PROPOSAL FOR DEVELOPMENT OF ENHANCED *IN-VIVO* AND EX-VIVO SYNTHETIC BIOLOGY METHODS THAT ADDRESS DEPARTMENT OF ENERGY PRIORITIES

Applicant Institution: Harvard Medical School
Street Address/City/State/Zip: 25 Shattuck St / Boston / MA / 02115
Investigator (PI): Professor George M. Church
PI Postal Address: 77 Avenue Louis Pasteur, NRB 238 / Boston / MA / 02115
PI Telephone Number: 617-432-7562
PI Email: church_office@genetics.med.harvard.edu
FOA Number: DE-FOA-0000412
DOE/OSP Office: BER Biological Systems Science
DOE/OSP Manager Contact Joseph Graber, Ph.D.
Office of Biology & Environmental Research SC-23.2
U.S. Department of Energy
1000 Independence Avenue, S.W. Washington, DC 20855-1290
DOE Grant Number: DE-FG02-02ER63445

SUBCONTRACTS

There is one subcontract:

Co-Investigator Institution: Yale University
Co-Investigator: Professor Farren Isaacs
Co-Investigator Postal Address: Dept. of MCDB / Systems Biology Institute / KBT 802 / 266 Whitney Avenue / P.O. Box 208103 / New Haven, CT 06520
Co-Investigator Telephone No. 203-737-3156/203-432-3783
Co-Investigator Email: farren.isaacs@yale.edu

Professor Isaacs will not be submitting a separate application.

Project Objectives

The goal of our proposed DOE Genomic Sciences Center renewal is to develop and demonstrate new Synthetic Biology (SB) techniques that will significantly advance the ability to engineer both living biocatalysts (*in vivo* biology) and non-living systems that functionally integrate biological with non-biological physico-chemical components (*ex vivo* biology), in support of DOE's priority areas: energy availability, environmental reclamation, and carbon cycle characterization and management. DOE has long emphasized applying systems biology, metagenomics, and omics-level metabolic analyses of microbes and plants for understanding biological contributions to these areas. However, because methods for manipulating biology are needed both to improve understanding and to engineer approaches for meeting human needs, DOE has also increasingly integrated SB into its research portfolio. This is most salient in DOE's Bioenergy Research Centers (BRCs), in which SB *in vivo* engineering methods are currently being very actively developed to modify microbes and plants for efficient biofuel production. However, SB has changed substantially over the past ~10 years as a suite of *in vitro* methods has emerged that has dramatically expanded *in vivo* engineering capabilities, and which is providing the basis for upcoming novel *ex vivo* biology methods (see Background and Significance). Several of these new *in vitro* methods have been developed by the Church Lab as part of its current DOE Genomic Sciences Center. For BRC and other DOE research to reap the benefits of these emerging SB methods, they must first mature and become available to the research community. Key steps along this path include the systematic evaluation of protocol options according to relevant performance metrics; convincing demonstration of the methods on important target problems that have resisted conventional solution; open source availability of complete protocols, reagents, and instrumentation; followed by commercialization. We propose to catalyze development, adoption, and application of new SB capabilities with focus on DOE priority areas with the following Specific Aims:

Aim 1. (*in vivo* biocatalyst engineering) We will generate new biocatalyst platforms, regulatory elements, and functions that expand options for carbon capture, environmental remediation, and product generation, as well as improve the productivity and safety of microbes used to address environmental concerns.

1.a. We will engineer an *E. coli* strain with remapped sense codons that will facilitate efficient generation of proteins with multiple non-standard amino acids, assess its ability to resist environmental pathogens, and develop and test measures that can ensure its safe usage.

1.b. To improve capabilities for engineering and directing microbial carbon capture, we will generate a MAGE-able cyanobacterium and associated tools for assembling MAGE-altered fragments into single genomes. With these tools we will then initiate creation of a codon-remapped strain for novel small molecule production and safe environmental usage.

1.c. We will improve the frequency of λ-Red recombination of long (≥ 1kb) DNA constructs into the *E. coli* chromosome by at least an order of magnitude (from ≤ 0.1% to ≥ 1%), and, with this, extend our Multiplex Automated Genome Engineering (MAGE) method to accommodate long constructs in addition to short oligos (MAGE-2).

1.d. We will develop suites of protein and RNA modules in *E. coli* that respond to small molecule ligands, and that can be used to sense and regulate metabolic pathways. By engineering ligand binding and allostery, we will expand the natural *E. coli* repertoire of ligands that can be sensed. We will demonstrate the use of these modules for optimizing production of useful products.

1.e. We will develop tools for identifying metabolic linkages between species in natural microbial communities, and methods for mathematically modeling and engineering microbial consortia to

efficiently produce useful molecules or remediate environmental toxins.

Aim 2. (enhanced *in vitro* DNA and protein synthesis) We will develop a coordinated set of *in vitro* methods for generating complex, high fidelity libraries of DNA constructs, and *in vitro* methods for generating proteins using modified ribosomes that will support enhanced translation control and integration of non-standard protein features. We will demonstrate the use of these capabilities for implementing pathways and processes important to biofuel production and carbon capture.

2.a. We will develop an automated DNA construct synthesis pipeline built from off-the-shelf components for generating high-fidelity libraries of complex constructs from oligonucleotide arrays, and develop algorithms for specifying construct libraries to be synthesized that enable optimization of biofuel-, carbon capture-, and remediation-relevant pathways and enzymes.

2.b. We will develop an efficient and completely defined *in vitro* transcription and translation system that possesses novel properties including: reduced dependency on rRNA modifications, support for orthogonal expression systems, and increased ability to incorporate D amino acids or other non-standard features.

Aim 3. (*ex vivo* biology) We will advance the ability to operate biological pathways *ex vivo* by enabling the *optimization by genetic selection* of biologically functionalized nanoparticles. Towards this end:

3.a. We will enhance *ex vivo* selection methods to accommodate new classes of enzymes, and improve compartmentalization, automation, and precision.

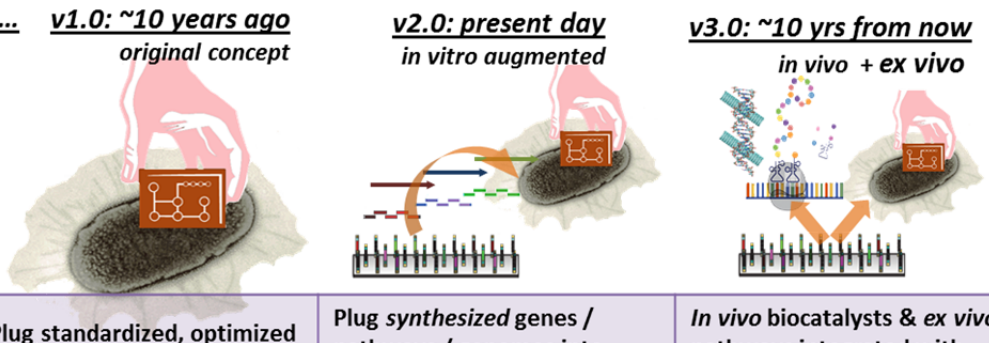
3.b. We will develop complementary 'high precision' *ex vivo* selection methods based on arraying and analysis of aqueous compartments that will enable isolation of rarer variants exhibiting smaller relative improvements in performance *vs.* emulsion methods.

3.c. We will develop the ability to optimize multi-enzyme, multi-step reactions by *ex vivo* screening and selection.

The Church Lab is ideally positioned to realize these Aims and to help DOE avail itself of powerful new SB capabilities generally. Professor Church has been associated with DOE since 1986 and has led a highly productive DOE Genomic Sciences Center since 2003. In addition to his pivotal role in development of next generation sequencing, Professor Church has been a foundational figure in the development of SB, where he has key roles in iGEM and SynBERC, and has developed many new SB techniques, including MAGE (213) and synthesis of DNA constructs from arrays (104; 199). He has extensive experience with development and commercialization of new technologies, and close connections with the biotechnology industry: Currently he is on Scientific Advisory Boards of 29 companies, and 9 companies have licensed Church Lab patents or software (33). His vast contributions to the advancement of genomics were recently recognized by his receipt of the 2011 Bower Award for Achievement in Science and election to the National Academies of Science. Lastly, the Church Lab is tightly interwoven with and has access to the resources of the rich array of leading scientific institutions in Boston, including the Wyss Institute, Broad Institute, MIT, and Boston University. Professor Isaacs has pioneered several biomolecular-genomic engineering technologies. In his thesis work, he developed novel systems of synthetic RNA components for probing and programming cellular function. As a post-doc in the Church Lab, he developed MAGE (213) and CAGE while leading the *E. coli* recoding project (87) (further developed in Aim 1 of this proposal). His lab focuses on further advancing these foundational technologies and applying them to development of new genetic codes and of engineered cell factories for chemical, biofuel, and drug production. In 2008, he was selected as a "Rising Young Star of Science" by Genome Technology Magazine (51; 167).

Background and Significance

(A)



Engineering	Plug standardized, optimized natural parts into chassis organisms	Plug *synthesized* genes / pathways / genomes into organisms; non-standard amino acids (NSAAs)	*In vivo* biocatalysts & *ex vivo* pathways integrated with nanotechnology, chemistry, physics, electronics
Optimization	Modeling, selections, screens	MAGE, transcript structure and optimization	Co-selection of bio- & inorganic components
Security	Control cell cycle, couple activity to stimulus, suicide capability	Genetic and nutritional containment	*Ex vivo* systems incapable of replication; genetic isolation *via* codon reassignment
Scaling	Separate growth and biofunction modes		Secure systems supporting containmentless operation

(B)

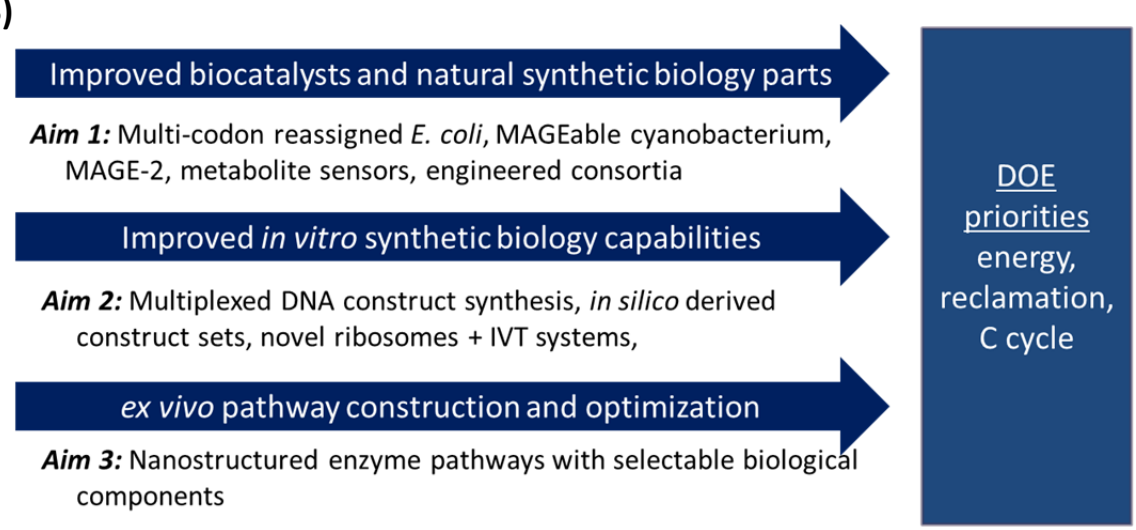


Figure 1: Changing capabilities of Synthetic Biology (SB) (A) and relation to DOE priorities (B). (A) Development of improved *in vitro* DNA synthesis capabilities coupled with genome engineering techniques has widened the capabilities of Synthetic Biology and may make possible the development of operational *ex vivo* biology systems that can complement use of *in vivo* biocatalysts (see text for details). (B) The Aims of our proposal will develop and apply new *in vivo*, *in vitro*, and *ex vivo* SB tools that support DOE priorities.

SB came of age ~10 years ago as a movement whose goal was to enable the engineering of biological organisms along lines similar to those routinely used in electronics and other areas by developing standards and repositories for "parts" with predictable behavior that could be readily combined in organisms (132). More recently, a suite of *in vitro* methods has emerged that have dramatically expanded the capabilities of such *in vivo* engineering, several of which were developed by the Church

Lab. Their further development is changing the emphasis of SB and promises to enable the building of operational *ex vivo* biology systems in the next ~10 years (see Figure 1A). For instance, improved DNA synthesis methods (104; 130; 199) have partially freed SB from its original focus on standardized *natural* sequences by enabling the engineering of expression-optimized pathways drawn across the entire tree of life directly from sequence databases, removing limits formerly imposed by genetic distance, the unavailability of physical sequences, and the unculturability of most organisms (14; 104). Our new MAGE method of genome engineering (213) not only leverages array-based oligo synthesis but, by its high multiplexity and ease of use, partially relieves SB of its original concern with prior precise characterization and modeling of parts. Meanwhile, ongoing work developing novel protein synthesis methods is synergizing with DNA synthesis and genome engineering advances, portending a new suite of SB *in vivo* and *in vitro* protein production methods. For instance, incorporation of non-standard amino acids (NSAAs) and peptide bonds into proteins (120) will be greatly advanced by completion of our *rEcoli* project (87) which, by making a free codon available for incorporation of NSAAs, will relieve the need to use nonsense suppressors to read stop codons. Advances in *in vitro transcription and translation* systems, and generation of artificial ribosomes (92) may enable efficient protein synthesis entirely *ex vivo*, and also allow development of orthogonal and mirror ribosome systems. Efficient and flexible *ex vivo* DNA and protein synthesis capabilities will enable construction of operational *ex vivo* pathways. Already, 12 enzyme *ex vivo* systems with redox cofactor recharge systems have been demonstrated and preliminary analyses suggest that such systems can be cost effective compared to *in vivo* biocatalysts (188; 227; 235). Finally, recent advances in integrating enzymes with nanomaterials promise not only to alleviate chemical engineering inefficiencies related to mass transfer, enzyme loss, and enzyme degradation, but to enable tight integration of enzymes with other physico-chemical processes (55; 93; 99; 144; 145; 216). This holds open the possibility of combining the unique advantages of both biological and non-biological processes without incurring the overhead of maintaining modified organisms and securing them from release into natural environments.

Our renewed Center will further these advances, bring them to bear on DOE priorities (see Figure 1B), and, by interacting with the DOE research community, promote their use and integration into DOE research at large. Our MAGE-able (Aim 1.c) cyanobacterium (Aim 1.b) will create a flexible and metabolically powerful carbon-fixing chassis that will not only simplify research aimed at improving carbon capture and conversion of solar energy into biofuels, but will also initiate the creation of a genetically recoded cyanobacterium whose enhanced genetic and nutritional containment features (Aim 1.a) will facilitate its eventual use in open environments. Our protein and riboregulator engineering (Aim 1.d) will produce new metabolic sensors that will broaden the ability to optimize pathways not only *in vivo* in organisms (Aim 1.d), but *ex vivo* (Aim 3.a). We will create novel tools not only for analyzing, but also for engineering, microbial consortia at the metabolic level (Aim 1.e). On the *in vitro* side, with greatly enhanced multiplexity of long DNA construct synthesis from arrays (Aim 2.a), it will be possible to design and generate precise high-complexity construct libraries that incorporate domains sampled widely across phylogeny (Aim 2.a), and evaluate and evolve them at vast throughput using emulsion and microdroplet methods (Aims 3.a, 3.b). Precise control over *in vitro* translation (Aim 2.b) will not only speed development of enhanced ribosomes (Aim 2.b), but will enable pathways to be synthesized and explored *ex vivo* that are incompatible with organism life (Aim 3). Finally, new *ex vivo* genetic selection methods will enable such pathways to be optimized by directed evolution (Aim 3.c).

Relationship to Prior Center and Center Organization

The central Aims of our 2007-2011 Center were (1) genome analysis and engineering and gene synthesis technology, (2) metabolic analysis and engineering, (3) proteomics and protein synthesis technology, and (4) DNA sequencing improvement. The relationship of our current and proposed Centers is shown

in Table 1. Broadly speaking, in this proposal we have sharpened our focus on Synthetic *vs.* Systems Biology and increased emphasis on development of new *in vitro* and *ex vivo* techniques.

Focus area	2007-2011 Center research areas	In this proposal
(1)	• MAGE (213) + MAGE improvements (28; 214) • Eliminate TAG codon in *E. coli* (87) • Initial work long-construct MAGE (137) • Improved gene construct synthesis (104; 130) • genome spatial conformation analysis (206)	• Cyanobacterial MAGE (Aim 1.b) • Non-standard amino acids + eliminate AGR codons (Aim 1.a) • MAGE-2 (Aim 1.c) • Highly multiplexed construct synthesis (Aim 2.a) • *Not pursued in this proposal*
(2)	• Engineering / systems biology of *C. phytofermentans* (200) • Metabolic modeling for pathway engineering (122); evolved syntrophs (180) • Protein fatty acid sensor + improved fatty acid synthesis (see Aim 1.d *Preliminary Results*) • Initial work riboregulators (86) • Functional metagenomics (189)	• *Not pursued in this proposal* • Metabolic modeling, engineering, and experimental analysis and evolution of consortia (Aim 1.e) • Protein sensors for several small molecules (Aim 1.d) • Riboregulator selections, ligand-sensing riboregulators (Aim 1.d), RNA aptamer molecular beacons (Aim 3.a) • Experimental analysis / evolution of consortia (Aim 1.e)
(3)	• Development of improved IVT systems ((92) + Aim 2.b *Preliminary Results*) • D-amino acid incorporation • Proteome analysis *C. phytofermentans* (200) • Proteomic motif analysis (179)	• Continued and expanded Aim 2.b • Continued and expanded Aim 2.b • *Not pursued in this proposal* • *Not pursued in this proposal*
(4)	• Polonator sequencer improvements (see *Preliminary Results* Aim 3)	• Polonator enhanced to support *in vitro* selections (Aim 3.b)
New		• Containment and genetic isolation of genome-recoded strains (Aim 1.a) • Improvement of *in vitro* selections and development of *ex vivo* synthetic biology methods (Aim 3) • Optimization of GH9 enzymes (Aim 2.a, 3.a)

Table 1: Relationship between our current and our proposed DOE Genomic Science Centers

All activities in our current Center have been performed in the laboratory of PI Professor Church. In our proposed Center, this will remain true except that Professor Farren Isaacs, who, while a post-doc in the Church Lab, led research into MAGE development, the recoding of the *E. coli* genome, and riboregulators, will continue to work on these in his new lab at Yale University with principal focus on riboregulator and RNA aptamer development.

The current Center will close with estimated unspent funds in the amount of $121,554.30 direct ($206,042.99 total). We request that these be carried forward to the new Center.

We now describe each of our proposed Aims in detail.

Aim 1: New *in vivo* biocatalyst engineering tools and methods

Aim 1: We will generate new biocatalyst platforms, regulatory elements, and functions that expand options for carbon capture, environmental remediation, and product generation, as well as improve the productivity and safety of microbes used to address environmental concerns.

Two key achievements of our Center have been the development of the MAGE technique for efficiently

making many changes across a genome (28; 85; 213), and the application of MAGE to the elimination of all 315 instances of the TAG codon in *E. coli* (85) and to the optimization of a biosynthetic pathway (213). Based on this latter study, in which an *E. coli* lycopene production was improved 3.9-fold by simultaneously mutagenizing 24 specific genomic loci (creating 4.3e9 variants in a 3 day period), MAGE was called a "major advance in synthetic biology" (155). Here will we press forward on several fronts.

Our TAG-free *E. coli* strain is now nearly complete, and we expect it to be finished by the end of 2011 when our renewed Center starts, putting us in position to test and document the many advantages of an *E. coli* strain with a free codon that have been anticipated. In Aim 1.a we will address this task by assessing the expected improved performance of this strain in the generation of proteins containing non-standard amino acids (NSAAs), and the potential for viral resistance and genetic isolation of recoded strains. These potentials will only increase with additional free codons, and in Aim 1.a we will begin the process of generating a new *E. coli* strain with two more free codons (AGR). From a biotechnology perspective, both viral resistance and genetic isolation are highly desirable characteristics in genetically modified organisms (GMOs), promising improved productivity and inability to express unwanted DNA picked up from other organisms. But our recoded strains also offer unique new opportunities to create safe strains that will only survive in carefully controlled environments by engineering them to be dependent on NSAAs for translation of essential genes. This has important implications for safe use of GMOs in industry, agriculture, and bioremediation.

Due to its fast growth, metabolic versatility, genetic tractability, and the wealth of available knowledge and technique, *E. coli* is an organism of choice for SB development and engineered *E. coli* strains have been used in industrial production of therapeutics (72; 212) and chemicals for industrial processes (57). However, cyanobacteria are much better starting points than *E. coli* for manipulating photosynthesis and improving carbon capture, play critical roles in carbon biomineralization (91), and naturally generate biofuel molecules such as alkanes (178; 197). Many experimental and informational resources have been developed for cyanobacteria, thereby enabling experiments in pathway engineering for such products as ethanol, isobutyraldehyde, hydrogen, fatty alcohols, and isoprene (6; 44; 47; 117; 131; 197). However, genome-scale engineering of cyanobacteria similar to what has been accomplished in *E. coli* has not yet been possible. In Aim 1.b we will address this deficit by creating a MAGE-able cyanobacterium that will make possible efficient and large-scale pathway engineering, and also enable the generation of codon remapped strains that can be safely used for biological production.

Although we have already enhanced our initial formulation of MAGE (28; 85), so far all MAGE methods have relied on using 90-mer oligos to transfer genetic variations to the *E. coli* genome. MAGE utilizes the λ-Red recombination system to incorporate these oligos as Okazaki fragments into the *E. coli* lagging strand during genome replication. This use of short oligos limits the changes that can be efficiently incorporated into a single locus to 1-30 bp, although longer deletions are possible. In Aim 1.c, we plan to improve the incorporation of longer (>1kb) DNA fragments. The reason MAGE is limited to short oligos is that the frequency of λ-Red-mediated incorporation of longer ~1kb fragments is currently $\leq$ 0.1% (230). We plan to overcome this limitation by a combination of methods, including modifications that protect long constructs from degradation, and by directed evolution and overexpression of λ-Red components. Initially reported rates for the λ-Red-mediated incorporation of oligos were ~0.2% (230), and our improvement of ssDNA oligo incorporation rates by ~150x to about ~30% enabled our development of MAGE. A ~20x improvement to $\geq$ 1% will similarly enable a long-construct version of MAGE (MAGE-2). This could be used to readily create populations of *E. coli* expressing combinations of endogenous and exogenous gene variants associated with entire pathways which could then be selected on the basis of pathway productivity.

Optimization of biological production pathways for useful moleculse depends on being able to efficiently

and specifically measure production of desired product molecules. When the target organism contains internal sensors for the molecules, optimization can be accomplished efficiently *via* screens and selections. Unfortunately, while microbes typically have many specific and sensitive sensors for needed substrates, they contain relatively few for useful product molecules and intermediates. In Aim 1.d, we will build a suite of *E. coli* allosteric transcriptional regulators for useful molecules both by adapting existing internal sensors, and by systematic engineering and directed evolution of the binding pockets of substrate sensors that alter their specificity. Another route to building internal sensors is through riboregulators, developed initially by proposed Center co-investigator Farren Isaacs (86).

The central role of microbial communities in nature is widely recognized and characterization of their species and gene composition is under intensive study in numerous microbiome and metagenomics projects. In Aim 1.e, we will go beyond species identification *via* 16S rRNA, gene identification by sequencing read assembly, and functional metagenomic screens for assayable enzyme classes (cf. (189; 190) from our Lab), by developing methods for dissecting species interactions within communities based on use of gel microdroplets (GMDs) and picotiter plates. We will also develop tools for engineering interactions between distinct strains and species. Synthetic communities of genetically tractable 'pathway specialists' offer several opportunities: Specialists can be flexibly combined to meet differing conditions (cf. (14)), while the components of complex pathways can be distributed to optimize productivity. This work will leverage methods we developed as part of this Center for generating obligate interactors from *E. coli* tryptophan and tyrosine auxotrophs (170; 180) (for other approaches, see (21)), as well as Flux Balance Analysis (FBA)-based mathematical methods we have developed for pathway optimization (122). Here GMDs will overcome earlier limitations imposed by the need to conduct directed evolutions by iterative plating *vs.* liquid culture.

Aim 1.a: We will engineer an *E. coli* strain with remapped sense codons that will facilitate efficient generation of proteins with multiple non-standard amino acids, assess its ability to resist environmental pathogens, and develop and test measures that can ensure its safe usage.

1.a.1 Replacement of AGR codon from our TAG-eliminated *E. coli*

As noted above, eliminating redundant codons in our *E. coli* strain is expected to yield advantages in the ability to produce proteins containing NSAAs, and to confer genomic isolation and viral resistance on the strain. While these characteristics will be assessed on our initial TAG-free strain as parts of Aim 1.a.2-3 (see below), replacing additional codons should enhance these characteristics. Thus, elimination of a sense codon has potential to raise the number of NSAA combinations incorporable into protein from the current ~70 (120) to ~4900 (=~70^2). Meanwhile, viral resistance and other genomic isolation should increase with increasing codon usage incompatibility (36; 106) with other organisms. We note that viral resistance translates into increased productivity of the strain as viral predation is a major cause of bioproduction failure and measures to limit it are major expenses for large-scale bioprocessing applications (193). Finally, replacement of an additional codon is a natural test of the scalability of our MAGE methods, and will prove a useful test bed for additional improvements such as MAGE-2 (below). We will change all 3861 AGR (Arg) codons to CGH (Arg; H=A|C|T) codons as this is the smallest number of changes that frees up a tRNA (counts from the *E. coli* K12 MG1655 genome; for CGH, see (60)).

Preliminary results: Our approach to codon reassignment and elimination of the TAG codon, currently in press (87), is summarized in Figure 2. In brief, we split the genome into 32 segments each containing ~10 TAG codons and used MAGE (213) to create a strain for each segment in which all 10 TAGs were replaced with TAAs. We also developed hierarchical Conjugative Assembly Genome Engineering (CAGE) to assemble the 32 recoded segments into a single fully recoded chromosome. At this time we have made both of the half genomes (~160 TAG $\rightarrow$ TAA).

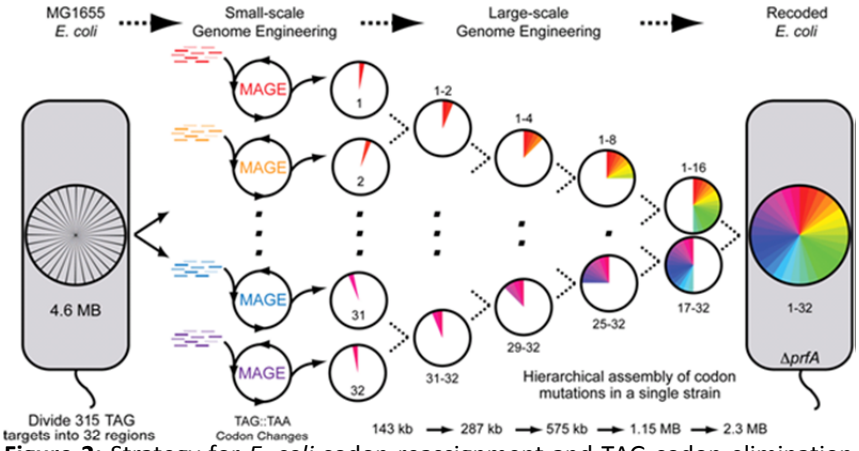


Figure 2: Strategy for *E. coli* codon reassignment and TAG codon elimination (see text).

Research Design: (i) We will complete the assembly of the TAG-replaced strain and eliminate Release Factor 1 (RF1) to completely abrogate TAG function, and then (ii) proceed to creation of a strain in which the AGR codon is eliminated. (iii) To explore potential for additional codon replacements, we will replace native copies of essential genes with versions lacking 10 unique codons (TAG, AGR, CTY, CCC, ACC, ATA, GTC, GCC, TCC, CGG) to check for viability. We have already generated constructs for these genes using our DNA construct synthesis capability (see Aim 2.a.1). In very early results, we have replaced one essential gene in each of five strains, suggesting that extreme codon flexibility is possible.

Timeline: (i) Year 1. (ii) Years 2-4. (iii) In order to use MAGE-2 (Aim 1.c) to expedite essential gene replacement, we will schedule the bulk of this work for Years 2-4.

Potential problems and alternatives: If (iii) meets with continued success, we may reprioritize (iii) in front of (ii) and replace all essential genes with highly reduced codon sequences first, and then move to eliminate AGR (or other) codons from remaining genes afterwards. Regarding (ii), it is possible that we will find that some AGR codons are refractory to reassignment; indeed, there are indications that AGR codons may play roles in ribosomal frameshifts and the regulation of expression (31; 110; 219). If we encounter such problems, we will use MAGE to generate diversity in the region of the gene containing the refractory codon and attempt to evolve a variant that does not require the AGR codon.

1.a.2 Demonstration of use of recoded *E. coli* strains for NSAA incorporation

Generation of proteins containing NSAAs is of interest because increasing the repertoire of residues incorporated into proteins beyond the natural 20 greatly increases the number and variety of conformations and functions that can be effected by synthesized proteins. Previous reports have demonstrated that NSAAs can improve enzyme efficiency (205), expand transcriptional control (119), and improve performance of therapeutics (32). Here we will develop and demonstrate new functionalities of NSAAs in our TAG-less *E. coli* strain (Aim 1.a.1), including (ii) photoisomerisable NSAAs that could be used to control switching between metabolic pathways, or between product generation and growth, and (iii) bio-orthogonal reactive NSAAs that could couple proteins *via* click chemistry to each other in novel ways, or to derivatized surfaces or inorganic catalysts (see Aim 3). These demonstrations will also establish that our recoded strain does not require use of nonsense suppressors to incorporate NSAAs, as do standard NSAA incorporation techniques. We will also use NSAA incorporation to (i) detect whether our recoded *E.coli* from Aim 1.a.1 has actually eliminated all use of the TAG codon (e.g., that frameshifts or unrecognized coding regions do not generate coding transcripts containing TAG codons).

Preliminary results: In the context of Aim 2.b.1 we have successfully used the Flexizyme system (140) to load 10 D-amino acids onto tRNAs. In preparation for task (i) below, we have already received aaRS/tRNA pairs for *p*-acetylphenylalanine (p-AcF), *p*-azidophenylalanine (p-AzF), and *p*-azophenyl-phenylalanine (p-AzPF).

Research Design: (i) To diagnose the presence of unpredicted TAG stop codons in our recoded *E. coli* strain, we will create a version of this strain containing p-AcF or p-AzF aaRS/tRNA pairs, grow it in the presence of the corresponding NSAA, extract proteins containing this NSAA by binding to aminooxy or alkyne-functionalized beads, and perform tandem mass spectrometry on the bound proteins. Any new TAG codons will be handed back to Aim 1.a.1 to be eliminated *via* MAGE. (ii) We will use p-AzPF, a photoisomerizable amino acid (11), to toggle enzyme activity in response to light wavelength. p-AzPF can be engineered into proteins of interest and those variants selected where only one cis/trans isomer of p-AzPF leads to enzyme inactivity (42). Initially, (ii.a) we will focus on metabolic flux control of amino acid synthesis by finding cis-inactivated (cis-ina) and trans-inactivated (trans-ina) variants of serine acetyltransferase (SATase; gene = cysE) and serine hydroxymethyl transferase (SHMT; gene = glyA), responsible for the conversion of Ser to Cys precursor *O*-acetylserine, and Ser to Gly, respectively. Combining the trans-ina SATase and cis-ina SHMT into a single strain will allow the control of metabolic flux, as can be assayed through light induced auxotrophy for Cys under 334 nm light and for Gly under 420 nm light. (ii.b) We will then attempt to engineer a strain that overproduces glycine under 334 nm light when provided cysteine, and that also overproduces cysteine at 420 nm when provided glycine. We will start by abrogating negative feedbacks in cysteine synthesis by mutagenizing the CysE cysteine binding pocket (196) with photoisomerizable P-AzPF, and in glycine synthesis by MAGE-mutating away PurR repressor binding sites regulating glyA (192). We will then explore using MAGE to optimize flux through these pathways by adjusting ribosome binding sites for source and production pathways, and by knocking out competing pathways, as we did for lycopene production (213). To enrich for cysteine and glycine overproducers at their respective wavelengths, we will exploit the fact that both amino acids are used to produce glutathione, which is itself needed for transition metal detoxification (81; 166). By providing glycine and excess glutamate but no cysteine at 420nm, and cysteine and excess glutamate but no glycine at 334 nm, in the presence of metal toxins, mutants that produce more cysteine and glycine at high levels at their respective wavelengths will experience a growth advantage. We hypothesize that by alternating between these conditions at each MAGE cycle, we can generate a single strain that overproduces both amino acids at their respective wavelengths. (ii.c) Finally, we will explore light-inactivated variants of essential *E. coli* genes (63), wherein inactivation is the result of one or multiple p-AzPF residues. Light-inactivated essential genes that do not interfere with metabolism and are bacteriostatic may be used to cycle between growth and metabolite production. We will also explore light-inactivated essential genes containing multiple p-AzPF residues as a control measure for engineered organisms, whereby cells will be enabled to grow under a particular wavelength of light, while growth under the entire light spectrum will inhibit growth. (iii) We will experiment with use of p-AzF or p-AcF to generate cytoplasmic double bond mimics on addition of bifunctional alkyne or aminooxy linkers. Our long-term goal will be to use NSAA-enabled interprotein bonds to physically couple enzymes in a pathway so as to improve pathway flux in the same way that synthetic protein scaffolds have been used (52) – except, here, without the need for separate scaffold proteins and domains. However, this depends on the ability to engineer distinct NSAAs into distinct proteins, for if two proteins use the same NSAA they will likely assemble promiscuously *vs* form stoichiometric aggregates. Thus, work on NSAA-enabled stoichiometric protein aggregates will need to wait until sufficient progress has been made on our AGR-eliminated strain (1.a.1). In the meantime, we will experiment with creating *intra*-protein double bond mimics that can occlude the active sites of enzymes, a device we expect to employ in 1.a.4 below (see 1.a.4.(i) for details).

Timeline: (i) Year 1; (ii) Years 2-5, (iii) Years 4-5.

Potential problems and alternatives: If suppression of an NSAA-containing protein is not 100% when NSAAs are withheld, or incorporation is not 100% when they are provided, we can perform directed evolution on the aaRS/tRNAs providing undesired suppression to better approach these targets. (ii)

Instead of using of photoisomerisable NSAAs for pathway switching, we could use the methods of (iii) to close off active sites of pathway enzymes on addition of appropriate click chemistry. This approach would be less easily reversible than photoisomerisable NSAAs, but could potentially be more efficient and would be unaffected by light scattering, absorbance, and turbidity. (iii) Efficient azide-alkyne click chemistry requires either Cu catalysts which could be toxic to cells, or strained-ring alkynes that are not available as NSAAs. A possible solution is to create proteins with multiple p-AzF residues and supplement cells with a bifunctional strained-ring alkyne linker to attach them. While ketone-aminooxy click chemistry is not subject to such requirements, there is a risk that the aminooxy could react with other ketones or aldehydes in the cell.

1.a.3 Assessment of viral resistance and genomic isolation of recoded *E. coli* strains

Because phages typically have small genomes that may not contain all codons, elimination of a codon and its processing apparatus from a viral host may not be sufficient to cause resistance to any given virus, and even if it does, resistance might be eliminated by small numbers of mutations that change the non-functional to functional codons. Nevertheless, experimental and theoretical evidence already indicates that even without codon elimination, altered codon usage can have large effects on viral productivity (36; 106). Similar considerations should apply to the ability of a recoded strain to express genes transferred horizontally from other species. Here we will assess both computationally and experimentally the degree of viral resistance of our TAG-eliminated *E. coli* strain and, prospectively, our AGR-eliminated strain.

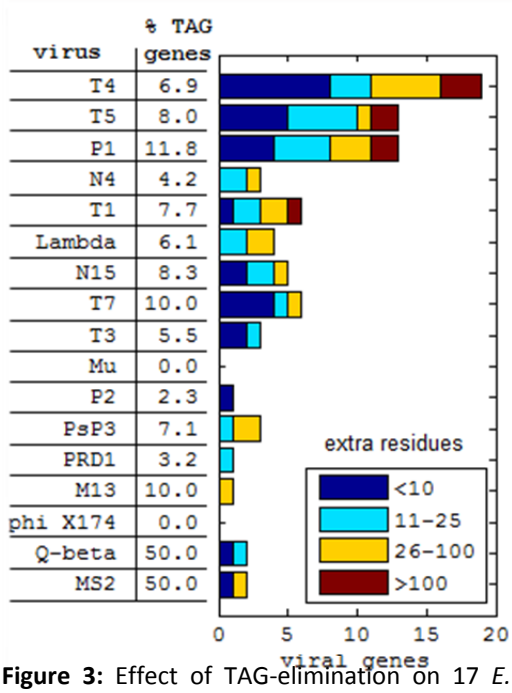


Figure 3: Effect of TAG-elimination on 17 *E. coli* phages.

Preliminary results: We have computationally analyzed the TAG codon in 17 *E. coli* phage genomes to assess how elimination of its function as a STOP codon might impact viral function (see Figure 3). For the Q-beta and MS2 phages, 50% of genes have TAG STOPs, but they each contain only four genes. If TAG codons mutate or the *E. coli* strain is NSAA-enabled, TAG-terminated genes are extended to the next functional stop codons. Over all viruses, 35% of TAG-terminated genes would have short extensions of < 10 extra residues which may have only small effects on protein function. We have not yet analyzed the impact of AGR codon elimination. We generated an *E. coli* strain in which essential genes were recoded and RF1 was eliminated (cf. (139)) and found that both T4 and T7 could propagate in that strain. However, viral propagation phenotypes may have been masked in this strain, which was of complex construction and may have possessed unidentified suppression.

Research Design: We will reassess the viral growth phenotypes with a larger panel of phages than T4 and T7 when our TAG-eliminated *E. coli* is completely assembled. We will also computationally assess the impact of AGR codon elimination. Ultimately we will test the phages on our AGR-eliminated strain.

Timeline: Years 1-2, after completion of the TAG-eliminated *E. coli* strain.

Potential problems and alternatives: Small modifications may be required to confer initial sensitivity to certain viruses (e.g. M13 bacteriophage requires an F+ host) before we can test our strain against them. Additionally, we expect that the impact of TAG elimination may be small and that viral growth phenotypes may only emerge when additional codons are removed or swapped.

1.a.4 Nutritional and genetic containment of recoded *E. coli* strain

Researchers have explored nutritional containment strategies for bacteria based on diaminopimelate (DAP) auxotrophy since the 1970s (129), and have more recently combined this with 'addiction' strategies that set up expression of a cell-killing toxin unless an exogenous suppressor compound is supplied (172). We will build similar systems into our recoded *E. coli* strain and establish their effectiveness as containment mechanisms. We will assess the degree to which the genetic isolation of recoded strains provides additional containment over wild type strains in three ways. (i) We will engineer essential genes so that they can be expressed only when an NSAA not naturally found in the environment is provided. This should overcome limitations in 'addiction' methods based on natural promoters and regulators whereby containment can be escaped by developing a few mutations or by encountering natural environmental substances similar to those that suppress toxin expression. We will then extend this method by using NSAAs not only to control growth, but as handles with which engineered organisms can be rapidly killed by addition of appropriate reactive chemistry. (ii) We will explore the degree to which the genetic isolation of our recoded strains prevents the acquisition of exogenous genetic elements that might otherwise overcome auxotrophy or genetic containment. (iii) When our AGR-recoded strain is ready (1.a.1), we will explore the degree to which containment is improved by addiction to two NSAAs. The preliminary evidence cited in 1.a.1 suggests that as many as 10 codons may be eliminable from essential genes, suggesting that a large space can eventually be generated for multiple NSAA 'addictions'.

Research Design: (i) To test containment strategies, we will build strains which incorporate combinations of DAP-auxotrophy and inducible cell-killing toxins in wild type *E. coli*, and also in a variety of strains built on our TAG-eliminated, NSAA-enabled *E. coli* in which we have engineered multiple TAG codons in essential genes. We will measure and compare growth rates, rates of escape from containment, and the lengths of time during which the strains remain viable after withdrawal of NSAAs. For simple NSAA 'addiction', the TAGs can be placed anywhere in the genes where the corresponding NSAAs do not affect protein function, and any NSAAs can be used. To explore the use of NSAAs as 'killing handles', we will use NSAAs that can participate in click chemistry (see 1.a.2.(iii)) and identify locations *via* structural analysis and experimentation where NSAA incorporation does not affect function unless the cognate click substrate is provided. Specifically, we will look for single locations near enzyme active sites where a click-appended NSAA could occlude the site, or for pairs of locations that straddle the site where the bifunctional linkers of 1.a.2.(iii) could effectively close the site. (ii) To test the effect of genetic isolation, we will add F+ *E. coli* with suitable conjugative plasmids bearing constitutive inducers of toxin escape genes, constitutive DAP generation, or both, and measure and compare the degree to which these lead to escape from containment of our different wild type and recoded *E. coli* strains. (iii) We will employ the methods of (i) to compare the growth rates, rates of escape from containment, and lengths of time of viability, of strains developed in (i) against strains in which some essential genes have been additionally engineered to incorporate two distinct NSAAs.

Timeline: (i) Years 1-2; (ii) Year 3; (iii) Year 4-5.

Potential problems and alternatives: NSAAs at important functional positions in essential genes may not be tolerated. To reduce this problem, we will use structural information to identify codons that are likely to have little impact on gene function (such as solvent-exposed Tyr codons), and use a subset of these as candidates for replacement with TAGs. Also, many TAGs will be introduced in many essential genes, and only a subset of these will need to be viable in order to establish dependence on the NSAA.

Aim 1.b: To improve capabilities for engineering and directing microbial carbon capture, we will generate a MAGE-able cyanobacterium and associated tools for assembling MAGE-altered fragments into single genomes. With these tools we will then initiate creation of a codon-

remapped strain for novel small molecule production and safe environmental usage.

Development of a MAGE-able cyanobacterium involves incorporation of a functioning λ-Red recombinase followed by optimization of MAGE efficiency (1.b.1). Our preferred target strain will be the naturally competent cyanobacterium *Synechocystis sp*. PCC 6803 (12; 143; 152; 232). Working with a competent strain will simplify the engineering required to introduce the recombinase system that will enable MAGE, but beyond that, the resulting strain will be endowed with useful complementary functionalities. For instance, MAGE efficiently recombines short oligos, while competence incorporates longer fragments by means of recA-dependent homologous recombination within long flanking homology arms. We will also explore the possibility that MAGE and competence can be synergized so that MAGE takes advantage of the enhanced DNA uptake capabilities inherent in competence. In support of this goal, we will undertake a study to improve our knowledge of natural cyanobacterial competence pathways (1.b.2). As photosynthetic organisms, cyanobacteria have evolved extensive light level- and cycle-dependent regulation of their functionality (including competence (232)) – as many as 25% of *Synechocystis sp.* PCC 6803 genes are light responsive (67) – and we will take these light dependencies into account in designing the MAGE cycle and production MAGE strain (1.b.3), which we will demonstrate by optimizing a pathway such as poly-hydroxybutyrate production. Finally, we will initiate generation of a codon-remapped strain (1.b.4).

1.b.1 Incorporation of a phage-based recombinase and optimization of MAGE efficiency

Our first task is (i) to engineer a phage-based recombinase into *Synechocystis sp.* PCC 6803. Success will be gauged by finding a statistically significant increase in genomic sequence replacement frequency in the engineered strain *vs* controls. We are principally interested in multiplexed oligo-mediated replacement of genomic sequence, but we will also pursue improvements in homologous recombination that lower homology requirements, support high multiplexity, and increase frequency of replacement. We will then proceed (ii) to optimize genome engineering efficiency along the lines used to optimize MAGE in *E. coli* (see below). Our goal is to get recombination efficiency to $\geq$ 1%. Finally (iii), we will develop Conjugative Assembly Genome Engineering (CAGE) (87) for use with *Synechocystis sp.* PCC 6803. CAGE is a protocol that enables the merging into a single strain of long MAGE-recoded segments of a genome maintained in different strains (see Figure 2).

Preliminary Results: To provide recombinase options beyond λ-Red itself, we have identified potential alternatives. Nine homologous recombinases were recently identified and found active in *E. coli* (39). To extend this list, we performed λ-Red *bet* homology searches and scanned for protein signatures for single stranded annealing domains (InterPro IPR018330), in 16 cyanophage genomes (194). Cyanophage P-SS2 orf102 (195) was a significant hit in both searches. No hits for λ-Red *exo* or *gam* were found.

Research Design: (i) *Incorporation and initial assessments of phage-based recombinases:* (i.a) Candidate recombinase operons/*bet* homologs will be placed under an inducible promoter, and a cassette containing the recombinase, chloramphenicol acetyl transferase (CAT), and a marker that has been made defective by means of a point mutation (such as kanamycin), will be targeted to a neutral site, the *ssl*0410 locus (141) using long homology regions (300bp), creating a test strain for the recombinase. We will then assess oligo-mediated correction of the kanamycin defect after provision of a correcting oligo in both recombinase-induced and -uninduced cells, and assess allele replacement frequency as the fraction of cells with restored marker function. As we expect that disabling mismatch repair will have a significant impact on efficiency (as it did in *E. coli* (213)), we will create a test strain in which mutS is disrupted *via* a targeted insertion and repeat the experiment; we expect all follow-on experiments to use the mutS-disabled test strain. (i.b) As we will also pursue improvements in recombination (*vs* oligo-mediated replacement), we will perform similar tests that assess integration of a marker cassette with

homology arms of various lengths (50-250bp). (i.c) Natural competence is optimized in *Synechocystis* cells that have incubated in high light for 5 hours after transfection (232). To look for synergies between competence and MAGE, we will perform experiments with the best recombinase in which transfections are performed at a range of times relative to the beginning of incubation at high light (250 $\mu mol_{quanta}/m^2$-sec) and which are otherwise maintained in low light (2 $\mu mol_{quanta}/m^2$-sec), as well as in constant light controls. (ii) *Optimization of MAGE:* Using the best recombinase and light induction regime from (i), we will (ii.a) improve MAGE performance by altering variables such as oligo length and number and positioning of phosphorothioate bonds (213; 214). These assessments will be performed mainly by correcting the defective marker in the test strain above, but we will also assess the frequency of replacements that result in small insertions, multi-base replacements, and both small and large deletions, using genomic (or test cassette) regions expected not to generate a phenotype. (ii.b) Once we have determined conditions that optimize MAGE in the test strain, we will then test oligo-mediated alteration of a variety of locations across the genome to scan for location-dependent biases in replacement frequency. These experiments will use single oligos targeted to generate synonymous point mutations that can be assessed by allele-specific PCR. (ii.c) We will then assess our ability to multiplex point mutation correction by providing sets of these oligos simultaneously and using our multiplexed allele-specific PCR protocols (MASC-PCR and MASC-qPCR) to measure replacement efficiency (87). (ii.d) Finally, we will experiment with improving MAGE efficiency using our "co-selection" technique (28), in which replacement of arbitrary target sites is improved by the simultaneous correction of a defective antibiotic resistance element in the vicinity of the target site. (iii) *CAGE* works by flanking a segment of a donor genome with a selection marker and a conjugative origin of transfer (*oriT*), and flanking the segment to be retained in the recipient with a positive marker and a positive/negative marker: Appropriate selections after conjugation yields genomes in which the donor and recipient segments are seamlessly joined without any scars at their junction (87). Developing CAGE for *Synechocystis* PCC 6803 will require (iii.1) establishing a conjugation system and (iii.2) identifying appropriate markers. (iii.1) While conjugative transfer from *E. coli* to *Synechocystis* PCC 6803 is well established (102; 127; 187), we have found no reports indicating that *Synechocystis* PCC 6803 can be a conjugative donor. CAGE could be performed by transforming *E. coli* with *Synechocystis* genome fragments for subsequent conjugative transfer (or assembly in *E. coli* followed by transfer) to a recipient *Synechocystis*, but this would be cumbersome and constrained by size limitations that would interfere with efficient hierarchical assembly (see Figure 2). Therefore we will attempt to establish a direct *Synechocystis* donor capability by exploring endogenous and engineered plasmids known to replicate in *Synechocystis* PCC 6803 (127; 169; 224) and other broad host range conjugative plasmids (127; 169; 224). (iii.2) Many antibiotic markers that work in *E. coli* also work in *Synechocystis* although their dosages and light-sensitivity must be considered (102). We will assess these and expect to find suitable sets for *E. coli* to *Synechocystis* transfer, and for *Synechocystis* to *Synechocystis* transfer if (iii.1) is successful. The positive/negative marker will be the key issue. The *tolC* positive/negative marker (46) currently used in *E. coli* CAGE (87) has no strong homolog in *Synechocystis* PCC 6803, which also lacks the *BtuB* colicin E1 co-receptor (29; 46) (however, homologs of the *acrAB* efflux system that confer SDS resistance with *tolC* (202) appear to be present). We will see if use of the *E. coli* genes for these factors restores function, and also similarly explore the *galK* (217) system. The inducible I-Sce (90) system and fusions of positive and negative markers provide yet other alternatives.

Timeline: (i) Years 1-2. (ii) Years 2-3. (iii) Years 3-5.

Potential problems and alternatives: (i) If no phage recombinase appears to support oligo-mediated recombination, we can try the recombinases in another competent cyanobacterium (such as *Synechococcus sp.* PCC 7002 (62) or *Thermosynechococcus elongatus* BP-1 (89)). If we are unable to reach our goal of $\geq$ 1% efficiency, we can try using directed evolution to improve the recombinase. If

these steps fail, we can transfer *Synechocystis* genome fragments to *E. coli* where they can be easily MAGE-altered, and then transfer them back to *Synechocystis* using (iii). (iii.1) If *Synechocystis* cannot be made to function as a conjugative donor, we can explore assembly of *Synechocystis* genome segments in another organism using large fragment methods developed by JCVI (65; 66) and the Fujita Lab (88), use *Synechocystis*'s natural competence (1.b.2) to re-introduce them into a recipient, and then apply the selections of (iii.2).

1.b.2 Improve understanding of natural cyanobacterial competence pathways

Research Design: We will transform an existing pooled *Synechocystis sp* PCC 6803 transposon insertion library (19; 22) with an antibiotic resistance cassette targeted to *ssl*0410 (141), and measure the abundance of insertion mutants in cultures grown with and without antibiotic using high-throughput sequencing (111) of transposon insertion sites in the populations. Genes with transposon sites with lower relative abundance in antibiotic *vs* no antibiotic will be candidate competence genes whose functions will be confirmed in independent targeted knockouts. A list of 25 *Synechocystis sp* PCC 6803 candidates based on reports in the literature (152; 228; 231) and by our own searches for *Synechocystis* homologs of *H. influenza* and *E. coli* competence genes (184) will serve as positive controls.

Timeline: Year 2.

Potential problems and alternatives: The assay may not find *essential* genes involved in competence.

1.b.3 Setting up the MAGE cycle and production MAGE strain and demonstrating pathway optimization

Research Design: Our information on the dependence of oligo-mediated replacement frequency on light intensity and cycles will have been obtained in 1.b.1, prior to MAGE optimization in 1.b.2, so we will first reassess what lighting conditions yield the best frequency. We will then convert the test MAGE strain into a production strain taking these dependencies into account. As part of this we will explore options for building in MAGE-optimal light regulation; for instance, if we find that efficiency is greatest when induction of the recombinase takes place at the end of the high light period, we will test incorporation of a light-sensitive promoter element such as the PE1 element of psAB P1 (141). We will also delete the defective marker used to test recombinase efficiency. We will then set up instrumentation to automate the *Synechocystis* MAGE cycle by modifying our *E. coli* MAGE configuration (213) to integrate the required lighting control. We will then demonstrate the MAGE cycle by optimizing a pathway in *Synechocystis*, as we did lycopene production in *E. coli* (213). Poly-hydroxybutryate (PHB) production is a plausible target because *Synechocystis* has several PHB pathway enzymes (see Table 2) also present in PHB-production organism *Ralstonia eutropha* (161). As PHB inclusions affect cell density, strains can be screened for improved PHB production by Percoll gradient centrifugation (7; 108).

gene	function
slr1830	PHB synthase
slr1829	PHB synthase component
slr1993	β-ketothiolase
slr1994	NADPH acetoacetyl-CoA reductase
slr0229	PHB dehydrogenase
slr0060	PHB hydrolase homolog

Table 2: PHB pathway enzymes in *Synechocystis sp.* PCC 6803

Timeline: Year 4-5.

Potential problems and alternatives: PHB inclusions may also change cellular refractive index. If so, this would enable use of flow cytometry for screening.

1.b.4 Initial steps towards a codon remapped *Synechocystis* strain

Once we have a production MAGE-able *Synechocystis* strain and automated cycle, we will begin removal of a codon following the procedures we used for E. coli (87) and carry it as far forward as possible within the remaining period of the Center. Based on our success in *E. coli*, we do not anticipate problems with

this step.

Aim 1.c We will improve the frequency of λ-Red recombination of long (≥ 1kb) DNA constructs into the *E. coli* chromosome by at least an order of magnitude (from ≤ 0.1% to ≥ 1%), and, with this, extend our Multiplex Automated Genome Engineering (MAGE) method to accommodate long constructs in addition to short oligos (MAGE-2).

Preliminary results: We developed an efficient system for assaying the λ-Red-mediated insertion of a ~1.2 kb construct consisting of a *kanR* gene flanked by *lacZ* homology segments (*lacZ::kanR*). After electroporating the construct into inducibly λ-Red+ *E. coli* cells (strain EcNR2 (213)), recombination frequency is easily assessed as the ratio of white kanamycin-resistant colonies to the total number of viable cells; white colonies indicate not only that *kanR* is active but also that it has inserted into the correct location. With this assay, we found that: (a) the single-stranded *lacZ::kanR* construct is highly recombinogenic, (b) single-stranded insertion constructs preferentially target the lagging strand, and (c) designed mismatch mutations within a *lacZ::kanR* dsDNA cassette typically co-segregate in a manner that indicates that a single strand provides all of the genetic information during a given recombination (137). These findings suggest that, contrary to previously published models (37; 165), λ-Red-mediated incorporation of large dsDNA cassettes proceeds through a fully single stranded intermediate. These results also suggest that such ssDNA intermediates are incorporated as Okazaki fragments, in a similar manner as short ssDNA oligonucleotides. Thus, the frequency of λ-Red-mediated ~1 kb insertion can likely be improved by similar strategies as those used in the development of MAGE. Table 3 shows one preliminary result in which the strategic incorporation of phosphorothioate bonds markedly improves the insertion frequency of a ~1 kb construct.

Strand phosporothioation		Mean ± Stdev Recomb. Freq. × 10^{-5} (N=2)	
Lagging	Leading		
Y	Y	6.3 ± 1.2	
Y	N	6.1 ± 3.0	p = .0068
N	Y	0.3 ± 0.2	
N	N	0.4 ± 0.3	

Table 3: λ-Red-mediated incorporation of ~1kb dsDNA cassettes is improved by four phosphorothioate bonds on the 5' end of the lagging-targeting strand, but not the leading-targeting strand (137).

Research Design: To achieve our goal of increasing the recombination frequency of gene insertion to > 1%, we will both (i) optimize the gene insertion DNA cassettes and (ii) engineer the bacterial strains used to perform these recombinations. To address (i) we will (i.a) optimize the length of insertion cassettes' flanking homologies, (i.b) protect the desired lagging-targeting strand by strategic placement of phosphorothioate linkages, and (i.c) create designed overhangs to bias λ Exo so that it degrades the undesired leading-targeting strand while sparing the recombinogenic lagging-targeting strand. A scheme for achieving this is shown in Figure 4. To achieve (ii), we will (ii.a) use directed evolution in order to optimize the λ Red proteins for recombination of ~1 kb fragments. To this end, we will use error prone PCR in order to diversify episomal λ-Red genes. Highly recombinogenic variants will then be enriched by sequentially recombining antibiotic resistance insertion cassettes and selecting for their incorporation. λ

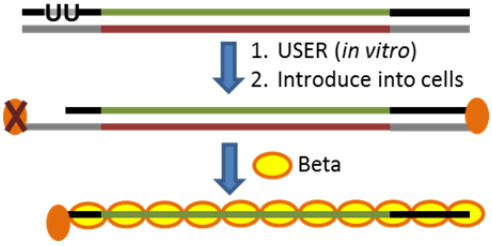


Figure 4: Generation of recombination cassettes for optimized Exo interaction. The cassette is PCR-generated with a primer containing deoxyuridines (dUs). The USER enzyme is then used to excise the dUs, leaving behind a 3' overhang. This overhang hinders Exo (orange oval) from degrading the lagging-targeting strand (green/black), thus restricting Exo to degrading the leading-targeting strand (red/gray) and recruiting Beta to the remaining (lagging-targeting) strand.

Red genes from promising variants will be sequenced, and MAGE will be used to transfer the detected mutations to the corresponding chromosomal genes of EcNR2, in order to confirm the desired phenotype. We also plan to (ii.b) test the overexpression of λ-Red components (particularly Beta and Exo), (ii.c) generate and assess mutations which improve the accessibility of the lagging strand of the replication fork, and (ii.d) assess whether knocking out several endonucleases and exonucleases augments recombination frequency. The latter is motivated by preliminary results that suggest that ~1 kb dsDNA cassettes undergo substantial degradation by nucleases before and/or after being processed into ssDNA by λ Exo (137; 176); thus, removing such nucleases may have a beneficial effect on recombination frequency. (ii.c) is motivated by our new mechanistic insight that ~1 kb dsDNA cassettes recombine by annealing to exposed ssDNA at the lagging strand of the replication fork. Mutations in *dnaG* primase and associated helicase (*dnaB*) have previously been shown (30; 201) to increase the amount of exposed ssDNA at the lagging strand of the replication fork, and to increase the length of Okazaki fragments. We expect that increasing the accessibility of ssDNA at the replication fork in this way will enhance recombination frequency.

Timeline: (i) and (ii): Years 1-3. In Years 4-5 we will extend this method to the transfer of multi-gene pathways and multiplexed or combinatorial insertion of several heterologous genes.

Potential Problems and Alternatives: Recombination bottlenecks could arise due to inefficient entry of large DNA cassettes into the cell or saturation of the λ Red machinery. These problems could be addressed by utilizing or developing a highly competent bacterial chassis, or by overexpressing the λ Red proteins, respectively. However, given our previous success in optimizing oligonucleotide MAGE (213), we are confident we will similarly improve large cassette recombination.

Aim 1.d: We will develop suites of protein and RNA modules in *E. coli* that respond to small molecule ligands, and that can be used to sense and regulate metabolic pathways. By engineering ligand binding and allostery, we will expand the natural *E. coli* repertoire of ligands that can be sensed. We will demonstrate the use of these modules for optimizing production of useful products.

We approach the problem of building molecular sensors as an SB engineering problem: Cells offer only limited repertoires of sensors to natural metabolic or environmental molecules that are tightly and non-modularly integrated into regulatory processes. How can we extract and enlarge upon these natural sensor domains and embed them in a modular framework? We explore two tracks, both of which build on foundations laid in our current Center (see Table 1). As part of our project on increasing fatty acid production in *E. coli*, we designed a fatty acid reporter based on the natural FadR repressor, and we have recently expanded and generalized this effort to the direct engineering of other repressors to both expand their sensor capabilities and improve their properties (1.d.1). In the other we pursue development of a modular system of riboregulators initially developed by co-investigator Professor Farren Isaacs (1.d.2).

<u>1.d.1 Protein sensors of small molecules</u>

Preliminary results: (a) We have developed a fatty acid sensor based on FadR and the fadB promoter (59) in connection with a novel flux analysis ((122), and see 1.e.3) that identifies genetic targets for increasing fatty acid flux. FadR binds the fadB promoter and inhibits downstream transcription in the absence of acyl-CoA. This inhibition is relieved in the presence of acyl-CoAs with chain length ≥ 12 (see Figure 5). We have done initial work on a citrate sensor based on similar principles (98; 225). (b) Figure 6 presents initial results on the system described in *Research Design* below, with which we will mutate the binding pockets of transcriptional repressors to respond to other ligands. Shown are results of a selection with this system in which ribose repressor RbsR was mutated to respond to IPTG.

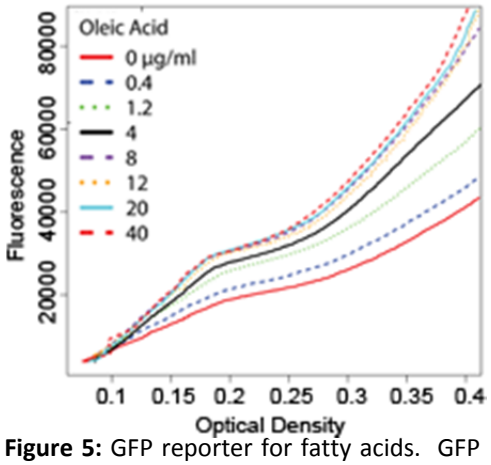


Figure 5: GFP reporter for fatty acids. GFP was placed downstream of a copy of the fadB promoter in a fadE knockout strain grown in minimal media supplemented with acetate and different amounts of oleic acid (C18:1).

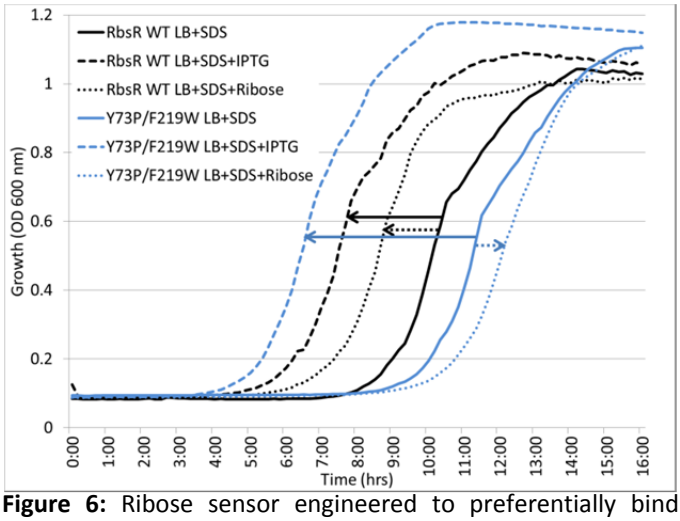


Figure 6: Ribose sensor engineered to preferentially bind IPTG. RbsR WT (black lines) or Y73P/F219W mutant (blue lines) were expressed in cells harboring tolC under control of the RbsR operator sequence. Growth curves are shown above for cells grown in LB + SDS and no inducer (solid curves), IPTG (dashed curves) or 100 µM ribose (dotted curves). Dashed or dotted curves shifted to the left of their corresponding solid lines indicate SDS resistance due to TolC expression controlled by their inducers. For WT RbsR, TolC expression is induced by either IPTG or ribose (black solid and dotted arrows). For the RbsR Y73P/F219W mutant, IPTG strongly induces TolC expression (blue solid arrow) but ribose fails to increase growth and may even decrease it (blue dotted arrow). These results suggest that RbsR binding preferences have shifted from ribose to IPTG in the Y73P/F219W mutant.

Research Design: (i) *Selection system for allosteric transcriptional regulators:* We have built and are testing (see above) the construct in Figure 7, which can be built with any inducible *E. coli* transcriptional regulator and cognate binding site. For a repressor, in the absence of inducer, the repressor represses selectable marker tolC and the cells are sensitive to SDS. When inducer is supplied, tolC is expressed and the cells are resistant to SDS. In tests with LacI, provision of IPTG yields growth in SDS at the same levels as growth without IPTG and SDS. A variant of the protocol is possible for activators. (ii) *Changing the binding specificity of regulators to respond to molecules of interest:* Redesigning a protein's specificity to even bind molecules that are close to its natural ligand has proven to be a hard problem. On the one hand, experimental methods must rely on large protein design libraries of order 1e10 or greater to identify a hit. Alternatively, computational protein design methods have advanced significantly to engineer active enzymes, but they require assaying one enzyme at a time. We will combine the atomic-level accuracy of computational design algorithms with our high throughput DNA construct synthesis method (Aim 2.b.1) to develop highly targeted libraries for redesigning transcription factor specificity, and then screen them using the selectable marker system in (i). Preliminary results are encouraging (see above). We will use the lacI/GalR family of transcription regulators as initial candidates of specificity change against a library of sugar and sugar-derivative molecule library. (iii) *Demonstration of use of sensor proteins to optimize metabolic pathway:* Regulators that are engineered to respond to molecules that are not naturally sensed create opportunities to optimize production of those molecules. All that is needed is to arrange for the regulator to drive expression of a resistance element, and the

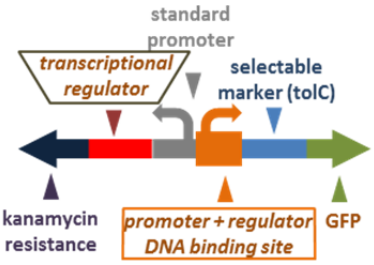


Figure 7: Construct design for evolution of binding pockets of transcriptional repressors.

selection conditions for a directed evolution for production are met. Here we will demonstrate use of natural and engineered sensors for optimizing production of a high-value molecule. Many high-value targets related to central metabolism have been identified (45). We will initially focus on production of cis,cis-muconate, a nylon precursor currently synthesized from petroleum derivatives. Cis,cis-muconate synthesis in *E. coli* has been achieved by expression of three exogenous genes (147). Here we will engineer natural transcription factors BenM and CatM from *Acinetobacter baylyi* that are regulated by cis,cis-muconate to operate in *E.coli*, and then optimize production in *E. coli* by directed evolution and MAGE. Notably, this will be the first demonstration of selection coupled with MAGE, as our MAGE optimization of lycopene production (213) required visual screens to pick production mutants. (iv) *Real-time sensors:* Finally, we will pursue the development and testing of real-time protein sensors for a subset of our target molecules by coupling our allosteric sensors to GFP (18). These sensors are produced by inserting an allosteric protein into a fluorescent reporter so that fluorescence is enabled only by the allosteric change of the sensor protein. We will use our high-throughput DNA construct synthesis method in Aim 2.a. to create libraries of GFP-sensor protein chimeras and screen for designs that have the most dynamic range.

Timeline: Optimization of system: Years 1-2. Application to cis, cis-muconate production: Years 2-3. GFP reporters: Years 3-5.

Potential problems and alternatives: Many molecules sensed by transcriptional regulators are metabolized. If this complicates selections, we will perform them in knockouts that abrogate such metabolism. For instance (cf. Figure 6), a ribulokinase knockout would block ribose consumption (121).

1.d.2 RNA sensors of small molecules

Preliminary results: We previously developed powerful and versatile RNA-based systems that regulate target protein translation (riboregulators, see Figure 8) (86). Because they are programmed by sequence, it is easy to calibrate specificity, regulate many proteins at once, and tune response; and because they operate at the translational level their operation is rapid and tunable (27). Moreover, we have now shown that riboregulators can be used to drive selections (see Figure 9).

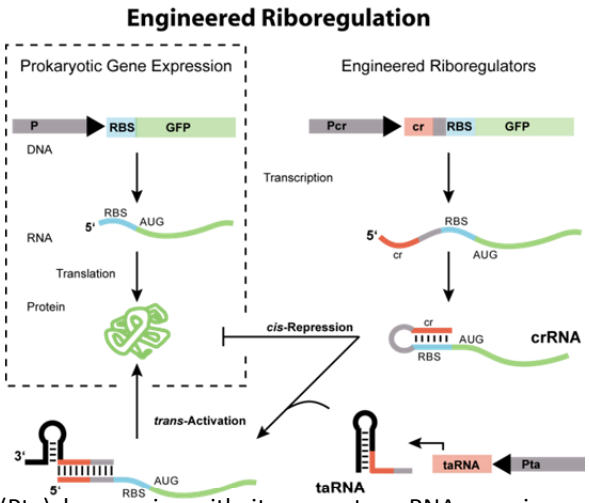


Figure 8: The riboregulator system uses short sequences (*cis*-repressed, cr, red) inserted between a promoter (Pcr) and ribosome binding site (RBS, blue), that cause transcripts to fold up and occlude the RBS (crRNA). A small non-coding *trans-activating* RNA (taRNA) produced at a second promoter (Pta) base pairs with its cognate crRNA, causing a structural rearrangement that exposes the RBS and enables translation.

Because, in addition to the useful control features above, RNAs can be used to sense small molecules (see *Research Design*), this sets the stage for development of ribosensors and riboselectors that will enable high-throughput selections and screens for organisms that produce many new small molecules, and new ways to use endogenous and exogenous small molecules to control pathway processing. Finally, these RNA control systems should be portable to many organisms with small modifications.

Research Design: We will generate a suite of *E. coli* translational regulatory, sensory and selectable modules that will modulate expression from regulatory networks, sense changes in small molecule concentrations, and select for desired phenotypes. (i) *Library of orthogonal riboregulator pairs*: To develop a large and general basis for multiplexed riboregulation, we will generate thousands of

independent cognate crRNA / taRNA pairs from a published set of 240,000 orthogonal 25mer DNA sequences (223), and characterize their dynamics using our existing GFP reporter system (86). We will focus on sequences not present in the *E. coli* genome and synthesize sequences on a DNA microchip. In theory, $>10^{15}$ independent cognate crRNA / taRNA pairs are possible (86). (ii) *Develop riboregulators that respond to molecules of interest*: Both natural (20) and synthetic (56; 222) RNA aptamers can sense and respond to changes in small molecule levels in cells. We will employ a subset of known aptamers (112) to construct a library of ligand-responsive riboregulators for target small molecules. Aptamer domains can be incorporated into either crRNAs or taRNAs, and different positionings of the domain are possible within each. We will explore the space of options so as to uncover key design principles, capabilities, and limitations of these different configurations of RNA switches. (iii) *Develop riboselectors for small-molecule phenotypes*: We will next convert a set of the small molecule-responsive riboregulators developed in (ii) into riboselectors by coupling them with trans-activated selectable markers and calibrating them for different levels of stringency using the methods indicated in Figure 9. We will also expand our suite of selectable markers beyond the *cat* gene to include other positive (e.g., kanamycin, zeocin) and positive-negative (e.g., tolC) markers. Ultimately, this strategy will allow us to regulate, sense, and select for multiple biomolecular processes in a single cell. (iv) *Demonstrate use of riboregulators in biosynthetic pathway engineering*: We will then integrate advances from aims (i)-(iii) to deploy both riboregulators and ligand-responsive riboregulators to effect screens and selections for small molecule production phenotypes in MAGE (213) optimizations of selected pathways. Riboselections should greatly improve MAGE efficiency compared to our optimization of lycopene production in (213), where screening steps in each MAGE cycle required plating and manual picking of colonies based on color. The S-adenosyl-L-methionine (SAM), tryptophan and tyrosine biosynthesis pathways are promising candidates that also possess existing aptamers (82; 124; 162; 226) that we can use immediately. We will focus initially on tryptophan and tyrosine production as these are not only important industrially, but because the genotypes developed by optimizing these individual pathways can be applied to and compared against the genotypes developed in our directed evolution of engineered syntrophic *E. coli* auxotrophs (see (180) and Aim 1.e).

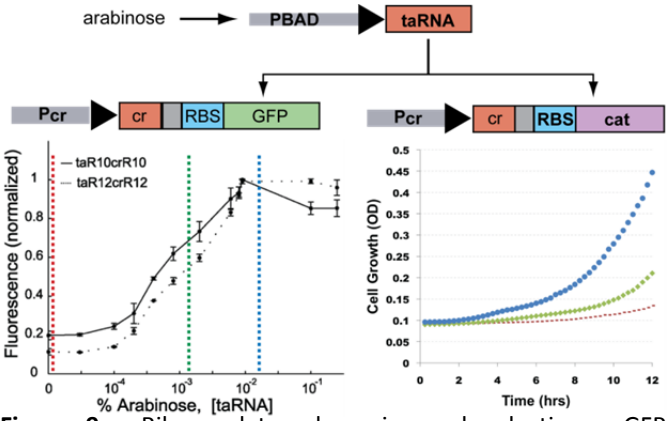


Figure 9: Riboregulator dynamics and selection. GFP expression from riboregualtor system can be tuned by modulating the level of taRNA as depicted by the arabinose-dependent promoter (left plot). By replacing GFP with the *cat* gene, population viability can be linked to riboregulator expression (right plot). Only at elevated levels of taRNA (blue) are cells able to survive 20 ug/mL of chloramphenicol.

Timeline: (i) Year 1. (ii) and (iii): Years 1-3. (iv) Years 4-5.

Potential problems and alternatives: Aptamers generated *via in vitro* methods may not work *in vivo*: thus, in (iv) we will focus on aptamers obtained from *in vivo* sources (20). We will have opportunities to experiment with *in vitro*-designed aptamers in *ex vivo* systems in Aim 3.a. *In vitro*-designed aptamers for *exogenous* molecules may also be used to institute exogenous control over pathways; tetramethylrhodamine and theophylline are examples that have successfully been used (13; 26).

Aim 1.e: We will develop tools for identifying metabolic linkages between species in natural microbial communities, and methods for mathematically modeling and engineering microbial consortia to efficiently produce useful molecules or remediate environmental toxins.

1.e.1 Novel method for identification of metabolic interactions within microbial communities

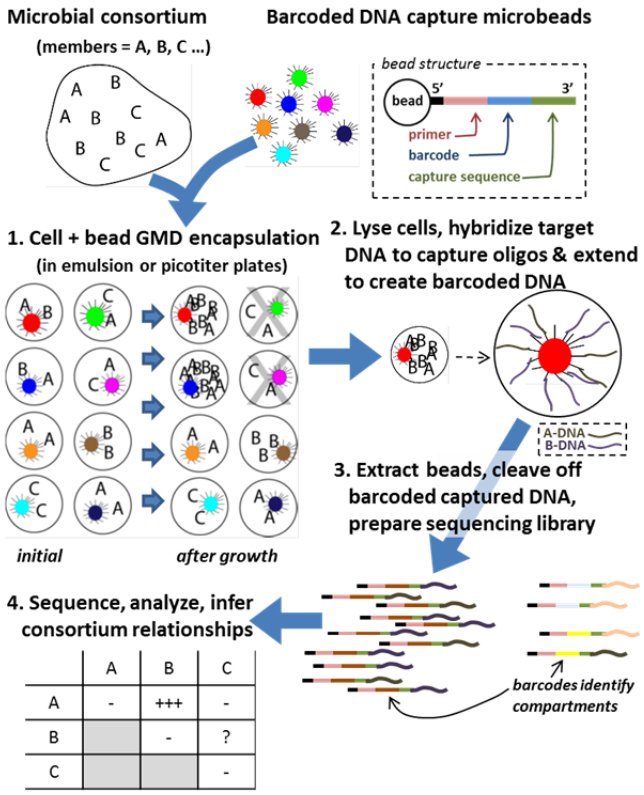
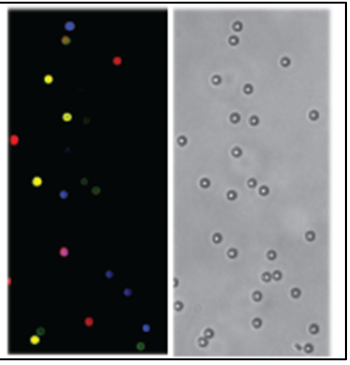


Figure 10: Cell-Cell Interaction Mapping (CCIM) (see text).

Gel microdroplets (GMDs) have been used since the 1980s (218) for analysis of individual microbial cells from complex mixtures. Recently, through integration with flow cytometry and fluorescence activated cell sorting, GMDs are being used to enable sorting, selective isolation, and growth of organisms previously considered to be uncultivatable (2; 38; 160; 233; 234). Here we plan to use GMDs to identify and characterize sets of metabolically interacting microbes from within microbial consortia, a technique we call Cell-Cell Interaction Mapping (CCIM) (see Figure 10). Small numbers of cells from a microbial consortium will be encapsulated into agarose GMDs along with microbeads covered with DNA sequences containing barcodes unique to each bead upstream of 3' capture sequences targeted to specific genomic sequences of interest (see *Preliminary Results*). The GMDs are generated so as to contain, with high probability, at most one barcoded bead and at most two bacteria, and are maintained either in emulsion or in picotiter plates (see *Research Design*). The bacteria in the GMDs are then allowed to grow for a small number of cell divisions (Figure 10 (1)). After growth, using protocols that ensure that GMDs remain their stability and mutual isolation, the GMDs are then treated so as to cause the bacteria to be lysed and enable their genomic DNA to hybridize to the capture sequences. The capture sequences are then extended by a primer extension reaction to create barcoded copies of the targeted genomic sequences (Figure 10 (2)). The beads are then extracted from their GMDS, and their barcoded captured sequences are cleaved off and prepared for sequencing (Figure 10 (3)). By sequencing the barcodes and captured sequences from the entire sample, we can reconstruct the contents of all compartments (Figure 10 (4)). In some experiments, universal capture sequences targeting 16S rRNA genes will be used to interrogate compartments for taxon membership, and taxon growth in compartments will be assessed by the number of corresponding 16S rRNA sequences. Beads that were compartmentalized with single bacteria will provide baseline data on non-consortial taxon growth. Together these data will enable identification of taxon growth interactions. In other experiments we will target representative genes in metabolic pathways so as to interrogate metabolic interrelationships of consortia. We will vary medium and temperature conditions to explore how these factors modulate interactions, and we will confirm and further characterize detected interactions by ordinary co-culturing of taxa and analyzing growth, mRNA and protein expression, metabolite production *via* LC-MS.

Preliminary results: We have developed barcoded microbeads (see Figure 11) in the context of another grant (NIH CEGS P50 HG005550-0) in which we are developing single cell expression assays. Briefly, the barcoded beads are created by emulsion microbead PCR (180) with DNA templates containing a degenerate sequence stretch. We have filed a patent on this technology. Tara Gianoulis on our team has had considerable experience with functional and sequence metagenomic analysis of microbial

Figure 11: 1 μm microbeads containing 20nt degenerate barcode sequence. *Left:* The first base of the degenerate barcode was sequenced on all beads where A bases were detected with Cy5, T with Cy3, C with TxRed, and G with FITC. Shown is a composite image indicating the clonality of barcode sequences on individual beads. *Right:* White light image of the same beads.



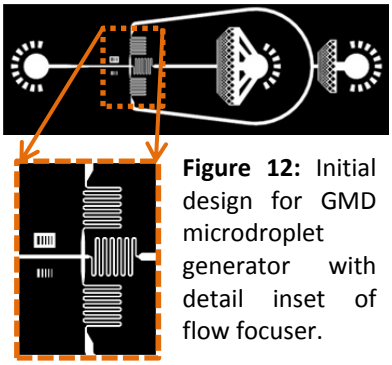


Figure 12: Initial design for GMD microdroplet generator with detail inset of flow focuser.

communities (64; 116; 154; 186). She has been working with the Weitz lab (see Letter of Support, Appendix 6) on the design of a droplet generator to create the GMDs (see Figure 12) and with protocols they are developing for in-compartment cell lysis and sequence capture.

Research Design: (i) We will start work with mixtures of engineered obligate *E. coli* syntrophs that were previously developed in the Church Lab (180) or are being developed in Aim 1.e.2 below, and with GMD emulsions generated by the protocols of (58). Our initial object will be to establish protocols that assure that (i.a) GMDs are generated efficiently with the required numbers of bacteria and microbeads (see above); (i.b) GMD compartmental integrity and mutual isolation are maintained throughout growth, cell lysis, and DNA capture procedures; (i.c) GMDs are not contaminated by free bacteria or microbeads in the bulk medium (cf. (58)) (i.d) growth, lysis and DNA capture are optimized, and (i.e) sequencing results are concordant with actual GMD syntroph population abundances. Using syntrophs that will be created in Aim 1.e.2 with distinct fluorescent markers, we will be able to test i.a-i.c by microscopy and FACS analysis, while i.e-i.f can be assessed by using beads with capture primers specific to the distinct marker genes. (ii) When our procedures have been developed, we will apply them directly to the obligate syntroph mutagenesis and selection described in 1.e.3. Meanwhile, here in 1.e.1 we will apply CCIM to map interactions between members of at least one environmental community whose major elements have been identified, such as the acid mine drainage (43), waste water treatment (123; 146), or fuel spillage (54; 79; 101) communities. As part of this work, we will deliberately vary the context of interaction (temperature, pH, nutrients, small molecules, etc.) in order to characterize how these factors modulate interactions.

Timeline: (i) Years 1-3. (ii) Years 4-5.

Potential problems and alternatives: It may prove difficult to identify conditions in which compartment integrity is maintained during cell lysis and genomic DNA capture (i.d). Based on the ease of perfusing GMDs with nutrients, enzymes, and agents like SDS, the Weitz lab has developed procedures (unpublished) for lysing cells and probing DNA in stable compartments; however capture of DNA targets by oligos on microbeads may require DNA fragmentation and/or denaturation. If we cannot develop workable protocols, we will create and manipulate GMD compartments in picotiter plates instead of emulsion, in which compartment isolation can be assured.

1.e.2 Engineering of obligately syntrophic consortia

As a substrate to engineering microbial consortia with enhanced metabolic capabilities, we first develop synthetic consortia with well-defined metabolic interactions. Expanding on the strategy of our earlier evolved syntrophy between *E. coli* tryptophan and tyrosine auxotrophs (180), we consider *E. coli* with engineered auxotrophies for any number of essential metabolites (nucleotides, vitamins, and cofactors in addition to amino acids). Each member can propagate only in the presence of another consortium

member that supplies the necessary metabolite. We will investigate the detailed mechanism of the syntrophic metabolite exchange through quantitative modeling and experimental analysis. Previous studies (53) have shown that auxotrophic strains grown in sugar-limiting media supplemented with the deficient metabolite can outcompete prototrophic strains because they do not expend energy on synthesis of the metabolite. Thus, it may be possible to divide the labor of executing long biosynthesis pathways into dedicated strains where each pathway can be individually optimized. Through directed evolution and MAGE (213), we will optimize each pathway to augment the synthrophic interaction.

Preliminary results: Using MAGE we constructed a library of strains with single auxotrophy in one of 14 amino acids, 91 strains with all combinations of double auxotrophies, and a small set of 4 strains with triple auxotrophies. Experiments in which these strains are allowed to cross-feed in pairs, triplets, or quadruplet sets shows synergistic interactions that indicate that some amino acids are better shared in co-cultures (Figure 13). Long-term co-culturing of these cells generates mutants with improved co-culture growth rate, which we are characterizing by full genome sequencing.

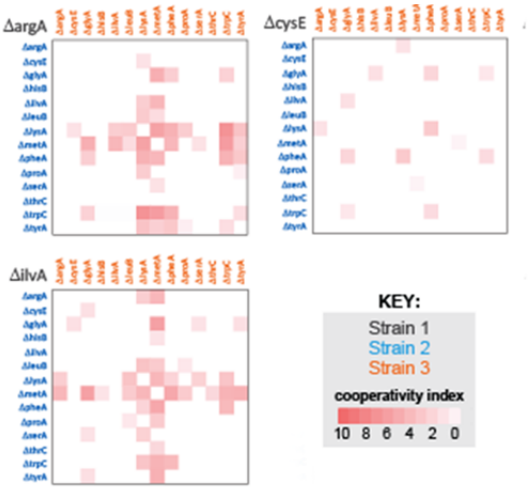


Figure 13: Three of 14 heat maps obtained representing degrees of cooperation of all 364 cross feeding triplets of 14 auxotrophs grown in minimal media plus glucose. Each heat map represents a Strain 1 auxotroph and each map coordinate represents its partnered Strain 2 and 3 auxotrophs. Growth of each culture depends on each partners' provision of nutrients needed by the others. The color of a coordinate represents the triplet's "cooperativity index," a measure of cell density at the end of a 9 day growth period that started with 1e7 cells/ml of each strain. Judged by overall color, some auxotrophs were superior sharers. From best to worst, the order is M K F W R I G L Y P S T H C.

Research Design: We will study these systems experimentally and computationally at various time scales. In the short term (5-20 generations), we expect the kinetics of bioproduction and import/export to dominate and determine a short-term steady state stoichiometry. Preliminary ODE models show that the relative export ability of the exchanged metabolites sets the system more than intrinsic growth rates or other parameters. At a longer time scale (20-400 generations), selective pressures will result in evolutionary changes to export ability (improved cooperativity), import or utilization (improved scavenging), as well as bioproduction. At this evolutionary time scale the steady state of the system becomes less deterministic. However, we hypothesize that the selective pressures revealed by mutations observed in the various members of the system will help us identify syntrophic networks compatible with traits we wish to enhance and engineer. We will use models from evolutionary dynamics and game theory proposed for the evolution and maintenance of cooperativity in a system, to design bacterial consortia engineered for these traits. The "snowdrift" game theory (71) is particularly relevant to our designs in terms of explaining development of cheaters. In our networks cheaters can develop to improve their fitness by reducing their export ability thus keeping more for themselves, as well as by developing prototrophy through cryptic pathways. We propose network complexity and spatial structuring as solutions to overcome these escapee problems. We hope to study how increased cross-feeding network complexity will reduce the chance that that any strain will be able to escape dependence on its partners. We will also design systems that tie cooperating strains together physically resulting in greater benefit to cooperating groups (77). Finally, we will compare our evolved syntrophs with strains engineered singly to overproduce amino acids, such as the tryptophan and tyrosine overproducers we will evolve using riboregulators in 1.d.2. We expect strains evolved here and in 1.e.3

that share amino acids with partner auxotrophs to exhibit mutations that upregulate export compared to single strain overproducers.

Timeline: Evolution of auxotrophs: Years 1-3; Characterization of partnerships and pathway relationships: Years 2-3.

Potential problems and alternatives: When 1.e.1 sufficiently developed, we will perform some evolutions with GMDs and compare results with those obtained by our current co-culturing methods.

Gene Group	Adjustment	Enzyme function	Pathway	Dependence
fabH	enhance	actyl-CoA-acp transacylase	fatty acid biosynthesis 1 and elongation	None
fabBF	enhance	b-ketoacyl-acp synthase	fatty acid elongation	None
fabAZ	enhance	hydroxydecanoyl -acp-dehydrase	fatty acid elongation	None
acnAB	reduce	aconitate hydratase 1	TCA cycle, glycoxylate cycle	fabBF / fabAZ / fabH
aceEF,lpd	enhance	pyruvate dehydrogenase (complex)	glycolysis, acetyl-CoA synthesis,	fabBF / fabAZ / fabH
putA	enhance	proline dehydrogenase	proline degredation 1	fabBF / fabAZ / fabH
fadI	reduce	fatty acid oxidation complex	fatty acid oxidation I	(fabBF or fabAZ) and (putA or fabH) / fabG and (acnAB or glnA)
glnA	reduce	glutamine synthetase	glutamine biosynthesis 1	(fabBF or fabAZ) and (aceEF+lpd or acnAB)
fabG	enhance	b-ketoacyl-acp reductase	fatty acid elongation	(fadI or fabBF) and glnA / acnAB and (fabBF or acnAB fadI)

Table 4: Redirector recommended engineering changes predicted to improve *E. coli* fatty acid production. The output describes gene groups whose enzyme pathway fluxes need to be enhanced or reduced. In addition to genes in fatty acid synthesis and degradation pathways, Redirector has identified genes in other pathways that have significant indirect effects, such as proline degradation and glutamine biosynthesis.

1.e.3 Design and optimization of metabolic pathways distributed across microbial consortia

Compared to 'blind' mutagenesis, optimization of metabolic pathway fluxes in microbes is more efficient when mutagenesis is targeted to elements in the genome with direct or indirect effects on pathway operation. Several targeted mutagenesis methods have been developed, including MAGE (213), error-prone PCR of target genes or regulatory machinery (4), structure-guided domain shuffling (153), and synthesis of mutated gene libraries using degenerate sequences (14; 73; 104). However, targeted mutagenesis requires pathway modeling tools to indicate what genetic elements need to be modified. To that end, we (122) and others (24; 168) have developed Flux Balance Analysis (FBA)-based methods that predict such elements, while ordinary differential equation (ODE)-based methods have also proved effective (198; 203). However, few tools are available to design and optimize pathways distributed over microbial consortia. One problem is that FBA allows optimization of only a *single* objective function, which, if specified as a pathway distributed over the microbes in a consortium, usually results in unrealistic solutions in which sub-networks of some microbes essentially operate as slaves to the other microbes. Another problem is that, as our experience with evolved syntrophs has shown (170; 180), transport processes prove to be critical to syntrophy. These are often not well characterized – a problem long recognized by the biotechnology industry (25; 107) – and thus not represented well in models. We take a multi-faceted approach to these issues.

Preliminary results: To optimize whole organism metabolic networks at the gene level, we developed the Redirector Framework, an enhancement of our earlier published GDLS algorithm (122). This method models metabolic flux pressures in a way that generates *targeted tuning levels* for fluxes through enzyme groups (see Table 4), a feature that takes advantage of tuning capabilities inherent in MAGE (213), rather than simpler but coarser recommendations to overexpress or knock out genes. Redirector also has favorable performance scaling with metabolic network sizes, is built on freely available solving packages (GLPK and SCIP), and will be made available as open source.

Research Design: (i) We will extend our Redirector system to accommodate multiple organisms. The ability of Redirector to balance complex products will form the foundation of dividing pathways between hosts. Our method combines bi-level optimization (23) with objective functions that weight

biomass contributions of the organisms independently along with pathway components for the product in question. The algorithm then operates iteratively to adjust the weights of components in the objective function as it identifies further elements to optimize. We propose a design that will explore the space of production pathway divisions between hosts and, using our compound objective functions, simultaneously optimize the production and metabolite transport costs and synergies on the growth of each host. Strain coexistence will be induced in the model through a balance of outer problem incentives and strain growth ratio constraints. Outer problem objective factors have the advantage of creating a system where strain growth equality comes naturally from each strain maximizing its own growth while under the stress from the parts of the production pathway it is hosting. Equality incentives can also use a stepped coefficient system, where differences beyond a threshold are more significantly penalized. We believe that this scheme of weighting and connecting the biomass of each organism in the objective will also act to overcome the tendency to make some organisms slaves of the others, since each organism will be optimized to grow at some level. Incentivized strain growth equality should induce division of production pathways by the outer problem control variables; such that production induced stress is divided relatively evenly depending on availability of metabolite transport. The benefits of distributing production stress would then be balanced against the transportation energy costs by the inclusion of well modeled transportation reactions. As an alternative, we will also explore multi-objective optimization as a means of coupling the growth objectives of the organisms to pathway productivity within a linear flux framework (40; 142; 210). (ii) We will then use the GMD methods developed in 1.e.1, in conjunction with MAGE and/or synthetic gene library construction (see Aim 2.a), to target mutations to genes predicted to be important by our syntrophy optimization modeling. In brief, mutagenized populations of each of a pair of syntrophs will be sorted into GMDs so that most GMDs contain one or two bacteria, and the bacteria will be allowed to grow within their GMDs. Here, instead of identifying cooperating strains by sequence analysis as in 1.e.1, we will use fluorescent markers engineered into the strains to sort out those GMDs exhibiting maximal growth *via* FACS. A subset of ~100-~1000 of these GMDs will be directed to multi-well plates for grow-out, metabolic characterization, and sequencing of the engineered mutations, while the remainder of the population will be released from their GMDs and then recombined for additional rounds of growth, selection, and sequencing. We expect by these means to identify sets of gene variants that improve the growth of the consortium, and also to determine whether and how iteration of GMD selection leads to continued improvement of consortium growth.

Timeline: (i) Years 1-3. (ii) Years 3-5.

Potential problems and alternatives: (i) Beyond 'master-slave', a second kind of artifactual solution arises when optimum production is best achieved by having the pathway completely in one strain, but, to satisfy objective function incentives for co-growth, the other stain is altered to grow at a minimal level. Comparisons against single strain optimizations, and careful modeling of dual strain transport and energy costs, will be essential to distinguishing these cases. (ii) Diversity created *via* mutagenesis must be calibrated to the capacity of the system. E.g., if 1e8 GMDs are occupied by organisms O1 and O2, each MAGEd with 50% efficiency to diversities $\delta 1$ and $\delta 2$, to get 90% of all combinations requires $\delta 1 \cdot \delta 2 = \sim 1.1e7$ (from $\mu = 1e8/(4 \cdot \delta 1 \cdot \delta 2)$ and $e^{-\mu} = .1$). As $\delta 1 = \delta 2 \Rightarrow \delta 1 = \delta 2 = \sim 3300$, we may need to alternate heavy diversifications between organisms. We can also try to couple resistance elements with fluorescence and perform selections *vs* screens to overcome capacity bottlenecks due to screening.

Aim 2: Enhanced *in vitro* DNA and protein synthesis

Aim 2. We will develop a coordinated set of *in vitro* methods for generating complex, high fidelity libraries of DNA constructs, and *in vitro* methods for generating proteins using modified

ribosomes that will support enhanced translation control and integration of non-standard protein features. We will demonstrate the use of these capabilities for implementing pathways and processes important to biofuel production and carbon capture.

Improvements in *in vitro* DNA construct and protein synthesis are having an enabling effect on all systems biology and bioengineering. *In vitro* protein synthesis, in the context of library display methods such as ribosome and mRNA display, allows creation of libraries of much higher diversity (~1e12-1e14) than display methods involving microbes (up to ~1e9) or phage (~1e12) (49; 97; 118; 156), largely because they avoid inefficient transformation steps. Meanwhile, *in vitro* DNA synthesis is uncoupling the sequence spaces explored by research from the spaces available in repositories and accessible natural populations. It is now possible to synthesize DNA sequences from across the tree of life directly from information in sequence databases, including sequences from metagenomic databases or from unculturable organisms (14); and to freely adjust codon usage, secondary structure, and ribosome binding sequences to optimize expression (109; 159) – all while building in diversity at precisely targeted locations and at high multiplexity. Together these capabilities enable the exploration of much larger and more specifically targeted conformational spaces than can be obtained by *in vivo* methods.

Our Center has contributed to this progress by developing efficient methods for synthesizing DNA constructs from oligonucleotide arrays (104; 130; 199) (see Figure 14), and improving *in vitro* ribosome assembly procedures (92). But we foresee that *additional significant advances will be achieved by replacing in vivo with synthetic in vitro components and functionality to the extent possible* (61). The rationale for this is that any steps taken in this direction will free systems used in research and (ultimately) production from the need to conform to the requirements of living cells. Conflicts between SB-enhanced functions and their living chassis can be roughly classed into three (related) types: implementation barriers, resource conflicts, and toxicity. Inefficient transformation, discussed above, is an example of an implementation barrier that complicates engineering and limits library sizes in *in vivo* screens. Resource conflicts arise when native and synthetic functions compete for the same components. Examples are found in attempts to modify translation machinery within living cells: For instance, the use of NSAAs in non-TAG recoded *E. coli* (see Aim 1.a) requires use of TAG nonsense suppressors that compete with native release factors for coded TAGs, leading to inefficient NSAA incorporation. Similar conflicts arise in attempts to develop and use alternative ribosomes in living cells, e.g., ribosomes that can accept D-amino acids (41), as the need to maintain translation by native ribosomes limits productivity of the modified ribosomes. Finally, toxicity may arise whenever an SB-enhanced function is inappropriately recognized by native proteins or regulators in a cell, or disrupts a required physical structure or parameter such as membrane integrity or potential.

Several strategies have been pursued to address these limitations. Genome engineering, such as our *E. coli* recoding efforts (Aim 1.a) can effectively deal with conflicts for codons, but cannot in itself address conflicts for translation machinery or toxicity without additional engineering specific for these. "Minimal genome" projects (60; 68; 103) have potential to reduce conflicts between engineered and native functionality by reducing the extent of the latter; initial results are encouraging (113; 164). In Aim 1 and elsewhere, Church Lab is pursuing both of these strategies, but our focus here in Aim 2 is the third approach of *developing and optimizing a completely defined in vitro transcription and translation (IVT) system*, eliminating the living chassis entirely. Our system is motivated by PURE system (182; 183) based on purified ribosome components, but differs from it by being completely defined. In addition to the benefits already achieved by IVT in maximizing library diversity (noted above), our defined synthetic system will provide a high-resolution analytical test bed and development platform for the *engineering of new translation functionality* that will complement and extend genome engineering methods (see Aim 1). Examples of new translation functionalities under development include ribosomes that

efficiently incorporate D-amino acids, and construction of orthogonal ribosomes (10; 215). D-amino acids are of interest because their inclusion in proteins expands the space of conformations attainable by proteins; they also appear to be refractory to biodegradation (17; 94; 95). The broader promise of re-engineering translation is to convert it from its evolved function as a biologically regulated producer of cellular proteins to a general purpose template-directed assembly factory for human-designed chemical components. Ultimately, we would like to develop an *in vitro* self-replicating entity with minimal dependence on cells. Here, however, we focus on how enhanced *in vitro* translation not only enables new functionality in the manners just noted, but also positions us to develop *ex vivo* systems that implement pathways of integrated biochemical with physico-chemical components. Such systems will inevitably need biological elements that operate in regimes employed by industrial chemistry that are incompatible with cellular life. Such elements are unlikely to be effectively screened or selected *in vivo vs. in vitro*, and our plans for *in vitro* optimization are laid out in Aim 3.

Meanwhile, here in Aim 2, sub-Aim 2.a is focused on further improvement and application of our DNA construct synthesis capability, while sub-Aim 2.b develops and applies our defined synthetic IVT system.

Aim 2.a We will develop an automated DNA construct synthesis pipeline built from off-the-shelf components for generating high-fidelity libraries of complex constructs from oligonucleotide arrays, and develop algorithms for specifying construct libraries to be synthesized that enable optimization of biofuel-, carbon capture-, and remediation-relevant pathways and enzymes.

The goal of Aim 2.a.1 is to further improve the DNA construct synthesis capability previously developed by our Center. Aim 2.a.2 develops a method for intelligently targeting the diversity of the large construct libraries that can be built through this capability and demonstrating this on a DOE-relevant application.

2.a.1 Methods for parallel and multiplexed synthesis of large numbers of DNA constructs

Preliminary Results: Our proof-of-concept study first demonstrated assembly of genes from DNA microchips in order to reduce costs of gene synthesis (199). We and others have found this process difficult to scale up because as the number of constructs in an oligo pool increases, the reliability of assembly of each gene decreases. To address this, we developed an improved DNA synthesis protocol that reduces oligo pool complexity prior to the DNA assembly steps (104). Briefly, we design array oligos by bracketing sequences for construct assemblies between computationally designed orthogonal primers so that we can amplify out DNA subpools containing only the oligos needed for individual assemblies (see Figure 14). These primers are then removed using Type IIs restriction enzymes, and the subpool is assembled using an optimized PCA (polymerase cycling assembly) protocol. Using these protocols, we performed assemblies on 47 sequences: 45 produced correct length assemblies, perfect sequences were found among 18/20 cloned genes, and error rates were 1/1350bp-1/1500bp for the simpler and ~1/315bp for the more challenging assemblies (104). We are in the process of setting up a pipeline for a shared DNA synthesis capability based on this method called "SynBioSIS" and designed along the lines of MOSIS

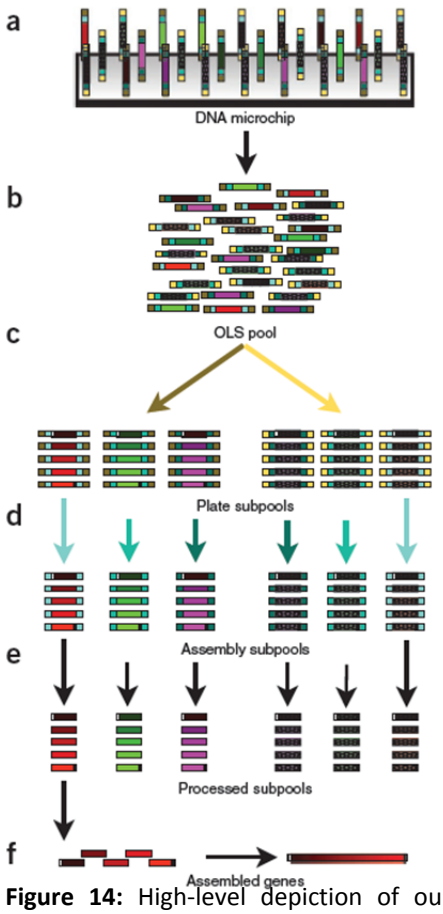


Figure 14: High-level depiction of our protocol for scalable DNA construct synthesis by selective amplification and assembly of subpools of oligos generated on oligo array (104).

(138). SynBioSIS will schedule purchase of oligo arrays for construct synthesis, collect information from users about oligos to be combined into constructs, batch user projects into scheduled arrays, and coordinate construct synthesis and distribution. SynBioSIS will allow researchers to explore new ways of using large numbers of synthetic DNA constructs, and provide us with data on the scalability, reliability, and costs of next-generation gene synthesis technologies. This pipeline could be of interest as a possible prototype of a DOE User Facility function.

Research Design: We will develop two methods for scaling up DNA construct synthesis: (i) We will automate and parallelize synthesis of constructs into 96 well plates using off-the-shelf and Open Source components with initial focus on integrating liquid handling robots. This will be practical for synthesizing batches of several thousand constructs. Once we have optimized reliability, error rates, and scalability, we will integrate them into SynBioSIS and begin discussions with DOE about how this capability could be adapted as a DOE User Facility function. (ii) We will then develop a highly multiplexed protocol for assembly of long DNA libraries from the microchip using emulsions (see Figure 15). Array oligos will be PCR-amplified and made single-stranded using protocols we developed previously (114), and then microbeads, each loaded with assembly pool-specific primers, will be used to pull down only those oligos from individual assembly subpools. After primer extension, the beads are emulsified in a mixture of hotstart polymerase and Type IIs restriction enzyme. The emulsified beads will be incubated at 37˚C to start the digestion and then assembled by PCR, which will both activate the hotstart polymerase and inactivate the restriction enzyme. Finally, the emulsion is broken and the resultant DNA can be size-selected and re-amplified. Library construction methods will be assayed for accuracy and distribution by Illumina or 454 sequencing depending on the read-length desired (up to 300 or 750 bp respectively). We will attempt to construct assemblies in this manner of up to 5 oligos, resulting in constructs up to 530bp in length.

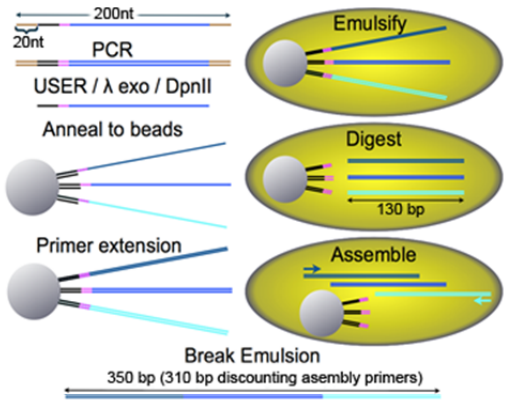


Figure 15: Proposed process for multiplexed assembly of oligos in emulsions. See text for details.

We will test and develop these capabilities by generating construct libraries that will be employed by other Aims of the Center, but to demonstrate its potential as a DOE User Facility function, we will also synthesize and make available one or more construct libraries of interest to the DOE research community, such as a previously published set of computed artificial C-fixation pathways (9) assembled from enzymes from organisms genes across the tree of life.

Timeline: (i) Year 1, with our prototype pipeline operational in Year 2. (ii) Years 2-4.

Potential problems and alternatives: Based on our extensive experience with automation and emulsion methods, we do not anticipate difficulties with this Aim.

2.a.2 Niche- and structure-guided construct libraries and mutation methods for enzyme improvement

Our strategy, which we call “chimeragenesis” (depicted in Figure 16), is to identify and use *niche-selected domains* as the basis of construct libraries for evolving new enzyme and protein properties. We propose to derive enzymes that operate within sets of conditions – for instance, at low pH and high temperature – by shuffling and otherwise varying sequences of conserved structural elements in proteins from organisms that operate in these *individual* conditions. Based on our high throughput DNA construct synthesis capability (2.a.1 above), our strategy significantly scales up the diversity of the *source* sequences that are recombined compared to other methods of structure-guided recombination.

For instance, while (153) generated 6561 shuffled proteins, these originated from only 3 source proteins divided into 8 domains ($6561 = 3^8$). Our goal is rather to generate chimeras from protein families with hundreds to thousands of members that are segmented into smaller numbers of domain elements. To develop and demonstrate our method, we will generate chimeragenic libraries for a gene family important to cellulosic biofuel generation: Glycosyl-Hydrolase family 9 (GH9), which has ~258 members across a wide range of phylogenetic space. We have recently identified a secreted GH9 from *Clostridium phytofermentans* that is essential for this organism's degradation of cellulose (200). This effort will also employ and help drive development of our proposed improvements to *in vitro* selection methods in Aim 3. We will attempt to generate GH9s that operate in pH ranges and temperatures used for culturing *E. coli* or *S. cerevisiae* modified for bioethanol production (pH 6-7, 30°C for *E. coli* (151)); 4.7-5.0, 30°C for *S. cerevisiae* (177; 191)), and we will also attempt to generate thermostable enzymes (>55°C, (80; 209)); the former will support Simultaneous Saccharification and Fermentation and Consolidated Bioprocessing applications, while the latter will demonstrate the advantages of *in vitro* methods that operate outside the optimum physiological ranges of organisms. Our high-throughput methods will complement the work of other groups who are attempting to modify and evaluate cellulases (8; 80).

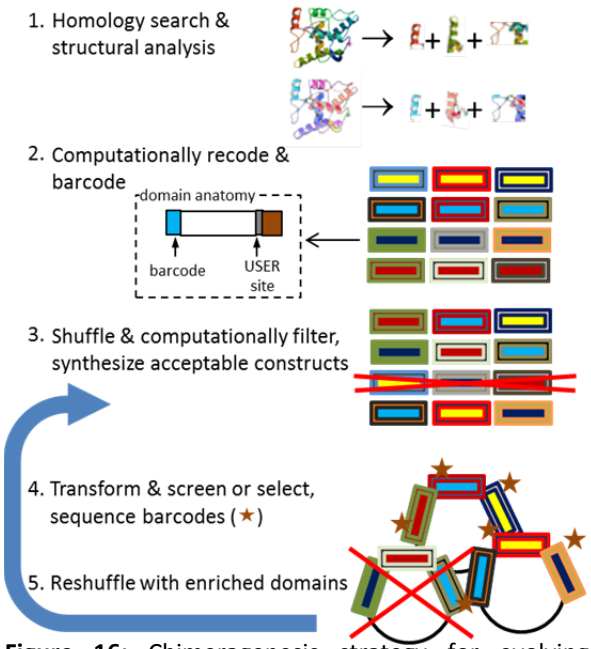


Figure 16: Chimeragenesis strategy for evolving enzymes with desired combinations of properties as chimeras of enzymes with individual properties.

Research Design: Starting with a set of genes of interest, the general workflow we will develop for chimeragenesis is (Figure 16): (i) We will use bioinformatics to identify homologs, and use research literature and metagenomic databases to identify organisms in target environmental niches. We will then structurally analyze the proteins to identify conserved structural elements and breakpoints. (ii) We then design DNA constructs from the elements identified in (i): Codons will be picked compatible with *E. coli* codon usage, sequence barcodes built in, and construct ends designed with unique restriction or priming sites to accommodate shuffling and assembly. (iii) We then shuffle these elements *in silico*, filter transcript sequences for structural compatibility with *E. coli* expression (109; 159), and filter protein sequences for correct folding by free energy minimization. Diversity in regions predicted to be functionally important will be designed into remaining sequences through degenerate sequence regions, and varying inter-domain spacer regions will be integrated as needed. Statistical samples of these modified sequences will be rechecked computationally. (iv) The final construct sequence library will then be decomposed into oligos for construct assembly, and actual constructs will be built using our construct synthesis pipeline (2.a.1). (v) Assembled constructs will be generated either *in vivo* or *in vitro* and screened with high multiplexity. In initial experiments we will simply sequence and reverify "winners," and analyze these sequences for features that might explain their improved performance. In later experiments, we will incorporate additional rounds of selection, diversification (through shuffling, error-prone PCR, or *in vivo* with MAGE) and rescreening.

In developing and applying chimeragenesis for GH9 modification, we will perform metagenomic and literature searches to identify enzymes operative in our temperature and pH niches. We will perform high-throughput selections using the emulsion ribosome display-based *in vitro* selection method of (128)

with improvements described in Aim 3.a, using modified oligosaccharide substrates such as resorufin or 4-methylumbelliferyl lactopyranoside (8; 80) that generate fluorescent signals upon enzyme activity. Our methods will generate a set of enzyme variants enriched for improved performance in our target pH and temperature ranges. We will independently verify a sample of top candidates, characterize their performance on an actual cellulosic substrate, and compare them against (and in combination with) standard cellulases.

Timeline: Years 1-2: Set up and testing of system without extra rounds of diversification. Years 3-4: Applications to GH9. Year 5: Incorporate second round of diversification.

Potential problems and alternatives: (a) To our knowledge, detection of fluorescence from cleaved resorufin or 4-methylumbelliferyl lactopyranoside in ~5-50 fl emulsion compartment volumes (from (73) and (128), respectively) has not been reported. However, as (128) successfully demonstrated fluorescence-based selection in emulsion in a similar reaction (using fluorescein di-β-galactopyranoside), we expect this to work here (see Aim 3). (b) The operative pH of intracellular enzymes depends on cytoplasmic pH, which is not usually well documented. This may complicate identification of source GH9s for pH ranges. We will start our list with secreted GH9s and extend it using organism niche pH as a proxy, as it has been found that the cytoplasms of acidophiles is often more acidic than that of neutrophiles (185).

Aim 2.b We will develop an efficient and completely defined *in vitro* transcription and translation system that possesses novel properties including: reduced dependency on rRNA modifications, support for orthogonal expression systems, and increased ability to incorporate D amino acids or other non-standard features.

2.b.1 Development of completely defined IVT system and synthetic ribosomes with enhanced function

Our goal includes not only development of a completely defined and optimized IVT capability, but also protocols for complete definition and optimized assembly of the ribosomes themselves. This is a major difference from the PURE system (182; 183), which, while it uses defined synthetic initiation factors, elongation factors, release factors, and amino-acyl tRNA synthetases (aaRS), requires natural assembled ribosomes extracted from *E. coli.* Towards this goal, we are attempting to build modified functional ribosomes that operate with *in vitro* generated rRNAs, an effort we will eventually also extend to tRNAs. Native rRNAs are known to be modified at 6 positions within a 79 nucleotide critical region (60); thus, to use *in vitro* generated rRNAs will require either identifying and adding defined factors that make these modifications, or developing ribosomes that have reduced need for them. Hand in hand with these developments, we also attempt to develop ribosomes with enhanced functionality, with initial focus on generating ribosomes capable of efficiently incorporating D-amino acids. Later we will develop orthogonal pairs of ribosomal subunits and, based on them, completely orthogonal expression systems. A long term application of this work will be the development of template-directed D-peptide polymerases that can make functional or D/L hybrids.

Preliminary Results: As steps towards our goal of efficiently generating synthetic ribosomes, we have successfully expressed all 54 ribosomal proteins using the PURE system, all 32 translation factors from 6 *E. coli* strains, and have developed simplified assembly procedures from individual proteins and rRNAs (92). We also identified *E. coli* elongation factor 4 (EF4) as a supplementary component of our in IVT that can increase product fidelity and yield by about 30%. Our experiment also confirmed that EF4 can increase the Mg^{2+} tolerance of our IVT, which serves as an important requirement for 30S and 50S ribosome subunit assembly. As expected, our defined IVT system is already conferring advantages over the commercially available PURE system (182; 183), as it is proving more flexible and easier to manipulate. For instance, it has simplified buffer screening as we work towards replacing traditional

reconstitution methods that include temperature and Mg^{2+} concentration shifts. Regarding the assembly and operation of ribosomes, earlier work in our Center established that reconstituted 50S *E. coli* ribosomal subunits assembled from purified disaggregated 50S proteins and *in vitro* transcribed (unmodified) 23S rRNA, when supplemented with a ribosome-free *E. coli* cytoplasm extract (S150) and purified 30S complexes, were capable of translating a test mRNA transcript at only 18-fold smaller levels than native 50S complexes (92). The S150 extract presumably contains factors that process, fold, or facilitate 50S assembly around the 23S rRNA. Finally, regarding development of ribosomes that efficiently incorporate D-amino acids, we have created a working construct for a 23S rRNA gene library and have designed a library of ribosomal proteins mutated in 11 positions (~1e6 complexity) based on a recent ribosome crystal structure (211).

Research Design: Development of our synthetic defined IVT system involves three main activities: (a) efficient generation of the individual components (ribosomal proteins and translation factors), (b) optimization of ribosome reconstitution, and (c) optimization of protein production. These activities need to be coordinated and iterated as we engineer the ribosome itself to accommodate D-amino acids. Given the progress in (a) reported in *Preliminary Results*, our initial focus will be on (b) and (c): (i) We will optimize ribosome reconstitution and protein productivity based on our synthetic versions of *native* ribosomal proteins and translation factors, by addition of various lengths of flanking sequences, use of chaperonins and RNA-helicases, and by using different RNA polymerases (e.g., *E. coli* RNAP *vs.* phage T7-RNAP). Our focus on chaperonins and RNA-helicases is based on our hypothesis that such factors are components of S150 (see *Preliminary Results*). Meanwhile, (ii) we will develop ribosomes with improved D-amino acid incorporation according to the scheme depicted in Figure 17. Rather than generate modified libraries in living cells as in (41), or with limited *in vitro* multiplexity as in (35), we will generate them in our defined IVT system (cf. (i) above), mutating protein components of the ribosome peptidyl transferase center in addition to the 23S rRNA, and select from ribosomes *via* ribosomal display by forcing them to incorporate a D-amino acid at a codon in a transcript that precedes an affinity tag sequence. Sequencing will identify the variant 23S rRNA sequences that improve D-amino acid incorporation. We will perform this experiment both with native ribosomal proteins and with a set generated with mutations in key positions in the peptidyl-transferase center (see *Preliminary Results*). In the latter case, we will use mass spectrometry to identify mutated positions which exhibit non-wild type residues, and follow these up as needed with position-specific screens to identify which residues are most effective at improving D-amino acid incorporation. To charge the tRNAs with D-amino acids, we will use the Flexizyme system (140); we have already successfully activated and loaded 10 of 20 D-amino acids in this way. (iii) Using a variant of the method above, we will evolve the interface between the 30S and 50S subunits so as to create orthogonal pairs of ribosomal subunits. We will then modify each pair to employ distinct Shine Delgarno sequences to achieve completely orthogonal protein expression pathways.

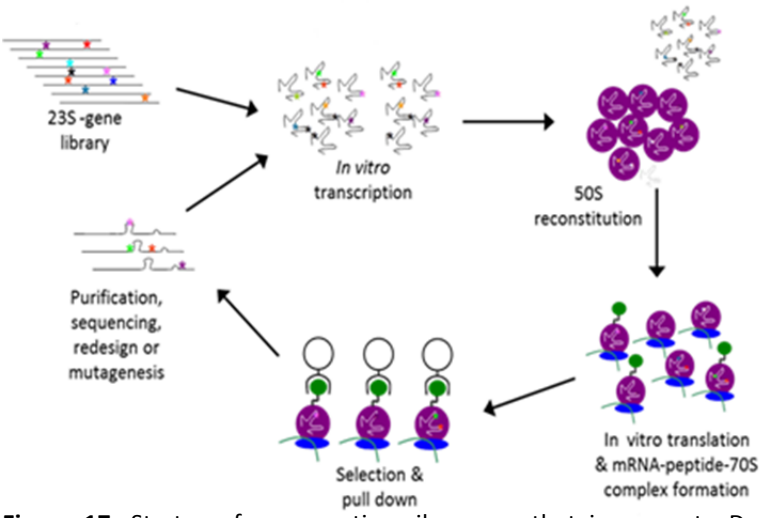


Figure 17: Strategy for generating ribosomes that incorporate D-amino acids efficiently. See text for details.

Timeline: (i) Optimization of 30S reconstitution: Years 1-2; Optimization of 50S reconstitution: Years 2-3.

(ii) Isolation of mutant rRNAs for D-amino acid incorporation: Year 2-3. Isolation of mutated ribosomal proteins for D-amino acid incorporation: Years 3-4. (iii) Orthogonal ribosomal subunits and SD sequences: Years 3-5.

Potential problems and alternatives: (i) If simply adjusting buffer, temperature, or incubation time proves insufficient to enable efficient 30S and 50S reconstitution, we will attempt to identify and include the enzymes in the S150 fraction that contribute to rRNA modification, folding, and subunit assembly. (ii) If, in the scheme in Figure 17, 50S reconstitution is too inefficient, or pulldown is too non-specific, we will revert to a published *in vivo* method in which mutant rRNAs are tagged with an MS2 stem loop that is exposed in assembled ribosomes. Ribosomes containing mutant rRNAs can then be pulled down by affinity purification (229).

2.b.2 Novel ribosome display system for characterizing protein-protein and ligand-mediated interactions

We are leveraging our work on IVT systems to develop a high-throughput method for interrogating protein functions and interactions by coupling ribosome display with next generation DNA sequencing. Briefly, (see Figure 18) proteins generated by ribosomal display will be allowed to interact and be assayed with ligands in solution, and then they will be captured on the surface of a Polonator flow cell. Functions of the proteins can be analyzed by multiplexed biochemical assays that give single molecule readouts. For example, by adding fluorescently labeled ligands and substrates, we can use this system to identify proteins interacting with these molecules in a massively parallel fashion. The mRNAs attached to the ribosomes will then be converted into cDNAs and amplified into rolling circle amplicons (which we term "rolonies," see Aim 3) on the flow cell surface. These will be sequenced *in situ* on the Polonator sequencing platform, and interacting proteins will be detected by their statistical colocalization patterns. There is potential to interrogate ~1e8 protein molecules in a single assay (1000s of different protein species). This project couples two technologies that have been well developed individually but not effectively combined to date: ribosome display and next generation sequencing.

Preliminary Results: We have established conditions under which proteins with an enzyme tag generated *via* ribosome display can be captured on a flow cell surface, and we have generated rolonies from their mRNAs on a flow cell surface. We are working with an updated Polonator instrument that can sequence such rolonies (see *Preliminary Results*, Aim 3).

Research Design: The key elements of this system are: (i) *Design of ribosome display library:* This involves creating construct libraries that append appropriate affinity tags, spacers, and ribosome stalling elements to the proteins to be displayed. (ii) *Ribosome display protocols:* We use a newly commercialized PURE protein synthesis system with all purified components (enzymes, tRNA, ribosome, and small molecules) which can avoid interferences from endogenous enzymes in crude cell extracts (204). (iii) *Surface capture of protein-ribosome-mRNA complexes:* The surface of a slide or Polonator flow cell is coated with an enzyme ligand that covalently captures the enzyme tag on the generated proteins. (iv) *Rolony generation:* A

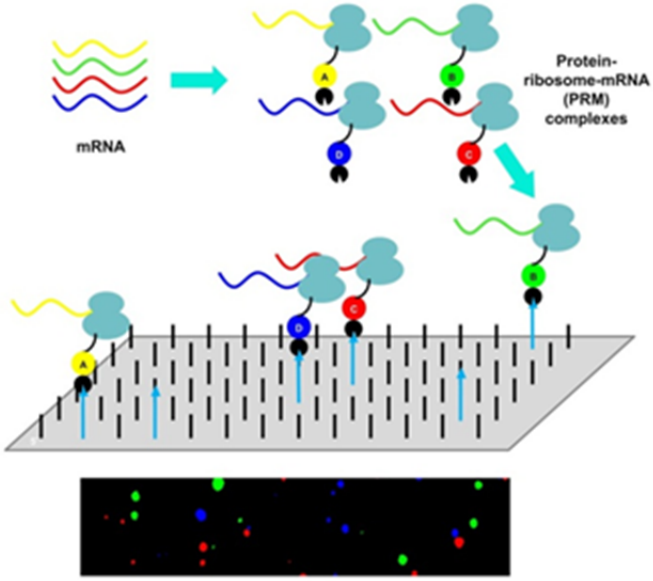


Figure 18: Strategy for novel ribosome display system for identifying protein functions. See text for details.

polyacrylamide gel containing the reagents needed for *in vitro* reverse transcription and isothermal amplification is formed over the immobilized complexes and rolonies are generated from the mRNA sequences using procedures previously published by our Lab (134; 135). (v) *Sequencing* (180) is then used to identify the proteins. (vi) *Statistical analysis:* Co-localization will be assessed with measures developed in (133) or (157).

Timeline: Year 1: Completion and publication of a small-library proof-of-concept. Year 2-3: Optimization of protocols and Polonator instrumentation for libraries with thousands of proteins. Year 4-5: Large library experiments.

Potential problems and alternatives: Based on our *Preliminary Results*, we expect no major problems.

Aim 3: Development and optimization of *ex vivo* biology

Aim 3: We will advance the ability to operate biological pathways *ex vivo* by enabling the *optimization by genetic selection* of biologically functionalized nanoparticles.

Interest is growing in *ex vivo* systems that integrate enzymes with inorganic chemistry and electrochemistry. These systems promise to incorporate the functionality and molecular precision of biochemistry into the physical and chemical regimes exploited by industrial chemistry, without worrying about the toxicity to or overhead of maintaining whole organisms (76). Many ways of co-structuring biological and physical elements are now being explored, including: coupling of enzymes and redox carriers to electrodes (221), co-loading of entire pathways of enzymes and redox factors onto microbeads (55), trapping and caging of enzymes in polymers (99) or self-assembling functionalized gels (220), and scaffolding of inorganic catalysts on viral coat proteins (144; 145). Structured nano-environments combining enzymes and other factors hold particular promise, as they have potential to: improve enzyme stability and reduce loss compared to bulk solution (99), overcome mass transfer limitations of surface-arrayed enzymes (93), improve pathway flux by reducing diffusion of reaction intermediates away from the enzymes (55), and offer new ways to anchor and functionalize inorganic factors for interaction with enzymes (220; 221). However, most work in this area neglects a key problem: how can the biological elements be optimized for best function in these environments? Most studies focus on the nano-environments and simply use natural enzymes modified for environmental attachment; the enzymes themselves are not systematically studied and are not subjects of further optimization. In large part this is because a key aspect of life-derived biochemistry that makes *in vivo* systems readily optimizable – *genetics* – has not yet been incorporated into current *ex vivo* systems.

The goal of Aim 3 is to develop this capability. The central problem is to extend and enhance *in vitro* selection methods to accommodate broader classes of enzymes and also multi-enzyme pathways. Our starting point is a pair of methods developed by the Tawfik and Griffiths Labs for *in vitro* selection of single enzymes (73; 74; 128). (A related bacterial display in emulsion system (1) is not pursued here because it is not truly *ex vivo*.) Beginning with a DNA library coding for mutated forms of the enzyme, aqueous emulsion compartments are created *in vitro* that contain at most one molecule of the DNA library and multiple copies of the cognate enzyme. The enzymes then operate on their substrates in their individual compartments, and FACS is used to sort compartments with the highest reaction product yields. The two methods differ mainly in that in one, ribosome display of the enzyme and the enzymatic reaction itself take place in the same compartments (128), while in the other, the reaction is conducted in an emulsion loaded with previously generated microbeads on which both the DNA and its enzyme are attached. These methods have successfully yielded a *Flavobacterium* sp. phosphotriesterase variant with a 63-fold increase in k_{cat} over wild type (73), and an Ebg enzyme (a defective β-galactosidase) comparable to *in vivo* evolved variants (128). However, general applicability of these

methods is limited because: (A) They rely on *modified substrates* to generate the signal by which enzyme activity is measured, and these are very reaction-specific. In (128), the substrate is FDG, which releases fluorescein with β-galactosidase activity; in (73), the phosphotriesterase is provided with a caged-biotin-conjugated substrate, to allow capture of product on the microbead for later quantitation *via* a labeled antibody. (B) Use of emulsions and FACS limits measurement of compartments to a small number of parameters that can be assessed in a single pass. This limits the ability to measure multiple molecular species such as would be needed for assessment of multi-enzyme reactions in addition to sources of variance such as compartment volumes and numbers of copies of enzyme molecules.

We plan to address (A) by developing ways of assessing reaction performance in compartments that do not require modified substrates. Here we will leverage our work in Aim 1.d on engineering proteins and riboregulators that respond to small molecules, as well as our work in Aim 2 on enhancing IVT. Our response to (B) will be to develop a method of arraying aqueous reaction compartments on our Polonator (48) platform that will enable multi-pass measurements of compartments (as well as sequencing), and the sorting of compartmentalized DNAs based on these measurements. Here we will leverage not only the Polonator itself but the Church Lab's general expertise in protocol automation (also exemplified by MAGE (213)). However, in concert with the development of this arrayed capability, we will also work to improve *in vitro* selection in emulsion. This is because emulsions will be a convenient form in which to work out our approaches to (A), and as shown by the successes mentioned above, emulsion *in vitro* selections can still prove very useful. Emulsion and arrayed compartment methods differ in that the former are "low precision" while the latter are "high precision" systems (see Figure 19), each of which can be useful. Figure 20 depicts the range of options we will consider for both emulsions and for the arrayed compartment capability. Details are described in Aims 3.a and 3.b.

A primary goal will be to improve *ex vivo* selection to a point where we can demonstrate optimization of a multi-step reaction (Aim 3.c): Specifically, we aim to be the first group to use high-throughput directed evolution methods to optimize a two-step enzymatic reaction in ~5-50 fl volume nano-compartments. We will also demonstrate the application of directed evolution methods to enzymes attached to "rolonies" (see Aim 3.a), DNA structures between ~300nm (50) and 500nm in diameter (Polonator sequencing rolonies). This would be a first instance of optimization of an enzyme mounted on a nano-

Performance of "good" enzymes relative to noise required for selection

initial frequency	rounds of selection				
	6	10	15	20	25
1E-06	3.6	2.2	1.7	1.5	1.4
1E-07	5.3	2.7	2	1.7	1.5
1E-08	7.7	3.4	2.3	1.9	1.6
1E-09	12.5	4.3	2.6	2.1	1.8

$$sensitivity = P(selected|good); fpr = P(selected|bad)$$

$$P(good|selected) = \frac{sensitivity \cdot P(good)}{sensitivity \cdot P(good) + fpr \cdot \left(1 - P(good)\right)}$$

Model:

$$P(good_{i+1}) = P(good_i|selected)$$

$$N = \min_i(P(good_{i+1})) \geq 1/500$$

Figure 19: "Low" *vs* "high precision" selection systems. The model on the left describes selection of high performance "good" enzymes from a mixture of "good" and noise-level "bad" enzymes. If selection noise is represented by the false positive rate (*fpr*) and "good" enzyme detection efficiency by *sensitivity*, then the performance exhibited by a "good" enzyme relative to noise is *sensitivity* / *fpr.* Shown on the right is the minimum *sensitivity* / *fpr* of a "good" enzyme that can be raised from low initial frequencies to 1/500 in specified numbers of selection rounds. At the target frequency 1/500, a "good" enzyme could be isolated *via* cloning. A maximum of 6 rounds of selection (first column of table) were performed in the emulsion methods of (73; 128). These "low precision" methods can only isolate enzymes with very high signal *vs* noise: 3.6 for a 1e-6, and 12.5 for a 1e-9 initial abundance (equivalent to 10 "good" enzymes in a 1e10 complexity ribosome display library). To isolate enzymes with lesser improvements in performance, or from larger libraries, requires "high precision": Either *sensitivity / fpr* must be increased by lowering *fpr* (noise reduction), or the number of cycles must be increased (automation).

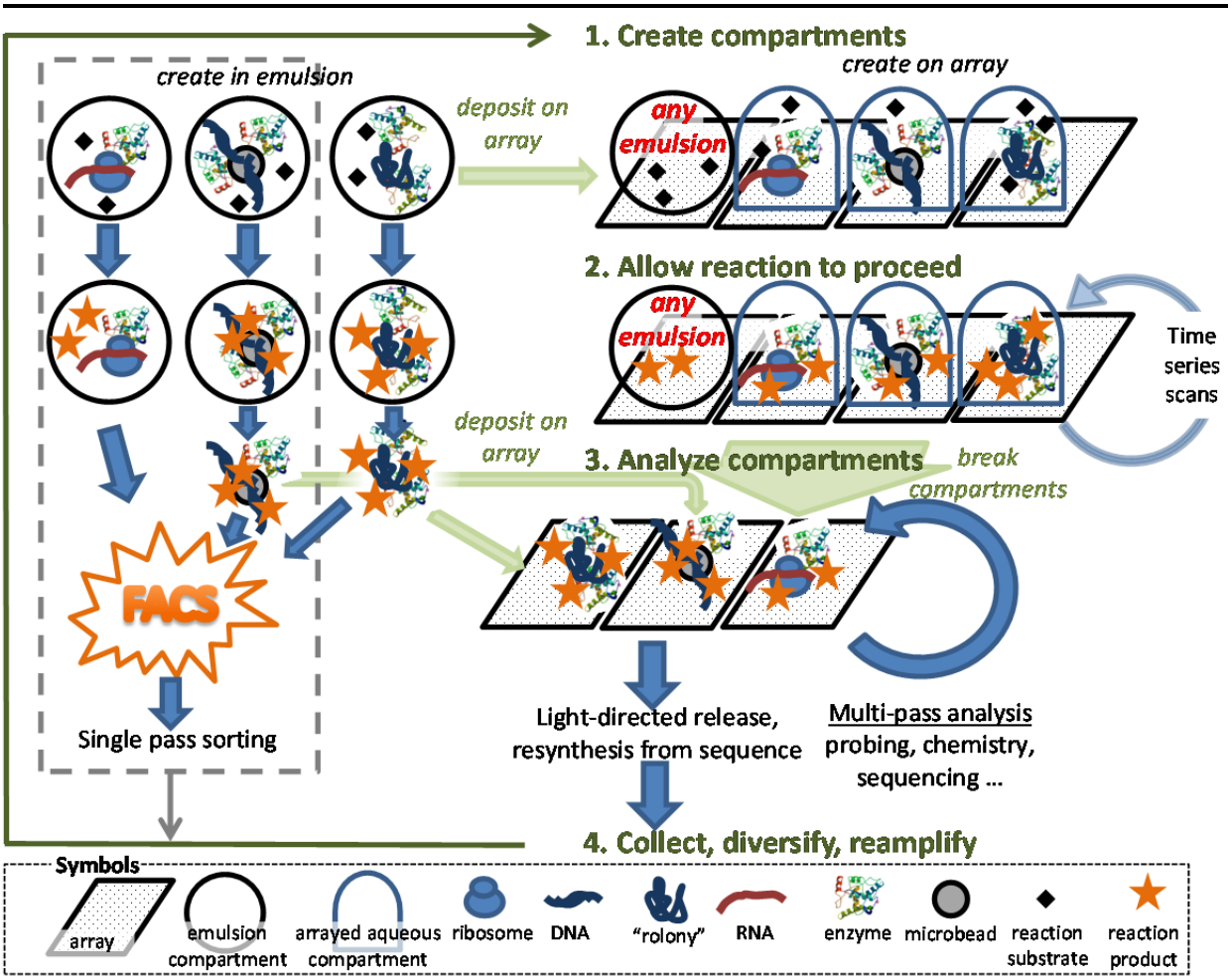


Figure 20: Range of options for *in vitro* selections in emulsions and arrayed aqueous compartments. *In vitro* selection involves four steps labeled 1-4. The dashed box at the left represents existing "low precision" emulsion methods (73; 128). The arrayed capability is shown on the right. We will use emulsion systems as testing grounds and inputs for the arrayed capacity, but we will also seek to improve these systems themselves (Aim 3.b). *Step 1:* Compartments may be created in emulsion (left) or directly on arrays (right), or created in emulsion and deposited on the array (center). Shown are the ribosome-display and microbead-display compartments from (73; 128), and a "rolony" emulsion method we will develop (see *Preliminary Results* and 3.a.1). *Step 2*: The reaction is allowed to proceed and reaction products accumulate in the isolated compartments. Conducting reactions on the array enables gathering of reaction time-course data unavailable with unarrayed emulsions. *Step 3:* After reaction termination, arrays can be analyzed in multiple passes as long as the products to be analyzed are anchored to the surface, yet for emulsion compartments processed by FACS, only a single pass analysis is possible. For emulsion compartments in which DNA, enzyme, and reaction product are captured on microbeads (73) or rolonies (3.a.1), however, the beads or rolonies can be deposited on an array, allowing multi-pass analysis of these forms of emulsion reactions. Once reaction products are captured on the array surface, the isolated aqueous compartments can be broken and the entire array processed in bulk solutions to detect and quantify enzymes, RNAs, and reaction products, and DNAs can be sequenced. This multi-pass capability will support analysis of multi- as well as single enzyme pathways. *Step 4:* Using capabilities already tested on the Polonator, DNA from high-performing compartments can be released by light-directed cleavage of labile attachment chemistry, or DNAs can be resynthesized from their sequences using the technology of Aim 2.a.

structure. We envision that these methods will be extendable to other nano-environments and nano-structures as technologies develop to create them within individual emulsion or array droplets.

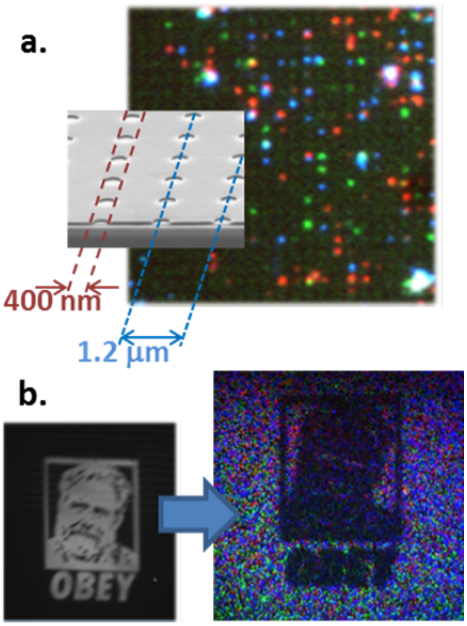


Figure 21: Recent Polonator enhancements. (a) Rolonies arrayed on a silicon surface patterned with photoresist. Colors in the background represent the base determined for each rolony from one cycle of sequencing performed over the entire array. (b) Light-directed release of DNA attached to microbeads with light-labile linkers. The beads were immobilized on a surface, and UV light focused on the array through a mask (left). Colors (right) represent the base determined for each bead after washing and a cycle of Polonator sequencing. A negative image appears where beads lost their DNA.

Preliminary Results: The Church Lab has considerable experience with emulsion methods (180; 208), IVT ((92), also see Aim 2.b.), directed evolutions (180; 213) and device development and automation (e.g., our Polonator and MAGE devices). Indeed, our Polonator (48) instrument is a generalized programmable platform for integrating cycles of microscopy, image acquisition (34), onboard computation (e.g., autofocus), and reagent handling (100; 180), all of which functions are relevant to *in vitro* selection. An upcoming release of the Polonator supports sequencing of rolling circle amplicons of DNA (which we call RCA "polonies" (100; 163; 180) or "rolonies") in place of the clonal DNA-coated microbeads in our original protocols (180) (see Figure 21a). Rolonies are a simple version of the "DNA nanoballs" used by Complete Genomics for sequencing (50). We have found that rolonies are simpler to generate, more efficient, and more cost-effective than bead preparation, and they can be arrayed easily and at higher densities than beads. The new Polonator release also supports fast "on the fly" scanning that improves sequencing time ~4-fold. Finally, we have integrated a Digital Micromirror Device with a Polonator, giving us the ability to project UV light onto a Polonator flow cell. In flow cells in which we loaded and immobilized microbeads with synthetic DNA templates containing photo-cleavable linkers, we demonstrated partial sequencing of templates, and released templates from the beads based on their sequences using UV light at single pixel resolution. We reamplified the wash buffer collected during release and verified that the amplicons were consistent with the selected sequences. Figure 21b gives an illustration of light-directed DNA release.

Aim 3.a: We will enhance *ex vivo* selection methods to accommodate new classes of enzymes, and improve compartmentalization, automation, and precision.

Aim 3.a has two parts: In (i) we extend the range of reactions that can be conducted and screened in emulsion, while in (ii) we improve emulsion methods in themselves. Each of these components integrates with other projects in our proposed Center. (i) will draw on small molecule sensors that are being developed in Aim 1.d, while the chimeragenesis project of Aim 2.a.2, in which we attempt to evolve better GH9 enzymes, will be a test bed for driving the emulsion improvements we develop in (ii).

Research Design: (i) We will devise a variety of sensors based on transcriptional, translational, or protein reporters from Aims 1.d for *unmodified* small molecule reaction products that can be included in reaction compartments. Three possibilities are illustrated in Figure 22. To develop small molecule molecular beacons from RNA aptamers, we will use existing methods (75; 115; 207) to affix FRET-pair dyes to the ends of ligand-responsive RNA aptamers explored and developed in Aim 1.d, so that ligand binding will cause conformational changes that either quench (115; 207) or dequench (75) a fluorescent signal. We will also explore additional options: (i.a) We will develop and include in the compartments

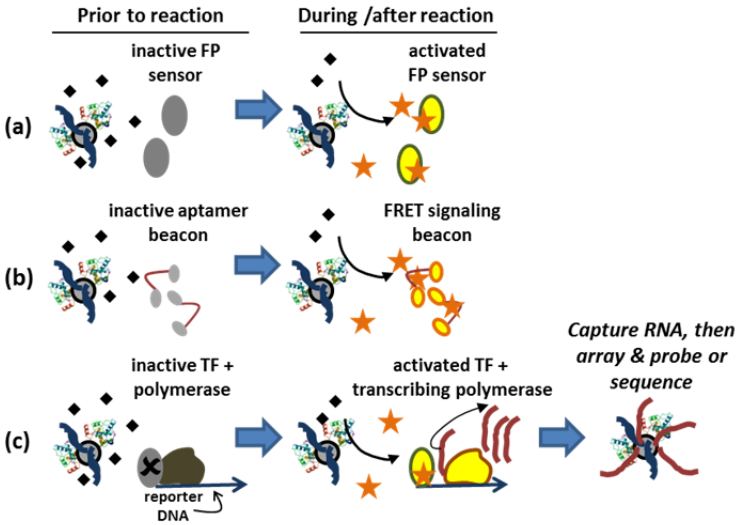


Figure 22: Sensors of unmodified reaction products. **(a)** Metabolite-activated fluorescent protein (FP) sensors developed in Aim 1.d will be included with the reaction mixture (left). Accumulating reaction product activates the FP which becomes fluorescent. **(b)** RNA aptamers (also see 1.d) that bind reaction products can be conjugated to pairs of dyes that can deliver a FRET signal in the product-bound conformation. **(3)** Engineered allosteric transcription activators responsive to the product (Aim 1.d) are included in the reaction mix along with an RNA polymerase and a DNA reporter construct to which the TF will bind. Upon reaction product formation, the TF and the polymerase become active and the reporter construct is transcribed. The RNA accumulates in the compartment where it can be detected by any of several means. As depicted here, the RNA is captured by capture primers on the microbead in the compartment and reverse transcribed to cDNA. The microbeads can then be arrayed on the Polonator and probed or sequenced. This method could be applied to multi-step reactions with several products if reporters and product-specific TFs were included. Methods **(1)** and **(2)** are applicable to emulsions as well as arrayed compartments. For (3), if compartments contain all IVT reagents and not just transcriptional apparatus, the RNA could code for a fluorescent protein *vs* just a reporter transcript.

sets of enzymes and RNA aptamer-controlled sequences (Aim 1.e) that will effect *metabolite-dependent PCR amplification* of the gene used to generate the enzyme variant. Instead of simply *measuring* per-compartment enzyme abundance, this method will effect *direct selection* of successful enzyme variants. (i.b) For some classes of reaction, we can explore adaptation of chemistries used in next-generation sequencing – e.g. pyrophosphate (PPi) detection as used pyrosequencing in 454 (125; 174) for reactions that release PPi, or local pH change as used by Ion Torrent for reactions that acidify the media. (i.c) We will explore "reactome" technology (15) as a source of chemical frameworks for attaching reactants that can generate detectable signals upon enzymatic activity. (Although the initial reactome article was retracted (16), it has been defended, notably by Richard Roberts (3).)

(ii) As an alternative to the microbead *in vitro* selection method of (73), we will explore the use of "rolonies." As noted in *Preliminary Results* above, we have found that rolonies are comparable or better than beads for use in sequencing. Moreover, rolonies represent stable DNA structures that can be further modified with various attachment chemistries (e.g. biotin- or amine-group) and so can be functionalized in the same way that beads are used in (73). Finally, as the genes of interest are available in amplified form in rolonies, they are easy to probe or sequence *in situ*. (ii) As noted in Figure 19, selections can be improved by controlling noise and by increasing cycles. (ii.a) We will significantly reduce the noise associated with the Poisson distributions of molecules per bead and beads per compartment ratio through the use of microfluidic emulsion technology. Our current setup uses Dolomite Microfluidics' pressure-based small droplet system to deliver microdroplets with high performance and precision (www.dolomite-microfluidics.com). (ii.b) We will also develop methods of quantitating protein copies in compartments. One alternative is to generate proteins fused with fluorescent proteins, or, if protein folding might be compromised by fusion, by using constructs in which the coding regions are separated by IRES sites. In our own IVT development work in Aim 2.b.1, we have successfully performed IVT reactions using plasmids that bear 4 IRES sites. For calibration purposes, the relative levels of proteins generated from coding sequences upstream *vs* downstream the IRES can be determined with separate control experiments using dual fluorescent protein constructs (e.g., GRP-IRES-

mCherry and *vice versa*). We will also investigate the use of NSAA FRET pairs in enzymes (cf. 1.a.2). We will explore the *p*-cyanophenylalanine (PheCN) / tryptophan (Trp) (69) and (PheCN/5-hydroxytryptophan (5HW) (171) FRET pairs, whose corresponding aaRS/tRNA are already available. The PheCN/Trp pair will give a starting point for the TAG knockout strain where only one NSAA can be incorporated, while the superior PheCN/5HW FRE-pair will be explored for strains allowing the incorporation of multiple NSAAs. Multi-FRET systems composed of PheCN, Trp, and 5HW will also be explored. This will improve quantitation over antibody labeling (73; 128) by delivering a signal that can be better distinguished from autofluorescence.

Timeline: (i) aptamer beacons and transcriptional reporters: Yrs 1-3; GFP sensors: Yrs 2-4; molecule-dependent PCR: Yrs 4-5. (ii) emulsion improvements, rolonies: Year 1-2; enzyme quantitation: Yrs 2-5.

Potential problems and alternatives: While control over droplet size and formation in (i) will generally reduce noise, a key question for 3.a is whether the variance associated with our proposed reporters will confound our attempts to better quantitate compartment contents. However, Figure 19 indicates that noise can be overcome by increasing selection cycles. The other confounding possibility is bias. To control bias, we will mix experimental with internal control emulsions to calibrate reporter signals and selections, and use mathematical modeling of in-droplet enzyme and reporter reactions to assess the sensitivity and dynamic ranges of our measurements.

Aim 3.b: We will develop complementary 'high precision' *ex vivo* selection methods based on arraying and analysis of aqueous compartments that will enable isolation of rarer variants exhibiting smaller relative improvements in performance *vs.* emulsion methods.

Research Design: We will explore the options in Figure 20. (i) The option most immediately implementable on the Polonator is arraying and analysis of microbeads created using the protocols of (73), or arraying of the rolonies we propose to develop in in 3.a.(ii). In the protocols of (73), derivatized microbeads displaying DNA and enzyme are created in an IVT emulsion reaction, and the enzyme reaction is conducted in a second emulsion containing the beads using caged-biotin modified substrates. The reaction products are then also captured on the beads after decaging the biotin. In 3.a.(ii) we propose to similarly capture enzymes and reaction products on rolonies. Here we propose (i.1) to array such beads or rolonies on the Polonator and analyze them with multiple scans to measure captured products and copy numbers of enzymes. We will then adjust protocols as needed for (i.2) sequencing the DNA, and for (i.3) light-directed release of DNA from beads or rolonies judged to have high-performing enzymes. For (i.2), we expect rolonies to be immediately sequenceable on the Polonator; however, the bead generation protocols of (73) will need to be adjusted because they bear only a single dsDNA molecule. Thus we will extend protocols of (73) to insert an emulsion PCR step that generates many clonal copies of DNA on the bead (180), prior to the in-emulsion IVT that captures generates the enzymes captured on the bead. For (i.3) we have already demonstrated light-directed release of DNA from beads (see *Preliminary Results*). However, we will need to adjust protocols for light-directed release of rolony DNA because rolonies comprise long single molecule DNA concatamers with numerous attachments to the surface *vs* the numerous separate DNA strands with single attachments for DNA-coated beads; thus, even though the attachments are light-labile, it may prove difficult to release them all and free an entire rolony. We will improve rolony DNA release by adjusting the relative frequency of rolony attachments, and by designing rolony concatamers to contain standard restriction sites, so that, when DNA release is to be effected, a restriction enzyme is used to cut *all* rolonies on the array into monomer fragments that remain attached, and subsequent light-direction then releases monomers of the particular sequences that are desired.

(ii) We will then explore the two approaches depicted in Figure 20 for arraying compartments in Polonator flow cells: Our preferred approach will be (ii.1) to form isolated aqueous compartments

directly on the surfaces of flow cells; however, we will also explore (ii.2) methods of depositing monolayers of emulsion droplets on these surfaces. (ii.1) *Forming arrays of droplets on surfaces:* Several approaches have been developed to generate isolated femtoliter scale reaction volumes on surfaces (158; 173; 175). We will focus on adapting the droplet arraying method of (175) in which hydrophilic regions are patterned on a hydrophobic surface; a bulk aqueous phase over the surface is then exchanged with an oil phase leaving behind an array of isolated aqueous droplets at densities of up to 1e6/cm^2. Here, using procedures we have already demonstrated on the Polonator (see *Preliminary Results*), we will first array rolonies into the hydrophilic regions, and then form the microdroplets around them using the droplet methods of (175). As an alternative to (175), we will explore the method of (173) in which a bulk droplet is partitioned into isolated microdroplets by a micropatterned PDMS overlay to densities of ~1.7e6/cm^2. Here we will need to rely on Poisson statistics to place single rolonies or microbeads into compartments, making it a less favored option. (ii.2) *Surface arraying of emulsion droplets:* (ii.2.1) Although many methods have been published for *physically* micropatterning surfaces for trapping droplets (84; 181), we will focus rather on *chemical* patterning of surfaces as this supports higher densities and does not require complex microfluidic control of droplet movements. Patterning of charge is highly effective for surface arraying of microbeads and rolonies. However, maintenance of a stable charge difference between phases in oil/water (o/w) or water/oil/water (w/o/w) emulsions may be confounded by charge redistribution, and aqueous phase charge imbalances may also affect in-compartment reaction chemistry. We will explore whether emulsions generated using anionic detergents or other means (70; 78; 126) can be stably arrayed. However, an attractive alternative is to attempt to create emulsions with fluorous surfactants (96) and then capture the compartments on fluorous patterned surfaces (83; 105), as fluorous surface to fluorous phase aggregation would not depend on charge. For promising approaches (ii.2.2), we will fabricate microwells with the appropriate chemistry *via* photolithography to see if droplet attachment or stability can be enhanced by improving surface geometrical fit. Finally, (ii.2.3) we will explore the generation and use of liposomes. Liposomes offer several advantages over o/w and w/o/w emulsions, such as higher stability and the ability to include proteins at the lipid bilayer that could interact with surface-arrayed ligands. Liposomes could also potentially be designed to dock with each other to transfer contents. Notably, IVT reactions and gene expression have been conducted in liposomes (148-150). Indeed, the emulsion IVT and enzymatic reaction of the *in vitro* selection method (128) was recapitulated in liposome emulsions and the liposomes were sorted by FACS based on reaction fluorescence signal (5) just as in (128). For all options within (ii), we will evaluate and devise methods to ensure not only stability of the droplets, but the absence of diffusion of enzymes, reaction intermediates, or products between them.

Timeline: (i) Year 1-3. (ii) Years 2-5.

Potential problems and alternatives: If we cannot resolve the technical issues of (ii) on the Polonator, another possibility is to conduct reactions with DNA and enzyme-displaying microbeads in 454-style picotiter plates. These are already designed to support isolated individual reactions (125; 174) (albeit only rapid sequencing reactions). To effect selection in this environment, robotics can be used to extract microbeads from 454 plates based on reaction performance data (130).

Aim 3.c: We will develop the ability to optimize multi-enzyme, multi-step reactions by *ex vivo* screening and selection.

Research Design: Our ultimate goal is to demonstrate *ex vivo* selection of a two-step reaction involving two enzymes E_1 and E_2, where E_1: S $\rightarrow$ I for substrate S and intermediate I and E_2: I $\rightarrow$ P for product P, and where selection is aimed at optimizing production of P. We further aim to conduct these experiments on our arrayed capacity of 3.b to make multi-pass measurements of the E_1 and E_2 reactions, and to take advantage of our molecular sensors of 3.a to measure I and P without need for modified

substrates. (S need not be measured in each droplet as it can be assumed part of the reaction mix out of which the droplets are formed and so to have a constant initial concentration; this also applies to any second substrates used by E_1 or E_2, if these are bimolecular.) However, because this would depend on having a pair of consecutive reactions with operating molecular sensors by Year ~3, which cannot be guaranteed, we formulate our plans around a reaction already used by the Griffiths Lab: Ebg (or LacZ): FDG $\rightarrow$ fluorescein + 2 galactose (128), which we can aim at coupling to a measureable follow-on reaction. If *via* 3.a, we develop a molecular sensor that allows us to measure this reaction without the modified substrate FDG, or replace this reaction by another sensor-measureable reaction, we will do so.

(i) We will first verify that we can conduct and measure the performance the Ebg reaction *and* a second *uncoupled* reaction in the same compartment. The Tawfik Lab reaction phosphotriesterase: (caged-biotin)-ethyl-paraoxon $\rightarrow$ (caged-biotin)-product (73) is a natural candidate as the methods of (73) render it measureable, but it is possible that these reactions might interfere with each other: If they do we will use a different second reaction that does not interfere with Ebg, and focus on verifying that we can measure the Ebg reaction while the second (possibly unmeasured) reaction also takes place in the compartment. (ii) We will then verify that we can effectively conduct *in vitro* selection on Ebg while it is operating in the same compartment with the second (uncoupled) reaction. The Ebg reaction is ideal for this purpose as it is defective and has been shown to be evolvable both *in vitro* and *in vivo*. (iii) The next step is to move to a second *measureable* reaction that can be *coupled* to Ebg, and attempt to measure the two reactions concurrently *with no attempt at selection*. Specifically, any reaction that *consumes galactose* and yields a detectable product can be considered. (Note that the coupled reaction need not use an *E. coli* enzyme.) If no other ATP-consuming reaction is taking place in the compartment, galactokinase could be attractive as a coupled reaction because we could measure it using an existing ADP/ATP ratio GFP sensor (18). Galactose dehydrogenase could be used if no NADH/NAD+-consuming reactions were present, and the NADH/NAD+ ratio can be measured: this could be done by adapting an existing RNA aptamer to be a riboregulator or NADH molecular beacon (see 1.d.2, 3.a.1). A third alternative would be to use galactose oxidase as a second reaction, as the H_2O_2 byproduct of the reaction can be quantitatively measured using horseradish peroxidase (136). A useful feature of these alternatives is that all three of these enzymes can be purchased commercially and included in the reaction mixture for initial experiments, and then expressed *via* IVT with Ebg in later ones. In setting up these experiments, we will ensure that dyes and fluorophores can be sensed independently, possibly by using appropriate GFP variants. A complication in using the galactose produced by Ebg is that all follow-on enzymatic reactions will use only specific anomers of galactose and so overall reaction rates will depend on mutarotation. If spontaneous mutarotation proves limiting, we will include a galactose mutarotase in the reaction mixture.

Finally, (iv) we would evolve the coupled enzymes using *in vitro* selections, starting with Ebg. For the second enzyme, we could either deliberately mutate the enzyme, or pick a homolog that operates best at a different pH or temperature. Finally, we will explore alternating or concurrent enzyme evolutions. Here we will explore different forms of sequence diversification (e.g., error-prone PCR *vs* variations confined specific sequence positions) that take into account the combinatorial capacity of our methods.

Timeline: (i) Year 2, (ii) Year 3, (iii) Year 4, (iv) Years 4-5.

Potential problems and alternatives: Our strategy above has been designed to be flexible and to consider and evaluate alternatives at every phase. One general issue is that it is possible that compartments may be found not be isolated against exchange of both reaction intermediate I and final product P. If such leakage is found, we will explore adjustments to compartment chemistry or structure to ensure containment.

Appendix 1: Biographical sketches

Biographical Sketch: George Church

Education and Training:

Duke University, Durham, NC	B.A.	1974	Zoology & Chemistry
Harvard University, Cambridge, MA	Ph.D.	1984	Biochemistry & Molecular Biology

Research and Professional Experience:

2006-present Senior Associate of Broad Institute (1990 Genome Center Co-founder)
2004-present Director of the Harvard/MIT/WashU NHGRI CEGS
2002-present Director of the Harvard/MIT DOE Genomes-to-Life Center
1998-present Professor of Genetics, Harvard Medical School, Boston, MA
1997-present Director of the Lipper Center for Computational Genetics, Boston, MA
1986-1998 Assistant/Associate Professor of Genetics, Harvard Medical School, Boston, MA
1985-1986 Research Fellow, Anatomy, Univ. Calif., San Francisco, CA
1984 Scientist, Biogen Research Corporation, Cambridge, MA

Publications: (also see: http://arep.med.harvard.edu/gmc_pub.html#patents)

1. Wang HH, Xu G, Vonner A, Church GM (2011) Modified Bases Enable High-efficiency Oligonucleotide-Mediated Allelic Replacement via Mismatch Repair Evasion. Nucleic Acids Res. doi:10.1093/nar/gkr18.
2. Isaacs FJ, Carr PA, Wang HH, Lajoie MJ,Sterling B, Kraal L, Tolonen A, Gianoulis T, Goodman D, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM(2011) Precise Manipulation of Chromosomes in vivo Enables Genome-wide Codon Replacement (Science, in press)
3. Tolonen AC, Haas W, Chilaka AC, Aach J, Gygi SP, Church GM (2010) Proteome-wide systems analysis of a cellulosic biofuel producing microbe. Nature MSB 7:461. PMID: 21245846
4. Kosuri S, Eroshenko N, LeProust E, Super M, Way J, Li JB, Church GM (2010) A Scalable Gene Synthesis Platform Using High-Fidelity DNA Microchips. Nature Biotech. 28(12):1295-9. PMID: 21113165.
5. Mosberg JA, Lajoie MJ, Church GM (2010) Lambda Red Recombination of Double-Stranded DNA in Escherichia coli Proceeds Through a Fully Single-Stranded Intermediate. Genetics 10.1534/genetics.110.120782 PMID: 20813883
6. Robertson DE, Jacobson S, Morgan F, Berry D, Church GM, Afeyan N (2010) A New Dawn for Industrial Photosynthesis. Photosyn Res. DOI 10.1007/s11120-011-9631-7.
7. Sommer MOA, Church GM, Dantas G (2010) A functional metagenomic approach for expanding the synthetic biology toolbox for biomass conversion. Molecular Systems Biology 6:360. PMID: 20393580
Tolonen A, Chilaka AC, Church GM (2009) Targeted gene inactivation in Clostridium phytofermentans shows that cellulose degradation requires the family 9 hydrolase Cphy3367. Mol Microbiol Dec;74(6):1300-13. PMID: 19775243
8. Friedland AE, Lu TK, Wang X, Shi D, Church GM, Collins J (2009) Synthetic Gene Networks that Count. Science. Jun 14; 324(5931):1199-202. PMID: 19478183 PMC2690711
9. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature. Jul 26; 460(7257):894-8. PMID: 19633652
10. Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Galagan JE, Church GM (2009) Large-Scale Identification of Genetic Design Strategies using Local Search. Nature EMBO MSB 5:296. PMID: 19690565

Synergistic Activities

1. Founder of LS9 Inc. (E.coli alkane production)

2. Founder of Joule Unlimited (Cyanobacterial alkanes)
3. Co-inventor of various Next-generation sequencing technologies.
4. Development and application of MAGE method and device.
5. Member of Wyss Inst., Broad Inst and NSF SynBERC Genomics & Synthetic Biology efforts.

Identification of Potential Conflicts of interest or Bias ini Selection of Reviewers.
Collaborators and Co-editors: 569

Graduate and Postdoctoral Advisors and Advisees:
List the names and current organizational affiliation of your graduate students and postdoctoral associates during past five years.

A partial list is provided here. A more complete list is provided in Appendix 6.

Person	Church Lab Status	Position	Affiliation
Aach, John	Current	Lecturer	Harvard Med School
Adesokan, Adeyemi	Former	CEO	Pathogenica
Ahlford, Annika	Former	Visiting fellow	Uppsala University
Azizi, Elham	Current	Rotation student	Harvard Med School
Bachelet, Ido	Current	Post doc	Harvard Med School
Bang, Duhee	Former	Assistant Professor	Yonsei University
Barrera, Luis	Current	Rotation student	Harvard Med School
Batada, Nizar	Former	Assistant Professor	U of Toronto
Bobe, Jason	Current	Consultant	Personal Genome Project
Briggs, Adrian	Current	Post doc	Harvard Med School
Brown, Chris	Former	Rotation student	Harvard Med School
Busskamp, Volker	Current	Post doc	Harvard Med School
Byrne, Susan	Current	Post doc	Harvard Med School
Cai, Long	Former	Post doc	?
Chari, Raj	Current	Post doc	Harvard Med School
Chew, Wei Leong	Current	Grad student	Harvard Med School
Chilaka, Amanda	Former	Undergraduate Intern	Northeastern University
Chou, Michael	Current	Post doc	Harvard Med School
Church, George	Current	Post doc	Harvard Med School
Cong, Le	Current	Grad student	Harvard Med School
Cox, David	Current	Rotation student	Harvard Med School
Curtis, Wayne	Former	Visiting scientist	U Penn
Dai, Mingjie	Current	Rotation student	Harvard Med School
Daniels, Rachel	Former	Rotation Student	Harvard Med School
Dantas, Gautam	Former	Assistant Professor	Wash. U. St Louis
DiCarlo, James	Current	Grad student	Harvard Med School
Complete list provided in Appendix 6			

Biographical Sketch: Farren J. Isaacs

Molecular, Cellular & Developmental Biology, Systems Biology Institute, Yale University

Education and Training

2005-2010 **Postdoctoral Training**, Harvard Medical School, Department of Genetics, Boston

2000-2003 **Ph.D.**, Bioinformatics, Boston University, College of Engineering and Graduate School of Arts and Science, Boston, MA

1997-2000 **M.S.**, Biomedical Engineering, College of Engineering, Boston, MA

1992-1996 **B.S.E.**, Bioengineering (Minors: Chemistry and Mathematics), University of Pennsylvania, School of Engineering and Applied Science, Philadelphia, PA

Research and Professional Experience

2011- **Assistant Professor**, Molecular, Cellular and Developmental Biology, Systems Biology Institute, Yale University

2010-2011 **Research Scientist**, Molecular, Cellular and Developmental Biology, Systems Biology Institute, Yale University
Established independent lab focused on developing foundational genomic and biomolecular engineering technologies to understand and engineer biological systems. The Isaacs Lab integrates engineering and evolution through the construction of genes, networks and whole genomes alongside quantitative models to gain a better understanding of whole biological systems. In turn, we utilize these insights to design and evolve organisms with new and desired function with applications directed to address global challenges in medicine, energy supply and the environment.

2005-2010 **Research Fellow**, Department of Genetics, Harvard Medical School
Led team of undergraduate & graduate students to develop genome engineering technologies for strain-pathway engineering and the construction of new genetic codes; Co-inventor of MAGE and CAGE; led team on the *rE.coli* project for the design and construction of a new genetic code in *E. coli.*

2003-2004 **Research Associate**, Department of Biomedical Engineering, Boston University

1998-2003 **Graduate Student**, Department of Biomedical Engineering, Center for BioDynamics, Center for Advanced Biotechnology, Boston University
Pioneered the design and development of synthetic RNA componenets capable of probing and programming cellular function; combined theory and experiment to study dynamic expression of gene regulatory networks.

2002-2005 **Consultant**, Systems Biodynamics Lab, Bioengineering Department, UCSD
Headed the set-up of a molecular biology lab with $500,000 budget; trained students and postdocs in experimental techniques; co-wrote NIH RO1 grant.

1990-1997 **Research Intern**, Neuro-oncology Lab, Barrow Neurological Inst, Phoenix, AZ
Studied tumor population, heterogeneity, and chromosomal abnormalities of human gliomas; headed the design and development of a PDGF-B DNA probe to be used in molecular cytogenetic technique, Fluorescent *In Situ* Hybridization.

1995-1996 **Research Assistant**, Neurosurgery Laboratory, Department of Bioengineering, University of Pennsylvania: Independent Study: “Altered Gene Expression Following *In Vitro* Mechanical Injury of Primary Neuronal Cultures.”

Publications (*=equal contributions, corresponding author)

1. Isaacs FJ*, Carr P*, Wang HH*, Lajoie M, Sterling B, Kraal L, Tolonen A, Gianoulis, TA, Goodman, DB, Reppas N, Emig C, Bang D, Hwang S, Jewett M, Jacobson J, Church G. Precise Manipulation of Chromosomes *in vivo* Enables Genome-wide Codon Replacement. *Science* (in press).

2. Callura JM, Dwyer DJ, Isaacs FJ, Cantor CR, Collins JJ (2010) Tracking, Tuning, and terminating microbial physiology using synthetic riboregulators. *Proceedings of the National Academy of Sciences USA* early addition, PMID: 20713708.
3. Wang HH*, Isaacs FJ*, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature***,** Jul 26; 460(7257):894-8.
4. Isaacs FJ, Dwyer DJ, Collins JJ (2006) RNA Synthetic Biology. *Nature Biotechnology*, 24:545-554.
5. Blake WJ, Balazsi G, Kohanski M, Isaacs FJ, Murphy K, Kuang Y, Cantor CR, Walt D, Collins JJ (2006) Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. *Molecular Cell*, 24:853-865.
6. Isaacs FJ, Collins JJ (2005) Plug-and-Play with RNA. *Nature Biotechnology*, 23:306-307 (2005).
7. Isaacs FJ, Dwyer DJ, Ding C, Pervouchine D, Cantor C and Collins JJ (2004) Engineered Riboregulators Enable Post-Transcriptional Control of Gene Expression. *Nature Biotechnology*, 22: 841-847.
8. Isaacs FJ*, Hasty J*, Cantor CR, Collins JJ (2003) Prediction and Measurement of an Autoregulatory Genetic Module. *Proceedings of the National Academy of Sciences USA*, 100:7714-7719.
9. Patent pending: G.M. Church, H. Wang, F.J. Isaacs. "Multiplex Automated Genome Engineering (MAGE)."
10. Patent pending: F.J. Isaacs, D.J. Dwyer, C.R. Cantor and J.J. Collins. "*Cis/Trans* Engineered Riboregulators."

Synergistic Activities

Yale Faculty Attendee, Society for Advancement of Chicanos and Native Americans in Science (SACNAS) 2011 National Conference, San Jose, CA (10/2011)
Faculty Mentor, Yale University int'l Genetically Engineered Machines (iGEM) team
DOE Expert Peer Review Panel: Biol and Env Research Genomic Science Program (12/2010)
Invited Participant, 7th Annual National Academies Keck Futures Initiative (NAKFI) conference, Synthetic Biology: Building on Nature's Inspiration, Irvine, CA (11/2009).
Development and application of MAGE technology and device

Identification of Potential Conflicts of Interest or Bias in Selection of Reviewers

Collaborators and Co-editors:

Bang, Duhee (Yonsei U.), Bader, Joel (Johns Hopkins), Boeke, Jef (Johns Hopkins), Callura, Jarred (Boston U.), Cantor, Charles (Boston U.), Carr, Peter (MIT), Church, George (Harvard), Collins, James (Boston U.), Dwyer, Daniel (Boston U.), Emig, Chris (Stanford), Forest, Craig (Georgia Tech), Gianoulis, Tara (Harvard), Goodman, Dan (Harvard), Hao, Haiping (Johns Hopkins), Hwang, Sam (MIT), Jacobson, Joseph (MIT), Jewett, Michael (Northwestern), Kraal, Laurens (UCSF), Lajoie, Marc (Harvard), Ostermeier, Marc (Johns Hopkins), Reppas, Nikos (Joule Unlimited), Söll, Dieter (Yale), Sterling, Bram (MIT), Sun, Zachary (Caltech), Tolonen, Andrew (Harvard), Tu, Benjamin (UTSW), Wang, Harris (Harvard), Xu, George (Harvard)

Graduate and Postdoctoral Advisors and Advisees:

Graduate Advisors: James Collins (Boston University) & Charles Cantor (BU & Sequenom)
Postdoctoral Advisor: George Church (Harvard Medical School)

Graduate Students:

Edward Barbieri, Ryan Gallagher, Adrian Haimovich, Alexis Rovner (Yale University)

Appendix 2: Current and Pending Support

Current and Pending Support: George Church

ACTIVE:

DE-FG02-02ER63445 (GTL) 12/1/2007 – 11/30/2011 3.72 calendar
DOE-GTL $8,933,939 total/award
PI: George Church
Title: Microbial Ecology, Proteogenomics & Computational Optima.
Project studies proteomics and cell models for E Coli.

1P50 HG005550 (CEGS)
NIH- NHGRI 9/13/10-7/31/15 4.32 calendar
PI: George Church $19,585,426 total/award
Title: "Center for Causal Transcriptional Consequences of Human Genetic Variation"
Role: The Center for Transcriptional Consequences of Human Genetic Variation (CTCHGV) will develop innovative and powerful genetic engineering methods and use them to identify genetic variations that causally control gene transcription levels.

SA5283-11210 (NSF) 7/1/2006 – 6/30/2014 0.36 calendar
NSF-(SynBERC) Sub: Church: $1,234,153 total/award
PI: Jay Keasling (UC Berkeley)
Title: Synthetic Biology Engineering Research Center
Project role is to develop synthetic bacterial genome “chasses” for safety, BIOFAB and mammalian systems

RO1 HL 094963 (NHLBI) 9/30/2008-6/30/2011 (NCE) as needed
NIH - NHLBI $4,017,742 total/award
PI: George Church
Title: Targeted 2nd generation sequencing in phenotyped Framingham & PGP populations.

AG-SS-2084-08 (Ellison) 10/01/08 - 9/30/12 0.36 calendar
The Ellison Medical Foundation $1,003,100 total/award
PI: George Church
Title: Establishment and Functional Characterization of a Large DNA Fragment Resource from the Long-lived Naked Mole-Rat for Comparison with Mice.
Project: The identification and characterization of naked mole-rat genes that contributed to the evolution of a long lifespan in this species.

RC2 HG005592 (NHGRI) 9/30/09-7/31/11 0.36 calendar
NIH-NHGRI - Halcyon $2,453,703 total/award
PI: George Church
Title: Development of Electron Microscopy-based Nucleic Acid Polymer Sequencing
Project: We aim to provide a comprehensive foundation for development of an ultra-low-cost, ultra-fast nucleic acid polymer sequencing technology based on single-atom resolution transmission electron microscopy (TEM) of heavy atom-labeled nucleic acid polymers.

RC2 HL102815 (NHLBI) 9/30/09-8/31/11 0.12 calendar
NIH- NHLBI Sub: Church: $261,360 total/award
PI: George Daley (CHB)

Title: Comparative phenotypic, functional, and molecular analysis of ESC and iPSC

ONRBAA09-001 (ONR) 4/1/10- 3/31/13 0.12 calendar
Office of Naval Research $488,998 total/award
PI: George Church
Title: Multiplexed Pathway and Organism Engineering.

RC1 HG005482 (NCRR) 9/22/09-6/30/11 as needed
NIH - NCRR Sub: Church: $43,656 total/award
PI: Peter Park
Title: Statistical Methods for Estimation of Copy Number from Next – Generation Sequencing

DOE- DE-AR0000079 (ARPA-E) 7/1/10-06/30/13 0.36 calendar
ARPA-E Sub: Church: $73,264 total/award
PI: Pamela Silver
Title: Engineering a Bacterial Reverse Fuel Cell

CBET1033397 (NSF) 1/1/11-1/31/13 0.12 calendar
PI: Ryan Gill (U. Colorado) Sub: Church: $261,045 total/award
Title: A new approach for directed genome engineering

DARPA 11-23-CCM-DT-FP-006 6/1/11-5/31/15 0.12 calendar
PI: Jim Collins (BU) Sub: Church: $1,000,000 total/award
Title: Synthetic Mammalian Gene Regulatory Circuits for In Vivo Biomedical Applications

ONR-MURI 6/1/11-5/31/16 0.12 calendar
PI: Jim Collins (BU) Sub: Church/Silver: $1,350,000 total/award
Title: Utilizing Synthetic Biology to Create Programmable Micro- Bio- Robots

<u>Pending:</u>
NIH 8/1/11-7/31/16 0.12 calendar
PI: Ryan Gill (U Colorado) Sub: Church: $847,509 total/award
Title: Development of a multiplex recombineering based technology platform

DARPA 11/1/11-10/31/15 0.36 calendar
PI: Don Ingber (Wyss) Co-I: Church:
Title: Sepsis-on-a-Chip Sepsis Therapeutics with Continuous Pathogen Detection

U19AI 089992-01 7/15/11-7/14/12 0.36 calendar
NIH- NIAID Sub: Church: $369,581 total/award
PI: David Hafler (Yale)
Title: Immune Sequencing profiling of the B cell antibody repertoire in response to influenza vaccination

DARPA 9/01/11-2/02/14 0.96 calendar
PI: John Reif (Duke University) Sub: Church: $1,439,999 total/award
Title: Crypto-Secure Attribution of Microorganisms via SNV Signatures

CEGS 4/1/12-3/31/17 0.48 calendar
NIH- NHGRI Sub: Church: $1,384,742 total/award
PI: Deirdre R. Meldrum (ASU)

Title: Center for Biosignatures Discovery of Exomic Variants in Human Disease

Current and Pending Support: Farren Isaacs (Yale University)

PENDING :

Agency:	Johns Hopkins Univ. subcontract (Prime: DARPA)
Total Award Amount:	$3,112,442.00
Award Dates:	01-Jul-11 - 31-Dec-13
Project Title:	CLIO Gene Guard Thrust
Person-Months Per Year Committed to the Project:	Cal: 0 Acad: 0 Sumr: 2

Agency:	Bill and Melinda Gates Foundation
Total Award Amount:	$100,000.00
Award Dates:	01-Nov-11 - 31-Oct-12
Project Title:	Bio-production of Mosquito Repellant via Genome Engineering
Person-Months Per Year Committed to the Project:	Cal: 0 Acad: 0 Sumr: 1

Agency:	Harvard Medical School subcontract (Prime: DOE)
Total Award Amount:	$750,000.00
Award Dates:	01-Dec-11 - 30-Nov-16
Project Title:	Microbial Ecology, Proteogenomics and Computational Optima (this proposal)
Person-Months Per Year Committed to the Project:	Cal: 0 Acad: 0 Sumr: 1

Appendix 3: Bibliography and references cited

1. Aharoni A, Amitai G, Bernath K, Magdassi S, Tawfik DS. 2005. High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chem Biol* 12:1281-9
2. Akselband Y, Cabral C, Castor TP, Chikarmane HM, McGrath P. 2006. Enrichment of slow-growing marine microorganisms from mixed cultures using gel microdrop (GMD) growth assay and fluorescence-activated cell sorting. *J Exper Marine Biol Ecol* 329:196-205
3. Akst J. 2010. Q&A: Why the reactome is real (http://www.the-scientist.com/blog/display/57615/).
4. Alper H, Stephanopoulos G. 2007. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng* 9:258-67
5. Amidi M, de Raad M, de Graauw H, van Ditmarsch D, Hennink WE, Crommelin DJ, Mastrobattista E. 2010. Optimization and quantification of protein synthesis inside liposomes. *J Liposome Res* 20:73-83
6. Atsumi S, Higashide W, Liao JC. 2009. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat Biotechnol* 27:1177-80
7. Baldwin WW, Kubitschek HE. 1984. Buoyant density variation during the cell cycle of Saccharomyces cerevisiae. *J Bacteriol* 158:701-4, PMCID: 215486
8. Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Walton JD. 2010. Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. *Biotechnol Biofuels* 3:22, PMCID: 2964541
9. Bar-Even A, Noor E, Lewis NE, Milo R. 2010. Design and analysis of synthetic carbon fixation pathways. *Proc Natl Acad Sci U S A* 107:8889-94, PMCID: 2889323
10. Barrett OP, Chin JW. 2010. Evolved orthogonal ribosome purification for in vitro characterization. *Nucleic Acids Res* 38:2682-91, PMCID: 2860124
11. Bartels E, Wassermann NH, Erlanger BF. 1971. Photochromic activators of the acetylcholine receptor. *Proc Natl Acad Sci U S A* 68:1820-3, PMCID: 389300
12. Barten R, Lill H. 1995. DNA-uptake in the naturally competent cyanobacterium, Synechocystis sp. PCC 6803. *FEMS Microbiology Letters* 129:83-7
13. Bayer TS, Smolke CD. 2005. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* 23:337-43
14. Bayer TS, Widmaier DM, Temme K, Mirsky EA, Santi DV, Voigt CA. 2009. Synthesis of methyl halides from biomass using engineered microbes. *J Am Chem Soc* 131:6508-15
15. Beloqui A, Guazzaroni ME, Pazos F, Vieites JM, Godoy M, Golyshina OV, Chernikova TN, Waliczek A, Silva-Rocha R, Al-Ramahi Y, La Cono V, Mendez C, Salas JA, Solano R, Yakimov MM, Timmis KN, Golyshin PN, Ferrer M. 2009. Reactome array: forging a link between metabolome and genome. *Science* 326:252-7
16. Beloqui A, Guazzaroni ME, Pazos F, Vieites JM, Godoy M, Golyshina OV, Chernikova TN, Waliczek A, Silva-Rocha R, Al-Ramahi Y, La Cono V, Mendez C, Salas JA, Solano R, Yakimov MM, Timmis KN, Golyshin PN, Ferrer M. 2010. Retraction. *Science* 330:912
17. Benner R, Kaiser K. 2003. Abundance of amino sugars and peptidoglycan in marine particulate and dissolved organic matter. *Limnol. Oceanogr.* 48:118-28
18. Berg J, Hung YP, Yellen G. 2009. A genetically encoded fluorescent reporter of ATP:ADP ratio. *Nat Methods* 6:161-6, PMCID: 2633436
19. Bhaya D, Takahashi A, Shahi P, Grossman AR. 2001. Novel motility mutants of synechocystis strain PCC 6803 generated by in vitro transposon mutagenesis. *J Bacteriol* 183:6140-3, PMCID:

99694
20. Breaker RR. 2010. Riboswitches and the RNA World. *Cold Spring Harb Perspect Biol*
21. Brenner K, You L, Arnold FH. 2008. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol* 26:483-9
22. Bricker TM, Zhang S, Laborde SM, Mayer PR, 3rd, Frankel LK, Moroney JV. 2004. The malic enzyme is required for optimal photoautotrophic growth of Synechocystis sp. strain PCC 6803 under continuous light but not under a diurnal light regimen. *J Bacteriol* 186:8144-8, PMCID: 529084
23. Burgard AP, Maranas CD. 2003. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol Bioeng* 82:670-7
24. Burgard AP, Pharkya P, Maranas CD. 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 84:647-57
25. Burkovski A, Kramer R. 2002. Bacterial amino acid transport proteins: occurrence, functions, and significance for biotechnological applications. *Appl Microbiol Biotechnol* 58:265-74
26. Buskirk AR, Landrigan A, Liu DR. 2004. Engineering a ligand-dependent RNA transcriptional activator. *Chem Biol* 11:1157-63
27. Callura JM, Dwyer DJ, Isaacs FJ, Cantor CR, Collins JJ. 2010. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc Natl Acad Sci U S A* 107:15898-903, PMCID: 2936621
28. Carr PA, Wang HH, Sterling B, Isaacs FJ, Xu G, Church GM, Jacobson JM. 2011. Enhanced Multiplex Genome Engineering through Oligonucleotide Co-selection (in revision).
29. Cascales E, Buchanan SK, Duche D, Kleanthous C, Lloubes R, Postle K, Riley M, Slatin S, Cavard D. 2007. Colicin biology. *Microbiol Mol Biol Rev* 71:158-229, PMCID: 1847374
30. Chang P, Marians KJ. 2000. Identification of a region of Escherichia coli DnaB required for functional interaction with DnaG at the replication fork. *J Biol Chem* 275:26187-95
31. Chen GT, Inouye M. 1994. Role of the AGA/AGG codons, the rarest codons in global gene expression in Escherichia coli. *Genes Dev* 8:2641-52
32. Cho H, Daniel T, Buechler YJ, Litzinger DC, Maio Z, Putnam AM, Kraynov VS, Sim BC, Bussell S, Javahishvili T, Kaphle S, Viramontes G, Ong M, Chu S, Gc B, Lieu R, Knudsen N, Castiglioni P, Norman TC, Axelrod DW, Hoffman AR, Schultz PG, Dimarchi RD, Kimmel BE. 2011. Optimized clinical performance of growth hormone with an expanded genetic code. *Proc Natl Acad Sci U S A* 108:9060-5, PMCID: 3107295
33. Church GM. 2011. *George M. Church's Tech Transfer, Advisory Roles, and Funding Sources (http://arep.med.harvard.edu/gmc/tech.html)*.
34. Church GM, Porreca GJ, Terry RC, Lares M. 2008. High-speed imaging for DNA sequencing (http://www.photonics.com/Article.aspx?AID=33989). *Biophotonics*
35. Cochella L, Green R. 2004. Isolation of antibiotic resistance mutations in the rRNA by using an in vitro selection system. *Proc Natl Acad Sci U S A* 101:3786-91, PMCID: 374322
36. Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S. 2008. Virus attenuation by genome-scale changes in codon pair bias. *Science* 320:1784-7, PMCID: 2754401
37. Court DL, Sawitzke JA, Thomason LC. 2002. Genetic engineering using homologous recombination. *Annu Rev Genet* 36:361-88
38. Czechowska K, Johnson DR, van der Meer JR. 2008. Use of flow cytometric methods for single-cell analysis in environmental microbiology. *Curr Opin Microbiol* 11:205-12
39. Datta S, Costantino N, Zhou X, Court DL. 2008. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc Natl Acad Sci U S A* 105:1626-31, PMCID: 2234195

40. Deb K. 2001. *Multi-Objective Optimization using Evolutionary Algorithms*. West Sussex, England: John Wiley & Sons Ltd.
41. Dedkova LM, Fahmi NE, Golovine SY, Hecht SM. 2003. Enhanced D-amino acid incorporation into protein by modified ribosomes. *J Am Chem Soc* 125:6616-7
42. Deiters A, Groff D, Ryu Y, Xie J, Schultz PG. 2006. A genetically encoded photocaged tyrosine. *Angew Chem Int Ed Engl* 45:2728-31
43. Denef VJ, Mueller RS, Banfield JF. 2010. AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *ISME J* 4:599-610
44. Deng MD, Coleman JR. 1999. Ethanol synthesis by genetic engineering in cyanobacteria. *Appl Environ Microbiol* 65:523-8, PMCID: 91056
45. Department of Energy. 2004. Top Value Added Chemicals From Biomass: Volume 1, Results of Screening for Potential Candidates from Sugars and Synthesis Gas (http://www1.eere.energy.gov/biomass/pdfs/35523.pdf) NREL and PNNL
46. DeVito JA. 2008. Recombineering with tolC as a selectable/counter-selectable marker: remodeling the rRNA operons of Escherichia coli. *Nucleic Acids Res* 36:e4, PMCID: 2248734
47. Dexter J, Fu P. 2009. Metabolic engineering of cyanobacteria for ethanol production. *Energy & Environmental Science* 2:857-64
48. Dover Systems. 2008. *The Polonator (http://www.polonator.org/)*.
49. Dreier B, Pluckthun A. 2011. Ribosome display: a technology for selecting and evolving proteins from large libraries. *Methods Mol Biol* 687:283-306
50. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, Dahl F, Fernandez A, Staker B, Pant KP, Baccash J, Borcherding AP, Brownley A, Cedeno R, Chen L, Chernikoff D, Cheung A, Chirita R, Curson B, Ebert JC, Hacker CR, Hartlage R, Hauser B, Huang S, Jiang Y, Karpinchyk V, Koenig M, Kong C, Landers T, Le C, Liu J, McBride CE, Morenzoni M, Morey RE, Mutch K, Perazich H, Perry K, Peters BA, Peterson J, Pethiyagoda CL, Pothuraju K, Richter C, Rosenbaum AM, Roy S, Shafto J, Sharanhovich U, Shannon KW, Sheppy CG, Sun M, Thakuria JV, Tran A, Vu D, Zaranek AW, Wu X, Drmanac S, Oliphant AR, Banyai WC, Martin B, Ballinger DG, Church GM, Reid CA. 2010. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327:78-81
51. Dublin M. 2008. *The Bioengineer and the RNA Switch*. http://www.genomeweb.com/bioengineer-and-rna-switch
52. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KL, Keasling JD. 2009. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* 27:753-9
53. Dykhuizen D. 1978. Selection for Tryptophan Auxotrophs of Escherichia coli in Glucose-Limited Chemostats as a Test of the Energy Conservation Hypothesis of Evolution. *Evolution* 32:125-50
54. Edmonds P, Cooney JJ. 1967. Identification of microorganisms isolated from jet fuel systems. *Appl Microbiol* 15:411-6, PMCID: 546914
55. El-Zahab B, Donnelly D, Wang P. 2008. Particle-tethered NADH for production of methanol from CO(2) catalyzed by coimmobilized enzymes. *Biotechnol Bioeng* 99:508-14
56. Ellington AD, Szostak JW. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818-22
57. Emptage M, Haynie SL, Laffend LA, Pucci JP, Whited G. 2003. *U.S.A.*
58. Eun YJ, Utada AS, Copeland MF, Takeuchi S, Weibel DB. 2011. Encapsulating bacteria in agarose microparticles using microfluidics for high-throughput cell analysis and isolation. *ACS Chem Biol* 6:260-6, PMCID: 3060957
59. Feng Y, Cronan JE. 2010. Overlapping repressor binding sites result in additive regulation of Escherichia coli FadH by FadR and ArcA. *J Bacteriol* 192:4289-99, PMCID: 2937390

60. Forster AC, Church GM. 2006. Towards synthesis of a minimal cell. *Mol Syst Biol* 2:45, PMCID: 1681520
61. Forster AC, Church GM. 2007. Synthetic biology projects in vitro. *Genome Res* 17:1-6
62. Frigaard NU, Sakuragi Y, Bryant DA. 2004. Gene inactivation in the cyanobacterium Synechococcus sp. PCC 7002 and the green sulfur bacterium Chlorobium tepidum using in vitro-made DNA constructs and natural transformation. *Methods Mol Biol* 274:325-40
63. Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, Bhattacharya A, Kapatral V, D'Souza M, Baev MV, Grechkin Y, Mseeh F, Fonstein MY, Overbeek R, Barabasi AL, Oltvai ZN, Osterman AL. 2003. Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J Bacteriol* 185:5673-84, PMCID: 193955
64. Gianoulis TA, Raes J, Patel PV, Bjornson R, Korbel JO, Letunic I, Yamada T, Paccanaro A, Jensen LJ, Snyder M, Bork P, Gerstein MB. 2009. Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proc Natl Acad Sci U S A* 106:1374-9, PMCID: 2629784
65. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA, 3rd, Smith HO. 2008. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* 319:1215-20
66. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, 3rd, Smith HO, Venter JC. 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329:52-6
67. Gill RT, Katsoulakis E, Schmitt W, Taroncher-Oldenburg G, Misra J, Stephanopoulos G. 2002. Genome-wide dynamic transcriptional profiling of the light-to-dark transition in Synechocystis sp. strain PCC 6803. *J Bacteriol* 184:3671-81, PMCID: 135141
68. Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, 3rd, Smith HO, Venter JC. 2006. Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A* 103:425-30, PMCID: 1324956
69. Glasscock JM, Zhu Y, Chowdhury P, Tang J, Gai F. 2008. Using an amino acid fluorescence resonance energy transfer pair to probe protein unfolding: application to the villin headpiece subdomain and the LysM domain. *Biochemistry* 47:11070-6
70. Goldstein D, Sader O, Benita S. 2007. Influence of oil droplet surface charge on the performance of antibody--emulsion conjugates. *Biomed Pharmacother* 61:97-103
71. Gore J, Youk H, van Oudenaarden A. 2009. Snowdrift game dynamics and facultative cheating in yeast. *Nature* 459:253-6, PMCID: 2888597
72. Graumann K, Premstaller A. 2006. Manufacturing of recombinant therapeutic proteins in microbial systems. *Biotechnol J* 1:164-86
73. Griffiths AD, Tawfik DS. 2003. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO J* 22:24-35, PMCID: 140064
74. Griffiths AD, Tawfik DS. 2006. Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol* 24:395-402
75. Hamaguchi N, Ellington A, Stanton M. 2001. Aptamer beacons for the direct detection of proteins. *Anal Biochem* 294:126-31
76. Hambourger M, Moore GF, Kramer DM, Gust D, Moore AL, Moore TA. 2009. Biology and technology for photochemical fuel production. *Chem Soc Rev* 38:25-35
77. Harcombe W. 2010. Novel cooperation experimentally evolved between species. *Evolution* 64:2166-72

78. Harnsilawat T, Pongsawatmanit R, McClements DJ. 2006. Influence of pH and ionic strength on formation and stability of emulsions containing oil droplets coated by beta-lactoglobulin-alginate interfaces. *Biomacromolecules* 7:2052-8
79. Head IM, Jones DM, Roling WF. 2006. Marine microorganisms make a meal of oil. *Nat Rev Microbiol* 4:173-82
80. Heinzelman P, Komor R, Kanaan A, Romero P, Yu X, Mohler S, Snow C, Arnold F. 2010. Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination. *Protein Eng Des Sel* 23:871-80
81. Helbig K, Bleuel C, Krauss GJ, Nies DH. 2008. Glutathione and transition-metal homeostasis in Escherichia coli. *J Bacteriol* 190:5431-8, PMCID: 2493246
82. Heppell B, Blouin S, Dussault AM, Mulhbacher J, Ennifar E, Penedo JC, Lafontaine DA. 2011. Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nat Chem Biol* 7:384-92
83. Holt DJ, Payne RJ, Chow WY, Abell C. 2010. Fluorosurfactants for microdroplets: interfacial tension analysis. *J Colloid Interface Sci* 350:205-11
84. Huebner A, Bratton D, Whyte G, Yang M, Demello AJ, Abell C, Hollfelder F. 2009. Static microdroplet arrays: a microfluidic device for droplet trapping, incubation and release for enzymatic and cell-based assays. *Lab Chip* 9:692-8
85. Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen A, Gianoulis T, Goodman D, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM. 2011. Precise Manipulation of Chromosomes in vivo Enables Genome-wide Codon Replacement (in revision). *Science*
86. Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ. 2004. Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* 22:841-7
87. Isaacs FJC, P.A., Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen A, Gianoulis T, Goodman D, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM. 2011. Precise Manipulation of Chromosomes in vivo Enables Genome-wide Codon Replacement (in print). *Science*
88. Itaya M, Tsuge K, Koizumi M, Fujita K. 2005. Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. *Proc Natl Acad Sci U S A* 102:15971-6, PMCID: 1276048
89. Iwai M, Katoh H, Katayama M, Ikeuchi M. 2004. Improved genetic transformation of the thermophilic cyanobacterium, Thermosynechococcus elongatus BP-1. *Plant Cell Physiol* 45:171-5
90. Jamsai D, Orford M, Nefedov M, Fucharoen S, Williamson R, Ioannou PA. 2003. Targeted modification of a human beta-globin locus BAC clone using GET Recombination and an I-Scei counterselection cassette. *Genomics* 82:68-77
91. Jansson C, Northen T. 2010. Calcifying cyanobacteria--the potential of biomineralization for carbon capture and storage. *Curr Opin Biotechnol* 21:365-71
92. Jewett MC, Church GM. 2011. *In vitro* integration of ribosomal RNA synthesis, ribosome self-assembly and protein synthesis. (in revision). *Nature*
93. Jia H, Zhu G, Wang P. 2003. Catalytic behaviors of enzymes attached to nanoparticles: the effect of particle mobility. *Biotechnol Bioeng* 84:406-14
94. Jiao N, Herndl GJ, Hansell DA, Benner R, Kattner G, Wilhelm SW, Kirchman DL, Weinbauer MG, Luo T, Chen F, Azam F. 2010. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat Rev Microbiol* 8:593-9
95. Kaiser K, Benner R. 2008. Major bacterial contribution to the ocean reservoir of detrital organic carbon and nitrogen. *Limnol. Oceanogr.* 53:99-112
96. Kandadai MA, Mohan P, Lin G, Butterfield A, Skliar M, Magda JJ. 2010. Comparison of

surfactants used to prepare aqueous perfluoropentane emulsions for pharmaceutical applications. *Langmuir* 26:4655-60, PMCID: 2866627

97. Karatan E, Han Z, Kay B. 2006. *Molecular Display Technologies*. Wiley-VCH Verlag GmbH & Co. KGaA
98. Kaspar S, Bott M. 2002. The sensor kinase CitA (DpiB) of Escherichia coli functions as a high-affinity citrate receptor. *Arch Microbiol* 177:313-21
99. Kim J, Grate JW, Wang P. 2008. Nanobiocatalysis and its potential applications. *Trends Biotechnol* 26:639-46
100. Kim JB, Porreca GJ, Song L, Greenway SC, Gorham JM, Church GM, Seidman CE, Seidman JG. 2007. Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. *Science* 316:1481-4
101. Kleinsteuber S, Riis V, Fetzer I, Harms H, Muller S. 2006. Population dynamics within a microbial consortium during growth on diesel fuel in saline environments. *Appl Environ Microbiol* 72:3531-42, PMCID: 1472369
102. Koksharova OA, Wolk CP. 2002. Genetic tools for cyanobacteria. *Appl Microbiol Biotechnol* 58:123-37
103. Kolisnychenko V, Plunkett G, 3rd, Herring CD, Feher T, Posfai J, Blattner FR, Posfai G. 2002. Engineering a reduced Escherichia coli genome. *Genome Res* 12:640-7, PMCID: 187512
104. Kosuri S, Eroshenko N, Leproust EM, Super M, Way J, Li JB, Church GM. 2010. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat Biotechnol* 28:1295-9
105. Krafft MP, Riess JG. 2007. Perfluorocarbons: Life sciences and biomedical uses Dedicated to the memory of Professor Guy Ourisson, a true RENAISSANCE man. *Journal of Polymer Science Part A: Polymer Chemistry* 45:1185-98
106. Krakauer DC, Jansen VA. 2002. Red queen dynamics of protein translation. *J Theor Biol* 218:97-109
107. Krämer R. 1994. Secretion of amino acids by bacteria: physiology and mechanism. *FEMS Microbiol Rev* 13:75-94
108. Kubitschek HE, Baldwin WW, Graetzer R. 1983. Buoyant density constancy during the cell cycle of Escherichia coli. *J Bacteriol* 155:1027-32, PMCID: 217795
109. Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in Escherichia coli. *Science* 324:255-8
110. Laine S, Thouard A, Komar AA, Rossignol JM. 2008. Ribosome can resume the translation in both +1 or -1 frames after encountering an AGA cluster in Escherichia coli. *Gene* 412:95-101
111. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK. 2009. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res* 19:2308-16, PMCID: 2792183
112. Lee JF, Hesselberth JR, Meyers LA, Ellington AD. 2004. Aptamer database. *Nucleic Acids Res* 32:D95-100, PMCID: 308828
113. Lee JH, Sung BH, Kim MS, Blattner FR, Yoon BH, Kim JH, Kim SC. 2009. Metabolic engineering of a reduced-genome strain of Escherichia coli for L-threonine production. *Microb Cell Fact* 8:2, PMCID: 2634754
114. Li JB, Gao Y, Aach J, Zhang K, Kryukov GV, Xie B, Ahlford A, Yoon JK, Rosenbaum AM, Zaranek AW, LeProust E, Sunyaev SR, Church GM. 2009. Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. *Genome Res* 19:1606-15, PMCID: 2752131
115. Li JJ, Fang X, Tan W. 2002. Molecular aptamer beacons for real-time protein recognition. *Biochem Biophys Res Commun* 292:31-40
116. Li X, Gianoulis TA, Yip KY, Gerstein M, Snyder M. 2010. Extensive in vivo metabolite-protein

interactions revealed by large-scale systematic analyses. *Cell* 143:639-50, PMCID: 3005334

117. Lindberg P, Park S, Melis A. 2010. Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism. *Metab Eng* 12:70-9
118. Lipovsek D, Pluckthun A. 2004. In-vitro protein evolution by ribosome display and mRNA display. *J Immunol Methods* 290:51-67
119. Liu CC, Qi L, Yanofsky C, Arkin AP. 2011. Regulation of transcription by unnatural amino acids. *Nat Biotechnol* 29:164-8
120. Liu CC, Schultz PG. 2010. Adding new chemistries to the genetic code. *Annu Rev Biochem* 79:413-44
121. Lopilato JE, Garwin JL, Emr SD, Silhavy TJ, Beckwith JR. 1984. D-ribose metabolism in Escherichia coli K-12: genetics, regulation, and transport. *J Bacteriol* 158:665-73, PMCID: 215481
122. Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, Galagan JE, Church GM. 2009. Large-scale identification of genetic design strategies using local search. *Mol Syst Biol* 5:296, PMCID: 2736654
123. Lykidis A, Chen CL, Tringe SG, McHardy AC, Copeland A, Kyrpides NC, Hugenholtz P, Macarie H, Olmos A, Monroy O, Liu WT. 2011. Multiple syntrophic interactions in a terephthalate-degrading methanogenic consortium. *ISME J* 5:122-30
124. Mannironi C, Scerch C, Fruscoloni P, Tocchini-Valentini GP. 2000. Molecular recognition of amino acids by RNA aptamers: the evolution into an L-tyrosine binder of a dopamine-binding RNA motif. *RNA* 6:520-7, PMCID: 1369933
125. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-80, PMCID: 1464427
126. Marinova KG, Alargova RG, Denkov ND, Velev OD, Petsev DN, Ivanov IB, Borwankar RP. 1996. Charging of Oil−Water Interfaces Due to Spontaneous Adsorption of Hydroxyl Ions. *Langmuir* 12:2045-51
127. Marraccini P, Bulteau S, Cassier-Chauvat C, Mermet-Bouvier P, Chauvat F. 1993. A conjugative plasmid vector for promoter analysis in several cyanobacteria of the genera Synechococcus and Synechocystis. *Plant Mol Biol* 23:905-9
128. Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT, Griffiths AD. 2005. High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chem Biol* 12:1291-300
129. Maturin L, Sr., Curtiss R, 3rd. 1977. Degradation of DNA by nucleases in intestinal tract of rats. *Science* 196:216-8
130. Matzas M, Stahler PF, Kefer N, Siebelt N, Boisguerin V, Leonard JT, Keller A, Stahler CF, Haberle P, Gharizadeh B, Babrzadeh F, Church GM. 2010. High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing. *Nat Biotechnol* 28:1291-4
131. McNeely K, Xu Y, Bennette N, Bryant DA, Dismukes GC. 2010. Redirecting reductant flux into hydrogen production via metabolic engineering of fermentative carbon metabolism in a cyanobacterium. *Appl Environ Microbiol* 76:5032-8, PMCID: 2916493
132. MIT. 2010. Registry of Standard Biological Parts (http://partsregistry.org/Main_Page).
133. Mitra RD, Butty VL, Shendure J, Williams BR, Housman DE, Church GM. 2003. Digital genotyping

and haplotyping with polymerase colonies. *Proc Natl Acad Sci U S A* 100:5926-31, PMCID: 156303

134. Mitra RD, Church GM. 1999. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res* 27:e34, PMCID: 148757
135. Mitra RD, Shendure J, Olejnik J, Edyta Krzymanska O, Church GM. 2003. Fluorescent in situ sequencing on polymerase colonies. *Anal Biochem* 320:55-65
136. Mohanty JG, Jaffe JS, Schulman ES, Raible DG. 1997. A highly sensitive fluorescent micro-assay of H2O2 release from activated human leukocytes using a dihydroxyphenoxazine derivative. *J Immunol Methods* 202:133-41
137. Mosberg JA, Lajoie MJ, Church GM. 2010. Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. *Genetics* 186:791-9, PMCID: 2975298
138. MOSIS. 2011. *The MOSIS Service (http://www.mosis.com/)*.
139. Mukai T, Hayashi A, Iraha F, Sato A, Ohtake K, Yokoyama S, Sakamoto K. 2010. Codon reassignment in the Escherichia coli genetic code. *Nucleic Acids Res* 38:8188-95, PMCID: 3001078
140. Murakami H, Ohta A, Ashigai H, Suga H. 2006. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat Methods* 3:357-9
141. Muramatsu M, Hihara Y. 2006. Characterization of high-light-responsive promoters of the psaAB genes in Synechocystis sp. PCC 6803. *Plant Cell Physiol* 47:878-90
142. Nagrath D, Avila-Elchiver M, Berthiaume F, Tilles AW, Messac A, Yarmush ML. 2007. Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. *Ann Biomed Eng* 35:863-85
143. Nakasugi K, Svenson CJ, Neilan BA. 2006. The competence gene, comF, from Synechocystis sp. strain PCC 6803 is involved in natural transformation, phototactic motility and piliation. *Microbiology* 152:3623-31
144. Nam YS, Magyar AP, Lee D, Kim JW, Yun DS, Park H, Pollom TS, Jr., Weitz DA, Belcher AM. 2010. Biologically templated photocatalytic nanostructures for sustained light-driven water oxidation. *Nat Nanotechnol* 5:340-4
145. Neltner B, Peddie B, Xu A, Doenlen W, Durand K, Yun DS, Speakman S, Peterson A, Belcher A. 2010. Production of hydrogen using nanocrystalline protein-templated catalysts on m13 phage. *ACS Nano* 4:3227-35
146. Nielsen PH, Mielczarek AT, Kragelund C, Nielsen JL, Saunders AM, Kong Y, Hansen AA, Vollertsen J. 2010. A conceptual ecosystem model of microbial communities in enhanced biological phosphorus removal plants. *Water Res* 44:5070-88
147. Niu W, Draths KM, Frost JW. 2002. Benzene-free synthesis of adipic acid. *Biotechnol Prog* 18:201-11
148. Noireaux V, Bar-Ziv R, Godefroy J, Salman H, Libchaber A. 2005. Toward an artificial cell based on gene expression in vesicles. *Phys Biol* 2:P1-8
149. Noireaux V, Libchaber A. 2004. A vesicle bioreactor as a step toward an artificial cell assembly. *Proc Natl Acad Sci U S A* 101:17669-74, PMCID: 539773
150. Noireaux V, Maeda YT, Libchaber A. 2011. Development of an artificial cell, from self-organization to computation and self-reproduction. *Proc Natl Acad Sci U S A* 108:3473-80, PMCID: 3048108
151. Ohta K, Beall DS, Mejia JP, Shanmugam KT, Ingram LO. 1991. Genetic improvement of Escherichia coli for ethanol production: chromosomal integration of Zymomonas mobilis genes encoding pyruvate decarboxylase and alcohol dehydrogenase II. *Appl Environ Microbiol* 57:893-900, PMCID: 182819
152. Okamoto S, Ohmori M. 2002. The cyanobacterial PilT protein responsible for cell motility and

transformation hydrolyzes ATP. *Plant Cell Physiol* 43:1127-36

153. Otey CR, Landwehr M, Endelman JB, Hiraga K, Bloom JD, Arnold FH. 2006. Structure-guided recombination creates an artificial family of cytochromes P450. *PLoS Biol* 4:e112, PMCID: 1431580
154. Patel PV, Gianoulis TA, Bjornson RD, Yip KY, Engelman DM, Gerstein MB. 2010. Analysis of membrane proteins in metagenomics: networks of correlated environmental features and protein families. *Genome Res* 20:960-71, PMCID: 2892097
155. Pennisi E. 2009. Genetic engineering. Two steps forward for synthetic biology. *Science* 325:928-9
156. Pepper LR, Cho YK, Boder ET, Shusta EV. 2008. A decade of yeast surface display technology: where are we now? *Comb Chem High Throughput Screen* 11:127-34, PMCID: 2681324
157. Philimonenko AA, Janacek J, Hozak P. 2000. Statistical evaluation of colocalization patterns in immunogold labeling experiments. *J Struct Biol* 132:201-10
158. Pla-Roca M, Leulmi RF, Djambazian H, Sundararajan S, Juncker D. 2010. Addressable nanowell arrays formed using reversibly sealable hybrid elastomer-metal stencils. *Anal Chem* 82:3848-55
159. Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 12:32-42, PMCID: 3074964
160. Podar M, Abulencia CB, Walcher M, Hutchison D, Zengler K, Garcia JA, Holland T, Cotton D, Hauser L, Keller M. 2007. Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl Environ Microbiol* 73:3205-14, PMCID: 1907129
161. Pohlmann A, Fricke WF, Reinecke F, Kusian B, Liesegang H, Cramm R, Eitinger T, Ewering C, Potter M, Schwartz E, Strittmatter A, Voss I, Gottschalk G, Steinbuchel A, Friedrich B, Bowien B. 2006. Genome sequence of the bioplastic-producing "Knallgas" bacterium Ralstonia eutropha H16. *Nat Biotechnol* 24:1257-62
162. Poiata E, Meyer MM, Ames TD, Breaker RR. 2009. A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. *RNA* 15:2046-56, PMCID: 2764483
163. Porreca GJ, Shendure J, Church GM. 2006. Polony DNA sequencing. *Curr Protoc Mol Biol* Chapter 7:Unit 7 8
164. Posfai G, Plunkett G, 3rd, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR. 2006. Emergent properties of reduced-genome Escherichia coli. *Science* 312:1044-6
165. Poteete AR. 2008. Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol Microbiol* 68:66-74
166. Preveral S, Gayet L, Moldes C, Hoffmann J, Mounicou S, Gruet A, Reynaud F, Lobinski R, Verbavatz JM, Vavasseur A, Forestier C. 2009. A common highly conserved cadmium detoxification mechanism from bacteria to humans: heavy metal tolerance conferred by the ATP-binding cassette (ABC) transporter SpHMT1 requires glutathione but not metal-chelating phytochelatin peptides. *J Biol Chem* 284:4936-43
167. PRNewsWire. 2007. *Genome Technology Magazine Names Top Young Investigators of 2007*. http://www.thefreelibrary.com/Genome+Technology+Magazine+Names+Top+Young+Investigators+of+2007.-a0172527944
168. Ranganathan S, Suthers PF, Maranas CD. 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol* 6:e1000744, PMCID: 2855329
169. Rawlings DE, Tietze E. 2001. Comparative biology of IncQ and IncQ-like plasmids. *Microbiol Mol Biol Rev* 65:481-96, table of contents, PMCID: 99038
170. Reppas NB, Lin X, Church GM. 2004. Experimental and Theoretical Studies of Artificial Microbial Symbiosis (http://www.nt.ntnu.no/users/skoge/prost/proceedings/aiche-2004/pdffiles/papers/011e.pdf). In *AIChE 2004 Annual Meeting*. Austin, TX

171. Rogers JM, Lippert LG, Gai F. 2010. Non-natural amino acid fluorophores for one- and two-step fluorescence resonance energy transfer applications. *Anal Biochem* 399:182-9, PMCID: 2830288
172. Ronchel MC, Ramos JL. 2001. Dual system to reinforce biological containment of recombinant bacteria designed for rhizoremediation. *Appl Environ Microbiol* 67:2649-56, PMCID: 92920
173. Rondelez Y, Tresset G, Tabata KV, Arata H, Fujita H, Takeuchi S, Noji H. 2005. Microfabricated arrays of femtoliter chambers allow single molecule enzymology. *Nat Biotechnol* 23:361-5
174. Rothberg JM, Leamon JH. 2008. The development and impact of 454 sequencing. *Nat Biotechnol* 26:1117-24
175. Sakakihara S, Araki S, Iino R, Noji H. 2010. A single-molecule enzymatic assay in a directly accessible femtoliter droplet array. *Lab Chip* 10:3355-62
176. Sawitzke JA, Costantino N, Li XT, Thomason LC, Bubunenko M, Court C, Court DL. 2011. Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J Mol Biol* 407:45-59, PMCID: 3046259
177. Schell DJ, Dowe N, Ibsen KN, Riley CJ, Ruth MF, Lumpkin RE. 2007. Contaminant occurrence, identification and control in a pilot-scale corn fiber to ethanol conversion process. *Bioresour Technol* 98:2942-8
178. Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB. 2010. Microbial biosynthesis of alkanes. *Science* 329:559-62
179. Schwartz D, Chou MF, Church GM. 2009. Predicting protein post-translational modifications using meta-analysis of proteome scale data sets. *Mol Cell Proteomics* 8:365-79, PMCID: 2634583
180. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309:1728-32
181. Shi W, Qin J, Ye N, Lin B. 2008. Droplet-based microfluidic system for individual Caenorhabditis elegans assay. *Lab Chip* 8:1432-5
182. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T. 2001. Cell-free translation reconstituted with purified components. *Nat Biotechnol* 19:751-5
183. Shimizu Y, Kanamori T, Ueda T. 2005. Protein synthesis by pure translation systems. *Methods* 36:299-304
184. Sinha S, Cameron AD, Redfield RJ. 2009. Sxy induces a CRP-S regulon in Escherichia coli. *J Bacteriol* 191:5180-95, PMCID: 2725579
185. Slonczewski JL, Fujisawa M, Dopson M, Krulwich TA. 2009. Cytoplasmic pH measurement and homeostasis in bacteria and archaea. *Adv Microb Physiol* 55:1-79, 317
186. Smith MG, Gianoulis TA, Pukatzki S, Mekalanos JJ, Ornston LN, Gerstein M, Snyder M. 2007. New insights into Acinetobacter baumannii pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. *Genes Dev* 21:601-14, PMCID: 1820901
187. Sode K, Tatara M, Takeyama H, Burgess JG, Matsunaga T. 1992. Conjugative gene transfer in marine cyanobacteria: Synechococcus sp., Synechocystis sp. and Pseudanabaena sp. *Appl Microbiol Biotechnol* 37:369-73
188. Sokic-Lazic D, Minteer SD. 2008. Citric acid cycle biomimic on a carbon electrode. *Biosens Bioelectron* 24:945-50
189. Sommer MO, Church GM, Dantas G. 2010. A functional metagenomic approach for expanding the synthetic biology toolbox for biomass conversion. *Mol Syst Biol* 6:360, PMCID: 2872612
190. Sommer MO, Dantas G, Church GM. 2009. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* 325:1128-31
191. Sonderegger M, Sauer U. 2003. Evolutionary engineering of Saccharomyces cerevisiae for anaerobic growth on xylose. *Appl Environ Microbiol* 69:1990-8, PMCID: 154834
192. Steiert JG, Kubu C, Stauffer GV. 1992. The PurR binding site in the glyA promoter region of

Escherichia coli. *FEMS Microbiol Lett* 78:299-304

193. Sturino JM, Klaenhammer TR. 2006. Engineered bacteriophage-defence systems in bioprocessing. *Nat Rev Microbiol* 4:395-404
194. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osburne MS, Henn MR, Chisholm SW. 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* 12:3035-56, PMCID: 3037559
195. Sullivan MB, Krastins B, Hughes JL, Kelly L, Chase M, Sarracino D, Chisholm SW. 2009. The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. *Environ Microbiol* 11:2935-51, PMCID: 2784084
196. Takagi H, Kobayashi C, Kobayashi S, Nakamori S. 1999. PCR random mutagenesis into Escherichia coli serine acetyltransferase: isolation of the mutant enzymes that cause overproduction of L-cysteine and L-cystine due to the desensitization to feedback inhibition. *FEBS Lett* 452:323-7
197. Tan X, Yao L, Gao Q, Wang W, Qi F, Lu X. 2011. Photosynthesis driven conversion of carbon dioxide to fatty alcohols and hydrocarbons in cyanobacteria. *Metab Eng* 13:169-76
198. Tan Y, Rivera JG, Contador CA, Asenjo JA, Liao JC. 2011. Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metab Eng* 13:60-75
199. Tian J, Gong H, Sheng N, Zhou X, Gulari E, Gao X, Church G. 2004. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* 432:1050-4
200. Tolonen AC, Chilaka AC, Church GM. 2009. Targeted gene inactivation in Clostridium phytofermentans shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Mol Microbiol* 74:1300-13, PMCID: 2810439
201. Tougu K, Marians KJ. 1996. The interaction between helicase and primase sets the replication fork clock. *J Biol Chem* 271:21398-405
202. Touze T, Eswaran J, Bokma E, Koronakis E, Hughes C, Koronakis V. 2004. Interactions underlying assembly of the Escherichia coli AcrAB-TolC multidrug efflux system. *Mol Microbiol* 53:697-706
203. Tran LM, Rizk ML, Liao JC. 2008. Ensemble modeling of metabolic networks. *Biophys J* 95:5606-17, PMCID: 2599852
204. Ueda T, Kanamori T, Ohashi H. 2010. Ribosome display with the PURE technology. *Methods Mol Biol* 607:219-25
205. Ugwumba IN, Ozawa K, Xu ZQ, Ely F, Foo JL, Herlt AJ, Coppin C, Brown S, Taylor MC, Ollis DL, Mander LN, Schenk G, Dixon NE, Otting G, Oakeshott JG, Jackson CJ. 2010. Improving a Natural Enzyme Activity through Incorporation of Unnatural Amino Acids. *J Am Chem Soc*
206. Umbarger MA, Wright MA, Porreca GJ, Bay D, Toro E, Shapiro L, Marti-Renom MA, Dekker J, Church GM. 2011. The 3D Architecture of a Bacterial Genome (submitted).
207. Vicens MC, Sen A, Vanderlaan A, Drake TJ, Tan W. 2005. Investigation of molecular beacon aptamer-based bioassay for platelet-derived growth factor detection. *Chembiochem* 6:900-7
208. Vigneault F, Laserson U, Simen BB, Lieberman-Aiden E, Egholm M, Church GM. 2010. Tracking the dynamics of a human antibody repertoire. (in revision)
209. Viikari L, Alapuranen M, Puranen T, Vehmaanpera J, Siika-Aho M. 2007. Thermostable enzymes in lignocellulose hydrolysis. *Adv Biochem Eng Biotechnol* 108:121-45
210. Vo TD, Greenberg HJ, Palsson BO. 2004. Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *J Biol Chem* 279:39532-40
211. Voorhees RM, Weixlbaumer A, Loakes D, Kelley AC, Ramakrishnan V. 2009. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol* 16:528-33, PMCID: 2679717

212. Walsh G. 2005. Therapeutic insulins and their large-scale manufacture. *Appl Microbiol Biotechnol* 67:151-9
213. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460:894-8
214. Wang HH, Xu G, Vonner A, Church GM. 2011. Modified Bases Enable High-efficiency Oligonucleotide-Mediated Allelic Replacement via Mismatch Repair Evasion (submitted)
215. Wang K, Neumann H, Peak-Chew SY, Chin JW. 2007. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat Biotechnol* 25:770-7
216. Wang P. 2006. Nanoscale biocatalyst systems. *Curr Opin Biotechnol* 17:574-9
217. Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG. 2005. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res* 33:e36, PMCID: 549575
218. Weaver JC, Williams GB, Klibanov A, Demain AL. 1988. Gel Microdroplets: Rapid Detection and Enumeration of Individual Microorganisms by their Metabolic Activity. *Nat Biotech* 6:1084-9
219. Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, Gustafsson C. 2009. Design parameters to control synthetic gene expression in Escherichia coli. *PLoS One* 4:e7002, PMCID: 2736378
220. Wheeldon IR, Gallaway JW, Barton SC, Banta S. 2008. Bioelectrocatalytic hydrogels from electron-conducting metallopolypeptides coassembled with bifunctional enzymatic building blocks. *Proc Natl Acad Sci U S A* 105:15275-80, PMCID: 2563127
221. Willner B, Katz E, Willner I. 2006. Electrical contacting of redox proteins by nanotechnological means. *Curr Opin Biotechnol* 17:589-96
222. Win MN, Liang JC, Smolke CD. 2009. Frameworks for programming biological function through RNA parts and devices. *Chem Biol* 16:298-310, PMCID: 2713350
223. Xu Q, Schlabach MR, Hannon GJ, Elledge SJ. 2009. Design of 240,000 orthogonal 25mer DNA barcode probes. *Proc Natl Acad Sci U S A* 106:2289-94, PMCID: 2631075
224. Xu W, McFadden BA. 1997. Sequence analysis of plasmid pCC5.2 from cyanobacterium Synechocystis PCC 6803 that replicates by a rolling circle mechanism. *Plasmid* 37:95-104
225. Yamamoto K, Matsumoto F, Minagawa S, Oshima T, Fujita N, Ogasawara N, Ishihama A. 2009. Characterization of CitA-CitB signal transduction activating genes involved in anaerobic citrate catabolism in Escherichia coli. *Biosci Biotechnol Biochem* 73:346-50
226. Yang X, Bing T, Mei H, Fang C, Cao Z, Shangguan D. 2011. Characterization and application of a DNA aptamer binding to L-tryptophan. *Analyst* 136:577-85
227. Ye X, Wang Y, Hopkins RC, Adams MW, Evans BR, Mielenz JR, Zhang YH. 2009. Spontaneous high-yield production of hydrogen from cellulosic materials and water catalyzed by enzyme cocktails. *ChemSusChem* 2:149-52
228. Yoshihara S, Geng X, Okamoto S, Yura K, Murata T, Go M, Ohmori M, Ikeuchi M. 2001. Mutational analysis of genes involved in pilus structure, motility and transformation competency in the unicellular motile cyanobacterium Synechocystis sp. PCC 6803. *Plant Cell Physiol* 42:63-73
229. Youngman EM, Green R. 2005. Affinity purification of in vivo-assembled ribosomes for in vitro biochemical analysis. *Methods* 36:305-12
230. Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* 97:5978-83, PMCID: 18544
231. Yura K, Toh H, Go M. 1999. Putative mechanism of natural transformation as deduced from genome data. *DNA Res* 6:75-82
232. Zang X, Liu B, Liu S, Arunakumara KK, Zhang X. 2007. Optimum conditions for transformation of Synechocystis sp. PCC 6803. *J Microbiol* 45:241-5

233. Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, Keller M. 2002. Cultivating the uncultured. *Proc Natl Acad Sci U S A* 99:15681-6, PMCID: 137776
234. Zengler K, Walcher M, Clark G, Haller I, Toledo G, Holland T, Mathur EJ, Woodnutt G, Short JM, Keller M. 2005. High-throughput cultivation of microorganisms using microcapsules. *Methods Enzymol* 397:124-30
235. Zhang YH. 2010. Production of biocommodities and bioelectricity by cell-free synthetic enzymatic pathway biotransformations: challenges and opportunities. *Biotechnol Bioeng* 105:663-77

Appendix 4: Facilities and other resources

FACILITIES: Harvard University

Laboratory:
Facilities and Resources:
The Harvard/MIT/BU intellectual environment is excellent for multidisciplinary, collaborative and functional genomics research. The Church laboratory provides some of the glue with students from all three universities and a location n several adjacent buildings at the heart of the

HMS campus:

1. The New Research Building is home to the Genetics department; and the Lipper Center for Computational Genetics.
2. The Seeley-Mudd building is home to the Harvard Institute of Proteomics (HIP), the Harvard Institute of Chemistry and Cell Biology (ICCB)
3. The Thorn Building of the BWH Genomics & Bioinformatics Center.
4. Wyss Institute of Biologically inspired Engineering.
   Harvard has recently made considerable endowment commitments to the above and the university-wide Center for Genomics Research. We work closely with our departmental Biopolymers facility, which has a staff of five; departmental computer facility with a staff of four.
   We have direct computer network and CAD links to the HMS machine shop, which coordinates with several other university and commercial machining and design facilities.

Clinical
n/a

Animal:
n/a

Computer:
The Group has an extensive computer facilities and CAD-PAM software for design of DNA constructs. Computers are connected via LAN to the HMS campus network for access to scientific literature.

Office:
PI's office space is in NRB building, at 77 Avenue Louis Pasteur. The lab has two large rooms, NRB 232, and NRB 238, and an office area for PI, administration, summer students and rotation students.

Other:

The lab has two large rooms, NRB 232, and NRB 238, and an office area for administration, summer students and rotation students.

FACILITIES: Yale University

The Isaacs lab is affiliated with the Department of Molecular, Cellular and Developmental

Biology and the Systems Biology Institute with office locations in the Kline Biology Tower and at West Campus, which is comprised of 20 buildings across 136 acres and home to five interdisciplinary life science institutes at Yale. Yale's West Campus houses the Isaacs Lab, which occupies ~1300 ft^2 of space on the second floor of Building B-31. Most facilities for basic research are on this floor: the Isaacs laboratory, shared equipment (centrifuges, HPLC, phosphoimager), dark room, cold room, autoclave and dishwashing facilities (on the first floor), and seminar room. Located at West Campus, the Isaacs lab also has unique access to four adjacent core facilities: Yale Center for Genome Analysis, Small Molecule Discovery Center, High Throughput Cell Biology and High Performance Computing Center.

Appendix 5: Equipment

Equipment: Harvard

2 Affymetrix Chip Scanners (HP & MD) and fluidics stations
1 Microarrayer prototype (Anorad stages),150 slide capacity, 16 piezoheads (GeSim)
1 microarray scanner (General Scanning 5000)
1 Automated DNA and protein sequencers, synthesizers and related items
(ABI 3700, 377, 373S, 391, 1000S, 394, 380B, 270A, 477A, 430A, 420A, 130A)
1 FPLC and Phast systems (Pharmacia)
1 LCQ HPLC-MSn Ion Trap mass spectrometer (Finnigan)
1 Storm Fluorimager with 29 exposure plates (Molecular Dynamics)
Numerous PCR machines with 96, 384-well, and slide heads (MJR)
1 Microfluidics development platform (Caliper)
5 -20 C freezers and two -80 freezers
7 low-speed centrifuges and ultra-centrifuges (IEC, Sorvall, Beckmann)
2 Oscilloscopes and 2 electrophysiological amplifiers 70 femtoamp rms (Axon)
1 micropipette puller and microforge (Narishige)
5 high voltage (500V to 6000V) power supplies (Biorad and EC)
2 Ultra-thin gel Direct Transfer Electrophoresis (HMS shop, Cykal)
1 96-pin array oligonucleotide synthesizer Primer Station 960 (IAS & HMS)
3 electrotransfer devices (Polytech)
1 pulsed-field CHEF boxes (Genplex)
1 UV crosslinker (HMS shop)
1 Capillary array electrophoresis prototype (HMS shop)
1 Laser-induced fluorescent 4-color capillary electrophoresis (ABI 310)
2 DEC alpha fileservers running Ultrix
1 dual Intel PII, RAID level 5 based Linux fileserver
15 computers running under WinNT, 10 Linux, 6 Linux&NT, 5 MacOS
1 Silicon Graphics Octane computer
1 Linux Celeron Cluster (Beowulf-type) for parallel & associative processing
1 Terabyte tape jukebox server running Arkeia
1 Confocal Microscope (Biorad)
1 Automatic film processor
1 Bioflo 3000 mammalian and microbial cell culture chemostat (New Brunswick)
1 EPICS ALTRA flow sorter with Autoclone multiwell plates option (Beckman-Coulter)
1 M5 plate reader (Molecular Devices)
1 KBiosciences high-capacity thermal cycler
3 Danaher Motion Dover Polonator DNA sequencers
Biotek Synergy H4 multimode plate reader
1 M2 plate reader (Molecular Devices)

Appendix 6: Other Attachments

Contents of Appendix 6

Complete list of Church Lab Post docs, Grad Students, and other affiliated people for the past five years

This list continues the partial list provided in Appendix 1.

Person	Church Lab Status	Position	Affiliation
Aach, John	Current	Lecturer	Harvard Med School
Adesokan, Adeyemi	Former	CEO	Pathogenica
Ahlford, Annika	Former	Visiting fellow	Uppsala University
Azizi, Elham	Current	Rotation student	Harvard Med School
Bachelet, Ido	Current	Post doc	Harvard Med School
Bang, Duhee	Former	Assistant Professor	Yonsei University
Barrera, Luis	Current	Rotation student	Harvard Med School
Batada, Nizar	Former	Assistant Professor	U of Toronto
Bobe, Jason	Current	Consultant	Personal Genome Project
Briggs, Adrian	Current	Post doc	Harvard Med School
Brown, Chris	Former	Rotation student	Harvard Med School
Busskamp, Volker	Current	Post doc	Harvard Med School
Byrne, Susan	Current	Post doc	Harvard Med School
Cai, Long	Former	Post doc	?
Chari, Raj	Current	Post doc	Harvard Med School
Chew, Wei Leong	Current	Grad student	Harvard Med School
Chilaka, Amanda	Former	Undergraduate Intern	Northeastern University
Chou, Michael	Current	Post doc	Harvard Med School
Church, George	Current	Post doc	Harvard Med School
Cong, Le	Current	Grad student	Harvard Med School
Cox, David	Current	Rotation student	Harvard Med School
Curtis, Wayne	Former	Visiting scientist	U Penn
Dai, Mingjie	Current	Rotation student	Harvard Med School
Daniels, Rachel	Former	Rotation Student	Harvard Med School
Dantas, Gautam	Former	Assistant Professor	Wash. U. St Louis
DiCarlo, James	Current	Grad student	Harvard Med School
Douglas, Shawn	Current	Post doc	Harvard Med School
Eroshenko, Nikolai	Current	Grad student	Harvard Med School

Estep, Pete	Current	Visting scientist	TeloMe, Inc.
Forest, Craig	Former	Assistant Professor	Georgia Inst Tech
Gianoulis, Tara	Current	Post doc	Harvard Med School
Goddard, Noel	Former	Assistant Professor	Hunter College
Goodman, Daniel Bryan	Current	Grad student	Harvard Med School
Gregg, Christopher	Current	Post doc	Harvard Med School
Gu, LiangCai	Current	Post doc	Harvard Med School
Guell Cargol, Marc	Current	Post doc	Harvard Med School
Guido, Nicholas	Current	Post doc	Harvard Med School
Hall, Heather	Current	Grad student	Harvard Med School
Huang, Po-Yi	Current	Grad student	Harvard Med School
Isaacs, Farren	Former	Assistant Professor	Yale University
Jajoo, Rishi	Current	Rotation student	Harvard Med School
Janse, Dan	Former	Post doc	?
Jewett, Michael	Former	Assistant Professor	Northwestern U
Kim, Jong-Hyun	Current	Post doc	Harvard Med School
Kim, Dae	Former	Post doc	?
Kosuri, Sriram	Current	Post doc	Harvard Med School
Lajoie, Marc	Current	Grad student	Harvard Med School
Laserson, Uri	Current	Grad student	Harvard Med School
Lee, Jay	Current	Post doc	Harvard Med School
Levanon, Erez	Former	Assistant Professor	Bar-Ilan University
Levner, Daniel	Current	Post doc	Harvard Med School
Lewis, Nathan	Current	Post doc	Harvard Med School
Li, Jun	Current	Grad student	Harvard Med School
Li, Chao	Current	Post doc	Harvard Med School
Li, Xin	Current	Rotation student	Harvard Med School
Li, Jin Billy	Former	Assistant Professor	Stanford
Liang, Fan	Former	Assistant Professor	?
Lieber, Daniel	Former	Rotation Student	Keck School Medicine
Lun, Desmond	Former	Assitant Professor	University South Australia
Lunshof, Jeantine	Current	Consultant	Maastricht University, VU University
Maguire, Yael	Current	Visting scientist	MIT
Mali, Prashant	Current	Post doc	Harvard Med School
Mandell, Dan	Current	Post doc	Harvard Med School
Marblestone, Adam	Current	Grad student	Harvard Med School
Maxwell, Evan	Current	Rotation student	Harvard Med School
Mayshar, Yoav	Current	Visting scientist	Harvard Med School
Mee, Michael	Current	Grad student	Boston University
Mosberg, Joshua	Current	Grad student	Harvard Med School

Mukhtar, Hamid	Current	Visiting Fulbright Scholar	Harvard Med School
Nguyen, Anthony	Current	Rotation student	Harvard Med School
Pe'er, Dana	Former	Assistant Professor	Columbia University
Porreca, Greg	Former	Director	Good Start Genetics
Price Ball, Madeleine	Current	Post doc	Harvard Med School
Raman, Srivatsan	Current	Post doc	Harvard Med School
Reppas, Nikos	Former	Scientist	Joule
Rios, Xavier	Current	Grad student	Harvard Med School
Robasky, Kimberly	Current	Grad student	Boston University
Rockwell, Graham	Current	Grad student	Boston University
Rogers, Jameson	Current	Grad student	Harvard Med School
Rosa Girald, Willie	Current	Post doc	Harvard Med School
Rosenbaum, Abraham	Former	Scientist	Ion Torrent
Schwartz, Dan	Former	Assistant Professor	University of Connecticut
Shendure, Jay	Former	Assistant Professor	Washington U. St. Louis
Sismour, Michael	Current	Post doc	Harvard Med School
Sommer, Morten	Former	Assistant Professor	DTU Denmark
Strong, Michael	Former	Assistant Professor	University of Colorado
Sun, Arthur	Current	Rotation student	Harvard Med School
Sunguroff, Alex	Current	Visting scientist	Synvivo Corp
Taylor, Noah	Current	Grad student	Harvard Med School
Ter-Ovanesyan, Dmitry	Current	Rotation student	Harvard Med School
Terry, Richard	Current	Scientist	Harvard Med School
Thakuria, Joe	Current	Post doc	Harvard Med School
Tolonen, Andy	Current	Post doc	Harvard Med School
Umbarger, Mark	Former	Scientist	Good Start Genetics
Vigneault, Francois	Current	Post doc	Harvard Med School
Vonner, Ashley	Current	Rotation student	Harvard Med School
Wait Zaranek, Alexander	Current	Post doc	Harvard Med School
Wang, Harris	Current	Post doc	Harvard Med School
Wong, Shou	Current	Visting scientist	Merck-KGaA
Yang, Luhan	Current	Grad student	Harvard Med School
Yang, Joyce	Current	Grad student	Harvard Med School
Yaung, Stephanie	Current	Grad student	Harvard Med School
Yu, Chuanfei	Former	Post Doc	?
Yu Sun, Arthur	Current	Grad student	Harvard Med School
Zhang, Feng	Former	Assistant Professor	Broad Institute
Zhang, Kun	Former	Assistant Professor	UCSD

HARVARD UNIVERSITY

DEPARTMENT OF PHYSICS
SCHOOL OF ENGINEERING AND APPLIED SCIENCES

PIERCE HALL, 29 OXFORD ST.
CAMBRIDGE, MA 02138

DAVID A. WEITZ
MALLINCKRODT PROFESSOR OF PHYSICS AND APPLIED PHYSICS

June 25, 2011

Prof. George M. Church
Department of Genetics
Harvard Medical School
New Research Building, Room 238
77 Avenue Louis Pasteur
Boston, MA 02115

Dear George,

I am writing you this letter to express my enthusiastic support of your GTL grant submission.

My lab has developed microfluidic methods for single cell encapsulation, controlled merging, and automated sorting of droplets. We would be happy to share protocols and consulting on the design of your device.

Methods that do not rely on imaging to interrogate the interactions between bacteria are critical in being able to understand the structure of natural environmental samples.

Given your expertise in the development of new library protocols, I am confident that you will succeed in producing a method for interrogating the interactions between bacteria through bead capture.

Best regards,

David Weitz

David A. Weitz

VOICE: (617) 496-2842 **FAX (617) 495-0426** **EMAIL: weitz@seas.harvard.edu**