ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Correlation Analysis of Targeted Proteins and Metabolites to Assess and Engineer Microbial Isopentenol Production

Kevin W. George,[1,2] Amy Chen,[1,3] Aakriti Jain,[1,4] Tanveer S. Batth,[1,2] Edward E.K. Baidoo,[1,2] George Wang,[1,2] Paul D. Adams,[1,2,3] Christopher J. Petzold,[1,2] Jay D. Keasling,[1,2,3,4] Taek Soon Lee[1,2]

[1]Joint BioEnergy Institute, Emeryville, California 94608; telephone: +1-510-495-2469; fax: +1-510-495-2437; e-mail: tslee@lbl.gov

[2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California

[3]Department of Bioengineering, University of California, Berkeley, California

[4]Department of Chemical & Biomolecular Engineering, University of California, Berkeley, California

ABSTRACT: The ability to rapidly assess and optimize heterologous pathway function is critical for effective metabolic engineering. Here, we develop a systematic approach to pathway analysis based on correlations between targeted proteins and metabolites and apply it to the microbial production of isopentenol, a promising biofuel. Starting with a seven-gene pathway, we performed a correlation analysis to reduce pathway complexity and identified two pathway proteins as the primary determinants of efficient isopentenol production. Aided by the targeted quantification of relevant pathway intermediates, we constructed and subsequently validated a conceptual model of isopentenol pathway function. Informed by our analysis, we assembled a strain which produced isopentenol at a titer 1.5 g/L, or 46% of theoretical yield. Our engineering approach allowed us to accurately identify bottlenecks and determine appropriate pathway balance. Paired with high-throughput cloning techniques and analytics, this strategy should prove useful for the analysis and optimization of increasingly complex heterologous pathways.

Biotechnol. Bioeng. 2014;111: 1648–1658.

Author contributions: K.W.G., J.D.K., and T.S.L. designed the experiments. K.W.G., A.C., A.J., T.S.B., E.B., G.W., P.D.A., C.J.P. performed the experiments. K.W.G., J.D.K., and T.S.L. wrote the manuscript.

Competing financial interests: J.D.K. has financial interest in Amyris, LS9, and Lygos.

Correspondence to: T.S. Lee

Contract grant sponsor: US Department of Energy

Received 22 November 2013; Revision received 23 January 2014; Accepted 18 February 2014

Accepted manuscript online 25 February 2014;

Article first published online 1 May 2014 in Wiley Online Library (http://onlinelibrary.wiley.com/doi/10.1002/bit.25226/abstract).

DOI 10.1002/bit.25226

KEYWORDS: biofuel; correlation analysis; isopentenol; proteomics; metabolic engineering; isoprenoid

Introduction

Metabolic engineering has the potential to produce a large variety of chemicals—many of which are currently derived from limited resources—from simple, renewable starting materials (Keasling, 2010). In the last decade, the heterologous over-production of pharmaceuticals, commodity chemicals, alternative transportation fuels, and other natural products in microbial hosts has been convincingly demonstrated with a variety of pathways (Ajikumar et al., 2010; Atsumi et al., 2008; Dellomonaco et al., 2011; Martin et al., 2013; Peralta-Yahya et al., 2011; Ro et al., 2006; Yim et al., 2011). Unfortunately, progress towards the large-scale production of these compounds in microbial biorefineries (Kamm and Kamm, 2004) has been slow. Although advances in systems and synthetic biology have greatly expanded the tools available to metabolic engineers (Boyle and Silver, 2012; Mukhopadhyay et al., 2008), the identification of pathway bottlenecks and establishment of appropriate pathway balance remains challenging due to confounding factors such as feedback regulation, product toxicity, and strain instability. The continued development of systematic methods to assess and engineer heterologous pathways (Yadav et al., 2012) is necessary to address these challenges and improve the efficacy of metabolic engineering.

The collection of metabolomics and proteomics data provides essential insight into microbial metabolism and complex physiological behavior. On the genome scale, large sets of "omics" data have facilitated continued advances in

systems biology and related disciplines. Unfortunately, obtaining such information is generally time and resource intensive. Since heterologous pathway optimization often requires the generation of numerous pathway variants, the collection of comprehensive metabolomics and proteomics data is not always feasible. Consequently, analytical approaches that are rapid, selective, and easily accessible are particularly valuable for pathway engineering. Selected reaction monitoring (SRM) (Lange et al., 2008) for targeted proteomics is one such method that has found use in metabolic engineering efforts. Compared to traditional shotgun proteomics, this method allows for more rapid, reliable quantification of selected peptides in complex mixtures. Recently, this technique was applied to the heterologous mevalonate pathway in *Escherichia coli* to aid in the identification of pathway bottlenecks based on poor protein expression (Redding-Johanson et al., 2011). It was shown that increased expression of mevalonate kinase (MK) and phosphomevalonate kinase (PMK), two poorly-expressed enzymes, facilitated a two- to threefold increase in the production of amorphadiene, an antimalarial drug precursor. Although this approach was successful in improving titer, quantification of protein levels alone reveals little about enzymatic activity or complex behaviors such as feedback inhibition that are better elucidated by metabolite profiling.

Here, we develop a systematic approach to pathway analysis and apply it to a modified mevalonate pathway capable of producing isopentenol (3-methyl-3-butenol), a potential biofuel (Chou and Keasling, 2012). Our approach explores correlations between targeted proteomics data and diagnostic metabolites to facilitate the rapid identification of bottlenecks and potential engineering targets. Coupled with the selective quantification of pathway intermediates, this strategy is capable of providing comprehensive, quantitative insight into pathway and strain behavior.

Materials and Methods

All chemicals, solvents and media components were purchased and used without modification from Sigma-Aldrich (St. Louis, MO), Fisher Scientific (Pittsburgh, PA), or VWR (West Chester, PA) unless otherwise noted. *E. coli* strains DH10B (Invitrogen, Carlsbad, CA) and DH1 (ATCC) were used for plasmid construction and production experiments, respectively. For targeted proteomics experiments, mass spectrometric-grade trypsin was obtained from Sigma-Aldrich and prepared according to manufacturer's instructions.

Plasmid and Strain Construction

E. coli DH10B was used as the cloning host for all plasmid manipulations. Plasmids were assembled based on the BglBrick standard as described previously (Anderson et al., 2010). The plasmid pBbA5c, a member of BglBrick plasmid library (Lee et al., 2011), was used as the vector backbone for mevalonate pathway genes up to *PMK*, while pTrc99A was the backbone for *nudB* and *PMD*. *E. coli* DH1 was used as the host for production assays, while *E. coli* DH1(DE3) was the host for plasmids which required T7 polymerase. Transformations were performed using chemically competent cells. A complete list of plasmids and strains is provided in Table I.

Growth Conditions and Production of Isopentenol

Starter cultures of *E. coli* DH1 harboring production plasmids were grown overnight in LB medium containing appropriate antibiotics at 37°C and shaken at 200 rpm in rotary shakers. Chloramphenicol, ampicillin, and kanamycin were provided at final concentrations of 25 mg/L, 100 mg/L, and 25 mg/L, respectively. Production assays were performed in EZ-Rich defined medium (Teknova, Hollister, CA) containing 1% glucose. Briefly, starter cultures were used to inoculate 5 mL of production media in a culture tube to an OD_{600} of 0.1. Production cultures were grown in rotary shakers (200 rpm) at 37°C to an OD_{600} of 0.6 and induced with 500 μM isopropyl β-D-1-thiogalactopyranoside (IPTG). Strains harboring pJBEI-6835 were also induced with 20 mM arabinose. Following induction, cultures were moved to 30°C for the duration of the assay. At 24 and 48 h, samples were taken for isopentenol analysis by GC-FID as described previously (Chou and Keasling, 2012). For time course analyses, cultures were grown in 50 mL volumes in 250 mL non-baffled Erlenmeyer flasks.

Quantification of Glucose, Organic Acids, and Pathway Intermediates

Glucose and organic acids were quantified in filter-sterilized supernatant by high performance liquid chromatography (HPLC) at set time points using an Agilent 1200 Series HPLC system. Intracellular concentrations of mevalonate and IPP were measured by liquid chromatography mass spectrometry (LC-MS). Please see the Supplemental Methods Section for a detailed description of metabolite quantification.

Targeted Proteomics Analysis

At 24 h, 2 mL of production culture was collected and pelleted by centrifugation at 8,000*g* (4°C). Supernatant was decanted, and pellets were frozen in liquid nitrogen prior to storage at −80°C. Sample preparation was performed as previously described (Batth et al., 2012; Redding-Johanson et al., 2011). For complete details, please see the Supplemental Methods Section.

Correlation Analysis

A correlation analysis was performed to search for relationships between measured variables. For Spearman rank correlations (Myers and Well, 2003), values for protein area and metabolite concentrations were ranked in ascending order (i.e., the lowest value received a rank of 1). A scatter plot comparing the ranks of two variables was generated to search for monotonic relationships. Plots with raw data are provided

Table I. Description of plasmids and strains.

Plasmids	Description[a]	Reference
pJBEI-6818	pBbA5c-MevTo-MK-PMK	This study
pJBEI-6819	pBbA5c-MevTo-PMK-MK	This study
pJBEI-6820	pBbA5c-MevTo-T1002 -ptrc-MK-PMK	This study
pJBEI-6821	pBbA5c-MevTo-T1002 -ptrc-PMK-MK	This study
pJBEI-6822	pBbA5c-MevTco-MK-PMK	This study
pJBEI-6823	pBbA5c-MevTco-PMK-MK	This study
pJBEI-6824	pBbA5c-MevTco-T1002 -ptrc-MK-PMK	This study
pJBEI-6825	pBbA5c-MevTco-T1002 -ptrc-PMK-MK	This study
pJBEI-6826	pBbA5c-MevTco-T1002 -T7-MK-PMK	This study
pJBEI-6827	pBbA5c-MevTco-T1002 -T7-PMK-MK	This study
pJBEI-6828	pBbA5c-MevTco-rPMK-rMK	This study
pJBEI-6829	pBbA5c-MevTsa-MK-PMK	This study
pJBEI-6830	pBbA5c-MevTsa-PMK-MK	This study
pJBEI-6831	pBbA5c-MevTsa-T1002 -ptrc-MK-PMK	This study
pJBEI-6832	pBbA5c-MevTsa-T1002 -ptrc-PMK-MK	This study
pJBEI-4574	pTrc99A-NudB	Chou and Keasling (2012)
pJBEI-6833	pTrc99A-NudB-PMD	This study
pJBEI-6834	pTrc99A-NudB-PMD-MK	This study
pJBEI-6835	pBbB8k-NudB	This study
Strains		
1A	pJBEI-6818 + pJBEI-6833	This study
1B	pJBEI-6819 + pJBEI-6833	This study
1C	pJBEI-6820 + pJBEI-6833	This study
1D	pJBEI-6821 + pJBEI-6833	This study
2A	pJBEI-6822 + pJBEI-6833	This study
2B	pJBEI-6823 + pJBEI-6833	This study
2C	pJBEI-6824 + pJBEI-6833	This study
2D	pJBEI-6825 + pJBEI-6833	This study
3A	pJBEI-6829 + pJBEI-6833	This study
3B	pJBEI-6830 + pJBEI-6833	This study
3C	pJBEI-6831 + pJBEI-6833	This study
3D	pJBEI-6832 + pJBEI-6833	This study
2A-T7	pJBEI-6826 + pJBEI-6833	This study
2B-T7	pJBEI-6827 + pJBEI-6833	This study
2B-rv	pJBEI-6828 + pJBEI-6833	This study
1A-mk	pJBEI-6818 + pJBEI-6834	This study
2A-mk	pJBEI-6822 + pJBEI-6834	This study
3A-mk	pJBEI-6829 + pJBEI-6834	This study
1A-NudB	pJBEI-6818 + pJBEI -6833 + pJBEI-6835	This study
2A-NudB	pJBEI-6822 + pJBEI -6833 + pJBEI-6835	This study
3A-NudB	pJBEI-6829 + pJBEI -6833 + pJBEI-6835	This study
1A-mk-NudB	pJBEI-6818 + pJBEI -6834 + pJBEI-6835	This study
2A-mk-NudB	pJBEI-6822 + pJBEI -6834 + pJBEI-6835	This study
3A-mk-NudB	pJBEI-6829 + pJBEI -6834 + pJBEI-6835	This study

[a]MevTo, *E. coli atoB* coupled with *HMGS* and *HMGR* from *S. cerevisiae*. MevTco, *E. coli atoB* paired with *E. coli* codon-optimized *HMGS* and *HMGR*. MevTsa, *E. coli atoB* paired with *HMGS* and *HMGR* from *S. aureus*. rPMK, *PMK* with a variant ribosomal binding site. rMK, *MK* with a variant ribosomal binding site.

as supplemental material. Spearman correlation coefficients (ρ) were calculated between ranked variables (x_i, y_i) to determine the degree of correlation according to the equation below. Two-tailed P values are reported for significant Spearman correlations.

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

Results

Engineering Strategy and Strain Construction

The heterologous mevalonate pathway in *E. coli* consists of six reactions to convert acetyl-CoA to isopentenyl pyrophosphate (IPP), the universal precursor for isoprenoid compounds. Co-expression of NudB, an endogenous *E. coli* phosphatase, converts IPP into isopentenol. Previous work on isopentenol production utilized a three-plasmid system: pMevT and pMevB contained the top and bottom portions of the mevalonate pathway, while a third plasmid expressed NudB. In the initial demonstration of this pathway (Chou and Keasling, 2012), yields of isopentenol were low (~8% theoretical) and extensive pathway modification was not attempted. To assess and engineer isopentenol pathway function, we adopted the systematic strategy depicted in Figure 1. The first step was the generation of pathway variants that express different amounts of mevalonate pathway proteins through changes in promoter identity, operon organization, and codon-usage. For 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase (*HMGS*), and HMG-CoA reductase (*HMGR*), gene variants from *Staphylococcus aureus* were also used. Next, strains expressing these differentially expressed pathways were assayed for concentrations of pathway proteins and selected metabolites. We quantified glucose, acetate (an unwanted side product) and isopentenol (the desired product). Following quantification, a correlation analysis was performed to search for broad, monotonic relationships between protein levels and the measured metabolites to identify critical pathway components. Next, diagnostic strains were selected to clarify these suspected correlations. For these strains, glucose, acetate, isopentenol, and pathway metabolites (mevalonate and IPP) were quantified in a time-course to better understand the dynamics of pathway function. Finally, a conceptual model of pathway behavior was developed and validated with the construction of additional strains.

We utilized a two-plasmid system for our engineering efforts: plasmid 1 contained mevalonate pathway genes from thiolase (*atoB*) to *PMK* on a medium copy BglBrick vector (pBbA5c) (Lee et al., 2011), and plasmid 2 contained genes encoding the phosphatase (*nudB*) and diphosphomevalonate decarboxylase (*PMD*) on a high copy vector (pTrc99A; Amann et al., 1988; Fig. 2). We chose this arrangement for *PMD* to focus primarily on varied expression of *MK* and *PMK*, genes

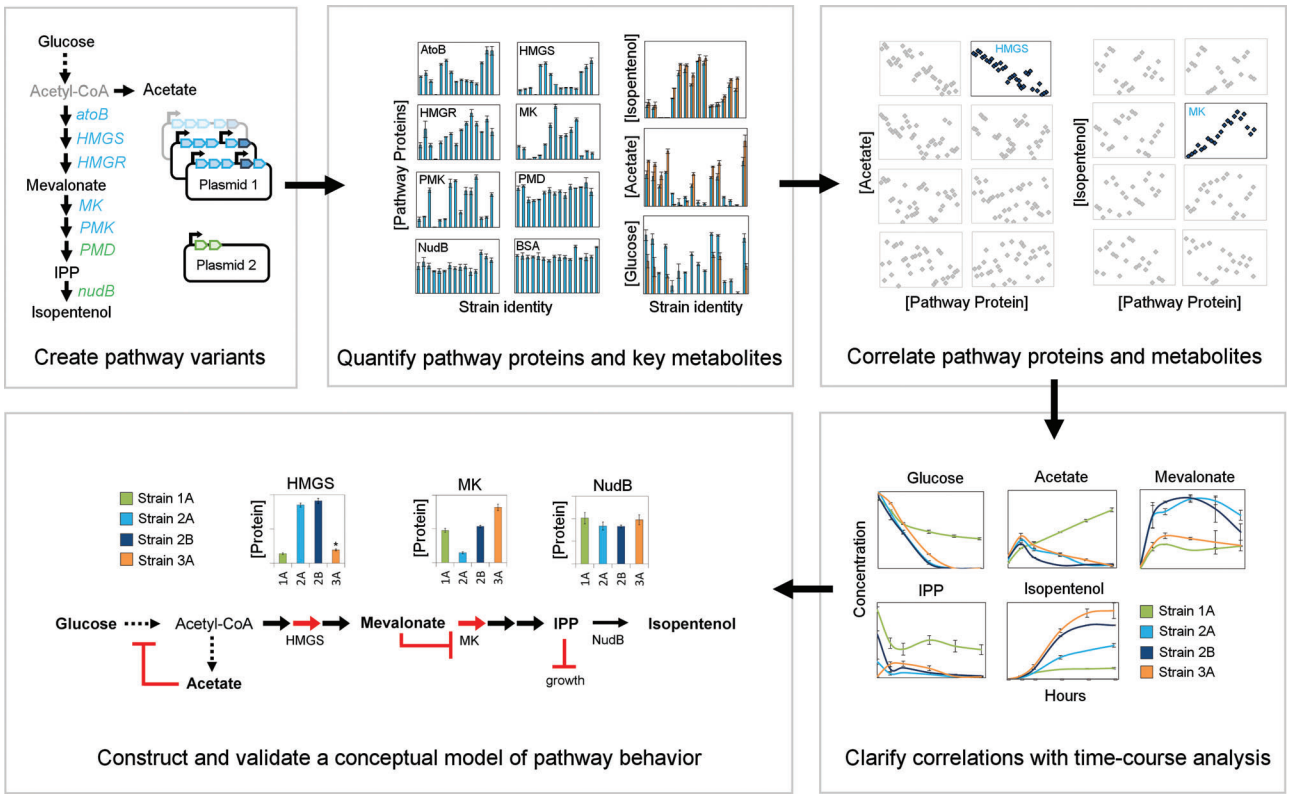

Figure 1. Engineering strategy as applied to isopentenol production in *E. coli*. Our approach centered on the use of proteomics-aided correlations to clarify pathway behavior. The first step was the creation of pathway variants that expressed different amounts of pathway proteins through standard techniques. Next, strains expressing these variants were assayed for pathway proteins along with glucose, acetate, and isopentenol. With quantification complete, a correlation analysis was performed to look for broad relationships between pathway proteins and measured metabolites. With a general understanding of key pathway correlations, strains were carefully picked to further examine pathway behavior in time-course experiments. Following the measurement of mevalonate and IPP, we developed a conceptual understanding of pathway function that was validated with the construction and analysis of additional strains.

that were previously shown to limit sesquiterpene production (Redding-Johanson et al., 2011). For the purposes of this work, we designated *atoB*, *HMGS*, and *HMGR* as the mevalonate-producing "top" of the pathway, while *MK* and *PMK* comprised the mevalonate-consuming "bottom."

Given the importance of pathway balance in metabolic engineering, part of our strategy was to vary "top" and "bottom" expression strength independently, combining a particular flux to mevalonate with variable flux to IPP. To achieve this goal, we initially created 12 plasmids that combined three separate "tops" (1–3) with a range of *atoB*, *HMGS*, and *HMGR* expression or activity with four "bottoms" (A–D) with various levels of *MK* and *PMK* expression. Additional variants constructed to probe *MK* and *nudB* expression adhered to the organization described above. A list of plasmids and strains used in this study is provided in Table I.

Quantification of Isopentenol, Acetate, Glucose, and Pathway Proteins

E. coli DH1 was transformed with each pathway variant and assayed for isopentenol production at 24 and 48 h post-induction. In addition, we quantified glucose, the primary carbon source, and acetate, a competing product derived primarily from acetyl-CoA (Fig. 2A), using a standard HPLC protocol. Importantly, we also quantified pathway proteins at 24 h. The aggregate data set is shown in Figure 3A.

The engineered strains produced a range of isopentenol titers, from ~200 to 1,200 mg/L (36% theoretical) at 48 h. Even greater variation was observed in acetate secretion, where concentrations ranged from <100 mg/L at 48 h to nearly 3 g/L. Substantial differences in glucose utilization were also apparent: although some strains completely catabolized the available glucose, others left up to 30% of the initially added glucose remaining in the culture medium at 48 h.

Pathway protein levels varied significantly between strains with the exception of NudB and PMD, which were relatively constant. Consequently, we anticipated that the wide distribution in isopentenol titer, acetate, and glucose reflected variability in the other five pathway proteins. Though protein levels matched expectations in some cases, unexpected variation in protein concentration was often apparent. For instance, levels of AtoB varied markedly despite the conserved identity of the gene, and the effects of switching *MK-PMK*

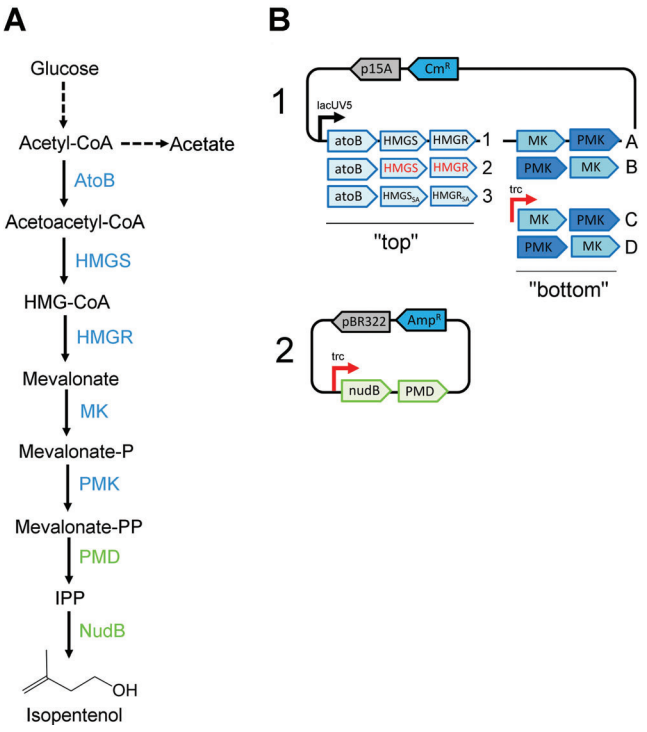

Figure 2. Pathway and strain organization. **A**: Isopentenol pathway. The heterologous mevalonate pathway in *E. coli* consists of six reactions to convert acetyl-CoA to IPP, the universal precursor for isoprenoid compounds. To produce isopentenol (3-methyl-3-butenol), an endogenous *E. coli* phosphatase (NudB) is expressed. **B**: Plasmid organization. A two-plasmid system was utilized for isopentenol production. Plasmid 1 contained mevalonate pathway genes (*atoB* through *PMK*) on a medium copy vector (pBbA5c, p15A ori) and plasmid 2 contained *nudB* and *PMD* on a high copy vector (pTrc99A, pBR33 ori). Plasmid 1 contained one of 3 top portions (1, 2, 3) coupled with 4 bottom portions (A, B, C, D). Top 1 consisted of original versions of *HMGS* and *HMGR* from *S. cervisiae*, top 2 incorporated *HMGS* and *HMGR* codon-optimized for *E. coli*, and top 3 used *HMGS* and *HMGR* from *S. aureus*. Bottom A contained *E. coli* codon-optimized *MK* and *PMK* in sequential order, B was the reverse order (*PMK-MK*), C added an upstream trc promoter with *MK-PMK*, and bottom D contained the same promoter with *PMK-MK*. Plasmid 2 remained constant.

order differed depending on the upstream context. These discrepancies highlight both the unpredictable, context-dependent nature of gene expression (Mutalik et al., 2013) and the value of proteomics as a diagnostic tool.

Correlation Analysis

We performed a simple correlation analysis to investigate potential relationships between the measured parameters. Given the presence of significant variation between variables—and the unlikelihood of perfectly linear relationships—we calculated Spearman rank correlation coefficients (ρ) to search for broad, monotonic relationships between parameters (Myers and Well, 2003). Correlations discussed in the main text are shown with raw data in . Following this initial analysis, we aimed to clarify these relationships with comprehensive metabolite data.

Without incorporating proteomics data, correlations between the diagnostic metabolites were apparent. For

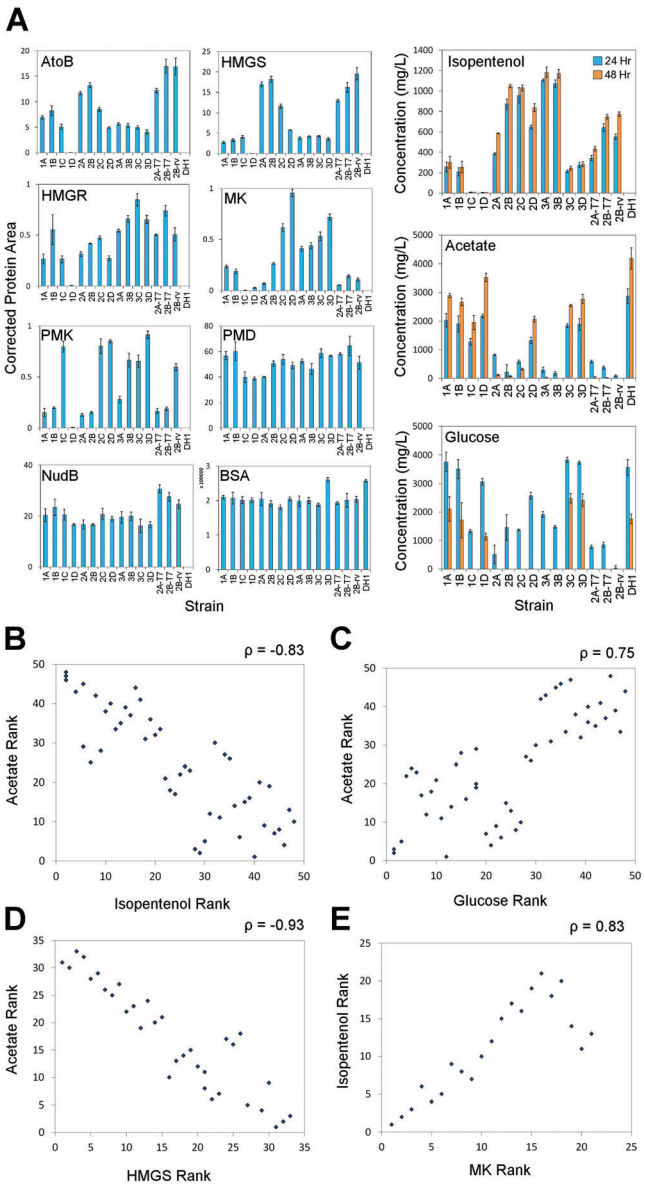

Figure 3. Correlation analysis. **A**: Aggregate data set. Pathway proteins were measured through a targeted SRM method at 24 h. Bovine serum albumin (BSA) internal standards are also shown. At 24 and 48 h, isopentenol, acetate, and glucose were quantified. Error bars represent standard deviations (n = 3). **B**: Acetate is negatively correlated with isopentenol titer. A scatter plot comparing acetate secretion and isopentenol production at 24 h by rank order resulted in a significant negative correlation ($\rho = -0.83$, $P < 0.000001$; n = 48). Individual replicates are shown for each strain. **C**: Acetate is positively correlated with residual glucose. A plot of acetate and glucose in the culture medium at 24 h yielded a positive correlation, suggesting that acetate accumulation inhibits glucose catabolism ($\rho = 0.75$, $P < 0.000001$) (n = 48). **D**: HMGS expression is negatively correlated with acetate secretion. In strains containing yeast-derived HMGR and HMGS (tops 1 and 2), increased HMGS expression was strongly correlated with decreased acetate secretion ($\rho = -0.93$, $P < 0.000001$) (n = 33). **E**: MK expression is positively correlated with isopentenol titer in strains containing top 2. In strains containing top 2 only, there was a strong positive correlation between increased MK expression and improved isopentenol titer ($\rho = 0.83$, $P = 0.000003$; n = 21).

instance, the accumulation of acetate was negatively correlated with isopentenol titer ($\rho = -0.83$, $P < 0.000001$; Fig. 3B) and positively correlated with the presence of uncatabolized glucose in the culture medium ($\rho = 0.75$, $P < 0.000001$) (Fig. 3C). High-producing strains such as 2B and 3A were characterized by low acetate accumulation and complete glucose consumption, while low-producing strains such as 1A accumulated acetate and residual glucose in the culture medium.

We next tested whether a correlation analysis could identify which proteins of the mevalonate pathway were associated with the measured metabolites. For acetate secretion, strong negative correlations with levels of AtoB, HMGS, and HMGR were observed (Fig. S2). Since direct comparisons between yeast-derived and *S. aureus*-derived HMGS and HMGR were not possible, and since substantial variation in top portion expression was not observed in strains harboring *S. aureus*-derived genes, our primary analysis made use of plasmids containing the yeast-derived proteins ("top 1" and "top 2"). Intriguingly, the strongest correlation was with HMGS ($\rho = -0.93$, $P < 0.000001$; Fig. 3D). The apparent effects of this correlation were striking: the strain with the highest HMGS protein level (19.54 relative protein area) produced ~24-fold less acetate at 24 h than the strain with the lowest detectable protein level (2.69 relative protein area).

Correlations between protein expression and isopentenol titer were more subtle and not immediately obvious from bulk correlations (Fig. S3). By grouping strains into comparable sets, however, stronger correlations were apparent. With strains containing "top 2" (codon-optimized *HMGS* and *HMGR*) a positive correlation between MK protein levels and isopentenol titer was observed ($\rho = 0.83$, $P = 0.000003$; Fig. 3E). Despite the significance of this correlation, strain 2D, which expressed the highest level of MK, produced less isopentenol than strain 2C. This may have indicated an upper limit to beneficial MK expression. However, a definitive conclusion could not be reached since levels of HMGS in strain 2D also decreased considerably. Concentrations of the other pathway proteins also varied among strains, but significant correlations with isopentenol titer were not evident (Fig. S4).

Though a positive correlation between MK expression and isopentenol titer was observed in strains containing "top 2," there was little evidence for similar correlations in strains containing "top 1" and "top 3." In strains 3A–3D, seemingly small changes in protein expression resulted in significant changes in isopentenol titer. Strains 3C and 3D, for instance, produced fivefold less isopentenol at 24 h than strains 3A and 3B with only minor changes in protein levels. Intriguingly, the strong correlation between HMGS levels and acetate observed with the yeast-derived protein was also absent: a 10-fold variation in acetate secretion occurred in strains 3A–3D with only minor changes in HMGS levels. Since MK levels gradually increased from strain 3A–3D, we suspected they might be a potential cause of these behaviors, though additional data were required to support this conjecture. Strains 1A–1D appeared to show a similar trend whereby introduction of a Trc promoter 5′ of *MK* and *PMK* (strains 1C and 1D) led to reduced isopentenol production. However, further analysis showed that strain 1C lacked MK protein and strain 1D had undetectable amounts of mevalonate pathway proteins from "plasmid 1" (Fig. 2B), suggesting unexpected plasmid instability or recombination. As with strains 3A–3D, additional data were necessary to determine the relationship between MK expression and isopentenol titer.

Metabolite Analysis and Development of a Conceptual Model

Our correlation analysis presented a simplified view of the pathway. Though seven proteins comprised the functional pathway, most behavior could be explained by variations in HMGS and MK levels. Drawing on this information, we conducted additional assays to clarify and validate these correlations. Following the collection of metabolite data, we aimed to develop a conceptual understanding of pathway function.

We first sought to clarify the role of HMGS and other "top" portion proteins in acetate secretion and glucose consumption. While the correlation analysis established links between these parameters, it was based on one or two time points. To better understand the dynamics of these behaviors, we quantified acetate, glucose, and isopentenol over a 48-h time-course. We also quantified commonly excreted organic acids such as lactate, pyruvate, and formate to further account for carbon flow through central metabolism. Since we were interested in observing the effect of each "top" in isolation, we used strains with similar "bottom" protein levels. Consequently, we chose to analyze strains 1A, 2B, and 3A, which contained unique "tops," but expressed similar levels of other pathway proteins (Fig. 3A).

Total acetate accumulation decreased with improvements in isopentenol titer, yielding clearly different product profiles at the end of the fermentation (Fig. 4). Succinate, formate, and pyruvate were detectable, but contributed little to the aggregate extracellular carbon profiles. Similar to DH1 (Fig. 4A), strain 1A—characterized by low levels of HMGS—accumulated acetate throughout the measured time-course (Fig. 4B). Conversely, acetate was rapidly assimilated in strain 2B, previously shown to express among the highest concentrations of HMGS protein (Fig. 4C). In strain 3A, acetate was assimilated at an intermediate rate compared to strain 2B (Fig. 4D).

Glucose consumption generally improved with reductions in acetate, underscoring the previously noted correlation between these parameters (Fig. 3C). Strains 2B and 3A consumed all available glucose by 36 h, while strain 1A left over 2 g/L uncatabolized. Intriguingly, strain 1A consumed even less glucose than DH1 despite accumulating 30% less acetate at 48 h, suggesting that factors other than acetate were partially responsible for this behavior.

We next quantified intracellular mevalonate and IPP, products of the top and bottom pathway, respectively. In

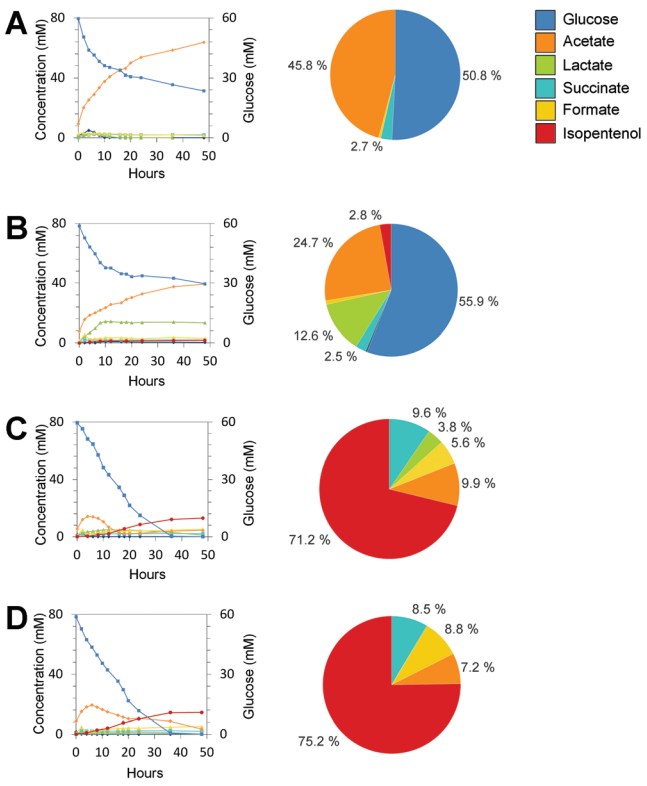

Figure 4. Organic acid analysis. Cultures of wild-type DH1 (**A**), strain 1A (**B**), strain 2B (**C**), and strain 3A (**D**) were grown in 50 mL volumes of EZ-Rich 1% glucose in 250 mL non-baffled Erlenmeyer flasks at 30°C (200 rpm). Line graphs show the results of a fermentation time-course. Extracellular concentrations of acetate, pyruvate, lactate, succinate, formate, and isopentenol were measured on the left axis (mM) while glucose concentrations are represented on the right (mM). Pie charts show measured extracellular product profiles at the end of the fermentation run in Cmol/L and thus account for differences in carbon content. Note that these profiles compare measured products only and do not constitute a total carbon inventory. Metabolites are colored according to the legend at right.

addition to determining the effect of HMGS expression on mevalonate production, we also wanted to explore the impact of MK expression on IPP concentrations. Accordingly, we added strain 2B to the analysis. Proteomics measurements showed that strains 1A and 2B differed primarily in levels of HMGS, while strains 2A and 2B differed only in MK concentrations. Thus, comparisons between these three strains could clarify the effects of HMGS and MK in isolation. We hypothesized that the prominent effect of HMGS expression on acetate secretion (Figs. 3D and 4) would be reflected in mevalonate production. Specifically, strains with low acetate accumulation and high rates of acetate assimilation would shunt more carbon into the pathway, leading to high mevalonate concentrations. Given the previously observed correlation between MK levels and isopentenol titer, we suspected that MK was the primary determinant of flux to IPP. Thus, we anticipated that improved MK expression would yield increased concentrations of IPP in our analysis. In addition to mevalonate and IPP, we measured cell growth, acetate secretion, glucose production, and isopentenol titer to construct a more complete model of isopentenol strain behavior (Fig. 5).

Quantification of mevalonate gave results consistent with our predictions. Strain 1A expressed the lowest level of HMGS and produced the least mevalonate, yielding concentrations of ~4 mM after 12 h. Strains 2A and 2B, containing ~6-fold higher levels of HMGS, yielded comparably high mevalonate concentrations of 12–15 mM. As predicted based on rates of acetate assimilation (Fig. 4D), strain 3A produced an intermediate amount of mevalonate.

IPP accumulated to high concentrations in all tested strains, exceeding levels of mevalonate in most cases (Fig. 5). As expected, increasing MK expression alone improved flux to IPP: strain 2B, containing fourfold more MK than strain 2A, accumulated ~100% more IPP and isopentenol with nearly identical levels of other pathway proteins. Concentrations of mevalonate, however, were similar between strains 2A and 2B despite the disparity in MK protein levels. Unexpectedly, strains 1A, 2A, and 2B accumulated high levels of IPP prior to induction. Since levels of IPP/DMAPP were negligible in DH1, it was unlikely that this was due to endogenous *E. coli* metabolism. Intriguingly, strain 1A produced the highest concentration of IPP despite yielding the lowest concentration of mevalonate. Since MK levels were not particularly high in strain 1A, we hypothesized that this inverse relationship between levels of mevalonate and IPP could be explained by substrate inhibition of MK (Ma et al., 2011). Given that excessive mevalonate inhibits MK, we suspected that maintenance of a lower steady state concentration may lead to increased flux to IPP as observed in strain 1A.

Strain 1A produced the least isopentenol among tested strains despite accumulating the highest concentration of the precursor IPP. This suggested that excessive IPP might be deleterious to isopentenol production or exert toxic effects. Indeed, strain 1A exhibited reduced growth and glucose consumption relative to DH1 (Fig. 5). Although these data were suggestive, further investigation was necessary to definitively implicate IPP as a cause of this phenotype.

Conceptual Model Validation

We hypothesized that high HMGS expression directed more carbon into the mevalonate pathway, preventing excess acetate accumulation and leading to high concentrations of mevalonate. In this scenario, subsequent inhibition of MK attenuated flux to IPP and limited isopentenol production. Low HMGS expression resulted in higher acetate accumulation and lower steady state concentrations of mevalonate. Counter-intuitively, this appeared to lead to increased downstream flux to IPP, which was toxic at high concentrations. Strains that balanced mevalonate production and consumption were characterized by low acetate accumulation, intermediate flux to IPP, and high isopentenol yields. This conceptual understanding of pathway function is applied to strains 1A, 2A, and 3A in Figure S5.

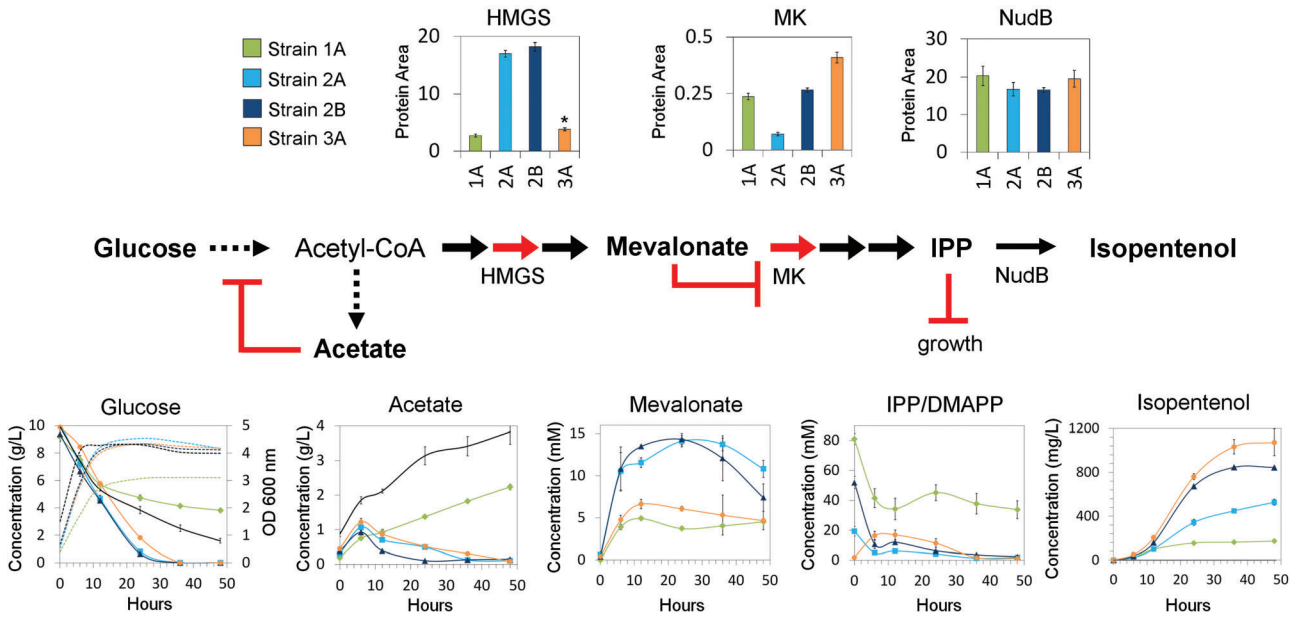

Figure 5. Conceptual model of pathway function. Strains 1A, 2A, 2B, and 3A were analyzed for glucose, growth (dotted lines in glucose panel), acetate, mevalonate, IPP, and isopentenol over 48 h. Strains are color-coded according to the legend on the left. DH1 levels of growth, glucose, and acetate production are shown as markerless black lines (growth is shown as a dotted line). DH1 levels of mevalonate, IPP, and isopentenol are not displayed as they were below detection level. Proteomics data are presented above the pathway. Error bars represent standard deviations (n = 3). Note that HMGS in strain 3A (*) is derived from *S. aureus* and thus protein levels are not directly comparable to HMGS in strains 1A, 2A, and 2B. According to our conceptual model, strain behavior is determined primarily by expression of HMGS and MK proteins (in red). Suspected inhibitory effects are shown as solid lines in red.

To evaluate our model, we compared strains 1A, 2A, and 3A with versions engineered to express substantially more MK protein (strains 1A-mk, 2A-mk, and 3A-mk; Fig. 6A). In strains 1A-mk and 3A-mk, we anticipated that MK overexpression would exacerbate the already high flux to IPP, reducing isopentenol titer due to IPP toxicity. In strain 2A-mk, we hypothesized that substrate inhibition of MK would prevent deleterious IPP accumulation, and thus MK overexpression would yield improvements in isopentenol titer. Since we expressed the additional copy of MK on plasmid 2, we anticipated that protein levels from plasmid 1 would be unaffected. An analysis of protein levels in strain 2A-mk revealed minimal changes in pathway proteins compared to strain 2A with the exception of MK, where an approximately 50-fold increase in protein expression was observed (Fig. S6). Isopentenol titer was significantly reduced in strains 1A-mk and 3A-mk relative to strains 1A and 3A, respectively (Fig. 6A). Strain 3A-mk yielded similar amounts of isopentenol as strains 3C and 3D, supporting our previous identification of increased MK expression as a primary cause for reduced titer in these strains. Matching expectations, deleterious effects on titer were absent in strain 2A-mk. Titers at 48 h reached ~700 mg/L, a 40% improvement over strain 2A, but a significant reduction compared to strain 2B.

Quantification of mevalonate and IPP yielded trends generally consistent with our conceptual model. In all strains with additional MK, a massive increase in IPP accumulation was observed (Fig. 6A). As anticipated, IPP accumulation was highest in strain 1A-mk, the strain predicted to have the largest imbalance between HMGS and MK expression, and second highest in strain 3A-mk. Although IPP still accumulated to high levels in strain 2A-mk, concentrations were threefold lower than in 1A-mk. IPP concentrations were highest at 6 h and rapidly declined in strains 1A-mk and 3A-mk. In strain 2A-mk, the decline in IPP was more gradual.

Significant growth defects were present in each strain that expressed a supplemental copy of *MK* (Fig. 6B). In strains 1A-mk and 3A-mk, reductions in growth rate were correlated with spikes in IPP. Surprisingly, growth recovered almost immediately following the disappearance of IPP. In strain 2A-mk, the initial impact of IPP accumulation on growth was less severe, but growth remained repressed longer in accordance with the more gradual decline in IPP levels in this strain. Glucose catabolism was also significantly reduced during periods of IPP accumulation (Fig. 6C). As with growth, glucose consumption rapidly resumed following a drop in IPP levels.

The toxicity of IPP at high concentrations suggested that NudB was limiting isopentenol yields in some strains. To determine if increased NudB expression could reduce IPP accumulation and improve production, we introduced a third plasmid containing *nudB* driven by an arabinose-inducible promoter into selected strains. Supplemental NudB expression relieved the growth inhibition present in strains

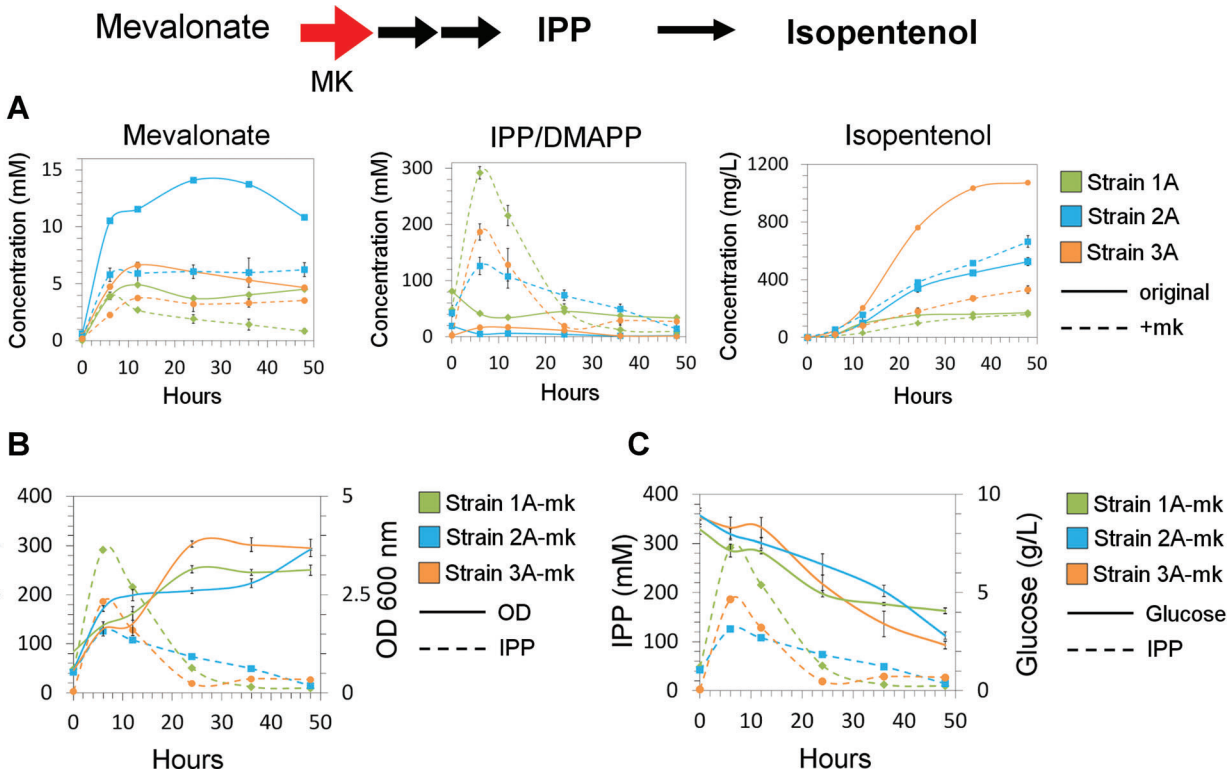

Figure 6. Model validation and identification of IPP toxicity. **A**: Impact of supplemental MK. Strains engineered to express high levels of MK (+mk, shown as dotted lines) were assayed for mevalonate, IPP, and isopentenol over a 48-h time course. Original strains are shown as solid lines for comparison. Strains are color coded according to the legend at right. Error bars represent standard deviations (n = 3). **B**: IPP inhibits cell growth. Spikes in IPP accumulation (dotted lines) aligned with reductions in cell growth measured as OD_{600nm} (solid lines). Strains are color coded according to the legend at right. Error bars represent standard deviations (n = 3). **C**: IPP inhibits glucose consumption. IPP accumulation (dotted lines) appeared inhibitory to glucose consumption (solid lines). Strains are color coded according to the legend at right. Error bars represent standard deviations (n = 3).

1A-mk, 2A-mk, and 3A-mk (Fig. 7A), providing evidence that higher NudB expression resulted in decreased IPP levels. In all tested strains, increased NudB expression resulted in higher isopentenol titer, though the degree of improvement varied substantially (Fig. 7B). In strain 3A-NudB, additional phosphatase expression yielded a 30% increase in isopentenol titer to 1.5 g/L. More significant improvements were observed in strains that accumulated high amounts of IPP: expression of additional NudB protein in strain 3A-mk-NudB yielded a sixfold increase in titer relative to strain 3A-mk. Quantification of residual glucose (Fig. 7C) and acetate (Fig. 7D) in these strains provided further support for the primary role of HMGS in the secretion of acetate and subsequent inhibition of glucose catabolism. Although NudB levels changed significantly, all strains containing "top 1"—characterized by low HMGS levels—accumulated acetate and residual glucose after 48 h. Conversely, expression of additional NudB in strains 2A-mk-NudB and 3A-mk-NudB resulted in complete glucose consumption, implying that IPP rather than acetate was chiefly responsible for inhibiting glucose catabolism in these strains. Intriguingly, strains 2A-mk and 3A-mk accumulated significantly more acetate than strains 2A and 3A, suggesting that IPP may also impact acetate secretion. Though we identified IPP as a cause of toxicity in

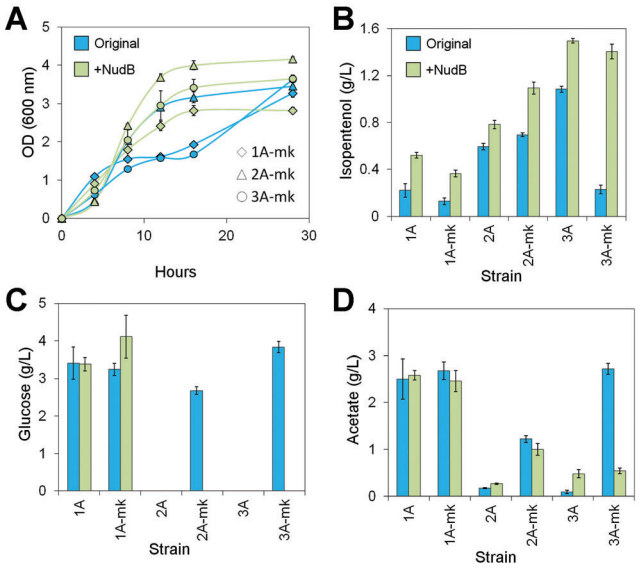

Figure 7. Impact of supplemental NudB expression. Strains 1A, 2A, 3A, 1A-mk, 2A-mk, and 3A-mk (in blue) were assayed for growth (**A**), isopentenol (**B**), glucose (**C**), and acetate (**D**) alongside strains engineered to express additional NudB protein (+NudB, in green). Error bars are standard deviations (n = 3).

our engineered strains, additional experiments are clearly required to better understand the impacts of IPP accumulation on both the mevalonate pathway and endogenous *E. coli* metabolism.

Discussion

The strongest correlation in our analysis was a negative relationship between HMGS protein levels and secreted acetate. Since the shunting of acetyl-CoA into the mevalonate pathway is initially mediated by AtoB, this dependency on the second reaction (the conversion of acetoacetyl-CoA into HMG-CoA) may reflect differences in reaction kinetics. Although the AtoB-catalyzed conversion of acetyl-CoA into acetoacetyl-CoA proceeds first, this reaction is highly reversible (Duncombe and Frerman, 1976). Consequently, it is likely that some carbon that initially enters the pathway is converted back to acetyl-CoA and thus available for conversion to acetate. In this scenario, the conversion of acetoacetyl-CoA to HMG-CoA by HMGS represents the first committed step in our synthetic pathway, "trapping" the carbon and thus preventing the formation of acetate. The accumulation of acetate has been previously identified as a major carbon sink in engineered pathways and a potential cause of acid stress, reduced growth, and impeded carbon utilization (Wolfe, 2005). Not surprisingly, acetate was correlated with reduced isopentenol production (Fig. 3B) and decreased glucose utilization (Fig. 3C) in the present study. Though acetate formation may be blocked through changes in host genetics (De Mey et al., 2007), these manipulations often have deleterious effects on cell growth or viability. In the current work, acetate accumulation was avoided through simple changes in pathway protein levels, highlighting the importance of pathway balance and the strong relationship between pathway architecture and strain behavior.

The correlation between increased MK expression and improved isopentenol yield supports a previous analysis that identified poor expression of this enzyme as a potential bottleneck in sesquiterpene production (Redding-Johanson et al., 2011). In the previous study, however, both MK and PMK were implicated on the basis of poor expression. Our correlation analysis allowed us to definitively identify MK as the primary determinant of downstream flux to IPP. More importantly, our analysis revealed that the effect of increased MK varies significantly depending on pathway context: increased MK expression was only beneficial in strains containing "top 2," where high HMGS expression resulted in greater flux to mevalonate and attenuated MK activity through substrate inhibition. Suggestively, mevalonate accumulated to ~15 mM in these strains, a concentration shown to reduce MK activity by 30–40% in in vitro studies (Ma et al., 2011). In strains containing "top 1" and "top 3," mevalonate accumulated to concentrations that resulted in significantly higher MK activity. Accordingly, increasing MK expression in these strains led to substantially reduced titer due to IPP accumulation and toxicity.

Reductions in growth and glucose consumption were evident in strains that accumulated high levels of IPP. Intriguingly, IPP levels rapidly decreased following peak accumulation, resulting in the resumption of glucose consumption and restoration of growth. It is unclear if this rapid drop in intracellular IPP concentration is due to cell lysis, conversion to longer chain terpenes, or another, uncharacterized response. Though previous studies have implicated prenyl diphosphate accumulation as a cause of reduced growth (Martin et al., 2003; Sivy et al., 2011; Withers et al., 2007), the mechanism of IPP-related toxicity is currently unknown. The strong correlation between glucose consumption and IPP suggests that glucose uptake may be inhibited, though this may be a nonspecific consequence of reduced membrane integrity or inhibited cell metabolism. Further investigation to interrogate membrane integrity is an important first step towards addressing this uncertainty. Increasing NudB expression facilitated an improvement in isopentenol titer and partially relieved IPP toxicity. Although NudB was identified as a probable bottleneck based on metabolite data, our proteomics correlation data offered an explanation for the differential effects of NudB overexpression. Given the correlation between HMGS expression and acetate, strains with low HMGS lost substantial carbon to acetate regardless of downstream flux to IPP and isopentenol. This likely caused the low isopentenol titer in strain 1A-NudB, a strain that might be expected to produce the highest titers of isopentenol based solely on pathway flux to IPP. According to our analysis, a delicate balance between pathway "pull," acetate secretion, and downstream flux to IPP is necessary to attain the highest isopentenol titers.

With a properly balanced pathway, 1.5 g/L of isopentenol was produced at 46% theoretical yield. This represents a fivefold improvement over the original pathway (Chou and Keasling, 2012) and a greater than threefold improvement over more recent work (Zheng et al., 2013) (Fig. S7). More importantly, we achieved a more complete understanding of pathway function that will inform future engineering efforts. Although our approach is similar to multivariate modular metabolic engineering (Yadav et al., 2012; MMME), there are clear advantages to our methodology. In MMME, pathways are first reorganized into "modules" to allow for the simple titration of expression levels between each synthetic operon. In the current strategy, the expression of each individual protein, rather than each "module," is used in a correlation analysis. This allows for a more extensive assessment of the pathway, revealing the impact of individual genes on overall pathway function. Though we performed the current work with a relatively modest number of strains and pathway variations, this method should be easily paired with techniques for high-throughput strain construction (Engler and Marillonnet, 2011; Engler et al., 2008; Quan and Tian, 2011). As the ability to monitor increasing numbers of proteins or metabolites also improves, future applications of this strategy should allow for the analysis and clarification of increasingly complex metabolic pathways.

We would like to thank Howard Chou for providing pTrc99A-NudB and Michelle Chang for generously providing DH1 (DE3). We also thank Aindrila Mukhopadhyay, Jorge Alonso-Gutiérrez, and Jeffrey Peacock for constructive comments on the manuscript. This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

References

Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G. 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. Science 330:70–74.

Amann E, Ochs B, Abel KJ. 1988. Tightly regulated Tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene 69:301–315.

Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA, Arkin AP, Keasling JD. 2010. BglBricks: A flexible standard for biological part assembly. J Biol Eng 4:1.

Atsumi S, Hanai T, Liao JC. 2008. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451:86–89.

Batth TS, Keasling JD, Petzold CJ. 2012. Targeted proteomics for metabolic pathway optimization. Methods Mol Biol 944:237–249.

Boyle PM, Silver PA. 2012. Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab Eng 14:223–232.

Chou HH, Keasling JD. 2012. Synthetic pathway for production of five-carbon alcohols from isopentenyl diphosphate. Appl Environ Microbiol 78:7849–7855.

De Mey M, De Maeseneire S, Soetaert W, Vandamme E. 2007. Minimizing acetate formation in *E. coli* fermentations. J Ind Microbiol Biotechnol 34:689–700.

Dellomonaco C, Clomburg JM, Miller EN, Gonzalez R. 2011. Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. Nature 476:355–359.

Duncombe GR, Frerman FE. 1976. Molecular and catalytic properties of the acetoacetyl-coenzyme A thiolase of *Escherichia coli*. Arch Biochem Biophys 176:159–170.

Engler C, Kandzia R, Marillonnet S. 2008. A one pot, one step, precision cloning method with high throughput capability. PLoS ONE 3:e3647.

Engler C, Marillonnet S. 2011. Generation of families of construct variants using golden gate shuffling. Methods Mol Biol 729:167–181.

Kamm B, Kamm M. 2004. Principles of biorefineries. Appl Microbiol Biotechnol 64:137–145.

Keasling JD. 2010. Manufacturing molecules through metabolic engineering. Science 330:1355–1358.

Lange V, Picotti P, Domon B, Aebersold R. 2008. Selected reaction monitoring for quantitative proteomics: A tutorial. Mol Syst Biol 4:222.

Lee TS, Krupa RA, Zhang F, Hajimorad M, Holtz WJ, Prasad N, Lee SK, Keasling JD. 2011. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. J Biol Eng 5:12.

Ma SM, Garcia DE, Redding-Johanson AM, Friedland GD, Chan R, Batth TS, Haliburton JR, Chivian D, Keasling JD, Petzold CJ, Lee TS, Chhabra SR. 2011. Optimization of a heterologous mevalonate pathway through the use of variant HMG-CoA reductases. Metab Eng 13:588–597.

Martin CH, Dhamankar H, Tseng HC, Sheppard MJ, Reisch CR, Prather KL. 2013. A platform pathway for production of 3-hydroxyacids provides a biosynthetic route to 3-hydroxy-gamma-butyrolactone. Nat Commun 4:1414.

Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD. 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat Biotechnol 21:796–802.

Mukhopadhyay A, Redding AM, Rutherford BJ, Keasling JD. 2008. Importance of systems biology in engineering microbes for biofuel production. Curr Opin Biotechnol 19:228–234.

Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, Mai Q-A, Tran AB, Paull M, Keasling JD, Arkin AP, Endy D. 2013. Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods 10:354–360.

Myers JL, Well AD. 2003. Research design and statistical analysis. 2nd edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD, Lee TS. 2011. Identification and microbial production of a terpene-based advanced biofuel. Nat Commun 2:483.

Quan J, Tian J. 2011. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc 6:242–251.

Redding-Johanson AM, Batth TS, Chan R, Krupa R, Szmidt HL, Adams PD, Keasling JD, Lee TS, Mukhopadhyay A, Petzold CJ. 2011. Targeted proteomics for metabolic pathway optimization: Application to terpene production. Metab Eng 13:194–203.

Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MC, Withers ST, Shiba Y, Sarpong R, Keasling JD. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440:940–943.

Sivy TL, Fall R, Rosenstiel TN. 2011. Evidence of isoprenoid precursor toxicity in *Bacillus subtilis*. Biosci Biotechnol Biochem 75:2376–2383.

Withers ST, Gottlieb SS, Lieu B, Newman JD, Keasling JD. 2007. Identification of isopentenol biosynthetic genes from *Bacillus subtilis* by a screening method based on isoprenoid precursor toxicity. Appl Environ Microbiol 73:6277–6283.

Wolfe AJ. 2005. The acetate switch. Microbiol Mol Biol Rev 69:12–50.

Yadav VG, De Mey M, Lim CG, Ajikumar PK, Stephanopoulos G. 2012. The future of metabolic engineering and synthetic biology: Towards a systematic practice. Metab Eng 14:233–241.

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.

Zheng Y, Liu Q, Li L, Qin W, Yang J, Zhang H, Jiang X, Cheng T, Liu W, Xu X, Xian M. 2013. Metabolic engineering of *Escherichia coli* for high-specificity production of isoprenol and prenol as next generation of biofuels. Biotechnol Biofuels 6:57.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.