Published in final edited form as:

Biotechnol J. 2013 May ; 8(5): 545–555. doi:10.1002/biot.201200371.

Protein engineering for metabolic engineering: current and next-generation tools

Ryan J. Marcheschi¹, Luisa S. Gronenberg¹, and James C. Liao^{1,2}

¹Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095, USA

²Institute of Genomics and Proteomics, UCLA, Los Angeles, CA 90095, USA

Abstract

Protein engineering in the context of metabolic engineering is increasingly important to the field of industrial biotechnology. As the demand for biologically-produced food, fuels, chemicals, food additives, and pharmaceuticals continues to grow, the ability to design and modify proteins to accomplish new functions will be required to meet the high productivity demands for the metabolism of engineered organisms. This article reviews advances of selecting, modeling, and engineering proteins to improve or alter their activity. Some of the methods have only recently been developed for general use and are just beginning to find greater application in the metabolic engineering community. We also discuss methods of generating random and targeted diversity in proteins to generate mutant libraries for analysis. Recent uses of these techniques to alter cofactor use, produce non-natural amino acids, alcohols, and carboxylic acids, and alter organism phenotypes are presented and discussed as examples of the successful engineering of proteins for metabolic engineering purposes.

Keywords

Biofuels; Industrial Biotechnology; Modeling; Screening; Synthetic Biology

1 Introduction

Protein engineering and metabolic engineering are synergistic, linked fields in the broader research arena of industrial biotechnology and synthetic biology [1]. The purpose of metabolic engineering is to build organisms that efficiently produce compounds of interest. To this effect metabolic engineers minimize the resources an organism diverts to growth and unwanted metabolites, and maximize the flux through a desired pathway [2]. A key component of this is engineering of proteins such that they facilitate flux through said pathway. Protein engineering involves altering protein structure, either by targeted or

Corresponding author: James, C. Liao (liaoj@seas.ucla.edu), Correspondence may be addressed to: Prof. James C. Liao, 5531 Boelter Hall, 420 Westwood Plaza, Los Angeles, CA 90095.

Competing financial interests

J.C.L. is a cofounder and a board member of Easel Biotechnologies, which has licensed biofuel technology from the University of California, Los Angeles.

random mutagenesis, to achieve functional changes such as decreased product inhibition, better substrate sensitivity, higher catalytic rates, desired cofactor use, and reduced substrate competition [3, 4]. Often the engineering process involves several rounds of directed evolution [5, 6]. These modifications can significantly improve both titers and yields of metabolically produced compounds.

In many ways, even the simplest bacterial systems are so complex in their regulation that engineering of single proteins or pathways is unable to alter their behavior significantly. In such cases, libraries of artificial transcriptional regulators have been used to alter gene expression on an organism-level scale, with the ability to screen or select for desired phenotypes [7].

In this review, we first discuss tools of engineering proteins and of identifying which proteins in a metabolic pathways would make the best targets for engineering. We then highlight selected examples from the recent literature of protein engineering being used for the purpose of improving a metabolic pathway. We believe this discussion will highlight the power of protein engineering to metabolic engineers and explain some of the key concepts of the field.

2 Protein Engineering and Metabolic Engineering tools

2.1 Metabolic pathway analysis

The goal of metabolic engineering is generally to maximize production of a specific chemical; occasionally the goal may coincide with maximizing growth rate. Pathway analysis approaches help scientists compare various routes towards a desired product on account of factors such as thermodynamic favorability of each step, reaction mechanisms and substrate binding specificity. These computational tools help metabolic engineers make informed choices as to which proteins should be altered to build an optimal pathway, i.e. where predicted steps require an enzymatic function that is not naturally available, or where enzymatic activity must be increased or cofactor specificity altered (Figure 1).

Metabolic network reconstructions combine genomic and metabolic data for a pathway or a whole organism. When the genome of an organism is newly sequenced, metabolic reconstructions can be built based on databases of previously annotated sequences. Local alignment tools, such as BLAST, can be used to align unknown sequences to ones that have been annotated [8, 9]. Much of the information used for these reconstructions is retrieved from databases such as KEGG, BRENDA, or Biocyc [10–12]. There are also programs available to automate the network reconstruction process. Examples include KAAS [13] or GLOBUS, which provides quantitative probabilities for each annotation [14]. The reconstructed network is then visualized, most often through a metabolic reaction network that depicts the stoichiometry of biochemical interactions. Some databases make the network reconstructions available online, these include KEGG Pathway and Metacyc [10, 11]. These tools provide initial selection of metabolic engineering targets based on biochemical and genomic information. Data from preliminary experiments can be used to develop and refine models using strategies such as ensemble modeling [15]. These in turn

can predict further engineering targets. A more thorough discussion of all the computational tools available to metabolic engineers was recently published [16].

2.2 Modeling chemical reactions

Details of chemical reactions can be examined by quantum mechanical (QM) modeling to investigate if the predicted or desired enzymatic reactions are likely to occur, what the energy barriers may be for the reaction, and what the chemical transition state may look like [17, 18]. These are important factors for metabolic engineers, since reactions with high energetic barriers require additional considerations. In such a case there are two options: choosing a different pathway to achieve the same transformation, or coupling the unfavorable reaction to positive driving forces such as high substrate concentrations, use of ATP [19], or release of CO₂ [20]. Because enzymes lower the energy of the transition state, a model of this structure is useful in subsequent protein engineering methods to generate proteins with increased activity [21, 22] (Figure 1). In cases where several sequential chemical reactions are involved in a metabolic pathway, the use of scaffold proteins, reviewed in this issue [23], can lead to enhanced metabolic flux through the pathway and reduced production of undesired byproducts.

2.3 Protein Structural Modeling

One common method of protein modeling is the use of sequence-based evolutionary approaches to determine which sites are conserved. A further comparison of the sequences, structures, and functions of related proteins is used to identify residues important for substrate specificity. This method has been recently reviewed [23, 24] and will not be further discussed here. Another tool is structure-based molecular modeling of enzymes and their binding of substrates (Figure 1). In the past, this was done by either obtaining a co-crystal structure of an enzyme with its substrate or by docking a substrate into the active site to create a model using tools like GOLD [25] or AUTODOCK [26]. Protein mutations could then be rationally designed to target the substrate-binding site to allow use of different substrates. Other sites on the protein, such as allosteric feedback-inhibition sites can also be rationally mutated. While rational design does enable more specific targeting than random mutagenesis or error-prone PCR, even with site-saturation mutagenesis, these rationally designed mutants do not always function as predicted, either by rendering the protein inactive, interfering with folding or solubility, or having no desired effect. This represents a significant loss of time and resources invested in making and testing mutants. From a practical standpoint, only a certain number of mutants can be tested, particularly if there is no selection pressure for altered protein function.

Rosetta, a molecular modeling program, has been developed to use structural information to examine *in silico* the effects of protein mutations on stability, folding and substrate binding [27, 28]. This program enables examination of site-saturation mutagenesis of multiple sites in a protein simultaneously, resulting in potentially billions of protein mutants to be examined. This process limits the candidates for experimental testing to only those mutants predicted to maintain protein folding, stability and substrate-binding. In effect, this increases the success rate of rationally designed mutations by allowing objective comparison of many more theoretical mutants than a biologist could individually examine. Thanks to improved

gene synthesis capabilities, these theoretical mutants can be made directly and tested for function, saving time and effort. Rosetta and similar programs can also be used to design enzymes with novel functions that have not been observed in nature. This, coupled with QM modeling, enables a large expansion of the available chemical reaction space, potentially reaching the point where any theoretically-possible reaction could be catalyzed by an enzyme; either one found in nature, mutated, or designed *de novo* [29, 30]. While metabolic engineering uses of Rosetta have not been published to date, Rosetta has progressed to a point where it could be very beneficial to the field.

2.4 Diversity generation

Protein engineering for metabolic engineering often relies on cell growth-coupled selections and assay-based screens to identify proteins with desired characteristics (as mentioned in Figure 1 and the examples section below). These methods are referred to as directed evolution and have been recently reviewed [5, 6]. For effective protein engineering, diverse libraries of proteins are needed in order to adequately sample sequence space. Here we describe several methods of generating random and targeted genetic diversity for protein libraries, including use of error-prone PCR, chemical mutagens, site-directed and site-saturation mutagenesis, DNA shuffling [31], artificial transcription factors, and Multiplex Automated Genome Engineering (MAGE) [32] (Figure 2). The primary difference amongst these techniques is whether they generate random diversity or targeted diversity. Random diversity is generated by techniques such as error-prone PCR, use of chemical mutagens, or use of artificial transcription factor libraries and is most useful when the structure and/or identity of the protein to be engineered is not known. Targeted diversity, in contrast, is useful when the protein structure is known, and is used to specifically alter sites in a protein that are the most likely to affect function.

One method for generating random diversity in a gene or pathway is error-prone PCR. This technique makes use of either mutated DNA polymerases with reduced fidelity or reaction conditions that promote an increased error rate and has been recently reviewed in detail [33]. Another method of random diversity generation at a phenotypic level, discussed in further detail below, is the use of artificial transcription factor libraries, *e.g.* where zinc-finger DNA binding domains are fused together with an RNA-polymerase-recruiting transactivation domain, to enable complex whole-cell phenotype changes. The last method discussed here is whole-cell mutagenesis. This method is particularly useful when the relevant metabolic pathway is unknown or poorly defined. The most common general mutagenesis methods are chemical mutagenesis with the alkylating agent *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (NTG) or UV irradiation. In cases where adaptive selection pressure can be applied, spontaneous mutations may also arise over time. This method has the downside that causative mutations are often hard to identify amongst the high background rate of mutations and therefore cannot be transferred to new strains. However, a recent sequencing and computation approach alleviates this problem by allowing identification of relevant mutations amongst a high background [34]. Another disadvantage of randomly generated diversity is the number of mutants that must be analyzed for full library coverage. For example, there are about 1.1×10^5 possibilities for 1 mutation anywhere in a 300 amino acid protein with 1.6×10^7 and 1.6×10^9 possibilities for 2 and 3 mutations, respectively, and three

times as many randomly-generated mutants must be subjected to selection or screening to guarantee around 95% library coverage.

Targeted diversity generation can dramatically reduce library sizes by defining specific residues in the protein to mutate and limiting mutation to a selected subset of amino acids, rather than all 20. Site-saturation mutagenesis (SSM) and site-directed mutagenesis (SDM) have been previously described [35, 36] and are used to specifically modify one or more sites in a protein sequence by encoding the desired mutation(s) into PCR primers and using PCR to generate a small library of degenerate sequence mutations by using the degenerate DNA codon NNK (which codes for all 20 amino acids) as part of the primer used for PCR (in SSM), or to produce a single mutation by encoding specified DNA mutations into the primer used for PCR (in SDM) (Figure 2). Unlike the libraries mentioned above, an SSM-produced library has up to 19 possibilities for 1 mutation, up to 361 possibilities for 2 mutations and up to 6859 possibilities for 3 mutations at the specified sites in the protein, regardless of protein sequence length. Another method of generating targeted sequence diversity is MAGE [32]. With this technique, multiple specific targets for engineering protein expression and function in a metabolic pathway can be mutated simultaneously, with up to 30 DNA base-pairs (encoding 10 amino acids) mutated in the center of each targeted region. These mutations occur and accumulate by repeated addition of targeted primers with degenerate DNA mutations to successive generations of continuously evolving cells, allowing rapid attainment of genetic diversity. MAGE has been used in a co-selection strategy to alter protein-coding sequences, thereby turning protein function on and off and allowing for selection of mutant genome sequences that co-segregate with the modified protein-coding sequences [37].

DNA shuffling techniques [38] are at the interface between random and targeted diversity generation. These techniques use homologous DNAs with different mutations (generated by either random or targeted diversity techniques) and recombine the mutations into a new combinatorial library, which is then analyzed for the optimal mutation combination. Since random mutation techniques generate a diverse set of mutations, it is often useful to take beneficial mutants identified during initial selections or screens and combine them randomly by fragmenting, melting, re-annealing, and PCR-amplifying the mutated DNA. Use of DNA shuffling, coupled with further selection or screening, enables rapid identification of beneficial mutation combinations that improve function. This can result in significant time-savings over continued rounds of mutation and directed evolution. This technique can be used with rationally-designed mutant libraries as well.

Once sequence diversity has been generated, appropriate screening and/or selection(s) must be used to determine which sequence mutants result in protein(s) with the desired function (Figure 1). In general, a selection is preferred, but is limited to growth-coupled enzymes and pathways. A recent example is the rescue of anaerobic growth in *E. coli* by directed evolution of an enzyme in a pathway that can consume reducing equivalents [39]. For screening, it is important to target enzyme or pathway activity in as direct a way as possible while minimizing the assay time and number of steps. Details of screening and selection methods have been recently reviewed [40] and successful examples have been highlighted [41].

2.5 Transcription factor engineering for complex phenotypes

In many cases single site mutations are not enough to elicit the desired cellular phenotype. Instead multiple mutations, often distant on the chromosome, are required to achieve the desired response. Apart from chemical mutagenesis, which carries some risks, as discussed above, there are not many methods of imparting multigenic modifications on the cell to achieve a complex change of phenotype. However, a one relatively simple method is mutating a transcription factor (TF). The advantage of this method is that the expression of many genes can be altered simultaneously, potentially producing dramatic changes in cell phenotypes. This requires mutagenesis of only one protein, the TF, which in turn acts on many distal genes. An approach termed global transcription machinery engineering (gTME) involves creating mutant TF libraries, screening for the desired phenotype and iteratively repeating this process for directed evolution [42]. The chosen TF is mutated to alter DNA binding or effects, most commonly by error-prone PCR. The library is transformed into the strain of interest and a screen or selection for the desired phenotype is applied. Several rounds of directed evolution can be applied if necessary. If the transcriptional changes underlying the resulting phenotype are of interest, the final strain can be characterized by qPCR and other quantitative methods. This technique has successfully been applied to prokaryotic and eukaryotic systems. Examples include engineering of yeast for increased ethanol tolerance [43] or xylose consumption [44], as discussed below. Interestingly, engineering of transcription factors in eukaryotic cells can be easier than in prokaryotes because eukaryotic transcription can be controlled using well-characterized zinc-finger proteins for DNA binding [45]. Additional methods of transcription factor engineering are reviewed in this issue [23].

2.6 Non-natural amino acids (NNAAs) for protein engineering

The ability to site-specifically incorporate NNAAs into proteins has recently been developed as a very powerful tool for protein engineering. It allows for incorporation of extensive chemical functionality into proteins since NNAAs can be synthesized with a great variety of side chains. Here we focus on one method, *in vivo* incorporation of NNAAs using orthogonal tRNA/aminoacyl-tRNA synthetase pairs. A recent review gives a detailed overview of this technology [44]. Even though there are no examples to date of NNAAs used directly for metabolic engineering, we highlight this tool here since the functions that can be achieved with NNAAs have many potential advantages for metabolic engineers.

Fidelity in protein translation is governed by the tRNA synthetase that loads an AA onto the correct tRNA and the correct pairing of the mRNA codon to the tRNA anticodon during translation by the ribosome. By manipulating the tRNA charging steps, novel AAs can be introduced into proteins. Peter Schultz and co-workers pioneered the strategy of using alternating positive and negative selections to engineer a tRNA/tRNA synthetase pair that is completely orthogonal to the organism where it is used [46]. Initial experiments used the tyrosyl tRNA/synthetase pair of *Methanococcus jannaschii* and the mRNA amber stop codon (UAG) in *E. coli* [46]. Many variations have since been published [44]. To date, over 70 NNAAs have successfully been incorporated [44]. These techniques have been expanded to yeast, mammalian cells and multicellular organisms [44, 47]. One disadvantage is that incorporation of NNAAs by replacing natural codons necessarily competes with natural

processes. Therefore, the recent progress in manipulating cellular genomes as a whole will be very useful for completely freeing up codons for NNAAAs [48]. Another development is the use of an evolved orthogonal ribosome as well as tRNA and synthetases to recognize several quadruplet codons [49].

Most relevant to the field of metabolic engineering are NNAAAs that introduce new chemical functionality to directly participate in the enzymatic mechanism. Examples include metal-ion binding, photoreactive and photocaged amino acids [44]. Also relevant are AAs that represent a posttranslationally modified version of a natural AA. Examples include sulfated, methylated, nitrosylated and, recently, phosphorylated amino acids [50].

One example of NNAA use to affect enzyme catalysis is modification of an enzyme that hydrolyzes a pesticide with the NNAA *L*-(7-hydroxycoumarin-4-yl)ethylglycine [51]. The modification increased the already high catalytic rate of the enzyme [51]. This type of work is of interest to metabolic engineers trying to improve the efficiency of a rate-limiting step; however, it is dependent on understanding the enzyme reaction mechanism at a structural level to make informed modifications.

3 Selected Examples of Protein Engineering for Metabolic Engineering

3.1 Metabolic engineering for production of alternative amino acids

Biosynthesis of the non-natural amino acid L-homoalanine was achieved by engineering mutants of glutamate dehydrogenase (GDH) from *E. coli* (Figure 3). This was accomplished with a metabolically-engineered transaminase-deficient valine auxotrophic strain of *E. coli* (*avtA ilvE*) [52]. Based on the crystal structure of *Clostridium symbiosum* GDH (PDB: 1BGV) [53] and sequence alignment with the *E. coli* GDH, residues within 6 Å of the gamma carbon of the glutamate substrate were identified as Lys92, Thr195, Val377, and Ser380. These sites were randomly mutated using site-saturation mutagenesis. The resulting library of mutants was selected for growth rescue of valine auxotrophy. Growth was rescued by selection of a GDH mutant capable of aminating 2-ketoisovalerate and producing valine. Since 2-ketobutyrate and 2-ketoisovalerate only differ by one carbon, this selection design produced GDH mutants that also worked to aminate 2-ketobutyrate, thereby producing L-homoalanine. One of these mutants, GDH (K92V/T195S) was found to be 50-fold more specific for 2-ketobutyrate than 2-ketoglutarate, with a 2-fold increased catalytic rate constant (k_{cat}) for 2-ketobutyrate and a 13-fold decreased k_{cat} for 2-ketoglutarate relative to wild-type GDH. To increase the yield of homoalanine, the strain overexpressing the GDH mutant was further engineered. Threonine to 2-ketobutyrate conversion was increased by overexpression of threonine dehydratase (IlvA) from *Bacillus subtilis* and use of a modified threonine-overproducing strain (ATCC98082), with the threonine transporter (*rhtA*) knocked out, led to production of 5.4 g L⁻¹ (26% of the theoretical maximum) at a productivity of 2.7 g L⁻¹ day⁻¹.

3.2 Engineering proteins in metabolic pathways for production of alcohols

Alcohol producing strains of *E. coli* have been constructed in several ways. Linear-chain alcohols can be produced either from the threonine biosynthesis pathway, which is native to *E. coli*, or via the citramalate pathway, which is not (Figure 3). For production of alcohols

using the citramalate pathway, citramalate synthase (CimA) from *Methanococcus jannaschii* was cloned into *E. coli* and engineered to increase its activity. CimA was mutated via error-prone PCR, followed by a growth-based selection in an *E. coli* isoleucine auxotroph (*ilvA tdcB*), which cannot synthesize 2-ketobutyrate from threonine [54]. Native *E. coli* enzymes isopropylmalate isomerase (LeuCD) and isopropylmalate dehydrogenase (LeuB) were also cloned to perform the isomerization and dehydrogenation steps for conversion of citramalate into 2-ketobutyrate (Figure 3). CimA mutants that rescued growth were identified and DNA shuffling was used to combine mutations, followed by testing for growth rescue. Initial mutants were subjected to further random mutation, shuffling and selection in a *ilvA tdcB ilvI* *E. coli* strain, thereby increasing selection pressure for high levels of 2-ketobutyrate production. Alcohols are produced by decarboxylation of 2-ketobutyrate by ketoisovalerate decarboxylase (KIVD) and dehydrogenation by alcohol dehydrogenase (ADH) to produce 1-propanol and by elongation of 2-ketobutyrate to 2-ketovalerate by the action of isopropylmalate synthase (LeuA), LeuCD, and LeuB, followed by action of KIVD and ADH to produce 1-butanol (Figure 3). CimA mutants that rescued growth in the *ilvA tdcB ilvI* *E. coli* strain were then screened for high levels of alcohol production. The best CimA mutant (I47V/E114V/H126Q/T204A/M250V/ 374-491) resulted in production of approximately 2.8 g L⁻¹ 1-propanol and 0.4 g L⁻¹ 1-butanol in a metabolically-engineered *ilvI ilvB E. coli* strain.

A similar approach to alcohol production used the threonine biosynthesis pathway with IlvA-mediated dehydration of threonine to produce 2-ketobutyrate (Figure 3). To enable recursive elongation of keto-acid substrates, which was recently reviewed [55], *E. coli* LeuA was rationally engineered to increase the size of the substrate-binding pocket based on the crystal structure of LeuA from *Mycobacterium tuberculosis* (PDB: 1SR9) [56]. Initial mutant LeuA enzymes allowed elongation of branched-chain substrates *in vivo* [57] and further mutation led to recursive elongation of linear-chain substrates *in vivo* [58]. The mutant with the largest binding pocket, LeuA (H97A/S139G/N167G/P169A/G462D) enabled 5 cycles of elongation for linear-chain substrates (Figure 3). Engineering of KIVD based on homology-modeling to the crystal structure of pyruvate decarboxylase from *Zymomonas mobilis* (PDB: 1ZPD) [59] was also done to enlarge its substrate binding site to enable decarboxylation of the elongated ketoacids produced by LeuB, LeuCD and the mutant LeuA [57]. Using the threonine-overproducing strain described above (ATCC98082 *rhtA*) with overexpression of IlvA from *Bacillus subtilis*, the engineered mutant KIVD (F381L/V461A), and a copy of the original feedback-resistant mutant of LeuA (G462D), a range of alcohols from 1-propanol to 1-octanol was produced (approximately 1.4 g L⁻¹ total alcohols). Use of the LeuA mutant-mediated recursive elongation in a phenylalanine-overproducing *E. coli* strain (ATCC31884) resulted in approximately 0.66 g L⁻¹ phenylethanol and 0.004 g L⁻¹ phenylpropanol production [58].

A recent example of branched-chain alcohol production involved engineering ketol-acid reductoisomerase (IlvC) to balance the NAD(P)H cofactor use and generation in the production of isobutanol via altering its cofactor specificity from NADPH to NADH by modifying key residues identified from a related crystal structure [60]. This enzyme is involved in the valine biosynthesis pathway, which can be manipulated for the production of

isobutanol [61] (Figure 3). Site-saturation mutagenesis was used to make mutant libraries for each targeted residue and libraries were screened for NADH preference. All beneficial mutants were then combinatorially recombined and screened, resulting in a mutant with a 54000-fold change in cofactor preference from the starting enzyme. Additionally, an alcohol dehydrogenase with NADH preference from *Lactococcus lactis* (AdhA) [62] was engineered to increase enzyme activity for conversion of isobutanol to isobutanol [53]. Here, a random mutagenesis library was screened for mutant enzymes with increased sensitivity towards low isobutanol concentrations. The most promising mutations were recombined and resulted in a sufficiently active enzyme with a 4-fold increased catalytic rate constant (k_{cat}) and 40-fold increased catalytic efficiency for isobutanol as compared to the wild-type enzyme. These protein engineering steps allowed for production of isobutanol in *E. coli* at 100% theoretical yield.

3.3 Protein engineering for production of branched-chain acids

The recursive elongation capability of the LeuA mutant has also been used to produce high levels of the carboxylic acids isovalerate and isocaproate *in vivo* [63]. In the valine and leucine biosynthesis pathways, LeuA, LeuCD, and LeuB elongate ketovaline to ketoleucine, with the LeuA mutant enabling further elongation to ketohomoleucine (Figure 3).

Decarboxylation by action of KIVD or indolepyruvate decarboxylase (IPDC) from *Salmonella typhimurium*, followed by oxidation using an aldehyde dehydrogenase (ALDH), then yields the branched-chain acids (Figure 3). To attain higher production of isovalerate and isocaproate *in vivo*, structure-based engineering of KIVD (PDB: 2VBG) [64] and IPDC was performed. Sequence alignment of IPDC to KIVD indicated amino acids in the active site to target for mutation. Resulting mutants were screened for gains in branched-chain acid production. One KIVD mutant (V461A/F542L) decreased isobutyrate production and increased isocaproate production, but had no effect on isovalerate production. In contrast, wild-type IPDC allowed production of 8.9 g L⁻¹ isovalerate (58% of the theoretical maximum) and a IPDC mutant (L544A) increased the isocaproate titer to 5 g L⁻¹ [63]. While this IPDC (L544A) mutation decreased k_{cat} by 6-39 fold depending on substrate, it also increased selectivity for ketohomoleucine 10-fold over ketoleucine and 567-fold over ketovaline, leading to increased production of isocaproate *in vivo* [63].

3.4 Protein engineering of hormones to treat disease

A recent use of protein engineering to alter complex metabolic processes is the conversion of a mammalian paracrine fibroblast growth factor (FGF) into an endocrine FGF [65]. This was done by structure-guided mutation of residues in the heparin sulfate binding site and generation of a chimeric protein composed of the mutated paracrine FGF2 and the C-terminal end of one of two endocrine FGFs, either FGF21, which is involved in glucose and lipid metabolism, or FGF23 which is involved in phosphate and vitamin D homeostasis [65]. The mutated chimeric growth factor with the FGF23 C-terminal domain was capable of decreasing serum phosphate levels and decreasing gene expression of CYP27B1, which catalyzes the conversion of vitamin D to its active form *in vivo* in mice. The chimeric growth factor with the FGF21 C-terminal domain was able to potentiate the hypoglycemic effect of insulin *in vivo*. This research provides an example of protein engineering for

treatment of metabolic disorders such as hyperphosphatemia and type-2 diabetes or other obesity-related diseases.

3.5 Protein engineering for tracing genetically-modified organism-sourced compounds

Transgenic peppermint (*Mentha x piperita*) plants used to produce essential oils were engineered to produce higher levels of (+) limonene to enable tracing of these oils through purification [66]. To accomplish this, the (–) limonene synthase (LS) from *Mentha spicata* was mutated and screened for (+) limonene production *in vitro*. The identified LS (M548C) mutant produced (+) limonene at 48.4% and (–) limonene at 8.9% of the total products, as compared to the wild-type LS, which produced 0.2% (+) limonene and 95.9% (–) limonene. When the LS (M548C) mutant was placed into the transgenic peppermint plants, the (+) limonene level increased by approximately 50% over that produced by non-transgenic plants, thus enabling tracing of transgenic-organism produced essential oils.

3.6 Transcription factor engineering for metabolic engineering

As described above, mutating transcription factors can have the beneficial effect of modifying expression levels of a large number of proteins, thereby causing a global change of cellular phenotype. A recent example used transcription factor libraries to improve *E. coli* tolerance to 1-butanol [67]. The cyclic AMP receptor protein (CRP) was mutated using error-prone PCR and four mutants with elevated tolerance were isolated [67]. A second library was then constructed by DNA shuffling of the first four mutants, and screening yielded a strain that grew twice as fast as wild-type in the presence of 1.2% 1-butanol. The expression profile of a number of CRP-controlled genes in the mutant strain was different from wild-type when exposed to 1-butanol. Separately, a library of zinc-finger DNA-binding proteins fused to CRP was used to select for a strain of *E. coli* with butanol tolerance up to 1.5% [68].

Another example is the improvement of lycopene production and ethanol tolerance in *E. coli* by screening a library of RpoD (the sigma70 transcription factor) mutants [69]. Similarly, an RpoD mutant library was screened in *L. plantarum* to identify a strain with tolerance to low pH and high lactic acid concentration [70]. Additionally, mutations in two transcription factors, RpoD and RpoS led to a phenotype of high hyaluronic acid production in *E. coli* [71]. In yeast, a library of mutants of the Spt15 transcription factor was screened to identify a mutant strain that used xylose more effectively than wild-type [44]. Screening a library of Spt15 mutants also produced a strain with increased ethanol tolerance [43].

Precise control of expression of one or several proteins can be attained by engineering of specific, heterologous, transcriptional regulators that are orthogonal to cellular metabolism. Engineering of such regulatory proteins can adjust their activity with regard to activator or inhibitor concentration. For example, variants of LuxR were engineered by directed evolution to respond to decanolyhomoserine lactone instead of the natural effector 3-oxo-hexanoylhomoserine lactone [72]. Separately, residues in the binding pocket of AraC were mutated to switch the binding specificity of this transcriptional regulator from L-arabinose to mevalonate, creating a reporter system that enabled screening for increased mevalonate production in metabolically-engineered *E. coli* [73].

4 Concluding Remarks

The examples detailed above have shown the promise of protein engineering for metabolic engineering, synthetic biology, and industrial biotechnology. Certain techniques, such as changing cofactor specificity or substrate specificity of enzymes have been wholeheartedly embraced by the field of metabolic engineering since balance of cofactors along a pathway and recursive action of proteins to elongate a product have a number of applications. Other methods of protein engineering are not yet as developed, but will likely find greater application by metabolic engineers over the next few years. Continued development of modeling tools such as Rosetta [28], engineered organisms allowing for more effective use of NNAs [48], and development of artificial transcription factors as global phenotype modification tools [42] reveals a strong potential for growth and are among the tools that will greatly expand the number of possible products that can be produced by engineered organisms.

Acknowledgments

We would like to thank Prof. K. N. Houk and Elizabeth L. Noey for critical reading and suggestions. This work was supported in part by the Electrofuel program of Advanced Research Projects Agency–Energy (ARPA-E; DE-AR0000085) and NSF MCB-0903955. R.J.M. is supported by an NRSA Kirschstein Postdoctoral Fellowship (5F32GM099277-02).

Abbreviations

PCR	Polymerase Chain Reaction
CoA	coenzyme A
PDB	Protein Data Bank
NAD(P)	nicotinamide adenine dinucleotide (phosphate)

References

1. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, et al. Engineering the third wave of biocatalysis. *Nature*. 2012; 485:185–194. [PubMed: 22575958]
2. Lee JW, Na D, Park JM, Lee J, et al. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nature chemical biology*. 2012; 8:536–546.
3. Pleiss J. Protein design in metabolic engineering and synthetic biology. *Curr Opin Biotechnol*. 2011; 22:611–617. [PubMed: 21514140]
4. Foo JL, Ching CB, Chang MW, Leong SS. The imminent role of protein engineering in synthetic biology. *Biotechnol Adv*. 2012; 30:541–549. [PubMed: 21963685]
5. Cobb RE, Si T, Zhao H. Directed evolution: an evolving and enabling synthetic biology tool. *Curr Opin Chem Biol*. 2012; 16:285–291. [PubMed: 22673064]
6. Brustad EM, Arnold FH. Optimizing non-natural protein function with directed evolution. *Curr Opin Chem Biol*. 2011; 15:201–210. [PubMed: 21185770]
7. Rao CV. Expanding the synthetic biology toolbox: engineering orthogonal regulators of gene expression. *Curr Opin Biotechnol*. 2012; 23:689–694. [PubMed: 22237017]
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215:403–410. [PubMed: 2231712]
9. Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, et al. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008; 36:W5–W9. [PubMed: 18440982]

10. Caspi R, Altman T, Dreher K, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 2012; 40:D742–D753. [PubMed: 22102576]
11. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 2006; 34:D354–D357. [PubMed: 16381885]
12. Chang A, Scheer M, Grote A, Schomburg I, Schomburg D. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res.* 2009; 37:D588–D592. [PubMed: 18984617]
13. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007; 35:W182–W185. [PubMed: 17526522]
14. Plata G, Fuhrer T, Hsiao TL, Sauer U, Vitkup D. Global probabilistic annotation of metabolic networks enables enzyme discovery. *Nature chemical biology.* 2012
15. Tan Y, Liao JC. Metabolic ensemble modeling for strain engineers. *Biotechnol J.* 2012; 7:343–353. [PubMed: 22021171]
16. Copeland WB, Bartley BA, Chandran D, Galdzicki M, et al. Computational tools for metabolic engineering. *Metab Eng.* 2012; 14:270–280. [PubMed: 22629572]
17. Mulholland AJ. Computational enzymology: modelling the mechanisms of biological catalysts. *Biochem Soc Trans.* 2008; 36:22–26. [PubMed: 18208378]
18. Acevedo O, Jorgensen WL. Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc Chem Res.* 2010; 43:142–151. [PubMed: 19728702]
19. Lan EI, Liao JC. ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proc Natl Acad Sci U S A.* 2012; 109:6018–6023. [PubMed: 22474341]
20. Shen CR, Lan EI, Dekishima Y, Baez A, et al. Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Appl Environ Microbiol.* 2011; 77:2905–2915. [PubMed: 21398484]
21. Siegel JB, Zanghellini A, Lovick HM, Kiss G, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science.* 2010; 329:309–313. [PubMed: 20647463]
22. Lassila JK, Baker D, Herschlag D. Origins of catalysis by computationally designed retroaldolase enzymes. *Proceedings of the National Academy of Sciences of the United States of America.* 2010; 107:4937–4942. [PubMed: 20194782]
23. Chen Z, Zeng A-P. Protein design in systems metabolic engineering for industrial strain development. *Biotechnol. J.* 2013:8.
24. Chen Z, Wilmanns M, Zeng AP. Structural synthetic biotechnology: from molecular structure to predictable design for industrial strain development. *Trends Biotechnol.* 2010; 28:534–542. [PubMed: 20727604]
25. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. *Proteins.* 2003; 52:609–623. [PubMed: 12910460]
26. Morris GM, Huey R, Lindstrom W, Sanner MF, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.* 2009; 30:2785–2791. [PubMed: 19399780]
27. Simons KT, Bonneau R, Ruczinski I, Baker D. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins.* 1999; (Suppl 3):171–176. [PubMed: 10526365]
28. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D. De novo enzyme design using Rosetta3. *PloS one.* 2011; 6:e19230. [PubMed: 21603656]
29. Jiang L, Althoff EA, Clemente FR, Doyle L, et al. De novo computational design of retro-aldol enzymes. *Science.* 2008; 319:1387–1391. [PubMed: 18323453]
30. Privett HK, Kiss G, Lee TM, Blomberg R, et al. Iterative approach to computational enzyme design. *Proceedings of the National Academy of Sciences of the United States of America.* 2012; 109:3790–3795. [PubMed: 22357762]
31. Stemmer WP. Rapid evolution of a protein in vitro by DNA shuffling. *Nature.* 1994; 370:389–391. [PubMed: 8047147]

32. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. 2009; 460:894–898. [PubMed: 19633652]
33. McCullum EO, Williams BA, Zhang J, Chaput JC. Random mutagenesis by error-prone PCR. *Methods Mol Biol*. 2010; 634:103–109. [PubMed: 20676978]
34. Harper MA, Chen Z, Toy T, Machado IM, et al. Phenotype sequencing: identifying the genes that cause a phenotype directly from pooled sequencing of independent mutants. *PloS one*. 2011; 6:e16517. [PubMed: 21364744]
35. Edelheit O, Hanukoglu A, Hanukoglu I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol*. 2009; 9:61. [PubMed: 19566935]
36. Labrou NE. Random mutagenesis methods for in vitro directed enzyme evolution. *Curr Protein Pept Sci*. 2010; 11:91–100. [PubMed: 20201809]
37. Carr PA, Wang HH, Sterling B, Isaacs FJ, et al. Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res*. 2012; 40:e132. [PubMed: 22638574]
38. Werkman JR, Pattanaik S, Yuan L. Directed evolution through DNA shuffling for the improvement and understanding of genes and promoters. *Methods Mol Biol*. 2011; 754:325–342. [PubMed: 21720962]
39. Machado HB, Dekishima Y, Luo H, Lan EI, Liao JC. A selection platform for carbon chain elongation using the CoA-dependent pathway to produce linear higher alcohols. *Metab Eng*. 2012; 14:504–511. [PubMed: 22819734]
40. Dietrich JA, McKee AE, Keasling JD. High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu Rev Biochem*. 2010; 79:563–590. [PubMed: 20367033]
41. Arnold, FH.; Georgiou, G. Directed Enzyme Evolution: Screening and Selection Methods. Humana Press; 2010.
42. Lanza AM, Alper HS. Global strain engineering by mutant transcription factors. *Methods Mol Biol*. 2011; 765:253–274. [PubMed: 21815097]
43. Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science*. 2006; 314:1565–1568. [PubMed: 17158319]
44. Liu CC, Schultz PG. Adding new chemistries to the genetic code. *Annu Rev Biochem*. 2010; 79:413–444. [PubMed: 20307192]
45. Zhang F, Cong L, Lodato S, Kosuri S, et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol*. 2011; 29:149–153. [PubMed: 21248753]
46. Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of *Escherichia coli*. *Science*. 2001; 292:498–500. [PubMed: 11313494]
47. Bianco A, Townsley FM, Greiss S, Lang K, Chin JW. Expanding the genetic code of *Drosophila melanogaster*. *Nature chemical biology*. 2012; 8:748–750.
48. Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science*. 2011; 333:348–353. [PubMed: 21764749]
49. Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*. 2010; 464:441–444. [PubMed: 20154731]
50. Park HS, Hohn MJ, Umehara T, Guo LT, et al. Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science*. 2011; 333:1151–1154. [PubMed: 21868676]
51. Ugwumba IN, Ozawa K, Xu ZQ, Ely F, et al. Improving a Natural Enzyme Activity through Incorporation of Unnatural Amino Acids. *J Am Chem Soc*. 2011; 133:326–333. [PubMed: 21162578]
52. Zhang K, Li H, Cho KM, Liao JC. Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:6234–6239. [PubMed: 20332210]

53. Stillman TJ, Baker PJ, Britton KL, Rice DW. Conformational flexibility in glutamate dehydrogenase. Role of water in substrate recognition and catalysis. *J Mol Biol.* 1993; 234:1131–1139. [PubMed: 8263917]
54. Atsumi S, Liao JC. Directed evolution of *Methanococcus jannaschii* citramalate synthase for biosynthesis of 1-propanol and 1-butanol by *Escherichia coli*. *Appl Environ Microbiol.* 2008; 74:7802–7808. [PubMed: 18952866]
55. Felnagle EA, Chaubey A, Noey EL, Houk KN, Liao JC. Engineering synthetic recursive pathways to generate non-natural small molecules. *Nat Chem Biol.* 2012; 8:518–526. [PubMed: 22596203]
56. Koon N, Squire CJ, Baker EN. Crystal structure of LeuA from *Mycobacterium tuberculosis*, a key enzyme in leucine biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America.* 2004; 101:8295–8300. [PubMed: 15159544]
57. Zhang K, Sawaya MR, Eisenberg DS, Liao JC. Expanding metabolism for biosynthesis of nonnatural alcohols. *Proceedings of the National Academy of Sciences of the United States of America.* 2008; 105:20653–20658. [PubMed: 19064911]
58. Marcheschi RJ, Li H, Zhang K, Noey EL, et al. A synthetic recursive "+1" pathway for carbon chain elongation. *ACS Chem Biol.* 2012; 7:689–697. [PubMed: 22242720]
59. Dobritzsch D, König S, Schneider G, Lu G. High resolution crystal structure of pyruvate decarboxylase from *Zymomonas mobilis*. Implications for substrate activation in pyruvate decarboxylases. *J Biol Chem.* 1998; 273:20196–20204. [PubMed: 9685367]
60. Bastian S, Liu X, Meyerowitz JT, Snow CD, et al. Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*. *Metab Eng.* 2011; 13:345–352. [PubMed: 21515217]
61. Atsumi S, Hanai T, Liao JC. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature.* 2008; 451:86–89. [PubMed: 18172501]
62. Atsumi S, Wu TY, Eckl EM, Hawkins SD, et al. Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by comparison of three aldehyde reductase/alcohol dehydrogenase genes. *Appl Microbiol Biotechnol.* 2010; 85:651–657. [PubMed: 19609521]
63. Xiong M, Deng J, Woodruff AP, Zhu M, et al. A bio-catalytic approach to aliphatic ketones. *Sci Rep.* 2012; 2:311. [PubMed: 22416247]
64. Berthold CL, Gocke D, Wood MD, Leeper FJ, et al. Structure of the branched-chain keto acid decarboxylase (KdcA) from *Lactococcus lactis* provides insights into the structural basis for the chemoselective and enantioselective carbonylation reaction. *Acta Crystallogr D Biol Crystallogr.* 2007; 63:1217–1224. [PubMed: 18084069]
65. Goetz R, Ohnishi M, Kir S, Kurosu H, et al. Conversion of a paracrine fibroblast growth factor into an endocrine fibroblast growth factor. *J Biol Chem.* 2012; 287:29134–29146. [PubMed: 22733815]
66. Lange BM, Mahmoud SS, Wildung MR, Turner GW, et al. Improving peppermint essential oil yield and composition by metabolic engineering. *Proceedings of the National Academy of Sciences of the United States of America.* 2011; 108:16944–16949. [PubMed: 21963983]
67. Zhang H, Chong H, Ching CB, Song H, Jiang R. Engineering global transcription factor cyclic AMP receptor protein of *Escherichia coli* for improved 1-butanol tolerance. *Appl Microbiol Biotechnol.* 2012; 94:1107–1117. [PubMed: 22466954]
68. Lee JY, Yang KS, Jang SA, Sung BH, Kim SC. Engineering butanol-tolerance in *Escherichia coli* with artificial transcription factor libraries. *Biotechnol Bioeng.* 2011; 108:742–749. [PubMed: 21404248]
69. Alper H, Stephanopoulos G. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng.* 2007; 9:258–267. [PubMed: 17292651]
70. Klein-Marcuschamer D, Stephanopoulos G. Assessing the potential of mutational strategies to elicit new phenotypes in industrial strains. *Proceedings of the National Academy of Sciences of the United States of America.* 2008; 105:2319–2324. [PubMed: 18252819]
71. Yu H, Tyo K, Alper H, Klein-Marcuschamer D, Stephanopoulos G. A high-throughput screen for hyaluronic acid accumulation in recombinant *Escherichia coli* transformed by libraries of engineered sigma factors. *Biotechnol Bioeng.* 2008; 101:788–796. [PubMed: 18500764]

72. Collins CH, Leadbetter JR, Arnold FH. Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat Biotechnol.* 2006; 24:708–712. [PubMed: 16715074]
73. Tang SY, Cirino PC. Design and application of a mevalonate-responsive regulatory protein. *Angew Chem Int Ed Engl.* 2011; 50:1084–1086. [PubMed: 21268200]
74. Rasila TS, Pajunen MI, Savilahti H. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment. *Anal Biochem.* 2009; 388:71–80. [PubMed: 19454214]
75. Tyagi R, Duquerroy S, Navaza J, Guddat LW, Duggleby RG. The crystal structure of a bacterial class II ketol-acid reductoisomerase: domain conservation and evolution. *Protein Sci.* 2005; 14:3089–3100. [PubMed: 16322583]

Figure 1. Scheme of protein engineering for metabolic engineering

(a) Computational pathway prediction and metabolic network analysis. 2-ketobutyrate (2KB), homoalanine (HAAla), and amino acids (three-letter codes) are indicated. Enzymatic pathways can be identified from defined inputs to outputs. After pathway identification, enzymes must be selected for engineering (example circled). (b) QM modeling of enzyme reaction. The chemical reaction from substrate(s) to transition state(s) to product(s) should be analyzed to determine the feasibility of the desired reaction. Reprinted and adapted with permission from Copyright 2012 American Chemical Society [58]. (c) If available, the enzyme structure (PDB: 1SR9 [56] used here) can be modeled with the transition state model using programs such as Rosetta [28] or PyMol (The PyMOL Molecular Graphics System, Version 1.3r1, Schrodinger, LLC) to determine which amino acid residues to mutate. Modeling the designed mutations and analyzing the protein *in silico* provides a check of their effects on protein folding, stability, and electrostatic potential. (d) Mutant proteins must be generated either individually or in a combinatorial library and analyzed by either a growth-based selection or by using an assay-based screen. After identifying a functional mutant with the desired properties, this protein can finally be used to produce the desired metabolic product *in vivo*.

Figure 2. Methods for generating diversity in protein coding sequences

A selection of methods in the areas of Random Diversity, Targeted Diversity, and their interface are shown. Random Diversity generation is useful when the structure of a protein is not known. In error-prone PCR a low-fidelity DNA polymerase is used to mis-incorporate nucleotides at a rate of approximately 1-5% [74]. Whole cell mutagenesis and evolution can be used to generate random diversity in gene expression and protein function. Artificial transcription factor libraries can also be used to generate random global diversity in protein expression. In contrast, Targeted Diversity is most useful when the protein structure is known. Site-saturation mutagenesis targets a specific site or sites in the protein for random mutation, while site-directed mutagenesis directly defines the amino acid mutation(s). MAGE is a targeted technique that can be used on a genome-level scale. This method is particularly useful for mutating many proteins at the same time. The Interface between Random and Targeted Diversity is DNA Shuffling. This technique is a way to leverage previously generated mutant libraries to create new combinatorial libraries rapidly.

Figure 3. Examples of protein engineering for metabolic production of useful compounds

Mutant proteins are indicated with asterisks and are circled. The citramalate pathway (upper left) and the Threonine pathway (center left) both produce the metabolic intermediate 2-ketobutyrate, which can be converted into L-homoalanine by a mutant GDH (GDH*, model based on PDB: 1BGV [53]). 2-ketobutyrate can also be converted into alcohols (pathways on lower left) by action of LeuA, LeuCD, LeuB, KIVD, and ADH. A mutant LeuA (LeuA*, model based on PDB: 1SR9 [56]) is required for recursive "+1" chain elongation of ketoacid substrates, and a mutant KIVD (KIVD*) is required to produce 5-8 carbon alcohols. Isobutanol and branched-chain acids (pathways on right) are made by co-opting the valine and leucine biosynthesis pathways. For production of isobutanol, a mutant IlvC (IlvC*, model based on PDB: 1YRL [75]) was engineered to use NADH instead of NADPH. For branched-chain acid production, KIVD, IPDC, and an ALDH were engineered.