J Ind Microbiol Biotechnol (2015) 42:349–360
DOI 10.1007/s10295-014-1541-1

SIMB
Society for Industrial Microbiology and Biotechnology

An integrated biotechnology platform for developing sustainable chemical processes

Nelson R. Barton · Anthony P. Burgard · Mark J. Burk · Jason S. Crater · Robin E. Osterhout · Priti Pharkya · Brian A. Steer · Jun Sun · John D. Trawick · Stephen J. Van Dien · Tae Hoon Yang · Harry Yim

Received: 6 October 2014 / Accepted: 7 November 2014 / Published online: 22 November 2014

Abstract Genomatica has established an integrated computational/experimental metabolic engineering platform to design, create, and optimize novel high performance organisms and bioprocesses. Here we present our platform and its use to develop *E. coli* strains for production of the industrial chemical 1,4-butanediol (BDO) from sugars. A series of examples are given to demonstrate how a rational approach to strain engineering, including carefully designed diagnostic experiments, provided critical insights about pathway bottlenecks, byproducts, expression balancing, and commercial robustness, leading to a superior BDO production strain and process.

Keywords 1,4-butanediol · Metabolic engineering · Modeling · Fermentation

Introduction

Industrial biotechnology strives to leverage the advantages of biological processes for the manufacture of a wide range of valuable products, including chemicals and fuels. Conventional production processes based on fossil feedstocks (oil, natural gas, coal) suffer from insecure raw material supply and high price volatility, high capital and energy costs, and poor sustainability and safety profiles. Alternative processes involving the use of engineered microorganisms can provide solutions to these industry-wide challenges [6]. For example, biotechnology can offer broader feedstock flexibility (sugars, C1 feedstocks, glycerin, fats and oils, etc.), direct and selective conversion of primary feedstock to product in a single unit operation (fermentation), lower capital and energy costs, lower by-products and waste, and safer more sustainable processes. Despite these promising benefits and the burgeoning interest in biotechnology, widespread commercial use of these technologies and processes has been constrained by long timelines and high costs for process development. Many different tools and methods are required to engineer microorganisms and processes that perform well at commercial scale. Moreover, the interdisciplinary nature of this research requires scientists and engineers with disparate skills to work together effectively.

In an effort to address these challenges and unleash the potential of biotechnology, we have built a fully integrated technology platform that enables an efficient workflow for engineering microorganisms and developing processes for industrial production of chemicals. Our platform (Fig. 1) combines predictive computational modeling with advanced laboratory technology and process engineering in a uniquely integrated workflow that accelerates development of economically attractive bio-manufacturing processes. We previously demonstrated how a model-based approach can guide the initial strain and pathway design [24]. Here we emphasize the importance of a focused systems biology effort that can provide a detailed understanding of metabolic and physiological changes that occur as strains are engineered for improved performance. 'Omics' data provides global information on gene and protein expression, metabolite expression, and intracellular fluxes. Together, this data enables a diagnosis of the strain: what is working as intended, what is not working well, and what

Special Issue: Metabolic Engineering.

N. R. Barton · A. P. Burgard · M. J. Burk · J. S. Crater · R. E. Osterhout · P. Pharkya · B. A. Steer · J. Sun · J. D. Trawick · S. J. Van Dien (✉) · T. H. Yang · H. Yim
Genomatica, Inc., 4757 Nexus Center Drive, San Diego, CA 92121, USA
e-mail: svandien@genomatica.com

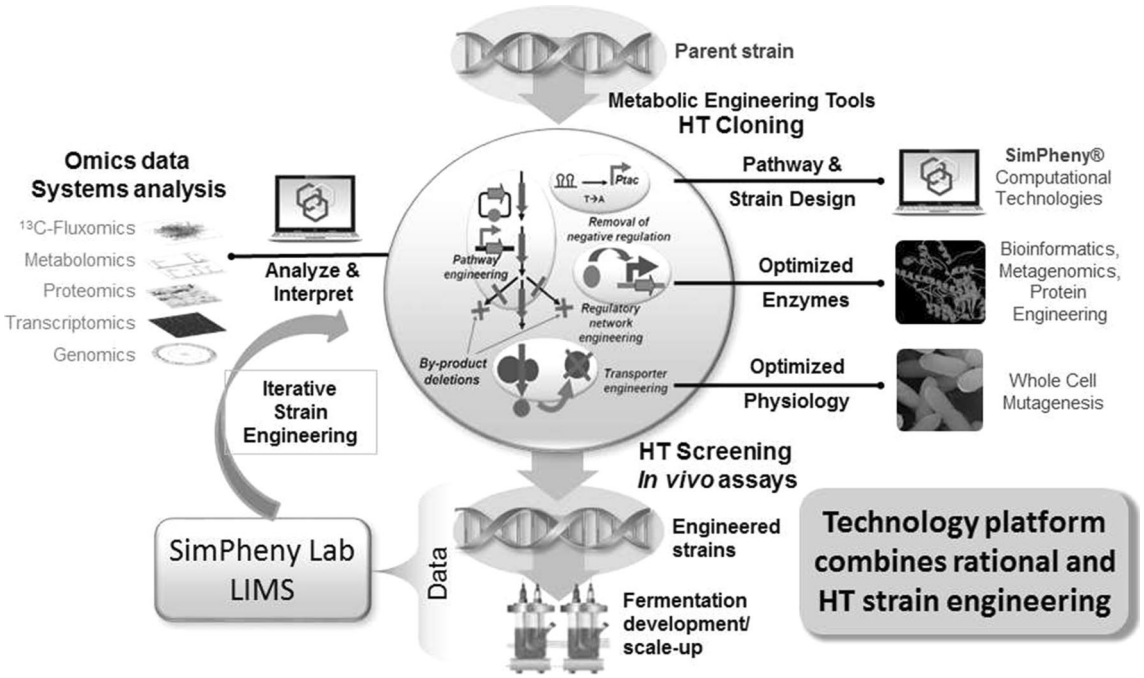

Fig. 1 Systems-based technology platform for strain engineering leads to rapid advances towards commercial metrics

might be limiting performance. In many cases, diagnostic experiments are designed to answer specific questions and while these can take time, the information often is invaluable for overcoming metabolic obstacles to high level production. In addition, robotic high-throughput cloning methods [9] and enzyme engineering technologies create thousands of strain variants, which are screened for product titer and rate improvements in small-scale 96-well plate assays. The best performing strains are validated and optimized in fermentation experiments; the resulting data provide a complete accounting of all carbon and electrons, which can be used as constraints for enhancing the accuracy of our metabolic models. The combined results can lead to a new set of proposed manipulations for strain improvement, which can be implemented using our high-throughput cloning tools, and the iterative cycle starts again. Below we describe the key elements and demonstrate the utility of our technology platform through selected examples encountered along the path to a commercial BDO strain and process.

Developing a pathway for 1,4-butanediol production in *E. coli*

Early development of the BDO production strain emphasized computational capabilities that enabled the achievement of several g/L BDO in fermentations that lasted approximately five days [24]. The work involved identifying and constructing new biochemical pathways to BDO as well as modeling and engineering a host strain that ideally would optimize carbon flux to our pathway, balance energy and redox, and help to eliminate unwanted by-products. The pathway was designed using in-house SimPheny™ Biopathway Predictor software that found all potential pathways to BDO. These pathways were then prioritized on the basis of factors that would increase the probability of success such as maximizing theoretical yield, minimizing the length of the pathway, having the fewest non-native steps, reducing novel steps, and ensuring the pathway was thermodynamically favorable. The selected route to BDO first required making 4-Hydroxybutyrate (4HB) from glucose via the Tricarboxylate acid (TCA) cycle intermediates α-ketoglutarate and/or succinyl-CoA (Fig. 2). The pathway utilized CoA-dependent succinate semialdehyde dehydrogenase (reducing succinyl-CoA) to generate succinyl-semialdehyde, which was then reduced by 4-hydroxybutyrate dehydrogenase to generate 4HB. In the next step, CoA was added to activate 4HB via 4-hydroxybutyrate-CoA transferase. The final two reduction steps required an aldehyde dehydrogenase and an alcohol dehydrogenase to reduce 4HB-CoA to produce BDO. For these last two steps, no enzymes had previously been identified to function on the particular pathway intermediates, so a small panel of aldehyde and alcohol dehydrogenases was tested and the best enzymes chosen. The genes for the heterologous pathway enzymes were cloned into two, compatible low copy plasmids under the control of IPTG inducible promoters.

Because the pathway from succinyl-CoA to BDO requires several reduction steps and requires a net input

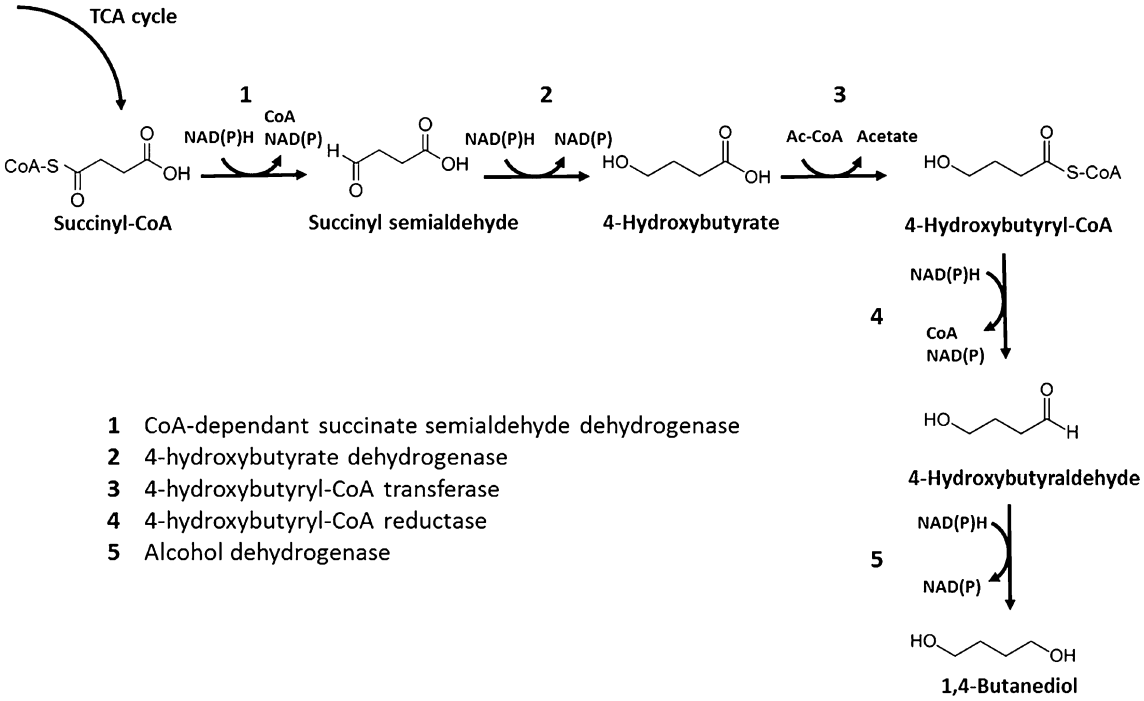

Fig. 2 BDO biosynthetic pathway used in *E. coli* BDO production strains

of energy, a strain that would balance both redox and ATP while channeling carbon flux through central metabolism and into the pathway was needed. Metabolic modeling and simulation indicated that making gene knockouts in alcohol dehydrogenase (*adhE*), pyruvate formatelyase (*pfl*), lactate dehydrogenase (*ldh*), and malate dehydrogenase (*mdh*) would couple BDO production to growth in an oxygen-limited environment. These deletions reduce formation of the byproducts ethanol, formate, lactate and succinate, as well as conserve redox for use in the BDO pathway. Additionally, the pathway required a more anaerobic environment to provide the redox necessary for the pathway. However, high NADH levels tend to inhibit enzymes that allow flux into and through the TCA cycle. To solve this issue, the *E. coli lpdA* gene was replaced with *Klebsiella pneumonia lpdA* that contained the mutation D345K. The analogous mutation in the *E. coli* gene had been reported to engender better activity at high NADH levels [12], which are predicted to occur at low aeration conditions that would favor BDO production. LpdA is a component of pyruvate dehydrogenase and α-ketoglutarate dehydrogenase; therefore, this was predicted to increase flux into the TCA cycle (pyruvate to acetyl-CoA) as well as improve the step from α-ketoglutarate to succinyl-CoA. Likewise, the native *gltA* gene encoding citrate synthase was mutated with a R163L change to make that enzyme more active when exposed to high NADH levels. To further increase flux into the oxidative TCA cycle, the transcriptional repressor of many TCA genes, *arcA*, was inactivated [11, 17].

The BDO pathway plasmids were transformed into the engineered host strain and assayed for BDO production in vivo. In small-scale experiments, we demonstrated feedstock flexibility and BDO was successfully made using a variety of feedstocks including glucose, sucrose, xylose, and biomass sugars (mixed C5/C6). There were differences in growth with the various feedstocks (glucose and sucrose providing the best growth), but the concentrations of BDO produced from each feedstock were roughly equivalent. When the best performing strain was tested in fed-batch fermentations, we were able to produce over 18 g/L BDO in a little over five days. In these fermentations, BDO was the primary product of the strain with acetate, pyruvate and 4-HB making up a majority of the by-products. The remainder of this review provides examples of how bottlenecks were overcome in order to improve performance beyond that previously reported and achieve commercially viable performance.

A systems biology approach to pathway optimization

Genomatica's systems biology platform uses extensive 'omics' tools in conjunction with small-scale screening and fermentation, interpreted through our computational models of metabolism to derive important biological knowledge and to drive experimental design. The most fundamental

'omics' dataset is the genome sequence. Next-generation sequencing technologies (e.g., the Illumina Miseq, http://systems.illumina.com/systems/miseq.ilmn) provide rapid and inexpensive methods to collect valuable data on all major milestone strains at regular intervals throughout the strain engineering process. The genome sequence verifies that the intended genetic manipulations were made and also detects unintended mutations that occurred during the cloning process. Transcriptomics (RNAseq, microarray, qPCR) and proteomics (iTRAQ, MRM) [7, 14, 20] show the expression levels of endogenous, heterologous and manipulated genes and proteins. Combined with promoter libraries covering a large range of expression strengths, and ribosome binding site prediction software [16], expression levels can be tuned for optimal performance and bottlenecks can be identified as those steps most sensitive to expression changes. Furthermore, global transcriptome and proteome profiling can be used to assess physiology and infer overall cell health, based on the expression of regulons associated with global stress responses. ^{13}C-flux analysis (fluxomics) measures the overall carbon flux distribution, thereby demonstrating whether strain manipulations had their desired metabolic effect. Finally, metabolomics provides relative concentrations of key pathway metabolites to diagnose bottlenecks as well as redox cofactor ratios and energy charge. Each of these data types has particular advantages, and all have been used in an integrative fashion to provide key global insights that are used to drive strain improvements. Overall, the systems-based platform is managed by a focused Systems Team that applies consistent data analysis workflows to provide high-quality information leading to a deeper understanding of the biological system. Systems results and interpretations are then integrated with computational models and phenotypic data to generate rational strain engineering targets for subsequent rounds of strain improvement.

One application of the systems platform was to guide the rational design of industrially-relevant properties in lab-scale BDO production organisms. First generation BDO production strains relied on plasmid-based and IPTG-inducible expression of BDO pathway genes [14]. Plasmid-based expression systems are undesirable for industrial-scale production because plasmid copy number control is not as tight as that of the chromosome and selection for maintenance of the plasmid typically requires addition of costly antibiotics [8]. The cellular cost of expression of plasmid-based genes can also introduce undesirable variation in gene copy number and strain stability [3]. The dependency of pathway expression on the inducer, Isopropyl β-D-1-thiogalactopyranoside (IPTG) also is not scalable from a cost standpoint. Developing an organism that is suitable for industrial-scale production required reconfiguring the BDO pathway to eliminate the requirement for costly fermentation additives, while optimizing enzyme levels to improve BDO production and long-term strain stability. BDO pathway reconfiguration entailed: (1) chromosomal integration of all pathway genes; (2) replacement of inducible promoters with constitutive promoters; and (3) expression and RBS balancing of pathway transcripts and proteins, respectively.

Chromosomal integration of pathway genes was pursued to improve organism genetic robustness to long-term passage, remove the costly addition of antibiotics, and reduce variation in gene copy number. Integration provided the additional advantage of freeing space on plasmids for testing overexpression of new gene targets to reduce byproducts and further improve BDO production. Chromosomal integration sites were identified by a combination of transposon-based insertion and rational targeting based on expression profiles of neighboring genes. Sites were also chosen where genes had previously been deleted. The integration sites varied in terms of chromosomal location and genomic context, as did the resulting expression levels of genes integrated at different loci under identical promoters. Replacement of inducible promoters with constitutive promoters was undertaken to increase strain stability and reduce cost during scale-up. Non-inducer-based conditional gene expression systems also were evaluated but ultimately not pursued. Promoter libraries were generated by random mutation of the −10 and −35 sequences of a parent constitutive promoter, and the strength tested by measurement of a reporter gene. Individual pathway genes and synthetic operons of multiple genes were constructed under the control of a range of constitutive promoters, ribosome binding sites and terminators. Constitutive overexpression of unregulated, heterologous genes can sometimes confer a selective disadvantage, resulting in poor growth and genetic instability. Strain stability, measured by ability to maintain BDO production after short-term (2 weeks) serial passaging of clonal isolates, was thus applied as a preliminary screen for identifying suitable promoters to express each pathway gene. Screening for strain stability effectively narrowed down the range of promoter levels for testing in the next round of screening, and also provided insight into expression sensitivity of each gene. For example, after 10 days of passaging a single colony of a strain expressing *ald* from a strong constitutive promoter, >95 % of clones lost the ability to produce BDO, apparently due to toxicity of high *ald* expression levels. Strains with weaker constitutive promoters were stable and advanced to the next round of screening.

Integrated and constitutively expressed genes of top-performing strains were then optimized for expression and protein levels to improve BDO production with a minimal metabolic burden on the host. A systematic and combinatorial approach was required as altering expression of one

gene frequently had unintended consequences on other genes that were difficult to predict. For example, early constitutive strains produced less BDO and more byproducts than their inducible counterparts, supporting the hypothesis that the expression level of one or more pathway genes was flux-limiting. A comparative transcriptomics experiment demonstrated that replacement of two inducible promoters with constitutive promoters, driving succinic semialdehyde dehydrogenase (035) and 4HB dehydrogenase (036), increased the expression level of the corresponding downstream genes 50- and 8- fold, respectively. Coincident with this change, the expression level of a downstream pathway gene (*cat2*) was reduced 4-fold. Replacement of the 036 promoter with a weaker promoter, in combination with a stronger *cat2* promoter, resulted in a significant performance improvement. The protein level of 035 was subsequently reduced 50-fold with no detrimental effect on BDO production, freeing up cell resources, presumably, for cell maintenance and expression of other heterologous proteins.

Routine assessment of strain-to-strain differences using systems biology tools such as whole-genome sequencing, metabolite profiles, transcriptomics and proteomics has enabled a top-down approach to BDO pathway optimization. The addition of targeted intracellular metabolomics and multiple reaction monitoring (MRM) proteomics [7, 14] to the systems biology platform helped to identify bottlenecks in later strains, and better quantify the effect genotype on phenotype.

^{13}C isotopic methods

The use of isotopic labeling including ^{13}C-based metabolic flux analysis has become a standard tool for metabolic engineering and industrial strain development. From a systems biology standpoint, any metabolic process comprises various network components such as genes, proteins, metabolites, and enzyme reactions, with interactions between the network components. Among those, in vivo metabolic reactions are the ultimate products of the highly complex metabolic processes, which can be quantitatively and comprehensively described by metabolic fluxes and intracellular metabolite concentrations. In this sense, gaining quantitative information on metabolic reactions in vivo and characterizing metabolic phenotypes at the reaction level is of importance for industrial strain engineering.

The most advanced form of isotopic labeling application is ^{13}C-based metabolic flux analysis (^{13}C-MFA), of which detailed computational and experimental protocols have been reported [15]. This technology can be applied to steady-state metabolic pathway quantification; e.g., to identify key branch point flux distributions, detect undesired carbon loss through competing pathways, and evaluate fluxes through native, engineered, and heterologous pathways. Also, ^{13}C-MFA can be applied to elucidate unknown, novel pathways [15] which may play a key role in strain engineering. To get the most accurate information on metabolic reaction phenotypes, implementing parallel ^{13}C-labeling experiments using multiple ^{13}C-substrates is highly recommended [22, 23] as is careful integration with other systems level data.

Direct implementation of ^{13}C-MFA through sampling an industrial fed-batch bioreactor may not always be feasible due to the high cost associated with ^{13}C-labeled compounds and significant ^{13}C-label dilution during the process. To circumvent these challenges, one can employ a highly sensitive method such as isotope-ratio monitoring mass spectrometry and apply a small amount of ^{13}C-labeled compound [25] or implement ^{13}C-MFA in an indirect way using smaller scale bioreactors [19, 22].

Even without employing the complex model-based ^{13}C-MFA, crucial information about metabolic pathways for strain optimization can be obtained. For instance, universally ^{13}C-labeled compounds can be applied in order to unveil dynamic characteristics of metabolic pathways, an application example is depicted in Fig. 3. Herein, we conducted a simple ^{13}C-experiment using [U-$^{13}C_6$]-glucose in the early phase of BDO production strain development to identify metabolic bottleneck steps in the BDO pathway. Immediately after spiking [U-$^{13}C_6$]-glucose during fed-batch fermentation, intracellular ^{13}C propagation dynamics from glucose to BDO were measured by taking time-course samples. The samples were rapidly quenched for intracellular metabolite extraction and subsequent ^{13}C labeling analysis using LC/MS [2]. The intracellular ^{13}C propagation to 4HB-CoA was not significantly different from other upstream metabolites such as alanine and glutamate within the sampling interval (Fig. 3c). In contrast, the ^{13}C propagation to intracellular BDO was significantly slower than that of 4HB-CoA. A meaningful detection of ^{13}C glucose carbon in BDO occurred 50 min after the [U-$^{13}C_6$]-glucose pulse into the bioreactor. This experiment clearly suggested that the last two enzyme steps for BDO production, the Ald and Adh-catalyzed steps were the bottleneck steps for BDO production. These steps therefore became the key targets for enzyme engineering efforts; the Ald case is described below.

Enzyme discovery and engineering to maximize pathway flux and reduce byproducts

The route to BDO from succinyl-CoA requires 5 sequential steps, none of which were known to be catalyzed by native *E. coli* enzymes. Moreover, the final two steps to BDO, a non-naturally occurring chemical, require enzymes to catalyze transformations involving non-natural metabolites. As such, enzymes had to be discovered from nature

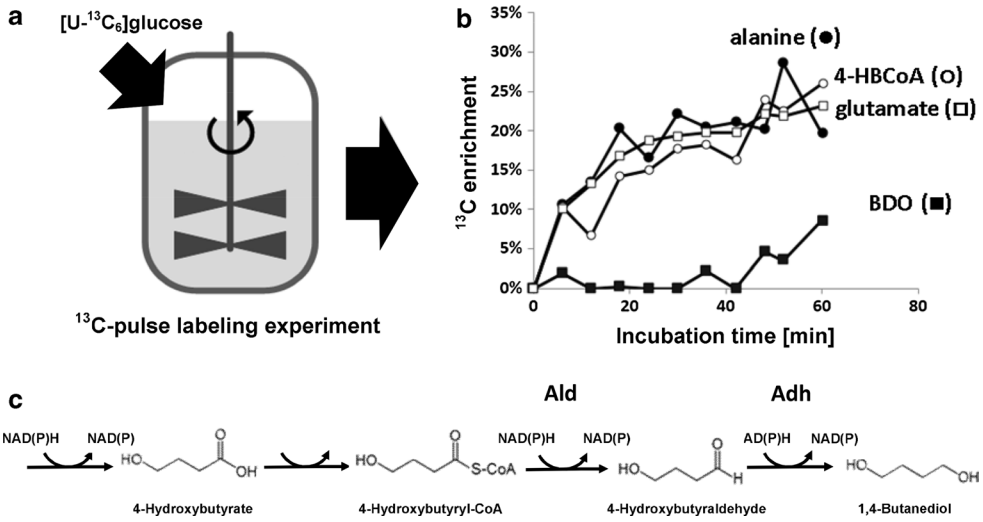

Fig. 3 **a** ^{13}C-pulse labeling experiment using [U-$^{13}C_6$] glucose in 2L bioreactor to measure ^{13}C propagation dynamics from glucose to BDO, **b** Intracellular dynamics of glucose carbon enriched in alanine, glutamate, 4HB-CoA, and BDO, **c** BDO pathway including the last two steps of aldehyde dehydrogenase Ald and alcohol dehydrogenase Adh

and heterologously expressed in *E. coli*. Multiple, and in some cases hundreds, of genes were identified, cloned and the gene products tested for catalytic activity for each step in the pathway. In general, these "pipelines" are developed to acquire, validate and characterize the genes for each step of the desired pathways.

In general, genes from sequenced genomes and metagenomic libraries are cloned from genomic DNA or constructed with codon optimization using gene synthesis. Gene expression is validated using SDS-PAGE or the Caliper–LabChip from Advanced Molecular Vision (http://www.perkinelmer.com/catalog/product/id/127229). Candidates are then tested in a small-scale in vivo screen for performance in a strain expressing part or all of a pathway that can produce the chemical of interest. Alternatively, the enzyme is expressed and directly assayed in vitro for activity on a specific pathway substrate. In many cases, the best enzymes found from nature still require further engineering to provide the performance needed to achieve maximum flux through the pathway. Performance improvements may include enhancing enzyme stability inside the cell, improving rate (k_{cat}) and substrate affinity (K_m), increasing specificity to reduce pathway by-products and alleviating inhibition by products or other pathway intermediates.

This complementary process of enzyme discovery and optimization was applied to one of our pathway steps, catalyzed by the enzyme aldehyde dehydrogenase (Ald). The Ald enzyme selected initially for our production strain was discovered by screening over 150 Ald gene candidates from nature followed by protein engineering using directed evolution. A strain (3368) which expresses all genes for a complete BDO pathway (with the exception of Ald) was constructed to screen gene candidates in a small-scale in vivo assay. Of the Ald enzymes tested, ten were found to produce BDO at significant levels, in the same range as

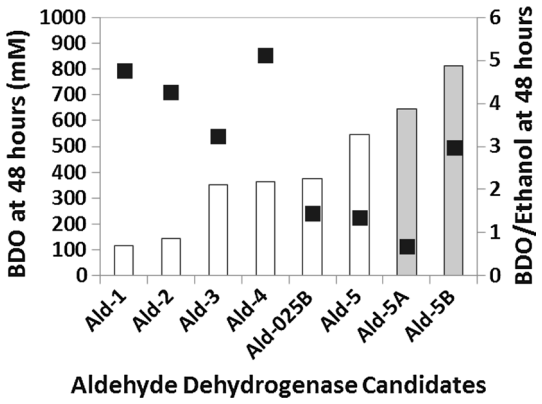

Fig. 4 BDO titer (*bars*) and BDO to ethanol ratio (*black squares*) in strain 3368 in fermentation at 2L scale for 6 heterologous Ald enzymes from nature, Ald-1, Ald-2, Ald-3, Ald-4, and Ald-025B, (*white bars*) and two engineered Ald enzymes Ald-5A and Ald-5B (*gray bars*)

the Ald used in the pathway described by Yim et al. [24] (designated Ald-025B). Six lead candidates from the small-scale screen were tested in fermentation at 2L-scale and BDO production was compared to Ald-025B. One Ald in particular (designated Ald-5) significantly outperformed Ald-025B by achieving 30 % higher BDO titer (Fig. 4).

However, Ald-5 also produced more ethanol (an unwanted by-product) than Ald-025B and even though Ald-5 made 30 % more BDO, the BDO to ethanol ratio of Ald-5 was slightly lower than the ratio for Ald-025B. Due to the *adhE* deletion in the host, Ald in the BDO pathway is the only enzyme in BDO production strains capable of converting acetyl-CoA (Ac-CoA) to acetaldehyde and it is thus responsible for ethanol production. These results were confirmed by ^{13}C-flux and metabolomics studies. To reduce ethanol and improve the ratio of BDO to ethanol, an Ald with higher selectivity for 4HB-CoA over Ac-CoA was

needed and therefore our protein engineering capabilities were implemented to create a more selective enzyme.

Ald-5 was engineered utilizing Genomatica's protein engineering technology platform which encompasses directed evolution, automated high-throughput screening and rational design based on sequence and structure information. When protein structural information is available, directed evolution combined with rational design is used to design small "smart" directed evolution libraries that can be screened very quickly. Smart libraries also enable the construction and screening of mutagenesis libraries where more than one site can be mutated simultaneously without creating an intractable library size. The high-throughput screening platform includes robotics for automated colony picking, liquid handling (plate to plate transfer, plate filling and aspiration), absorbance and fluorescence plate reading automated with robotic arms. The entire system utilizes barcoding and is integrated with our custom-built Laboratory Information Management System (LIMS).

In the case of Ald, which is an NAD(P)H-dependent enzyme, no structural information was available and therefore a random approach involving saturation mutagenesis combined with combinatorial recombination was used along with an automated high-throughput screen. Initially, an Ald variant (Ald-5A) that improved BDO production was identified but it also made more ethanol in the production strain (Fig. 4). However, using Ald-5A as a template for further directed evolution, a variant capable of production of even higher BDO titers along with significantly reduced ethanol was discovered. A production strain expressing the variant Ald-5B showed a significant decrease in ethanol production (Fig. 5a) and a noteworthy improvement in BDO production (Fig. 5b). At 48 h fermentation time, a 20 % improvement in BDO titer and an increase in the BDO to ethanol ratio from 0.5 to almost 5 compared to a strain expressing the parent Ald gene (Ald-5A) was measured (Fig. 4).

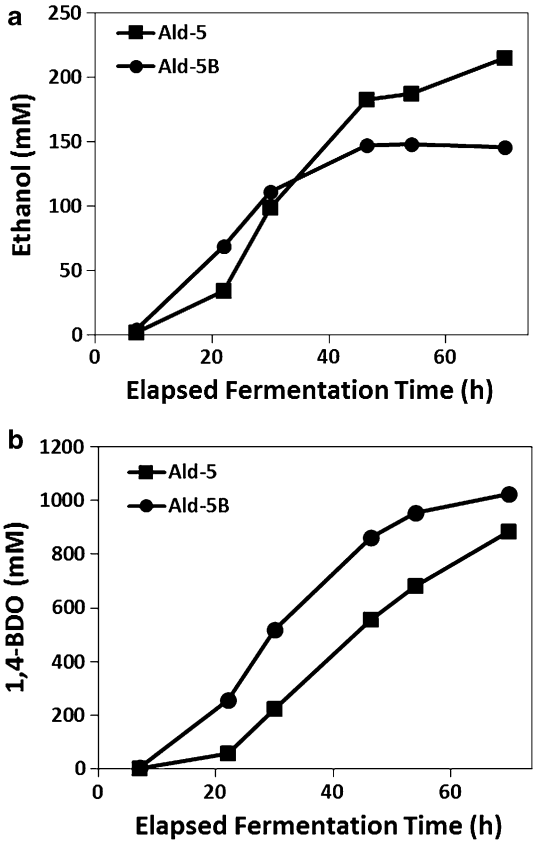

Fig. 5 Comparison of wild-type (Ald-5) and engineered (Ald-5B) performance in fermentation at 2L-scale. **a** Ethanol production verses elapsed fermentation time; **b** BDO production verses elapsed fermentation time

Host strain development and optimization

Hundreds of genetic changes have been made to the wild type *E. coli* K12 strain throughout the lifetime of the BDO project, and more than 50 have been retained in the current production strain. Host strain engineering work was conducted to achieve one or more of the following objectives: (1) reduction of byproducts, (2) increase of redox availability and energy efficiency, (3) introduction of phage resistance, and (4) diagnostic studies to understand the bottleneck(s) to product formation. It is illustrative to note that in the initial phases of the project, strain changes to reduce typical byproducts such as ethanol, lactate, and formate were critical for increasing BDO product yields. However, as the BDO yield and titers increased, strain engineering was not as straightforward. Frequently, multiple strain modifications had to be introduced to observe the desired effect. The redundancy in *E. coli* metabolism provides alternate enzymes and pathways that potentially can circumvent any manipulation, allowing cells to find alternate solutions to function without improving BDO production. Production of BDO (or any product of a heterologous pathway) also requires precise balancing of carbon, redox, and energy for cell biomass and product formation. It was particularly challenging to eliminate some of the byproducts such as 4HB, glutamate and ethanol because their production is more energetically favorable to the strain than BDO production. A few examples are described below to illustrate how multiple host strain modifications were required to accomplish the desired step-change in performance.

When the BDO pathway was first introduced into *E. coli* and strain changes made to delete genes associated with typical byproduct pathways, it was discovered that the strain had high respiration rates, producing more CO_2 than required by stoichiometry. Furthermore, ^{13}C labeling experiments demonstrated that carbon could flow

backwards through the pathway and into the TCA cycle. A transcriptomics study with the BDO-producing strain in the presence of high 4HB concentrations revealed very high transcript levels of two succinate semialdehyde dehydrogenases: (1) NAD-dependent dehydrogenase encoded by *sad* and (2) NADP-dependent dehydrogenase encoded by *gabD*. These could both convert succinate semialdehyde to succinate. Deletion of both dehydrogenases led to reduced respiration and increased production of 4HB. This reduction could not be accomplished by the deletion of just one dehydrogenase because the other one could carry out the same function.

Similarly, it was discovered that multiple reactions could lead to the formation of a byproduct. Even if one reaction is not as efficient as the other in forming the byproduct, it could still carry out the function. A case in point is gamma-butyrolactone (GBL), a byproduct formed from 4HB-CoA downstream of 4HB. Early in the course of BDO strain development it was discovered that GBL accumulated during fermentations. Unfortunately, GBL is a poor substrate for any of the downstream BDO pathway enzymes; accumulation of GBL is undesirable because it diverts carbons and reduces the achievable BDO yield, and also the accumulated GBL can interfere with BDO recovery as the two products have similar physical properties. It was hypothesized that the unstable intermediate 4HB-CoA decays to GBL in the cell. It was also possible that 4HB undergoes low pH-dependent lactonization spontaneously, though literature data indicates the lactonization rate to be very slow [5].

One course of action was to delete thioesterases that potentially could act on 4HB-CoA to facilitate cyclization to GBL. *E. coli* possesses CoA thioesterases reported to have broad substrate specificity on CoA substrates of C4 to C6 carbon chain lengths. Analogous activities can be found in other organisms. The top candidates for deletion were prioritized using data from a transcriptomics experiment measuring relative expression of thioesterases that were also reported in the literature to work on C4 to C6 substrates. Three gene candidates were initially deleted individually; two of these deletions, Δ*ybgC* and Δ*tesB*, showed efficacy in reducing GBL formation and were combined. Deleting both genes eliminated approximately 60 % of the GBL formation via the pathway (Fig. 6). It was also shown that the overexpression of *ybgC* increased GBL production significantly, further reinforcing the contribution of these thioesterases to GBL production.

To further reduce the concentration of GBL in the fermentation broth, we proposed to use enzymes that could convert GBL back into 4HB. The non-enzymatic GBL ↔ 4HB interconversion is too slow at neutral pH to rely upon spontaneous conversion during the timescales of BDO fermentation [5].Therefore, an enzyme to facilitate hydrolysis

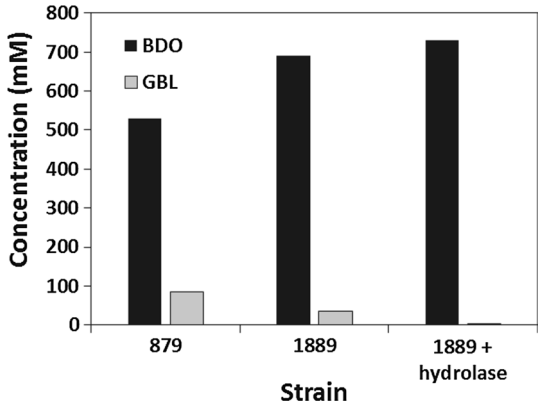

Fig. 6 The deletion of thioesterases contributing to GBL formation led to a more than 60 % reduction in GBL formation (strain 1889). A further integration of lactonase reduced the GBL formation to low mM quantities (strain 1889 + hydrolase)

of GBL to 4HB was needed. Literature searches led to the identification of a candidate lactonase from *Agrobacterium tumefaciens str.* C58 [4], and subsequent bioinformatics analysis provided additional candidates. The lead candidate was highly active and was integrated into a known non-essential site within the *E. coli* production strain genome behind a weak promoter. The resultant BDO strains produced less than 15 % of the GBL compared to the control strain in small-scale assays. The integration of the lactonase, combined with the deletions of both thioesterases, led to a strain that makes less than 1 mM of GBL in fermentations (Fig. 6).

Diagnostic host strain modifications can be important for identifying the most promising path to performance improvements; in the example here, to reduce accumulation of acetate, one of the key byproducts in the BDO pathway. Acetate is a challenging byproduct to eliminate since it is intimately associated with BDO production by virtue of the transferase enzyme (*cat2*) that converts 4HB to 4HB-CoA and uses Ac-CoA as the donor. Therefore, for every mole of BDO formed, one mole of acetate is formed. This acetate needs to be recycled efficiently back into Ac-CoA. *E. coli* has two natural mechanisms for acetate recycling [21]. Acetyl-CoA synthetase (Acs) requires two ATP equivalents to convert acetate to Ac-CoA, while acetate kinase/phosphotransacetylase (AckA-Pta) activates acetate to Ac-CoA by a net loss of one ATP molecule. AckA first converts acetate to acetyl-phosphate and then Pta converts the latter to Ac-CoA.

In order to determine whether Acs or AckA-Pta was responsible for acetate recycling in our BDO production strain, we deleted genes encoding these proteins separately and compared strain performance in fermentations relative to the parent strain. The *acs* deletion strain performed very similar to the parent, while the *ackA-pta* deletion strain

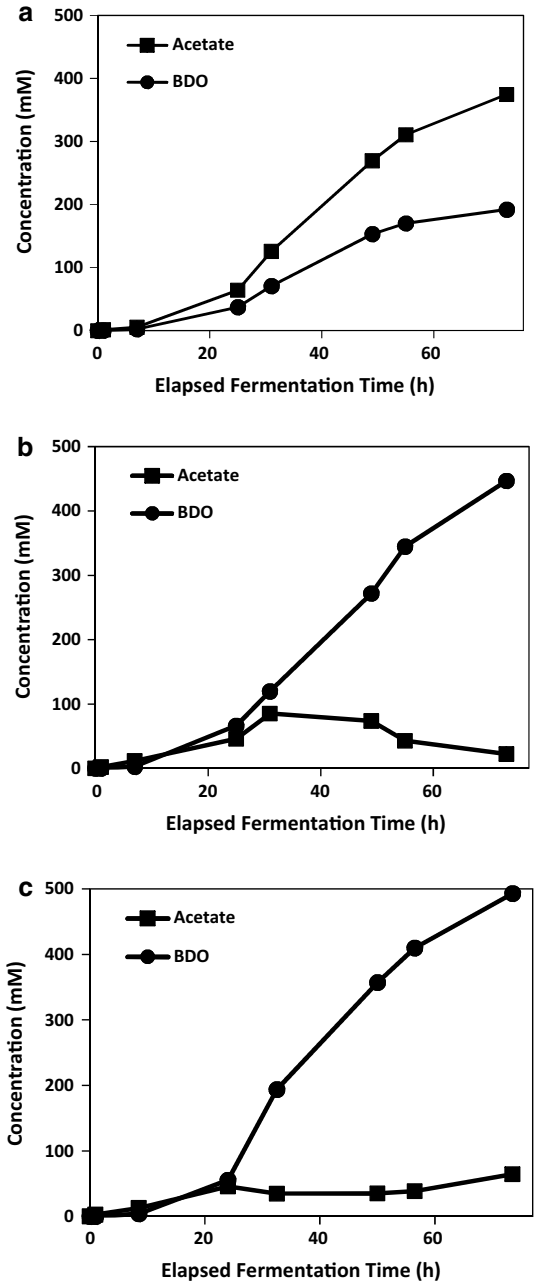

Fig. 7 A diagnostic fermentation experiment with individual deletions of *ackA-pta* (**a**) and of *acs* (**b**) showed that *ackA-pta* was responsible for the majority of acetate recycling in the BDO strains. Results for a similar experiment using the parent strain (**c**) is shown for comparison

accumulated three-fold more acetate and reduced BDO production by half (Fig. 7). Therefore, at least for this particular strain, AckA-Pta was responsible for the majority of acetate recycling to acetyl-CoA. The same result was validated for some of the more recent strains. With this information in hand, we optimized expression of *ackA-pta* to minimize acetate production and increase BDO yield.

Fermentation process optimization

Optimization of Genomatica's fermentation process is driven by economics, as the performance of the fermentation dictates the overall design and competitiveness of the BDO production process. Several key performance indicators, including titer, productivity, yield, and conductivity, are used to assess and optimize the cost position of the BDO process through the use of techno-economic analyses.

Titer is the concentration of BDO product in the fermentation broth, typically expressed in units of grams of product per liter of fermentation broth (g BDO/L). Higher fermentation titers lead to significantly reduced costs in both the upstream and downstream process sections due to reduced hydraulic load with correspondingly smaller equipment sizes and lower energy consumption.

Productivity is the rate of BDO production per unit fermentor volume, typically expressed in units of grams of product per liter of fermentation broth per hour (g BDO/L/hr). The productivity determines the total fermentation volume required to achieve target production capacity. Increasing productivity reduces the size or number of fermentors required and the associated capital costs. Further increases in fermentation productivity after a plant is designed or built can result in increased plant capacity.

Yield is the fraction of feedstock that is converted to product, typically expressed as a mass or mole fraction (g BDO/g sugar or mol BDO/mol sugar). The yield dictates how much feedstock is required to produce a given amount of product. The maximum theoretical yield is determined by the overall stoichiometry of the metabolic pathway, which is 0.50 grams of BDO per gram of glucose consumed. Since a portion of the carbohydrate feedstock is required for cell growth and maintenance energy, it is not possible to achieve maximum theoretical yields. Fermentations with low yields do not fully utilize the feedstock, or convert it to unwanted byproducts, both of which increase net feedstock cost. Low yields also increase separations costs in the downstream process since unused feedstock or unwanted byproducts must be removed from the BDO product.

Conductivity approximately reflects the total load of salts, organic acids, amino acids, peptides, and other small charged molecules. These byproducts need to be removed from the fermentation broth downstream, which directly impacts the economics of the downstream separations process. Though conductivity provides a good proxy for total byproduct load of charged species, it does not lend any insight into the concentration of uncharged byproducts, such as ethanol.

Understanding cost sensitivity with regards to these key indicators allows for optimization of fermentation process parameters to minimize the total production cost.

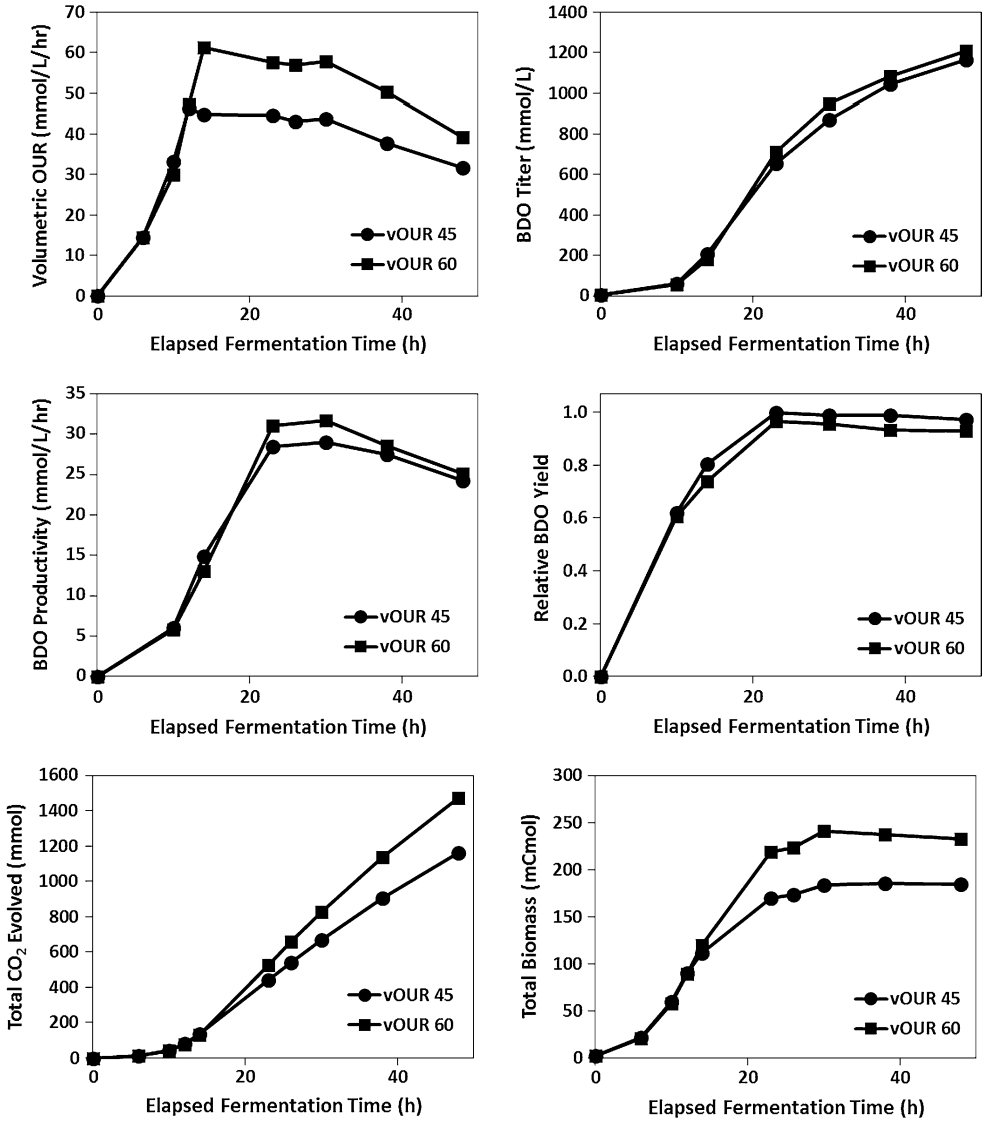

Fig. 8 Effect of oxygen uptake rate on BDO titer, productivity, and yield in strain 7,607

Genomatica's fermentation optimization process, which is conducted for every new milestone strain, involves assessing the effects of numerous process parameters; these include medium formulation, seed train, inoculum ratio, pH, temperature, oxygen transfer/uptake rate (OTR/OUR), substrate feeding strategy, and total fermentation time. Furthermore, scale-dependent process conditions, such as partial pressure of carbon dioxide (pCO_2), oxygen transfer rate gradients, and broth heterogeneity, are modeled to assist in the design of commercial scale-down experiments. These conditions are investigated to assess the production organism's sensitivity to anticipated commercial-scale conditions, which is essential for ensuring predictable performance across scales.

An example of fermentation process optimization involves the optimization of oxygen uptake rate (OUR). Due to the oxygen-limited nature of Genomatica's BDO fermentation process, the peak and shape of the OUR profile determine the amount of cell biomass propagated and the specific oxygen consumption rate, which influence the key performance metrics outlined above. Lowering peak OUR reduces the amount of biomass propagated, which improves BDO yield by reducing the fraction of carbohydrate feedstock used to make biomass. In addition, lower oxygen uptake reduces respiration and thus the amount of carbon lost to CO_2 via respiration (Fig. 8). Lowering the cell biomass concentration in the fermentation broth also reduces downstream biomass separation and disposal costs. Furthermore, reducing OUR lowers the amount of air required and heat generated which lowers energy consumption and reduces the required air compressor and chiller capacities. On the other hand, reducing OUR decreases BDO productivity as fewer cells are producing BDO per unit volume. The reduction in BDO productivity requires a longer fermentation time to achieve the same product titer, and larger fermentor volumes or

more fermentors to achieve the same production capacity at commercial scale. The use of techno-economic modeling allows for the determination of the optimum OUR, delivering the most favorable capital and operating cost scenario.

Conclusion and future perspectives

BDO is a challenging fermentation product for various reasons. First, it is a non-natural compound requiring multiple heterologous enzymes working on non-native substrates. Second, it requires high flux through the TCA cycle under oxygen-limited conditions to generate the necessary reducing equivalents to drive the final pathway steps. Third, acetate is made in stoichiometric proportion relative to BDO, and must be recycled into Ac-CoA. Furthermore, Ac-CoA is a substrate in two pathway steps (citrate synthase and *cat2*), as well as a substrate for the byproduct ethanol. Overcoming these challenges required a careful balance of pathway intermediates, ATP, and redox cofactors. The examples provided here demonstrate how Genomatica's technology platform was leveraged to surmount some of the critical issues in BDO strain development; ultimately, leading to commercial-scale production within 5 years of project start (http://www.genomatica.com/news/press-releases/successful-commercial-production-of-5-million-pounds-of-bdo/). We have achieved economically viable metrics (titer > 120 g/L, rate > 3 g/L–h, yield > 100 % of commercial target), while the only other reported microbial production of BDO from sugars is with final titer of less than 1 g/L [10].

The scientific literature provides many examples of microbes engineered to produce chemicals and fuels. With some notable exceptions (e.g., [1, 13]), these reported levels are far from the titer, productivity, and yield needed for an economically-viable process. Historically, the slow road to commercialization is done by classical mutagenesis, fermentation process optimization, and brute force screening that requires a large amount of time and resources. The fully integrated technology platform described here leverages a rational approach to strain development coupled to next-generation tools in molecular biology, omics, enzyme engineering, high-throughput screening, and precision fermentation. The importance of technology integration should not be underestimated: integration = efficiency. In addition, early guidance provided by process engineering regarding commercially viable design criteria is a vital component of any technology platform; many blind alleys and false starts can be avoided by looking at the end first. Our integrated platform provides a streamlined workflow for the design-build-test-learn cycle that encourages multidisciplinary collaboration and develops a fundamental biological understanding of the process. Emerging methods, technologies, and biological data can be rapidly assimilated and are being used to further enhance this platform, including automated cloning, lower cost DNA synthesis, high-throughput fermentation, and metagenomic data mining [18]. Technology platforms are plentiful, but only the best truly create value by expediting commercialization timelines and lowering costs. Herein we have described the main elements of our integrated technology platform and how we have applied it to drive the successful development of strains and processes for commercial scale production of BDO in fewer than 5 years. It is our hope that this review inspires others to develop technologies that continue to drive down timelines and costs, increase probabilities of success, and elevate the biotechnology industry to new heights of achievement.

References

1. Banner T, Fosmer A, Jessen HJ, Marasco E, Rush B, Veldhouse J, de Souza M (2011) Microbial bioprocesses for industrial-scale chemical production. In: Tao J, Kazlauskas R (eds) Biocatalysis for green chemistry and chemical process development. Wiley, Hoboken
2. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nat Chem Biol 5(8):593–599
3. Boyd D, Weiss DS, Chen JC, Beckwith J (2000) Towards single-copy gene expression systems making gene cloning physiologically relevant: Lambda InCh, a simple *Escherichia coli* plasmid-chromosome shuttle system. J Bacteriol 182(3):842–847
4. Carlier A, Chevrot R, Dessaux Y, Faure D (2004) The assimilation of gamma-butyrolactone in *Agrobacterium tumefaciens* C58 interferes with the accumulation of the *N*-acyl-homoserine lactone signal. Mol Plant Microbe Interact 17(9):951–957
5. Ciolino LA, Mesmer MZ, Satzger RD, Machal AC, McCauley HA, Mohrhaus AS (2001) The chemical interconversion of GHB and GBL: forensic issues and implications. J Forensic Sci 46(6):1315–1323
6. Erickson B, Nelson JE, Winters P (2012) Perspective on opportunities in industrial biotechnology in renewable chemicals. Biotechnol J7:176–185
7. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci USA 100:6940–6945
8. Goh S, Good L (2008) Plasmid selection in *Escherichia coli* using an endogenous essential gene marker. BMC Biotechnol 8:61
9. Hillson NJ, Rosengarten RD, Keasling JD (2012) j5 DNA assembly design automation software. ACS Synth Biol 1:14–21
10. Hwang HJ, Park JH, Kim JH, Kong MK, Kim JW, Park JW, Cho KM, Lee PC (2014) Engineering of a butyraldehyde dehydrogenase of *Clostridium saccharoperbutylacetonicum* to fit an engineered 1,4-butanediol pathway in *Escherichia coli*. Biotechnol Bioeng 111(7):1374–1384
11. Iuchi S, Lin EC (1988) *arcA* (*dye*), a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways. Proc Natl Acad Sci USA 85:1888–1892
12. Kim Y, Ingram LO, Shanmugam KT (2008) Dihydrolipoamide dehydrogenase mutation alters the NADH sensitivity of pyruvate dehydrogenase complex of *Escherichia coli* K-12. J Bacteriol 190:3851–3858

13. Nakamura CE, Whited GM (2003) Metabolic engineering for the microbial production of 1,3-propanediol. Curr Opin Biotechnol 14:454–459
14. Picotti P, Aebersold R (2012) Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. Nat Methods 9:555–566
15. Risso C, Van Dien SJ, Orloff A, Lovley DR (2008) Elucidation of an alternate isoleucine biosynthesis pathway in *Geobactersulfurreducens*. J Bacteriol 190(7):2266–2274
16. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27:946–950
17. Silverman PM, Rother S, Gaudin H (1991) Arc and Sfr functions of the *Escherichia coli* K-12 *arcA* gene product are genetically and physiologically separable. J Bacteriol 173(18):5648–5652
18. Van Dien S (2013) From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals. Curr Opin Biotechnol 24(6):1061–1068
19. Wahl A, El Massaoudi M, Schipper D, Wiechert W, Takors R (2004) Serial 13C-based flux analysis of an L-phenylalanine-producing *E. coli* strain using the sensor reactor. Biotechnol Prog 20(3):706–714
20. Wiese S, Reidegeld KA, Meyer HE, Warscheid B (2007) Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics 7:340–350
21. Wolfe AJ (2005) The acetate switch. Microbiol Mol Biol Rev 69(1):12–50
22. Yang TH, Heinzle E, Wittmann C (2005) Theoretical aspects of 13C metabolic flux analysis with sole quantification of carbon dioxide labeling. Comput Biol Chem 29(2):121–133
23. Yang TH, Wittmann C, Heinzle E (2006) Respirometric 13C flux analysis—Part II: in vivo flux estimation of lysine-producing *Corynebacterium glutamicum*. Metab Eng 8(5):432–446
24. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, Estadilla J, Teisan S, Schreyer HB, Andrae S, Yang TH, Lee SY, Burk MJ, Van Dien S (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat Chem Biol 7:445–452
25. Yuan Y, Yang TH, Heinzle E (2010) 13C metabolic flux analysis for larger scale cultivation using gas chromatography-combustion-isotope ratio mass spectrometry. Metab Eng 12(4):392–400