sealing film (Roche, 04729757001). D-Ribose, D-Xylose, L-Rhamnose, and D-Psicose were obtained from Sigma. β-D-Ribopyranosyl amine, D-Allose, Methyl-β-D-ribopyranose, and 2 deoxy ribose were obtained from Carbosynth. L Lyxose and D Arabinose were obtained from Acros Organics.

2.4.3 Gel Based Screening

10 uL purified protein per lane was run on 26 well precast 4-12% Bis-tris midi gels (Invitrogen, WG1403) at 180V. Benchmark pre-stained protein ladder (Invitrogen, 10748) was used a molecular size marker. Gels were stained using GelCode colloidal Coomassie stain (Thermo, 24592) according to manufacturer instructions. Gels were imaged on an Alpha Innotech MultiImage II, and scored for expression by visual inspection.

§ The true chemical identity of the SYPRO dye is not divulged by the manufacturer nor is the concentration or extinction coefficient provided. Therefore concentrations are reported relative to the 1X concentration recommended by the manufacturer for use as a gel stain.
3. Quantitative Cysteine Reactivity using ABD

3.1 Introduction

The genomic revolution has greatly influenced the scope and scale of quantitative experiments in biology [19]. High-throughput genomic and proteomic initiatives have explored transcription [132], translation [133], post-translational modification [134,135,136,137,138], and protein degradation [139,140] and turnover [141] in response to changes in cellular physiology, and have expanded our knowledge of protein structure [142,143,144]. Less emphasis has been placed on developing methods that use small quantities of material to rapidly quantify protein thermodynamics and folding kinetics. Such advancements are essential to furthering our understanding of the relationship between protein sequence, structure, stability and function that underlies the molecular basis of disease and evolution. Here we present a miniaturized thermodynamic technique, fast determination of quantitative cysteine reactivity (fQCR), that measures protein stability (ΔGU), thermal stability (Tm), ligand-binding affinity (Kd), and estimates (un)folding kinetics. By exploiting the fundamental linkage between protein thermodynamics and function, fQCR experiments can be used to identify and assess many aspects of biological function that affect protein conformational stability. These functions include the strength of biomolecular interactions [145,146,147,148], the impact of deleterious mutations[149,150], and the effects of post-translational modifications [151,152,153,154,155].
Traditional approaches for measuring protein stability (e.g., temperature and chemical denaturation) and folding kinetics are low-throughput (requiring > 100 μg of protein) and time-consuming (taking hours to days). By contrast, the emergence of miniaturized techniques for determining $T_m$ [121,131,156], $\Delta G_{ii}$ [131,157,158,159], and $K_d$ [121,127,158,160] are beginning to enable the thermodynamic characterization of proteins using less than a few micrograms of purified material. One such technique, quantitative cysteine reactivity (QCR) [131], uses cysteine protection to measure protein stability in a manner analogous to amide protection experiments [161]. Cysteine sidechains protected by burial in a folded protein structure are inaccessible to solvent and cannot be modified by a thiol-reactive probe. However, these residues can be modified upon exposure to solvent by transient unfolding reactions. Consequently, the kinetics of cysteine modification are coupled to the thermodynamic stability and (un)folding kinetics of the protein.

Here we present a development of the QCR technique that enables rapid determination of cysteine modification using fluorescence: fast determination of quantitative cysteine reactivity (fQCR). The fQCR approach was validated using variants of Staphylococcal nuclease (SN) and E. coli ribose-binding protein (ecRBP) engineered to contain single, buried cysteines. In fQCR experiments, the kinetics of cysteine labeling are determined using the fluorogenic thiol-probe, 4-(aminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole (ABD). The fluorescence readout facilitates the rapid
quantification of the degree of fractional cysteine labeling as a function of temperature. From these observations ΔGU, Tm, and, in some cases, folding kinetics can be determined. Additionally, fQCR experiments can be used to quantify the affinity of biomolecular interactions by measuring changes in ΔGU in the presence of a binding partner. This approach was demonstrated for SN and ecRBP by determining the binding affinities of these proteins for inhibitor and ribose respectively. Unlike traditional methods for measuring protein stability, folding kinetics, and ligand binding affinity, fQCR data can be acquired on non-specialized equipment within minutes, requires only micrograms of protein, and is not restricted to proteins that exhibit ideal two- or three-state unfolding behavior.
3.2 Results

3.2.1 Measuring Protein Stability by QCR.

Cysteine sidechains are attractive targets for site-directed chemistry in proteins given the wide variety of commercially available compounds that preferentially label thiols (e.g., methanethiosulfonate (MTS), 5,5′-dithiobis(2-nitrobenzoate) (DTNB), iodoacetamide (IAM) and ABD). These reagents are most frequently used to covalently modify cysteine sidechains to introduce chromophores or fluorophores for determining protein concentration [162,163], measuring binding affinity [164,165], and assessing sidechain accessibility [166,167]. Such reagents can also be employed as site-specific probes of protein stability [168,169,170] and folding kinetics [171,172], but rarely are. Cysteine reactivity to characterize protein thermodynamics and function has been underutilized.

In the QCR method, protein stability is quantified by exploiting the relationship between conformational free energy and the reactivity of a buried cysteine [131]. In a folded protein (Closed-SH), a buried cysteine is concealed by protein structure and is inaccessible to bulk solvent. To be covalently modified by a thiol-reactive probe (P), the cysteine must become exposed to solvent (Open-SH) in an unfolding reaction

$$\text{Closed-SH} \xrightleftharpoons[k_{\text{open}}]{k_{\text{close}}} \text{Open-SH} \rightarrow \text{Open-S-P} \quad [2]$$
where $k_{\text{open}}$ and $k_{\text{close}}$ are the rate constants for opening (unfolding) and closing (folding), and $k_{\text{int}}$ is the rate constant for cysteine modification in the unprotected (open) state. Under the appropriate conditions (see below), the observed rate constant for labeling a buried cysteine ($k_{\text{label}}$) can be expressed as [173]

$$k_{\text{label}} = \frac{k_{\text{open}} k_{\text{int}}}{(k_{\text{open}} + k_{\text{close}} + k_{\text{int}})}$$ \[3\]

If $k_{\text{close}} \gg k_{\text{int}}$, cysteine modification is said to proceed under EX2 conditions [173] and $k_{\text{int}}$ can be eliminated from the denominator of Eq. 3, thereby directly relating observed cysteine reactivity with protein thermodynamic stability

$$k_{\text{label}} \equiv {\text{EX2}} \frac{k_{\text{open}} k_{\text{int}}}{(k_{\text{open}} + k_{\text{close}})} = \frac{k_{\text{int}}}{(1 + e^{\Delta G_U/RT})}$$ \[4\]

where $\Delta G_U$ is the conformational free energy of unfolding, $\Delta G_U = RT \ln(k_{\text{open}}/k_{\text{close}})$. Rearrangement of Eq. 4 yields

$$\Delta G_U = RT \ln \left( \frac{k_{\text{int}} - k_{\text{label}}}{k_{\text{label}}} \right)$$ \[5\]
which is used to derive $\Delta G_u$ from the labeling of single, buried cysteines in QCR-based experiments.

### 3.2.2 The Mechanism of Cysteine Labeling.

A protected cysteine can be exposed to bulk solvent by transient local or global unfolding events. Deeply buried sites are usually highly protected and only label as a consequence of global unfolding. Sites that are in proximity to the solvent-accessible surface of the protein can be labeled as a consequence of either local or global unfolding, depending on their structural context and experimental conditions. In the QCR approach, the effects of global unfolding are measured by manipulating the global thermodynamic stability to achieve a range of 99% to 1% native protein, typically by altering temperature. Under these conditions, which we refer to as the global unfolding window of observation (GUWO)[131], global unfolding is the dominant mechanism for the exposure of protected cysteines, and measured values for $k_{\text{label}}$ can be analyzed using Eq. 5 to give global protein thermodynamic stability. Outside of the GUWO, because the labeling of a buried cysteine can result from a variety of conformational fluctuations (i.e., transient local, subglobal or global unfolding events), $\Delta G_u$ cannot be guaranteed to exclusively reflect the free energy of global unfolding.

In a typical fQCR experiment, the GUWO comprises a 20-30 degree range of temperatures that span $T_m$. Under these conditions, $\Delta G_u$ is highly dependent on temperature and the equilibrium fraction of folded and unfolded protein is variable.
Consequently, the observed rate of cysteine labeling can be affected by kinetics of conformational equilibration \( k_{\text{conf}} = -(k_{\text{open}} + k_{\text{close}}) \) upon exposure to the temperature gradient. In cases where \( k_{\text{conf}} \) and \( k_{\text{close}} \) are at least an order of magnitude faster than \( k_{\text{int}} \) (EX2 conditions), the fractional labeling of a protected cysteine is described by a single kinetic phase \[174\]

\[
f_{\text{label}}(T) = 1 - e^{-k_{\text{label}}(T)t}
\]  

[6]

at a given temperature \( T \). Such steady-state conditions break down when the magnitude of \( k_{\text{int}} \) approaches \( k_{\text{conf}} \) and \( k_{\text{close}} \). In such cases, fractional cysteine labeling is described by two kinetic phases

\[
f_{\text{label,PSS}}(T) = 1 - A_f e^{-k_f(T)t} - A_s e^{-k_s(T)t}
\]  

[7]

where the fast (amplitude \( A_f \), rate constant \( k_f \)) and slow (amplitude \( A_s \), rate constant \( k_s \)) phases describe the kinetics of cysteine labeling in the unfolded and folded fraction of protein established at time \( t \) \[173,174\]. Equation 7 makes no assumptions regarding the kinetics of conformational equilibration, nor the relative magnitudes of

\* All SI reference refer to Mathematica Notebooks containing simulation code can be found hosted online at http://dezymerwiki.biochem.duke.edu/wiki/index.php/Simulations
$k_{\text{close}}$, $k_{\text{open}}$, and $k_{\text{int}}$. It does, however, assume that labeling of protected cysteines in the closed state is negligible.

Even though the possibility of multiple kinetic phases makes for a relatively complex analysis, $k_{\text{label}}$ can always be extracted under EX2 conditions. To assist in the understanding of this process and to help identify appropriate experimental conditions (temperatures and label concentrations), we have developed some simulation tools (SI-2). By simulating Eq. 7, combinations of $k_{\text{close}}$, $k_{\text{open}}$ and $k_{\text{int}}$ can be identified that eliminate the fast phase of the reaction. These simulations reveal that the fast phase is negligible under conditions where $k_{\text{int}}$ is an order of magnitude less than $k_{\text{close}}$ (i.e., traditional EX2 conditions), regardless of the relative values of $k_{\text{close}}$ and $k_{\text{open}}$. This generalization holds true even if the protein is 99% unfolded at equilibrium and $k_{\text{open}}$ is $>100$ times $k_{\text{close}}$. Furthermore, these simulations demonstrate that QCR-based experiments can be designed to measure $\Delta G_U$ at temperatures beyond $T_m$, providing EX2 conditions persist.

In the QCR method, the use of sidechain protection to measure protein stability at temperatures that promote unfolding is quite different from methods such as hydrogen-deuterium exchange (HDX) that exploit the differential reactivity of protected groups under native conditions (i.e. $>99\%$ folded, $k_{\text{close}} > 100 k_{\text{open}}$). HDX cannot be implemented at temperatures that promote unfolding because the long timescale and high protein concentration lead to protein aggregation [175,176]. QCR and HDX-like
experiments can therefore be regarded as complementary ways for determining ΔGU under quite different experimental conditions.

### 3.2.3 Determinants of $k_{int}$

The reactivity of an unprotected cysteine, $k_{int}$, can be affected by a number of factors including the cysteine pK$_a$ value in the unfolded state, solution pH, temperature, and probe concentration. To investigate the effects of pH and temperature on $k_{int}$, we used the reaction between ABD and reduced glutathione (L-GSH), a tripeptide that serves as a model for the reaction of an unprotected cysteine (Figure 18).

The negatively charged form of a cysteine sidechain (thiolate) is the nucleophile that reacts with thiol-specific probes[162]. The second-order rate constant for labeling an unprotected cysteine ($k'$) is therefore pH dependent (Figure 18B):

$$
\ln k' = \ln k'_{ref} - \ln \left(1 + 10^{pK_a - pH}\right)
$$

[8]

where $k'_{ref}$ is the pH-independent second-order rate constant measured over a pH range where cysteine is present in >99% thiolate form (in this case of L-GSH at pH values > 9).

The pK$_a$ values of unprotected cysteines can also be influenced by residual electrostatic interactions in the open (unfolded) state of proteins, and therefore may vary by location and protein. The pK$_a$ values of solvent-exposed ionizable sidechains are

78
rarely perturbed from their model compound values by more than one pH unit [177];
the pKₐ values of most unprotected cysteines are expected to be within 8 ± 1. Any
significant variation in the pKₐ of a cysteine in the unprotected state is observed directly
in the fQCR experiment, and may report on residual structure and electrostatic
interactions in the unfolded state [178,179].

The temperature dependence of $k'$ is described by an Arrhenius relationship

$$\ln k' = \ln A + \frac{E_a}{R} \cdot \frac{1}{T}$$  \hspace{1cm} [9]

where $A$ is a pre-exponential factor and $E_a$ is the activation energy for the
reaction. Analysis of the temperature dependence of the reaction between L-GSH and
ABD (Figure 18C) shows that $\ln A$ (27.3 ± 0.3) is independent of both pH and
temperature, and consequently remains fixed in the analysis of the temperature
dependence of $k_{int}$ (Eq. 16). However, $E_a$ is dependent upon pH (Figure 18C inset) and
depends on the pKₐ of the specific cysteine under consideration. From this analysis, a
general expression for $k_{int}$ at a given pH, temperature and probe concentration ([P]) can
be derived

$$k_{int}(T) = [P]k'(T) = [P]e^{(27.3+E_a/RT)}$$  \hspace{1cm} [10]
Figure 18: pH and temperature dependence of the reaction of ABD with L-GSH. (A) Representative progress curves for the pseudo-first order reaction of 10 mM L-GSH with 2.5 mM ABD at pH 8.0 and (●) 20°C, (○) 30°C, (△) 40°C and (▪) 50°C. Solid lines represent a fit of a first-order exponential function to obtain $k_{int}$. Such progress curves were used to obtain the rate constants in B and C. (B) pH dependence of the second-order rate constants, $k' = k_{int}/[ABD]$, for the reaction of ABD with L-GSH at (●) 20°C, (○) 30°C, (△) 40°C and (▪) 50°C. Solid lines represent a fit of Eq. 8 to derive the apparent pKₐ of L-GSH as a function of temperature: (●, 8.03 ± 0.01), (○, 7.9 ± 0.1), (△, 7.88 ± 0.08), (▪, 7.75 ± 0.08). (C) Temperature dependence of $k'$ at pH (●) 6, (○) 7, (△) 8 and (▪) 8.5. Solid lines represent a fit of the Arrhenius relationship (Eq. 10) to obtain $E_a/R$ as a function of pH: (●, -9300 ± 200 K), (○, -8300 ± 200 K), (△, -7800 ± 200 K), (▪, -7400 ± 200 K). \( \ln A \) is independent of pH and temperature: (●, 28.4 ± 0.8), (○, 27.1 ± 0.9), (△, 27.3 ± 0.5), (▪, 26.2 ± 0.8). The average value and standard error of \( \ln A \) (27.3 ± 0.3) is the same for any cysteine and was used in all fQCR simulations and data analysis.
3.2.4 Measuring Protein Stability Using fQCR.

In the original implementation of the QCR experiment, \( k_{\text{label}} \) was derived for protected cysteines from full kinetic traces at a given temperature [131]. The experimental throughput of fQCR experiments is dramatically increased by determining \( k_{\text{label}} \) from single timed reaction endpoints multiplexed across a gradient of temperatures (“temperature slice”) and probe concentrations (“probe slice”). In a temperature slice, \( i \) parallel reactions are simultaneously initiated and incubated at different temperatures \( T_i \) in a gradient thermocycler, and quenched by rapid cooling or addition of acid after an incubation time, \( \Delta t \). The observed temperature dependence of cysteine labeling is an indirect measure of the temperature dependence of protein stability (Figure 19 and SI-3), and can be analyzed in terms of the Gibbs-Helmholtz relationship

\[
\Delta G_{\text{U},i} = \Delta H_m \left( 1 - \frac{T_i}{T_m} \right) - \Delta C_p \left( T_m - T_i \right) + T_i \ln \frac{T_i}{T_m} \]  

[11]

where \( \Delta H_m \) is the enthalpy of unfolding, \( \Delta C_p \) the change in heat capacity of unfolding, \( T_m \) the midpoint of thermal denaturation, and \( \Delta G_{\text{U},i} \) is obtained experimentally from Eq. 5 as

\[
\Delta G_{\text{U},i} = RT_i \ln \left( \frac{k_{\text{int},i} - k_{\text{label},i}}{k_{\text{label},i}} \right) \]  

[12]
with $k_{\text{label},i}$ given by

$$k_{\text{label},i} = \frac{-\ln(1 - f_i)}{\Delta t} \quad [13]$$

and $k_{\text{int},i}$ given by

$$k_{\text{int},i} = [P]k'_i = [P]e^{(27.3 + E_a / RT_i)} \quad [14]$$

where $f_i$ is the fractional cysteine labeling observed at $\Delta t$ and $T_i$, $[P]$ is probe concentration, $E_a$ is the activation energy of the labeling reaction, and $k'_i$ is the second-order rate constant for the labeling reaction at $T_i$ (see Eq. 10). Combining 11-14 we obtain

$$f_i = 1 - e^{\left(1 + \frac{k_{\text{int},i} \Delta t / RT_i}{e^{(27.3 + E_a / RT_i)}}\right)} \quad [15]$$

which links the temperature dependence of fractional cysteine labeling at a set of constant endpoints to protein stability (Figure 19). Rewriting Eq. 15 in terms of relative fluorescence measurements observed in an instrument gives

$$F_i = F_{\text{max}}f_i \quad [16]$$

where $F_{\text{max}}$ is the fluorescence of a fully labeled sample.
The principal idea behind the fQCR approach is that the temperature dependence of $\Delta G_U$ can be determined solely from experimentally measured values of $f_i$ and $k_{int,i}$.

Without further analysis, conformational free energies determined in this way quantify the temperature dependence of protein stability and can be used to measure ligand binding affinities. However, analysis of a single fQCR dataset by a least-squares fit of Eq. 16 requires the Gibbs-Helmholtz relationship (Eq. 11) to model the temperature dependence of protein stability. Robust determination of the thermodynamic parameters $\Delta H_m$ and $T_m$ require an accurate value of $\Delta C_p$, which is poorly determined by fQCR data collected over a 20-30 degree temperature interval that spans $T_m$[180,181]. To circumvent this issue, we fixed the value of $\Delta C_p$ at an average value of 3, based on known values of $\Delta C_p$ from structure-based calculations [182] and experiments [183,184,185]. This procedure has no significant impact on the interpretation of the temperature dependence of $\Delta G_U$ with the GUWO.
Figure 19: The relationships between fractional labeling and protein stability. (A) Simulation of the Gibbs-Helmholtz relationship (Eq. 11) using $C_p$, $T_m$ and $H_m$ of 3 kcal mol$^{-1}$ K$^{-1}$, 329 K (56°C) and 105 kcal/mol respectively. Colors in panels A, B and C indicate a set of temperatures within a typical GUWO: (●) 60°C, (●) 57°C, (●) 55°C, (●) 53°C, (●) 51°C, (●) 49°C, (●) 47°C, (●) 45°C and (●) 43°C. (B) Simulated QCR kinetic traces calculated from $k_{label,i}$ derived from $G_{i,i}$ in panel A using Eq. 5. Values of $k_{int}$ were simulated using $\ln A$, $E_a/R$ and ABD concentration of 27.3, -8300 K and 2.6 mM respectively. (C) Simulated temperature dependence of fractional labeling at 300 s (indicated by the arrow in panel B). The solid line is a simulation of Eq. 15 using $\ln A$, $E_a/R$ and ABD concentration of 27.3, -8300 K and 2.6 mM respectively.
3.2.5 EX2 Conditions.

Analysis of fQCR experiments by Eq. 15 and 16 is valid only under EX2 conditions, when \( k_{\text{close}} > 10k_{\text{int}} \) (SI-2). A critical and simple diagnostic to identify whether cysteine labeling occurs under EX2 conditions is to measure the dependence of \( k_{\text{label}} \) on probe concentration (“probe slice”). At a given temperature \( T_i \), if the corresponding \( k_{\text{label},i} \) at different probe concentrations \( j \) exhibits a linear dependence (usually presented in a double log plot), EX2 conditions hold (Figure 20). Furthermore, if EX2 conditions do not hold, the transformation of the \( T_{ij}\phi_i \) relationship into \( T_i\Delta G_{Ui} \) (Eq. 15) no longer gives the parabolic shape characteristic of the Gibb-Helmholtz relationship (Figure 19A), and instead “peels” off (Figure 20B and SI-4). Visualization of these peeling regions indicates the combination of probe concentrations and temperatures where no inference can be drawn about protein stability. However, it may be possible to extract information about folding kinetics in these regions.
Figure 20: Simulation of EX2 behavior. (A) Dependence of $\log(k_{\text{label}})$ on $\log([\text{ABD}])$ at (●) 40°C, (●) 45°C, (●) 50°C, (▼) 55°C, (▼) 60°C and (▼) 70°C (refer to SI-4 for the details of this simulation). EX2 conditions (closed circles and solid lines) hold in the linear portion of these graphs (slope of 1.0 ± 0.2). (B) Apparent unfolding free energies derived by transforming the simulated rate constants in panel A using Eq. 5. The subset of free energy profiles (●, ●, ●) that correspond to an EX2 mechanism give a coincident free energy profile. The subset of free energy profiles that correspond to a loss of EX2 conditions at higher ABD concentrations and temperatures (▼, ▼, ▼) diverge from the coincident free energy profile “peeling”.

3.2.6 Estimation of Folding Rate Constants.

Even though no thermodynamic inferences can be drawn in the absence of EX2 conditions, it is possible to obtain kinetic information in this regime under certain conditions. Under non-EX2 conditions, $k_{\text{int}}$ is similar to, or faster than $k_{\text{close}}$ as temperature and probe concentration are increased. As discussed above, the situation is readily identifiable as nonlinearity in a plot of $\log(k_{\text{label}})$ versus $\log(P)$ obtained from a
probe slice (Figure 20 and SI-4). In this nonlinear region, $k_{\text{int}}$ therefore serves as a proxy for protein (un)folding rate constants as the fQCR experiment transitions out of EX2 labeling conditions. Beyond this nonlinear region, $k_{\text{label}}$ becomes independent of probe concentration and $k_{\text{int}} \gg k_{\text{close}}$.

In order to access the folding kinetics of a particular protein, the intrinsic reaction rate of the probe has to match $k_{\text{close}}$ within the range that is accessible by manipulation of probe concentration. Because protein folding rates vary over five orders of magnitude [186], a given probe can access folding kinetics of only a subset of proteins. For instance, ABD can be used to measure folding kinetics for proteins with folding rate constants between $\sim 3 \times 10^{-4}$ s$^{-1}$ and $\sim 1$ s$^{-1}$.

3.2.7 Experimental fQCR Results.

In a typical fQCR experiment, cysteine labeling is monitored as a function of temperature and probe concentration (Figure 21A, D and G). Global analysis of these data using Eq. 16 provides $k_{\text{int}}$, $F_{\text{max}}$, $T_m$, the temperature dependence of protein stability, estimates of folding rate constants, and confirms EX2 labeling conditions. Analysis of Eq. 16 (SI-3) using experimental parameters for ABD shows that $k_{\text{int,i}}$ is accurately determined under EX2 conditions at relatively low probe concentrations and short labeling times (typically 100-400 μM and 300 s). Outside of EX2 labeling conditions, $k_{\text{int}}$ can only be determined if $k_{\text{open}}$ is much faster than $k_{\text{int}}$ (SI-2). To determine $F_{\text{max}}$ in order to derive $f_i$, relatively high probe concentrations need to be used (typically 400-2000 μM).
The fQCR experiments presented here used variants of SN and ecRBP that contain single, buried cysteines and exemplify three classes of behavior (Figure 21): a system that is always in EX2 (SN.L36C, Figure 21A-C); one that switches out of EX2 at some elevated probe concentrations and temperatures (ecRBP.A188C, Figure 21D-F); and one that is never fully in EX2 (ecRBP.L61C, Figure 4G-I). In the first system, only thermodynamic parameters can be determined; in the second, both thermodynamic parameters and estimates of folding rate constants can be obtained; and the third, estimates of folding rate constants and $T_m$ can be made (Table 2).

Figure 21: fQCR results for SN and ecRBP.A, D, G: Temperature dependence of fractional labeling at single timed reaction endpoints monitored by ABD fluorescence fit with Eq. 16. Different curves represent different ABD concentrations or labeling times: SN.L36C, 300 s, (▼) 200 μM (▼) 400 μM, (●) 800 μM and (●) 1600 μM; ecRBP.A188C, 300 s (closed circles and triangles) or 600 s (closed squares), (▼, ▼) 200 μM (▼, ▼) 400 μM, (●, ▼) 800 μM, (▼) 1200 μM, and (●, ▼) 1600 μM; ecRBP.L61C, 300 s, (▼) 100 μM (▼) 200 μM, (●) 400 μM, (●) 600 μM, (●) 800 μM and (●) 1200 μM. B, E, H: Dependence of labeling rate constant on ABD concentration to identify EX2 conditions: SN.L36C, (●) 31.5°C, (●) 31.9°C, (●) 32.7°C, (●) 34.1°C, (●) 36.2°C, (●) 38.7°C and (●) 41.5°C; ecRBP.A188C, (●) 48.1°C, (●) 49.6°C, (●) 50.8°C, (●) 52.4°C, (●) 53.9°C, (●) 55.6°C and (●) 57.7°C; ecRBP.L61C, (▼) 45.6°C, (▼) 47.1°C, (▼) 49.1°C, (▼) 51.7°C, (▼) 54.3°C and (▼) 56.9°C. EX2 conditions (closed circles and solid lines) hold where the slope of the linear fit exceeds ~0.8. C, F, I: Transformation of fractional labeling to conformational free energy using Eqs. 13 and 5, and using the temperature dependence of $k_{int}$ obtained from Eq. 14. Gibbs-Helmholtz relationships (dashed line, Eq. 11 with $\Delta C_p$ fixed at 3.0 kcal mol$^{-1}$ K$^{-1}$) can be fit to cases where EX2 conditions can be observed for at least some combinations of temperature and protein concentration (SN.L36C, ecRBP.A188C). Non-EX2 behavior is seen as peeling (F). The transformation of $f_i$ to $\Delta G_{ii}$ is done only for $f_i$ values between 0.02 and 0.95 to minimize artifacts associated with the experimental uncertainty encountered at very low and high levels of cysteine labeling. (see figure on following page)
(see legend on preceding page)
Table 2: Experimentally derived fQCR parameters for SN and ecRBP

<table>
<thead>
<tr>
<th>Protein</th>
<th>$\Delta t$</th>
<th>[ABD] μM</th>
<th>$E_a/R$ K</th>
<th>$F_{max}$</th>
<th>$\Delta H_m$ kcal/mol</th>
<th>$T_m$ °C</th>
<th>EX2</th>
<th>$-k_{close}$ at $T_m$ s$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN.L36C</td>
<td>300</td>
<td>200</td>
<td>-7760 ± 8</td>
<td>41 ± 1</td>
<td>Yes</td>
<td>$\geq 1 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>400</td>
<td>-7875 ± 7</td>
<td>42 ± 1</td>
<td>Yes</td>
<td>$\geq 1 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>800</td>
<td>4100 ± 30</td>
<td>80 ± 3</td>
<td>Yes</td>
<td>$\geq 1 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>1600</td>
<td>4100 ± 20</td>
<td>78 ± 1</td>
<td>Yes</td>
<td>$\geq 1 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-7820$^a$</td>
<td>4100$^a$</td>
<td>83 ± 1$^b$</td>
<td>41 ± 1$^b$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ecRBP.A188C</td>
<td>300</td>
<td>200</td>
<td>-8163 ± 6</td>
<td>56 ± 1</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>400</td>
<td>-8110 ± 14</td>
<td>57 ± 1</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>800</td>
<td>4800 ± 40</td>
<td>121 ± 4</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>1600</td>
<td>4700 ± 50</td>
<td>117 ± 8</td>
<td>No</td>
<td>$4 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>200</td>
<td>56 ± 1</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>400</td>
<td>56 ± 1</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>800</td>
<td>4800 ± 40</td>
<td>128 ± 3</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>1200</td>
<td>4700 ± 40</td>
<td>125 ± 5</td>
<td>No</td>
<td>$4 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>1600</td>
<td>4700 ± 40</td>
<td>121 ± 4</td>
<td>No</td>
<td>$4 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-8140$^a$</td>
<td>4740$^a$</td>
<td>138 ± 1$^b$</td>
<td>57 ± 1$^b$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ecRBP.L61C</td>
<td>300</td>
<td>100</td>
<td></td>
<td>No</td>
<td>$&lt; 2 \times 10^{-2}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>200</td>
<td></td>
<td>No</td>
<td>$&lt; 2 \times 10^{-2}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>400</td>
<td>3926 ± 40</td>
<td>53 ± 1$^c$</td>
<td>No</td>
<td>$&lt; 2 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>600</td>
<td>3930 ± 40</td>
<td>53 ± 1$^c$</td>
<td>No</td>
<td>$&lt; 2 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>800</td>
<td>3845 ± 30</td>
<td>53 ± 1$^c$</td>
<td>No</td>
<td>$&lt; 2 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>1200</td>
<td>3824 ± 30</td>
<td>53 ± 1$^c$</td>
<td>No</td>
<td>$&lt; 2 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3881$^a$</td>
<td>53 ± 1$^c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$Average values for these parameters.
$^b$Thermodynamic parameters derived from a fit of the coincident free energy profiles in Figure 21C and F using the Gibbs-Helmholtz relationship (Eq. 11)
$^c$Quantified graphically
$^d$The folding rate constant is at least faster than the greatest value of $k_{int,i}$ encountered over this range of temperatures and probe concentrations
$^e$In the absence of EX2 conditions, $k_{int}$ represents the upper limit of $k_{close}$
3.2.8 $T_m$ Values.

At $T_m$, $\Delta G_{i\ell}$ is zero and Eq. 5 reduces to $k_{\text{int},i} = 2k_{\text{label},i}$. Under EX2 conditions, $T_m$ can therefore be identified as the temperature at which this relation is true. Alternatively, $T_m$ can be derived from fQCR datasets collected under EX2 conditions by identifying the temperature at which the fitted Gibbs-Helmholtz relationship is zero (Figure 21C and F). If cysteine labeling never conforms to an EX2 mechanism (Figure 21G), the determination of $T_m$ is more complex. In these cases, under the appropriate combination of labeling time, probe concentration, and where $k_{\text{open}} < k_{\text{int}}$, the midpoint of a single fQCR curve approximates $T_m$ (Figure 21G). These situations can only be identified by simulation (SI-5). In the case of ecRBP.L61C (Figure 21G), a labeling time of 300 s, probe concentrations between 400-1200 $\mu$M, and folding rate constants between $10^{-2}$-$10^{-3}$ give a set of fQCR curves with midpoints that estimate $T_m$. The $T_m$ values reported by these different classes of fQCR data (Table 2) are in agreement with those reported by unfolding monitored by CD (Figure 22).
Figure 22: Thermal unfolding of SN and ecRBP. Temperature unfolding of SN.L36C (▼), ecRBP.L61C (●) and ecRBP.A188C (•) at pH 7 monitored by CD signal at 222 nm. A fit of Eq. 1 to the data (solid lines) gives apparent $T_m$ values of 40 ± 1°C, 53 ± 1°C and 56 ± 1°C respectively.

3.2.9 Assessing Protein Function by fQCR.

Biomolecular function is the consequence of proteins interacting with other macromolecules (e.g., nucleic acids and other proteins) and small molecules (e.g., ligands, substrates and inhibitors) [145,146,148]. All of these interactions affect protein stability, typically by stabilizing the folded state [147]. Consequently, stability measurements in the absence and presence of a binding partner serve as a general tool for identifying and quantifying protein function. Figure 23 demonstrate how fQCR experiments can be used to assess protein function in terms of binding interactions. Temperature slices were acquired for SN.L36C (Figure 23A) and ecRBP.A188C (Figure 23B) in the absence (Apo) and presence of binding partner: 100 μM nucleotide inhibitor.
(SN) and 20 μM D-ribose (ecRBP). In both cases, addition of binding partner increases protein stability and shifts the endpoint labeling curves to higher temperatures. The free energy profiles in the absence ($\Delta^\text{apo}G_U$) and presence ($\Delta^iG_U$) of binding partner (see insets of Figure 23), and the relative difference in these free energies (i.e., relative stability, $\Delta\Delta G_U$) can be interpreted in terms of the linkage relationship

$$\Delta\Delta G_U = \Delta^i G_U - \Delta^\text{apo} G_U = RT \ln \left( 1 + \frac{[L]}{K_D} \right)$$

[17]

to obtain the equilibrium constant of binding ($K_d$) as function of temperature and free ligand concentration ([L]) within a GUWO (Table 3). These values are consistent with previous experiments[131].
Figure 23: Ligand binding by fQCR. Temperature slices were collected in the absence (black) and presence (purple) of 100 μM pdTp and 20 μM ribose for SN.L36C (A) and ecRBP.A188C (B) respectively. The insets in panels A and B correspond to the free energy profiles derived from the primary fQCR data for SN.L36C in the absence (black, $\Delta H_m = 83 \pm 1$ kcal/mol, $T_m = 41 \pm 1$) and presence (purple, $\Delta H_m = 94 \pm 3$ kcal/mol, $T_m = 43 \pm 1$) of pdTp, and for ecRBP.A188C in the absence (black, $\Delta H_m = 138 \pm 1$ kcal/mol, $T_m = 57 \pm 1$) and presence (purple, $\Delta H_m = 160 \pm 8$ kcal/mol, $T_m = 59 \pm 1$) of ribose.

Table 3: $K_d$ of SN and ecRBP as a function of temperature

<table>
<thead>
<tr>
<th>Protein</th>
<th>$T$ °C</th>
<th>$\Delta^{\text{up}} G_U$ kcal/mol</th>
<th>$\Delta^i G_U$ kcal/mol</th>
<th>$K_d$ μM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN.L36C</td>
<td>40</td>
<td>0.2</td>
<td>0.8</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>0.7</td>
<td>1.3</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>1.1</td>
<td>1.9</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>1.6</td>
<td>2.3</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>1.9</td>
<td>2.8</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>2.3</td>
<td>3.2</td>
<td>27</td>
</tr>
<tr>
<td>ecRBP.A188C</td>
<td>56</td>
<td>0.1</td>
<td>1.5</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>0.9</td>
<td>2.4</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>1.6</td>
<td>3.2</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2.3</td>
<td>4.0</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>3.0</td>
<td>4.8</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>3.7</td>
<td>5.5</td>
<td>1.2</td>
</tr>
</tbody>
</table>
3.3 Discussion

In the 1950’s Linderstrøm-Lang demonstrated how protein stability and folding kinetics are inextricably linked [187]. It is therefore possible to determine protein stability from reaction rates of probes with groups that are differentially accessible in the native and denatured states [161]. Here we show that these general principles can be applied to cysteine reactivity using ABD, a probe that becomes fluorescent only upon reaction with a thiolate. This fast determination of quantitative cysteine reactivity technique (fQCR) enables protein stability and folding kinetic properties to be determined using small amounts of material in short period of time with readily accessible instrumentation.

In an fQCR experiment, data is acquired as fluorescence intensities obtained at single timed reaction endpoints following addition of probe. A single protein sample is multiplexed into a two-dimensional set of parallel points, in which the first dimension samples different temperatures using a gradient PCR thermocycler, and the second dimension alters probe concentration. From this block of data we can extract probe labeling rate constants, and identify different exchange regimes within which $T_m$, $\Delta G_{\text{II}}$, and folding kinetics can be determined in various combinations. For systems fully in EX2, $T_m$ and $\Delta G_{\text{II}}$ can be determined; for systems that switch out of EX2 under some conditions, folding rate constants can additionally be extracted; for systems that never attain EX2, it may be possible to estimate folding rate constants and $T_m$. In all cases,
biological function as defined by the effects of molecular interactions can be evaluated by virtue of thermodynamic linkage between binding, stability, and folding kinetics. These relatively straightforward experiments can provide a remarkable amount of information regarding protein thermodynamic stability, folding kinetics, and biological function.
3.4 Materials and Methods

3.4.1 Protein Engineering and Purification.

Linear DNA encoding SN.L36C was assembled from synthetic oligonucleotides [110] and cloned into pET-21a (EMD Biosciences; 69740) at the XbaI and XhoI sites. A C-terminal GGSHEHHHHHHHK tag was introduced by QuickChange® (Stratagene; 200519).

Linear DNA fragments encoding ecRBP.L61C and ecRBP.A188C variants were assembled from synthetic oligonucleotides[110] and cloned into pET-21a at NdeI and XhoI sites. These mutations were introduced into a variant of the wild-type E. coli ribose-binding protein in which rare codons have been replaced, the N-terminal signal sequence removed, and a C-terminal GGSHEHHHHHH tag added. All sequence-verified cloned variants were transformed into E. coli KRX strain (Promega L3002) and stored as glycerol stocks at -80°C.

Starter cultures from glycerol stocks were diluted 1:500 into auto-induction media ZYM-5052[188] supplemented with 0.04% L-Rhamnose for delayed induction of the KRX T7 expression system. Cells were grown for 12-16 hours at 30°C with shaking at 225-270 rpm and pelleted by centrifugation at 6,000 g. ecRBP variants were lysed by addition of 1 mL BugBuster® Master Mix (EMD Biosciences; 71456) for each 20 mL of culture. The SN.L36C culture was lysed on ice by sonication for two minutes after resuspension in TBS (20 mM Tris, 136 mM NaCl, pH 7.4). In all cases 5 mM β-mercaptoethanol was added to prevent disulfide bond formation during lysis. Lysed
samples were centrifuged at 16,000 g for 10 min. The tagged recombinant proteins were purified by batch immobilized metal affinity chromatography by mixing the supernatant with an equal volume of binding buffer (15 mM imidazole, 20 mM MOPS, 500 mM NaCl, pH 7.5) before addition to His-Select Nickel affinity gel (Sigma P6611). Non-specifically bound protein was removed by four successive washes with 14 resin bed volumes of a 50:50 mixture of TBS and binding buffer supplemented with 5 mM β-mercaptoethanol. The tagged proteins were eluted in one bed volume of elution buffer (400 mM imidazole, 20 mM MOPS, 500 mM NaCl, pH 7.5), followed by buffer exchange on 10DG gel-filtration columns (Bio-Rad 732-2010) that were pre-equilibrated with 25 mM potassium phosphate, 100 mM KCl, pH 7. Samples were flash-frozen by dropwise addition (~20uL / drop) into liquid nitrogen, and stored as frozen beads at -80°C. No difference in stability (assessed by thermal melt) was observed between samples that underwent a single freeze-thaw cycle and aliquots stored overnight at 4°C. Protein concentration was measured spectrophotometrically using absorbance at 280 nm using extinction coefficients ($\varepsilon_{SN} \approx 16,000 \text{ M}^{-1}\text{cm}^{-1}$, $\varepsilon_{RBP} \approx 4,000 \text{ M}^{-1}\text{cm}^{-1}$) calculated from amino acid sequence[189]. Protein purity was assessed by SDS-PAGE.

3.4.2 Preparation of ABD Stock Solutions.

A ~0.5 M stock solution was prepared by dissolving 200 mg of 4-(aminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole (ABD; TCI America; A5597) in ~1.8 mLs of DMSO. The
final concentration was determined spectroscopically \((\varepsilon_{313\text{nm}} = 4200 \text{ M}^{-1}\text{cm}^{-1})\) [190]. This stock solution was stored in 30 μL aliquots at –20ºC, and was used over multiple freeze-thaw cycles. Working solutions of ABD were mixed immediately before an experiment by combining the appropriate volume of ABD stock solution and phosphate buffer (25 mM potassium phosphate and 100 mM KCl at pH 7).

### 3.4.3 Reaction of ABD with Glutathione.

The pseudo-first-order rate constants for the reaction of ABD with reduced glutathione \((L\text{-GSH}; \text{Sigma-Aldrich}; G4251)\) were determined from full kinetic traces measured at combinations of four temperatures (20, 30, 40, and 50°C) and pH values (6, 7, 8, and 8.5) in buffers containing 25 mM potassium phosphate and 100 mM KCl. The progress curve of the reaction between 10 μM \(L\text{-GSH}\) and 2.5 mM ABD was monitored in a SLM Series 2 Luminescence Spectrometer by recording the emission intensity of the fluorescent ABD-\(L\text{-GSH}\) adduct \((\lambda_{\text{ex}} 389 \text{ nm}; \lambda_{\text{em}} 513 \text{ nm}; \text{excitation bandwidth 8 nm}; \text{emission bandwidth 16 nm}; \text{PMT sensitivity 700})\). To ensure thermal equilibration, 1.45 mL of phosphate buffer was pre-incubated (with stirring) in a quartz cuvette for ten minutes in a temperature-controlled sample cell, after which 35.1 μL of 107 mM ABD was added (final concentration of 2.5 mM). The reaction was initiated after a further two minutes by addition of 15 μL of 1 mM \(L\text{-GSH}\) (final concentration of 10 μM).
3.4.4 The fQCR Experiment.

For a single fQCR experiment, 89 μL of 5-10 μM protein in phosphate buffer (25 mM potassium phosphate and 100 mM KCl at pH 7) was combined with 1 μL of 5 mM tris(2-carboxyethyl)phosphine (TCEP; Fluka; GA12963) and 10 μL of 10X ABD working stock solution (typically 1-20 mM). The final concentration of ABD was always at least 10-fold greater than the protein concentration to ensure that cysteine labeling followed pseudo-first-order kinetics. Following initial mixing of the 100 μL reaction, 7 μL aliquots were distributed into a PCR strip tube (12 tubes per strip) and incubated for 5 or 10 minutes (depending on the experiment) in a BioRad DNA Engine® Peltier gradient thermal cycler using a heated lid offset by +10ºC. After labeling, the samples were immediately cooled in an aluminum block kept at room temperature. Average fluorescence and standard error values determined from three independent readings of 1.5 μL aliquots of each sample on a NanoDrop 3300 fluorospectrophotometer (Thermo Scientific).

The temperatures of the thermocycler gradient were calibrated using a digital thermometer (Cole-Parmer; 7001H) wired to a high-precision thermister (Measurement Specialties; 44033RC) embedded in a PCR strip tube using a high thermal conductance compound (Artic Silver Inc; artic silver II). The difference between the temperatures reported by the thermocycler software and the recorded temperature could vary by as much as 5°C, depending on the difference between the heat of the lid and the sample.
Temperatures in each column of a gradient thermocycler were recorded after an equilibration time of one minute with a closed, heated lid (i.e. conditions identical to an fQCR incubation).

The fQCR experiments for ecRBP.L61C were modified to account for the relatively slow conformational equilibration of this protein over the experimental temperature range. To ensure conformational equilibrium was established for this variant, 9 μL aliquots of protein sample were transferred to PCR tubes and placed in the temperature gradient for 2 minutes prior to labeling. Following equilibration, 1 μL of ABD working stock solution was added to each aliquot using a multi-channel pipette to initiate the labeling reaction. After labeling for 5 minutes, the samples were processed as described above.

One fQCR data set (Figure 6B) was collected using a Tecan Genios fluorescence plate reader. In this case, 40 μL samples (5 μM protein, 400 μM ABD, 50 μM TCEP, 25 mM potassium phosphate, 100 mM KCl, pH 7) were incubated in the gradient thermocycler for 10 min and 30 μL aliquots transferred to 384-well black plates (Corning; 3821) after labeling. Fluorescence was measured at 30°C using 405 nm and 465 nm excitation and emission filters respectively.

3.4.5 Thermal Unfolding Monitored by Circular Dichroism.

Thermal unfolding experiments were carried out by preparing 3 mL of 3 μM protein in phosphate buffer (25 mM potassium phosphate and 100 mM KCl at pH 7) and
monitoring the change in CD signal at 222 nm as a function of temperature in a Jasco J-815 CD spectrometer. Samples were heated at a rate of 1°C per minute and allowed to equilibrate for 2 minutes following each temperature change. The midpoint of thermal unfolding, \( T_m \), was determined for each protein by fitting a two-state model of unfolding to the temperature dependence of normalized far-UV CD signal

\[
\theta_{222\text{ nm}}(T) = \frac{\theta_N + \theta_D e^{\Delta G_{GH}(T)/RT}}{1 + e^{\Delta G_{GH}(T)/RT}}
\]  

[18]

where \( \theta_N \) and \( \theta_D \) are linear relationships that describe the native (N) and denatured (D) baselines (e.g., \( \theta_N = m_N T + b_N \), describe the native baseline), and \( \Delta G_{GH}(T) \) is the Gibbs-Helmholtz relationship (Eq. 11). In the process of using Eq. 18 to derive \( T_m \), \( \Delta C_p \) was fixed at a value of 3 kcal mol\(^{-1}\) K\(^{-1}\).
4. Site Specificity and Local Unfolding in QCR

4.1 Introduction

As demonstrated in the previous chapter, QCR can be used to determine a wealth of thermodynamic properties: protein stability (Δ$G_u$), thermal stability ($T_m$), ligand-binding affinity ($K_d$), and in some cases (un)folding kinetics. These properties are inexorably coupled to biological function. At the cellular level, thermodynamic stability plays a key role in the turnover and posttranslational modification of proteins. Sequons that mark the protein for modification (e.g., phosphorylation, glycosylation, etc.) or render it susceptible to degradation (e.g., ubiquitination and proteolytic sites) may be hidden within the fold of the protein. Stability and folding kinetics will dictate the frequency and duration of transient unfolding events that expose these sequons and make them available for biochemistry. In this way, the thermodynamic stability of proteins reflects a facet of biological regulation encoded at the level of protein physical chemistry. At the molecular level, protein stability and (un)folding kinetics are directly affected by biomolecular interactions. In this way, the relationship between biomolecular interactions and protein stability may provide an additional level of functional control in the overall context of cellular physiology. Understanding a protein stability and its modulation by binding partners is therefore critical to understanding the protein’s role in its biological context.
In the previous chapter we focused primarily on global stability, and experiments were designed to remain within the global unfolding window of observation (GUWO). However, it has been well established that within the dynamic motions of proteins, certain localized regions can unfold with much greater frequency than the global whole. Because local unfolding also exposes previously buried protein structure to solvent, the degree of unfolding can be quantified by measuring the reaction rate of a buried moiety using HDX, and should likewise be possible with QCR. Site specificity afforded by a single cysteine site (QCR) or peak assignment (HDX) enables mapping the degree of local unfolding across regions of protein structure.

The existence of local unfolding raises an important question when considering the role stability plays in modulating cellular processes. Namely, is it global or local thermodynamics that controls the rate of stability-dependent reactions such as turnover and post-translational modification? Intuitively, it should depend on whether the modification site is found in a region of structure that can be appreciably exposed by local unfolding. However, to our knowledge this question has not been formally addressed and the possible link between local protein stability and the rates of cellular processes has not been demonstrated. It has been established that protein modification sites are frequently localized to structural regions that are classified as “intrinsically disordered,” and that protein disorder is accompanied by turnover. We hypothesize that so-called disordered regions simply represent an extreme on the continuum of local
unfolding, and that due to their extreme nature the impact of local unfolding is more apparent in these cases. To investigate the links between intrinsic disorder, local unfolding, and reaction rates, however, better tools for quantifying and mapping local unfolding are needed. In this chapter we evaluate QCR using ABD as a tool for determining local unfolding and quantifying its stability.

Pioneering work in the Englander lab established a mathematical framework for labeling by local unfolding, which builds upon the basic equations used for HDX and QCR presented in the previous chapter. The critical addition is that the observed labeling rate is, in fact, the sum of the labeling rate caused by local unfolding and the labeling rate caused by global unfolding. The apparent free energy of unfolding ($\Delta G_{u,app}$) is therefore also a composite function of global and local effects described by Equation 19.

$$
\Delta G_{u,app} = -RT \ln \left( e^{\frac{\Delta G_{global}}{RT}} + e^{\frac{\Delta G_{local}}{RT}} \right)
$$

[19]

The temperature dependence of global unfolding ($\Delta G_{global}$) is still defined by Gibbs-Helmholtz relationship (Eq. 11 of the previous chapter). However, the empirical observation made by hydrogen exchange is that the enthalpy and entropy of local unfolding are independent of temperature. Therefore in the Gibbs-Helmholtz equation for local unfolding $\Delta C_P$ is set to zero. To better visualize how the temperature
dependence of the local and global stability contributes to the overall temperature
dependence of the apparent free energy of unfolding, they are simulated individually
and together in Figure 24. From these simulations it quickly becomes apparent that with
the exception of a small transition region, either global or local unfolding dominates the
observed stability. The region where the global stability dominates, and thus the region
where the global stability curve overlaps with the apparent curve, defines the GUWO.

![Graph showing apparent free energy of unfolding](image)

Figure 24: Simulated global and local contributions to apparent free energy of
unfolding. The blue global stability curve is calculated with the standard Gibbs-
Helmholtz equation. The red local stability curve is calculated with a Gibbs-
Helmholtz equation where the change in heat capacity is set to zero. The apparent
stability curve is calculated using Eq. 19.

The study presented here had two main aims. The first was to determine
whether QCR could be used to measure local unfolding in different regions of a test
protein. The second was to assess to what degree other site specific properties of each
cysteine varied by location. Such properties include fluorescence intensity and spectra,
impact cysteine mutation on stability, and intrinsic reaction rate.
4.2 Results

5.2.1 Planning a Cysteine Scan in E. Coli Ribose Binding Protein

The L61C and A188C mutations made to ecRBP and described in the previous chapter were specifically selected on the basis of their burial deep within each lobe of the ecRBP structure. As such, they were not expected to be highly susceptible to labeling by a local mechanism. Indeed, no evidence of local unfolding was observed within the temperature range tested, indicating that the GUWO was never exceeded. Although, to be fair, we made no attempt to exit the GUWO when testing these proteins. In the present study we explicitly sought to observe local unfolding, and therefore desired sites with smaller GUWOs. Rather than attempt to predict which sites would demonstrate local unfolding without being entirely surface exposed, we took a blind approach to site selection that assumed no structural knowledge.

Making cysteine mutations can significantly destabilize a protein— an unfortunate but unavoidable consequence of the method. Naturally, we and the majority of potential users of QCR are most interested in information from variants that closely approximate the behavior of the wild type protein. Therefore, we restricted ourselves to sites where cysteine mutations are expected to be least disruptive and applied the Blosum62 alignment weights as a general benchmark. Accordingly, mutations were only made to Alanine, Isoleucine, Leucine, Methionine, Serine, Threonine, and Valine residues, all of which have no penalty or minimal mutational penalty in the Blosum
table (score of 0 or -1). This class of side chains represents 130 of the 270 residues in ecRBP. To maintain our blind approach, every third site of each side chain was selected for mutation to cysteine for a total of 46 variants (Figure 25).

![Cysteine Scan of ecRBP](image)

**Figure 25: Ca of sites selected for a cysteine scan of ecRBP color coded by side chain.** As expected from the selection method, sites are fairly randomly distributed across the structure. There is, however, a noticeable absence of binding pocket side chains. That was not by design, but neither is it entirely coincidental, as the ecRBP pocket features a number of ionizable and bulky hydrophobic side chains which were excluded from mutagenesis on the basis of the Blosum62 penalty.

Mutations were made to cloned ecRBP using QuikChange mutagenesis. Protein from sequence verified colonies was expressed in KRX cells, purified in the presence of reductant by affinity resin, and desalted before being aliquotted and stored at -80°C. SDS-PAGE demonstrated a disulfide bond formation by some variants despite the limited time (~3 hours) and mild conditions (4°C) between desalting and the gel (Figure 26)

108
Figure 26: Some ecRBP cysteine mutants form covalent dimers under mild conditions. SDS-PAGE demonstrated that purified protein was free of contaminants with the exception that some variants showed bands at approximately twice the molecular weight of ecRBP. Given the presence of cysteine, these bands were suspected to be disulfide-bonded dimers. An examination of the 5 variants with the most dimer (boxed in red) demonstrates that without exception, the cysteine sites are located on the surface and oriented towards solvent (bottom). It is suspected there was an error loading variant V269C because other experiments demonstrated protein.
5.2.2 Site Dependent Properties of Cysteine Variants

Despite restricting ourselves to more “conservative” mutations, we anticipated that many of the cysteines inserted would destabilize ecRBP compared to wild type. As an initial measure of thermal stability, SYPRO melts were performed (Figure 27).

Figure 27: $T_m$ of cysteine variants determined by SYPRO thermal melts. $T_m$ values shown are the average of three independent thermal melts (error bars are SD), (top). With one exception (L170C), all variants were destabilized compared to wild type, whose SYPRO $T_m$ was previously determined to be 57.2°C. Classification by structural location (bottom, left) shows that surface mutations are the less likely to be destabilizing, and core mutations are the more likely to be destabilizing. Classification by original residue identity shows that Leucine and Serine mutants were generally the least destabilizing, although care should be taken attempting to generalize from these results, given the small number of data points for each side chain. Numeric values for each site can be found in Table 4.
In addition to the destabilizing influence of cysteine insertion, covalent labeling can potentially further destabilize the protein. In essence, ABD labeling is the functional equivalent of mutating a residue to a non-natural amino acid larger than tryptophan. Additional destabilization upon labeling is important for two reasons. The first regards proteins containing multiple cysteines; labeling at one position could promote labeling at other positions, resulting in cooperative labeling. Care should therefore be taken when interpreting QCR results from proteins that naturally contain multiple cysteines. If cooperative labeling occurs, it is likely that the apparent GUWO would only be as large as that of the cysteine site most susceptible to local unfolding. Experimental validation is needed to test this hypothesis. The second reason, which is more germane to this study, is that ABD fluorescence is determined in part by the local environment. If the protein is further destabilized into the temperature regime where measurements are made, small changes in temperature could result in large changes in signal. By following the ABD fluorescence of fully labeled protein, we can investigate this phenomenon and determine a transition temperature as a function of temperature (Figure 28). Although physical chemistry dictates that fluorescence generally decreases with increasing temperature (due to increased molecular collisions) for most variants a transition region featuring either a sharper decrease or absolute increase relative to the baseline can be defined. Low post-labeling stability also raises concerns about protein aggregation during
prolonged incubation at temperatures which might be below the original $T_m$ but are above the stability of the labeled state.

Figure 28: Examples of temperature-dependent transitions in ABD labeled proteins. Fluorescence can dramatically decrease as is the case for L265C, dramatically increase as exemplified by M210, increase slightly, which is interpreted as a transition in the context of the decreasing baseline, as exemplified by V71C, or show no clear transition, as is the case for I86C. Numerical values of transition temperatures for the entire the set of variants can be found in Table 4.
Given the structural sensitivity of ABD, as shown in Figure 28 by the response to temperature, it is not surprising that absolute fluorescence intensity also varies from site to site as each residue location offers a unique microenvironment. Differences in fluorescence were investigated by recording the fluorescence intensity and emission spectra for identical concentrations of fully labeled variants (Figure 29). Despite expecting variability, the span of signal intensity was somewhat surprising with the brightest variant signaling ~75fold brighter than the weakest variant. This means that dimmer sites might have to be used at greater protein concentrations to avoid noisy signal. It also means that if ABD based QCR were to be applied in conjunction with protein engineering and cell-free expression (as in Chapter 2), it would probably be worth screening cysteine locations in order select a variant that will produce high signal at low concentration. Notably, 13 of the 14 dimmest variants demonstrate an increase in fluorescence upon thermal transition. This raises the possibility that for specific variants optimal read temperatures could be defined that maximize signal. A protocol where the reaction was quenched by low pH would allow such heating without further labeling. In addition to variability in intensity, the peak fluorescence wavelength also varied from 465nm to 563nm and exhibits a somewhat bimodal distribution with variants clustered around 490nm and 515nm. Practically, wavelength variability may be less of an issue than low intensity, but it could impact filter selection for maximal signal in certain instruments.
Full numerical values for each variant can be found in Table 4. These values can be used to help determine which sites are the most interesting and practical for initial QCR study. For example we could initially focus on sites that: destabilize ecRBP by less than 3°C, were in the upper 50% percentile of fluorescence intensity, did not form disulfide dimers from the folded state and where not drastically destabilized by labeling. This set includes L265C, A92C, L37C, M77C, and L170C.
Figure 29: Wavelength and maximal fluorescence distribution across ecRBP cysteine variants. From the wavelength distribution it would appear that core residues are somewhat blueshifted as compared to boundary and surface residues, but there is no clear by trend by residue identity (top). From maximal fluorescence distribution it would appear that surface residues are generally dimmer than boundary and core and those mutations from Leucine are amongst the brightest (middle). There is also a weak apparent correlation between shorter wavelengths and increased fluorescence intensity(bottom). Site by site details can be found in Table 4. In all histograms bin labels represent the upper bin limit.
Table 4: Site-dependent properties of ecRBP Cysteine variants

<table>
<thead>
<tr>
<th>Position of Cysteine Mutant</th>
<th>Core, Boundary, Surface</th>
<th>Mutant T&lt;sub&gt;m&lt;/sub&gt;</th>
<th>ABD SYPRO Transition Temperature</th>
<th>AFU Change upon Transition</th>
<th>AFU Max</th>
<th>λ max (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V008</td>
<td>C</td>
<td>55.3</td>
<td>41</td>
<td>decrease</td>
<td>7360</td>
<td>487</td>
</tr>
<tr>
<td>L019</td>
<td>C</td>
<td>54.2</td>
<td>35</td>
<td>decrease</td>
<td>5542</td>
<td>479</td>
</tr>
<tr>
<td>V036</td>
<td>S</td>
<td>55.3</td>
<td>NA</td>
<td>NA</td>
<td>4839</td>
<td>491</td>
</tr>
<tr>
<td>L037</td>
<td>B</td>
<td>56.1</td>
<td>55</td>
<td>decrease</td>
<td>8705</td>
<td>483</td>
</tr>
<tr>
<td>S039</td>
<td>C</td>
<td>53.9</td>
<td>NA</td>
<td>NA</td>
<td>6106</td>
<td>483</td>
</tr>
<tr>
<td>A044</td>
<td>S</td>
<td>56.1</td>
<td>53</td>
<td>increase</td>
<td>694</td>
<td>519</td>
</tr>
<tr>
<td>T054</td>
<td>C</td>
<td>56.9</td>
<td>52</td>
<td>increase</td>
<td>699</td>
<td>495</td>
</tr>
<tr>
<td>I060</td>
<td>C</td>
<td>56.3</td>
<td>NA</td>
<td>NA</td>
<td>3682</td>
<td>479</td>
</tr>
<tr>
<td>L062</td>
<td>C</td>
<td>55.0</td>
<td>31</td>
<td>increase</td>
<td>4381</td>
<td>487</td>
</tr>
<tr>
<td>V071</td>
<td>C</td>
<td>56.2</td>
<td>48</td>
<td>increase</td>
<td>1983</td>
<td>507</td>
</tr>
<tr>
<td>A074</td>
<td>C</td>
<td>55.2</td>
<td>52</td>
<td>increase</td>
<td>1270</td>
<td>471</td>
</tr>
<tr>
<td>M077</td>
<td>B</td>
<td>56.6</td>
<td>55</td>
<td>decrease</td>
<td>6529</td>
<td>487</td>
</tr>
<tr>
<td>I086</td>
<td>C</td>
<td>54.7</td>
<td>NA</td>
<td>NA</td>
<td>3280</td>
<td>487</td>
</tr>
<tr>
<td>T087</td>
<td>C</td>
<td>51.3</td>
<td>37</td>
<td>increase</td>
<td>2546</td>
<td>487</td>
</tr>
<tr>
<td>A092</td>
<td>C</td>
<td>55.4</td>
<td>49</td>
<td>decrease</td>
<td>12057</td>
<td>467</td>
</tr>
<tr>
<td>V097</td>
<td>B</td>
<td>54.3</td>
<td>52</td>
<td>increase</td>
<td>1887</td>
<td>515</td>
</tr>
<tr>
<td>S103</td>
<td>C</td>
<td>54.3</td>
<td>NA</td>
<td>NA</td>
<td>6958</td>
<td>487</td>
</tr>
<tr>
<td>L107</td>
<td>S</td>
<td>53.6</td>
<td>54</td>
<td>decrease</td>
<td>3648</td>
<td>507</td>
</tr>
<tr>
<td>I116</td>
<td>C</td>
<td>44.5</td>
<td>NA</td>
<td>unclear</td>
<td>479</td>
<td>515</td>
</tr>
<tr>
<td>A117</td>
<td>C</td>
<td>53.1</td>
<td>53</td>
<td>increase</td>
<td>237</td>
<td>515</td>
</tr>
<tr>
<td>V126</td>
<td>C</td>
<td>52.6</td>
<td>NA</td>
<td>NA</td>
<td>2217</td>
<td>491</td>
</tr>
<tr>
<td>A133</td>
<td>S</td>
<td>54.2</td>
<td>48</td>
<td>decrease</td>
<td>5027</td>
<td>519</td>
</tr>
<tr>
<td>S136</td>
<td>B</td>
<td>55.2</td>
<td>52</td>
<td>increase</td>
<td>1193</td>
<td>515</td>
</tr>
<tr>
<td>A148</td>
<td>C</td>
<td>49.7</td>
<td>41</td>
<td>increase</td>
<td>748</td>
<td>503</td>
</tr>
<tr>
<td>A158</td>
<td>B</td>
<td>52.0</td>
<td>51</td>
<td>increase</td>
<td>1456</td>
<td>491</td>
</tr>
<tr>
<td>I167</td>
<td>S</td>
<td>55.7</td>
<td>54</td>
<td>increase</td>
<td>787</td>
<td>519</td>
</tr>
<tr>
<td>L170</td>
<td>S</td>
<td>58.1</td>
<td>60</td>
<td>decrease</td>
<td>4364</td>
<td>495</td>
</tr>
<tr>
<td>V172</td>
<td>B</td>
<td>52.2</td>
<td>NA</td>
<td>NA</td>
<td>3623</td>
<td>495</td>
</tr>
<tr>
<td>M173</td>
<td>S</td>
<td>50.5</td>
<td>NA</td>
<td>NA</td>
<td>4362</td>
<td>483</td>
</tr>
<tr>
<td>T178</td>
<td>S</td>
<td>55.5</td>
<td>52.5</td>
<td>increase</td>
<td>1995</td>
<td>519</td>
</tr>
<tr>
<td>A185</td>
<td>C</td>
<td>51.2</td>
<td>41</td>
<td>increase</td>
<td>454</td>
<td>487</td>
</tr>
<tr>
<td>M193</td>
<td>C</td>
<td>45.6</td>
<td>NA</td>
<td>NA</td>
<td>2432</td>
<td>491</td>
</tr>
<tr>
<td>A197</td>
<td>C</td>
<td>53.0</td>
<td>NA</td>
<td>NA</td>
<td>3231</td>
<td>463</td>
</tr>
<tr>
<td>S207</td>
<td>S</td>
<td>56.1</td>
<td>53</td>
<td>increase</td>
<td>1250</td>
<td>519</td>
</tr>
<tr>
<td>V209</td>
<td>C</td>
<td>55.1</td>
<td>43</td>
<td>increase</td>
<td>1274</td>
<td>503</td>
</tr>
<tr>
<td>M210</td>
<td>S</td>
<td>56.6</td>
<td>50</td>
<td>increase</td>
<td>1282</td>
<td>495</td>
</tr>
<tr>
<td>A223</td>
<td>B</td>
<td>51.2</td>
<td>51</td>
<td>decrease</td>
<td>4748</td>
<td>483</td>
</tr>
<tr>
<td>V224</td>
<td>C</td>
<td>49.1</td>
<td>47</td>
<td>decrease</td>
<td>2752</td>
<td>495</td>
</tr>
<tr>
<td>T232</td>
<td>C</td>
<td>47.0</td>
<td>43</td>
<td>decrease</td>
<td>3861</td>
<td>491</td>
</tr>
<tr>
<td>V257</td>
<td>C</td>
<td>52.4</td>
<td>47</td>
<td>decrease</td>
<td>1772</td>
<td>507</td>
</tr>
<tr>
<td>M259</td>
<td>S</td>
<td>54.4</td>
<td>55</td>
<td>decrease</td>
<td>1701</td>
<td>515</td>
</tr>
<tr>
<td>L265</td>
<td>C</td>
<td>56.9</td>
<td>49</td>
<td>decrease</td>
<td>14539</td>
<td>479</td>
</tr>
<tr>
<td>V269</td>
<td>S</td>
<td>54.3</td>
<td>53</td>
<td>increase</td>
<td>191</td>
<td>523</td>
</tr>
</tbody>
</table>
5.2.3 QCR and Local Unfolding

Performing QCR for the purposes of measuring local unfolding is nearly identical to running a QCR experiment to determine global stability (as described in the previous chapter). The main difference is that a larger temperature region is covered in order to exit the GUWO region. In most cases this will require additional temperature ranges for incubation and/or more times incubation per variant. A short (~5-10 min) and hot (spans the $T_m$) incubation range is still required in order to empirically determine intrinsic reaction rate, as described previously. Recall that intrinsic reactivity can differ on a site-to-site basis resulting from differences in cysteine pKa which determine the fraction of the side chain found in the thiolate form. Intrinsic reactivity therefore needs to be established for each variant individually. A longer (~one to several hours) and cooler (below the $T_m$) incubation extends the stability information into a regime where appreciable labeling can only be observed if the side chains are labeled by a local mechanism.

The inability to observe global labeling over this second temperature range is the result of the increasingly infrequent exposure of cysteines at greater stability and the practical limits on parameters used to speed up the labeling rate. For example, labeling reagent concentration can be limited by solubility, background fluorescence of unconjugated dye, and inner filter effects. Incubation time can be limited by propensity
for protein aggregation and label hydrolysis. pH can be limited by protein destabilization effects resulting from ionizable side chains. In the case of ABD and ecRBP these limits place the upper bound of stability determination roughly between 5 and 6 kcal/mol. Therefore, if local unfolding is not observed by QCR it does not imply the side chain can only label through a global mechanism. The conclusion from this scenario is that, regardless of labeling mechanism, stability is in excess of ~5 kcal/mol. Indeed, hydrogen exchange experiments have shown that only a very small minority of side chains are absolutely restricted to exchange via global mechanism. Do note, however, that a stability of 5.7 kcal/mol at 37°C implies that for every protein molecule in the unfolded, labeling competent, state there are over ten thousand protein molecules in the fully folded, protected, state.

Data from three cysteine sites is shown that demonstrates the capacity for QCR to measure local unfolding. Raw data, the conversion to stability values over the experimental range, and the EX2 tests for each incubation are shown for L265C in Figure 30. Mutation locations for S99C, L170C, and L265C as well as the stability profile for the three variants are show in Figure 31. As was the observation by hydrogen exchange, once labeling is dominated by the local mechanism, energetics can be approximated by a linear relationship, re-enforcing the notion that change in heat capacity for the local unfolding reaction is indeed negligible.
Figure 30: Raw QCR data, $\Delta G$ conversion, and the EX2 test for L265C. (A) Raw data are the fluorescence intensity as a function of temperature, incubation time, and dye concentration. These data are converted to fraction labeled and used in the calculation of stability as a function of temperature (B). EX2 tests for the long incubation (C) and short incubation (D) show that all temperatures below 325K are in EX2 because the slope of label concentration vs. labeling rate is above 0.8.
Figure 31: Three examples of local unfolding in ecRBP measured by QCR. S99C (top) features the largest GUWO and the stability of local unfolding at 37°C is within 1kcal/mol of global stability. L170C (middle) features the smallest GUWO and also demonstrates extreme loss of EX2 behavior at high temperatures as indicated by the curve peeling. L265C, which is located very near the C terminus, exhibits an intermediate GUWO. ABD concentrations and incubation times are shown in the legends.
The specific energetic values calculated for local unfolding rely on the value of the intrinsic reaction rate. In the QCR protocol, this rate is calculated empirically from the data set that spans the $T_m$. The calculation assumes that the $pK_a$ of the cysteine in the fully denatured form is identical to the $pK_a$ of the cysteine in the locally unfolded form. We recognize that this assumption may not be 100% accurate, but note that hydrogen exchange suffers from a similar issue. It is regularly observed that actual intrinsic hydrogen exchange rate is slower than that predicted by the model peptides used to estimate the rate as a result of residual blocking structure in the unfolded form.

Error in the intrinsic reaction rate can also occur as a function of erroneous incubation time values. Although it is straightforward to record the duration that the physical tubes remain in the incubation block, this value does not consider the dead time in bringing the solution up to temperature. We note that when a thermister is used to record the time course of equilibration from 25°C to 65°C, even after a full minute of incubation the temperature is still below 64°C. In the context of a 5 minute incubation, a minute of equilibration time results in an appreciable underestimation of the reaction rate. When only a single incubation time and temperature range is considered (as in the previous chapter) this error is irrelevant for stability calculations. This is because errors in time are compensated for by errors in the intrinsic reaction rate. In fact, artificially changing the value of incubation time from the true value of 5 minutes down to 1 minute or up to 1 hour during data fitting changes intrinsic reaction rate accordingly but
has absolutely no impact on the calculated $T_m$ or enthalpy values. However, if the erroneous reaction rate value is used to process data from an incubation range where dead time is a small relative to total time, energetic values calculated are indeed perturbed. This error manifests as a gap between the stability curves generated by different ranges of incubation temperature and can be resolved by adding a dead time correction factor to the fast incubation to bring the data sets in agreement with one another.
4.3 Discussion

QCR is capable of measuring local stability while maintaining the speed and ease-of-use advantages outlined in the previous chapter. The burial of a cysteine within the folded structure and the protection from reactivity afforded by burial is not fundamentally different than the burial of a modification site being protected from post-translational reactions in the cellular context. We therefore feel there is no reason why stability-dependent cellular processes should not be directly influenced by local unfolding as well as global unfolding. Given the results presented here, we are hopeful that QCR can play a central role in the exploration of these relationships.

The cysteine scan in the ecRBP model system can also be used to probe the role that ligand binding can play in local unfolding by repeating these experiments in the presence of ribose. Ligand has a clear and predictable impact on global stability, as shown by the SYPRO thermal shift experiments (Figure 11) and the calculation of $K_d$ by QCR (Figure 23). However, this impact is predicated on the fact the folded state binds the ligand with higher affinity than the unfolded state. It is not intuitively obvious whether a given locally unfolded state will suffer a loss of affinity for ligand, and whether ligand can therefore shift the equilibrium of local unfolding. Ligand may increase local stability in some regions (particularly those involved in ligand recognition) but not others, and QCR experiments can be used to elucidate and map these effects.
Understanding the way in which ligands can modulate local stability will be important in the field of disease treatment by “pharmacological chaperones”. Protein misfolding diseases and protein deficiency diseases caused by increased turnover of unstable allele variants have been shown to be amenable to treatment with ligand or co-factor supplementation [191]. These diseases include: Fabry’s Disease, Tay-Sachs, Sandhoff Disease, GM1-Gangliosidosis, Pompe Disease, Phenylketonuria, Cystic Fibrosis, and certain cancers. If local unfolding is a dominant factor in protein turnover, then the success of therapeutic intervention may depend more on whether the ligand increases local stability than its impact on global stability.
4.4 Materials and Methods

4.4.1 Variant Generation and Protein Production

Sites were selected as described in the text. QuikChange oligos were designed using the online primer design tool (available at www.genomics.agilent.com) and ordered from Integrated DNA Technologies. QuikChange was performed according to manufacturer’s instructions and plasmids transformed into KRX cells (Promega). Sequence verified ecRBP variants with HIS tags were purified and concentration determined as in Chapter 3. SDS-PAGE gels were run as in Chapter 2.

4.4.2 Characterization Of Site Dependent Properties

SYPRO thermal melts were performed as in Chapter 2 on the LightCycler using 5 µM protein. Fully ABD labeled protein was produced by incubating 10 µM ecRBP with 1mM ABD and 50 µM TCEP for 8 minutes at 65°C. At this temperature all ecRBP variants are fully unfolded, and the reaction proceeds to completion. Fluorescence intensity and spectra were recorded on a Nanodrop-3300. Transition temperature of the fully labeled state was determined by recording fluorescence while heating in a FluoDiaT70 plate reader using a 389nm excitation filter and a 475nm emission filter.

4.4.3 fQCR for Local Unfolding

QCR was performed as described in Chapter 3. All reactions were performed in buffer at pH 7. L265C and S99C were incubated over two temperature ranges 30-46°C and 43-63°C for ABD concentrations and incubation times as listed in Figure 31. L170C
was incubated over three temperature ranges 30-50°C, 40-60°C, and 47-67°C for ABD concentrations and incubation times as listed in Figure 31. For each variant, all conditions that were determined to both be in EX2 and represent global unfolding were globally fit in Origin software using a master fluorescence equation that combines Equations 5, 6, 9, and 11 from Chapter 3.

\[
F = F_0 + \Delta F - \Delta F \left(1 + \frac{1}{e^{\left(-\frac{E_a}{RT}\right) t}}\right)
\]

In this protocol, fluorescence (F), temperature (T) and label concentration ([ABD]₀), and time (t) are experimentally determined and can vary from trace to trace. Pre-exponential factor (A), heat capacity (ΔCₚ) baseline fluorescence (F₀) and fluorescence range (ΔF) are global constants that are held fixed. Activation energy (Eₐ), enthalpy (ΔHₘ) and melting temperature (Tₘ) are global constants that are determined from the fitting process.
5. Oscillations by Minimal Bacterial Suicide Circuits Reveal Hidden Facets of Host-Circuit Physiology

5.1 Introduction

Synthetic biology [23,192,193,194,195,196] seeks to enable predictable engineering of cells and biological systems with altered or expanded function. Critical to this effort is the (re)design of information processing that establishes the timing and execution of cellular operations. At the molecular level, interpretation of particular internal and external stimuli is determined by the concentration and activity of cellular components (such as proteins, nucleic acids, and metabolites). Cellular responses, in turn, are executed by dynamically modulating these components in accordance with the cell’s genetic program. The task of engineering synthetic systems therefore requires an understanding of not only the system components, but also their interactions, and the control mechanisms that adjust concentration and activity. In principle, a perfect understanding of these factors enables the development of models that accurately predict behavior for a proposed design. In reality, the scale and scope of cellular physiology, coupled with an imperfect understanding of the system and host components, make the construction of such models quite challenging. For this reason, simplified models that assume a well-defined interface between the circuit and host, while ignoring the background processes of host metabolism, are generally employed. If the predictions generated by the simple models deviate from the experimental
implementation it indicates that the model insufficiently encompassed the critical components and interactions of the system. In this manner, the validity of common simplifying assumptions can be partially addressed and refined during the design process.

In this work we present a genetic circuit (ePop) that causes oscillations in bacterial population density over time. The oscillations arise through the unanticipated interplay of growth conditions, the host cell, and one of the circuit’s “background processes” – plasmid replication. Specifically, a density-dependent rise in plasmid copy number leads to an increase in gene dosage and concomitant increase in the expression of a plasmid-borne toxin gene. Toxin expression causes cell lysis and decreases the population density, allowing growth recovery, and generating oscillations as multiple cycles proceed. Although this mechanism was not intended during original circuit design, a refined mathematical model that incorporates plasmid copy number control captures the circuit dynamics, and predictions based on the model are validated by experimental results. Conditional plasmid amplification as a control mechanism for population density has not previously been described and evokes its potential for other applications. Furthermore, these results emphasize that copy number deserves increased attention when designing plasmid-based synthetic gene circuits.

One synthetic biology focus has been exhaustive documentation and standardization of individual biological parts [197]. While these efforts are valuable,
circuits such as ePop exemplify the context-dependence of parts and devices and highlight the intrinsic difficulty in attempting to anticipate every possible circuit/host/condition interaction - even when using previously described and well-characterized parts, as is the case for ePop. Elucidating basic control mechanisms and improving tools that enable the assessment of parts in their final context is therefore important to the advancement of synthetic biology.
5.2 Results

5.2.1 ePop Causes Oscillations In Cell Density

The gene circuit (ePop) contains two modules (Figure 32). The first module confers cell killing and consists of a lysis gene \((E)\) from phage \(\phi X174\) \([198,199]\) placed behind the \(luxI\) promoter \((P_{luxI})\) from \(Vibrio fischeri\), recently reclassified as \(Aliivibrio fischeri\) \([200]\). The \(E\) protein is an inhibitor of MraY, an enzyme that catalyzes the production of the first lipid intermediate in \(E. coli\) cell wall synthesis. Cells deficient in MraY activity lyse during septation, a process that requires newly synthesized cell wall \([201]\). The second module was intended to confer density sensing and consists of an inadvertently mutated \(luxR\) gene \((luxR^*)\) and an intact \(luxI\) gene under control of an isopropyl \(\beta\)-D-1-thiogalactopyranoside (IPTG) inducible promoter \((P_{lacIa-1})\). The \(luxR\) and \(luxI\) genes originate from \(A. fischeri\) and constitute a quorum sensing pair. LuxI produces a quorum signal, an acyl- homoserine lactone (AHL), from cellular precursors. LuxR is a transcription factor that responds to AHL and activates the \(P_{luxI}\) promoter. In ePop, \(luxR^*\) contains a frame-shift mutation that introduces an early stop codon and completely abolishes the \(luxR\) DNA-binding domain, as determined by homology to \(TraR\) \([202]\). ePop was implemented in a single plasmid that carries a chloramphenicol resistance gene and a ColE1 origin of replication that lacks the Rom/Rop protein.
Figure 32: ePop dynamics in liquid culture. MC4100z1 cells containing the ePop circuit grown in liquid culture exhibited regular oscillations in cell density. Each trace represents a culture started from an individual colony. Cells were grown in the absence (top) or presence (bottom) of 1mM IPTG and OD (600nm) was measured. IPTG did not drastically change the nature of the oscillations but did affect the synchronization across different colonies and starter cultures (different colored traces).

When transformed into MC4100z1 cells, ePop generated multiple cycles of population oscillations during long-term batch cultures. These oscillations were highly reproducible and showed only minor difference upon induction by IPTG (Figure 32). MC4100z1 cells consistently displayed many cycles of oscillations, proceeding at times
for five rounds of lysis. Oscillations were also observed in TOP10F’, DH5α, BW25113 and MG1655. Colony forming unit (CFU) experiments showed that optical density reflected changes in viable cell density (Figure 33), although loss of viability appeared to precede lysis – consistent with previous characterization of E protein function. MC4100z1 cultures that did not carry ePop did not oscillate under these conditions (Figure 34).

![Graph showing optical density measurements](image)

**Figure 33:** Optical density measurements are consistent with cell density. Colony forming unit (CFU) experiments were performed every three hours and show that viable cell density correlated with OD for two separate colonies.

The increasing baseline observed in optical density can be attributed to the accumulation of lysed ‘ghost’ cells. Solid phase measurements with BW25113 show synchronous growth arrest and lysis (Figure 35). The MC4100z1 oscillations were among the most robust oscillations generated in a population of cells by synthetic gene circuits. Oscillations from past studies were less robust [14,15,16], operated among a sub-population of individual cells [14,15], or required highly specialized devices to be
observed [14,15,63,203]. Generation of oscillations by synthetic circuits has recently been
extended to mammalian cells [204].

Figure 34: ePop plasmid is required for oscillations. MC4100z1 cells that do not
carry any plasmids (red) do not show the oscillations exhibited by MC4100z1 cells
that carry the ePop plasmid (green) demonstrating that strain and culturing
conditions were insufficient to produce oscillations.

To further characterize the oscillations, we investigated effects of culturing
conditions specific to our growth assay, including media, temperature, and oxygenation.
Complex media (Luria Broth) might have contributed to oscillations as shifts in
preferred nutrient sources during growth could impact E protein expression. However,
cells grown in defined M9 minimal media supplemented with glycerol exhibited
oscillations similar to those grown in LB supplemented with glycerol (Figure 36).
Culturing temperature affects many parameters, such as growth rate and rates of
cellular reactions – all of which might impact oscillations. Cultures grown at 33°C
 oscillated, but both the oscillation period and threshold OD were increased relative to
cells grown at 37°C (Figure 37). Because the majority of experiments were performed
under mineral oil, and oxygen availability is a factor that affects luminescence in *V. fischeri*, anaerobic growth conditions might have contributed to oscillations. Aerobically grown cells underwent one or two rounds of lysis, but did not oscillate as well. These results indicate that oxygen availability contributed to but was insufficient to account for the observed oscillations.

Figure 35: Solid phase growth of BW25113 carrying ePop. Although the microcolony shows synchronous growth arrest and lysis, oscillations were not observed as no cells at this location recovered.
Figure 36: ePop oscillates in defined media as well as complex media. Cells grown in buffered LB supplemented with glycerol (left) showed similar oscillations to cells grown in M9 minimal media supplemented with glycerol (right). Different colored traces represent individual colonies. Oscillations are therefore not due to some unknown component in complex media or shifts in preferred media source.

Figure 37: Effect of culturing temperature on ePop oscillations. Cells grown at 33°C (black) exhibit an elevated lysis density and longer period than those grown at 37°C (red).
5.2.2 Oscillations do not Depend on *luxRI* Quorum Sensing or Promoter-Level Regulation

For oscillations to occur, the lysis phenotype conferred by the *E* gene must be triggered at elevated cell density. However, the *luxR* truncation and the low impact of IPTG induction suggested that the *luxR* and *luxI* genes were not responsible for density sensing. To test their role, we constructed two new versions of the ePop circuit. By design, both have the same *P_{lux}-E* module as ePop. ePop-*lite* contains functional *luxR* but no *luxI* (Figure 38, middle); ePop-*mini* contains neither *luxR* nor *luxI* (Figure 38, bottom). Both ePop-*lite* and ePop-*mini* generated oscillations similar to those generated by ePop. Furthermore, exogenously added AHL did not affect oscillations by ePop (Figure 38, top) or ePop-*mini* (Figure 38, bottom) but suppressed growth of cells containing ePop-*lite*, in a dose-dependent manner (Figure 38, middle). These results confirmed that, unlike cells carrying ePop-*lite*, cells carrying ePop did not produce functional LuxR.
Figure 38: ePop based circuits treated with AHL. (A-C) Cultures treated with AHL: red (1000nM), green (100 nM), yellow (10nM), and black (0 nM). (A) Cells containing ePop oscillated independent of AHL concentrations. (B) Cells containing ePop-lite oscillated, but showed dose-dependent sensitivity to AHL (C) Cells containing ePop-mini oscillated independent of AHL concentrations (although only two cycles of lysis were observed here, other experiments with this plasmid showed up to four cycles).
More significantly, luxR, luxI and AHL-mediated cell-cell communication were shown dispensable for population oscillations. We conclude that LuxR* plays no role in oscillations given the truncation of the DNA-binding domain, insensitivity to AHL, and ability of circuits lacking it (ePop-lite) to oscillate.

Oscillations require at least one negative feedback coupled with time delay (Figure 39A). In our circuits, time delay can be accounted for by the non-instantaneous rates of reactions affecting the level and function of the E protein, including transcription, translation, cell killing by the E protein, and E protein degradation. It is unclear, however, what control mechanism established feedback in ePop given the non-functional LuxR. Because the minimal oscillatory circuit was determined to be the PluxI promoter preceding the E gene, density-sensing must occur at either the level of E production or at the level of E protein activity (Figure 39B).
We hypothesized that regulation of mRNA production was a major cause of oscillations. This hypothesis predicts the existence of some signal X, whose level is a function of cell density, which either induces gene expression from the \( P_\text{luxI} \) promoter or modulates gene dosage. Several candidates for X might be capable of inducing \( E \) gene transcription, such as the host cell’s quorum sensing systems. Although \( E. coli \) do not possess a \( luxI \) homolog, their genome contains a \( luxR \) homolog, \( sdiA \), which might
activate the $P_{lux}$ promoter. Two lines of evidence argued against this explanation. First, cells containing the ePop circuit did not respond to AHL (Figure 38, bottom), which has been shown to bind SdiA [205] and activate gene expression. Second, oscillations were not significantly affected by indole (2mM), a stationary phase signal that has been reported to act through SdiA [206]. *E. coli* possesses a second quorum sensing system, the *luxS*-based AI-2 system [207], whose signal is derived from 4, 5-Dihydroxy-2, 3-pentanedione (DPD). Enzymatically produced DPD had a minor effect on oscillation dynamics, but did not inhibit growth in the manner that might be expected were it the feedback signal X (Figure 40).

![Figure 40: Effect of DPD on oscillations. Cultures treded with DPD exhibited one of two phenotypes in response. Cells either oscillated with a similar period but recovered more quickly from an initial round of lysis (left) or had oscillation period significantly increased by DPD (right). The differences between the two types of colonies, presumably of genetic origin, have not been determined. DPD and AI-2 did not appear to be the feedback signal X, however, because neither prevented growth or caused increased lysis.](image)

It has been shown that cyclic adenosine monophosphate (cAMP) and cAMP receptor protein (CRP) play a significant role in *V. fischeri* luminescence [208,209]. In its
natural genetic context, the lux promoter is located between the divergently expressed genes luxR and luxI. The literature suggests that cAMP-CRP activates the left-operon (luxR) and represses the right operon (luxICDAEB). In ePop, the original CRP site is present and the E gene is located at the position of right operon. E expression and lysis was therefore predicted to be repressed by cAMP and stimulated by glucose (which reduces cAMP concentrations). Paradoxically, glucose supplementation abolished oscillations and allowed elevated culture densities (Figure 39C, blue lines). cAMP supplementation only slightly increased oscillation period and lysis density (Figure 39C, red lines). When added concurrently with glucose, cAMP limited growth to intermediate densities, and some cycles of lysis were recovered (Figure 39C, green lines); cAMP thus seemed necessary for oscillations to occur. Nevertheless, exogenous cAMP neither prevented growth, nor increased lysis severity when added in the absence of glucose. These data suggest that cAMP is required for the manifestation of oscillations but is not the feedback signal X. Furthermore, these results were inconsistent with the previously reported promoter-level effects of cAMP on transcription in the lux operon.
Figure 41: Feedback is not at the promoter level. A promoter deletion series demonstrates that removal of the cAMP receptor site ($\Delta$CRP), Lux box ($\Delta$LUX), or both ($\Delta$CRP/LUX) did not abolish oscillations. Deletion of the full promoter ($\Delta$PROMOTER) did abolish oscillations, but this can be explained by deletion of the RNA polymerase binding site and ribosome binding site ($\Delta$RBS).

To address whether glucose and cAMP operated at the promoter level, as well as whether specific promoter regions required for oscillations could be isolated, a promoter deletion series was constructed (Figure 41). ePop constructs that removed the lux box, cAMP site, or both maintained oscillations. Only those that deleted the core promoter site and the ribosome binding site abolished oscillations. These results demonstrate that transcriptional activation from upstream promoter elements, such as the lux box or CRP site, were not required for oscillations.
5.2.3 Plasmid Amplification Can Account for Negative Feedback

In the apparent absence of transcriptional regulation, we surmised that changes in gene dosage at higher cell density could cause oscillations. Modulation of DNA levels has been implicated in effects of iron deprivation on luminescence in *V. fischeri* as well as *E. coli* transformed with the *lux* operon from *V. fischeri* [208]. To investigate plasmid amplification, MC4100z1 cells carrying plasmid pNewTet.E, which is a backbone vector of ePop containing a resistance marker and origin of replication identical to that of ePop, were grown in the presence and absence of glucose. Plasmid content at various time points was measured by the miniprep yield of culture samples resuspended to similar ODs. Final values were normalized to OD readings to account for slight difference in the density of resuspended cultures. pNewTet.E was used rather than ePop because cells in the process of lysing are fragile and cannot withstand the miniprep protocol. Given the identical plasmid backbone, plasmid amplification should impact copy number similarly. This allows pNewTet.E to serve as proxy for ePop for the purposes of addressing copy number as a function of culture phase under different media conditions. Miniprep yield/OD was found to increase upon transition to stationary phase in the LB grown cultures, but not those supplemented with glucose (Figure 42). Together with the existing understanding of ColE1 plasmid copy number control, this data suggested a mechanism for both density sensing and inhibition of oscillations by glucose – plasmid amplification.
Figure 42: Plasmid level as a function of growth phase and media composition. OD (triangles) and miniprep yield (circles) from cells grown in LB (open symbols, blue) and LB + 0.2% glucose (closed symbols, red). Miniprep yield increases upon entry to stationary phase in the LB culture, but not in the glucose supplemented culture.

5.2.4 ColE1 Replication

Regulation of replication in ColE1-type plasmids has been well characterized, for reviews see [210,211]. Briefly, the origin codes for two constitutively expressed regulatory RNAs. One is an RNA primer (RNA II) that initiates replication after recognition and cleavage by RNase H. The other (RNA I) is short-lived, more highly expressed, antisense RNA that associates with RNA II and inhibits processing by RNase H, and therefore replication. In ColE1-type origins that lack the Rom/Rop protein (like ePop), basal copy number is increased because Rom/Rop stabilizes the RNA I / RNA II interaction. This feedback provides a means of copy number control because levels of RNA I increase with copy number and serve to inhibit further replication (Figure 43). A consequence of
this mechanism is that copy number is very dependent on factors that affect RNA I production and degradation.

Relevant to the ePop dynamics, studies have identified several environmental factors that increase ColE1 copy number. Chloramphenicol causes the arrest of protein synthesis and chromosomal replication while stimulating replication from the ColE1 origin resulting in drastic plasmid amplification [212]. Amino acid starvation can result in amplification similar to chloramphenicol and was found to be particularly pronounced in relaxed strains [213] invoking a role for the stringent response mediators (p)ppGpp, relA, and spoT. Also, growth rate is inversely related to copy number across cells strains as well as by altering media composition for a given cell strain [214,215]. This relationship is observed in both relA+ and relA strains (Lin-Chao & Bremer, 1986) and is somewhat enhanced for plasmid lacking the Rom/Rop protein [216]. Finally, IPTG can lead to increased copy number of ColE1 plasmids, but only when protein expression caused a decreased growth rate and drop in ppGpp levels [217].
Figure 43: Stringent control, plasmid replication, and a possible mechanistic link. The host stringent response prepares E. coli cells for prolonged periods of nutritional limitation through the control of ppGpp levels [56,57]. Although the ppGpp response is multifaceted, for simplicity, only the regulation of tRNAs is shown. In wild-type cells ppGpp is either produced by RelA as a consequence of uncharged tRNAs resulting from amino acid starvation, or by SpoT in response to other nutritional stresses. However, because MC4100 cells are relaxed (relA1 allele) ppGpp is not produced in response to amino acid starvation and uncharged tRNA levels can accumulate to a greater degree. Uncharged tRNAs have been shown to degrade RNA I, the negative regulator of plasmid replication, in vitro, and lead to plasmid amplification when overexpressed in vivo [218]. Indeed, relaxed hosts experience ColE1 plasmid amplification when starved for amino acids. Our observations on ePop are consistent with a model where uncharged tRNAs accumulate and plasmid is amplified at high cell density and nutrient limitation. Low nutrient goes unacknowledged by the cell because RelA is not present to sense uncharged tRNAs and chloramphenicol is present to inhibit ppGpp accumulation. Plasmid amplification leads to increased E expression, cell lysis, decreased population density, and subsequent release of nutrient limitation. Although this model can account for the observations, we cannot exclude the possibility that other interactions exist to provide alternate or additional linkage between host metabolism and plasmid replication. Details on a mechanistic link between environmental factors and ColE1 replication have recently emerged and suggest that uncharged tRNAs directly catalyze the degradation of RNA I thereby dis-inhibiting replication [218]. Accumulation of uncharged tRNAs can be mitigated by the stringent response and ppGpp accumulation, which is perturbed in MC4100z1 [219], and prevented by chloramphenicol respectively [220,221].
5.2.5 Construction of a Mathematic Model that Captures ePop Dynamics

The complexity and number of inputs inherent to the both the stringent response and ColE1-type plasmid regulation (only partially encompassed in Figure 43) challenge the construction and parameterization of a comprehensive and accurate mathematical model. To this end, we have constructed a drastically simplified model to capture the key aspects of observed dynamics. The purpose of our model is not to intricately simulate ColE1 copy number control or its interaction with metabolism, and more detailed models of these types have been built [222,223,224,225,226]. Rather the model provides a simplified framework to interrogate effects of the perturbations on plasmid copy number amplification with regards to ePop oscillations.

![Diagram](image_url)

**Figure 44:** A simplified model for ePop function. Solid lines indicate positive and negative regulation. Dashed lines represent the effect of cell growth on component dilution.

- a. Increasing cells density causes RNA I degradation (possibly through uncharged tRNAs).
- b. RNA I inhibits plasmid replication (through its interaction with RNA II).
- c. RNA I is produced from the origin; elevated plasmid levels increase RNA I production.
- d. E protein is produced from ePop plasmid by basal expression from the luxI promoter in the absence of functional LuxR. Elevated plasmid levels increase E protein production.
- e. E protein decreases cell density by blocking cell-wall synthesis and lysing cells.
We propose the following: *E. coli* growth proceeds until culture density is sufficiently high. At high cell density, RNA I levels and division rate decrease, causing plasmid amplification, leading to increased basal production of E. Lysis of the majority of the cells results in nutrient release and loss of cell density signals allowing growth recovery and a subsequent cycle of growth and lysis. A schematic of these major reactions is shown (Figure 44). A system of ordinary differential equations describes key reactions for the four basic model components: cell growth and killing, E protein accumulation, plasmid amplification, and modulation of RNA I levels (Figure 45A). By broadly considering the impact of cell density on RNA I, the model is equivalently valid whether uncharged tRNAs provides the feedback or some other molecular mechanisms are involved in coupling cell density to plasmid amplification.
Figure 45: Simplified model equations and parameters. (A) Dimensionless ODE model of the circuit. Changes in cell density ($n$) are modeled as logistic growth with an intrinsic growth rate, $\alpha$. We assume that killing of cells by the E protein is cooperative and describe it using a Hill-type function (Hill coefficient, $p$). We note that cooperatively of E protein-mediated killing is not required for generating oscillations. The E protein is produced from a plasmid ($y$) with a rate $\beta_1$ and degraded with a rate $\gamma_1$; both processes follow first-order kinetics with regards to the amount of plasmid and E protein, respectively. Plasmid replication is inhibited by RNA I ($s$), and replication inhibition follows a power of hyperbolic function where $r$ is the effective number of reaction steps in the inhibitory scheme [225]. $\beta_2$ sets the maximum plasmid replication rate and $\gamma_2$ the intrinsic decay rate. RNA I is produced from the plasmid with a rate $\beta_3$ whereas its degradation rate is dependent on the cell density. Degradation of RNA I is described by a Hill-type function (Hill coefficient, $v$) to account for possible cooperativity. E protein, plasmid and RNA II are subject to dilution with cell growth. (B) The base parameter set that can generate sustained oscillations. Rate coefficients are normalized to a maximum killing rate (i.e. the maximum cell killing rate by E protein is 1). Biologically relevant parameter values have been chosen to illustrate the basic dynamics. E protein production rate $\beta_1$ is set to be small to reflect leaky expression. Plasmid decay rate $\gamma_3$ is set small to reflect the stability of plasmid molecules, and under oscillatory conditions plasmid dilution dominates.
5.2.6 Perturbations that Modulate Copy Number Impact Oscillations

Parameterized appropriately (Figure 45B), the model can generate sustained or damped oscillations in the absence of promoter-level or post-transcriptional regulation. It cannot be determined from batch culture experiments whether the oscillations caused by ePop are damped or sustained because conditions change as nutrients are depleted. However, the model can indicate parametric space that promotes sustained oscillations. A key parameter that determines the strength of feedback between cell density and copy number control is $\delta_1$, the half-maximal constant for RNAI cleavage by cell density. Bifurcation analysis indicates that oscillations become sustained for when $\delta_1$ is sufficiently small (Figure 46).

![Figure 46: Impact of $\delta_1$ value on oscillations. Bifurcation diagram showing a region of sustained oscillations over varying ‘half-maximal constant for RNAI cleavage’ ($\delta_1$). Insets show simulated time courses of cell density for three $\delta_1$ values. Damped oscillations can be generated outside the bifurcation region.](image)
Given the established link between chloramphenicol and plasmid amplification (Figure 43), we hypothesized that if ePop oscillated through a plasmid amplification mechanism, increasing chloramphenicol should mirror the effect of decreasing $\delta_1$. According to our model, decreasing $\delta_1$ would lead to an increasing number of cycles of damped oscillations (Figure 47, left column) or sustained oscillations (corresponding to an infinite number of cycles, Figure 46). Experimentally, we indeed observed an increasing number of oscillations increasing chloramphenicol concentrations (Figure 47, right column), consistent with the model prediction.
Similarly, the model provides an interpretation for the effects of glucose. If $E$ synthesis rate ($\beta_1$) is set to zero, lysis does not occur and the model reduces to a simplified treatment CoIE1 copy number control. Increasing $\delta_1$ in this context minimizes the plasmid amplification that occurs as stationary phase is reached and mirrors the impact of glucose on the accumulation of the ePop backbone pNewTet.E (Figure 48). If

Figure 47: Model predictions and experimental responses to system perturbations. Model predictions of increasing $\delta_1$ outside the bifurcation region on oscillations (left) match the result of decreasing chloramphenicol concentration (right), providing further support for the plasmid amplification mechanism. All chloramphenicol concentrations tested completely inhibited the growth of MC4100z1 cells and are therefore sufficient to prevent the growth of plasmid free segregates.
we consider the impact of glucose on ePop dynamics as being mediated through an increase in δi then the observation that glucose abolishes oscillations (Figure 39C) is not surprising. On a molecular level it is more difficult to attribute the precise role played by glucose, which could exert an impact through growth rate, ppGpp production/degradation by SpoT, overall energetic state or any number of host pathways under catabolite control.

![Graphs showing data analysis](image)

**Figure 48: Impact of chloramphenicol on plasmid amplification in pNewTet.E.** Experimental data of plasmid amplification (plotted as DNA/cell as a function of OD) demonstrate adding glucose or lowering chloramphenicol concentrations have the apparent effect of increasing δi. Glucose when present was used at 0.2% and chloramphenicol concentrations were 30.6μg/mL or 106 μg/mL. Coloring of traces is meant to demonstrate the trend and should not imply a direct quantitative agreement of specific model values with specific culturing conditions. The two traces at 106 μg/mL are from the same data as Figure 42.
5.3 Discussion

Given the frequent application of the quorum sensing in gene circuits [37,62,65,227,228,229,230,231], it is perhaps surprising that quorum sensing-like behavior resulting from plasmid amplification has not been described or encountered in synthetic systems until now. One possible explanation is due to differences in the threshold at which downstream genes become effective, underscoring the importance of matching the dynamic range of input/output elements in a circuit (Figure 49). Cell density induced increases in mRNA production should result in comparable expression of either the E protein or a typical reporter protein, such as the green fluorescent protein (GFP). Whereas low levels of the E protein can cause loss of viability [232], at the same level, the reporter protein may be below the detection threshold of standard methods [197]. Our results also suggest that plasmid amplification is most pronounced when the ColE1 origin and chloramphenicol are used in tandem. For this reason the chloramphenicol resistance marker could be considered partially incompatible with the ColE1 origin when minimal plasmid copy number change is desired. In addition, it appears that culturing conditions and cell strain have a significant impact on oscillations. Perhaps the confluence of needed conditions for significant plasmid amplification was not met in the past. In this regards, ePop itself can be valuable as a probe of cell physiology to interrogate what culturing and genetic conditions must be met for plasmid amplification. For example, observations on the impact of glucose on ePop oscillations
motivated the experiments demonstrating that glucose inhibited plasmid amplification (Figure 42). As a probe, ePop has the advantage of a simple observable (OD), high sensitivity, and rich information from complex dynamics. A better understanding of the factors that influence ColE1 control in a particular genetic and environmental context will enable the exploitation of plasmid amplification as a feature in future circuits.

The ease with which plasmids are isolated, genetically engineered, and re-introduced has led to their near ubiquitous use in synthetic biology, despite the potential consequences of variable copy number. Theoretical work has shown that even small changes to copy number values can result in significant non-linear effects on simple network motifs [233]. Nevertheless, plasmid copy number is often ignored under the assumption that any variation (basal or dynamic) will not have a large impact on the overall performance of the genetic device. In the case of ePop, this assumption was violated as plasmid amplification appears to be the cause of population oscillations. Placement of a circuit subsystem like copy control in a black box while ignoring molecular details is not faulty *per se*, and often necessary in order to initiate the engineering of large-scale systems [20]. However, it is most appropriate when the input/output relationship of the boxed subsystem is well defined. Unfortunately, no equation exists that takes any combination of strain genome, media composition, temperature, and chloramphenicol concentration into account while accurately returning copy number as a function of cell density.
Figure 49: Gate matching and unexpected feedbacks. (top) Total PluxI activity is the combination of plasmid amplification and quorum sensing. In ePop, defective LuxR prevents the contribution from quorum sensing - leaving only that of plasmid amplification. Low PluxI activity is sufficient to cause lysis, due to the extreme toxicity of the E gene. A more typical reporter used for promoter characterization (such as GFP) may be undetectable at this level, causing the effects of plasmid amplification to be missed. (bottom) A gene circuit can be designed as an open loop to process a series of inputs into defined outputs. When circuits are placed into host cells, however, hidden interactions between circuit and cellular components can introduce feedback that significantly impacts circuit dynamics. In ePop, the interaction between cell density and plasmid amplification is an unanticipated feedback that allows the circuit output (cell density) to serve as an input by modulating gene dosage.

ePop therefore underscores a fundamental challenge in standardizing cellular parts for synthetic biology [103,195]. The value of information from standardized parts is critically dependent on how closely characterization conditions mirror implementation conditions. When choosing parts for engineering a circuit, literature data are the primary resource, but such data are often available for the parts only in their natural context or a narrow range of characterization conditions. Because biological parts are influenced by,
and exert their influence through, interactions with other parts and their host cell, their behavior likely changes with their context. For example, even identical parts and network motifs can behave differently depending on their physical DNA configuration [234]. Circuit-host interactions (Figure 49) can drastically influence dynamics as evidenced by this work and the recent example of circuit-induced growth retardation leading to bistability [235]. Furthermore, it is impossible to assume that all of a part’s functions and interactions have been determined, even for well-characterized systems. While the definition of “standard” biological parts and the concept of parts abstraction and hierarchical composition can often simplify circuit design and analysis, these strategies can drastically underestimate the potential complexity of circuit dynamics. The likelihood that circuit behavior will deviate from predictions derived from characterization information scales non-linearly with the size and complexity of circuit as each new part can have unexpected interactions with every other circuit component and the host strain.

These issues call for greater emphasis on the development of methods to produce and monitor systems using parts for which data are incomplete, and whose behavior may change when placed in a new setting. Every circuit whose real-world behavior varies dramatically from our best predictions represents an opportunity to better understand the components, interactions, and control mechanisms of both the system and the host. The advantages of a standard parts-based approach should not obscure the
fundamental biological insight that can be gained from a more holistic analysis of synthetic systems and their emergent properties – especially when those systems “fail” to behave as anticipated.
5.4 Materials and Methods

5.4.1 Strains, Growth Conditions, Chemicals and Media

MC4100Z1 cells were the gift of M Elowitz. BW25113 and MG1655 were obtained from the *E. coli* stock center. Unless otherwise noted Luria Broth (LB) buffered with 100mM MOPS (pH = 7.0) was used for cell growth. 3oxoC6 homoserine lactone (AHL) was synthesized by the Duke Small Molecule Synthesis Facility. DPD was produced using *pfs* and *luxS* genes cloned from MG1655, expressed in BL21 (DE3) cells using pET-21 and purified using an immobilized metal affinity chromatography column (GE Healthcare). The enzymatic reaction was performed as previously described [236] and 5,5'-Dithio-bis(2-nitrobenzoic acid) (DTNB, Ellman’s reagent) was used to calculate yield on the basis of homocysteine released. All other chemicals were purchased from Sigma-Aldrich.

5.4.2 Liquid Culture Monitoring

Cells were grown in 48 well Tissue Culture Plates (Falcon 353078, BD Labware) inside a Perkin-Elmer VICTOR3 plate reader heated to 37°C unless otherwise noted. Wells containing 500µL fresh media and the appropriate chemicals were inoculated with 50µL of a starter culture (grown aerobically in a 5mL culture tube for 2-4h from glycerol stocks or fresh colonies). 250µL mineral oil was layered above each well to prevent evaporation over the course of the experiment. Care was taken to ensure starter cultures did not reach densities high enough to induce lysis. Prior to each measurement, plates
were shaken for 5 sec in an orbital pattern. OD at 600nm was measured every 15 min. Data presented are the raw OD values that result from 500uL culture in the plates used. They have not been modified to account for path length.

5.4.3 Plasmids

ePop (ColE1 origin, chloramphenicol®) was constructed using the lux box region (140 bp upstream of luxI in V. fischeri) from pluxGFPuv [37] and E gene coding sequence from φX174 (NEB). Each region was PCR-amplified and then joined together in an overlap PCR reaction. The ‘lux box-E gene’ fragment was inserted into the AatII site of host vector pLuxRI2 [62]. ePop-mini was constructed by inserting the lux box-E gene fragment (AatII digest of ePop) into pLuxR2 [62] at the AatII site. ePop-lite was constructed by digesting ePop with HindIII (deleting most of luxR and all of luxI) and re-ligating the larger fragment from an agarose gel purification. Promoter deletion mutants were generated using divergent primers flanking the deletion of interest. Each primer introduced a terminal NheI site. Digestion and intramolecular ligation generated plasmids where the region targeted for deletion was replaced by a 6 base pair NheI “scar”. pNewTet.E was generated by introducing a 40bp oligo with NdeI and NheI sites into the AatII site of pProTet.E (Clontech).

5.4.4 Plasmid Content Determination

A starter culture of MC4100z1 cells carrying the pNewTet.E plasmid was grown until it had entered early logistic growth (OD = 0.28). 750μL of starter culture was added
to 2L of LB or LB + glucose (0.2% final concentration) in 6L non-baffled flasks and grown at 37C and 180rpm. Chloramphenicol was added to the media to a final concentration of 30.6 µg/mL or 106µg/mL. An equivalent OD of cells was removed at various time points using a previously determined calibration curve to account of the non-linearity of OD measurements. Cultures were centrifuged for 10min at 5000g and re-suspended in 2mL Tris-buffered Saline. 240µL of the resuspended culture was added to 2mL LB and OD was recorded. Resuspended OD values were all within 20% of one another, enabling linear normalization. The remaining re-suspended culture was used to perform triplicate 500µL minipreps using the Zyppy miniprep kit (Zymo Research). Miniprep yield was measured on a ND-1000 UV spectrophotometer (Nanodrop).

5.4.5 Solid Phase Monitoring

Microscope slides were prepared by allowing warm LB agarose (1%) containing the appropriate antibiotics and inducers (1mM IPTG, 100ng/uL aTc) to be drawn under a suspended cover slip. After cooling, the coverslip was removed and the agarose pad cut to size with a clean razor blade. 1 µL low density starter culture (BW25113, ePop) was placed on the agarose and a fresh coverslip was applied. Slides were sealed with mineral oil and nailpolish to prevent evaporation. Cells were placed on a Leica inverted microscope in an environmentally controlled enclosure set to 37 °C. After 1-2 hours, a region for monitoring was selected on the basis of exhibiting cell division. Images were
captured using a Hamamatsu 1384 ORCA-ERA camera at 4 frames per minute overnight using autofocus feature to maintain the focal plane.
References


http://dezymerwiki.biochem.duke.edu/wiki/index.php/Manuals


Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system. BIOCHEMISTRY 37: 6718-6726.


Biography

Philippe Robert Marguet was born in Grenoble, France on May 26, 1983. He earned a B.A. in Biology with a concentration in Biochemistry from Cornell University in 2005. While at Cornell, Philippe performed ribozyme research under the direction of Dr. David Usher and was an active member of Chi Phi fraternity and the Cornell University Sprint Football team. It was also at Cornell that he met the North Carolina native who would later become his wife. Following graduation, Philippe matriculated to Duke in the Cell and Molecular Biology training program and was awarded a James B. Duke fellowship. The dissertation research described here was performed under the supervision of Dr. Homme Hellinga in the Biochemistry department and Dr. Lingchong You in the Biomedical Engineering department.

Publications:

