A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories

Amanda K. Fisher\textsuperscript{a,b,c,1}, Benjamin G. Freedman\textsuperscript{a,1}, David R. Bevan\textsuperscript{b,c}, Ryan S. Senger\textsuperscript{a,*}

\textsuperscript{a} Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA, United States
\textsuperscript{b} Department of Biochemistry, Virginia Tech, Blacksburg, VA, United States
\textsuperscript{c} Interdisciplinary Genetics, Bioinformatics, and Computational Biology Program, Virginia Tech, Blacksburg, VA, United States

\textbf{Abstract}

Microbial cell factories (MCFs) are of considerable interest to convert low value renewable substrates to biofuels and high value chemicals. This review highlights the progress of computational models for the rational design of an MCF to produce a target bio-commodity. In particular, the rational design of an MCF involves: (i) product selection, (ii) de novo biosynthetic pathway identification (i.e., rational, heterologous, or artificial), (iii) MCF chassis selection, (iv) enzyme engineering of promiscuity to enable the formation of new products, and (v) metabolic engineering to ensure optimal use of the pathway by the MCF host. Computational tools such as (i) de novo biosynthetic pathway builders, (ii) docking, (iii) molecular dynamics (MD) and steered MD (SMD), and (iv) genome-scale metabolic flux modeling all play critical roles in the rational design of an MCF. Genome-scale metabolic flux models are of considerable use to the design process since they can reveal metabolic capabilities of MCF hosts. These can be used for host selection as well as optimizing precursors and cofactors of artificial de novo biosynthetic pathways. In addition, recent advances in genome-scale modeling have enabled the derivation of metabolic engineering strategies, which can be implemented using the genomic tools reviewed here as well.

© 2014 Fisher et al. Published by Elsevier B.V. on behalf of the Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

\section*{Contents}

1. Introduction ............................................................... 91
2. Selecting components of an MCF ...................................................... 92
   2.1. Defining the approach: native, heterologous, or artificial ......................... 92
   2.2. Selecting the MCF Chassis ..................................................... 92
   2.3. Designing an artificial de novo biosynthetic pathway ............................. 93
3. Building a functional MCF: enzyme engineering .............................................. 94
   3.1. Producing functional enzymes ................................................... 94
   3.2. Rational enzyme engineering tools ................................................ 94
4. Building a functional MCF: metabolic engineering .............................................. 96
   4.1. Experimental tools ........................................................ 96
   4.2. Computational tools ......................................................... 96
5. Conclusions ............................................................... 97
Acknowledgments .................................................................. 97
References ........................................................................ 97

\section{1. Introduction}

In traditional chemical processes, a low-value starting material is converted into a high-value product through a series of unit operations. Initial operations may concentrate or refine the starting material by
separating it from contaminants. The processed starting material is reacted with additional substrates in the presence of a catalyst, and the product of interest is separated from unreacted substrates and byproducts. Advances in catalysis and process optimization maximize single-pass conversion and profitability. Microbial cell factories (MCFs) have emerged as a revolutionary platform for combining traditional unit operations and complex multi-step catalysis into a single self-replicating microbe [1–3]. Reactors filled with billions of microbes can now replace much of the traditional chemical factory. Each cell can selectively uptake a low value substrate and use its vast metabolic network (and compartmentalization if necessary) to produce desired products. This review covers recent advances in (i) how chassis microbes are selected and engineered to serve as an MCF, (ii) how new catalytic properties are added to the metabolic network, and (iii) how the cell is engineered to use new metabolic pathways to maximize yield of a desired product. Methods are often grouped into combinatorial (i.e., evolutionary) and rational (i.e., informed design) approaches. This review specifically targets rational approaches that are informed by computational models and demonstrates how computational approaches are advancing the design of a complete, custom MCF.

2. Selecting components of an MCF

2.1. Defining the approach: native, heterologous, or artificial

Designing an MCF begins with defining the product of interest. The desired product could be a native metabolite of the chassis organism (i.e., wild-type host), or additional metabolic capabilities may be required for a chosen chassis to produce the product of interest. It is important to note that simply the presence of a biosynthesis pathway does not guarantee that a particular chassis is the optimum choice. Even if a biosynthetic pathway is already present, often metabolic and/or enzyme engineering strategies may be required to increase metabolic flux through the pathway to arrive at yields needed for industrial production. In a recent example, an MCF was created using Saccharomyces cerevisiae along with a computationally-derived metabolic engineering strategy for succinate overproduction. Even though succinate is produced naturally by wild-type S. cerevisiae, it is consumed by the TCA cycle. An engineered strain of S. cerevisiae capable of producing industrially-relevant quantities of succinate (>40-fold yield improvement over wild-type) was created by deleting the succinate dehydrogenase (responsible for succinate depletion) and the 3-phosphoglycerate dehydrogenase isoenzymes. The resulting mutant up-regulated isocitrate conversion to succinate and glyoxylate to counteract serine and glycine deficiency [4]. Additional computationally-derived metabolic engineering strategies are discussed throughout.

Another common approach to creating an MCF is to install a heterologous or artificial de novo biosynthetic pathway in a chassis organism to arrive at a new product. The desired product could be (i) native to a microbe that is difficult to culture/engineer, (ii) from a higher organism (e.g., a plant) whose industrial production is not cost effective, or (iii) non-native to all microbes and a product of artificial metabolism. In addition, the MCF has also provided a convenient way of producing new derivatives of a compound of interest. As an example, phenylpropanoids, including resveratrol, are natural plant secondary metabolites that have demonstrated therapeutic benefits and commercial value. These and more bioavailable derivatives of resveratrol were sought from an MCF. A de novo biosynthetic pathway for the formation of resveratrol in Escherichia coli was constructed using heterologous enzymes from bacteria and plants [5], and it was later expanded by the addition of a glycosyltransferase (from Bacillus), which enabled synthetic production of resveratrol glucoside derivatives (i.e., resveratrol 3-O-glucoside and resveratrol 4’-O-glucoside) in an E. coli MCF [6]. The use of enzymes here in their natural function, with natural substrates toward the production of phenylpropanoids, is an example of a heterologous biosynthetic pathway. However, the use of the glycosyltransferase to produce new compounds relies on enzyme promiscuity (i.e., the ability of an enzyme to accept multiple substrates [7,8]). It is with promiscuous enzymes that novel arrangements of enzymes can give rise to artificial de novo biosynthetic pathways that allow MCFs to produce new chemicals. While there are many published accounts, some examples include the production of: (i) isobutanol [9,10], (ii) hydrocarbons [11], (iii) styrene [12], (iv) 3-hydroxybutyric acid [13], (v) native silk protein [14], and (vi) isoprenoids [15,16]. Most naturally occurring enzymes maintain a spectrum of substrate promiscuity to maximize evolutionary fitness and that promiscuity can be engineered [17]. Computational tools for enzyme engineering along with tools for artificial pathway synthesis and assembly are discussed below. However, first, the guidelines for selecting/engineering an optimal MCF chassis (i.e., host organism) are discussed.

2.2. Selecting the MCF Chassis

The choice of MCF chassis can vary greatly and is generally made according to: (i) the difficulty of metabolic engineering needed (and available toolsets), (ii) the nature and toxicity of the product, and (iii) the metabolic requirements (i.e., pathways, precursors, and cofactors) needed to produce the product. A list of common MCF chasses and their advantages/disadvantages is given in Table 1. While E. coli and yeast still dominate as popular chasses due to their well-developed genomic tools, this is expected to change. In the near-term, new genomic toolsets will allow the MCF chassis to take advantage of biodiversity, natural capabilities, and synergies. Ultimately, the minimal

<table>
<thead>
<tr>
<th>Organism*</th>
<th>Advantages/disadvantages of chassis</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clostridium sp.</td>
<td>Sporulating obligate anaerobes; gene knockout and over-expression tools available but can be very difficult to grow and engineer; ability to use a wide variety of complex substrates including lignocellulose and CO₂ to sustain growth</td>
<td>[3,20–22]</td>
</tr>
<tr>
<td>Corynebacterium glutamicum</td>
<td>A well-established industrial workhorse; genetic tools are available</td>
<td>[23]</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>Most well-characterized prokaryote; already used broadly in industry; genomic tools and systems biology datasets are widely available</td>
<td>[24]</td>
</tr>
<tr>
<td>Myxococcus xanthus</td>
<td>Effective host for myxobacterial, polyketide, and delftacteobacterium synthesis pathways</td>
<td>[25]</td>
</tr>
<tr>
<td>Pseudomonas putida</td>
<td>Ease of cultivation and well established transformation techniques; capable of rapid growth, homologous recombination, and post-translational modifications; swapable genetic elements with E. coli</td>
<td>[26,27]</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>Well characterized and widely used in industry; genomic tools and systems biology datasets are widely available. Difficulties with anaerobic fermentation</td>
<td>[4]</td>
</tr>
<tr>
<td>Streptomyces sp.</td>
<td>Synthesis of polyketide derivatives</td>
<td>[1,28]</td>
</tr>
<tr>
<td>Chinese Hamster Ovary (CHO)</td>
<td>Production of sialylated and glycosylated proteins, recombinant human proteins, and high value pharmaceutical therapeutics; large production and cultivation costs</td>
<td>[29]</td>
</tr>
<tr>
<td>Taxus plant cells</td>
<td>Effective synthesis of toxic secondary plant metabolites; slow growing and low yields</td>
<td>[30]</td>
</tr>
</tbody>
</table>

* CHO and plant cells are included for comparison with traditional MCFs.
1. Metabolic resources
For high levels of expression, a target pathway of an MCF requires abundant precursors and cofactors, including ATP and NAD(P)H. Strategies for allocating reducing power to biosynthetic pathways have been reviewed recently [31], and genome-scale metabolic flux modeling represents a computational method of exploring metabolic capabilities [32,33]. While the topic of genome-scale modeling is discussed in more detail below, it can be used to assess biodiversity to select an appropriate MCF host. New tools have enabled the evaluation of multiple potential MCF hosts to see which best accommodates a biosynthetic pathway of interest. In particular, automated methods of genome-scale model building, such as the Model SEED [34] and Path2Models [35], have produced models for thousands of potential MCF hosts. Database resources such as MetaNetX [36] enable the direct incorporation of new de novo biosynthetic pathways into existing genome-scale models, where their interactions and use by the metabolic network can be studied.

2. Minimization of metabolic adjustment
When a microbe is metabolically engineered, its metabolic network use is usually compromised, adhering to the form and function of the wild-type [37] until evolutionary pressure enables optimality [38]. There may be advantages to choosing an MCF host that requires minimal metabolic perturbations and further adjustment through evolution; although, this topic remains under investigation. Some strains considered as an MCF host may require specialized metabolic capabilities such as: (i) photosynthesis, (ii) CO₂ fixation using the Wood–Ljungdahl pathway, (iii) lignocellulosic substrate utilization through a cellulose complex, (iv) N₂ fixation by a nitrogenase, or (v) methanogenesis. Currently, to take advantage of these pathways often requires adopting the host as the expression platform. An example is Clostridium ljungdahlii, a native host of the Wood–Ljungdahl pathway, which was recently engineered with genes for 1-butanol synthesis [20]. However, success was also recently seen in transferring the Wood–Ljungdahl pathway into a non-native, but related, Clostridium acetobutylicum host to enable production of biofuels and specialty chemicals from CO₂ [21]. Computational tools for MCF host selection for different de novo biosynthetic pathways are emerging. Two recent toolsets involve the use of agent-based modeling and global sensitivity analysis to identify critical components [39] and genome-scale metabolic flux modeling to identify compatible metabolic networks and medium formulations to maximize expression of a biosynthetic pathway.

3. Secretion of products
It is known that Bacillus spp. are preferred for protein secretion over E. coli; however, considerable research is ongoing to identify additional protein secretion hosts [40], including halophiles that enhance solubility [41] as well as Streptomyces [42] and yeast [43] (among others) with genetically encoded capabilities for post-translational modifications. Of course, the type of product can influence the MCF host as well. For example, production/secretion of fatty acids has been optimized in E. coli [44] and cyanobacteria [45] while the Gram-positive soil bacterium Corynebacterium glutamicum has demonstrated effective secretion of amino acids and other biobased fuels, chemicals, and materials [46].

4. Toxicity of pathway products or intermediates
Many over-produced metabolic products are toxic to the host organism, which is a serious consideration when choosing an MCF host for a particular product [47,48]. Species naturally tolerant to alcohols tend to maintain membrane fluidity under stress and yield increased amounts of osmoprotectants, redox, and ion regulators. Other species, such as Clostridium and E. coli, have been rationally or combinatorially engineered to increase alcohol tolerance 8-fold over wild-type levels [49–51]. Many synthetic pathways employ methods of preventing the accumulation of harmful intermediates using gene regulation. The production of fatty acids for biodiesel in yeast was made more effective and showed increased genetic stability by a synthetic pathway that incorporated regulation of the fabD gene by available fatty acids [52]. Further, Dahl et al. [53] increased production of the isoprenoid aromadiene using gene promoters that respond to the presence of farnesyl pyrophosphate and prevent it from rising to inhibitory concentrations.

5. Genomic toolsets and cultivation considerations
E. coli is a popular MCF chassis due to the relative ease of genetic manipulations; however, rapid developments have been seen in toolsets for other organisms. Furthermore, exonuclease-based “Gibson” cloning strategies [54] have greatly enabled the ease and reliability of DNA plasmid assembly. For culture growth, the choice of anaerobic fermentation versus aerobic cultivation must also be addressed. Aerobic organisms in culture must be controlled for oxygen limitation, excess heat generation, and a rapid growth rate that often results in high biomass conversion and low yields for secondary products [55]. The general advantages of anaerobic fermentation include simplified fermenter mass transfer considerations, higher yield of product over biomass, and a non-O₂ terminal electron acceptor that enables production of several different biofuels and chemicals from pathways requiring significant reducing power [31,56].

6. Proper enzyme folding and function
Expression of active heterologous enzymes in an MCF chassis is dependent on a number of factors including proper transcript reading, availability of necessary chaperonins, and proper post–translation modifications [57–59]. Eukaryotic systems such as (i) the yeast Pichia pastoris [60], (ii) Chinese Hamster Ovary (CHO) and human cell lines [61,62], (iii) highly versatile baculovirus-based insect cell lines [63], and (iv) Taxis and other plant cells [30] enable host-dependent post-translational modifications that are required for activity of some enzymes. Of course there is a significant interest in transferring these capabilities to other microbes. Recent advances using combinatorial libraries, codon optimization, and shotgun proteomics have enabled N-linked glycosylation and proper folding of AcrA and IgG in E. coli [64]. While these considerations are important, there is a critical need to quickly determine whether heterologous mRNAs are properly translated in an MCF. A translation-coupling cassette has been developed to aid troubleshooting by quickly determining whether large multi-domain enzymes are translated in MCF hosts [65].

2.3. Designing an artificial de novo biosynthetic pathway
The concept of an artificial de novo biosynthetic pathway design involves the re-arrangement of characterized enzymes and the reliance on enzyme promiscuity to enable the production of new products. This is also an area of much ongoing research activity. In some cases, familiarity with a heterologous pathway that produces a desired product naturally can provide a means to link the chassis metabolism to the installed pathway. Such a “plug-in” heterologous pathway was recently used with the chassis microbe C. glutamicum to produce an MCF capable of making the chemical chaperone ectoine. The responsible gene cluster was taken from the natural ectoine-producer Pseudomonas stutzeri and uncoupled from its normal expression dependence on high-salinity surroundings by placing it under control of the tuf gene promoter in C. glutamicum. Because the tuf gene itself is an elongation factor, the created MCF produced ectoine consistently, without the need for a corrosive high-salinity medium [23].

In several cases, a desired product will not exist in a heterologous pathway or more economical routes to that product will be desired. Here, artificial de novo biosynthetic pathways based on known reactions and enzyme promiscuity are explored. Valuable resources for
biosynthetic pathway synthesis are listed and discussed in Table 2. They include truly novel approaches such as the Biochemical Network Integrated Computational Explorer (BNICE) [66–68], which encodes enzymes as computational rules that target functional groups of compounds and catalyze chemical reactions. Recently, the BNICE algorithm was used in conjunction with docking studies to test a method for generating and screening novel pathways, using the production of 1-butanol from pyruvate in C. acetobutylicum as an example. Docking studies, which will be discussed in a later section, were used following pathway prediction to evaluate whether the proposed substrate-enzyme pairs would be viable. Nine novel biosynthetic pathways were predicted and evaluated in silico to be thermodynamically feasible and metabolically favorable [69]. When enzyme promiscuity is assumed, novel and economical pathways between substrate and product can be located. Further optimization can even look for the use of particular precursors and/or cofactors preferred by the metabolic network. This approach provides possibilities that must be further explored through enzyme and metabolic engineering, which are the next two topics of this review.

### 3. Building a functional MCF: enzyme engineering

#### 3.1. Producing functional enzymes

With informed MCF chassis selection, a functional biosynthetic pathway is needed to produce the product of interest. In the case of an installed heterologous or artificial de novo biosynthetic pathway, deficiencies may exist in the physical and chemical structure of the enzyme(s) that can affect the catalytic rate, stability, specificity, and cofactor requirements to such an extent to render the pathway inoperable.

The ratio of the catalytic to Michaelis constants ($k_{cat}/K_m$) can be readily measured and is often used to determine the effectiveness of an enzymatic reaction, accounting for both the dependence of reaction rate on the substrate concentration and the intrinsic rate of conversion of the substrate to product. This ratio is affected by adjustments in temperature, pH, solvent, choice of substrate, and mutations to the enzyme structure. Thus, in the case of de novo biosynthetic pathways, especially artificial pathways relying on enzyme promiscuity, enzyme engineering may be required to produce a fully functional pathway. Methods and applications of evolutionary approaches to enzyme engineering have been reviewed extensively [85]. Here, we examine the rational design-driven enzyme engineering strategies and outline some informed combinatorial strategies that exist to produce a functional de novo biosynthetic pathway capable of producing a novel product in an MCF.

A standard toolbox of techniques related to rational enzyme engineering exists. However, many new tools are under development, and recent expansions of computing (and supercomputing) power and resources are now enabling tremendous growth in the field. If the crystal structure of an enzyme is available and the residues that form the active site (or catalytic residues) are known, standard methods can generate mutant libraries based on rational knowledge. This is an informed combinatorial approach that involves random mutation and screening, and these methods are discussed elsewhere [86,87]. In a directed approach, one amino acid is replaced by another based on prior knowledge of the structure or function of the enzyme. If structure analysis indicates region(s) of an enzyme vital for substrate conversion, cofactor binding, or thermostability, a saturation mutagenesis approach can yield significant information if little is known about specific amino acid substitutions. In this technique, mutagenic oligonucleotides are used with degenerate or partially degenerate three-base-pair substitutions which generate mutant libraries expressing all natural amino acids simultaneously and evenly at a given mutation site [88,89]. Once a fully representative library is generated, a directed evolution or high-throughput screening method is used to identify the effect of the particular mutations. Screening techniques potentially measure (i) the activity of mutants by measuring substrate/product levels directly, (ii) absorbance caused by cofactor usage, (iii) production of optically active compounds, or (iv) auxotrophy resolution [86, 90–92].

#### 3.2. Rational enzyme engineering tools

The two most common rationally-based enzyme engineering approaches are (i) rational design and (ii) rational redesign. Rational design is not as commonly used as evolutionary approaches, and it is unique in its ability to completely design new (i.e., never before seen in nature) functional enzymes. This is done through computationally driven design using the Rosetta suite of programs [93] to design a minimal active site given (i) the transition state for a reaction to be catalyzed and (ii) a stable scaffold to support the active site. Thus far, enzymes designed in this way have had sub-optimal catalytic rates [94]. It has recently been proposed that by focusing on only one critical aspect of catalysis during rational design, the catalytic rate of the engineered enzyme should approach that of natural enzymes that...
catalyze similar reactions. A serine hydrolase was recently developed possessing comparable activity to native enzymes through a focus on obtaining the correct serine-containing catalytic triad design. The concept of synthesizing rationally designed enzymes for MCFs based on a “bottom-up” approach from computational design represents a significant step forward for generating novel products from MCFs. However, this application of rational design in MCFs remains years away, and immediate applications are more poised to focus on rational redesign methods to engineer enzyme promiscuity. In rational redesign, the natural catalytic activity of an existing enzyme is altered through rational selection of mutations. The selection process is also computationally driven, and specific aspects are discussed below. In addition, a set of useful computational tools is described in Table 3, and these tools are referenced in the following sections.

1. Static analysis

The most direct way to rationally select mutations for enzyme redesign, which requires the least amount of information about the target enzyme, is by sequence comparison. In one example, it was desired that a glucose oxidase be constructed to have the stability of its homolog from Aspergillus niger and the catalytic activity of its homolog from Penicillium amagasakiense to better carry out glucose oxidation in industrial applications. By comparing the amino acid sequences of these two homologs, 15 residues in the active site of the more stable homolog were rationally supplanted with the residues from the more catalytically active homolog. Separately and then combinatorially, the 6 most heterogeneous residues among the glucose oxidase family in the active site were subjected to random mutagenesis. High-throughput directed evolution techniques determined that the most stable and most active mutants contained mutations both rationally selected and acquired randomly. This combinatorial approach of both rational redesign and directed evolution produced four mutant enzymes with slightly improved stability and 3 to 4-fold increase in specificity. One mutant showed a 4.5-fold improvement in catalytic rate over the homolog from A. niger and a slight improvement in catalytic rate over the homolog from P. amagasakiense.

Beyond sequence information, three-dimensional crystal structure information (if available) can also be compared among homologs through overlay analysis of two protein crystal structures and root mean square deviation (RMSD) between configurations of bound substrates. Mannosyl binding was rationally redesign in the glycosidase endo-β-1,4-mannanase from Cellulomonas fimii through structural comparison between it and its homolog, another endo-β-1,4-mannanase from Cellvibrio japonicus. An important phenylalanine was substituted with an alanine (F325A), and the resulting space created was used to accommodate an arginine substitution (A323R).

These rationally selected residues were redesigned to mimic the C. japonicus enzyme structure, and this resulted in enzymatic activity similar to that of the C. japonicus enzyme.

In some cases, a three-dimensional structure of the enzyme of interest may not be available. Utilizing the known structures of enzymes with comparable sequences, a homology model of the protein of interest can be constructed. Several programs and online web servers provide homology modeling services as well as tools useful in evaluating the quality of the produced homology model (102,103,105,106). A homology model of the extremely thermostable homolog of E. coli penicillin acylase from Thermus thermophilus was created in order to investigate both its thermostability and substrate specificity. To exchange the preference of T. thermophilus penicillin acylase for penicillin K substrates with a preference for penicillin G substrates residues were rationally selected for mutation to mimic the aromatic binding site of the E. coli penicillin acylase.

Several single and combinatorial mutations were investigated in a low-throughput format, arriving at a method for improving the catalytic efficiency for penicillin G by T. thermophilus penicillin acylase by up to 6.6-fold with the single L24F rationally selected mutation. Beyond comparison among enzyme structures, three-dimensional enzyme models can also be used for more computationally intensive analysis. The most intuitive of these analyses are docking studies. Three-dimensional structures of natural or non-natural ligands are allowed to adjust their conformations in order to best fit themselves within a binding pocket of a protein three-dimensional structure. Docking studies can be conducted and visualized with a choice of computer programs, such as Autodock and DOCK (97,100,111). Both enzyme redesign and drug design can benefit from the insight provided by docking studies. Docking was recently used to study the active site interactions taking place between newly synthesized HIV-1 inhibitors thiazolidin-4-ones and the non-nucleoside binding site of a HIV-1 reverse transcriptase. The tested compounds had been verified to inhibit HIV-1 replication, and the docking studies elucidated their methods of inhibition. By collecting such information about docking poses, further analogs of inhibitors or activators can be designed (112). These techniques have direct applicability to engineering enzyme promiscuity to enable conversion of new substrates to desired products.

2. Dynamic analysis

A weakness of molecular docking studies is that the protein usually is held static during the analysis. While this is computationally advantageous, accuracy may be sacrificed. Another, more computationally intensive method called molecular dynamics (MD), requiring supercomputing resources and the ability to learn and implement scripting languages, allows both enzyme and ligand(s) to adjust their conformations in the complex. MD studies simulate proteins behaving according to physical laws defined in user-supplied force field, such as those provided by the AMBER suite of programs.
or the CHARMM development team [98], and usually span a single nanosecond up to several microseconds of simulation time. Many different force fields and programs for conducting molecular dynamics have been published and refined, each with different strengths, such as having very accurate parameters for modeling lipid-membrane systems, and weaknesses, such as a propensity to over stabilize alpha-helices [96,98,101,104]. The data obtained from MD can be analyzed in several different ways and can provide significant insight into ways to modify enzyme behavior. These methods are introduced and discussed next.

Among the different measures for interpreting the results of MD, the RMSD analysis is a simple measure of the change in molecular conformation as compared to a reference conformation. RMSD analysis can be used to (i) quantify the change in ligand conformation among different docking studies, (ii) determine when a protein has equilibrated during MD simulations, and (iii) sort protein conformations throughout MD simulations into clusters for further analysis. The percent amount of secondary structures formed throughout MD simulations can also be measured quantitatively using the industry standard DSSP algorithm [113] for assigning secondary structure to each residue in a protein; this provides a timeline of structural changes in the enzyme during the simulation. The amount of solvent-accessible surface area on the enzyme is also quantifiable throughout the simulation, allowing the identification of when and under what circumstances quaternary structures are formed or broken. Detecting and quantifying the bond pairs, frequency, and duration of hydrogen bonds throughout an MD simulation can provide insight into ligand stability in the active site.

Steered MD (SMD) can also be conducted to generate information about enzyme structure organization and how ligands enter or leave the active site [114]. A directional force is applied to an atom or to the center of mass of a group of atoms in order to calculate potential of mean force and binding affinities in silico. Qualitatively analyzing SMD trajectories can provide clues on how to increase or decrease the amount of steric clashes between the proposed ligand and the active site. MD and SMD simulations have recently been used to encourage substrate specificity of the lipase from Bacillus thermocatenulatus (BTL2). C4 and C8 triacylglycerols were dynamically modeled in the BTL2 active site and allowed to equilibrate. The ligands were then pulled from the active site using SMD to estimate the potential of mean force. In order to encourage specificity for C4 triacylglycerols, the three-dimensional structure of BTL2 used in the study was altered to reflect the proposed mutations that would decrease the volume of the active site cleft while preserving hydrophobicity. The mutation L360F made the most progress towards decreasing activity with C8 triacylglycerols while increasing activity with C4 triacylglycerols. This change in specificity was confirmed both in silico and in vitro [115].

4.1. Experimental tools

Following the installation of de novo biosynthetic pathway(s), metabolic engineering is used to ensure their use by the host and enhance productivity. Often, metabolic engineering strategies designed to increase production of a desired product involve over-expression of the synthesis pathway. This strategy has been found to be effective in certain cases, but it often fails when used alone. Enzyme copy number is optimized by evolution based on the availability substrates and certain cases, but it often fails when used alone. Enzyme copy number

4.2. Computational tools

A major focus in genome-scale metabolic flux modeling is the rational design of metabolic engineering strategies. While multiple algorithms have been published, OptKnock [125] was the first to demonstrate widespread success. It effectively tied product formation to cellular growth through selected reaction knockouts. The gene knockout was the solely required experimental tool to implement OptKnock derived strategies. This was updated by OptForce [126], which enabled both up- and down-regulation strategies to optimize product formation. The tool OptORF [127] takes transcriptional regulatory networks into consideration in designing strategies, and k-OptForce [128] is the latest development to incorporate kinetic constants to enhance metabolic prediction and improve metabolic engineering strategies. The approach Flux Balance Analysis with Flux Ratios (FBrAtio) [129,130] is another attempt to effectively model wild-type metabolism and design metabolic engineering strategies. FBrAtio considers how multiple enzymes compete for the same limited metabolite pool. Factors such as enzyme
and cofactor availability as well as reaction thermodynamics determine how metabolic flux is distributed at critical branch points (i.e., nodes) in the metabolic network. FBratio can be used to locate the critical nodes in a metabolic network and design how available experimental tools (i.e., over-expression, knockout, knock-down) can be implemented to enhance expression of a product-forming pathway elsewhere in the metabolic network.

5. Conclusions

MCFs can accommodate specialized metabolic pathways that allow the conversion of renewable substrates into valuable products. Design of an MCF is a multifaceted optimization problem that consists of several challenges in enzyme and metabolic engineering. While research is ongoing on all fronts, it is clear that MCF design originates with the critical choice of product, which will ultimately be the most important factor in determining profitability. Multiple tools exist for designing either heterologous or artificial de novo biosynthetic pathway(s) to a product of interest, and the nature of the product and process is critical in host selection for an MCF.

Computational modeling can now be used to determine if metabolic precursors and cofactors are available in certain hosts, given a biosynthetic pathway, and host selection is strongly tied to: (i) the substrate(s) to be converted, (ii) product secretion machinery required, (iii) genomic tools available for metabolic engineering, (iv) toxicity of the product, (v) cell growth considerations, and (vi) biosynthetic pathway activity in the host. Following these selections, enzyme and metabolic engineering are required to develop a productive MCF. Metabolic engineering can improve flux through a chosen biosynthetic pathway, but will eventually be stymied by enzymatic limitations. Enzyme engineering, through rational design, is the most important factor in determining productivity. While research is ongoing on all fronts, it is clear that MCF design allows the conversion of renewable substrates into valuable products.

5. Conclusions

MCFs can accommodate specialized metabolic pathways that allow the conversion of renewable substrates into valuable products. Design of an MCF is a multifaceted optimization problem that consists of several challenges in enzyme and metabolic engineering. While research is ongoing on all fronts, it is clear that MCF design originates with the critical choice of product, which will ultimately be the most important factor in determining profitability. Multiple tools exist for designing either heterologous or artificial de novo biosynthetic pathway(s) to a product of interest, and the nature of the product and process is critical in host selection for an MCF.

Computational modeling can now be used to determine if metabolic precursors and cofactors are available in certain hosts, given a biosynthetic pathway, and host selection is strongly tied to: (i) the substrate(s) to be converted, (ii) product secretion machinery required, (iii) genomic tools available for metabolic engineering, (iv) toxicity of the product, (v) cell growth considerations, and (vi) biosynthetic pathway activity in the host. Following these selections, enzyme and metabolic engineering are required to develop a productive MCF. Metabolic engineering can improve flux through a chosen biosynthetic pathway, but will eventually be stymied by enzymatic limitations. Enzyme engineering, through rational design, is the most important factor in determining productivity. While research is ongoing on all fronts, it is clear that MCF design allows the conversion of renewable substrates into valuable products.

98


Ferreira et al. / Computational and Structural Biotechnology Journal 11 (2014) 91–99


Becerra et al. / Computational and Structural Biotechnology Journal 11 (2014) 91–99


