CHAPTER NINETEEN

Metabolic Pathway Flux Enhancement by Synthetic Protein Scaffolding

Weston R. Whitaker* *and* John E. Dueber*,†

Contents

Abstract

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

* Department of Bioengineering, University of California, Berkeley, California, USA
† Energy Biosciences Institute, University of California, Berkeley, California, USA

Methods in Enzymology, Volume 497
ISSN 0076-6879, DOI: 10.1016/B978-0-12-385075-1.00019-6

1. Introduction

Metabolic engineering has the potential to provide environmentally safe and cost-effective routes for synthesizing a range of compounds, from high-value specialty compounds such as therapeutics to bulk commodities including plastics and biofuels. Particularly for the latter class of compounds, a complement of strategies will be needed to achieve the production yields, near theoretical maximum, necessary to achieve industrial viability. These stringent requirements will likely inspire improvements across many technologies: modeling metabolic and cellular behavior (Price *et al.*, 2004), predictable control over gene expression (Pfleger *et al.*, 2006; Salis *et al.*, 2009; Win and Smolke, 2007), and directed evolution approaches for improved enzyme characteristics (Dougherty and Arnold, 2009; Zhang *et al.*, 2008). In this chapter, we focus on ongoing efforts to improve pathway efficiency through engineered enzyme complex formation using synthetic scaffolds; however, all strategies discussed here must eventually be performed in concert with existing proven methodologies to achieve optimal yields. Since our mechanistic understanding of scaffold function is still at an early stage, we describe here the parameters empirically derived thus far to be important and describe a suggested process for applying scaffolding strategies to a new pathway.

There are numerous natural examples of enzymes forming complexes for optimal metabolic pathway performance. For excellent in-depth reviews on this topic, please see those written by Conrado *et al.* (2008) and Miles *et al.* (1999). The most striking examples of improved pathway efficiency via complex formation are those that have evolved structures capable of physically channeling substrates. Tryptophan synthase, carbamoyl phosphate synthase, and glutamine phosphoribosylpyrophosphate amidotransferase are three examples described in detail by Miles *et al.* (1999) whose structures reveal tunnels connecting catalytic sites that are capable of protecting reactive intermediates from the bulk solution. Another mechanism of channeling substrates is through electrostatic channeling. Thymidylate synthase and dihydrofolate reductase are two enzymatic activities found in a single polypeptide in some plants and some protozoa, including *Leishmania major* for which the crystal structure has been solved. The surface of this structure is predominantly positively charged, suggesting a mechanism of an electrostatic "highway" spanning the 40 Å between the two active sites across which the negatively charged dihydrofolate intermediate would travel (Stroud, 1994). Recently, evidence has grown for the dynamic assembly of complexes, perhaps as a feedback mechanism to achieve a precise concentration of metabolite product (An *et al.*, 2008; Narayanaswamy *et al.*, 2009). These dynamic complexes have been difficult

to observe biochemically *in vitro*. For example, purine biosynthesis in eukaryotes involves six enzymes. Despite early anticipation of potential interactions between these enzymes, only recently was it understood, by fluorescently tagging these enzymes *in vivo*, that all six proteins coassemble (An *et al.*, 2008). Interestingly, these proteins dynamically assemble and disassemble depending on purine concentration. Narayanaswamy *et al.* (2009) similarly showed that numerous metabolic enzyme complexes dynamically assemble depending on culture conditions, suggesting these phenomena are considerably more common than would be predicted by *in vitro* biochemical experiments. Likely, many of these complexes are not detected due to characterization under conditions incompatible with complex formation.

Drawing inspiration from natural pathways, engineers have begun assembling synthetic enzyme complexes to improve pathway performance. For degradation of cellulose and hemicellulose *in vitro*, various enzyme combinations have been corecruited to cellulose substrate to include synergistic combinations of activities as found in natural cellulosome complexes (Fierobe *et al.*, 2001, 2005). Recently, our lab has expressed scaffolds built from modular protein–protein interaction domains (Table 19.1) to optimize flux of engineered metabolic pathways *in vivo* (Dueber *et al.*, 2009). Enzymes were tagged with peptide ligands specific for these scaffold protein–protein interaction domains. The modular composition of the scaffolds was used to build various architectures that were critical for optimizing flux as discussed later.

The mevalonate biosynthetic pathway presents an interesting model system for synthetic complex engineering in that it suffers from a flux imbalance between HMG-CoA synthase (HMGS) and HMG-CoA reductase (HMGR) that results in the accumulation of the cytotoxic HMG-CoA intermediate (Martin *et al.*, 2003; Pitera *et al.*, 2007). Scaffolding this pathway improves efficiency, producing higher product titers even at considerably lower enzyme inducer concentrations (Dueber *et al.*, 2009). It should be noted that the relationship between scaffold architecture and titer improvement is not predictable for this pathway, as discussed later. The same set of scaffold architectures was applied to a second pathway engineered by Moon *et al.* (2009, 2010) for the biosynthesis of glucaric acid. This pathway presents an interesting test case in that it is a relatively high titer-producing pathway, on the order of 1 g/L, with a flux bottleneck enzyme, MIOX, that appears to be substrate activated (Moon *et al.*, 2009). Varying the number of domains that recruit the enzyme upstream of MIOX, Ino1, resulted in gradually increased product titers to a maximum of almost fivefold improvement with four Ino1 recruitment domains, whereas varying the number of MIOX recruiting domains had little impact (Moon *et al.*, 2010). This observation is consistent with a model in which the local concentration of substrate for the limiting MIOX activity is modulated by upstream enzyme recruitment via scaffold domain stoichiometry within the synthetic complex.

Table 19.1 Protein–protein interaction domain families potentially useful for scaffold construction

Part family	Tightest affinity K_d	Domain/ligand size (AAs)	Source (Accession #)	Features and issues	Confirmed orthogonal pairs
SH3 domain/ peptide	1×10^{-1} u*M* (Posern *et al.*, 1998)	57/11	AAH31149: 196–274 PPPALPPKRRR (Posern *et al.*, 1998)	Relatively context independent and well-characterized. Ideal for internal insertion. Natural peptides tend to have micromolar affinities	Specificity observed within species (Zarrinpar *et al.*, 2003)
PDZ domain/ peptide	1×10^{0} u*M* (Tonikian *et al.*, 2008)	96/6	EDL06069: 77–171 GVKESLV (Harris *et al.*, 2001)	Generally PDZ peptide must be C-terminal. nNOS domain can be used for non-C-terminal ligands	2 Natural (Fuh *et al.*, 2000)
GBD domain/ peptide	1×10^{0} u*M* (Kim *et al.*, 2000)	80/32	BAA21534: 196–274 P42768: 466–497	Less well-characterized than SH3 or PDZ. Longer linker sequence.	1
Leucine Zippers	6.1×10^{-3} u*M* (Acharya *et al.*, 2002) and 8.3×10^{-2} u*M* (Grünberg *et al.*, 2010)	43/43	ITIRAAFLEKENTALRTEIAE LEKEVGRCENIVSKYETRYGPL LEIRAAFLEKENTALRTRAAEL RKRVGRCRNIVSKYETRYGPL	Significant likelihood of homodimerization, particularly important to test for intramolecular pairs	3 Synthetic (Reinke *et al.*, 2010)
PhyB/Pif3 light switchable binding	2×10^{-2}–1×10^{-1} u*M* (Levskaya *et al.*, 2009)	908/91	AAW56577: 1–908 NP_172424: 120–210	Light dependent binding activated at 720 nm and deactivated at 660 nm light	1

FKBP/FRB	1.2×10^{-3} u*M* (Banaszynski *et al.*, 2005)	107/93	AAI19733: 39–145 EAW71681: 1972–2064	Interaction is inducible with the small molecule Rapamycin at $K_d = 0.2$ n*M* (Bierer *et al.*, 1990). A FRB (T2098L) mutation allows use of a nontoxic rapamycin analog for T2098L characterization see (Grünberg *et al.*, 2010)	1
Cohesin/ Dockerin	$<1 \times 10^{-5}$ u*M* (Fierobe *et al.*, 2001)	~150/~70	YP_001039466 YP_001038489	Calcium-dependent binding activity is likely not functional at free cellular Ca^{2+} levels Calcium $K_d = 2.5 \times 10^{-7}$ 1/M^2, half binding at 500 u*M* Ca^{2+} (Leibovitz *et al.*, 1997)	5 Natural (Haimovitz *et al.*, 2008)

Even without direct substrate channeling guiding intermediates between active sites as observed in the natural examples discussed previously, product titers may be improved by colocalizing consecutive metabolic enzymes to produce a higher local concentration of metabolite in close proximity to the downstream enzyme (Conrado *et al.*, 2008; Welch, 1977). This has been the subject of debate in early papers in which consecutive enzymes, β-galactosidase and galactose dehydrogenase, were tethered with a translational fusion, generating a higher product titer (Ljungcrantz *et al.*, 1989). The authors suggested the mechanism for increased flux was substrate channeling; however, a subsequent paper performing a detailed kinetic analysis of the fusion protein challenged this conclusion (Pettersson and Pettersson, 2001). Local concentration effects have been modeled for engineered pathways of heterologous enzymes by simulating native and engineered pathway reaction rates within an *Escherichia coli* discretized into subvolumes, localizing the engineered pathway within a single subvolume, and accounting for metabolite diffusion to simulate compartmentalization (Conrado *et al.*, 2007). Though direct measurements of local intermediate concentrations within enzyme complexes have remained elusive, this mechanism seems to be an attractive explanation for some of the successes observed with scaffolding (Dueber *et al.*, 2009; Moon *et al.*, 2010) and other colocalization engineering examples reviewed in Conrado *et al.* (2008). However, local concentration effects may be acting in conjunction with other mechanisms discussed later in this chapter. Enzyme colocalization may allow achievement of a specific local intermediate concentration with a lower concentration of upstream enzyme than would be possible with freely diffusing enzymes, thus retaining high flux while reducing the metabolic load on the cells (Fig. 19.1A). Reduction of intermediate in the bulk of the cell may also be beneficial if the intermediate is toxic or undergoes undesired reactions through competing pathways (Fig. 19.1A).

We employ modular protein–protein interaction machinery tethered with long flexible synthetic linkers to colocalize enzymes in engineered metabolic pathways. While this approach currently lacks the ability to precisely control the three-dimensional positioning of recruited enzymes, it has the advantage that each protein–protein interaction domain should be capable of targeting its interaction partner in a manner independent of composition context, provided neighboring targeted enzymes do not sterically block a physical interaction. When a binding domain is incorporated into a scaffold using long linkers, it generally will retain the ability to bind its target ligand regardless of where it is located on the scaffold and what domains are encoded up or downstream. This provides a highly designable platform where matrices of scaffolds can be generated in which key parameters are varied while interaction functionality is maintained (Fig. 19.1B). Additionally, once interaction tags have been successfully added to pathway enzymes, the pathway can be used with a variety of scaffold architectures.

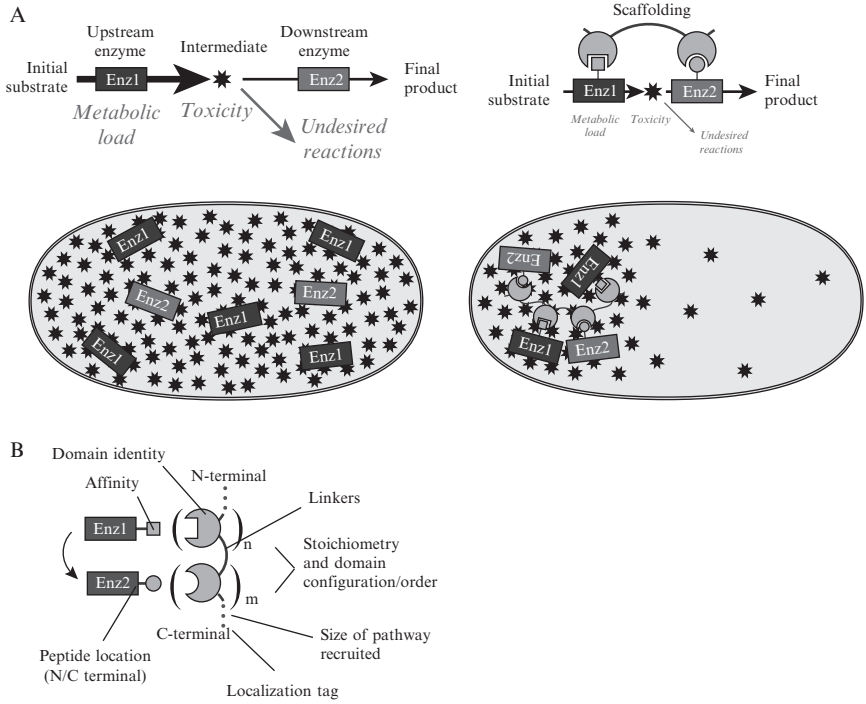

Figure 19.1 (A) High local substrate concentrations may be achieved with low enzyme expression, thereby reducing the cellular burden by using scaffolding to colocalize enzymes. (B) Scaffolding based on modular protein–protein interaction domains provides a highly designable control point with multiple parameters available for optimization.

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

In this chapter, we describe methodologies by which synthetic complexes can be engineered from metabolic pathways using modular protein scaffolds. Although the synthetic scaffolds built to date, and certainly for the foreseeable future, do not rival the elegance of natural systems, they offer designable control over several assembly parameters, most notably enzyme

stoichiometry, incorporation of heterologous enzymes, and potential scalability for increased numbers of enzymes. We hope to provide a practical guide for how we would approach scaffolding a pathway *de novo* together with a discussion of our considerations and experiences thus far. We conclude this chapter with a discussion of several mechanisms that may play a role in observed titer improvements for pathways tested thus far and of systems that may benefit from these possible effects.

2. Method—How to Build Modular Protein Scaffolded Systems for Metabolic Engineering Applications

2.1. Selecting protein–protein interaction domains and ligands for scaffold construction

The first decision to be made for scaffolding a metabolic pathway is the choice of colocation components. Each enzyme is translationally fused to a ligand specific for a protein–protein interaction domain. A translational fusion of these domains will compose a scaffold capable of colocalizing the ligand-fused enzymes. The structural modularity of the protein–protein interaction domains is of primary importance, as they will need to retain binding activity in the nonnative context of the translational fusions. A number of modular protein–protein interaction domains have been characterized and employed in various applications, a partial list of which is compiled in Table 19.1. In our experience, the members of SH3, PDZ, GBD, and leucine zipper families tend to retain binding activity as N-, C-terminal, or internal fusions and, given sufficient linker lengths, often do not require linker optimization to achieve binding activity. However, as discussed later, despite robustness of binding activity, overall flux improvement is likely also influenced by scaffold architecture including parameters such as linker length/composition and number/arrangement of protein–protein interaction domains.

Another parameter of importance is association affinity, particularly under low expression regimes. We generally design our scaffolds to target enzymes in the low micromolar range or tighter. To date, all targetings have been executed with the tightest affinity ligands available; however, many of these domains, as listed in Table 19.1, include lower affinity ligands that could be employed if transient interactions are desired.

Protein–protein interaction domains belonging to families with many members are particularly attractive choices for use in scaffolding, as they may offer a set of domains that potentially recognize specific ligands orthogonally (i.e., minimal cross talk with other ligands used as enzyme tags), yet have conserved folds and can more likely be used interchangeably.

For example, individual SH3 domains appear to have undergone negative selection such that they do not measurably interact with other SH3 domain family ligands within that organism (Zarrinpar *et al.*, 2003). Zarrinpar *et al.* showed that a peptide ligand was highly specific for a single SH3 domain within its native host, *Saccharomyces cerevisiae*, whereas this same peptide ligand interacted with a high percentage of non-*S. cerevisiae* SH3 domains. This selection for reduced cross talk should considerably increase the number of orthogonal domain/ligand pairs available for simultaneous use, particularly of domains recognizing small ligands such as the SH3 and PDZ domain family. Additionally, these interaction domain families often appear to have evolved physical and functional modularity, including characteristics such as robust-independent domain folding and surface-exposed N- and C-termini that are located close together to permit domain functioning as either terminal or internal fusions.

Leucine zipper and synthetic coiled-coil domain folds share many of these characteristics and are attractive targets for expanding the available number of orthogonal interaction partners. Works such as those by Havranek and Harbury (2003), where eight residues between leucine zipper pairs were altered based on computational prediction to create new pairs of either homodimers or heterodimers, show promise for rationally engineering new domains. More recently, Reinke *et al.* (2010) investigated the interaction specificities of a large set of synthetic coiled-coils that do not exhibit measurable self-association providing up to three orthogonal pairs that do not cross talk. These large libraries of structurally similar but orthogonally binding pairs provide excellent candidates for scaffold parts, as presumably they may be interchanged to switch specificity with minimal perturbation. However, as generating very large libraries of orthogonal parts has proven challenging, limited to sets with only several experimentally verified orthogonal pairs, taking parts from different families to minimize likelihood of cross talk is still likely to be fruitful for producing larger numbers of orthogonal protein–protein interaction pairs.

Cohesin–dockerin interaction modules have been successfully used to scaffold multienzyme complexes to function as synthetic cellulosomes *in vitro*. Up to three cellulose degrading enzymes were translationally fused to dockerins that localize to specific cohesins on a synthetic scaffold, which itself localizes to cellulose substrate via a carbohydrate-binding module. The resultant complex enhanced cellulose degradation in the complex substrate of straw sixfold over free enzyme (Fierobe *et al.*, 2005). An in-depth review of a number of applications that have taken advantage of cohesin–dockerin domains to provide controlled extracellular binding has recently been published (Nordon *et al.*, 2009). A study of cohesin–dockerin specificities has demonstrated up to five cohesin–dockerin pairs exhibiting orthogonal-binding specificity, providing a set of modules for further application (Haimovitz *et al.*, 2008). A unique feature of cohesin–dockerin interactions

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

2.2. Assembling scaffolds from domains and tagging enzymes for corecruitment

Domain/ligand choice is particularly important for proteins whose activity, stability, and/or solubility are sensitive to translational fusion. Particular peptide sequences and fusion locations may decrease the flux through these sensitive enzymes beyond the capability of the scaffolding effect to surpass. We attempt to minimize the perturbation to the enzyme of interest by selecting the smaller member of the binding pair to use as a tag. For proteins known to be problematic, we often use either an 11-amino acid peptide with a $K_d = 0.1\ \mu M$ for the Crk SH3 domain on the N- or C-terminus or a six-amino acid peptide with a $K_d = 8\ \mu M$ for the syntrophin PDZ domain as a C-terminal fusion. PDZ peptides must be used as C-terminal fusions since the carboxyl group is critical for binding. In the case of enzymes that are already experiencing solubility problems, it is possible that adding a larger, well-folded, binding domain may increase solubility similar to the oft-used strategy of tethering folding problematic proteins to maltose-binding protein or other highly soluble motifs (Di Guan *et al.*, 1988). The coiled-coil motifs, due to their high solubility, may be good candidates to try for this purpose, although all of these efforts are protein-dependent and currently unpredictable.

Linkers connecting domains of the scaffold are likely to provide another parameter that could be explored for flux optimization. To date, we have limited experience in the effect of different linker types. To connect domains, we have been using linkers expected to be of sufficient length (nine or more amino acids) to avoid sterically obstructing neighboring domains from binding and with compositions predicted to be unstructured and flexible (glycine–serine repeats). We have observed a small improvement in performance when linkers separating blocks of domains (i.e., the linkers in the scaffold GBD_1_linker_$(SH3)_2$_linker_$(PDZ)_2$) were further increased from 9 to 25 total amino acids of Gly-Ser repeats (Dueber, unpublished observation). However, we have not investigated varying linker lengths or the composition of these linkers beyond this initial characterization. Robinson and Sauer (1998) investigated the effect of linker length and composition on the stability of single-chain Arc repressor. For this protein, both linker length and the amino acid composition had a large impact on stability. Initial work on a simple tethering of two enzymes in the mevalonate biosynthetic pathway (Fig. 19.2) also suggested that a linker of adequate length must be used to

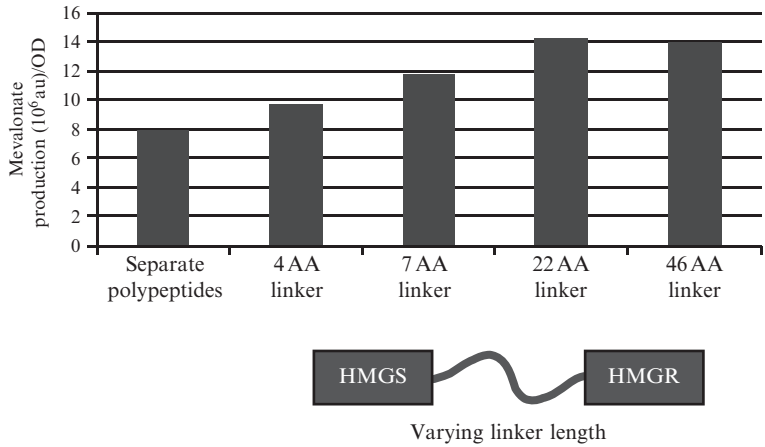

Figure 19.2 Two consecutive enzymes in a synthetic mevalonate pathway were tethered, C- to N-termini, and the length of a simple synthetic linker of alternating glycine and serine residues was varied. Tethering with a short linker length showed small improvements, which were increased to a higher titer with increased linker lengths. Mevalonate production was measured as GC/MS peak area as described in Dueber *et al.* (2009).

achieve highest flux improvement and, in this case, that the improvement is maintained through linkers of increasing lengths.

When making translational fusions of binding domains/ligands and enzymes, it is important to assay enzyme activity and interaction domain functionality. A GST pull-down assay can be employed to ensure binding activity remains functional. When domains from the same family are being employed with the intention of orthogonal function, GST pull-down assays often are very useful for avoiding unexpected intermolecular interactions, though low affinity and high effective concentration intramolecular interactions may be missed. For *in vivo* assays of enzyme activity, untagged enzyme function can be tested against tagged enzymes in absence of scaffolding. However, care must be taken to ensure expression rates are not being altered by domain addition, particularly for N-terminal tags or polycistronic systems which, in prokaryotes, should be expected to alter expression rates through RNA secondary structure interactions with the ribosome-binding site (Mathews *et al.*, 1999). For example, the addition of a C-terminal peptide targeting an enzyme might affect the expression level of an immediately downstream enzyme in a polycistronic message. This effect can be estimated or corrected for with the RBS calculator (Salis *et al.*, 2009). To confirm that expression levels are not changing, enzyme concentration should be carefully measured or expressed on independent transcripts.

2.3. Balancing the scaffold and enzyme concentrations

Balancing concentrations of enzymes has been shown to be an important consideration for optimizing metabolic pathways. Balancing scaffold concentrations is also important, as there is a theoretical optimal concentration for maximizing fully occupied scaffold molecules. This effect was modeled by Levchenko *et al.* (2000) for scaffolding in the MAPK-signaling pathway, where low concentrations of scaffold result in insufficient scaffold to colocalize the targets, while concentrations of scaffold considerably higher than enzymes result in segregation of components and a high percentage of scaffold molecules with low occupancy. We believe this biphasic trend also exists for synthetic scaffolding of multienzyme pathways. We simulated a simple mathematical model of equilibrium-binding reactions for a varying number of different enzymes that bind to different single sites on a scaffolding protein (Fig. 19.3). Differential equations for simple binding kinetics were generated with code written and simulated in MATLAB (The MathWorks, Natick, MA), code available upon request. Enzyme concentrations were held constant at 10 μM with each binding to a single site on the scaffold with a 100 nM K_d, while scaffold concentrations were varied. As expected, the optimal scaffold performance occurs when scaffold concentrations are approximately equal to enzyme concentrations, and concentration optimization becomes increasingly important as pathway size increases. In agreement with these modeled predictions, we observed a strong dependence of production titers on the relative expression levels of both scaffold and metabolic enzymes (Dueber *et al.*, 2009; Moon *et al.*, 2010). Thus, it may be helpful to drive expression of pathway enzymes and scaffold with independent promoters to independently tune expression to the optimal levels.

In addition to optimizing scaffolding levels, it may be beneficial to simultaneously adjust the pathway enzyme concentrations. High induction was found to be the optimal expression level for mevalonate biosynthetic pathway enzymes in the absence of scaffolding. However, when the most effective scaffold was present, low enzyme induction produced optimal production titers, giving higher titers than the maximum achievable in the absence of scaffold, even at the uninduced background expression level of the promoter (Fig. 19.4). Although it would be interesting to independently optimize the expression of the three enzymes scaffolded in this pathway, optimization beyond polycistronic expression level has not yet been carried out. Scaffold architecture, expression level, and enzyme expression levels are all interconnected variables that must be optimized.

2.4. Varying scaffold stoichiometry

Often when metabolic pathways are engineered, one pathway enzyme exhibiting relatively low activity creates a bottleneck in the pathway. In many cases, this limitation can be alleviated by increasing expression of

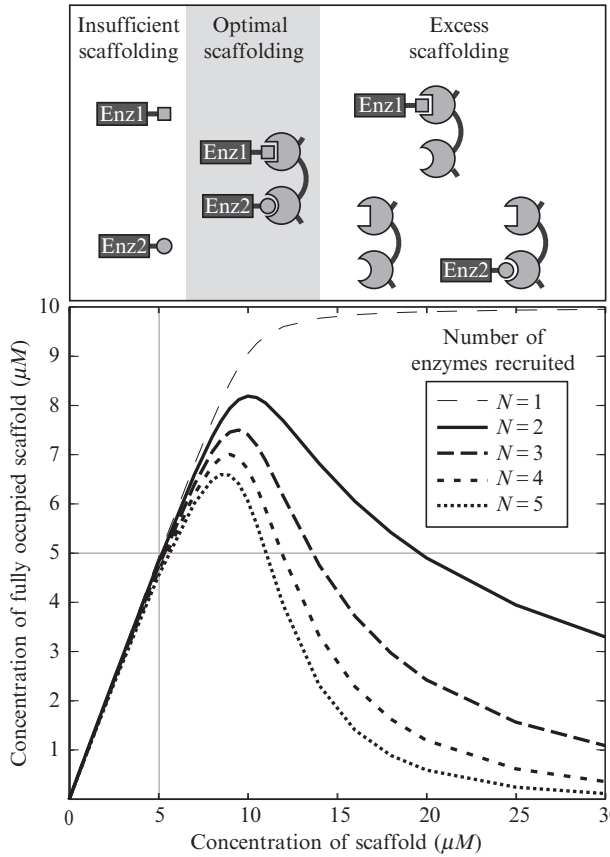

Figure 19.3 A mathematical model of equilibrium binding of scaffolding recruiting enzymes shows an optimal scaffold concentration for maximizing full occupancy. Five scaffolds consisting of a varying number of enzyme recruitment domains are independently simulated. Enzyme levels are held constant at 10 μ*M* each, and each scaffold molecule recruits a specified number of different enzymes each with a dissociation constant of 100 n*M*. Initially, as scaffold concentration is increased, excess enzymes are recruited to fully occupy the scaffold. As enzymes become limiting, scaffold competition for enzyme recruitment leads to low occupancy of scaffold.

that enzyme (Pitera *et al.*, 2007). Pathway scaffolding presents another strategy for addressing this problem at lower enzyme expression regimes. Ability to achieve high product titers with low concentration of enzymes should prove particularly advantageous for systems with enzymes prone to aggregate. Additionally, improved pathway efficiency is likely to become increasingly important as the number of enzymes in the pathway is increased, although this remains to be empirically tested.

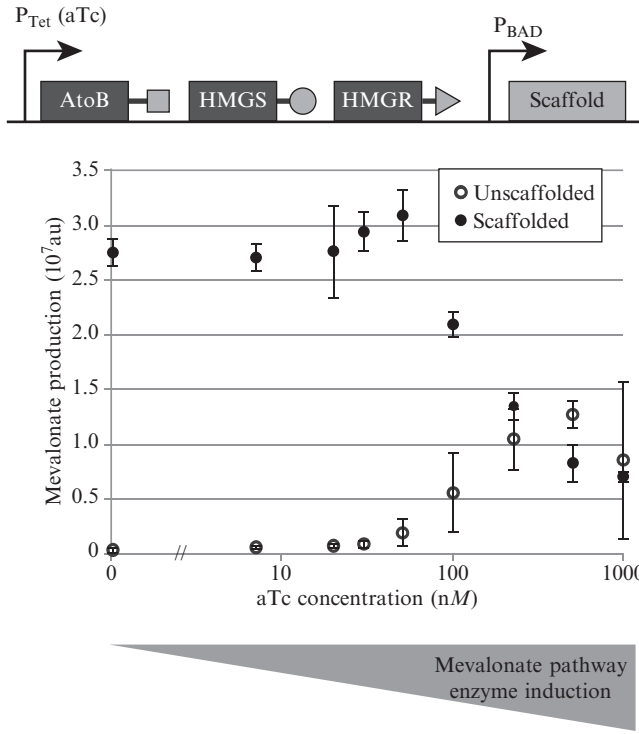

Figure 19.4 Mevalonate titers are measured at varying aTc concentrations, corresponding to induction of P_{Tet} driving peptide-tagged AtoB, HMGS, and HMGR polycistronically. The P_{BAD} promoter either drives the optimal scaffold, GBD_1–$SH3_2$–PDZ_2 (filled circles), or GFP, representing unscaffolded pathway (unfilled circled). Mevalonate production was measured as GC/MS peak area. Error bars represent one standard deviation from three separate experiments (figure adapted from Dueber *et al.*, 2009).

Due to the modular nature of the described scaffold strategy, the relative stoichiometry of enzymes cocomplexed can be controlled by varying the number of repeats of each protein–protein interaction domain. For the mevalonate biosynthetic pathway, a matrix of nine scaffold architectures was assembled with one, two, or four protein–protein interaction domains recruiting the enzymes for the bottleneck intermediate transfer, HMGS and HMGR (Fig. 19.5). Within this matrix, the optimal architecture produced a 77-fold improvement in product titers relative to the unscaffolded pathways. Importantly, although all scaffolds improved titers relative to the unscaffolded pathway, the results were difficult to explain based on stoichiometry alone.

We reapplied the same matrix of scaffolds used to improve flux of the mevalonate biosynthetic pathway to a second pathway, the glucaric acid

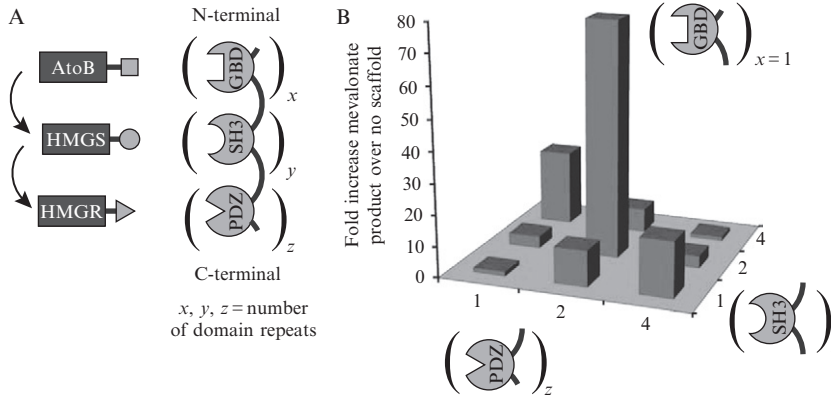

Figure 19.5 Three enzymes in the synthetic mevalonate pathway were tagged with binding peptides and recruited to a synthetic scaffolding protein with varying stoichiometry of binding domains. (A) Schematic of scaffolding with varying stoichiometry, where the number of GBD domains is held constant at one, while the number of SH3 and PDZ domains is varied to be one, two, or four. (B) The different scaffold stoichiometries gave very different results, with the best performing scaffold resulting in a 77-fold increase in mevalonate production compared to when no scaffold was expressed (figure adapted from Dueber *et al.*, 2009).

biosynthetic pathway. This pathway was previously engineered by Moon and Prather from three heterologous enzymes, Ino1, MIOX, and Udh (Moon *et al.*, 2009) and expressed at high levels under the T7 promoter to maximize glucaric acid titers. Interestingly, the Prather group measured higher activities of the limiting enzymatic activity, MIOX, in the presence of high concentrations of *myo*-inositol substrate (Moon *et al.*, 2009, 2010). The three heterologous enzymes were tagged with recruitment peptides and expressed from the P_{Lac} promoter. Similar to our findings with the mevalonate biosynthetic pathway, the scaffolds showed various degrees of titer improvements dependent on architecture. Consistent with a hypothesis of increasing the local concentration of *myo*-inositol at the resultant synthetic complex, titer improvements were dependent on the number of Ino1-recruiting domains producing *myo*-inositol, whereas there was no strong dependence on the number of domains recruiting MIOX enzyme. Titers of 2.3 g/L glucaric acid were produced with the optimal scaffold, giving an almost fivefold increase over the control lacking scaffold expression, a 50% improvement over highest titers previously reached (Fig. 19.6).

There are a few considerations to be made when constructing sequence for scaffolds with domains repeated multiple times. Cloning strategies relying on PCR or homologous overhangs such as recombination-based methods (Shao *et al.*, 2009), sequence and ligation-independent cloning (SLIC; Li and Elledge, 2007), or isothermal enzymatic assembly

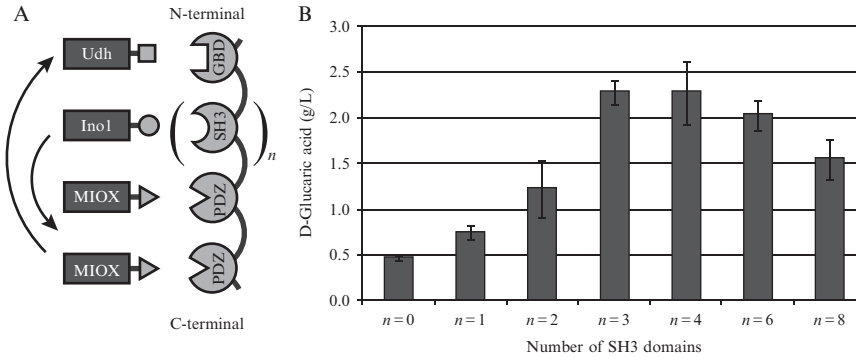

Figure 19.6 A pathway engineered to produce glucaric acid was scaffolded with optimized stoichiometry to improve production. (A) Schematic of scaffolded pathway where the number of SH3 domains corresponding to Ino1 recruitment is varied. Ino1 produces the substrate for MIOX, which MIOX converts to the substrate for Udh. (B) Glucaric acid concentrations were measured under conditions where only the number of SH3 domains on the scaffolding is varied, showing around three or four SH3 domains provided optimal production (figure adapted from Moon *et al.*, 2010).

(Gibson *et al.*, 2009) may result in misannealing for products containing repeated sequences. One construction strategy that deals particularly well with repeated domains is the BioBrick-based cloning strategies (Anderson *et al.*, 2010; Shetty *et al.*, 2008), particularly the *Bgl*II/*Bam*HI-based strategy (Anderson *et al.*, 2010) that leaves generally innocuous and often useful glycine–serine scars that can be used as part of the linker sequence. Interestingly, we have observed problems coming from recombination arising from greater than four identical repeats for both domains (~200 bases) and peptide ligands (~60 bases). A solution to this problem is to design multiple domain "parts" with degenerate codon usage such that repeated domains, linkers, and ligands are sufficiently different to prevent recombination. For making multiple SH3 domain repeats, using six degenerate SH3 parts, we were able to make constructs with 10 repeats without observing a significant number of incorrect products due to recombination (Dueber, unpublished observation; Moon *et al.*, 2010).

2.5. Scaffold composition effects

The three-dimensional structure of the scaffolded complex will determine the efficiency of improving flux. Although these structures are determined by the domain architecture, these architecture/structure relationships are not currently predictable. This is highlighted by the importance of not only the total number of each protein–protein interaction domain but also by the arrangement of these domains. The number and identity of SH3 and PDZ

interaction domains in a scaffold was held constant but the order of these domains was rearranged (Dueber *et al.*, 2009). GBD–$(SH3)_2$–$(PDZ)_2$, GBD–$(SH3)_1$–$(PDZ)_2$–$(SH3)_1$, and GBD–$(SH3)_1$–$(PDZ)_1$–$(SH3)_1$–$(PDZ)_1$ scaffolds showed dramatically varied abilities to improve mevalonate titers despite each having the same number of recruitment domains (Fig. 19.7).

Work on synthetic cellulosomes has also supported the importance of scaffold composition (Mingardon *et al.*, 2007). In this study, scaffolds were designed to recruit other scaffolds, creating complexes containing up to four different scaffolds, each in turn recruiting, or directly fused to, a cellulose degrading enzyme. One issue that arose was the importance of enzyme mobility, as redundant binding that was likely to limit flexibility decreased degradation efficiency (Mingardon *et al.*, 2007). Natural cellulosome protein sequences suggest that they are physically flexible complexes, since they generally have long linkers (tens up to 550 residues) predicted to result in highly mobile enzymes (Xu *et al.*, 2003). Another issue was the importance

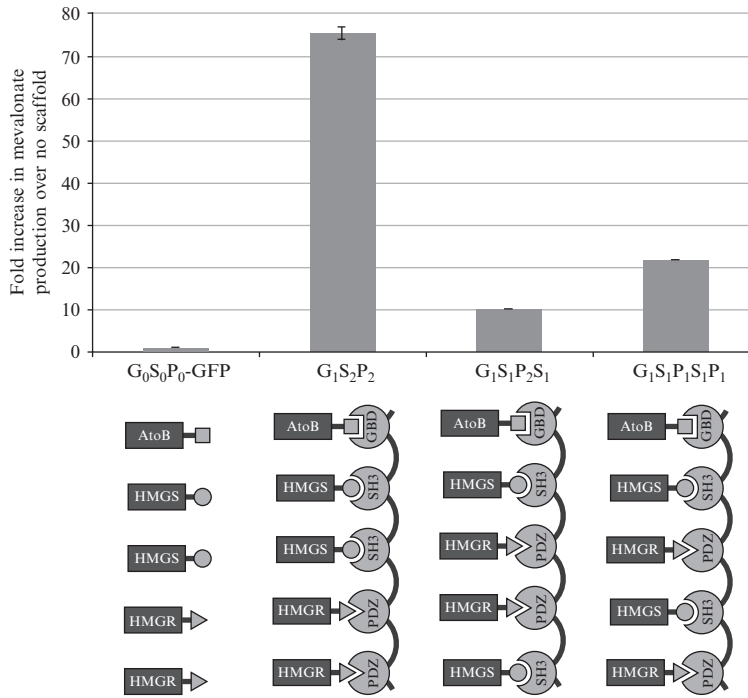

Figure 19.7 Mevalonate titers are measured for ligand tagged enzymes with expression of scaffolds of differing architecture and compared to the titers in the absence of scaffold expression. Rearranging domain order, while retaining the one GBD, two SH3, and two PDZ stoichiometry gives significant changes in titer improvement (figure adapted from Dueber *et al.*, 2009).

of scaffolding complex mobility, suggested from lowered activity with the addition of more than one carbohydrate-binding module. These issues of mobility and flexibility are likely to be particularly relevant to the catalysis of cellulose, an immobilized substrate, and may have limited pertinence to readily diffusing metabolic intermediates.

Interestingly, it has been suggested that when individual enzymes are tethered together, if these enzymes exist in oligomeric form, they may multimerize, forming even larger complexes (Bülow and Mosbach, 1991; Conrado *et al.*, 2008). A scaffold with repeated domains that recruits oligomeric enzymes, as is the case in several applications thus far (Dueber *et al.*, 2009; Moon *et al.*, 2010), may also form large multimeric complexes (Fig. 19.8). These multimeric complexes may improve titers by further increasing local concentrations beyond those achievable with individual scaffolds. This potential phenomenon is another reason we recommend taking a library approach to optimize scaffolded pathway flux, varying as many parameters as practical to empirically determine the optimal combination of architectural parameters.

Scaffold configuration may become an increasingly important variable as pathways are scaled to consist of larger numbers of enzymes. One practical concern is the increasing scaffold protein size to target increasing numbers of enzymes while also achieving stoichiometry control. As explored in the development of synthetic cellulosomes (Mingardon *et al.*, 2007), a potential solution to this problem is building multiple scaffolds that can coassemble either directly or through a separate adaptor molecule. This approach might also prove to be a convenient method for modularizing various sections of a pathway as well as increasing the combinatorial architecture possibilities.

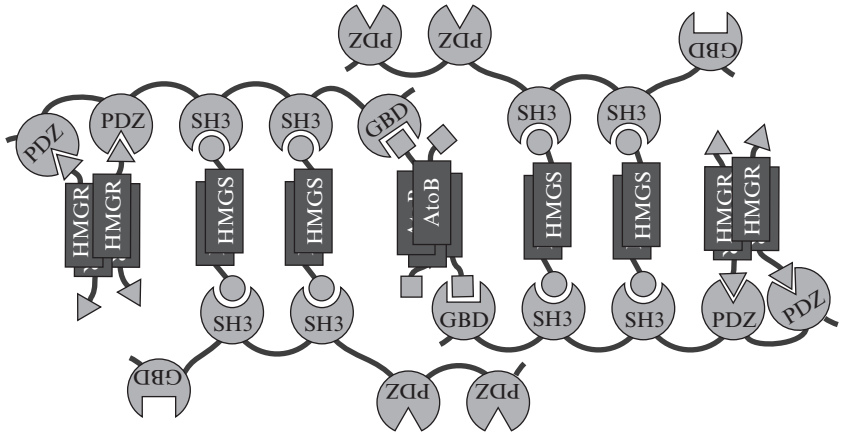

Figure 19.8 Schematic of a multimeric enzyme complex. Enzymes with oligomeric structures could potentially bind multiple scaffolds, resulting in large complexes and difficult to predict positioning in three-dimensional space.

3. Systems that May Benefit from Scaffolding

There are a number of potential mechanisms that may contribute to the increased titers shown in the scaffolded systems discussed. We have thus far focused on the increased local concentration effect by which scaffolding may reduce toxic intermediates, reduce the load on the cell by reducing the necessary enzyme expression levels, and increase pathway efficiency. Another potential benefit of scaffolding is the prevention of enzyme aggregation by lowering the necessary protein expression levels in addition to sequestering individual enzyme molecules. It is also possible that scaffold could be used to increase enzyme stability or activity, though this was shown to not be the case for the glucaric acid pathway (Moon *et al.*, 2010). Substrate-activated enzymes, such as MIOX in the glucaric acid pathway discussed earlier, are also particularly good candidates for scaffolding. Scaffold complexes with higher numbers of protein–protein interaction domains can be used to recruit substrate-producing upstream enzyme (Moon *et al.*, 2010).

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

4. Concluding Remarks

In conclusion, scaffolding provides a posttranslational tool that may help increase production yields and deal with problematic enzymes as well as reduce the cellular burden and unintended interactions that may become an issue when scaling to engineer the next generation of biosynthetic

metabolic pathways. The methods we describe in this chapter have the advantage of being highly designable and easily adaptable to library approaches. Future research must be done to determine the mechanisms by which these scaffolded systems function such that they can be more rationally applied to other pathways in a predictable manner.

ACKNOWLEDGMENTS

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

REFERENCES

Acharya, A., Ruvinov, S. B., Gal, J., Moll, J. R., and Vinson, C. (2002). A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: Amino acids I, V, L, N, A, and K. *Biochemistry* **41,** 14122–14131.

An, S., Kumar, R., Sheets, E., and Benkovic, S. (2008). Reversible compartmentalization of *de novo* purine biosynthetic complexes in living cells. *Science* **320,** 103–106.

Anderson, J. C., Dueber, J. E., Leguia, M., Wu, G. C., Goler, J. A., Arkin, A. P., and Keasling, J. D. (2010). BglBricks: A flexible standard for biological part assembly. *J. Biol. Eng.* **4,** 1.

Banaszynski, L. A., Liu, C. W., and Wandless, T. J. (2005). Characterization of the FKBP–Rapamycin–FRB ternary complex. *J. Am. Chem. Soc.* **127,** 4715–4721.

Bierer, B. E., Mattila, P. S., Standaert, R. F., Herzenberg, L. A., Burakoff, S. J., Crabtree, G., and Schreiber, S. L. (1990). Two distinct signal transmission pathways in T lymphocytes are inhibited by complexes formed between an immunophilin and either FK506 or rapamycin. *Proc. Natl. Acad. Sci. USA* **87,** 9231–9235.

Bülow, L., and Mosbach, K. (1991). Multienzyme systems obtained by gene fusion. *Trends Biotechnol.* **9,** 226–231.

Conrado, R. J., Mansell, T. J., Varner, J. D., and DeLisa, M. P. (2007). Stochastic reaction–diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway. *Metab. Eng.* **9,** 355–363.

Conrado, R. J., Varner, J. D., and DeLisa, M. P. (2008). Engineering the spatial organization of metabolic enzymes: Mimicking nature's synergy. *Curr. Opin. Biotechnol.* **19,** 492–499.

Di Guan, C., Li, P., Riggs, P. D., and Inouye, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* **67,** 21–30.

Dougherty, M. J., and Arnold, F. H. (2009). Directed evolution: New parts and optimized function. *Curr. Opin. Biotechnol.* **20,** 486–491.

Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., Prather, K. L. J., and Keasling, J. D. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27,** 753–761.

Fierobe, H., Mechaly, A., Tardif, C., Bélaïch, A., Lamedi, R., Shoham, Y., Bélaïch, J., and Bayer, E. A. (2001). Design and production of active cellulosome chimeras: Selective incorporation of dockerin-containing enzymes into defined functional complexes. *J. Biol. Chem.* **276,** 21257–21261.

Fierobe, H., Mingardon, F., Mechaly, A., Bélaïch, A., Rincon, M. T., Pagés, S., Lamed, R., Tardif, C., Bélaïch, J., and Bayer, E. A. (2005). Action of designer cellulosomes on homogeneous versus complex substrates: Controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. *J. Biol. Chem.* **280,** 16325–16334.

Fuh, G., Pisabarro, M. T., Li, Y., Quan, C., Lasky, L. A., and Sidhu, S. S. (2000). Analysis of PDZ domain–ligand interactions using carboxyl-terminal phage display. *J. Biol. Chem.* **275,** 21486–21491.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345.

Grünberg, R., Ferrar, T. S., Van der Sloot, A. M., Constante, M., and Serrano, L. (2010). Building blocks for protein interaction devices. *Nucleic Acids Res.* **38,** 2645–2662.

Haimovitz, R., Barak, Y., Morag, E., Voronov-Goldman, M., Shoham, Y., Lamed, R., and Bayer, E. A. (2008). Cohesin–dockerin microarray: Diverse specificities between two complementary families of interacting protein modules. *Proteomics* **8,** 968–979.

Harris, B. Z., Hillier, B. J., and Lim, W. A. (2001). Energetic determinants of internal motif recognition by PDZ domains. *Biochemistry* **40,** 5921–5930.

Havranek, J. J., and Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nat. Struct. Biol.* **10,** 45–52.

Kim, A. S., Kakalis, L. T., Abdul-Manan, N., Liu, G. A., and Rosen, M. K. (2000). Autoinhibition and activation mechanisms of the Wiskott–Aldrich syndrome protein. *Nature* **404,** 151–158.

Leibovitz, E., Ohayon, H., Gounon, P., and Béguin, P. (1997). Characterization and subcellular localization of the *Clostridium* thermocellum scaffoldin dockerin binding protein SdbA. *J. Bacteriol.* **179,** 2519–2523.

Levchenko, A., Bruck, J., and Sternberg, P. W. (2000). Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc. Natl. Acad. Sci. USA* **97,** 5818–5823.

Levskaya, A., Weiner, O. D., Lim, W. A., and Voigt, C. A. (2009). Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **461,** 997–1001.

Li, M. Z., and Elledge, S. J. (2007). Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nat. Methods* **4,** 251–256.

Ljungcrantz, P., Carlsson, H., Minsson, M., Buckel, P., Mosbach, K., and Biilow, L. (1989). Construction of an artificial bifunctional enzyme, β-galactosidasel/galactose dehydrogenase, exhibiting efficient galactose channeling. *Biochemistry* **28,** 8786–8792.

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

Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288,** 911–940.

Miles, E. W., Rhee, S., and Davies, D. R. (1999). The molecular basis of substrate channeling. *J. Biol. Chem.* **274,** 12193–12196.

Mingardon, F., Chanal, A., Tardif, C., Bayer, E. A., and Fierobe, H. (2007). Exploration of new geometries in cellulosome-like chimeras. *Appl. Environ. Microbiol.* **73,** 7138–7149.

Moon, T. S., Yoon, S., Lanza, A. M., Roy-Mayhew, J. D., and Prather, K. L. J. (2009). Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli*. *Appl. Environ. Microbiol.* **75,** 589–595.

Moon, T. S., Dueber, J. E., Shiuea, E., and Prather, K. L. J. (2010). Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab. Eng.* **12,** 298–305.

Narayanaswamy, R., Levy, M., Tsechansky, M., Stovall, G. M., O'connell, J. D., Mirrielees, J., Ellington, A. D., and Marcotte, E. M. (2009). Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proc. Natl. Acad. Sci. USA* **106,** 10147–10152.

Nordon, R. E., Craig, S. J., and Foong, F. C. (2009). Molecular engineering of the cellulosome complex for affinity and bioenergy applications. *Biotechnol. Lett.* **31,** 465–476.

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

Pfleger, B. F., Pitera, D. J., Smolke, C. D., and Keasling, J. D. (2006). Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* **24,** 1027–1032.

Pitera, D. J., Paddon, C. J., Newman, J. D., and Keasling, J. D. (2007). Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab. Eng.* **9,** 193–207.

Posern, G., Zheng, J., Knudsen, B. S., Kardinal, C., Müller, K. B., Voss, J., Shishido, T., Cowburn, D., Cheng, G., Wang, B., Kruh, G. D., Burrell, S. K., *et al.* (1998). Development of highly selective SH3 binding peptides for Crk and CRKL which disrupt Crk-complexes with DOCK180, SoS and C3G. *Oncogene* **16,** 1903–1912.

Price, N., Reed, J., and Palsson, B. (2004). Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nat. Rev. Microbiol.* **2,** 886–897.

Reinke, A. W., Grant, R. A., and Keating, A. E. (2010). A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J. Am. Chem. Soc.* **132,** 6025–6031.

Robinson, C. R., and Sauer, R. T. (1998). Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proc. Natl. Acad. Sci. USA* **95,** 5929–5934.

Salis, H. M., Mirsky, E. A., and Voigt, Christopher A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27,** 946–952.

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

Shetty, R. P., Endy, D., and Knight, T. F. (2008). Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2,** 5.

Stroud, R. (1994). An electrostatic highway. *Nat. Struct. Biol.* **1,** 131–134.

Tonikian, R., Zhang, Y., Sazinsky, S. L., Currell, B., Yeh, J., Reva, B., Held, H. A., Appleton, B. A., Evangelista, M., Wu, Y., Xin, X., Chan, A. C., *et al.* (2008). A specificity map for the PDZ domain family. *PLoS Biol.* **6,** 2043–2059.

Welch, G. R. (1977). On the role of organized multienzyme systems in cellular metabolism: A general synthesis. *Prog. Biophys. Mol. Biol.* **32,** 103–191.

Win, M., and Smolke, C. (2007). A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc. Natl. Acad. Sci. USA* **104,** 14283.

Xu, Q., Gao, W., Ding, S., Kenig, R., Shoham, Y., Bayer, E. A., and Lamed, R. (2003). The cellulosome system of *Acetivibrio cellulolyticus* includes a novel type of adaptor protein and a cell surface anchoring protein. *J. Bacteriol.* **185,** 4548–4557.

Zarrinpar, A., Park, S., and Lim, W. A. (2003). Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **426,** 676–680.

Zhang, K., Sawaya, M. R., Eisenberg, D. S., and Liao, J. C. (2008). Expanding metabolism for biosynthesis of nonnatural alcohols. *Proc. Natl. Acad. Sci. USA* **105,** 20653–20658.