Synthetic Metabolism: Engineering Biology at the Protein and Pathway Scales

Collin H. Martin, David R. Nielsen, Kevin V. Solomon and Kristala L. Jones Prather*

Department of Chemical Engineering, Synthetic Biology Engineering Research Center (SynBERC), Massachusetts Institute of Technology
Cambridge, MA, 02139

* Corresponding author,
77 Massachusetts Avenue, Room 66-458
Cambridge, MA 02139, USA
kljp@mit.edu
Phone: 617-253-1950
Fax: 617-258-5042

Keywords: Synthetic Biology, Metabolism, Proteins, Pathways

Running title: *Engineering Biology at the Protein and Pathway Scales*

1 **Abstract**

2 Biocatalysis has become a powerful tool for the synthesis of high value compounds,
3 particularly so in the case of highly functionalized and/or stereoactive products. Nature
4 has supplied thousands of enzymes and assembled them into numerous metabolic
5 pathways. While these native pathways can be use to produce natural bioproducts, there
6 are many valuable and useful compounds which have no known natural biochemical
7 route. Consequently, there is a need for both unnatural metabolic pathways and novel
8 enzymatic activities upon which these pathways can be built. Here, we review the
9 theoretical and experimental strategies for engineering synthetic metabolic pathways at
10 the protein and pathway scales and highlight the challenges that this subfield of synthetic
11 biology currently faces.

12

13 **Introduction**

14 Synthetic biology has emerged as a powerful discipline for the creation of novel
15 biological systems (Endy, 2005; Pleiss, 2006), particularly within the subfield of
16 metabolic pathway and product engineering (Keasling, 2008; Savage et al., 2008).
17 Continuing efforts to characterize and understand natural enzymes and pathways have
18 opened the door for the building of synthetic pathways towards exciting and beneficial
19 compounds such as the anti-malarial drug precursor artemisinic acid (Ro et al., 2006) and
20 several branched-chain alcohols for use as biofuels (Atsumi et al., 2007). The need for
21 synthetic metabolic routes is a consequence of the fact that the array of compounds of
22 interest for biosynthesis vastly outnumbers the availability of characterized pathways and
23 enzymes. Several key building blocks can be made biologically (Patel et al., 2006);
24 however, a recent report from the U.S. Department of Energy highlighted twelve
25 biomass-derived chemical targets, only half of which have known biochemical routes
26 (Werpy and Petersen, 2004).

27

28 With the lack of characterized natural pathways to synthesize many high-value
29 compounds, we must learn to forge our own metabolic routes towards these molecular
30 targets. Logically, it follows that for unnatural pathways, we will need new, unnatural
31 enzymes from which these pathways can be composed. The parts-devices framework of
32 synthetic biology lends itself well to this dual-sided problem of synthetic pathway
33 creation (Endy, 2005); that is, pathways can be thought of as metabolic devices
34 composed of individual enzyme-catalyzed reaction parts. Implicit within this framework
35 is the idea that the challenges of pathway creation are best approached at both the part

36 and device levels. In this review, we first discuss efforts at the protein-level for
37 broadening the array of enzyme parts that can be recruited for use in synthetic pathways.
38 The discussion is then expanded to pathway-level synthetic biology, where we review the
39 tools available for designing metabolic pathways from enzyme-level parts and the
40 implementation strategies for realizing these pathways experimentally. The overall
41 process of pathway creation (Figure 1) combines experimental and theoretical
42 components of synthetic biology at both scales.
43
44 **Synthetic Biology at the Protein Scale**
45 Through natural evolution, organisms have acquired the capacity to catalyze a multitude
46 of diverse chemical reactions as a means to proliferate in a wide range of unique
47 microenvironments. Although only a small fraction of the earth's biodiversity (and an
48 even smaller subset of its composite enzymes) has been characterized, the identification
49 and isolation of novel proteins with unique properties or enzymatic function is a
50 laborious procedure. One particularly promising source of new enzymes and enzymatic
51 activities is the emerging field of metagenomics (Handelsman, 2004). Nonetheless, the
52 physical and catalytic properties of natural enzymes often render them as incompatible
53 or, at the very least, unoptimized for use in engineered pathways and strains. In cases
54 where natural evolution has fallen short of industrial needs, the tools and practices of
55 synthetic biology can be applied to aid in the creation of designer enzymes and cellular
56 phenotypes. The challenge of building new enzymes and reengineering natural ones has
57 been approached with the development of predictive theoretical frameworks and a range
58 of experimental techniques (Figure 2).

Theoretical Approaches

Computational tools exist to adapt the natural array of proteins for use in an increasing number of applications. For example, the effects of codon bias on expression levels (Kane, 1995; Gustafsson et al., 2004) can be resolved by design tools such as Gene Designer (Villalobos et al., 2006). Other effects such as Shine-Dalgarno sequences, promoter strength, and mRNA stability can be similarly optimized. Nonetheless, the application of these tools is still limited to the biochemical diversity found in nature. To increase the number and efficiency of biologically-catalyzed reactions, more sophisticated *in silico* techniques are needed. While full-scale protein folding and *ab initio* protein design and modeling are neither trivial nor currently practical, the use of solved protein structures, strong physical models and experimentally derived libraries allow for the design and improvement of enzymes. These theoretically designed proteins in turn have significant potential to impact pathway-level synthetic metabolism (Yoshikuni et al., 2008).

An empirical approach to synthetic protein design includes an understanding of the protein sequence/function relationship. One example is the use of a linguistic metaphor to describe a protein sequence (Searls, 1997; Searls, 2002). In language, a sentence is composed of a sequence of words whose parsed meaning is a function of not only their individual definitions but their connotations which are encoded by their type (part of speech) and their relative location to other words. Similarly, a protein 'sentence' is composed of residues that have not only a definitive identity but also possess chemical

82 properties and a relative position that affect the subsequent fold and function of the
83 resultant protein (Przytycka et al., 2002). Building on the successes of a putative protein
84 grammar (Przytycka et al., 2002; Naoki and Hiroshi, 1997), Loose et al. (2006) recently
85 demonstrated its use in the design of new antimicrobial peptides. Using the TEIRESIAS
86 algorithm (Rigoutsos and Floratos, 1998), a library with homology to known sequences
87 restricted to below 60% was generated with approximately 50% of designs showing some
88 antimicrobial activity. An alternative approach to modeling protein sequence/function
89 relationships involves the use of folded protein scaffolds and quantum transition state
90 models. Through detailed crystal structures and transition state models, Hederos et al.
91 (2004) noted that the active site of a glutathione transferase was of the appropriate size
92 and structure to stabilize the transition state complex of the hydrolytic degradation of a
93 thioester. By introducing a histidine residue within the active site they were able to
94 impart significant thioesterase activity. Finally, physics based free energy approaches
95 have been developed to predict protein structure/function relationships in the context of
96 antibody binding strength. While total free energy models were not a good predictor,
97 Lippow et al. (2007) found that the electrostatic interaction contributions to total energy
98 were well correlated with antibody binding affinity. Using this relationship, they were
99 able to generate an improved lysozyme antibody design which demonstrated a 140-fold
100 increase in binding. While neither of these examples fully describe protein
101 structure/function relationships, each does offer a unique insight into the problem.
102 Namely, they drastically reduced the sequence space of potential modifications to a
103 manageable subset with a high probability of success. In this manner, such empirical
104 models serve as an important tool in the design and improvement of enzymes.

105

106 Using a quantum transition state framework, great strides have been made in the *in silico*
107 development of enzyme activities (Jiang et al., 2008; Rothlisberger et al., 2008; Kaplan
108 and Degrado, 2004). At the heart of these efforts is a strong understanding of the desired
109 catalytic mechanism and its associated transition states and reaction intermediates. Once
110 compiled, this information can be used to generate an active site of the appropriate
111 dimensions with critical residues incorporated into appropriate locations for catalysis. At
112 this point, the designer has two options: try to identify a suitable folded scaffold that can
113 accommodate the active site with minimal mutations or generate a protein backbone with
114 correctly folded active site *de novo*. Each method has its inherent advantages and
115 challenges. While finding a host scaffold would appear to be the simpler of the two, it
116 requires extensive searches of protein structure libraries with tools such as RosettaMatch
117 (Zanghellini et al., 2006). Nonetheless, this approach has had some success with the
118 catalysis of unnatural reactions such as the retro-aldol catalysis of 4-hydroxy-4-(6-
119 methoxy-2-naphthyl)-2-butanone (Jiang et al., 2008) and the Kemp elimination
120 (Rothlisberger et al., 2008). Coupled with experimental techniques, *in silico* designed
121 enzymes can have activity levels comparable to that of evolved natural enzymes
122 (Rothlisberger et al., 2008). In contrast, *de novo* protein scaffold development requires
123 significant computational effort to not only consider the stability of the desired
124 conformation of the backbone and active site but also the likelihood of destabilization.
125 Nonetheless, Kaplan and DeGrado (2004) have successfully used such an approach to
126 generate an O_2-dependent phenol oxidase. Despite the computational overhead
127 associated with these methods, their feasibility points to an improving and functional

128 understanding of protein structure/function relationships, leading to increased
129 possibilities for the rational design of enzymes and proteins.
130
131 In the absence of rational insight, theoretical tools can assist experimental techniques in
132 generating new and improved proteins. One common technique is protein recombination
133 or *in vitro* shuffling which combines the best traits of two or more individual enzymes
134 (Stemmer, 1994a; Stemmer, 1994b). However, successful recombination is contingent
135 on shuffling at domain boundaries to ensure proper folding of each domain. The
136 predictive algorithm SCHEMA, developed by Voigt et al. (2002), was designed to aid in
137 the screening process of such chimeric proteins. By analyzing the nature and number of
138 the disruptions of the intermolecular interactions, Voigt et al. were able to generate a
139 metric correlated with the probability of active β-lactamase hybrids of TEM-1 and PSE-4
140 (2002). Subsequent studies by Meyer et al. (2003) have confirmed this correlation and
141 used SCHEMA-guided recombination to derive functional and diverse libraries of
142 cytochrome P-450s (Otey et al., 2004) and β-lactamases (Meyer et al., 2006). Another
143 available predictive algorithm is FamClash (Saraf et al., 2004), which analyzes chimeras
144 for the conservation of charge, volume and hydrophobicity at a given residue. Generated
145 sequence scores have been demonstrated to be well correlated with the activities of
146 hybrid dihydrofolate reductases. While experimental techniques are important generators
147 of diverse protein libraries, tools such as FamClash, SCHEMA and other related
148 sequence analysis programs enrich such chimeric libraries and vastly improve their value
149 in the development of new and improved proteins. Currently, these tools are incapable of

150 predicting hits *a priori*; however, their importance in successful protein design should not
151 be underestimated.
152
153 **Experimental Approaches**
154 Rather than focusing on the prediction of protein structure and function, experimental
155 techniques allow the improvement or modification of existing enzymes, in some
156 instances creating entirely new enzymes and enzyme activities. These techniques include
157 mutagenesis, enzyme engineering and evolution, and gene synthesis technology, with
158 each boasting their own distinct advantages and inherent limitations (Bonomo et al.,
159 2006; Alper and Stephanopoulos, 2007). Collectively, they comprise a powerful set of
160 tools for the efficient generation of enzymes with user-specified properties. Protein
161 recombination, for example, provides a means by which secondary structural elements,
162 from natural or evolved proteins, can be rationally assembled in a modular fashion to
163 integrate domains featuring desired attributes (Otey et al., 2004).
164
165 The construction of synthetic pathways typically involves the recruitment of genes from
166 an array of sources to provide the required enzymatic function and activity (Figure 1).
167 However, heterologously expressed proteins, particularly those originating from a source
168 organism belonging to a different kingdom than that of the expression host, often suffer
169 from poor activity as a result of dissimilarities in codon usage. In such cases, the use of
170 synthetic genes with codon optimized sequences has been frequently employed to
171 achieve sufficient levels of functional expression. Synthesis of a codon optimized
172 xylanase gene from *Thermomyces lanuginosus* DSM 5826 led to a 10-fold improvement

173 in expression level in *E. coli* (Yin et al., 2008). Plant genes are often found to be poorly
174 expressed in *E. coli* (Martin et al., 2001). Martin et al. (2003) synthesized a codon
175 optimized variant of amorpha-4,11-diene synthase from the *Artemisia annua* to catalyze
176 the conversion of farnesyl pyrophosphate to amorphadiene, a precursor used for the
177 production of the anti-malarial drug artemisinin. As the cost associated with gene
178 synthesis continues to decrease, imaginable applications of synthetic genes and artificial,
179 designer proteins to include increased elements of rational design become increasingly
180 plausible.

181

182 The versatility of directed evolution for engineering desired enzyme attributes is
183 highlighted by a multitude of recent works employing this approach for a diverse
184 assortment of applications, including the enhancement of thermal stability (Asako et al.,
185 2008; Shi et al., 2008) and acid tolerance (Liu et al., 2008); promoting higher chemo-,
186 regio-, and enantio-selectivity towards substrates (Asako et al., 2008); elimination of
187 undesired biochemical activities (e.g., side reactions; Kelly et al., 2008); and improving
188 heterologous expression (Mueller-Cajar et al., 2008). In the example of the stereospecific
189 reduction of 2,5-hexanedione to (2S,5S)-hexanediol by alcohol dehydrogenase (AdhA)
190 from the thermophillic bacteria *Pyrcoccus furiosus*, laboratory evolution was used by
191 Machielsen et al. (2008) to alter the enzyme's optimum temperature and improve its
192 activity in recombinant *E. coli* under moderate culture conditions. Meanwhile, Aharoni
193 et al. (2004) have achieved functional expression of mammalian paraoxonases PON1 and
194 PON3 in *E. coli* through a directed evolution scheme that incorporated family DNA
195 shuffling (shuffling of DNA encoding homologous genes from different genetic sources)

196 and random mutagenesis to achieve the first active microbial expression of recombinant
197 PON variants. As a tool, directed evolution continues to benefit from refinements aimed
198 at improving the efficiency at which desired mutations can be obtained from a minimal
199 number of iterations while also reducing screening efforts (Reetz et al., 2007; Reetz et al.,
200 2008).
201
202 In addition to improving expression and altering the thermal properties of heterologous
203 enzymes, novel biochemical activities can be similarly engineered by the aforementioned
204 strategies. For example, cytochrome P450 BM3 from *Bacillus megaterium* has been
205 engineered via directed evolution using several sequential rounds of mutagenesis to alter
206 its regioselectivity for the hydroxylation of n-alkanes from subterminal positions to that
207 of the terminus (Meinhold et al., 2006). The approach has been employed to convert
208 several different n-alkanes to their corresponding n-alcohols, including the hydroxylation
209 of ethane to ethanol as a means for producing more tractable transportation fuels from
210 petrochemical feedstocks (Meinhold et al., 2005). To promote high end-product
211 specificity while maximizing metabolite flux, the preferential activity of an enzyme
212 between multiple competing substrates can also be tailored. For instance, the substrate
213 specificity of pyruvate oxidase (PoxB) from *E. coli* was altered via localized random
214 mutagenesis to decrease its activity on pyruvate in favor of an alternative endogenous
215 metabolite, 2-oxo-butanoate (Chang and Cronan, 2000). Synthetic pathways
216 incorporating this PoxB mutant will accordingly display preferential synthesis of
217 products from the four-carbon precursor. Meanwhile, Tsuge et al. (2003) utilized site
218 directed mutagenesis to shift the substrate specificity of PhaJ, an *R*-specific enoyl-CoA

219 hydratase from *Aeromonas caviae* from short-chain 3-hydroxyacyl-CoA precursors
220 towards those with longer carbon chain lengths (8 to 12). When incorporated into an
221 engineered polyhydroxyalkanoate (PHA) synthesis pathway in *E. coli*, increased molar
222 fractions of C_8 and C_{10} 3-hydroxyacid monomer units were found to be incorporated into
223 PHA. In this case, the capacity to distinctly manipulate the composition of PHAs makes
224 possible the synthesis of novel bio-plastics with customizable physical properties to meet
225 commercial requirements. The ability to finely tune the substrate specificity of an
226 engineered enzyme is of particular importance for promoting high selectivity and product
227 yield, as well as for reducing the ill-effects of molecular cross-talk between engineered
228 and endogenous pathways.
229
230 At the protein level, synthetic biology aims to expand the catalog of well-characterized
231 enzymes while also engineering novel biochemistries. Subsequent incorporation of
232 engineered enzymes into synthetic pathways leads to the construction of devices that can
233 be implemented to achieve a user-specified function, such as the production of biofuels or
234 high-value pharmaceutical compounds. The design and construction of new metabolic
235 routes from individual enzymes represents synthetic biology at the next scale, the
236 pathway scale, and has unique challenges of its own.
237

238 Synthetic Biology at the Pathway Scale

239 Pathway-scale synthetic biology aims to create novel metabolic routes towards both
240 existing metabolites and unnatural compounds. Traditionally, pathway engineering has
241 been synonymous with metabolic engineering and its toolbox has been composed of the

242 same tools: gene knockouts, flux optimization, gene overexpression, and the like. The
243 ability to manipulate natural metabolism has seen many useful applications, such as
244 improving ethanol production in *Saccharomyces cerevistiae* (Bro et al. 2006),
245 solventogenesis in *Clostridium acetobutylicum* (Mermelstein et al., 1993; Woods, 1995),
246 and penicillin production in *Penicillium chrysogenum* (Casqueiro et al., 2001). A key
247 limitation in all of these examples is the confinement of pathway engineering to the
248 manipulation of natural metabolism. Continuing advances in characterizing, modifying,
249 and even creating enzymes (several of them discussed in the previous section of this
250 review) now allow us to build unnatural pathways for the biological production of
251 compounds. Understanding synthetic biology at the protein scale affords us the
252 opportunity to apply it at the pathway scale.
253
254 As at the protein scale, pathway-level synthetic biology has been approached from both
255 theoretical and experimental fronts. The theoretical work centers on the concept of
256 pathway design – assembling a logical series of enzyme-catalyzed reactions to convert an
257 accessible substrate into a valued final compound. Theoretical pathway design probes
258 what conversions are possible and what enzyme parts need to be assembled to create a
259 functional metabolic device. In contrast, experimental efforts focus on the construction
260 and application of unnatural pathways and serve as powerful real-world examples of what
261 these pathways can accomplish. Experimental approaches enable the exploration of
262 enzyme behaviors such as substrate promiscuity and activity, both useful properties for
263 creating unnatural pathways that cannot readily be predicted with theoretical approaches.
264

Theoretical Approaches

Before an unnatural metabolic pathway can be built in the laboratory, it must first be designed. The goal of pathway design is to use a series of biochemically-catalyzed reactions to connect a target product molecule to either a cellular metabolite (such as acetyl-CoA, α-ketoglutarate, or L-alanine) or to a feasible feedstock (such as glucose or glycerol). This can be accomplished using either natural enzymes or engineered ones. The sheer number of known enzymes (both natural and engineered) and enzyme-catalyzed reactions available means that there will almost certainly exist many possible theoretical pathways towards a given target compound (Li et al., 2004; Hatzimanikatis et al., 2005). Identifying and ranking these different possibilities are the central challenges in pathway design.

One of the first steps in pathway design is obtaining knowledge of the enzymes and enzyme-catalyzed reactions available for use in a pathway. Comprehensive protein and metabolism databases, such as BRENDA (Schomburg et al., 2004), KEGG (Kaneshisa et al., 2006), Metacyc (Capsi et al., 2006), and Swiss-Prot (Wu et al., 2006), provide a wealth of information on the pool of natural, characterized enzymes that can be recruited. More importantly, these databases reveal chemical conversions that are achievable with enzymes. As of the preparation of this manuscript, there are approximately 398,000 protein entries in Swiss-Prot (build 56.2), from which the enzymes are organized into 4757 four-digit enzyme classification (E.C.) groups in the most recent version of BRENDA (build 2007.2). Because of the large number of characterized enzymes, those performing similar reaction chemistries are typically organized into generalized enzyme-

288 catalyzed reactions for the purposes of pathway construction (Li et al., 2004). A
289 generalized enzyme-catalyzed reaction is defined as the conversion of one functional
290 group or structural pattern in a substrate into a different group or structure in its product
291 (Figure 3). Structural information about the non-reacting portions of the substrate is
292 ignored, making the identification of enzymes to carry out a desired chemical conversion
293 a much more tractable problem. However, the logical rules for assigning enzymes to a
294 generalized reaction can be subjective (Figure 3). One could for instance differentiate
295 between reactions solely on the reacting functional groups (i.e. aldehyde to alcohol) as Li
296 and coworkers (2004) did, or one could also include information about conserved
297 patterns of molecular structure between similar enzyme-catalyzed reactions.
298 Furthermore, generalized enzymatic reactions do not all fall cleanly into the existing E.C.
299 system (Figure 3c).
300

301 Despite the need for a universal standard in reaction generalization, several publically-
302 available tools utilize this approach to address the problem of pathway design. The
303 BNICE (*B*iochemical *N*etwork *I*ntegrated *C*omputational *E*xplorer) framework allows for
304 the discovery of numerous possible metabolic routes between two compounds (Li et al.,
305 2004; Hatzimanikatis et al., 2005). This framework was applied to aromatic amino acid
306 biosynthesis to find over 400,000 theoretical biochemical pathways between chorismate
307 and phenylalanine, tyrosine, or tryptophan (Hatzimanikatis et al., 2005) and it was used
308 to explore thousands of novel linear polyketide structures (González-Lergier et al., 2005).
309 Our group has developed a database of over 600 conserved structure generalized enzyme-
310 catalyzed reactions called ReBiT (Retro-Biosynthesis Tool, http://www.retro-

311 biosynthesis.com) which accepts as input a molecular or functional group structure and
312 returns as output all 3-digit E.C. groups capable of reacting with or producing that
313 structure. The University of Minnesota Biocatalysis/Biodegradation Database (UM-
314 BBD) uses a series of generalized reaction rules to propose pathways step by step, with
315 particular emphasis on analyzing the degradation trajectories of xenobiotics (Ellis et al.,
316 2006; Fenner et al., 2008).
317
318 Typically multiple, and indeed in some cases, several thousand, metabolic routes can be
319 proposed for a given compound. How does one distinguish logical, feasible pathways
320 from frivolous, improbable ones? What metrics can be applied to judge one
321 computationally-generated pathway as superior (i.e. more likely to be functionally
322 constructed) to another? One way of narrowing the choice of pathways is to apply
323 natural precedent to filter out unlikely pathway steps. In this strategy, a large set of
324 experimentally validated enzyme-catalyzed reactions are examined for patterns of
325 structural change and a series of rules are developed to give preference to reaction steps
326 containing structural changes that follow these rules. This methodology is implemented
327 in the UM-BBD to avoid the "combinatorial explosion" that results when considering all
328 the possible pathways that any given compound can take (Fenner et al., 2008). Another
329 ranking strategy is to calculate the thermodynamic favorability of the steps and to
330 penalize pathways involving steps which are energetically unfavorable. This approach is
331 taken by the BNICE framework (Hatzimanikatis et al., 2005) using a functional group
332 contribution method (Jankowski et al., 2008) to compute the overall change in Gibbs
333 energy for each individual pathway step. A new pathway modeling tool, DESHARKY,

334 quantifies and employs metabolic burden as a metric for judging unnatural pathways and,
335 in particular, how they are connected to cellular metabolism (Rodrigo et al., 2008).
336 DESHARKY is a Monte Carlo-based algorithm that estimates the transcriptomic and
337 metabolic loads on cells expressing unnatural pathways and calculates the decrease in
338 specific growth rate as a result of these additional burdens. There are still other
339 possibilities for pathway ranking, such as the number of pathway steps taken, the known
340 substrate specificities (or lack thereof) of the enzymes involved in each pathway, or the
341 availability and diversity of homologous enzymes to test at each pathway step. One of
342 the key challenges in pathway design is scoring pathways in a robust and balanced
343 manner, and only as more non-natural pathways are designed and built will there be a
344 better understanding as to which of these metrics are relevant and useful.
345

346 **Experimental Approaches**

347 With a target compound and a proposed metabolic route to reach that compound in hand,
348 one is now ready to begin experimental implementation of that pathway. Synthetic
349 pathway construction occurs over several shades of novelty – from recreating natural
350 pathways in heterologous hosts to creating synthetic pathways that parallel natural ones
351 to building completely novel metabolic routes towards unnatural compounds from
352 multiple, ordinarily unrelated enzymes (Figure 4). Here we discuss the situations in
353 which non-natural pathways prove useful and several general strategies for creating these
354 pathways.
355

356 Through the course of evolution, nature has assembled many pathways towards several
357 useful compounds, such as the biofuel and solvent 1-butanol in *Clostridium*
358 *acetobutylicum* (Jones and Woods, 1986; Dürre et al., 2002; Lee et al., 2008), the C_5
359 terpenoid building block isopentenyl pyrophosphate (IPP) in *Saccharomyces cerevisiae*
360 (Seker et al., 2005), and the biopolymer polyhydroxybutyrate (PHB) in *Ralstonia*
361 *eutropha* (Wang and Yu, 2007). These pathways have physiological roles within their
362 native hosts; for example, the butanol pathway from acetyl-CoA in *C. acetobutylicum*
363 serves as an electron sink to regenerate NAD^+ for glycolysis while deacidifying its
364 environment (Jones and Woods, 1986). Pathways in nature are optimized through
365 evolution to accomplish their physiological objectives, yet in most cases of pathway
366 engineering, it is desired to maximize the production of a target molecule in a pathway
367 rather than to accomplish a physiological goal. Butanol production in *C. acetobutylicum*,
368 for instance, is constricted by cellular regulation tying it to pH, redox conditions, and
369 sporulation (Dürre et al., 2002; Lee et al., 2008). The transference of natural pathways
370 into heterologous hosts isolates these pathways from their regulatory elements and
371 represents a first small step towards the creation of non-natural metabolism. While
372 heterologous pathway expression is limited to only pathways found in nature, it
373 nonetheless has proven effective in enhancing product titers and/or deregulating
374 compound production for a wide array of products, including the compounds in the
375 examples above (Atsumi et al., 2007; Kang et al., 2008; Martin et al., 2003; Pitera et al.,
376 2007).
377

378 The next level of novelty in synthetic pathway construction is creating metabolic routes
379 that parallel natural pathways, typically by capitalizing on enzymatic promiscuity or
380 enzyme engineering to operate natural or near-natural pathways on non-natural
381 substrates. This pathway construction strategy allows for the biosynthesis of truly
382 unnatural compounds. Returning to the PHB example, recombinant *R. eutrophia* have
383 been shown to incorporate sulfur-containing short- and medium-chain length thioacids
384 into polythioester co-polymers (Ewering et al., 2002). The synthesis of these completely
385 unnatural polymers was made possible by taking advantage of the relatively broad
386 substrate specificity of polyhydroxyalkanoate (PHA) synthases (Hazer and Steinbüchel,
387 2007), and because of that broad substrate specificity, hundreds of different monomer
388 units of various sizes (C_3-C_{16}) and substituents have been incorporated into PHA co-
389 polymers (Steinbüchel and Valentin, 1995). Another example of parallel pathway
390 construction is the synthesis of triacetic acid lactone from acetyl-CoA by expressing an
391 engineered fatty acid synthase B from *Brevibacterium ammoniagenes* (Zha et al., 2004).
392 This multifunctional enzyme has many domains designed to catalyze the various
393 reductions and condensations necessary for fatty acid synthesis (Meurer et al., 1991). By
394 specifically inactivating the ketoacyl-reductase domain of this fatty acid synthase, the
395 enzyme could no longer use NADPH to reduce its acetyl-CoA condensation products,
396 causing them to circularize into triacetic acid lactone rather than forming linear fatty
397 acids. Finally, natural products can be synthesized by arranging whole or partial
398 pathways to form a mixed, synthetic metabolic route. For example, the theoretical yield
399 of L-glutamate was improved from 1 mol glutamate per mol glucose to 1.2 mol per mol
400 by augmenting the native *Corynebacterium glutamicum* pentose phosphate pathway with

401 a phosphoketolase from *Bifidobacterium lactis* (Chinen et al., 2007). This strategy
402 allowed for the production of acetyl-CoA without the loss of carbon caused by pyruvate
403 decarboxylation to acetyl-CoA and resulted in increased glutamate titers and
404 productivity.
405
406 One of the most promising (and challenging) strategies for building synthetic pathways is
407 *de novo* pathway construction: the creation of pathways using disparate enzymes to form
408 entirely unnatural metabolic routes towards valuable compounds. This method of
409 pathway building does not rely upon natural precedent, but rather allows one to build
410 entirely new metabolite conduits from individual enzymatic pieces. As a result, this
411 approach allows for the biosynthesis of the widest array of compounds. On the other
412 hand, this strategy is the most difficult to realize given that for a completely unnatural
413 pathway, there may not be a complete set of appropriate known enzymes in nature to
414 build it. *De novo* pathway construction illustrates the need for a more complete set of
415 enzymatic tools for use in building synthetic pathways, and frequently this strategy is
416 coupled with enzyme engineering or the exploitation of enzymatic promiscuity to
417 compensate for the absence of a natural enzyme to execute a desired conversion step.
418
419 Because of the challenge in creating functional *de novo* pathways, few examples exist.
420 However, those that are available describe the biosynthesis of a wide range of useful
421 compounds and illustrate the utility of the approach. For instance, a pathway for the
422 biosynthesis of 1,2,4-butanetriol from D-xylose and L-arabinose was assembled using
423 pentose dehydrogenases and dehydratases from *Pseudomonas fragi* and *E. coli* and

424 benzoylformate decarboxylase from *Pseudomonas putida* (Nui et al., 2003). In this case,
425 multiple decarboxylases were screened to find a promiscuous decarboxylase from *P.*
426 *putida* capable of acting on a 3-deoxy-glyceropentulosonic acid intermediate in the
427 pathway. Another example of exploiting substrate promiscuity in *de novo* pathway
428 design is in the synthesis of several higher biofuels such as 2-methyl-1-butanol,
429 isobutanol, and 2-phenylethanol from glucose in *E. coli* (Atsumi et al., 2007). Here,
430 several 2-keto-acid decarboxylases were screened to identify one from *Lactococcus lactis*
431 for use in creating alcohols from 2-ketoacids (when combined with native *E. coli* alcohol
432 dehydrogenase activity). In a third example, a synthetic pathway for the unnatural
433 aminoacid phenylglycine from phenylpyruvate was made by combining
434 hydroxymandelate synthase, hydroxymandelate oxidase, and D-(4-
435 hydroxy)phneylglycine aminotransferase activities from *Amycolatopsis orientalis*,
436 *Streptomyces coelicolor*, and *P. putida* (Müller et al., 2006). Finally, engineered
437 enzymes can be employed to create *de novo* pathways, as in the recent case of the
438 synthesis of 3-hydroxypropionic acid from alanine in *E. coli* (Liao et al., 2007). Here, a
439 lysine 2,3-aminomutase from *Porphyromonas gingivalis* (Brazeau et al., 2006) was
440 evolved to have alanine 2,3-aminomutase activity, allowing for the biosynthesis of β-
441 alanine. Combining this evolved enzyme with β-alanine aminotransferase and
442 endogenous alcohol dehydrogenase activities afforded the final 3-hydroxypropionic acid
443 product. Another very recent work utilizes engineered pyruvate decarboxylase and 2-
444 isopropylmalate synthase for the synthesis of non-natural alcohols from 2-ketoacids in *E.*
445 *coli* (Zhang et al., 2008). By engineering the enzymes responsible for elongating 2-

446 ketoacids and carrying out their decarboxylation and reduction, the production of a
447 broader array of longer-chain alcohols was enabled.
448
449 **Conclusions**
450 The design and assembly of unnatural metabolic pathways represents a young and
451 exciting field with the potential to supplement, expand upon, or even replace current
452 industrial processes for the production of fine and commodity chemicals. Synthetic
453 pathway engineering integrates many components and consequently is highly
454 interdisciplinary (Figure 1). Key issues that need to be overcome in pathway design are
455 (1) establishing a standard for generalized enzyme-catalyzed reactions, (2) capturing
456 enzyme substrate preferences in these generalized reactions, and (3) determining the
457 pathway metrics that correlate with successful pathway construction. Overcoming the
458 first two challenges will allow for the creation of the next generation of pathway design
459 tools that better account for enzyme behavior, while conquering the last challenge will
460 afford us the ability to rank and choose metabolic pathways and refine the results from
461 design tools. For experimentally implementing unnatural pathways, the central challenge
462 is the limited number of characterized enzymes for the construction of new pathways. In
463 particular, there is great demand for both promiscuous natural enzymes and engineered
464 enzymes to perform specific desired reactions.
465
466 The need for new enzymes has given rise to several theoretical frameworks for relating
467 protein sequence, structure, and function. These frameworks each address a piece of the
468 problem – energetics, active site catalysis, and protein backbone structure, etc. – but the

469 ability to routinely build whole enzymes is still in the distant future. In the meantime,
470 mimicking active sites, backbones, and protein linguistics from nature has proven fruitful
471 in creating novel proteins. Experimental evolution and chimeragenesis of enzymes are
472 standard ways of imparting unnatural properties, particularly in the absence of detailed
473 information about the protein. The power of these experimental techniques is primarily
474 limited by the size of the resulting enzyme libraries and the throughput of the screen to
475 analyze them. Computational tools such as SCHEMA (Voigt et al., 2002) and Famclash
476 (Saraf et al., 2004) can assist in focusing and enriching these libraries.
477
478 As biotechnology is increasingly relied upon as a means for chemical production,
479 progress on the creation of new enzymes and unnatural pathway design and construction
480 will flourish. These new pathways must still be expressed within a cellular context, thus
481 improving and understanding unnatural pathway efficacy at a systems level will be
482 important for shattering barriers in pathway expression and product titer. For example,
483 application of flux balance analysis (Edwards et al., 2002) can guide systems-level
484 integration of non-natural pathways with host metabolism. Furthermore, redox balancing
485 and cofactor regeneration with respect to new pathways are critical to minimize their
486 burden on the host cell (Endo and Koizumi, 2001). Systems-level functionality can also
487 be coupled with unnatural pathways, for instance in the delivery of recombinant microbes
488 to a cancerous tumor (Anderson et al., 2006). Such microbes could be engineered to
489 simultaneously produce and deliver a drug. Established and recent advances in metabolic
490 engineering, such as global transcription machinery engineering (Alper and

491 Stephanopoulos, 2007), can complement synthetic biology in this regard, leading to
492 improved performance of novel pathways.
493
494 **Acknowledgments**

495 This work was supported by the Synthetic Biology Engineering Research Center
496 (SynBERC) funded by the National Science Foundation (Grant Number 0540879). We
497 are also grateful to Effendi Leonard for helpful discussions.
498

499 **References**

500 Aharoni, A., Gaidukov, L., Tagur, S., Toker, L., Silman, I., and Tawfik, D.S. (2004).
501 Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression
502 and catalytic specialization. Proc. Natl. Acad. Sci. USA, 101, 482-487.
503
504 Alper, H. and Stephanopoulos, G. (2007). Global transcription machinery engineering: A
505 new approach for improving cellular phenotype. Metab. Eng., 9, 258-267.
506
507 Anderson, J.C., Clarke, E.J., Arkin, A.P., and Voigt, C.A. (2006). Environmentally
508 Controlled Invasion of Cancer Cells by Engineered Bacteria. J. Mol. Biol., 355, 619-627.
509
510 Asako, H., Shimizu, M., and Itoh, N. (2008). Engineering of NADPH-dependent aldo-
511 keto reductase from *Penicillium citrinum* by directed evolution to improve
512 thermostability and enantioselectivity. Appl. Microbiol. Biotechnol., 80, 805-812.
513
514 Atsumi, S., Cann, A.F., Connor, M.R., Shen, C.R., Smith, K.M., Brynildsen, M.P., Chou,
515 K.J.Y., Hanai, T., and Liao, J.C. (2008). Metabolic engineering of *Escherichia coli* for 1-
516 butanol production. Metab. Eng.. Doi:10.1016/j-ymben.2007.08.003.
517
518 Atsumi, S., Hanai, T., and Liao, J.C. (2007). Non-fermentive pathways for synthesis of
519 branched-chain higher alcohols as biofuels. Nature, 451, 86-90.
520

521 Bonomo, J., Warnecke, T., Hume, P., Marizcurrena, A., and Gill,R.T. (2006). A
522 comparative study of metabolic engineering anti-metabolite tolerance in *Escherichia coli*.
523 Metab. Eng., 8, 227-239.
524
525 Brazeau, B.J., Gort, S.J., Jessen, H.J., Andrew, A.J., and Liao, H.H. (2006). Enzymatic
526 Activation of Lysine 2,3-Aminomutase from *Porphyromonas gingivalis*. Appl. Environ.
527 Microbiol., 72, 6402-6404.
528
529 Bro C., Regenberg B., Förster J., and Nielsen, J (2006). In silico aided metabolic
530 engineering of *Saccharomyces cerevisiae* for improved bioethanol production. Metab.
531 Eng., 8, 102-111.
532
533 Capsi, R., Foerster, H., Fulcher, C.A., Hopkinson, R., Ingraham, J., Kaipa, P.,
534 Krummenacker, M., Paley, S., Pick, J., Rhee, S.Y., Tissier, C., Zhang, P., and Karp, P.D.
535 (2006). MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nuc.
536 Acids Res., 34, D511-D516.
537
538 Casqueiro, J., Bañuelos, O., Gutiérrez, S., and Martín, J.F. (2001). Metabolic
539 Engineering of the Lysine Pathway for β-Lactam Overproduction in *Penicillium*
540 *chrysogenum*. In Focus on Biotechnology, Van Broekhoven, A., Shapiro, F., and Anne,
541 J. eds. (Dordrecht, Netherlands: Kluwer Academic Publishers), pp. 147-159.
542

543 Chang, Y.Y. and Cronan, J.E., Jr. (2000). Conversion of *Escherichia coli* pyruvate
544 oxidase to an ‘alpha-ketbutyrate oxidase’. Biochem. J., 352 Pt. 3, 717-724.
545
546 Chinen, A., Kozlov, Y.I., Hara, Y., Izui, H., and Yasueda, H. (2007). Innovative
547 Metabolic Pathway Design for Efficient L-Glutamate Production by Suppressing CO_2
548 Emission. J. Biosci. Bioeng., 103, 262-269.
549
550 Dürre, P., Böhringer, M., Nakotte, S., Schaffer, S., Thormann, K., and Zickner, B.
551 (2002). Transcriptional Regulation of Solventogenesis in *Clostridium acetobutylicum*. J.
552 Mol. Microbiol. Biotechnol., 4, 295-300.
553
554 Edwards, J.S., Covert, M., and Palsson, B. (2002). Metabolic modeling of microbes: the
555 flux-balance approach. Environ. Microbiol., 4, 133-140.
556
557 Ellis, L.B.M., Roe, D., and Wackett, L.P. (2006). The University of Minnesota
558 Biocatalysis/Biodegradation Database: the first decade. Nuc. Acids Res., 34, D517-
559 D521.
560
561 Endo, T. and Koizumi, S. (2001). Microbial Conversion with Cofactor Regeneration
562 using Genetically Engineered Bacteria. Adv. Synth. Catal., 343, 521-528.
563
564 Endy, D. (2005). Foundations for engineering biology. Nature, 438, 449-453.
565

566 Ewering, C., Lütke-Eversloh, T., Luftmann, H., and Steinbüchel, A. (2002).
567 Identification of novel sulfur-containing bacterial polyesters: biosynthesis of poly(3-
568 hydroxy-S-propyl-ω-thioalkanoates containing thioether linkages in the side chains.
569 Microbiology, 148, 1397-1406.

570

571 Fenner, K., Gao, J., Kramer, S., Ellis, L., and Wackett, L. (2008). Data-driven extraction
572 of relative reasoning rules to limit combinatorial explosion in biodegradation pathway
573 prediction. Bioinformatics, 24, 2079-2085.

574

575 González-Lergier, J., Broadbelt, L.J., and Hatzimanikatis, V. (2005). Theoretical
576 Considerations and Computational Analysis of the Complexity in Polyketide Synthesis
577 Pathways. J. Am. Chem. Soc., 127, 9930-9938.

578

579 Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous
580 protein expression. Trends Biotechnol., 22, 346-353.

581

582 Handelsman, J. (2004). Metagenomics: Application of Genomics to Uncultured
583 Microorganisms. Microbiol. Mol. Biol. Rev., 68, 669-685.

584

585 Hatzimanikatis, V., Li, C., Ionita, J.A., Henry, C.S., Jankowski, M.D., and Broadbelt, L.J.
586 (2005). Exploring the diversity of complex metabolic networks. Bioinformatics, 21,
587 1603-1609.

588

589 Hazer, B. and Steinbüchel, A. (2007). Increase diversification of polyhydroxyalkanoates
590 by modification reactions for industrial and medical applications. Appl. Microbiol.
591 Biotechnol., 74, 1-12.
592
593 Hederos, S., Broo, K.S., Jakobsson, E., Kleywegt, G.J., Mannervik, B., and Baltzer, L.
594 (2004). Incorporation of a single His residue by rational design enables thiol-ester
595 hydrolysis by human glutathione transferase A1-1. Proc. Natl. Acad. Sci.USA, 101,
596 13163-13167.
597
598 Jankowski, M.D., Henry, C.S., Broadbelt, L.J., and Hatzimanikatis, V. (2008). Group
599 Contribution Method for Thermodynamic Analysis of Complex Metabolic Networks.
600 Biophys. J., 95, 1487-1499.
601
602 Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., Rothlisberger, D., Zanghellini, A.,
603 Gallaher, J.L., Betker, J.L., Tanaka, F., Barbas, C.F., III, Hilvert, D., Houk, K.N.,
604 Stoddard, B.L., and Baker, D. (2008). *De novo* Computational Design of Retro-Aldol
605 Enzymes. Science 319, 1387-1391.
606
607 Jones, D.T. and Woods, D.R. (1986). Acetone-Butanol Fermentation Revisited.
608 Microbiol. Rev., 50, 484-524.
609
610 Liao, H.H., Gokarn, R.R., Gort, S.J., Jensen, H.J., and Selifonova, O (2007). Alanine
611 2,3-aminomutase. U.S. Patent 7,309,597.

612

613 Kane, J.F. (1995). Effects of rare codon clusters on high-level expression of
614 heterologous proteins in *Escherichia coli*. Curr. Opin. Biotechnol., 6, 494-500.

615

616 Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S.,
617 Katayama, T., Araki, M., and Hirakawa, M. (2006). From genomics to chemical
618 genomics: new developments in KEGG. Nuc. Acids Res., 34: D354-D357.

619

620 Kang, Z., Wang, Q., Zhang, H., and Qi, Q. (2008). Construction of a stress-induced
621 system in *Escherichia coli* for efficient polyhydroxyalkanoates production. Appl.
622 Microbiol. Biotechnol., 79, 203-208.

623

624 Kaplan, J., and DeGrado, W.F. (2004). *De novo* design of catalytic proteins. Proc. Natl.
625 Acad. Sci. USA, 101, 11566-11570.

626

627 Keasling, J.D. (2008). Synthetic Biology for Synthetic Chemistry. ACS Chem. Biol.,
628 2008, 3, 64-76.

629

630 Kelly, R.M., Leemhuis, H., Rozeboom, H.J., van Oosterwijk, N., Dijkstra, B.W., and
631 Dijkhuizen, L. (2008). Elimination of competing hydrolysis and coupling side reactions
632 of a cyclodextrin glucanotransferase by directed evolution. Biochem. J., 413, 517-525.

633

634 Lee, S.Y., Park, J.H., Jang, S.H., Nielsen, L.K., Kim, J., and Jung, K.S. (2008).
635 Fermentive Butanol Production by Clostridia. Biotechnol. Bioeng., 101, 209-228.
636
637 Li, C., Henry, C.S., Jankowski, M.D., Ionita, J.A., Hatzimanikatis, V., and Broadbelt, L.J.
638 (2004). Computational discovery of biochemical routes to specialty chemicals. Chem.
639 Eng. Sci., 59, 5051-5060.
640
641 Lippow, S.M., Wittrup, K.D., and Tidor, B. (2007). Computational design of antibody-
642 affinity improvement beyond *in vivo* maturation. Nat. Biotechnol., 25, 1171-1176.
643
644 Liu, Y.H., Lu, F.P., Li, Y., Wang, J.L., and Gao, C. (2008). Acid stabilization of *Bacillus*
645 *licheniformis* alpha amylase through introduction of mutations. Appl. Microbiol.
646 Biotechnol., 80, 795-803.
647
648 Loose, C., Jensen, K., Rigoutsos, I., and Stephanopoulos, G. (2006). A linguistic model
649 for the rational design of antimicrobial peptides. Nature, 443, 867-869.
650
651 Machielsen, R., Leferink, N.G., Hendriks, A., Brouns, S.J., Hennemann, H.G.,
652 Daubetamann, T., and van der Oost, J. (2008). Laboratory evolution of *Pyrococcus*
653 *furiosus* alcohol dehydrogenase to improve the production of (2S,5S)-hexanediol at
654 moderate temperatures. Extremophiles, 12, 587-594.
655

656 Martin, V.J., Yoshikuni, Y., and Keasling, J.D. (2001). The *in vivo* synthesis of plant
657 sesquiterpenes by *Escherichia coli*. Biotechnol. Bioeng., 75, 497-503.
658
659 Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D., and Keasling, J.D. (2003).
660 Engineering a mevalonate pathway in *Escherichia coli* for the production of terpenoids.
661 Nat. Biotechnol., 21, 796-802.
662
663 Meinhold, P., Peters, M.W., Chen, M.M., Takahashi, K., and Arnold, F.H. (2005). Direct
664 conversion of ethane to ethanol by engineered cytochrome P450 BM3. Chembiochem, 6,
665 1765-1768.
666
667 Meinhold, P., Peters, M.W., Hartwick, A., Hernandez, A.R., and Arnold, F.H. (2006).
668 Engineering Cytochrome P450 BM3 for Terminal Alkane Hydroxylation. Adv. Synth.
669 Catal., 348, 763-772.
670
671 Mermelstein L.D., Papoutsakis E.T., Petersen D.J., and Bennett G.N. (1993). Metabolic
672 Engineering of *Clostridium acetobutylicum* ATCC 824 for Increased Solvent Production
673 by Enhancement of Acetone Fromation Enzyme Activities Using a Synthetic Acetone
674 Operon. Biotechnol. Bioeng., 42, 1053-1060.
675
676 Meurer, G., Biermann, G., Schütz, A., Harth, S., and Schweizer, E. (1991). Molecular
677 structure of the mutlifunctional fatty acid synthetase gene of *Brevibacterium*

678 *ammoniagenes*: its sequence of catalytic domains is formally consistent with a head-to-
679 tail fusion of the two yeast genes *FAS1* and *FAS2*. Mol. Gen. Genet., 232, 106-116.
680
681 Meyer, M.M., Hochrein, L., and Arnold, F.H. (2006). Structure-guided SCHEMA
682 recombination of distantly related β-lactamases. Protein Eng., Design and Selection, 19,
683 563-570.
684
685 Meyer, M.M., Silberg, J.J., Voigt, C.A., Endelman, J.B., Mayo, S.L., Wang, Z.-G., and
686 Arnold, F.H. (2003). Library analysis of SCHEMA-guided protein recombination.
687 Protein Sci., 12, 1686-1693.
688
689 Mueller-Cajar, O. and Whitney, S.M. (2008). Evolving improved *Synechococcus rubisco*
690 functional expression in *Escherichia coli*. Biochem. J., 414, 205-214.
691
692 Müller, U., van Assema, F., Gunsior, M., Orf, S., Kremer, S., Schipper, D., Wagemans,
693 A., Townsend, C.A., Sonke, T., Bovenberg, R., and Wubbolts, M. (2006). Metabolic
694 engineering of the *E. coli* phenylalanine pathways for the production of D-phenylglycine
695 (D-Phg). Metab. Eng., 8, 196-208.
696
697 Naoki, A., and Hiroshi, M. (1997). Predicting Protein Secondary Structure Using
698 Stochastic Tree Grammars. Mach. Learn., 29, 275-301.
699

700 Nui, W., Molefe, M.N., and Frost, J.W. (2003). Microbial Synthesis of the Energetic
701 Material Precursor 1,2,4-Butanetriol. J. Am. Chem. Soc., 125, 12998-12999.
702
703 Otey, C.R., Silberg, J.J., Voigt, C.A., Endelman, J.B., Bandara, G., and Arnold, F.H.
704 (2004). Functional evolution and structural conservation in chimeric cytochromes p450:
705 calibrating a structure-guided approach. Chem. Biol., 11, 309-318.
706
707 Patel, M., Dornburg, V., Hermann, B., Roes, L., Hüsing, B., Overbeek, L., Terragni, F.,
708 and Recchia, E. (2006). Medium and Long-term Opportunities and Risks of the
709 Biotechnological Production of Bulk Chemicals from Renewable Resources – The
710 Potential of White Biotechnology. Utrecht, Netherlands: Utrecht University, September
711 2006.
712
713 Pitera, D.J., Paddon, C.J., Newman, J.D., and Keasling, J.D. (2007). Balancing a
714 heterologous mevalonate pathway for improved isoprenoid production in *Escherichia*
715 *coli*. Metab. Eng., 9, 193-207.
716
717 Pleiss, J. (2006). The promise of synthetic biology. Appl. Microbiol. Biotechnol., 73,
718 735-739.
719
720 Przytycka, T., Srinivasan, R., and Rose, G.D. (2002). Recursive domains in proteins.
721 Protein Sci. 11, 409-417.
722

723 Reetz, M.T. and Carballeira, J.D. (2007). Iterative saturation mutagenesis (ISM) for
724 rapid directed evolution of functional enzymes. Nat. Protoc., 2, 891-903.
725
726 Reetz, M.T., Kahakeaw, D., and Lohmer, R. (2008). Addressing the numbers problem in
727 directed evolution. Chembiochem, 9, 1797-1804.
728
729 Rigoutsos, I. and Floratos, A. (1998). Combinatorial pattern discovery in biological
730 sequences: The TEIRESIAS algorithm [published erratum appears in Bioinformatics
731 1998;14(2):229]. Bioinformatics, 14, 55-67.
732
733 Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho.,
734 K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C.Y., Withers, S.T., Shiba, Y.,
735 Sarpong, R., and Keasling, J.D. (2006). Production of the antimalarial drug precursor
736 artemisinic acid in engineered yeast. Nature, 440, 940-943.
737
738 Rodrigo, G., Carrera, J., Prather, K.J., and Jaramillo, A. (2008). DESHARKY:
739 Automatic design of metabolic pathways for optimal cell growth. Bioinformatics, 24,
740 2554-2556.
741
742 Rothlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J.,
743 Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O., Albeck, S., Houk, K.N., Tawfik,
744 D.S., and Baker, D. (2008). Kemp elimination catalysts by computational enzyme
745 design. Nature, 453, 190-195.

746

747 Saraf, M.C., Horswill, A.R., Benkovic, S.J., and Maranas, C.D. (2004). FamClash: A
748 method for ranking the activity of engineered enzymes. Proc. Natl. Acad. Sci. USA, 101,
749 4142-4147.

750

751 Savage, D.F., Way, J., and Silver, P.A. (2008). Defossiling Fuel: How Synthetic Biology
752 Can Transform Biofuel Production. ACS Chem. Biol., 3, 13-16.

753

754 Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., and Schomburg,
755 D. (2004). BRENDA, the enzyme database: updates and major new developments. Nuc.
756 Acids Res., 32, D431-D433.

757

758 Searls, D.B. (1997). Linguistic approaches to biological sequences. Comput. Appl.
759 Biosci., 13, 333-344.

760

761 Searls, D.B. (2002). The language of genes. Nature 420, 211-217.

762

763 Seker, T., Møller, K., and Nielsen, J. (2005). Analysis of acyl CoA ester intermediates of
764 the mevalonate pathway in *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol., 67,
765 119-124.

766

767 Shi, C., Lu, X., Ma, C., Ma, Y., Fu, X., and Yu, W. (2008). Enhancing the
768 thermostability of a novel beta-agarase AgaB through directed evolution. Appl.
769 Biochem. Biotechnol., 151, 51-59.
770
771 Steinbüchel, A. and Valentin, H.E. (1995). Diversity of microbial polyhydroxyalkanoic
772 acids. FEMS Microbiol. Lett., 128, 219-228.
773
774 Stemmer, W.P. (1994a). DNA shuffling by random fragmentation and reassembly: in
775 vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. USA, 91, 10747-
776 10751.
777
778 Stemmer, W.P.C. (1994b). Rapid evolution of a protein in vitro by DNA shuffling.
779 Nature, 370, 389-391.
780
781 Tsuge, T., Hisano, T., Taguchi, S., and Doi, Y. (2003). Alteration of chain length
782 substrate specificity of *Aeromonas caviae* R-enantiomer-specific enoyl-coenzyme A
783 hydratase through site-directed mutagenesis. Appl. Environ. Microbiol., 69, 4830-4836.
784
785 Villalobos, A., Ness, J.E., Gustafsson, C., Minshill, J., and Govindarajan, S. (2006).
786 Gene Designer: a synthetic biology tool for construction artificial DNA segments. BMC
787 Bioinformatics, 7, 285.
788

789 Voigt, C.A., Martinez, C., Wang, Z.-G., Mayo, S.L., and Arnold, F.H. (2002). Protein
790 building blocks preserved by recombination. Nat. Struct. Mol. Biol., 9, 553-558.
791
792 Wang, J. and Yu, H.-Q. (2007). Biosynthesis of polyhydroxybutyrate (PHB) and
793 extracellular polymeric substances (EPS) by *Ralstonia eutropha* ATCC 17699 in batch
794 cultures. Appl. Microbiol. Biotechnol., 75, 871-878.
795
796 Werpy, T. and Petersen, G.: Top value added chemicals from biomass, Vol 1: results of
797 screening for potential candidates from sugars and synthesis gas. Oak Ridge, TN: U.S.
798 Department of Energy, August 2004.
799
800 Woods, D.R. (1995). The genetic engineering of microbial solvent production. Trends
801 biotechnol., 13, 259-264.
802
803 Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro,
804 S., Gasteiger, E., Huang, H., Lopez, R., et al. (2006). The Universal Protein Resource
805 (UniProt): an exapanding universe of protein information. Nuc. Acids Res., 34, D187-
806 D191.
807
808 Yin, E., Le, Y.L., Pei, J.J., Shao, W.L., and Yang, Q.Y. (2008). High-level expression of
809 the xylanase from *Thermomyces lanuginosus* in *Escherichia coli*. World. J. Microbiol.
810 Biotechnol., 24, 275-280.
811

812 Yoshikuni, Y., Dietrich, J.A., Nowroozi, F.F., Babbitt, P.C., and Keasling, J.D. (2008).
813 Redesigning Enzymes Based on Adaptive Evolution for Optimal Function in Synthetic
814 Metabolic Pathways. Chem. Biol., 15, 607-618.
815
816 Zanghellini, A., Jiang, L., Wollacott, A.M., Cheng, G., Meiler, J., Althoff, E.A.,
817 Rothlisberger, D., and Baker, D. (2006). New algorithms and an in silico benchmark for
818 computational enzyme design. Protein Sci 15, 2785-2794.
819
820 Zha, W., Shao, Z., Frost, J.W., and Zhao, H. (2004). Rational Pathway Engineering of
821 Type I Fatty Acid Synthase Allows the Biosynthesis of Triacetic Acid Lactone from D-
822 Glucose in Vivo. J. Am. Chem. Soc., 2004, 4534-4535.
823
824 Zhang, K., Sawaya, M.R., Eisenberg, D.S., and Liao, J.C. (2008). Expanding metabolism
825 for biosynthesis of nonnatural alcohols. Proc. Natl. Acad. Sci., 105, 20653-20658.
826

Figure Captions

Figure 1: Overall scheme for pathway creation. The creation process includes protein-level recruitment and reengineering of enzymes and pathway-level efforts to design and assemble these enzymes into an unnatural pathway.

Figure 2: Flowchart for the creation of new enzymes with experimental techniques and computational tools. New enzymes generated with these methods are examined for desired properties and either further reengineered or adapted for use in unnatural pathways.

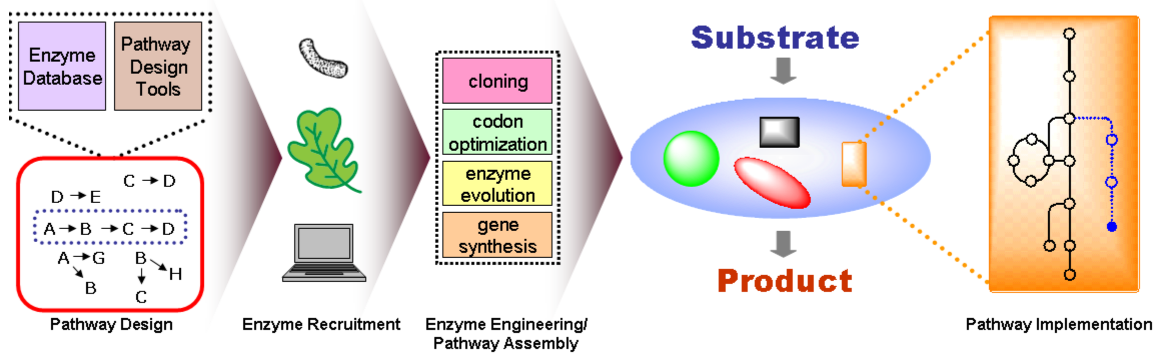
Figure 3: Generalized enzyme-catalyzed reactions for a subset of E.C. 1.1.1 alcohol dehydrogenases (3a), E.C. 4.3.1 ammonia-lyases (3b), and E.C. 2.5.1 synthases (3c). The “A” atoms present in the molecular structures are wildcards. In Figure 3a, two different methods of assigning generalized reactions, one considering only the reacting parts of the molecule (core generalized reaction) and one identifying patterns of conserved molecular structure in addition to the reacting structural elements (conserved structure generalized reaction), arrive at the same generalized reaction. In Figure 3b, the two methods arrive at different generalized reactions, illustrating the need for a generalization standard. In Figure 3c, a set of five enzymes within a three-digit E.C. class result in two different sets of generalized reactions, illustrating that the E.C. system does not necessarily correlate with reaction generalization.

Figure 4:

Strategies for synthetic pathway creation arranged in increasing degrees of departure from nature. A, B, C, D, F, α, β, γ, and Δ represent metabolites, E represents an enzyme catalyzing a reaction, and ε represents an engineered enzyme catalyzing a reaction. In (1), a natural pathway in its native host is transferred to a heterologous host, decoupling it from native regulation. This strategy is limited to the production of natural products using natural pathways. In (2), new pathways are made in parallel to natural ones through the use of promiscuous enzymes (2a), enzyme engineering (2b), or combinations of natural pathways (2c). Strategies 2a and 2b allow for the synthesis of new, non-natural products, while 2c allows for the creation of new metabolic routes between natural metabolites. Strategy 3 represents *de novo* pathway construction,

855 where individual unrelated enzymes are recruited to form entirely unnatural pathways. This can be done
856 using native enzyme activities (3a), promiscuous enzymes (3b), engineered enzymes (3c), or combinations
857 thereof.
858

859 **Figure 1.**

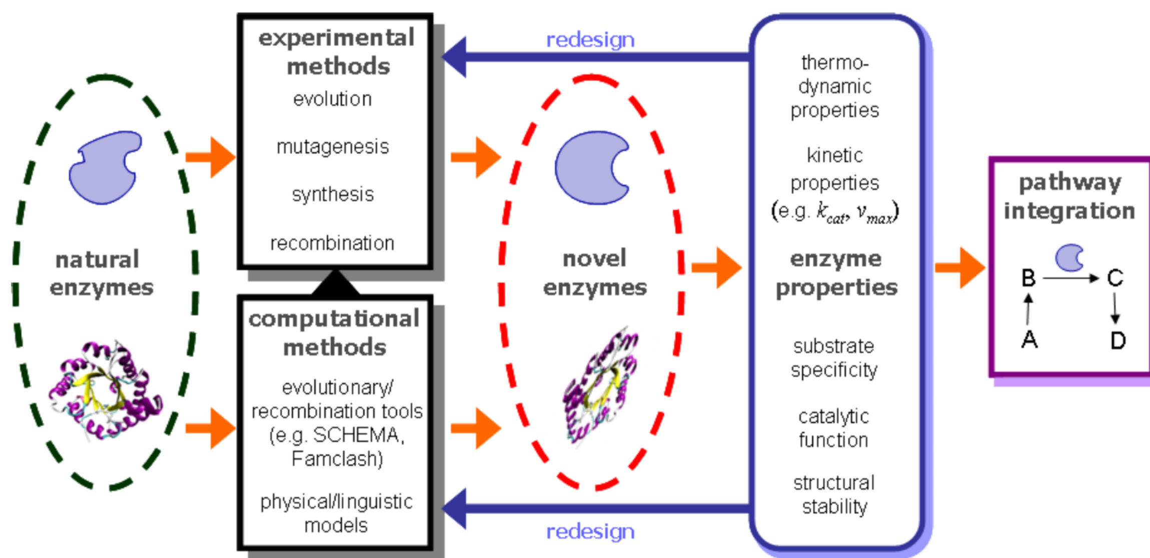


860

861

862

863 **Figure 2.**

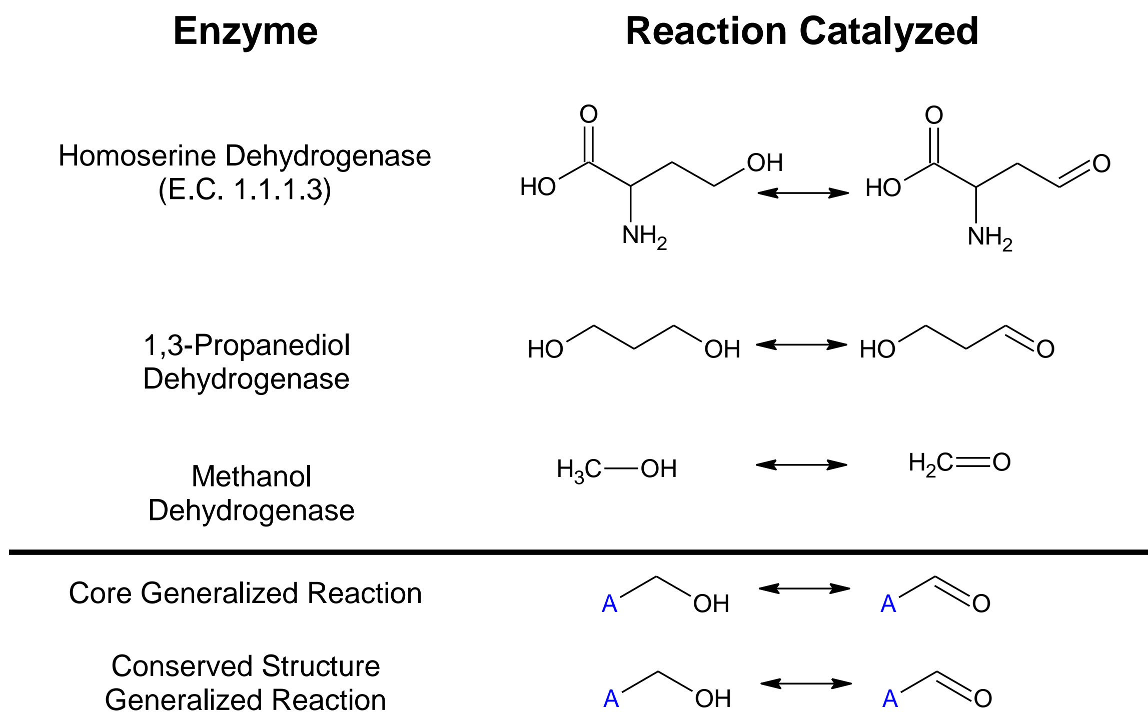


864

865

866

867 **Figure 3a.**

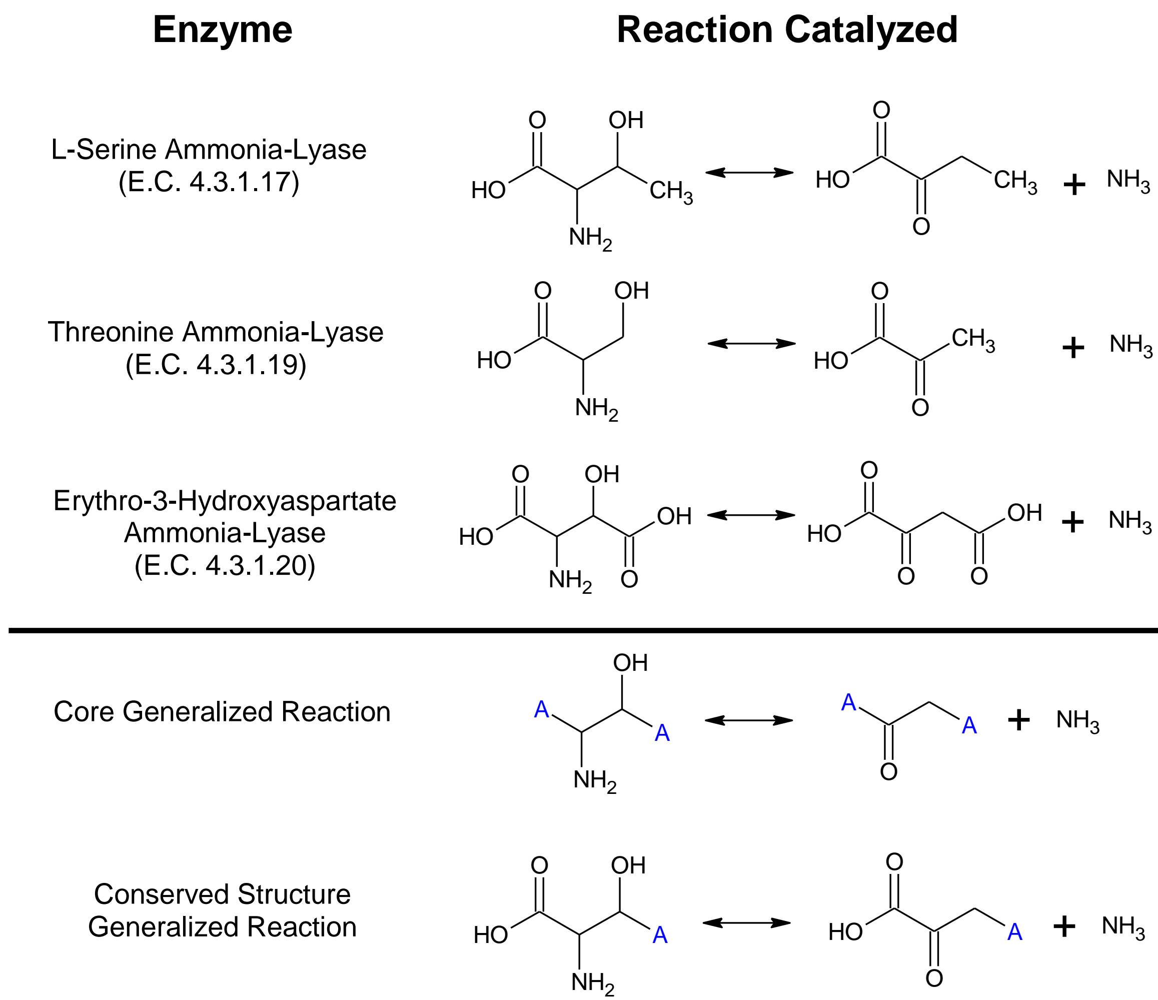


868

869

870

871 **Figure 3b.**

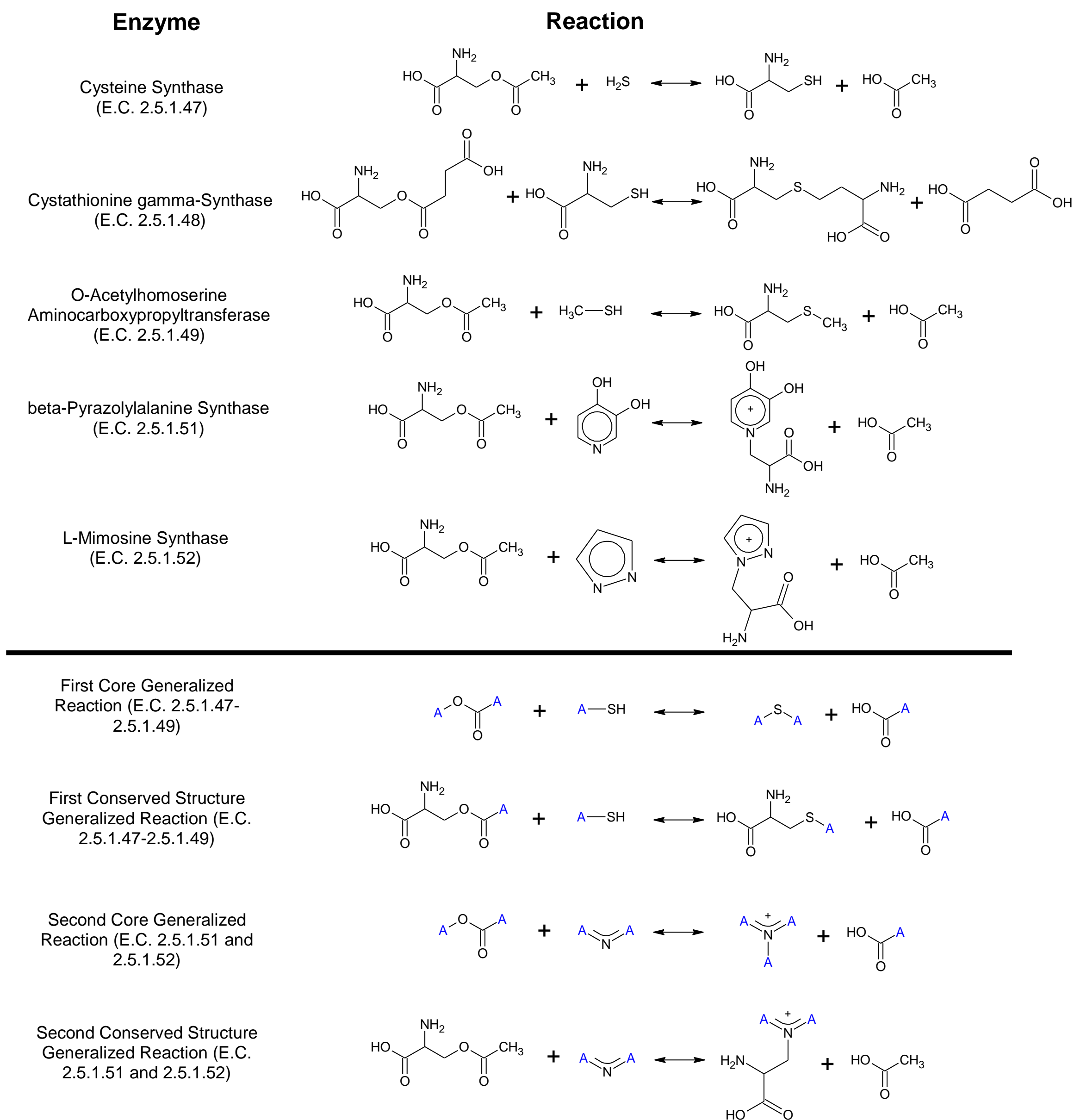


872

873

874

875 **Figure 3c.**

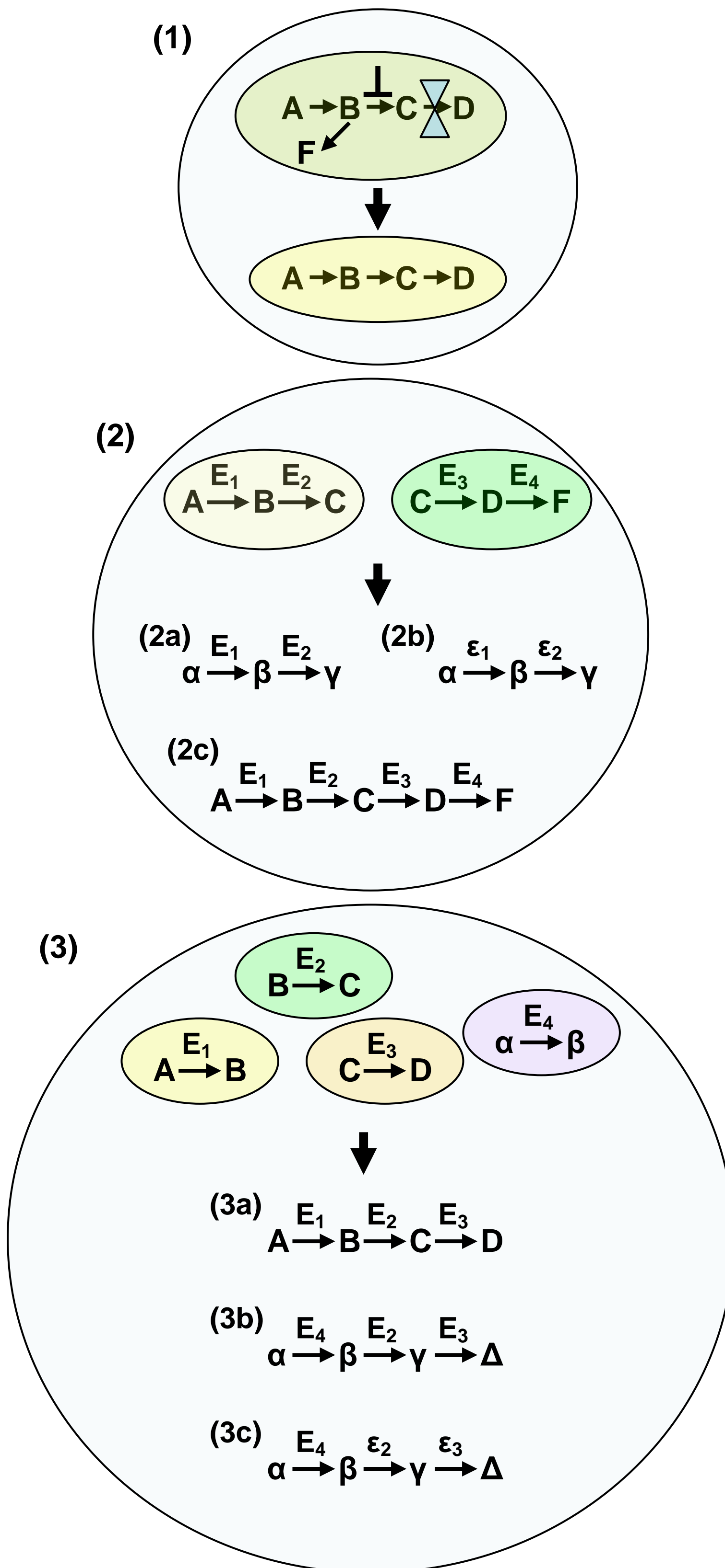


876

877

878

879 **Figure 4.**



880