AEM Accepts, published online ahead of print on 31 August 2012
Appl. Environ. Microbiol. doi:10.1128/AEM.01175-12

Five-Carbon Alcohols are Produced from Isopentenyl Diphosphate using a Synthetic Pathway

Howard H. Chou[1,2,3], and Jay D. Keasling[1,2,3,4,5,6,*]

[1]UCSF-UCB Joint Graduate Group in Bioengineering, University of California, Berkeley, CA 94720, USA; [2]Joint BioEnergy Institute, Emeryville, CA 94720; [3]Synthetic Biology Engineering Research Center, University of California, Berkeley, CA 94720, USA; [4]Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA 94720; [5]California Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720; [6]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

*Corresponding author. Email address: keasling@berkeley.edu

ABSTRACT

Synthetic biological pathways could enhance the development of novel processes to produce chemicals from renewable resources. Based on models that describe the evolution of metabolic pathways and enzymes in nature, we developed a framework to rationally identify enzymes able to catalyze reactions on new substrates that overcomes one of the major bottlenecks in assembling a synthetic biological pathway. We verified the framework by implementing a pathway with two novel enzymatic reactions to convert isopentenyl diphosphate into 3-methyl-3-butenol, 3-methyl-2-butenol, and 3-methyl-butanol. To overcome competition with native pathways that share the same substrate, we engineered two bifunctional enzymes that redirect metabolic flux towards the synthetic pathway. Taken together, our work demonstrates a new approach to engineering novel synthetic pathways in the cell.

INTRODUCTION

The chemical and transportation industries currently use limited nonrenewable resources to produce raw materials that could instead be synthesized from renewable resources using metabolic engineering (18). Natural biological pathways have traditionally been the source of industrially important chemicals produced using fermentation (40). Some of the regulatory mechanisms that limit production in the native hosts are circumvented by reconstructing pathways in heterologous hosts (23). To synthesize chemicals not produced by natural pathways, enzymes are over-expressed in novel combinations in order to construct synthetic biological pathways (29,50).

One of the rate-limiting steps in assembling a synthetic pathway is the identification of enzymes that can catalyze new reactions of interest. Although the reliance on known

enzymatic reactions limits the number of synthetic pathways that could be assembled, the task of engineering an enzyme to catalyze a novel reaction remains a challenge. Strategies that incorporate directed evolution or metagenomic libraries require product-specific high-throughput screens or selections in order to identify enzymes able to catalyze a desired reaction (17). Rational and *in silico* protein engineering strategies generate smaller libraries to alleviate the need for high-throughput technologies, but require *a priori* knowledge about the protein being engineered (6). Therefore, the effort to discover an enzyme that can catalyze a desired reaction is often hindered by the lack of high-throughput screens or selections, and insufficient knowledge about how to rationally engineer a protein.

We provide a new framework to identify proteins able to catalyze a desired reaction on a new substrate based on models that describe pathway and enzyme evolution. The patchwork evolution model suggests that a new biological pathway is assembled by recruiting enzymes from existing pathways (16). The ability to assemble new pathways could be facilitated by the high degree of promiscuity demonstrated by enzymes. The promiscuity could be a source of divergence during enzyme evolution, since members of the same superfamily typically share the same fold or a common catalytic strategy despite dissimilarities in their sequences, substrates, and chemical transformations (19). Combining the patchwork evolution model with hypotheses that address enzyme promiscuity, our framework constructs targeted libraries from enzyme families, which could be regarded as a collection of evolutionarily related proteins able to catalyze a specific chemistry on various substrates, to discover enzymes able to catalyze the desired reactions on new substrates. Enzyme families might already serve as libraries in nature

and provide an organism with a pool of enzymes from which to quickly evolve new reactions that might be useful, for example during adaptation to xenobiotics (36).

We reasoned enzyme families could be used in the laboratory to assemble synthetic biological pathways in order to develop new processes to convert biomass to biofuels and biomaterials. As an example, we chose to develop a metabolic pathway to produce 3-methyl-butanol, 3-methyl-3-butenol, and 3-methyl-2-butenol, which have better combustion efficiencies and more similar research octane numbers (RONs; 102, 102, and 92, respectively) to gasoline than ethanol, which is widely used as a gasoline oxygenate and substitute (48,49). 3-Methyl-2-butenol can be readily converted to citral, an intermediate in the synthesis of pharmaceuticals, vitamin A, vitamin E, some widely-used carotenoids, and certain flavors and fragrances (2). 3-Methyl-butanol has been produced in *Escherichia coli* (8) modified with the Ehrlich pathway (14) from *Saccharomyces cerevisiae*, but no pathway exists to synthesize either 3-methyl-3-butenol or 3-methyl-2-butenol. We used enzyme families as libraries to screen for novel enzymatic reactions, and explored *E. coli*'s metabolic potential for the assembly of a synthetic pathway to produce 3-methyl-butanol, 3-methyl-3-butenol, and 3-methyl-2-butenol from isopentenyl diphosphate (IPP), the central metabolite in isoprenoid biosynthesis (21).

MATERIALS AND METHODS

Oligonucleotides and DNA sequencing. All oligonucleotides were obtained from Integrated DNA Technologies with standard purification. Primer sequences mentioned here are presented in Supplementary Fig. 3. DNA sequencing to confirm cloning products was performed by Quintara Biosciences.

Strains and plasmids availability. Strains, plasmids, and plasmid sequences (in Genbank format) constructed in this study are deposited in the private instance of the JBEI registry and will be moved to the public instance (https://public-registry.jbei.org) after publication. Strains and plasmids are physically available from Addgene (http://www.addgene.org).

Strains and plasmids. All of the experiments were performed in either *E. coli* JM109 (endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB$^+$ Δ(lac-proAB) e14- [F' traD36 proAB$^+$ lacI^q lacZΔM15] hsdR17($r_K^-m_K^+$)) or DH1 (endA1 recA1 gyrA96 thi-1 glnV44 relA1 hsdR17(r_K^- m_K^+) λ^-). The mevalonate pathway was encoded by the plasmids pMevT and pMevB (23). pMevT contains the genes *atoB* (acetoacetyl-CoA thiolase), *hmgs* (HMG-CoA synthase), and *hmgr* (truncated HMG-CoA reductase). pMevB contains the genes *mk* (mevalonate kinase), *pmk* (phosphomevalonate kinase), and *pmd* (mevaloante pyrophosphate decarboxylase). All genes amplified from *E. coli* were from MG1655.

Construction of plasmids for phosphatase and reductase libraries. pHADX (where X is a number from 1 to 23) were constructed by amplifying *had*X from the *E. coli* chromosome using the primers HADX-F and HADX-R, and cloning into pTrc99A using NcoI and EcoRI. pNUDY (where Y is a letter from A to M) were constructed by amplifying *nud*Y from the *E. coli* chromosome using the primers NUDY-F and NUDY-R, and cloning into pTrc99A using NcoI and EcoRI. pNemA was constructed by amplifying *nemA* from the *E. coli* chromosome using the primers NEMA-F and NEMA-R, and cloning into pTrc99A using EcoRI and KpnI. PCR, restriction digest, and ligation reactions were performed using standard cloning protocols following the manufacturers' instructions.

Screening for enzymes able to catalyze the phosphatase reaction. pHADX (where X is a number from 1 to 23) and pNUDY (where Y is a letter from A to M) were transformed into

105 *E. coli* JM109 co-expressing pMevT and pMevB, and plated on LB agar plates with
106 ampicillin, chloramphenicol, and tetracycline. Cell cultures were grown and tested for 3-
107 methyl-3-butenol production using GC-MS as described in Growth and GC protocols for
108 production and detection of C5 Alcohols.
109 **Screening for enzymes able to catalyze the reductase reaction.** pNemA was
110 transformed into *E. coli* DH1, and plated on LB agar plates with ampicillin. Growth and
111 experimental protocols were the same as other experiments except 1 g/L of 3-methyl-2-
112 butenol was added to the medium after induction with isopropyl β-D-1-
113 thiogalactopyranoside (IPTG) (Sigma-Aldrich). Cell cultures were grown and tested for 3-
114 methyl-butanol production using GC-MS as described in Growth and GC protocols for
115 production and detection of C5 Alcohols.
116 **Protein purification of HAD-like phosphatases and Nudix hydrolases and comparison**
117 **of relative protein concentrations.** We amplified each enzyme of the HAD-like
118 phosphatases using the primers HADX-F and HADX-R (where X is a number from 1 to 23),
119 and Nudix hydrolases using the primers NUDY-F and NUDY-R (where Y is a letter from A to
120 M). The PCR products were cloned into the NcoI and BamHI sites of pPro29b (22) using,
121 which was transformed into JM109. An overnight culture was inoculated into a liter of LB
122 medium with ampicillin to an initial optical density at a wavelength of 600 nm (OD_{600}) of
123 0.05. We grew the culture at 37°C until the OD_{600} reached 0.6, induced it with 20 mM
124 propionate, and grew it overnight at 30°C. The cells were pelleted and lysed using 10X Bug
125 Buster (Novagen) diluted with binding buffer (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 0.1 mM
126 dithiothreitol) following the directions provided by the manufacturer.

The soluble and insoluble fractions were collected and separated on NuPAGE 4-12% Bis Tris Gels 1.5 mm, 15 well (Invitrogen) following the manufacturer's instructions. A Novex Sharp Protein Standard (Invitrogen) ladder was used. The proteins were transferred to PVDF membranes using the iBlot Dry Blotting Transfer System (Invitrogen). Each membrane was blocked with 1% BSA for 1 hour, washed, and incubated overnight with 3 µL S-Protein AP Conjugate (Novagen) in 12 mL of TBST (Tris-buffered saline, pH 7.5 with 0.05% Tween-20). The following day, the membranes were washed and stained with BCIP/NBT Liquid Substrate System (Sigma).

Protein purification of NudB and *in vitro* assay for phosphatase activity. We amplified NudB using the primers NUDB-F and NUDB-R3 from pNUDB, and cloned the PCR product into pPro29b using NcoI and BamHI to make pPro29b-NUDB. *E. coli* BLR(DE3) was transformed with pPro29b-NudB, and an overnight culture was inoculated into a liter of LB medium with ampicillin to an initial OD_{600} of 0.05. We grew the culture at 37°C until the OD_{600} reached 0.6, induced it with 20 mM propionate, and grew it overnight at 20°C. The cells were pelleted, suspended in binding buffer (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 0.1 mM dithiothreitol), sonicated, and centrifuged. The tagged protein was purified from the supernatant using the S•Tag Thrombin Purification Kit (Novagen) following the manufacturer's instructions. Protein concentration was determined using the Pierce BCA Protein Assay Kit (Thermo Scientific).

The standard reaction mixture for studying NudB kinetics contained the following in 100 µL: 50 mM Tris, pH 8.0, 10 mM $MgCl_2$, 1 mM DTT, 200 µM 7-methylguanosine (7-MEG) (Sigma-Aldrich), 0.5 U/mL bacterial purine nucleoside phosphorylase (PNPase) (Sigma-Aldrich), 0.5 U/mL yeast inorganic pyrophosphatase (PPase) (Sigma-Aldrich), and 0.5 µM

Phosphate Sensor (Invitrogen). Phosphate release was measured using a Spectramax M2 (Molecular Devices) exciting at 426 nm and measuring emission at 464 nm. A standard curve comparing the concentration of P_i to fluorescence was made by incubating different concentrations of P_i to the standard mixture.

To determine NudB kinetics, 15 different concentrations of isopentenyl pyrophosphate triammonium salt solution (IPP salt) (Sigma-Aldrich) ranging from 0 to 25 µM were tested. The assay was performed at 30°C with 1 µM of purified NudB, and IPP salt was added to start the reaction. The reaction was monitored in a Spectramax M2 for 3 minutes, and the change in fluorescence was used to calculate the velocity of the reaction after being normalized to the sample with 0 µM IPP salt. The change in fluorescence was converted to a change in P_i concentration using the standard curve. SigmaPlot and its Enzyme Kinetics module were used to analyze the velocity data to calculate V_{max} and K_m.

Construction of fusion proteins Idi-NudB and Idi1-NudB. The gene for Idi was amplified from the *E. coli* chromosome using the primers IDI-F2 and IDI-NUDB-SOE-R, and the gene for NudB was amplified using the primers IDI-NUDB-SOE-F and NUDB-R2. The PCR products were used as templates in a second PCR with the primers IDI-F2 and NUDB-R2 to amplify the fusion protein Idi-NudB, which was cloned into pTrc99A using EcoRI and KpnI to construct pIdi-NudB. pIdi1-NudB was constructed similarly, except Idi1 was amplified from the *S. cerevisiae* chromosome using the primers IDI1-F and IDI1-NUDB-SOE-R, NudB was amplified using the primers IDI1-NUDB-SOE-F and NUDB-R2, and Idi1-NudB was amplified using IDI1-F and NUDB-R2 in the second PCR. The 19- and 15-amino acid linkers used to construct Idi-NudB and Idi1-NudB, respectively, were based on previous work (34).

Construction of plasmids with isomerase, phosphatase, and reductase activities. *idi* was amplified from the *E. coli* chromosome using the primers IDI-F and IDI-R, and cloned using EcoRI and KpnI into pTrc99A and pNudB to construct pIdi and pNudB-s-Idi, respectively. *nemA* was amplified from pNemA using the primers NEMA-F2 and NEMA-R2, and cloned using BamHI and XbaI into pIdi-NudB to construct pIdi-NudB-s-NemA. *nemA* was amplified from pNemA using the primers NEMA-F3 and NEMA-R2, and cloned using BamHI and XbaI into pIdi1-NudB to construct pIdi1-NudB-s-NemA.

Verification of IPP conversion to 3-methyl-2-butenol and 3-methyl-butanol. pIdi, pNudB-s-Idi, pIdi-NudB, pIdi1-NudB, pIdi-NudB-s-NemA, and pIdi1-NudB-s-NemA were transformed into *E. coli* JM109 co-expressing pMevT and pMevB, and plated on LB agar plates with ampicillin, chloramphenicol, and tetracycline. 3-methyl-3-butenol, 3-methyl-2-butenol, and 3-methyl-butanol production were analyzed using GC-FID.

Growth and GC protocols for production and detection of C5 Alcohols. Three colonies were picked for each construct and grown overnight at 37°C in 5 mL of LB medium with antibiotic. The overnight cultures were inoculated into 5 mL of EZ Rich Defined Medium (Teknova) with 0.2% glucose and antibiotic to an initial OD_{600} of 0.05. Cultures were grown at 37°C for 3 hours, induced with 0.5 mM IPTG, and grown for 20 hours at 30°C. 700 µL of culture was sampled, mixed with a solution of chloroform:methanol (80:20) spiked with 50 µg/mL of butanol (Sigma-Aldrich) as an internal standard, vortexed for 15 minutes, and centrifuged for 1 minute at 13,000 × g. The chloroform layer was removed for analysis on a GC.

The GC-MS data were collected in full-scan mode (m/z 50-300) using a Tr-Wax column (0.25 mm x 30 m, 0.25 µm film thickness; Thermo Electron) on a PolarisQ GC-MS

with TriPlus autosampler (Thermo Electron). The carrier flow was 1.2 ml min^{-1}, and the inlet temperature was set to 200°C. The oven program was as follows: 40°C (1.20 min hold), 40-130°C (25°C min^{-1}), 130-220°C (35°C min^{-1}). The solvent delay was set at 3.40 min. Samples were normalized using the butanol internal standard and quantified using authentic standards. 3-Methyl-butanol standards were purchased from Sigma-Aldrich. 3-Methyl-3-butenol and 3-methyl-2-butenol were purchased from Tokyo Chemical Industry.

The GC-FID data were collected using a Tr-Wax column (0.25 mm x 30 m, 0.25 µm film thickness; Thermo Electron) on a Focus GC with TriPlus autosampler (Thermo Electron). The carrier was set at constant pressure for 300 kPa, and the inlet temperature was set to 200°C. The oven program was as follows: 40°C (1.50 min hold); 40-110°C (15°C min^{-1}). Samples were normalized using the butanol internal standard and quantified using authentic standards.

RESULTS

A synthetic pathway converts IPP into three five-carbon alcohols. IPP is naturally synthesized from either the mevalonate or deoxyxylulose 5-phosphate (DXP) pathway, and is the primary substrate of the synthetic biological pathway used to produce three C5 alcohols (Fig. 1). In the synthetic pathway, IPP is first isomerized to dimethylallyl diphosphate (DMAPP) (Reaction 1), and both IPP and DMAPP are dephosphorylated to 3-methyl-3-butenol and 3-methyl-2-butenol, respectively (Reaction 2). Finally, 3-methyl-2-butenol is reduced to 3-methyl-butanol (Reaction 3). Reaction 1 is catalyzed by an enzyme with IPP isomerase activity. No enzymes have been discovered yet to catalyze either Reaction 2 or 3, so we identified three enzyme families to screen for enzymes that might catalyze these novel reactions.

IPP phosphatase activity is discovered by screening two enzyme families. We identified the Haloacid Dehalogenase (HAD) (4) and Nudix (25) superfamilies as likely to contain enzymes that can catalyze the dephosphorylation of IPP and DMAPP from the table of mechanistically diverse superfamilies provided in Glasner *et al.* (11). The prevalence of both superfamilies in all three domains of life (bacteria, archaea, and eukaryotes) suggests that their members are used extensively throughout evolution to catalyze mechanistically diverse reactions on various substrates, and both superfamilies of enzymes have the catalytic residues required to recognize and dephosphorylate the phosphate moieties of IPP and DMAPP. We limited our search to the family of 23 HAD-like phosphatases in *E. coli* (20) within the HAD superfamily as the first library, and the family of 13 Nudix hydrolases in *E. coli* (24) within the Nudix superfamily as the second library.

We overexpressed each enzyme from the two libraries and screened for the ability to catalyze the conversion of IPP to 3-methyl-3-butenol in *E. coli* that expressed the mevalonate pathway genes. Gas chromatography (GC) was used to detect 3-methyl-3-butenol production. A control that expressed the mevalonate pathway genes in the absence of any phosphatases produced a detectable quantity of 3-methyl-3-butenol. The background production indicates that the catalytic activity to convert IPP to 3-methyl-3-butenol exists in *E. coli*. Out of the thirty-six enzymes screened, overexpression of two HAD-like phosphatases (HAD4 and HAD10) and five Nudix hydrolases (NudB, NudF, NudI, NudJ, NudM) led to a more than two-fold increase in production of 3-methyl-3-butenol compared to the control (Fig. 2). NudF, whose isozyme from *Bacillus subtilis* was identified to have prenyl alcohol biosynthesis activity (46), demonstrated an almost three-fold improvement in production over the control. The overexpression of NudB led to the

241 greatest conversion of IPP to 3-methyl-3-butenol *in vivo* at five-fold more C5 alcohol than
242 the control.

243 The production of 3-methyl-3-butenol did not correlate well with the concentration
244 of the phosphatase produced by overexpression. To determine the phosphatase
245 concentration produced by the cell, each of the HAD-like phosphatases and Nudix
246 hydrolases screened was tagged with a N-terminal S•Tag (Fig. S2). Even though
247 overexpression of NudB led to the highest 3-methyl-3-butenol production, its protein
248 concentration was lower than many of Nudix hydrolases tested. Similarly, the protein
249 concentration of HAD4 and HAD10 was not greater than any other HADs even though these
250 two HAD-like phosphatases produced more 3-methyl-3-butenol than the other members of
251 the family.

252 **IPP phosphatase activity is verified *in vitro*.** We developed an assay to detect the
253 phosphatase reaction with IPP as a substrate to verify NudB's ability to catalyze Reaction 2
254 *in vitro*. The assay measures pyrophosphate release kinetics in real-time using a coupled
255 enzyme assay that produces a fluorescent output (13). We purified NudB using a N-
256 terminal S•Tag. We determined the apparent V_{max} of NudB to be 0.63 nmole/min/mg and
257 its apparent K_m to be 4.7 μM for IPP (Fig. S1). NudB's apparent K_m is comparable to that of
258 *E. coli* IPP isomerase Idi, but its apparent V_{max} is 1000-fold less than that of Idi (12). The
259 poor kinetics of NudB for IPP exemplifies the natural evolution of an enzyme, which
260 generally begins with a protein having low affinity for a new substrate or low rates of
261 catalytic action on the bound substrate (32). However, we recognize that the N-terminal
262 S•Tag might also affect the activity and apparent kinetic values measured. Overall, the *in*

263 *vitro* data support the *in vivo* data that NudB binds IPP and catalyzes the phosphatase
264 reaction to convert it to 3-methyl-3-butenol.
265 **3-methyl-2-butenol reductase activity is discovered from one enzyme family.** The
266 reduction of xenobiotics with carbonyl groups commonly uses NADPH as a co-factor (30).
267 We identified a family of NADPH dehydrogenases that we hypothesized could catalyze the
268 reduction of 3-methyl-2-butenol to 3-methyl-butanol from the list of enzyme superfamilies
269 presented by Todd *et al.* (41). These dehydrogenases are part of the FMN-oxidoreductase
270 superfamily called Old Yellow Enzyme (OYE) (44). The OYE family is represented in both
271 prokaryotes and eukaryotes, and several members are associated with xenobiotic
272 metabolism (44,45), suggesting that natural systems exploit the propensity of these
273 enzymes to catalyze reduction reactions on novel substrates. We limited our search to the
274 OYE family in *E. coli* that can reduce α/β-unsaturated carbonyl functionalities.
275 NemA (28), the only documented member of the OYE from *E. coli* that fit our criteria,
276 was overexpressed in the presence of 1 g/L of 3-methyl-2-butenol in the medium, and
277 catalysis of Reaction 3 was confirmed by measuring 3-methyl-butanol production using GC.
278 The overexpression of NemA led to 3-methyl-butanol production, and the conversion of 3-
279 methyl-2-butenol to 3-methyl-butanol was 40% after 24 hours. The control that did not
280 overexpress any enzyme did not produce any 3-methyl-butanol. Even though 3-methyl-2-
281 butenol lacks the α/β-unsaturated carbonyl functional group present in most of the
282 substrates of enzymes in the OYE family, overexpression of NemA still demonstrated
283 increased 3-methyl-butanol production from 3-methyl-2-butenol.
284 **Idi-NudB fusion overcomes competition with native pathways for IPP and DMAPP.**
285 We overexpressed *E. coli* IPP isomerase (Idi) to catalyze Reaction 1 and NudB to catalyze

286 Reaction 2 separately in *E. coli* that also overexpressed the mevalonate pathway in order to
287 synthesize 3-methyl-2-butenol from IPP. However, co-expression of Idi and NudB
288 decreased 3-methyl-3-butenol production by five fold compared to overexpression of NudB
289 by-itself, and no 3-methyl-2-butenol production was observed (Figure 3A). No 3-methyl-3-
290 butenol or 3-methyl-2-butenol production was observed when Idi was overexpressed by-
291 itself.

292 We hypothesized that a novel protein to restore 3-methyl-3-butenol production and
293 synthesize 3-methy-2-butenol from IPP could be engineered by fusing Idi and NudB
294 together joined by a peptide linker to construct a single bifunctional polypeptide (Idi-
295 NudB). Idi-NudB consists of two functional domains that each catalyzes a different reaction
296 (isomerization and dephosphorylation) and is part of a consecutive reaction (conversion of
297 IPP to DMAPP and DMAPP to 3-methyl-2-butenol). This fusion protein mimics natural
298 bifunctional enzymes, such as *Arabidopsis* lysine-ketoglutarate reductase/saccharopine
299 dehydrogenase (51), human 5-amino-4-imidazolecarboxamide ribonucleotide
300 transformylase/inosine 5'-monophosphate cyclohydrolase (3), *Abies grandis* abietadiene
301 synthase (31), and lysine-ketoglutarate reductase/saccharopine dehydrogenase from
302 developing soybean seeds (27). We joined Idi and NudB with a peptide linker, because
303 linkers can control favorable and unfavorable interactions between adjacent protein
304 domains (47).

305 The overexpression of the Idi-NudB enzyme fusion increased C5 alcohol production
306 by more than two-fold compared to the overexpression of Idi and NudB separately, and led
307 to 3-methyl-2-butenol production (Fig 3B). We also fused the *S. cerevisiae* IPP isomerase
308 Idi1 (37) to NudB (Idi1-NudB). The overexpression of Idi1-NudB increased 3-methyl-3-

butenol and 3-methyl-2-butenol production by an additional 40% compared to Idi-NudB. The production of 3-methyl-2-butenol from IPP by the fusion protein indicates that the catalytic domains of Idi and NudB are functional, the single polypeptides are bifunctional, and Idi-NudB and Idi1-NudB can compete with the native enzymes in *E. coli* for IPP and DMAPP.

Three five-carbon alcohols are produced from IPP. We assembled the entire synthetic pathway by cloning either i*di-nudB* or i*di1-nudB* 5' of *nemA* to construct the plasmid pIdi-NudB-s-NemA or pIdi1-NudB-s-NemA, respectively. We observed production of 3-methyl-3-butenol, 3-methyl-2-butenol, and 3-methyl-butanol from *E. coli* that overexpressed the genes of the mevalonate pathway and harbored one of these two plasmids. The total amount of C5 alcohols produced remained the same compared to production in the absence of NemA, and the amount of 3-methyl-3-butenol remained unchanged. The decrease in 3-methyl-2-butenol production observed in the presence of NemA is accounted for by the increase in 3-methyl-butanol production. The production of three 5C alcohols from IPP verifies our synthetic pathway and framework of using enzyme families to identify enzymes able to catalyze new reactions.

DISCUSSION

One major challenge in the construction of synthetic biological pathways is the identification of enzymes able to catalyze each new desired reaction in the pathway. The patchwork evolution model (16) provides a foundation for strategies that screen libraries of natural enzymes to identify those that can catalyze a novel reaction of interest. However, these strategies typically involve large library sizes and high-throughput screens or selections able to recognize the desired product. For example, a metagenomic library of

almost 10^6 variants constructed from environmental DNA was screened to identify enzymes with 4-hydroxybutyrate dehydrogenase activity, whose identification was made possible by the use of growth on 4-hydroxybutyrate as a selection (15). Unfortunately, most industrially important compounds are not readily amenable to high-throughput screens and selections, which makes large library sizes unwieldy. Alternative screening methods, such as phage display (25,37) and mRNA display (34,37), can screen 10^9-10^{12} variants and detect protein-protein, protein-peptide, and protein-DNA interactions. However, these techniques are hampered by practical limitations, such as whether the proteins are folded in a functional form either *in vitro* or when displayed on the phage, and have not been used to engineer enzymes to catalyze new chemical reactions involving small molecules.

In this work, we decreased the size of the library that needs to be screened based on our understanding of how enzymes naturally evolve new activities, and the body of work that categorizes enzymes into families based on their mechanistic activities (7). We combined the patchwork evolution model with hypotheses that explain the role of enzyme promiscuity during protein evolution and formulated a framework that uses protein families as small, targeted libraries. We screened thirty-six enzymes from the HAD-like phosphatase and Nudix hydrolase families and identified seven able to dephosphorylate IPP and DMAPP. Overexpression of NudB led to the highest 3-methyl-3-butenol production even though its *in vivo* protein concentration was lower than most of the other enzymes screened, which suggests that the increased alcohol production is due to its superior kinetic properties. Furthermore, we identified a member of the Old Yellow Enzyme family able to reduce 3-methyl-2-butenol.

355 Even though the overexpression of an enzyme with the desired activity on the new
356 substrate can lead to a sufficient increase in reaction rate to produce the desired product,
357 overexpression alone is insufficient when engineering a new combination of enzymes to
358 work in concert and produce the desired final product. Overexpression is insufficient,
359 because multiple native pathways often compete for a common intermediate, which is a
360 recurring problem in metabolic engineering (41). For example, the overexpression of Idi
361 and NudB did not lead to any significant increase in 3-methyl-2-butenol production. One
362 explanation for this observation is that when Idi is overexpressed, the low kinetic efficiency
363 of NudB for catalyzing the novel phosphatase reaction compared to the enzymes in the
364 native pathways that consume IPP and DMAPP prevents any significant amount of 3-
365 methyl-2-butenol from accumulating. We reasoned that the IPP and DMAPP produced by
366 the heterologous mevalonate pathway in the presence of Idi were consumed by native
367 enzymes, since *in vitro* analysis of NudB indicated it is relatively slow at catalyzing the new
368 phosphatase reaction. For example, *E. coli* farnesyl pyrophosphate (FPP) synthase, IspA
369 (9), uses IPP and DMAPP to synthesize FPP, which is used in quinone and cell wall
370 biosynthesis. Therefore, we engineered Idi-NudB and Idi1-NudB, which successfully
371 competed with native enzymes in *E. coli* for IPP and DMAPP, and led to 3-methyl-2-butenol
372 production and the production of 3-methyl-butanol from IPP when they were co-expressed
373 with NemA.

374 The decrease in C5 alcohol production only occurred when Idi was overexpressed,
375 because DMAPP is required by the native pathways in *E. coli* to synthesize isoprenoids.
376 Therefore, the IPP produced by the mevalonate pathway could have been channeled away
377 from C5 alcohol production and towards native isoprenoid production when Idi and NudB

were overexpressed. Even though overexpression of the fusion proteins Idi-NudB and Idi1-NudB recovered some of the C5 alcohol production, total production was still lower compared to that from the overexpression of only NudB. The difference in production might be due to some of the DMAPP produced by the Idi and Idi1 domains of the fusion enzymes being used to synthesize native isoprenoids before it could be converted to 3-methyl-2-butenol by the NudB domain. Since both Idi and NudB are *E. coli* proteins, it is unlikely that their solubilities would change significantly by fusing them together. Idi1 is from *S. cerevisiae* and is unlikely to be more soluble than Idi in *E. coli*. Therefore, the increase in C5 alcohol production by Idi1-NudB compared to Idi-NudB might be explained by the fact that even though Idi and Idi1 have similar catalytic efficiencies, *E. coli* Idi has a lower V_{max} and lower K_m compared to yeast Idi1 (12). Overexpression of Idi1-NudB led to higher production of 3-methyl-2-butenol than Idi-NudB, suggesting that under conditions where abundant quantities of IPP is available from the overexpression of the mevalonate pathway a higher V_{max} can lead to more DMAPP produced, which compensates for the catalytic inefficiency of NudB for converting DMAPP to 3-methyl-2-butenol.

The approach of fusing metabolic enzymes together to control metabolic flux in a new pathway is similar in principle to the use of protein scaffolds to control flux (9), and complements existing metabolic engineering strategies based on controlling gene expression using gene knockouts (4), controlling promoter strength and plasmid copy number (1), and changing RBS or promoter strengths (37). The synthetic pathway we constructed provides a new process for producing 3-methyl-3-butenol, 3-methyl-2-butenol, and 3-methyl-butanol from IPP. The maximum theoretical production of C5 alcohols from glucose using the mevalonate pathway is 0.33 g/g. Overexpression of NudB achieved 55

401 mg/L of 3-methyl-3-butenol from 2 g/L of glucose, which is 8.3% of the theoretical
402 maximum, and the overexpression of Idi1-NudB and NemA achieved 4.8% of the
403 theoretical maximum. These titers do not account for the volatility of C5 alcohols and their
404 loss to the headspace during production, similar to the production of other IPP-derived
405 compounds such as isoprene and amorphadiene (23). Additionally, the yield from this new
406 pathway could be improved by building on previous engineering efforts that increased
407 isoprenoid yields (33).

408 The shift from nonrenewable to renewable resources for the synthesis of chemical
409 building blocks used by various industries will require the invention of new processes to
410 produce molecules that are functionally identical to those synthesized from nonrenewable
411 resources. The field of synthetic chemistry facilitated the development of processes to
412 produce novel compounds from basic laboratory materials. Similarly, metabolic
413 engineering has the potential to produce novel compounds from renewable resources by
414 engineering cells. To achieve that potential, new enzymatic reactions are necessary for
415 assembling synthetic pathways, so that engineers are not limited to producing compounds
416 from natural pathways nor reliant on known enzymatic reactions. The framework we
417 describe here for discovering enzymes able to catalyze new reactions not discovered in
418 nature yet could be used to assemble a large number of synthetic biological pathways for
419 producing chemicals with industrial applications.

420 ACKNOWLEDGEMENTS

421 This work was supported by the Joint BioEnergy Institute, contract DE-AC02-05CH11231
422 between Lawrence Berkeley National Laboratory and the US Department of Energy, and by

the Synthetic Biology Engineering Research Center (SynBERC) through National Science Foundation Grant BES-0439124.

REFERENCES

1. 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.
2. Babler JH. August 2001. Methods for Conversion of Isoprene to Prenyl Alcohol and Related Compounds. U.S. patent 6,278,016.
3. Bulock KG, Beardsley GP, Anderson KS. 2002. The kinetic mechanism of the human bifunctional enzyme ATIC (5-amino-4-imidazolecarboxamide ribonucleotide transformylase/inosine 5'-monophosphate cyclohydrolase). A surprising lack of substrate channeling. J. Biol. Chem. **277**:22168-22174.
4. Burgard AP, Pharkya P, Maranas CD. 2003. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. **84**:647-657.
5. Burroughs AM, Allen KN, Dunaway-Mariano D, Aravind L. 2006. Evolutionary Genomics of the HAD Superfamily: Understanding the Structural Adaptations and Catalytic Diversity in a Superfamily of Phosphoesterases and Allied Enzymes. J. Mol. Biol. **361:**1003-1034.
6. Chen, R. 2001. Enzyme engineering: rational redesign versus directed evolution. Trends Biotechnol. **19**:13-14.

444 7. Chiang RA, Sali A, Babbitt PC. 2008. Evolutionarily Conserved Substrate Substructures
445 for Automated Annotation of Enzyme Superfamilies. PLoS Computational Biology
446 **4:**e1000142.
447 8. Connor MR, Liao JC. 2008. Engineering of an *Escherichia coli* Strain for the Production of
448 3-methyl-1-Butanol. Applied and Environmental Microbiology **74:** 5769-5775.
449 9. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KLJ,
450 Keasling JD. 2009. Synthetic protein scaffolds provide modular control over metabolic
451 flux. Nature Biotechnol. **27**:753-759.
452 10. Fujisaki S, Hara H, Nishimura Y, Horiuchi K, Nishino T. 1990. Cloning and nucleotide
453 sequence of the ispA gene responsible for farnesyl diphosphate synthase activity in
454 *Escherichia coli*. J. Biochem **108:**995-1000.
455 11. Glasner ME, Gerlt JA, Babbitt PC. 2006. Evolution of enzyme superfamilies. Current
456 Opinion in Chemical Biology **10:**492-497.
457 12. Hahn FM, Hurlburt AP, Poulter CD. 1999. *Escherichia coli* Open Reading Frame 696 Is
458 *idi*, a Nonessential Gene Encoding Isopentenyl Diphosphate Isomerase. Journal of
459 Bacteriology **181:**4499-4504.
460 13. Hanes JW, Johnson KA. 2008. Real-time measurement of pyrophosphate release
461 kinetics. Analytical Biochemistry **372:**125-127.
462 14. Hazelwood LA, Daran J-M, van Maris AJA, Pronk JT, Dickinson JR. 2008. The Ehrlich
463 Pathway of Fusel Alcohol Production: a Century of Research on *Saccharomyces*
464 *cerevisiae* Metabolism. Applied and Environmental Microbiology **74:**2259-2266.
465 15. Henne A, Daniel R, Schmitz RA, Gottschalk G. 1999. Construction of Environmental DNA
466 Libraries in *Escherichia coli* and Screening for the Presence of Genes Conferring

467 Utilization of 4-Hydroxybutyrate. Applied and Environmental Microbiology **65**:3901-
468 3907.
469 16. Jensen RA. 1976. Enzyme Recruitment in Evolution of New Function. Annu. Rev.
470 Microbio. **30:**409-425.
471 17. Kazlauskas RJ, Bornscheuer UT. 2009. Finding better protein engineering strategies.
472 Nature Chemical Biology **5:**526-529.
473 18. Keasling JD. 2010. Manufacturing Molecules Through Metabolic Engineering. Science
474 **330:**1355-1358.
475 19. Khersonsky O, Tawfik DS. 2010. Enzyme Promiscuity: A Mechanistic and Evolutionary
476 Perspective. Annu. Rev. Biochem. **79:**471-505.
477 20. Kuznetsova E, Proudfoot M, Gonzalez CF, Brown G, Omelchenko MV, Borozan I, Carmel
478 L, Wolf YI, Mori H, Savchenko AV, Arrowsmith CH, Koonin EV, Edwards AM, Yakunin AF.
479 2006. Genome-wide Analysis of Substrate Specificities of the *Escherichia coli* Haloacid
480 Dehalogenase-like Phosphatase Family. Journal of Biological Chemistry **281:**36149-
481 36161.
482 21. Lange BM, Rujan T, Martin W, Croteau R. 2000. Isoprenoid biosynthesis: The evolution
483 of two ancient and distinct pathways across genomes. Proc. Natl. Acad. Sci. USA
484 **97:**13172-13177.
485 22. Lee SK, Keasling JD. 2008. Heterologous protein production in *Escherichia coli* using the
486 propionate-inducible pPro system by conventional and auto-induction methods.
487 Protein Expression and Purification **61:**197-203.

23. Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD. 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nature Biotechnology **7:**796-802.

24. McLennan AG. 2006. The Nudix hydrolase superfamily. Cellular and Molecular Life Sciences **63:**123-143.

25. Meyer SC, Shomin CD, Gaj T, Ghosh I. 2007. Tethering Small Molecules to a Phage Display Library: Discovery of a Selective Bivalent Inhibitor or Protein Kinase A. J. Am. Chem. Soc. **129**:13812-13813.

26. Mildvan AS, Xia Z, Azurmendi HF, Saraswat V, Legler PM, Massiah MA, Gabelli SB, Bianchet MA, Kang L-W, Amzel LM. 2005. Structures and mechanisms of Nudix hydrolases. Archives of Biochemistry and Biophysics **433:**129-143.

27. Miron D, Ben-Yaacov S, Reches D, Schupper A, Galili G. 2000. Purification and characterization of bifunctional lyzine- ketoglutarate reductase/saccharopine dehydrogenase from developing soybean seeds. Plant Physiol. **123:**655 – 663.

28. Miura K, Tomioka Y, Suzuki H, Yonezawa M, Hishinuma T, Mizugaki M. Molecular cloning of the *nemA* gene encoding N-ethylmaleimide reductase from *Escherichia coli.* Biol. Pharm. Bull. **20:**110-112.

29. Moon, TS, Yoon S-H, Lanza AM, Roy-Mayhew JD, Prather, KLJ. 2009. Production of Glucaric Acid from a Synthetic Pathway in Recombinant *Escherichia coli*. Applied and Environmental Microbiology **75:**589-595.

30. Oppermann U. 2007. Carbonyl Reductases: The Complex Relationships of Mammalian Carbonyl- and Quinone-Reducing Enzymes and Their Role in Physiology. Annual Review of Pharmacology and Toxicology **47**:293-322.

511 31. Peters RJ, Ravn MM, Coates RM, Croteau RB. 2001. Bifunctional abietadiene synthase:
512 free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct
513 active sites. J. Am. Chem. Soc. **123:**8974 – 8978.
514 32. Pettersson G. 1989. Effect of evolution on the kinetic properties of enzymes. Eur. J.
515 Biochem. **184:**561-566.
516 33. Pitera DJ, Paddon CJ, Newman JD, Keasling JD. 2007. Balancing a heterologous
517 mevalonate pathway for improved isoprenoid production in *Escherichia coli*. Metabolic
518 Engineering **9:**193-207.
519 34. Roberts RW, Szostak JW. 1997. RNA-peptide fusions for the *in vitro* selection of peptides
520 and proteins. Proc. Natl. Acad. Sci. USA **94**:12297-12302.
521 35. Robinson CR, Sauer RT. 1998. Optimizing the stability of single-chain proteins by linker
522 length and composition mutagenesis. Proc. Natl. Acad. Sci. USA **95:**5929-5934.
523 36. Russell RJ, Scott C, Jackson CJ, Pandey R, Pandey G, Taylor MC, Coppin CW, Liu J-W,
524 Oakeshott JG. 2011. The evolution of new enzyme function: lessons from xenobiotic
525 metabolizing bacteria versus insecticide-resistant insects. Evolutionary Applications
526 **4:**225-248.
527 37. Salis HM, Mirsky EA, Voigt CA. 2009. Automated design of synthetic ribosome binding
528 sites to control protein expression. Nat. Biotechnol. **27**:946-950.
529 38. Seelig B, Szostak JW. 2007. Selection and evolution of enzymes from a partially
530 randomized non-catalytic scaffold. Nature **448**:828-831.
531 39. Smith GP. 1985. Filamentous fusion phage: novel expression vectors that display cloned
532 antigens on the virion surface. Science **228**:1315-1317.

533 40. Street IP, Poulter CD. 1990. Isopentenyl diphosphate:dimethylallyl diphosphate
534 isomerase: construction of a high-level heterologous expression system for the gene
535 from *Saccharomyces cerevisiae* and identification of an active-site nucleophile.
536 Biochemistry **29:**7531-7538.
537 41. Tepper N, Shlomi T. 2010. Predicting metabolic engineering knockout strategies for
538 chemical production: accounting for competing pathways. Bioinformatics **26**:536-543.
539 42. Todd AE, Orengo CA, Thorton JM. 2001. Evolution of Funtion in Protein Superfamilies,
540 from a Structural Perspective. Journal of Molecular Biology **307**:1113-1143.
541 43. Villadsen J, Nielsen J, Liden G. 2011. Chemicals from Metabolic Pathways, p. 7-62. *In*
542 Bioreaction and Engineering Principles, 3rd ed. Springer-Verlag, New York, NY.
543 44. Williams RE, Bruce NC. 2002. 'New uses for an Old Enzyme' – the Old Yellow Enzyme
544 family of flavoenzymes. Microbiology **148:**1607-1614.
545 45. Williams RE, Rathbone DA, Scrutton NS, Bruce NC. Biotransformation of Explosives by
546 the Old Yellow Enzyme Family of Flavoproteins. Applied and Environmental
547 Microbiology **70:**3566-3574.
548 46. Withers ST, Gottlieb SS, Lieu B, Newman JD, Keasling JD. 2007. Identification of
549 Isopentenol Biosynthesis Genes from *Bacillus subtilis* by a Screening Method Based on
550 Isoprenoid Precursor Toxicity. Applied and Environmental Microbiology **73:**6277-6283.
551 47. Wriggers W, Chakravarty S, Jennings PA. 2005. Control of Protein Functional Dynamics
552 by Peptide Linkers. Biopolymers **80:**736-746.
553 48. Prof. Robert Dibble, Department of Mechanical Engineering, University of California,
554 Berkeley, personal communication.

555 49. Yang Y, Dec J, Dronniou N, Simmons B. 2010. Characteristics of Isopentanol as a Fuel for
556 HCCI Engines. SAE Int. J. Fuels. Lubr. **3:**725-741.
557 50. Zhang K, Sawaya MR, Eisenberg DS, Liao JC. 2008. Expanding metabolism for
558 biosynthesis of nonnatural alcohols. Proc. Natl. Acad. Sci. USA **105:**20653-20658.
559 51. Zhu X, Tang G, Galili G. 2002. The activity of the *Arabidopsis* bifunctional lysine-
560 ketoglutarate reductase/saccaropine dehydorgenase enzyme of lysine catabolism is
561 regulated by functional interaction between its two enzyme domains. J. Biol. Chem.
562 **277:**49655-49661.
563

564 **FIGURE CAPTIONS**

565 **Fig. 1.** Synthetic pathway converts isopentenyl diphosphate (IPP) to 3-methyl-3-butenol, 3-
566 methyl-2-butenol, and 3-methyl-butanol. IPP is produced by co-expression of the
567 mevalonate pathway genes (*atoB*, *hmgS*, *hmgR*, *mk*, *pmk*, *pmd*) in *E. coli*. Three different
568 reactions convert IPP into three different 5C alcohols. Reaction 1 catalyzes the
569 isomerization of IPP to dimethylallyl diphosphate (DMAPP). Reaction 2 catalyzes the
570 dephosphorylation of IPP into 3-methyl-3-butenol and DMAPP into 3-methyl-2-butenol.
571 Reaction 3 catalyzes the reduction of 3-methyl-2-butenol to 3-methyl-butanol. The
572 synthetic pathway competes with native pathways in *E. coli* that also use IPP and DMAPP as
573 substrates, such as those for quinone and cell membrane synthesis (shown in red).
574 **Fig. 2.** Enzymes catalyzing Reaction 2 are identified from two libraries. (**A**) The family of 23
575 haloacid dehalogenase (HAD)-like phosphatases from *E. coli* (black bars) was screened for
576 an enzyme able to catalyze the dephosphorylation of IPP to 3-methyl-3-butenol.
577 Overexpression of HAD4 and HAD10 led to a more than two-fold increase in 3-methyl-3-
578 butenol production compared to the control (white bar). (**B**) The family of 13 Nudix
579 hydrolases from *E. coli* (black bars) was screened for an enzyme able to catalyze the
580 dephosphorylation of IPP to 3-methyl-3-butenol. Overexpression of NudB, NudF, NudI,
581 NudJ, and NudM led to a more than two-fold increase in 3-methyl-3-butenol production
582 compared to the control (white bar).
583 **Fig. 3.** 3-Methyl-3-butenol, 3-methyl-2-butenol, and 3-methyl-butanol are produced from
584 IPP. (**A**) No C5 alcohol production was observed when Idi was overexpressed by-itself
585 compared to the controls (99A, NudB). Co-expression of NudB and Idi produced five fold
586 less 3-methyl-3-butenol (gray bars) than the positive control. (**B**) Overexpression of the

587 fusion proteins Idi-NudB and Idi1-NudB increased C5 alcohol production by more than two
588 fold compared to co-expression of NudB and Idi, and 3-methyl-2-butenol production (white
589 bars) was observed. Production of 3-methyl-butanol (black bars) was observed by co-
590 expressing NemA with either Idi-NudB or Idi1-NudB. No production of either 3-methyl-2-
591 butenol or 3-methyl-butanol was observed when only NemA was overexpressed.
592

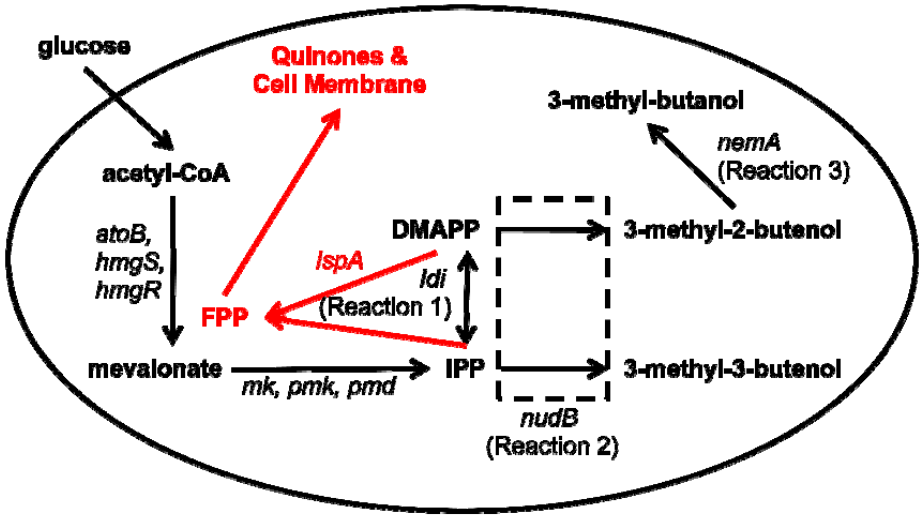
593

594

595 **FIGURES**

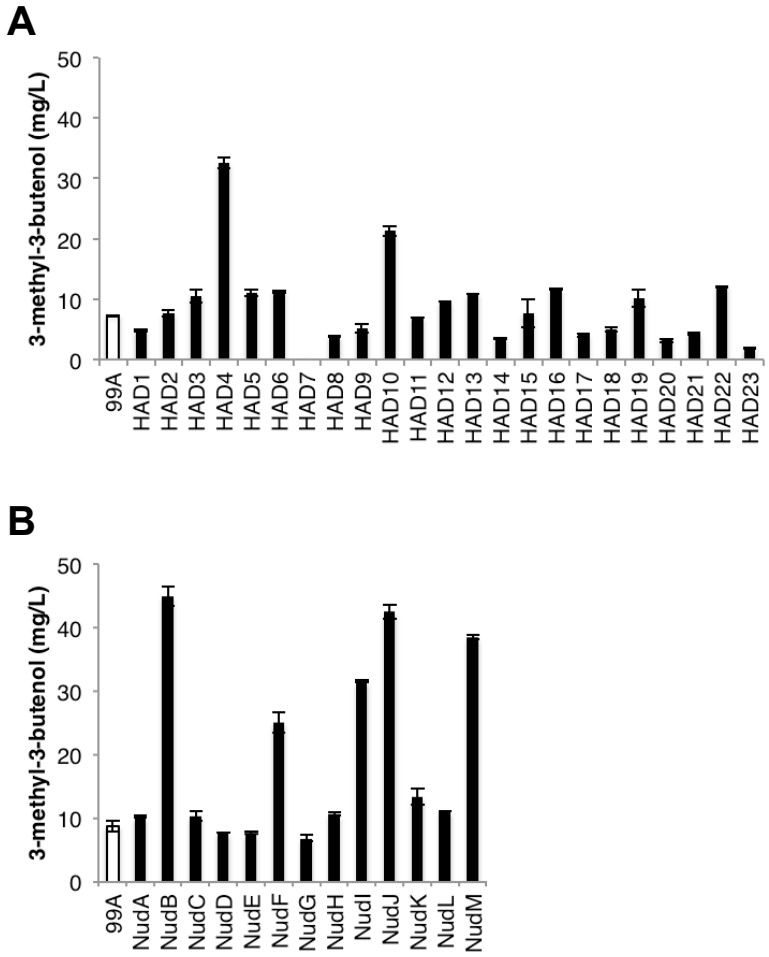
596 **Fig. 1.**

597



598 1.5 glucose → 3 Acetyl-CoA → 1 IPP → 1 C5 Alcohol + 4 CO_2

599 **Fig. 2.**



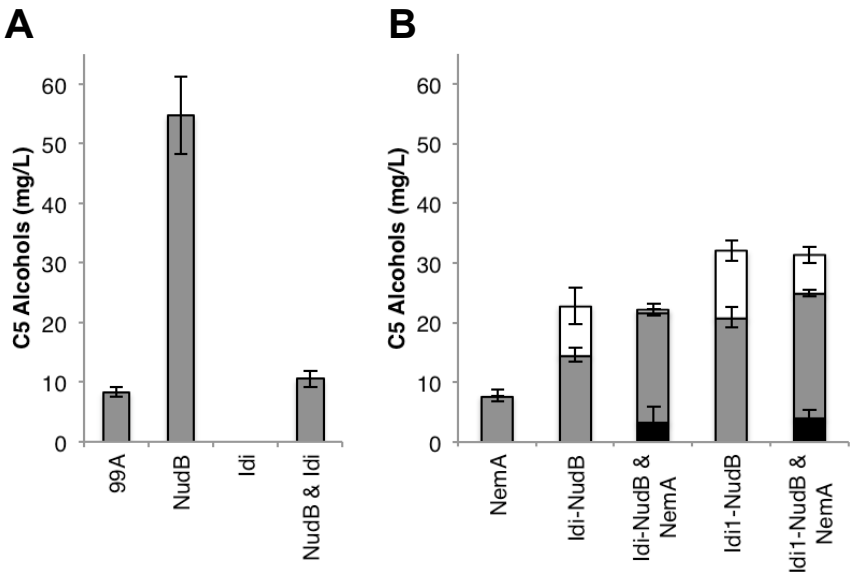
600

601

602

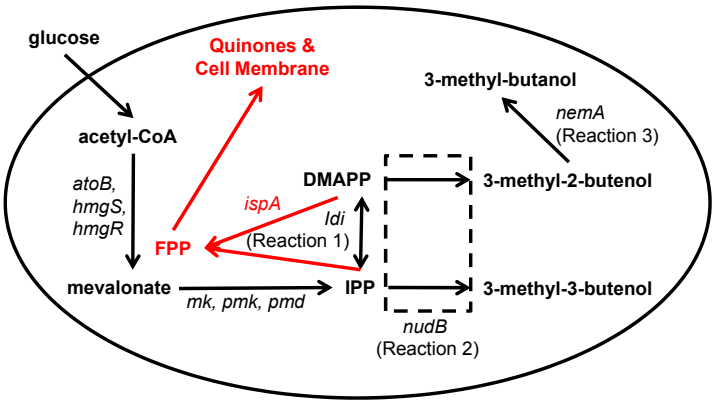
603

604 **Fig. 3.**



605

606



1.5 glucose → 3 Acetyl-CoA → 1 IPP → 1 C5 Alcohol + 4 CO_2