

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Engineering Modular Post-Translational Control Strategies in Prokaryotes

Permalink

<https://escholarship.org/uc/item/5009j96w>

Author

Whitaker, Weston Robert

Publication Date

2012

Peer reviewed|Thesis/dissertation

Engineering Modular Post-Translational Control Strategies in Prokaryotes

By

Weston Robert Whitaker

A dissertation submitted in partial satisfaction of the requirements for the degree of

Joint Doctor of Philosophy

with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Adam Arkin, Co-Chair
Professor John Dueber, Co-Chair
Professor Kathleen Ryan
Professor Wendell Lim

Fall 2012

Abstract

Engineering Modular Post-Translational Control Strategies in Prokaryotes

by

Weston Robert Whitaker

Joint Doctor of Philosophy in Bioengineering with UCSF

University of California, Berkeley

Professor Adam Arkin, Co-Chair

Professor John Dueber, Co-Chair

Synthetic Biology seeks to apply engineering principles to the design of biological systems, such that new behaviors can be predictably and reliably designed. Increasingly, systems are being engineered with well-characterized, modular, independently tunable parts. This bottom-up approach allows us to test design principles as well as engineer systems that may be adapted for use in other contexts. Modular design, in which systems are connected through a small number of well-defined interactions, is an important engineering principle that aids the design of complex systems. As we work toward the design of sophisticated systems necessary to exist in complex environments, it will become important to utilize many layers of regulation, as is done ubiquitously in natural systems. Post-translational control is relatively under-utilized in Synthetic Biology, in part due to the often complex interactions distributed across protein-protein interfaces that determine connectivity. Eukaryotic signal transduction pathways, however, are generally organized around modular domains and scaffolding proteins that improve evolvability by allowing reuse of signaling parts and potential for pathway rewiring via recombination¹. The power of modular pathway design has been demonstrated by recent successes in engineering eukaryotic scaffolds to serve as signaling hubs in order to create sophisticated behavior². Using well-characterized parts, we seek to apply the modular strategies evolved in higher organisms to control signal transduction in prokaryotic systems.

We begin our discussion by detailing the characteristics of biological parts families that are ideal for engineering. We then explain how protein-protein interaction parts that display these characteristics can be composed to create scaffolds capable of spatially organizing enzymes. Systems organized around scaffolds appear to function efficiently at relatively low component concentrations. We found that at low expression levels, the eukaryotic parts we employ often contain internal translation initiation sites that are selected against in prokaryotes and can be removed with the introduction of mRNA secondary

structure. Next we used simple synthetic scaffolds, built entirely from well-characterized protein-protein interaction motifs, to demonstrate that an increased local concentration effect, via tethering, is sufficient to direct signaling specificity. Scaffold-directed signaling was sensitive to expression levels of each component, thus additional regulation is necessary for robust behavior. We describe a strategy for introducing a peptide ligand into a polypeptide fold without perturbing activity that can be used for engineering an intra-molecular, autoinhibitory interaction. Robustness to varying component expression levels was gained by engineering an autoinhibitory interaction into the kinase protein such that the scaffold both colocalizes and activates components. Such domain-based allosteric regulation is frequently seen in natural eukaryotic signaling proteins. We also discuss a number of other strategies that used simple parts, which were able to improve some aspects of system behavior but not overall robustness. Interestingly, many natural systems that exhibit robust behavior make use of bifunctional components, and we found that designing a bifunctional scaffold (i.e. tethering and activation) improved robustness, while strategies that involved the expression of additional simple components failed to increase overall robustness. Taken together these results are among the first steps in designing highly modular, reliable signal transduction pathways in prokaryotes and demonstrate that principles governing pathway control of natural systems in higher organisms can be generalized and applied with well-characterized parts to prokaryotic systems.

Table of Contents

Table of Contents.....	i
Table of Figures.....	iv
Table of Tables.....	vii
Nomenclature.....	viii
Acknowledgements.....	ix
 Chapter 1 – Introduction.....	 1
1.1 – Modular design in Synthetic Biology.....	1
1.2 – Toward scalable parts families for predictable design of biological circuits.....	1
1.3 – Metabolic pathway flux enhancement by synthetic protein scaffolding.....	2
1.4 – Engineering robust control of two-component system phosphotransfer using modular scaffolds.....	2
1.5 – Internal ribosome binding sites within heterologous and synthetic protein coding sequences.....	3
1.6 – Additional findings and future directions.....	3
 Chapter 2 – Toward scalable parts families for predictable design of biological circuits. 4	
2.1 – Abstract.....	4
2.2 – Introduction.....	4
2.3 – Independence.....	7
2.4 – Reliability.....	8
2.5 – Tunability.....	9
2.6 – Orthogonality.....	9
2.7 – Composability.....	10
2.8 – Conclusions: the road to scalability and predictive design.....	11
2.9 – Acknowledgements.....	12
 Chapter 3 – Metabolic Pathway Flux Enhancement by Synthetic Protein Scaffolding... 13	
3.1 – Abstract.....	13
3.2 – Introduction.....	13
3.3 – Method—How to build modular protein scaffolded systems for metabolic engineering applications.....	17
3.3.1 – Selecting protein–protein interaction domains and ligands for scaffold construction.....	17
3.3.2 – Assembling scaffolds from domains and tagging enzymes for corecruitment.....	19
3.3.3 – Balancing the scaffold and enzyme concentrations.....	21
3.3.4 – Varying scaffold stoichiometry	23
3.3.5 – Scaffold composition effects.....	25
3.4 – Systems that May Benefit from Scaffolding.....	27
3.5 – Concluding Remarks.....	28
3.6 – Acknowledgements.....	28

Chapter 4 – Engineering Robust Control of Two-Component System Phosphotransfer Using Modular Scaffolds.....	29
4.1 – Abstract.....	29
4.2– Introduction.....	30
4.3 – Results.....	31
4.3.1 – Colocalization amplifies weak natural cross-talk.....	31
4.3.2 – Synthetic scaffolding directs histidine kinase specificity.....	34
4.3.3 – Specificity control with scaffolding is sensitive to component concentration.....	35
4.3.4 – Engineering autoinhibition for scaffold dependent activation of histidine kinase.....	37
4.3.5 – Autoinhibition improves robustness of scaffold-directed specificity.....	40
4.4 – Discussion.....	41
4.5 – Materials and Methods.....	43
4.5.1 – Strains.....	43
4.5.2 – Plasmid construction.....	43
4.5.3 – Culture method and fluorescence assays.....	44
4.5.4 – Estimating relative component concentrations.....	44
4.5.5 – Engineering colocalization and scaffolding machinery.....	45
4.5.6 – Engineering autoinhibition into Taz.....	45
4.6 – Acknowledgements.....	45
4.7 – Supplemental Information.....	46
 Chapter 5 – Internal Ribosome Binding Sites within Heterologous and Synthetic Protein Coding Sequences.....	 54
5.1 – Abstract.....	54
5.2 – Introduction.....	54
5.3 – Results and Discussion.....	55
5.3.1 – Negative selection against internal RBS sites within E. coli ORFs.....	55
5.3.2 – Internal RBS sites produce truncated proteins.....	57
5.3.3 – Balancing transcription and translation rates.....	59
5.4 – Discussion.....	61
5.5 – Materials and Methods.....	61
5.6 – Supplemental Figures.....	63
 Chapter 6 – Additional findings and future directions.....	 64
6.1 – Introduction.....	64
6.2 – The engineering and optimization process.....	64
6.2.1 – Key design variables.....	64
6.2.2 – Cloning strategy.....	66
6.2.3 – Mathematical modeling of to inform system design.....	68
6.2.4 – Optimization process.....	69

6.3 – Additional findings on two-component system signaling.....	73
6.3.1 – <i>In vitro</i> scaffolding results.....	74
6.3.2 – Broadening TCS phosphotransfer specificity.....	75
6.3.3 – Improving scaffold-directed phosphotransfer robustness.....	76
6.4 – Future directions.....	83
6.4.1 – Directions for improving scaffold-controlled signal transduction.....	83
6.4.2 – Expanding the utility of scaffold-controlled signal transduction.....	84
6.4.3 – Applications of scaffold-controlled signal transduction.....	84
6.5 – Conclusions.....	85
Chapter 7 – References.....	86
Chapter 8 – Appendix.....	100
8.1 – AUTOMODEL code.....	100
8.1.1 – Matlab code for “setup_MIEfig3.m”.....	100
8.1.2 – Matlab code for “varyX.m”	113
8.1.3 – Matlab code for “run.m”	114
8.2 – Example plasmid sequences.....	115
8.2.1 – pWW2179 : kinase plasmid: Pcon Taz-(SH3) ₄	115
8.2.2 – pWW2021: reporter plasmid: PBAD CusR-LZa, PCusR RFP, PRham -LZb, PCpxR RFP.....	119
8.2.3 – pWW2181: scaffold plasmid: Ptet peptide-LZB, PsaI peptide-LZA.....	127
8.2.4 – pWW2254: autoinhibited kinase plasmid: Ptet AiTaz-(SH3) ₄	132

Table of Figures

Figure 2.1. Biological circuits designed with parts from scalable families interconnecting specialty purpose parts.....	7
Figure 2.2. Specifics of a scalable parts family.....	11
Figure 3.1. Schematic of potential colocalization effects.....	16
Figure 3.2. Two consecutive enzymes in a synthetic mevalonate pathway were tethered, C to N termini, and the length of a simple synthetic linker of alternating Glycine and Serine residues was varied.....	21
Figure 3.3. A mathematical model of equilibrium binding of scaffolding recruiting enzymes shows an optimal scaffold concentration for maximizing full occupancy...	22
Figure 3.4. Mevalonate titers are measured at varying aTc concentrations, corresponding to induction of PTet driving peptide-tagged AtoB, HMGS and HMGR polycistronically.....	23
Figure 3.5. Three enzymes in the synthetic mevalonate pathway were tagged with binding peptides and recruited to a synthetic scaffolding protein with varying stoichiometry of binding domains.	24
Figure 3.6. A pathway engineered to produce glucaric acid was scaffolded with optimized stoichiometry to improve production.....	25
Figure 3.7. Mevalonate titers are measured for ligand tagged enzymes with expression of scaffolds of differing architecture and compared to the titers in the absence of scaffold expression.....	26
Figure 3.8. Schematic of a multimeric enzyme complex.....	27
Figure 4.1. Phosphotransfer to a noncognate response regulator can be directed by synthetic assembly with a histidine kinase.....	33
Figure 4.2. Phosphotransfer flux can be programmably directed with synthetic scaffolded HK/RR assemblies.....	35
Figure 4.3. Colocalization controlled phosphotransfer is sensitive to component concentrations.....	36

Figure 4.4. Engineering scaffold controlled activation of HK, via competitive autoinhibitory interaction, can be used to reduce non-targeted signaling.....	39
Figure 4.5. Autoinhibition decreased sensitivity to HK concentrations.....	41
Figure S4.1. Diagrams for plasmids used to demonstrate scaffold directed phosphotransfer.	46
Figure S4.2. Effect of native TCSs on in vivo crosstalk.....	46
Figure S4.3. Western blots for relative protein concentration estimation.....	47
Figure S4.4. Engineering an autoinhibitory SH3 interaction into Taz.....	48
Figure 5.1. Frequency of IRBS sites within natural and recoded sequences.	56
Figure 5.2. IRBS sequences can result in truncated proteins that are not inhibited by upstream translation.....	58
Figure 5.3. Optimizing for reduction of stochastic expression and IRBS translation requires balance of transcription and translation rates, particularly at low expression.....	60
Figure S5.1. Similar to Figure 5.2C, the PDZ IRBS is not inhibited by upstream translation.....	63
Figure 6.1. The importance of component concentrations.....	65
Figure 6.2. Nested BioBrick cloning strategy allows for easily composing domains and operons.....	66
Figure 6.3. Cells naturally make use of a wide range of protein expression levels.....	70
Figure 6.4. Two alternative strategies for concentration optimization.....	71
Figure 6.5. Varying the length of an unstructured glycine-serine linker.....	73
Figure 6.6. An in vitro phosphotransfer assay between colocalized non-cognate TCSs is shown.....	74
Figure 6.7. Attempting to create dually specific Taz.....	76
Figure 6.8. Engineering phosphatases to reduce undesired signaling.....	77
Figure 6.9. Balancing kinase and phosphatase expression improves colocalization- controlled phosphotransfer.....	78

Figure 6.10. Switching specificity via colocalization and the effect of phosphatase activity...	79
Figure 6.11. Expression of a competitive inhibitor can be used to expand the range of colocalization-controlled phosphotransfer.....	80
Figure 6.12. Competitive inhibitor reduces both colocalized and non-colocalized phosphotransfer.....	81
Figure 6.13. Colocalization of TCSs can serve as an assay for binding affinity.....	85

Table of Tables

Table 2.1. An informal glossary of synthetic biological part properties.....	6
Table 3.1. Protein-protein interaction domain families potentially useful for scaffold construction.....	19
Table S4.1 Summary of strains.....	50
Table S4.2 Primers of interest.....	50
Table S4.3 Protein-protein interaction machinery.....	51
Table S4.4 Summary of plasmids.....	52
Table S4.5 AiTaz characterization.....	53

Nomenclature

AiTaz	autoinhibited Taz histidine kinase
aTc	anhydrous tetracycline
GFP	green fluorescent protein
HK	histidine kinase
HMGR	HMG-CoA reductase
HMGS	HMG-CoA synthase
IRBS	internal ribosome binding site
OD	optical density: absorbance at 600 nm
ORF	open reading frame
P_{BAD}	arabinose-inducible araBAD promoter
P_{con}	constitutive promoter
P_{rham}	rhamnose-inducible rhaB promoter
P_{sal}	salicylate-inducible promoter
P_{tet}	tetracycline or anhydrous-tetracycline inducible promoter
RBS	ribosome binding site
RFP	red fluorescent protein (mRFP1)
RR	response regulator
SH3	SRC Homology 3 Domain
STP	Signal transduction pathway
TCS	two-component system

Acknowledgements

This work would not be possible without a number of people who have helped me. First I would like to thank my advisors Adam Arkin and John Dueber who were both essential to this work coming together. Adam Arkin has played a large part in guiding my understanding of Synthetic Biology. I feel lucky to have experienced the lab full of very diverse and intelligent people that Adam has brought together. John Dueber has taught me almost everything I know about doing research. The time and energy he spent mentoring me are very much appreciated. All the members of the Arkin, Anderson, Dueber and other labs here at Berkeley that I have been fortunate enough to know have provided a fun and collaborative atmosphere. My co-authors on the paper used in Chapter 2 Julius Lucks, Stanley Qi, and Adam Arkin were great to work with and I've learned a lot while discussing these ideas with them. Okoia Uket, Rami El-Kweifi and Aobo Wang did excellent work contributing to the data for Chapter 5. I enjoyed working with Alan Masinter, Jamie Yassif and Julius Lucks who helped me on this project early in my graduate work. Stephanie Davis really pushed the autoinhibition work in Chapter 4 forward and this dissertation is much better because of her help. I would also like to thank the Joint Graduate Group in Bioengineering, SynBERC and the NSF for funding this work.

Finally, I would like to thank my family and friends who have been very supportive while I worked on this. My parents have been very patient and encouraging, despite not being given any evidence that I would eventually finish graduate school. Doug Brownfield and Hayley Rucker in particular have helped me tremendously during my time at Berkeley and have also had more discussions with me about this project than I can say without embarrassment. I've enjoyed my time in graduate school and all of these people were a big part of that.

Chapter 1 – Introduction

1.1 – Modular design in Synthetic Biology

Synthetic biology seeks to apply engineering principles to the design of novel biological behavior. Although impressive successes have already been achieved in this field³⁻⁵, substantial work is needed to move the field in the direction of other engineering disciplines, in which complex systems can be rapidly designed and constructed with predictable and reliable behavior. Systems of modest complexity (e.g. ten genes) rarely function as intended when first tested and can take years of tinkering to achieve desired behavior⁵. Current efforts to improve this process are focused primarily around a parts-based approach, in which a set of parts with well-characterized functions are available and may be connected with other parts to interact with well-defined rules. Modular design is a central principle in this approach, with the goal of simplifying this process by working with context-independent parts or systems that interact with other parts or systems through a small number of well-defined interactions, rather than having a large number of subtle interactions following many components in the system. Towards this end, we aim to employ parts that are inherently less context dependent, such as those that have evolved to function in many different contexts, and design modular systems with connections that are robust and generalizable.

Cells have achieved impressive adaptive capabilities through the integration of transcriptional, translational, and post-translational control. The ability to rationally engineer post-translational processes will enable the design of more sophisticated systems when these are integrated with transcriptional and translational control. Post-translational control is particularly critical for applications that require fast time-scales or subcellular spatial localization. In this work we focus on the modular engineering of post-translational processes. We begin by discussing the qualities important to parts families and how we expect these parts to aid the engineering process. Next we discuss a methodology for taking the protein-protein interaction parts families and organizing them into scaffolds that serve as a modular control point for metabolic enzyme spatial organization. Building on this, we show that modular protein scaffolds can robustly and programmably control redirection of signal flow in prokaryotic two-component systems through both the activation of an autoinhibited kinase and colocalization with target response regulator. We discuss the unexpected occurrence of internal translation start sequences that are apparently common in the eukaryotic parts we employ. Finally, we discuss the design process and compare the strategies we employed with varying degrees of success to gain modular, specific and robust control of signaling pathways.

1.2 – Toward scalable parts families for predictable design of biological circuits

Chapter 2 discusses the importance of part families and argues that the critical properties to consider when choosing or developing parts families are: independence, reliability, tunability, orthogonality, composability, and scalability. Efforts to design biological

systems are often confounded by complex interactions between biological elements, resulting in system behavior that is difficult to predict. Employing parts families that meet the criteria discussed in Chapter 2 would be expected to significantly improve the predictability of design. Protein-protein interaction parts that fit all of the critical properties to some degree are discussed and employed in Chapters 3 and 4. Eukaryotic protein-protein interaction domains from natural signal transduction pathways provide a number of parts have been successfully employed in different context. These domains are thought to have evolved to be modular as they are naturally found in a number of different protein contexts yet maintain conserved fold and function. Chapter 3 lists parts families that exhibit independence, reliability, and composability, as confirmed by use in a number of synthetic biology applications. Several parts families, such as SH3, also exhibit a high degree of tunability. However, the scalability of these parts is primarily limited by the modest number of orthogonal parts. Although different binding domain families are highly orthogonal to other families, a larger number of orthogonal parts within each family would be beneficial and allow parts with similar folds and properties to be used more interchangeably. The rational engineering of orthogonal sets of leucine zipper has already provided a useful parts families for post-translational control⁶.

1.3 – Metabolic pathway flux enhancement by synthetic protein scaffolding

Chapter 3 discusses the use of synthetic scaffolds to serve as a modular control point for the spatial organization of enzymes. Colocalizaion of metabolic enzymes, via scaffolding, has been shown for several pathways to increase product titers^{7,8}. This is particularly beneficial for pathways that benefit from some degree of insulation from host processes, such as pathways that generate toxic intermediates. By employing protein-protein interaction with evolved modularity, which largely meets the criteria detailed in Chapter 2, scaffolds proteins of varying architecture can be easily generated. One of these scaffold sets has been directly applied to at least one other metabolic pathway and maintained colocalization function and even improved titers without the need to redesign or adjust the scaffolds for the new context, illustrating the benefit of modular design. Chapter 3 details some of the practical challenges and outlines a methodology for applying scaffold-directed spatial organization to metabolic pathways of interest. Many of these challenges hold for the application of scaffolding to signal transduction discussed in Chapter 4.

1.4 – Engineering robust control of two-component system phosphotransfer using modular scaffolds

Chapter 4 demonstrates the feasibility of achieving modular rewiring in post-translational circuits, an underutilized layer of control in prokaryotes. As discussed in Chapter 3, scaffold proteins, built from modular protein-protein interaction domains, can be used to specifically co-target proteins tagged with corresponding interaction ligands. By switching expression between two synthetic scaffolds, flux can be directed from a histidine kinase to either of two non-cognate response regulators. Finally, since this effect was highly sensitive to the balance of histidine kinase and response regulator concentrations, robustness to kinase

concentration was gained by designing an allosterically-regulated kinase switch such that the ligand on the scaffold serves the dual functions of directing assembly and switching activation. These results are first steps towards a generalizable strategy for designing modular prokaryotic signal transduction.

1.5 – Internal ribosome binding sites within heterologous and synthetic protein coding sequences

Chapters 3 and 4 describe the use of eukaryotic binding domains and express these domains at relatively low levels. Under these conditions we find that internal translation initiation sites are relatively common. In Chapter 5 we discuss internal ribosome binding sites (IRBSs) and investigate their prevalence, as estimated by a biophysical algorithm⁹, in natural and synthetic sequence. These IRBS sequences appear to be selected against in *E. coli*, suggesting a physiological relevance. Additionally, we demonstrate that these sites result in truncated protein products and that IRBS strength can be reduced by introduction of secondary mRNA structure. At the low expression levels used in Chapters 3 and 4 both transcription and translation must be balance to avoid stochastic transcription and truncated translation.

1.6 – Additional findings and future directions

During the process of adapting TCSs for scaffold-directed phosphotransfer a number of strategies were employed, with varying degrees of success. These strategies are summarized in Chapter 6. We discuss the design, construction and optimization process we employed to engineer the system used to demonstrate scaffold-directed phosphotransfer, with emphasis on parameters we found to be particularly important. We discuss preliminary results towards broadening specificity, results for two additional strategies intended to improving robustness, and hypothesize as to the most important distinguishing factors between the effective and ineffective strategies for improving robustness. Finally, we discuss future directions to advance the broad application of scaffold-directed TCS phosphotransfer.

Chapter 2 – Toward scalable parts families for predictable design of biological circuits

Authors:

Julius B Lucks

Department of Bioengineering, University of California, Berkeley CA, United States
The Miller Institute for Basic Research in Science, Berkeley CA, United States

Lei Qi

Department of Physics, University of California, Berkeley CA, United States

Weston R Whitaker

Department of Bioengineering, University of California, Berkeley CA, United States

Adam P Arkin (corresponding author: aparkin@lbl.gov)

Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley CA, United States

Reprinted from *Current Opinion in Microbiology*, JB Lucks, L Qi, WR Whitaker, AP Arkin, "Toward scalable parts families for predictable design of biological circuits", Vol. 11, 567–573, (2008) with permission from Elsevier.

2.1 – Abstract

Our current ability to engineer biological circuits is hindered by design cycles that are costly in terms of time and money, with constructs failing to operate as desired, or evolving away from the desired function once deployed. Synthetic biologists seek to understand biological design principles and use them to create technologies that increase the efficiency of the genetic engineering design cycle. Central to the approach is the creation of biological parts — encapsulated functions that can be composited together to create new pathways with predictable behaviors. We define five desirable characteristics of biological parts — independence, reliability, tunability, orthogonality and composability, and review studies of small natural and synthetic biological circuits that provide insights into each of these characteristics. We propose that the creation of appropriate sets of families of parts with these properties is a prerequisite for efficient, predictable engineering of new function in cells and will enable a large increase in the sophistication of genetic engineering applications.

2.2 – Introduction

Microbes are profoundly intertwined with our environment and our lives. They metabolize a wide variety of chemicals including elemental metals, and can synthesize an

equally broad array of molecules including complex organics and drugs. Microbial communities play a critical role in global cycles of important elements such as carbon, nitrogen, and sulfur. For these reasons, understanding the molecular basis of microbial metabolic mechanisms and their regulatory control has been a central goal of molecular biology.

It was recognized early on that once genes and networks responsible for the broad array of microbial function were identified and understood, they could be exploited for technological benefit. Bacteria have been genetically engineered to produce commodity chemicals, pharmaceuticals^{10,11}, and recently fuel molecules^{11–13}. Beyond the controlled culture environment of an industrial production bioreactor, bacteria have been cautiously engineered to aid in bioremediation¹⁴, to support agriculture¹⁵, and act as vehicles for macromolecular delivery¹⁶, immunotherapy¹⁷, and even cancer¹⁸. Despite its name, however, genetic engineering still remains an inefficient tool rather than an engineering science, and projects are plagued by multiple, costly cycles of design and testing as constructs fail to operate as desired¹⁹, or evolve away from the desired behavior once generated²⁰. To address these difficulties, synthetic biologists have recently designed and synthesized biological circuits aimed at uncovering, exploiting, and optimizing cellular components for use in predictable design and more safe and efficient construction of new complex function in organisms.

To construct biological circuits, synthetic biologists have focused on biological ‘parts’ — distinct encapsulations of biological function that may be wired together in different contexts to create new and sometimes predictable behaviors. A classic example that predates Synthetic Biology is the ability to attach promoters to arbitrary genes to express them in heterologous hosts which demonstrates a kind of discrete, rewirable containment of biological function that forms the basis of most metabolic engineering. In this case, a metabolic pathway may be viewed as a connection between enzyme gene parts that are used to produce specific metabolites, and gene expression control parts (promoters) that coordinate expression of the enzymes. Other examples of biological parts include regions of DNA (e.g. operator sites), RNA (e.g. ribosome binding sites), protein (e.g. domains), and even whole complex subsystems (e.g. secondary metabolic pathways) that can all be connected in synthetic circuits displaying myriad functions.

It might seem, then, that the major obstacle to engineering new function in microbes is assembling DNA encoding specific parts in the desired configuration into the cell. This is indeed a challenge, but technological advances in DNA synthesis²¹, and methods for transforming large DNA fragments into cells²² are rapidly solving them. Instead, the main challenge seems to be uncertainty: uncertainty in our understanding of the precise mechanisms of internal part operation and part–part interaction, uncertainty in our understanding of environmental influence on part function, uncertainty in how to identify and measure characteristic part properties relevant to circuit design, and uncertainty in our understanding of the environment in which the cell must operate and survive.

Term	Definition	Illustrative examples
Independence	The degree to which a part does not interfere with other parts in the host or with the host machinery, and vice versa.	Nitrogen fixation in different hosts ²³ , Multiple repressors in the same cell ^{24,25}
Reliability	The degree to which a part functions as intended with respect to variability in its components and its environment. Implies independence.	Small RNAs increasing noise ²⁶ , feedback loops decreasing noise ²⁷
Tunability	The degree to which the function of a part may be controllably adjusted.	Ribosome-binding sites ^{28–31} , mRNA stability elements ³² , histidine kinases ³³ , protein–protein-binding domains ³⁴
Orthogonality	The degree to which parts derived from a parent part can be tuned to the point of noninterference while maintaining the same basic conceptual function.	Designed ribosome–RBS pairs ³⁵
Composability	The degree to which parts can be combined together to form units with composite function.	Zinc finger domains ³⁶ , chimeric proteins ³⁷ , Aptamer–ribozyme composites ^{38,39} (Figure 2.2), protein-binding domains ⁴⁰ , repressor networks ⁴¹
Scalability	The degree to which the confluence of independence, reliability, tunability, orthogonality, and composability in a family of parts can be exploited to create many distinct, noninterfering instances of a function all otherwise having similar operating characteristics.	Aptamer–ribozyme composite parts ^{38,39} (Figure 2.2)

Table 2.1. An informal glossary of synthetic biological part properties. Below we suggest definitions of commonly used terms in synthetic biology^{11,42–45}. While a consensus for a formal definition of these terms is still being developed, we present one interpretation of the meaning of these emerging concepts.

Synthetic biologists seek to address these uncertainties directly as outlined in a number of reviews^{11,42–45}. At the heart of these efforts are attempts to develop families of characterized biological parts that would interoperate in a predictable manner, even in complex biological circuits and across a variety of biological environments. In this paper, we define five fundamental properties that we seek in optimal biological parts: independence, reliability, tunability, orthogonality, and composability, which when combined, lead to scalable parts families that can be used in synthetic biological design (Table 2.1, Figure 2.1). Here, we emphasize recent work that is beginning to uncover the principles behind each of these fundamental properties. We then discuss how we can move these studies forward to

produce a conceptual and physical infrastructure around a parts-based biological circuit design cycle that will create a dramatic increase in the efficiency of genetic engineering, and will improve our understanding of the principles behind biological design.

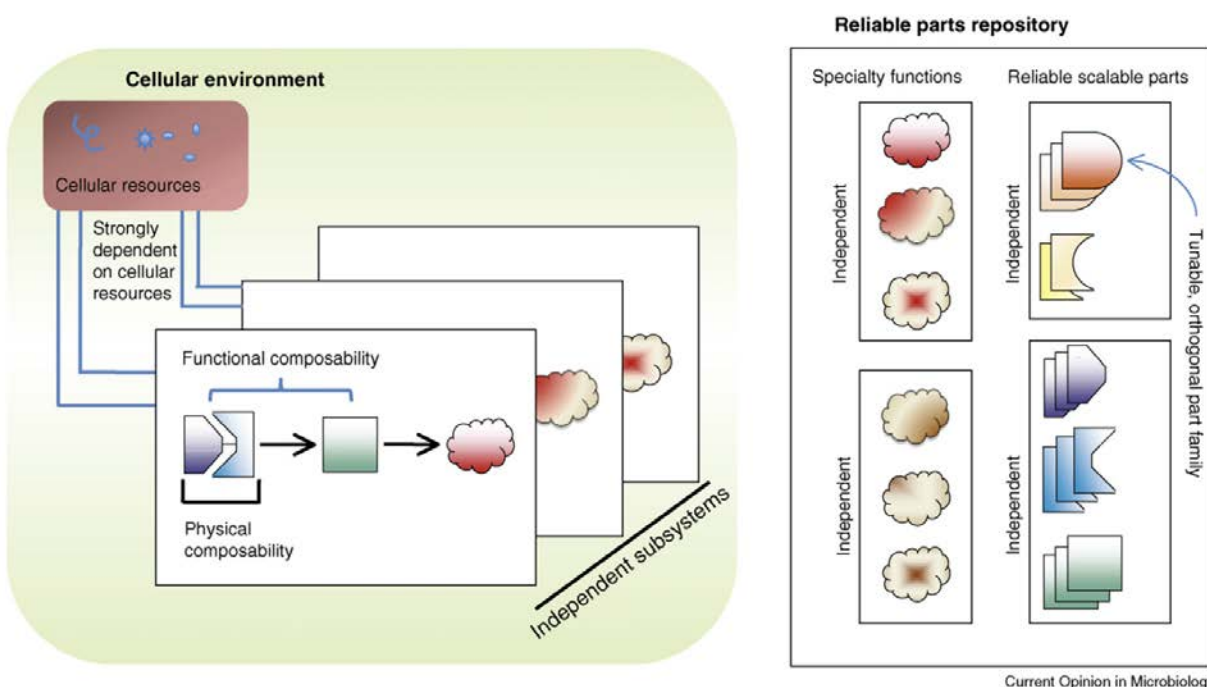


Figure 2.1. Biological circuits designed with parts from scalable families interconnecting specialty purpose parts. Multiple independent biological circuits are placed inside a single cellular environment (left panel), all sharing the same common pool of cellular resources and functioning more or less reliably with respect to variability in their components and the environment. The circuits are composed of interconnected parts (left panel, lines). Ideally, parts for circuits are drawn from a repository or knowledgebase (right panel) and consist of two types — parts that perform specialty activities, controlled by parts from scalable parts families (stacked groups). Within a single part family, there can be many orthogonal members indicated by shading. There can be many groups of scalable parts families, and those families whose parts function independently from each other are enclosed in rectangles.

2.3 – Independence

Independent parts do not interfere with their host circuitry and *vice versa* (Table 2.1). In an early example involving complex function, the multigene nitrogen fixation system from *Klebsiella pneumoniae* was shown to operate when transplanted into *E. coli*, albeit with somewhat diminished function²³. That is, the transformed *E. coli* was able to fix nitrogen, implying that the system can function with some degree of independence from the context of the host in which it evolved. Independent parts also do not interfere with each other. Repressors that affect different promoters and do not interact with each other are independent, for example. Part independence is far from guaranteed however. For example, different plasmid origins of replication can interfere with each other and form plasmid

incompatibility groups that prohibit multiple members of the same incompatibility group from stably coexisting in the same cell⁴⁶.

2.4 – Reliability

A reliable biological part functions as intended in a biological circuit in a suitable host (Table 2.1). Independence of part function is one aspect of reliability. Another aspect of reliability is robustness in the face of noise in the cellular circuitry and fluctuations in the cellular environment. Because of the fundamentally discrete and stochastic nature of chemistry, there is intrinsic noise in the dynamics of biochemical networks⁴⁷. These effects can be quite large in certain cellular systems such as those involved in gene expression, and lead to very different dynamics that would be predicted from the classical deterministic picture⁴⁸. In a study of noise propagation through gene expression cascades, the sharpness of the transition between off and on states, as well as the variability in the proportion of the cellular population that made the transition, was shown to increase with the length of the cascade⁴¹. As a possible route to engineering this aspect of part reliability, a number of articles have demonstrated that engineering different molecular features and feedback mechanisms into gene expression circuits can alter the noise profile of parts and circuits^{26,27,49}. This sort of intrinsic noise, however, does not always lead to unreliable function but can actually be a source of reliability^{50–52}. Much like diversifying a stock portfolio, intrinsic noise in physiological function can be leveraged to bank against uncertainty in the environment. Blake et al. built synthetic circuits to test this experimentally in which noisy promoters connected to an antibiotic resistance gene were shown to confer an advantage over more stable promoters for cells exposed to acute bursts of the antibiotic⁵³.

The function of a synthetic part can also be affected by cell-to-cell variation in key cellular resources required for transcription, translation, and replication that, in turn, can be affected by changes in the cellular environment. A recent example that illustrates this problem is a synthetic genetic AND gate that utilizes both control of transcription and translation and requires two inputs to turn on gene expression⁵⁴. The gate was shown to have a lower gain at low cell densities, an undesirable coupling to growth phase.

Finally, the load, either energetic or toxic, that a reliable part places on the host should be well understood and optimized so that it is not selected against over generations and does not add a diffuse 'metabolic' coupling among components in the cell. Addition of extra circuitry to a cell places an extra burden on the cell⁵⁵. Use of selective markers can maintain a burdensome part in the right environment. In the absence of such markers, mutants of the engineered cell that inactivate or rid the cell of the part will outgrow the original. For instance, in one recent study, You et al. discovered that cells started to escape their population-controlled cell death circuit three to six days after introducing the circuit into the cells⁵⁶, and Canton *et al.* found that the functioning of a gene expression controller at high induction decayed after 56 generations of being present inside the cells⁵⁷. While these examples may represent an actual resource load on the cell, other examples include high-production metabolic pathways that produce intermediates toxic to the cell⁵⁸. In this case,

reliable function may require additions and tuning of other parts such as enzymes immediately downstream of the toxic compound, or changing environmental conditions such as adding protective chemicals⁵⁹.

2.5 – Tunability

‘Tunability’ refers to the ability to make controlled adjustment to a part’s function (Table 2.1). For simple parts such as ribosome-binding sites (RBSs) on mRNA transcripts upstream of protein coding regions, this takes the form of varying the sequence of the RBSs to change both the structure of the transcript around the translation initiation site, and its interaction with the ribosome such that translational efficiency is affected. A number of groups have used this trick to adjust gene expression up to 1000 times the normal level^{28–30} and some have tried to model the effect³¹. In addition to RBSs, the tuning of mRNA degradation was studied by Carrier and Keasling, who constructed a library of mRNA stability structures shown to confer mRNA half-lives in the range of 2–20 min³². Similar approaches have highlighted the role of intergenic regions in tuning the expression of multiple genes within a single operon⁶⁰. The tuning of a molecular part function can, of course, alter the entire function of the circuit in which it is embedded. In early work, Gardner and Collins experimentally demonstrated that tuning RBSs in certain versions of a genetic toggle switch can affect whether the switch displays graded or bistable behavior²⁴. Dueber *et al.* exploited the differential affinity of SH3, PDZ, and GBD peptide-binding domains to different peptide targets to construct proteins that function conditionally in the presence or absence of multiple environmental inputs⁴⁰. Voigt *et al.* describe a theoretical ‘evolvable’ circuit motif in which it is possible to tune the strength of a promoter to switch the behavior of the circuit from a graded switch, to a bistable switch, to an oscillator, to a pulse generator⁶¹.

In some cases, tuning a part’s strength of function amounts to tuning its specificity toward its interaction target, and away from the multitude of other molecular species with which it could potentially interact (see Figure 2.2a). Skerker *et al.* demonstrated such specificity tuning by showing how the specific activity of a histidine kinase (HK) for a response regulator (RR) could be tuned by the mutation of rationally chosen amino acids³³. In some cases, it was possible to tune the interaction between two parts in the extreme and switch the specificity of one HK to a different RR and away from its natural partner. This type of tuning leads to parts whose specificities are so extremely tuned toward their target as to make them functionally orthogonal to each other.

2.6 – Orthogonality

Orthogonal parts families are derived from parent parts that can be tuned to the point of noninterference with each other, while maintaining the same basic conceptual function (Table 2.1). The HK/RR example mentioned above demonstrates the possibility of designing multiple synthetic independent/orthogonal HK/RR pairs. Similarly, Reina *et al.* were able to tune three different PDZ domains to bind to new targets two orders of magnitude higher than to their cognate peptides³⁴. Recently, there have been multiple examples illustrating the

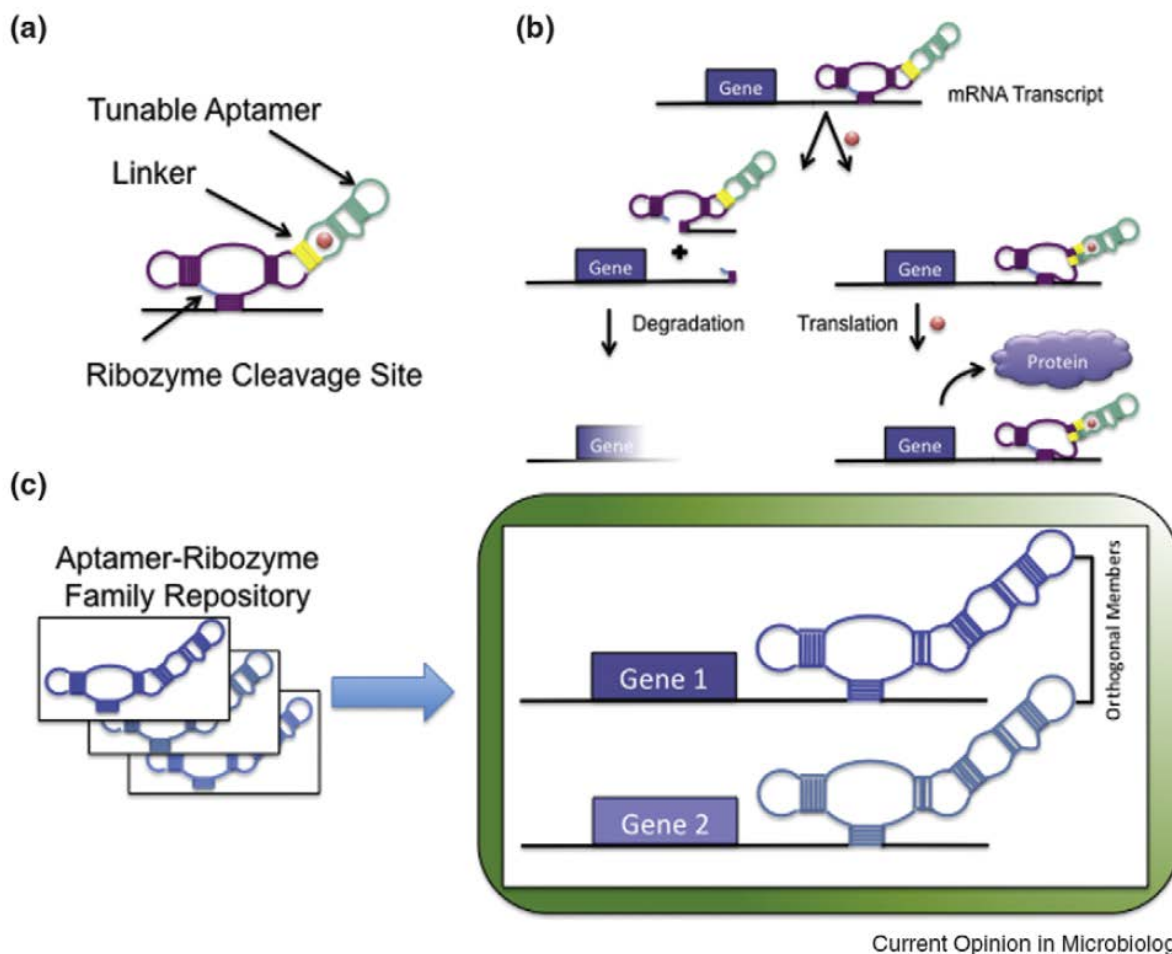
power of RNA designability for creating orthogonal parts families. By mutating the 16S ribosomal RNA sequence, Rackham and Chin created mutant ribosomes that each acted on a specific target RBS, independently of each other and the natural host ribosomes³⁵. Bayer and Smolke, and Isaacs *et al.* used RNA complementarity rules in their design of orthogonal translational locksystems that block translation initiation and can be unlocked by specific small molecules, or complementary RNA molecules expressed *in trans*^{62,63} [48,49]. Although achieved in mammalian cells, Rinaudo *et al.* were even able to exploit similar mechanisms to design RNAi logics that relied on orthogonally acting siRNA molecules to target multiple RNAi sites on mRNA transcripts to perform several combinations of Boolean operations controlling the expression of a reporter gene⁶⁴. Each of these examples demonstrates the ability to create families of function in which multiple parts from each family can be used in the same cell without interference.

2.7 – Composability

Composability is the property of certain parts that allows them to be combined to form a predictably functioning biological circuit (Table 2.1). Parts that are physically composable can be placed on the same molecule (DNA, mRNA transcript, protein, or membrane) thereby displaying a composite behavior predictable from the individual functions. Promoters, RBSs, and genes are one of the most familiar examples of physically composable parts that can be placed in series on a DNA molecule to control expression of the gene. However, these have to be carefully chosen to prevent the formation of interfering secondary structures on the transcript, among other things, that could lead to poor expression. In proteins, Mandell and Barbas have shown programmable specificity of synthetic zinc finger binding domains, which can be physically composed together to create proteins that bind to desired target DNA sequences³⁶. Physically composable parts can also be combined to form a composite part with chimeric function. Taz is a protein which combines the sensing domain from the aspartate receptor Tar from *E. coli*, with the kinase domain of EnvZ. EnvZ activity is normally regulated in response to osmolarity. *E. coli* cells containing Taz respond to the presence of aspartate with the activation of promoters that are normally regulated in response to osmolarity³⁷. Win and Smolke composed self-cleaving RNA ribozymes with small-molecule sensing RNA aptamers^{38,39,65} (see Figure 2.2). When the aptamer domain is placed in the middle of the ribozyme sequence, changes in conformation of the aptamer domain upon binding a specific small molecule, either allow, or prohibit ribozyme cleavage. When this composite part is composed downstream of a gene, cleavage by the ribozyme results in transcript destabilization, thus controlling gene expression. Proper functioning of this composite part required the creation of a linking element that could preserve the functioning of the individual parts while coupling their function together^{38,65}.

When two parts are not on the same molecule they can still be functionally composed. For example, a promoter– repressor encoded on one DNA molecule can couple to a promoter on another DNA molecule to repress its activity²⁵. This requires matching part–part interaction parameters, a currently heterogeneous process requiring a great deal of tuning. Anderson *et al.* resorted to screening RBS libraries to find one that allowed leakless, inducible expression

of a target gene above a target amount from a commonly used promoter over a desired range of inducer concentration³⁰.



Current Opinion in Microbiology

Figure 2.2. Specifics of a scalable parts family. (a) Self-cleaving ribozymes (purple) can be physically composed (yellow) to aptamer-sensing domains (green) which cause the inhibition of the self-cleaving reaction in the presence of a specific small molecule (red) (following³⁸). Aptamer sequences can be tuned to show specificity for different metabolites. (b) When the aptamer-ribozyme is placed downstream of a gene, a self-cleavage reaction in the absence of the small molecule destabilizes the transcript leading to degradation. Transcript destabilization is prohibited by the small molecule, which prevents self-cleavage through binding and RNA secondary structure rearrangement, enabling translation of the upstream gene. (c) Since they are part of the physical transcript, different aptamer-ribozymes composed behind different genes could act orthogonally. The combination of tunability, functional composability, and orthogonality make these parts an example of a scalable parts family for controlling gene expression³⁹.

2.8 – Conclusions: the road to scalability and predictive design

In most engineering disciplines, there are specialty parts and there are generic parts. The specialty parts carry out application-specific function such as LCD screens and CCD cameras, and they are interconnected by generic parts — key parts that are used in nearly every design, such as transistors in electronics. There are many specialty parts in biology, for example, enzymes and molecular machines such as photosynthesis, motility, protein secretion, and nitrogen fixation, evolved over billions of years. For now, we should essentially use these ‘as-is’, using our understanding of how to tune these parts⁶⁶ to match particular biological circuit designs.

In contrast, for the generic parts, we need an engineering science that can provide predictable and scalable design. The confluence of parts that are all independent, reliable, tunable, orthogonal, and generally composable, as outlined above, leads to scalable families of parts that can be readily combined together to form predictable and possibly complex new functions in cells (Figures 2.1 and 2.2). A central challenge of synthetic biology is how best to choose these scalable parts families to form a powerful basis set of biological function.

An initial effort ought to be the generation of scalable parts families that control transcription, translation, and the generic features of protein–protein interaction. These processes are central to nearly every application and generally provide the logic by which the application’s key activities are deployed. There is evidence that early success will come from nucleic-acid-based gene expression regulators where Watson–Crick base pairing rules are a good starting point for design^{28,32,35,38,39,60,62–64,67} (Figure 2.2). However, an organized program to characterize such parts in the seemingly immense and diverse number of contexts in which they may appear is needed to understand the fundamental principles behind part independence, reliability, tunability, orthogonality, and composability.

With sufficiently deep parts families covering a broad, but carefully selected array of function, we should have the tools to finally enable a predictable biological circuit design cycle, thereby dramatically increasing the efficiency, safety and sophistication of genetic engineering. This will make the small forays into design principles for pattern forming circuits⁶⁸, growth controllers⁵⁶, and other higher level designs⁵⁴ of greater general interest and use to other designers and scientists. Eventually, understanding these basic concepts will help us transit from the engineering of small biological circuits and pathways, to genome scale designs that operate beyond the bioreactor across the population, and ultimately ecological levels, all the while enabling a deeper identification and understanding of the design principles of biology.

2.9 – Acknowledgements

JBL acknowledges the support of the Miller Institute for Basic Scientific Research. JBL, LQ, WRW, and APA acknowledge the support of the Synthetic Biology Engineering Research Center under NSF grant number 04-570/0506186. The authors also thank Dustin Rubinstein, Jason Stajich, and Sera Young for their helpful comments during the preparation of the manuscript.

Chapter 3 – Metabolic Pathway Flux Enhancement by Synthetic Protein Scaffolding

Authors:

Weston R. Whitaker

Department of Bioengineering, University of California, Berkeley CA, United States

John E. Dueber (corresponding author: jdueber@berkeley.edu)

Department of Bioengineering, University of California, Berkeley CA, United States

Energy Biosciences Institute, University of California, Berkeley, CA, United States

Reprinted from *Methods in Enzymology*, WR Whitaker, JE Dueber, "Metabolic Pathway Flux Enhancement by Synthetic Protein Scaffolding" Vol. 497, 447-468, (2011) with permission from Elsevier.

3.1 – Abstract

Spatial control over enzyme organization presents a promising post-translational strategy for improving metabolic flux. Directly tethering enzyme polypeptides has had inconsistent success. Use of a separate scaffold molecule, built from modular protein-protein interaction domains, provides designable control over enzyme assembly parameters, including stoichiometry, as well as providing scalability for multiple enzymes. Thus, metabolic flux can be optimized by expression of these scaffolds in vivo. It is important to note that exploration of the use of synthetic scaffolds for improving metabolic flux is in its early stages. Accordingly, in this manuscript we describe efforts to date, hypotheses for scaffold function, and parameters to consider for application to new pathways.

3.2 – Introduction

Metabolic engineering has the potential to provide environmentally safe and cost-effective routes for synthesizing a range of compounds, from high-value specialty compounds such as therapeutics to bulk commodities including plastics and biofuels. Particularly for the latter class of compounds, a complement of strategies will be needed to achieve the production yields, near theoretical maximum, necessary to achieve industrial viability. These stringent requirements will likely inspire improvements across many technologies: modeling metabolic and cellular behavior⁶⁹, predictable control over gene expression^{9,38,60}, and directed evolution approaches for improved enzyme characteristics^{70,71}. In this manuscript, we focus on ongoing efforts to improve pathway efficiency through engineered enzyme complex formation using synthetic scaffolds; however, all strategies discussed here must eventually be performed in concert with existing proven methodologies to achieve optimal yields. Since our mechanistic understanding of scaffold function is still at

an early stage, we describe here the parameters empirically derived thus far to be important and describe a suggested process for applying scaffolding strategies to a new pathway.

There are numerous natural examples of enzymes forming complexes for optimal metabolic pathway performance. For excellent in-depth reviews on this topic, please see those written by Conrado *et al.*⁷² and Miles *et al.*⁷³. The most striking examples of improved pathway efficiency via complex formation are those that have evolved structures capable of physically channeling substrates. Tryptophan synthase, carbamoyl phosphate synthase, and glutamine phosphoribosylpyrophosphate amidotransferase are three examples described in detail by Miles *et al.* whose structures reveal tunnels connecting catalytic sites that are capable of protecting reactive intermediates from the bulk solution⁷³. Another mechanism of channeling substrates is through electrostatic channeling. Thymidylate synthase and dihydrofolate reductase are two enzymatic activities found in a single polypeptide in some plants and some protozoa, including *Leishmania major* for which the crystal structure has been solved. The surface of this structure is predominantly positively charged, suggesting a mechanism of an electrostatic “highway” spanning the 40Å between the two active sites across which the negatively charged dihydrofolate intermediate would travel⁷⁴. Recently, evidence has grown for the dynamic assembly of complexes, perhaps as a feedback mechanism to achieve a precise concentration of metabolite product^{75,76}. These dynamic complexes have been difficult to observe biochemically *in vitro*. For example, purine biosynthesis in eukaryotes involves six enzymes. Despite early anticipation of potential interactions between these enzymes, only recently was it understood, by fluorescently tagging these enzymes *in vivo*, that all six proteins co-assemble⁷⁶. Interestingly, these proteins dynamically assemble and disassemble depending on purine concentration. Narayanaswamy *et al.* similarly showed that numerous metabolic enzyme complexes dynamically assemble depending on culture conditions⁷⁵, suggesting these phenomena are considerably more common than would be predicted by *in vitro* biochemical experiments. Likely, many of these complexes are not detected due to characterization under conditions incompatible with complex formation.

Drawing inspiration from natural pathways, engineers have begun assembling synthetic enzyme complexes to improve pathway performance. For degradation of cellulose and hemicellulose *in vitro*, various enzyme combinations have been co-recruited to cellulose substrate to include synergistic combinations of activities as found in natural cellulosome complexes^{77,78}. Recently, our lab has expressed scaffolds built from modular protein-protein interaction domains (Table 2.1) to optimize flux of engineered metabolic pathways *in vivo*⁷. Enzymes were tagged with peptide ligands specific for these scaffold protein-protein interaction domains. The modular composition of the scaffolds was used to build various architectures that were critical for optimizing flux as discussed later.

The mevalonate biosynthetic pathway presents an interesting model system for synthetic complex engineering in that it suffers from a flux imbalance between HMG-CoA synthase (HMGs) and HMG-CoA reductase (HMGR) that results in the accumulation of the cytotoxic HMG-CoA intermediate^{58,79}. Scaffolding this pathway improves efficiency, producing

higher product titers even at considerably lower inducer concentrations⁷. It should be noted that the relationship between scaffold architecture and titer improvement is not predictable for this pathway, as discussed later. The same set of scaffold architectures was applied to a second pathway engineered by Moon *et al.* for the biosynthesis of glucaric acid^{8,80}. This pathway presents an interesting test case in that it is a relatively high titer-producing pathway, on the order of 1 g/L, with a flux bottleneck enzyme, MIOX, that appears to be substrate activated⁸⁰. Varying the number of domains that recruit the enzyme upstream of MIOX, Ino1, resulted in gradually increased product titers to a maximum of almost five-fold improvement with four Ino1 recruitment domains, whereas varying the number of MIOX recruiting domains had little impact⁸. This observation is consistent with a model in which the local concentration of substrate for the limiting MIOX activity is modulated by upstream enzyme recruitment via scaffold domain stoichiometry within the synthetic complex.

Even without direct substrate channeling guiding intermediates between active sites as observed in the natural examples discussed previously, product titers may be improved by co-localizing consecutive metabolic enzymes to produce a higher local concentration of metabolite in close proximity to the downstream enzyme^{72,81}. This has been the subject of debate in early papers in which consecutive enzymes, β -galactosidase and galactose dehydrogenase, were tethered with a translational fusion, generating a higher product titer⁸². The authors suggested the mechanism for increased flux was substrate channeling; however, a subsequent paper performing a detailed kinetic analysis of the fusion protein challenged this conclusion⁸³. Local concentration effects have been modeled for engineered pathways of heterologous enzymes by simulating native and engineered pathway reaction rates within an *E. coli* discretized into subvolumes, localizing the engineered pathway within a single subvolume, and accounting for metabolite diffusion to simulate compartmentalization⁸⁴. Though direct measurements of local intermediate concentrations within enzyme complexes have remained elusive, this mechanism seems to be an attractive explanation for some of the successes observed with scaffolding^{7,8} and other co-localization engineering examples reviewed in⁷². However, local concentration effects may be acting in conjunction with other mechanisms discussed later in this manuscript. Enzyme co-localization may allow achievement of a specific local intermediate concentration with a lower concentration of upstream enzyme than would be possible with freely diffusing enzymes, thus retaining high flux while reducing the metabolic load on the cells (Figure 3.1A). Reduction of intermediate in the bulk of the cell may also be beneficial if the intermediate is toxic or undergoes undesired reactions through competing pathways (Figure 3.1A).

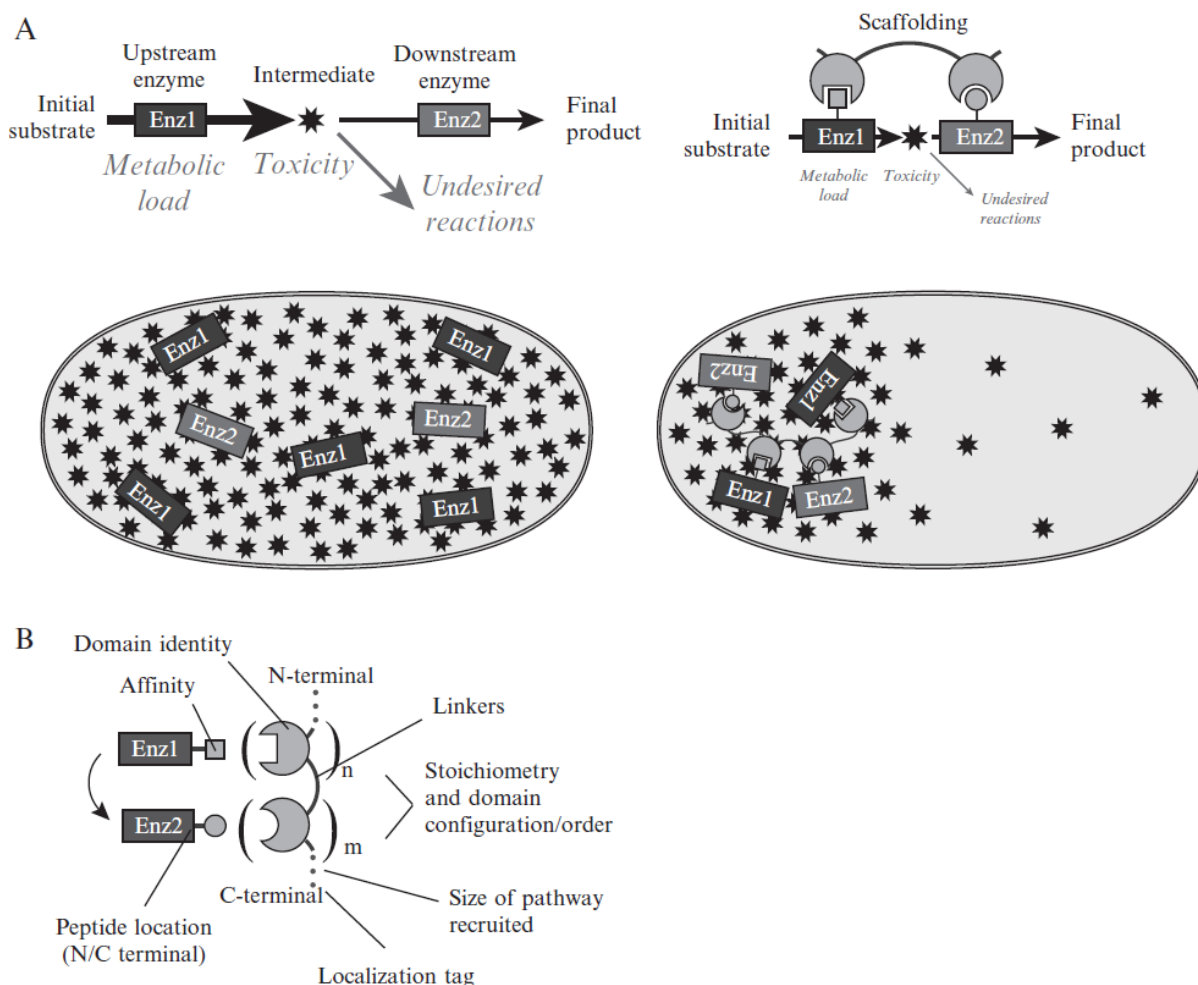


Figure 3.1. Schematic of potential colocalization effects. (A) High local substrate concentrations may be achieved with low enzyme expression, thereby reducing the cellular burden by using scaffolding to co-localize enzymes. (B) Scaffolding based on modular protein-protein interaction domains provides a highly designable control point with multiple parameters available for optimization.

We employ modular protein-protein interaction machinery tethered with long flexible synthetic linkers to colocalize enzymes in engineered metabolic pathways. While this approach currently lacks the ability to precisely control the three dimensional positioning of recruited enzymes, it has the advantage that each protein-protein interaction domain should be capable of targeting its interaction partner in a manner independent of composition context, provided neighboring targeted enzymes do not sterically block a physical interaction. When a binding domain is incorporated into a scaffold using long linkers, it generally will retain the ability to bind its target ligand regardless of where it is located on the scaffold and what domains are encoded up or downstream. This provides a highly designable platform where matrices of scaffolds can be generated in which key parameters are varied while interaction functionality is maintained (Figure 3.1B). Additionally, once interaction tags have been successfully added to pathway enzymes, the pathway can be used with a variety of

scaffold architectures. However, as discussed later in the manuscript, applying tags to enzymes can have difficult-to-predict effects on enzyme activity and concentration, both of which must be assessed. To scale recruitment to include additional enzymes, additional binding domains can be fused to existing scaffolds or a second scaffold molecule can be co-targeted to the original scaffold. For interaction domains with a set of ligands ranging in recruitment affinity, binding strength can usually be modulated by point mutations to the ligand without the need to redesign scaffold libraries. Furthermore, different enzyme packing structures can be explored by shuffling the configurations of the scaffold domains.

In this manuscript, we describe methodologies by which synthetic complexes can be engineered from metabolic pathways using modular protein scaffolds. Although the synthetic scaffolds built to date, and certainly for the foreseeable future, do not rival the elegance of natural systems, they offer designable control over several assembly parameters, most notably enzyme stoichiometry, incorporation of heterologous enzymes, and potential scalability for increased numbers of enzymes. We hope to provide a practical guide for how we would approach scaffolding a pathway *de novo* together with a discussion of our considerations and experiences thus far. We conclude this manuscript with a discussion of several mechanisms that may play a role in observed titer improvements for pathways tested thus far and of systems that may benefit from these possible effects.

3.3 – METHOD: How to build modular designable protein scaffolded systems for metabolic engineering applications

3.3.1 – Selecting protein-protein interaction domains and ligands for scaffold construction

The first decision to be made for scaffolding a metabolic pathway is the choice of modular interaction motifs. Each enzyme is translationally fused to a ligand specific for a protein-protein interaction domain. A translational fusion of these domains will compose a scaffold capable of co-localizing the ligand-fused enzymes. The structural modularity of the protein-protein interaction domains is of primary importance, as they will need to retain binding activity in the non-native context of the translational fusions. A number of modular protein-protein interaction domains have been characterized and employed in various applications, a partial list of which is compiled in the Table 2.1. In our experience, the members of SH3, PDZ, GBD, and leucine zipper families tend to retain binding activity as N-, C-terminal, or internal fusions and, given sufficient linker lengths, rarely require linker optimization to achieve binding activity. However, as discussed later, despite robustness of binding activity, flux improvement is likely also influenced by scaffold architecture including parameters such as linker length/composition and number/arrangement of protein-protein interaction domains.

Another parameter of importance is association affinity, particularly under low expression regimes. We generally design our scaffolds to target enzymes in the low micromolar range or tighter. To date, all targetings have been executed with the tightest

affinity ligands available; however, many of these domains, as listed in Table 2.1, include lower affinity ligands that could be employed if transient interactions are desired.

Protein-protein interaction domains belonging to families with many members are particularly attractive choices for use in scaffolding, as they may offer a set of domains that potentially recognize specific ligands orthogonally (i.e. minimal cross-talk with other ligands used as enzyme tags), yet have conserved folds and can more likely be used interchangeably. For example, individual SH3 domains appear to have undergone negative selection such that they do not measurably interact with other SH3 domain family ligands within that organism⁸⁵. Zarrinpar *et al.* showed that a peptide ligand was highly specific for a single SH3 domain within its native host, *S. cerevisiae*, whereas this same peptide ligand interacted with a high percentage of non-*S. cerevisiae* SH3 domains. This selection for reduced cross-talk should considerably increase the number of orthogonal domain/ligand pairs available for simultaneous use, particularly of domains recognizing small ligands such as the SH3 and PDZ domain family. Additionally, these interaction domain families often appear to have evolved physical and functional modularity, including characteristics such as robust independent domain folding and surface-exposed N and C termini that are located close together to permit domain functioning as either terminal or internal fusions.

Leucine zipper and synthetic coiled-coil domain folds share many of these characteristics and are attractive targets for expanding the available number of orthogonal interaction partners. Works such as those by Havranek *et al.*⁸⁶, where eight residues between leucine zipper pairs were altered based on computational prediction to create new pairs of either homodimers or heterodimers, show promise for rationally engineering new domains. More recently, Reinke *et al.* investigated the interaction specificities of a large set of synthetic coiled-coils that do not exhibit measurable self-association providing up to three orthogonal pairs that do not cross-talk⁸⁷. These large libraries of structurally similar but orthogonally binding pairs provide excellent candidates for scaffold parts, as presumably they may be interchanged to switch specificity with minimal perturbation. However, as generating very large libraries of orthogonal parts has proven challenging, limited to sets with only several experimentally verified orthogonal pairs, taking parts from different families to minimize likelihood of crosstalk is still likely to be fruitful for producing larger numbers of orthogonal protein-protein interaction pairs.

Cohesin-dockerin interaction modules have been successfully used to scaffold multi-enzyme complexes to function as synthetic cellulosomes in vitro. Up to three cellulose degrading enzymes were translationally fused to dockerins that localize to specific cohesins on a synthetic scaffold, which itself localizes to cellulose substrate via a carbohydrate binding module. The resultant complex enhanced cellulose degradation in the complex substrate of straw 6-fold over free enzyme⁷⁸. An in-depth review of a number of applications that have taken advantage of cohesin-dockerin domains to provide controlled extracellular binding has recently been published⁸⁸. A study of cohesin-dockerin specificities has demonstrated up to five cohesin-dockerin pairs exhibiting orthogonal binding specificity, providing a set of modules for further application⁸⁹. A unique feature of cohesin-dockerin interactions is that

they bind with a very tight affinity in a calcium ion dependent manner. This makes them ideal candidates for extracellular scaffolding but likely limits their application in vivo due to the low concentration of free calcium in the cytoplasm.

Part Family	Tightest Affinity K_D	Domain/ Ligand Size (AAs)	Source	Features and Issues	Confirmed Orthogonal Pairs
SH3 domain / peptide	$1 \times 10^{-1} \mu M^{90}$	57/11	AAH31149: 196-274 PPPALPPKRRR ⁹⁰	Relatively context independent and well-characterized. Ideal for internal insertion. Natural peptides tend to have micromolar affinities.	Specificity observed within species ⁸⁵
PDZ domain / peptide	$1 \times 10^0 \mu M^{91}$	96/6	EDL06069: 77-171 GVKESLV ⁹²	Generally PDZ peptide must be C-terminal. nNOS domain can be used for non-C-terminal ligands.	2 natural ⁹³
GBD domain / peptide	$1 \times 10^0 \mu M^{94}$	80/32	BAA21534: 196-274 P42768: 466-497	Less well-characterized than SH3 or PDZ. Longer linker sequence.	1
Leucine Zippers	6.1×10^{-3} and $8.3 \times 10^{-2} \mu M^{95,96}$	43/43	ITIRAAFLEKENTALR TEIAELEKEVGRCENI VSKYETRYGPL, LEIRAAFLEKENTAL RTRAAELRKRVGRC RNIVSKYETRYGPL	Significant likelihood of homodimerization, particularly important to test for intramolecular pairs.	3 synthetic ⁹⁷
PhyB/Pif3 light switchable binding	2×10^{-2} – $1 \times 10^{-1} \mu M^{98}$	908/91	AAW56577: 1-908 NP_172424: 120-210	Light dependent binding activated at 720nm and deactivated at 660nm light.	1
FKBP/FRB	$1.2 \times 10^{-3} \mu M^{99}$	107/93	AAI19733: 39-145 EAW71681: 1972–2064	Interaction is inducible with the small molecule Rapamycin at $K_D=0.2nM^{100}$. A FRB(T2098L) mutation allows use of a non-toxic rapamycin analogue for T2098L characterization ⁹⁶ .	1
Cohesin / Dockerin	$<1 \times 10^{-5} \mu M^{77}$	$\sim 150/\sim 70$	YP_001039466 YP_001038489	Calcium dependent binding activity is likely not functional at free cellular Ca^{+2} levels. Calcium $K_D=2.5 \times 10^{-7} M^{-2}$, half binding at $500 \mu M Ca^{+2}$. ¹⁰¹ .	5 natural ⁸⁹

Table 3.1. Protein-protein interaction domain families potentially useful for scaffold construction.

3.3.2 – Assembling scaffolds from domains and enzyme tagging for co-recruitment

Domain/ligand choice is particularly important for proteins whose activity, stability and/or solubility are sensitive to translational fusion. Particular peptide sequences and fusion locations may decrease the flux through these sensitive enzymes beyond the capability of the scaffolding effect to surpass. We attempt to minimize the perturbation to the enzyme of interest by selecting the smaller member of the binding pair to use as a tag. For proteins known to be problematic, we often use either an eleven-amino-acid peptide with a $K_D = 0.1 \mu M$ for the Crk SH3 domain on the N or C-terminus or a six-amino-acid peptide with a $K_D = 8$

μM for the syntrophin PDZ domain as a C-terminal fusion. PDZ peptides must be used as C-terminal fusions since the carboxyl group is critical for binding. In the case of enzymes that are already experiencing solubility problems, it is possible that adding a larger, well folded, binding domain may increase solubility similar to the oft-used strategy of tethering folding problematic proteins to maltose binding protein or other highly soluble motifs¹⁰². The coiled-coil motifs, due to their high solubility, may be good candidates to try for this purpose, although all of these efforts are protein-dependent and currently unpredictable.

Linkers connecting domains of the scaffold are likely to provide another parameter that could be explored for flux optimization. To date, we have limited experience in the effect of different linker types. To connect domains, we have been using linkers expected to be of sufficient length (nine or more amino acids) to avoid sterically obstructing neighboring domains from binding and with compositions predicted to be unstructured and flexible (glycine-serine repeats). We have observed a small improvement in performance when linkers separating blocks of domains (i.e. the linkers in the scaffold $\text{GBD}_1\text{-linker}_{(\text{SH3})_2}\text{-linker}_{(\text{PDZ})_2}$) were further increased from nine to 25 total amino acids of Gly-Ser repeats (Dueber unpublished observation). However, we have not investigated varying linker lengths or the composition of these linkers beyond this initial characterization. Robinson *et al.* investigated the effect of linker length and composition on the stability of single-chain Arc repressor¹⁰³ (Robinson *et al.*, 1998). For this protein, both linker length and the amino acid composition had a large impact on stability. Initial work on a simple tethering of two enzymes in the mevalonate biosynthetic pathway (Figure 3.2) also suggested that a linker of adequate length must be used to achieve highest flux improvement and, in this case, that the improvement is maintained through linkers of increasing lengths.

When making translational fusions of binding domains/ligands and enzymes, it is important to assay enzyme activity and interaction domain functionality. A GST pull-down assay can be employed to ensure association activity remains functional. When domains from the same family are being employed with the intention of orthogonal function, GST pull-down assays often are very useful for avoiding unexpected inter-molecular interactions, though low affinity and high effective concentration intra-molecular interactions may be missed. For *in vivo* assays of enzyme activity, untagged enzyme function can be tested against tagged enzymes in absence of scaffolding. However, care must be taken to ensure expression rates are not being altered by domain addition, particularly for N-terminal tags or polycistronic systems which, in prokaryotes, should be expected to alter expression rates through RNA secondary structure interactions with the ribosome binding site¹⁰⁴. For example, the addition of a C-terminal peptide targeting an enzyme might affect the expression level of an immediately downstream enzyme in a polycistronic message. This effect can be estimated or corrected for with the RBS calculator⁹. To confirm that expression levels are not changing, enzyme concentration should be carefully measured or expressed on independent transcripts.

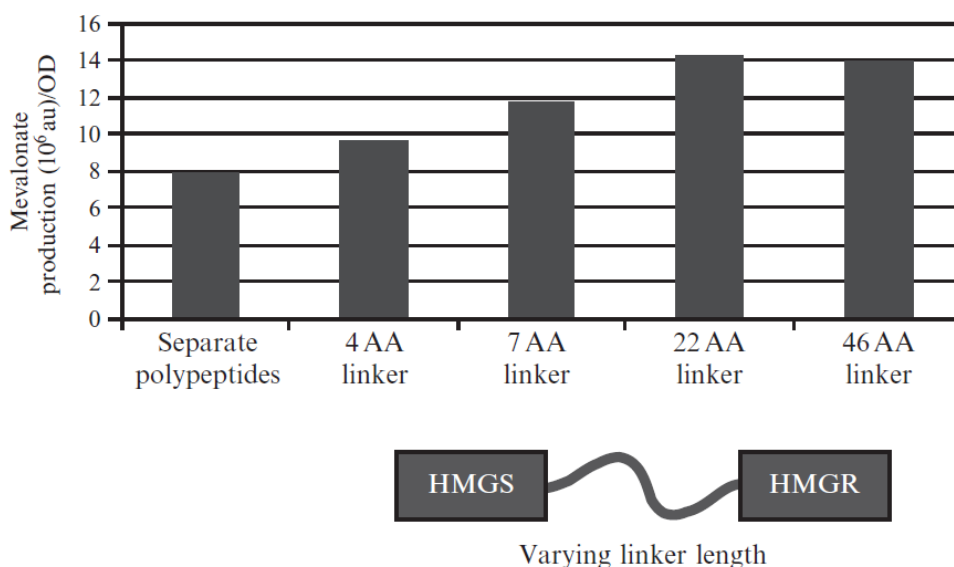


Figure 3.2. Two consecutive enzymes in a synthetic mevalonate pathway were tethered, C to N termini, and the length of a simple synthetic linker of alternating Glycine and Serine residues was varied. Tethering with a short linker length showed small improvements, which were increased to an optimized titer with increased linker lengths. Mevalonate production was measured as GC/MS peak area as described in detail⁷.

3.3.3 – Balancing the scaffold and enzyme concentrations

Balancing concentrations of enzymes has been shown to be an important consideration for optimizing metabolic pathways. Balancing scaffold concentrations is also important, as there is a theoretical optimal concentration for maximizing fully occupied scaffold molecules. This effect was modeled by Levchenko *et al.* for scaffolding in the MAPK signaling pathway¹⁰⁵, where low concentrations of scaffold result in insufficient scaffold to co-localize the targets, while concentrations of scaffold considerably higher than enzymes result in segregation of components and a high percentage of scaffold molecules with low occupancy. We believe this biphasic trend also exists for synthetic scaffolding of multi-enzyme pathways. We simulated a simple mathematical model of equilibrium binding reactions for a varying number of different enzymes that bind to different single sites on a scaffolding protein (Figure 3.3). Differential equations for simple binding kinetics were generated with code written and simulated in MATLAB (The MathWorks, Natick, MA), code available upon request. Enzyme concentrations were held constant at 10 μ M with each binding to a single site on the scaffold with a 100 nM K_D , while scaffold concentrations were varied. As expected, the optimal scaffold performance occurs when scaffold concentrations are approximately equal to enzyme concentrations, and concentration optimization becomes increasingly important as pathway size increases. In agreement with these modeled predictions, we observed a strong dependence of production titers on the relative expression levels of both scaffold and metabolic enzymes^{7,8}. Thus, it may be helpful to drive expression of pathway enzymes and scaffold with independent promoters to independently tune expression to the optimal levels.

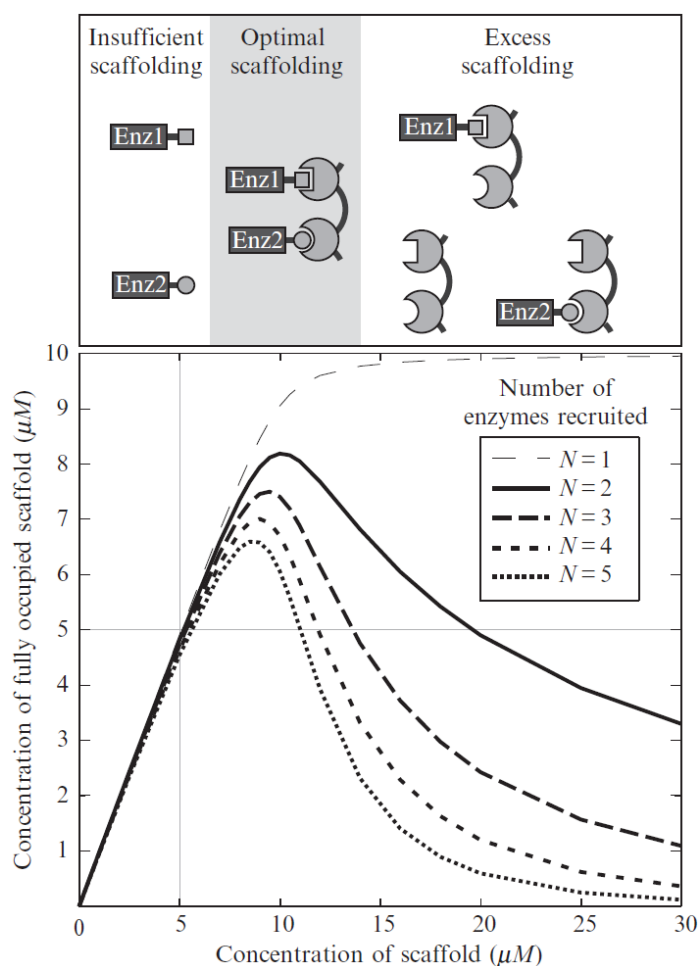


Figure 3.3. A mathematical model of equilibrium binding of scaffolding recruiting enzymes shows an optimal scaffold concentration for maximizing full occupancy. Five scaffolds consisting of a varying number of enzyme recruitment domains are independently simulated. Enzyme levels are held constant at 10 μM each, and each scaffold molecule recruits a specified number of different enzymes each with a dissociation constant of 100 nM. Initially, as scaffold concentration is increased, excess enzymes are recruited to fully occupy the scaffold. As enzymes become limiting, scaffold competition for enzyme recruitment leads to low occupancy of scaffold.

In addition to optimizing scaffolding levels, it may be beneficial to simultaneously adjust the pathway enzyme concentrations. High induction was found to be the optimal expression level for mevalonate biosynthetic pathway enzymes in the absence of scaffolding. However, when the most effective scaffold was present, low enzyme induction produced optimal production titers, giving higher titers than the maximum achievable in the absence of scaffold, even at the uninduced background expression level of the promoter (Figure 3.4). Although it would be interesting to independently optimize the expression of the three enzymes scaffolded in this pathway, optimization beyond polycistronic expression level has not yet been carried out. Scaffold architecture, expression level, and enzyme expression levels are all interconnected variables that must be optimized.

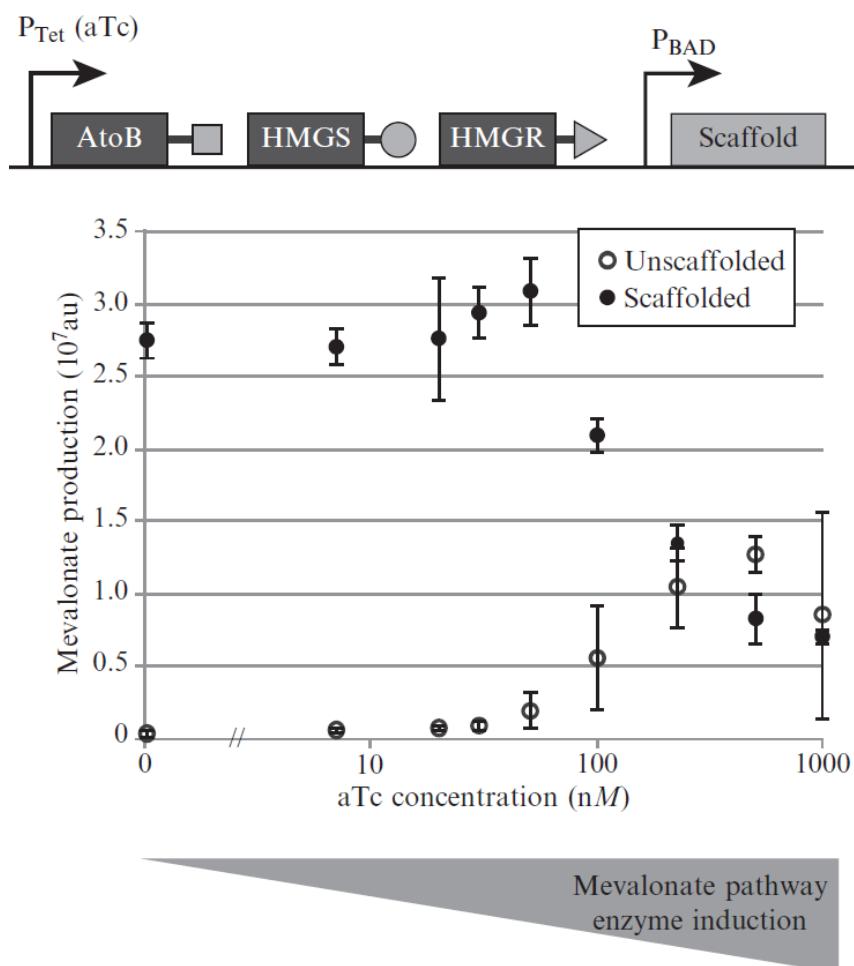


Figure 3.4. Mevalonate titers are measured at varying aTc concentrations, corresponding to induction of P_{Tet} driving peptide-tagged AtoB, HMGS and HMGR polycistronically. The P_{BAD} promoter either drives the optimal scaffold, GBD₁-SH3₂-PDZ₂, (filled circles), or GFP, representing unscaffolded pathway (unfilled circled). Mevalonate production was measured as GC/MS peak area. Error bars represent one standard deviation from three separate experiments (figure adapted from Dueber et al., 2009).

3.3.4 – Scaffolding stoichiometry

Often when metabolic pathways are engineered, one pathway enzyme exhibiting relatively low activity creates a bottleneck in the pathway. In many cases, this limitation can be alleviated by increasing expression of that enzyme. Pathway scaffolding presents another strategy for addressing this problem at lower enzyme expression regimes. Ability to achieve high product titers with low concentration of enzymes should prove particularly advantageous for systems with enzymes prone to aggregate. Additionally, improved pathway efficiency is likely to become increasingly important as the number of enzymes in the pathway is increased, although this remains to be empirically tested.

Due to the modular nature of the described scaffold strategy, the relative stoichiometry of enzymes co-complexed can be controlled by varying the number of repeats of each protein-protein interaction domain. For the mevalonate biosynthetic pathway, a matrix of nine scaffold architectures was assembled with one, two or four protein-protein interaction domains recruiting the enzymes for the bottleneck intermediate transfer, HMGS and HMGR (Figure 3.5). Within this matrix, the optimal architecture produced a 77-fold improvement in product titers relative to the unscaffolded pathways. Importantly, although all scaffolds improved titers relative to the unscaffolded pathway, the results were difficult to explain based on stoichiometry alone.

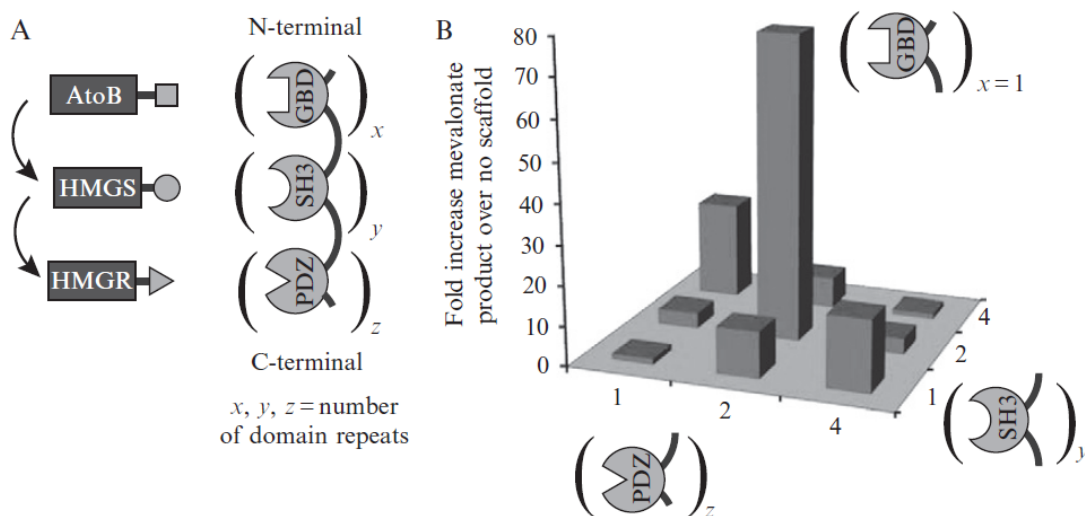


Figure 3.5. Three enzymes in the synthetic mevalonate pathway were tagged with binding peptides and recruited to a synthetic scaffolding protein with varying stoichiometry of binding domains. (A) Schematic of scaffolding with varying stoichiometry, where the number of GBD domains is held constant at one, while the number of SH3 and PDZ domains are varied to be one, two or four. (B) The different scaffold stoichiometries gave very different results, with the best performing scaffold resulting in a 77-fold increase in mevalonate production compared to when no scaffold was expressed (figure adapted from Dueber *et al.*, 2009).

We reapplied the same matrix of scaffolds used to improve flux of the mevalonate biosynthetic pathway to a second pathway, the glucaric acid biosynthetic pathway. This pathway was previously engineered by Moon and Prather from three heterologous enzymes, Ino1, MIOX, and Udh⁸⁰ and expressed at high levels under the T7 promoter to maximize glucaric acid titers. Interestingly, the Prather group measured higher activities of the limiting enzymatic activity, MIOX, in the presence of high concentrations of *myo*-inositol substrate^{8,80}. The three heterologous enzymes were tagged with recruitment peptides and expressed from the PLac promoter. Similar to our findings with the mevalonate biosynthetic pathway, the scaffolds showed various degrees of titer improvements dependent on architecture. Consistent with a hypothesis of increasing the local concentration of *myo*-inositol at the

resultant synthetic complex, titer improvements were dependent on the number of Ino1-recruiting domains producing myo-inositol, whereas there was no strong dependence on the number of domains recruiting MIOX enzyme. Titters of 2.3 g/L glucaric acid were produced with the optimal scaffold, giving an almost five-fold increase over the control lacking scaffold expression, a 50% improvement over highest titers previously reached (Figure 3.6).

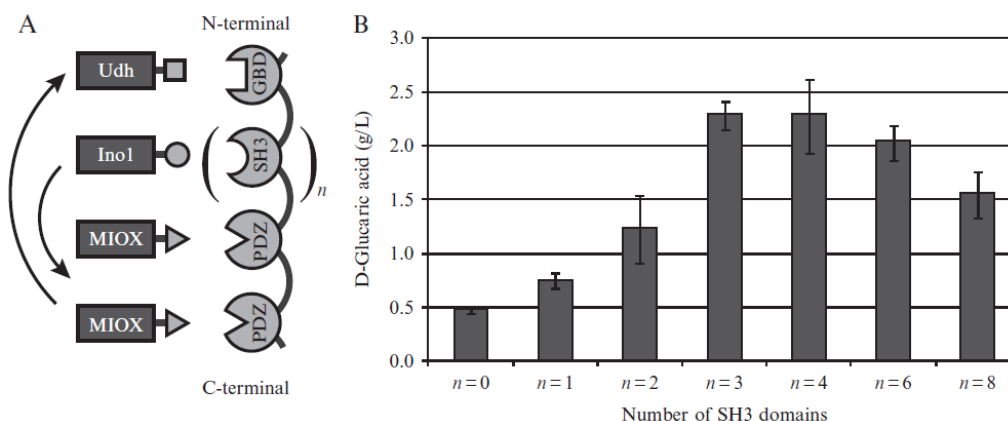


Figure 3.6. A pathway engineered to produce glucaric acid was scaffolded with optimized stoichiometry to improve production. (A) Schematic of scaffolded pathway where the number of SH3 domains corresponding to Ino1 recruitment is varied. Ino1 produces the substrate for MIOX, which MIOX converts to the substrate for Udh. (B) Glucaric acid concentrations were measured under conditions where only the number of SH3 domains on the scaffolding is varied, showing around 3 or 4 SH3 domains provided optimal production (figure adapted from Moon *et al.*, 2010).

There are a few considerations to be made when constructing sequence for scaffolds with domains repeated multiple times. Cloning strategies relying on PCR or homologous overhangs such as recombination-based methods¹⁰⁶, sequence and ligation independent cloning (SLIC)¹⁰⁷, or isothermal enzymatic assembly¹⁰⁸ may result in misannealing for products containing repeated sequences. One construction strategy that deals particularly well with repeated domains is the BioBrick-based cloning strategies^{109,110}, particularly the BglII/BamHI based strategy¹¹⁰ that leaves generally innocuous and often useful glycine-serine scars that can be used as part of the linker sequence. Interestingly, we have observed problems coming from recombination arising from greater than 4 identical repeats for both domains (~200 bases) and peptide ligands (~60 bases). A solution to this problem is to design multiple domain “parts” with degenerate codon usage such that repeated domains, linkers and ligands are sufficiently different to prevent recombination. For making multiple SH3 domain repeats, using six degenerate SH3 parts, we were able to make constructs with ten repeats without observing a significant number of incorrect products due to recombination⁸.

3.3.5 – Scaffold composition

The three dimensional structure of the scaffolded complex will determine the efficiency of improving flux. Although these structures are determined by the domain

architecture, these architecture/structure relationships are not currently predictable. This is highlighted by the importance of not only the total number of each protein-protein interaction domain but also by the arrangement of these domains. The number and identity of SH3 and PDZ interaction domains in a scaffold was held constant but the order of these domains was rearranged⁷. GBD-(SH3)₂-(PDZ)₂, GBD-(SH3)₁-(PDZ)₂-(SH3)₁, and GBD-(SH3)₁-(PDZ)₁-(SH3)₁-(PDZ)₁ scaffolds showed dramatically varied abilities to improve mevalonate titers despite each having the same number of recruitment domains (Figure 3.7).

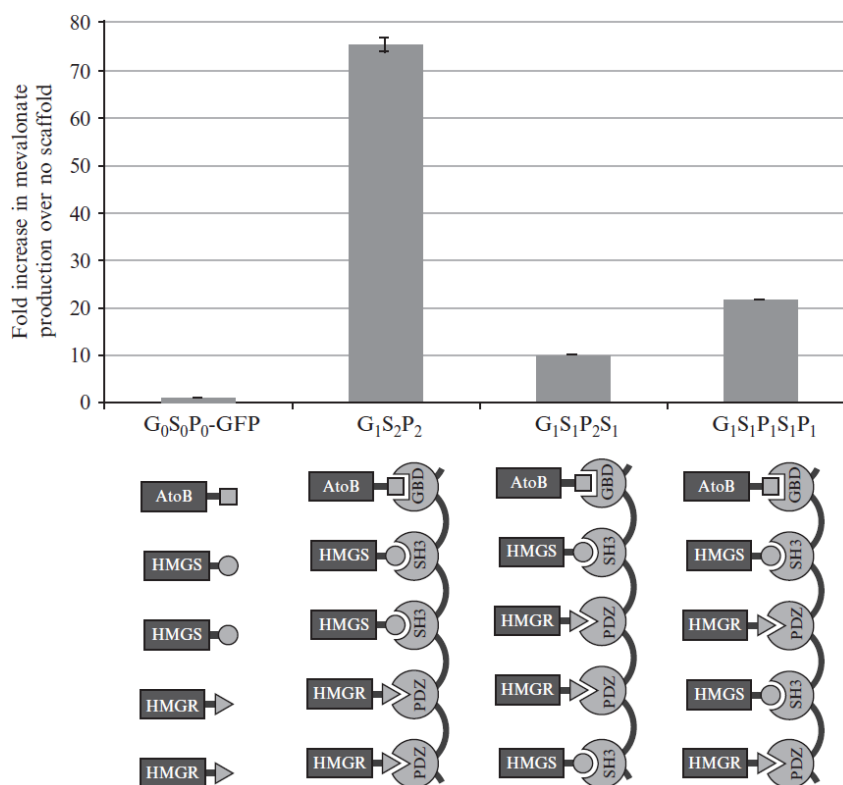


Figure 3.7. Mevalonate titers are measured for ligand tagged enzymes with expression of scaffolds of differing architecture and compared to the titers in the absence of scaffold expression. Rearranging domain order, while retaining the one GBD, two SH3 and two PDZ stoichiometry gives significant changes in titer improvement (figure adapted from Dueber *et al.*, 2009).

Work on synthetic cellulosomes has also supported the importance of scaffold composition¹¹¹. In this study, scaffolds were designed to recruit other scaffolds, creating complexes containing up to four different scaffolds, each in turn recruiting, or directly fused to, a cellulose degrading enzyme. One issue that arose was the importance of enzyme mobility, as redundant binding that was likely to limit flexibility decreased degradation efficiency¹¹¹. Natural cellulosome protein sequences suggest they are physically flexible complexes, since they generally have long linkers (tens up to 550 residues) predicted to result in highly mobile enzymes¹¹². Another issue was the importance of scaffolding complex mobility, suggested from lowered activity with the addition of more than one carbohydrate

binding module. These issues of mobility and flexibility are likely to be particularly relevant to the catalysis of cellulose, an immobilized substrate, and may have limited pertinence to readily diffusing metabolic intermediates.

Interestingly, it has been suggested that when individual enzymes are tethered together, if these enzymes exist in oligomeric form, they may multimerize, forming even larger complexes^{72,113}. A scaffold with repeated domains that recruits oligomeric enzymes, as is the case in several applications thus far^{7,8}, may also form large multimeric complexes (Figure 3.8). These multimeric complexes may improve titers by further increasing local concentrations beyond those achievable with individual scaffolds. This potential phenomenon is another reason we recommend taking a library approach to optimize scaffolded pathway flux, varying as many parameters as practical to empirically determine the optimal combination of architectural parameters.

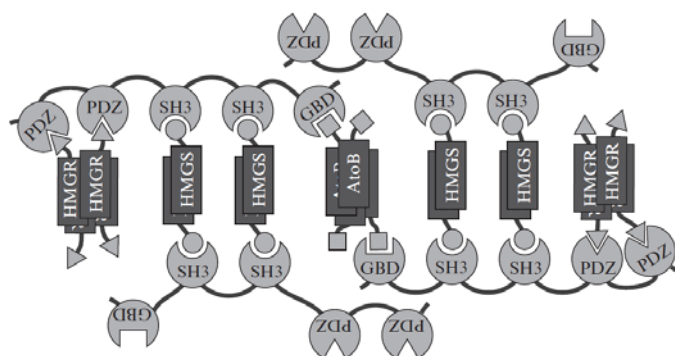


Figure 3.8. Schematic of a multimeric enzyme complex. Enzymes with oligomeric structures could potentially bind multiple scaffolds, resulting in large complexes and difficult to predict positioning in three-dimensional space.

Scaffold configuration may become an increasingly important variable as pathways are scaled to consist of larger numbers of enzymes. One practical concern is the increasing scaffold protein size to target increasing numbers of enzymes while also achieving stoichiometry control. As explored in the development of synthetic cellulosomes¹¹¹, a potential solution to this problem is building multiple scaffolds that can co-assemble either directly or through a separate adaptor molecule. This approach might also prove to be a convenient method for modularizing various sections of a pathway, as well as increasing the combinatorial architecture possibilities.

3.4 – Systems that may benefit from scaffolding

There are a number of potential mechanisms that may contribute to the increased titers shown in the scaffolded systems discussed. We have thus far focused on the increased local concentration effect by which scaffolding may reduce toxic intermediates, reduce the load on the cell by reducing the necessary enzyme expression levels and increase pathway efficiency. Another potential benefit of scaffolding is the prevention of enzyme aggregation

by lowering the necessary protein expression levels in addition to sequestering individual enzyme molecules. It is also possible that scaffold could be used to increase enzyme stability or activity, though this was shown to not be the case for the glucaric acid pathway⁸. Substrate-activated enzymes, such as MIOX in the glucaric acid pathway discussed earlier, are also particularly good candidates for scaffolding. Scaffold complexes with higher numbers of protein-protein interaction domains can be used to recruit substrate-producing upstream enzyme⁸.

Scaffolding a complex of efficient enzymes may also provide a means of limiting intermediate loss. Increased local concentration, or corresponding reduction in bulk cytoplasmic concentrations, would reduce the rate of intermediate loss to competing pathways. Additionally, the reduced enzyme expression also reduces the rate of those enzymes metabolizing unintended substrates. Insulation from unintended interactions is clearly a desired engineering characteristic as it is difficult to predict interactions with the natural cellular metabolism. These undesired interactions may be reduced with co-assembled enzyme complexes. This may be increasingly important as engineered metabolic pathways continue to scale in size and complexity. These larger pathways may necessitate building and characterizing different modules of a pathway independently, which when assembled would benefit from reduction of unintended interactions between shared intermediates. Another potential mechanism for reducing unwanted interactions is the localization to different subsections or compartments of the cell. The ability to target the scaffold to a particular location may facilitate pathway localization strategies.

3.5 – Concluding Remarks

In conclusion, scaffolding provides a post-translational tool that may help increase production yields and deal with problematic enzymes, as well as reduce the cellular burden and unintended interactions that may become an issue when scaling to engineer the next generation of biosynthetic metabolic pathways. The methods we describe in this paper have the advantage of being highly designable and easily adaptable to library approaches. Future research must be done to determine the mechanisms by which these scaffolded systems function such that they can be more rationally applied to other pathways in a predictable manner.

3.6 – Acknowledgements

We thank R. Conrado and M. DeLisa, as well as K. Tipton and other members of the Dueber lab for comments and discussion regarding the preparation of the manuscript. This work was supported by funding from NSF grant no. CBET-0756801 (W.R.W., J.E.D.).

Chapter 4 – Engineering robust control of two-component system phosphotransfer using modular scaffolds

Authors:

Weston R Whitaker

Department of Bioengineering, University of California, Berkeley CA, United States

Stephanie A Davis

Department of Chemistry, University of California, Berkeley CA, United States

Adam P Arkin (corresponding author: aparkin@lbl.gov)

Department of Bioengineering, University of California, Berkeley CA, United States

Energy Biosciences Institute, University of California, Berkeley, CA, United States

Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley CA, United States

John E Dueber (corresponding author: jdueber@berkeley.edu)

Department of Bioengineering, University of California, Berkeley CA, United States

Energy Biosciences Institute, University of California, Berkeley, CA, United States

Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley CA, United States

4.1 – Abstract

The field of synthetic biology seeks to predictably forward engineer sophisticated biological systems. Remarkable successes towards this goal have been demonstrated with genetic circuits, aided by the clear modularity between transcription control elements and genes allowing new input control to be reliably achieved for a given gene. We demonstrate the feasibility of achieving similar modular rewiring in post-translational circuits, an underutilized layer of control in prokaryotes. We synthetically control enzyme-substrate co-assembly to direct phosphotransfer of two component systems. Scaffold proteins built from modular protein-protein interaction domains specifically co-target proteins tagged with corresponding interaction ligands. Flux can be directed from a histidine kinase to either of two non-cognate response regulators depending on the scaffold expressed. Finally, since this effect was highly sensitive to the balance of histidine kinase and response regulator concentrations, robustness to kinase concentration was gained by designing an allosterically-regulated kinase switch such that the ligand on the scaffold serves the dual functions of directing assembly and switching activation. These results are first steps towards a generalizable strategy for designing modular prokaryotic signal transduction.

4.2 – Introduction

Living cells exhibit a remarkable ability to process multiple signals to maintain homeostasis in a dynamic environment, to diversify or specialize behavior, and to organize complex structures. Signal transduction pathways (STPs) play a key role in these processes as essential elements in sensing, processing, and transmitting environmental information, often at rapid speeds and with the potential to hold different states at different locations within the same cell. As synthetic biology is challenged to design more sophisticated systems, it will become increasingly important to match these abilities in a designable and predictable manner. Impressive successes towards these goals have been made in the design of genetic circuits, achieving a diverse set of desired behaviors by varying the arrangement of common transcriptional regulators and promoters^{68,114–116}. The physical and functional modularity of genes and promoters is essential for allowing parts to be easily rearranged with relatively predictable behavior¹¹⁷. While connectivity at the transcriptional level can be rewired via adjacency on DNA, connectivity of STPs are generally determined through interactions between many interface residues that are difficult to manipulate with predictable behavior. One recent study, however, successfully engineers modular control into the yeast mating STP through colocalization of positive or negative signaling modulators to the Ste5 scaffold via introduced binding domains². By rearranging the recruitment machinery to alter the feedback acting on the pathway, an impressive range of sophisticated temporal and steady state behavior is achieved².

Eukaryotic signaling pathway proteins are often composed of a catalytic output domain regulated by physically distinct input domains¹¹⁸. Domain recombination events allow novel input/output linkages as well as reshaping the transfer function of a given signaling response^{1,40}. Scaffolds, which spatially organize catalytic activities, represent an extreme example of the evolutionary benefits of input modularity by allowing catalysis and input control to be separated into distinct molecules^{119,120}. Prokaryotic organisms, however, make far less extensive use of such modular organization, with the notable exception of CheW in the chemotaxis complex, which has already demonstrated utility in synthetic biology applications¹²¹. Generally prokaryotic signaling proteins encode both catalytic activity and binding specificity into the same cooperative fold. This raises the question of whether engineering strategies mimicking eukaryotic modular organization would be effective for directing signaling flux in the form of phosphate transfer in prokaryotes.

Two-component systems (TCSs) represent an attractive target for engineering signal transduction as they are the predominant signaling strategy used by prokaryotes for sensing a wide array of environmental signals, have a relatively simple and well-characterized phosphotransfer scheme, and have highly conserved structures suggesting lessons learned from one pathway could be applied to many others. The TCS signal transduction architecture is generally organized around two conserved proteins, a histidine kinase (HK) that senses a stimulus and autophosphorylates, and a response regulator (RR) to which the phosphate is transferred. Response regulators subsequently actuate a response, usually transcriptional regulation. Phosphotransfer routing and the coupling of stimulus to activity are both

important aspects of signal transduction, but in this work we focus primarily on the former. In the last decade, significant progress has been made to understand TCS phosphotransfer routing and the precise mechanisms by which phosphotransfer specificity is encoded in TCS^{122–124}. In particular, while detectable phosphotransfer cross-talk between non-cognate components is somewhat common for purified components in, a clear kinetic preference between cognate pairs appears to be a major determinant of specificity¹²³. We propose that TCS kinetic preference is, at least in part, due to specific binding (i.e., a K_M effect) and, therefore, an increased local concentration through synthetic assembly could serve as a basis for controlling signaling for non-cognate components. It is unclear how responsible the local concentration effect is for controlling flux in the scaffolded pathways of higher organisms, but the frequency of enzyme co-assembly suggests it does play a significant role. We demonstrate HK/RR colocalization to be sufficient for scaffolding to impart specificity, explore sensitivity to the relative expression level of each signaling component that occurs when a simple scaffold designed solely for co-assembly is employed, and reduce this sensitivity by engineering an autoinhibited HK that is activated upon scaffold binding.

4.3 – Results

4.3.1 – Colocalization amplifies weak natural cross-talk

We have designed an in vivo assay to examine phosphotransfer specificity consisting of one HK exhibiting weak natural crosstalk to two non-cognate RRs. In vitro phosphotransfer assays show that although some crosstalk to one or a few non-cognate RRs is common for most HKs, EnvZ appears to phosphotransfer to a particularly large number of non-cognate RRs^{123,125}. Additionally, the large number of studies on Taz, a HK chimera of the Tar sensing domain and the EnvZ cytoplasmic domain, giving the phosphotransfer specificity of EnvZ but with a more well-defined input control³⁷, led us to choose Taz as our model HK. To study the phosphotransfer specificity in vivo, Taz as well as noncognate RRs were expressed on independent, inducible promoters and fluorescent transcriptional reporters for both target RRs were introduced on a plasmid (see Table S4.4 for summary of plasmids and Figure S4.1 for example plasmid schematics). Crosstalk between EnvZ and CpxR has been well characterized^{33,126}, making it an obvious choice for targeted crosstalk. Additionally, EnvZ exhibits in vitro crosstalk to CusR¹²³, and although previously not characterized in vivo, we have detected significant activation of CusR by Taz. Over-expression of Taz was sufficient to detect some increased CusR and CpxR reporter activity even in an E. coli strain containing all native TCSs (Figure S4.2). Removing the native OmpR, CusS and CpxA would be expected to increase crosstalk since OmpR may act as a competitive inhibitor of non-cognate RRs and the kinase and phosphatase activities CusS and CpxA may reduce the relative effect of changes in Taz crosstalk^{127,128}. Deletions of the chromosomal OmpR, EnvZ, CusR, CusS, CpxR and CpxA genes from the E. coli strain BW27783¹²⁹ were made resulting in strain WW62 used for all following experiments in this study. Reporter plasmids were constructed with CusC or CpxP promoters driving fluorescent proteins as well as inducible expression of CusR or CpxR respectively. As expected, the deletion of native TCS pathways led to increased Taz-CpxR and

Taz-CusR crosstalk, displaying significant Taz-dependent activation of RRs at lower Taz expression levels (Figure S4.2).

To amplify the low level of non-cognate crosstalk in a controllable manner, we translationally fused protein-protein interaction domains to the TCSs to increase local concentrations and thus phosphotransfer rates (Figure 4.1A). Though a large number of domains are available for this purpose^{130,131} we chose the Crk SH3 domain and ligand due to its independent folding, tight affinity as low as $K_D=100$ nM, and successful use in previous synthetic biology applications^{7,40,132}. Rather than attempting to precisely control component orientation, the domain or ligand was fused to the C-terminus of the HK or RR, respectively, with linkers composed of 12 residues of glycine and serine residues. These linkers are predicted to be unstructured and flexible, thus the engineered assembly is designed solely for the purpose of increasing the local concentration of colocalized components. At low expression levels of Taz and CusR or CpxR, reporter activation can be seen only in the presence of engineered interaction(s) between SH3-fused Taz and ligand-fused CusR (Figure 4.1B) and CpxR (Figure 4.1C), tested independently. As the number of binding domains fused to Taz increased and as the affinity of the peptide ligand increased, the reporter fluorescence, and thus the phosphorylation, of both CusR (Figure 4.1B) and CpxR (Figure 4.1C) also increased. These results are consistent with our hypothesis that colocalization can be used to increase phosphotransfer rates between otherwise weakly cross-talking TCS pairs.

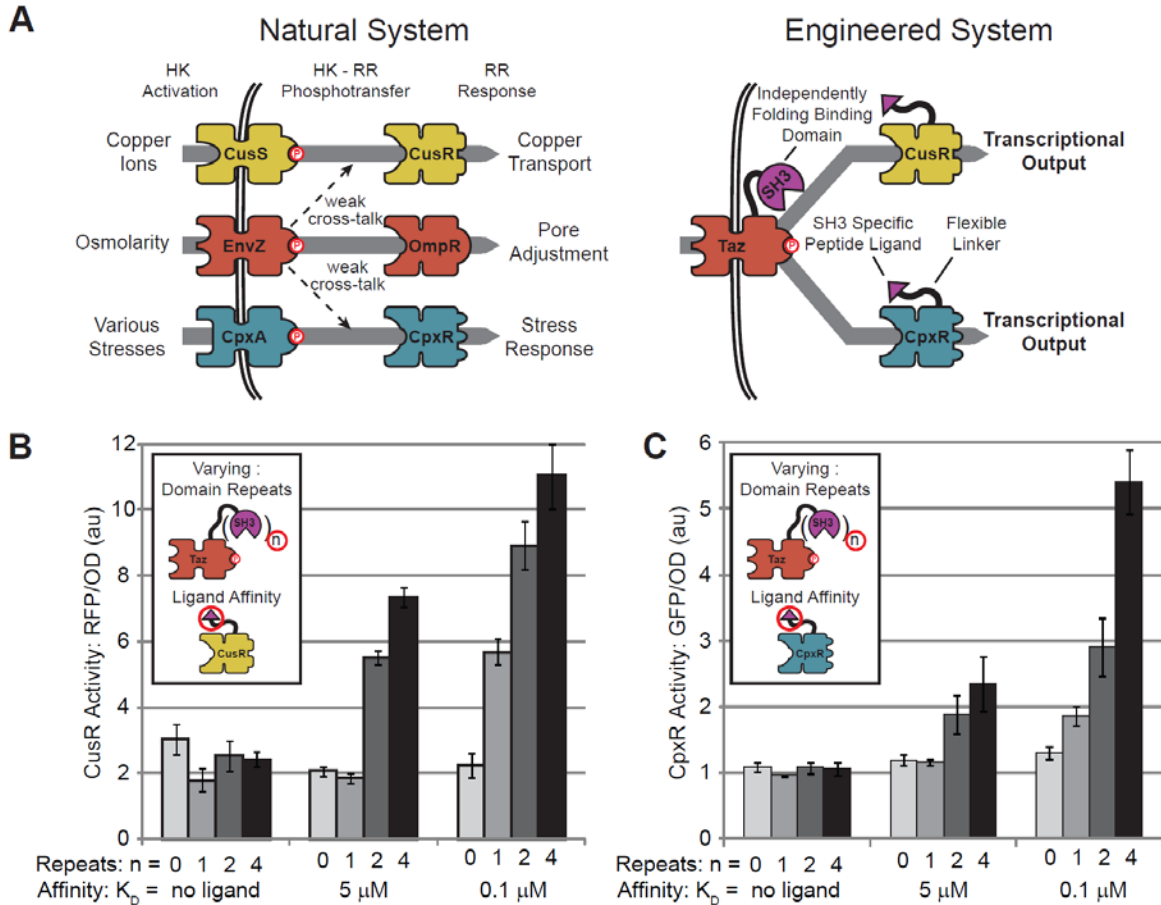


Figure 4.1. Phosphotransfer to a noncognate response regulator can be directed by synthetic assembly with a histidine kinase. (A) Highly specific natural TCSs exhibit weak phosphotransfer crosstalk to non-cognate RRs. Colocalization can increase non-cognate phosphotransfer to physiologically relevant levels through a local concentration effect. To demonstrate this effect in vivo one or several SH3 domains were fused to the Taz HK, while a corresponding SH3 binding peptide was fused to either CpxR or CusR. Knockouts of endogenous OmpR, EnvZ, CusR, CusS, CpxR, CpxA were made and components were expressed from plasmids. Phosphorylated CusR or CpxR drive transcription of fluorescent reporters encoded on plasmids. (B) Activation of CusR was measured by via fluorescence from a CusC promoter driven RFP at late log phase and normalized such that background cell autofluorescence measures 1au. Error bars represent standard error from three independent experiments, each with four biological replicates. Taz was expressed with 0, 1, 2, or 4 SH3 domains tethered with repeating glycine-serine flexible linkers. CusR was expressed fused either to no ligand, a single weak ($K_D = 5 \mu\text{M}$) ligand, or a single tight ($K_D = 0.1 \mu\text{M}$) ligand. CusR activation increased in a manner dependent on increased affinity and number of recruitment domains. (C) Activation of CpxR was similarly measured via fluorescence from a CpxP promoter driven GFP. As in (B), Taz was expressed with a varying number of SH3 domains and CpxR was expressed with no ligand, a single weak ($K_D = 5 \mu\text{M}$) ligand, or a single tight ($K_D = 0.1 \mu\text{M}$) ligand. CpxR activation similarly increased in a manner dependent on increased affinity and number of recruitment domains.

4.3.2 – Synthetic scaffolding directs histidine kinase specificity

Scaffold proteins were designed to recruit either CusR or CpxR to Taz, allowing Taz specificity to be redirected through expression of alternative scaffolds (Figure 4.2A). Each scaffold contained an SH3 peptide ligand, targeting it to a Taz that has been fused to four SH3 domains, each of which was separated by 12-residue glycine–serine linkers, Taz-(SH3)₄. Each scaffold additionally includes one of two different synthetic leucine zippers fused to the SH3 peptide also with a 9-residue glycine–serine linker. Leucine zippers were chosen due to the large number of available pairs⁸⁷, availability of high affinity pairs⁹⁵, and successful use in previous synthetic biology applications^{2,133}. Each scaffold leucine zipper, here termed LZA and LZB (see Table S4.4 for detailed information on interaction parts) incorporated into the scaffold was specific for a corresponding zipper, LZa and LZb, that was fused to either CusR or CpxR, respectively, with a 10 or 12-residue glycine–serine linker. Taz-(SH3)₄, CusR-LZa, and CpxR-LZb were simultaneously expressed at constant levels while the expression of the two scaffolds, SH3pep-LZA and SH3pep-LZB, were induced as indicated. To monitor pathway flux, CusR and CpxR reporters driving RFP and GFP expression were used, respectively. Figure 4.2B shows alternating between scaffold expression switches specificity of the HK; reporter activation occurred only when the scaffold capable of colocalization to the HK is expressed. Scaffolds were also made with two point mutations that changed the PxxP motif in the SH3 peptide to AxxA, which severely decreases affinity for the SH3 domain¹³⁴. These scaffolds were incapable of activating their corresponding RR (Figure 4.2B). Switching expression between scaffolds resulted in a 17-fold change in the ratio of RFP/OD to GFP/OD (56-fold with background autofluorescence, i.e. 1 au, subtracted), showing that even with long flexible linkers colocalization can dramatically influence phosphotransfer signaling rates.

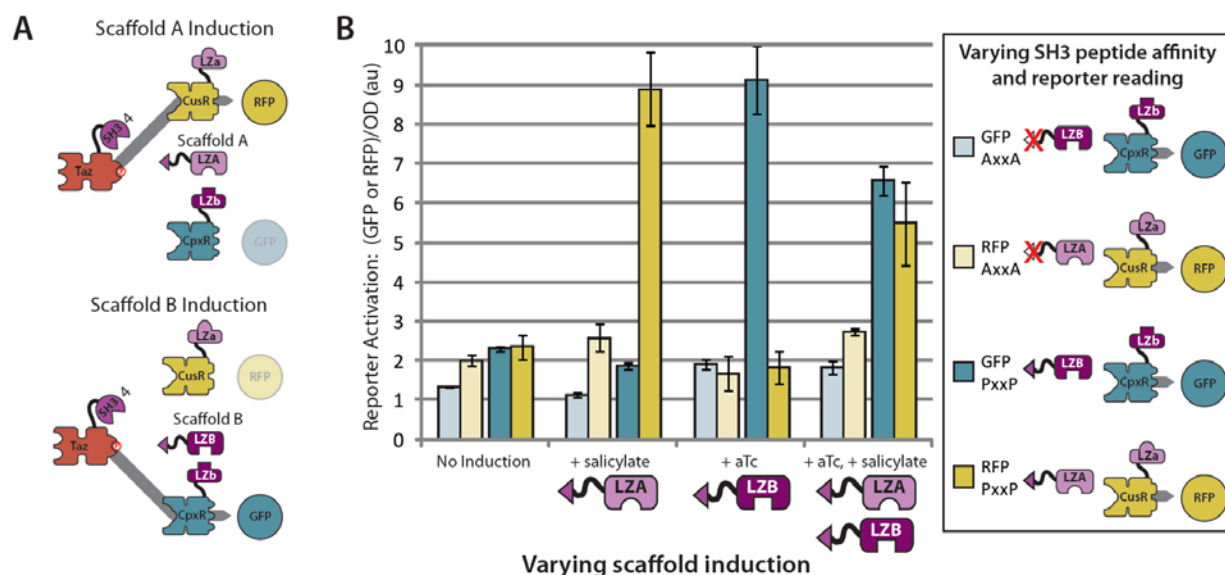


Figure 4.2. Phosphotransfer flux can be programmably directed with synthetic scaffolded HK/RR assemblies. (A) Scaffolds were designed to colocalize Taz with either CusR or CpxR, which enabled directed phosphotransfer depending on which scaffold was expressed. (B) Scaffolds were composed of an SH3 peptide fused to one of two different leucine zippers, LZA and LZB, each specific to a corresponding leucine zipper, LZa or LZb, that was fused to either CusR or CpxR respectively. Non-functional control scaffolds were created by mutating two proline residues in the SH3 peptide to alanine, changing the PxxP motif important for binding the SH3 domain to AxxA. Taz-(SH3)₄, CusR-LZa, and CpxR-LZb were expressed at constant levels. One strain expressed functional scaffolds for localization of CusR and CpxR to Taz under either salicylate or anhydrous-tetracycline (aTc) inducible promoters respectively. Another strain similar expressed both non-functional (AxxA) scaffolds. Inductions of the two scaffolds were varied and CpxR-driven GFP and CusR-driven RFP were measured for the strain with functional scaffold (dark blue and dark yellow) and for strain with non-functional scaffold (light blue and light yellow). High RFP and GFP signal was only observed with induction of the corresponding functional scaffold.

4.3.3 – Specificity control with scaffolding is sensitive to component concentration

Achieving scaffold-directed control of phosphotransfer required careful expression optimization of each signaling component. To investigate this dependency, expression levels were varied for both the HK, Taz-(SH3)₄, and RR, CusR, either with or without a SH3 peptide for colocalization. CusR activation was measured via the expression of GFP reporter. Three expression levels, representing low (i), moderate (ii), and high (iii) induction levels are plotted in Figure 4.3A. Colocalization produced a large increase in signaling only at moderate induction. To illustrate the concentration regime in which colocalization was effective, the fold change in reporter activity between without and with SH3 peptide fused to CusR is plotted for each HK and RR induction level (see source data for individual fluorescent values). As expected from Figure 4.1, colocalization via direct recruitment substantially increased signaling, but this effect was only observed within a limited range of HK and RR

concentrations (Figure 4.3B). This sensitivity to component concentrations is consistent with the hypothesis that phosphotransfer is modulated by an increased local concentration effect. At very low induction of Taz and CusR, colocalization was not sufficient to achieve high activity, while at high induction of Taz and CusR signaling occurred even in the absence of colocalization.

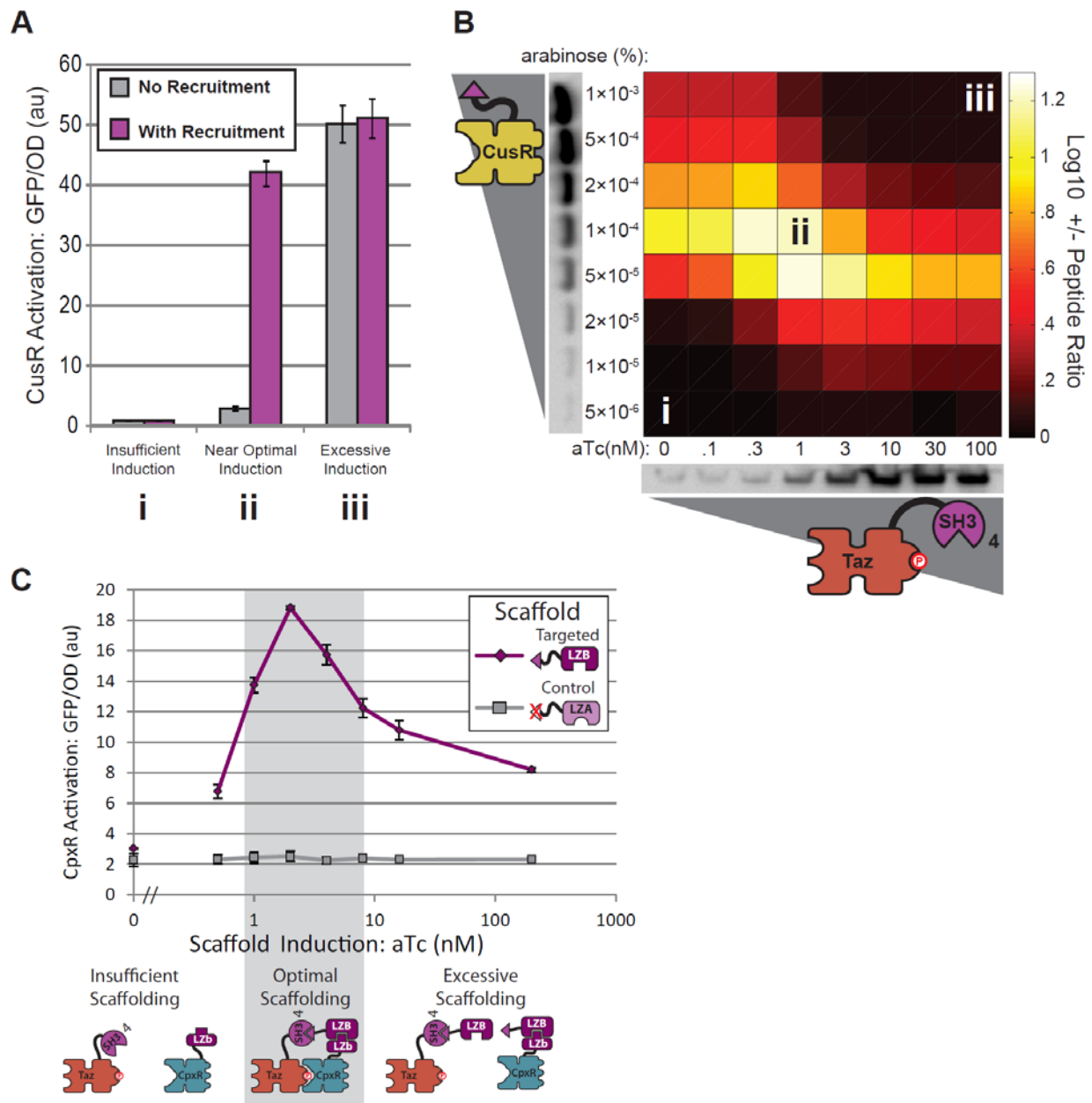


Figure 4.3. Colocalization controlled phosphotransfer is sensitive to component concentrations. (A) Expression of Taz-(SH3)₄ was titrated with the aTc-inducible promoter, P_{tet^r}, and expression of CusR both with and without SH3 peptide fusion was titrated with the arabinose-inducible promoter, P_{BAD}. Representative conditions including under-expression, near optimal expression, and over-expression of components are marked (i), (ii) and (iii),

respectively. (B) Expanding on plot (A), for each induction level, the ratio representing the corecruitment effect was calculated by dividing the GFP/OD measured for CusR with SH3 peptide by without SH3 peptide fusion, with a maximal 17-fold increase in GFP/OD with peptide addition (see source data for individual GFP/OD measurements). To estimate relative concentrations of components, Western blots of epitope tagged HK and RR versions are included along the axes. (C) Taz-(SH3)₄ and CpxR-LZb were expressed at a constant level, while scaffold expression from the P_{tet} promoter was varied via aTc titration. As induction of scaffold designed to colocalize Taz and CpxR was increased, CpxR activity displayed a biphasic response, giving a maximum efficiency at an intermediate scaffold concentration. As induction of a control scaffold (not capable of colocalizing CpxR or Taz) was increased, CpxR activity remained low as expected.

In addition, scaffold concentrations must also be optimized to achieve efficient corecruitment of HK and RR. As scaffold induction is increased we expect activity to display a biphasic behavior, initially increasing up to a maximum followed by a decrease as excess scaffold competition for components results in sequestration of pathway components, (i.e., fewer ternary complexes of scaffold:HK:RR) as predicted for scaffolding of MAP kinase STPs¹⁰⁵ and experimentally verified specifically for the Ste5 scaffold¹³⁵. We tested this prediction in our system with constant expression of Taz-(SH3)₄ and CpxR-LZb, while SH3pep-LZB scaffold was induced to different levels. As predicted, CpxR activation initially increased as scaffold expression was induced but a maximum effect was reached after which further scaffold expression resulted in a decrease in CpxR reporter fluorescence (Figure 4.3C). This biphasic response was consistent with an initially scaffold-limited regime in which additional scaffold colocalized more Taz and CpxR but increasingly sequestered these components as scaffold was expressed in excess, thereby decreasing colocalization. Thus, achieving near maximal scaffold-dependent signaling required carefully balanced scaffold:targeted protein concentrations.

4.3.4 – Engineering autoinhibition for scaffold dependent activation of histidine kinase

Robustness is a critical feature of natural TCSs, ensuring that despite variation in component concentration: specificity is maintained so that the sensory machinery acts on the correct target RR, and that the RR is set to the phosphorylation level appropriate for the stimulus. Natural TCSs meet these goals primarily through high-kinetic preference of the HK for cognate RRs¹²³ and bifunctional – kinase and phosphatase – activities^{122,136}. We have added the additional goal of scaffold dependent control of phosphotransfer (Figure 4.2A), but Figure 4.3 shows that this quality was not robust to changes in component concentrations. Thus, although scaffold-directed specificity as demonstrated in Figure 4.2B was achievable, programmable control was sensitive to the balance of Taz, RR, and scaffold concentrations. Looking to natural systems, in addition to localizing components, scaffolding can serve to activate components, such as the activation and localization of the Ras kinase, by the KSR scaffold in the mammalian ERK pathway¹³⁷. Combining activation and assembly should prevent undesired signaling in the absence of scaffolding. We have engineered autoinhibition of Taz kinase activity by the addition of an intramolecular SH3/ligand

interaction similar to that seen in many eukaryotic signaling proteins^{40,132,138}. In this design (Figure 4.4A), the HK bore a ligand in addition to a flexibly linked SH3 scaffold domain. Intramolecular interaction of the SH3 domain with the ligand competitively inhibited activation of the RR until displaced by an intermolecularly supplied target peptide. This ensured that high expression of autoinhibited Taz did not contribute to undesired activation of RRs in the absence of targeting of RR to the HK.

Engineering an SH3-dependent autoinhibitory interaction necessitated the addition of both an SH3 domain as well as a corresponding SH3 ligand to the polypeptide fold of Taz. SH3 peptide addition was targeted for the turn between helices of the four-helix bundle in the HisKA (dimerization/phosphoacceptor) domain¹³⁹. This region, located near the catalytic histidine residue, contains residues that come within close proximity to the RR when complexed. Thus binding of SH3 to a peptide inserted into this region has potential to inhibit HK-RR signaling through several potential mechanisms including steric hindrance of RR binding and/or autophosphorylation, or altering phosphotransfer rates through shifting the conformation of the four-helix bundle. Several attempts at directly inserting ligand sequence flanked by flexible linkers or replacing residues at this location eliminated kinase function even in the absence of an SH3 domain. We reasoned that a structured linker might allow incorporation of the eleven amino acid SH3 peptide sequence into this sensitive region. As described in detail in the supplementary information (Figure S4.4), we inserted the peptide flanked by a linker sequence of degenerate residues constrained to the polar/non-polar pattern that would be expected to serve as a structured linker extending the four-helix bundle (Figure S4.4A-C). Linker library patterns were first screened in the absence of engineered colocalization for the ability to retain phosphotransfer activity to CusR at high Taz and CusR concentrations, such that colocalization was not required for signaling. Although several libraries with varied polar/non-polar constraints and unstructured linker libraries failed to produce functional kinases, one library pattern surprisingly returned over 30% of its members with retention of at least 50% of wild-type Taz-CusR activity – measured at HK and RR high expression levels sufficient to phosphotransfer in the absence of colocalization (Figure S4.4D).

Several structured linker library variants, which displayed activity at near wild type levels when lacking SH3 domains, were then tested for SH3 domain-dependent inhibition. One of the most promising candidates, AiTaz(29A), had an inserted high affinity SH3 ligand ($K_D = 0.1 \mu\text{M}$ measured in trans), and with the addition of four C-terminal SH3 domains displayed low basal activity (Figure 4.4B). Without the SH3 domains, AiTaz(29A) displayed near wild-type activity demonstrating that the engineered interaction is required for autoinhibition (Table S4.5). To further verify that the degree of autoinhibition is determined by the affinity of the intramolecular SH3 interaction, we replaced the high affinity peptide ($K_D = 0.1 \mu\text{M}$) with a lower affinity peptide ($K_D = 5 \mu\text{M}$) and a non-functional peptide¹³⁴ (AxxA) to make switches AiTaz(29B) and AiTaz(29C), respectively. All three switches tested without SH3 domain fusions retained near wild type activation of CusR, indicating that the structured linker for AiTaz(29A) could tolerate altered SH3 peptide length and sequence without disrupting kinase function (Table S4.5). The basal activity of AiTaz(29B)-SH3 was slightly

higher than that of the tighter affinity version, AiTaz(29A)-SH3, indicating that the tight affinity was necessary for more complete autoinhibition. The control construct with a non-functional peptide inserted, AiTaz(29C), retained near wild-type activity with zero, one, or four SH3 domains (Table S4.5) indicating that functional SH3 peptide was necessary for autoinhibition. Expression of a separate, external SH3 peptide was capable of releasing autoinhibition fully for the low affinity AiTaz(29B)-SH3 construct and partially for the high affinity, AiTaz(29A)-SH3 versions (Figure S4.4F). However, release of autoinhibition via a single SH3 peptide required a higher expression of this peptide than would be presented by scaffold at levels previously (Figure 4.3C) found to be optimal.

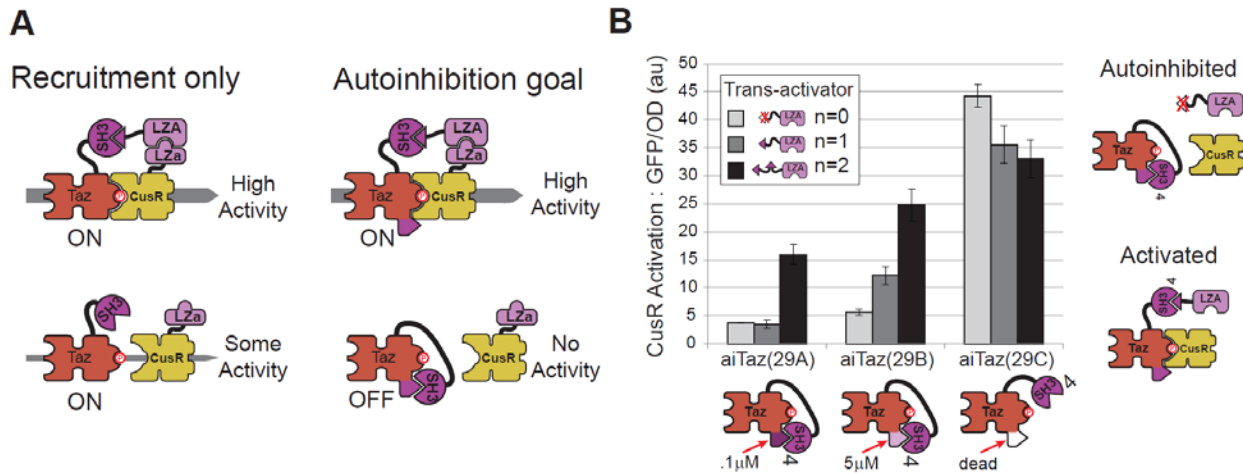


Figure 4.4. Engineering scaffold controlled activation of HK, via competitive autoinhibitory interaction, can be used to reduce non-targeted signaling. (A) Although colocalization alone is sufficient to increase phosphotransfer rates, significant phosphotransfer may still occur in absence of colocalization, particularly at high component concentrations. To limit phosphotransfer in the absence of an engineered interaction, we introduced an autoinhibitory SH3 interaction, such that scaffold may both localize HK-RR and activate HK. (B) CusR, lacking any engineered interaction motifs, was highly expressed to test scaffold dependent activation of HK in the absence of engineered colocalization. Activity of Taz switches AiTaz(29A)-(SH3)₄, AiTaz(29B)-(SH3)₄ and AiTaz(29C)-(SH3)₄ with SH3 ligands of tight ($K_D=0.1 \mu\text{M}$), intermediate ($K_D=5 \mu\text{M}$) or non-functional (AxxA) affinity respectively were tested with scaffolds with various input ligand sequences. In the absence of functional trans-activating peptide (n=0) the activity of the HK decreased in a manner dependent on the affinity of the engineered interaction. Expression of scaffold containing one SH3 ligand or especially two SH3 ligand peptide(s) competitively activated the switch. See Table S4.5 for data on all combinations of kinases, number of fused domains, and identity of trans-acting peptide.

To achieve trans-activation of the autoinhibited HK at lower levels of scaffold, we constructed a scaffold containing two SH3 peptides, rather than the single peptide version used previously, and tested trans-activation of an autoinhibited kinase with four SH3 domains, AiTaz(29A)-(SH3)₄, so that multivalent peptide binding may better compete with the

intra-molecular interaction. Figure 4.4B demonstrates scaffold dependent activation of the autoinhibited Taz. The scaffolds expressed were composed of a leucine zipper fused to one non-functional, one tight affinity ($K_D=0.1 \mu\text{M}$), or two tight affinity SH3 peptides. The AiTaz(29B)-(SH3)₄ was activated by scaffolds with one or two SH3 high affinity peptides, while the scaffold with two peptides, potentially through a cooperative multivalency effect, were considerably more effective at releasing autoinhibition of the AiTaz(29A)-(SH3)₄. The AiTaz(29C)-(SH3)₄, containing a non-functional peptide insertion and displaying no apparent autoinhibition, was not affected by scaffold expression. In these experiments, scaffold was used solely as an activator as CusR did not have a tethered scaffold-targeting ligand. These results demonstrate that we have engineered an autoinhibited histidine kinase that exhibits scaffold-dependent activation.

4.3.5 – Autoinhibition improves robustness of scaffold-directed specificity

Scaffold-dependent HK activity should improve robustness to varying expression levels by lowering phosphotransfer in the absence of a synthetic assembly. For the purpose of testing a wide range of Taz concentrations, expression was driven with the Ptet promoter with either of two different strength ribosome binding sites allowing for a titration spanning more than a 300-fold range as approximated by Western blot intensity quantification (Figure S4.3B). The scaffold consisted of two SH3 peptides and a leucine zipper, LZB, for CpxR-LZb recruitment. Without the incorporation of autoinhibition, scaffold expression significantly increased activation of the targeted CpxR-LZb at intermediate expression of Taz-(SH3)₄ (Figure 4.5A). However, the fold change in activity due to scaffold varied substantially as Taz expression was changed (Figure 4.5B), consistent with previously characterized sensitivity to HK concentration (Figure 4.3B). Particularly problematic was the activation of CpxR in the absence of scaffold when Taz expression levels are high. When autoinhibition was introduced, both with and without scaffold, a striking insensitivity to changes in Taz concentrations was shown: a fixed amount of pathway flux was achieved in the presence of scaffold while only near background fluorescence was measured in the absence of scaffolding even as Taz expression was increased by over 300-fold (Figure 4.5B). While the maximum ratio of GFP/OD for +/- scaffold (Figure 4.5C) was similar at 6.5 and 5.9 fold without and with autoinhibition, respectively (11 and 20 fold, respectively, with background fluorescence subtracted), the effect was far more sensitive to kinase expression levels in the absence of autoinhibition. These results show that by incorporating the autoinhibitory interaction into Taz, although we have increased the sophistication of a component in this system, we were able to reduce the complexity of the system response, as the device output was sensitive to fewer component concentrations. Rather than carefully balancing kinase concentrations to achieve scaffold directed specificity, any autoinhibiting kinase expression level within a wide range should be sufficient for scaffold control, simplifying the design process for future applications.

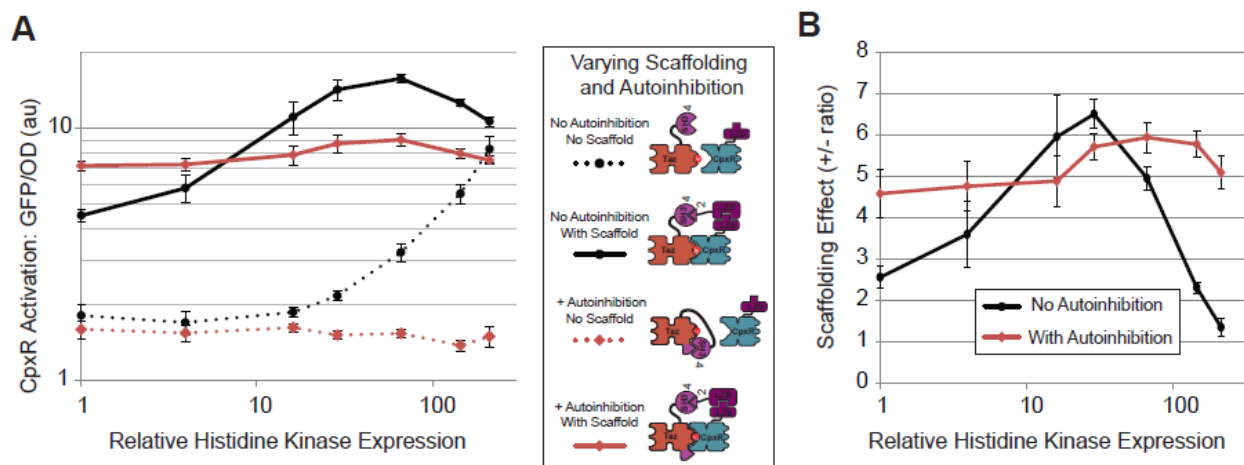


Figure 4.5. Autoinhibition decreased sensitivity to HK concentrations. (A) Titration of HK concentrations across a more than a 200-fold range was achieved by titrating P_{tet} inducer, aTc, for two expression vectors including either a weak or a strong RBS upstream of Taz-(SH3)₄ or AiTaz(29A)-(SH3)₄. Relative kinase concentrations, as estimated by Western blot densitometry (Figure S4.3 B), are plotted on the x-axis. CpxR-LZb was expressed at a constant level and CpxR activity is measured via GFP/OD in the presence of Taz with and without autoinhibitory interaction and with (SH3pep)₂-LZB scaffold expression and without scaffold. In absence of autoinhibition, CpxR activation increased for the scaffolded assembly as HK concentrations were increased (solid black) while the unscaffolded system also began to phosphotransfer, though at the higher HK expression levels (dotted black). With the addition of autoinhibition, the GFP/OD remained near background levels in the absence of scaffolding (dotted red) even as the high strength RBS construct was fully induced, and CpxR activation with scaffold expression (solid red) also showed reduced variation. (B) The fold change in GFP/OD upon expression of scaffold is plotted for the systems with and without autoinhibition. The effect of scaffolding was far more variable in the system lacking autoinhibition – producing a high scaffold effect only at intermediate HK concentrations, while the scaffolding effect remained steady for the autoinhibited HK as its concentration was varied.

4.4 – Discussion

This work demonstrates that although TCS signaling components have not evolved to employ scaffolding, synthetic scaffolds made from well-characterized, modular interaction domains can be used to control TCS specificity. Switching expression between two synthetic scaffolds was used to selectively direct phosphate from Taz to either the CusR or CpxR RR depending on the identity of the scaffold expressed. Thus, the simple act of colocalizing signaling components, thought of as a primary function of most eukaryotic scaffolds, was sufficient to control phosphotransfer specificity of TCSs. However, to improve the robustness of scaffold control to varying component concentrations, it was necessary to also incorporate activation upon colocalization, a characteristic often present in natural scaffolded pathways^{137,140}. This work demonstrates that the principles underlying the robust, yet highly evolvable, signaling pathways in higher organisms may be generalized, abstracted, and

applied to prokaryotes using well-characterized, modular signaling enzymes, and protein-protein interaction domains.

There are several constraints that still must be addressed before applying this approach more generally to other two component systems. We have focused on two non-cognate pairs shown to weakly crosstalk in vitro: EnvZ-CpxR and EnvZ-CusR¹²³. Other non-cognate pairs lacking detectable in vitro crosstalk may also lack physiologically relevant phosphotransfer rates even when co-assembled. Increasing the promiscuity of HKs to allow signaling to a wider range of RRs would be ideal. A solution may come from studies that show mutations in the specificity determining residues^{123,141} can be used to tune RR specificity. Through a series of mutations to the specificity-determining residues of EnvZ, phosphotransfer specificity shifts from a cognate to non-cognate RR, including intermediates with increased promiscuity¹⁴¹. In addition to increasing phosphotransfer to a particular non-cognate target, this offers a promising approach to increase the promiscuity of components either generally or at least to a predefined set of targets. Another challenge is that it is common, as was observed for both Taz/CusR and Taz/CpxR tested in this study, for non-cognate pairs displaying crosstalk to display constitutive activity rather than coupling activity with stimulus. In one study, the CpxA-OmpR non-cognate HK-RR pair is shown to exhibit weak crosstalk, but this crosstalk consists of constitutive phosphotransfer regardless of the environmental stimulus of CpxA¹²². Siryaporn *et al.* further demonstrates that with a series of mutations to CpxA, they achieve a stimulus-dependent switch for phosphorylation of non-cognate OmpR, including dephosphorylation in the absence of stimulus. The remarkable successes uncovering the mechanisms governing natural TCS signal transduction specificity provide a number of promising methods for tuning TCS components for engineering applications.

As engineers are challenged to design cells that dynamically respond to environmental and intracellular stimuli, reliably developing sophisticated signaling pathways will become increasingly important. Organizing signaling pathways with scaffold control allows for the dynamic rewiring of pathways so that stimuli having different meaning in different contexts can be processed appropriately. Additionally, scaffolding may facilitate the incorporation of novel sensory function found in TCSs by providing a tool for routing signal to the system of interest. Although simply switching the TCS output would be most easily accomplished by switching the gene expressed by a cognate RR-driven promoter, there are cases where phosphotransfer control is advantageous or necessary. Changing between signaling outputs that control transcription (most RRs) and protein localization (such as CheY) will need to occur at the post-translational level. Directing output to protein localization, instead of transcription control, could provide a powerful tool for actuating a wide variety of responses, even in organisms with different transcriptional control machinery. Since localization can be both an input and output of TCS pathways, several scaffold-controlled pathways can be linked for multi-layer processing at fast time-scales. Although our system followed a canonical structure, with the stimulus-detecting HK directly phosphorylating the transcription-regulating RR, the application of scaffolding to create longer phosphotransfer relays, as seen in natural hybrid TCSs, is a promising extension, particularly if more

sophisticated processing or additional control points are desired. The *B. subtilis* sporulation pathway uses TCS phosphotransfer structures to build a sophisticated signal transduction pathway with multiple points of feedback control¹⁴². Multiple phosphotransfer steps may also offer an alternative to maintain stimulus sensitivity, where the stimulus-dependent phosphorylation / dephosphorylation control could occur upstream with a cognate pair used for the first phosphotransfer interaction, while constitutive phosphotransfer may be sufficient for downstream scaffold controlled interactions. Phosphorelays built from TCS components offer a particularly promising direction for scaling to larger circuits as the highly conserved structure aids incorporation of the large number of natural parts, these parts share phosphate as a common currency, each component may be connected to many others, and scaffolding can serve as modular, even dynamic, wires to connect components.

Natural systems have achieved powerful processing capabilities through the integration of transcriptional, translational, and post-translational control. Building signaling systems from well-characterized, modular, independently tunable parts will facilitate rational and predictable design of novel signaling pathways and post-translational control modules. Incorporating the post-translational layers of control into genetic circuits will bring us closer to the ultimate goal of designing the complex regulatory networks necessary for adaptive behavior vital to living cells.

4.5 – Materials and Methods

4.5.1 – Strains

The WW62 strain containing knockouts of OmpR, EnvZ, CpxR, CpxA, CusR, CusR was derived from the BW27783¹²⁹ an *E. coli* K12 strain, with standard procedures for creating markerless knockouts¹⁴³ and P1 phage transduction. BW27783 was chosen as a base due to its ability to support titratable mono-modal PBAD induction and its derivation from the widely used BW25113 strain¹⁴⁴. See Table S4.1 for detailed strain modification process and Table S4.2 for primer sequences used to generate knockout strains.

4.5.2 – Plasmid construction

See supplemental information for plasmid sequence files, Table S4.4 for summary of plasmids used, and Figure S4.1 for illustration of plasmid expression machinery. Plasmids in this study encoded up to seven operons on two or three independent plasmids. One plasmid with a BBR1 origin¹⁴⁵ and Kanamycin resistance pertained to the response regulators CusR and/or CpxR expressed under a unique inducible promoter, either P_{BAD} or $P_{Rham'}$, as well as a corresponding reporter CusC or CpxP promoter driving a fluorescent protein upon RR phosphorylation. Another plasmid with the p15A origin and Chloramphenicol resistance pertained to HK expression with Taz expression driven by either P_{tet} or a constitutive promoter. A final plasmid with the ColE1 origin and Ampicillin resistance was used for scaffold expression with one or two unique promoters, P_{tet} or $P_{salicylate}$ driving expression of a particular scaffold.

Plasmids were constructed with a hybrid BglBrick¹¹⁰ strategy, with the open reading frame of HK, RR or scaffold proteins lacking a stop codon between the BglII and BamHI restriction sites so that a BglBrick assembly could be used to add domains creating a glycine serine scar. Additionally, unique restriction sites were maintained for altering expression machinery. All expression levels were tuned with RBS libraries with degenerate oligo-nucleotide insertion (see Table S4.2 for primer sequences). All operons are flanked with SpeI and XbaI sites similar to the BioBrick strategy¹⁰⁹ so that operons could be easily composed, with the exception of the dual reporter plasmids necessary to prevent terminator bleed-through.

4.5.3 – Culture method and fluorescence assays

Plasmids were transformed into chemically competent WW62 strains using standard methods. LB/agar/antibiotic plates containing transformed strains were stored at 4 degrees for up to five days and inoculated into 300 µL of MOPS rich defined media (MRDM) (Teknova, Hollister, CA) with 0.4% glycerol as the carbon source and appropriate antibiotic in a 96 well plate for experimentation. This culture was grown for 10 hours at 37 degrees in an ATR Multitron plate shaker at 1000 RPM, then diluted 1:40 into 400µL of pre-warmed MRDM and similarly grown for 2 hours, and then diluted 1:10 into MRDM containing inducers for the experimental condition inducer concentrations. Strains were then similarly grown for 3 to 4 hours until late log phase and immediately measured. Measurement was done in a TECAN Safire2 machine with OD absorbance at 600 nm, GFP and RFP excitation/emission of 481/507 nm and 584/607 nm respectively with a 5 nm band width. Fluorescent reading were normalized by OD, though were generally within two-fold between sample OD readings, and then the arbitrary fluorescence per OD value was normalized by the fluorescence per OD of cells lacking fluorescent reporters, such that a value of 1 au would be equal to the cell auto-fluorescence. Although fold-change in GFP or RFP protein production would be theoretically estimated after the subtraction of background cell auto-fluorescence, the auto-fluorescence was not subtracted to minimize exaggerating fold-change calculations with very small denominator. This would be expected to underestimate the effect of scaffolding. All error bars represent standard error between experimental replicates performed on different days, each with at least three biological replicates. For detailed description of experimental conditions and plasmids used see Supplementary Table S4.6.

4.5.4 – Estimating relative component concentrations

Relative component concentrations were estimated from densitometry analysis of Western blots of with epitope-tagged components. Experimental conditions were replicated but with a C-terminal fusion of the triple FLAG tag epitope to one of the components to be quantified. Protein gels were run under standard conditions for 10% Bis-Tris NuPAGE denaturing gels, followed by transfer to nitrocellulose membrane, then labeled with Sigma's monoclonal ANTI-FLAG M2- peroxidase (HRP) antibody using standard procedure. Western blots were exposed on an ImageQuant LAS 4000 (GE Healthcare) and densitometry analysis

was performed on Image J analysis software (NIH) to estimate relative component concentrations.

4.5.5 – Engineering colocalization and scaffolding machinery

See Table S4 for interaction machinery sequences and affinities, and plasmid sequence files for the nucleotide sequences of fusions. Interaction machinery was added to Taz, CusR and CpxR via C-terminal BglBrick fusion. A direct BglBrick fusion between two domains will create a two amino acid glycine-serine linker, but longer flexible linker of glycine-serine repeats totaling 8 or more amino acids was found to be more effective than the scar alone, so a linker sequence of glycine-serine repeats is generally fused via BglBrick addition before fusing the interaction machinery to the TCS components.

4.5.6 – Engineering autoinhibition into Taz

To engineer an autoinhibitory peptide sequence into Taz the seven amino acids, MSEQDGY 293-299, near the turn of the 4-helix bundle were targeted for replacement with a sequence containing an SH3 specific peptide sequence. A two-step PCR was used to incorporate the peptide sequence flanked by linker sequences containing degenerate nucleotides so that a library of linker sequences would be created. The linker sequence was designed such that the degeneracy would allow for sampling of between four and seven different amino acids per linker position from a set of either polar, non-polar or flexible residues (see Table S4.2 for primer sequences used to generate linker library). The structured linker residues were restricted to sample either only polar or non-polar residues in a pattern consistent with a continuation of the 4-helix bundle, as was similarly employed recently for creation of new 4-helix bundle proteins¹⁴⁶. Two additional amino acids from a degenerate set giving small flexible residues directly flanked the SH3 peptide sequence (Figure S4.4 A-C). Several different such degenerate library patterns were designed and inserted into Taz lacking SH3 domains, and one pattern produced members were an insertion did not eliminate autophosphorylation and phosphotransfer function, exhibiting detectable crosstalk to CusR when highly expressed (Figure S4.4 D). Library members found to display near wild-type activity were then fused to either one or four SH3 domains and one member displayed inhibited activity towards CusR only when lacking SH3 peptide (Figure S4.4 E). Further characterization of the most tightly inhibited sequence, MSEQDGY 293-299 FTNSAPPPALPPKRRRTSIEVVE, displayed reversible SH3 peptide dependent autoinhibition (Figure 4.4 and Figure S4.4 F).

4.6 – Acknowledgements

We thank Doug Brownfield, Robin Prince and members of the Dueber, Arkin, and Anderson labs for discussion and advice. We thank Michael Laub for plasmids used in preliminary work. This work was supported by funding from National Science Foundation (NSF) Synthetic Biology Engineering Research Center grant no. EEC-0540879, and NSF grant no. CBET-0756801.

4.7 – Supplemental Information

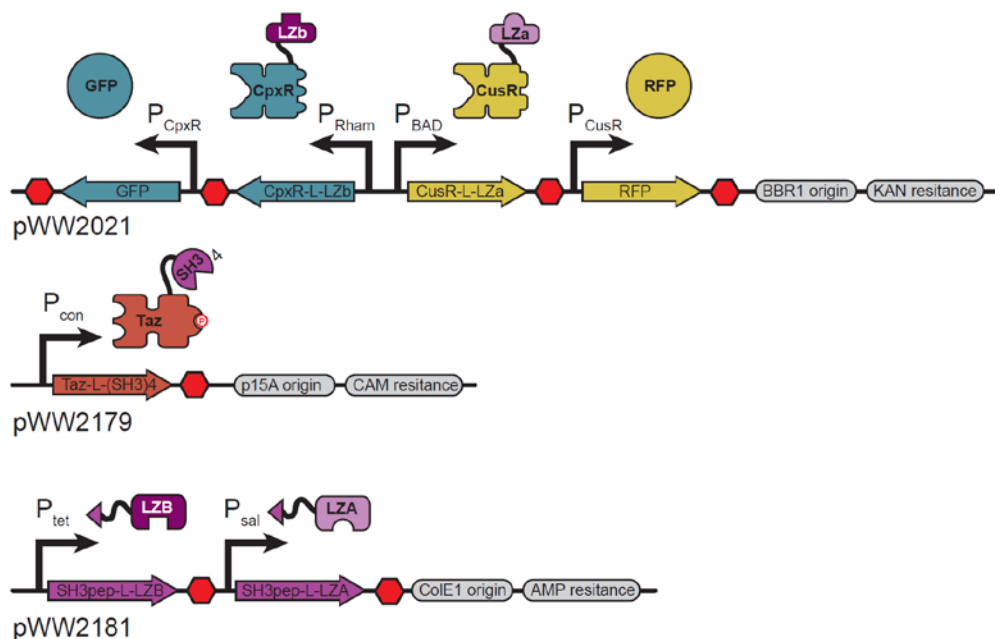


Figure S4.1. Diagrams for plasmids used to demonstrate scaffold directed phosphotransfer (results shown in Figure 4.2). All signaling components are expressed from different promoters such that modification or induction of one component would not be expected to impact other components. Both RRs and both scaffolds can be independently induced with small molecules. Phosphorylated CusR and CpxR drive RFP and GFP expression respectively.

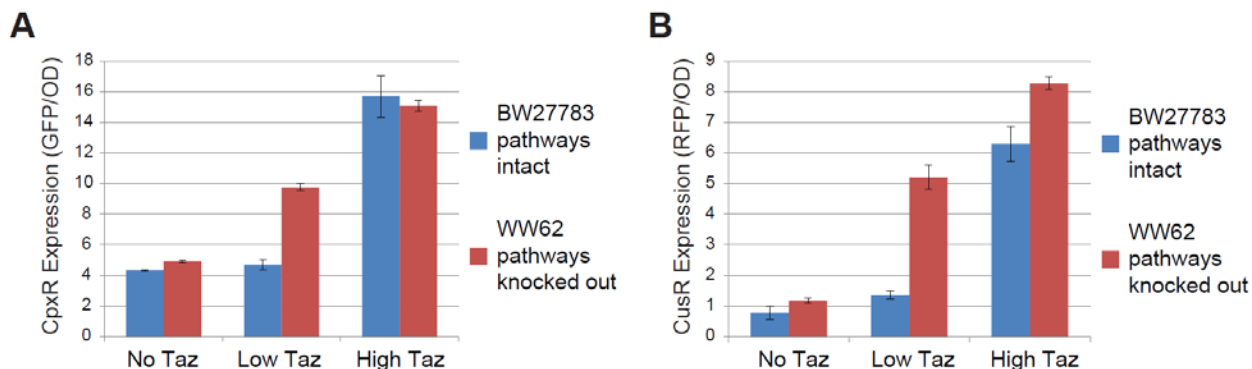


Figure S4.2. Effect of native TCSs on in vivo crosstalk. (A) Activation of the CusR reporter was measured with varying concentrations of Taz in a strain containing all TCS pathways, BW25113, and a strain with the relevant TCS pathways knocked out, WW62. A large increase in activation of CusR by increased Taz expression was seen at lower expression levels in strains lacking the native TCS pathways, which is consistent with native pathway components buffering against crosstalk. (B) CpxR activation was measured with increasing Taz concentrations in BW25113 and WW62 strains. Similarly, activation of CpxR was seen at lower levels of Taz in the WW62 strain lacking native TCS pathways.

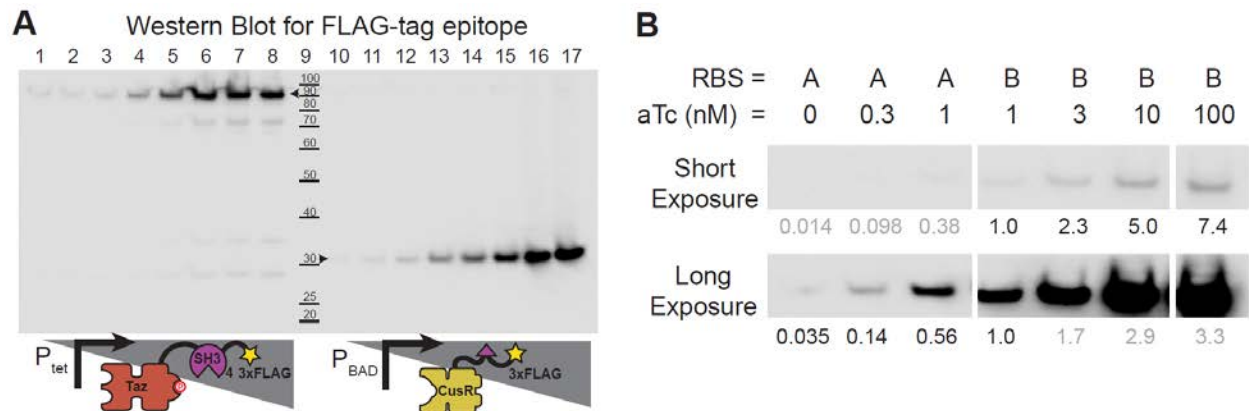


Figure S4.3. Western blots for relative protein concentration estimation. (A) Full Western blot that appears in Figure 4.3B is shown. Western blots of 3xFLAG-tagged Taz or CusR were performed for conditions corresponding to induction in Figure 4.3B to estimate relative concentrations. Prominent bands appear at expected sizes for Taz-(SH3)₄-3xFLAG and CusR-pep-3xFLAG of 90.1 kDa and 31.7 kDa respectively, marked by arrow heads, and increases with small molecule induction. A ponceau stained membrane used as loading control showed even loading. (B) To span a wide range of kinase expression levels for demonstrating reduced concentration dependent variation in scaffold effect with autoinhibition (Figure 4.5), a low and a high strength RBS for the HK are used termed “A” and “B” respectively. Induction of Ptet driven Taz-(SH3)₄-3xFLAG or AiTaz(29A)-(SH3)₄-3xFLAG was varied, via aTc concentration, and densitometry measurements of Western blots were used to estimate concentrations. Comparison of Taz and AiTaz densities for the same RBS and induction levels consistently gave slightly higher Taz than AiTaz values, though with less than a two-fold difference and spanning a nearly identical concentration range. On a single blot all RBS and induction levels were run for Taz-(SH3)₄-3xFLAG with a short and a long exposure used so that high induction points were not saturating the detector and low induction point could be detected. To relate short and long exposure time intensities, intensity was normalized to the RBS = B aTc = 1 nM band which appeared to be within an acceptable intensity range for both exposures. Values obtained from densitometry analysis performed using Image J analysis software (NIH) are shown below each band. Black values correspond to the more reliable reading (grey values were ignored) and were used as an estimation of relative kinase concentration.

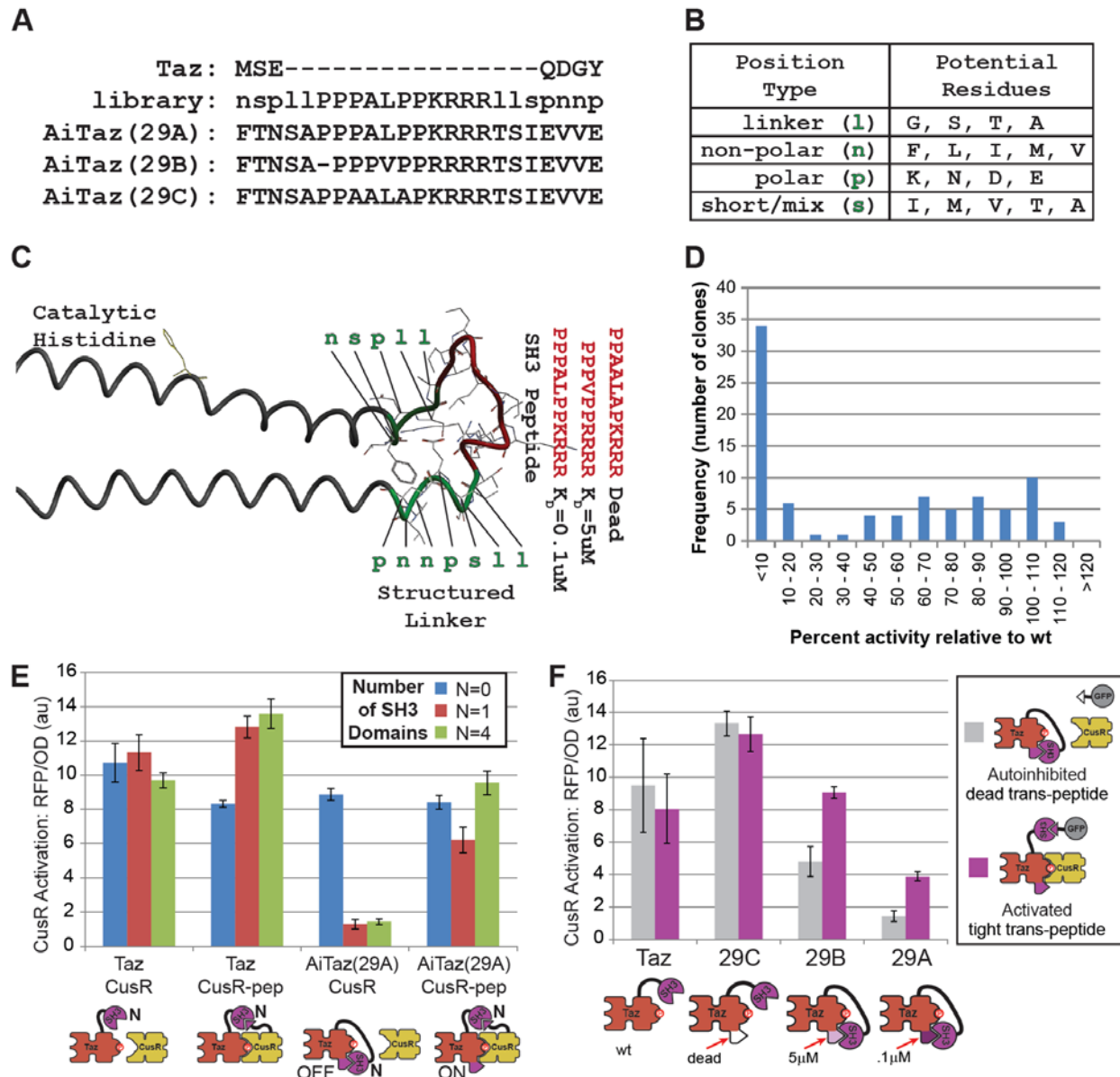


Figure S4.4. Engineering an autoinhibitory SH3 interaction into Taz. (A) The sequence near the turn between helices of the 4-helix bundle of the dimerization/phosphotransfer domain of Taz is shown. This sequence was replaced with a library that included an SH3 ligand flanked by a library of degenerate residues. This library was screened, discussed below, producing a linker sequence capable of incorporating the insertion of three different affinity SH3 ligand. (B) Degenerate nucleotides were chosen to provide different types of residues, each with 4 or 5 possible amino acids per residue. (C) A phylogeny based structure prediction, from the Phyre web server¹⁴⁷ for AiTaz(29A) is shown with degenerate residues labeled. The structured linker library was designed with degenerate residues expected to continue the alpha-helices of the four helix bundle, which is consistent with the predicted structure shown. Different residue positions were constrained to a pattern expected to place non-polar residues in the center of the 4-helix bundle, polar residues at surface-exposed positions, non-polar or short

residues for several difficult to predict positions, and short flexible residues directly flanking the ligand. The degenerate residue pattern found to produce the most successful library is labeled. (D) The histidine kinase linker library, lacking SH3 domains, and CusR were highly expressed, such that non-congrate phosphotransfer in the absence of colocalization or autoinhibition could be screened to reveal linkers that allowed ligand insertions that were not disruptive to kinase function. A remarkably high fraction of library members, 41 of 88, were capable of reaching at least 50% of the full CusR activation reached by non-modified Taz. In contrast, similar libraries made with different polar/non-polar constraints or flexible linkers gave little or no hits. (E) To test for autoinhibition zero, one or four SH3 domains were fused to several of the library members that displayed near wild-type activity, with the expectation that a loss of activity from fusion of SH3 domain to the HK would be indicative of autoinhibition. Unmodified Taz as well as a high performing candidate, AiTaz(29A), are plotted with a varying number of domains, and activation of CusR with or without SH3 ligand fusion was measured. At the high expression levels used, un-modified Taz was capable of fully activating CusR with or without recruitment. AiTaz(29A) lacking SH3 domains was similarly able to activate CusR, but activation substantially decreased with fusion of SH3 domains. Fusion of SH3 ligand to CusR restored activation by the AiTaz(29A) with SH3 domains. These results are consistent with AiTaz(29A) exhibiting reversible autoinhibition. (F) To test the dependency of the SH3 domain fusion on the interaction with the inserted ligand, the tight affinity ($K_D = 0.1 \mu\text{M}$) sequence from AiTaz(29A) was replaced with a weaker affinity ($K_D = 5 \mu\text{M}$) ligand, for AiTaz(29B), and a non-functional ligand where the PxxP motif critical to binding was mutated to AxxA, for AiTaz(29C). All three of these kinases displayed near wild-type activity when lacking an SH3 domain fusion (see Table S4.5). Activation of CusR lacking ligand by HK-SH3 for wild-type Taz and all three ligand versions of AiTaz was measured to determine the autoinhibitory interaction effect on kinase activity without the complicating effect of colocalization. Additionally, to test reversibility of the autoinhibitory interaction, either functional (PxxP) or non-functional (AxxA) trans-acting ligands were expressed as GFP fusions to improve ligand stability and allow easy quantification. The inhibitory effect of SH3 domain fusion was found to be weaker for the weak ligand version, AiTaz(29B), and absent for the non-functional peptide version, AiTaz(29C). Expression of functional SH3 ligand activated the tight and weak affinity AiTaz-SH3, indicating that SH3 peptide presented on scaffold would be capable of activating the autoinhibited kinase.

Strain Name	Genotype	Reference	Base Cell Line	Parent Strain	Modification	Resistance
BW27783	DE(araFGH), Pcp8-araE	(Khlebnikov et al, 2001)	BW25113 (CGSC#: 12119)	-	-	-
WW29	DE(OmpR, EnvZ, araFGH), Pcp8-araE	This study	BW25113 (CGSC#: 12119)	BW27783	Datsenko-Wanner: PCR	KAN
WW32	DE(CusR, CusS, araFGH), Pcp13-araE	This study	BW25113 (CGSC#: 12119)	BW27786	Datsenko-Wanner: PCR	CAM
WW33	DE(CpxR, CpxA, araFGH), Pcp8-araE	This study	BW25113 (CGSC#: 12119)	BW27783	Datsenko-Wanner: PCR	KAN
WW39	DE(CpxR, CpxA, araFGH), Pcp8-araE	This study	BW25113 (CGSC#: 12119)	BW27783	P1 transduction: WW33	KAN
WW48	DE(CusR, CusS, CpxR, CpxA, araFGH), Pcp8-araE	This study	BW25113 (CGSC#: 12119)	WW39	P1 transduction: WW32	KAN, CAM
WW54	DE(CusR, CusS, CpxR, CpxA, araFGH), Pcp8-araE	This study	BW25113 (CGSC#: 12119)	WW48	pCP20/curing	-
WW58	DE(OmpR, EnvZ, CusR, CusS, CpxR, CpxA, araFGH), Pcp8-araE	This study	BW25113 (CGSC#: 12119)	WW54	P1 transduction: WW29	KAN
WW62	DE(OmpR, EnvZ, CusR, CusS, CpxR, CpxA, araFGH), Pcp8-araE	This study	BW25113 (CGSC#: 12119)	WW58	pCP20/curing	-

Table S4.1 - Summary of strains.

Primer name	Sequence	Notes
Ww.F05. EnvOmp KO F	Cacgcttacaaattgttgcgaaacctttggg agtacaacaGTGTAGGCTGGAGCTGCTTC	used to knockout OmpR-EnvZ from BW25113
Ww.F17. EnvOmp KO R	Gtttatttacccttcttttgcgtgccctg cgcccgcgttaccgGAATTAGCCATGGTCC	used to knockout OmpR-EnvZ from BW25113
Ww.F09. CusRS KO F	Cgctactctagaattgccgggcaacatg cggaggaaatGTGTAGGCTGGAGCTGCTTC	used to knockout CusR-CusR from BW25113
Ww.F10. CusRS KO R	Cgttattttttactggttataaaagtgtgc cgtttgcgtgaaggGAATTAGCCATGGTCC	used to knockout CusR-CusR from BW25113
Ww.G39. CpxRA KO F	Tcaggggggtgtaaaacaacgtaaagtcag gattagcgacgtGTAGGCTGGAGCTGCTTC	used to knockout CpxR-CpxA from BW25113
Ww.G40. CpxRA KO R	Gatgcggcgtaaacgccttatcctgcctgc aaatgcgaagttaGAATTAGCCATGGTCC	used to knockout CpxR-CpxA from BW25113
Ww.G01. RBSlib3 F	CTAGCggtaccNNNNNNNNNNNNNNNNNNNNgA	oligos that are annealed to generate a low-moderate strength RBS library
Ww.G02. RBSlib3 R	GATCTcNNNNNNNNNNNNNNNNNNNNnggtaccG	oligos that are annealed to generate a low-moderate strength RBS library
Ww.K60. HLL4 F	TGCCACCGAAACGTCGCCGTRSYRSYRNR ANNTNNTNRANctggcagaatcgatcaata	internal primer for a 2-step PCR to generate structured linker libraries
Ww.K61. HLL4 R	CGTTTCGGTGGCAGCGCTGGTGGCGGRSYR SYNTYRNYNANcatctcagtcgcaaggcga	internal primer for a 2-step PCR to generate structured linker libraries

Table S4.2 – Primers of interest.

Shorthand	Domain Type	Binding Partner (Affinity)	Sequence	Reference
SH3	SH3 domain	SH3pep ($K_D=0.1 \mu\text{M}$), SH3pep5uM ($K_D=5 \mu\text{M}$), AxxA (NA)	MAEYVRALFDFNGNDEEDLP FKKGDILRIRDKPEEQWWNA EDSEGKRGMIIPVYPYVEKY	Posern, G. et al. (1998). Oncogene 16, 1903–1912. Knudsen, B. et al. (1995). EMBO J. 14, 2191–2198.
SH3pep	SH3 peptide	SH3 ($K_D=0.1 \mu\text{M}$)	PPPALPPKRRR	Posern, G. et al. (1998). Oncogene 16, 1903–1912.
SH3pep 5uM	SH3 peptide	SH3 ($K_D=5 \mu\text{M}$)	PPPVPPRRRR	Knudsen, B. et al. (1995). EMBO J. 14, 2191–2198
AxxApep	SH3 peptide	SH3 (NA)	PPAALAPKRRR	Knudsen, B. et al. (1995). EMBO J. 14, 2191–2198
LZA	Leucine Zipper	LZa (NA)	SVKELEDKNEELLSELYHLKN EVARLKKLVGER	Reinke, A. et al. (2010). J. Am. Chem. Soc. 132, 6025–6031.
LZa	Leucine Zipper	LZA (NA)	SVKELEDKNEELLSRIYHERNE VARLKKLVGER	Reinke, A. et al. (2010). J. Am. Chem. Soc. 132, 6025–6031.
LZB	Leucine Zipper	LZb ($K_D=6 \text{ nM}$)	ITIRAAFLEKENTALRTEIAELE KEVGRGENIVSKYETRYGPL	Acharya, A. et al. (2002). Biochemistry 41, 14122–14131.
LZb	Leucine Zipper	LZB ($K_D=6 \text{ nM}$)	LEIRAAFLEKENTALRTRAAEL RKRVGRCRNVSKYETRYGPL	Acharya, A. et al. (2002). Biochemistry 41, 14122–14131.
L	Linker	-	GSGSGSGSGSGS	

Table S4.3 – Protein-protein interaction machinery.

Plasmid #	Description	Resistance	Origin
pWW2179	pCon (low rbs) Taz-L-(L-S)4	CAM	p15A
pWW2204	pCon (low rbs) aiTaz(HL29a)-L-(L-S)4	CAM	p15A
pWW2420	pCon (high rbs) Taz-L-(L-S)4	CAM	p15A
pWW2314	pCon (high rbs) aiTaz(HL29a)-L-(L-S)4	CAM	p15A
pWW2076	pTet (low rbs) Taz	CAM	p15A
pWW2080	pTet (low rbs) Taz-L-(L-S)	CAM	p15A
pWW2081	pTet (low rbs) Taz-L-(L-S)2	CAM	p15A
pWW2037	pTet (low rbs) Taz-L-(L-S)4	CAM	p15A
pSD145	pTet (low rbs) Taz-L-(L-S)4	CAM	p15A
pSD173	pTet (low rbs) aiTaz(library)	CAM	p15A
pSD176	pTet (low rbs) aiTaz(HL29a)	CAM	p15A
pSD205	pTet (low rbs) aiTaz(HL29a)-L-S	CAM	p15A
pSD181	pTet (low rbs) aiTaz(HL29a)-L-(L-S)4	CAM	p15A
pSD209	pTet (low rbs) aiTaz(HL29b)	CAM	p15A
pSD213	pTet (low rbs) aiTaz(HL29b)-L-S	CAM	p15A
pSD217	pTet (low rbs) aiTaz(HL29b)-L-(L-S)4	CAM	p15A
pSD211	pTet (low rbs) aiTaz(HL29c)	CAM	p15A
pSD215	pTet (low rbs) aiTaz(HL29c)-L-S	CAM	p15A
pSD219	pTet (low rbs) aiTaz(HL29c)-L-(L-S)4	CAM	p15A
pWW2224	pTet (high rbs) Taz-L-(L-S)4	CAM	p15A
pWW2254	pTet (high rbs) aiTaz(HL29a)-L-(L-S)4	CAM	p15A
pWW1525	pBad (low rbs) CusR-L, pCusR-RFP	KAN	BBR1
pWW1526	pBad (low rbs) CusR-L-SH3pep(KD=.1uM), pCusR-RFP	KAN	BBR1
pWW1764	pBad (low rbs) CusR-L-SH3pep(KD=5uM), pCusR-RFP	KAN	BBR1
pWW1829	pBad (low rbs) CusR-L-LZc, pCusR-RFP	KAN	BBR1
pWW2244	pBad (high rbs) CusR, pCusR-GFP	KAN	BBR1
pWW2241	pBad (high rbs) CusR-L-SH3pep, pCusR-GFP	KAN	BBR1
pWW2245	pBad (high rbs) CusR-L-LZc, pCusR-GFP	KAN	BBR1
pWW1638	pRham CpxR-L, pCpxR-GFP	KAN	BBR1
pWW1623	pRham CpxR-L-SH3pep(KD=.1uM), pCpxR-GFP	KAN	BBR1
pWW2131	pRham CpxR-L-SH3pep(KD=5uM), pCpxR-GFP	KAN	BBR1
pWW2006	pRham CpxR-L-LZx, pCpxR-GFP	KAN	BBR1
pWW2021	pBad (low rbs) CusR-L-LZc, pCusR-RFP // pRham CpxR-L-LZx, pCpxR-GFP	KAN	BBR1
pWW2263	pBad (high rbs) CusR-L-LZc, pCusR-RFP // pRham CpxR-L-LZx, pCpxR-GFP	KAN	BBR1
pWW2005	pTet SH3pep-L-LZX	AMP	ColE1
pWW1835	pTet SH3pep-L-LZC	AMP	ColE1
pWW2172	pTet AxxApep-LZX (dead scaf)	AMP	ColE1
pWW2173	pTet AxxApep-LZC (dead scaf)	AMP	ColE1
pWW2232	pTet (SH3pep-L)2-LZX	AMP	ColE1
pWW2231	pTet (SH3pep-L)2-LZC	AMP	ColE1
pWW2321	pSal SH3pep-L-LZC	AMP	ColE1
pWW2322	pSal AxxApep-L-LZC	AMP	ColE1
pWW2181	pTet SH3pep-L-LZX, pSal SH3pep-L-LZC	AMP	ColE1
pWW2353	pTet AxxApep-L-LZX, pSal AxxApep-L-LZC	AMP	ColE1
pWW2401	pTet (low rbs) Taz-L-(L-S)4-3xFLAG	CAM	p15A
pWW2402	pTet (low rbs) aiTaz(HL29a)-L-(L-S)4-3xFLAG	CAM	p15A
pWW2422	pTet (high rbs) Taz-L-(L-S)4-3xFLAG	CAM	p15A
pWW2423	pTet (high rbs) aiTaz(HL29a)-L-(L-S)4-3xFLAG	CAM	p15A
pWW2403	pRham CpxR-L-LZx-3xFLAG, pCpxR-GFP	KAN	BBR1
pWW2405	pBad (high rbs) CusR-L-SH3pep-3xFLAG, pCusR-GFP	KAN	BBR1
pWW2406	pTet SH3pep-L-LZC-3xFLAG	AMP	ColE1
pWW2407	pTet SH3pep-L-LZX-3xFLAG	AMP	ColE1

Table S4.4 – Summary of plasmids.

	Avg: Taz	Avg. 29A	Avg. 29B	Avg. 29C	SE: Taz	SE: 29A	SE: 29B	SE: 29C
SH3=0, trans n=0	28.2	34.7	28.5	34.0	1.2	1.8	0.9	1.2
SH3=0, trans n=1	28.7	30.4	29.8	33.2	1.8	1.6	0.6	0.2
SH3=0, trans n=2	21.1	21.7	25.7	24.2	1.0	0.9	1.6	0.5
SH3=1, trans n=0	27.6	2.6	13.8	33.6	1.8	0.3	0.5	1.4
SH3=1, trans n=1	26.9	3.5	22.6	31.6	1.0	0.2	0.5	0.9
SH3=1, trans n=2	18.9	11.4	19.6	29.1	0.7	0.9	0.8	0.7
SH3=4, trans n=0	26.0	2.0	5.1	27.8	2.0	0.3	1.6	1.0
SH3=4, trans n=1	21.6	3.1	9.1	25.6	1.1	0.8	0.5	2.8
SH3=4, trans n=2	14.9	11.9	13.3	19.3	0.4	0.8	0.5	0.7

Table S4.5 – AiTaz characterization. The averages (columns 2-5) and standard errors (columns 6-9) for the activation CusR lacking localization tags are given in the presence of different autoinhibited kinases and scaffolds. Taz corresponds to the HK without insertion of peptide. 29A, 29B and 29C correspond to the Taz engineered to contain a tight, weak or dead intra-molecular peptide. SH3=X, corresponds to the number of SH3 domains fused to the Taz HK. Trans n = X, corresponds to the number of peptides on the scaffold, where zero corresponds to a dead (AxxA) peptide.

Chapter 5 – Internal Ribosome Binding Sites within Heterologous and Synthetic Protein Coding Sequences

5.1 – Abstract

Genetic sequences taken out of their native context for employment in synthetic biology applications can gain unexpected properties. In this study we explore ribosome binding sites (RBS) found within protein coding sequences as predicted from biophysical calculations⁹ and as manifested in N-terminal truncations of protein products. These truncated proteins may be particularly problematic for applications in which the target RBS strength is low and the coding sequence is heterologous. Genome wide prediction of RBS sites based on biophysical calculation suggests a negative selection against RBSs within *E. coli* protein coding sequences, in agreement with previous analysis based solely on the frequency of Shine-Dalgarno consensus sequence occurrences. C-terminal tagged proteins analyzed with Western blots and relative fluorescence assays show that increasing the strength of upstream an RBSs has little influence on the downstream RBSs responsible for truncation. However, software designed to silencing these RBS sites, often with the addition of mRNA secondary structure, effectively removes the N-terminal truncation products. Synthetic biology applications that require expressing protein such that stochastic noise is low and internal RBS truncated product remains a small fraction of total protein, particularly at low expression levels, fundamentally requires a balance between transcription and translation rates for optimal performance.

5.2 – Introduction

Open reading frames (ORFs), in addition to encoding protein sequence, can have a number of embedded regulatory elements that may begin or end transcription or translation. Evidence for the importance of these signals can be seen in recent studies analyzing the representation in natural sequence, which suggested significant selective pressure acts on these signals within ORFs¹⁴⁸. ORFs taken from heterologous organisms or generated synthetically that have not experienced selective pressure against representation of problematic sequence may exhibit behavior that is difficult to predict or interpret when incorporated into an engineered system. In this work we focus on prokaryotic translation initiation sites encoded within ORFs, here termed internal ribosome binding sites (IRBSs).

Alternative translation initiation sites are apparently relatively uncommon naturally, but are utilized by both prokaryotes and eukaryotes despite very different translation machinery. Internal ribosome entry sites (IRESs), which drive cap-independent translation in eukaryotes, appear to be used primarily as regulatory control point¹⁴⁹, but are also used in some cases as a means of generating alternative isoforms of genes¹⁵⁰. In prokaryotes alternative translation initiation sites have been less extensively studied but do appear to play roles for generating alternative isoforms in natural systems for several documented cases^{151–154}. While these alternative translation sites may be exploited to express alternative isoforms,

unintended IRBS sites may be problematic, as documented with several cryptic RBS sites^{155,156} particularly if these sites were not selected for in the source organism. Recent work has also suggested these sites may lead to translational stalling, resulting in altered translation rates¹⁵⁷.

The Shine-Dalgarno sequence (SD) is thought to be the primary indicator of a strong translation initiator sequence. Recent work however, has expanded on this indicator to quantify RBS strength based on free energy of both mRNA-rRNA hybridization and displacement of mRNA secondary structure⁹. That work was prompted in part by observations that the strength of RBSs was highly context dependent, and quantitative accuracy was needed for synthetic biology applications. By employing a statistical thermodynamic model of RNA hybridization and folding to predict the translation initiation rate and an optimization algorithm to vary potential RBS sequence, the RBS calculator was capable of forward engineering RBS strengths spanning a range of five orders of magnitude that were generally within two-fold of the targeted level. This provides a powerful tool for probing natural sequence for translation initiation sites with potentially higher accuracy based on biophysical calculations rather than identity to SD. It has been reported¹⁵⁸ that a significant proportion of translated proteins do not contain a SD, which may be in part due to inaccuracy in relying on percent identity alone as a metric of translation strength. Additionally, this tool provides a means of altering codon usage to reduce IRBS strengths even if the SD sequence is found in highly constraining residues (e.g. arginine: A-G-G/A). This becomes increasingly important at low expression levels of non-prokaryotic or synthetic sequences, where intended translational initiation sites may be significantly lower than IRBS strengths expected to be frequent by random chance.

5.3 – Results and Discussion

5.3.1 – Negative selection against internal RBS sites within *E. coli* ORFs

We analyzed the frequency of IRBSs as predicted by the RBS calculator⁹ within the genomes of *E. coli* and *S. cerevisiae* – analyzed as representative prokaryotic and eukaryotic organisms. All ORFs, excluding hypothetical genes, were analyzed to determine the calculated translation strength associated with each potential start codon. Start codons within 35 base pairs (bps) of the annotated translation start and stop codons were discarded, as they may over-represent biologically relevant initiation sites and present difficulty accurately representing mRNA transcript length needed for mRNA structure calculations (Figure 5.1A). The *S. cerevisiae* ORFs were found to contain a substantially higher frequency of high strength IRBS predictions than *E. coli*, with more than a 200-fold increased likelihood of containing very high IRBS sites, over 10^4 au (Figure 5.1B).

The *E. coli* and *S. cerevisiae* genome ORFs were each shuffled to preserve amino acid sequence, codon usage, and bi-codon pair frequency, as described by Itzkovitz and co-workers¹⁴⁸. Additionally, *E. coli* and *S. cerevisiae* ORFs were each recoded randomly with the codon usage frequencies associated with *E. coli* and *S. cerevisiae*, not preserving bi-codon pairs. The *S. cerevisiae* ORFs both shuffled and recoded with *S. cerevisiae* codon usage all

produced similar, high frequencies of strong IRBSs (Figure 5.1B). The *S. cerevisiae* ORFs recoded with *E. coli* codon usage displayed significantly lower strong IRBS frequency. The *E. coli* ORFs displayed significantly higher frequencies of strong IRBSs when shuffled, and slightly higher still when recoded with *E. coli* codon not preserving bi-codon pairs. The *E. coli* ORFs displayed the highest frequencies of strong IRBSs when recoded with *S. cerevisiae* codon usage. Together, this data suggest there is a significant negative selection against strong IRBS sequences in the *E. coli* genome, and that this selective pressure is evident in both the amino acid sequence and the codon usage.

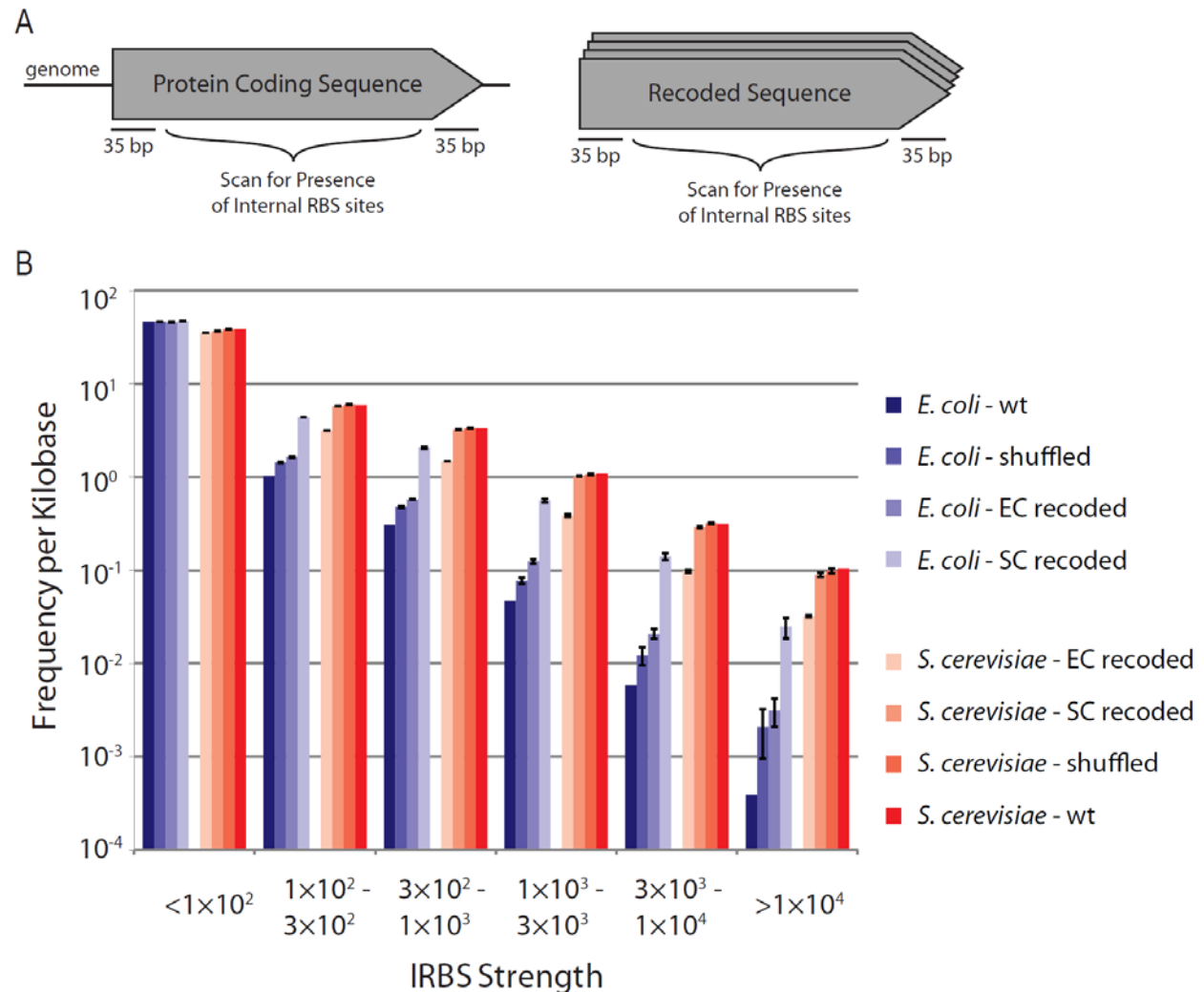


Figure 5.1. Frequency of IRBS sites within natural and recoded sequences. (A) Each protein coding sequence in the *E. coli* and *S. cerevisiae* genomes was analyzed with the RBS calculator⁹, with the exception genes annotated as hypothetical or predicted. The codons for these protein coding sequences were recoded three independent times and analyzed with the RBS calculator. RBSs occurring within the first and last 35 base pairs were ignored. (B) Comparison of internal RBS frequency for *E. coli* (blue) and *S. cerevisiae* (red) protein coding sequences for natural and recoded genomes. The “shuffled” genomes represent recoding that preserves both codon and bi-codon frequencies. EC recoded and SC recoded are ORFs

that have been randomly recoded with codon frequencies associated with *E. coli* and *S. cerevisiae* respectively. Native *E. coli* ORFs (dark blue) show significant reduction in high strength IRBS frequency compared to recoded or *S. cerevisiae* variants, suggesting a selective pressure against these sequences in *E. coli*.

5.3.2 – Internal RBS sites produce truncated proteins

We investigated an in-frame IRBS in a common mRFP1¹⁵⁹ to determine if truncated protein products are present as would be expected from an internal translation start site. The protein was predicted to have a single in-frame IRBS site of 362au, and no other in-frame sites above 100au. The sequence encoding the IRBS, GGTGAA, only bears some resemblance to the canonical SD sequence, AGGAGG, and is missed by algorithms that identify RBS sites by sequence identity alone. By introducing eight silent mutations, only 2 within the canonical RBS sequence giving GGCGAG, which slightly increase percent identity to AGGAGG, this sequence designed to have increased secondary structure to sequester the IRBS is predicted to have the IRBS strength decreased from 362au to 43au. Software is openly available on the Salis Lab website for using the RBS calculator⁹ to recode sequences with IRBSs sites removed.

Both the unmodified and the IRBS-silenced version of mRFP1 were translationally fused to the N-terminus of super-folding GFP¹⁶⁰ with a 12-residue linker composed of repeating glycines and serines that is predicted to be unstructured. A triple FLAG-tag epitope was fused to the C-terminal of the GFP. Translation initiated at the N-terminal of the RFP, corresponding to translation of the full protein, is expected to produce protein fluorescent in both the red and green channels that can be identified at 57 kDa with an anti-FLAG western blot. Translation initiated at the IRBS is expected to produce a truncated protein fluorescent only in the green channel that can be identified at 39 kDa with an anti-FLAG western blot.

To vary the strength of the N-terminal RBS, driving translation at the start of RFP, an RBS library was created for both the unmodified and the IRBS-silenced fusion proteins. A wide range of RBS strengths were selected spanning more than a thousand-fold range as measured by RFP fluorescence after background cell autofluorescence was subtracted. 24 constructs, varying only in upstream RBS, were chosen for the unmodified and the IRBS-silenced fusion proteins, and were analyzed with a Western blot (Figure 5.2B). Lanes were run with alternating unmodified and IRBS-silenced constructs in increasing upstream RBS strength. As expected, as the upstream RBS strength increases, a band corresponding to the full length protein size increases. A band also appears at a size corresponding to the truncated product from IRBS initiation and is consistently less intense from the IRBS-silenced constructs. Notably the intensity of the IRBS band changes very little as the upstream RBS is increased, suggesting that for these constructs, increased upstream RBS strength does not inhibit IRBS translation initiation.

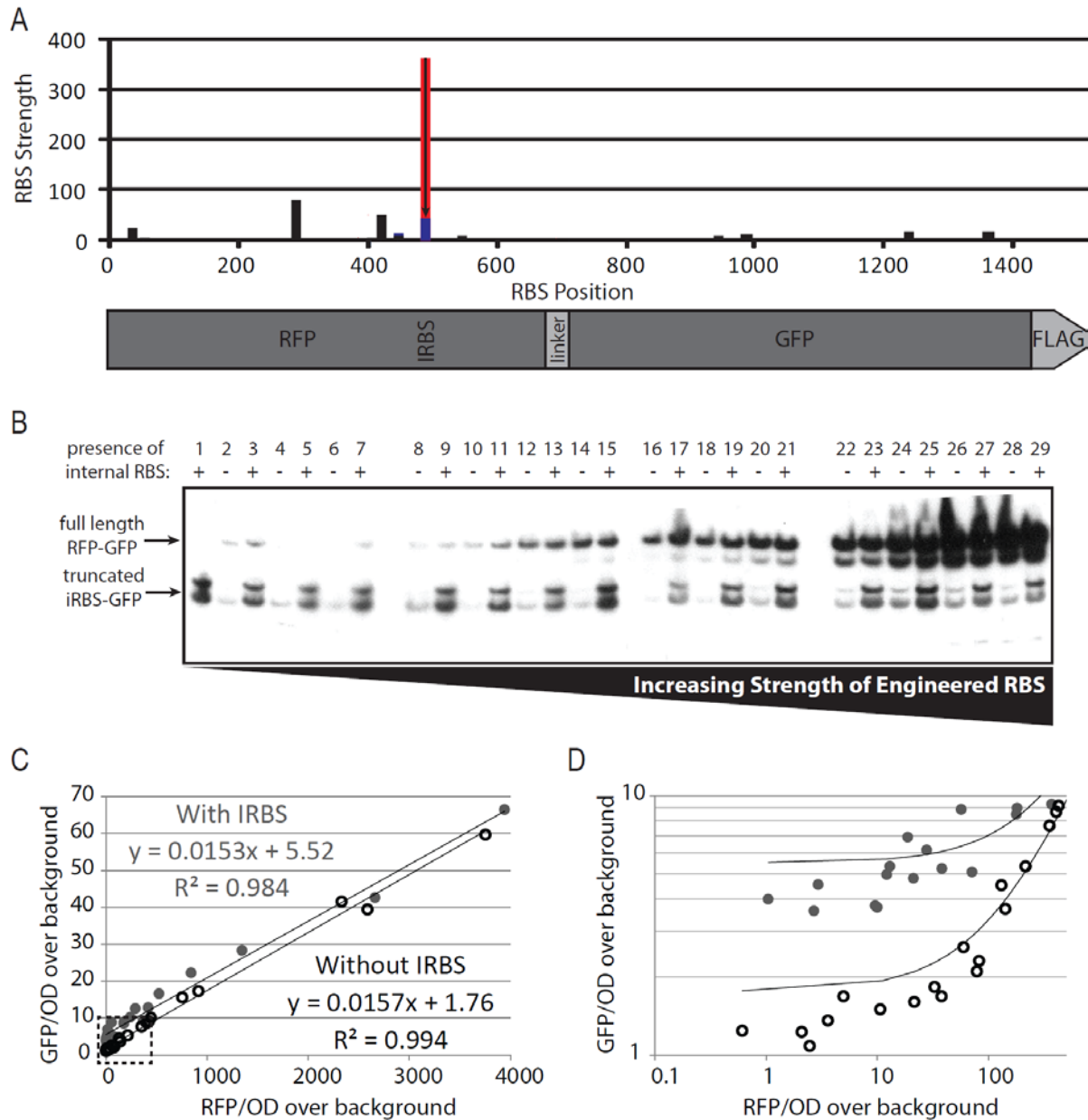


Figure 5.2. IRBS sequences can result in truncated proteins that are not inhibited by upstream translation. (A) Schematic of a RFP/GFP fusion protein with C-terminal 3xFLAG tag and the predicted internal RBS strength calculated for in-frame RBSs at its corresponding position. All predicted RBS strengths are plotted. The RFP sequence contains an internal RBS with a calculated strength of 362au at nucleotide 486 (red bar) that is removed with alternative codon usage to decrease the RBS strength to 43au (blue bar). (B) A library was produced of varying strength RBSs corresponding to the production of the full length fusion protein for both with and without the reduction in strength of the nucleotide 486 internal RBS. A western blot for the FLAG tag shows bands corresponding to proteins with sizes expected from the full length fusion protein and the truncated protein if translation were to begin at the internal RBS site. Lane 1 corresponds to a construct containing a sequence beginning only 50 base

pairs upstream of the internal RBS which is expected to produce only a truncated product. Lanes 2 through 29 correspond to an increasing strength RBS for full length protein as estimates from RFP fluorescence. Even lanes correspond to constructs in which the internal RBS strength has been decreased. (C) Members of the RBS libraries are plotted according the measured GFP and RFP levels. Members containing the full strength IRBS are plotted (grey filled circles) with the linear fit (grey) producing have a similar slope and higher intercept when compared to members with reduced IRBS strength (black circles) and its corresponding linear fit (black). The increase in GFP relative to RFP, due to the IRBS remains relatively constant as upstream RBS strength is increased. (D) An expanded plot of the low upstream RBS region (dashed box in C) shows consistently higher GFP/RFP for constructs that do not have the IRBS site removed.

RFP and GFP fluorescence measured for each constructs produce a linear GFP-RFP relationship ($R^2 > 0.98$ for each construct), as expected. The linear regression shows the GFP values for the unmodified fusion protein were consistently higher across a number of N-terminal RBS constructs when compared to the IRBS-silenced fusion protein of an equal RFP value, resulting in more than a 200% increase in the y-intercept, while the slopes of the two constructs are nearly identical, within 3%. This result is consistent with the Western data, showing the IRBS site contributing a fixed expression rate of truncated protein, fluorescent only in the green channel, that acts apparently independently of the upstream RBS.

We similarly repeated with process with IRBSs found in the syntrophin PDZ¹⁶¹ domain that has been commonly used^{7,8,40,162}, and found similar results (Figure S5.1). Two in-frame IRBS in close proximity with predicted strengths of 2431 and 1345 au were altered with 4 silent mutations which decreased the predicted value to less than 10 au. This PDZ was translationally fused downstream of the IRBS-silenced RFP and upstream of the GFP to give a similar assay as described above. Again the removal of the IRBS decreased the GFP by relatively constant value (Figure S5.1). Together these results suggest that sequences predicted to have high strength IRBS sites, even when not closely matching the canonical SD, may contribute truncated protein products, and that high strength upstream RBS is not sufficient to eliminate this effect.

5.3.3 – Balancing transcription and translation rates

The proportion of protein translated at full length is determined by the N-terminal RBS strength relative to the IRBS strengths. The IRBS strength may be reduced substantially with software, but imperfect reduction, as is occasionally the case with highly constrained sequences, may require high N-terminal RBS strength and lower translation rates to drive down the proportion of truncated protein expressed while keeping expression level constant. However, this balance of high translation and low transcription rates is expected to increase gene expression stochasticity^{163,164}. Although stochastic expression may be useful in some cases, as is believed to be exploited naturally as a means of diversifying cell behavior⁵⁰, the increased complexity may be detrimental in for engineered systems. Thus stochastic

expression and IRBS translation represent forces fundamentally requiring a balance of transcription and translation rates.

We investigated the balance between stochastic expression and relative IRBS expression as related to transcription and translation rates. To vary translation rates, several RBS variants RFP-GFP fusion protein from above (Figure 5.2) were chosen spanning a wide expression range. These constructs were driven by the P_{BAD} promoter so that transcription rates may be varied with arabinose titration. For each RBS variant both unmodified and IRBS-silenced RFP fusion protein were compared to so that increased GFP fluorescence resulting from the IRBS could be easily assayed. Protein expression, as measured by RFP fluorescence, is shown for variation in both transcription and translation rates (Figure 5.4A). Constant expression levels can be seen as transcription and translation are inversely varied, following from both transcription and translation rates are proportional to protein expression. Three point producing similar, but with different transcription and translation rates are marked (Figure 5.4A). Although points (i) and (iii) have within a 1% difference in RFP when the IRBS is removed, the GFP/RFP ratio for point (iii) is over 50 times that of point (i) when the IRBS is not removed. This suggests that at high transcription, low translation rates, constructs are more vulnerable to the effect of IRBS sequences. Conversely, we expect that at low transcription, high translation rates, high cell-to-cell variance associated with stochastic expression would be problematic^{163,164}. If highly stochastic gene expression and significant relative IRBS expression were to be avoided, optimization of both transcription and translation rates would be necessary to avoid problematic regions, as illustrated in the Figure 5.4C schematic. Of the three marked points with similar expression values, expect the point with moderate transcription and translation, point (ii), would not suffer from high relative truncated products or cell-to-cell variability.

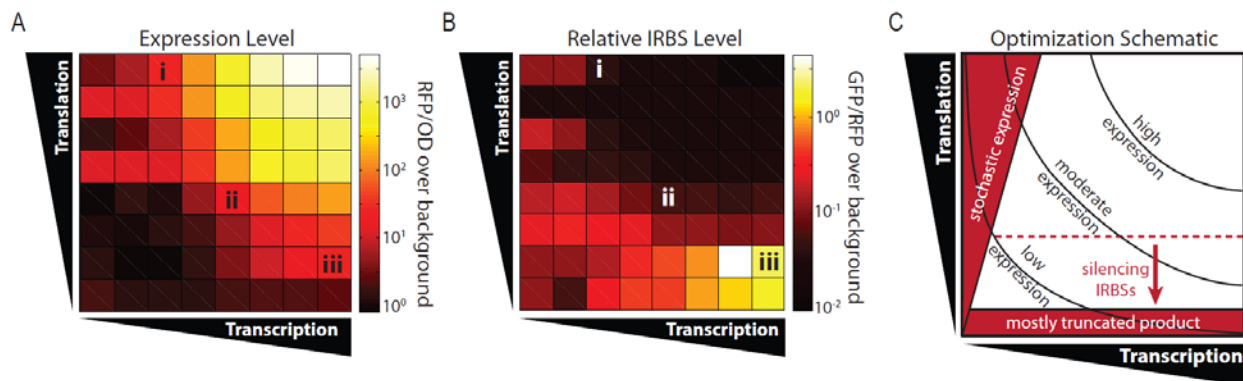


Figure 5.3. Optimizing for reduction of stochastic expression and IRBS translation requires balance of transcription and translation rates, particularly at low expression. (A) Flow cytometer measurement of RFP -fluorescence of two samples with approximately equal RFP expression levels. Sample A, with high translation and low transcription rates, displays higher variance than sample B, with low translation and high transcription rates. (B) Western blots for C-terminal FLAG tags show truncated product is being produced in addition to full length product. The low translation rate sample, B, expresses more truncated product than sample A. Samples A and B are labeled on plots C, D and E. (C) The relative full length expression level is

approximated by RFP fluorescence and shows increased expression in response to increased induction and RBS strength. (D) The relative effect of internal RBS translation is measured with the GFP to RFP ratio. A low strength upstream RBS result in a larger relative level of truncated product. (E) The variance of RFP fluorescence shows the highest noise at high translation and low transcription levels, as would be expected for low rates of stochastic expression with large bursts. (F) A schematic comparing expression levels as transcription and translation are varied includes regions with noisy expression as the transcription to translation ratio is decreased and alternatively shows truncated product as the ratio is increased. Engineering applications that require low expression noise and low fraction of truncated product fundamentally require balance between transcription and translation rates to meet this specification.

5.4 – Discussion

High strength IRBSs are under-represented in *E. coli* ORFs relative to frequencies in shuffled or *S. cerevisiae* ORFs, which is consistent with observations of selection against SD sequences in ORFs. IRBS sequences, including ones missed by SD sequence identity, result in truncated protein expression when in-frame. IRBS appears to not be influenced by upstream RBS strength with the two cases investigated here. Accounting for IRBSs during gene optimization provides a means of reducing the truncated products from IRBS translation. Although encoding a high strength N-terminal RBS can be used to reduce the relative proportion of truncated products, transcription and translation rates must both be balanced, particularly for low expression levels if cell-to-cell variability from stochastic expression is to be avoided.

IRBS sequences can be particularly problematic for heterologous sequences expressed at low levels. One of the proteins investigated here (Figure S5.1), the syntrophin PDZ domain¹⁶¹, represents the type of candidate that would benefit significantly from IRBS-silencing codon optimization. This domain comes from a eukaryotic organism, not likely to select against IRBS sequences, and if used to organize signaling machinery, as used naturally, would likely be expressed at lower levels than typical synthetic biology applications, and would represent a challenge to balance transcription and translation. This IRBS issue arising when machinery from eukaryotic systems is used in prokaryotes shows yet another way in which the context in which a part operates, and has evolved, can be important to account for. These gain-of-function characteristics are particularly troubling since most assays are more likely to rapidly identify loss of function. Building these checks into optimization software offers promising solutions when possible.

5.5 – Materials and Methods

IRBS strength calculations

All RBS strength calculations were performed using the algorithm described by Salis and co-workers⁹. Code adapted for use with the Vienna RNA package¹⁶⁵ for improved portability was provided by Ying-Ja Chen and Chris Voigt.

Genomic shuffling and recoding and IRBS frequency calculation

The ORFs of the entire *E. coli* str. K-12 substr. MG1655 and *S. cerevisiae* S288c genomes were downloaded from NCBI. Any ORFs annotated as “hypothetical” or “predicted” were excluded. All ORFs were shuffled three independent times as described by Itzkovitz and co-workers¹⁴⁸ to preserve bi-codon pairs. Additionally all ORFs were recoded three independent times according to either *E. coli* or *S. cerevisiae* codon usage, preserving amino acid sequence but not bi-codon frequencies. The RBS calculator was run on the entire ORF and RBSs in all reading frames were considered except those within the first or last 35 base pairs that ignored. RBS frequencies were normalized by sequence length analyzed, excluding the first and last 35 base pairs.

Fluorescent assays

Plasmids were transformed into chemically competent BW227783¹²⁹ cells strains using standard methods. LB/agar/antibiotic plates containing transformed strains were stored at 4 degrees for up to five days and inoculated into 300 μ L of MOPS rich defined media (MRDM) (Teknova, Hollister, CA) with 0.4% glycerol as the carbon source and appropriate antibiotic in a 96 well plate for experimentation. This culture was grown for 7 to 8 hours at 37 degrees in an ATR Multitron plate shaker at 1000 RPM, then diluted 1:30 into 300 μ L of pre-warmed MRDM containing inducers for the experimental condition inducer concentrations. Strains were then similarly grown for 4 hours until late log phase and immediately measured. Measurement was done in a TECAN Safire2 machine with OD absorbance at 600 nm, GFP and RFP excitation/emission of 481/507 nm and 584/607 nm respectively with a 5 nm band width. Fluorescent reading were normalized by OD, though were generally within two-fold between sample OD readings. Cell auto-fluorescence per OD, as determined by strains with empty vectors grown under identical conditions, was subtracted.

Western blots

Western blots were prepared with protein gels were run under standard conditions for 10% Bis-Tris NuPAGE denaturing gels, followed by transfer to nitrocellulose membrane, then labeled with Sigma’s monoclonal ANTI-FLAG M2- peroxidase (HRP) antibody using standard procedure. Western blots were exposed on an ImageQuant LAS 4000 (GE Healthcare).

5.6 – Supplemental Figures

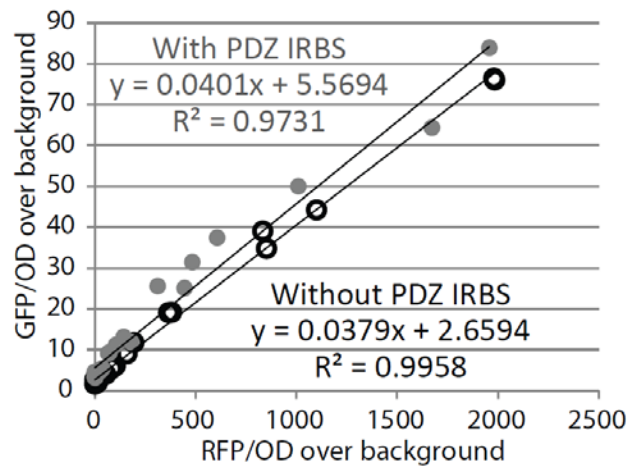


Figure S5.1. Similar to Figure 5.2C, the PDZ IRBS is not inhibited by upstream translation. A commonly used PDZ is predicted to have an IRBS. The RFP with IRBSs removed was fused to a PDZ domain, with or without the IRBS removed, followed by a GFP. A library is made for the upstream RBS preceding RFP and are plotted according the measured GFP and RFP levels. Members containing the full strength PDZ IRBS are plotted (grey filled circles) with the linear fit (grey) producing have a similar slope and higher intercept when compared to members with reduced IRBS strength (black circles) and its corresponding linear fit (black). The increase in GFP relative to RFP, due to the IRBS remains relatively constant as upstream RBS strength is increased as found in Figure 5.2C for the RFP IRBS.

Chapter 6 – Additional findings and future directions

6.1 – Introduction

Synthetic biology is currently at a stage where the process of system design is largely *ad hoc*. Here we discuss the strategies we employed for developing the scaffold-dependent phosphotransfer control of two-component system (TCSs) presented in Chapter 5. In particular, we discuss the cloning strategy, the role of simulation, and the experimentally driven optimization process. We continue with the presentation of additional findings and strategies employed with mixed successes that were ultimately not included in the finished system presented in Chapter 5. Finally, we discuss challenges remaining to developing a broadly applicable phosphotransfer control strategy, and discuss specific strategies that may be employed to this end.

6.2 – The engineering and optimization process

The construction and optimization methods employed were determined by the strengths and challenges associated with this system. Two notable strengths of this system are that the components are well characterized and that the fluorescent readouts allow for fairly high throughput assays. This permits a fairly detailed characterization of how specific parameters affect the system level behavior. A challenge associated with this project is the simultaneous protein and system engineering. Our design process seeks to maximize flexibility of manipulating components at the protein and system level with a relatively rapid turn-around time between experiments.

6.2.1 – Key design variables

Many of the first generation synthetic biology papers showed that by varying network architecture a wide range of behaviors could be achieved for a gene circuit. In focusing on these impressive successes achieved by varying architecture, it is easy to underestimate the importance of carefully balancing component concentrations. Many components do not even qualitatively exhibit characterized or native behavior when expressed orders of magnitude above or below the optimal concentration. For example, CusR exhibits unexpected behavior at high concentrations (Figure 6.1). Maximizing scaffold-dependent signaling required careful control of all components employed. This sensitivity to concentration was a key parameter that had to be considered at every stage of the design process.

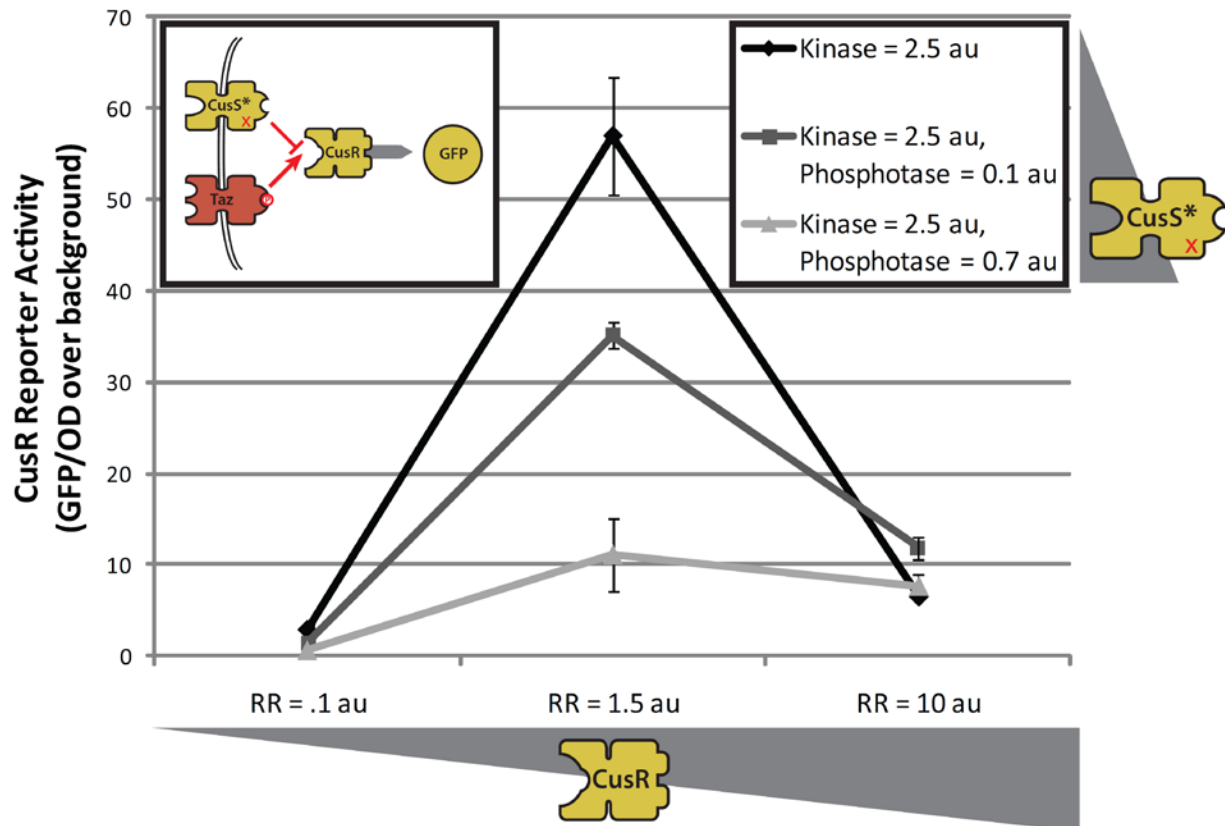


Figure 6.1. The importance of component concentrations. A kinase (Taz), a phosphatase (CusS G448A), and a response regulator (CusR) are expressed at different concentrations. The response regulator activation drives GFP expression, which would be expected under high kinase and response regulator concentrations and low phosphatase concentrations. The relative component concentrations are given in arbitrary units determined by fluorescence over background when fused to a RFP. At low CusR concentrations, the reporter activity is very weak. At intermediate concentrations, the CusR activity correlates with the ratio of kinase to phosphatase as expected. However, at high concentrations CusR behaves contrary to our expectations. CusR is not characterized to have any repression capability, yet increasing CusR expression beyond the intermediate level results in a decrease in fluorescence which is not expected according to simple models of kinase and response regulator activities. Additional interactions not been explicitly characterized in literature would be necessary to explain behavior at expression levels not normally accessed under physiological conditions.

Additionally, the design process for this system should account for many rounds of modification to the proteins. In our study, we frequently varied the translational fusion of interaction machinery including types and number of binding domains, linkers, and ligand affinities. Variable number of components and repeating sequences presented a challenge to the construction method in particular. We considered these design variables when developing a cloning strategy to minimize time between design iteration and finished constructs experimentation.

6.2.2 – Cloning strategy

We have employed a variation on the BioBrick¹⁰⁹ and BglBrick¹¹⁰ cloning strategy for our plasmid construction. Unlike standard multiple cloning site based cloning, this BioBrick based strategy has the advantage of allowing the assembly of any number and order of components without prior knowledge or restriction site rearrangement. To reduce the number of cloning steps required to modify a protein composition, expression level or number of operons, we have used a nested compatible sites to allow plasmid assembly at either the domain level (Figure 6.2 A-C) or the operon level (Figure 6.3 D-F). The scar made with the BglBrick assembly (Figure 6.2 A-C) encodes innocuous residues, glycine-serine, which makes it ideal for translational fusions. Unique restriction sites are also maintained to allow for promoter or RBS sequences to be easily swapped, which is particularly beneficial for construction of promoter or RBS libraries. This strategy allowed for the rapid alteration of protein composition or expression level, often in a single step and without the need for ordering unique oligos.

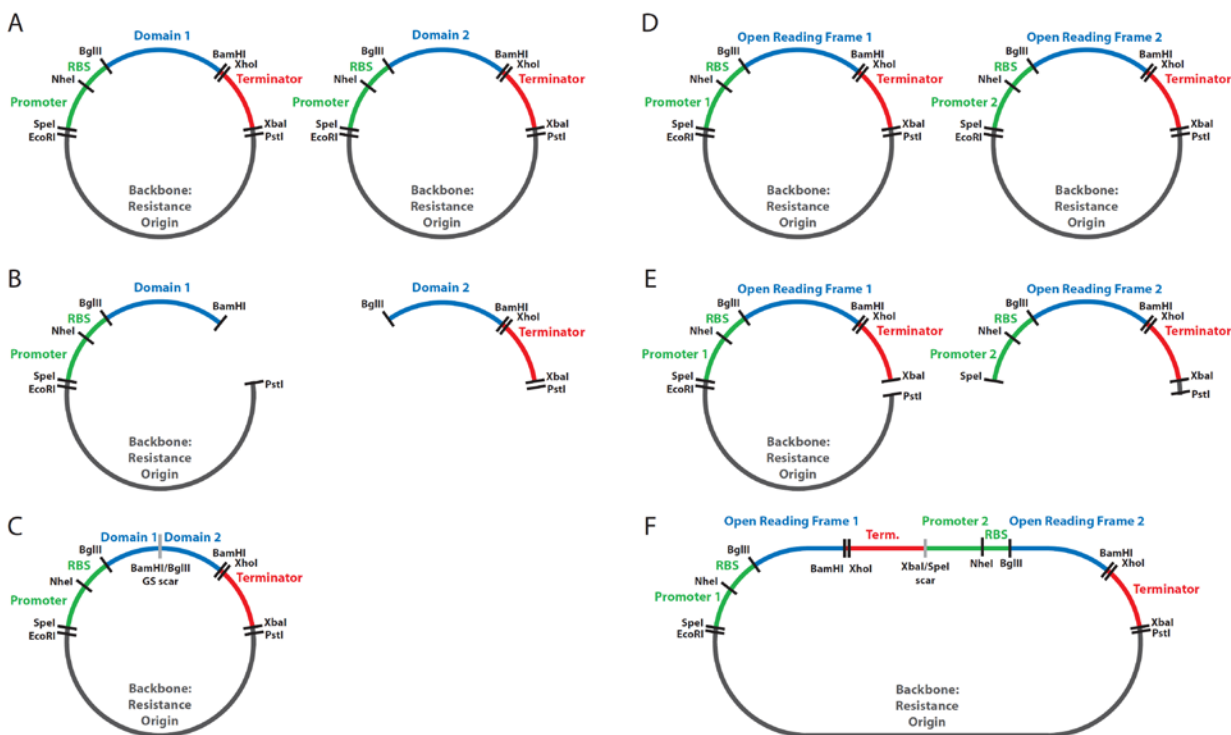


Figure 6.2. Nested BioBrick cloning strategy allows for easily composing domains and operons. (A) The open reading frame parts intended to be translated, including enzymes, binding domains and linkers, are located between the BglIII and BamHI restriction sites. The parts expected to be N-terminal contain a start codon and most parts lack a stop codon. This is designed such that translation begins directly after the BglIII, translates the entire BglIII to BamHI sequence and then stops translation at a stop codon located between the BamHI and XhoI restriction sites. (B) To fuse domains in this format, the plasmid encoding the N-terminal domain is cut with BamHI and PstI and the plasmid encoding the C-terminal domain is cut

with BglII and PstI. BglII and BamHI share compatible sticky ends and can be ligated. (C) BamHI-BglII ligated plasmids will contain a translational fusion of the two domains separated by a scar encoding for a generally innocuous two-residue, glycine-serine, linker. This six nucleotide scar sequence is not recognized by either the BglII or BamHI enzymes used to create it, so this process may be repeated indefinitely to create fusion proteins containing a large number of domains. (D) A similar approach can be applied to combine multiple operons onto the same plasmid using the SpeI and XbaI compatible restriction enzymes. In this strategy, the SpeI and XbaI sites are switched compared to the original BioBrick assembly method because in previous applications the XbaI-SpeI scar encodes a more innocuous linker residue, though it is no longer used for translational fusions. (E) To fuse two operons, plasmids are cut with SpeI/PstI and XbaI/PstI. (F) These SpeI-XbaI composite plasmids contain unique SpeI and XbaI sites and may be indefinitely expanded with the SpeI-XbaI subcloning. However, these composite plasmids now contain duplicate restriction sites used to alter open reading frame composition and expression levels, which would have to be altered at an earlier stage as is the case with standard BioBrick assembly.

To further aid in the construction of plasmids, code was written in Matlab to guide the process. Although the process of plasmid design and construction method was performed manually, code aided in construction validation, record keeping and generating construction protocols. The code functions by reading excel sheets describing design process in which the user specifies the plasmids, oligos (if necessary) and assembly type (e.g. subcloning, 2-step PCR, etc.). The code then accesses the user's plasmids and oligos sequences, checks for a number of common cloning errors, generates genbank files for plasmids and generates detailed cloning protocols for the creating batch of plasmids specified. Through the course of this dissertation, over one thousand uniquely cloned plasmids were planned with this code, which most significantly served to increase the success rate of cloning by checking for a large number of low-probability cloning errors and improving record keeping.

Although largely successful for the scope of this dissertation, our cloning strategy presents some limitations, particularly when scaling to larger systems. As the size of the system employing this cloning strategy scaled to include up to 7 operons, it was necessary to include additional, non-standard restriction sites to aid in the tuning of expression levels and protein composition. We found it particularly useful to be able to screen multiple libraries simultaneously with all relevant operons present. However, tuning components via restriction based methods becomes challenging because unique restriction sites are limiting in large systems. Future DNA assembly techniques would ideally allow for short assembly time, ability to accommodate a large numbers of repeated sequences, ability to simultaneously include a number of different libraries, and expansion to a large number of engineered components.

6.2.3 – Mathematical modeling of to inform system design

Although biological designs were guided primarily through empirical means, mathematical modeling was also employed as an early means of design validation. Many of the difficulties encountered are not easily identified with modeling, such as components exhibiting unexpected properties, but initial effort spent on a simple model often has significant returns. In particular, modeling was useful for ruling out designs if simulations found desired behavior to be either not possible or highly sensitive to parameters that could not be easily controlled.

Software was written to aid modeling of signal transduction and other protein-protein interaction focused systems (see appendix for the "AUTOMODEL" code written for Matlab). The core of the software (the "setup.m" function) takes higher level information about proteins that may be quickly specified by the user and then generates functions for differential equation or stochastic simulation. Users specify a matrix representing affinities between domains present in the system, a matrix representing what domains are present on each protein in the system, and a matrix for each reaction undergone by a domain other than binding, such as phosphorylation. Based on this easily specified information, the model will automatically infer all possible protein complexes (up to a limit) and all transitions between complexes. A limitation must be imposed on the number of complexes considered because even some simple systems could theoretically contain an infinite number of different polymerization states. The user specifies the maximum number of times any protein may be repeated in a complex, and all complexes meeting this criteria are considered (this has only been tested up to $N=2$ repeated proteins per complex). A differential equation for each complex is generated and includes a term for each possible transition into and out of that complex. Functions are included for easy conversion between the set of inferred complexes and basic protein stoichiometry. The software accounts for an increased local concentration of elements present within the same complex. This version of the code only considers a single user specified local concentration, and does not take into account effects of changing complex size or steric interactions, but this parameter is fairly modularly encoded and more sophisticated treatment should be possible. However, for the purpose of this project, intra-molecular interactions were generally not limiting, and ten-fold changes in local concentration had almost no impact on the behavior of all the systems tested. A key advantage of this software is that models can be generated very rapidly and with reduced human error. For example, in simulation of scaffold occupancy discussed earlier (Figure 3.3), in which the scaffold binding five different proteins, the user needs to only specify a simple affinity matrix and a simple matrix description of the proteins, which contained only 20 non-zero numbers between those two matrices. The code then automatically infers the 37 different proteins and protein complexes and the 160 different reactions that may occur and can run a differential equation simulation to equilibrium on a standard laptop (as of 2012) in under a second.

This software generates a model in a convenient modular form so that higher order qualities, such as robustness, can be assessed with additional scripts. The model generating

software is designed to not only allow easy manipulation of starting conditions, but also manipulation of reaction rates and other parameters, and even protein composition. Scripts were written to guide intuition about behavior of specific models at varying expression levels, robustness of a specified behavior to changes in kinetic parameters, and optimization of parameters to improve the robustness of specified behavior. This provided a means of ruling out designs that were found to be highly sensitive to parameters even when using optimized parameters. However, for the purpose of this project, the computational analysis was used more as a means of informing basic intuition and ruling out designs that were highly likely to fail. Empirical experimentation played a far greater role in the design and optimization of biological systems.

6.2.4 – Optimization process

A significant amount of optimization was required to develop scaffold-controlled signal transduction. The choice of signaling components and binding domains, assembly of multi-domain fusion proteins, and tuning expression level were important considerations. The concentration of the components expressed was an important parameter throughout the design process, and was often done while tuning other parameters. The optimization process employed in this work is discussed here, which was effective for developing a fairly complex system by 2012 synthetic biology standards including systems with up to five interacting engineered proteins and two additional reporter proteins. Although our design process was effective for systems of this modest complexity, scaling up to more complex systems will likely require a more sophisticated design process that more heavily incorporates computer aided design.

We chose signaling components using the simplest assays feasible. The selection of the primary histidine kinase used, based on *E. coli* EnvZ, was guided initially by published in vitro phosphotransfer crosstalk assays¹²³. Before testing if scaffold-dependent colocalization was sufficient to increase phosphotransfer between non-cognate pairs, we monitored in vivo the activation of a RR of interest (e.g. CusR) in response to overexpression by non-cognate HK (e.g. Taz). Native RR levels were used, unmodified HK was over-expressed on a plasmid, and RR activation was measured with a plasmid containing fluorescent protein driven by a promoter sequence known to be transcribed by phosphorylated RR. Although this simple assay could be confounded by a number of variables, including insufficient over-expression of HK or buffering by native TCSs resulting in false negatives, a positive result for this simple assay provides a system that is sensitive to our perturbation and may offer a starting point for optimization.

Protein Concentration (nM)

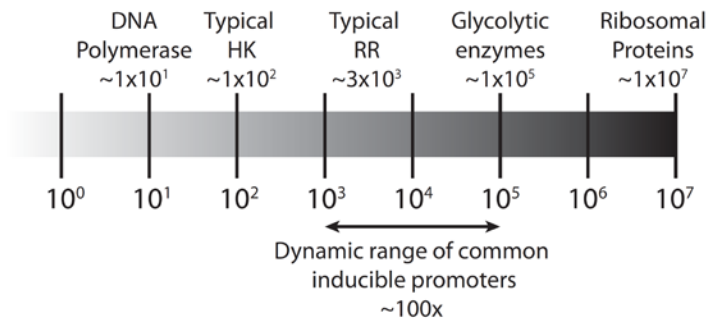


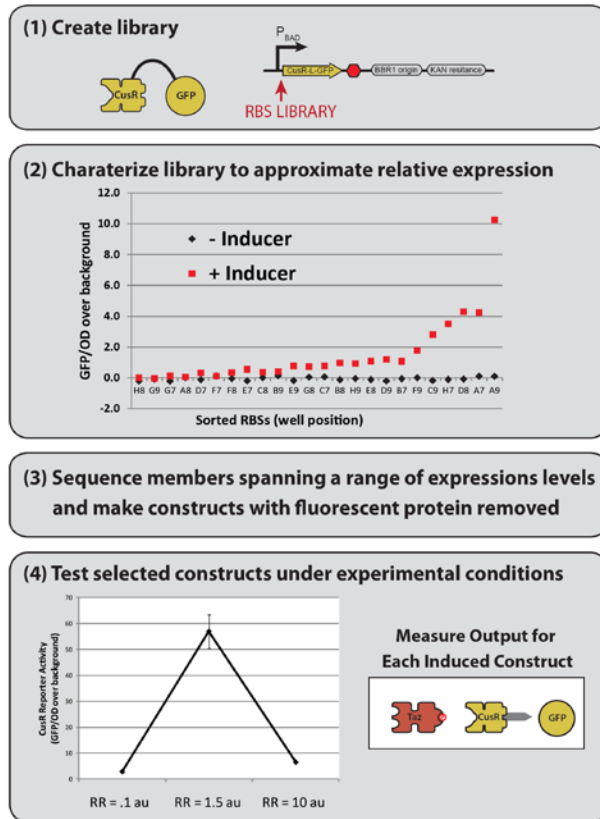
Figure 6.3. Cells naturally make use of a wide range of protein expression levels. System and even protein behavior can vary substantially depending on expression levels. Although inducible promoters provide a convenient means of tuning expression levels, the dynamic range of promoters is generally not large enough to cover the range of physiologically relevant concentrations. Expression levels often must be tuned by other means in addition to promoter induction, such as copy number or translation strength, to include optimal concentrations within its range.

Every engineered protein was tested across a range of expression levels *in vivo* to achieve a physiologically relevant concentration. Up to four compatible inducible promoters (i.e. arabinose, rhamnose, tetracycline, and salicylate inducible) were used, while other components were expressed on constitutive promoters. In addition to varying the induction level of each promoter, it was often necessary to vary other factors relating to expression level so that the inducible promoter would include the optimal expression level within its range (Figure 6.3). During this research, the most commonly employed method was generating variation through RBS libraries. Several methods of optimizing concentrations were used including rationally setting relative expression level and generating random libraries and screening for function.

Generally the optimal expression level of system components is unknown, so screening a range of expression levels to empirically determine the optimal is our preferred method. Rationally setting expression levels can be done either by using computational methods, such as the RBS calculator⁹, to design a range of expression levels, or by screening a large number of library variants and selecting a subset of the library to give a range of expression levels (Figure 6.4 A). The current computational methods for selecting a range of expression levels generally require screening component concentration to guarantee accuracy. This can require more effort, particularly at the cloning level, than simply screening a library and selecting members based on empirical data. However, computational methods are particularly useful if screens are not amenable to fairly high throughput assays. Computational methods are also very useful for selecting very high or low expression levels that require very large screens if the libraries are not biased to over-represent the desired range. For example, an RBS library composed of totally degenerate residues (e.g. 25 Ns) expressed on a high strength promoter on a medium copy number plasmid (e.g. around 20

copies per cell) is likely to require the screening of less than one hundred colonies to find an acceptable expression level. However, finding an RBS sequence that will either express very low levels or very high levels from totally degenerate nucleotides would be very unlikely to produce an acceptable sequence from screen of one hundred library members. When computational methods improve to the point where accuracy is high enough to not require empirical confirmation at the initial stage, they will provide an incredibly useful tool.

(A) Characterizing A Range Of Expression Levels



(B) Screening For Desired Behavior

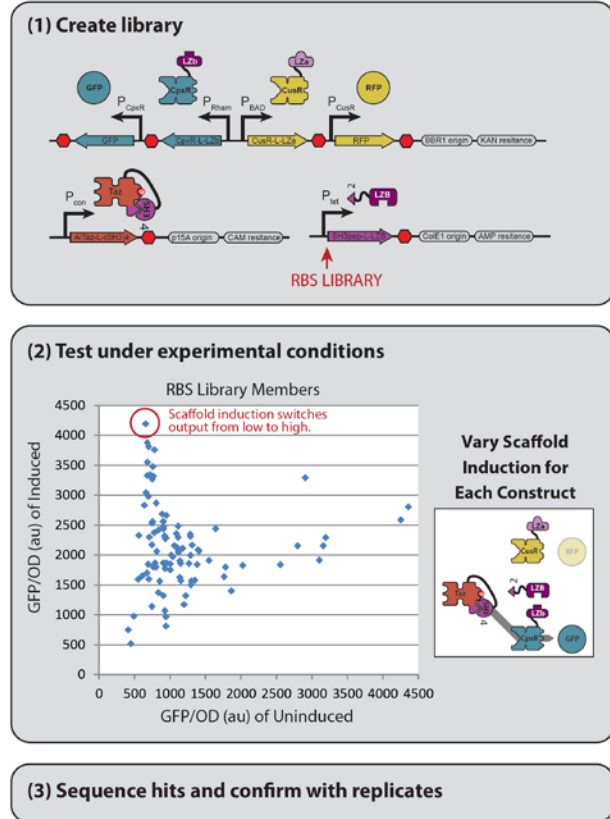


Figure 6.4. Two alternative strategies for concentration optimization. (A) The protein of interest is translationally fused to a reporter, such as a fluorescent protein. An expression library is generated and members are characterized. A set of members spanning a range of expression levels is selected and sequenced. The translationally fused reporter protein is removed if it is not compatible with the final assay, which may result in some inaccuracy in protein estimation. C-terminally tagged components are expected to have a smaller perturbation than an N-terminally tagged protein which would alter RBS context. The selected set of library members is characterized in the system of interest and optimal expression level based on device behavior is chosen. (B) An expression library is generated directly in the system of interest. If inducible promoters are present, several induction conditions for each library member can be used to help inform the behavior of the system under the range of expression levels. Library members producing desired behavior are isolated, sequenced and confirmed with replicates.

Directly generating and screening a library in the experimental context offers an alternative approach to creating a set of different expression level constructs and then testing them in the experimental context (Figure 6.4 B). Directly screening the library in the experimental context can be more efficient and can offer higher resolution, allowing the best strength out of potentially hundreds of candidates to be found with a single cloning step and experiment. Although not initially measuring protein concentrations can lead to less intuition or misinterpretation of system function, this can be rectified by measuring concentrations afterwards. If possible, short epitope tags can even be included in the functional system, allowing relative concentrations to be quantified via Western blot.

Assembly of protein-protein interaction machinery was also an important parameter for optimization in this work. Protein-protein interaction domains with modular folding, orthogonal interaction to with other interaction domains, and high affinity interaction were selected. For detailed discussion of domain selection criteria see Chapter 3. Translational fusion of binding domains to the C-terminal of enzymatic components was generally preferred, due to minimal impacts on expression level, though N-terminal fusion was also considered. Translational fusions were first tested with long flexible linkers. For the systems tested in this dissertation, a linker above around 10 amino acids is sufficient for most proteins, and extending linkers beyond their optimal length is not a significant problem, as seen in Figure 6.5. If the enzyme cannot tolerate terminal addition, such as if extending the terminals would sterically inhibit a dimerization interface, internal peptide insertion could be considered. A first step for internal peptide insertion may be targeting regions that have variable size in homologs for insertion of peptide with flexible linkers. An additional method for peptide insertion is targeting a structured region, if the crystal structure is known, and designing linkers that extend the natural structure as shown in Chapter 5. For engineering simple colocalization via tethering, the components used in this work were generally very amenable to C-terminal fusion with a long glycine-serine linker, predicted to be unstructured, and optimization of linker length was often not required.

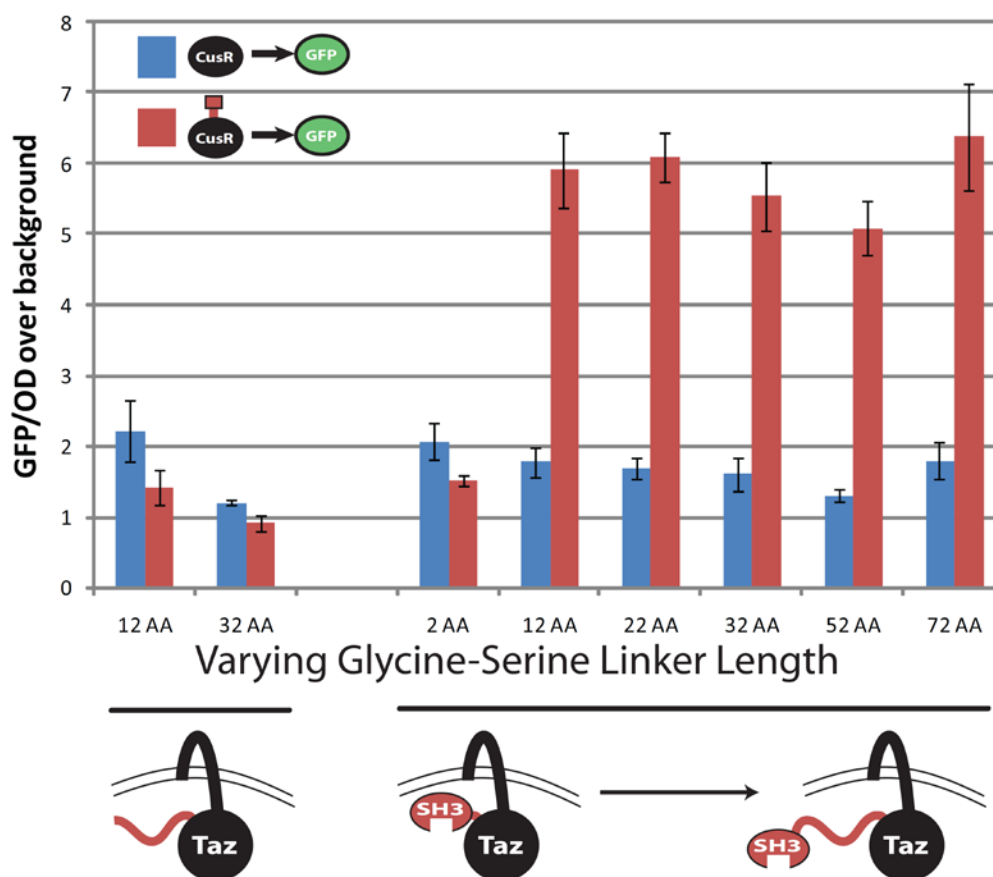


Figure 6.5. Varying the length of an unstructured glycine-serine linker. Blue bars represent signaling from Taz to the non-cognate CusR lacking a SH3 ligand. Red bars represent signaling to CusR with a SH3 ligand fused, which would be expected to be higher than the corresponding blue bar if colocalization was effective at amplifying phosphotransfer. The x-axis labels the length of an unstructured linker composed of repeating glycine serine residues. The left two pairs of bars represent Taz fused only to a linker, while the right five pairs of bars represent Taz fused to a linker followed by an SH3 domain. The very short, two residue, linker is not capable of increasing phosphotransfer, while all longer linker lengths are effective.

6.3 – Additional findings on two-component system signaling

Here we discuss strategies and findings related to but not directly included in Chapter 5. Informative but non-essential in vitro phosphotransfer assays are presented. Efforts to broaden specificity of HK with limited success are discussed. A number of strategies were pursued to improve the robustness of scaffold-directed phosphotransfer. Autoinhibition of HK proved to be the most effective strategy for improving scaffold-directed phosphotransfer, but competitive inhibition and phosphatase buffering also showed promise. We discuss these strategies below.

6.3.1 – In vitro scaffolding results

Phosphotransfer assays were conducted in vitro with purified proteins, using methods described by Skerker and coworkers¹²³. The HK can be autophosphorylated with gamma-32P-ATP, which can be subsequently transferred to RR when co-incubated. Phosphotransfer can be quantified by SDS PAGE separation and autoradiography imaging. The truncated cytoplasmic EnvZ displays constitutive autophosphorylation and a modest phosphotransfer activity to non-cognate CusR¹²³. In our work, SH3 ligand tagged EnvZ and PDZ ligand tagged CusR were colocalized with a synthetic scaffold composed of an SH3 domain and PDZ domain. Increased phosphotransfer rates were observed when incubated in the presence of scaffold, as indicated by a larger P32-band corresponding to CusR when compared to the band produced when lacking scaffold protein (Figure 6.6). Several scaffolds of varying linker length were used, and all scaffolds had a similar effect. The effect also appeared to be detectable for CitB, a non-cognate not previously show to crosstalk with EnvZ, though the phosphotransfer was much weaker. This in vitro evidence that an increased local concentration effect could be used to alter phosphotransfer rates between non-cognate components led us to develop the in vivo system, amenable to higher throughput assays.

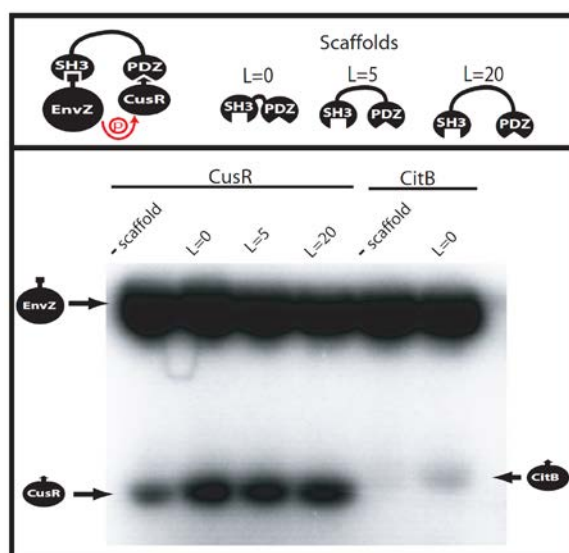


Figure 6.6. An in vitro phosphotransfer assay between colocalized non-cognate TCSs is shown. Purified SH3 ligand tagged EnvZ is incubated with gamma-32P-ATP, autophosphorylates, and is then incubated with a purified PDZ ligand tagged non-cognate RR, CusR or CitB. Components are incubated with or without a synthetic scaffold protein composed of an SH3 and PDZ domain tethered with a flexible, glycine-serine, linker of varying length. The inclusion of a scaffold in co-incubation increased phosphotransfer as analyzed by SDS-PAGE and autoradiography.

6.3.2 – Broadening TCS phosphotransfer specificity

Scaffolding has provided a means of increasing a weak phosphotransfer rate to a physiologically relevant level. One hurdle limiting general application to a wide range of RRs is that any given HK appears to display undetectable phosphotransfer to most non-cognate RRs. Ideally scaffold-dependent colocalization would be capable of directing phosphotransfer between any HK-RR pair. Towards this end, we attempted to find or engineer components with increased phosphotransfer promiscuity.

We hypothesized that TCSs may have undergone negative selection to reduce crosstalk, thus a heterologous TCS would be more likely to exhibit increased promiscuity. This negative selection against interaction has been suggested for eukaryotic signaling components, based on increased promiscuity between heterologous SH3 domains⁸⁵. A limited investigation of this hypothesis, including testing several *B. subtilis* HKs (CitS, KinB, and PhoR) against several *E. coli* RRs (CusR, OmpR, and CpxR), revealed only modest promiscuity. Though only pursued with a small sample size, preliminary results suggest that negative selection is unlikely to be the major factor in shaping TCS specificity.

Recent work, identifying the residues responsible for TCS specificity, has demonstrated that kinetic preference may be tuned through a series of mutations to include HK mutants that have dual specificity for a cognate and non-cognate RR^{33,141}. Though it is not clear how broadly HK specificity can be tuned, it provides a means of increasing promiscuity for at least a set of target RRs. We attempted to create a mutant Taz that exhibits dual specificity for cognate OmpR and non-cognate CusR through a series of mutations to the specificity determining residues (Figure 6.7). We targeted residues found previously to be important³³, however we performed this assay *in vivo*, which allowed us to determine mutation effect on phosphotransfer in response to stimulus. While we succeeded in increasing phosphotransfer to CusR, the limited number of mutations tested did not significantly decrease activation of OmpR, suggesting that it may still have too high of an affinity to be used for scaffold-controlled phosphotransfer. Additionally, most mutants exhibited decreased stimulus sensitivity. Ultimately, this was not pursued further since CpxR proved to be compatible with CusR signaling and presented a more direct path to our goal.

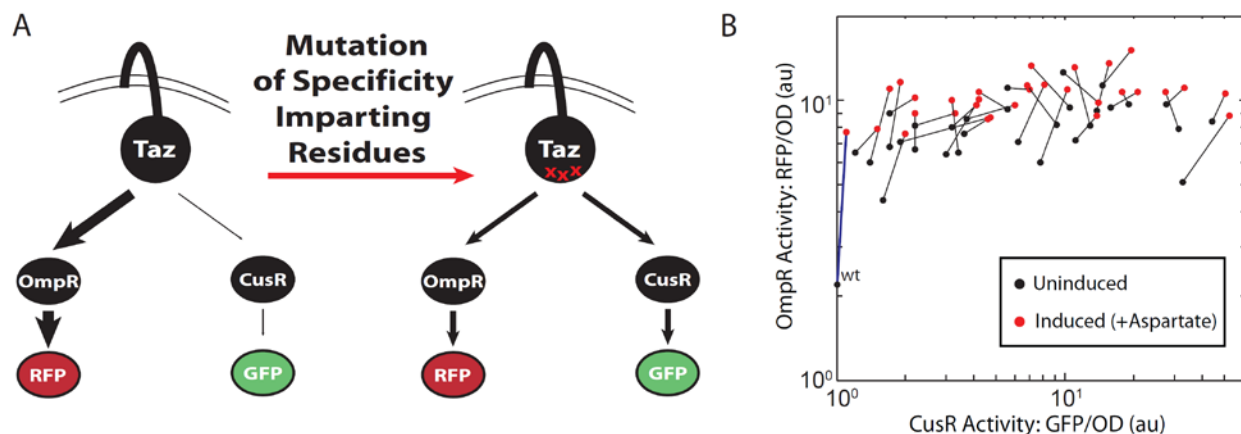


Figure 6.7. Attempting to create dually specific Taz. (A) A set of mutations found to be important for determining HK specificity³³ were made to Taz with the goal of tuning Taz specificity to achieve a similar kinetic preference for both cognate OmpR and non-cognate CusR. (B) A dual reporter for OmpR and CusR driving different red and green fluorescent proteins, respectively, was created. Five residues in Taz were varied between the native and the corresponding CusS residue identities. Activation of OmpR and CusR, simultaneously expressed, was measured in vivo for a number of Taz mutants. The assay was performed with induction of Taz expression and with both with and without addition aspartate, the stimulus of Taz. Each line represents a different mutant variant, and wild type Taz is marked and shown with a blue line. As Taz residues are changed to the corresponding CusS residue identities, the specificity for CusR increases, though the effect of mutations was non-additive.

Additionally, we attempted to mutate the RR to increase phosphotransfer promiscuity. We focused on increasing Taz-CitB crosstalk since it displayed very weak crosstalk (Figure 6.6). Rather than mutating specificity determining residues, we examined the effect of a mutation to a core hydrophobic residue. Mutating hydrophobic core residues of the SpoF response regulator was shown to increase crosstalk between SpoF and non-cognate KinC and KinD in the *B. subtilis* sporulation cascade¹⁶⁶. The authors hypothesized that this mutation may have increased the structural flexibility and thus the promiscuity of SpoF. We hypothesized that this may be used to create a more universal RR that can be phosphorylated by a wide range of HKs when colocalized via scaffolding. To mimic the mutation seen to increase promiscuity of SpoF, three mutant CitB variants were made with mutations to the following hydrophobic core residues: L69A, V93A and Y104A. Unfortunately the mutations to CitB did not result in detectable Taz-CitB phosphotransfer and this strategy was not further pursued.

6.3.3 – Improving scaffold-directed phosphotransfer robustness

We found that scaffold-directed phosphotransfer was highly sensitive to component concentrations, as discussed in Chapter 5. Although engineering autoinhibition into the histidine kinase proved to be a successful method of improving robustness, several other strategies for improving robustness were also pursued. Specifically, phosphatase activity and competitive inhibition have been recognized as important in natural systems, and were

employed here with limited success. Here we describe these efforts and offer a hypothesis as to why some methods for enhancing robustness are more effective than others.

We used phosphatases specific to each RR of interest to improve colocalization-controlled phosphotransfer. Although Taz exhibits both kinase and phosphatase activities for its cognate RR, OmpR, Taz was found to display only constitutive kinase activity to non-cognates CusR and CpxR. We reasoned that a low level of phosphatase activity towards CusR and CpxR may help buffer against the weak, non-colocalized phosphotransfer that becomes significant only at high component concentration. This hypothesis was based in part on the observation that when the natural TCSs are knocked out, the activation of the non-recruited RR increased substantially (Figure 6.8 A and B). The natural HKs may be serving as a phosphate sink, thus our engineered system may benefit by expression of phosphatases. To gain independent tunable control of phosphatase activity we mutated CusS, G448A, and CpxA, G420A, in a position corresponding to the G405A mutation in EnvZ that has been shown to disrupt ATP binding¹⁶⁷. These mutations are expected to eliminate kinase autophosphorylation but have minimal impact on phosphatase activity, as seen in EnvZ G405A¹⁶⁷. By reducing the non-colocalized CusR activation, CusS G448A phosphatase was capable of restoring the colocalization dependent phosphotransfer that was lost when the native CusR/CusS TCS was knocked out (Figure 6.8 C).

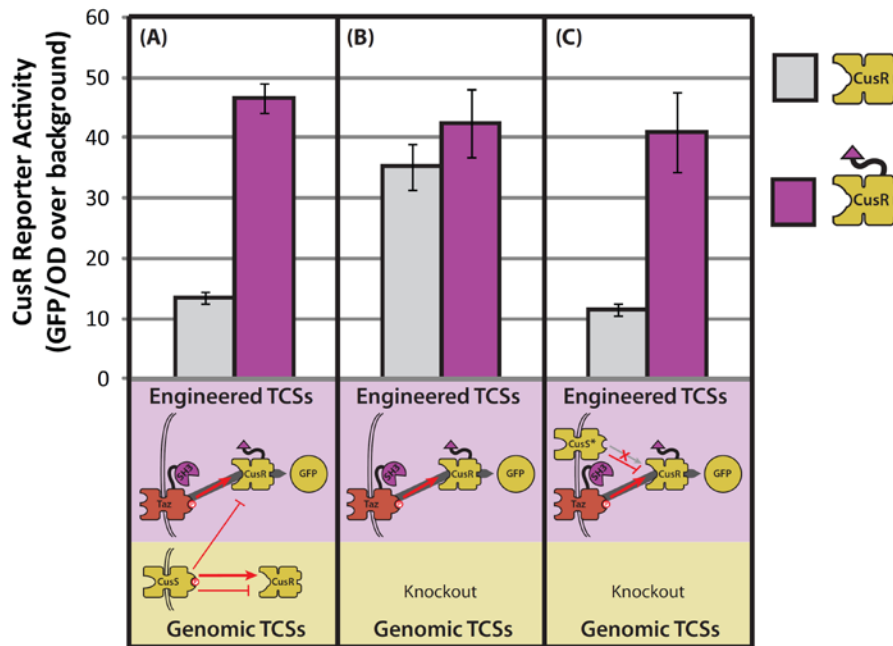


Figure 6.8. Engineering phosphatases to reduce undesired signaling. (A) Expression levels of Taz fused to an SH3 domain and CusR either with or without fusion of an SH3 ligand are expressed at a constant level while a GFP reporter of CusR activation is measured. If native pathways are left intact, they can function as a phosphate sink. Corecruitment overcomes the phosphatase activity of the native CusRS with engineered recruitment. (B) Once the endogenous phosphatase is removed by knocking out CusRS, the RR CusR is active even without corecruitment. (C) The cognate bifunctional CusS was mutated, G448A, to function

only as a phosphatase, providing a necessary phosphate sink to gain colocalization dependent activation.

These mutant HKs give independent tunable control of the phosphatase activities, such that phosphatase rates may be tuned to the optimal level by varying the phosphatase expression level. To show that a moderate phosphatase activity was optimal at a high CusR expression level, we simultaneously varied Taz-SH3 and CusS G448A (Figure 6.9). As expected we see THE activation of CusR, both with (Figure 6.9 A) and without (Figure 6.9 B) SH3 peptide, when only kinase is highly expressed, and no activation when only phosphatase is highly expressed. At intermediate kinase and phosphatase expression levels the CusR with SH3 peptide is highly activated while CusR lacking peptide is very low. The fold change in activity due to colocalization is largest when kinase and phosphatase expression levels are balanced at an intermediate concentration (Figure 6.9 C).

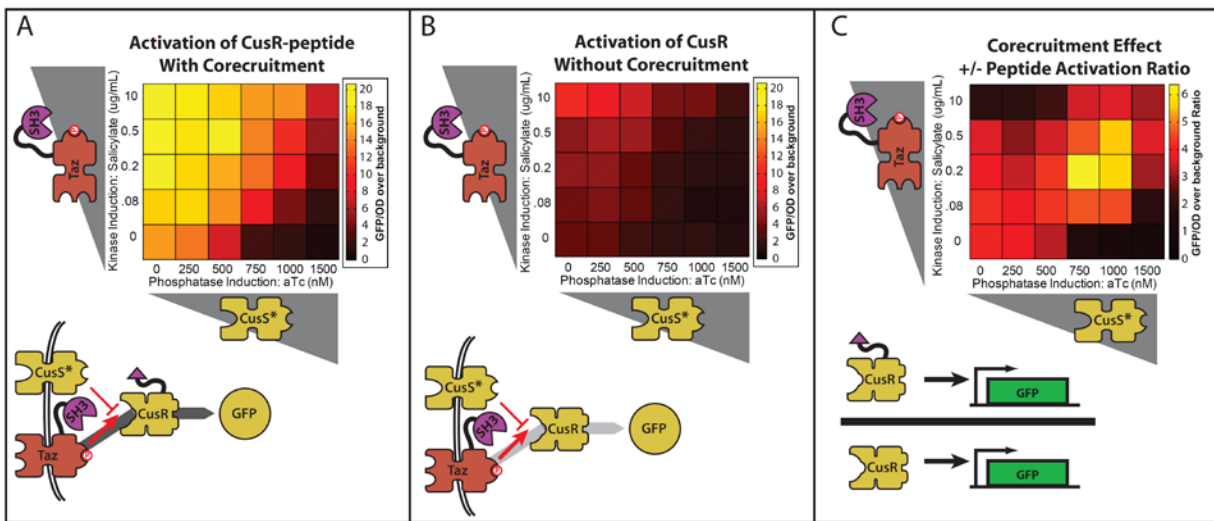


Figure 6.9. Balancing kinase and phosphatase expression improves colocalization-controlled phosphotransfer. (A) Expression levels are titrated for both the SH3 domain fused the Taz HK, and for the CusS G449A phosphatase. CusR, fused to an SH3 ligand capable of colocalization with Taz, is expressed at a constant level and GFP corresponding to its activation is reported. Activation of CusR occurs at most concentrations except with high phosphatase and low kinase concentrations. (B) Similarly Taz-SH3 and CusS G448A are titrated and activation CusR lacking the colocalization ligand is reported. High activation of CusR not capable of colocalizing is seen only with high kinase and low phosphatase expression. (C) The fold change due to colocalization, comparing CusR with and without ligand, is plotted. A balance of kinase and phosphatase expression is needed to maximize the colocalization effect.

To test if colocalization is capable of rerouting signal from a HK to either of two RRs, an SH3 domain was fused to Taz and an SH3 ligand peptide was fused to either CusR or CpxR. Phosphorylation of CusR or CpxR would drive expression of RFP or GFP reporters respectively. The CusS G448A and CpxA G420A phosphatases were incorporated into this system and all five components were expressed at a constant level, while the peptide was present on either

CusR or CpxR. Switching the peptide fusion from CpxR to CusR was sufficient to switch from high GFP to high RFP (Figure 6.10 A). As a comparison of device function with and without phosphatase presence, expression of CusR with peptide and CpxR without peptide was varied as Taz concentrations were held constant and either both or neither phosphatase were expressed. Interestingly, the non-localized CpxR had no impact on the peptide-fused CusR activity when phosphatases were not present (Figure 6.10 B), but exhibited significant inhibition from the over-expression of CpxR in the presence of phosphatase (Figure 6.10 C). This behavior would be consistent with components in this system exhibiting relatively transient association with HK and competition for binding not being significant unless a phosphate sink is present.

This phosphatase control allowed some high expression regimes, which otherwise produced too much background phosphorylation, to exhibit scaffold-directed phosphotransfer. However, phosphatase control ultimately proved to be unwieldy since it required a careful balance of phosphatase to kinase ratios and the phosphatase rate could easily overcome the kinase rate even for a colocalized HK-RR. While in some limited concentration regimes the phosphatase may have expanded the tolerance of the system to changes in concentrations of HK or RR, this effect was not significant enough to warrant introducing this additional sensitivity to phosphatase concentrations. Alternative means of improving the system performance and robustness were pursued.

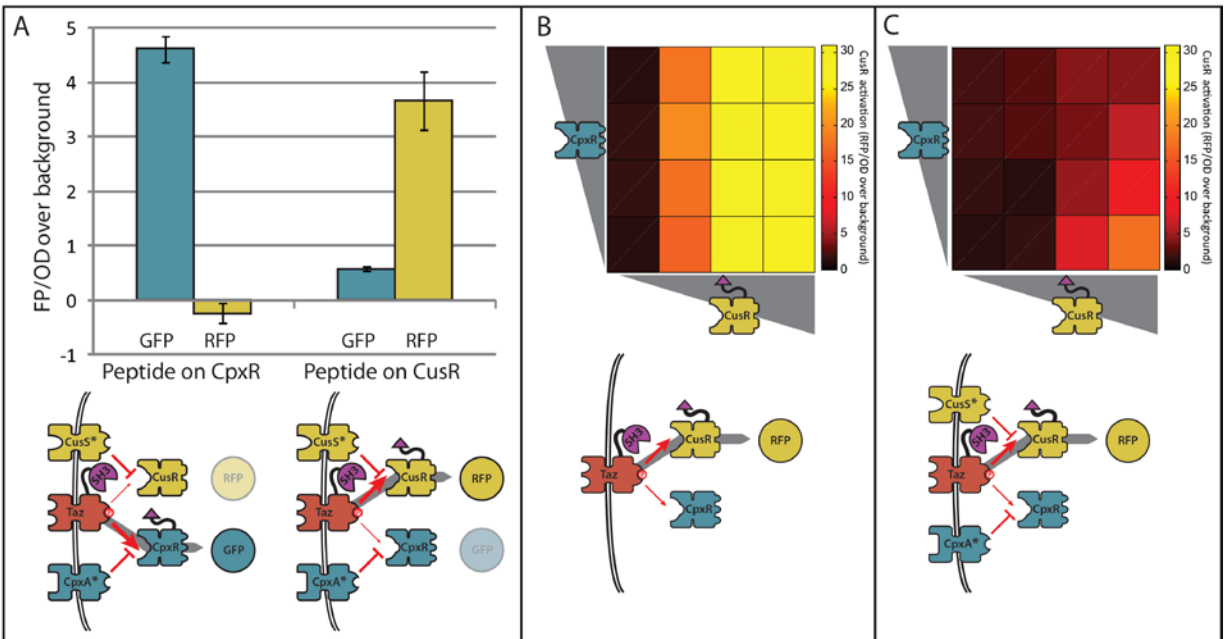


Figure 6.10. Switching specificity via colocalization and the effect of phosphatase activity. (A) Expression level is held constant for all signaling components: CusS G448A and CpxR G420A phosphatases, the Taz HK fused to SH3 domain, and CusR and CpxR with one of the two RRs fused to an SH3 ligand for colocalization. CusR and CpxR activity are measured via RFP and GFP, respectively. High GFP and RFP is seen only when CpxR or CusR, respectively, are fused to the SH3 ligand, indicating colocalization can direct signaling. (B) To test the effect of

competition between RRs, CpxR and ligand-fused CusR concentrations are varied, while CusR output is measured. CpxR does not compete with ligand fused CusR when phosphatases are not expressed. (C) When phosphatases are expressed, CusR and CpxR compete for phosphate, as evident by the decreased activation of CusR with the increased expression of CpxR.

Expression of a competitive inhibitor to reduce unwanted non-targeted signaling was investigated as an alternative means of improving the performance and robustness of our system. We hypothesized that a competitive inhibitor with a low affinity SH3 peptide may localize to Taz-SH3 and reduce phosphotransfer to non-localized RR, while a RR interacting with Taz-SH3 with a tight affinity peptide could competitively displace the inhibitor and effectively interact with the HK. Expression of a catalytically inactive version of the Taz cognate RR, OmpR D55A, fused to a weak affinity ($K_D=5\mu\text{M}$) SH3 peptide was used as a means of tuning the amount of competitive inhibition. Simulation of this strategy predicts that expression of the competitive inhibitor could expand the range of concentrations over which colocalization could effectively control phosphotransfer (Figure 6.11 A). We found that when expressed at the correct levels, OmpR D55A fused to a weak affinity peptide was capable of restoring colocalization-controlled phosphotransfer at high HK and RR expression levels (Figure 6.11 B).

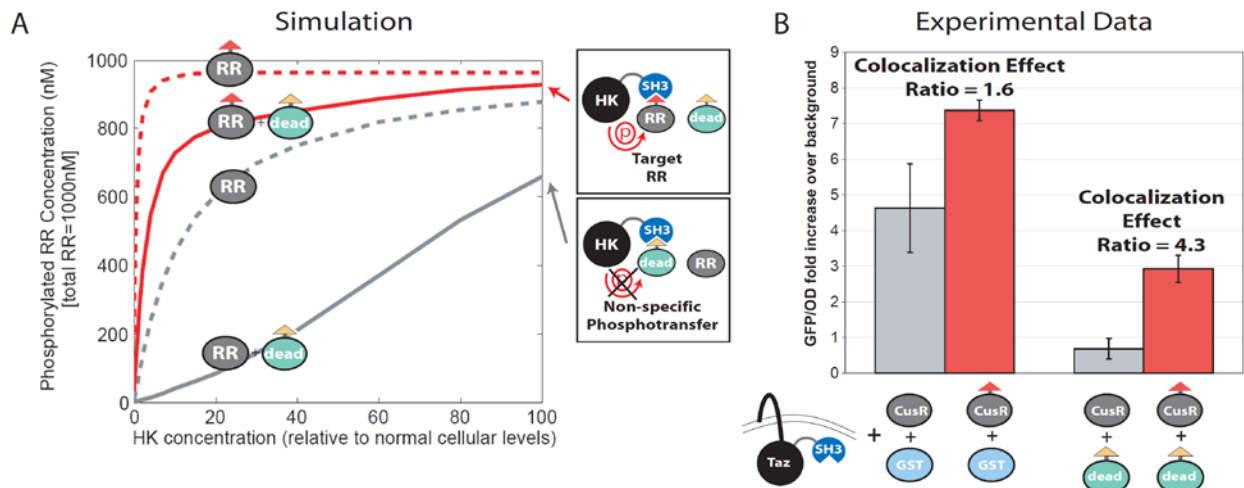


Figure 6.11. Expression of a competitive inhibitor can be used to expand the range of colocalization-controlled phosphotransfer. (A) A steady state simulation in which concentrations are varied for a HK containing a binding domain, with constant concentrations of a RR containing either a tight affinity ligand or no ligand, and a catalytically inactive RR containing a weak affinity ligand acting as a competitive inhibitor. Without expression of the competitive inhibitor, RR containing a tight affinity ligand (red dashed) is more highly activated than RR lacking the ligand (grey dashed), but the difference decreases quickly as kinase concentrations increase. Expression of a competitive inhibitor allows the ligand fused RR (red solid) to remain substantially more active than the RR containing no ligand (grey solid). (B) We found that expression of a competitive inhibitor was capable of improving the colocalization effect at high HK and RR expression levels. Here Taz is fused to one SH3 domain, CusR is fused to a tight ($K_D=0.1\mu\text{M}$) affinity ligand or no ligand, and a catalytically

inactive OmpR (D55A) is fused to a weak affinity ligand ($K_D=5\mu\text{M}$). Without a competitive inhibitor, colocalization only increases signal to 1.6 fold of non-colocalized levels, while competitive inhibitor increases the effect of colocalization to 4.3 fold of non-colocalized levels.

Similar to our findings with the phosphatase expression, expression of a competitive inhibitor was capable of restoring scaffold-controlled phosphotransfer at high HK and RR expression, but this effect was sensitive to the expression level of the competitive inhibitor. Although we intended for the competitive inhibitor to preferentially inhibit activation of non-recruited CusR, since the recruited CusR would dissociate the inhibitor (Figure 6.12 A), we found that at high OmpR D55A concentrations the CusR with peptide was also significantly inhibited (Figure 6.12 B). The degree of inhibition of CusR with peptide by competitive inhibitor was dependent on the affinity of the OmpR D55A fused peptide. However, even in the absence of peptide, CusR with peptide was inhibited by the high expression of competitive inhibitor. This sensitivity to the expression level of the competitive inhibitor diminishes its utility as a means of reducing HK or RR sensitivity.

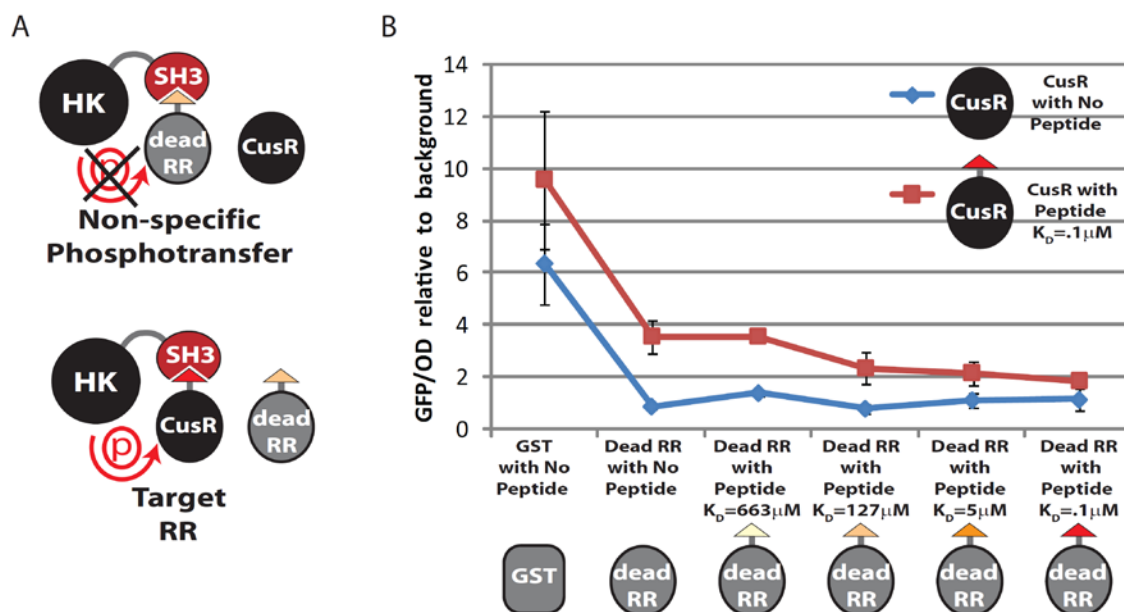


Figure 6.12. Competitive inhibitor reduces both colocalized and non-colocalized phosphotransfer. (A) The intended function of the competitive inhibitor is illustrated. CusR with high affinity ligand is expected to dissociate competitive inhibitor, freeing the HK for phosphotransfer. (B) At high expression of the HK, RR and competitive inhibitor, phosphotransfer is inhibited even with the high affinity ligand fused to the RR. The degree of inhibition is dependent on the affinity of the ligand fused to the competitive inhibitor. However, even with no ligand fused, the catalytically inactive OmpR D55A is capable of substantially reducing activation of targeted RR.

Three different strategies were investigated for improving the robustness of scaffold-dependent signaling: incorporating autoinhibition, expressing phosphatases, and expressing

a competitive inhibitor. All three of these strategies are thought to contribute to insulation from crosstalk in natural TCSs or eukaryotic signal transduction pathways. Phosphatase activity of cognate HK^{126,127} and competitive inhibition from cognate RR¹²⁷ play an important role in TCS insulation from crosstalk. Naturally scaffolded systems often make use of components that are activated upon binding to scaffold possibly as a means of increasing signaling efficiency and specificity^{137,140}. We found that all three strategies allowed scaffold-controlled signaling in a higher expression regime than otherwise possible. However, in all three cases we also found that overly inhibiting activity, such that even the targeted RR was not activated, was easily achieved and problematic. In the case of difficult to reverse, tight autoinhibition, the genetically encoded solutions of tuning the inhibitory peptide affinity or using two peptides for improved release of autoinhibition proved to be very effective and resulted in a more robust system. In the case of expressing phosphatases or competitive inhibitors the expression level had to be carefully tuned, which only shifted the concentration sensitivity to a different component, rather than reducing the over-all sensitivity to concentrations.

One clear difference between the autoinhibition strategy, which was successful, and the phosphatase and competitive inhibitor strategies, which were unsuccessful, is that the autoinhibition strategy introduces a new regulatory function, via protein engineering, without increasing the number of components. The phosphatase and competitive inhibitor activities are present in natural TCSs, but are encoded within the HK-RR pair, rather than on a separate protein. Implementing these activities with a separate protein allowed the degree of competitive inhibition of phosphatase activity to be easily tuned by expression level, thus facilitating our investigation of function but resulting in an overly sensitive system. It is tempting to hypothesize that the number of activities per component determines an upper limit on the maximum global robustness to varying concentrations of a system. If this hypothesis were proven to be correct it may be analogous to the waterbed effect used to describe the frequency sensitivity of electric systems¹⁶⁸, where sensitivity may be shifted to different frequencies by altering a system, but the integral of the system's sensitivity across all frequencies must remain constant. This idea that sensitivity cannot be removed, only shifted, may have an analog for genetic systems where the sensitivity can only be shifted, but it may be possible to shift the sensitivity from concentration related parameters to genetically encoded parameters to result in a system that is less sensitive to varying concentrations. For example, consider the system described above employing AiTaz(29A)-SH3 (Table S4.5), where the affinity of the SH3 domain to its ligand impacts both the autoinhibitory interaction and the scaffold recruitment. A mutation in the SH3 domain that has a small impact on affinity may have a large impact on system level behavior, while changes that effect component concentrations have a relatively small impact on this system. If indeed there is a trade-off between sensitivity to parameters that are primarily genetically encoded (e.g. affinities) and parameters that are dependent on a number of variables (i.e. protein concentrations), we may be able to designing systems with a small number of sophisticated components rather than a large number of single-function components in order to shift sensitivities away from the concentration dependent parameters.

6.4 – Future directions

We have demonstrated that scaffolding can be used to programmably control the phosphotransfer routing of TCSs. Additionally we have characterized the sensitivity of the system and demonstrated that autoinhibition can be used to improve the robustness of scaffold-controlled signaling to varying component concentrations. Due to the large body of work on TCSs, a number of opportunities are available for expand the utility and generalizability of scaffold-directed TCS signaling.

6.4.1 – Directions for improving scaffold-controlled signal transduction

This work focuses on the routing of phosphotransfer specificity, such that a HK only phosphorylates the targeted RR. Another important aspect of signal transduction is that the phosphorylation state of the RR target is set to the level appropriate for the stimulus. A characteristic common to non-cognate interactions is that phosphotransfer is constitutive and independent of stimulus^{122,126}, unlike cognate pairs that adjust the kinase to phosphatase ratio dependent on stimulus. We also found the non-cognates Taz-CusR and Taz-CpxR to exhibit this constitutive phosphotransfer. One solution to this stimulus insensitivity has recently been demonstrated for CpxA-OmpR, where, with a small set of mutations, the authors were able to gain stimulus dependent phosphotransfer between the non-cognate pair. Due to the high degree of structural and mechanistic conservation between TCS pathways, we expect the approach developed by Siryaporn and coworkers to be generalizable to other non-cognate pairs. Another means of incorporating scaffold-directed specificity and stimulus sensitivity may come from mimicking hybrid TCSs. Hybrid TCS systems generally contain one phosphotransfer domain directly linked to the transmembrane or stimulus sensing domain, as well as a number of downstream phosphotransfer domains that make up a multistep signal transduction pathway. The phosphotransfer domain adjacent to the stimulus-sensing domain would be expected to exhibit stimulus dependent kinase:phosphatase ratios, while downstream domains may be decoupled from stimulus and exhibit constitutive phosphotransfer. By organizing the engineered pathway such that the first phosphotransfer step exhibits stimulus sensitivity, constitutive downstream phosphotransfer would be expected to be sufficient for maintaining the appropriate phosphorylation state. Simply including the native RR may also be sufficient for gaining stimulus dependent signal transduction, as previous results suggest competition for phosphate can be limiting and native elimination of phosphate would likely be substantial (Figure 6.7 and Figure 6.10).

Another important consideration for broadly applying a scaffold-based strategy is the generalizability to other TCSs not investigated here. When we focused on two of the non-cognate pairs shown to weakly crosstalk in vitro, EnvZ-CpxR and EnvZ-CusR, scaffolding was sufficient to increase non-cognate phosphotransfer rate to a physiologically relevant level. However, this may not be directly applied to all non-cognate pairs, as was found with a brief attempt to direct EnvZ-CitB phosphotransfer. Several preliminary efforts at solving this problem are discussed above (Chapter 6.2.2). In particular, mutations to specificity-

determining residues seem to be a particularly promising approach for tuning kinase activity. Recent work has demonstrated that not only can specificity be tuned to increase phosphotransfer from a HK to a non-cognate RR target, but that these mutations are capable of broadening the specificity of a kinase¹⁴¹.

6.4.2 – Expanding the utility of scaffold-controlled signal transduction

We have shown that transcription of scaffolds designed to control colocalization can be used as an input to a pathway, which then outputs also to transcriptional regulation. Rather than driving a fluorescent reporter as an output, a scaffold protein can be driven as an output to either feedback on the pathway or to control another pathway. This has potential to scale, as a number of pathways may be chained with this transcriptional link. However, using transcription as both an input and an output to the signal transduction system fails to take advantage of the fast time-scales exhibited by signal transduction pathways. TCS pathways have the potential to rapidly process inputs, which is necessary for fast processes such as chemotaxis, and may be an important quality for future applications. If, rather than outputting to transcription, TCS pathways output to control colocalization directly through post-translational means, pathways could communicate on a rapid time scale. Some RRs respond to phosphorylation by changing affinity to a target protein, such as CheY-FliM. If a system were designed in which a CheY served as the target of one HK, then, in response to phosphorylation by this HK, it could be used to direct a target RR to a different HK through colocalization with a scaffold incorporating FliM. In this way, post-translational control of colocalization could serve as both an input and an output to a TCS pathway. Using colocalization as an output may also improve the portability of this strategy so that it may be applied more easily to eukaryotic organisms, since it does not rely on conservation of transcriptional machinery.

6.4.3 – Applications of scaffold-controlled signal transduction

One direct application of colocalization-controlled phosphotransfer is used in assaying protein-protein interaction strength in prokaryotes. Binding domains may be tethered to a HK while ligands are tethered to a RR with a transcriptional reporter indicating interaction strength. Here we demonstrate the feasibility of this assay with different PDZ domains and ligands fused to Taz and CusR respectively (Figure 6.13). This is analogous to a yeast-two-hybrid assay, but may be used in prokaryotes where prokaryotic specific physiology or selection is advantageous.

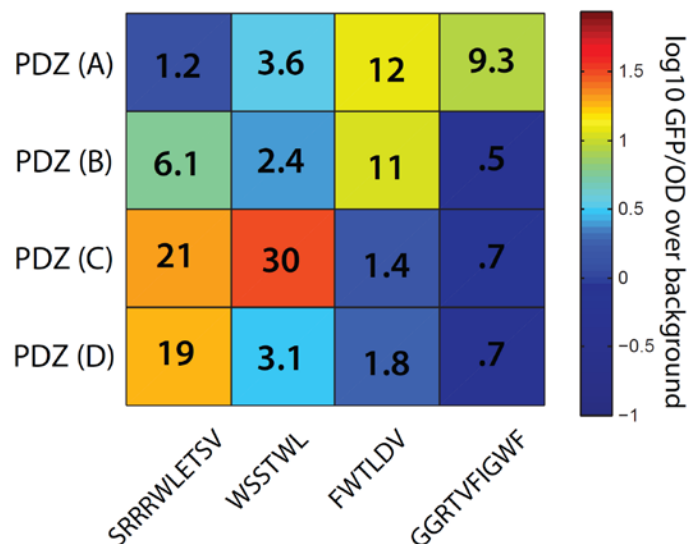


Figure 6.13. Colocalization of TCSs can serve as an assay for binding affinity. Four different PDZ domains are fused to Taz (x-axis) and four different ligands are fused to CusR (y-axis). The fluorescent output from every combination is plotted indicating the strength of the interaction between the PDZ domain and ligand.

As cells are engineered to function in complex and changing environments, the rational engineering of signal transduction will become increasingly important. Stimuli can change meaning different contexts, and scaffold-directed signal transduction provides an ideal means of multiplexing signaling machinery to respond to changing conditions. Designing systems from the bottom-up with individually tunable interactions will facilitate their adaptation for use in other synthetic systems. We hope the strategy and parts developed here will be broadly applicable to a number of future applications.

6.5 – Conclusions

We have demonstrated that design principles learned from higher organisms can be abstracted, generalized and implemented with well-characterized, modular parts. Inspired by the modular, highly evolvable, yet robust natural signal transduction systems of higher organisms, we designed a simple scaffold to colocalize prokaryotic two-component signaling components. By building a system from relatively simple, well-characterized parts we gain a clear understanding of design principles for achieving robust signaling pathways with high fidelity. Although a simplified version of a natural scaffold designed solely to tether components was sufficient to direct signaling, we found the signaling specificity to be highly sensitive to component concentrations. However, including activation of kinase upon localization to the scaffolded complex, often a property of natural scaffolded pathways, improved the robustness of the signaling pathway. This illustrates how, although refactoring a system to function with simple modular parts can lead to increased sensitivity, engineering more sophisticated parts, though methods such as protein engineering, can be used to achieve a modular, tunable system with robust behavior.

Chapter 7 – References

1. Peisajovich, S.G., Garbarino, J.E., Wei, P. & Lim, W.A. Rapid diversification of cell signaling phenotypes by modular domain recombination. *Science* **328**, 368-72 (2010).
2. Bashor, C.J., Helman, N.C., Yan, S. & Lim, W.A. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science* **319**, 1539-43 (2008).
3. Bashor, C.J., Horwitz, A. a, Peisajovich, S.G. & Lim, W.A. Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annual review of biophysics* **39**, 515-37 (2010).
4. Keasling, J.D. Synthetic biology and the development of tools for metabolic engineering. *Metabolic engineering* **14**, 189-195 (2012).
5. Purnick, P.E.M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nature reviews. Molecular cell biology* **10**, 410-22 (2009).
6. Thompson, K.E., Bashor, C.J., Lim, W.A. & Keating, A.E. SYNZIP Protein Interaction Toolbox: in Vitro and in Vivo Specifications of Heterospecific Coiled-Coil Interaction Domains. *ACS synthetic biology* **1**, 118-129 (2012).
7. Dueber, J.E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnology* **27**, 753-9 (2009).
8. Moon, T.S., Dueber, J.E., Shiue, E. & Prather, K.L.J. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metabolic engineering* **12**, 298-305 (2010).
9. Salis, H.M., Mirsky, E. a & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nature biotechnology* **27**, 946-50 (2009).
10. Ro, D.-K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-3 (2006).
11. Keasling, J.D. Synthetic biology for synthetic chemistry. *ACS chemical biology* **3**, 64-76 (2008).
12. Fortman, J.L. *et al.* Biofuel alternatives to ethanol: pumping the microbial well. *Trends in biotechnology* **26**, 375-81 (2008).
13. Atsumi, S. & Liao, J.C. Metabolic engineering for advanced biofuels production from *Escherichia coli*. *Current opinion in biotechnology* **19**, 414-9 (2008).

14. Cases, I. & De Lorenzo, V. Genetically modified organisms for the environment: stories of success and failure and what we have learned from them. *International microbiology the official journal of the Spanish Society for Microbiology* **8**, 213-222 (2005).
15. Penyalver, R., Vicedo, B. & Lopez, M.M.-M. Use of the genetically engineered *Agrobacterium* strain K1026 for biological control of crown gall. *European Journal of Plant Pathology* **106**, 801-810 (2000).
16. Parsa, S. & Pfeifer, B. Engineering bacterial vectors for delivery of genes and proteins to antigen-presenting cells. *Molecular pharmaceutics* **4**, 4-17 (2007).
17. Brockstedt, D.G. & Dubensky, T.W. Promises and challenges for the development of *Listeria monocytogenes*-based immunotherapies. *Expert Review of Vaccines* **7**, 1069-1084 (2008).
18. Pawelek, J.M., Low, K.B. & Bermudes, D. Bacteria as tumour-targeting vectors. *The lancet oncology* **4**, 548-56 (2003).
19. Torres, N.V. & Voit, E.O. Pathway Analysis and Optimization in Metabolic Engineering. *System* 1-17 (2002).doi:10.1017/CBO9780511546334
20. Kirk, T.K. *et al.* Biological Confinement of Genetically Engineered Organisms. *Biological confinement of genetically engineered organisms* **2009**, xx + 255 pp. (2004).
21. Newcomb, J., Carlson, R. & Aldrich, S.C. *Genome synthesis and design futures: implications for the U.S. economy*. Bio economic research associates only exec summary (Bio Economic Research Associates (bio-era): 2006).
22. Itaya, M., Tsuge, K., Koizumi, M. & Fujita, K. Combining two genomes in one cell: Stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15971-15976 (2005).
23. Dixon, R.A. & Postgate, J.R. Genetic transfer of nitrogen fixation from *Klebsiella pneumoniae* to *Escherichia coli*. *Nature* **237**, 102-103 (1972).
24. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339-342 (2000).
25. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-8 (2000).
26. Tabor, J.J., Bayer, T.S., Simpson, Z.B., Levy, M. & Ellington, A.D. Engineering stochasticity in gene expression. *Molecular bioSystems* **4**, 754-61 (2008).

27. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590-593 (2000).
28. Barrick, D. *et al.* Quantitative analysis of ribosome binding sites in E.coli. *Nucleic Acids Research* **22**, 1287-1295 (1994).
29. Yokobayashi, Y., Weiss, R. & Arnold, F.H. Directed evolution of a genetic circuit. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16587-16591 (2002).
30. Anderson, J.C., Clarke, E.J., Arkin, A.P. & Voigt, C. a Environmentally controlled invasion of cancer cells by engineered bacteria. *Journal of molecular biology* **355**, 619-27 (2006).
31. De Smit, M.H. & Van Duin, J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 7668-7672 (1990).
32. Carrier, T. a & Keasling, J.D. Library of synthetic 5' secondary structures to manipulate mRNA stability in Escherichia coli. *Biotechnology progress* **15**, 58-64 (1999).
33. Skerker, J.M. *et al.* Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043-54 (2008).
34. Reina, J. *et al.* Computer-aided design of a PDZ domain to recognize new target sequences. *Nature Structural Biology* **9**, 621-627 (2002).
35. Rackham, O. & Chin, J.W. A network of orthogonal ribosome mRNA pairs. *Nature Chemical Biology* **1**, 159-166 (2005).
36. Mandell, J.G. & Barbas, C.F. Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Research* **34**, W516-W523 (2006).
37. Jin, T. & Inouye, M. Ligand binding to the receptor domain regulates the ratio of kinase to phosphatase activities of the signaling domain of the hybrid Escherichia coli transmembrane receptor, Taz1. *Journal of molecular biology* **232**, 484-92 (1993).
38. Win, M.N. & Smolke, C.D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14283-14288 (2007).
39. Win, M.N. & Smolke, C.D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456-460 (2008).

40. Dueber, J.E., Yeh, B.J., Chak, K. & Lim, W.A. Reprogramming control of an allosteric signaling switch through modular recombination. *Science* **301**, 1904-8 (2003).
41. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 3581-3586 (2005).
42. Yeh, B.J. & Lim, W.A. Synthetic biology: lessons from the history of synthetic organic chemistry. *Nature Chemical Biology* **3**, 521-525 (2007).
43. Voigt, C.A. Genetic parts to program bacteria. *Current Opinion in Biotechnology* **17**, 548-557 (2006).
44. Endy, D. Foundations for engineering biology. *Nature* **438**, 449-53 (2005).
45. McDaniel, R. & Weiss, R. Advances in synthetic biology: on the path from prototypes to applications. *Current Opinion in Biotechnology* **16**, 476-483 (2005).
46. Selzer, G., Som, T., Itoh, T. & Tomizawa, J. The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids. *Cell* **32**, 119-129 (1983).
47. McAdams, H.H. & Arkin, A. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 814-819 (1997).
48. Samoilov, M.S. & Arkin, A.P. Deviant effects in molecular reaction pathways. *Nature Biotechnology* **24**, 1235-1240 (2006).
49. Lu, T., Ferry, M., Weiss, R. & Hasty, J. A molecular noise generator. *Physical Biology* **5**, 036006 (2008).
50. Wolf, D.M., Vazirani, V.V. & Arkin, A.P. Diversity in times of adversity: probabilistic strategies in microbial survival games. *Journal of Theoretical Biology* **234**, 227-253 (2005).
51. Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075-8 (2005).
52. Thattai, M. & Van Oudenaarden, A. Stochastic gene expression in fluctuating environments. *Genetics* **167**, 523-530 (2004).
53. Blake, W.J. *et al.* Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. *Molecular Cell* **24**, 853-865 (2006).

54. Anderson, J.C., Voigt, C.A. & Arkin, A.P. Environmental signal integration by a modular AND gate. *Molecular Systems Biology* **3**, 133 (2007).
55. Bailey, J.E. Host-vector interactions in Escherichia coli. *Advances in biochemical engineering/biotechnology* **48**, 29-52 (1993).
56. You, L., Cox, R.S., Weiss, R. & Arnold, F.H. Programmed population control by cell-cell communication and regulated killing. *Nature* **428**, 868-871 (2004).
57. Canton, B., Labno, A. & Endy, D. Registry of Standard Biological Parts Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology* **26**, 787-793 (2008).
58. Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D. & Keasling, J.D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology* **21**, 796-802 (2003).
59. Kizer, L., Pitera, D.J., Pfleger, B.F. & Keasling, J.D. Application of Functional Genomics to Pathway Optimization for Increased Isoprenoid Production. *Applied and Environmental Microbiology* **74**, 3229-3241 (2008).
60. Pfleger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnology* **24**, 1027-1032 (2006).
61. Voigt, C.A., Wolf, D.M. & Arkin, A.P. The Bacillus subtilis sin operon: an evolvable network motif. *Genetics* **169**, 1187-1202 (2005).
62. Bayer, T.S. & Smolke, C.D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature Biotechnology* **23**, 337-343 (2005).
63. Isaacs, F.J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology* **22**, 841-847 (2004).
64. Rinaudo, K. *et al.* A universal RNAi-based logic evaluator that operates in mammalian cells. *Nature biotechnology* **25**, 795-801 (2007).
65. Link, K.H. *et al.* Engineering high-speed allosteric hammerhead ribozymes. *Biological Chemistry* **388**, 779-786 (2007).
66. Temme, K. *et al.* Induction and relaxation dynamics of the regulatory network controlling the type III secretion system encoded within Salmonella pathogenicity island 1. *Journal of Molecular Biology* **377**, 47-61 (2008).

67. Beisel, C.L., Bayer, T.S., Hoff, K.G. & Smolke, C.D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Molecular Systems Biology* **4**, 224 (2008).
68. Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130-4 (2005).
69. Price, N.D., Reed, J.L. & Palsson, B.O. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology* **2**, 886-897 (2004).
70. Dougherty, M.J. & Arnold, F.H. Directed evolution: new parts and optimized function. *Current Opinion in Biotechnology* **20**, 486-91 (2009).
71. Zhang, K., Sawaya, M.R., Eisenberg, D.S. & Liao, J.C. Expanding metabolism for biosynthesis of nonnatural alcohols. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20653-20658 (2008).
72. Conrado, R.J., Varner, J.D. & DeLisa, M.P. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Current opinion in biotechnology* **19**, 492-9 (2008).
73. Miles, E.W., Rhee, S. & Davies, D.R. The Molecular Basis of Substrate Channeling. *Journal of Biological Chemistry* **274**, 12193-12196 (1999).
74. Stroud, R.M. An electrostatic highway. *Nature Structural Biology* **1**, 131-134 (1994).
75. Narayanaswamy, R. *et al.* Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10147-10152 (2009).
76. An, S., Kumar, R., Sheets, E.D. & Benkovic, S.J. Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. *Science* **320**, 103-106 (2008).
77. Fierobe, H.P. *et al.* Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional complexes. *The Journal of biological chemistry* **276**, 21257-61 (2001).
78. Fierobe, H.-P. *et al.* Action of designer cellulosomes on homogeneous versus complex substrates: controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. *The Journal of biological chemistry* **280**, 16325-34 (2005).
79. Pitera, D.J., Paddon, C.J., Newman, J.D. & Keasling, J.D. Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metabolic Engineering* **9**, 193-207 (2007).

80. Moon, T.S., Yoon, S.-H., Lanza, A.M., Roy-Mayhew, J.D. & Prather, K.L.J. Production of Glucaric Acid from a Synthetic Pathway in Recombinant *Escherichia coli*. *Applied and Environmental Microbiology* **75**, 540-548 (2009).
81. Welch, G.R. On the role of organized multienzyme systems in cellular metabolism: a general synthesis. *Progress in biophysics and molecular biology* **32**, 103-91 (1977).
82. Ljungcrantz, P. *et al.* Construction of an artificial bifunctional enzyme, beta-galactosidase/galactose dehydrogenase, exhibiting efficient galactose channeling. *Biochemistry* **28**, 8786-8792 (1989).
83. Pettersson, H. & Pettersson, G. Kinetics of the coupled reaction catalysed by a fusion protein of beta-galactosidase and galactose dehydrogenase. *Biochimica et Biophysica Acta* **1549**, 155-160 (2001).
84. Conrado, R.J., Mansell, T.J., Varner, J.D. & DeLisa, M.P. Stochastic reaction-diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway. *Metabolic Engineering* **9**, 355-363 (2007).
85. Zarrinpar, A., Park, S.H. & Lim, W. A. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **426**, 676-80 (2003).
86. Havranek, J.J. & Harbury, P.B. Automated design of specificity in molecular recognition. *Nature Structural Biology* **10**, 45-52 (2003).
87. Reinke, A.W., Grant, R. A. & Keating, A.E. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *Journal of the American Chemical Society* **132**, 6025-31 (2010).
88. Nordon, R.E., Craig, S.J. & Foong, F.C. Molecular engineering of the cellulosome complex for affinity and bioenergy applications. *Biotechnology letters* **31**, 465-76 (2009).
89. Haimovitz, R. *et al.* Cohesin-dockerin microarray: Diverse specificities between two complementary families of interacting protein modules. *Proteomics* **8**, 968-79 (2008).
90. Posern, G. *et al.* Development of highly selective SH3 binding peptides for Crk and CRKL which disrupt Crk-complexes with DOCK180, SoS and C3G. *Oncogene* **16**, 1903-12 (1998).
91. Tonikian, R. *et al.* A Specificity Map for the PDZ Domain Family. *PLoS Biology* **6**, 17 (2008).
92. Harris, B.Z., Hillier, B.J. & Lim, W.A. Energetic determinants of internal motif recognition by PDZ domains. *Biochemistry* **40**, 5921-5930 (2001).

93. Fuh, G. *et al.* Analysis of PDZ domain-ligand interactions using carboxyl-terminal phage display. *The Journal of biological chemistry* **275**, 21486-91 (2000).
94. Kim, a S., Kakalis, L.T., Abdul-Manan, N., Liu, G. a & Rosen, M.K. Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature* **404**, 151-8 (2000).
95. Acharya, A., Ruvinov, S.B., Gal, J., Moll, J.R. & Vinson, C. A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: amino acids I, V, L, N, A, and K. *Biochemistry* **41**, 14122-31 (2002).
96. Grünberg, R., Ferrar, T.S., van der Sloot, A.M., Constante, M. & Serrano, L. Building blocks for protein interaction devices. *Nucleic acids research* **38**, 2645-62 (2010).
97. Grigoryan, G., Reinke, A.W. & Keating, A.E. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* **458**, 859-64 (2009).
98. Levskaya, A., Weiner, O.D., Lim, W.A. & Voigt, C. a Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **461**, 997-1001 (2009).
99. Banaszyński, L. a, Liu, C.W. & Wandless, T.J. Characterization of the FKBP.rapamycin.FRB ternary complex. *Journal of the American Chemical Society* **127**, 4715-21 (2005).
100. Bierer, B.E. *et al.* Two distinct signal transmission pathways in T lymphocytes are inhibited by complexes formed between an immunophilin and either FK506 or rapamycin. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 9231-9235 (1990).
101. Leibovitz, E., Ohayon, H., Gounon, P. & Béguin, P. Characterization and subcellular localization of the Clostridium thermocellum scaffoldin dockerin binding protein SdbA. *Journal of bacteriology* **179**, 2519-23 (1997).
102. Di Guan, C., Li, P., Riggs, P.D. & Inouye, H. Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene* **67**, 21-30 (1988).
103. Robinson, C.R. & Sauer, R.T. Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 5929-34 (1998).
104. Mathews, D.H., Sabina, J., Zuker, M. & Turner, D.H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology* **288**, 911-940 (1999).

105. Levchenko, A., Bruck, J. & Sternberg, P. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proceedings of the National Academy of Sciences* **97**, 5818-5823 (2000).
106. Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Research* **37**, e16 (2009).
107. Li, M.Z. & Elledge, S.J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature Methods* **4**, 251-256 (2007).
108. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 12-16 (2009).
109. Shetty, R.P., Endy, D. & Knight, T.F. Engineering BioBrick vectors from BioBrick parts. *Journal of biological engineering* **2**, 5 (2008).
110. Anderson, J.C. *et al.* BglBricks: A flexible standard for biological part assembly. *Journal of biological engineering* **4**, 1 (2010).
111. Mingardon, F., Chanal, A., Tardif, C., Bayer, E. a & Fierobe, H.-P. Exploration of new geometries in cellulosome-like chimeras. *Applied and environmental microbiology* **73**, 7138-49 (2007).
112. Xu, Q. *et al.* The Cellulosome System of *Acetivibrio cellulolyticus* Includes a Novel Type of Adaptor Protein and a Cell Surface Anchoring Protein. *Journal Of Bacteriology* **185**, 4548-4557 (2003).
113. Bülow, L. & Mosbach, K. Multienzyme systems obtained by gene fusion. *Trends in biotechnology* **9**, 226-231 (1991).
114. Tamsir, A., Tabor, J.J. & Voigt, C. a Robust multicellular computing using genetically encoded NOR gates and chemical “wires”. *Nature* **469**, 212-5 (2011).
115. Friedland, A.E. *et al.* Synthetic gene networks that count. *Science* **324**, 1199-202 (2009).
116. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516-9 (2008).
117. Lucks, J.B., Qi, L., Whitaker, W.R. & Arkin, A.P. Toward scalable parts families for predictable design of biological circuits. *Current opinion in microbiology* **11**, 567-73 (2008).

118. Dueber, J.E., Yeh, B.J., Bhattacharyya, R.P. & Lim, W. a Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry. *Current opinion in structural biology* **14**, 690-9 (2004).
119. Bhattacharyya, R.P., Reményi, A., Yeh, B.J. & Lim, W.A. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annual review of biochemistry* **75**, 655-80 (2006).
120. Delebecque, C.J., Lindner, A.B., Silver, P. a & Aldaye, F. a Organization of intracellular reactions with rationally designed RNA assemblies. *Science (New York, N.Y.)* **333**, 470-4 (2011).
121. Moon, T.S. *et al.* Construction of a genetic multiplexer to toggle between chemosensory pathways in Escherichia coli. *Journal of molecular biology* **406**, 215-27 (2011).
122. Siryaporn, A., Perchuk, B.S., Laub, M.T. & Goulian, M. Evolving a robust signal transduction pathway from weak cross-talk. *Molecular systems biology* **6**, 452 (2010).
123. Skerker, J.M., Prasol, M.S., Perchuk, B.S., Biondi, E.G. & Laub, M.T. Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. *PLoS biology* **3**, e334 (2005).
124. Casino, P., Rubio, V. & Marina, A. Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell* **139**, 325-36 (2009).
125. Yamamoto, K. *et al.* Functional characterization in vitro of all two-component signal transduction systems from Escherichia coli. *The Journal of biological chemistry* **280**, 1448-56 (2005).
126. Groban, E.S., Clarke, E.J., Salis, H.M., Miller, S.M. & Voigt, C. a Kinetic buffering of cross talk between bacterial two-component sensors. *Journal of molecular biology* **390**, 380-93 (2009).
127. Laub, M.T. & Goulian, M. Specificity in two-component signal transduction pathways. *Annual review of genetics* **41**, 121-45 (2007).
128. Siryaporn, A. & Goulian, M. Cross-talk suppression between the CpxA-CpxR and EnvZ-OmpR two-component systems in E. coli. *Molecular microbiology* **70**, 494-506 (2008).
129. Khlebnikov, A., Datsenko, K.A., Skaug, T., Wanner, B.L. & Keasling, J.D. Homogeneous expression of the P(BAD) promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology (Reading, England)* **147**, 3241-7 (2001).

130. Grünberg, R., Ferrar, T.S., Van Der Sloot, A.M., Constante, M. & Serrano, L. Building blocks for protein interaction devices. *Nucleic Acids Research* **38**, 2645-2662 (2010).
131. Whitaker, W.R. & Dueber, J.E. Metabolic pathway flux enhancement by synthetic protein scaffolding. *Methods in enzymology* **497**, 447-68 (2011).
132. Dueber, J.E., Mirsky, E. a & Lim, W. a Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nature biotechnology* **25**, 660-2 (2007).
133. Lohmueller, J.J., Armel, T.Z. & Silver, P. a A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic acids research* 1-8 (2012).
134. Knudsen, B.S. *et al.* Affinity and specificity requirements for the first Src homology 3 domain of the Crk proteins. *The EMBO journal* **14**, 2191-8 (1995).
135. Chapman, S. a & Asthagiri, A.R. Quantitative effect of scaffold abundance on signal propagation. *Molecular systems biology* **5**, 313 (2009).
136. Batchelor, E. & Goulian, M. Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 691-6 (2003).
137. Rajakulendran, T., Sahmi, M., Lefrançois, M., Sicheri, F. & Therrien, M. A dimerization-dependent mechanism drives RAF catalytic activation. *Nature* **461**, 542-5 (2009).
138. Yeh, B.J., Rutigliano, R.J., Deb, A., Bar-Sagi, D. & Lim, W. a Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors. *Nature* **447**, 596-600 (2007).
139. Tanaka, T. *et al.* NMR structure of the histidine kinase domain of the E. coli osmosensor EnvZ. *Nature* **396**, 88-92 (1998).
140. Bhattacharyya, R.P. *et al.* The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Science* **311**, 822-6 (2006).
141. Capra, E.J. *et al.* Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. *PLoS genetics* **6**, e1001220 (2010).
142. Goulian, M. Two-component signaling circuit structure and properties. *Current opinion in microbiology* **13**, 184-9 (2010).

143. Datsenko, K. a & Wanner, B.L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6640-5 (2000).
144. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology* **2**, 2006.0008 (2006).
145. Miller, W.G., Leveau, J.H. & Lindow, S.E. Improved gfp and inaZ broad-host-range promoter-probe vectors. *Molecular plant-microbe interactions : MPMI* **13**, 1243-50 (2000).
146. Fisher, M. a, McKinley, K.L., Bradley, L.H., Viola, S.R. & Hecht, M.H. De novo designed proteins from a library of artificial sequences function in Escherichia coli and enable cell growth. *PloS one* **6**, e15364 (2011).
147. Kelley, L.A. & Sternberg, M.J.E. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363-371 (2009).
148. Itzkovitz, S., Hodis, E. & Segal, E. Overlapping codes within protein-coding sequences. *Genome research* **20**, 1582-9 (2010).
149. Baird, S.D., Turcotte, M., Korneluk, R.G. & Holcik, M. Searching for IRES. *RNA (New York, N.Y.)* **12**, 1755-85 (2006).
150. Ray, P.S., Grover, R. & Das, S. Two internal ribosome entry sites mediate the translation of p53 isoforms. *EMBO reports* **7**, 404-10 (2006).
151. Thomas, J.-C., Ughy, B., Lagoutte, B. & Ajlani, G. A second isoform of the ferredoxin:NADP oxidoreductase generated by an in-frame initiation of translation. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 18368-73 (2006).
152. Cot, S.S.-W., So, A.K.-C. & Espie, G.S. A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria. *Journal of bacteriology* **190**, 936-45 (2008).
153. Kofoed, E. Tandem translation starts in the cheA locus of Escherichia coli. *Journal of bacteriology* **173**, 2116-2119 (1991).
154. Matsumura, P., Silverman, M. & Simon, M. Synthesis of mot and che gene products of Escherichia coli programmed by hybrid ColE1 plasmids in minicells. *Journal of bacteriology* **132**, 996-1002 (1977).

155. Cabantous, S. *et al.* Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from *Mycobacterium tuberculosis*. *Journal of structural and functional genomics* **6**, 113-9 (2005).
156. Sachadyn, P., Stanisławska-Sachadyn, A., Kabat, E.M., Zielińska, A. & Kur, J. A cryptic ribosome binding site, false signals in reporter systems and avoidance of protein translation chaos. *Journal of biotechnology* **143**, 169-72 (2009).
157. Li, G.-W., Oh, E. & Weissman, J.S. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* (2012).doi:10.1038/nature10965
158. Chang, B., Halgamuge, S. & Tang, S.-L. Analysis of SD sequences in completed microbial genomes: non-SD-led genes are as common as SD-led genes. *Gene* **373**, 90-9 (2006).
159. Campbell, R.E. *et al.* A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7877-82 (2002).
160. Tran, T., Terwilliger, T.C., Waldo, G.S. & Pe, J.-denis Engineering and characterization of a superfolder green fluorescent protein. **24**, 79-89 (2006).
161. Schultz, J. *et al.* Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels. *Nature Structural Biology* **5**, 19-24 (1998).
162. Park, S.-H., Zarrinpar, A. & Lim, W. a Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science* **299**, 1061-4 (2003).
163. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a single cell. *Science* **297**, 1183-6 (2002).
164. Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature genetics* **31**, 69-73 (2002).
165. Gruber, A.R., Lorenz, R., Bernhart, S.H. & Neubo, R. The Vienna RNA Websuite. **36**, 70-74 (2008).
166. Jiang, M., Tzeng, Y.-L., Feher, V. a., Perego, M. & Hoch, J. a. Alanine mutants of the Spo0F response regulator modifying specificity for sensor kinases in sporulation initiation. *Molecular Microbiology* **33**, 389-395 (1999).
167. Zhu, Y. & Inouye, M. The role of the G2 box, a conserved motif in the histidine kinase superfamily, in modulating the function of EnvZ. *Molecular microbiology* **45**, 653-63 (2002).

168. Megretski, A. The Waterbed Effect. *Control* **23**, 1-9 (2004).

Chapter 8 – Appendix

8.1 – AUTOMODEL code

Included below is the code discussed in Chapter 6.2.3 and used to generate Figure 3.3. This code was validated for MATLAB 7.10.0 (R2010a) on Windows 7. Included is the core of the code “setup_MIEfig3.m” where the model is specified (in a designated section), and the differential and stochastic equations are inferred. Two other scripts “varyX.m” and “run.m” are included. To generate Figure 3.3, create these three files and run the file “run.m”.

8.1.1 – Matlab code for “setup_MIEfig3.m”

```
function system = setup_MIEfig3(model)
% This code specifies the model and automatically generates the
% differential and stochastic equations to simulate the model.
%
% The model specified below is a simple scaffold that has 5 binding
% domains, each of which binds a specific protein.
% The original version was written by Weston Whitaker in 2010.
% Email me at westonwhitaker@gmail.com if you need help with it.
%
% Example of how to use this code:
% system = setup(); % this creates a default model object specified below
% X0=[100, 500, 200, 0, 0, 0]; % sets the initial concentrations at t=0 of the basic
    components
% X0=system.fixX(X0, system.model.pairs); % this adds on all the combinations of
    basic components and sets them to zero
% t_max=24*60*60; %this sets the maximal simulations time to be run. I generally just
    give it way more time than it should need to be sure it reaches equilibrium
% SimulationResults = system.deModel(X0, t_max); % outputs the time and
    concentrations from the simulation
% loglog(SimulationResults.t, SimulationResults.X); %plot the results
% Xfinal=SimulationResults.X(end,:); %just take the last time point for the equilibrium
    concentrations
% Xsimple=system.simplifyX(Xfinal) % this just simplifies the components to their
    stoichiometries for plotting, ignoring all the various binding states
% legend(Xsimple.key);
%
% How to specify a model with this code:
% I wrote this code so that I could input a model the way that I think
% about it, rather than through a tedious set of equations. I find that not
% only does this setup save me time, but it greatly minimizes the mistakes
% I would make if I tried to just write out the equations. It was designed
% with modeling the scaffolding of bacterial signaling proteins in mind.
```

```
% Basically the program sees each protein as a set of domains, each of
% which can bind to other domains with a specified affinity. Also,
% additional (non-binding) reactions can be added (such as
% phosphotransfer). To model a system, you just need to specify how many of
% each domain every protein in your model has, fill out the affinity matrix
% and then add in any non-binding interactions that occur between domains.
% Each of the original explicitly specified proteins will be associated
% with a differential equation. Then the program will automatically
% calculate all the combinations of binding interactions that can occur and
% will associate each one of these basic protein / binding states
% combinations and with a differential equation. The program automatically
% generates the equations for differential and stochastic simulation.
```

```
%%
```

```
if (exist('model')) % This allows external editing of parameters. Do not edit
```

```
    key.x=model.key.x;
```

```
    E=model.E;
```

```
    Kds=model.Kds;
```

```
    Kfs=model.Kfs;
```

```
    Krs=model.Krs;
```

```
    Rxns=model.Rxns;
```

```
    Kcat=model.Kcat;
```

```
else
```

```
%% % -----DEFAULT MODEL: EDIT SECTION BELOW TO DEFINE MODEL-----
```

```
    % note: All parameters in units of nM and seconds unless otherwise specified
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% PROTEIN KEY %%%%%%%%%%%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% >> MODIFY THIS (start) << %%%%%%%%%%%
```

```
key.x{1} = 'Scaffolding'; %
```

```
key.x{2} = 'enzA'; %
```

```
key.x{3} = 'enzB'; %
```

```
key.x{4} = 'enzC'; %
```

```
key.x{5} = 'enzD'; %
```

```
key.x{6} = 'enzE'; %
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% >> MODIFY THIS (end) << %%%%%%%%%%%
```

```
%%%%%%%%%% E = ELEMENTS : DEFINED BY DOMAIN COMPONENTS %%%%%%%%%%%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
% Note: each domain with a unique set of binding affinities must be
```

```
% individually specified (add another column).
```

```
%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% >> MODIFY THIS (start) << %%%%%%%%%%%
```

```
%    1  2  3  4  5  6  7  8  9  10
```

```

%   A   a   B   b   C   c   D   d   E   e
E(1,:)= [ 1   0   1   0   1   0   1   0   1   0 ]; % Scaffolding
E(2,:)= [ 0   1   0   0   0   0   0   0   0   0 ]; % enzA
E(3,:)= [ 0   0   0   1   0   0   0   0   0   0 ]; % enzB
E(4,:)= [ 0   0   0   0   0   1   0   0   0   0 ]; % enzC
E(5,:)= [ 0   0   0   0   0   0   0   1   0   0 ]; % enzD
E(6,:)= [ 0   0   0   0   0   0   0   0   0   1 ]; % enzE
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% >> MODIFY THIS (end) << %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% CONSTANT SPECIFICATION %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% >> MODIFY THIS (start) << %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
kd=100; %nM < THIS IS THE PARAMETER TO MODIFY TO CHANGE BINDING AFFINITY
diffusionLimited = .083; % nM-1 s-1
linkLength=15; %nm this is used to approximate the effective concentration effect
                    for intramolecular interactions
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% >> MODIFY THIS (end) << %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% coliDivTime= 20*60; % (s)
% Kdeg=log(2)/coliDivTime; %Ref 1 shows CusR deg is very slow -> assume dilution
                    of RRP and replacement with RR expression
v=(4/3)*pi*linkLength^3; %nm^3 volume in which linked RR can be relavite to HK
f=1.6667e9; %conversion constant for nM per molecule/nm^3
effectiveConcentration=f*1/v; %nM effective concentration when linked

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Complex Size Limit %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Sets the number of the same type of interaction allowed per molecule.
% Prevents consideration of long unlikely polymers of components.
% NOTE: seems to work for >1, but only well tested for PolymerizationLimit=1
PolymerizationLimit=1; % for the MIE paper, this should just be kept at
                    % 1, but if you had dimers and you wanted to
                    % allow for simulation of complexes where the
                    % dimer is bound to two scaffolds, you could
                    % change this to 2

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Kds = DOMAIN Kd SPECIFICATIONS %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Note: each domain with a unique set of binding affinities must be
% individually specified (add another column and row).
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% >> MODIFY THIS (start) << %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%       1       2       3       4       5       6       7       8       9       10
%       A       a       B       b       C       c       D       d       E       e
Kds(1,:)= [ 0.0   kd   0   0   0   0   0   0   0   0 ]; % HK
Kds(2,:)= [ kd   0.0   0   0   0   0   0   0   0   0 ]; % RR

```

```

Kds(3,:)= [ 0 0 0.0 kd 0 0 0 0 0 0 ]; % RRp
Kds(4,:)= [ 0 0 kd 0.0 0 0 0 0 0 0 ]; % SH3
Kds(5,:)= [ 0 0 0 0 0.0 kd 0 0 0 0 ]; % Spew
Kds(6,:)= [ 0 0 0 0 kd 0.0 0 0 0 0 ]; % PDZ
Kds(7,:)= [ 0 0 0 0 0 0 0.0 kd 0 0 ]; % Ppew
Kds(8,:)= [ 0 0 0 0 0 0 kd 0.0 0 0 ]; % D
Kds(9,:)= [ 0 0 0 0 0 0 0 0 0.0 kd ]; % dR
Kds(10,:)= [ 0 0 0 0 0 0 0 0 kd 0.0 ]; % Speww
%%%%%%%%%% >> MODIFY THIS (end) << %%%%%%%%%%
Kfs=(Kds>0)*diffusionLimited;
Krs=Kds*diffusionLimited;

%%%%%%%%%% Reaction Specifications %%%%%%%%%%
% Note: specify the reaction stoichiometry and rates for each reaction
% that isn't automatically calculated based on binding.
%

Rxns=[];
dStoic=[];
Kcat=[];
%%%%%%%%%% >> MODIFY THIS (start) << %%%%%%%%%%
% here's an example I left in from my phosphotransfer simulations, but
% for the MIE figure 3, the only reactions modeled are binding
% reactions, so this section was unnecessary for that
% %(1) Phosphotransfer - noncognate
% %      subs prod
% i=1;
% Rxns(i,1:2)= [ 1 1 ]; % HK
% Rxns(i,2,1:2)= [ 1 0 ]; % RR
% Rxns(i,3,1:2)= [ 0 1 ]; % RRp
% Rxns(i,4,1:2)= [ 0 0 ]; % SH3
% Rxns(i,5,1:2)= [ 0 0 ]; % Spew
% Rxns(i,6,1:2)= [ 0 0 ]; % PDZ
% Rxns(i,7,1:2)= [ 0 0 ]; % Ppew
% Rxns(i,8,1:2)= [ 0 0 ]; % D
% Rxns(i,9,1:2)= [ 0 0 ]; % dR
% Rxns(i,10,1:2)= [ 0 0 ]; % Speww
%
% %      forw rev
% Kcat(i,1:2)= [Khr_kcat 0 ]; %note: should have a reverse for cognate
%
% % change in elemental proteins as a result of the reaction
% dStoic(i,:)=zeros(1,length(key.x));

```



```
% FUNCTIONS
% these can be called by external programs to do simulations
system.deModel=@deModel;
system.stocModel=@stocModel;
system.stocModel_cheat=@stocModel_cheat;
system.simplifyX=@simplifyX;
system.fixX=@fixX;
```

105


```

        && min(pairs.boundDoms(xind,:) == boundDoms),
        xIndex=xind;
        break;
    end
end
end

```

```

function Xfixed = fixX(X, pairs)
% appends zeros to X for non-fundamental elements
% intended specifically to adjust X0
    Xfixed=zeros(1,length(pairs.names));
    Xfixed(1:length(X))=X;
end

```

```

function simpleX = simplifyX(x)
% sorts/condenses and returns concentrations by stoichiometry
    shortNames=system.model.pairs.shortNames;
    simpleX.x=[];
    simpleX.key={};
    simpleX.stoic=[];
    for i=1:length(shortNames)

        if i==1
            simpleX.x(1)=x(1);
            simpleX.key=shortNames(1);
            simpleX.stoic(1,:)=system.model.pairs.stoic(1,:);
        else
            new=1;
            for j=1:length(simpleX.key)
                if strcmp(shortNames(i),simpleX.key(j))
                    new=0;
                    simpleX.x(j)=simpleX.x(j)+x(i);
                end
            end
            if new
                simpleX.x(end+1)=x(i);
                simpleX.key(end+1)=shortNames(i);
                simpleX.stoic(end+1,:)=system.model.pairs.stoic(i,:);
            end
        end
    end
    simpleX.elements=simpleX.x*simpleX.stoic;
end

```

```

function [pairs] = generatePairs(E, Kds, Kfs, Krs, key, PolymerizationLimit)

s=size(E); %s(1) is number of elements, s(2) is number of domains
n=s(1);

names=key.x;
shortNames=names;
stoic=eye(s(1));
freeDoms=E;
boundDoms=zeros(s(1),s(2));
rxn=[];
k=[];

for i=1:s(1)
    Kd(i)={Kds*diag(freeDoms(i,:))};
end

i=1;j=1;
while i<=n %i= element1
    while j<=n %j= element2
        ki=Kd{i}&Kd{j}';
        [inter1,inter2,val]=find(ki);
        for newElm=1:length(inter1)
            if i==j %NOTE: ASSUMES NO SELF-SELF POLYMERIZATION (should be very
                unlikely due to local concentration)
                intramolecular=1;
                testStoic=stoic(i,:);
                testBondDoms=boundDoms(i,:);
                testFreeDoms=freeDoms(i,:);
            else
                intramolecular=0;
                testStoic=stoic(i,:)+stoic(j,:);
                testBondDoms=boundDoms(i,:)+boundDoms(j,:);
                testFreeDoms=freeDoms(i,:)+freeDoms(j,:);
            end
            testBondDoms(inter1(newElm))=testBondDoms(inter1(newElm))+1;
            testBondDoms(inter2(newElm))=testBondDoms(inter2(newElm))+1;
            testFreeDoms(inter1(newElm))=testFreeDoms(inter1(newElm))-1;
            testFreeDoms(inter2(newElm))=testFreeDoms(inter2(newElm))-1;

            %Complex Uniqueness Test - ASSUME UNIQUE RXN PARAMETERS IS ALL THAT
            MATTERS
            uniqueComplex=1;

```

```

u=1;
matchingComplex=n+1;%assuming unique
while max(testBondDoms)<= PolymerizationLimit && uniqueComplex==1 &&
u<=n
    if
min([testStoic==stoic(u,:),testBondDoms==boundDoms(u,:),testFreeDoms==freeDoms(u,:)])
        uniqueComplex=0;
        matchingComplex=u;
    end
    u=u+1;
end

if uniqueComplex && max(testBondDoms)<= PolymerizationLimit
%CREATE A NEW ELEMENT
    if i==j

names(end+1)={['(',names{i},['(',num2str(inter1(newElm)),':',num2str(inter2(new
Elm)),',']]);
        else

names(end+1)={['(',names{i},['(',num2str(inter1(newElm)),':',num2str(inter2(new
Elm)),','],names{j},')']]);
        end
        stoic(end+1,1:s(1))=testStoic;
        freeDoms(end+1,1:s(2))=testFreeDoms;
        boundDoms(end+1,:)=testBondDoms;
        Kd(end+1)={Kds*diag(freeDoms(end,:))};
        n=length(stoic);
        shortNames(n)={' '};

        for sn=1:s(1)
            if stoic(end,sn)
                if stoic(end,sn)>1

shortNames(n)=[shortNames{n},':',key.x{sn},['(',num2str(stoic(end,sn)),')']];
                    else
                        shortNames(n)=[shortNames{n},':',key.x{sn}];
                    end
                    if shortNames{n}(1)=='('
                        shortNames(n)={shortNames{n}(2:end)};
                    end
                end
            end
        end
    end
end

```

```

end

%Calc Rxn, check Uniqueness Test and update
if max(testBondDoms)<= PolymerizationLimit

    %Calculate RXN
    testrxn=zeros(1,n);
    testrxn(i)=-1;
    testrxn(j)=-1;
    testrxn(matchingComplex)=1;

    %check uniqueness
    uniqueRxn=1;
    u=1;
    sR=size(rxn);
    if sR(2)~=n & sR(1)
        rxn(1:end,end+1)=0;
    end

    while max(testBondDoms)<= PolymerizationLimit && uniqueRxn==1 &&
u<=sR(1)
        if min(testrxn==rxn(u,:))
            uniqueRxn=0;
        end
        u=u+1;
    end

    if uniqueRxn
        %UPDATE RXN
        rxn(end+1,:)=testrxn; %forward rxn
        rxn(end+1,:)=-testrxn; %reverse rxn
        k(end+1)=Kfs(inter1(newElm),inter2(newElm));
        kmod(length(k))=intramolecular;
        %if intramolecular, i, j, length(k), end
        k(end+1)=Krs(inter1(newElm),inter2(newElm));
        kmod(length(k))=0; %the reverse direction doesn't need to be modified
    end
end
end
j=j+1;
end
i=i+1;
j=1;
end

```

```

pairs.names=names;
pairs.shortNames=shortNames;
pairs.stoic=stoic;
pairs.freeDoms=freeDoms;
pairs.boundDoms=boundDoms;
pairs.bindingRxn=rxn;
pairs.bindingK=k;
pairs.kmod=kmod;
pairs.PolymerizationLimit=PolymerizationLimit;
end

% RATE CONSTANTS - Append non-linear and time dependant functions to constants
% here.
% Also append enzyme concentrations (i.e. components that are not used up)
% function k_rxn = calc_k_rxn(x)
% %
% k_rxn=k.*(kmod*system.model.effectiveConcentration+~kmod);
%
% % %binding interactions
% % for i=1:length(k)
% % k_rxn(i)=k(i);
% % if kmod(i)
% % k_rxn(i)=k_rxn(i)*system.model.effectiveConcentration;
% % end
% % end
% end

%-----
%-----

% PDEs
% automatically multiplied by the concentrations used in the reaction
% and adjust concentrations of components used and produced
function dXdt = Xp(t,x)
% do not alter this function
concentrationMatrix=system.dxcalc.r*diag(x)+system.dxcalc.nr;
dXdt = (sum(diag(prod(concentrationMatrix,2))*system.dxcalc.rd,1))';
end

function [R] = deModel(X_0, t_max)
tspan = [0 t_max];
[t, X] = ode23s(@Xp, tspan, X_0, odeset('RelTol', 1e-2));
R.t=t;
R.X=X;

```

end

```
function [R] = stocModel(X_0, t_max, maxRunTime, res)
```

```
% Inputs:
```

```
% X_0 - initial concentrations (nM)
```

```
% t_max - the maximum reaction time (not real time) (s)
```

```
% maxRunTime - the maximum real time the script is allowed to run (s)
```

```
% res - the number of reactions between each saved concentration
```

```
% for res=1 all reaction will be output. To reduce the amount of
```

```
% memory used by X and t, increase res.
```

```
% NOTE: this model assumes the conversion from nM to # molecules = 1 for E
```

```
% coli. The model needs to be updated to consider other reaction volumes
```

```
    t0_real = cputime;
```

```
    t_real=0;
```

```
    x=round(X_0)'; % the current # of molecules
```

```
    t=0; % the current time
```

```
    x_(1,:)=x; % a record of the concentrations
```

```
    t_=0; % a record of the time
```

```
    n=1; % reaction #
```

```
    while t<t_max & t_real<maxRunTime
```

```
        k_rxn=calc_k_rxn(x,t);
```

```
        concentrationMatrix=(rxn<0)*diag(x)+(rxn>=0);
```

```
        a = prod(concentrationMatrix,2)'.*k_rxn;
```

```
        a0=sum(a);
```

```
        tao=-log(rand)/a0;
```

```
        mu=1;
```

```
        r=rand;
```

```
        while sum(a(1:mu))<a0*r
```

```
            mu=mu+1;
```

```
        end
```

```
        n=n+1;
```

```
        t=t+tao;
```

```
        x=x+rxn(mu,:)';
```

```
        if mod(n,res)==0, x_(end+1,:)=x(:); t_(end+1)=t; end
```

```
        t_real=cputime-t0_real;
```

```
    end
```

```
    R.X=x_;
```

```
    R.t=t_;
```

```
end
```

```
end
```

8.1.2 – Matlab code for “varyX.m”

```
function info = varyX(setup, X0, xi, dX, doPlot)
% setup should be a setup function that produces system
% X0 is the initial concentration
% xi is the index of the concentration to vary
% dX the numbers to multiply the concentration by
% todo: make time limit an input, then run the system at X0 and adjust

system=setup();
t_max=24*60*60;

progress(1:length(dX))=' ';
progress=[progress, '|'];

for i=1:length(dX)

    c=clock;
    display(['Varying x(',num2str(xi),'). Progress: |',progress,'
            (',num2str(c(4)),',',num2str(c(5)),',',num2str(c(6)),',')']);
    progress(i)='=';

    X_0=X0;
    X_0(xi)=X_0(xi)*dX(i);
    [R] = system.deModel(system.fixX(X_0, system.model.pairs), t_max);
    Xfinal(i,:)=R.X(end,:);
    Xsimple=system.simplifyX(Xfinal(i,:));
    x(i,:)=Xsimple.x;
    xe(i,:)=Xsimple.elements;
    key=Xsimple.key;
end

info.setup=setup;
info.X0=X0;
info.xi=xi;
info.dX=dX;
info.Xfinal=Xfinal;
info.x=x;
info.key=key;
info.xe=xe;

if doPlot==1
    loglog(dX,x);
    legend(key);
```

```

end
if doPlot==2
    loglog(dX,x);
    legend(key);
    ylim([1e-4, 1e4]);
end
if doPlot==3
    plot(dX,x);
    legend(key);
end
if doPlot==4
    loglog(dX,x);
    legend(key);
    figure;
    loglog(dX,x);
    legend(system.model.key.x)
end
if doPlot==5
    loglog(dX,x);
    legend(system.model.key.x)
    note=['xi=',num2str(xi)];
    title([note,' K=',num2str(X0(1)),' Ro=',num2str(X0(2)),' Phos=',num2str(X0(4))])
    axis([min(dX) max(dX) 1 10^6]);
end
end

```

8.1.3 – Matlab code for “run.m”

```

setup=@setup_MIEfig3;
system=setup();

xi=1;
dX=[ 0 .05 .1 .25 .5 .7 .8 .85 .9 .95 1 1.05 1.1 1.2 1.4 1.6 1.8 2 2.5 3 4 5 ];
xKd=100;
x=dX*100*xKd;

display('n=1');
figure;
X0=[100, 100, 0, 0, 0, 0]*xKd;
doPlot=0;
vS1 = varyX(setup, X0, xi, dX, doPlot);
plot(x,vS1.Xfinal(:,7));

display('n=2');
hold on;

```

```

X0=[100, 100, 100, 0, 0, 0]*xKd;
doPlot=0;
vS2 = varyX(setup, X0, xi, dX, doPlot);
plot(x,vS2.Xfinal(:,12));

display('n=3');
X0=[100, 100, 100, 100, 0, 0]*xKd;
doPlot=0;
vS3 = varyX(setup, X0, xi, dX, doPlot);
plot(x,vS3.Xfinal(:,19));

display('n=4');
X0=[100, 100, 100, 100, 100, 0]*xKd;
doPlot=0;
vS4 = varyX(setup, X0, xi, dX, doPlot);
plot(x,vS4.Xfinal(:,28));

display('n=5');
X0=[100, 100, 100, 100, 100, 100]*xKd;
doPlot=0;
vS5 = varyX(setup, X0, xi, dX, doPlot);
plot(x,vS5.Xfinal(:,37));
legend(vS5.key{7},vS5.key{12},vS5.key{19},vS5.key{28},vS5.key{37})
axis([min(x) max(x) 1 xKd*100]);

```

8.2 – Example plasmid sequences

Four plasmid sequences are given below that include the kinase plasmid, pWW2179, the RR expression and reporter plasmid, pWW2021, the scaffold expression plasmid, pWW2181, and an autoinhibited kinase expression plasmid, pWW2254. For illustration see Figure S4.1, and for summary of all plasmids see Table S4.4. The sequences are in GenBank format and were created using the plasmid editor software: ApE - A plasmid Editor v2.0.38 (Copyright 2003-2009 M.Wayne Davis).

8.2.1 – pWW2179 : kinase plasmid: P_{con} Taz-(SH3)₄

```

LOCUS   pWW2179           4973 bp ds-DNA   circular   10-FEB-2012
DEFINITION
ACCESSION
VERSION
SOURCE
  ORGANISM
COMMENT
COMMENT

```

COMMENT ApEinfo:methylated:1
FEATURES Location/Qualifiers
misc_feature 1604..1777
 /label=SH3
 /ApEinfo_fwdcolor=#800080
 /ApEinfo_revcolor=#800080
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
misc_feature 1814..1987
 /label=SH3(1)
 /ApEinfo_label=SH3
 /ApEinfo_fwdcolor=#800080
 /ApEinfo_revcolor=#800080
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
misc_feature 2024..2197
 /label=SH3(2)
 /ApEinfo_label=SH3
 /ApEinfo_fwdcolor=#800080
 /ApEinfo_revcolor=#800080
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
misc_feature 2234..2407
 /label=SH3(3)
 /ApEinfo_label=SH3
 /ApEinfo_fwdcolor=#800080
 /ApEinfo_revcolor=#800080
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
misc_feature 13..41
 /label=P_Con
 /ApEinfo_fwdcolor=#008000
 /ApEinfo_revcolor=#008000
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
misc_feature 1538..1603
 /label=GS linker
 /ApEinfo_fwdcolor=#c0c0c0
 /ApEinfo_revcolor=#c0c0c0
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
misc_feature 1778..1813
 /label=GS linker(1)
 /ApEinfo_label=GS linker

```

    /ApEinfo_fwdcolor=#c0c0c0
    /ApEinfo_revcolor=#c0c0c0
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 1988..2023
    /label=GS linker(2)
    /ApEinfo_label=GS linker
    /ApEinfo_fwdcolor=#c0c0c0
    /ApEinfo_revcolor=#c0c0c0
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 2198..2233
    /label=GS linker(3)
    /ApEinfo_label=GS linker
    /ApEinfo_fwdcolor=#c0c0c0
    /ApEinfo_revcolor=#c0c0c0
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 2429..2796
    /label=Terminator (rrnB)
    /ApEinfo_fwdcolor=#800000
    /ApEinfo_revcolor=#800000
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 3987..4827
    /label=p15A ORI
    /ApEinfo_fwdcolor=#3f3f3f
    /ApEinfo_revcolor=#3f3f3f
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 2918..3574
    /label=CamR
    /ApEinfo_fwdcolor=#3f3f3f
    /ApEinfo_revcolor=#3f3f3f
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 86..1537
    /label=Taz
    /ApEinfo_fwdcolor=#d82727
    /ApEinfo_revcolor=#d82727
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
ORIGIN
1 gaattcacta gtttgacagt tagctcagtc ctaggAATTA TGCTAGCGGT ACCTCCCGGT

```

61 GCCCTCTACC AGGGCCAAGA gatctatgat taaccgtatc cgcgtagtca cgctgttgg
 121 aatggtgctg ggggtattcg cactgttaca gcttatttcc ggcagtctgt tttttcttc
 181 ccttcacat agccagaaga gtttgtgggt ttccaatcaa ttacgggaac agcagggcga
 241 gctgacgtca acctgggatt taatgctgca aacgcgcatt aacctgagtc gttcagcgg
 301 acggatgatg atggattcct ccaatcaaca aagtaacgcc aaagttgaat tgctcgatag
 361 cgccaggaaa acattggcgc aggcagcgac gcattataaa aaattcaaaa gcatggcacc
 421 gttacctgaa atggtcgcta ccagtcgtaa tattgatgaa aaatataaaa actattacac
 481 agcgtaact gaactgattg attatctaga ttatggcaat actggagctt atttcgctca
 541 gccaacccag ggaatgcaaa atgcaatggg cgaagcggtt gtcagtagc ccctcagcag
 601 tgaaaaactg tatcgcgata tcgtcactga caacgcagat gattaccgat ttgccagtg
 661 gcaactggcg gttatcgcg tggtgggtgg attgattctg ctgggtggcg ggtacggcat
 721 tcgccgatg ttgcttact cgctggcaaa aattattgct cacattcgcg aaatcgccg
 781 tggtaacctg gcgaataccc tgaccattga cgggcgcagt gaaatgggag acctggcgca
 841 gagcgtttca catatggcgg ctggtgttaa gcaactggcg gatgaccgca cgctgctgat
 901 ggcgggggta agtcacgact tgcgcacgcc gctgacgcgt attcgccttg cgactgagat
 961 gatgagcgag caggatggct atctggcaga atcgatcaat aaagatatcg aagagtgcaa
 1021 cgccatcatt gagcagtta tcgactacct gcgcaccggg caggagatgc cgatggaaat
 1081 ggcggaatct aatgcagtac tcggtgaggt gattgctgcc gaaagtggct atgagcggga
 1141 aattgaaacc gcgctttacc ccggcagcat tgaagtgaat atgcaccgcg tctcgatcaa
 1201 acgcgcggtg gcgaatatgg tggtaacgc cgcccggtt ggcaatggct ggaatcaaag
 1261 cagcagcgga acggagccga atcgcgctg gttccaggtg gaagatgacg gtccgggaat
 1321 tgcgccgga caacgtaagc acctgttcca gccgttgtc cgcgcgaca gtgcgcgac
 1381 cattagcggc acgggattag ggctggcaat tgtgcagcgt atcgtaggata accataacgg
 1441 gatgctggag ctggcacca gcgagcgggg cgggctttcc attcgcgctt ggctgccagt
 1501 gccgtaacg cgggcgcagg gcacgacaaa agaaggggga tctggtagtg gcagcggcag
 1561 cggtagtgga tctggtagtg gcagcggcag cggtagtgga tctatggcag agtatgtgcg
 1621 ggccctgtt gactttaatg ggaatgatga agaagacctt cccttaaga aaggagacat
 1681 cctgagaatc cgggataagc ctgaagagca gtggtggaat gcagaggaca gcgaaggaaa
 1741 gagggggatg attcctgtcc ctacgtgga gaagtatgga tctggtagtg gcagcggcag
 1801 cggtagtgga tctatggcag agtatgtgcg ggccctgtt gactttaatg ggaatgatga
 1861 agaagacctt cccttaaga aaggagacat cctgagaatc cgggataagc ctgaagagca
 1921 gtggtggaat gcagaggaca gcgaaggaaa gagggggatg attcctgtcc ctacgtgga
 1981 gaagtatgga tctggtagtg gcagcggcag cggtagtgga tctatggcag agtatgtgcg
 2041 ggccctgtt gactttaatg ggaatgatga agaagacctt cccttaaga aaggagacat
 2101 cctgagaatc cgggataagc ctgaagagca gtggtggaat gcagaggaca gcgaaggaaa
 2161 gagggggatg attcctgtcc ctacgtgga gaagtatgga tctggtagtg gcagcggcag
 2221 cggtagtgga tctatggcag agtatgtgcg ggccctgtt gactttaatg ggaatgatga
 2281 agaagacctt cccttaaga aaggagacat cctgagaatc cgggataagc ctgaagagca
 2341 gtggtggaat gcagaggaca gcgaaggaaa gagggggatg attcctgtcc ctacgtgga
 2401 gaagtatgga tctgactcg agtaatgaga agcttgggcc cgaacaaaaa ctatctcag
 2461 aagaggatct gaatagcgcc gtcgaccatc atcatcatca tcattgagtt taaacggtct
 2521 ccagcttggc tgtttggcg gatgagagaa gatttcagc ctgatacaga ttaaatcaga
 2581 acgcagaagc ggtctgataa aacagaattt gcttggcggc agtagcgcgg tgggtccacc
 2641 tgaccccatg ccgaactcag aagtgaacg ccgtagcgcc gatggtagtg tgggtctcc

2701 ccatgcgaga gtagggaact gccaggcatc aaataaaacg aaaggctcag tcgaaagact
 2761 gggcctttcg tttatctgt tgtttgtcgg tgaactaatt tctagactgc agttgatcgg
 2821 gcacgtaaga ggttccaact ttcaccataa tgaaataaga tctactaccgg gcgtattttt
 2881 tgagttatcg agattttcag gagctaagga agctaaaatg gagaaaaaaa tctactggata
 2941 taccaccgtt gatatatccc aatggcatcg taaagaacat ttgaggcat ttcagtcagt
 3001 tgctcaatgt acctataacc agaccgttca gctggatatt acggcctttt taaagaccgt
 3061 aaagaaaaat aagcacaagt tttatccggc cttattcac attcttggcc gcctgatgaa
 3121 tgctcatccg gaatttcgta tggcaatgaa agacgggtgag ctgggtgat gggatagtg
 3181 tcacccttgt tacaccgtt tccatgagca aactgaaacg tttcatcgc tctggagtga
 3241 ataccacgac gatttccggc agtttctaca catatattcg caagatgtgg cgtgttacgg
 3301 tgaaaacctg gcctatttcc ctaaaagggtt tattgagaat atgtttttcg tctcagccaa
 3361 tccctgggtg agtttcacca gttttgattt aaacgtggcc aatatggaca acttcttcgc
 3421 ccccgttttc accatgggca aatattatac gcaaggcgac aaggtgctga tgccgctggc
 3481 gattcagggt catcatgccg ttgtgatgg cttcatgtc ggcagaatgc ttaatgaatt
 3541 acaacagtac tgcgatgagt ggcaggcgcg ggcgtaatt gatatcgagc tcgcttggac
 3601 tcctgttgat agatccagta atgacctcag aactccatct ggatttgttc agaacgctcg
 3661 gttgccgccc ggcgtttttt attggtgaga atccaagcct cggtgagaat ccaagcctcg
 3721 atcaacgtct cattttcgcc aaaagttggc ccagggttc ccggtatcaa caggacacc
 3781 aggatttatt tattctgca agtgatctc cgtcacaggt attattcgg cgaaagtgc
 3841 gtcgggtgat gctgccaact tactgattta gtgtatgat gtgttttga ggtgctccag
 3901 tggcttctgt ttctatcagc tgtcctctc gttcagctac tgacggggtg gtgcgtaacg
 3961 gcaaaagcac cgccggacat cagcggtagc ggagtgtata ctggcttact atgttggcac
 4021 tgatgagggg gtcagtgaag tgcttcatgt ggcaggagaa aaaaggctgc accggtgcgt
 4081 cagcagaata tgtgatacag gatatttcc gttcctcgc tctactgact gctacgctcg
 4141 gtcgttcgac tgcggcgagc ggaaatggct tacgaacggg gcggagattt cctggaagat
 4201 gccaggaaga tacttaacag ggaagtgaga gggccgcggc aaagccgttt ttcataggg
 4261 tccgcccccc tgacaagcat caggaatct gacgtcaaa tcagtgggtg cgaaaccga
 4321 caggactata aagataccag gcgtttcccc ctggcggctc cctcgtgctc tctctgttc
 4381 ctgcctttcg gttaccgggt gtcattccgc tgttatggcc gcgttgtct cattccacgc
 4441 ctgacactca gttccgggta ggcagttcgc tccaagctgg actgtatgca cgaaccccc
 4501 gttcagtcg accgctgctc cttatccggt aactatcgtc ttgagtccaa cccggaaaga
 4561 catgcaaaag caccactggc agcagccact ggtaattgat ttagaggagt tagtctgaa
 4621 gtcatgcgcc ggtaaggct aaactgaaag gacaagttt ggtgactgct ctctccaag
 4681 ccagttacct cgttcaaag agttggtagc tcagagaacc ttcgaaaaac cgccctgaa
 4741 ggcgggtttt tcgttttcag agcaagagat tacgcgcaga ccaaacgat ctcaagaaga
 4801 tcatcttatt aatcagataa aatatttcta gatttcagt caatttatct cttcaaatgt
 4861 agcacctgaa gtcagcccca tacgatataa gttgtaattc tcatgttga cagcttatca
 4921 tcgataagct tccgatggcg cgccgagagg cttacactt tatgcttccg gct

8.2.2 – pWW2021: reporter plasmid: P_{BAD} CusR-LZa, P_{CusR} RFP, P_{Rham} CpxR-LZb, P_{CpxR} RFP

LOCUS pWW2021_labeled 11539 bp ds-DNA circular 10-FEB-2012
 DEFINITION
 ACCESSION

```

VERSION
SOURCE
ORGANISM
COMMENT
COMMENT    ApEinfo:methylated:1
FEATURES    Location/Qualifiers
    CDS      complement(6328..7119)
              /label=Kan/neoR
              /ApEinfo_fwdcolor=yellow
              /ApEinfo_revcolor=yellow
              /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
              width 5 offset 0
    misc_feature  4910..5000
              /label=P_CusR
              /ApEinfo_fwdcolor=#37c8a0
              /ApEinfo_revcolor=#37c8a0
              /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
              width 5 offset 0
    misc_feature  4769..4897
              /label=Terminator (dblTerm)
              /ApEinfo_fwdcolor=#be0c10
              /ApEinfo_revcolor=#be0c10
              /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
              width 5 offset 0
    misc_feature  complement(1453..1581)
              /label=Terminator (dblTerm)(1)
              /ApEinfo_label=Terminator (dblTerm)
              /ApEinfo_fwdcolor=#be0c10
              /ApEinfo_revcolor=#be0c10
              /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
              width 5 offset 0
    misc_feature  3944..4624
              /label=CusR
              /ApEinfo_fwdcolor=#ff8040
              /ApEinfo_revcolor=#ff8040
              /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
              width 5 offset 0
    misc_feature  9372..10034
              /label=pBBR1 ORI
              /ApEinfo_fwdcolor=#de9b69
              /ApEinfo_revcolor=#de9b69
              /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
              width 5 offset 0
    misc_feature  10814..10994

```

```

        /label=T1 terminator from pPROBE
        /ApEinfo_fwdcolor=#c2a0a6
        /ApEinfo_revcolor=#c2a0a6
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 10995..11175
        /label=T1 terminator from pPROBE(1)
        /ApEinfo_label=T1 terminator from pPROBE
        /ApEinfo_fwdcolor=#c2a0a6
        /ApEinfo_revcolor=#c2a0a6
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 11176..11356
        /label=T1 terminator from pPROBE(2)
        /ApEinfo_label=T1 terminator from pPROBE
        /ApEinfo_fwdcolor=#c2a0a6
        /ApEinfo_revcolor=#c2a0a6
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 11357..11537
        /label=T1 terminator from pPROBE(3)
        /ApEinfo_label=T1 terminator from pPROBE
        /ApEinfo_fwdcolor=#c2a0a6
        /ApEinfo_revcolor=#c2a0a6
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature complement(44..411)
        /label=Terminator (rrnB)
        /ApEinfo_fwdcolor=#9191ff
        /ApEinfo_revcolor=#9191ff
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 1190..1383
        /label=P_CpxR
        /ApEinfo_fwdcolor=#00ff80
        /ApEinfo_revcolor=#00ff80
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature complement(415..1131)
        /label=GFP
        /ApEinfo_fwdcolor=#00ff00
        /ApEinfo_revcolor=green
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0

```

misc_feature 2721..3599
 /label=AraC
 /ApEinfo_fwdcolor=#808000
 /ApEinfo_revcolor=#808000
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature 3627..3864
 /label=P_BAD
 /ApEinfo_fwdcolor=#008040
 /ApEinfo_revcolor=#008040
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature complement(1762..2454)
 /label=CpxR
 /ApEinfo_fwdcolor=#043c11
 /ApEinfo_revcolor=#043c11
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature complement(2504..2670)
 /label=P_Rham
 /ApEinfo_fwdcolor=#008000
 /ApEinfo_revcolor=#008000
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature 5721..5750
 /label=Terminator (t500)
 /ApEinfo_fwdcolor=#800000
 /ApEinfo_revcolor=#800000
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature 4652..4750
 /label=LZa
 /ApEinfo_fwdcolor=#f5800a
 /ApEinfo_revcolor=#f5800a
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature 5043..5714
 /label=mRFP1
 /ApEinfo_fwdcolor=#cd0532
 /ApEinfo_revcolor=#cd0532
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature complement(1597..1725)
 /label=LZb

```

/ApEinfo_fwdcolor=#8000ff
/ApEinfo_revcolor=#8000ff
/ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
width 5 offset 0
misc_feature 4625..4651
/label=GS linker
/ApEinfo_fwdcolor=#c0c0c0
/ApEinfo_revcolor=#c0c0c0
/ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
width 5 offset 0
misc_feature complement(1726..1761)
/label=GS linker(1)
/ApEinfo_label=GS linker
/ApEinfo_fwdcolor=#c0c0c0
/ApEinfo_revcolor=#c0c0c0
/ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
width 5 offset 0

```

ORIGIN

```

1 ggatcggaag ctgaattcgg gaacaaaagc tggctagaa attagttcac cgacaaacaa
61 cagataaaac gaaaggccca gtcttcgac tgagccttc gttttatttg atgcctggca
121 gttccctact ctgcatggg gagacccac actaccatcg gcgctacggc gtttcacttc
181 tgagttcggc atggggtcag gtgggaccac cgcgctactg ccgccaggca aattctgttt
241 tatcagaccg cttctgcgtt ctgatttaat ctgtatcagg ctgaaaatct tctctcatcc
301 gccaaaacag ccaagctgga gaccgttaa actcaatgat gatgatgatg atggctgcagc
361 gcgctattca gatcctcttc tgagatgagt tttgttcgg gcccaagctt ctactattt
421 gtatagtcca tccatgccat gtgtaatccc ggcagctgtt acaaactcaa gaaggacat
481 gtggtctctc tttcgttg gtatcttcga aagggcagat tgtgtggaca ggtaatggtt
541 gtctggtaaa aggacagggc catcgccaat tggagtattt tgttgataat ggtctgctag
601 ttgaacgctt ccatcttcaa tgtgtgtct aatttgaag ttaacttga ttccattctt
661 ttgtttgtct gccatgatgt atacattgtg tgagttatag ttgtattcca acttgtgtcc
721 aagaatgttt ccatcttctt taaaatcaat acctttaac tcgattctat taacaagtgt
781 atcaccttca aacttgactt cagcacgtgt cttatagttc ccgtcatctt tgaaaaatat
841 agttctttcc tgtacataac ctcggggcat ggcactcttg aaaaagtcac gccgtttcat
901 atgatctggg tatctgaaa agcattgaac accataagtc aaagtagtga caagtgttg
961 ccatggaaca gtagtcttc cagtagtgca aataaattta agggtaagtt ttccgtatgt
1021 tgcacacct tcaccctctc cactgacaga aaatttgtgc ccattaacat caccatctaa
1081 ttcaacaaga attgggacaa ctccagtga aagttctct ctttactca tctaacttc
1141 ctgcgcccg tttctacgaa tgatgagct ctcagcagc gtggcttaat gaactgactg
1201 ccagcgttga ggccatgacg gcagcgtaa ctatgcgcat cattgtctcc caaaatcttt
1261 ctgtcgcat tcaacgatag agagtttacg attcaggctg caaacatgcg tcagggggtg
1321 taaaacaacg taaagtcag gattagcgac gtctgatgac gtaatttctg cctcggaggt
1381 atttaacaa tgaataaaat cctgttagt gatgatgacc gagagctgac ttccctatta
1441 cctaggccgc ggtataaacg cagaaaggcc caccgaagg tgagccagtg tgactctagt
1501 agagagcgtt caccgacaaa caacagataa aacgaaaggc ccagtcttc gactgagcct

```

1561 ttcgttttat ttgatgcctg gctcgagtta ggatccaagc gggccgtatc gggtttcgta
1621 tttggataca atgttgccgc accggcccac acgtttaccg agttcggccg cgcggttacg
1681 cagtgcggtg tttcttttt caagaaacgc tgcgcggatt tccagAGATC cactaccgct
1741 gccgctgcca ctaccagatc ctgaagcaga aaccatcaga tagccgcgac cacgcaaggt
1801 tttaaaccac gggtgaccat ctttacgatc cggcagttta cgacgcaggt tggaatgtg
1861 catatcaata gcgcggtcga aaggcgtcag gcgtttgccc aacacttcct ggcttaaatg
1921 ttcacgggaa accacctgac ccagatgctg tgccagcaaa tagagcaggg taaactcagt
1981 accggttaac tccagcggtt gcccgctgaa gctggcttcc tgacggcctg gattcagcac
2041 taaggcatca acttcagtg tcggtgaacc gttgtcgttg ttttgctgtt gctcgtcca
2101 gtgcgaacgg cgcaggatcg cgcgaatacg tgccaccagc tcacgatcat taaacggtt
2161 cgggagatag tcatctgctc ccagctcaag gccgagaacg cgatcaagtt cactgccgcg
2221 cgccgtcaac ataatgacag gcgtctggtg tgtctggcga agtgcttta atgtgtcgat
2281 accatttttc ttcggcatca ttacgtcaag caaaagtaaa tcaatgctgt cgtccagaag
2341 atcaagcgcc tgttcccat cgtgggcaac aatcacgttg aagccttcca tctcaagcag
2401 ctcctttaat agggaagtca gctctcgctc atcatcaact aacaggattt tattcatAGA
2461 TCTCACTAGA CAATATCACG GCAGTATTG CGGTACCGCT AGCatcctgc
tgaatttcat
2521 tacgaccagt ctaaaaagcg cctgtattcg cgaccttc gttactgaca ggaaaatggg
2581 ccattggcaa ccagggaag atgaacgtga tgatgttac aatttgctga attgtggtga
2641 tgtgatgctc accgcaTTTC CTGAAAATTC ACTAGTCTAT GCTACTCCAT CGAGccgtca
2701 attgtctgat tcgttaccaa ttatgacaac ttgacggcta catcattcac ttttcttca
2761 caaccggcac ggaactcgct cgggctggcc ccggtgcatt ttttaatac ccgcgagaaa
2821 tagagttgat cgtcaaaacc aacattgcga ccgacgggtg cgataggcat ccgggtggtg
2881 ctcaaaagca gttcgcctg gctgatacgt tggctctcgc gccagcttaa gacgctaact
2941 cctaactgct ggcggaaaag atgtgacaga cgcgacggcg acaagcaaac atgtgtgctg
3001 acgctggcga tatcaaaatt gctgtctgcc aggtgatcgc tgatgtactg acaagcctcg
3061 cgtacccgat tatccatcgg tggatggagc gactcgtaa tcgcttccat gcgccgcagt
3121 aacaattgct caagcagatt tatcgccagc agtccgaat agcgcccttc cccttgccc
3181 gcgttaatga ttgcccaca caggctcgtg aaatgcggct ggtgcgcttc atccgggcca
3241 aagaaccccg tattggcaa tattgacggc cagttaagcc attcatgcca gtaggcgcgc
3301 ggacgaaagt aaaccactg gtgataccat tcgcgagcct ccgcatgacg accgtagtga
3361 tgaatctctc ctggcgggaa cagcaaaata tcaccgggc ggcaaaaaa ttctgctcc
3421 tgatttttca ccacccctg accgcgaatg gtgagattga gaataaacc tttcattccc
3481 agcggtcggt cgataaaaaa atcgagataa ccgttggcct caatcggcgt taaaccggc
3541 accagatggg cattaaacga gtatcccggc agcaggggat cttttgcgc ttacgccata
3601 ctttcatac tcccgcatt cagagaagaa accaattgtc catattgcat cagacattgc
3661 cgtactgctg tctttactg gctctctcg ctaaccaa acgtaacccc gcttataaa
3721 agcattctgt acaaaagcg gaccaagcc atgacaaaaa cgcgtaaaa aagtgtctat
3781 aatcacggca gaaaagtcca cattgattat ttgcacggcg tcacacttg ctatgccata
3841 gcatttttat ccataagatt agcggatctt acctgacgt ttttatcgca actcttact
3901 gtttctcat acccgctagc ggtacctcac gcgatacaga tctatgaaac tgttgattgt
3961 cgaagatgaa aagaaaaccg gagaatactt gaccaaaggg ttaaccgaag ccggtttgt
4021 ggtcgatttg gccgacaacg ggctgaatgg ctaccatctg cggatgaccg gtgattatga
4081 tctgataatc ctcgatatta tgctgccgga cgtgaacggc tgggatatcg tgcgcatgtt

4141 acgctccgcc aataaaggga tgccgattct gttgcttacc gcgcttggca ccattgaaca
 4201 tcgctgaag gggctggagt tgggggcaga tgactacctg gtgaagccat tcgctttgc
 4261 tgaactgctg gcgcgggtgc gcacattact gcggcgcggg gcggcgggtga ttatcgaaag
 4321 tcagtttcag gttgccgatt tgatgggtcga tctcgtcagc cgcaaagtca cccgcagcgg
 4381 cacgcgcac acttgacca gtaaagagtt tactctgctg gagttcttcc ttcgccatca
 4441 gggcgaagtg ctgccccgct cgcttatcgc ctcgcaggta tgggacatga atttgacag
 4501 cgataccaat gctattgatg tggcgggtgaa gcggctgcgc ggcaaaatcg acaacgactt
 4561 tgagccgaag ctaattcaga ccgtgcgcgg cggtgggttac atgcttgagg tgccggatgg
 4621 tcagggatct ggggtccggca gcggtggatc tagcgtaaa gaactggaag ataaaaacga
 4681 agagctcttg agcgaacttt atcatttgaa aaacgaagtg gcgcgtctga aaaagcttgt
 4741 tggcgaacgt taaggatcct gactcgagcc aggcatacaa taaaacgaaa ggctcagtcg
 4801 aaagactggg ctttcgttt tatctgttgt ttgtcgggtga acgctctcta ctgagtcac
 4861 actggctcac ctgcgggtgg gcctttctgc gtttataccg cggcctagga ttggcaaaat
 4921 gacaattttg tcatttttct gtcaccggaa aatcagagcc tggcgagtaa agttggcggc
 4981 ataaatcac cagaaattat gagctgcgc ttacgttagc gagtattaag gaggtacca
 5041 tggcttctc cgaagacgtt atcaaagagt tcatgcgttt caaagttcgt atggaagggt
 5101 ccgttaacgg tcacgagttc gaaatcgaag gtgaagggtga aggtcgtccg tacgaaggta
 5161 cccagaccgc taaactgaaa gttaccaaag gtggtccgct gccgttcgct tgggacatcc
 5221 tgtccccgca gttccagtac ggttcaaag ctacgttaa acaccggct gacatcccg
 5281 actacctgaa actgtcttc ccggaagggt tcaaattgga acgtgttatg aacttcgaag
 5341 acggtggtgt tgttaccgtt acccaggact cctccctgca agacggtgag ttcattaca
 5401 aagtaaaact gcgtgttacc aactcccgt ccgacggtcc ggttatgcag aaaaaacca
 5461 tgggttggga agcttcacc gaacgtatgt acccgaaga cgggtgctcg aaaggtgaaa
 5521 tcaaatgcg tctgaaactg aaagacggtg gtcactacga cgctgaagtt aaaaccacct
 5581 acatggctaa aaaaccggtt cagctgccg gtgcttaca aaccgacatc aaactggaca
 5641 tcacctcca caacgaagac tacaccatcg ttgaacagta cgaacgtgct gaaggtcgtc
 5701 actccaccgg tgcttaataa caaagccgc cgaaaggcgg gcttttctgt tctagaccag
 5761 cttttgttcc cttagtgag ggtaattgc gcgcttgcg taatcatggt catagctgtt
 5821 tcctgtgtga aattgttatc cgctcacaat tccacacaac atacgagccg gaagcataaa
 5881 gtgtaaagcc tgggtgcct aatgagtgag ctaactaca ttaattgcgt tgcgtcact
 5941 gcccgttct cagtcgggaa acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc
 6001 ggggagaggc ggtttgcgt tgggcgcat gcataaaaac tgttgaatt cattaagcat
 6061 tctgccgaca tgaagccat cacaacggc atgatgaacc tgaatcgcca gcggcatcag
 6121 cacctgtcgc cttgcgtat aatatttgc catgggggtg ggcgaagaac tccagcatga
 6181 gatccccgc ctggaggatc atccagccgg cgtccggaa aacgattccg aagccaacc
 6241 tttcatagaa ggcggcgggt gaatcgaaat ctcgtgatgg caggtgggc gtcgcttgg
 6301 cgtcatttc gaacccaga gtcccgtca gaagaactcg tcaagaaggc gatagaaggc
 6361 gatgcgctgc gaatcgggag cggcgatacc gtaaagcacg aggaagcggg cagccattc
 6421 gccgcaagc tcttcagcaa tatcagggt agccaacgt atgtctgat agcgggtccgc
 6481 cacaccagc cgccacagt cgatgaatcc agaaaagcgg ccattttcca ccatgatatt
 6541 cggcaagcag gcatcgccat gggtcacgac gagatcctcg ccgtcgggca tgcgcgcctt
 6601 gagcctggcg aacagttcgg ctggcgcgag ccctgatgc tcttcgtcca gatcatcctg
 6661 atcgacaaga ccggttcca tccgagtacg tgctcgtcg atgcgatgtt tcgcttgggt
 6721 gtcgaatggg caggtagccg gatcaagcgt atgcagccgc cgattgcat cagccatgat

6781 ggatactttc tcggcaggag caaggtgaga tgacaggaga tctgccccg gcacttcgcc
6841 caatagcagc cagtccttc ccgcttcagt gacaacgtcg agcacagctg cgcaaggaac
6901 gcccgtcgtg gccagccacg atagccgcgc tgcctcgtcc tgcagttcat tcagggcacc
6961 ggacaggctg gtcttgacaa aaagaaccgg gcgcccctgc gctgacagcc ggaacacggc
7021 ggcacagag cagccgattg tctgttgctc ccagtcatag ccgaatagcc tctccacca
7081 agcggccgga gaacctgctg gcaatccatc ttgttcaatc atgcgaaacg atcctcatcc
7141 tgtctcttga tcagattttg atcccctgcg ccatcagatc cttggcggca agaaagccat
7201 ccagtttact ttgcagggtt tccaacctt accagagggc gcccagctg gcaattccgg
7261 ttcgcttctg gtccataaaa ccgcccagtc tagctatcgc catgtaagcc cactgcaagc
7321 tacctgtttt ctctttgcgc ttgcgtttc cttgtccag atagcccagt agctgacatt
7381 catcccaggt ggcaactttc ggggaaatgt gcgcgcccgc gttcctgctg gcgctgggcc
7441 tgtttctggc gctggacttc ccgctgttcc gtcagcagct ttccgcccac ggccttgatg
7501 atcgcgccgg ccttgccctg catatcccga ttcaacggcc ccagggcgctc cagaacgggc
7561 ttcaggcgct cccgaaggct tcggggccgtc tcttgggctt gatcgccctt cttgcgcac
7621 tcacgcgctc ctgcggcgcc ctgtagggca ggctcatacc cctgccgaac cgctttgtc
7681 agccggtcgg ccacggcttc cggcgtctca acgcgcttg agattcccag ctttcggcc
7741 aatccctgcg gtgcataggc gcgtggctcg accgcttgcg ggctgatggt gacgtggccc
7801 actggtggcc gctccagggc ctgtagaac gcctgaatgc gcgtgtgacg tgcttgctg
7861 cctcgatgc cccgttcgag cctagatcg gccacagcgg ccgcaaactg ggtctggtcg
7921 cgggtcatct gcgcttggc gccgatgaac tcttggccg acagcctgcc gtctgcgct
7981 agcggcacca cgaacgcgtt catgtgcggg ctggtttcgt cacggtggat gctggccgtc
8041 acgatgcgat ccgccccgta cttgtccgcc agccactgt gcgcttctc gaagaacgc
8101 gcctgctgtt cttggctggc cgactccac cattccggc tggccgtcat gacgtactcg
8161 accgccaaca cagcgtcctt gcgccgttc tctggcagca actcgcgcag tcggcccatc
8221 gcttcatcgg tgctgctggc cgcccagtc tcgttctctg gcgtcctgct ggcgtcagcg
8281 ttgggctgtc cgcgctcgcg gtaggcgtgc ttgagactgg ccgccacgtt gccattttc
8341 gccagcttct tgcacgcat gatcgctat gccgcatgc ctgccctcc ctttggtgt
8401 ccaaccggct cgacgggggc agcgcaaggc ggtgcctccg gcgggccact caatgcttga
8461 gtatactcac tagactttgc ttcgaaagt cgtgaccgcc tacggcggct gcggcgccct
8521 acgggcttgc tctccgggct tcgccctgcg cggtcgtgc gctccttgc cagcccggtg
8581 atatgtggac gatggccgcg agcggccacc ggctggctcg cttcgtcgg cccgtggaca
8641 accctgctgg acaagctgat ggacaggctg cgctgcca cgagcttgac cacagggatt
8701 gccaccggc taccagcct tcgaccat acccaccggc tcaactgcg cggcctgcg
8761 cttgcccc tcaattttt taattttctc tggggaaaag cctccggcct gcggcctgcg
8821 cgcttcgtt gccggttga caccaagtgg aaggcgggtc aaggctcgcg cagcgaccgc
8881 gcagcggctt ggccttgacg cgcttgaac gaccaagcc tatgcgagt ggggcagctg
8941 aaggcgaag ccgcccgcct gcccccgag cctcacggcg gcgagtgcg gggttccaag
9001 ggggcagcgc caccttggc aaggccgaag gccgcgcagt cgatcaacaa gccccggagg
9061 ggccactttt tgccggaggg ggagccgcgc cgaaggcgtg ggggaacccc gcaggggtgc
9121 ctttcttgg gcaccaaaga actagatata gggcgaaat cgaaagactt aaaaatcaac
9181 aacttaaaaa aggggggtac gcaacagctc attgcggcac ccccgcaat agctcattgc
9241 gtaggttaa gaaaatctgt aattgactgc cacttttac caacgataa ttgtgtcgc
9301 gctgccgaaa agttgcagct gattgcgcat ggtgccgcaa ccgtgcggca cctaccgca
9361 tggagataag catggccacg cagtccagag aaatcgcat tcaagccaag aacaagccc

9421 gtcactgggt gcaaacggaa cgaaaagcgc atgaggcgtg ggccgggctt attgcgagga
 9481 aaccacggc ggcaatgctg ctgcatcacc tcgtggcgca gatgggccac cagaacgccg
 9541 tgggtgtcag ccagaagaca cttccaagc tcacggacg ttcttgcg acggtccaat
 9601 acgcagtcaa ggacttggtg gccgagcgt ggatctccgt cgtgaagctc aacggccccg
 9661 gcaccgtgtc ggctacgtg gtcaatgacc gcgtggcgtg gggccagccc cgcgaccagt
 9721 tgcgcctgtc ggtgttcagt gccgccgtg tggttgatca cgacgaccag gacgaatcg
 9781 tgttggggca tggcgacctg cgccgcatcc cgaccctgta tccgggagag cagcaactac
 9841 cgaccggccc cggcgaggag ccgccagcc agcccggcat tccgggcatg gaaccagacc
 9901 tgccagcctt gaccgaaacg gaggaatggg aacggcgcg gcagcagcgc ctgccgatgc
 9961 ccgatgagcc gtgtttctg gacgatggcg agccgttgga gccgccgaca cgggtcacgc
 10021 tgccgcgccg gtagcacttg ggtgctgcag caaccgtaa gtgcgctgtt ccagactatc
 10081 ggctgtagcc gctcgccgc cctatacctt gtctgcctcc ccgcttgcg tcgcggtgca
 10141 tggagccggg ccacctgac ctgaatggaa gccggcgga cctcgtaac ggattaccg
 10201 ttttatcag gctctgggag gcagaataaa tgatcatatc gtcaattatt acctccacgg
 10261 ggagagcctg agcaactgg cctcagcat ttgagaagca cacggtcaca ctgctccg
 10321 tagtcaataa accggtaaac cagcaataga cataagcggc tatttaacga ccctgcctg
 10381 aaccgacgac cgggtcgaat ttgcttcga atttctgcca tcatccgct tattatcact
 10441 tattcaggcg tagcaccagg cgttaaggg caccaataac tgcctaaaa aaattacgcc
 10501 ccgccctgcc actcatcgca gtcggcctat tggtaaaaa atgagctgat ttaacaaaa
 10561 ttaacgcga atttaacaa aatattaacg ctacaattt ccattcgcca ttcaggctgc
 10621 gcaactgttg ggaagggcga tcggtgcggg cctctcgct attacgccag ctggcgaaag
 10681 ggggatgtgc tgcaaggcga ttaagttgg taacgccagg gtttccag tcacgacgtt
 10741 gtaaacgac ggccagttag cgcgcgtaac acgactact atagggcgaa ttggagctgg
 10801 ggccgcaatt ccaattcca ggcataaat aaaacgaaag gtcagtcga aagactgggc
 10861 ctttcgttt atctgtgtt tgcggtgaa cgctctctg agtaggaca atccgccgg
 10921 agcggatttg aacgttgca agcaacggc cggaggggtg cgggcaggac gcccgcata
 10981 aactgccagg aattaattc aggcataaa taaaacgaaa ggctcagtcg aaagactggg
 11041 ctttcgttt tatctgtgt ttgtcggtga acgctctct gagtaggaca aatccgccg
 11101 gagcggattt gaacgttgcg aagcaacggc ccggaggggtg gcgggcagga cgccgccat
 11161 aaactgccag gaattaattc caggcatcaa ataaaacgaa aggctcagtc gaaagactgg
 11221 gccttcgtt ttatctgtt tttgtcggt aacgctctc tgagtaggac aaatccgccg
 11281 ggagcggatt tgaacgttg gaagcaacgg cccggaggtt ggccggcagg acgcccgcca
 11341 taaactgcca ggaattaatt ccaggcatca aataaacga aaggctcagt cgaaagactg
 11401 ggccttcgt ttatctgtt gttgtcggt gaacgctct ctgagtagga caaatccgcc
 11461 gggagcggat ttgaacgtt cgaagcaac gcccgagggt tggcgggcag gacgcccgc
 11521 ataaactgcc aggaattgg

//

8.2.3 – pWW2181: scaffold plasmid: P_{tet} peptide-LZB, P_{sal} peptide-LZA

LOCUS pWW2181 5597 bp ds-DNA circular 10-FEB-2012
 DEFINITION
 ACCESSION
 VERSION

SOURCE
 ORGANISM
 COMMENT
 COMMENT
 COMMENT ApEinfo:methylated:1
 FEATURES Location/Qualifiers
 rep_origin complement(3720..4402)
 /label=ColE1 origin
 /ApEinfo_fwdcolor=#3f3f3f
 /ApEinfo_revcolor=#3f3f3f
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0
 width 5 offset 0
 CDS complement(4500..5159)
 /label=AmpR
 /ApEinfo_fwdcolor=#3f3f3f
 /ApEinfo_revcolor=#3f3f3f
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0
 width 5 offset 0
 misc_feature 1260..1627
 /label=Terminator (rrnB)
 /ApEinfo_fwdcolor=#800000
 /ApEinfo_revcolor=#800000
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0
 width 5 offset 0
 misc_feature 3178..3545
 /label=Terminator (rrnB)(1)
 /ApEinfo_label=Terminator (rrnB)
 /ApEinfo_fwdcolor=#800000
 /ApEinfo_revcolor=#800000
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0
 width 5 offset 0
 misc_feature 1841..2743
 /label=nahR
 /ApEinfo_fwdcolor=#9ae4a4
 /ApEinfo_revcolor=#9ae4a4
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0
 width 5 offset 0
 misc_feature 2822..2885
 /label=P_Sal
 /ApEinfo_fwdcolor=#008000
 /ApEinfo_revcolor=#008000
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0
 width 5 offset 0
 misc_feature complement(138..761)

```

        /label=tetR
        /ApEinfo_fwdcolor=#9ae4a4
        /ApEinfo_revcolor=#9ae4a4
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 876..945
        /label=P_Tet
        /ApEinfo_fwdcolor=#008000
        /ApEinfo_revcolor=#008000
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 993..1010
        /label=6xHis Tag
        /ApEinfo_fwdcolor=#c0c0c0
        /ApEinfo_revcolor=#c0c0c0
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 2941..2958
        /label=6xHis Tag(1)
        /ApEinfo_label=6xHis Tag
        /ApEinfo_fwdcolor=#c0c0c0
        /ApEinfo_revcolor=#c0c0c0
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 1050..1082
        /label=SH3pep (.1uM)
        /ApEinfo_fwdcolor=#d900d9
        /ApEinfo_revcolor=#d900d9
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 2998..3030
        /label=SH3pep (.1uM)(1)
        /ApEinfo_label=SH3pep (.1uM)
        /ApEinfo_fwdcolor=#d900d9
        /ApEinfo_revcolor=#d900d9
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 1011..1049
        /label=Tev cleavable linker
        /ApEinfo_fwdcolor=#c0c0c0
        /ApEinfo_revcolor=#c0c0c0
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 2959..2997

```

```

        /label=Tev cleavable linker(1)
        /ApEinfo_label=Tev cleavable linker
        /ApEinfo_fwdcolor=#c0c0c0
        /ApEinfo_revcolor=#c0c0c0
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 3058..3156
        /label=LZA
        /ApEinfo_fwdcolor=#800080
        /ApEinfo_revcolor=#800080
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 1110..1238
        /label=LZB
        /ApEinfo_fwdcolor=#800080
        /ApEinfo_revcolor=#800080
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 1083..1109
        /label=GS linker
        /ApEinfo_fwdcolor=#c0c0c0
        /ApEinfo_revcolor=#c0c0c0
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0
misc_feature 3031..3057
        /label=GS linker(1)
        /ApEinfo_label=GS linker
        /ApEinfo_fwdcolor=#c0c0c0
        /ApEinfo_revcolor=#c0c0c0
        /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
        width 5 offset 0

```

ORIGIN

```

1 gaattcacta gtgatccggt tccatttagg tgggtacgtt ggagccgcat tatttcgct
61 ttatgaatct aaaggggtgt taactcgaca tcttggttac cgtgaagtta ccatcacgga
121 aaaagggtat gctgcttta agaccactt tcacatttaa gttgttttc taatccgcat
181 atgatcaatt caaggccgaa taagaaggct ggctctgcac cttggtgatc aaataattcg
241 atagcttgtc gtaataatgg cggcatacta tcagtagtag gtgtttcct ttctcttta
301 gcgacttgat gctcttgatc ttccaatacg caacctaaag taaaatgccc cacagcgctg
361 agtgcataata atgcattctc tagtgaaaaa ccttgttggc ataaaaaggc taattgattt
421 tcgagagttt catactgttt ttctgtaggc cgtgtaccta aatgtacttt tgctccatcg
481 cgatgactta gtaaagcaca tctaaaactt ttagcggtat tacgtaaaaa atcttgccag
541 ctttcccctt ctaaagggca aaagtgagta tgggtgcctat ctaacatctc aatggctaag
601 gcgctgagca aagcccgtt attttttaca tgccaataca atgtaggctg ctctacacct
661 agcttctggg cgagtttacg gggtgttaaa ccttcgattc cgacctcatt aagcagctct

```

721 aatgcgctgt taactacttt acttttatct aatcttgaca tcattaattc ctaatttttg
781 ttgacactct atcggtgata gagttatfff accactccct atcagtgata gagaaaagtg
841 aaaatccata tgactagcag atccactaga gtcgactaag aaaccattat tatcatgaca
901 ttaacctata aaaataggcg tatcacgagg ccctttcgtc ttcacgctag cggtagctga
961 agtgtggccc aacgcaaaact aagagatcta tgcacacca tcaccatcac gactacgaca
1021 tcccgactac cgaaaacctg tacttccagc ctccacctgc attaccgcca aagagaaggg
1081 gtggatctgg gtccggcagc ggtggatcta tcacaattcg ggcagctttt ctggagaagg
1141 aaaacaccgc actgcgtact gagattgctg aactggagaa agaagtaggc cgctgcgaga
1201 atattgttag caaatatgaa acccgctacg gcccctggg atcctaactc gagtaatgag
1261 aagcttgggc ccgaacaaaa actcatctca gaagaggatc tgaatagcgc cgctgacat
1321 catcatcatc atcattgagt ttaaagggtc tccagcttgg ctgttttggc ggatgagaga
1381 agattttcag cctgatacag attaaatcag aacgcagaag cggctctgata aaacagaatt
1441 tgcctggcgg cagtagcgcg gtggtccac ctgaccccat gccgaactca gaagtgaac
1501 gccgtagcgc cgatggtagt gtggggtctc cccatgcgag agtagggaac tgccaggcat
1561 caataaaaac gaaaggctca gtcgaaagac tgggcctttc gttttatctg ttgtttgtcg
1621 gtgaactaat ttctagtgcg atcccgcgaa gaacaaaaaa agctcgacag agggcgcggt
1681 catttaggt cgggcgatc ggcgcccgcg gctcggctgg tgtgccgcac agcaccgct
1741 acgtgagctg ccagttgatg aactcccc gttgccagct agggcgcaag cgggctgtat
1801 aagatcactg cccatcacat tgatcggctc ggatttttc tcaatccgta aacagggtcaa
1861 acatcagttg ccgcaaccaa atattggcta ggtccttgtg gtacttcgca tgccagaaca
1921 tgttgatggc tatttcaggc aagacgactg ggtgcggcaa ggcgcttagg ccgaagggt
1981 ccacgcagca gtcggctaaa cgtatcgga cagtggcgag cagatcgggt cgctggagga
2041 tgtggccaac ggcggcgaag tgcggcactt ccagacggat gtcgcgccgg atgccgaccc
2101 gtgtcatgta cgtgtccacc tcgccgtggc cggtgccagc ggcgatgaca cgcacgtggc
2161 cgtaggaaca gaagcgctcc agatgcagg gttcgcgggt gactggatgg tccttgcgac
2221 ataggcacac gtagtgattc tggagcagcc ggcgctgaaa gaagccagtt tgagattgg
2281 gaagcaggcc cacggccaag tccacggtc cgttctgcaa ggcctgcatc aggtcatcg
2341 aactgtcgcg caccgtactg atcacgcaat tgggggcctg gtgagccagc acatccatca
2401 gccgcggcat gaagtagatt tcgcaatgt cggatcatggc cagggtgaag gtacgctcgc
2461 tggtagcggc atcgaagctt tcatggtgct gtagggcggt gcgcagtgcg tgcattggccg
2521 aagtgcggg ctccggcaga tgcgcggcat aggggtgtggg ttccattccc tgatgtgtgc
2581 gcacgaagag tgggtcctgt agcaggtgc gcaggcggtt cagcgattg ctacggcag
2641 gctgggtcag gccaggttc tccgagtga tagagacgcg tctgtcgacc agcaattggt
2701 tgaacaccac cagcaggtt aaatccaggt cagcagttc catggggcct cgctggggt
2761 attgctggtg cccggccggg cgcaatattc atgttgatga ttattatat atcagtggt
2821 gtatttatca atattgttg ctccgttatc gttattaaca agtcatcaat aaagccatca
2881 cgagtacat agagctagcG GTACCACGGG AAATCCACCA TATGGGGCAC
GagatctATG
2941 catcaccatc accatcacga ctacgacatc ccgactaccg aaaacctgta cttcagcct
3001 ccacctgcat taccgcaaaa gagaaggcgt ggtatctgggt ccggcagcgg tggatctagc
3061 gttaaagaac tggaagataa aaacgaagag ctcttgagcc gtatttatca tgaacgcaac
3121 gaagtggcgc gtctgaaaaa gcttgttggc gaacgtggat cctgactcga gtaatgagaa
3181 gcttggggcc gaacaaaaac tcattcaga agaggatctg aatagcgccg tcgacatca
3241 tcattcatcat cattgagttt aaacggtctc cagcttggct gttttggcgg atgagagaag

3301 attttcagcc tgatacagat taaatcagaa cgcagaagcg gtctgataaa acagaatttg
 3361 cctggcgcca gtagcgcggt ggtccacac gaccccatgc cgaactcaga agtgaaacgc
 3421 cgtagcgccg atggtagtgt ggggtctccc catgagagag tagggaactg ccaggcatca
 3481 aataaaacga aaggctcagt cgaagactg ggccttcgt tttatctgtt gttgtcgggt
 3541 gaactaattt ctgactgca ggcttcctcg ctactgact cgctgcgctc ggtcgttcgg
 3601 ctgcgcgag cggtatcagc tactcaaag gcggaatac gggtatccac agaactcaggg
 3661 gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag
 3721 gccgcgttgc tggcgtttt ccataggctc gccccctg acgagcatca caaaaatcga
 3781 cgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc gttccccct
 3841 ggaagctccc tcgtgcgctc tctgttcg accctgccgc ttaccggata cctgtccgc
 3901 tttctccctt cggaagcgt ggcgcttct catagctcac gctgtaggta tctcagttcg
 3961 gtgtaggctg ttcgtccaa gctgggctgt gtgcacgaac cccccgtca gcccgaccgc
 4021 tgcgccttat ccgtaacta tcgtcttgag tccaacccgg taagacacga cttatcgcca
 4081 ctggcagcag cactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag
 4141 ttctgaagt ggtggcctaa ctacggctac actagaagga cagtatttg tatctgcgt
 4201 ctgctgaagc cagttacctt cggaaaaaga gttgtagct cttgatccgg caaacaacc
 4261 accgctgga gcggtggtt tttgtttgc aagcagcaga ttacgcgag aaaaaagga
 4321 tctcaagaag atccttgat ctttctacg ggtctgacg ctactggaa cgaaaactca
 4381 cgtaaggga tttggatc gagattatca aaaaggatct tcactagat cttttaaat
 4441 taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaactggtc tgacagttac
 4501 caatgcttaa tcagtgggc acctatctca gcgatctgtc tatttcgttc atccatagtt
 4561 gcctgactcc ccgtcgtga gataactacg atacgggagg gttaccatc tggccccagt
 4621 gctgcaatga taccgcgaga ccacgctca ccggctccag atttatcagc aataaaccag
 4681 ccagccggaa gggccgagcg cagaagtgg cctgcaactt tatccgcctc catcagttc
 4741 attaattgtt gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacggt
 4801 gttgccattg ctacaggcat cgtgggtgta cgctcgtcgt ttggtatggc ttactcagc
 4861 tccggttccc aacgatcaag gcgagttaca tgatcccca tgttggtgaa aaaagcgggt
 4921 agtccttcg gtccctcgat cgtgtcaga agtaagttgg ccgagtggt atcactcatg
 4981 gttatggcag cactgcataa ttcttact gtcatgcat ccgtaagatg ctttctgtg
 5041 actggtgagt actcaacaa gtcattctga gaatagtga tgcggcgacc gagttgctc
 5101 tgcccgcggt caatacggga taataccgcg ccacatagca gaactttaaa agtgctcatc
 5161 attggaaaac gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt
 5221 tcgatgtaac cactcgtgc acccaactga tcttcagcat ctttacttt caccagcgtt
 5281 tctgggtgag caaaaacagg aaggcaaaa gccgcaaaa aggaataag ggcgacacgg
 5341 aaatgttgaa tactcact ctctctttt caatattatt gaagcattta tcagggttat
 5401 tgttcatga gcggatacat attgaatgt atttagaaaa ataaacaaat aggggttccg
 5461 cgcacatttc cccgaaaagt gccacgtgac gtctaagaaa ccattattat catgacatta
 5521 acctataaaa ataggcgtat cagcaggcag aatttcagat aaaaaaatc cttagcttc
 5581 gctaaggatg atttctg

//

8.2.4 – pWW2254: autoinhibited kinase plasmid: P_{tet} AiTaz-(SH3)₄

LOCUS pWW2254 5940 bp ds-DNA circular 17-FEB-2012

DEFINITION
 ACCESSION
 VERSION
 SOURCE
 ORGANISM
 COMMENT
 COMMENT
 COMMENT ApEinfo:methylated:1
 FEATURES Location/Qualifiers
 misc_feature 3396..3763
 /label=TrrnB
 /ApEinfo_fwdcolor=#9191ff
 /ApEinfo_revcolor=#9191ff
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature 4954..5794
 /label=p15A ORI
 /ApEinfo_fwdcolor=#808080
 /ApEinfo_revcolor=#808080
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature 2511..2534
 /label=GS linker
 /ApEinfo_fwdcolor=#c0c0c0
 /ApEinfo_revcolor=#c0c0c0
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature 2541..2564
 /label=GS linker(1)
 /ApEinfo_label=GS linker
 /ApEinfo_fwdcolor=#c0c0c0
 /ApEinfo_revcolor=#c0c0c0
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature 2751..2774
 /label=GS linker(2)
 /ApEinfo_label=GS linker
 /ApEinfo_fwdcolor=#c0c0c0
 /ApEinfo_revcolor=#c0c0c0
 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
 width 5 offset 0
 misc_feature 2961..2984
 /label=GS linker(3)
 /ApEinfo_label=GS linker

```

    /ApEinfo_fwdcolor=#c0c0c0
    /ApEinfo_revcolor=#c0c0c0
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 3171..3194
    /label=GS linker(4)
    /ApEinfo_label=GS linker
    /ApEinfo_fwdcolor=#c0c0c0
    /ApEinfo_revcolor=#c0c0c0
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 2574..2744
    /label=SH3
    /ApEinfo_fwdcolor=#8000ff
    /ApEinfo_revcolor=#8000ff
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 2784..2954
    /label=SH3(1)
    /ApEinfo_label=SH3
    /ApEinfo_fwdcolor=#8000ff
    /ApEinfo_revcolor=#8000ff
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 2994..3164
    /label=SH3(2)
    /ApEinfo_label=SH3
    /ApEinfo_fwdcolor=#8000ff
    /ApEinfo_revcolor=#8000ff
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 3204..3374
    /label=SH3(3)
    /ApEinfo_label=SH3
    /ApEinfo_fwdcolor=#8000ff
    /ApEinfo_revcolor=#8000ff
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 2022..2489
    /label=EnvZ KD
    /ApEinfo_fwdcolor=#62b0ff
    /ApEinfo_revcolor=#62b0ff
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0

```

```

misc_feature 993..1757
    /label=Tar receptor
    /ApEinfo_fwdcolor=cyan
    /ApEinfo_revcolor=cyan
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 138..761
    /label=tetR
    /ApEinfo_fwdcolor=#ffff00
    /ApEinfo_revcolor=#ffff00
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 876..945
    /label=P_TET
    /ApEinfo_fwdcolor=#008000
    /ApEinfo_revcolor=#008000
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 3885..4541
    /label=CamR
    /ApEinfo_fwdcolor=#3f3f3f
    /ApEinfo_revcolor=#3f3f3f
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 1761..2006
    /label=HL29a DD
    /ApEinfo_fwdcolor=cyan
    /ApEinfo_revcolor=cyan
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0
misc_feature 990..2489
    /label=aiTaz(HL29a)
    /ApEinfo_fwdcolor=#d84b4b
    /ApEinfo_revcolor=#d84b4b
    /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
    width 5 offset 0

```

ORIGIN

```

1 gaattcacta gtgatccgtt tccatttagg tgggtacgtt ggagccgcat tattttcgct
61 ttatgaatct aaaggggtgt taactcgaca tcttggttac cgtgaagtta ccatcacgga
121 aaaaggttat gctgcttta agaccactt tcacatttaa gttgttttc taatccgcat
181 atgatcaatt caaggccgaa taagaaggct ggctctgcac cttggtgatc aaataattcg
241 atagcttgtc gtaataatgg cggcatacta tcagtagtag gtgtttcct ttctcttta
301 gcgacttgat gctcttgatc ttccaatacg caacctaaag taaaatgccc cacagcgctg
361 agtgcatata atgcattctc tagtgaaaaa ccttgttggc ataaaaaggc taattgattt

```

421 tcgagagttt catactgttt ttctgtaggc cgtgtaccta aatgtacttt tgctccatcg
 481 cgatgactta gtaaagcaca tctaaaactt ttagcggtat tacgtaaaaa atcttgccag
 541 ctttcccctt ctaaagggca aaagtgagta tggcgctat ctaacatctc aatggctaag
 601 gcgtcgagca aagcccgtt attttttaca tgccaataca atgtaggctg ctctacacct
 661 agcttctggg cgagtttacg ggttgtaaaa ccttcgattc cgacctcatt aagcagctt
 721 aatgcgctgt taatcacttt acttttatct aatcttgaca tcattaattc ctaatttttg
 781 ttgacactct atcggtgata gagttatfff accactccct atcagtata gagaaaagt
 841 aaaatccata tgactagcag atccactaga gtcgactaag aaaccattat tatcatgaca
 901 ttaacctata aaaataggcg tatcacgagg cccttcgctc ttcacgctag cggtagccga
 961 tctatgactc agttacttcc gtgagatcta tgattaaccg tatccgcgta gtcacgctgt
 1021 tggtaatggg gctgggggta ttcgactgt tacagcttat ttccggcagt ctgtttttt
 1081 cttcccttca ccatagccag aagagctttg tggtttcaa tcaattacgg gaacagcagg
 1141 gcgagctgac gtcaacctgg gatttaatgc tgcaaagcg cattaacctg agtcgttcag
 1201 cggtagcgat gatgatggat tctccaatc acaaaagtaa cgcaaagt gaattgctcg
 1261 atagcgccag gaaaacattg gcgcaggcag cgacgcatta taaaaaattc aaaagcatgg
 1321 caccgttacc tgaaatgggc gctaccagtc gtaattattga tgaaaaatat aaaaactatt
 1381 acacagcgtt aactgaactg attgattatc tagattatgg caatactgga gcttatttcg
 1441 ctacgccaac ccagggaatg caaatgcaa tgggcgaagc gtttgctcag tacgccctca
 1501 gcagtgaaaa actgtatcgc gatatcgtca ctgacaacgc agatgattac cgatttgccc
 1561 agtggaact ggcggtatc gcgctgggtg tggattgat tctgctggg gcgtggtacg
 1621 gcattcgccg tatgttgctt actccgctgg caaaaattat tgctcacatt cgcgaaatcg
 1681 ccggtggtaa cctggcgaat accctgacca ttgacgggag cagtgaatg ggcgacctgg
 1741 cgagagcgt ttcacatatg gcggtgggtg ttaagcaact ggcggatgac cgcacgctgc
 1801 tgatggcggg ggtaagtcac gacttgcgca cgccgctgac gcgtattcgc cttgcgactg
 1861 agatgttcac aaacagcgcc ccgccaccag cgctgccacc gaaacgtcgc cgtactagca
 1921 tcgaagtagt tgaactggca gaatcgatca ataaagatat cgaagagtgc aacgccatca
 1981 ttgagcagtt tatcgactac ctgcgaccg ggcaggagat gccgatggaa atggcggatc
 2041 ttaatgcagt actcggtagg gtgattgctg ccgaaagtgg ctatgagcgg gaaattgaaa
 2101 ccgcgcttta ccccgccagc attgaagtga aaatgcaccc gctgtcgtac aaacgcgcgg
 2161 tggcgaatat ggtggtcaac gccgcccgtt atggcaatgg ctggatcaaa gtcagcagcg
 2221 gaacggagcc gaatcgcgcc tggttccagg tggaagatga cggtagccgga attgcgcccg
 2281 aacaacgtaa gcacctgtc cagccgtttg tccgcggcga cagtgcgcgc accattagcg
 2341 gcacgggatt agggctggca attgtgcagc gtatcgtgga taaccataac gggatgctgg
 2401 agcttggcac cagcgagcgg ggcgggctt cacttcgctc ctggctgcca gtgcccgtta
 2461 cgcgggagca ggcacgaca aaagaagggg cgccgcagc ttcgggatct ggtagtggca
 2521 gcggcagcgg tagtggtatc ggtagtggca gcggcagcgg tagtggtatc atggcagagt
 2581 atgtgcgggc cctgtttgac ttaatggga atgatgaaga agacctccc ttaagaaag
 2641 gagacatcct gagaatccgg gataagcctg aagagcagtg gtggaatgca gaggacagcg
 2701 aaggaaagag ggggatgatt cctgtccctt acgtggagaa gtatggatct ggtagtggca
 2761 gcggcagcgg tagtggtatc atggcagagt atgtgcgggc cctgtttgac ttaatggga
 2821 atgatgaaga agacctccc ttaagaaag gagacatcct gagaatccgg gataagcctg
 2881 aagagcagtg gtggaatgca gaggacagcg aaggaaagag ggggatgatt cctgtccctt
 2941 acgtggagaa gtatggatct ggtagtggca gcggcagcgg tagtggtatc atggcagagt
 3001 atgtgcgggc cctgtttgac ttaatggga atgatgaaga agacctccc ttaagaaag

3061 gagacatcct gagaatccgg gataagcctg aagagcagtg gtggaatgca gaggacagcg
 3121 aaggaaagag ggggatgatt cctgtccctt acgtggagaa gtatggatct ggtagtgga
 3181 gcggcagcgg tagtgatct atggcagagt atgtgcgggc cctgtttgac ttaaatgga
 3241 atgatgaaga agaccttccc ttaagaaag gagacatcct gagaatccgg gataagcctg
 3301 aagagcagtg gtggaatgca gaggacagcg aaggaaagag ggggatgatt cctgtccctt
 3361 acgtggagaa gtatggatcc tgactcgagt aatgagaagc ttgggcccga aaaaaactc
 3421 atctcagaag aggatctgaa tagcgccgtc gaccatcatc atcatcatca ttgagtttaa
 3481 acggtctcca gcttggtgt tttggcggat gagagaagat tttcagcctg atacagatta
 3541 aatcagaacg cagaagcggc ctgataaaac agaatttgcc tggcggcagt agcgcggtg
 3601 tccacactga ccccatgccg aactcagaag tgaaacgccg tagcgccgat ggtagtggtg
 3661 ggtctcccca tgcgagagta gggaaactgcc aggcataaaa taaaacgaaa ggctcagtcg
 3721 aaagactggg cctttcggtt tatctgtgt ttgtcgggtga actaatttct agactgcagt
 3781 tgatcgggca cgtaagaggt tccaacttc accataatga aataagatca ctaccgggcg
 3841 tatttttga gttatcgaga tttcaggag ctaaggaagc taaaatggag aaaaaatca
 3901 ctggatatac caccgtgat atatccaat ggcacgtaa agaacatttt gaggcatttc
 3961 agtcagttgc tcaatgtacc tataaccaga ccgttcagct ggatattacg gccttttaa
 4021 agaccgtaaa gaaaaataag cacaagtttt atccggcctt tattcacatt ctgcccgcc
 4081 tgatgaatgc tcatccggaa ttcgtatgg caatgaaaga cggtagagctg gtgatatggg
 4141 atagtgttca ccctgttac accgtttcc atgagcaaac tgaaacgttt tcatcgctct
 4201 ggagtgaata ccacgacgat ttccggcagt ttctacacat atattcgcaa gatgtggcgt
 4261 gttacgggtga aaacctggcc ttttcctta aagggtttat tgagaatatg ttttcgtct
 4321 cagccaatcc ctgggtgagt ttcaccagtt tgatttaaa cgtggccaat atggacaact
 4381 tcttcgcccc cgttttcacc atgggcaaat attatacgca aggcgacaag gtgctgatgc
 4441 cgctggcgat tcaggttcat catgccgttt gtgatggctt ccatgtcggc agaatgctta
 4501 atgaattaca acagtactgc gatgagtggc agggcggggc gtaatttgat atcgagctcg
 4561 cttggactcc tgttgataga tccagtaatg acctcagaac tccatctgga ttgttcaga
 4621 acgctcggtt gccgccgggc gtttttatt ggtgagaatc caagcctcg tgagaatcca
 4681 agcctcgatc aacgtctcat tttcgcaaaa agttggccca gggcttccc gtatcaacag
 4741 ggacaccagg atttatttat tctgcgaagt gatcttccgt cacagggtatt tattcggcgc
 4801 aaagtgcgtc gggtagtgct gccaacttac tgatttagtg tatgatggtg ttttgaggt
 4861 gctccagtgg cttctgttc tatcagctgt cctcctgtt cagctactga cgggggtggtg
 4921 cgtaacggca aaagcaccgc cggacatcag cggtagcga gtgtatactg gcttactatg
 4981 ttggcactga tgagggtgtc agtgaagtgc ttcagtggc aggagaaaaa aggctgcacc
 5041 ggtgcgtcag cagaatatgt gatacaggat atattccgct tcctcgctca ctgactcgct
 5101 acgctcggtc gttcgactgc ggcgagcga aatggcttac gaacggggcg gagatttcct
 5161 ggaagatgcc aggaagatac ttaacaggga agtgagaggg ccgcggcaaa gccgttttc
 5221 cataggctcc gccccctga caagcatcac gaaatctgac gctcaaatca gtggtggcga
 5281 aacccgacag gactataaag ataccaggcg tttccccctg gcggctcct cgtgcgctct
 5341 cctgttcctg ctttcggtt taccggtgtc attccgctgt tatggccgcg tttgtctcat
 5401 tccacgcctg aactcagtt ccgggtaggc agttcgctcc aagctggact gtatgcacga
 5461 acccccggtt cagtcggacc gctgcgcctt atccggtaac tatcgtctg agtccaaccc
 5521 ggaaagacat gaaaagcac cactggcagc agccactgg aattgattta gaggagttag
 5581 tcttgaagtc atgcgccggt taaggctaaa ctgaaaggac aagttttggt gactgcgctc
 5641 ctccaagcca gttacctcg ttcaaagagt tggtagctca gagaaccttc gaaaaaccgc

5701 cctgcaaggc ggTTTTTCg tttcagagc aagagattac gcgcagacca aaacgatctc
5761 aagaagatca tcttattaat cagataaaat atttctagat ttcagtgcaa tttatctctt
5821 caaatgtagc acctgaagtc agccccatac gatataagtt gtaattctca tgtttgacag
5881 cttatcatcg ataagcttcc gatggcgcg c gagaggctt tacactttat gcttccggct

//