ENGINEERING ENZYME COLOCALIZATION WITHIN THE CYTOPLASM OF ESCHERICHIA COLI

by Robert John Conrado

ENGINEERING ENZYME COLOCALIZATION WITHIN THE CYTOPLASM OF *ESCHERICHIA COLI*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Robert John Conrado

August 2010

ENGINEERING ENZYME COLOCALIZATION WITHIN THE CYTOPLASM OF *ESCHERICHIA COLI*

Robert John Conrado, Ph. D.

Cornell University 2010

There is ever increasing interest in developing biological routes for the production of valuable chemicals, especially for those products whose current manufacture is from non-renewable resources or whose synthesis has proven difficult or intractable using traditional synthetic routes. One common approach to this problem is to harness the synthetic potential of living cells that carry engineered metabolic pathways. However, even after a functional biosynthetic pathway is created in a metabolically engineered microorganism such as *Escherichia coli*, many bottlenecks can ensue including (i) diverted metabolic flux away from the desired product, and (ii) accumulation of toxic intermediates that can inhibit host cell growth. This work reviews the entire history of engineered colocalized enzymes, both *in vitro* and *in vivo*, and posits the necessary components for successful engineered systems. Next, a stochastic multidimensional model was developed to study the impact of enzyme compartmentalization to measure the effect of increased local concentrations versus substrate diffusivity. As a first step to engineer these systems *in vivo*, fusion proteins and various post translational interactions were employed to improve the production of both 1,2-propanediol and 1,2,4-butanetriol in *E. coli*. The difficulty in resolving true synergy indicated the importance of independent colocalization techniques that would provide better engineering controls and permit multiple parameters for further optimization. In response, DNA scaffolds were developed to colocalize a series of sequential reactions

along a plasmid DNA surface. This was accomplished for both 1,2-propanediol and mevalonate production in *E. coli* and resulted in 5-fold and 4-fold increases respectively. Future work is needed to further optimize these systems and understand the inherent complexity of coupled, multimeric enzymes. The use of synthetic DNA scaffolds *in vivo* provides a powerful new methodology that can be used alongside conventional engineering methods. In the long term, the possibility of integrating several of these strategies simultaneously to balance pathway flux, eliminate metabolic bottlenecks, reduce cell stress, and prevent accumulation of unwanted intermediates and/or byproducts could pave the way to produce commercially viable levels of diverse metabolic products such as biofuels, specialty chemicals and therapeutics.

BIOGRAPHICAL SKETCH

Robert John Conrado enrolled at Dartmouth College in September of 2001, where he graduated *cum laude* with a Bachelors of Arts in Engineering in June of 2005. Additionally, he graduated with a Bachelors of Engineering in Biochemical Engineering from the Thayer School of Engineering at Dartmouth College in June of 2005. Later, he enrolled at Cornell University in August of 2005 and received his doctorate of philosophy in Chemical and Biomolecular Engineering in August of 2010 under the guidance of Dr. Matthew P. DeLisa.

To Mom, Dad, Colleen, Allison, and Harper

ACKNOWLEDGMENTS

This work was supported by the Cornell University College of Engineering's Sustainable Energy Systems Seed Grant Program, and the Office of Naval Research under a Young Investigator Award (to MPD) and also Grant N00014-07-1-0027.

In addition, technical support was provided by: Thomas Mansell, Hansen Xu, Parry Grewal, Catherine Manix, Douglas Reed, Andrea Yonge, and Preeti Panda.

I have been supported by a cast of family, friends, mentors, and teachers throughout the Ph. D. and in my life up to this point. In particular, Dr. Matt DeLisa has been a wonderful mentor and friend. My family has always provided unwavering support, Colleen as great friend and confidante, my dad as my biggest promoter, and my mom as my biggest believer. Allison has been a daily source of support and friendship. Longtime friend, Nate Gorence, has constantly provided both confidence and humility. Finally, I would like to thank several mentors and teachers including: Dr. Andreas Wankerl, Prof. Tillman Gerngross, Dr. Gavin Bernard, Mr. Tom Alessandri, Mrs. Sullivan, Mrs. Ann Freitas, Mrs. Kirkland, Sister Aileen, and Sister Mary.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Metabolites

1,2-PD	1,2-propanediol
1,3-BPG	1,3-diphosphoglycerate
1,2,4-BT	1,2,4-butanetriol
2PG	2-phosphoglycerate
3PG	3-phosphoglycerate
6PG	6-phosphogluconate
AcCoA	acetyl-coenzyme A
ADP	adenosindiphosphate
aKG	a-ketoglutarate
AMP	adenosinmonophosphate
ATP	adenosintriphosphate
Cho	chorismate
DHAP	dihydroxyacetone-phosphate
Dipim	diaminopimelate
E4P	erythrose-4-phosphate
F6P	fructose-6-phosphate
FAD	flavin-adenine-dinucleotide, oxidized
$FADH_2$	flavin-adenine-dinucleotide, reduced
FBP	fructose-1,6-bisphosphate
G1P	glucose-1-phosphate
G6P	glucose-6-phosphate
GAP	glyceraldehyde-3-phosphate
Gly	glycine
Ile	isoleucine
Kival	a-ketoisovalerate
LAla	L-alanine
Met	methionine
MG	methylglyoxal
Mur	mureine
NAD	diphosphopyridindinucleotide, oxidized
NADH	diphosphopyridindinucleotide, reduced
NADP	diphosphopyridindinucleotide-phosphate, oxidized
NADPH	diphosphopyridindinucleotide-phosphate, reduced
OAA	oxaloacetate
PEP	phosphoenolpyruvate
Poly Sacc	polysaccharide
PRPP	phosphoribosylpyrophosphate
PYR	pyruvate
R-1,2-PD	*R*-1,2-propanediol
Rib5P	ribulose-5-phosphate
Ribu5P	ribose-5-phosphate
R-Lac	*R*-lactaldehyde
Sed7P	sedoheptulose-7-phosphate

Ser	serine
Ser	serine
SuccCoA	succinyl-coenzymeA
Trp	tryptophan
Xy5P	xylulose-5-phosphate

Enzymes

AASynth	amino acid synthesis 1
AASynth2	amino acid synthesis 2
AASynth3	amino acid synthesis 3
AC	adenylate cyclase
AdhP	ethanol dehydrogenase
aKGDH	a-ketoglutarate dehydrogenase
ALDI	alcohol dehydrogenase
AtoB	acetyl-CoA acetyltransferase
CFSynth	cholesterol, fatty acid synthesis
CI1	cometabolite interconversion 1
CI2	cometabolite interconversion 2
CI3	cometabolite interconversion 3
DHAD	glycerol dehydrogenase
DkgA	2,5-diketo-D-gluconate reductase A
DkgB	2,5-diketo-D-gluconate reductase B
FADH2cons	$FADH_2$ conservation
Fmase	fumerase
FucO	*L*-1,2-propanediol oxidoreductase
GldA	*D*-aminopropanol dehydrogenase
HMGR	hydroxymethylglutaryl-CoA reductase
HMGS	hydroxymethylglutaryl-CoA synthase
ISCDH	isocitrate dehydrogenase
MAE	malic enzyme
MALDH	malate dehydrogenase
MdlC	benzoylformate decarboxylase
MGS	methylglyoxal synthase
NADHcons	NADH conservation
NADPHcons	NADPH conservation
OxPhos	oxidative phosphorylation reaction 1
OxPhos2	oxidative phosphorylation reaction 2
SucCoAS	succinyl-coenzyme A synthetase
SUCDH	succinate dehydrogenase
TCA1	tricarboxylic acid cycle reaction 1
YagE	predicted dihydropicolinate synthase-like protein
YagF	predicted dehydratase
Xdh	xylose dehydrogenase

Organisms

A. thaliana	*Arabidopsis thaliana*
A. niger	*Aspergillus niger*
A. shirousamii	*Aspergillus shirousamii*
B. stearothermophilus	*Bacillus stearothermophilus*
B. helvolum	*Brevibacterium helvolum*
C. crescentus	*Caulobacter crescentus*
C. cellulolyticum	*Clostridium cellulolyticum*
C. stercorarium	*Clostridium stercorarium*
C. thermosulfurogenes	*Clostridium thermosulfurogenes*
E. coli	*Escherichia coli*
M. Musculus	*Mus musculus*
M. gastri MB19	*Mycobacterium gastri MB19*
P. fluorescens	*Pseudomonas fluorescens*
P. syringae	*Pseudomonas syringae*
R. eutropha	*Ralstonia eutropha*
S. cerevisiae	*Saccharomyces cerevisiae*
S. solfataricus MT4	*Sulfolobales solfataricus MT4*
T. thermophilus	*Thermus thermophilus*

Others

DNA	Deoxyribonucleic acid
DSE	Degree of synergistic effect
EMSA	Electrophoretic mobility shift assay
PID	Protein interacting domain
TCA	Tricarboxylic acid

CHAPTER 1

ENGINEERING THE SPATIAL ORGANIZATION OF METABOLIC ENZYMES[1]

Introduction

Many years of engineering microorganisms have provided scientists with an expansive toolkit for the production of a vast array of valuable compounds, including commodity and specialty chemicals (e.g., biodegradable plastics), biofuels (e.g., ethanol, butanol), and therapeutic molecules (e.g., anticancer drugs, antimicrobial compounds). Since these products can be derived from renewable feedstocks, there is increasing interest in developing and optimizing biological routes for their production, especially for complex molecules whose manufacture has proven difficult or intractable using traditional synthetic routes. While numerous opportunities exist for biochemical conversion using natural and engineered microorganisms, a large number of metabolic reactions are characterized by low productivity and yield [1], as well as by undesirable side reactions [2]. Overcoming these issues presents a formidable challenge for the following reasons: first, metabolic intermediates are often unstable or toxic [3]; second, native and especially non-native enzymes can have low substrate specificity [2]; and third, competing reactions are frequently required or ill-defined [2]. Thus, when considering the use of microorganisms for chemical manufacture, several key questions arise such as: (1) how many metabolic processes would be economically viable if product titers were increased; and (2) are there techniques available that provide a simple, generic framework for increasing productivity of the desired product while reducing the formation of unwanted byproducts? In searching

[1] Adapted with permission from: Conrado RJ, Varner JD, DeLisa MP (2008) Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. Curr Opin Biotechnol 5:492-9 and DeLisa MP and Conrado RJ (2009) Synthetic metabolic pipelines. Nat Biotechnol 8:728-9.

for the answers to these questions, it is important to consider how natural biological systems employ spatial organization of metabolic enzymes as a strategy to overcome barriers to metabolite production. From an understanding of this organization, new strategies for efficiently assembling metabolic enzymes into structures that yield high product titers can be developed that will effectively expand the metabolic engineering toolkit.

Naturally occurring metabolic channels

In stark contrast to the dilute solutions in which classical enzymology studies are conducted, the interior of cells is a highly volume-occupied physico-chemical environment with macromolecules comprising ~20-30% of the cytoplasmic volume of *Escherichia coli* [4]. This high concentration of macromolecules results in crowding conditions that can have a profound effect on the structure, function, and evolution of biological processes. For instance, crowding can drastically alter equilibrium and biochemical rate constants and can impose increased tortuosity on molecular motion. As a result, small molecule diffusion measured in cells is 3-4-fold slower than in pure water and this effect dramatically increases for proteins of high molecular weight [5,6]. Crowding is also known to exert a generalized pressure for the reduction of the surface to volume ratio that drives the formation of well-defined oligomeric complexes via specific associations [4]. There are also a number of kinetic benefits [7] that have provided an evolutionary driving force towards multifunctional enzyme systems that couple enzyme reactions for sequential conversion steps, a process commonly referred to as metabolic channeling [8,9]. One of the most immediate consequences of metabolic channeling is that enzyme active sites for consecutive reactions are brought into close proximity, which reduces the intermediate diffusion distance and therefore increases the probability that a metabolic intermediate undergoes a sequential reaction step before diffusing away. This serves to increase the

catalytic efficiency for channeled substrates by 1-2 orders of magnitude as observed for the *Salmonella typhimurium* tryptophan synthase channel [10]. In addition, metabolic channels relieve cellular constraints as they maintain high local concentrations while keeping total cellular levels low, which is especially important for toxic or unstable intermediates [8,11,12]. Lastly, multifunctional enzyme systems enable metabolic control by regulating the assembly of enzymes that act on multiple substrates and preventing metabolic cross-talk between competing pathways [13]. In a remarkable example of metabolic control, Benkovic and coworkers observed that all six enzymes needed for purine biosynthesis colocalized to form clusters in the cytoplasm of HeLa cells termed "purinosomes" [14]. Unexpectedly, the enyzmes associated when cells were depleted of purine but dissociated when the demand for purine was low, providing clear evidence for dynamic regulation of metabolism at the level of post-translational assembly of enzymes into multifunctional complexes.

Natural multifunctional enzyme systems have been discovered in both primary and secondary metabolism and span a wide range of organisms. In the case of primary metabolic pathways, channeling has been observed in central carbon metabolism [15], fatty acid oxidation [16], the Calvin cycle [17], amino acid biosynthesis [18], the carboxysome [19], and the proteasome [20]. One notable example of channeling is seen for tryptophan biosynthesis, which is mediated by the polyaromatic pathway that is comprised of 13 total enzyme reactions, 7 of which constitute the polyaromatic branch. In fungi, this pathway includes a multifunctional enzyme known as the AROM complex that has evolved to link five distinct enzymatic activities into a single polypeptide (Figure 1.1A). The AROM complex is encoded by a continuous 4,812-bp open reading-frame without introns that is believed to have arisen by multiple gene fusions [9]. The output of this locus is a single mRNA transcript that specifies a

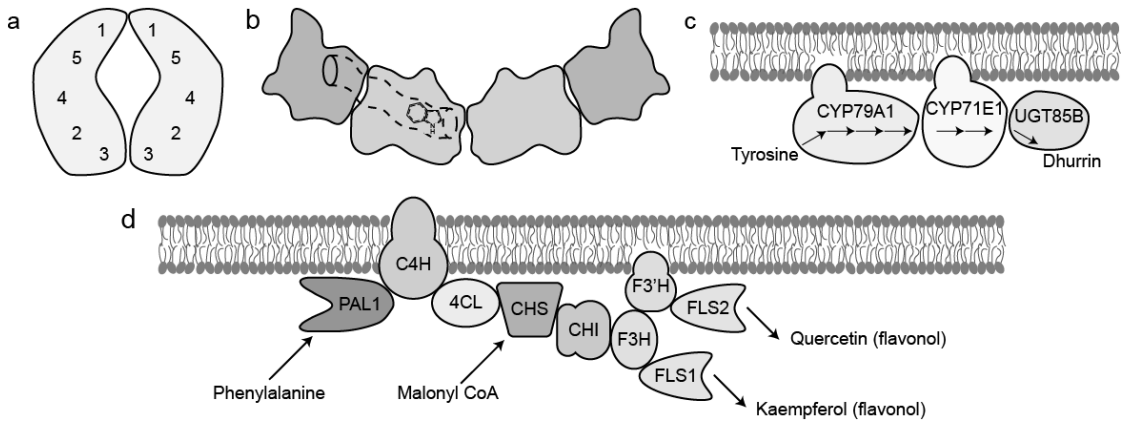


Figure 1.1. Nature's solutions to enzyme colocalization. (A) Fungal AROM complex with the five active sites labeled on this single polypeptide chain (adapted from [18]); (B) Tryptophan synthase αββα complex showing the physical tunnel channeling the indole intermediate (adapted from [10]); (C) Dhurrin synthesis from a multifunctional enzyme complex where arrows indicate a reaction step; (D) Model of phenylpropanoid metabolism in plants where several endoplasmic reticulum membrane bound enzymes allow multienzyme complex assembly for metabolite transfer between sequential reaction steps. Key branch points exist resulting from a common set of intermediates between the specific phenylpropanoid pathways (adapted from [21]).

pentafunctional polypeptide catalyzing five consecutive steps leading to the production of 5-enolpyruvylshikimate 3-phosphate in the shikimate pathway [18,22]. Interestingly, the same enzymatic activity in prokaryotes arises from five unlinked and monofunctional enzymes [23]. Evidence that the five active sites of the AROM protein behave as a coordinated multienzyme system come from kinetic studies in which higher catalytic activity and lower K_m values were observed for the individual catalytic sites within wild type complexes over complexes with mutations to residues in a single active site [18]. In addition to the AROM protein, the polyaromatic pathway also includes the tryptophan synthase complex, which catalyzes two sequential reactions to produce tryptophan (Figure 1.1B). Tryptophan synthase, an a_2b_2 complex, is thought to channel the metabolic intermediate indole from the active site of the a subunit to the active site of the b subunit. Channeling is conferred by a physical tunnel, with a diameter matching that of indole, that connects the adjacent active sites and prevents indole from diffusing into the bulk [10,24]. This mode of substrate channeling

effectively prevents indole from freely diffusing out of the cell and increases the reaction rates by 1-2 orders of magnitude over the free, uncomplexed subunits [10,25].

In contrast to the high-affinity protein interactions typical of the tryptophan biosynthetic machinery, channeling in secondary metabolism is often characterized by dynamic, low-affinity enzyme interactions such as has been observed for modular multifunctional polyketide synthases in bacteria [26]. Similarly, a plethora of studies in plants have demonstrated a role for metabolic channeling and colocalization in the formation of numerous secondary products [7], including the biosynthesis of isoprenoids [27], alkaloids [28], flavonoids [29], cyanogenic glucosides [11] (Figure 1.1C), and phenylpropanoids [30] (Figure 1.1D). Cyanogenic glucosides such as dhurrin represent a class of plant secondary compounds produced via a multienzyme system for protection against herbivores. In *Sorghum bicolor*, 7 catalytic steps yield dhurrin from L-tyrosine, but only 3 membrane-bound enzymes are required for this biotransformation [31] (Figure 1.1C). According to the proposed reaction mechanism, the first two enzymes of this system, CYP79A1 and CYP71E1, are both multifunctional, catalyzing 4 and 2 reactions respectively when expressed endogenously in *S. bicolor* and in recombinant *E. coli* [32]. As further evidence of channeling in this system, radioactive labeling experiments demonstrated that dhurrin production primarily came from L-tyrosine, even in the presence of saturating intermediates [33]. However, the channeling properties of this system were not fully appreciated until the *S. bicolor* 3-enzyme pathway was introduced into *Arabidopsis thaliana* [11]. In the engineered plant lines carrying only the first 2 enzymes for dhurrin biosynthesis, a stunted phenotype resulted from accumulation of the toxic intermediate, *p*-hydroxymandelonitrile. When the third enzyme of the dhurrin pathway was introduced into these transgenic plants, the normal growth phenotype was restored, dhurrin accumulated up to 4% by dry weight, and toxic byproducts were no

longer detectable within the plant material [11]. These dramatic results provide compelling evidence for the efficient coupling between the individual enzymes in dhurrin biosynthesis.

Engineering Multifunctional Enzyme Systems

The intentional engineering of metabolic pathways is a rapidly maturing field and much attention has already been given to cellular productivity and diverted flux. Indeed, a number of strategies have been reported for optimizing cellular metabolism including, for example, stabilizing mRNA transcript levels, balancing protein expression levels, evolving enzyme activity, and redirecting cellular metabolism [34]. These successes notwithstanding, very little attention has been given to the spatial organization of the recombinant metabolic pathway enzymes *in vivo*. Yet, based on numerous examples from nature (see above), it is clear that enzyme organization is a key design variable to be considered when engineering metabolism, as exemplified by the studies described below.

Immobilized Enzyme Systems. One approach to engineering multienzyme systems that mimic those found in nature is via immobilized enzymes whereby sequential chemical conversions are performed by tethering pathway enzymes on polymer particles (Figure 1.2A). In one notable example, a 140% increase in product formation was reported in comparison to the free enzyme system [35]. A similar effect has been observed for several two- and three-enzyme systems, either by immobilization [36,37,38] or chemical crosslinking of the enzymes [38]. While immobilized enzyme systems provide evidence that sequential pathway enzymes can be spatially organized to achieve kinetic benefits, the approach is often limited by the technical difficulty and cost associated with scale-up of cell-free systems. Thus, techniques for organizing enzymes inside of living cells would be preferable owing to the ease with which cell-based systems can be manipulated and scaled-up.

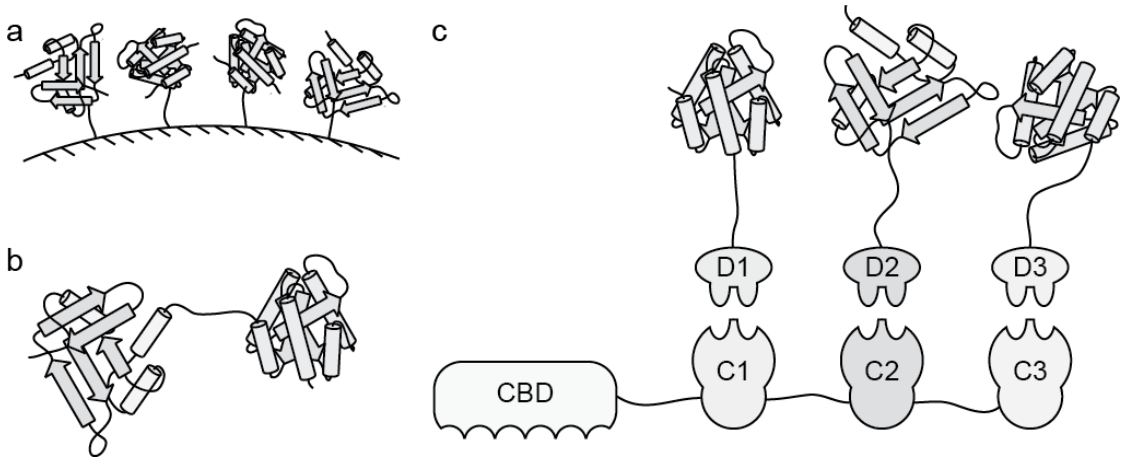


Figure 1.2. Engineering opportunities for enzyme colocalization. (A) Immobilized enzymes on a solid support. (B) Fusion proteins connected by a peptide linker. (C) Post-translational assembly of enzyme complexes where CBD represents the cellulose binding domain, C1-C3 represent cohesin domains, and D1-D3 represent dockerin domains (adapted from [39]).

Fusion Proteins. An alternative approach to artificial bifunctional and multifunctional enzymes is the creation of chimeric proteins whereby two or more distant genetic elements are combined by a short linker sequence to yield a single polypeptide that exhibits more than one activity (Figure 1.2B). For example, Mosbach and coworkers engineered an in-frame fusion between the structural gene for *E. coli* beta-galactosidase (LacZ), a tetrameric enzyme, and the gene encoding *E. coli* galactokinase (GalK), a monomeric enzyme, for the two-step conversion of lactose to galactose-1-phosphate. The resulting fusion protein displayed the enzymatic activity of both gene products, albeit with significantly reduced individual activities [40]. Shortly thereafter, the same group created an in-frame fusion between LacZ and the dimeric galactose dehydrogenase (GalDH) from *Pseudomonas fluorescens* for the sequential hydrolysis of lactose followed by the oxidation of the galactose, forming the corresponding lactone [41]. In this instance, the bifunctional enzyme displayed kinetic advantages (e.g., exhibited a 2-fold reduction in the saturation constant, K_m, for lactose and reached steady state 2-4 fold faster in solution) over the identical native

system in conversion of lactose to galactonolactone especially at low substrate concentrations. This suggests that the proximity conferred by the artificial fusion enabled some amount of synergistic coupling of the sequential reactions (i.e., substrate channeling) that was absent when the enzyme activities were distinct. Such an effect has been seen with other bifunctional and, in some cases multifunctional, enzyme systems (summarized in Table 1.1) as well as with immobilized enzyme systems above. It should be pointed out, however, that the notion of channeling for bifunctional enzyme fusions has not progressed without some controversy. For instance, Pettersson and coworkers developed an alternative analysis approach whereby inferences about the rate behavior of a fusion protein were drawn from the altered kinetic properties of the fusion protein itself [42,43], which they argued was more reliable than the analysis methods used previously. These authors found no tenable kinetic evidence to support the earlier claim [41] that the LacZ-GalDH fusion protein catalyses galactonolactone formation from lactose by a mechanism involving channeling of galactose [42]. Thus, care needs to be taken when evaluating the effect of enzyme fusion proteins on metabolic conversions.

In order to mimic the crowding and higher viscosity of the cytoplasm under *in vitro* conditions, crowding agents such as poly(ethylene glycol) have been added to the enzyme reactions to increase the viscosity and slow molecular diffusion [44,45]. For instance, when the *in vitro* viscosity was raised to 8 times that of water, the coupled enzyme activity of a choline dehydrogenase (CDH)/betaine aldehyde dehydrogenase (BADH) fusion protein increased by over 120% while the activity of the free enzyme system increased only 45%. Importantly, the activity of the BADH enzyme remained constant when studied in isolation, despite the increase in the viscosity of solution [60]. These studies suggest that the benefit of enzyme fusions may be even greater *in*

Table 1.1: Engineered Pathway Colocalization and Degree of Synergistic Effect (DSE)

Pathway	Method	Enzymes (Organism)	Assay Environ-ment[1]	DSE[2]	Ref.
Glycerol Production	Fusion	GPD1, GPP2 (*S. cerevisiae*)	*In vivo* (*E. coli*)	2	[8]
			In vitro	2	
Poly (hydroxybutyrate) (PHB) production	Fusion	PhaA, PhaB (*R. eutropha*)	*In vivo* (*E. coli*)	0.2	[46]
			In vivo (*A. thaliana*)	0.7	
Starch hydrolysis	Fusion	CelY, CelZ (*C. stercorarium*)	*In vitro*	1.9-3.5	[47]
	Fusion	α-amylase, glucoamylase (*A. shirousamii*)	*In vitro*	2.5	[48]
Glucose phosphotransferase	Fusion	$IICB^{Glc}$, IIA^{Glc}, HPR, I (*E. coli*)	*In vitro*	3-4	[49]
Trehalose synthesis	Fusion	TPS, TPP (*E. coli*)	*In vitro*	3.5-4.0	[50]
				1.0-1.3	[51]
		BvMTS, BvMTH (*B. helvolum*)			
	Fusion	TDFE, TFE (*S. solfataricus* MT4)	*In vitro*	2.6	[52]
	Fusion	BA (*C. thermo-sulfurogenes*), TS (*T. thermophilus*)	*In vitro*	1.2-1.4	[53]
Formaldehyde fixation	Fusion	HPS, PHI (*M. gastri* MB19)	*In vitro*	2	[54]
NADH recycling	Fusion	LDH (*B. stearothermophilus*), GalDH (*P. fluorescens*)	*In vitro*	2	[44]
Ferulic acid release	Fusion	FAEA, XYNB (*A. niger*)	*In vitro*	1.3-2.0	[55]
Cellulose hydrolysis	Scaffoldin	CelA, CelG (*C. cellulolyticum*)	*In vitro*	2.1-2.6	[56]
				2	[57]
Mevalonate	Protein Scaffold	AtoB (*E. coli*), HMGS, HMGR (*S. cerevisiae*)	*In vivo (E. coli)*	77	[58]

Table 1.1: Engineered Pathway Colocalization and Degree of Synergistic Effect (DSE) cont.

Pathway	Method	Enzymes (Organism)	Assay Environ-ment[1]	DSE[2]	Ref.
Glucaric Acid	Protein Scaffold	Ino1 (*S. cerevisiae*), MIOX (*M. musculus*), Udh (*P. syringae*)	*In vivo (E. coli)*	3	[58]
			In vivo (E. coli)	5	[59]

[1]For *in vivo* studies, host organism is given in parenthesis.
[2]DSE is measured as the fold improvement in the enzyme compartment activity or flux compared with the equimolar free enzyme system, where a value of 1.0 means no difference.

vivo, as the molecular crowding of the cell may further enhance the advantages seen *in vitro*. Interestingly, *in vitro* studies on purified fusion proteins typically report a lower specific activity, as low as 20% of the individual enzyme activities that make up the fusion [50,53,61], yet despite this the bifunctional enzyme often outperforms its free enzyme counterpart due to the effect of channeled enzyme kinetics [8].

While the kinetic properties of bifunctional and multifunctional enzymes offer several advantages that might be useful in metabolically engineered cells, most of these systems have not been systematically evaluated *in vivo.* In fact, there have only been a few reported cases exploring the utility of enzyme fusions for enhancing multi-step metabolic reactions inside living cells. One noteworthy example was reported by Soucaille and coworkers who sought to engineer a *Saccharomyces cerevisiae* glycerol pathway in *E. coli* via a traditional strain engineering approach [8] (see Table 1.1). *E. coli* cells transformed with a plasmid for the separate expression of the *S. cerevisiase* glycerol pathway enzymes GPD1 and GPP2 were grown in a chemostat and after several days a high glycerol producer was isolated. Serendipitously, the evolved strain carried a plasmid in which a deletion event had occurred between GPD1 and GPP2, creating an in-frame fusion protein containing both glycerol-3-P dehydrogenase and glycerol-3-P phosphatase activities. Fed-batch cultivation of *E. coli* cells expressing

this fusion exhibited a 2-fold improvement in glycerol biosynthesis. Consistent with this intracellular enhancement, the purified fusion displayed measurable kinetic improvements including a 7-fold reduced transient time, 5-fold reduced metabolite intermediate concentration, and a 2-fold increased production rate when studied under steady-state kinetics, which the authors ascribed to partial glycerol-3-P channeling between the two active sites [8]. This rather remarkable result provides support for the hypothesis that multifunctional enzymes in nature arose from gene fusions and that spatial organization of recombinant metabolic pathways can be intentionally designed to exhibit similar synergistic activity.

Along similar lines, Snell and coworkers explored enzyme fusions for enhancing bacterial and plant-based biosynthesis of polyhydroxybutyrate (PHB) granules by expressing a fusion between the *Ralstonia eutropha* thiolase (PhaA) and reductase (PhaB) enzymes in *E. coli* and separately in *Arabidopsis* [46]. A PhaA-PhaB fusion protein library, in which the linker sequence length and composition were randomized, was screened in *E. coli* and the highest polymer-producing clone had significantly reduced enzymes activities, indicating that a fusion was not well tolerated between these naturally homotetrameric enzymes (see Table 1.1). When the construct was transformed into *Arabidopsis*, the fusion performed better relative to the *E. coli* system but still underperformed relative to the free enzymes (see Table 1.1) due to problems with folding and solubility that adversely affected the enzyme activity [46].

The above examples underscore some of the key challenges associated with enzyme fusions. First, the fact that many metabolic enzymes are multimeric presents an assembly conundrum (Figure 1.3). That is, while a fusion of two normally monomeric enzymes is straightforward. Fusion of two normally multimeric enzymes requires subunit assembly for activity. These higher-ordered structures can interfere with the non-covalent interactions that are required for the activity of individual

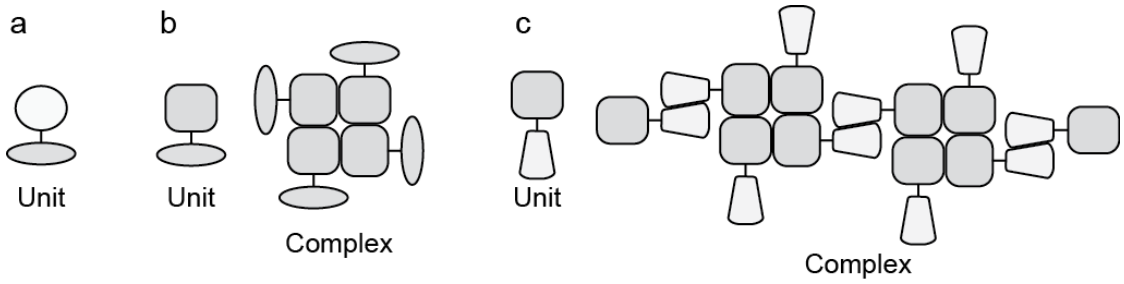


Figure 1.3. Complexes of multimeric fusion proteins. Tertiary and quaternary structures of a protein fusion between (A) two monomers, (B) a monomer and a tetramer, (C) a dimer and a tetramer, forming a multimeric protein aggregate in this last case (adapted from [62]).

enzyme subunits and can lead to a partial protein network that may or may not exhibit enzymatic activity. Along these lines, the LacZ-GalDH fusion yielded two major forms, consisting of four and six subunits, but other forms could also be identified and were characterized as aggregated material [41]. Second, folding of large multidomain proteins is typically inefficient in bacteria [63,64], thus numerous enzyme fusions may misfold or aggregate when expressed in bacteria. Indeed, this was the case for the LacZ-GalK fusion, which formed insoluble aggregates following *in vivo* expression [40].

Post-Translational Assembly. As an alternative to protein fusions, recent work has explored the use of post-translational assembly mechanisms for linking individual enzymes that then act coordinately on sequential reaction steps. The first-documented example is based on natural cellulosomes [56] (Figure 1.2C), complex macromolecular assembles that efficiently degrade cellulose and plant cell walls. The engineered complexes are composed of two main parts derived from natural cellulosomes: (1) the scaffoldin, containing a cellulose-binding domain (CBD) and several cohesin binding domains on a single polypeptide; and (2) the enzyme-dockerin fusion that binds to the scaffoldin post-translationally by means of a high-affinity dockerin-cohesin interaction [39,57,65]. Using the CBDs, cohesins, and dockerins as building blocks, Fierobe and colleagues created designer cellulosomes in which

selected enzymes were incorporated in specific locations within a multicomponent complex by means of high-affinity cohesin-dockerin interactions [39]. Using this strategy, the *Clostridium cellulolyticum* cellulosomal family-5 CelA and family-48 CelF cellulases were each fused to dockerins and the resulting fusions were assembled precisely and by design onto a scaffoldin backbone. Compared with the mixture of free cellulases, the resultant cellulosome chimeras exhibited enhanced synergistic action on crystalline cellulose. These same authors went on to construct a library of 75 different chimeric cellulosomes and identified complexes that conferred a 1.1-7.2-fold enhancement over free enzymes in the ability to degrade cellulose and a 2-fold enhancement resulted from the synergy when sequential enzymes were brought into close proximity on the cellulose substrate [57] (Table 1.1). Within this framework, a variety of geometries and enzyme arrangements have been explored with varying amounts of success [65], indicating that geometry may play an important role in enzyme colocalization, at least in the case of recombinant cellulosomes.

Recently, Dueber *et al*. [58] created intracellular protein scaffolds (pipelines) to which successive pathway enzymes can dock by means of small peptide ligands. The scaffolds contained multiple highly specific and modular docking domains from metazoan genomes: the GTPase binding domain (GBD) from the actin polymerization switch NWASP, the SH3 domain from the adaptor protein CRK, and the PDZ domain from the adaptor protein syntrophin. Three metabolic enzymes for mevalonate production—*E. coli* AtoB and *Saccharomyces cerevisiae* HMGS and HMGR—were targeted to the scaffold using specific ligands corresponding to each docking domain, a strategy reminiscent of the modular domain framework recently used to build ultrasensitive signaling switches [66].

The advantage of this strategy for enzyme assembly is that these ligands are much smaller than the dockerin domains employed above and in Figure 1.2C. By

noncovalently tethering the enzymes to the synthetic scaffold, Dueber *et al*. [58] showed that mevalonate yields could be increased by as much as 77-fold compared with the unscaffolded pathway, depending on both the number of interaction domain repeats and the domain orientation in the scaffold architecture. The same programmable scaffolds were then used to enhance the production of a structurally unrelated compound, glucaric acid, indicating that this approach might be extended to other metabolic pathways.

It appears that post-translational assembly of individual enzymes may overcome the issues of solubility, folding, and macromolecular complex assembly, and it has been successfully applied *in vivo* and *in vitro*. However, the fundamental reasons behind this increased yield of mevalonate remain mysterious, as it is difficult to decouple protein stability from activity, and the stability of the protein scaffold remains an important question. Regardless, the use of synthetic enzyme scaffolds can clearly improve metabolic performance and provides a powerful new method that can be used alongside conventional methods. In the long term, the possibility of integrating several of these strategies simultaneously to balance pathway flux, eliminate metabolic bottlenecks, reduce cell stress, and prevent accumulation of unwanted intermediates or by-products is limited only by one's imagination and should clear the way to produce commercially viable levels of diverse metabolic products such as biofuels, specialty chemicals and therapeutics.

Recent developments engineering enzyme colocalization

In this work, we present a progression of engineering enzyme colocalization in the cytoplasm of *E. coli* toward *R*-1,2-propanediol (1,2-PD) first, and later *D*-1,2,4-butanetriol (1,2,4-BT) and mevalonate. First, we develop a kinetic model of 1,2-PD diffusion-reaction kinetics using a 3-D stochastic algorithm that allows us to study coupled reaction kinetics under enzyme compartmentalization. Second, we

systematically evaluate the efficacy of fusion proteins with an array of proteins and peptide linkers within the 1,2-PD biosynthetic pathway. This represents an application of the current state of the art and provides motivation in moving towards more complex systems. Then, we apply well-known protein-interacting domains (PIDs) to assemble enzymes, allowing proteins to fold independently. The novel application of a leucine zipper to drive interactions of metabolic enzymes (both towards 1,2-PD and 1,2,4-BT) shows the benefits of post-translational assembly and highlights the importance of balancing flux within the context of a channeled biosynthetic pathway. We briefly test the protein scaffolds employed by Dueber *et al.* [58]. Finally we develop stable and modular DNA scaffolds with the use of DNA binding domains. In this manner, we generated a robust set of DNA scaffolds for tethering metabolic enzymes in a way that both balances flux as well as maximizes site occupancy, allowing us to improve the yield towards both 1,2-PD and mevalonate.

CHAPTER 2

STOCHASTIC REACTION-DIFFUSION SIMULATION OF ENZYME COMPARTMENTALIZATION REVEALS IMPROVED CATALYTIC EFFICIENCY FOR A SYNTHETIC METABOLIC PATHWAY[1]

Introduction

Enzyme-to-enzyme channeling (a.k.a. metabolic channeling) and compartmentalization of biochemical reaction modules have long been suggested as key components of cellular metabolism [67]. Metabolic channeling is a process whereby enzymes and their active sites are arranged in a manner that enables cells to effectively synthesize specific products without metabolic interference, diffusion limitations, or inhibition from intermediate steps [7,21,68]. The first direct evidence of this behavior came in the 1940's when David Green isolated all of the Krebs tricarboxylic acid cycle (TCA) enzymes as an aggregated system, which he termed a 'multienzyme complex' [69]. It is now apparent that metabolic channels are ubiquitous in nature and have evolved into several different physical forms including, for instance, tunneling between enzyme active sites that is clearly visible in the X-ray crystal structure [10], direct coupling between sequential pathway enzymes via protein-protein interactions [21], and compartmentalizing specific enzymes into small volumes within the cell in the form of subcellular organelles [70,71]. Likewise, engineered channels comprised of simple, coupled enzyme systems can reportedly improve the overall kinetic properties of the enzymes [72,73,74,75] as well as the

[1] Adapted with permission from: Conrado RJ, Mansell TJ, Varner JD, DeLisa MP (2007) Stochastic reaction-diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway. Metab Eng 4:355-63.

metabolic flux through a pathway while reducing concentrations of intermediate species [11].

Computer simulations of coupled enzyme networks have been instrumental in revealing the kinetic ramifications of enzyme channeling and compartmentalization [76,77,78,79,80,81]. The earliest models that explored channeling were deterministic and reported conflicting results both with respect to the ability of channels to exert greater metabolic control and to maintain low levels of pathway intermediates [76,77,78,81]; however, more recent approaches to modeling and simulation of metabolic networks unequivocally support the argument for channeling and its effect on cellular metabolism [79,80]. Specifically, Degenring et al. used a discrete event, multi-level model to analyze the interrelation between structural and functional characteristics of the static channeling that occurs in the tryptophan synthase complex. Maher et al. used Metabolic Control Analysis (MCA) and carefully accounted for enzyme compartmentalization to show that channeling in the urea cycle was required for the model output to reflect what was found in experiments. These two recent models highlight the importance of stochastic and spatial models, respectively, for metabolic pathway analysis and provide powerful mathematical tools to facilitate the analysis of channeling in biochemical pathways like the urea cycle or in multi-enzyme complexes like tryptophan synthase.

Despite the greater accuracy that these and other modeling efforts have provided, there is little reliable simulation analysis on the kinetic benefits offered by coupled enzyme systems. Earlier modeling attempts have fallen short of answering the evolutionary benefit of channeling for several reasons: (1) coupled enzyme systems rely on the spatial organization of sequential enzyme steps; the reaction-diffusion relationship is crucial to channeling kinetics; and (2) the concentration of intracellular species is often low, especially with biological enzymes or channeled intermediates,

which deterministic models cannot capture; and while late modeling efforts have taken a more apt approach, they have not studied the enzyme kinetics thoroughly. For these reasons, a spatial and stochastic modeling approach could provide a more realistic portrayal of the system dynamics and capture the competing events within a cell. Accordingly, we have explored the effect of channeling within the context of biologically relevant pathways and parameters since complex interactions in metabolic networks can influence output in a non-trivial way and a realistic biological context is necessary to understand the effect of enzyme compartmentalization on specific pathways.

Here, we present a spatial stochastic model of *E. coli* central carbon metabolism, including a description of energy metabolism within the cell, that relies on the combined recent advances in simulation methods, detailed kinetic biological models, and relevant biological pathways. Using this model, we simulate the addition of a pathway for the production of 1,2-PD to explore the kinetic consequences of enzyme compartmentalization on a synthetic pathway in the context of *E. coli* central carbon metabolism. Since bacterial production of 1,2-PD is hindered by the accumulation of the bactericidal intermediate methylglyoxal [82,83,84] and by the formation of undesired side products from competing pathways, it represents a unique opportunity for artificial enzyme compartmentalization as a means to increase production titers via the reduction of toxicity and side product formation. Furthermore, the techniques presented here provide a generic framework for simultaneously analyzing spatial and stochastic events in metabolism and, thus, should be useful in evaluating the contribution of enzyme compartmentalization within any recombinant pathway of interest.

Materials and methods

Model structure. The stochastic model of central carbon and energy metabolism of *E. coli* was described by the Markov reaction diffusion master equation. The intracellular volume of *E. coli* was spatially divided into homogeneous three-dimensional subvolumes. The spatially discretized model was solved using the Next Subvolume Method (NSM) [85]. NSM provides an efficient implementation of the Gillespie Direct Method [86] coupled to a description of Stokes-Einstein diffusion between subvolumes. NSM allowed the central carbon metabolism of *E. coli* to be recast as a spatially heterogeneous reaction system where potential reaction events were defined with respect to a given subvolume and diffusion events occurred freely between connected subvolumes.

Two computational controls were conducted to validate our implementation of the NSM. First, simulations of the spatially homogenous stochastic reaction system

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C \xrightarrow{k_3} D \quad (1)$$

were compared to the analytical solution of the deterministic system. As shown in Figure 2.1A, the NSM solution is consistent with the true solution indicating the NSM was correctly implemented. A second computational control was conducted to gauge the impact of the kinetics upon the solution predicted by the NSM. The Markov reaction-diffusion master equation, typical of stochastic simulations, is most often associated with mass action propensity functions. Mass action kinetics require two parameters to describe the association/disassociation of the enzyme substrate complex and a third to describe the catalytic rate of the enzyme. Enzymes, however, are more commonly characterized in terms of their Michaelis-Menten parameters: a reaction velocity, k_{cat}, and a saturation constant, K_{m}. We have employed Michaelis-Menten rate forms, which are based upon a pseudo steady-state assumption, in our subsequent

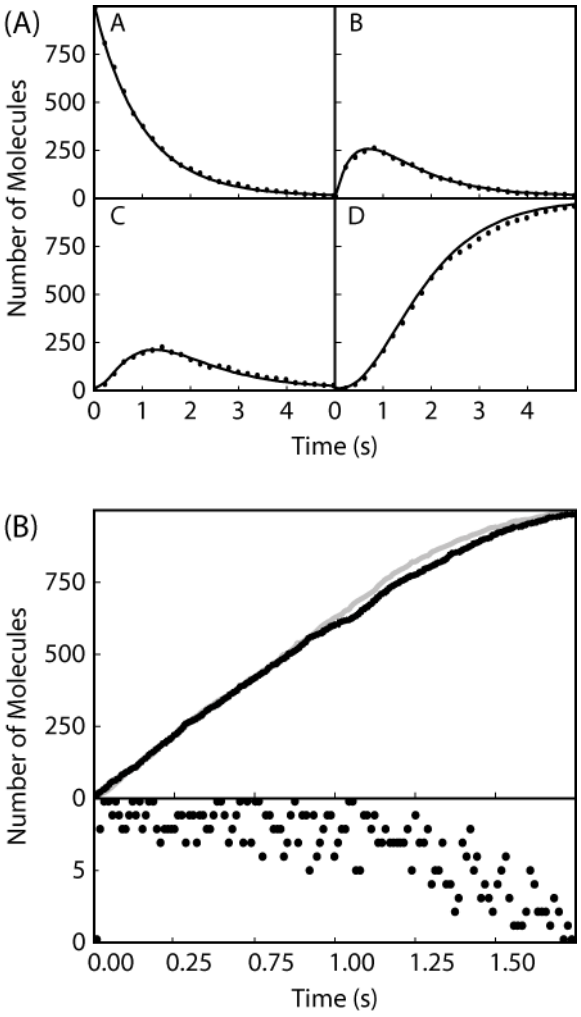


Figure 2.1. Comparison between (A) species concentration (A -> B -> C -> D) using stochastic simulation (points) and analytical solution (lines) where $k_1=1s^{-1}$, $k_{-1}=2s^{-1}$, $k_2=2s^{-1}$; and (B) product formation using mass action (grey) and Michaelis-Menten kinetics (black) where $k_1=20s^{-1}$, $k_{-1}=1000s^{-1}$, $k_2=1000s^{-1}$, $K_M=100$, $k_{cat}=1000s^{-1}$. Bound enzyme-substrate complexes in mass action kinetics are indicated by points.

Compared simulations of the enzymatic conversion of substrate (S) to product (P) using mass action kinetics (Equation 2) and Michaelis-Menten kinetics (Equation 3):

$$E+S \underset{}{\overset{k_1,k_{-1}}{\longleftrightarrow}} ES \xrightarrow{k_2} P+E, \quad r_1 = k_1[E][S],\ r_{-1} = k_{-1}[ES],\ r_2 = k_2[ES] \qquad (2)$$

$$E+S \xrightarrow{k_{overall}} P+E, \quad r_{overall} = k_{cat}[E][S]/(K_m+[S]) \qquad (3)$$

development. To test the impact of the pseudo steady-state assumption upon the NSM simulations, where enzyme was denoted by 'E', enzyme-substrate complex 'ES', $k_{cat}=k_2$, and $K_m=(k_2+k_{-1})/k_1$ (Figure 2.1B). The close similarity between Michaelis-Menten and mass action simulations justified the use of Michaelis-Menten kinetics

and demonstrated that the pseudo steady-state assumption was accurate over a stochastic regime.

Kinetic rate expressions, with their parameters (Table A.1 and A.2), followed the form of traditional Michaelis-Menten equations for a variety of reaction types: forward and reversible, two-substrate, and various activation and deactivation schemes. In all reactions, a freely diffusive enzyme was additionally considered as an input to the rate expressions.

***E. coli* reaction network model.** The *E. coli* reaction network, fluxes, rate expressions, and kinetic parameters used in this study (Figure 2.2) were based upon the earlier work of Reuss and coworkers [87]. The Reuss network described glucose uptake via the phosphotranferase system, glycolysis, and the pentose phosphate pathway. We modified the Reuss network to include the TCA cycle, oxidative phosphorylation, and other cellular energetic reactions [88]. Because we are exploring the network response on time scales shorter than gene expression, we have assumed the enzyme number density to be constant for all of the cytoplasmic reactions [89]. The engineered *R*-1,2-PD metabolic pathway was modeled as three sequential steps that consume dihydroxyacetone phosphate (DHAP) and reducing power to form *R*-1,2-PD [82]. In this study, *E. coli* methylglyoxal synthase (MGS; EC 4.2.99.11) converts DHAP to methylglyoxal (MG), which is then converted to *R*-lactaldehyde by a glycerol dehydrogenase (DhaD; EC 1.1.1.6) from *Klebsiella pneumoniae.* The intermediate, *R*-lactaldehyde is then converted to *R*-1,2-PD via an alcohol dehydrogenase (AdhI; EC 1.1.1.1) from *Saccharomyces cerevisiae.* As indicated in Figure 2.2, MGS releases inorganic phosphate, while DhaD and AdhI consume NADH.

Enzyme compartmentalization. Enzyme compartmentalization confines an enzyme within a small volume or localizes an enzyme within the cell. In this study,

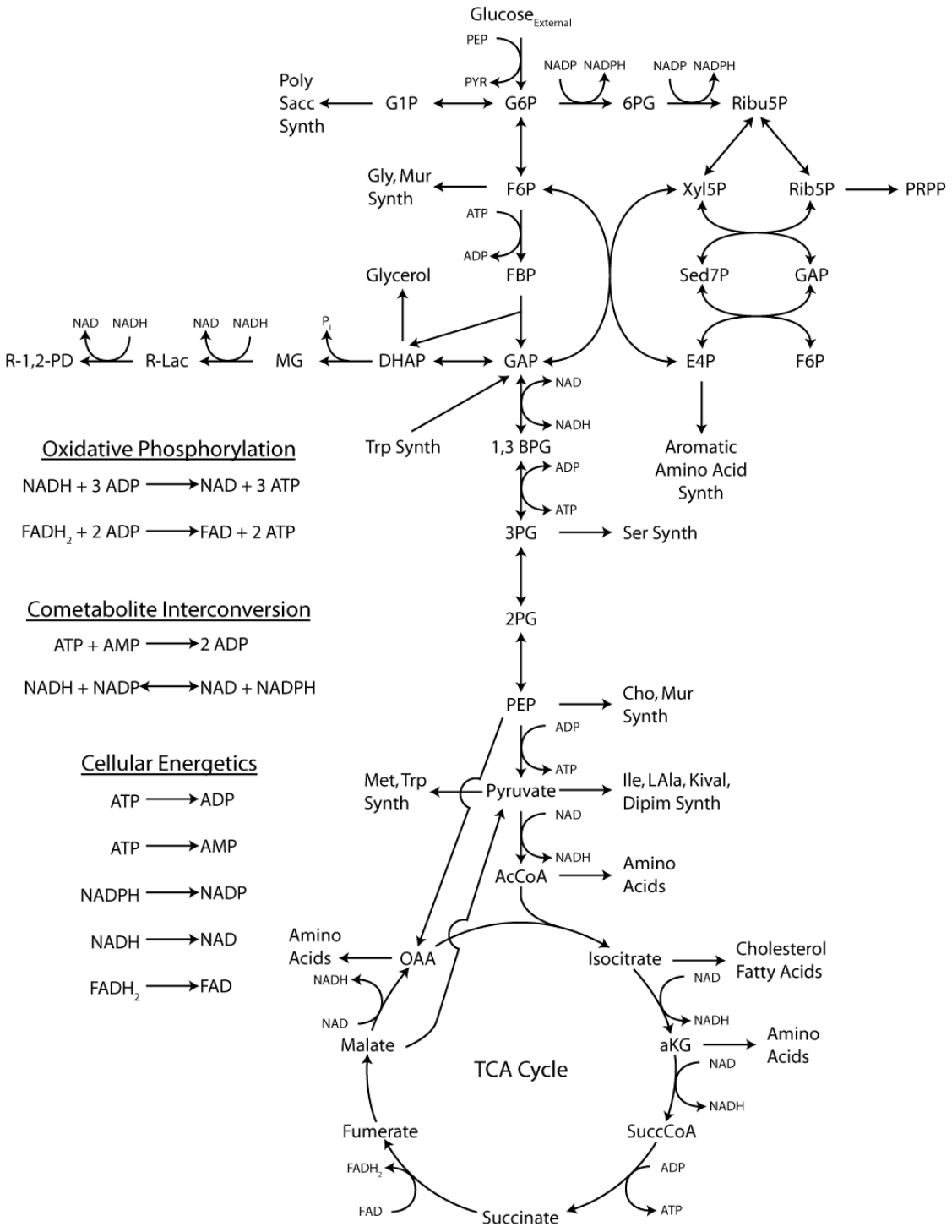


Figure 2.2. Metabolic model of glycolysis, pentose phosphate pathway and TCA cycle.

compartmentalized enzymes were prevented from freely diffusing, whereas diffusion of substrates was not constrained. All compartmentalized enzymes were confined to

the lower left hand subvolume, while all metabolites were allowed to freely diffuse (Figure 2.3). Non-compartmentalized enzymes were active and free to diffuse into any of the subvolumes. Under the NSM, the model of a single bacterium consisted of several well-mixed subvolumes in which species and enzyme populations were tracked. Analysis of enzyme compartmentalization proceeded by increasing the total number of subvolumes and concomitantly decreasing the size of the active enzyme compartment, as the total cell volume remained constant across simulations. Thus, as the subvolume dimensions were decreased, the local concentration of the confined enzymes increased, while their total cellular concentration remained constant.

Evaluation of Michaelis-Menten kinetic parameters. The Michaelis-Menten parameters k_{cat} and K_m for AdhI were extracted from simulation data: first, the total extracellular rate of 1,2-PD production (V) was estimated from simulation data by dividing the difference in product concentration at two time points by the respective time difference (Figure 2.4A). An example Michaelis-Menten plot (V with respect to substrate concentration) generated by the simulation is shown (Figure 2.4B). Next, the value of V and the total cellular substrate concentration were used to create Eadie-Hofstee plots (Figure 2.4C), from which V_{max} and K_m were determined. Finally, k_{cat} was estimated by dividing V_{max} by the total cellular enzyme concentration. The AdhI enzyme kinetics simulations were carried out assuming saturating amounts of NADH, thereby decoupling *R*-lactaldehyde conversion from reducing power limitations.

Results and Discussion

Comparison with experimental data. Stochastic simulations of the response of *E. coli* to a glucose pulse were compared to time-resolved intracellular metabolite measurements [87] obtained from fed-batch cultures of *E. coli* (Figure 2.5). The stochastic simulation tested for both one and two compartments captured the dynamic characteristics of the glucose pulse and described the observed trends for most

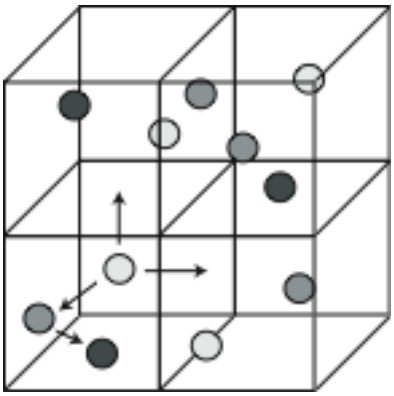


Figure 2.3. A sample discretization in a 2x2x1 subvolume as an example of enzyme compartmentalization. Here species (filled circle) can diffuse into any subvolume but, in this example, reactions are restricted to occur in the lower left subvolume.

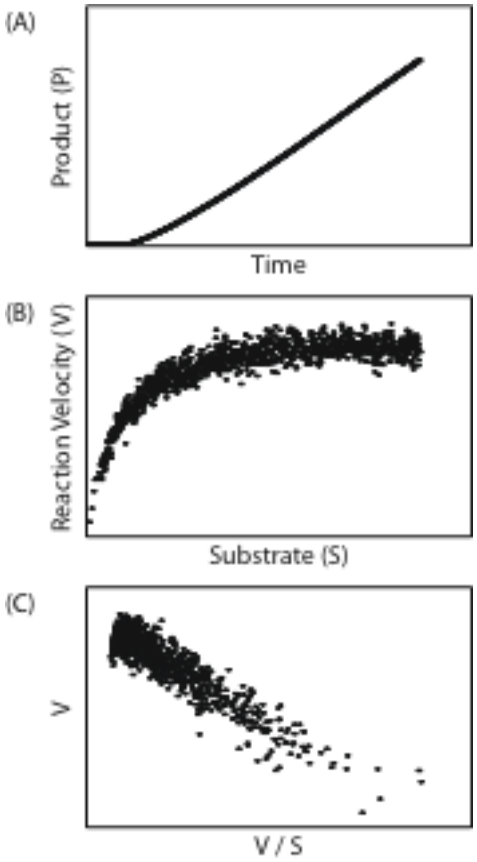


Figure 2.4. Calculation of Michaelis-Menten parameters from simulation data. Slope of product concentration with respect to time (A) is calculated to form a Michaelis-Menten plot (B) and fitted on an Eadie-Hofstee plot (C) to extract K_m and k_{cat}.

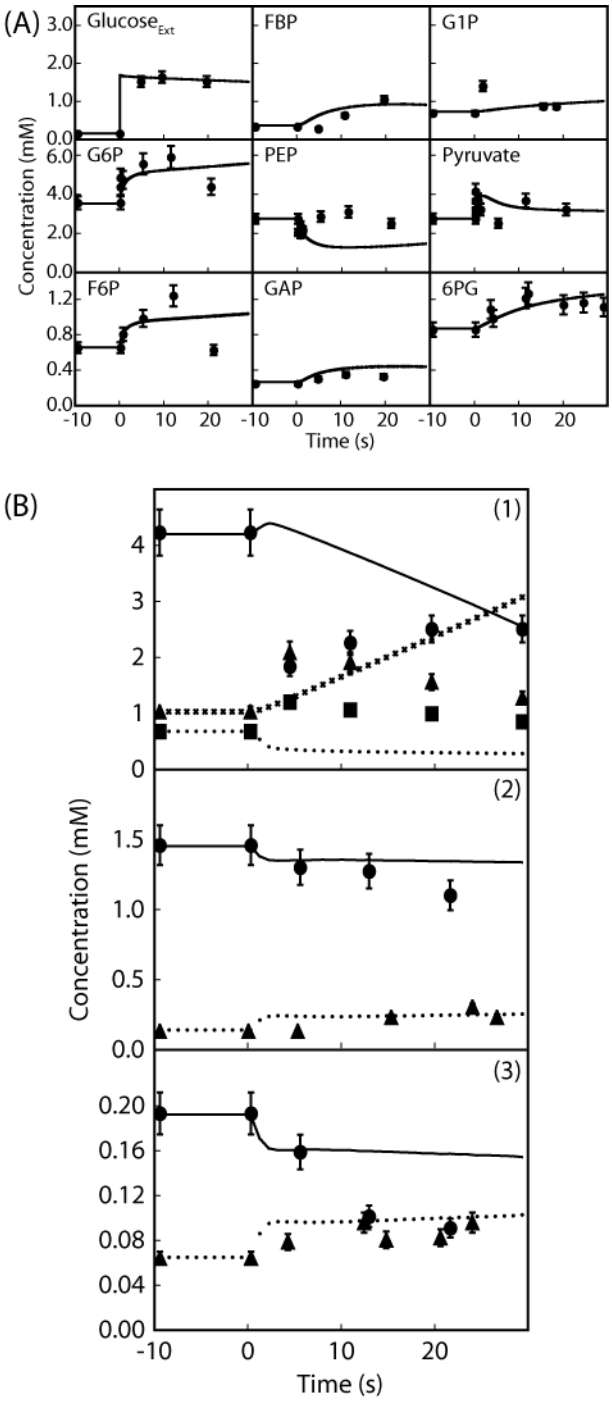
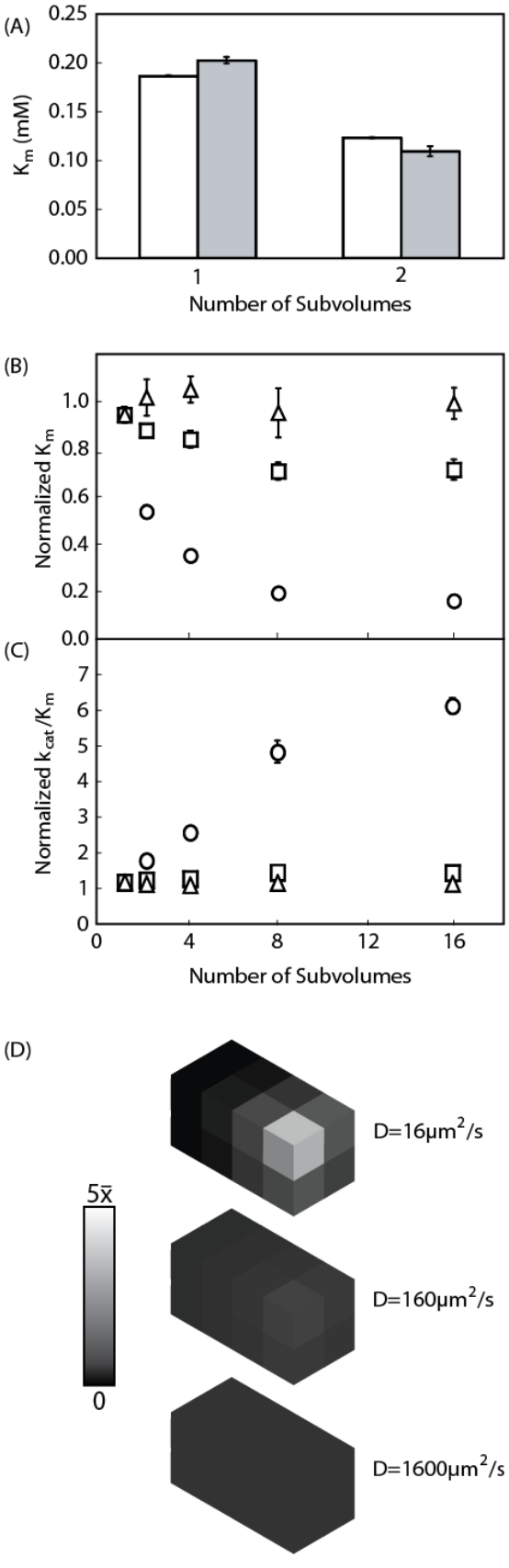


Figure 2.5. Comparison between experimental data from Chassagnole et al. [87] (points) and stochastic simulation (lines) after a glucose pulse at time zero. (A) Solid line: simulation; filled circle: experimental data. (B1) Solid line: ATP; bold dotted line: AMP; dotted line: ADP; filled circle: ATP; filled triangle: AMP; filled square: ADP. (B2) Solid line: NADP; dotted line: NADPH; filled circle: NADP; filled triangle: NADPH. (B3) Solid line: NAD; dotted line: NADH; filled circle: NAD; filled triangle: NADH.

metabolites over a range of extracellular conditions (Figure 2.5A) using the single parameter set. Differences between model predictions and experimental observations were seen for FBP and PEP, with PEP showing significant deviation. However, unlike earlier models which used data driven descriptions of energy and reducing power, the energy and reducing state of the network was computed as part of our model. The profiles for most of the energetic co-metabolites, especially NAD derivatives (Figure 2.5B), were correctly predicted by the model. Notable deviations from experimental data were observed primarily for ATP metabolism, which highlights the complexity of dynamically modeling central metabolism. The stochastic simulation results validate our description of *E. coli* central carbon metabolism and therefore lend the model to further engineering efforts, namely enzyme compartmentalization for an engineered metabolic pathway.

Modeling and simulation of enzyme compartmentalization. To probe the impact of compartmentalization on 1,2-PD production, the recombinant 1,2-PD pathway enzymes were introduced into the full model of *E. coli* central carbon metabolism. Initial kinetic studies using the full model augmented with the compartmentalized 1,2-PD pathway, for one and two subvolumes, revealed a sharp reduction in the Michaelis-Menten saturation constant, K_m, for the channeled substrate (Figure 2.6). That is, as the number of subvolumes increased, the saturation constant decreased. Here, a single subvolume system has well mixed enzymes, while in a two subvolume system, the recombinant enzyme reactions can occur in only one of the two subvolumes and substrates must diffuse into the compartmentalized subvolume before those enzymes may act on them. Although this study considers only the impact on kinetic properties for the third pathway enzyme, yeast AdhI, we expect this to be a general effect that is present at each step where the substrate is channeled. We next attempted to simulate the effect of further compartmentalization (i.e., subvolumes > 2) and the effects of

Figure 2.6. (A) Plot of K_m for yeast AdhI comparison full (white) and simplified (grey) kinetic models for a substrate diffusivity of 16 mm^2/s. Normalized (B) K_m and (C) k_{cat} / K_m for compartmentalized enzymes versus freely diffusive enzymes plotted against the number of subvolumes. (open circle) D = 16 mm^2/s; (open squard) D = 160 mm^2/s; (open triangle) D = 1600 mm^2/s. (D) Concentration gradient of *R*-lactaldehyde under enzyme compartmentalization with 16 subvolumes at each diffusivity tested. All data points represent the average of three simulation runs with error bars representing the standard error. Geometries for the various numbers of subvolumes are as follows: 1-1x1x1, 2-2x1x1, 4-2x2x1, 8-2x2x2, 16-4x2x2.


(A)
0.25
0.20
0.15
0.10
0.05
0.00
K_m (mM)
1
2
Number of Subvolumes
(B)
1.0
0.8
0.6
0.4
0.2
0.0
Normalized K_m
(C)
7
6
5
4
3
2
1
0
Normalized k_{cat}/K_m
0
4
8
12
16
Number of Subvolumes
(D)
5x̄
0
D=16μm^2/s
D=160μm^2/s
D=1600μm^2/s

different reaction-diffusion regimes; however, simulation times became prohibitively expensive. To circumvent the issue of computational burden, we reduced the model to only the recombinantly expressed enzymes of the 1,2-PD pathway. The reduced model allowed us to explore the impact of compartmentalization on the utilization of carbon in detail. Since the reduced model neglects the dependence upon NADH, we assumed saturating reducing power in order to decouple the Michaelis-Menten kinetics and analyze the impact of compartmentalization on the channeled carbon substrate. Simulation times for the reduced model decreased by 100-fold; yet despite the drastic simplification of the model, the predictions were within 15% of the predictions made using the full *E. coli* reaction network (Figure 2.6A, compare white vs. grey bars). That is, both the full reaction network and the simplified recombinant pathway yielded similar results regarding the ability of enzyme compartmentalization to reduce the Michaelis-Menten saturation constant. Based on this observation, we were highly confident that the simplified model was a reliable surrogate for the full *E. coli* reaction network.

Enzyme compartmentalization increases catalytic efficiency. Using the simplified model, we observed that compartmentalization of the 1,2-PD pathway led to a decrease in the overall K_m when compared to the evenly distributed enzyme case (Figure 2.6B). The decrease in K_m was magnified by increasing the number of subvolumes i.e., increasing the degree of enzyme compartmentalization. The decrease in K_m was accompanied by a small (≤10%) decrease in k_{cat} and, as a result, an overall increase in catalytic efficiency (k_{cat} / K_m) (Figure 2.6C) was observed. The increase in catalytic efficiency was dependent upon substrate diffusivity. Substrate diffusivity, *D*, for a wide range of biologically relevant species where *D* for ions > small molecules > enzymes was simulated; the effect of compartmentalization increased with decreasing substrate diffusivity (Figure 2.6). The impact of compartmentalization was greatest at

the lowest substrate diffusivity (D = 16 mm^2/s) [5,90], but was still present even at small-molecule-like diffusion rates (D = 160 mm^2/s) [91,92]. Values for K_m were shown to decrease up to 6-fold relative to the case of non-compartmentalized enzymes for slow diffusing substrates, and up to 20% for normal diffusing small molecules. There was little to no effect on the K_m for faster diffusing ion-like species (D = 1600 mm^2/s). It was expected that enzyme compartmentalization would impact the apparent K_m but not the apparent k_{cat}, since the relationship between substrate concentration and reaction velocity is intrinsic to K_m itself.

To explain the inverse relationship observed between the apparent K_m and enzyme compartmentalization we formulated the working hypothesis that the decrease in K_m was due to an increase in the local concentration of substrate in the compartmentalized subvolume. In other words, increased enzyme concentration in the active compartment leads to a local buildup of pathway intermediates. In the regime below K_m, where many engineered metabolic reactions lie, reaction velocity is directly proportional to the substrate concentration. Therefore, a local buildup of intermediates would be expected to increase the overall enzymatic reaction velocity. To test our working hypothesis, we plotted the concentration of the pathway intermediate *R*-lactaldehyde (the substrate of yeast AdhI) in each subvolume when the total cellular concentration of *R*-lactaldehyde was equal to one quarter of its K_m (Figure 2.6D). A significant concentration gradient of *R*-lactaldehyde was observed in the slowest diffusing case (D = 16 mm^2/s), with the localized intermediate concentrated in the compartmentalized subvolume by as much as 4-fold over the average cellular levels. However, this gradient was much smaller for faster diffusing cases ($D \geq$ 160 mm^2/s), demonstrating that this effect depended greatly on substrate diffusivity and verifying our initial hypothesis. Increasing the diffusivity of the substrates effectively rendered the cell well-mixed for the pathway intermediates and tended to negate the apparent

effect of local concentration, thus reducing the impact on K_m. Thus, the kinetic improvements demonstrated by enzyme concentration are likely to be much more evident for cases of low substrate diffusivity [75,93].

To our knowledge, the above results represent the first attempt to model substrate channeling for a synthetic metabolic pathway, comprised of enzymes that are expressed recombinantly in *E. coli*. Our finding that enzyme compartmentalization can yield improved catalytic efficiency agrees well with experiments where engineered fusion proteins were created to mimic enzyme compartmentalization [73]. Though not directly studied here, the kinetic benefits that we observed could have important consequences for product formation. Channeling has shown to have a dramatic impact on rate without a significant impact on the measured kinetic properties of enzymes involved [36,72,75]. Further, naturally occurring bacterial compartments exist for the degradation of small molecules like 1,2-PD [70] and therefore point to the important benefits provided by channeling even when the impact on kinetic parameters may be slight. This suggests that the kinetic shift seen during our simulations is significant from an evolutionary standpoint and is even more important for larger molecules. Examples of channeling for the production of slow diffusing species is seen in aromatic amino acid synthesis [10,18,94], plant phenylpropanoid metabolism [29] and the bacterial PKS system [95]. In addition to the results presented in this work, we hypothesize that channeling will contribute further benefits to the bacterial production titers of 1,2-PD. These include: (1) sequestration from competing reaction pathways which are present in the cell but not considered in our model; (2) reduction of kinetic constraints on the cell by lowered intermediate concentrations; and (3) rapid conversion of MG to *R*-lactaldehyde, which is significant because MG is bactericidal at sub mM concentrations [3]. For these reasons, the production of 1,2-PD is an appropriate target for the creation of synthetic metabolic channels.

In considering the marked improvements made possible by enzyme compartmentalization, there are several factors one must take into account before embarking on the engineering of a channeled recombinant pathway. Of primary consideration, channeling has evolved in cases where the native enzymes have poor efficiency or low substrate specificity, other constraints being absent. Those enzymes that have a low K_m or high catalytic activity never allow substrate to build up and are already operating at near maximal efficiency. Traditional recombinant metabolic pathways, however, suffer from poor translation efficiency, are present in highly variable copy numbers, and have low turnover for new substrates, which this study directly considers. Another key constraint is the limited availability of energy and reducing power in the cell. Although not considered here, further enzyme compartmentalization could be envisioned to produce both ATP and NADH. Enzyme compartmentalization presents an excellent opportunity for improving the kinetic properties for the channeled carbon substrate. Ongoing research is directed towards a more comprehensive approach for understanding the full impact of substrate channeling, including the changes in system flux, intermediate concentrations, and rate of product formation.

Acknowledgments

We thank Professors Claude Cohen and Abe Stroock for helpful discussions. This work was supported by the Office of Naval Research under a Young Investigator Award (to MPD) and also ONR Grant # N00014-07-1-0027 (to MPD) and ONR Grant # N00014-06-1-0293 (to JDV).

CHAPTER 3

ENGINEERING ENZYME COLOCALIZATION FOR PRODUCTION OF 1,2-PROPANEDIOL AND 1,2,4-BUTANETRIOL IN *E. COLI*

Introduction

Enzyme-to-enzyme channeling (a.k.a. metabolic channeling) and compartmentalization of biochemical reaction modules have long been suggested as key components of cellular metabolism [67]. Metabolic channeling is a process whereby enzymes and their active sites are arranged in a manner that enables cells to effectively synthesize specific products without metabolic interference, diffusion limitations, or flux imbalances [7,21,68]. The first direct evidence of this behavior came in the 1940's when David Green isolated all of the Krebs tricarboxylic acid cycle (TCA) enzymes as an aggregated system, which he termed a 'multienzyme complex' [69]. It is now apparent that metabolic channels are ubiquitous in nature and have evolved into several different physical forms including, for instance, tunneling between enzyme active sites that is clearly visible in the X-ray crystal structure [10], direct coupling between sequential pathway enzymes via protein-protein interactions [21], and compartmentalizing specific enzymes into small volumes within the cell in the form of subcellular organelles [70,71]. Likewise, engineered channels comprised of simple, coupled enzyme systems can reportedly improve the overall kinetic properties of the enzymes [72,73,74,75] as well as the metabolic flux through a pathway while reducing concentrations of intermediate species [11].

The intentional engineering of metabolic pathways is a rapidly maturing field and much attention has already been given to cellular productivity and diverted flux. Indeed, a number of strategies have been reported for optimizing cellular metabolism

including, for example, stabilizing mRNA transcript levels, balancing protein expression levels, evolving enzyme activity, and redirecting cellular metabolism [34]. These successes notwithstanding, very little attention has been given to the spatial organization of the recombinant metabolic pathway enzymes *in vivo*. Yet, based on numerous examples from nature, it is clear that enzyme organization is a key design variable to be considered when engineering metabolism.

Engineering efforts have made great strides at mimicking these natural systems, moving from immobilized enzymes [35] to fusion proteins [40] to post-translational assembly of metabolic enzymes along a scaffold chimera [39]. However, there have been few intentional and successful attempts to channel metabolic substrates within the cytoplasm of *E. coli* [96]. The high number of demonstrated improvements *in vitro* motivates the need for a generic toolkit for colocalizing a sequential series of enzymes for cellular production. Early engineering attempts have fallen short for several reasons: (1) fusion chimeras often fail to fold properly within the cell and are rapidly degraded; and (2) colocalized enzymes that do fold, frequently suffer from a reduction in activity that masks any potential benefits of channeling. While genetic techniques have improved dramatically since immobilized enzyme technology, these have largely not been applied to the creation of active compartmentalized enzyme pathways. Accordingly, we have explored the parameters involving colocalized enzymes and mimicked techniques thus far only successful *in vitro*, to create increasingly elegant methods to colocalize enzymes in the cytoplasm of *E. coli*.

Here, we present a study of enzyme colocalization for the production of both 1,2-propanediol and 1,2,4-butanetriol from three and four enzymatic steps that branch from *E. coli* central carbon metabolism. Since bacterial production of 1,2-PD is hindered by the accumulation of the bactericidal intermediate methylglyoxal

[82,83,84] and by the formation of undesired side products from competing pathways, it represents a unique opportunity for artificial enzyme compartmentalization as a means to increase production titers via the reduction of toxicity and side product formation. Furthermore, 1,2,4-BT production suffers from a myriad of competing reactions, both known and unknown, that dramatically limit its production *in vivo*. The techniques presented here provide a generic framework for creating colocalized enzymes and should be useful in evaluating the contribution of enzyme compartmentalization within any recombinant pathway of interest.

Materials and methods

1,2-PD biosynthesis pathway. Plasmids were constructed using molecular biology techniques according to standard practices [97]. The three sequential genes for encoding 1,2-PD biosynthesis, (1) *mgsA*, (2) *dkgA*, *dkgB*, *fucO*, or *ydjG* and (3) *gldA* were PCR amplified from *E. coli* MG1655 genomic DNA. These genes were then cloned into pBAD18 [98] in the order shown, using unique restriction sites within the multiple cloning site, resulting in a polycistron for 1,2-PD synthesis. The first gene was placed between NheI and XbaI, the second gene between XbaI and SphI, and the last gene between SphI and HindIII. The same strong ribosomal binding site was placed directly upstream of each gene in the polycistron and an NdeI site at each start codon. At times, the second biosynthetic gene was appended by an 3' HA epitope tag and *gldA* was appeneded by a 3' FLAG epitope tag.

1,2,4-butanetriol biosynthetic pathway. Plasmids were constructed using molecular biology techniques according to standard practices [97]. The four sequential genes for encoding 1,2,4-BT biosynthesis, (1) *xdh* (*C. crescentus*), (2) *yagF* or *yjhG* (*E. coli*), (3) *mdlC* (*P. putida*) and (3) *adhP* (*E. coli*) were PCR amplified and cloned into pBAD18 [98] in the order shown, using unique restriction sites within the multiple cloning site, resulting in a polycistron for 1,2,4-BT synthesis. *xdh* was PCR amplified

from pWN9.046A and placed between EcoRI and BglII restriction sites. *yjhG* was PCR amplified from pWN8.022 and *yagF* was PCR amplified from *E. coli* MG1655 genomic DNA and placed between BglII and XbaI restriction sites. *mdlC* was PCR amplified from pWN5.238A and placed between XbaI and HindIII restriction sites. *adhP* was PCR amplified from pML6.166 and placed between HindIII and NcoI restriction sites. The same strong ribosomal binding site was placed directly upstream of each gene in the polycistron and an NdeI site at each start codon.

Fusion proteins. The fusion proteins were generated using codon optimized peptide linkers {Chang, 2005 #220}, with the linkers flanked by SpeI and NotI restriction sites. The restriction sites flanking the biosynthesis genes were maintained. As a control for fusion proteins for 1,2-PD, MalE was cloned into the C-terminal position of the fusion with MgsA, allowing the gene in the second position to be freely expressed. At times, fusions were appended with a 3' HA epitope tag.

Post-translational interacting domains (PIDs). PIDs were generated using GCN4-GCN4 [99], cJun-cFos, and SH3-SH3lig [66] interacting pairs. GCN4, cJun were PCR amplified [99] and fused C-terminally to mgsA using an L16 polylinker [64]. GCN4 and cFos fragments were PCR amplified [99] and fused N-terminally to either dkgA or fucO using an L16 peptide linker [64]. The strong binding SH3 ligand (SH3ligS) were fused C-terminally to mgsA using an L16 linker [64]. The SH3 domain was PCR amplified from pET19(a): B.3.1 [66] and fused N-terminally to either DkgA or FucO using an L16 linker [64]. As controls to reduce the binding affinity of these ligands, GCN4 (7P14P) [99], cJun (L3V) and cFos (L2V) [99], and the weak binding SH3 ligand (SH3ligW) [66] were employed in the same context as the strong binding counterpart.

Protein Scaffolds. The SH3-PDZ-GBD protein scaffold was PCR amplified from pET19(a): B.3.1 [66] and fused to the C-terminus to MalE (*E. coli*) with an L16

peptide linker [64] and additionally contained a C-terminal HA epitope tag. This construct was cloned into the MCS of both pBAD18Cm and pBAD33Cm [98]. Scaffold ligands [66] were fused N-terminally to the 1,2-PD biosynthetic enzymes with an L16 peptide linker [64]: SH3 ligand fused to mgsA, PDZ ligand fused to dkgA, and GBD ligand fused to gldA.

Bacterial strains, media, and growth conditions. *E. coli* MC4100 [F^- *araD139*Δ(*argF-lac*)*U169 rspL150 relA1 flbB5301 fruA25 deoC1 ptsF25*] or its derivatives were used for 1,2-PD production unless otherwise noted. 1,2-PD fermentations were followed as described [82] with the following exceptions. L-arabinose was added to 0.2% (w/v) at the time of inoculation to induce gene expression. All fermentations were run at 37°C, either at 200 or 250 rpm, with tubes held vertically or at a 45 degree angle. The 10-mL fermentation mixtures were inoculated to an optical density of 0.05 (typically 1:50 to 1:100 dilution) with the overnight culture. The optical density (OD) was measured at 600 nm with a Thermo Spectronic BioMate 3 model spectrophotometer.

E. coli BW25113 [Δ*(araD-araB)567*, Δ*lacZ4787*(::rrnB-3), lambda$^-$, *rph-1*, Δ*(rhaD-rhaB)568*, *hsdR514*] or its derivatives were used for 1,2,4-BT production unless otherwise noted. BW25113 derivatives were constructed by P1 transduction from the Keio Collection [100]. 1,2,4-BT fermentations were followed as described [101,102,103] with the following exceptions. Overnight cultures were grown from a single colony into 5 mL of LB media. All fermentations were run in 20 mL of rich media (20 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 3.75 g/L K_2HPO_4) in 250-mL baffled flasks at 33°C, 250 rpm for 24 hours. D-xylose was added to 2% (w/v) at 4 hours. L-arabinose was added to 0.2% (w/v) to induce gene expression at 4 hours. The 20-mL fermentation mixtures were inoculated to an optical density of 0.05 (typically 1:50 to 1:100 dilution) from the overnight culture. Antibiotics were provided at the

following concentrations: ampicillin, 100 μg/mL; chloramphenicol, 20 μg/mL.

HPLC analysis. Products in the fermentation media were quantified as described [82] with the following exceptions. Compounds were measured with a Waters Breeze system. The mobile phase was a 0.03 N sulfuric acid solution, with a flow rate of 0.45 ml per min, and the column temperature was 50°C. All samples were filtered through 0.22-μm-pore-size membranes prior to analysis.

Western blot analysis. MC4100 cultures were grown overnight in LB media, inoculated as above, lysed via sonication after either 9 or 24 hours, and normalized on the amount of total soluble cell protein. Insoluble lysates were isolated by centrifugation at 16,000 g for 10 minutes. Pellets were washed twice with 1 mL of with Tris-Cl (50 mM) and EDTA (1 mM) pH 8.0. The pellet is then resuspended in 1x PBS and 2% SDS and then boiled for 10 minutes. The insoluble fraction is collected as the supernatant following a 10 minute centrifugation at 16,000 g. Western blotting of these soluble and insoluble lysates was performed with anti-HA primary antibody (1:1000, v/v; Sigma) and anti-rabbit secondary antibody (1:2500, v/v; Promega) on SDS-PAGE gels, typically 10 or 12%.

Results

Enzyme Compartmentalization Towards Production of 1,2-Propanediol. In order to construct a synthetic metabolic pathway from DHAP to *R*-1,2-PD, we first needed to identify three sequential pathway enzymes (Figure 3.1). Since natural reaction pathways consume *R* and *S*-lactaldehyde for reduction to lactate, a route through the acetol intermediate would suffer from less diverted flux. Along this path, the first and last enzymatic steps were well-defined, specifically the synthesis of methylglyoxal by methylgloxal synthase (MgsA, *E. coli*) and reduction of acetol by glycerol dehydrogenase (GldA, *E. coli*) [82]. However, the reduction of the intermediate, MG to acetol, can be performed by a number of NADH or NADPH dependent enzymes

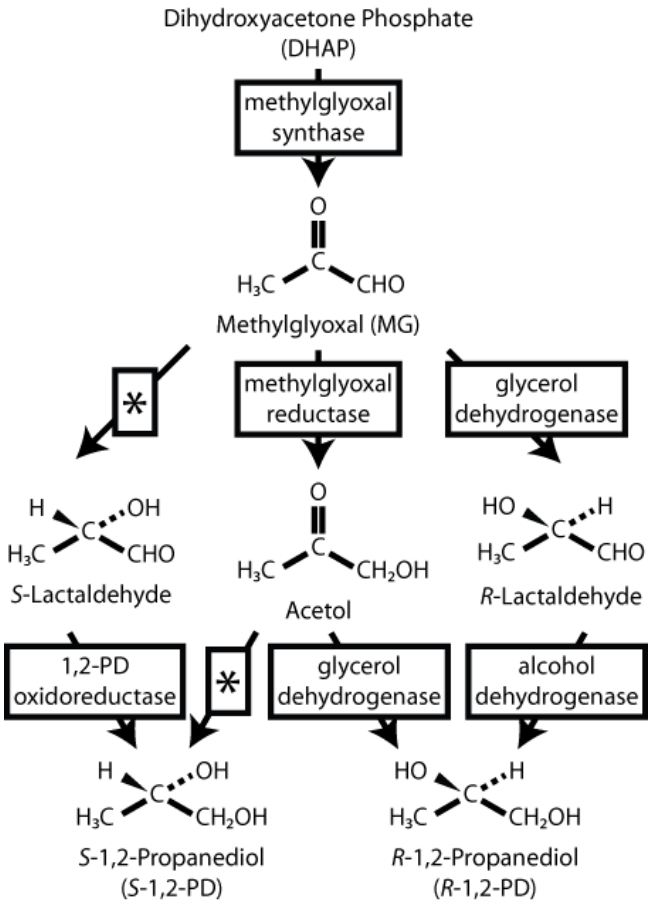


Figure 3.1. Metabolic pathway of 1,2-PD production from dihydroxyacetone phosphate in *E. coli*.

[104], most of which have not been tested *in vivo*. To test for activity on methylglyoxal *in vivo*, a high copy plasmid expressing MgsA, one of the following NADH (FucO, YdjG, *E. coli*) or NADPH dependent enzymes (DkgA, DkgB, *E. coli*), and GldA [104] was constructed. As shown in Figure 3.2, all genetic constructs produced significant levels of 1,2-PD, which allowed enzyme selected based upon our design criteria. In moving forward, DkgA was selected for primary focus because its kinetic parameters are well defined and it is active as a monomer, which would aid in the design and analysis of metabolic channels.

Fusion Proteins Towards 1,2-PD. Since fusion proteins are a well-studied example [96], and are a simple test for analyzing the impact of enzyme compartmentalization,

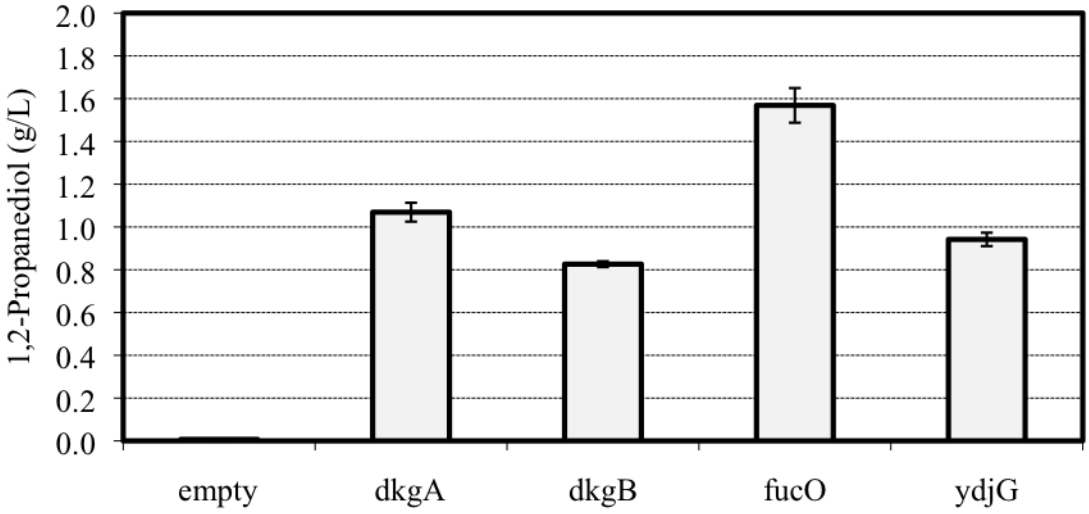


Figure 3.2. The effect of methyglyoxal reductase activity on 1,2-PD production. The above enzymes were overexpressed along with MgsA and GldA. Columns represent the average of three experiments with error bars showing one standard error.

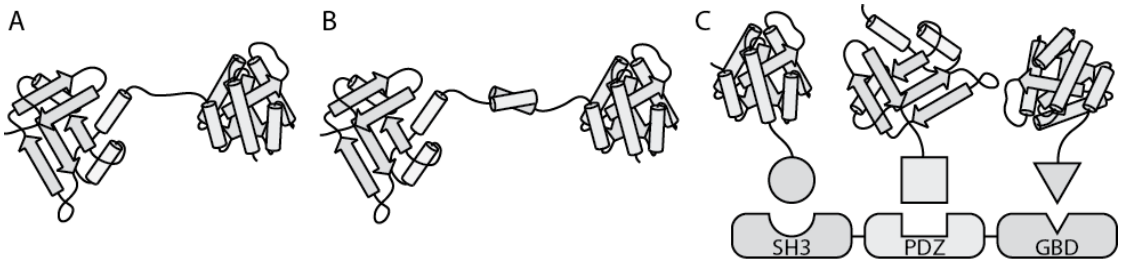


Figure 3.3. Engineering enzyme compartmentalization. (A) Fusion proteins connected by a peptide linker, (B) Chimeric proteins fused to leucine zippers connected by post-translational assembly of PIDs, (C) Post-translational assembly of enzyme complexes where the SH3-PDZ-GBD fusion represents the three specific docking sites for each enzyme-ligand fusion.

engineering our pathway enzymes by genetic fusion was a natural first step, to covalently attach two enzymes by encoding both active sites on the same polypeptide (Figure 3.3A). In this way, the active sites could be brought into close proximity and therefore allow for the proposed synergistic activity found in natural systems [68]. In order to systematically test this approach, the two design elements considered were (1) the linker length and composition and (2) the order of the protein fusions within the context of the entire pathway. Despite a large body of knowledge regarding natural

linkers within multidomain proteins [105], there is little consensus on the length and composition of synthetic linkers when connecting two proteins that are normally not fused [96].

In order to find a suitable candidate, we systematically applied a set of 6 well-studied linker peptides between fusion proteins in *E. coli*, which varied in composition and length, from 5 to 37 amino acids [64]. This approach allowed us to test role of the linker in terms of its effect on protein stability, as well as the effect of linker length on the degree of synergistic coupling, i.e., if the kinetic benefit disappeared with longer linkers due to the increased diffusional distance. This resulted in several in-frame genetic fusions between the first two enzymatic steps of the pathway, mgsA and dkgA, with the third enzyme freely expressed in a bicistronic message. As shown in Figure 3.4, a modest 30% increase in 1,2-PD production was found when the fusions were compared to the fusion of MgsA to a soluble but inactive protein partner, MalE. MgsA, a hexamer in its active state, is very sensitive to fusions at either its N or C-termini. In order to assess any synergy that occurs between MgsA or DkgA *in vivo*, this control was most appropriate. The modest gain in activity through the pathway is consistent with computational studies on compartmentalization in *E. coli* [12].

These results prompted us to determine whether or not a greater benefit could be seen when fusing only the second two pathway enzymes, thus channeling acetol towards 1,2-PD. Applying the same set of linkers, genetic fusions were created between *dkgA* and *gldA*, and with free *mgsA* encoded bicistronically. Interestingly, this set of fusion constructs (Figure 3.5) had little impact on yield with the best linkers tested, suggesting that no kinetic benefit was observable and/or that these fusions were not well tolerated.

As a final goal of compartmentalizing a set of sequential pathway enzymes, a three-gene fusion was constructed, to determine if a fold improvement could be

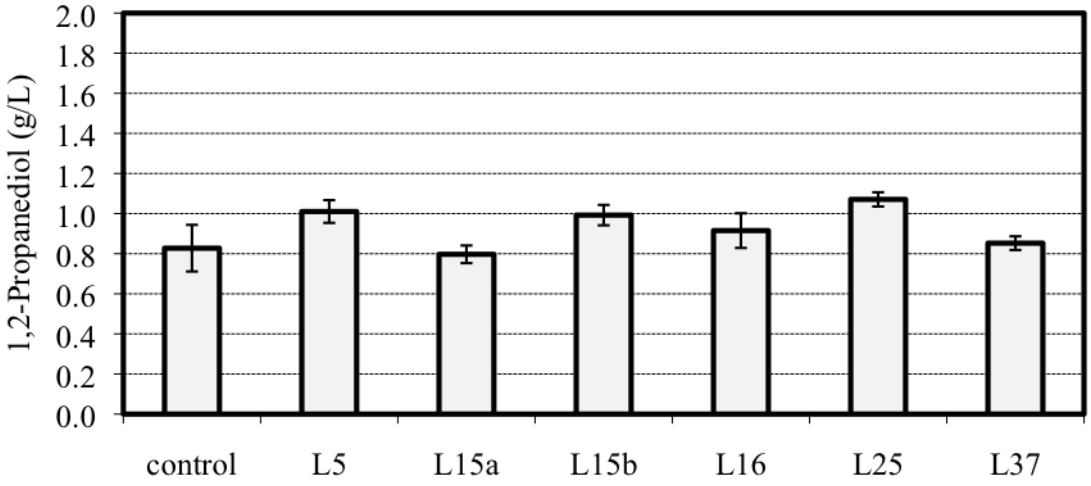


Figure 3.4. Effect of enzyme fusions between of MgsA and DkgA on 1,2-PD, production, while coexpressing GldA. The name of each sample indicates the linker employed with the number denoting the linker length in amino acids. The control sample is an MgsA-MalE enzyme fusion, coexpressing DkgA and GldA. Columns represent the average of three experiments with error bars showing one standard error.

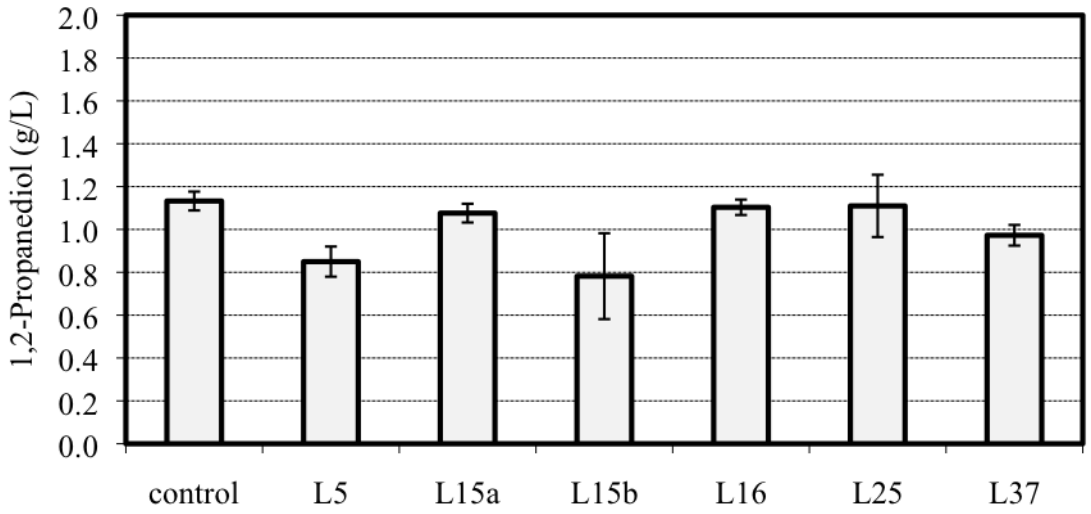


Figure 3.5. Effect of enzyme fusions between of DkgA and GldA on 1,2-PD production, while coexpressing MgsA. The name of each sample indicates the linker employed with the number denoting the linker length in amino acids. The control sample has MgsA, DkgA, and GldA coexpressed without fusion. Columns represent the average of three experiments with error bars showing one standard error.

detected over the *mgsA-dkgA* fusions due to altered kinetics over the free enzymes

(Figure 3.4). Selecting successful linkers in both cases, the L16 linker was used

between the first two pathway enzymes and the L37 linker between the last two enzymes, thus making an MgsA-L16-DkgA-L37-GldA fusion. After being constructed, these were compared to the corresponding free and fusion enzyme systems by their production of 1,2-PD (Figure 3.6). Combining the fusions between all pathway genes did not yield an improvement over the simple fusion of the first two enzymes. This was not altogether surprising since the fusion of the last two enzymes seemed to impair enzyme activity from the decrease in 1,2-PD production (Figure 3.5). In probing for the explanation for these disparate levels of production, a Western Blot was performed to analyze the intracellular protein levels, which might be impacted from these novel fusions. In comparing the protein levels of these selected fusions however, the soluble protein levels are very similar, despite fusions involving MgsA having significant insoluble fractions (Figure 3.7). This result suggests that these fusion proteins have significantly impaired individual activities despite having decent folding characteristics. It is also important to note that the first and last enzymes of the pathway enzymes are multimeric, and the active structures are denoted $MgsA_6$, DkgA, and $GldA_8$. Combining these polypeptides by gene fusion, especially when combining more than two genes as above, can have important consequences in terms of aggregation and achieving active enzyme units (Figure 1.3 and [96]). While aggregation may be a general problem with compartmentalizing multimeric enzymes, it is important to pursue alternative strategies that might allow these to form in a manner that they maintain activity.

Protein Interacting Domains (PIDs) Towards 1,2-PD. In moving beyond protein fusions as a result of the fusion loss in activity when colocalizing two or more, the enzyme active sites should be coupled in a way that would allow proper protein folding, and possibly subunit assembly, before compartmentalizing the pathway enzymes. Similar to the assembly of metabolic channels like the tryptophan synthase

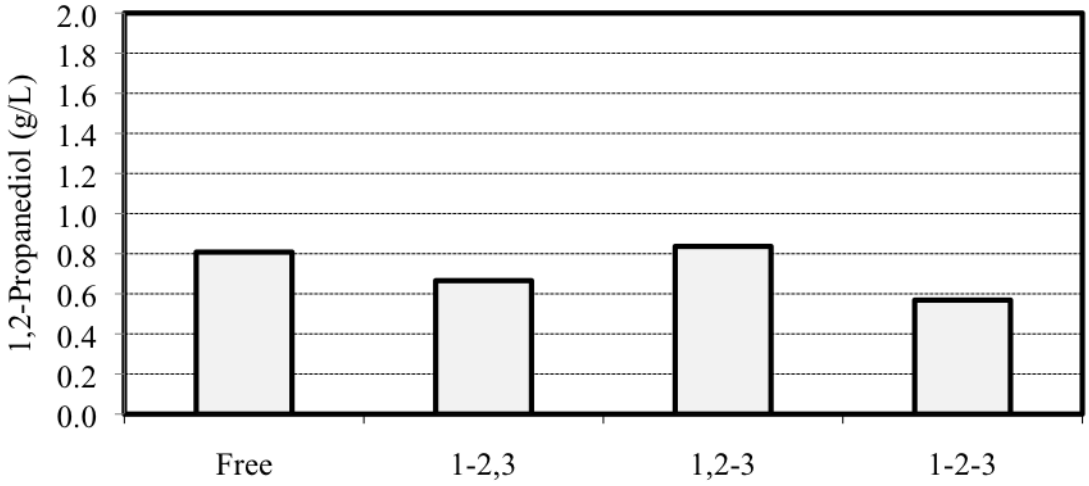


Figure 3.6. Effect of enzyme fusion on 1,2-PD production. Free denotes all enzymes coexpressed without fusion. 1-2,3 denotes a fusion of MgsA and DkgA with GldA freely expressed. 1,2-3 denotes a fusion of DkgA and GldA with MgsA freely expressed. 1-2-3 denotes a tripartite fusion between MgsA, DkgA, and GldA. Columns represent represent a single experiment.

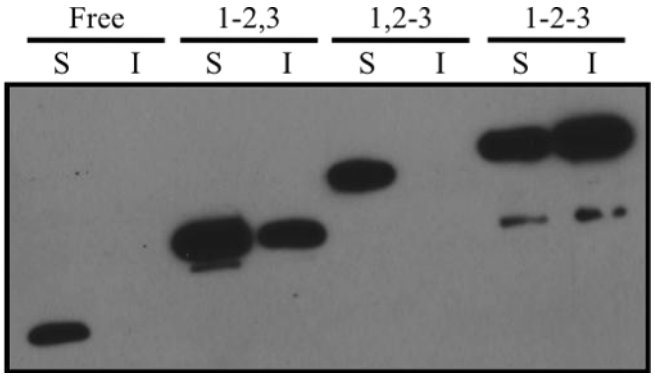


Figure 3.7. Western Blot of soluble (S) and insoluble (I) protein levels on samples of Figure 3.6.

αββα complex [10], a post-translational, non-covalent assembly of sequential pathway enzymes would likely allow for more stable enzyme formation of active multimers. To achieve this, known protein interacting domains were employed, that when fused N- or C-terminally to the pathway enzymes, would bind together and would thus bring the sequential active sites into close proximity (Figure 3.3B). This strategy offers more design flexibility than gene fusion above, because in addition to

design of the linker region, there are also considerations for the binding associations, homo- or heterodimerization, and their affinity in terms of the dissociation constant, K_D. Here our design criteria were domains that: (1) are well expressed in *E. coli*, (2)possess high affinity interactions in the μM to sub-μM range, given that natural proteins can reach 1-10 μM concentrations in *E. coli* [89], (3) are short in length, preferably <50 amino acids, (4) are characterized by highly specific interactions with little cross-reactivity. Based on these standards, we selected three sets of interacting domains for further analysis: GCN4 [106], cJun-cFos [107,108,109,110], and the SH3 domain and associating ligand [66]. These domains were then fused to each of the first two pathway enzymes with the third enzyme freely expressed. The degree of synergistic effect was measured by comparison to a polycistronic free enzyme system, where either the first or second pathway enzyme was fused to a PID. In this way, the domain would similarly impact activity, but not enable interaction. Additionally, the PIDs offer further controls in that in each of these interacting domains, (1) PIDs can be tested individually for their impact on enzyme activity and (2) point mutations can be made to reduce binding affinity. These additional controls should help to elucidate the mechanism of the fold-improvement seen in protein fusions, namely whether the increase was due to enzyme compartmentalization or an increase in protein stability as seen by Western Blot.

These constructs were analyzed as above for their extracellular 1,2-PD production levels following fermentation, shown in Figure 3.8, with each of the interacting domains and the inactivating control. Interestingly for the DkgA, the GCN4 leucine zippers were responsible for the largest increase in production, followed by either cJun/cFos or the SH3/SH3ligand pair. Looking towards the controls for each of these interacting domains, the inactivating mutations to GCN4 [111], had a slight decrease in 1,2-PD levels, and this trend is similar for each of the interacting

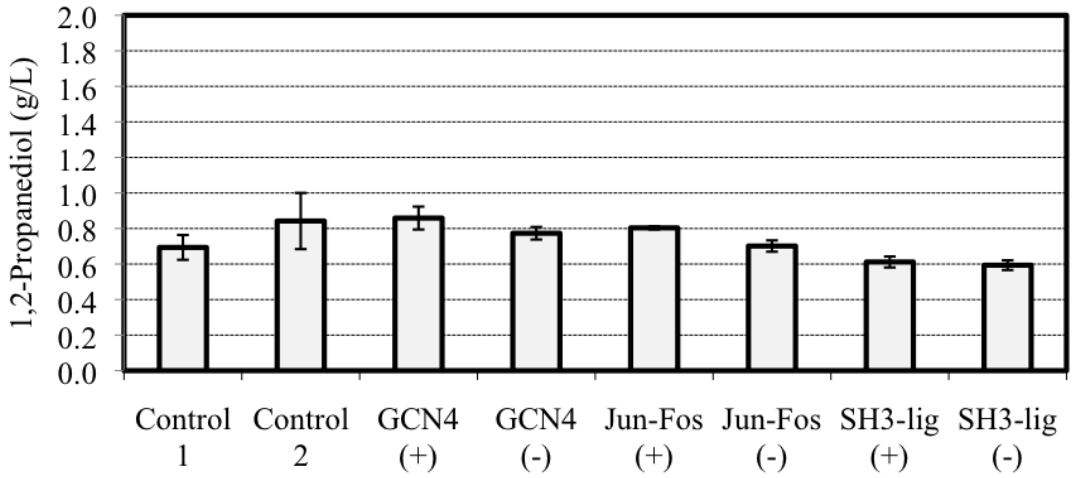


Figure 3.8. Effect of PIDs on both MgsA and DkgA on 1,2-PD production, while coexpressing GldA. The name of each sample indicates the PID pair employed where (+) denotes full interaction strength, and the (-) denotes mutated PIDs to reduce binding affinity. Control 1 has a PID on MgsA only, and control 2 has a PID on DkgA only. Columns represent the average of three experiments with error bars showing one standard error.

domain pairs. While none of the PIDs show a significant increase in production of 1,2-PD over the controls, each slightly outperforms its weakly interacting pair.

Orientation of Fusions and PIDs Towards 1,2-PD. When compared to the metabolic enzymes alone, both fusions and PIDs suffered from loss of activity. As a method to ameliorate this impact, the opposite orientation fusion was constructed, as well as PIDs at either termini of the protein, using the same domains and peptide linkers. As shown in Table 3.1, the fusion orientation matter greatly, both for protein fusions and post-translational interactions and with pathways expressing DkgA and FucO, however, none of the orientations tested showed improvements over the free enzyme system.

Protein Expression Level and Its Impact on Synergy Towards 1,2-PD. Since none of the systems tested to date showed marked improvements in levels of 1,2-PD, one explanation was that the over-expressed enzyme levels within the cell were too highly

Table 3.1. Effect of fusion and PID orientation on 1,2-PD production, while coexpressing GldA. Values given in grams per liter. Data represent a single experiment.

FucO		
Free (+GldA)	MgsA; FucO	1.59
Fusion (+GldA)	MgsA-L16-FucO	1.03
	FucO-L16-MgsA	0.76
PIDs (+GldA)	MgsA-L16-cJun	cJun-L16-MgsA
FucO-L16-cFos	1.21	1.23
cFos-L16-FucO	0.66	0.60
DkgA		
Free (+GldA)	MgsA; DkgA	1.22
Fusion (+GldA)	MgsA-L16-DkgA	0.87
	DkgA-L16-MgsA	0.54
PIDs (+GldA)	MgsA-L16-cJun	cJun-L16-MgsA
DkgA-L16-cFos	0.59	1.00
cFos-L16-DkgA	0.47	0.70

for the benefits of enzyme colocalization to have any impact. To test this, relevant plasmids were transformed into WM2949 cells harboring pLacY(A177C) [112]. This combination allows for leaky transport of arabinose and therefore titratable induction of the PBad promoter, instead of the on/off phenotype observed in MC4100. In this way, a comparison between free and colocalized enzymes could be made at wide range of induction levels, and the corresponding range of enzyme levels (Figure 3.9). Interestingly however, the free enzyme case performed better at all levels of induction, when compared to fusion proteins and PIDs. Even at very low expression levels, neither fusion proteins nor PIDs showed an increase in 1,2-PD production, indicating that no synergy was observable.

Coexpression of Free Enzymes Towards 1,2-PD. Within this work, enzyme colocalization has shown only very modest improvements, and only in very isolated cases. One explanation for this is the assembly conundrum that is present when multimeric enzymes are attached, by fusion of PID (Figure 1.3). The enzymes that participate in this matrix may have reduced activity if any. To circumvent this, the three metabolic enzymes were freely expressed from a separate plasmid and

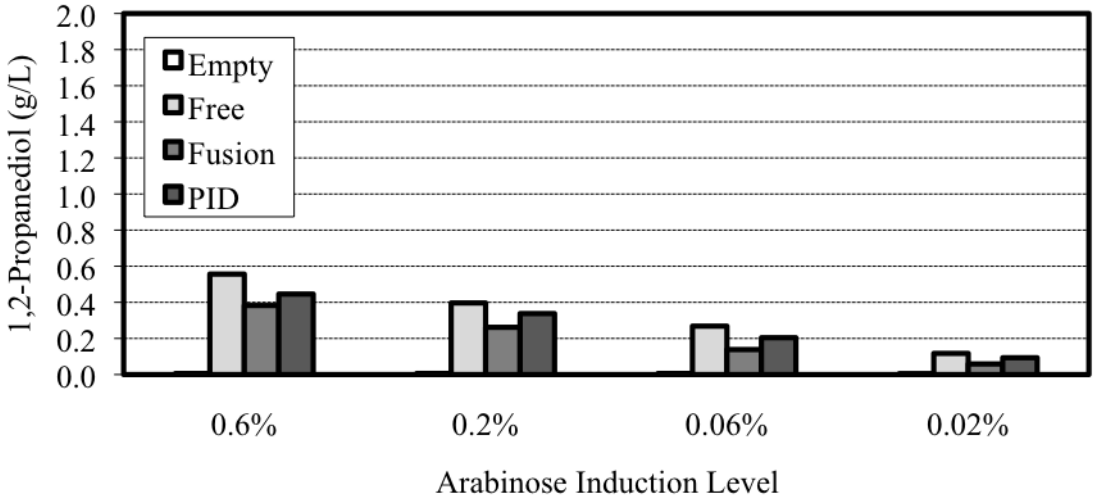


Figure 3.9. Effect of arabinose induction level on 1,2-PD production. Empty: no overexpressed enzymes; Free: MgsA, DkgA, and GldA overexpressed; Fusion: MgsA-L16-DkgA, and GldA overexpressed; and PID: cJun-L16-MgsA, DkgA-L16-cFos, and GldA overexpressed. Columns represent a single experiment.

colocalized enzyme constructs were coexpressed. This effect can be modulated by high expression of colocalized enzymes (from pBAD18) and low expression of free enzymes (pZA31) or vice versa. As show in Figure 3.10 for DkgA and Figure 3.11 for FucO, coexpressing free enzymes from two plasmids at the same time showed very little increase, although coexpressing free enzymes did improve the colocalized enzyme cases to the level of free enzymes. This final results gives further support that no synergy between colocalized enzymes is observable towards 1,2-PD.

Protein Scaffolds Towards 1,2-PD. Neither fusions proteins or interacting domains were very successful in increasing the yields of 1,2-PD for even two-enzyme channels, and this effect became more evident with colocalizing three enzymes. Insolubility of these complexes and loss of activity hinders the ability to benefit from even slight synergy between colocalized enzymes. Further, flux imbalances within the pathway have been overlooked and may be responsible for the lack of improvement. The move away from fusions and PIDs is motivated both by the membrane bound metabolons

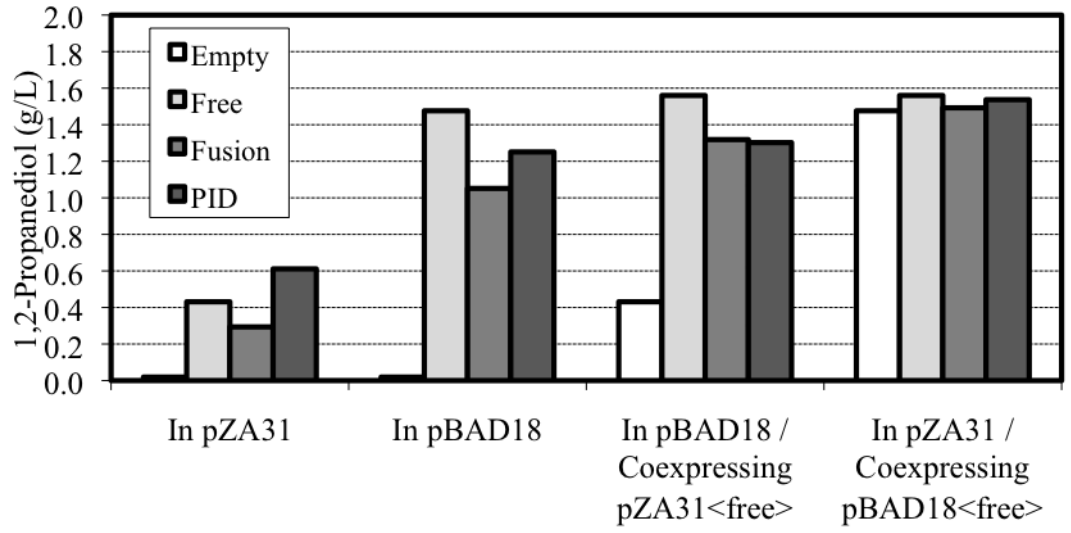


Figure 3.10. Effect of free enzyme coexpression on 1,2-PD production. Empty: no overexpressed enzymes; Free: MgsA, DkgA, and GldA overexpressed; Fusion: MgsA-L16-DkgA, and GldA overexpressed; and PID: cJun-L16-MgsA, DkgA-L16-cFos, and GldA overexpressed. Columns represent a single experiment.

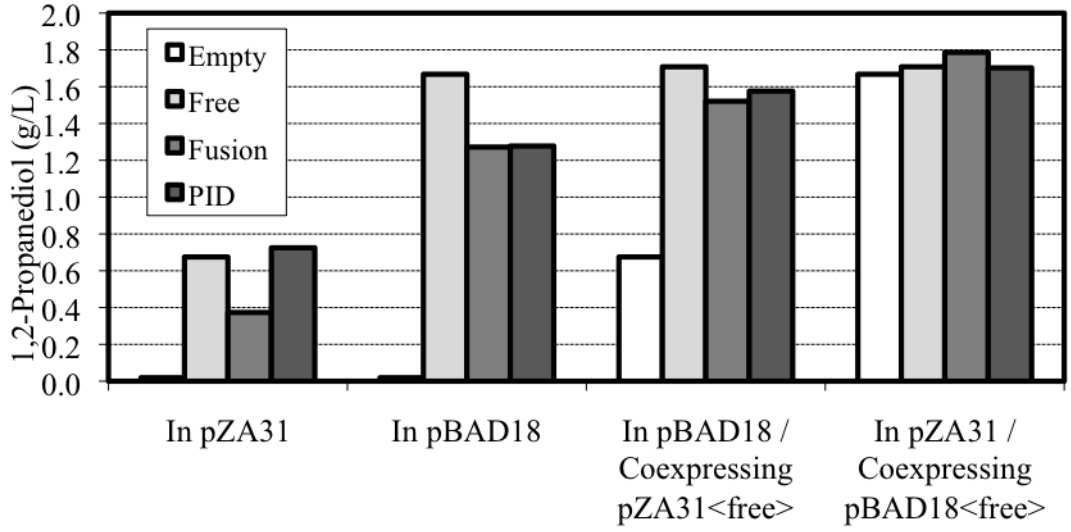


Figure 3.11. Effect of free enzyme coexpression on 1,2-PD production. Empty: no overexpressed enzymes; Free: MgsA, FucO, and GldA overexpressed; Fusion: MgsA-L16-FucO, and GldA overexpressed; and PID: cJun-L16-MgsA, FucO-L16-cFos, and GldA overexpressed. Columns represent a single experiment.

found in nature [21] and the chimeric cellulosomes tested *in vitro* [56]. Using the basic structure of these chimeric cellulosomes, a fusion protein consisting of several

Table 3.2. Effect of colocalizing enzymes to a protein scaffold on 1,2-PD production. Free: MgsA, DkgA, and GldA overexpressed; and Ligands: MgsA-L16-SH3ligand, DkgA-L16-PDZligand, and GldA-L16-GBDligand. Scaffold (MBP-L16-SH3-PDZ-GBD) was overexpressed in either pBAD18Cm or pBAD33Cm, where empty has no scaffold. Values given in grams per liter. Data represent a single experiment.

	pBAD18Cm		pBAD33Cm	
	Scaffold	Empty	Scaffold	Empty
Free	1.04	1.10	0.92	0.82
Ligands	0.22	0.17	0.19	0.23

docking domains connected by peptide linkers was constructed (Figure 3.3C). The pathway enzymes can then colocalize on this scaffold by several short ligands, which specifically target each of these docking domains. As the number and order of these scaffold domains was varied, so to would the metabolic enzymes, creating a mechanism to balance flux.

Since the docking and ligand domains from these engineered cellulosomes are rather large and were not tested *in vivo*, a protein scaffold recently developed to tightly control input/output signals was employed as a proof of concept [66]. Using the SH3-PDZ-GBD scaffold construct from this work, the corresponding strong ligands were fused to each of the individual pathway enzymes. When this scaffold was tested for levels of 1,2-PD, compared to same vector lacking the scaffold, no significant increase in 1,2-PD resulted at either high (pBAD18) or low (pBAD33) expression of the scaffold (Table 3.2). Looking into this further, the expression of the scaffold was then analyzed by Western Blot, along with its codon-optimized counterpart, with and without fusion to MalE. Not surprisingly the scaffold expresses very poorly in *E. coli* (Figure 3.12), is not detectable without fusion to MalE, a known solubility enhancer [113]. Due to the significant amount of protein degradation and lack of scaffold solubility, this approach was abandoned.

Enzyme Compartmentalization towards Production of 1,2,4-Butanetriol. In order to validate our approach towards colocalizing enzymes, it was important to test

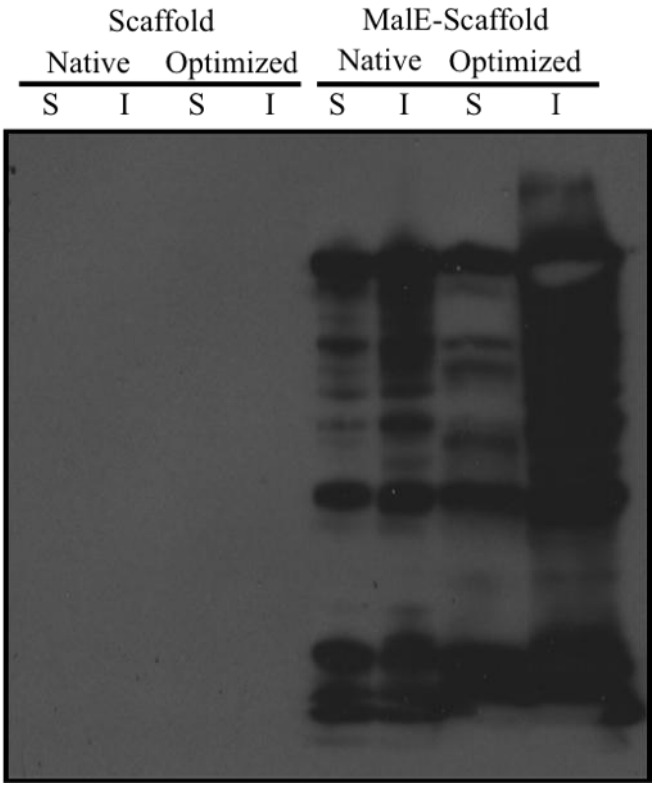


Figure 3.12. Western Blot of soluble (S) and insoluble (I) scaffold expression levels. From left to right: native and codon optimized constructs of SH3-PDZ-GBD scaffold, followed by native and codon optimized constructs of MalE-L16-SH3-PDZ-GBD scaffold.

another metabolic pathway, whose enzymes had different kinetic parameters and multimeric states and whose substrates had different diffusivities. The production of 1,2,4-BT stands as such an example and can be produced in *E. coli* from four enzymatic steps from xylose (Figure 3.13) [102]. Initial tests to produce 1,2,4-BT in *E. coli* failed to produce significant yields, despite overexpression of two version of the metabolic pathway, and knockout of *xylA*, which competes for the xylose substrate (Figure 3.14) [114]. In order to increase the levels of 1,2,4-BT so comparison to colocalized enzyme constructs were possible, several gene knockouts were tested (Figure 3.14). Both YagE and YjhH are hypothesized to compete for the second intermediate so it is [114] not surprising that we observed increases in yield as well when these enzymes were absent. In further testing, a *xylA yjhH* double mutant was

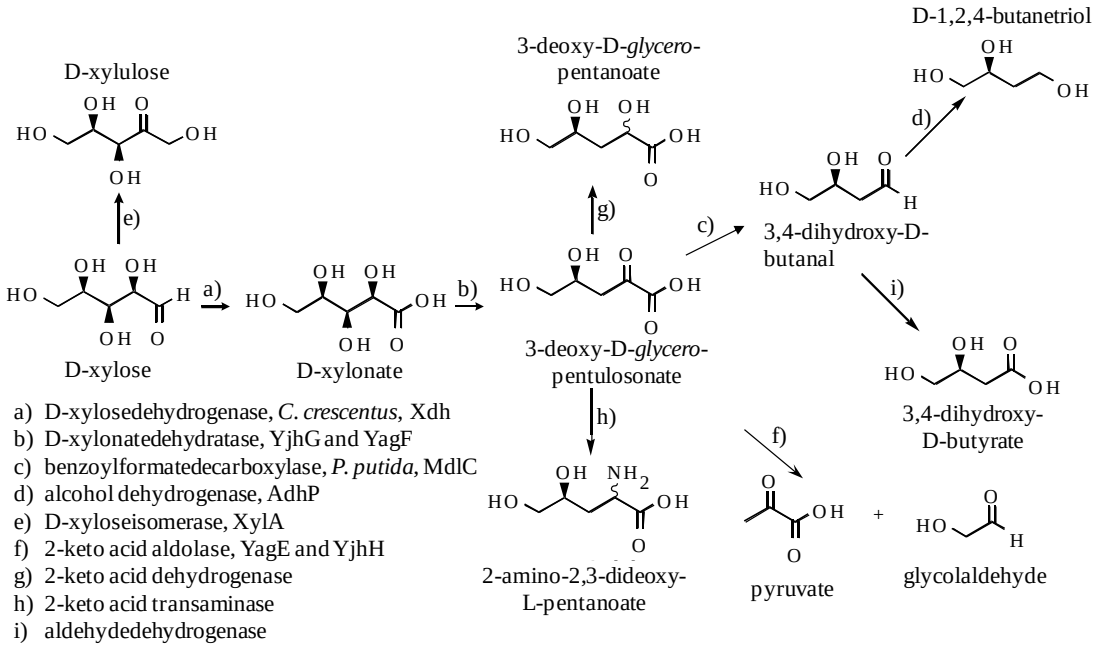


Figure 3.13. Metabolic pathway of 1,2,4-butanetriol production from xylose *E. coli*. Courtesy of John Frost (Department of Chemistry, Michigan State University).

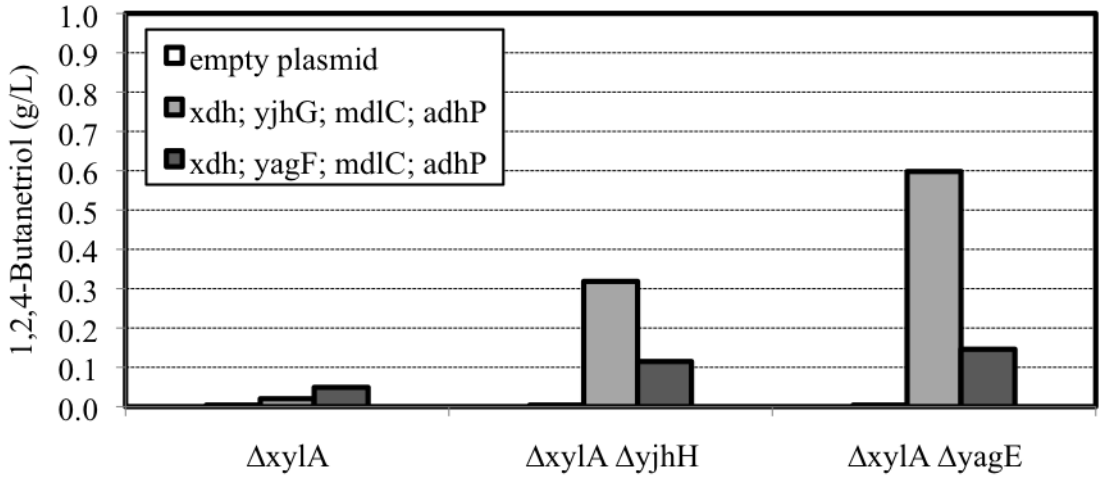


Figure 3.14. The effect of pathway and gene deletion on 1,2,4-butanetriol production. The above enzymes were overexpressed in the given *E. coli* knockout. Data represent a single experiment.

employed as it produced a high yield of 1,2,4-BT and by maintaining the competing enzyme, YagE, benefits of channeling could be readily observed.

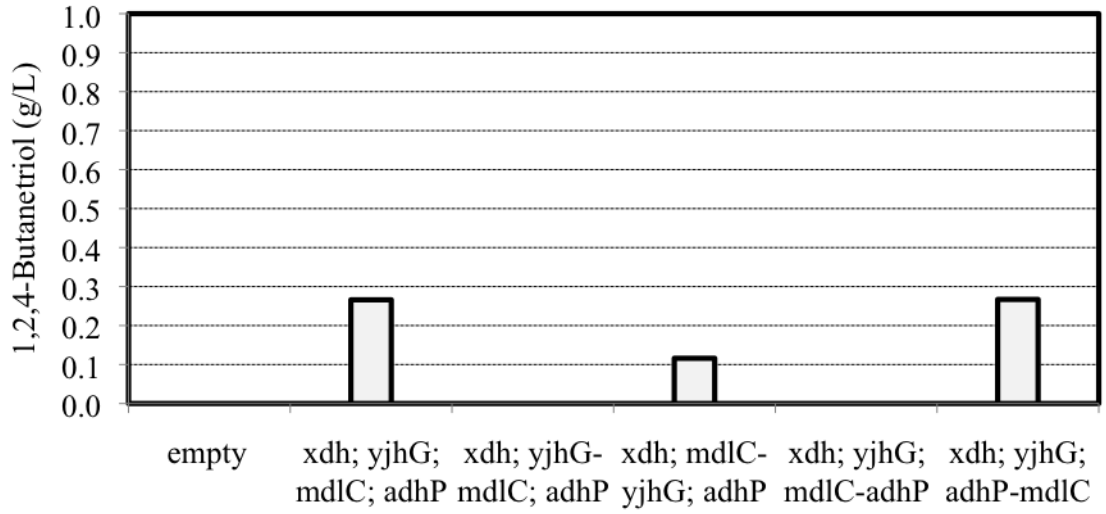


Figure 3.15. Effect of fusion and orientation on 1,2,4-butanetriol production. Columns represent a single experiment.

As a first test of enzyme colocalization, fusions between enzymes two and three, YjhG and MdlC, using a single L16 linker, were constructed in both orientations. The same was done for enzymes three and four, MdlC and AdhP. When coexpressed with the other free enzymes in the pathway, only two of the four fusions gave detectable levels of 1,2,4-BT production, and neither showed improvements over the free enzymes (Figure 3.15). Given that YjhG appeared intolerable to fusion, this approach was abandoned in lieu of pursuing PIDs.

For simplicity, a single PID pair was chosen for testing, cJun-cFos, since it performed similarly to GCN4, and formed only heterodimers. Additionally, since competing reactions are only well defined for the second intermediate, these PIDs were formed only between enzymes YjhG and MdlC. As shown in Figure 3.16, these PIDs reduced the levels of 1,2,4,-BT over the free enzyme control.

Discussion

Despite the promise of enzyme colocalization from the large number of successful *in vitro* studies, their intentional and successful engineering has proven much more challenging. Synthetic metabolic pathways suffer from a number of

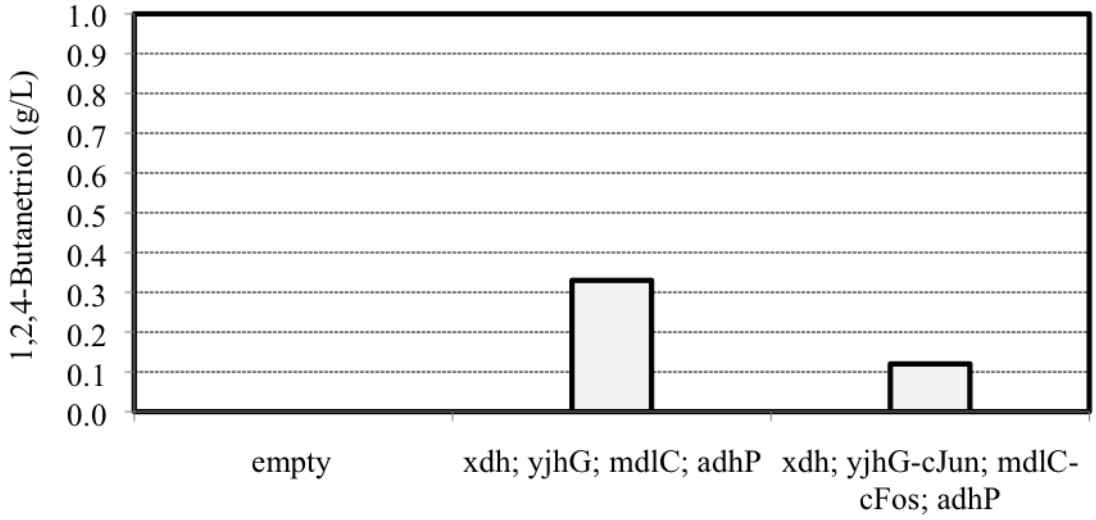


Figure 3.16. Effect of PIDs on 1,2,4-butanetriol production. Columns represent a single experiment.

limitations that curb engineering efforts towards this aim. (1) Proteins from heterologous hosts commonly suffer from low expression and poor folding properties in their new environment. (2) If they indeed fold, many, if not most, metabolic enzymes are multimeric in structure [104], which can lead to an assembly conundrum during colocalization and impairs activity. (3) If the proteins fold and are active, synthetic pathways have not been evolved to balance rates, and these flux imbalances offset the benefits that may result from enzyme colocalization. For example, the catalytic activity for DkgA is 10-fold below that of MgsA [115,116]. When considering this, it is no surprise that the greatest *in vivo* success resulted from an evolved, rather than designed system [8].

In order to address these issues, we set out to test the large set of parameters for colocalized enzymes, both in terms of the mechanisms employed as well as the reaction-diffusion kinetics involved. Fusion proteins proved intractable to modification, and led to both reductions in solubility and in activity over the free

enzymes, even when they performed better than the appropriate control. The relatively few design choices here may ultimately limit the application of this technology. PIDs represent a great step forward for enzyme colocalization. Although their performance was similar, the possibility of using solubility enhancing binding domains with small ligands may enhance both individual and couple kinetic properties of the system. Further, the ability to toggle binding affinity may prove useful in allowing the sequential enzymes to colocalize and allow their activity to be unperturbed by termporarily dissociating. Ultimately, the challenge facing fusions and PIDs is the inability to balance flux in a robust manner. With fusion and PIDs, the ratios of the sequential catalytic centers is dictated by the multimerization state of the enzymes, and this may be true whether or not one coexpresses free enzymes to populate these structures. This fact limits these systems to improvements from only reduced diffusional times, since a flux imbalance would prevent intermediate sequestration. To address this issue, we employed a protein scaffold that would allow each enzyme to bind independently and thus their ratios could be toggled. Unfortunately here, a tripartite fusion of protein docking domains is not well suited for soluble and stable expression in *E. coli*. Since this system provides the maximum design parameters, developing a stable intracellular scaffold will be an important improvement, especially for systems greater than two or three enzymes.

This study addressed two metabolic pathways, 1,2-PD and 1,2,4-BT, and this variety of enzymes, from three different hosts, begins to represent the diversity in engineering metabolic pathways. However, the diffusion properties of the intermediates are very similar for these pathways, with molecular weights all under 200 Daltons. If indeed, the benefits are seen with much slower diffusivities [12], other metabolic pathways will have to be pursued to analyze this effect experimentally. Nevertheless, this study highlights the challenges in engineering enzyme

colocalization and stresses the need for modular strategies with high degrees of freedom in order to optimize around the metabolic pathway of interest.

Acknowledgements

We thank John Dueber of UC Berkeley for providing us with the plasmids pET19(a): A.1.1.a, A.1.1.b, B.2.2, B.3.1. We thank for William Yang of Draths Corporation providing us with the plasmids pWN9.046A, pWN8.022, pWN5.238A, pML6.166. This work was supported by the Cornell University College of Engineering's Sustainable Energy Systems Seed Grant Program, and the Office of Naval Research under a Young Investigator Award (to MPD) and also Grant N00014-07-1-0027.

CHAPTER 4

REPROGRAMMING SYNTHETIC PATHWAYS IN BACTERIA USING ENGINEERED SCAFFOLDS

Introduction

The intentional engineering of metabolic pathways is a rapidly maturing field and much attention has already been given to cellular productivity and diverted flux. Indeed, a number of strategies have been reported for optimizing cellular metabolism including, for example, stabilizing mRNA transcript levels, balancing protein expression levels, evolving enzyme activity, and redirecting cellular metabolism [34]. Recently, an orthogonal strategy was developed for the intentional colocalization of metabolic enzymes along a synthetic intracellular scaffold for the production of both mevalonate and glucaric acid [58,59]. The authors reported a 77-fold improvement in mevalonate titers while maintaining very low enzyme levels, which served to reduce cell stress [58]. Due to the exceptionally low enzyme levels in this case, isolating the impact from changes in individual enzyme activity and protein levels were not possible [117]. In another study, despite high initial titers, glucaric acid levels increased 5-fold, due in large part to the substrate activation of the scaffolded MIOX enzyme, whose individual activity increased 18 to 49-fold [59]. This stands as the clearest case of locally increased intermediate concentrations, but this benefit cannot be extended towards all metabolic pathways. Nevertheless, these engineering metabolic pathways are remarkable examples of using the power of synthetic biology to mimic nature's synergy.

These successes notwithstanding, protein scaffolds employed above begin to break down when considering large metabolic pathways or even a high number of repeating scaffold sites. The prospect of a stable, well expressed 11-domain fusion in

E. coli as employed [59] should stand at the limit of protein folding, and even if not, problems during plasmid construction of repeating DNA units [59], indicate a need for a new type of molecular scaffold. Further, despite the promising results, the difficulty in isolating the impact of enzyme stability, individual, and coupled activity, begs the need for a new type of scaffold that permits these controlled experiments *in vivo*.

Here, we present a study of enzyme colocalization for the production of both 1,2-propanediol and mevalonate from three enzymatic steps that branch from *E. coli* central carbon metabolism. Since bacterial production of 1,2-PD and mevalonate are hindered by the accumulation of the bactericidal intermediates and by the formation of undesired side products from competing pathways, these represents a unique opportunity for artificial enzyme compartmentalization as a means to increase production titers via the reduction of toxicity and side product formation. In this study we investigate the use of a stable and well-ordered biological scaffold for successive docking of our metabolic pathway enzymes. Using the available components within a bacterial cytoplasm, plasmid DNA is employed to provide an ordered surface for modification. The use of modular DNA binding domains permits enzyme colocalization along its surface and the subsequent optimization of metabolite production within these pathways. The techniques presented here provide a generic framework for creating colocalized enzymes and should be useful in evaluating the contribution of enzyme compartmentalization within any recombinant pathway of interest.

Materials and methods

1,2-PD biosynthesis pathway. Plasmids were constructed using molecular biology techniques according to standard practices [97]. The three sequential genes for encoding 1,2-PD biosynthesis, (1) *mgsA*, (2) *dkgA*, and (3) *gldA* were PCR amplified from *E. coli* MG1655 genomic DNA. These genes were then cloned into pBAD18

[98] in the order shown, using unique restriction sites within the multiple cloning site, resulting in a polycistron for 1,2-PD synthesis. The first gene was placed between NheI and XbaI, the second gene between XbaI and SphI, and the last gene between SphI and HindIII. The same strong ribosomal binding site was placed directly upstream of each gene in the polycistron and an NdeI site at each start codon.
To the C-terminus of MgsA, the zinc finger triplet OZ052 (ZFa) [118] was codon optimized and fused by an L5 polylinker [64]. To the C-terminus of DkgA, the zinc finger triplet OZ300 (ZFb) [118] was codon optimized and fused by an L5 polylinker [64]. To the C-terminus of GldA, the zinc finger triplet OZ076 (ZFc) [118] was codon optimized and fused by an L5 polylinker [64]. Each zinc finger had a C-terminal HA epitope tag.

Mevalonate biosynthetic pathway. Plasmids were constructed using molecular biology techniques according to standard practices [97]. The mevalonate pathway enzymes were tethered to zinc fingers, based on the construct pRM178 [58]. Here the linker and ligand at the 3' end of each gene was excised and replaced exactly with the L5 linker and appropriate zinc finger as above. ZFa was fused to AtoB, ZFb fused to HMGS, and ZFc fused to HMGR, with a C-terminal HA epitope HA on all proteins. The resulting plasmid was named, pRM178-ZF.

DNA Scaffold Constructs. pUC19 [119] served as the basis for construction of the DNA scaffold. pUC19 was digested with AatII and PvuII and ligated with a polylinker containing the following restriction sites: AatII-SacI-SpeI-XbaI-SphI-ClaI-PvuII. The scaffold sequences were based on the following binding sites: (1) ZFa: 5'-GTCGATGCC-3', (2) ZFb: 5'-GCGGCTGGG-3', and (3) ZFc: OZ076 5'-GAGGACGGC-3'. DNA scaffolds were assembled using SpeI and XbaI cohesive ends compatible ends for ligation. Basic parts were made so that they would be flanked by an SpeI site on the 5' end and XbaI site on the 3' end. Composite scaffolds

were constructed by digesting the backbone with XbaI and adding an SpeI/XbaI digested insert at the 3' end, maintaining the SpeI at the 3' end and XbaI site at the 5' end. To separate the binding sites, a 12 base pair spacer was employed for linear constructs and a 4 base pair spacer for spiral constructs. For 1,2-PD production, the resulting scaffold were (1) subcloned into pBAD18, between the B-lactamase and pBAD promoters, (2) subcloned into pZE11-MCS1 at the AatII/XbaI sites, or (3) used directly from pUC19. For mevalonate production, the scaffolds were employed directly from pUC19.

Bacterial strains, media, and growth conditions. *E. coli* W3110 [F^- λ^- *rph-1 INV(rrnD, rrnE)*] or its derivatives were used for 1,2-PD production unless otherwise noted. 1,2-PD fermentations were followed as described [82] with the following exceptions. L-arabinose was added to 0.2% (w/v) at the time of inoculation to induce gene expression. All fermentations were run at 37°C, either at 200 or 250 rpm, with tubes held vertically or at a 45 degree angle. The 10-mL fermentation mixtures were inoculated to an optical density of 0.05 (typically 1:50 to 1:100 dilution) with the overnight culture. The optical density (OD) was measured at 600 nm with a Thermo Spectronic BioMate 3 model spectrophotometer.

Mevalonate production was followed as described [58]. Antibiotics were provided at the following concentrations: ampicillin, 100 μg/mL; chloramphenicol, 20 μg/mL; spectinomycin, 20 ug/mL.

HPLC analysis. Products in the fermentation media were quantified as described [82] with the following exceptions. Compounds were measured with a Waters Breeze system. The mobile phase was a 0.03 N sulfuric acid solution, with a flow rate of 0.45 ml per min, and the column temperature was 50°C. All samples were filtered through 0.22-μm-pore-size membranes prior to analysis.

Protein purification. Maltose binding protein was PCR amplified from *E. coli* MG1665 genomic DNA and cloned into pET28a(+) (Invitrogen). The 3' end was appended with a codon optimized L5 linker along with the appropriate zinc finger with or without epitope tags. Growth and purification were followed as described [120] with the following exceptions. Cell cultures were induced with 0.3 mM IPTG for 3 hours. Cell lysates were applied over amylose resin (NEB) using Poly-Prep Chromatopgraphy Columns (Bio-Rad) with WB1 buffer substituted for column buffers. Elutions were collected in fractions and the concentrations were estimated by absorbance at 280 nm.

Electrophoretic mobility shift assay. DNA oligos were 5' biotinylated (IDT) and annealed by slow cooling from 95°C. Binding assays were performed as described [120]. Complexes were run on 5% TBE-PAGE gels (Bio-Rad) at 4°C. Samples were transferred to Biodyne B Nylon Membranes (Pierce). Detection was performed using a LightShift Chemiluminescent EMSA Kit (Pierce).

Western blot analysis. W3110 cultures were grown overnight in LB media, inoculated as above, lysed via sonication after either 9 or 24 hours, and normalized on the amount of total soluble cell protein, with 0.2 ug total loaded in each lane. Insoluble lysates were isolated by centrifugation at 16,000 g for 10 minutes. Pellets were washed twice with 1 mL of with Tris-Cl (50 mM) and EDTA (1 mM) pH 8.0. The pellet is then resuspended in 1x PBS and 2% SDS and then boiled for 10 minutes. The insoluble fraction is collected as the supernatant following a 10 minute centrifugation at 16,000 g. Western blotting of these soluble and insoluble lysates was performed with anti-HA primary antibody (1:1000, v/v; Sigma) and anti-rabbit secondary antibody (1:2500, v/v; Promega) on SDS-PAGE gels, typically 10 or 12%.

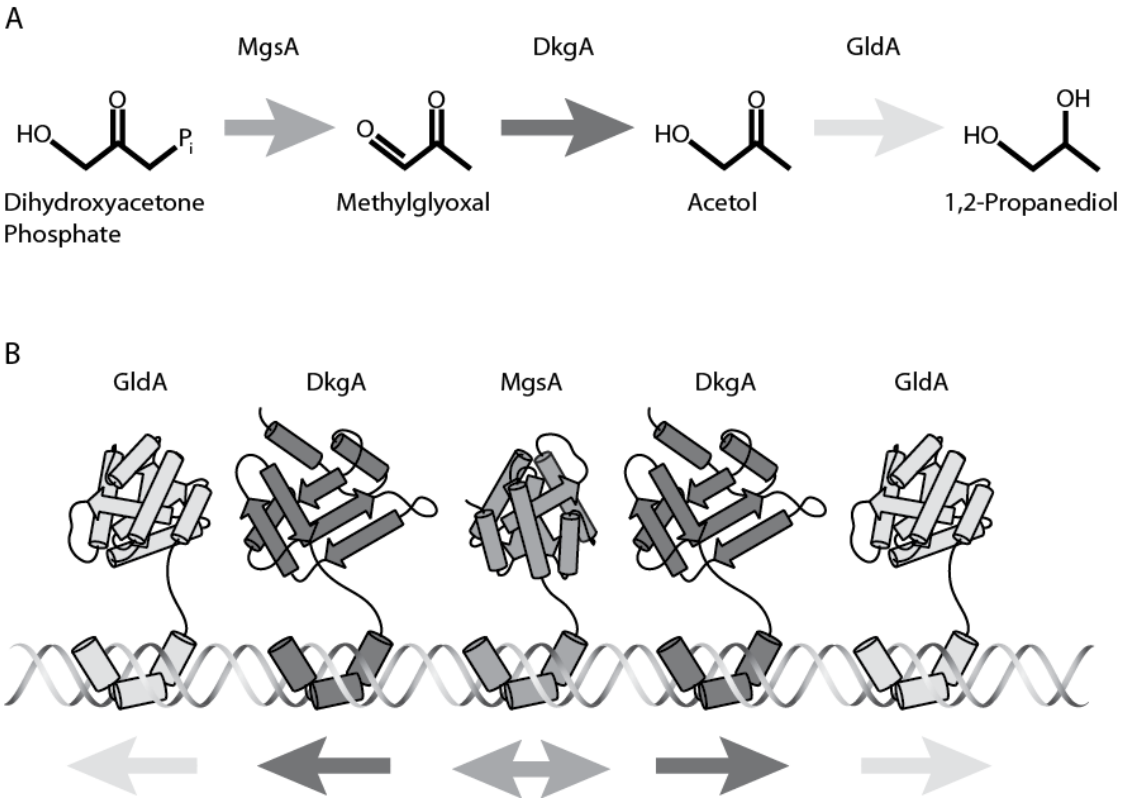


Figure 4.1. (A) Metabolic pathway of 1,2-PD production from dihydroxyacetone phosphate in *E. coli*. (B) Example of enzyme architecture along a plasmid DNA surface. Arrows indicate direction of reaction progress.

Results

Independent control of enzyme colocalization. A biosynthetic route towards 1,2-PD has been well defined ([82], and Chapter 3 above, and Figure 4.1A), and the compartmentalization of these three enzymes would enable sequestration of the toxic methylglyoxal intermediate [3], reduce diverted flux, and balance pathway flux. In order to create a modular and stable scaffold, capable of a wide array of defined geometries, we envisioned the use of a plasmid DNA scaffold surface to which successive enzymes could dock to form an efficient reaction sequence (Figure 4.1B). Control of enzyme position on the DNA surface was gained through the use of engineered zinc finger triplets [118], which have been developed for a large number of specific 9 base pair DNA sites. By simply fusing these to the C-terminal domain of

our metabolic enzymes, we could successively target each enzyme to a specific site on a plasmid molecule. In a relatively short distance, the series of enzymes could be spatially arranged along the DNA molecule (Figure 4.1B). In this manner, enzyme stoichiometry and spacing was determined through regular placement of these scaffold sites along the DNA molecule (Figure 4.1B). Individual control over the proteins and scaffolds are maintained by transcriptional control of protein levels, and by varying the plasmid origin of replication and numbers of repeating units to control DNA scaffold levels.

As a first test of the system components, it was important to quantify the binding affinities of the zinc finger-DNA complexes to determine how these parts would function *in vivo*. A gel shift assay, employing an MBP fusion was used to assess the dissociation constant of zinc finger triplets for short linear dsDNA. The first triplet tested, here ZFa, bound with a K_D in the low nM range, and further showed that the presence of an epitope tag would not dramatically hinder this interaction (Figure 4.2A). This finding is significant because comparison of intracellular protein levels is important in determining if enzyme channeling is present [117]. Testing other engineered triplets towards orthogonal DNA sites, the dissociation constant was found to be similarly in the low nM range. This may be a general feature of zinc finger triplets, engineered through *in vivo* selections [118] that may well permit their general use in expanding colocalized metabolic pathways.

Extending these results *in vivo*, we selected a small number of enzyme stoichiometries that had been previously successful [58], and were determined to be useful in balancing the flux given that the catalytic activity of MgsA is much greater than that of DkgA [115,116]. Given the well-defined local structure of plasmid DNA, it was necessary to identify possible enzyme arrangements on the scaffold surface. The motivation to employ bidirectional scaffolds is derived from the fact that plasmid

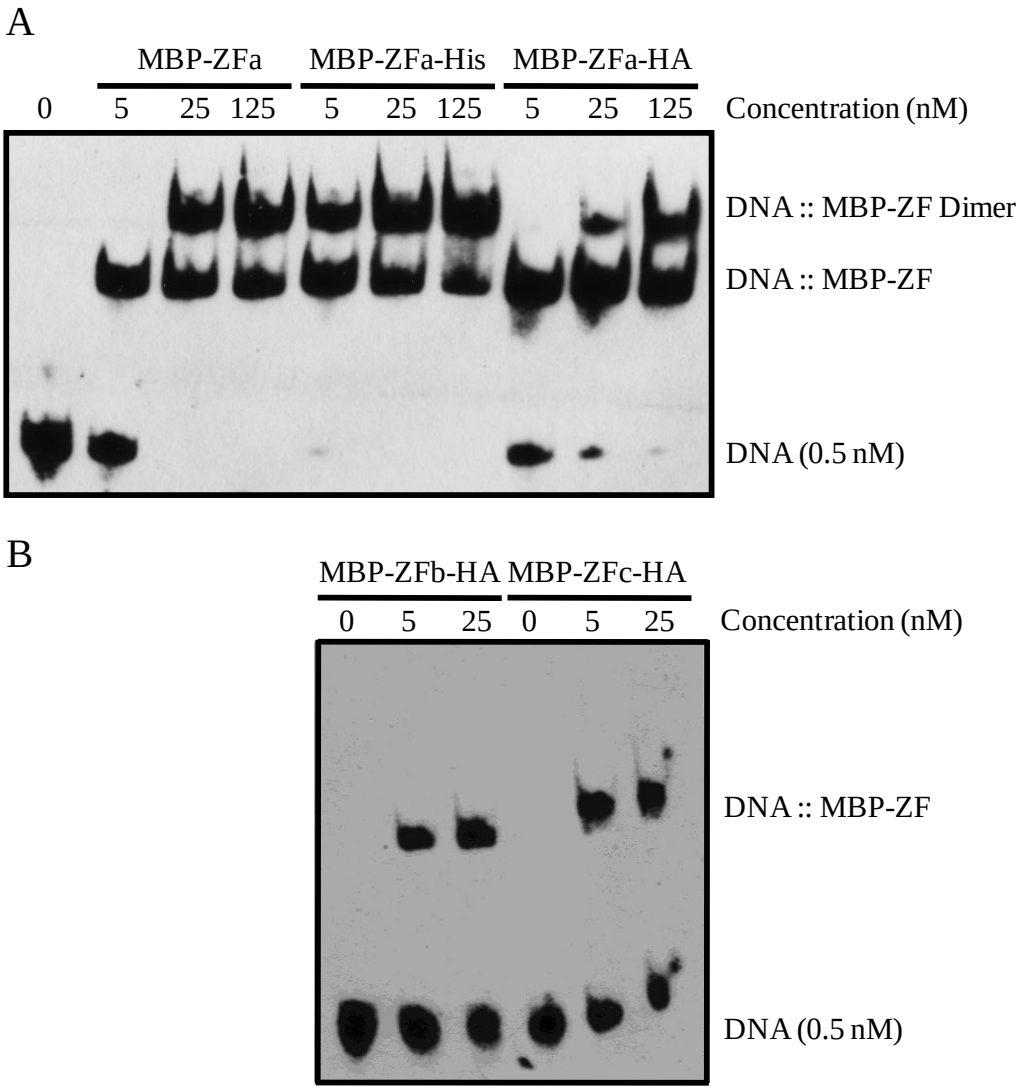


Figure 4.2. EMSA of MBP-zinc finger fusions and their biotinylated-DNA target. (A) Effect of epitope tag on dissociation constant of zinc finger A-DNA interaction and (B) Effect of various zinc fingers on the dissociation constant of zinc finger-DNA interaction. Detection with streptavidin-HRP against biotinylated DNA.

DNA copy numbers [121] are commonly far below that of overexpressed metabolic enzymes. With this in mind, a simple scaffold geometry, like the one shown in Figure 4.1B, could need to be repeated tens of times on a plasmid to achieve full site occupancy, so scaffold sites were designed to be compatible in this context.

DNA scaffolds increase 1,2-PD levels with precise control. 1,2-PD titers were measured when no scaffold sites were present on the plasmid DNA and fold

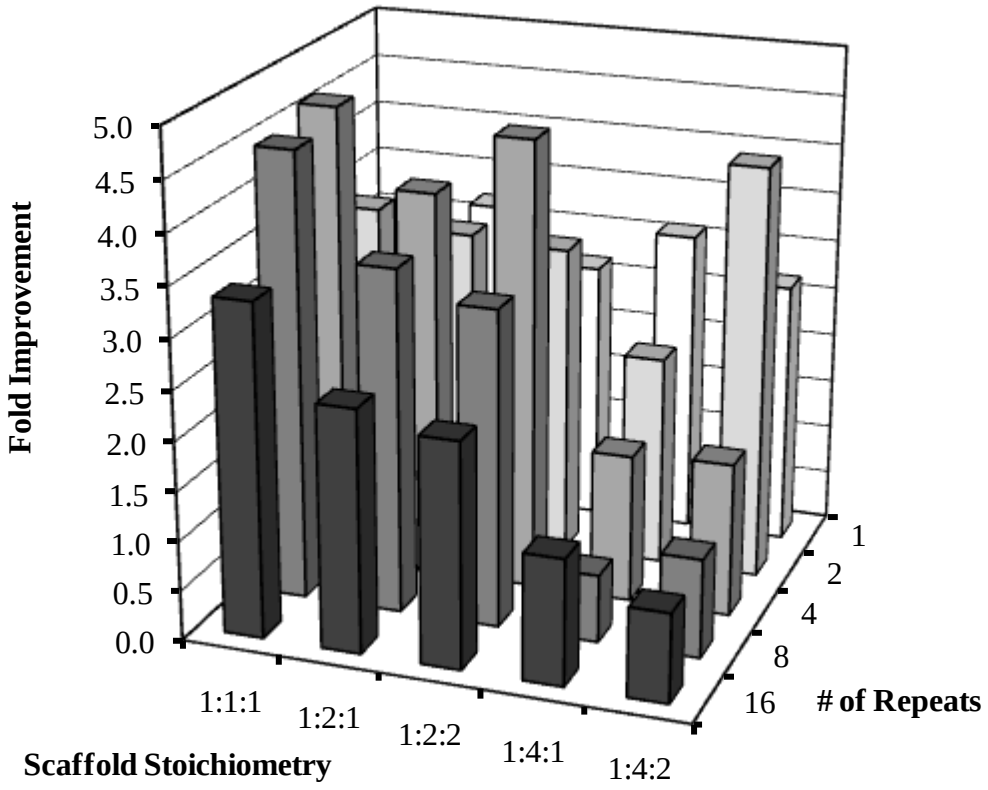


Figure 4.3. Fold improvement in 1,2-PD production comparing various linear DNA scaffold stoichiometries and numbers of repeats to the case where no DNA scaffold is present. The scaffold stoichiometries have the following geometries: 1:1:1 as $(ABC)_n$, 1:2:1 as $(BABC)_n$, 1:2:2 as $(CBABC)_n$, 1:4:1 as $(BBABC)_n$, 1:4:2 as $(CBBABBC)_n$, where A represents MgsA binding; B, DkgA binding; C, GldA binding; and n is the number of repeats of the defined unit. Columns represent the average of three experiments with standard error of less than 30 percent.

improvement was determined for each scaffold stoichiometry and number of repeating units against this control. A 4.5-fold improvement is achieved for scaffold harboring both a 1:1:1 and 1:2:2 stoichiometry (Figure 4.3). Titers were highly dependent on both the stoichiometry of binding sites as well as the number of repeating units of these sites. For example, while 4-fold improvements were achieved with both a 1:1:1 and a 1:4:2 scaffold, the titers for the 1:1:1 geometry is relatively independent of the number of repeating units, while the 1:4:2 scaffold shows a clear maximum, after which improvements drop off markedly. Four of the five scaffolds show maxima in

the numbers of repeats, akin to the optimal induction levels observed for protein scaffolds [58]. Here, the impact on the precise number of binding sites is more direct given the consistency of DNA levels observed (data not shown), where protein levels are harder to control since (1) few promoters show a linear response [121] and (2) aggregation and degradation rates depend on the amount of overexpressed protein.

To investigate the role of scaffold numbers for each pathway enzyme, we transformed Figure 4.3 into one that plots the fold improvements against the number of scaffold sites for each enzyme (Figure 4.4). Interestingly, each enzyme shows a maximum in 1,2-PD levels when plotted against the number of binding sites. This effect is most clear for DkgA although this trend may continue for MgsA and GldA if scaffold constructs were built that contained up to 64 binding. This is the logical outcome since there are several competing reactions for the methyglyoxal intermediate [104]. Important to note is that there is still a large variation between stoichiometries in the improvement towards 1,2-PD which suggests that enzyme stoichiometry plays an important role in increasing 1,2-PD titer.

Role of protein levels and DNA site spacing on improvements to 1,2-PD. Enzyme levels were measured when no scaffold sites were present on the plasmid DNA and compared to those of a 1:2:1 scaffold stoichiometry with varying numbers of repeats. Enzyme levels for both MgsA and DkgA are highly dependent on both the presence and number of repeating scaffold units (Figure 4.5). Surprisingly, the plasmid with a single MgsA binding site results in a 13-fold increase in its protein level, which is relatively constant until the number of binding sites increases to 8 or 16. The same trend is true for DkgA as well, though less dramatic, and the presence of DkgA binding sites increases its protein level up to 5-fold. Surprisingly, the simple act of binding to DNA does not drive this increase in protein levels, as fewer binding sites

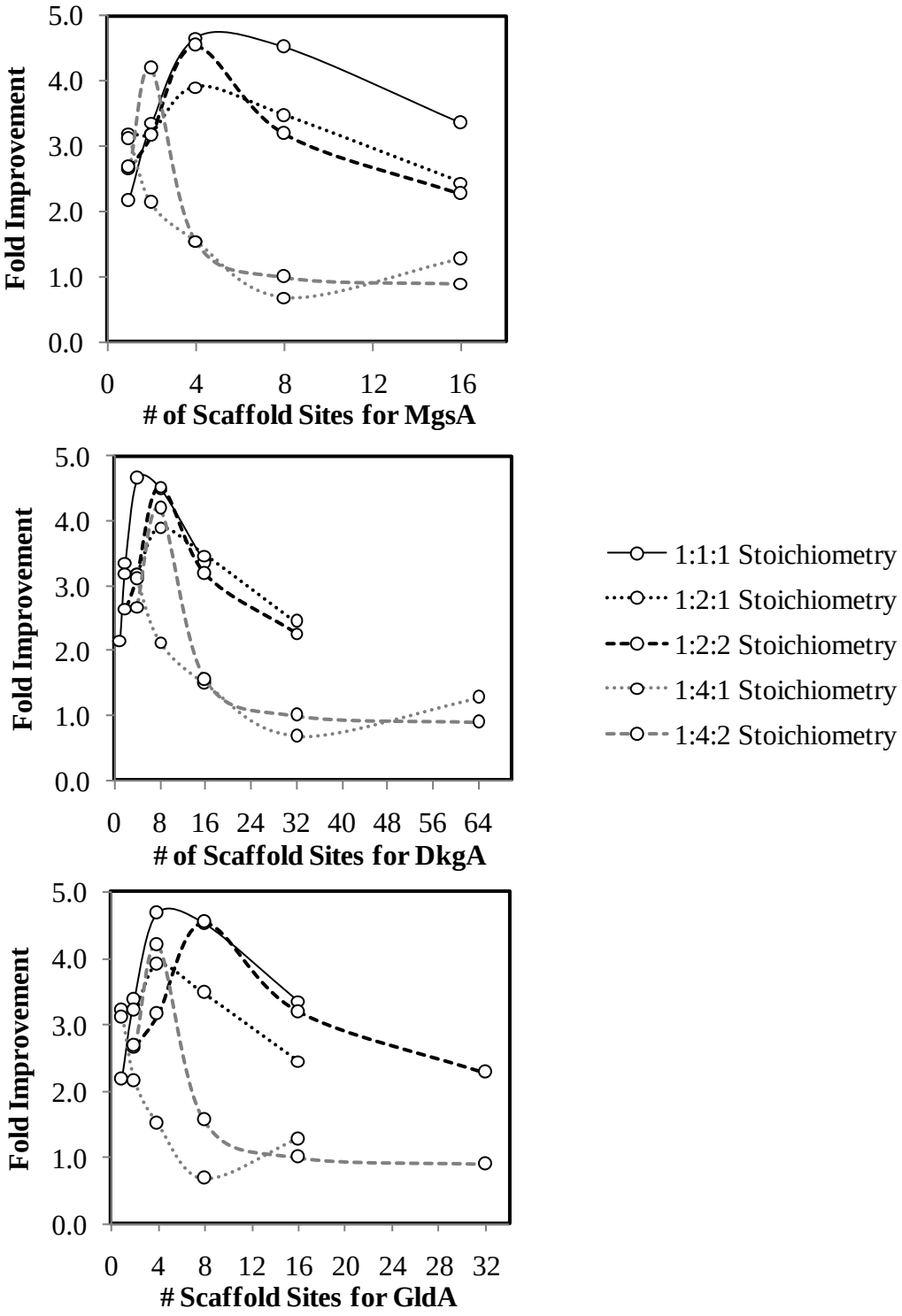


Figure 4.4. Fold improvement in 1,2-PD production for various numbers of DNA scaffold sites for each enzyme, compared to where no DNA scaffold is present.

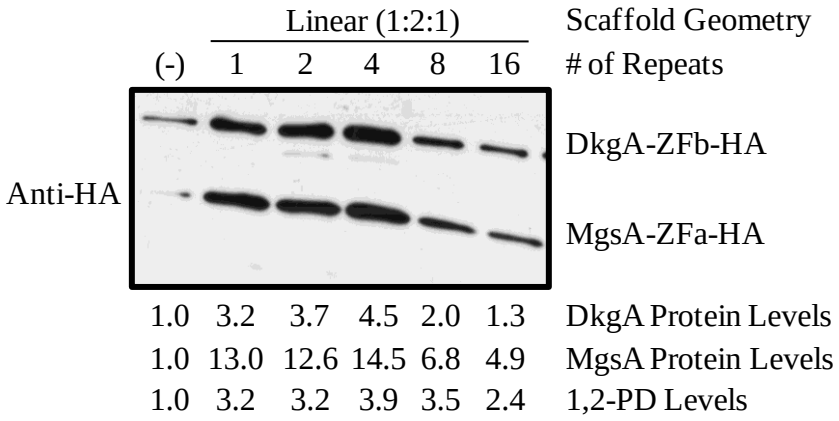


Figure 4.5. Western Blot of soluble protein levels of both MgsA and DkgA with no scaffold or varying repeats of the 1:2:1 scaffold geometry. Below, relative protein levels are presented (ImageJ, NIH), along with relative 1,2-PD production levels from Figure 4.3.

results in higher protein levels. It may be that a tightly packed scaffold may cause an increase in stability whereas this effect disappears when proteins sparsely populate the DNA surface. Additionally, the increase in protein levels, especially for DkgA, trends with the increase in 1,2-PD titers, an effect not seen for protein scaffolds and glucaric acid production [59]. As a result, further experiments to decouple binding from enzyme proximity will be especially important to determine the role of enzyme colocalization.

As a further investigation into the improvement of scaffolding enzymes, DNA binding sites were constructed with a varying geometry. Instead of the linear enzyme format shown in Figure 4.1B, a shorter 4 base pair spacing between binding sites was employed so that each enzyme rotated 450 degrees around the plasmid DNA, instead of 720 degrees in a linear structure. Although the resulting spiral led to tighter site packing, the distance in space between enzymes remained similar. 1,2-PD titers were measured when no scaffold sites were present on the plasmid DNA and fold improvement was determined for each scaffold stoichiometry and number of repeating units against this control. Here, a 4.0-fold improvement is achieved for scaffold

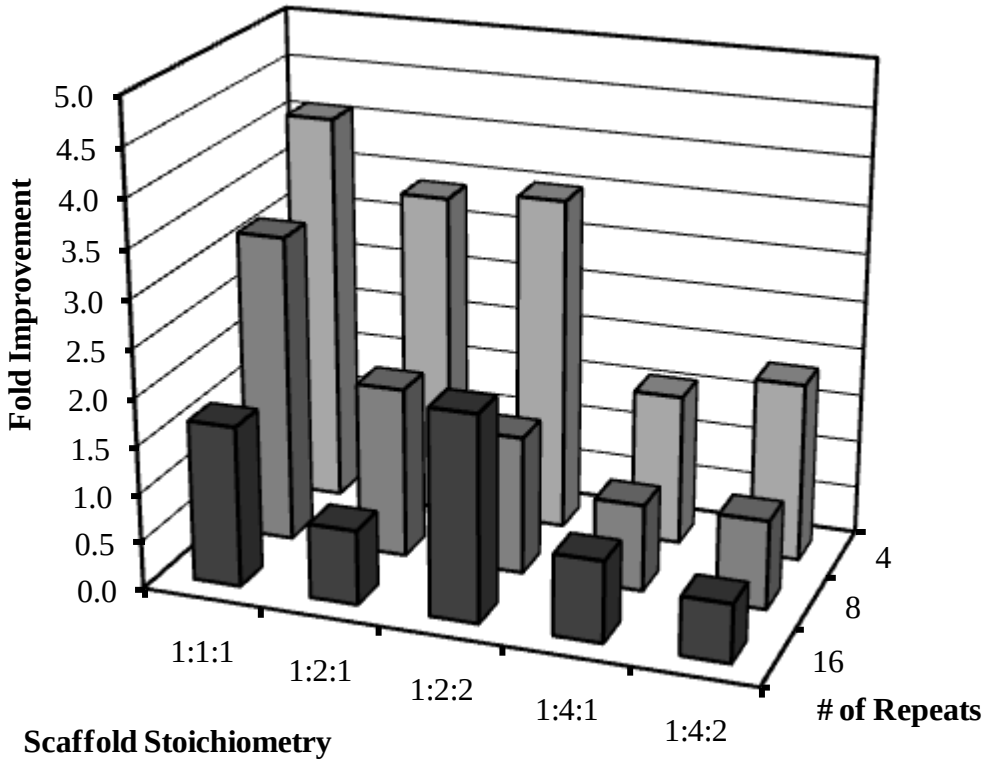


Figure 4.6. Fold improvement in 1,2-PD production comparing various spiral DNA scaffold stoichiometries and numbers of repeats to the case where no DNA scaffold is present. The scaffold stoichiometries have the following geometries: 1:1:1 as $(ABC)_n$, 1:2:1 as $(ABCB)_n$, 1:2:2 as $(ABCCB)_n$, 1:4:1 as $(ABCBB)_n$, 1:4:2 as $(ABBCCBB)_n$, where A represents MgsA binding; B, DkgA binding; C, GldA binding; and n is the number of repeats of the defined unit. Columns represent the average of three experiments with standard error of less than 55 percent.

harboring a 1:1:1 stoichiometry and other scaffolds showing much less improvement (Figure 4.6). Interestingly, the spiral scaffolds all performed the same or worse than the linear scaffold counterpart, implying that the enzyme geometry does matter. Even though this alternative arrangement did not perform as well, it suggests that the geometry of DNA binding sites is an important parameter that should be considered when engineering a metabolic pathway.

1,2-PD improvements are binding dependent. To demonstrate that the improvements in 1,2-PD titers results from enzyme docking and not simply the

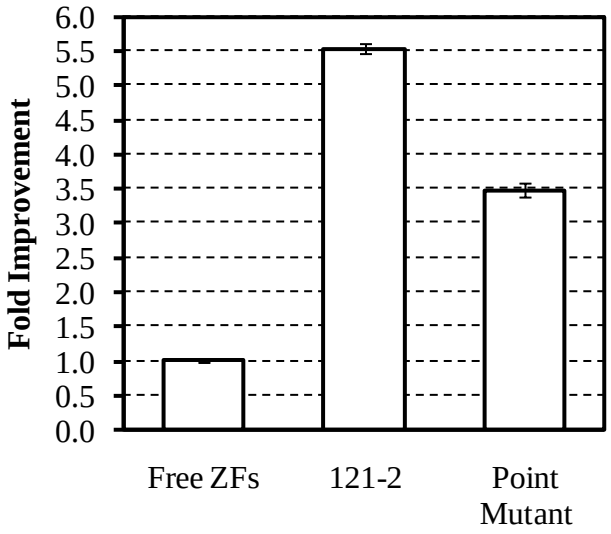


Figure 4.7. Fold improvement in 1,2-PD production comparing no scaffold to a 1:2:1 scaffold with two repeats to the same scaffold harboring point mutations. Columns represent the average of three experiments with standard error of less than 4 percent

Figure 4.8. Western Blot of soluble protein levels on samples of Figure 4.7.

presence of the DNA scaffold sequence, point mutations were made to the 1:2:1-2x scaffold sequence to reduce the binding affinity for the zinc finger triplets. It is important to note that no changes were made to the proteins themselves and fewerthan 10% of the base pairs were mutated over the original scaffold. When tested for the production of 1,2-PD, the scaffold harboring point mutations led to a 40% reduction in 1,2-PD levels (Figure 4.7), while maintaining similar protein levels (Figure 4.8). Surprisingly, the scaffold with the point mutation still led to the same high protein levels, while experiencing somewhat lower 1,2-PD levels. This suggests that the full synergy is retained only when enzymes have high affinity for the scaffold sites, though

even low binding affinity for the DNA is sufficient to increase protein levels compared to having no scaffold present. This is possible because the zinc fingers likely retained some affinity for the mutated sites and because nM affinity is not required when protein concentrations inside the cell greatly exceed that amount.

DNA scaffold benefits generalizable to mevalonate biosynthetic pathway. Mevalonate titers were measured when no scaffold sites were present on a separate plasmid DNA and fold improvement was determined for each scaffold stoichiometry and number of repeating units against this control. A 3.9-fold improvement is achieved for scaffold harboring a 1:4:2 (Figure 4.9). Titers were highly dependent on both the stoichiometry of binding sites as well as the number of repeating units of these sites. For example, the 1:2:1 stoichiometry has increasing yields with the number of repeats while the reverse is true for the 1:2:2 and 1:4:2 stoichiometries. Interestingly, the 1:4:1 stoichiometry was the worst performing arrangement, which is consistent with the second intermediate being cytotoxic [58]. The fact that the 1:2:2 ratio performed best using protein scaffolds is not surprising given the differences in site ordering, enzyme expression levels, and differences in the numbers of binding sites used here. The fact that similar improvements were achieved for mevalonate as 1,2-PD suggests that DNA scaffolds may be generalizable to increasing metabolite yields for a variety of synthetic pathways.

Discussion

While engineers have not yet realized the promise of enzyme compartmentalization in the form of a protected tunnel [10] or intracellular organelles [122], the use of DNA scaffolds is an important extension from the protein scaffolds developed for both mevalonate and glucaric acid production [58,59]. The ultra stable nature of DNA and its ability to permit long range scaffolds, here up to 2.4kb with over 100 individual binding sites, will enable the colocalization of large metabolic pathways where it will

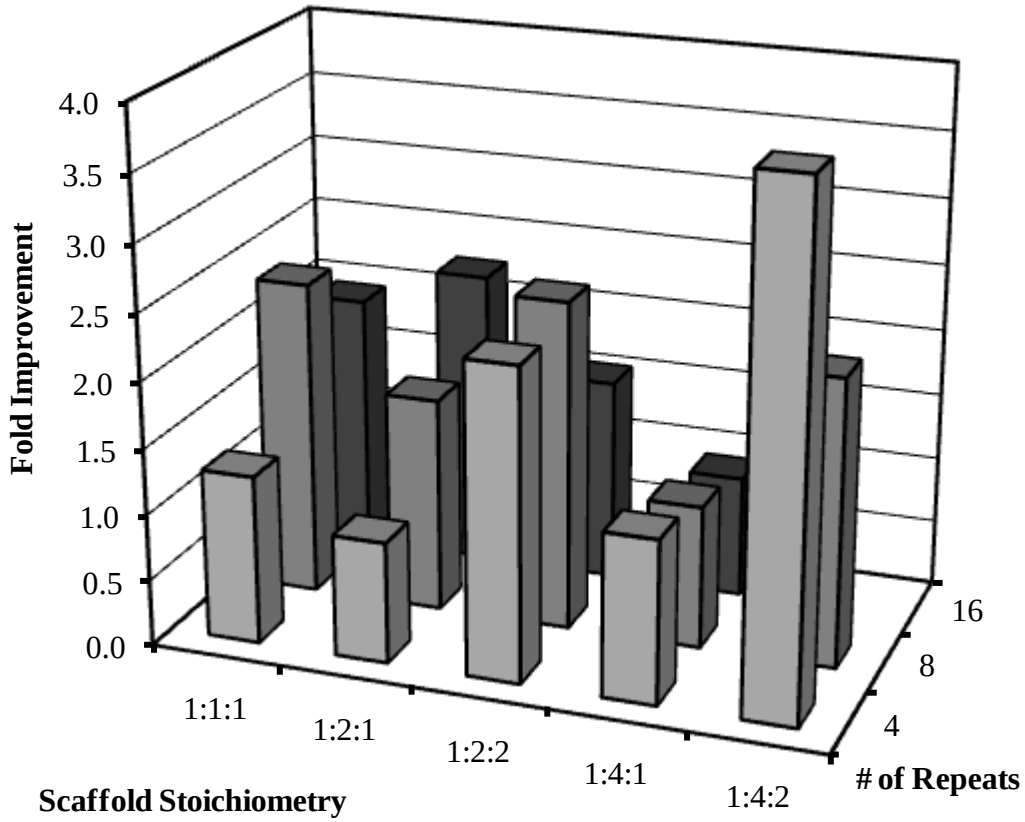


Figure 4.9. Fold improvement in mevalonate comparing various linear DNA scaffold stoichiometries and numbers of repeats to the case where no DNA scaffold is present. The scaffold stoichiometries have the following geometries: 1:1:1 as $(ABC)_n$, 1:2:1 as $(ABCB)_n$, 1:2:2 as $(ABCCB)_n$, 1:4:1 as $(ABCBB)_n$, 1:4:2 as $(ABBCCBB)_n$, where A represents AtoB binding; B, HMGS binding; C, HMGR binding; and n is the number of repeats of the defined unit. Columns represent a single experiment.

be important to create channels of more than three enzymes; a feat simply infeasible with protein scaffolds. Further, although the benefits of engineered metabolic channels exists with some controversy [123], DNA scaffolds permit the kind of controls necessary to assess these improvements in their natural environment. One can decouple binding from activity from synergy, and determine the role of each enzyme, all without making changes to the enzymes themselves or affecting the stability of the DNA scaffold itself. Finally, the binding to DNA scaffolds improves the levels of the

metabolic enzymes, not observed with protein scaffolds [59], which may itself be a useful tool in metabolic engineering.

The DNA scaffolds presented here are not without drawbacks. First, the very rigidity of the DNA scaffold over short ranges could emphasize this. Since many metabolic enzymes are multimers, here $MgsA_6$ and $GldA_8$ for 1,2-PD and $AtoB_4$, $HMGS_2$, and $HMGR_2$ for mevalonate, the flexibility of a protein scaffold may benefit by allowing the enzyme to form in their multimeric state on the same scaffold molecule. While this would be possible on DNA scaffold with the use of long peptide linkers connecting the enzymes to the zinc finger binding domains, this could offset the stability of these fusions. Second is that like with the protein scaffold, these zinc fingers impact native enzyme activity. Despite the larger size of these zinc finger triplets when compared to these peptide ligands, 100 versus 20 amino acids, both fusions have similar impact, implying this may be a general feature of enzyme activity. Similar to the studies of fusion and PID orientation in Chapter 3, however, this degree of freedom is retained with zinc finger triplets and may improve the individual enzyme activity back to its native levels. If the fold improvements from docking pathway enzymes are retained, real gains in metabolite levels could be observed.

It is interesting that 1,2-PD levels were at the highest with a 1:1:1 stoichiometry, keeping in mind the imbalances in catalytic activity of MgsA and DkgA. This result is complicated by the fact that GldA has catalytic activity towards methylglyoxal. In the 1:1:1 scaffold construct alone, repeating units juxtapose binding sites for MgsA and GldA, ABCABC as an example of two repeats. For this reason, this was avoided in the other stoichiometries.

DNA scaffolds should prove a useful tool for synthetic biologists and metabolic engineers resulting from their stability, modularity, and ease of

implementation given the library of constructs that we have assembled. DNA scaffold offer a very high degree of freedom and parameters to consider are: site spacing, number, geometry, location on the plasmid, origin of replication, and binding affinity for the DNA-zinc finger complex. These parameters increase dramatically once changes to the enzyme fusion are considered and complete optimization of a system should permit very high titers. Important in all of this is that this approach is complementary to the well-developed tools of directed evolution and strain engineering commonly applied to metabolic pathways. A successful strategy will likely apply enzyme colocalization to a previously optimized system as preferential stoichiometries will likely vary with enzyme activity and cellular metabolite levels. Of particular interest may be that Zn^{2+} is required for zinc finger binding to DNA sites, and while not tested here, could be a useful on/off switch for coupled metabolic reactions. Considering the demonstrated improvements using DNA scaffolds and its vast promising from unexplored parameters, DNA scaffolds should provide a useful platform for engineering metabolite production *in vivo*.

Acknowledgements

We thank Jay Keasling of UC Berkeley for providing us with the plasmid pRM178. This work was supported by the Cornell University College of Engineering's Sustainable Energy Systems Seed Grant Program, and the Office of Naval Research under a Young Investigator Award (to MPD) and also Grant N00014-07-1-0027.

CHAPTER 5

A GENERAL STRATEGY FOR ENGINEERING ENZYME COLOCALIZATION

A genenral strategy

The field of enzyme colocalization is ever growing [96] and recent results by ourselves and others [58,59] have begun to fulfill the promise of this nascent technology. While initial demonstrations were limited to *in vitro* settings, numerous mechanisms for colocalizing enzymes *in vivo* have been developed: fusions [40], post-translational interactions (Chapter 3), protein scaffolds [58], and DNA scaffolds (Chapter 4). The development of these techniques has provided engineers with increasingly greater control over these systems and greater confidence that the synergistic benefits are real. In particular, the use of DNA scaffolds offers great potential to maximize metabolite production from couple enzymatic reactions with its ability to independently colocalize large metabolic pathways on a stable intracellular scaffold. Since the DNA scaffolding technology is limited to the bacterial cytoplasm, the concurrent development of other methods is important.

The persistence in nature of colocalized enzyme systems [68], once evolved, demonstrates the powerful driving forces for their occurrence: (1) decrease the intermediate transit time and to increase the catalytic efficiency that results from a net decrease in the Km value for channeled substrates, (2) reduce the cellular levels of intermediates while keeping the effective local substrate concentrations high, (3) swift conversion of toxic or unstable intermediates, (4) prevent inhibitory compounds from accessing and deactivating the active site, (5) prevent metabolic cross talk between competing pathways, (6) rapid and reversible assembly and disassembly allows for

differential control over the synthesis of secondary products in response to the intra- and extracellular environment [68]. When engineering cellular metabolism, these are the shortcomings that we can hope to overcome, and should look to when considering engineering enzyme colocalization. Metabolic pathways that suffer from slow substrate uptake, are cofactor limited [44], or have product toxicity [124], will likely not benefit from enhanced pathway synergy.

When considering implementing DNA scaffolds, first one must select two or more sequential reactions to couple where the second reaction is limited by one of the aforementioned problems. All things being equal, this should come at the beginning of the synthetic metabolic pathway to maximize flux towards the end product of interest. An example of this is demonstrated in Chapter 3. Methylglyoxal suffers from several competing reactions and a flux imbalance, whereas acetol has no defined competing reactions, and the activity of GldA towards it is unknown. The next most critical feature is the impact of the zinc finger binding domain on native protein activity. The challenge observed thus far is that the deleterious modifications needed to colocalize enzymes outweigh the noted improvement from the couple reaction synergy (Chapter 4). As seen with many protein fusions [40,125], while it is possible to impact novel activity (here, binding) often this comes at the expense of catalytic activity and folding. This situation is no different. As observed in Chapters 3 and 4, domain fusions significantly impact the metabolic enzymes involved in 1,2-PD and 1,2,4-BT production, so much so that 1,2,4-BT production was unable to be measured using DNA scaffolds because the fusion of zinc finger triplets eliminated native enzyme activity. However, the basic techniques demonstrated here of optimizing linker and orientation (Chapter 3), as well as other developed methods for engineering better folded proteins [126,127], can help to overcome this fundamental limitation.

After engineering the metabolic enzymes, it is time to consider the myriad of parameters governing the effectiveness of the DNA scaffold. The most important of which is likely the balance between the number of DNA binding sites and the number of enzymes that will bind. A simple assay will help will to approximate this value, for a single scaffold geometry, vary the number of repeats on a given plasmid to determine which gives the optimum metabolite levels. In more detail, intracellular protein levels can be compared to a dilution series of a purified sample by Western Blot and plasmid DNA levels can be similarly quantified by RT-PCR or by plasmid miniprep, after determining the cell concentration (cfu/mL). When the number of DNA binding sites is off by more than an order of magnitude, it is best to vary the plasmid copy number to adjust the number of binding sites. For small changes in the number of DNA binding sites, simply increasing or decreasing the numbers of repeating units will suffice. Further optimization of the scaffolds themselves, includes parameters such as: (1) site stoichiometry, (2) site ordering, e.g. ABBCC vs. CBABC, (3) site spacing along the DNA, (4) interaction strength of the DNA-zinc finger complex, (5) site location on the plasmid. Many of these parameters have not been explored (Chapter 4) and the importance of each cannot be ascribed. Given the many degrees of freedom granted to the engineer, this should become a powerful technique for engineering metabolic pathway.

Future work

While the use of DNA scaffolds has been very successful in increasing metabolite production, the understanding behind the working of these systems is lacking. This stems from the challenges associated with multimeric enzymes and the difficulty in carefully balancing all of the components in these systems, even if all of the proper controls have been considered. Simulations of enzyme colocalization have been useful in furthering the theoretical understanding [12], but a gap between what is

being simulated and the actual system *in vivo*, and slow computational times, hinder the study of successful cases *in vivo*. In this way, *in vitro* assays could be instrumental in decoupling the effects of binding, activity, and synergy observed *in vivo*. The ability to measure individual enzyme activity, with and without fusion to zinc fingers, and with and without DNA binding will be important. The multimeric state of these complexes could be assessed by chromatography to determine (1) to what extent multiple DNA molecules are involved in the same enzyme complex and (2) if DNA binding prevents full complex formation. It would further be possible to study a variety of DNA scaffolds, enzyme stoichiometries, and flux balanced or imbalanced scenarios to determine when enzyme colocalization is most beneficial. This is turn will be very useful for the metabolic engineer in guiding future design of these systems and in envisioning new techniques for enzyme colocalization.

These challenges notwithstanding, there are untapped opportunities for enzyme colocalization moving beyond the channeling of small molecules. As our simulations have demonstrated [12], the impact of enzyme compartmentalization is greatest for slow diffusing molecules, such as proteins. As proteins serve as metabolic intermediate in pathways of post-translational modification, including protein folding, the ability to reduce diffusional limitations could have an enormous impact. Of particular interest are pathways for proper disulfide bond formation and glycan formation for protein glycosylation. Challenges exist for both as disulfide bond formation would need to employ an oxidizing cytoplasm, such as exists in *E. coli* Shuffle (NEB), and glycan formation may need to employ a membrane binding motif, to localize the DNA where the reactions occur. However, these illustrate the enormous potential for enzyme colocalization as engineers begin to consider the benefits from coupled enzyme kinetics and further refinement in the techniques allows for greater ease of implementation.

APPENDIX

Table A.1. Kinetic Rate Expressions

Rate expressions used in this study were borrowed from Reuss [87]. Additional rate equations used in this work are listed below.

$$r_{TCA1} = \frac{kcat_{TCA1} E_{TCA1} C_{OAA} C_{AcCoA}}{(K_{TCA1,oaa} + C_{OAA})(K_{TCA1,accoa} + C_{AcCoA})}$$

$$r_{ISCDH} = \frac{kcat_{ISCDH} E_{ISCDH} C_{ISC} C_{NAD}}{(K_{ISCDH,isc} + C_{ISC})(K_{ISCDH,nad} + C_{NAD})}$$

$$r_{aKGDH} = \frac{kcat_{aKGDH} E_{aKGDH} C_{aKG} C_{NAD}}{(K_{aKGDH,oaa} + C_{aKG})(K_{aKGDH,accoa} + C_{NAD})}$$

$$r_{SucCoAS} = \frac{kcat_{SucCoAS} E_{SucCoAS} C_{SucCoA} C_{ADP}}{(K_{SucCoAS,succoa} + C_{SucCoA})(K_{SucCoAS,adp} + C_{ADP})}$$

$$r_{SUCDH} = \frac{kcat_{SUCDH} E_{SUCDH} C_{SUCC} C_{FAD}}{(K_{SUCDH,succ} + C_{SUCC})(K_{SUCDH,fad} + C_{FAD})}$$

$$r_{FMase} = \frac{kcat_{FMase} E_{FMase} C_{FUM}}{(K_{FMase,fum} + C_{FUM})}$$

$$r_{MALDH} = \frac{kcat_{MALDH} E_{MALDH} C_{MAL} C_{NAD}}{(K_{MALDH,mal} + C_{MAL})(K_{MALDH,nad} + C_{NAD})}$$

$$r_{MAE} = \frac{kcat_{MAE} E_{MAE} C_{MAL}}{(K_{MAE,mal} + C_{MAL})}$$

$$r_{CFSynth} = \frac{kcat_{CFSynth} E_{CFSynth} C_{ISC}}{(K_{CFSynth,isc} + C_{ISC})}$$

$$r_{AASynth} = \frac{kcat_{AASynth} E_{AASynth} C_{aKG}}{(K_{AASynth,aKG} + C_{aKG})}$$

$$r_{AASynt2} = \frac{kcat_{AASynt2} E_{AASynt2} C_{OAA}}{(K_{AASynt2,oaa} + C_{OAA})}$$

$$r_{AASynt3} = \frac{kcat_{AASynt3} E_{AASynt3} C_{AcCoA}}{(K_{AASynt3,accoa} + C_{AcCoA})}$$

$$r_{CI1,f} = \frac{kcat_{CI1} E_{CI1} \frac{C_{NADH} C_{NADP}}{K_{CI1,eq}}}{(K_{CI1,nad}(1 + \frac{C_{NADH}}{K_{CI1,nadh}}) + C_{NAD})(K_{CI1,nadph}(1 + \frac{C_{NADP}}{K_{CI1,nadp}}) + C_{NADPH})}$$

$$r_{CI1,r} = \frac{kcat_{CI1} E_{CI1} C_{NADPH} C_{NAD}}{(K_{CI1,nad}(1 + \frac{C_{NADH}}{K_{CI1,nadh}}) + C_{NAD})(K_{CI1,nadph}(1 + \frac{C_{NADP}}{K_{CI1,nadp}}) + C_{NADPH})}$$

$$r_{OxPhos1} = \frac{kcat_{OxPhos1} E_{OxPhos1} C_{NADH} C_{ADP}}{(K_{OxPhos1,nadh} + C_{NADH})(K_{OxPhos1,adp} + C_{ADP})}$$

$$r_{OxPhos2} = \frac{kcat_{OxPhos2} E_{OxPhos2} C_{FADH_2} C_{ADP}}{(K_{OxPhos2,fadh_2} + C_{FADH_2})(K_{OxPhos2,adp} + C_{ADP})}$$

$$r_{CI2} = \frac{kcat_{CI2} E_{CI2} C_{ATP}}{(K_{CI2,atp} + C_{ATP})}$$

$$r_{CI3} = \frac{kcat_{CI3} E_{CI3} C_{ATP} C_{AMP}}{(K_{CI3,atp} + C_{ATP})(K_{CI3,amp} + C_{AMP})}$$

$$r_{AC} = \frac{kcat_{AC} E_{AC} C_{ATP}}{(K_{AC,atp} + C_{ATP})}$$

$$r_{NADPHcons} = \frac{kcat_{NADPHcons} E_{NADPHcons} C_{NADPH}}{(K_{NADPHcons,nadph} + C_{NADPH})}$$

$$r_{NADHcons} = \frac{kcat_{NADHcons} E_{NADHcons} C_{NADH}}{(K_{NADHcons,nadh} + C_{NADH})}$$

$$r_{FADH_2cons} = \frac{kcat_{FADH_2cons} E_{FADH_2cons} C_{FADH_2}}{(K_{FADH_2cons,fadh_2} + C_{FADH2})}$$

$$r_{MGS} = \frac{kcat_{MGS} E_{MGS} C_{DHAP}}{(K_{MGS,dhap} + C_{DHAP})}$$

$$r_{DhaD} = \frac{kcat_{DhaD} E_{DhaD} C_{MG} C_{NADH}}{(K_{DhaD,mg} + C_{MG})(K_{DhaD,nadh} + C_{NADH})}$$

$$r_{AdhI} = \frac{kcat_{AdhI} E_{AdhI} C_{R-lac} C_{NADH}}{(K_{AdhI,R-lac} + C_{R-lac})(K_{AdhI,nadh} + C_{NADH})}$$

Table A.2. Kinetic parameters.

Enzyme activities	Parameters	Parameter values
TCA1	$Kcat_{TCA1}$	176.63 s^{-1}
	$K_{TCA1,accoa}$	1 mM
	$K_{TCA1,oaa}$	1 mM
ISCDH	$Kcat_{ISCDH}$	87.85 s^{-1}
	$K_{ISCDH,isc}$	1 mM
	$K_{ISCDH,nad}$	0.2 mM
aKGDH	$Kcat_{aKGDH}$	85.35 s^{-1}
	$K_{aKGDH,aKG}$	1 mM
	$K_{aKGDH,nad}$	0.2 mM
SucCoAS	$Kcat_{SucCoAS}$	100.38 s^{-1}
	$K_{SuCCoAS,succoa}$	1 mM
	$K_{SucCoAS,adp}$	0.2 mM
SUCDH	$Kcat_{SUCDH}$	150.24 s^{-1}
	$K_{SUCDH,succ}$	1 mM
	$K_{SUCDH,fad}$	0.2 mM
Fmase	$Kcat_{FMase}$	75.12 s^{-1}
	$K_{FMase,fum}$	1 mM
MALDH	$Kcat_{MALDH}$	76.55 s^{-1}
	$K_{MALDH,mal}$	1 mM
	$K_{MALDH,nad}$	0.2 mM
MAE	$Kcat_{MAE}$	7.73 s^{-1}
	$K_{MAE,mal}$	1 mM
CFSynth	$Kcat_{CFSynth}$	11.06 s^{-1}
	$K_{CFSynth,isc}$	1 mM
AASynth	$Kcat_{AASynth}$	2.20 s^{-1}

	$K_{AASynth,aKG}$	1 mM
AASynth2	$Kcat_{AASynth2}$	6.63 s^{-1}
	$K_{AASynth2,oaa}$	1 mM
AASynth3	$Kcat_{AASynth3}$	23.20 s^{-1}
	$K_{AASynth3,accoa}$	1 mM
CI1	$Kcat_{CI1}$	1225.44 s^{-1}
	$K_{CI1,nadh}$	0.2 mM
	$K_{CI1,nadph}$	0.2 mM
	$K_{CI1,eq}$	0.214
	$K_{CI1,nad}$	0.2 mM
	$K_{CI1,nadp}$	0.2 mM
OxPhos	$Kcat_{OxPhos}$	486.91 s^{-1}
	$K_{OxPhos,nadh}$	0.2 mM
	$K_{OxPhos,adp}$	0.2 mM
OxPhos2	$Kcat_{OxPhos2}$	162.30 s^{-1}
	$K_{OxPhos2,fadh}$	0.2 mM
	$K_{OxPhos2,adp}$	0.2 mM
CI2	$Kcat_{CI2}$	418.59 s^{-1}
	$K_{CI2,atp}$	0.2 mM
CI3	$Kcat_{CI3}$	70.05 s^{-1}
	$K_{CI3,atp}$	0.2 mM
	$K_{CI3,amp}$	0.2 mM
AC	$Kcat_{AC}$	57.89 s^{-1}
	$K_{AC,atp}$	0.2 mM
	$K_{AC,g6p,ss}$	3.48 mM
NADPHcons	$Kcat_{NADPHcons}$	326.64 s^{-1}
	$K_{NADPHcons,nadph}$	0.2 mM
NADHcons	$Kcat_{NADHcons}$	397.93 s^{-1}
	$K_{NADHcons,nadh}$	0.2 mM
$FADH_2$cons	$Kcat_{FADH2cons}$	66.32 s^{-1}
	$K_{FADH2cons,fadh2}$	0.2 mM
MGS	$Kcat_{MGS}$	220 s^{-1}
	$K_{MGS,dhap}$	0.2 mM
DHAD	$Kcat_{DHAD}$	220 s^{-1}
	$K_{DHAD,mgs}$	0.4 mM
	$K_{DHAD,nadh}$	0.2 mM
ALDI	$Kcat_{ALDI}$	220 s^{-1}
	$K_{ALDI,R-lac}$	0.2 mM
	$K_{ALDI,nadh}$	0.2 mM

REFERENCES

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

2. Jiang H, Wood KV, Morgan JA (2005) Metabolic engineering of the phenylpropanoid pathway in *Saccharomyces cerevisiae*. Appl Environ Microbiol 71: 2962-2969.

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

4. Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci 26: 597-604.

5. Arrio-Dupont M, Foucault G, Vacher M, Devaux PF, Cribier S (2000) Translational diffusion of globular proteins in the cytoplasm of cultured muscle cells. Biophys J 78: 901-907.

6. Kao HP, Abney JR, Verkman AS (1993) Determinants of the translational mobility of a small solute in cell cytoplasm. J Cell Biol 120: 175-184.

7. Jorgensen K, Rasmussen AV, Morant M, Nielsen AH, Bjarnholt N, et al. (2005) Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. Curr Opin Plant Biol 8: 280-291.

8. Meynial Salles I, Forchhammer N, Croux C, Girbal L, Soucaille P (2007) Evolution of a *Saccharomyces cerevisiae* metabolic pathway in *Escherichia coli*. Metab Eng 9: 152-159.

9. Charles IG, Keyte JW, Brammar WJ, Smith M, Hawkins AR (1986) The isolation and nucleotide sequence of the complex AROM locus of *Aspergillus nidulans*. Nucleic Acids Res 14: 2201-2213.

10. Hyde CC, Ahmed SA, Padlan EA, Miles EW, Davies DR (1988) Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*. J Biol Chem 263: 17857-17871.

11. Kristensen C, Morant M, Olsen CE, Ekstrom CT, Galbraith DW, et al. (2005) Metabolic engineering of dhurrin in transgenic Arabidopsis plants with marginal inadvertent effects on the metabolome and transcriptome. Proc Natl Acad Sci U S A 102: 1779-1784.

12. Conrado RJ, Mansell TJ, Varner JD, DeLisa MP (2007) Stochastic reaction-diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway. Metab Eng 9: 355-363.

13. Kholodenko BN, Westerhoff HV, Cascante M (1996) Effect of channelling on the concentration of bulk-phase intermediates as cytosolic proteins become more concentrated. Biochem J 313 (Pt 3): 921-926.

14. An S, Kumar R, Sheets ED, Benkovic SJ (2008) Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. Science 320: 103-106.

15. Campanella ME, Chu H, Low PS (2005) Assembly and regulation of a glycolytic enzyme complex on the human erythrocyte membrane. Proc Natl Acad Sci U S A 102: 2402-2407.

16. Ishikawa M, Tsuchiya D, Oyama T, Tsunaka Y, Morikawa K (2004) Structural basis for channelling mechanism of a fatty acid beta-oxidation multienzyme complex. Embo J 23: 2745-2754.

17. Suss KH, Arkona C, Manteuffel R, Adler K (1993) Calvin cycle multienzyme complexes are bound to chloroplast thylakoid membranes of higher plants in situ. Proc Natl Acad Sci U S A 90: 5514-5518.

18. Welch GR, Gaertner FH (1980) Enzyme organization in the polyaromatic-biosynthetic pathway: the arom conjugate and other multienzyme systems. Curr Top Cell Regul 16: 113-162.

19. Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, et al. (2005) Protein structures forming the shell of primitive bacterial organelles. Science 309: 936-938.

20. Schmidt M, Hanna J, Elsasser S, Finley D (2005) Proteasome-associated proteins: regulation of a proteolytic machine. Biol Chem 386: 725-737.

21. Winkel BS (2004) Metabolic channeling in plants. Annu Rev Plant Biol 55: 85-107.

22. Hawkins AR, Smith M (1991) Domain structure and interaction within the pentafunctional arom polypeptide. Eur J Biochem 196: 717-724.

23. Hawkins AR (1987) The complex Arom locus of *Aspergillus nidulans*. Evidence for multiple gene fusions and convergent evolution. Curr Genet 11: 491-498.

24. Houben KF, Dunn MF (1990) Allosteric effects acting over a distance of 20-25 A in the *Escherichia coli* tryptophan synthase bienzyme complex increase ligand affinity and cause redistribution of covalent intermediates. Biochemistry 29: 2421-2429.

25. Yanofsky C (1960) The tryptophan synthetase system. Bacteriol Rev 24: 221-245.

26. Pfeifer BA, Khosla C (2001) Biosynthesis of polyketides in heterologous hosts. Microbiol Mol Biol Rev 65: 106-118.

27. Leivar P, Gonzalez VM, Castel S, Trelease RN, Lopez-Iglesias C, et al. (2005) Subcellular localization of Arabidopsis 3-hydroxy-3-methylglutaryl-coenzyme A reductase. Plant Physiol 137: 57-69.

28. Panicot M, Minguet EG, Ferrando A, Alcazar R, Blazquez MA, et al. (2002) A polyamine metabolon involving aminopropyl transferase complexes in Arabidopsis. Plant Cell 14: 2539-2551.

29. Burbulis IE, Winkel-Shirley B (1999) Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. Proc Natl Acad Sci U S A 96: 12929-12934.

30. Winkel-Shirley B (1999) Evidence for enzyme complexes in phenylpropanoid and flavonoid pathways. Physiol Plant 107: 142-149.

31. Kahn RA, Fahrendorf T, Halkier BA, Moller BL (1999) Substrate specificity of the cytochrome P450 enzymes CYP79A1 and CYP71E1 involved in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. Arch Biochem Biophys 363: 9-18.

32. Bak S, Kahn RA, Nielsen HL, Moller BL, Halkier BA (1998) Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) Moench by a PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. Plant Mol Biol 36: 393-405.

33. Moller BL, Conn EE (1980) The biosynthesis of cyanogenic glucosides in higher plants. Channeling of intermediates in dhurrin biosynthesis by a microsomal system from *Sorghum bicolor* (linn) Moench. J Biol Chem 255: 3049-3056.

34. Kern A, Tilley E, Hunter IS, Legisa M, Glieder A (2007) Engineering primary metabolic pathways of industrial micro-organisms. J of Biotechnol 129: 6-29.

35. Mosbach K, Mattiasson B (1970) Matrix Bound Enzymes, Part II: Studies on a Matrix-Bound Two-Enzyme System. Acta Chemica Scandinavica 24: 2093-2100.

36. Srere PA, Mattiasson B, Mosbach K (1973) An immobilized three-enzyme system: a model for microenvironmental compartmentation in mitochondria. Proc Natl Acad Sci U S A 70: 2534-2538.

37. Mansson MO, Siegbahn N, Mosbach K (1983) Site-to-site directed immobilization of enzymes with bis-NAD analogues. Proc Natl Acad Sci U S A 80: 1487-1491.

38. Koch-Schmidt AC, Mattiasson B, Mosbach K (1977) Aspects of microenvironmental compartmentation. An evaluation of the influence of restricted diffusion, exclusion effects, and enzyme proximity on the overall efficiency of the sequential two-enzyme system malate dehydrogenase--citrate synthase in its soluble and immobilized form. Eur J Biochem 81: 71-78.

39. Fierobe HP, Mechaly A, Tardif C, Belaich A, Lamed R, et al. (2001) Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional complexes. J Biol Chem 276: 21257-21261.

40. Bülow L, Ljungcrantz P, Mosbach K (1985) Preparation of a Soluble Bifunctional Enzyme by Gene Fusion. Bio/Technology 3: 821-823.

41. Ljungcrantz P, Carlsson H, Mansson MO, Buckel P, Mosbach K, et al. (1989) Construction of an artificial bifunctional enzyme, beta-galactosidase/galactose dehydrogenase, exhibiting efficient galactose channeling. Biochemistry 28: 8786-8792.

42. Pettersson H, Pettersson G (2001) Kinetics of the coupled reaction catalysed by a fusion protein of beta-galactosidase and galactose dehydrogenase. Biochim Biophys Acta 1549: 155-160.

43. Pettersson H, Olsson P, Bulow L, Pettersson G (2000) Kinetics of the coupled reaction catalysed by a fusion protein of yeast mitochondrial malate dehydrogenase and citrate synthase. Eur J Biochem 267: 5041-5046.

44. Prachayasittikul V, Ljung S, Isarankura-Na-Ayudhya C, Bulow L (2006) NAD(H) recycling activity of an engineered bifunctional enzyme galactose dehydrogenase/lactate dehydrogenase. Int J Biol Sci 2: 10-16.

45. Bulow L (1987) Characterization of an artificial bifunctional enzyme, beta-galactosidase/galactokinase, prepared by gene fusion. Eur J Biochem 163: 443-448.

46. Kourtz L, Dillon K, Daughtry S, Madison LL, Peoples O, et al. (2005) A novel thiolase-reductase gene fusion promotes the production of polyhydroxybutyrate in Arabidopsis. Plant Biotechnol J 3: 435-447.

47. Riedel K, Bronnenmeier K (1998) Intramolecular synergism in an engineered exo-endo-1,4-beta-glucanase fusion protein. Mol Microbiol 28: 767-775.

48. Shibuya I, Tamura G, Shima H, Ishikawa T, Hara S (1992) Construction of an alpha-amylase/glucoamylase fusion gene and its expression in *Saccharomyces cerevisiae*. Biosci Biotechnol Biochem 56: 884-889.

49. Mao Q, Schunk T, Gerber B, Erni B (1995) A string of enzymes, purification and characterization of a fusion protein comprising the four subunits of the glucose phosphotransferase system of *Escherichia coli*. J Biol Chem 270: 18295-18300.

50. Seo HS, Koo YJ, Lim JY, Song JT, Kim CH, et al. (2000) Characterization of a bifunctional enzyme fusion of trehalose-6-phosphate synthetase and trehalose-6-phosphate phosphatase of *Escherichia coli*. Appl Environ Microbiol 66: 2484-2490.

51. Kim YH, Kwon TK, Park S, Seo HS, Cheong JJ, et al. (2000) Trehalose synthesis by sequential reactions of recombinant maltooligosyltrehalose synthase and maltooligosyltrehalose trehalohydrolase from *Brevibacterium helvolum*. Appl Environ Microbiol 66: 4620-4624.

52. de Pascale D, Di Lernia I, Sasso MP, Furia A, De Rosa M, et al. (2002) A novel thermophilic fusion enzyme for trehalose production. Extremophiles 6: 463-468.

53. Wang JH, Tsai MY, Lee GC, Shaw JF (2007) Construction of a recombinant thermostable beta-amylase-trehalose synthase bifunctional enzyme for facilitating the conversion of starch to trehalose. J Agric Food Chem 55: 1256-1263.

54. Orita I, Sakamoto N, Kato N, Yurimoto H, Sakai Y (2007) Bifunctional enzyme fusion of 3-hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase. Appl Microbiol Biotechnol 76: 439-445.

55. Levasseur A, Navarro D, Punt PJ, Belaich JP, Asther M, et al. (2005) Construction of engineered bifunctional enzymes and their overproduction in *Aspergillus niger* for improved enzymatic tools to degrade agricultural by-products. Appl Environ Microbiol 71: 8132-8140.

56. Mingardon F, Chanal A, Lopez-Contreras AM, Dray C, Bayer EA, et al. (2007) Incorporation of fungal cellulases in bacterial minicellulosomes yields viable, synergistically acting cellulolytic complexes. Appl Environ Microbiol 73: 3822-3832.

57. Fierobe HP, Bayer EA, Tardif C, Czjzek M, Mechaly A, et al. (2002) Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. J Biol Chem 277: 49621-49630.

58. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, et al. (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27: 753-U107.

59. Moon TS, Dueber JE, Shiue E, Prather KLJ (2010) Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. Metab Eng 12: 298-305.

60. Yilmaz JL, Bulow L (2002) Enhanced stress tolerance in *Escherichia coli* and *Nicotiana tabacum* expressing a betaine aldehyde dehydrogenase/choline dehydrogenase fusion protein. Biotechnol Prog 18: 1176-1182.

61. Ljungcrantz P, Bulow L, Mosbach K (1990) Construction and characterization of a recombinant tripartite enzyme, galactose dehydrogenase/beta-galactosidase/galactokinase. FEBS Lett 275: 91-94.

62. Bulow L, Mosbach K (1991) Multienzyme systems obtained by gene fusion. Trends Biotechnol 9: 226-231.

63. Netzer WJ, Hartl FU (1997) Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature 388: 343-349.

64. Chang HC, Kaiser CM, Hartl FU, Barral JM (2005) *De novo* folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria. J Mol Biol 353: 397-409.

65. Mingardon F, Chanal A, Tardif C, Bayer EA, Fierobe HP (2007) Exploration of new geometries in cellulosome-like chimeras. Appl Environ Microbiol 73: 7138-7149.

66. Dueber JE, Mirsky EA, Lim WA (2007) Engineering synthetic signaling proteins with ultrasensitive input/output control. Nat Biotechnol 25: 660-662.

67. Srere PA (1987) Complexes of sequential metabolic enzymes. Annu Rev Biochem 56: 89-124.

68. Conrado RJ, Mansell TJ, DeLisa MP (2010) Engineering multifunctional enzyme systems for optimized metabolite transfer between sequential conversion steps. In: Smolke CD, editor. The Metabolic Pathway Engineering Handbook: Tools and Applications. Boca Raton: CRC Press.

69. Green D (1957) Studies in organized enzyme systems. The Harvey Lectures 52: 177-227.

70. Bobik TA (2006) Polyhedral organelles compartmenting bacterial metabolic processes. Appl Microbiol Biotechnol 70: 517-525.

71. Straight PD, Fischbach MA, Walsh CT, Rudner DZ, Kolter R (2007) A singular enzymatic megacomplex from *Bacillus subtilis*. Proc Nat Acad Sci U S A 104: 305-310.

72. Beaujean A, Ducrocq-Assaf C, Sangwan RS, Lilius G, Bulow L, et al. (2000) Engineering direct fructose production in processed potato tubers by expressing a bifunctional alpha-amylase/glucose isomerase gene complex. Biotechnol Bioeng 70: 9-16.

73. Lindbladh C, Rault M, Hagglund C, Small WC, Mosbach K, et al. (1994) Preparation and kinetic characterization of a fusion protein of yeast mitochondrial citrate synthase and malate dehydrogenase. Biochemistry 33: 11692-11698.

74. Shatalin K, Lebreton S, Rault-Leonardon M, Velot C, Srere PA (1999) Electrostatic channeling of oxaloacetate in a fusion protein of porcine citrate synthase and porcine mitochondrial malate dehydrogenase. Biochemistry 38: 881-889.

75. Tian L, Dixon RA (2006) Engineering isoflavone metabolism with an artificial bifunctional enzyme. Planta 224: 496-507.

76. Cascante M, Sorribas A, Canela EI (1994) Enzyme-enzyme interactions and metabolite channelling: alternative mechanisms and their evolutionary significance. Biochem J 298 (Pt 2): 313-320.

77. Cornish-Bowden A (1991) Failure of channelling to maintain low concentrations of metabolic intermediates. Eur J Biochem 195: 103-108.

78. Cornish-Bowden A, Cardenas ML (1993) Channelling can affect concentrations of metabolic intermediates at constant net flux: artefact or reality? Eur J Biochem 213: 87-92.

79. Degenring D, Rohl M, Uhrmacher AM (2004) Discrete event, multi-level simulation of metabolite channeling. Biosystems 75: 29-41.

80. Maher AD, Kuchel PW, Ortega F, de Atauri P, Centelles J, et al. (2003) Mathematical modelling of the urea cycle. A numerical investigation into substrate channelling. Eur J Biochem 270: 3953-3961.

81. Mendes P, Kell DB, Westerhoff HV (1992) Channelling can decrease pool size. Eur J Biochem 204: 257-266.

82. Altaras NE, Cameron DC (1999) Metabolic engineering of a 1,2-propanediol pathway in *Escherichia coli*. Appl Environ Microbiol 65: 1180-1185.

83. Altaras NE, Cameron DC (2000) Enhanced production of (R)-1,2-propanediol by metabolically engineered *Escherichia coli*. Biotechnol Prog 16: 940-946.

84. Altaras NE, Etzel MR, Cameron DC (2001) Conversion of sugars to 1,2-propanediol by *Thermoanaerobacterium thermosaccharolyticum* HG-8. Biotechnol Prog 17: 52-56.

85. Elf J, Ehrenberg M (2004) Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. Systems Biology 1: 230-236.

86. Gillespie DT (1977) Exact Stochastic Simulation of Coupled Chemical Reactions. J Phys Chem 81: 2340-2361.

87. Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M (2002) Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotech Bioeng 79: 53-73.

88. Sauer U, Eikmanns BJ (2005) The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. FEMS Microbiol Rev 29: 765-794.

89. Sceller LL, Ripoll C, Demarty M, Cabin-Flamand A, Nystrom T, et al. (2000) Modelling bacterial hyperstructures with cellular automata. Interjournal Paper 366.

90. Elowitz MB, Surette MG, Wolf PE, Stock JB, Leibler S (1999) Protein mobility in the cytoplasm of *Escherichia coli*. J Bacteriol 181: 197-203.

91. Mastro AM, Babich MA, Taylor WD, Keith AD (1984) Diffusion of a small molecule in the cytoplasm of mammalian cells. Proc Natl Acad Sci U S A 81: 3414-3418.

92. Smith LJ, Talon C, Brady JW, Chowdhuri Z, Copley JRD, et al. (2003) Dynamics of Glucose Solutions. CHRNS Science Highlights 20.

93. Chang MCY, Eachus RA, Trieu W, Ro D-K, Keasling JD (2007) Engineering *Escherichi coli* for production of functionalized terpenoids using plant P450s. Nat Chem Biol 3: 274-277.

94. Hrazdina G (1992) Compartmentation in aromatic metabolism. In: Stafford H, Ibrahim R, editors. Recent Advances in Phytochemistry. New York: Plenum Press. pp. 1-23.

95. Cane DE, Walsh CT, Khosla C (1998) Harnessing the biosynthetic code: combinations, permutations, and mutations. Science 282: 63-68.

96. Conrado RJ, Varner JD, DeLisa MP (2008) Engineering the compartmentalizaion of metabolic enzymes: mimicking nature's synergy. Curr Opin Biotech 19: 492-499.

97. Sambook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

98. Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol 177: 4121-4130.

99. Waraho D, DeLisa MP (2009) Versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism. Proc Nat Acad Sci U S A 106: 3692-3697.

100. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular Systems Biology 2.

101. Niu W, Molefe MN, Frost JW (2003) Microbial synthesis of the energetic material precursor 1,2,4-butanetriol. J Am Chem Soc 125: 12998-12999.

102. Frost JW, Niu W (2008) Microbial Synthesis of *D*-1,2,4-Butanetriol. In: Organization WIP, editor. United States.

103. Frost JW. Creating a Pathway for the Biosynthesis of 1,2,4-Butanetriol; 2008.

104. Keseler IM, Collado-Vides J, Gama-Castro S, Ingraham J, Paley S, et al. (2005) EcoCyc: a comprehensive database resource for *Escherichia coli*. Nucleic Acids Res 33: D334-337.

105. Argos P (1990) An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. J Mol Biol 211: 943-958.

106. Dragan AI, Frank L, Liu YY, Makeyeva EN, Crane-Robinson C, et al. (2004) Thermodynamic signature of GCN4-bZIP binding to DNA indicates the role of water in discriminating between the AP-1 and ATF/CREB sites. J Mol Biol 343: 865-878.

107. Pernelle C, Clerc FF, Dureuil C, Bracco L, Tocque B (1993) An Efficient Screening Assay for the Rapid and Precise Determination of Affinities between Leucine-Zipper Domains. Biochemistry 32: 11682-11687.

108. Patel LR, Curran T, Kerppola TK (1994) Energy-Transfer Analysis of Fos-Jun Dimerization and DNA-Binding. Proc Nat Acad Sci U S A 91: 7360-7364.

109. Oyama R, Takashima H, Yonezawa M, Doi N, Miyamoto-Sato E, et al. (2006) Protein-protein interaction analysis by C-terminally specific fluorescence labeling and fluorescence cross-correlation spectroscopy. Nuc Acids Res 34: e102.

110. Heuer KH, Mackay JP, Podzebenko P, Bains NPS, Weiss AS, et al. (1996) Development of a sensitive peptide-based immunoassay: Application to detection of the Jun and Fos oncoproteins. Biochemistry 35: 9069-9075.

111. Leder L, Berger C, Bornhauser S, Wendt H, Ackermann F, et al. (1995) Spectroscopic, calorimetric, and kinetic demonstration of conformational adaptation in peptide antibody recognition. Biochemistry 34: 16509-16518.

112. Morgan-Kiss RM, Wadler C, Cronan JE (2002) Long-term and homogeneous regulation of the *Escherichia coli* araBAD promoter by use of a lactose transporter of relaxed specificity. Proc Nat Acad Sci U S A 99: 7373-7377.

113. Kapust RB, Waugh DS (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 8: 1668-1674.

114. Frost JW (2006) Byproduct formation during *D*-BT biosynthesis.

115. Saadat D, Harrison DHT (1998) Identification of catalytic bases in the active site of *Escherichia coli* methylglyoxal synthase: Cloning, expression, and functional characterization of conserved aspartic acid residues. Biochemistry 37: 10074-10086.

116. Ko J, Kim I, Yoo S, Min B, Kim K, et al. (2005) Conversion of methylglyoxal to acetol by *Escherichia coli* aldo-keto Reductases. J Bacteriol 187: 5782-5789.

117. DeLisa MP, Conrado RJ (2009) Synthetic metabolic pipelines. Nat Biotechnol 27: 728-729.

118. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, et al. (2008) Rapid "Open-Source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31: 294-301.

119. Yanischperron C, Vieira J, Messing J (1985) Improved M13 Phage Cloning Vectors and Host Strains - Nucleotide-Sequences of the M13mp18 and Puc19 Vectors. Gene 33: 103-119.

120. Sander JD, Zaback P, Joung JK, Voytas DF, Dobbs D (2009) An affinity-based scoring scheme for predicting DNA-binding activities of modularly assembled zinc-finger proteins. Nuc Acids Res 37: 506-515.

121. Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I-1-I-2 regulatory elements. Nuc Acids Res 25: 1203-1210.

122. Havemann GD, Bobik TA (2003) Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. J Bacteriol 185: 5086-5095.

123. Petersson G (1991) No convincing evidence is available for metabolic channeling between enzymes forming dynamic complexes. J Theor Biol 152: 65-69.

124. Yomano LP, York SW, Ingram LO (1998) Isolation and characterization of ethanol-tolerant mutants of *Escherichia coli* KO11 for fuel ethanol production. J Indust Microbiol Biotechnol 20: 132-138.

125. Hartl FU, Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295: 1852-1858.

126. Fisher AC, Kim W, DeLisa MP (2006) Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. Protein Sci 15: 449-458.

127. Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. Science 324: 255-258.