COMPUTATIONAL
ANDSTRUCTURAL
BIOTECHNOLOGY
J O U R N A L

Review

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

Amanda K. Fisher [a,b,c,1], Benjamin G. Freedman [a,1], David R. Bevan [b,c], Ryan S. Senger [a,*]

[a] Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA, United States
[b] Department of Biochemistry, Virginia Tech, Blacksburg, VA, United States
[c] Interdisciplinary Genetics, Bioinformatics, and Computational Biology Program, Virginia Tech, Blacksburg, VA, United States

A R T I C L E   I N F O

A B S T R A C T

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.

Contents

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

* Corresponding author at: 301C HABB1, Blacksburg, VA 24061, United States. Tel.: +1 540 231 9501.
E-mail address: senger@vt.edu (R.S. Senger).
[1] These authors contributed equally to this work.

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 1
Common MCF chasses.

Organism[a]	Advantages/disadvantages of chassis	References
Clostridium sp.	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 CO_2 to sustain growth	[3,20–22]
Corynebacterium glutamicum	A well-established industrial workhorse; genetic tools are available	[23]
Escherichia coli	Most well-characterized prokaryote; already used broadly in industry; genomic tools and systems biology datasets are widely available	[24]
Myxococcus xanthus	Effective host for myxobacterial, polyketide, and deltaproteobacterium synthesis pathways	[25]
Pseudomonas putida	Ease of cultivation and well established transformation techniques; capable of rapid growth, homologous recombination, and post-translational modifications; swappable genetic elements with *E. coli*	[26,27]
Saccharomyces cerevisiae	Well characterized and widely used in industry; genomic tools and systems biology datasets are widely available. Difficulties with anaerobic fermentation	[4]
Streptomyces sp.	Synthesis of polyketide derivatives	[1,28]
Chinese Hamster Ovary (CHO)	Production of sialylated and glycosylated proteins, recombinant human proteins, and high value pharmaceutical therapeutics; large production and cultivation costs	[29]
Taxus plant cells	Effective synthesis of toxic secondary plant metabolites; slow growing and low yields	[30]

[a] CHO and plant cells are included for comparison with traditional MCFs.

cell [18], which will eventually be programmed with desired capabilities [19], has the potential to dominate. Given the current state of technology, six criteria are given consideration when choosing an MCF chassis and are discussed below.

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) CO_2 fixation using the Wood–Ljungdahl pathway, (iii) lignocellulosic substrate utilization through a cellulosome complex, (iv) N_2 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 CO_2 [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 fadD gene by available fatty acids [52]. Further, Dahl et al. [53] increased production of the isoprenoid amorphadiene 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, exonucleotide-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-O_2 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-translational 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) Taxus 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

Table 2
Tools for designing de novo biosynthetic pathways.

De novo pathway prediction program	Function	References
Biochemical Network Integrated Computational Explorer (BNICE)	Formulation of enzyme rules based on EC classifications; assumes enzyme promiscuity to develop novel pathways	[67,68,70]
BRENDA	Database of enzymatic information	[71]
DESHARKY	Monte Carlo-based pathway design algorithm based on a enzymatic reaction database and linking to host metabolism	[72]
From Metabolite to Metabolite (FMM)	Reconstruction of metabolic pathways based on KEGG mappings	[73]
L1SVM, L2SVM, BASELINE	Use of chemical fingerprints to generate reaction-filling framework to predict likeliness of reaction occurring between compounds	[74]
META	Predicts sites on molecules prone to enzyme catalyzed reactions	[75]
Metabolic Route Search/Design (MRSD)	Utilizes metabolic network of an organism to find all known pathways between two defined metabolites	[76]
Metabolic tinker	Large-scope heuristic search strategy for thermodynamically feasible paths between two compounds	[77]
METEOR	Metabolic fate of a chemical is calculated given known enzymatic capabilities	[78]
Minnesota Biocatalysts/Biodegradation Database (UM-BBD)	Predicts degradation pathways for environmental contaminants	[79]
PathPred	Predicts pathways based on chemical reaction group pattern matching and KEGG reactant pair library for xenobiotics and secondary metabolites	[80]
Rahnuma	Prediction, analysis, and comparison of metabolic networks focusing on phylogenetic differences between organisms	[81]
Retro-Biosynthesis Tool (ReBit)	Query of enzyme catalyzed reactions by molecular structure with links to protein databases	[82,83]
XTMS	Provides ranked pathways for use with an MCF based on Extended Metabolic Space allowed by Gibbs free energies, flux balance, enzyme sequence annotations, and toxicity of metabolites	[84]

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 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 $\left(\frac{k_{cat}}{K_M}\right)$ 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

Table 3
Computational tools used in rational enzyme engineering.

Tool	Use	Reference
AMBER	Popular force field for conducting MD simulations	[96]
Autodock	Several different ways of conducting docking studies and visualizing the results	[97]
CHARMM	Popular force field for conducting MD simulations	[98]
Chimera	Visualization and editing tools for molecular structures. Sequence alignment	[99]
DOCK	Docking studies	[100]
GROMACS	Conducting MD with a variety of force fields. Analysis of MD trajectories	[101]
Modeller	Homology model creation	[102]
Molecular Operating Environment (MOE)	Visualization of protein crystal structures. Homology model creation	[103]
NAMD	Popular force field for conducting MD simulations	[104]
PHYRE2	Online homology modeling server	[105]
Pymol	Visualizing protein crystal structures and homology models; makes publication-worthy figures	
Rosetta Suite	De novo rational protein design	[93]
SWISS-MODEL	Online homology modeling server; homology model analysis tools	[106]
Visual Molecular Dynamics (VMD)	Visualizing protein crystal structures, homology models, and MD trajectories	[107]

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 [95]. 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* [108].

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 redesigned in the glycosidase endo-β-1,4-mannanase from *Cellulomonas fimi* through structural comparison between it and its homolog, another endo-β-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 [109].

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 [110]. 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 [96]

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. Building a functional MCF: metabolic engineering

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 reaction thermodynamics. A balance is reached between activity requirements and protein production costs, and over-expression can tax the cell of resources and result in sacrificed cell growth and productivity [116]. As such, additional metabolic engineering strategies look to knockout gene expression and enzymatic activity of competing pathway(s) in effort to direct metabolic flux towards formation of the

desired product. Gene knockout is a well-established genomic tool used for substituting or removing sections of genomic DNA by engineering unique sites on an integration plasmid complementary to the gene of interest. Accessory proteins, either native to the host or supplied in the integration plasmid, catalyze the replacement of the targeted gene with a marker (e.g., antibiotic resistance gene) that can later be removed [117,118]. However, gene knockouts can cause host instability, resulting in decreased growth rate. Decreasing gene activity through engineered RNA regulation can resolve such instability. In this technique, small regulatory RNA (sRNA) molecules anneal and capture nascent transcripts and block translation and/or tag them for degradation. A library of synthetic sRNA molecules was screened and ultimately resulted in increased production of tyrosine and cadaverine in *E. coli*. The technique is advantageous because sRNAs can be fine-tuned through the calculation of sRNA-mRNA binding thermodynamics, and mRNAs from multiple genes can be targeted simultaneously without chromosomal integrations [119].

Though gene addition and over-expression can yield new products, enzyme activity deficiencies are more often the result of substrate limitation than low enzyme copy numbers. The formation of bottlenecks and buildup of toxic intermediates have been overcome using regulated or specialized promoters to balance substrate pools below toxic thresholds. This was seen in the engineered regulation of an artifical amorphadiene pathway that was designed to limit farnesyl pyrophosphate accumulation [53]. Even when an operon of multiple genes is controlled by a single promoter, variations in individual gene activities can be engineered by modifying secondary structure stability in the mRNA transcript between separate genes and altering the ribosome binding site. This can generate up to a 100,000-fold range in protein production from neighboring genes [120,121]. Substrate limitations due to membrane transport also limit the formation of product. For example, over-expression and mutation of appropriate transporter proteins in various experiments have resulted in a 70% increase in isoprenoid titer [122] and a 70% increase in xylose sugar utilization [123]. A recent approach involves protein scaffolds, which physically bind enzymes of a multi-step synthesis pathway into close proximity to enable substrate channeling. This technique has been used in the synthetic multi-step production of hydrogen from glucose, where the enzyme scaffold effectively reduced interference from competing enzymes and minimized kinetic limitations caused by the diffusion of substrates into the bulk cell environment [124]. Thus, several tools exist to implement a rational metabolic engineering strategy to effectively over-express a de novo biosynthetic pathway of interest. However, cellular resources are based on global demands for precursors, energy, and reducing power. Computational tools, based on genome-scale metabolic flux modeling, are now in place to gain holistic understandings of metabolic activity and design engineering strategies.

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 the biosynthetic pathways, but will eventually be stymied by enzymatic limitations. Enzyme engineering, through rational design, rational redesign, and directed evolution, can lessen or remove such limitations. Computational approaches have improved the design flow for MCFs, and as tools made specifically for the prediction and improvement of novel biosynthetic pathways evolve, it may become possible to produce any product from virtually any renewable resource.

Acknowledgments

References

[1] Dhamankar H, Prather KL. Microbial chemical factories: recent advances in pathway engineering for synthesis of value added chemicals. Curr Opin Struct Biol 2011;21:488–94.
[2] Du J, Shao Z, Zhao H. Engineering microbial factories for synthesis of value-added products. J Ind Microbiol Biotechnol 2011;38:873–90.
[3] Lee SY, Mattanovich D, Villaverde A. Systems metabolic engineering, industrial biotechnology and microbial cell factories. Microb Cell Fact 2012;11:156.
[4] Otero JM, Cimini D, Patil KR, Poulsen SG, Olsson L, et al. Industrial systems biology of Saccharomyces cerevisiae enables novel succinic acid cell factory. PLoS One 2013; 8:e54144.
[5] Choi O, Wu CZ, Kang SY, Ahn JS, Uhm TB, et al. Biosynthesis of plant-specific phenylpropanoids by construction of an artificial biosynthetic pathway in Escherichia coli. J Ind Microbiol Biotechnol 2011;38:1657–65.
[6] Choi O, Lee JK, Kang SY, Pandey RP, Sohng JK, et al. Construction of artificial biosynthetic pathways for resveratrol glucoside derivatives. J Microbiol Biotechnol 2014; 24:614–8.
[7] Babtie A, Tokuriki N, Hollfelder F. What makes an enzyme promiscuous? Curr Opin Chem Biol 2010;14:200–7.
[8] Khersonsky O, Tawfik DS. Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu Rev Biochem 2010;79:471–505.
[9] Atsumi S, Hanai T, Liao JC. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 2008;451:86–9.
[10] Atsumi S, Liao JC. Metabolic engineering for advanced biofuels production from Escherichia coli. Curr Opin Biotechnol 2008;19:414–9.
[11] Lennen RM, Braden DJ, West RM, Dumesic JA, Pfleger BF. A process for microbial hydrocarbon synthesis: overproduction of fatty acids in Escherichia coli and catalytic conversion to alkanes. Biotechnol Bioeng 2010;106:193–202.
[12] McKenna R, Nielsen DR. Styrene biosynthesis from glucose by engineered E. coli. Metab Eng 2011;13:544–54.
[13] Tseng HC, Martin CH, Nielsen DR, Prather KLJ. Metabolic engineering of Escherichia coli for enhanced production of (R)- and (S)-3-hydroxybutyrate. Appl Environ Microbiol 2009;75:3137–45.
[14] Xia XX, Qian ZG, Ki CS, Park YH, Kaplan DL, et al. Native-sized recombinant spider silk protein produced in metabolically engineered Escherichia coli results in a strong fiber. Proc Natl Acad Sci U S A 2010;107:14059–63.
[15] Anthony JR, Anthony LC, Nowroozi F, Kwon G, Newman JD, et al. Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene. Metab Eng 2009; 11:13–9.
[16] Asadollahi MA, Maury J, Schalk M, Clark A, Nielsen J. Enhancement of farnesyl diphosphate pool as direct precursor of sesquiterpenes through metabolic engineering of the mevalonate pathway in Saccharomyces cerevisiae. Biotechnol Bioeng 2010;106:86–96.
[17] Tawfik OK, S D. Enzyme promiscuity: a mechanistic and evolutionary perspective. Ann Rev Biochem 2010;79:471–505.
[18] Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science 2010;329:52–6.
[19] Gibson DG. Programming biological operating systems: genome design, assembly and activation. Nat Methods 2014;11:521–6.
[20] Köpke M, Held C, Hujer S, Liesegang H, Wiezer A, et al. Clostridium ljungdahlii represents a microbial production platform based on syngas. Proc Natl Acad Sci U S A 2010;107:13087–92.
[21] Papoutsakis E, Al-Hinai M, Jones SW, Indurthi DC, Mitchell DK, et al. Recombinant clostridia that fix CO₂ and co and uses thereof; 2013.
[22] Papoutsakis ET. Engineering solventogenic clostridia. Curr Opin Biotechnol 2008; 19:420–9.
[23] Becker J, Schafer R, Kohlstedt M, Harder BJ, Borchert NS, et al. Systems metabolic engineering of Corynebacterium glutamicum for production of the chemical chaperone ectoine. Microb Cell Fact 2013;12:110.
[24] Choi O, Lee JK, Kang SY, Pandey RP, Sohng JK, et al. Construction of artificial biosynthetic pathways for resveratrol glucoside derivatives. J Microbiol Biotechnol 2014; 24:614–8.
[25] Marchisio MA, Stelling J. Computational design tools for synthetic biology. Curr Opin Biotechnol 2009;20:479–85.
[26] Li XJ, Hayward C, Fong PY, Dominguez M, Hunsucker SW, et al. A blood-based proteomic classifier for the molecular characterization of pulmonary nodules. Sci Transl Med 2013;5.
[27] Wang CJ, Funk CC, Eddy JA, Price ND. Transcriptional analysis of aggressiveness and heterogeneity across grades of astrocytomas. PLoS One 2013;8:e76694.
[28] Chandrasekaran S, Price ND. Metabolic constraint-based refinement of transcriptional regulatory networks. PloS Comput Biol 2013;9:e1003370.
[29] Medema MH, van Raaphorst R, Takano E, Breitling R. Computational tools for the synthetic design of biochemical pathways. Nat Rev Microbiol 2012;10:191–202.
[30] Walker K. Application of plant and bacterial enzymes in the biocatalysis of paclitaxel (110.3). FASEB J 2014;28.
[31] Wang Y, San KY, Bennett GN. Cofactor engineering for advancing chemical biotechnology. Curr Opin Biotechnol 2013;24:994–9.
[32] Milne CB, Kim PJ, Eddy JA, Price ND. Accomplishments in genome-scale in silico modeling for industrial and medical biotechnology. Biotechnol J 2009;4: 1653–70.
[33] Goel A, Wortel MT, Molenaar D, Teusink B. Metabolic shifts: a fitness perspective for microbial cell factories. Biotechnol Lett 2012;34:2147–60.
[34] Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, et al. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 2010;28:977–82.
[35] Buchel F, Rodriguez N, Swainston N, Wrzodek C, Czauderna T, et al. Path2Models: large-scale generation of computational models from biochemical pathway maps. BMC Syst Biol 2013;7:116.
[36] Ganter M, Bernard T, Moretti S, Stelling J, Pagni M. MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. Bioinformatics 2013;29:815–6.
[37] Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A 2002;99:15112–7.
[38] Fong SS, Palsson BO. Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. Nat Genet 2004;36:1056–8.
[39] Apte AA, Senger RS, Fong SS. Designing novel cellulase systems through agent-based modeling and global sensitivity analysis. Bioengineered 2014;5:1–11.
[40] Ferrer-Miralles N, Villaverde A. Bacterial cell factories for recombinant protein production; expanding the catalogue. Microb Cell Fact 2013;12:113.
[41] Tokunaga H, Arakawa T, Tokunaga M. Novel soluble expression technologies derived from unique properties of halophilic proteins. Appl Microbiol Biotechnol 2010;88:1223–31.
[42] Okesli A, Cooper LE, Fogle EJ, van der Donk WA. Nine post-translational modifications during the biosynthesis of cinnamycin. J Am Chem Soc 2011;133: 13753–60.
[43] Mattanovich D, Branduardi P, Dato L, Gasser B, Sauer M, et al. Recombinant protein production in yeasts. Methods Mol Biol 2012;824:329–58.
[44] Lu X, Vora H, Khosla C. Overproduction of free fatty acids in E. coli: implications for biodiesel production. Metab Eng 2008;10:333–9.
[45] Liu X, Sheng J, Curtiss 3rd R. Fatty acid production in genetically modified cyanobacteria. Proc Natl Acad Sci U S A 2011;108:6899–904.
[46] Becker J, Wittmann C. Bio-based production of chemicals, materials and fuels — Corynebacterium glutamicum as versatile cell factory. Curr Opin Biotechnol 2012; 23:631–40.

[47] Royce LA, Liu P, Stebbins MJ, Hanson BC, Jarboe LR. The damaging effects of short chain fatty acids on *Escherichia coli* membranes. Appl Microbiol Biotechnol 2013; 97:8317–27.

[48] Jarboe LR, Royce LA, Liu P. Understanding biocatalyst inhibition by carboxylic acids. Front Microbiol 2013;4:272.

[49] Brown SD, Guss AM, Karpinets TV, Parks JM, Smolin N, et al. Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*. Proc Natl Acad Sci U S A 2011;108:13752–7.

[50] Tomas CA, Welker NE, Papoutsakis ET. Overexpression of groESL in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. Appl Environ Microbiol 2003;69:4951–65.

[51] Huffer S, Clark ME, Ning JC, Blanch HW, Clark DS. Role of alcohols in growth, lipid composition, and membrane fluidity of yeasts, bacteria, and archaea. Appl Environ Microbiol 2011;77:6400–8.

[52] Zhang F, Carothers JM, Keasling JD. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nat Biotechnol 2012;30:354–9.

[53] Dahl RH, Zhang F, Alonso-Gutierrez J, Baidoo E, Batth TS, et al. Engineering dynamic pathway regulation using stress–response promoters. Nat Biotechnol 2013;31: 1039–46.

[54] Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009;6:343–5.

[55] Mohr G, Hong W, Zhang J, Cui GZ, Yang Y, et al. A targetron system for gene targeting in thermophiles and its application in *Clostridium thermocellum*. PLoS One 2013;8:e69032.

[56] Senger RS, Yen JY, Fong SS. A review of genome-scale metabolic flux modeling of anaerobiosis in biotechnology; 2014 (under review).

[57] Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature 2011;475:324–32.

[58] Bernal V, Castaño-Cerezo S, Gallego-Jara J, Écija-Conesa A, de Diego T, et al. Regulation of bacterial physiology by lysine acetylation of proteins. New Biotechnol 2014; 31:586–95.

[59] Osterlehner A, Simmeth S, Göpfert U. Promoter methylation and transgene copy numbers predict unstable protein production in recombinant Chinese hamster ovary cell lines. Biotechnol Bioeng 2011;108:2670–81.

[60] Damasceno L, Huang Jr C, Batt C. Protein secretion in *Pichia pastoris* and advances in protein production. Appl Microbiol Biotechnol 2012;93:31–9.

[61] Swiech K, Picanço-Castro V, Covas DT. Human cells: new platform for recombinant therapeutic protein production. Protein Expr Purif 2012;84:147–53.

[62] Zhu J. Mammalian cell protein expression for biopharmaceutical production. Biotechnol Adv 2012;30:1158–70.

[63] Drugmand J-C, Schneider Y-J, Agathos SN. Insect cells as factories for biomanufacturing. Biotechnol Adv 2012;30:1140–57.

[64] Pandhal J, Woodruff LBA, Jaffe S, Desai P, Ow SY, et al. Inverse metabolic engineering to improve *Escherichia coli* as an N-glycosylation host. Biotechnol Bioeng 2013; 110:2482–93.

[65] Mendez-Perez D, Gunasekaran S, Orler VJ, Pfleger BF. A translation-coupling DNA cassette for monitoring protein translation in *Escherichia coli*. Metab Eng 2012; 14:298–305.

[66] Finley SD, Broadbelt LJ, Hatzimanikatis V. In silico feasibility of novel biodegradation pathways for 1,2,4-trichlorobenzene. BMC Syst Biol 2010;4:7.

[67] Henry CS, Broadbelt LJ, Hatzimanikatis V. Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. Biotechnol Bioeng 2010;106:462–73.

[68] Brunk E, Neri M, Tavernelli I, Hatzimanikatis V, Rothlisberger U. Integrating computational methods to retrofit enzymes to synthetic pathways. Biotechnol Bioeng 2012;109:572–82.

[69] Wu D, Wang Q, Assary RS, Broadbelt LJ, Krilov G. A computational approach to design and evaluate enzymatic reaction pathways: application to 1-butanol production from pyruvate. J Chem Inf Model 2011;51:1634–47.

[70] Finley SD, Broadbelt LJ, Hatzimanikatis V. Computational framework for predictive biodegradation. Biotechnol Bioeng 2009;104:1086–97.

[71] Schomburg I, Chang A, Schomburg D. BRENDA, enzyme data and metabolic information. Nucleic Acids Res 2002;30:47–9.

[72] Rodrigo G, Carrera J, Prather KJ, Jaramillo A. DESHARKY: automatic design of metabolic pathways for optimal cell growth. Bioinformatics 2008;24: 2554–6.

[73] Chou C-H, Chang W-C, Chiu C-M, Huang C-C, Huang H-D. FMM: a web server for metabolic pathway reconstruction and comparative analysis. Nucleic Acids Res 2009;37:W129–34.

[74] Kotera M, Tabei Y, Yamanishi Y, Tokimatsu T, Goto S. Supervised de novo reconstruction of metabolic pathways from metabolome-scale compound sets. Bioinformatics 2013;29:i135–44.

[75] Klopman G, Dimayuga M, Talafous J. META. 1. A program for the evaluation of metabolic transformation of chemicals. J Chem Inf Comput Sci 1994;34: 1320–5.

[76] Xia D, Zheng H, Liu Z, Li G, Li J, et al. MRSD: a web server for metabolic route search and design. Bioinformatics 2011;27:1581–2.

[77] McClymont K, Soyer OS. Metabolic tinker: an online tool for guiding the design of synthetic metabolic pathways. Nucleic Acids Res 2013;41:e113.

[78] Testa B, Balmat A-L, Long A, Judson P. Predicting drug metabolism — an evaluation of the expert system METEOR. Chem Biodivers 2005;2:872–85.

[79] Hou BK, Wackett LP, Ellis LBM. Microbial pathway prediction: a functional group approach. J Chem Inf Comp Sci 2003;43:1051–7.

[80] Moriya Y, Shigemizu D, Hattori M, Tokimatsu T, Kotera M, et al. PathPred: an enzyme-catalyzed metabolic pathway prediction server. Nucleic Acids Res 2010;38:W138–43.

[81] Mithani A, Preston GM, Hein J. Rahnuma: hypergraph-based tool for metabolic pathway prediction and network comparison. Bioinformatics 2009;25:1831–2.

[82] Prather KLJ, Martin CH. De novo biosynthetic pathways: rational design of microbial chemical factories. Curr Opin Biotechnol 2008;19:468–74.

[83] Hatzimanikatis V, Li C, Ionita JA, Henry CS, Jankowski MD, et al. Exploring the diversity of complex metabolic networks. Bioinformatics 2005;21: 1603–9.

[84] Carbonell P, Parutto P, Herisson J, Pandit SB, Faulon J-L. XTMS: pathway design in an eXTended metabolic space. Nucleic Acids Res 2014;42:W389–394.

[85] Dalby PA. Strategy and success for the directed evolution of enzymes. Curr Opin Struct Biol 2011;21:473–80.

[86] Maté D, García-Ruiz E, Camarero S, Alcalde M. Directed evolution of fungal laccases. Curr Genomics 2011;12:113–22.

[87] Wilderman PR, Gay SC, Jang HH, Zhang Q, Stout CD, et al. Investigation by site-directed mutagenesis of the role of cytochrome P450 2B4 non-active-site residues in protein–ligand interactions based on crystal structures of the ligand-bound enzyme. FEBS J 2012;279:1607–20.

[88] Kille S, Acevedo-Rocha CG, Parra LP, Zhang Z-G, Opperman DJ, et al. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. ACS Syn Biol 2012;2:83–92.

[89] Andrews FH, McLeish MJ. Using site-saturation mutagenesis to explore mechanism and substrate specificity in thiamin diphosphate-dependent enzymes. FEBS J 2013; 280:6395–411.

[90] Wang Z, Chen T, Ma X, Shen Z, Zhao X. Enhancement of riboflavin production with *Bacillus subtilis* by expression and site-directed mutagenesis of zwf and gnd gene from *Corynebacterium glutamicum*. Bioresour Technol 2011;102: 3934–40.

[91] Eriksen DT, Hsieh PCH, Lynn P, Zhao H. Directed evolution of a cellobiose utilization pathway in *Saccharomyces cerevisiae* by simultaneously engineering multiple proteins. Microb Cell Fact 2013;12:61.

[92] Li H, Liao JC. Development of an NADPH-dependent homophenylalanine dehydrogenase by protein engineering. ACS Syn Biol 2013;3:13–20.

[93] Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol 2011;487:545–74.

[94] Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels–Alder reaction. Science 2010;329:309–13.

[95] Rajagopalan S, Wang C, Yu K, Kuzin AP, Richter F, et al. Design of activated serine-containing catalytic triads with atomic-level accuracy. Nat Chem Biol 2014;10: 386–91.

[96] Case DA, Cheatham 3rd TE, Darden T, Gohlke H, Luo R, et al. The Amber biomolecular simulation programs. J Comput Chem 2005;26:1668–88.

[97] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 2009;30:2785–91.

[98] Brooks BR, Brooks 3rd CL, Mackerell Jr AD, Nilsson L, Petrella RJ, et al. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30:1545–614.

[99] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. UCSF Chimera —a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–12.

[100] Moustakas DT, Lang PT, Pegg S, Pettersen E, Kuntz ID, et al. Development and validation of a modular, extensible docking program: DOCK 5. J Comput Aided Mol Des 2006;20:601–19.

[101] Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 2013;29: 845–54.

[102] Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.

[103] Molecular Operating Environment (MOE) (2013) Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.

[104] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–802.

[105] Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 2009;4:363–71.

[106] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 2014;42:W252–258.

[107] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14(33–38):27–38.

[108] Holland JT, Harper JC, Dolan PL, Manginell MM, Arango DC, et al. Rational redesign of glucose oxidase for improved catalytic function and stability. PLoS One 2012;7: e37924.

[109] Hekmat O, Lo Leggio L, Rosengren A, Kamarauskaite J, Kolenova K, et al. Rational engineering of mannosyl binding in the distal glycone subsites of *Cellulomonas fimi* endo-beta-1,4-mannanase: mannosyl binding promoted at subsite-2 and demoted at subsite-3. Biochemistry 2010;49:4884–96.

[110] Torres LL, Cantero A, del Valle M, Marina A, Lopez-Gallego F, et al. Engineering the substrate specificity of a thermophilic penicillin acylase from *Thermus thermophilus*. Appl Environ Microbiol 2013;79:1555–62.

[111] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31:455–61.

[112] Murugesan V, Makwana N, Suryawanshi R, Saxena R, Tripathi R, et al. Rational design and synthesis of novel thiazolidin-4-ones as non-nucleoside HIV-1 reverse transcriptase inhibitors. Bioorg Med Chem 2014;22:3159–70.

[113] Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–637.

[114] Isralewitz B, Gao M, Schulten K. Steered molecular dynamics and mechanical functions of proteins. Curr Opin Struct Biol 2001;11:224–30.

[115] Durmaz E, Kuyucak S, Sezerman UO. Modifying the catalytic preference of tributyrin in *Bacillus thermocatenulatus* lipase through in-silico modeling of enzyme–substrate complex. Protein Eng Des Sel 2013;26:325–33.

[116] Li G-W, Burkhardt D, Gross C, Weissman Jonathan S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 2014;157:624–35.

[117] Heap JT, Ehsaan M, Cooksley CM, Ng Y-K, Cartman ST, et al. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. Nucleic Acids Res 2012;40:e59.

[118] St-Pierre F, Cui L, Priest DG, Endy D, Dodd IB, et al. One-step cloning and chromosomal integration of DNA. ACS Syn Biol 2013;2:537–41.

[119] Na D, Yoo SM, Chung H, Park H, Park JH, et al. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. Nat Biotechnol 2013;31:170–4.

[120] Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 2009;27:946–50.

[121] Pfleger BF, Pitera DJ, Smolke CD, Keasling JD. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol 2006;24:1027–32.

[122] Wang J-F, Xiong Z-Q, Li S-Y, Wang Y. Enhancing isoprenoid production through systematically assembling and modulating efflux pumps in *Escherichia coli*. Appl Microbiol Biotechnol 2013;97:8057–67.

[123] Young EM, Comer AD, Huang H, Alper HS. A molecular transporter engineering approach to improving xylose catabolism in *Saccharomyces cerevisiae*. Metab Eng 2012;14:401–11.

[124] You C, Myung S, Zhang YHP. Facilitated substrate channeling in a self-assembled trifunctional enzyme complex. Angew Chem Int Ed 2012;51:8787–90.

[125] Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 2003;84:647–57.

[126] Ranganathan S, Suthers PF, Maranas CD. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PloS Comput Biol 2010;6:e1000744.

[127] Kim J, Reed JL. OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst Biol 2010;4:53.

[128] Chowdhury A, Zomorrodi AR, Maranas CD. k-OptForce: integrating kinetics with flux balance analysis for strain design. PloS Comput Biol 2014;10:e1003487.

[129] McAnulty MJ, Yen JY, Freedman BG, Senger RS. Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism in silico. BMC Syst Biol 2012;6:42.

[130] Yen JY, Nazem-Bokaee H, Freedman BG, Athamneh AI, Senger RS. Deriving metabolic engineering strategies from genome-scale modeling with flux ratio constraints. Biotechnol J 2013;8:581–94.